

# UNSERE SCHULE ...

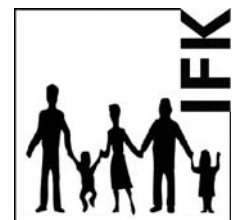
## SOZIALE SCHULQUALITÄT AUS SCHÜLERSICHT



Gefördert aus Mitteln des Bundesministeriums für Bildung und Forschung



Institut für angewandte  
Familien-, Kindheits- und Jugendforschung  
an der Universität Potsdam  
Institute for Applied Research on Childhood,  
Youth, and the Family

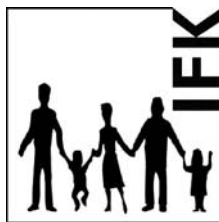




Detlef Landua, Christine Gerbich, Dorett Jenkel, Dietmar Sturzbecher

# Entwicklung und Erprobung eines Fort- und Weiterbildungsprogramms gegen Gewalt, Ausländerfeindlichkeit, politischen Extremismus und Antisemitismus

## Ergebnisbericht



**Institut für angewandte  
Familien-, Kindheits- und Jugendforschung  
an der Universität Potsdam  
Institute for Applied Research on Childhood,  
Youth, and the Family**

## **IMPRESSUM**

**Titel:** Ergebnisbericht des Projekts „Entwicklung und Erprobung eines Fort- und Weiterbildungsprogramms gegen Gewalt, Ausländerfeindlichkeit, politischen Extremismus und Antisemitismus“

**Autoren:** Detlef Landua, Christine Gerbich, Dorett Jenkel und Dietmar Sturzbecher

**Anschrift:** Institut für angewandte Familien-, Kindheits- und Jugendforschung e.V. (IFK)  
an der Universität Potsdam

IFK Vehlefanzenz

Burgwall 15

16727 Oberkrämer

Tel.: (03304) 39 70 10

Fax: (03304) 39 70 16

[www.ifk-vehlefanzenz.de](http://www.ifk-vehlefanzenz.de)

Das diesem Band zugrunde liegende Modellprojekt „Entwicklung und Erprobung eines Fort- und Weiterbildungsprogramms gegen Gewalt, Ausländerfeindlichkeit, politischen Extremismus und Antisemitismus“ wurde mit Mitteln des Bundesministerium für Bildung und Forschung gefördert (Förderkennzeichen SI 074). Die Realisierung des Projekts erfolgte in Zusammenarbeit mit dem Institut für berufliche Bildung und Weiterbildung (ibbw) in Göttingen.

Die Verantwortung für den Inhalt dieser Veröffentlichung liegt bei den Autoren.

Zitiervorschlag: Landua, D., Gerbich, C., Jenkel, D. & Sturzbecher, D. (2005).

Ergebnisbericht des Projekts „Entwicklung und Erprobung eines Fort- und Weiterbildungsprogramms gegen Gewalt, Ausländerfeindlichkeit, politischen Extremismus und Antisemitismus“. Vehlefanzenz: IFK an der Universität Potsdam.

## Inhaltsverzeichnis

1	Soziale Schulqualität aus Kundensicht – Schülerbefragungen als Mittel der Schulevaluation .....	7
1.1	„Soziale Schulqualität“ – eine empirische Annäherung an ein Schlagwort.....	7
1.2	Soziale Schulqualität als Erziehungs- und Bildungsfaktor .....	8
1.3	Soziale Schulqualität – eine multiperspektivische Betrachtung .....	11
1.4	Zusammenfassung.....	16
1.5	Literatur.....	17
2	Qualität von Schulen – ein Forschungsüberblick.....	22
2.1	Die Messung von Schulqualität.....	22
2.2	Überblick über Verfahren und Studien zur Bestimmung von Schulqualität.....	24
2.3	Unterrichtsforschung.....	24
2.4	Schulklimaforschung.....	28
2.5	Schuleffektivitätsforschung.....	35
2.6	Integrative Ansätze.....	42
2.7	Zusammenfassung.....	55
2.8	Literatur.....	57
3	Evaluation sozialer Schulqualität – eine Serviceleistung für Schulen .....	63
3.1	Methodische Anmerkungen zur Evaluation von Schulen .....	63
3.2	Ziele und Organisation des Projekts „Unsere Schule...“ .....	68
3.3	Zur Projektentwicklung in den einzelnen Bundesländern.....	71
3.4	Die Entwicklung des Erhebungsinstruments.....	77
3.5	Die Datenerfassung und -auswertung .....	88
3.6	Die Datenpräsentation .....	90
3.7	Ergänzende Maßnahmen und Leistungsangebote des Projekts.....	97
3.8	Zusammenfassung und Ausblick .....	100
3.9	Literatur.....	104
4	Erhebung sozialer Schulqualität – computerbasierte Befragungstechniken .....	105
4.1	Methoden zur Schulevaluation.....	105
4.2	Technische Möglichkeiten computergestützter Datenerhebungen .....	108
4.3	Methodische Aspekte computergestützter Befragungen.....	123
4.4	Zum Ablauf von Schulevaluationen mit Hilfe des Online-Verfahrens.....	127
4.5	Praxisbericht.....	133
4.6	Zusammenfassung und Ausblick .....	137
4.7	Literatur.....	139
5	Anhang .....	142



# 1 Soziale Schulqualität aus Kundensicht – Schülerbefragungen als Mittel der Schulevaluation

D. Sturzbecher, C. Gerbich

## 1.1 „Soziale Schulqualität“ – eine empirische Annäherung an ein Schlagwort

Mit der regelmäßigen Veröffentlichung immer neuer Schulleistungsstudien hat die Diskussion über die Qualität des deutschen Bildungswesens auch die breite Öffentlichkeit erreicht. Zusätzlich angetrieben wird diese Diskussion durch die anhaltenden Klagen von Vertretern der deutschen Wirtschaft über den Mangel an qualifizierten Auszubildenden und Arbeitskräften. Aber nicht nur das relativ schlechte Abschneiden der deutschen Schülerinnen und Schüler im internationalen Leistungsvergleich oder die bei einigen Schulabsolventen unzureichenden Voraussetzungen für eine Berufsausbildung geben Anlass zur Sorge. Immer wieder finden sich auch spektakuläre Berichte über Gewalt und Kriminalität an Schulen. Diese Berichte werden nicht selten als Indiz für eine Vernachlässigung der Erziehungsfunktion der Schule, mindestens jedoch als Beleg für Defizite im sozialen Zusammenleben an Schulen angesehen. In diesem Zusammenhang wird dann regelmäßig die „Soziale Schulqualität“ angesprochen. Was bedeutet dieser Begriff, der für die Bestimmung von Qualitätskriterien für die Bewertung von Schulen genutzt wird?

Betrachten wir zunächst den allgemeinen Begriff der „Schulqualität“, für den drei Qualitätsbereiche bedeutsam sind: die Strukturqualität, die Prozessqualität und die Ergebnisqualität (Donabedian, 1966). Unter die Dimension der Strukturqualität fallen alle rechtlichen, organisatorischen, personellen und finanziellen Rahmenbedingungen von schulischer Bildung. Hierzu gehören beispielsweise organisatorische und konzeptionelle Arbeitsgrundlagen wie die Lehrpläne ebenso wie die personelle und materielle Ausstattung der Schule oder die Aus- und Fortbildung des Personals. Zur Prozessqualität zählt man interaktionale Merkmale des Schulsystems. Diese umfassen sowohl die Merkmale der innersystemischen Interaktion, also beispielsweise die Qualität des sozialen Umgangs zwischen Lehrenden und Lernenden, als auch die Merkmale der systemübergreifenden Interaktion der Schule mit anderen Institutionen und der weiteren sozialen Umgebung, denken wir an die Zusammenarbeit mit den Eltern. Diese Qualitätsparameter stehen mit der didaktischen und sozialen Kompetenz der Lehrer im Zusammenhang und werden nicht zuletzt in der Art und Weise sichtbar, wie die Lehrer Schulmotivation und Lernspaß fördern, Kritik üben oder das Schülerverhalten sanktionieren. Die Ergebnisqualität schließlich resultiert aus der Überprüfung der Wirksamkeit von schulischen Dienstleistungen und zeigt sich einerseits in den Schulleistungen der Schüler und andererseits in ihren Wertvorstellungen und moralischen Orientierungen. Der letztgenannte, mit den Erziehungsaufgaben der Schule zusammenhängende Aspekt des schulischen „Outputs“ wird allerdings in der eingangs beschriebenen öffentlichen Diskussion häufig ausgeblendet, solange nicht spektakuläre Vorfälle wie Gewalt- oder Kriminalitätsdelikte zur Schulkritik ermuntern.

Mit der Unterscheidung der drei genannten Qualitätsdimensionen wird deutlich, dass der Begriff der „sozialen Schulqualität“ der Prozessqualität der schulischen Bildung zuzuordnen ist; es geht damit also um die sozialen Bedingungen, unter denen schulisches Lernen stattfindet. Die Beschäftigung mit diesen sozialen Bedingungen des schulischen Zusammenlebens ist nicht neu. Die Unterrichts- bzw. Schulklimaforschung beschreibt und analysiert seit langem die Interaktion schulischer Akteure und die sozialen Voraussetzungen,

die für ein günstiges Lernumfeld gegeben sein müssen (vgl. dazu im Überblick Lüders & Rauin, 2004). Tillmann und Meier (2001, S. 490) definieren Schulklima als „ein zentrales Qualitätsmerkmal des pädagogischen Prozesses, das sowohl das didaktisch-methodische Arrangement des Fachunterrichts durchdringt als auch darüber hinausgeht.“ Diesen Gedanken griffen Leiske, Sturzbecher und Keil (2001) auf, als sie versuchten, im Rahmen der Zeitreihenstudie „Jugend in Brandenburg“ den Begriff „soziale Schulqualität“ empirisch zu rekonstruieren. Dazu werteten sie die Einschätzungen von 3438 Jugendlichen aus 40 brandenburgischen Schulen und Oberstufenzentren zu sozialen Umfeldbedingungen an ihrer Schule aus und strukturierten diese Einschätzungen mit faktoren- und pfadanalytischen Methoden. Im Ergebnis fanden die genannten Autoren ein Modell der sozialen Schulqualität mit fünf Bestimmungsstücken; dieses Modell haben Sturzbecher und Hess (2002) auf der Grundlage von Längsschnittdaten hinsichtlich seiner Binnenstruktur, seiner Stabilität und seiner Zusammenhänge zu den schulischen Einstellungen und Verhaltensweisen von Schülern vertiefend analysiert und dargestellt.

Betrachten wir diese Bestimmungsstücke, die zugleich Indikatoren für die soziale Schulqualität darstellen, etwas genauer. Es handelt sich bei diesen Indikatoren um die „Fachliche Lehrqualität“, die „Soziale Lehrqualität“, die „Schulattraktivität“, die „Klassenkohäsion“ und die „Unterstützung durch die Lehrerschaft“. Vor allem die Komponente „Soziale Lehrqualität“, also eine an den individuellen Voraussetzungen der Schülerinnen und Schüler orientierte differenzierte Unterrichtsgestaltung, erscheint für eine hohe soziale Schulqualität als ausschlaggebend. Aber auch die Unterstützung durch Lehrer bei außerschulischen Problemen, wertschätzende soziale Beziehungen innerhalb der Klasse und zwischen den schulischen Interessengruppen sowie die Möglichkeiten von Schülern zur Mitsprache und Verantwortungsübernahme stehen in einem engen Zusammenhang zur sozialen Schulqualität. Damit ist ein weiteres Kennzeichen guter Schulen angesprochen, nämlich der Aspekt der „Schülerpartizipation“: Erst wenn den Schülerinnen und Schülern die Möglichkeit geboten wird, auf „gleicher Augenhöhe“ am schulischen Leben teilzunehmen, können sie lernen, für sich selbst und andere Verantwortung zu tragen. Dazu bzw. zu einer attraktiven Schule gehört ein vielfältiges und anregendes Schulleben, das den Schülern Handlungsspielräume und Wahlmöglichkeiten zur Verfügung stellt, in denen sie ihre Interessen formulieren und aushandeln können. Ähnliche Befunde fanden sich bei Posch und Altrichter (1999), die Fallstudien mit dem Ziel analysierten, Merkmale „guter“ Schulen zu beschreiben.

## 1.2 Soziale Schulqualität als Erziehungs- und Bildungsfaktor

Wir haben im vorangegangenen Kapitel die Einschätzungen von Schülerinnen und Schülern zum Zusammenleben an ihrer Schule strukturiert und Anhaltspunkte dafür gefunden, was ihnen in dieser Hinsicht wichtig erscheint. Aber warum sollte es sich lohnen, diese Merkmale bzw. Aspekte der sozialen Schulqualität aus Schülersicht an den einzelnen Schulen aufmerksam zu beobachten oder gar Anstrengungen zu ihrer Verbesserung zu unternehmen?

Die Aufgaben der Schule dürfen sich nicht darauf beschränken, Fachwissen zu vermitteln und die kognitiven Fähigkeiten der Schülerinnen und Schüler zu fördern. Darüber hinaus hat die Schule auch die Funktion, die psycho-soziale Entwicklung der Heranwachsenden und ihre politische Sozialisation zu unterstützen: In der Schule sollen die Kinder und Jugendlichen sich grundlegende Kulturfertigkeiten aneignen und sich zu verantwortlichen, urteilsfähigen und mündigen Bürgerinnen und Bürgern entwickeln. Es geht also bei der Forderung nach einer hohen sozialen Schulqualität nicht nur darum, ein günstiges soziales Umfeld für die Aneignung von Fachwissen zur Verfügung zu stellen, sondern es handelt sich nicht zuletzt



auch um die Vermittlung von Werten und Normen für das Zusammenleben in einer demokratischen Gesellschaft. Dazu gehört die Erziehung zur Selbständigkeit und Selbstbestimmtheit, ohne die demokratisches Handeln gar nicht erst möglich ist. Diese Besinnung auf die sozialisatorischen Aufgaben der Schule im weiteren Sinne findet sich neben dem Wunsch, förderliche Bedingungen für die Schulleistungen der Schüler zu schaffen, sowohl in den Ansätzen der Schulklimaforschung (Fend, 1998; Tillmann, 1994) als auch im noch stärkeren Maße in dem Ansatz der sozialen Schulqualität, der im vorliegenden Beitrag dargestellt wird.

Bleiben wir noch kurz beim Zusammenhang zwischen dem Schul- und Klassenklima bzw. der sozialen Schulqualität einerseits und den schulischen Leistungen andererseits. Hier finden sich durchaus widersprüchliche Forschungsergebnisse (Tillmann & Meier, 2001, S. 493) und zuweilen auch der Schluss, dass die Lernumgebung keinen signifikanten Einfluss auf die schulischen Leistungen ausüben würde. Derartige Befunde scheinen jedoch auf einer undifferenzierten Erfassung von Parametern der Lernumgebung zu beruhen und stehen im Kontrast zu einer Reihe von Forschungsbelegen dafür, dass Lernumgebungen über motivationale Faktoren (z.B. Lerneinstellungen) die schulischen Leistungen von Schülern beeinflussen (Helmke & Schrader, 1993). Im Rahmen der oben bereits genannten Zeitreihenstudie „Jugend in Brandenburg“ fanden Leiske, Sturzbecher und Keil (2001) im Rahmen einer Pfadanalyse sogar einen moderaten direkten Zusammenhang zwischen den Einschätzungen der Schüler zum bereits beschriebenen Konstrukt „soziale Schulqualität“ einerseits und ihren Schulleistungen andererseits; darüber hinaus beeinflussten die Einschätzungen zur sozialen Schulqualität die Lernmotivation, das Selbstwerterleben und vor allem den Spaß der Schüler an der Schule. Interessanterweise fanden sich diese Zusammenhänge nicht nur in den Individualdaten der Schüler, sondern auch, wenn die Daten schulbezogen aggregiert wurden. Dies deutet darauf hin, dass die soziale Schulqualität von den Schülern einer Schule ähnlich wahrgenommen wird und auch schultypische Effekte im Hinblick auf die Lerneinstellungen der Schüler erzeugt.

Die dargestellten Ergebnisse im Hinblick auf die Zusammenhänge zwischen der sozialen Schulqualität und der Lernmotivation der Schüler konnten Sturzbecher und Hess (2002) auf der Grundlage von Längsschnittdaten (N = 762) aus zwei Messzeitpunkten bestätigen. Sie untersuchten zunächst das Binnenverhältnis zwischen den oben genannten fünf einzelnen Bestimmungsstücken von sozialer Schulqualität aus Schülersicht (Schulattraktivität, fachliche und soziale Lehrqualität, Klassenkohäsion, Unterstützung durch die Lehrerschaft) einerseits und dem Gesamtkonstrukt andererseits, und zwar separat für die beiden Erhebungszeitpunkte 1999 und 2001. Dabei zeigte sich ein bemerkenswertes Ergebnis: Obwohl die Schulqualitätskomponenten an zwei Messpunkten unabhängig voneinander erfasst wurden, stimmte ihre Rangfolge hinsichtlich der Einflussstärke auf „soziale Schulqualität“ fast überein und ihre Pfadkoeffizienten ähnelten sich stark. Die Modellierung von „sozialer Schulqualität“ mit den dargestellten Qualitätsfaktoren erwies sich also als methodisch stabil.

Welcher Zusammenhang zeigte sich nun konkret zwischen der sozialen Schulqualität und der Lernmotivation der Schüler? Zunächst war im Ergebnis einer Pfadanalyse festzustellen, dass sowohl die Einschätzungen der Schüler zur eigenen Lernmotivation als auch ihre Einschätzungen zur sozialen Schulqualität über einen Zeitraum von zwei Jahren hinweg eine relativ hohe Stabilität aufwiesen: Vorhandene Abneigungen gegen das schulische Lernen abzubauen, ist also wie auch die Verbesserung der Schulqualität aufwändig und, wie Lehrer und Schulleitungen wissen, harte Arbeit. Diese Arbeit wird aber vermutlich bei vielen Schülern nicht vergeblich bleiben, denn sowohl 1999 als auch 2001 korrespondierte eine hohe soziale Schulqualität mit einer hohen Lernmotivation. Ein direkter Langzeiteffekt der sozialen Schulqualität im Jahr 1999 auf die Lernmotivation im Jahr 2001 fand sich allerdings nicht. Ein solcher Langzeiteffekt im Sinne einer Verbesserung der Lernmotivation zeigte sich erst

im Rahmen einer Regressionsanalyse mit einer besonders wichtigen Komponente der sozialen Schulqualität, nämlich der „Sozialen Lehrqualität“: Diejenigen Schüler, die im Jahr 1999 trotz geringer Lernmotivation daran glaubten,

- dass ihr Lehrer versuche, ihre individuellen Stärken und Schwächen zu erkennen, und
- dass der Lehrer dieses Wissen bei der Planung und Gestaltung seiner Lehr- und Unterstützungsangebote berücksichtigen würde (methodisch kann man hier von „Binnendifferenzierung“ sprechen),

wiesen zwei Jahre später auch eine deutlich höhere Lernmotivation und bessere Leistungen auf (Sturzbecher & Hess, 2002). Dieser Zuwachs an Lernmotivation wurde durch elterliche Unterstützung und internale Kontrollüberzeugungen zusätzlich verstärkt; ein Effekt, der sich auch in anderen Studien findet (Ryan, Stiller & Lynch, 1994).

Die Ergebnisse von Sturzbecher und Hess (2002) zeigen also, dass man mit einer Verbesserung der sozialen Schulqualität und vor allem mit einer verbesserten Binnendifferenzierung die Lernmotivation der Schüler erhöhen und auch ihre Überzeugung fördern kann, des „eigenen Glückes Schmied“ zu sein, d.h. eigene Interessen im Leben verwirklichen zu können („internale Kontrollüberzeugung“). Dies entspricht Befunden anderer Studien, nach denen sich soziale schulische Rahmenbedingungen wie die soziale Einbindung der Lernenden in eine Bezugsgruppe, Mitbestimmungsrechte von Schülern oder soziale Anerkennung auf die Lernmotivation von Schülern auswirken (Prenzel & Schiefele, 1999) sowie zu höheren Leistungsstandards und in der Folge auch zu höheren individuellen Fachleistungen von Schülern führen (Lehmann et al., 2000; Rösner, 2000).

Unsere bisherigen Darlegungen waren darauf gerichtet, die Bedeutung der sozialen Schulqualität mit der Sozialisationsfunktion der Schule einerseits und mit der Unterstützung der Aneignung von Fachwissen durch ein förderliches soziales Umfeld andererseits zu begründen. Den erstgenannten Aspekt wollen wir noch einmal aufgreifen; zu diesem Aspekt zählt nämlich auch die Gewalt- und Kriminalitätsprävention. Bereits Rutter, Maughan, Mortimore und Ouston (1979) konnten anhand einer Längsschnittanalyse nachweisen, dass soziale Schulqualität nicht nur als Garant für das „Wohlfühlen“ der Schüler in einer Schule und für gute Schulleistungen anzusehen ist, sondern gleichzeitig einen entscheidenden Schutzmechanismus gegen delinquentes Verhalten von Jugendlichen darstellt.

Auch die neuere Schulforschung misst der sozialen Schulqualität eine bedeutende Rolle bei der Erklärung und Prävention von Jugendgewalt und Jugendkriminalität zu. So untersuchten Schubarth und Melzer (1995), Holtappels et al. (1997), Melzer (1998) die Bedeutsamkeit von schulischen Faktoren (Schulklima, Lernkultur, Lernatmosphäre) für die Herausbildung abweichenden und straffälligen Verhaltens bei Schülern. Im Ergebnis dieser Studien tragen Sinndefizite und Distanzgefühle im Hinblick auf das schulische Lernen bei Jugendlichen zur Entstehung von gewalttätigen und kriminellen Verhaltensmustern bei. Funk (1995) konnte in seiner Studie weiterhin den Einfluss von „Schulklimavariablen“ (z.B. Lehrer-Schüler-Verhältnis und Partizipationsmöglichkeiten für die Schülerschaft) auf die gewaltbezogenen Einstellungen und das Gewalthandeln belegen. Nicht zuletzt konnten wir durch unsere eigenen Zeitreihenstudie (Sturzbecher, Landua & Shahla, 2001) bzw. Längsschnittanalyse (Sturzbecher, Hess & Them, 2002) empirisch belegen, dass eine gute soziale Schulqualität im Sinne von Mitgestaltungsmöglichkeiten für Schüler, einem guten Klassenklima und einer an den Bedürfnissen der Lernenden ausgerichteten Unterrichtsgestaltung die Gewaltbereitschaft und die Neigung Jugendlicher zum politischen Extremismus mindern. Daher unterstreicht auch der Gesichtspunkt der Gewalt- und Kriminalitätsprävention die Notwendigkeit, sich ein zutreffendes Bild von der sozialen Schulqualität an einer Schule zu verschaffen und sich auf dieser Grundlage um die Optimierung der verschiedenen Qualitätskomponenten zu bemühen.

### 1.3 Soziale Schulqualität – eine multiperspektivische Betrachtung

Am schulischen Zusammenleben sind unterschiedliche soziale Interessengruppen beteiligt: die Schülerschaft, die Lehrerschaft, die Schulleitungen und im weiteren Sinne auch die Elternschaft. Jede dieser Gruppen könnte man nach der Qualität des schulischen Zusammenlebens fragen. Die zahlenmäßig größte Gruppe stellen die Schülerinnen und Schüler dar. Schon allein deshalb erscheint eine Befragung der Schülerschaft am aufwändigsten; eine Befragung der Schulleitung oder der Lehrerschaft wäre sicher mit geringerem zeitlichen und finanziellen Kosten zu leisten. Neben diesem Aufwand spricht gegen Schülerbefragungen, dass sie in der Praxis vor allem von Lehrerinnen und Lehrern nicht selten als nutzlos eingeschätzt werden. Am häufigsten hört man bei der Durchführung von Schülerbefragungen von Lehrern die folgenden Einwände:

- Schüler könnten aufgrund ihrer fehlenden pädagogischen Expertise die Schulqualität nicht fachgerecht beurteilen, ihre Einschätzungen seien folglich nicht zutreffend bzw. „falsch“;
- Schülerbefragungen würden nur „Momentaufnahmen“ des schulischen Alltags darstellen, da die Befragten ihre Erfahrungen (noch) nicht verallgemeinern sowie Wesentliches nicht vom Unwesentlichen trennen könnten, und
- Schüler könnten die Schulqualität nicht beurteilen, weil für sie vor allem Aspekte wie Spaß und der Kontakt zu Gleichaltrigen wichtig seien, während für die Lehrer- und Elternschaft bei der Beurteilung von Qualität eher leistungsbezogene Kriterien eine Rolle spielen würden.

Um diesen Einwänden zu begegnen und auch das ökonomische Argument zu entkräften, gilt es die Frage zu beantworten, warum die Schülersicht auf die soziale Schulqualität von der Lehrersicht abweicht und trotzdem nicht weniger „richtig“ oder wichtig ist als diese.

#### ***Warum unterscheiden sich die Schüler- und Lehrersichten auf die Schüler-Lehrer-Interaktion bzw. die soziale Schulqualität?***

Fragen wir Schüler oder Lehrer nach ihren Einschätzungen über die soziale Wirklichkeit an ihrer Schule, dann erfassen wir ihr soziales Wissen bzw. Inhalte ihres sozialen Gedächtnisses. Solche (sozio-kognitiven) Wissensstrukturen sind keine passiven Abbilder der Realität. Die Merkmale und Vorgänge der Lebensumwelt werden nämlich während der Informationsaufnahme, -speicherung und -reaktivierung aktiv bewertet und selektiert (Hoffmann, 1982; Engelkamp, 1990), also individuell „verzerrt“. Diese Abweichungen führen zu einer subjektiv einzigartigen Sicht auf die objektive Realität.

Die Tatsache, dass Wahrnehmungs- und Gedächtnisinhalte von den subjektiven Bedürfnissen und Motivationen des jeweiligen Individuums wie auch von seinem vorhandenen Wissen und kognitiven Fähigkeiten beeinflusst werden, gilt insbesondere für soziale Sachverhalte. Daher kann der Vergleich von Beurteilungen des gleichen sozialen Gegenstands von verschiedenen Interaktionspartnern unterschiedliche Ergebnisse erbringen, ohne dass einer der Befragten aus seiner subjektiven Sicht die Unwahrheit sagen würde. Dieser insbesondere in der forensischen Psychologie gut dokumentierte Grundsatz kann auch für Beurteilungen der Lehrer-Schüler-Interaktion aus der Sicht der beteiligten Interaktionspartner Gültigkeit beanspruchen: Lehrer und Schüler berichten bei diesbezüglichen Befragungen Interaktionsmerkmale, die jeweils subjektive Relevanz besitzen und daher für die Speicherung in den Wissensstrukturen selektiert wurden. Diese Merkmale müssen aber für Lehrer und Schüler nicht identisch sein.

Obwohl es also unvermeidlich Unterschiede zwischen den Sichten von Lehrern und Schülern auf das schulische Zusammenleben geben muss, wurden systematische Perspektivenvergleiche im Rahmen der Lehr- und Lernforschung erst in jüngerer Zeit

vorgenommen (vgl. Clausen, 2002). Bei der familienpsychologischen Forschung zeigt sich in dieser Hinsicht eine bessere Forschungslage. Bereits seit den 1970er Jahren wurde eine Reihe von Forschungsarbeiten durchgeführt, bei denen anhand von Befragungen elterliche und kindliche Sichten auf die Erzieher-Kind-Interaktion verglichen wurden (z.B. Cox, 1970; Lukesch & Tischler, 1975; Sturzbecher und Freytag, 1999). Die Ergebnisse der vorliegenden Forschungsarbeiten zeigen ein weitgehend übereinstimmendes Bild: Die Angaben der Kinder über das elterliche Erziehungsverhalten weisen bei den meisten Untersuchungen nur einen relativ schwachen bzw. statistisch nicht zu sichernden Zusammenhang zu denen der Eltern auf. Die gefundenen Zusammenhänge betreffen vorrangig negative Dimensionen elterlichen Erziehungsverhaltens wie besondere Strenge oder unangemessene Kontrolle.

Durch die gefundene große Diskrepanz zwischen den elterlichen und kindlichen Sichten auf die Familienerziehung drängte sich die Frage auf, ob neben den bereits geschilderten subjektiven Besonderheiten bei der Gewinnung und Verarbeitung sozialer Informationen noch weitere Einflussfaktoren die Beurteilungen von Eltern und Kindern verzerren. Diese Frage gewann noch an Bedeutung, als man eine hohe Übereinstimmung zwischen den Einschätzungen von Ehepartnern zu ihrem Erziehungsverhalten fand (Cox, 1970; Hoff, Minsel, Minsel & Grüneisen, 1973; Lukesch & Tischler, 1975; McKenry, Price-Bonham & O'Bryat, 1981; Sturzbecher und Freytag, 1999). Lukesch und Tischler schreiben zu möglichen Ursachen solcher „Erziehereffekte“: „Am plausibelsten erscheint aber die Annahme, daß die Fragebogenantworten der Eltern im Sinne der sozialen Erwünschtheit [...] verändert sind. Mit dieser Verfälschungstendenz ist gemeint, daß die Itembeantwortung von der Anpassung an geltende Gruppennormen, z.B. an Mehrheitsmeinungen über die 'richtige' Erziehung, determiniert wird" (1975, S. 95, vgl. auch Gehring et al., 1994, S. 126). Verzerrungen dieser Art sind bei Kindern bzw. Schülern kaum zu erwarten, weil sie allenfalls grobe Einsichten in die pädagogischen Leitvorstellungen ihrer Erzieher haben und ihr Antwortverhalten deshalb von diesen Maximen nicht betroffen ist. Allerdings sind analoge Anpassungsmechanismen der Selbst- oder Fremdwahrnehmung entsprechend der sozialen Erwünschtheit auch bei Schülern denkbar, wobei in diesem Fall die sozialen Normen alterstypischer Umwelten einen Konformitätsdruck erzeugen können.

Fassen wir zusammen: Es erscheint außerordentlich wahrscheinlich, dass sich nicht nur die Sichten von Eltern und Kindern auf die Erzieher-Kind-Interaktion systematisch unterscheiden, sondern auch die Sichten von Lehrern und Schülern. Zu diesen Perspektivendiskrepanzen tragen Besonderheiten der subjektiven Aufnahme und Verarbeitung sozialer Informationen genauso bei wie soziale und pädagogische Normen, welche das Antwortverhalten der Befragten beeinflussen. Entsprechend weist Clausen (2000, S. 81) mit Blick auf die Validität von Schülerbefragungen darauf hin, dass Wahrnehmungsprozesse im Allgemeinen von Subjektivität und Selektivität geprägt sind. Er argumentiert im oben dargestellten Sinne, dass eine geringe Übereinstimmung der Einschätzungen von Schülern, Lehrern oder externen Beobachtern gegebenenfalls nicht als Fehler zu verstehen sei, sondern die Angaben aus der Sicht der jeweiligen Befragten durchaus als valide gelten könnten („perspektivenspezifische Validität“); entsprechend gäbe es keinen Grund, Schülereinschätzungen einen geringeren Aussagewert zuzusprechen. Die Frage, ob denn nun die Lehrersicht oder die Schülersicht richtig sei, erscheint also falsch gestellt und kann nicht beantwortet werden. Vielmehr sind beide Sichten richtig (bzw. „gültig“ oder „valide“) in dem Sinne, als dass sie (in der Regel) dem subjektiven Erleben der Befragten entsprechen und daher wahrheitsgemäß sind.

***Können (jüngere) Schülerinnen und Schüler aufgrund ihres kognitiven Entwicklungsstandes die soziale Schulqualität (auch schon) zutreffend einschätzen?***

Will man diese Frage beantworten, gilt es zu diskutieren, ob Kinder über stabile mentale Vorstellungen zur Erzieher-Kind-Interaktion in ihrem sozialen Gedächtnis verfügen, ab wann

man diese Vorstellungen ggf. abfragen kann und ob diese Vorstellungen, verglichen mit denen Erwachsener, vielleicht in besonderer Weise für situative Verzerrungen anfällig sind.

Die Beziehung von Kindern zu erwachsenen Bezugs- bzw. Erziehungspersonen ist eine zentrale Voraussetzung für ihre emotionale, affektive und kognitive Entwicklung. Bereits Kleinkinder verfügen daher über mentale Modelle ihrer Bindung an Bezugspersonen. Entsprechend dieser Modelle werden beispielsweise soziale Kontakte gesucht oder abgelehnt (Brambring, Rauh & Beelmann, 1996). Eine Schlüsselrolle beim Aufbau dieser internen Repräsentationen sozialer Beziehungen spielt die Entäußerung positiver und negativer Emotionen, die von Geburt an zum kindlichen Kommunikationsrepertoire gehört. Die moderne Emotionspsychologie geht davon aus, dass Emotionen nicht primär als interne Abbildung organismischer Zustände ihre Bedeutung erlangen, sondern durch ihre regulative Funktion im sozialen Zusammenleben (Campos et al., 1994; Thompson, 1993 und 1994). Die kindliche Erfahrung, dass Sozialpartner beispielsweise auf Signale von negativen Emotionen infolge von Hunger oder Langeweile prompt und sensitiv reagieren, erzeugt eine sozio-emotionale Bindung an diese Sozialpartner (Bowlby, 1984; Ainsworth et al., 1978). Diese Bindung begründet bei Kleinkindern das Vertrauen in die Kompetenz der Sozialpartner, beim Abbau von emotionalen Spannungen zu helfen (Cassidy, 1994; Cicchetti, Ganiban & Barnett, 1991). Sind derartige Interaktionserfahrungen insbesondere (aber nicht ausschließlich) in emotional bedeutsamen Situationen konsistent und stabil, werden sie bereits in den ersten Lebensjahren zu einem situationsübergreifenden mentalen Modell des Verhaltens von Bezugspersonen generalisiert und verdichtet (Bowlby, 1984; Bretherton, 1990), in dem die kindlichen Erwartungen an Verhaltenswahrscheinlichkeiten abgebildet und emotional unterlegt werden ("Bindung"). Diese mentalen Modelle variieren interindividuell und besitzen eine hohe intraindividuelle Stabilität und Entwicklungswirksamkeit. Es kann daher keinen Zweifel daran geben, dass nicht nur Kleinkinder, sondern auch Schülerinnen und Schüler jeglichen Alters über stabile Vorstellungen über ihre schulischen Betreuungs-, Erziehungs- und Bildungsarrangements verfügen.

Während in der einschlägigen Forschung also weitgehend Einigkeit herrscht, dass selbst sehr junge Kinder über mentale Modelle ihrer Sozialbeziehungen zu Erziehungspersonen verfügen, war lange Zeit umstritten, ab wann Kinder über diese Modelle angemessen reflektieren und Auskunft geben können, mithin als sichere Informationsquelle für die Einschätzung von (schulischen) Erziehungsumwelten angesehen werden können. Einige Forscher setzen diesen Zeitpunkt in sehr jungem Lebensalter an. So konnte in der Bindungsforschung nachgewiesen werden, dass man über den „Fremde-Situation-Test“ bereits im zweiten Lebensjahr die Beziehungsqualität zu Bezugspersonen (vornehmlich die Eltern) aus kindlicher Perspektive zuverlässig erfassen kann und diese Einschätzungen einen hohen prädiktiven Wert für die sozio-kognitive Entwicklung haben (Grossmann et al., 1989). Für Vorschulkinder haben Sturzbecher und Freytag (2000) nachgewiesen, dass man mit dem „Familien- und Kindergarten-Interaktions-Test“ die Einschätzungen von Vier- bis Achtjährigen zur Erziehungs- und Bildungsqualität in der Familie und im Kindergarten erheben kann. Dabei schätzten die befragten Kinder das Verhalten ihrer Eltern und Erzieherinnen differenziert, konsistent und für Kinder dieses Alters auch zufriedenstellend stabil ein. Die Stabilität der Einschätzungen entsprach in der Größenordnung den Stabilitätsparametern von üblichen Erziehungsstilfragebogen für Jugendliche oder von Fragebogen zur Partnerschaftszufriedenheit bei Erwachsenen. Bereits Vier- bis Achtjährige wissen also gegebenenfalls, wenn sie von ihren Eltern und Erzieherinnen "schlecht behandelt" werden, und dieses Wissen steht in Zusammenhang mit ihrer kognitiven und sozialen Entwicklung bis hin ins Schulalter.

Welche gedächtnispsychologischen Mechanismen sorgen dafür, dass bereits jüngere Kinder und daher auch Schülerinnen und Schüler zuverlässig und gültig ihre Erziehungsumwelt

einschätzen können? Für die Erfassung von (kindlichen) Einschätzungen der sozialen Erfahrungswelt mittels Befragungen spielt der „Skript-Ansatz“ von Schank und Abelson (1977) eine bedeutende Rolle. Dieser begründet, auf welche Wissensstrukturen bei der Befragung von jüngeren Kindern Bezug genommen werden kann. Kennzeichnend für den Skript-Ansatz ist die Annahme einer speziell strukturierten Speicherung sozialen Wissens, die es einerseits erlaubt, komplexe Erfahrungen und Handlungsabläufe so zu verarbeiten, dass sie als Handlungsschemata immer wieder abgerufen werden können und somit Handlungsroutinen ermöglichen. Andererseits begründen die Besonderheiten dieser Speicherung und ihrer Entwicklung in der Ontogenese, weshalb auch jüngere Kinder über präzises und abrufbares soziales Wissen über ihre Erziehungsumwelt verfügen.

Nach dem Skriptansatz (Schank, 1975) simulieren Skripte als kognitive Schemata die Abspeicherung der Bedeutung von sozialen Ereignissen bzw. unseres Alltagswissens: Der Bedeutungsgehalt eines Ereignisses „wird durch ein Skript insofern repräsentiert, als die Rollen der beteiligten Personen, die Eingangs- und Endzustände, beteiligte Sachverhalte sowie die Sequenz der Aktionen spezifiziert sind“ (Kluwe & Spada, 1981, S. 310f). Neben diesen konkreten Komponenten enthalten Skripte als „Klassen stereotypisierter sozialer Ereignisse“ (Silbereisen, 1987) „slots“ für situative Besonderheiten.

Der Aufbau und die Erweiterung von Skriptstrukturen (Nelson, 1986) vollzieht sich bereits beim Kleinkind, ausgehend von konkreten Situationserfahrungen und zunächst vorrangig auf der Grundlage zwischenbegrifflicher Relationen, durch Generalisierungs- und Differenzierungsprozesse (Rumelhart & Ortony, 1977), wie sie auch in der Bindungs-, der Kognitions- und der sozialkognitiven Forschung beschrieben werden. Es besteht daher in der gedächtnispsychologischen Forschung kein Zweifel darüber,

- dass bereits Klein- und Vorschulkinder genauso wie Erwachsene ihr soziales Wissen in Skriptform speichern (Nelson, 1981),
- dass bereits Vorschulkinder sich bei Schilderungen sozialer Sachverhalte mehr auf ihr soziales Gedächtnis als auf die unmittelbare Situation beziehen (French, 1986; French begründet dieses unerwartete Forschungsergebnis damit, dass Skriptstrukturen sich aus vielfach wiederholten persönlichen Erfahrungen ableiten und deshalb hoch stabil seien) und
- dass Vorschulkinder mehr Informationen über das geben, was im Allgemeinen in einer sozialen Umwelt geschieht, als über das, was dort "gestern" geschehen ist (Hudson, 1986).

Unabhängig davon, dass die Skriptstrukturen mit dem Alter komplexer werden, stellt French (1985) resümierend fest, dass die Strukturen sozialen Wissens bei Kindern und Erwachsenen übereinstimmen und bereits Vorschulkinder beachtliche soziokognitive Reflexionsleistungen erreichen. Übertragen auf die Beantwortung unserer eingangs aufgeworfenen Frage bedeutet dies, dass eine Befragung von (jüngeren) Schülern nach der sozialen Schulqualität kein grundsätzliches oder unlösbares Problem darstellt, sondern lediglich ein Nachdenken über angemessene Methoden erfordert. Falls es sich bei den Befragten um jüngere Kinder handelt, erscheinen spielbasierte Befragungsstrategien als angemessen (Sturzbecher, 2001).

### ***Ist die Lehrersicht oder die Schülersicht auf die soziale Schulqualität wertvoller?***

Die Antwort auf diese Frage ist nicht zuletzt davon abhängig, mit welchem Ziel man Einschätzungen zur sozialen Schulqualität erhebt: Will man das Belastungserleben der Lehrerschaft analysieren, muss man Lehrer befragen; will man dagegen die Entwicklung oder die Leistungen der Schüler prognostizieren, erscheint eher eine Schülerbefragung angezeigt. Für die Begründung dieser Position ist wieder ein Rückgriff auf familienpsychologische Forschungsergebnisse interessant.

Zu den wenigen vergleichenden Studien zum Erklärungswert der Interaktionseinschätzungen von Erziehern (Eltern) und Kindern für die kindliche Entwicklung gehört die Studie von Garbe und Strasser (1978) unter 9- bis 10jährigen Grundschulkindern (zit. nach Helmke & Kischkel, 1980). Die Autoren analysierten Zusammenhänge zwischen kind- und mütterperzipiertem mütterlichem Sanktionsverhalten und kindlichen leistungsbezogenen Persönlichkeitsmerkmalen (z.B. „Hoffnung auf Erfolg“, „Lernmoral“, „Begabungseinschätzungen“). Sie fanden, ähnlich wie Davids und Hainsworth (1967) oder Cox (1970), dass die Kindperzeption insbesondere negativ-sanktionierenden mütterlichen Erziehungsverhaltens einen Erklärungswert für die Ausprägung der genannten Merkmale besitzt; der Erklärungswert der Kindperzeptionen übertraf im Bereich negativen Sanktionsverhaltens immer den Prädiktionswert der Mütterperzeptionen. Diese Befunde stützen die bereits dargestellte These, dass die Erzieherbeurteilungen der Erzieher-Kind-Interaktion tatsächlich durch eine größere Vulnerabilität gegenüber Verfälschungstendenzen gekennzeichnet sind und sich dies in einem gegenüber kindlichen Einschätzungen geringeren Prognosewert für die kindliche Entwicklung niederschlägt. Sturzbecher und Freytag (1999) konnten diesen Effekt auch für Vorschulkinder nachweisen: Sie fanden, dass die Perzeptionen von Vorschulkindern zum elterlichen Erziehungsverhalten eine höhere prognostische Validität im Hinblick auf die kindliche Entwicklung aufwiesen als die Elternperzeptionen.

Zusammenfassend lässt sich also festhalten, dass die Qualität der Erzieher-Kind-Interaktion ihren Einfluss auf die kindliche Entwicklung nicht allein durch ihre objektive Existenz an sich gewinnt, sondern auch durch das subjektive Erleben dieser Qualität durch die Interagierenden. So dürfte beispielsweise das gleiche Niveau an Hilfe durch einen Lehrer in Problemsituationen in Abhängigkeit von der Kompetenzerwartung und Problemlösekompetenz des Schülers von diesem auch sehr unterschiedlich bewertet werden. Individuelle psychische Rahmenbedingungen bestimmen also, vermittelt über das Erleben des Lehrerverhaltens und des eigenen Verhaltens durch den Schüler, die Effizienz interaktionaler Entwicklungsbedingungen auf die Herausbildung kindlicher Persönlichkeitsstrukturen mit. Das Erleben der Lehrer-Schüler-Interaktion durch den Schüler ist deshalb ein unverzichtbares Bestimmungsstück zur Bewertung der Entwicklungs- bzw. Erziehungswirksamkeit der sozialen Schulqualität. Will man also etwas über die Wirksamkeit schulischer Lern- und Lebensbedingungen erfahren, führt kein Weg an einer Schülerbefragung vorbei.

Damit ist nicht gesagt, dass Lehrerbefragungen zur sozialen Schulqualität nutzlos wären. Im Gegenteil; es gibt eine Reihe von Gründen, diesbezügliche Schülerbefragungen durch Lehrerbefragungen zu ergänzen. Ein erster Grund ist, dass die Schülerschaft und die Lehrerschaft genauso wie die Schulleitung und die Eltern teilweise andere Schwerpunkte bei den Kriterien setzen, die von der jeweiligen Gruppe als bedeutsam für die Qualität einer Schule erachtet werden (Scottish Office, 1992). Eine multiperspektivische Befragung zur Schulqualität gleicht also einem Puzzle, das neben vergleichbaren Facetten auch Elemente enthält, die für einzelne schulische Interessengruppen als spezifisch anzusehen sind. Kohlmeyer (2002) weist darauf hin, dass zu den für alle Interessengruppen vergleichbar relevanten Merkmalen beispielsweise die finanzielle, räumliche und personelle Ausstattung der Schule genauso gehört wie die fachlichen und pädagogisch-didaktischen Fähigkeiten des Lehrpersonals. Gleiches gilt für das Streben nach guten Leistungen und die Praxisnähe der Ausbildung. Insgesamt wird also deutlich, dass die Schülerschaft keineswegs „abgehobene“ Erwartungen an qualitativ gute Schulen stellt, sondern allgemein anerkannte und sinnvolle Forderungen äußert: Eine gute Schule und der daraus meist resultierende individuelle Lernerfolg liegen nämlich, wie auch die meisten Schülerinnen und Schüler wissen, durchaus in ihrem Eigeninteresse, da durch eine gute Ausbildung ihre Lebenschancen steigen (Ditton, 2002; Sturzbecher & Hess, 2002).

Hinsichtlich der zwischen den schulischen Interessengruppen vergleichbaren Merkmale der sozialen Schulqualität treten natürlich Perspektivendiskrepanzen bzw. Beurteilungsunterschiede auf. Gerade die Aufdeckung solcher Beurteilungsdiskrepanzen zwischen Schüler- und Lehrerschaft bei der sozialen Schulqualität könnte aber die Suche nach Ursachen für eine schlechte Schulqualität bereichern und zur Verbesserung der schulischen Leistungsbilanz beitragen. Um noch einmal ein Beispiel aus der familienpsychologischen Forschung zu bemühen: Es konnte empirisch belegt werden, dass unbewußte pädagogische Selbstbild-Fremdbild-Diskrepanzen die Erziehungswirksamkeit der Mutter-Kind-Interaktion beeinträchtigen und positiv mit familialer Konfliktneigung korrelieren (Schneewind, 1991); eine Aufdeckung und Bearbeitung solcher Diskrepanzen durch Erziehungs- und Familienberatung führt zu höherer (familien-) pädagogischer Effizienz und einem geringerem (familialen) Konfliktniveau. Genau diese Effekte lassen sich auch von einer multiperspektivischen Befragung zur sozialen Schulqualität an einer Schule und einer partnerschaftlichen Diskussion der Befragungsergebnisse erwarten, wenn alle schulischen Interessengruppen in diese Diskussion einbezogen werden.

#### 1.4 Zusammenfassung

Schülermeinungen zur sozialen Schulqualität spiegeln die subjektive Wahrnehmung des schulischen Alltags und nicht zuletzt des Lehrerverhaltens durch die Schülerinnen und Schüler wider. Damit sind diese Meinungen eine soziale Realität, ein „Teil des pädagogischen Feldes“ (Steltmann, 1992, S. 571), das auf das Wohlbefinden und die Erziehungs- und Bildungsergebnisse der Schule einwirkt. Selbst wenn externe Beobachter bzw. pädagogische Experten dem Unterricht und dem sozialen Zusammenleben an einer Schule eine hohe Qualität bescheinigen würden, die Lehrerinnen und Lehrer von ihrer Kompetenz überzeugt wären und die Schulleitung das große Angebot an der Schule loben würde: Erst die Schülermeinungen könnten aufzeigen, ob die als gut empfundenen und sicherlich auch gut gemeinten Konzepte, Methoden und Aktivitäten bei den Schülerinnen und Schülern auch so ankommen und wirken, wie es beabsichtigt war. Dabei steht fest, dass Schülereinschätzungen zur sozialen Schulqualität bei sorgfältiger Einhaltung methodischer Befragungsstandards genauso zuverlässig und gültig sind wie Befragungen der Schulleiter oder Lehrer. Es gibt also weder einen Grund, an der kognitiven Kompetenz von Schülern zur Einschätzung der Schulqualität zu zweifeln, noch ist aufgrund von sozialisationsgeprägten Erwartungsmustern davon auszugehen, dass Schülereinschätzungen durch individuelle Präferenzen (z.B. den Wunsch nach Spaß und Unterhaltung) einseitig beeinflusst werden (Gerstenmaier, 1975). Vielmehr erlaubt eine Schülerbefragung durch die mögliche Vielzahl von Einschätzungen eine Bündelung und Wichtung von Einzelmeinungen, sodass die Mittelwerte der Einschätzungen einen Schluss auf die intersubjektiv geteilte Wahrnehmung der Schulqualität erlauben (Ditton, 2002).

Aus den genannten Gründen wäre ein Verzicht auf die Erfassung von Schülereinschätzungen zur sozialen Schulqualität ein schwerwiegender Fehler bei der Steuerung der Schulentwicklung: Keine andere Personengruppe verfügt über ein derartig detailliertes Wissen über die Strukturen und sozialen Interaktionen an der einzelnen Schule wie die eigentlichen „Kunden“ der Schule, die Schülerinnen und Schüler (Haider 2002). Gerade in Bezug auf die sozialen Aspekte des Schulalltags wäre beispielsweise bei einer Befragung des Lehrerkollegiums zum Klassenzusammenhalt der Schüler die Distanz zwischen Lehrerschaft und Schülerschaft wohl zu groß, um einen Einblick in relevante Vorfälle und Gegebenheiten zu ermöglichen. Weiterhin können Schüler Erfahrungen zur sozialen Schulqualität in ihre Einschätzungen einbringen, die Lehrern verschlossen sind; denken wir nur daran, dass



Schüler ihre Lehrer miteinander vergleichen können und sie diese in einer Vielzahl unterschiedlicher Situationen erleben (Ditton, 2002).

Das Auftreten von Beurteilungsdifferenzen zwischen den schulischen Interessengruppen hinsichtlich vergleichbarer Kriterien für die soziale Schulqualität erscheint normal; die entscheidende Frage besteht darin, wie solche Beurteilungsdifferenzen pädagogisch verarbeitet werden. Diskreditiert man die Schülersicht auf Schulqualität als unqualifiziert und ignoriert man die damit verbundenen Qualitätsbewertungen, verschenkt man die Chancen für einen Dialog über Schul- und Bildungsqualität, der die Schulentwicklung vorantreiben könnte. Man verspielt damit auch Gelegenheiten für die Demokratieerziehung, denn Kinder und Jugendliche erhalten durch die Möglichkeit, ihre Meinung zu äußern, auch einen Einblick in den Ablauf demokratischer Prozesse. Gerade bei der Untersuchung sozialer Schulqualität sollte dieser Aspekt besonders berücksichtigt werden, denn Partizipation stellt ein Medium und zugleich ein Ziel der Entwicklung im Kindes- und Jugendalter dar (Grundmann, Welskopf & Sturzbecher, 2001).

Vorsorglich sei bemerkt, dass die gegenwärtig wachsenden Anstrengungen, die Schulqualität durch Hospitationen bzw. „Schulvisitationen“ von pädagogischen Experten zu kontrollieren, keinen Anlass darstellen, auf Schülerbefragungen zur Schulqualität zu verzichten. Auch diese Experten haben (nur) eine subjektive Sicht auf die vorgefundene Schulqualität, selbst wenn sie ihre Beobachtungen (was dringend zu empfehlen wäre) auf standardisierte Beobachtungsinventare stützen sollten. Darüber hinaus ist zu berücksichtigen, dass externe Beobachter im Gegensatz zu den Schülern nur einen außerordentlich kurzen Ausschnitt des schulischen Zusammenlebens kennen lernen und dabei auch noch einen Teil des zu beobachtenden sozialen Interaktionssystems darstellen, was die psychometrische Güte der Beobachtungsergebnisse beeinträchtigen dürfte (Cierpka, 1987). Anders ausgedrückt: Externe Beobachter können von Schülern und Lehrern leicht als „Störfaktoren“ im Unterricht wahrgenommen werden; es ist nicht auszuschließen, dass die beobachteten Unterrichtssequenzen sich deshalb von den alltäglichen Unterrichtssituationen deutlich unterscheiden, und die außerunterrichtlichen Aspekte der sozialen Schulqualität dürften externen Beobachtern vielfach verschlossen bleiben.

Aus unserer Sicht kann also eine erfolgreiche Optimierung der sozialen Schulqualität nur im Diskurs von Schülerschaft, Lehrerschaft und Elternschaft einer Schule gelingen. Dabei ist zu akzeptieren, dass der Begriff „Schulqualität“ von verschiedenen Interessengruppen mit Inhalt gefüllt werden muss. Dies ist auch deshalb notwendig, weil vor dem Hintergrund der Forderungen nach einer neuen Lehr-Lernkultur an unseren Schulen (vgl. z.B. Fullan, 1999; Reusser, 1995) ein Bildungsbegriff zugrunde gelegt werden muss, der auf Authentizität und Individualität, auf Selbstbestimmung und moralische Verantwortung jeden Schülers gerichtet ist. Es ist daher unbedingt zu begrüßen, dass seit einiger Zeit in höherem Maße als bisher die Wahrnehmungen von Schülern zur Qualität ihrer Schule in den Mittelpunkt des Interesses gerückt sind und diese Wahrnehmungen auch systematisch durch Schülerbefragungen erfasst werden: Schließlich werden die Lebenschancen der Schüler als „Endverbraucher“ und Kunden des Schulsystems durch die Schulqualität nachhaltig beeinflusst.

## 1.5 Literatur

Ainsworth, M.D.S., Blehar, M.C., Waters, E. & Wall, S. (1978). Patterns of attachment. A psychological study of the strange situation. Hillsdale NJ: Lawrence Erlbaum.

Bowlby, J.B. (1984). Eine Analyse der Mutter-Kind-Beziehung. Frankfurt/Main: Fischer.

- Brambring, M., Rauh, H. & Beelmann, W. (Hrsg.) (1996). Early childhood intervention. Theory, evaluation and practice. Berlin: DeGruyter.
- Bretherton, I. (1990). Open communication and international working models. Their role in the development of attachment relationships. In Thompson, R.A. (Ed.), Socioemotional development. Nebraska Symposium on Motivation 1988 (57-113). Lincoln: University of Nebraska Press.
- Campos, J. J., Mumme, D. L., Kermoian, R. & Campos, R. G. (1994). A functionalist perspective on the nature of emotion. In Fox, N. A. (Ed.), The development of emotion regulation: Biological and behavioral considerations. Monographs of the Society for Research in Child Development. Serial No. 240, Vol. 59, 284-303.
- Cassidy, J. (1994). Emotion regulation: Influences of attachment relationships. In Fox, N.A. (Ed.), The development of emotion regulation: Biological and behavioral considerations. Monographs of the Society for Research in Child Development. Serial No. 240, Vol. 59, 228-249.
- Cicchetti, D., Ganiban, J. & Barnett, D. (1991). Contributions from the study of high risk populations to understanding the development of emotion regulation. In Garber, J. & Dodge, K. A. (Eds.), The development of emotion regulation and dysregulation (15-48). Cambridge, New York, Port Chester, Melbourne, Sydney: Cambridge University Press.
- Cierpka, M. (1987). Familiendiagnostik. Berlin, Heidelberg, New York, London, Paris, Tokyo: Springer.
- Clausen, M. (2000). Wahrnehmung von Unterricht. Übereinstimmung, Konstruktvalidität und Kriteriumsvalidität in der Forschung zur Unterrichtsqualität. Berlin: Freie Universität.
- Clausen, M. (2002). Unterrichtsqualität: Eine Frage der Perspektive? Empirische Analysen zur Übereinstimmung, Konstrukt- und Kriteriumsvalidität.. Münster, New York, München, Verlin: Waxmann.
- Cox, A.D. (1970). Intrafamily comparison of loving rejection child-rearing practices. Child Development, 41, 437-448.
- Davids, A. & Hainsworth, P. (1967). Maternal attitudes about family life and childrearing as avowed by mothers and perceived by their under-achieving and high-achieving sons. Journal of Consulting Psychology, 31, 29-37.
- Donabedian, A. (1966): Evaluating the quality of medical care. In: Milbank Memorial Fund Quarterly, 44, 166-206.
- Ditton (2002). Lehrkräfte und Unterricht aus Schülersicht. Zeitschrift für Pädagogik, 48 (2), 262-286.
- Engelkamp, J. (1990). Das menschliche Gedächtnis. Das Erinnern von Sprache, Bildern und Handlungen. Göttingen, Toronto, Zürich: Hogrefe.
- Fend, H. (1998). Qualität im Bildungswesen. Weinheim: Juventa.
- French, L.A. (1985). Children's understanding of relational terms. In Kuczaj, S.A. & Barrett, M.D. (Eds.), Development of word meaning. New York: Springer.
- French, L.A. (1986). The language of events. In Nelson, K. (Ed.), Event knowledge. Structure and function in development (S. 119-136). Hillsdale, NJ: Erlbaum.
- Fullan, M. (1999). Die Schule als lernendes Unternehmen: Konzepte für eine neue Kultur in der Pädagogik. Stuttgart: Klett-Cotta.
- Funk, W. (Hrsg.). (1995). Nürnberger Schüler-Studie 1994. Gewalt an Schulen. Regensburg: Roderer.

- Garbe, U. & Strasser, E.M. (1978). Schulische Leistung und ihre Bedingungen aus der Sichtweise von Schülern und ihren Müttern. Unveröff. Diplomarbeit. Universität Konstanz, Fachbereich Psychologie und Soziologie, Konstanz:
- Gehring, T.M., Marti, D. & Sidler, A. (1994). Family System Test (FAST): Are parents' and children's family constructs either different or similar, or both? *Child Psychiatry and Human Development*, 25 (2), 124-138.
- Gerstenmaier, J. (1975). Urteile von Schülern über Lehrer: Eine Analyse ausgewählter empirischer Untersuchungen. Weinheim: Beltz.
- Grossmann, K.E. et al. (1989). Die Bindungstheorie: Modell und entwicklungspsychologische Forschung. In Keller, H. (Hrsg.), *Handbuch der Kleinkindforschung* S. 31-55). Berlin: Springer-Verlag.
- Grundmann, M., Welskopf, R. & Sturzbecher, D. (2001). Wozu braucht man kindliche Einschätzungen von Betreuungskontexten? In: D. Sturzbecher (Hrsg.), *Spielbasierte Befragungstechniken* (S. 11 - 23). Göttingen, Bern, Toronto, Seattle: Hogrefe.
- Haider, G. (2002). Schulqualität und Schülerleistungen. In C. Wallner-Paschon & G. Haider (Hrsg.), *PISA Plus 2000. Thematische Analysen nationaler Projekte* (S. 125-132). Innsbruck: StudienVerlag.
- Helmke, A. & Kischkel, K.H. (1980). Zur Wahrnehmung elterlichen Erziehungsverhaltens durch die Eltern und ihre Kinder und dessen Erklärungswert für kindliche Persönlichkeitsmerkmale. In: H. Lukesch, H., Perez, M. & Schneewind, K.A. (Hrsg.), *Familiäre Sozialisation und Intervention* (S. 81-105). Bern: Huber.
- Helmke, A. & Schrader, F.-W. (1993). Was macht erfolgreichen Unterricht aus? Ergebnisse der Münchner Hauptschulstudie. *Praxis Schule 5-10* (1), 11-13.
- Hoff, E.H., Minsel, W.R., Minsel, B. & Grüneisen, V. (1973). Beziehungen zwischen Erziehverhalten und Persönlichkeitsmerkmalen von Eltern und ihren Kindern. *Psychologie in Erziehung und Unterricht*, 20, 145-155.
- Hoffmann, J. (1982). *Das aktive Gedächtnis. Psychologische Experimente und Theorien zur menschlichen Gedächtnistätigkeit*. Berlin: Deutscher Verlag der Wissenschaften.
- Holtappels, H.G., Heitmeyer, W., Melzer, W. & Tillmann, K.-J. (Hrsg.) (1997). *Forschung über Gewalt an Schulen: Erscheinungsformen und Ursachen, Konzepte und Prävention*. Weinheim: Juventa.
- Hudson, J.A. (1986). Memories are made of this: General event knowledge and development of autobiographic memory. In Nelson, K. (Ed.), *Event knowledge. Structure and function in development* (97-118). Hillsdale, NJ: Erlbaum.
- Kluwe, R. & Spada, H. (1981). Wissen und seine Veränderung: Einige psychologische Beschreibungsansätze. In Foppa, K. & Groner, R. (Hrsg.), *Kognitive Strukturen und ihre Entwicklung*. Bern: Huber.
- Kohlmeyer, K. (2002). *Evaluation als Instrument von Schulentwicklung*. Göttingen: Institut für berufliche Bildung und Weiterbildung.
- Lehmann, R.H., Peek, R., Gänsfuß, R., Lutkat, S., Mücke, S. & Barth, I. (2000). Quasum: Qualitätsuntersuchung an Schulen zum Unterricht in Mathematik. In Ministerium für Bildung, Jugend und Sport des Landes Brandenburg (Hrsg.), *Schulforschung in Brandenburg* (1). Potsdam: Ministerium für Bildung, Jugend und Sport des Landes Brandenburg.
- Leiske, M., Sturzbecher, D. & J.-G. Keil (2001). Soziale Schulqualität aus der Sicht von Jugendlichen in Brandenburg. In D. Sturzbecher (Hrsg.), *Jugend in Ostdeutschland: Lebenssituation und Delinquenz* (S. 210-248). Opladen: Leske + Budrich.

- Lüders, M. & U. Rauin (2004). Unterrichts- und Lehr-Lernforschung. In Helsper, W. & J. Böhme (Hrsg.), *Handbuch der Schulforschung* (S. 691-720). Wiesbaden: Verlag für Sozialwissenschaften.
- Lukesch, H. & Tischler, A. (1975). Selbst- und fremdperzipierter Erziehungsstil. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, VII, 88-99.
- McKenry, P.C., Price-Bonham, S. & O'Bryat, S.L. (1981). Adolescent discipline: Different family members' perception. *Journal of Youth and Adolescence*, 10/81, Vol. 10(5), 327-337.
- Melzer, W. (1998). *Gewalt als soziales Problem in Schulen*. Opladen: Leske + Budrich.
- Nelson, K. (1981). Social cognition in a script framework. In Flavell, J.H. & Ross, L. (Eds.), *Social cognitive development: Frontier and possible futures* (97-118). New York: Cambridge University Press.
- Nelson, K. (1986). *Event knowledge. Structure and function in development*. Hillsdale, NJ: Erlbaum.
- Posch, P. & Altrichter, H. (1999). *Schulqualität. Merkmale schulischer Qualität in der Perspektive verschiedener Bezugsgruppen*. [WWW Dokument]. URL <http://www.qis.at>.
- Prenzel, M. & Schiefele, H. (1999). *Motivation und Interesse*. In L. Roth (Hrsg.), *Pädagogik. Handbuch für Studium und Praxis* (2. Aufl.). München: Ehrenwirth.
- Reusser, K. (1995). *Lehr-Lernkultur im Wandel: Zur Neuorientierung in der kognitiven Lernforschung*. In R. Dubs & R. Dörig (Hrsg.), *Dialog Wissenschaft und Praxis*. St. Gallen: Universität St. Gallen, Institut für Wirtschaftspädagogik.
- Rösner, E. (2000). *Schulqualität konkret*. [WWW Dokument]. URL <http://www.gew-Berlin.de/blz/archiv/2000/blz0009/blz0906.htm>
- Rumelhart, D.E. & Ortony, A. (1977). The representation of knowledge in memory. In R.C. Anderson, R.C., Spiro, R.J. & Montague, W.E. (Eds.), *Schooling and the acquisition of knowledge*. Hillsdale, NJ: Erlbaum.
- Rutter, M., Maughan, B., Mortimore, P. & Ouston, J. (1979). *Fifteen thousand hours: Secondary schools and their effects on children*. London: Open Books.
- Ryan, R.M., Stiller, J. & Lynch, J.H. (1994). Representations of relationships to teachers, parents, and friends as predictors of academic motivation and self-esteem. *Journal of Early Adolescence*, 14, 226-249.
- Schank, R.C. (1975). Concepts for representing mundane reality in plans. In Bobrow, D.G. & Collins, A. (Eds.), *Representation and understanding*. New York: Academic Press.
- Schank, R.C. & Abelson, R.P. (1977). *Scripts, plans, goals and understanding*. Hillsdale, NJ: Erlbaum.
- Schneewind, K.A. (1991). *Familienpsychologie*. Stuttgart, Berlin, Köln: Kohlhammer.
- Schubarth, W. & Melzer, W. (1995). *Schule, Gewalt und Rechtsextremismus* (2. Aufl.). Opladen: Leske + Budrich.
- Scottish Office (1992). *Using ethos indicators in secondary school self evaluation – taking account of the views of pupils, parents and teachers*. The Scottish Office. Education Department: Edinburgh.
- Silbereisen, R.K. (1987). *Soziale Kognition. Entwicklung von sozialem Wissen und Verstehen*. In Oerter, R. & Montada, L. (Hrsg.), *Entwicklungspsychologie* (696-736). München: Psychologie Verlags Union.
- Steltmann, K. (1992). *Lehrerbeurteilung und Lehrereffektivität*. In Ingenkamp, K. u.a. (Hrsg.), *Empirische Pädagogik 1970-1990* (Bd. 2, S. 565-589), Weinheim: Beltz.

- Sturzbecher, D. (2001). Methodische Lösungsansätze zur Befragung jüngerer Kinder. In D. Sturzbecher (Hrsg.), *Spielbasierte Befragungstechniken* (S. 51 - 62). Göttingen, Bern, Toronto, Seattle: Hogrefe.
- Sturzbecher, D. & Freytag, R. (1999). Ein Vergleich elterlicher und kindlicher Einschätzungen der Eltern-Kind-Interaktion und ihres entwicklungsprognostischen Wertes. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 31 (1), 32-44.
- Sturzbecher, D. & Freytag, R. (2000). Familien- und Kindergarten-Interaktions-Test (FIT-KIT). Handanweisung, Testkasten und Protokollbogen. Göttingen, Bern, Toronto, Seattle: Hogrefe.
- Sturzbecher, D. & Hess, M. (2002). Soziale Schulqualität aus Schülersicht. In: D. Sturzbecher (Hrsg.). *Jugendtrends in Ostdeutschland: Bildung, Freizeit, Politik, Risiken* (S. 115-181). Opladen: Leske + Budrich.
- Sturzbecher, D., Hess, M. & Them, W. (2002). Jugendgewalt und Reaktionen des sozialen Umfelds. In D. Sturzbecher (Hrsg.), *Jugendtrends in Ostdeutschland: Bildung, Freizeit, Politik, Risiken* (S. 182-209). Opladen: Leske + Budrich.
- Sturzbecher, D., Landua, D. & Shahla, H. (2001). Jugendgewalt unter ostdeutschen Jugendlichen. In D. Sturzbecher (Hrsg.), *Jugend in Ostdeutschland - Lebenssituationen und Delinquenz* (S. 249 - 300). Opladen: Leske + Budrich.
- Thompson, R.A. (1993). Socioemotional development: Enduring issues and new challenges. *Developmental Review*, 13, 372-402.
- Thompson, R.A. (1994). Emotion regulation: A theme in search of definition. In Fox, N.A. (Ed.) *The development of emotion regulation: Biological and behavioral considerations*. Monographs of the Society for Research in Child Development, Serial No. 240, Vol. 59, 7-24. Chicago: The University of Chicago Press.
- Tillmann, K.-J. (Hrsg.). (1994). *Was ist eine gute Schule?* Hamburg: Bergmann + Helbig.
- Tillmann, K.J.& Meier, U. (2001): *Schule, Familie und Freunde – Erfahrungen von Schülerinnen und Schülern in Deutschland*. In Baumert, J. u.a. (Hrsg.), *PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich* (S. 468-510). Opladen: Leske + Budrich.

## 2 Qualität von Schulen – ein Forschungsüberblick

D. Jenkel

### 2.1 Die Messung von Schulqualität

Das mäßige Abschneiden deutscher Schülerinnen und Schüler bei der Third International Mathematics and Science Study (TIMSS) führte Mitte der 1990er Jahre zu einem Umschwung in der Bildungspolitik. Es wurde von einer „empirischen Wende“ oder einem „Paradigmenwechsel“ gesprochen. Nicht mehr die Kontrolle und Steuerung der Rahmenbedingungen des Lernens an deutschen Schulen standen im Mittelpunkt, sondern das Ergebnis des Bildungsprozesses. Nachdem im ersten Kapitel gezeigt wurde, warum es wichtig ist, den Konsumenten des Schulsystems – den Schülerinnen und Schülern – Aufmerksamkeit zu schenken, stellt sich nun die Frage nach der besten Herangehensweise. Ziel dieses Kapitels ist es, einen Überblick über bereits vorliegende Instrumente und Ansätze der empirischen Schulforschung zu geben, bevor im nachfolgenden Kapitel unser eigener Forschungsansatz beschrieben wird. In diesem Kapitel werden drei Forschungstraditionen skizziert, die für die Schulqualitätsdiskussion bedeutsam sind: die Unterrichtsforschung, die Schulklimaforschung und die Schuleffektivitätsforschung. Da neuere Studien die verschiedenen Ansätze miteinander kombinieren, werden anschließend ausgewählte integrative Ansätze vorgestellt. Zunächst sollen jedoch die fünf wichtigsten Methoden dargestellt werden, die für das Verständnis der verschiedenen Verfahren von Bedeutung sind. Hierzu gehören Dokumentenanalyse, Leistungstests, Befragung, Beobachtung und Experiment.

Nicht immer ist es notwendig neue Daten zu sammeln. Für eine Reihe von Fragen kann auf das in der Schule schon vorhandene Datenmaterial zurückgegriffen werden. Die zahlenmäßige Entwicklung von Übergangsquoten, Abschlüssen, Wiederholern und Abbrechern, Leistungsbewertungen, Kurswahlen, Bildungskosten, etc. bieten sich hierfür an. Mit den Mitteln der Dokumentenanalyse und Statistik können so wertvolle Informationen verschiedener Schulen gesammelt und vergleichend betrachtet werden. Voraussetzung ist deren Erhebung und die Vergleichbarkeit der Daten. Der Vorteil dieser Methode ist, dass sie relativ wenig kostet und die Ergebnisse schnell verfügbar sind, da diese Informationen von den Schulen und der Schulverwaltung ohnehin gesammelt werden. Mit Hilfe der Dokumentenanalyse können aber auch Unterrichtsmaterialien, Schulprogramme oder Aufsätze analysiert werden. Je nach Auswertungsart wird hierfür ein Kategorienschema erstellt. Ein Beispiel dafür ist die Aufsatzstudie (The Written Composition Study) der International Association for the Evaluation of Educational Achievement (IEA), die von 1980-1988 (ohne deutsche Beteiligung) durchgeführt wurde.

Eine ganz andere Vorgehensweise wird bei der Schulleistungsforschung gewählt. Die ersten Tests zur Erfassung von Persönlichkeitsmerkmalen wurden vor etwa 100 Jahren für die Diagnose von Intelligenzdefiziten entwickelt (z.B. Binet & Simon 1905). Das Interesse an solchen Tests führte zur Entstehung der klassischen Testtheorie, später auch der probabilistischen, auf deren Basis eine Vielzahl unterschiedlicher Tests entwickelt wurde. Schulleistungstests sind standardisierte Messinstrumente zur Erfassung individueller Schülerleistungen. Standardisiert heißt, dass die Testgütekriterien eingehalten werden und eine Normierung vorgenommen wird. So kann das Ergebnis eines Schülers mit dem Durchschnittswert einer Gruppe verglichen werden. Solche Tests gibt es für verschiedene Altersgruppen und Fächer. Gemeinsam ist ihnen eine starke Standardisierung in der Konzeption, Durchführung und Auswertung. Dabei sind Stichprobenerhebungen genauso möglich (z.B. IEA-Studien und PISA) wie eine Vollerhebung aller Schüler und Schülerinnen

einer Grundgesamtheit (z.B. die Lernausgangslagenuntersuchung in Hamburg oder MARKUS in Rheinland-Pfalz). Schulleistungstests sind teuer und zeitaufwändig. So wurden die Ergebnisse von PISA 2000 (die im Mai und Juni 2000 durchgeführt wurde) erst Ende 2001 veröffentlicht. Das heißt, eine Mehrheit der untersuchten 15-Jährigen hatte die Schule bereits verlassen, als die Untersuchungsergebnisse zu ihrer Schulleistung vorlagen.

Befragungen bieten sich immer dann an, wenn man Meinungen und Einstellungen erfassen will. Von besonderer Bedeutung ist dabei die Perspektive der Konsumenten des Schulsystems – der Schülerinnen und Schüler (vgl. Kapitel 1). Aber auch die Sichtweise der Eltern, der Lehrer/innen, der Schulleitung und anderer Personen kann durch Befragungen berücksichtigt werden. Somit ist es möglich, die unterschiedlichen Sichtweisen mehrerer Gruppen auf das System Schule zu erfassen. Durch die Aggregation der Antworten der Teilnehmer/innen einer solchen Untersuchung sind sowohl repräsentative Aussagen für die einzelne Schule (Erstellung eines Schulprofils), einer Schulform oder der deutschen Schülerschaft insgesamt möglich. Gleichzeitig kann man einzelne Analyseeinheiten betrachten (z.B. auf Klassenebene). Befragungen unterscheiden sich im Grad der Standardisierung. Man kann sie mündlich oder schriftlich durchführen. In der Schulforschung dominiert die schriftliche Befragung mittels eines Fragebogens. Für die Auswertung werden in der Regel Skalen genutzt, die mehr oder weniger empirisch bewährt sind. Der Vorteil der Verwendung bewährter Skalen ist die Existenz eines Vergleichsmaßstabs und die Kenntnis über die Güte der Skalen. Befragungen kann man in identischer Form wiederholen (Längsschnittuntersuchungen sind genauso möglich wie wiederholte Querschnitte). Der Nachteil solcher Befragungen ist, dass das Gefragte nicht ohne Einfluss auf die Antworten ist. Soziale Erwünschtheit, Meinungslosigkeit oder Antwortverzerrungen beeinflussen die Erhebung.

Klassisches und ältestes Beobachtungsinstrument im Bereich der Schule ist die Unterrichtshospitation. Die Hospitation ist eine Form der teilnehmenden Beobachtung. Hospitationen sind u.a. durch die Schulaufsicht, die Schulleitung, durch die Kollegen, durch die Eltern oder Forscher/innen möglich. Im Rahmen des Schul-TÜVs wird sie durch Schulinspektoren (bzw. wie in Brandenburg durch Visitatoren) durchgeführt. Allen ist gemeinsam, dass sie mit einer bestimmten Zielstellung das Unterrichtsgeschehen verfolgen und hinterher auswerten. Dies kann ein sehr umfassendes (standardisiertes) Beobachtungsinstrument voraussetzen, aber auch ohne ein bestimmtes System erfolgen. Ein Beispiel für eine große internationale Studie mit der Methode der Hospitation ist die „Classroom Environment Study“. Auch im Rahmen der TIMS-Studie wurde in Deutschland das Unterrichtsgeschehen beobachtet. Hierzu wurde der Unterricht mit einer Videokamera aufgezeichnet. Diese Vorgehensweise hat im Hinblick auf die Güte der Beurteilung erhebliche Vorteile. Mit Beobachtungen sind aber auch eine Reihe von Fehlern oder Einschränkungen verbunden. Unterrichtshospitationen – ob angekündigt oder nicht – beeinflussen den Unterricht. Gleichzeitig beobachtet man in der Regel nicht den gesamten Unterricht, sondern nur wenige Stunden oder gar Minuten.

Eine weitere wichtige Methode ist das Experiment. Mit ihm können kausale Beziehungen überprüft werden. In Experimenten werden eine oder mehrere Bedingungen verändert und deren Wirkungen auf eine abhängige Variable untersucht. In der Regel werden dazu zwei Gruppen gebildet. In der einen Gruppe wird eine Versuchsbedingung manipuliert und in der anderen nicht. So kann der Effekt der unabhängigen Variable auf die abhängige bestimmt werden. Die Kontrolle aller Randbedingungen ist in der empirischen Bildungsforschung aber schwierig. Häufig können keine Laborexperimente durchgeführt werden. Soll zum Beispiel überprüft werden, ob Projektunterricht (unabhängige Variable) zu besseren Lernergebnissen (abhängige Variable) führt als der herkömmliche Unterricht, wird man dies in der Schule tun. Somit wird ein Feldexperiment durchgeführt. Neben der Schwierigkeit der Äquivalenz der beiden Gruppen, kann die Versuchsanordnung von einer Reihe von Störfaktoren beeinflusst

werden. Sie reichen von Unterhaltungen der Schüler/innen untereinander, dem Anschauen einer bestimmten Fernsehsendung im Vorfeld des Experiments bis hin zu Lehrereffekten. Die Vielfalt der zu kontrollierenden Bedingungen führt dazu, dass in Soziologie und Pädagogik im Gegensatz zur psychologischen Forschung nur wenige Experimente durchgeführt werden.

Neben den fünf hier vorgestellten Forschungsmethoden stehen eine Reihe von alternativen Methoden zur Verfügung. Als Beispiele seien genannt: Gruppendiskussionen, die Einbeziehung von „Werkstücken“, Klassenarbeitsnoten, Arbeiten von Schülerinnen und Schülern, Bilder, Photo- und Videodokumentationen.

## 2.2 Überblick über Verfahren und Studien zur Bestimmung von Schulqualität

1771/72 wurde zum ersten Mal in der Schweiz eine flächendeckende und systematische Erhebung des Zustands der Landschulen vorgenommen. Von 110 der 150 Kirchengemeinden sind ausgefüllten Fragebögen erhalten. Die darin enthaltenen 81 Fragen beziehen sich beispielsweise auf die Infrastruktur, die Organisation und das Curriculum der Schulen, die angewandten Lehrmittel und Methoden, das Einkommen und den Status der Schulmeister. Die zum Teil sehr umfangreichen Ausführungen der Landpfarrer geben insgesamt einen differenzierten Einblick in den schulischen Alltag mit seinen Problemen und Konflikten, etwa bei der Durchsetzung des allgemeinen Schulbesuches, im Kräfteverhältnis zwischen Schulmeister, Pfarrer, Eltern und Dorfborgerschaft, aber auch in der Auswirkung ökonomischer Verhältnisse (Missernten und Teuerung von 1770–1772) auf Erziehung und Unterricht (Berner, 2002). Dieses Beispiel illustriert, dass die systematische Beschäftigung mit der schulischen Bildung keine Erfindung der Neuzeit ist. Eine wichtige Quelle der modernen Schulforschung war die Chicagoer Schule, eine soziologische Schule, die zwischen den beiden Weltkriegen in Chicago entstand. Stellvertretend seien die Arbeiten von Dewey und die Eight-Years-Study von 1933-1941 genannt (Aikin, 1942). Schon diese frühen Studien verdeutlichen, dass man die Qualität einer Schule aus unterschiedlichen Blickwinkeln betrachten kann. Im Folgenden werden drei Forschungstraditionen nacheinander dargestellt. Diese Trennung in der Darstellungsform wurde gewählt, um die Entstehung und die Schwerpunkte der einzelnen Forschungsansätze voneinander abzugrenzen. Die neueren Forschungen sind sehr viel integrativer ausgerichtet und nicht mehr einfach einer der beschriebenen Richtungen zuzuordnen. Der erste Teil der vorgestellten Verfahren stellt die Unterrichtsforschung in den Vordergrund. Dem schließt sich die Betrachtung der Schulklimaforschung an, gefolgt von der Schuleffektivitätsforschung. Den Abschluss bilden eine Reihe von neueren integrativen Verfahren der Schulevaluation und Leistungsforschung. Beginnen wir nun mit der Unterrichtsforschung.

## 2.3 Unterrichtsforschung

Das Kerngeschäft von Schulen ist der Unterricht. Im Mittelpunkt dieses Abschnitts stehen deshalb Verfahren, welche die Unterrichtsqualität erfassen. Durch die zunehmende Anzahl empirischer Untersuchungen in der Schulforschung wurde eine Vielzahl von Faktoren identifiziert, die für erfolgreiche schulische Lernprozesse von Schülern und Schülerinnen von Bedeutung sind (Helmke & Weinert, 1997). Unterschiedliche Forschungstraditionen haben dabei die Unterrichtsforschung beeinflusst. Bereits am Ende der 1940er Jahre wurden in den USA Effektivitätsstudien durchgeführt. Sie orientierten sich an behavioristischen Lerntheorien und versuchten, effektive Unterrichtsmethoden und Eigenschaften guter Lehrer und Lehrerinnen zu identifizieren (zusammenfassend: Gage, 1963). Mit ihnen gelang es allerdings nicht, eine allgemeine Unterrichtstheorie zu entwickeln. Hinzu kam, dass Laborbefunde nicht der komplexen interaktiven Wechselwirkung zwischen Lehrerinnen und Lehrern und Schülerinnen und Schülern entsprachen. So entwickelte sich die Lehr-Lern-Forschung (LLF), die sich mit der Analyse und Darstellung von Voraussetzungen, Prozessen



und Ergebnissen menschlichen Lernens in verschiedenen Lernsituationen beschäftigt (Lüders & Rauin, 2004). In diesem Forschungsprogramm gewann die Prozess-Produkt-Forschung an Bedeutung. Mit ihr konnten alltägliche Unterrichtssituationen untersucht werden. Dabei wurden einerseits Prozessvariablen wie die Schüler-Lehrer-Interaktion erhoben, andererseits aber auch Produktvariablen wie die Schülerleistung bzw. der Lernzuwachs als Maß der Effektivität des Unterrichts. Prozess- und Produktvariablen wurden anschließend zueinander in Beziehung gesetzt.

Erste Modellierungen der Unterrichtsqualität fanden in den 60er und 70er Jahren des letzten Jahrhunderts statt (z.B. Carroll, 1963; Bloom, 1971). Carroll ging davon aus, dass der Lernerfolg maßgeblich durch das Verhältnis von benötigter und aufgewendeter Lernzeit bestimmt wird. Bloom hingegen betont den sequenziellen und kumulativen Charakter von schulischen Lernprozessen. Wie Carroll entwickelte auch Bloom ein Lernmodell, das verschiedene Faktoren spezifiziert und deren Wirkungen beschreibt. Es bildet die Grundlage für das Konzept des zielgerichteten Lernens. In den 90er Jahren entwickelte Slavin ein Modell der Unterrichtsqualität – das QuAIT-Modell (Slavin, 1990, 1996). Er fand auf der Basis umfangreicher Literaturanalysen und eigener Studien vier wirksame Faktoren, die einen erfolgreichen Unterricht charakterisieren: die Qualität der Instruktion, die Angemessenheit des Unterrichts, die Motivierung und Anreize der Schülerinnen und Schüler und die Lernzeit. Für den Lernerfolg sind alle vier angeführten Merkmale gleichermaßen wichtig, d.h. sie sind nicht gegenseitig kompensierbar.

Wichtig für die Unterrichtsforschung war weiterhin die Erziehungs- und Unterrichtsstilforschung. Rosenthal und Jacobson beschrieben 1968 den Pygmalion-Effekt: Die Autoren der Studie bestimmten die Intelligenzwerte von 255 Schülerinnen und Schülern. Die Namen der Kinder, die am begabtesten waren, wurden den Lehrkräften mitgeteilt. Tatsächlich wurden diese Schülerinnen und Schüler aber per Zufall ausgewählt. Ein Jahr später wurden die Kinder wieder einem Intelligenztest unterzogen. Die „begabten“ Schülerinnen und Schüler erhielten dabei höhere Zugewinne als die „weniger begabten“ Kinder. Rosenthal und Jacobson wiesen damit nach, dass die Intelligenzentwicklung von Kindern positiv mit den Leistungserwartungen der Lehrerschaft korreliert. Die Autoren machten dafür die unterschiedliche Behandlung der Kinder durch die Lehrkräfte verantwortlich, die begabten Kindern mehr Aufmerksamkeit und Freundlichkeit schenkten. Somit gewannen psychologische Merkmale der Lehrerpersönlichkeit an Bedeutung. Wie nehmen Lehrerinnen und Lehrer ihre Schülerinnen und Schüler wahr, welche Ursachen schreiben sie schulischen Erfolg und Misserfolg zu? Die Kognitionspsychologie entwickelte sich somit zur Basis der Unterrichtsforschung.

Zusammenfassend lässt sich sagen, dass die Unterrichtsforschung durch eine Vielzahl von Forschungstraditionen beeinflusst wurde. Dazu gehören neben der Forschung zu Erziehungs- und Unterrichtsstilen auch Untersuchungen zum effektiven Lehrerverhalten, die Lehrmethodenforschung sowie Forschungen zur Klassenführung und pädagogischen Interaktion (Clausen, 2001). Als Ergebnis kann festgehalten werden, dass es „den“ guten Unterricht ebenso wenig gibt wie eine ideale Lehrmethode. Unterschiedliche Bildungsziele erfordern unterschiedliche Lernmethoden (z.B. Ditton, 2002b; Helmke, 2003). Es liegen etwa 100 Konstrukte zur Erfassung und Erklärung der Unterrichtsqualität vor (Clausen, Schnabel & Schröder, 2002). Die Unterrichtsqualität kann nach normativen Maßstäben gemessen (wie z.B. der Prozessaspekt: Wann ist Unterricht gut?) oder an den Wirkungen bzw. Effekten des Unterrichts abgelesen werden (siehe Schulleistungsforschung).

Weinert, Schrader und Helmke (1989) definieren Unterrichtsqualität als „jedes stabile Muster von Instruktionsverhalten, das als Ganzes oder durch einzelne Komponenten die substantielle Vorhersage und/oder Erklärung von Schulleistung erlaubt.“ (S. 899). Diese Fokussierung auf das Produkt „Leistung“ wurde vielfach kritisiert (Einsiedler, 2002). In neueren Arbeiten

werden auch Kriterien wie die emotionale und motivationale Entwicklung der Schülerinnen und Schüler einbezogen (Helmke & Schrader, 1998). Weitere Kritikpunkte waren die Betrachtung von Einzelmerkmalen des Unterrichts ohne Faktorenkombinationen zu berücksichtigen und der Black-Box-Charakter des Modells (Ditton, 2002b), denn es gelang bisher nicht, eine erklärende Theorie zu entwickeln, wie sich die ermittelten Effekte des Lehrerverhaltens auf die Schülerleistung niederschlagen.

Will man die Qualität einer Unterrichtsstunde erfassen, kann man dies auf unterschiedliche Art und Weise tun. Diese Methoden können von verschiedenen Akteuren und – damit eng verbunden – aus mehreren Perspektiven heraus eingesetzt werden. Eine Selbstbeurteilung der Lehrkräfte wird anders ausfallen als ein Schülerfeedback. Auch die Schulleitung, Peers, die Schulaufsicht oder Wissenschaftlerinnen und Wissenschaftler können Quellen der Qualitätsbewertung sein. Ein Beispiel für diese Forschungstradition ist die Classroom Environment Study (CES) der IEA, die weltweit in neun Ländern von 1978 bis 1989 durchgeführt wurde (Anderson, Ryan, Shapiro, 1989). Ihre Ziele waren die Beschreibung von Ähnlichkeiten und Unterschieden beim Unterrichten in verschiedenen Ländern und die Analyse des Lehrverhaltens, das mit hoher Schülerleistung korreliert. Zwei zentrale Kategorien wurden identifiziert: Managementpraktiken, mit denen Schülerinnen und Schüler an Lernaufgaben herangeführt werden und Instruktionspraktiken, die den Lernprozess der Schüler unmittelbar strukturieren.

Den deutschen Beitrag dieser Studie bildete die Münchener Hauptschulstudie, die von Weinert und Helmke geleitet wurde. Die zentrale Fragestellung war der Einfluss der Unterrichtsqualität auf das Lernverhalten von Hauptschülern und Hauptschülerinnen, wobei kognitive und motivationale Merkmale der Kinder, Besonderheiten des Elternverhaltens und schulische Kontextbedingungen berücksichtigt wurden (Helmke, Schneider & Weinert, 1986; Helmke & Schrader, 2001). Die Untersuchungen wurden 1985, die notwendigen Datenanalysen 1989 beendet. An der Untersuchung nahmen 69 Hauptschulklassen der 5. und 6. Jahrgangsstufe teil. Methodisch umgesetzt wurde diese Studie durch einen Mathematikleistungstest, einen Intelligenztest, einen Fragebogen für Schülerinnen und Schüler sowie für Lehrkräfte, durch Interviews mit Lehrkräften und Eltern sowie durch intensive Unterrichtsbeobachtung. Hierzu wurde ein Beobachtungssystem entwickelt, in dem der Unterrichtsverlauf und die Schüler-Lehrer-Interaktionen erfasst wurden. Die Leistungsentwicklung von ca. 850 Schülerinnen und Schülern wurde von Beginn der 5. Klasse an bis zum Ende der 6. Klasse analysiert. In neun Mathematikstunden wurde das Lehrer- und Schülerverhalten beobachtet. Viermal wurde die Lernmotivation und Leistungsangst der Schülerinnen und Schüler erhoben. In diesen Beobachtungen der Unterrichtsqualität werden eine Reihe von Aspekten mit berücksichtigt, die auch für „Soziale Schulqualität“ wichtig sind, denken wir beispielsweise an die Schüler-Lehrer-Interaktion. Es wurden folgende Unterrichtsaspekte bewertet: Die effektive Regelverwendung, das Vermeiden zeitlicher Fehleinschätzungen (das Eingreifen bei Störungen durch die Schüler und Schülerinnen), die Vermeidung von Adressierungsfehlern (Kennt der Lehrer/die Lehrerin die wahren Urheber der Störung?), die Vermeidung von Überreaktionen (Ist die Reaktion des Lehrers bzw. der Lehrerin bei Störungen angemessen?), das Aufnehmen von Schülerbeiträgen, das Verwenden von Stützmaßnahmen (der Umfang von Stützmaßnahmen bei festgestellten Lerndefiziten), die Verständlichkeit der Lehrerräuerungen, das Lehrerengagement (das Interesse, die Energie sowie die Beziehung zum Fach und zum Unterricht), die Unterrichtsorganisation, die Lehrstofforientierung (Wie viele fachliche Inhalte werden vermittelt, wie viel Zeit wird für organisatorische Fragen verwendet? Sind Diskussionen zielführend oder abschweifend?), die Verteilung der Antwortgelegenheiten, die Beschäftigung nach einer Aufgabenstellung, die Hausaufgabenkontrolle, das Aufrufen der Schüler und Schülerinnen nach einem festem Schema, das Bemühen des Lehrers/der Lehrerin,

die Schülerschaft durch interessante Beispiele zu motivieren, das Unterrichtsengagement der Lernenden, die Lehrerreaktion auf Schülerantworten, das Zeit lassen zum Überlegen, die Sprech- und Ausdrucksweise des Lehrers/der Lehrerin, der Lärmpegel in der Klasse und nicht zuletzt die Wärme, die Herzlichkeit und der Humor des Lehrers bzw. der Lehrerin (Helmke, 2003).

Unterrichtshospitationen, wie sie in der Münchener Studie durchgeführt wurden, sind sehr aufwändig. In der Regel müssen sie videographiert und anschließend von mehreren Beobachtern ausgewertet werden. So können sie nur einen kleinen Teil des Unterrichtsgeschehens untersuchen. Damit stellt sich die Frage, wie repräsentativ die Beobachtungen oder Aufzeichnungen für den Unterricht eines bestimmten Lehrers bzw. der Schule insgesamt sind. Die Hospitation selbst beeinflusst sowohl das Lehrer- als auch das Schülerverhalten. All dies muss man bedenken, wenn man die Unterrichtsqualität nur beobachtet. Eine ganz andere Vorgehensweise ist die Schülerbefragung zum Unterricht, die die explizite Sichtweise der Betroffenen beschreibt.

Schülerfeedbacks zum Unterricht erfassen die Kundensichtweise. Mit ihnen können die schülerspezifischen Wahrnehmungen des Unterrichts und des Lehrpersonals analysiert werden (Ditton, 2002a). Schülerinnen und Schüler haben einige tausend Stunden Unterrichtserfahrung. Hinzu kommt, dass sie in dieser Zeit von unterschiedlichen Lehrern und Lehrerinnen unterrichtet werden. Sie haben somit Vergleichsmöglichkeiten. Die individuellen Auffassungen zum Unterricht können zu einem Gesamtbild aggregiert werden. „Verlässliche Aussagen über die Unterrichtsqualität scheinen am ehesten über den von individuellen Präferenzen abstrahierenden gemeinsamen Anteil der Wahrnehmungen möglich zu sein.“ (Ditton, 2002a: 264f). Somit kann auf der Basis von Mittelwerten die durchschnittlich wahrgenommene Qualität des Unterrichts erfasst werden. Wahrnehmungsunterschiede innerhalb einer Klasse können dabei als Fehlervarianz und bzw. oder leistungsbedingter Wahrnehmungsbias betrachtet werden (Gruehn, 2000).

Am Beispiel des sächsischen Pilotprojekts „Unterrichtsbeurteilung durch Schüler“ soll kurz die Vorgehensweise erläutert werden (nach Helmke, 2003). Die Untersuchung, die im Auftrag des sächsischen Kultusministeriums durchgeführt wurde, begann 1999. An der Erhebung waren zwölf allgemein bildende und berufsbildende Schulen beteiligt. Befragt wurden Kinder und Jugendliche ab Klassenstufe 6. Dieses Schülerfeedback wurde ausgewertet und an die Schulen zurückgegeben. Die beteiligten Lehrerinnen und Lehrer erhielten ihre Ergebnisse separat in verschlossenen Umschlägen, was die Akzeptanz im Kollegium erhöhte. Weder die Schulaufsicht noch die Schulleitung erhielt Kenntnis von den individuellen Ergebnissen. Die Forschergruppe und die beteiligten Lehrerinnen und Lehrer kamen zu dem Schluss, dass Schülerinnen und Schüler sehr wohl in der Lage sind, den Unterricht valide zu beurteilen. Schülerbefragungen zum Unterricht können somit ein wichtiger Bestandteil schulinterner Evaluation sein. Für eine solche Vorgehensweise spricht, dass sich die Schülerinnen und Schülern aktiv einbezogen fühlen und eine gemeinsame Basis für die Reflexion des Unterrichts entsteht (Eikenbusch, 2001). Die Schülerbefragung eignet sich immer dann als Methode, wenn konkret wahrnehmbare Aspekte erfasst werden sollen, für die keine didaktischen Kenntnisse nötig sind (Clausen, 2002).

Ein anderes Instrument der Unterrichtsbeurteilung durch die Schülerschaft ist der „Stuttgarter Unterrichtsbeurteilungsbogen“, der im Internet kostenlos zur Verfügung gestellt wird. Er wurde vom Institut für Berufs-, Wirtschafts- und Technikpädagogik der Universität Stuttgart für Berufsschulen entwickelt, wird aber auch von allgemein bildenden Schulen genutzt (Mayer & Nickolaus, 2003). Diese Schülerfeedbacks sollen dazu dienen, das Klassenklima und den Unterricht zu verbessern. Die Autoren merken an, dass bei der Beurteilung des Lehrpersonals mit Verzerrungen zu rechnen ist, die durch den Leistungsstand der Schülerinnen und Schüler oder die Beliebtheit und Schwierigkeit der Fächer hervorgerufen werden können.

Der Fragebogen enthält 60 Items, die einerseits den Ist-Zustand erfassen, andererseits aber auch die Relevanz der Aussage messen. Sie sind überwiegend positiv formuliert und werden sechsstufig abgefragt. Daraus werden Skalen gebildet, die die Erzeugung von Aufmerksamkeit, die Angemessenheit der Anforderungen, die Vermittlung der inhaltlichen Bedeutsamkeit, das Ermöglichen von Erfolgserlebnissen, die inhaltliche Einbindung, die Gestaltung des sozialen Klimas sowie die Gerechtigkeit und Ausstrahlung des Lehrers oder der Lehrerin messen. Die Auswertung erfolgt mittels einer Excel-Datei, in die nur die Schülerantworten eingegeben werden müssen. Dann erhält man eine Reihe von Ergebnisdiagrammen, die die Relevanz jeder Aussagen ihrer Bewertung gegenüberstellt. Mit Excel können anschließend weitere Analysen durchgeführt werden. Der Fragebogen enthält keine demographischen Merkmale oder die Abfrage nach Noten, sodass keine Subgruppenanalysen möglich sind. Allerdings kann der Fragebogen modifiziert werden, man verzichtet damit aber auf die einfache und bequeme Ergebnisauswertung.

Zusammenfassend lässt sich festhalten, dass Schülerangaben zum Unterricht eine wichtige Perspektive sind, aber diese Vorgehensweise Grenzen aufweist. Sie ist besonders für formative Evaluationen und als Werkzeug der Unterrichtsentwicklung geeignet, aber kaum für summative Bewertungen der Unterrichtsqualität (Helmke, 2003). Der Vergleich unterschiedlicher Perspektiven zeigt, dass zwischen den Schüler-, Lehrer- und Beobachterangaben zum Unterricht eher eine geringe Übereinstimmung besteht. Clausen (2000) argumentiert, dass diese Unterschiede nicht als Fehler zu werten sind, sondern als perspektivenspezifische Wahrnehmungen. Weiter muss angemerkt werden, dass die Erfassung der Wahrnehmung des Unterrichts durch die Schülerinnen und Schüler eine naheliegende und zugleich ökonomische Alternative zur Unterrichtsbeobachtung ist. Bisher fehlte jedoch eine Evaluationskultur des Unterrichts wie sie beispielsweise in jüngerer Vergangenheit an deutschen Hochschulen Einzug gehalten hat (Helmke & Schrader, 2001). Um dies zu ändern, wurde beispielsweise im Rahmen des Programms „Bildungsqualität“ von Schule (BIQUA) u. a. der Effekt von Schülerrückmeldungen zur Qualitätssicherung des Unterrichts untersucht.

## 2.4 Schulklimaforschung

### 2.4.1 Überblick

Wenn man eine Schule betrachtet, wird deutlich, dass jede Schule einzigartig ist. Es sind nicht nur die Unterschiede in der Architektur und Gestaltung der Schulgebäude oder die geographische Lage der Schule, die diese Unterschiede ausmachen. Wenn man den Schulhof betritt, eröffnet sich ein eigener Mikrokosmos, es offenbart sich ein spezifisches „Klima“ einer Schule. Dort ein asphaltierter, liebloser Pausenhof, dort eine Abenteuerlandschaft oder selbstgebaute Holzstuhlguppen. Auch der Umgang der Schüler und Schülerinnen untereinander unterscheidet sich. Es drängt sich unwillkürlich die Frage auf: Woraus resultieren diese Besonderheiten und wie entstehen sie?

“School climate is much like the air we breathe – it tends to go unnoticed until something is seriously wrong.” (Freiberg, 1999: 1) Mit diesen Worten beschreibt Jerome Freiberg einen Wesenszug von Schulklima, der erklärt, weshalb die Erforschung des Schulklimas lange unterschätzt wurde: Schulklima ist eine Kontextvariable der allgemeinen Schulsituation, die erst dann Aufmerksamkeit erregt, wenn etwas schiefgegangen ist. Im englischen Sprachraum wird für den deutschen Begriff ‚Klassenklima‘ der Begriff ‚Classroom environment‘ verwendet. Die Schulklimaforschung basiert auf unterschiedlichen Ansätzen. Diese verbindet, dass die subjektive Wahrnehmung und Bewertungen der schulischen Umwelt durch die Betroffenen im Mittelpunkt stehen (Eder, 2001). Mittel der Wahl zur Erfassung des Schulklimas ist die Befragung. Je nach Analyseeinheit unterscheidet man zwischen dem

Klima einer Klasse, dem Unterrichtsklima, dem Klima im Lehrkörper und dem Schulklima. Für das Klima in Klassen und Schulen sind vier Dimensionen von zentraler Bedeutung: Die Schüler-Schüler-Beziehung, die Schüler-Lehrer-Beziehung, die Unterrichtsqualität und die Lernhaltung der Schülerinnen und Schüler.

Das Schulklima ist ein wichtiger Aspekt der Schulqualität. Eine nichtbedrohliche Umwelt ist die Voraussetzung für kreatives Lernen. Die Förderung bestimmter Dimensionen des Klassenklimas wie Kohäsion, Zielorientierung und Demokratie hat einen konsistent positiven Einfluss auf das Lernen der Schülerinnen und Schüler. Allerdings ist der Import des Klimabegriffs aus der Meteorologie in die Pädagogik nicht ganz unproblematisch. Analogien zu naturwissenschaftlichen Phänomenen sind häufig nur Annäherungen; so auch der Begriff des Klassen- und Schulklimas. Gemeinsam sind ihnen die rhythmischen Schwankungen (je nach Tageszeit bzw. Zeitpunkt im Schuljahr), die multifaktorielle Bedingtheit und die Messbarkeit mit Hilfe von Instrumenten. Es gibt aber auch Unterschiede. Der Begriff in der Meteorologie impliziert ein langandauerndes Phänomen, dass mittels Einmalmessungen valide bestimmt werden kann. Das soziale Klima einer Schule ist variabler und keineswegs zeitlich stabil. Die Messung selbst hat Einfluss auf das Schulklima, während das meteorologische Klima nicht durch die Messung beeinflussbar ist. Singuläre Ereignisse können das Schul- oder Klassenklima nachhaltig beeinflussen. Wiederholte Klimamessungen an Schulen bringen bei einem Teil der Untersuchungspopulation große Unterschiede zu Tage, andere Befragte schätzen das Schulklima konstant ein. Klimaschwankungen werden also häufig nur von einem Teil der Befragten wahrgenommen (v. Saldern, 1987).

Der Hauptvorwurf gegen die Schulklimaforschung wird in der mangelnden theoretischen Fundierung und in ihrer geringen Verzahnung zu anderen Aspekten der Schulqualitätsforschung gesehen. „In ihrer theoretischen Verankerung und hinsichtlich der von ihr verwendeten Analysemethoden ist die Schulklimaforschung hinter den entwickelten Standards der Unterrichtsforschung und der Evaluationsforschung so stark zurückgeblieben, dass die Forderung laut geworden ist, auf den Begriff des Klimas im Kontext des Unterrichts entweder ganz zu verzichten (Gruehn, 2000) oder sie theoretisch wesentlich enger mit der Lehr-Lernforschung zu verzahnen...“ (Helmke, 2003, S. 33). Ein weiterer Kritikpunkt dieser Forschungen ist die gemittelte Erfassung der Schüler-Lehrer-Beziehung und des Unterrichts. In der Regel beziehen sich die Fragen nicht auf einzelne Lehrerinnen und Lehrer, sondern auf die gesamte Lehrerschaft oder alle Fächer. Auch der Versuch der Quantifizierung des Urteils (alle, viele, wenige, gar keine Lehrerinnen und Lehrer) ist schwer zu interpretieren (Ditton, 2002b). Die Schulklimaunterschiede innerhalb einzelner Klassen einer Schule können beträchtlich sein. Der Mittelwert einer ganzen Schule ist deshalb nur eingeschränkt geeignet, das Klima einer ganzen Schule zu beschreiben (Helmke, 2002).

#### **2.4.2 Die Anfänge der Schulklimaforschung in den USA und Deutschland**

Die Erforschung der Lernumgebung als ein Kontextfaktor kann bis zu Murray (1938) zurückverfolgt werden, dessen frühe Klassenstudien sich auf Schülerwahrnehmungen der Klassenraumbedingungen bezogen. In den 1960er Jahren entwickelten in den USA mehrere Forschergruppen Fragebögen, um die Lernumwelt von Schülern und Schülerinnen zu erfassen. Die Untersuchungen bezogen sich auf das Klima in einzelnen Klassen. Am besten konnte dies aus Sicht der Forscher geschehen, indem man die Lernenden selbst befragte. Im Mittelpunkt standen deshalb die Schülerinnen und Schüler und ihre Wahrnehmung des Verhaltens von Mitschülerinnen und Mitschülern sowie Lehrerinnen und Lehrern. Das Anliegen der Klimaforschung war neben der Beschreibung des Klassenklimas auch, dessen Auswirkungen auf die Schulleistungen und die Schulmotivation festzustellen. Zwei US-Forschergruppen beeinflussten die Schulklimaforschung maßgeblich: auf der einen Seite eine Gruppe um Walberg und Anderson, auf der anderen Moos und Kollegen. Beide Studien wurden international sehr stark rezipiert, die entwickelte Forschungsmethodik häufig

angewendet bzw. auf die Gegebenheiten in dem betreffenden Land spezifiziert. Sie werden deshalb nachfolgend kurz vorgestellt.

#### *Learning Environment Inventory (LEI)*

Im Rahmen des Harvard Physics Project wurde von Walberg, Anderson und anderen die Einführung eines neuen Lehrplans in Physik untersucht. Hierfür wurde das soziale Klima als ein Bewertungskriterium herangezogen. Soziales Klima wurde definiert als Ergebnis des Zusammenspiels von institutionell beeinflussten Rollen der Schüler und ihren Persönlichkeiten im Rahmen des Klassenverbandes. Um dies zu messen, wurde das „Learning Environment Inventory“ (LEI) entwickelt (Walberg & Anderson, 1968). Das LEI wurde für die Sekundarstufe konzipiert. Mit ihm ist es möglich, sowohl die Einschätzung eines einzelnen Schülers als auch die Lernumwelt einer ganzen Klasse zu ermitteln. Die Skalen des Fragebogens sind auf die Beziehung der Schülerinnen und Schüler untereinander, das Schüler-Lehrer-Verhältnis, die Einstellung der Schüler zum Fach und die Lernmethodik ausgerichtet. Das Instrument umfasst 15 Skalen, die aus jeweils 7 Items bestehen. Vier Antwortmöglichkeiten dienen der Abstufung des Schülerurteils. So gehört zur Skala Klassenkohäsion z.B. die Aussage: „Die Schüler einer Klasse sind miteinander befreundet“. Das Klima einer Klasse ergibt sich aus den durchschnittlichen Angaben der Schülerinnen und Schüler einer Klasse. Inzwischen gibt es von diesem Instrument mehrere Fassungen. So misst das „Learning Environments Inventory“ von 1982 das Klassenklima mit 48 Items. Diese Klassenklimamessungen waren gute Prädiktoren des Lernerfolgs eines Schülers bzw. einer Schülerin (Anderson & Walberg, 1974). Für die Primarstufe wurde mit dem „My Class Inventory“ (MCI) ein eigenes Instrument entwickelt (Anderson, 1973). Dieses enthielt fünf Klimaskalen (Zufriedenheit, Konflikte („Friction“), Wettbewerbsfähigkeit, Schwierigkeit des Unterrichts bzw. Unterrichtsstoffes und Kohäsion).

#### *Classroom Environment Scale (CES)*

Ein Team unter der Leitung von R. H. Moos in Stanford entwickelten eine Reihe von Forschungsinstrumenten zur Erfassung der sozialen Umwelt bzw. des Klimas von Institutionen. Konzeptueller Ausgangspunkt war eine klare theoretische Trennung von Person und Umweltmerkmalen. Diesem Modell zufolge werden aber beide Seiten im Organismus durch kognitive Prozesse miteinander verbunden (Moos, 1979). Die Untersuchungen von Moos bezogen sich jedoch nicht ausschließlich auf Klassenräume; auch Gefängnisse, Kasernen, Studentenwohnheime und Arbeitsmilieus wurden untersucht. 1974 wurde ein Test publiziert, der das gesamte psycho-soziale Unterrichtsklima einer Schule erfasste und so das Lernumfeld eines Schülers bzw. einer Schülerin in den Blickpunkt der Betrachtung rückte. Berücksichtigt wurden dabei die Beziehungen innerhalb der Schülerschaft, das Schüler-Lehrer-Verhältnis und die organisatorische Struktur (Moos, 1979).

Die Entwicklung der CES erfolgte im Verlauf eines längeren Prozesses. In der Endfassung wurden zur Charakterisierung variierender Lernumgebungen drei allgemeine Kategorien verwendet, um gezielt auf unterschiedliche Kontextbedingungen eingehen zu können. Die drei erfassten Dimensionen sind:

1. die Beziehungsdimension (d.h. die Art und Intensität der persönlichen Beziehungen innerhalb der Umwelt und das Ausmaß, in dem Menschen innerhalb dieser Umwelt engagiert sind und sich gegenseitig unterstützen),
2. die individuelle Entwicklung und Zielorientierung (Konzentriertheit im Unterricht und Konkurrenzverhalten innerhalb der Klasse) und
3. Systemerhaltung und -veränderung, etwa erfasst durch Klasse-Umwelt-Skalen (Verhaltensregeln, Lehrerkontrolle und Aktivitäten).

Ziel dieser Studien war die Beschreibung des Klassenklimas („classroom environment“). Eine klare begriffliche Fassung des Klimabegriffs fehlt jedoch bei Moos. Er geht davon aus, dass den sozialen Umwelten ebenso Persönlichkeiten zu Grunde liegen wie Menschen (Moos & Tricket, 1974). Es wurde ein Fragebogen entwickelt, der verschiedene Dimensionen erfasst. Es wurden sowohl die Lehrer- als auch die Schülerperspektive berücksichtigt und einander gegenübergestellt. Den Schülerinnen und Schülern wurden zwei Versionen des Fragebogens vorgelegt, die dichotom beantwortet werden sollten. Die Besonderheit der Untersuchung lag darin, dass sowohl der Soll- als auch der Ist-Zustand erfasst wurde. An einem Beispiel soll diese Vorgehensweise kurz veranschaulicht werden:

Real form:

- |   |            |
|---|------------|
| A. Students put a lot of energy into what they do here.       | True/False |
| B. Students in this class get to know each other really well. | True/False |

Ideal and expectations form:

- |  |            |
|--|------------|
| A. Students will put a lot of energy into what they do.            | True/False |
| B. Students in this class will get to know each other really well. | True/False |

Beide Studien, das LEI wie auch die CES, beeinflussten die Schulklimaforschung maßgeblich. Der Hauptvorwurf gegen die beiden Instrumente wird in der empiristischen Herangehensweise gesehen (v. Saldern, 1987; Bessoth, 1989), denn sie stützen sich nicht auf eine konsistente Theorie schulischen Lernens. Auch die empirische Prüfung der Skalen zeigt methodische Mängel. So fand Kahl (1977) nur drei verschiedene Dimensionen und nicht 15, wie von Walberg behauptet. Und auch wenn beide Autorengruppen in den Schülerinnen und Schülern die besten Beurteiler des Klimas sehen, ist diese Perspektive aus heutiger Sicht jedoch einseitig.

#### *Der Lernsituationstest (LST)*

Die ersten deutschsprachigen Schulklimatests wurden in den 1970er Jahren entwickelt. Im Mittelpunkt dieser Tests stand das Unterrichtsklima. Das Anliegen des Lernsituationstests von Kahl war, das unterrichtliche Geschehen und das Klassenklima aus Sicht der Schülerinnen und Schüler zu erfassen. Dazu wurde ein Fragebogen entworfen, der Hamburger Sechstklässlern vorgelegt wurde (N=545, verschiedene Schulformen). Vier Dimensionen wurden dabei faktoranalytisch ermittelt: Kohäsion, Identifikation mit der Unterrichtsarbeit, Leistungsanforderungen im Unterricht, Betonung von Wettbewerb und Ordnung (Kahl, 1977, S. 252f). Kahl hat das Lernklima nicht explizit definiert, betont jedoch die Abhängigkeit des Lernklimas vom jeweiligen Fach, der Unterrichtsmethode und der Erfahrung des Lehrers oder der Lehrerin, der Lehrerpersönlichkeit und einer Vielzahl anderer Faktoren. Die ausschließliche Fokussierung auf die Schülersicht muss aufgrund ihrer Einseitigkeit kritisch hinterfragt werden. Da der Schwerpunkt der Untersuchung auf dem Unterricht lag, wäre die Erhebung der Lehrerperspektive von essentieller Bedeutung. Standop (2002) kritisiert, dass das Lernklima einer Klasse nicht umfassend abgebildet wurde und dass die Konstrukte nicht verhaltensnah operationalisiert wurden. Auch methodische Mängel sind zu nennen. So ist die Gesamtvarianzaufklärung mit 24% als eher mäßig einzustufen. Die Items sind sehr heterogen, schwer zu interpretieren (Dreesmann, 1979, S. 58f) und unvollständig (v. Saldern, 1987, S. 174).

### *Fragebogen zum Unterrichtsklima (FUK)*

Der „Fragebogen zum Unterrichtsklima“ wurde von Dreesmann Mitte der 1970er Jahre zur Erfassung von Schülerwahrnehmungen des psycho-sozialen Unterrichtsklimas entwickelt (Dreesmann, 1982). Die Untersuchung selbst fand 1977 statt. Grundlagen bildeten eine Reihe von internationalen Studien, insbesondere LEI, CES und der hier nicht dargestellte Classroom Activities Questionnaire. Ausgangspunkt war eine Untersuchung im Mathematikunterricht der 8. Klasse an Hauptschulen in Rheinland-Pfalz. Neben dem Fragebogen zum Unterrichtsklima kamen auch ein Intelligenztest, ein Rechentest sowie ein Fragebogen zur Attribuierungstendenz und zum Selbstkonzept zum Einsatz. Das komplexe Untersuchungsdesign erforderte mehrere Untersuchungstermine. Der Klimafragebogen wurde zusammen mit dem Intelligenztest durchgeführt, um eine unterrichtstypische Belastungssituation herzustellen (Dreesmann, 1979, S. 132). Das Unterrichtsklima wird dabei verstanden als eine relativ überdauernde Qualität der Umwelt des Unterrichtsgeschehens. Es bezieht sich auf bestimmte Merkmale, die Schülerinnen und Schüler erleben und die ihr Verhalten beeinflussen (Dreesmann, 1982, S. 44f). Klima entsteht, wenn die individuellen Beurteilungen in einer Gruppe ein Minimum an Gemeinsamkeit aufweisen. Dieser gemeinsame Erlebnisanteil der Schülerinnen und Schüler kann erfasst werden. Der Autor bildete mehrstufige Skalen. Die abschließende Fassung enthält 33 Items, die sieben Skalen zugeordnet werden können. Wichtige Dimensionen sind die Kooperation zwischen Schülerschaft und Lehrerschaft im Unterricht, die Kritikäußerungen an Lehrerinnen und Lehrer, die Lenkung durch Lehrpersonal sowie die Förderung von Selbstvertrauen. Deutlich wird die starke Betonung kognitiver Aspekte, insbesondere in Leistungssituationen, obwohl sich nach Dreesmann das Unterrichtsklima aus den Bereichen soziales, affektives, intellektuelles und organisatorisches Klima zusammensetzt. Kritisch anzumerken ist, dass keine Zusammenhänge zwischen den Konstrukten berücksichtigt wurden. Es treten erhebliche Validitäts- und Reliabilitätsprobleme auf (v. Saldern, 1987). Die Durchführung der Tests durch die Lehrer und Lehrerinnen führen zu einer Reihe von methodischen Problemen. So ist nicht auszuschließen, dass Hilfestellungen bei Rechentests gegeben wurden und Schülerantworten im Sinne der sozialen Erwünschtheit verzerrt wurden.

### **2.4.3 Wichtige deutsche Schulklimainstrumente**

#### *Landauer Skalen zum Sozialklima (LASSO)*

Die Landauer Skalen zum Sozialklima wurden in den 80er Jahren durch von Saldern, Littig und Ingenkamp im Rahmen des DFG-Projekts „Schulische Umwelt und Verhalten von Schülern“ entwickelt (v. Saldern & Littig, 1987). Ziel war es, das Sozialklima in Schulen zu beschreiben und zu erfassen. Wie nehmen Schülerinnen und Schüler ihre Lernumwelt wahr? Von Saldern verwendet den Begriff des Sozialklimas in Abgrenzung zum Schul- und Klassenklima. Das Sozialklima wird definiert als ein „auf die Schulklasse bezogenes, differenzierendes, relativ überdauerndes, molares und mehrdimensionales Aggregat subjektiver Wahrnehmung und kognitiver Verarbeitung von situationalen Reizen, das sich in der Beschreibung von Umwelten, Strukturen und Verhalten in der Schulklasse bzw. in einem ihrer Subsysteme (z.B. Cliquen) durch das Individuum widerspiegelt und die Bildung von Einstellungen zur Lernsituation sowie individuellen Einstellungen zur Lernsituation sowie individuellem Verhalten beeinflusst.“ (Ingenkamp & v. Saldern, 1983, S. 17)

Mit Hilfe eines Fragebogens sollten die Einstellungen der Schülerinnen und Schüler zu ihrer Lernumwelt erfasst werden. Die Autoren ordnen der Lernumwelt drei Dimensionen zu: die Lehrer-Schüler-Beziehung, die Schüler-Schüler-Beziehung und allgemeine Merkmale des Unterrichts. LASSO wurde in drei verschiedenen Versionen erstellt (v. Saldern, 1987): (a) als reales Klima, d.h.: Wie gestaltet sich das Sozialklima einer Schulklasse? (b) als ideales Klima



oder: Wie sollte sich das Klima gestalten? und (c) als Relevanzaussage (Wie wichtig sind die verschiedenen Klimadimensionen für einzelne Schüler?). Methodisch umgesetzt wurde dieses Anliegen durch die Sammlung von 380 Items vorangegangener Studien, die 25 Subskalen zugeordnet wurden. Das Sozialklima einer Klasse ergibt sich auch hier aus den Durchschnittswerten aller Schülerinnen und Schüler einer Klasse. Ursprünglich bezogen sich alle Aussagen nur auf den Klassenlehrer bzw. die Klassenlehrerin. Das schränkte die Einsatzmöglichkeiten des Instruments erheblich ein. In einer überarbeiteten Fassung beziehen sich die Fragen auf die Lehrerschaft allgemein.

LASSO ist als Schultest konzipiert, um die subjektive Meinung eines Klassenmitglieds und das Sozialklima einer ganzen Klasse zu erfassen sowie Vergleiche zwischen einzelnen Klassen zu ermöglichen (v. Saldern & Littig, 1987, S. 4). Der Test erscheint theoretisch und empirisch fundiert. Dennoch wird eine Reihe methodischer Fragen aufgeworfen: Der Fragebogen ist sehr komplex. Er enthält in der Realform immerhin 142 Items. Dies mag in der Sekundarstufe II unproblematisch sein, in der 4. Jahrgangsstufe ist die Testdurchführung jedoch mit einer hohen Belastung verbunden. Aussagen wie „Schüler wie wir können bei unserem Lehrer kaum erfolgreich sein“ oder „Die Schüler finden Wettbewerb notwendig um festzustellen, was der einzelne kann“ produzieren in dieser Altersgruppe wahrscheinlich Zufallsantworten. Manche Aussagen sind nur schwer als positiv oder negativ zu bewerten. Beispielsweise widerspricht die Aussage „Jeder in der Klasse wird gleich behandelt“ den Anforderungen an binnendifferenzierten Unterricht. Viele der Aussagen beziehen sich auf konkretes Lehrerverhalten. Die Durchführung durch Lehrkräfte der eigenen Schule verstärkt die Tendenz, sozial erwünschte Antworten zu geben. Auch die Auswertungen von einzelnen Schülerfragebögen durch an der Schule unterrichtende Lehrerinnen oder Lehrer erscheint aus diesem Grund problematisch.

#### *Der Linzer Fragebogen zum Schul- und Klassenklima (LFSK)*

Der LFSK (Eder, 1998; Eder & Mayr 2000) liegt in zwei Fassungen vor. Einmal für die Klassen 4 bis 8 (LFSK 4-8) und zum anderen für die Klassen 8-13 (LFSK 8-13). Beide dienen der Diagnose subjektiver Lernumwelten. Eder entwickelte ein eigenes Klimainstrument, weil aus seiner Sicht in der bisherigen deutschsprachigen Klimaforschung zu wenige Dimensionen berücksichtigt wurden und nicht die Gesamtheit der Lehrerschaft einer Schule erfasst werden konnte (Eder, 1996). Sein Ziel war es, die schulische Umwelt aus Schülersicht zu beschreiben. Bezugsrahmen sollte die Schulklasse sein. Wie schon bei den vorausgegangenen Instrumenten zur Schulklimamessung ist auch mit diesem Test die Diagnose der subjektiven Lernumwelt für die Individualberatung möglich. Gleichzeitig ist er aber auch ein Gruppenverfahren, sodass der Fragebogen als Evaluationsinstrument und im Rahmen von Forschungsprojekten zum Einsatz kommen kann. LFSK ermöglicht Vergleiche des sozialen Klimas zwischen einzelnen Klassen und Schulen.

Der Fragebogen für ältere Schülerinnen und Schüler umfasst rund 100 Items und baut auf den vorausgehenden Instrumenten der Schulklimaforschung auf. Dabei werden 14 Schulklimaelemente erfasst:

- Pädagogisches Engagement
- Restriktivität
- Mitsprache
- Gerechtigkeit
- Komparation
- Gemeinschaft
- Rivalität
- Lernbereitschaft
- Störneigung
- Leistungsdruck
- Unterrichtsdruck
- Vermittlungsqualität
- Schülerbeteiligung
- Kontrolle der Schülerarbeit

Durch faktoranalytische Berechnungen lassen sich vier Dimensionen bilden. Diese Dimensionen sind: „Sozial- und Leistungsdruck“, „Schülerzentriertheit“, „Kohäsion“ und „Disziplin“. Auf Schulebene werden zwei Klimabereiche erfasst: „Strenge“ und „Wärme“. Eder beschäftigt sich sehr ausführlich mit den Auswirkungen und Determinanten des Schulklimas (z.B.: Eder, 1996). Dazu nutzte er Längsschnittanalysen. Im Ergebnis lässt sich festhalten, dass die Auswirkungen des Schulklimas auf die Schulleistung relativ schwach sind (Eder, 1996). Hingegen ist ein positives Schulklima günstig für die Schulmotivation, die Schulzufriedenheit und die Mitarbeit im Unterricht. Es verringert Störungen des Unterrichts und aggressives Verhalten (Eder, 2002).

Der Nachteil der bisher vorgestellten Schulklimainstrumente ist, dass nur eine Perspektive erfasst wird – nämlich die der Schülerinnen und Schüler. Für das Schulklima ist aber die Lehrerperspektive ebenfalls von zentraler Bedeutung. Am Beispiel der IFS-Schulbarometer, dessen Herkunft der Titel verrät, wird nun gezeigt, dass moderne Schulklimaforschung durchaus dieses Defizit erkannt hat und unterschiedliche Perspektiven berücksichtigt.

#### *Das Schulbarometer des Instituts für Schulentwicklungsforschung Dortmund (IFS)*

Das Institut für Schulentwicklungsforschung Dortmund bietet ein Instrument an, um verschiedene Aspekte von Schulqualität mittels Befragungen zu erfassen. Der Untertitel des Schulbarometers verdeutlicht das Anliegen: „Ein mehrperspektivisches Instrument zur Erfassung von Schulwirklichkeit“ (Institut für Schulentwicklungsforschung, 1999). Es wurde aus einer alle zwei Jahre durchgeführten Elternbefragung entwickelt, die durch Lehrer- und Schülerbefragungen ergänzt wurde. Heute liegt es in der überarbeiteten Neufassung von 1999 vor. Das Schulbarometer ist ein Instrument der Selbstevaluation der Schulen und kann ab Klassenstufe 5 angewandt werden. Drei Nutzungsmöglichkeiten werden von den Autoren angegeben: Es dient erstens als Bestandsaufnahme, d.h. der Ist-Zustand kann erhoben werden. Zweitens dient es der Einordnung der eigenen Schule im Sinne eines Benchmarkings. Die dritte Anwendungsmöglichkeit besteht in der Evaluation von Maßnahmen oder Programmen im Längsschnitt oder Querschnitt.

Das IFS-Schulbarometer ist mehrperspektivisch angelegt, d.h. mit ihm können Eltern, Schülerinnen und Schüler sowie die Lehrerschaft einer oder mehrerer Schulen befragt werden. Des Weiteren ist es modular aufgebaut und kann von den Schulen verändert werden. Zur Einordnung der eigenen Ergebnisse dienen repräsentative Vergleichszahlen von 1995 bzw. 1999, die nach Schulformen untergliedert sind. Die Befragung erfolgt schriftlich und überwiegend standardisiert. In der Schülerbefragung schätzen die Jugendlichen die Schule zunächst allgemein ein. Diese Einschätzung umfasst die verschiedenen Aspekte des Schullebens. Auch eigene Gewalterfahrungen werden thematisiert. Ein zweiter Schwerpunkt ist das Schüler-Lehrer-Verhältnis. Hierbei ist hervorzuheben, dass sich einerseits die Antwortvorgaben auf die Anzahl der Lehrerinnen und Lehrer beziehen, auf die ein Merkmal zutrifft. Andererseits wird die Bedeutung des Merkmals für den Schüler bzw. die Schülerin erfasst. Es folgen Fragen zur Unterrichtsmethodik, Leistungsanforderungen und der Schülerpartizipation. Aus den Einzelaussagen können Skalen gebildet werden. Für die Schülerdaten werden zwei Skalen angegeben: Die Umweltqualität der eigenen Schule und die Schul- und Leistungsangst.

Insgesamt kann festgehalten werden, dass das Schulbarometer ein Instrument der Selbstevaluation ist. Eine klare Definition, was eine gute Schule ist, wird hier ebenso wenig gegeben wie Hinweise auf mögliche Interventionen. Welche Merkmale eine gute Schule auszeichnet leitet sich aber aus den erfassten Themen ab. Die Stärke des Instrumentariums ist die Vielfalt der Perspektiven, die bausteinartig benutzt und modifiziert werden können. Gleichzeitig offenbart sich hier auch das Problem dieser Vorgehensweise: Es ist sehr komplex und stellt die Schulen vor hohe Anforderungen. Wünschenswert wäre eine Vergrößerung der

Vergleichsstichprobe (N=2200, darunter: 268 Gesamtschulschülerinnen und Schüler) und die Erweiterung auf die Sekundarstufe II. Auf diese Weise würde man auch altersdifferenzierte Vergleichswerte erhalten.

## 2.5 Schuleffektivitätsforschung

### 2.5.1 Überblick

Neben der Schulklimaforschung entwickelte sich in den USA die Schuleffektivitätsforschung (für einen Überblick Reynolds & Teddlie, 2000). In den 1960er Jahren war in der Bildungsforschung der USA und Großbritanniens die Meinung weit verbreitet, dass die Schulmerkmale kaum einen Effekt auf die Leistungsentwicklung und das Verhalten der Schülerinnen und Schüler haben. Besonders einflussreich waren zwei große amerikanische Studien: „Equality of Educational Opportunity“ unter der Leitung von James S. Coleman (1966) und „Inequality: A Reassessment of the Effect of Family and Schooling in America“ unter der Leitung von Christopher Jencks (1972). Der von der britischen Regierung in Auftrag gegebene Plowden Report (Central Advisory Council for Education, 1967) kam zu ähnlichen Schlussfolgerungen. Statt der Schule wurden familiäre Herkunft, die Begabung und die kognitiven Eingangsvoraussetzungen der Schülerinnen und Schüler oder auch allgemeine gesellschaftliche Faktoren als Verursacher von Ungleichheit identifiziert. Noch 1975 formulierten Good, Biddle & Brophy: „Machen Schulen oder Lehrer eigentlich einen Unterschied aus? Es gibt keine sichere Antwort.“ (1975, S. 3)

Diesem Ergebnis standen zunehmend Befunde der Schulleistungsforschung entgegen, die in den angelsächsischen Ländern sehr viel intensiver entwickelt war als in Deutschland. Sie wies nach, dass sich die erreichten Schulleistungen von Schule zu Schule erheblich unterschieden. Um diese Unterschiede in der Lernwirksamkeit zu erklären, entwickelte sich die Schuleffektivitätsforschung. Im Laufe der Zeit wurden nicht nur Schulleistungen miteinander verglichen, sondern auch innerschulische Bedingungen sowie die Einschätzungen von Lehrerschaft, Eltern und Schülerschaft. Dabei wurden immer anspruchsvollere Kriterien guter Schulen herausgearbeitet (Wenzel, 2004). Hinzu kamen Studien wie die von Rutter et al. (1979), die den Anstoß zu einer Reihe weiterer Studien zur Schuleffektivität gab (für einen Überblick siehe Aurin, 1990; Steffens & Bargel, 1993).

In den vergangenen Jahren wurden in vielen Ländern der Welt Qualitätskriterien für gute Schulen entwickelt. In Deutschland hat insbesondere die Gruppe um Fend Qualitätsmerkmale guter Schulen herausgearbeitet. In diesen Studien wird nicht nur die Unterrichtsqualität betrachtet, sondern auch die Qualität der Lehrerbildung, schulstrukturelle Aspekte (z.B. Klassengröße), die Ausstattung der Schule und das Klassen- und Schulklima (Schnabel, 2001). Ihren Höhepunkt hatte diese Forschungstradition in den 1980er Jahren (Reynolds & Teddlie, 2000). Schuleffizienz ist nicht direkt beobachtbar. Er handelt sich um ein Konstrukt bei dem einzelne messbare Merkmale miteinander verknüpft werden (Arnold, 2002). In ihrem methodischen Design ähneln sich die meisten dieser Studien. Demnach zeichnen sich Schüler und Schülerinnen guter Schulen dadurch aus, dass sie in einem Leistungstest besonders gut abschneiden. Im Endeffekt ergibt sich ein sehr komplexes Bild mit vielen, zum Teil widersprüchlichen Details. Es bleibt offen, wie stark der Effekt der Einzelfaktoren jeweils ist und wie die Faktoren zusammenhängen. Schuleffektivität wurde in den englischsprachigen Ländern überwiegend im Hinblick auf den Abbau sozialer Ungleichheit verwendet. Schulen sollten allen Kindern und Jugendlichen gleiche Bildungschancen bieten. Schon die Studie von Coleman et al. (1966) bezieht sich auf die Integration von schwarzen und weißen Kindern. Chancengerechtigkeit ist bis heute ein entscheidendes Ziel der amerikanischen Bildungspolitik geblieben (National Center for Education Statistics, 2000).

Komplexe, empirisch fundierte Modelle, die die Schuleffizienz erklären, sind in dieser Forschungstradition die Ausnahme. Kritisch gesehen werden muss, dass die Beherrschung von „basic skills“ als alleiniger Qualitätsmaßstab angelegt wurde, während Lernziele höherer Ordnung vernachlässigt wurden. Überfachliche Kompetenzen wie zum Beispiel Kritikfähigkeit, autonomes Selbstbild, selbständiges Lernen oder Sozialverhalten werden ebenso wenig erfasst wie sozialstrukturelle Aspekte. Weiterhin bleibt offen, ob die Qualitätskriterien auf alle Alters- und Sozialgruppen von Schülerinnen und Schüler die gleiche Wirkung haben (Holtappels, 2003). Der Fokus auf der Schule als Ganzes verdeckt zudem die innerschulische Streuung des Leistungsvermögens (Steffens & Bargel, 1993).

Im nächsten Abschnitt werden sowohl Beispielstudien aus dem angelsächsischen Raum vorgestellt als auch die Arbeiten von Fend und INIS.

## **2.5.2 Wichtige englischsprachige Studien**

### *Fünftehtausend Stunden*

1970 führten Rutter und Kollegen eine Vergleichsstudie zur Entwicklung von Schulkindern durch (Rutter, Maughan, Mortimore & Ouston, 1979). Die Autoren wählten den Titel „Fünftehtausend Stunden“, weil Schulkinder Schätzungen zufolge von der Einschulung bis zum Schulabschluss ungefähr 15.000 Stunden in der Schule verbringen. In einer Längsschnittuntersuchung wiesen sie nach, dass sich trotz gleicher Eingangsvoraussetzungen zum Schulbeginn am Ende der Schulzeit erhebliche Unterschiede in der Schulleistung, bei devianten Verhaltensweisen und bei der Mitwirkung am Schulleben ergeben hatten. Diese Studie soll nun ausführlicher vorgestellt werden.

Vor dem Hintergrund der eingangs beschriebenen Skepsis zur Bedeutung der Einzelschule führten Rutter et al. 1970 eine Vergleichsstudie von zehnjährigen Schülern und Schülerinnen an Grundschulen in London und auf der Isle of Wight durch. An dieser Untersuchung nahmen alle Zehnjährigen eines Stadtbezirks im Süden Londons teil. Sie wurden Leistungstests unterzogen, und ihr Verhalten wurde von ihren Lehrerinnen und Lehrern in einem Fragebogen beurteilt. Zusätzlich wurden Informationen zum beruflichen Hintergrund der Eltern und zu familiären Problemen im Elternhaus gesammelt. Im Ergebnis dieser Studie zeigte sich, dass emotionale Probleme, Verhaltensauffälligkeiten und Leseschwierigkeiten in London doppelt so häufig auftraten wie auf der Isle of Wight. Die Studie zeigte zudem, dass diese Auffälligkeiten in starkem Maße mit verschiedenen Formen familiärer Probleme im Elternhaus der betroffenen Kinder in Zusammenhang standen. Obwohl familiäre Probleme die stärkste Erklärungskraft hatten, gab es jedoch Unterschiede, die nicht durch sie erklärt werden konnten. Die Autoren schlussfolgerten, dass ein signifikanter Einfluss der Schulen selbst die wahrscheinlichste Erklärung sei. Dies wurde zum Gegenstand der nächsten Befragung.

Ungefähr zwei Drittel der befragten Schülerinnen und Schüler gehörten zur Grundgesamtheit einer weiteren Studie im Jahre 1974. Diese Untersuchung wurde an den 20 nicht-selektiven Sekundarschulen im Süden Londons durchgeführt, an denen die Mehrzahl der Befragungsteilnehmer der ersten Studie nun zur Schule ging. Das Hauptziel der Studie lag darin, Unterschiede zwischen diesen Schulen zu identifizieren. Die Schülerinnen und Schüler wurden im Alter von 14 Jahren bzw. im dritten Jahr der Sekundarstufe erneut befragt. Es kamen andere, altersgerechte Leistungstests zum Einsatz. Die Lehrpersonen füllten erneut den gleichen Fragebogen zum Schülerverhalten aus. Wieder wurden umfassende Daten zum ethnischen Hintergrund und zum beruflichen Status der Eltern sowie Daten zur Delinquenz der Kinder gesammelt. Dieselben Tests wurden zusätzlich mit den Schülerinnen und Schülern des Jahrgangs 1971 durchgeführt, die sich jedoch nicht wesentlich von der Ausgangsstichprobe unterschieden. Es traten wiederum spürbare Unterschiede im

Schülerverhalten und in den Delinquenzraten der Kinder auf, die nicht durch die Testergebnisse am Ende der Primarstufe (1970er Studie) erklärt werden konnten. Genauso wenig kam der familiäre Hintergrund als erklärende Variable in Betracht. Es schien so, dass die Entwicklung der Kinder entscheidend durch ihre Erfahrungen an der Sekundarschule geprägt worden waren. Während einige Schulen einen positiven Einfluss auf die Schülerinnen und Schüler ausgeübt hatten, war an anderen das Gegenteil der Fall.

An diesem Punkt wurde entschieden, den Fokus von nun an auf Schulen zu richten – und nicht länger auf individuelle Schüler und Schülerinnen. Um die Funktionsweise jeder einzelnen Schule so detailliert wie möglich untersuchen zu können, wurde ihre Anzahl für die erneute Untersuchung 1975/76 auf zwölf reduziert. Rutter et al. (1979) wollten in ihrer Studie der Frage nachgehen, ob die Wahl einer bestimmten Schule, die Erfahrungen, die ein Kind an dieser Schule macht und die besonderen Merkmale dieser Schule einen Einfluss auf den Lernerfolg haben. Das Hauptziel der Untersuchung war es jedoch herauszufinden, warum es Unterschiede zwischen den Schulen in Bezug auf Schülerverhalten und Schulleistungen gab. Außerdem sollte festgestellt werden, wie Schulen die Entwicklung von Kindern beeinflussen.

Um dieses Ziel zu erreichen, mussten Maße für wenigstens vier Merkmalskombinationen entwickelt werden: 1. Maße für individuelle Eigenschaften der Schülerinnen und Schüler beim Eintritt in die Sekundarstufe (z.B.: der soziale Hintergrund, die kognitiven Fähigkeiten sowie das Verhalten); 2. Maße für den Schulprozess (die Schulorganisation, das Lernumfeld); 3. Maße für Schulergebnisse (das Erreichen von Lernzielen, die Schulleistungen) und schließlich 4. Maße für Umweltfaktoren (also all jene Faktoren, die einen Effekt auf 1.-3. hatten). Es wurden halbstrukturierte Interviews mit 219 Lehrerinnen und Lehrern durchgeführt, die durchschnittliche Interviewdauer betrug 40 Minuten. In dem Schülerfragebogen wurden folgende Themen behandelt:

- Stabilität vs. Veränderungen in der Zusammensetzung der Lehrerschaft
- Partizipation an schulischen, sportlichen und sozialen Aktivitäten
- Übernahme von Verantwortung an der Schule durch die Schülerinnen und Schüler
- Bestrafungen und Belohnungen innerhalb eines bestimmten Zeitraums
- Umfang an Hausaufgaben
- Wie oft werden Lehrerinnen und Lehrer bei Problemen um Hilfe gebeten?
- Anzahl von Schulausflügen
- Fragen zu Delinquenz: korrekte Schuluniform, Graffiti, Schulschwänzen
- Abfrage von Schulzielen, die den Schülerinnen und Schülern am wichtigsten sind vs. jene, die Lehrerinnen und Lehrern am wichtigsten sind.

Es zeigte sich, dass die Leistungen der Schulen erheblich variierten. Diese Varianz erwies sich über die Jahre als stabil. Dieser Unterschied blieb auch dann erhalten, wenn ausschließlich Kinder mit ähnlichen Merkmalen verglichen wurden. Im Ergebnis konnte also festgehalten werden, dass die Schule eine eindeutige Wirkung auf die Entwicklung eines Kindes hatte. Dementsprechend macht es einen Unterschied, welche Schule ein Kind besucht. Darüber hinaus gibt die Studie wichtige Hinweise darauf, was die Merkmale erfolgreicher Schulen sind. Entscheidend waren systematische Unterschiede in der Rolle der Schulen als soziale Institutionen. Zu den Faktoren, die die Ergebnisse statistisch signifikant beeinflussten, gehörten das Einfordern schulischer Leistung, das Lehrerverhalten im Unterricht, die Anwendung von Anreizen und Belohnungen, eine schülerzentrierte Atmosphäre, in der mehr gelobt als getadelt wird und sich die Schülerinnen und Schüler als Person akzeptiert fühlen sowie das Ausmaß, in dem Schülerinnen und Schüler Verantwortung innerhalb der Schule

und im sozialen Umfeld übernehmen konnten. Wichtig waren auch eine geringe Fluktuation im Lehrerkollegium und in der Lerngruppe sowie eine enge Zusammenarbeit in der Lehrerschaft.

Alle diese Faktoren konnten vom Lehrpersonal beeinflusst werden und wurden nicht durch äußere Umstände vorgegeben. Es gab jedoch auch Einflussfaktoren außerhalb der Kontrolle des Lehrpersonals. Dazu gehören bestimmte Jahrgangsmerkmale. Schulen mit einem stabilen Kern durchschnittlich begabter Schülerinnen und Schüler waren erfolgreicher als andere. Umgekehrt war delinquentes Verhalten an den Schulen weiter verbreitet, in denen dies nicht der Fall war. Den stärksten Effekt auf die Schulergebnisse hatte der kombinierte Index aller Schulprozess-Variablen – stärker als jede einzelne dieser Variablen. Das heißt, der kumulative Effekt dieser Faktoren war weitaus größer als die Einzeleffekte. Es konnte weiterhin festgestellt werden, dass die Schulgröße, die Klassengröße, das Alter und der Zustand der Schulgebäude, der Typ der Schulverwaltung oder der Schulorganisation keinen kausalen Einfluss auf die Variationen der Schülerleistung hatten.

#### *USA: Monitoring School Quality: An Indicators Report*

Das Thema Schuleffektivität (effective schooling) hat in der erziehungswissenschaftlichen Forschung der Vereinigten Staaten in den letzten Jahren einen hohen Stellenwert erlangt. Dementsprechend hat es in diesem Zeitraum eine beträchtliche Anzahl an Publikationen zu diesem Thema gegeben. Diese Studien zusammenfassend veröffentlichte das National Center for Education Statistics, das für die Datensammlung, Datenanalyse und datengestützte Berichterstattung zu Bildungsfragen zuständige Bundesinstitut der Vereinigten Staaten im Jahre 2000 einen umfassenden Forschungsbericht zur Schulqualität. Ziel des Berichts war es zu erklären, weshalb es manchen Schulen besser als anderen gelingt, Schüler beim Lernen zu unterstützen. Außerdem sollten auf der Grundlage einer umfassenden Bestandsaufnahme der nationalen und internationalen Forschung zur Schulqualität für alle Bundesstaaten der USA gültige, zuverlässige Indikatoren für Schulqualität identifiziert werden. Darauf aufbauend sollte der Zustand der amerikanischen Schulen auf Basis dieser umfassenden Kritik nationaler Indikatoren beschrieben werden.

Basierend auf einer umfassenden Analyse vorliegender Literatur und statistischer Daten zur Schulqualitätsforschung wurden 13 Indikatoren für Schulqualität identifiziert, die wiederum einem von drei Bereichen zuzuordnen sind (vgl. Abb. 2.1): (1) Lehrer, (2) Klassenraum und (3) allgemeine Kontextbedingungen der Schule (National Center for Education Statistics, 2000). Hintergrund ist die Feststellung, dass die Schulqualität entscheidend durch Ausbildung und Fähigkeit der Lehrkräfte, das Unterrichtsgeschehen und die allgemeine Schulkultur und -atmosphäre beeinflusst werden. Schulqualitätsfaktoren können direkt oder indirekt auf das schulische Lernen wirken. Hohe Schulqualität wird durch folgende Faktoren gefördert:

1. Einen wesentlichen Einfluss auf die Schulqualität hat demnach eine gute Ausbildung der Lehrkräfte. Wichtig ist zudem, dass Lehrerinnen und Lehrer in den Fachgebieten eingesetzt werden, für die sie ausgebildet wurden. Qualitätsfördernd machen sich außerdem mehrjährige Lehrerfahrung und die Teilnahme an qualitativ hochwertigen Weiterbildungsmaßnahmen bemerkbar.
2. Der Inhalt des Lehrplans, die Pädagogik sowie die im Unterricht verwendeten Materialien wirken entscheidend auf die Unterrichtsqualität ein. Jüngere Kinder und Angehörige ethnischer Minderheiten lernen in kleinen Klassen besser.
3. Schulische Kontextfaktoren, die man durch folgende Fragen illustrieren kann: Auf welche Art und Weise werden bildungsbezogene Führerschaft und die Umsetzung von Schulzielen angegangen, wie die professionelle Gemeinschaft ausgeprägt und ein Schulklima entwickelt wird, dass disziplinäre Probleme minimiert und herausragende Schulleistungen fördert, haben einen klaren Effekt auf die Schulqualität und

Schulleistungen. Diese Wirkungen sind jedoch weitaus schwieriger zu messen als die zuvor genannten.

**Abbildung 1 Schulqualitätsmerkmale aus den USA**



Quelle: National Center for Education Statistics, 2000, Figure ES.1

Zusammenfassend betonen die Autoren der Studie die Notwendigkeit der regelmäßigen Überwachung von Schulqualität. Diese Studie ist eine Metastudie, sie fasst den Stand der Forschung für die Vereinigten Staaten zusammen. Aus ihr leiten sich kaum konkrete Indikatoren ab. Einen Schritt weiter geht beispielsweise das Internationale Netzwerk Innovativer Schulen (INIS). Hier werden neben der Dimensionsbestimmung von Schulqualität auch Messinstrumente entwickelt, um die Schulqualität auch zu erfassen.

### 2.5.3 Deutsche Schuleffektivitätsstudien

In Deutschland fasste der Arbeitskreis „Qualität von Schule“ um Fend (deshalb wird die Schuleffektivitätsforschung auch „Schulqualitätsforschung“ genannt) die Ergebnisse der Schuleffektivitätsforschung zusammen und führte eigene Untersuchungen durch. Lange Zeit wurde in Westdeutschland die Frage diskutiert, was die bessere Schulform ist (für einen Überblick s. v. Friedenborg, 1989). Die Verfechter der seit Beginn der 1970er Jahre in den sozialdemokratisch regierten Bundesländern eingeführten Gesamtschulen argumentierten, dass an Gesamtschulen die Chancengleichheit besser gewährleistet sei. Durch das Nebeneinanderbestehen von Gesamtschulen einerseits und Haupt-, Real-, und Gymnasialschulen andererseits entstand die Notwendigkeit der Evaluation der pädagogischen und sozialen Vor- und Nachteile der jeweiligen Schulform (Horstkemper & Tillmann, 2004). Besonders Fend und Kollegen untersuchten in den 1970er Jahren in großen Stichprobenuntersuchungen die Wirkungen der beiden Schulsysteme (u.a. Fend, 1977; 1982). Er kam zu dem Schluss, dass sich beim Vergleich der Schulsysteme eher geringe Unterschiede ergeben und stellte fest, dass die Unterschiede zwischen den Schulen einer Schulform erheblich sind. Deshalb leitete er die Schlussfolgerung ab, die einzelne Schule als „pädagogische Handlungseinheit“ (Fend, 1986) zu begreifen.

Fend hat in seiner 1998 erschienen Studie „Qualität im Bildungswesen“ die Ergebnisse seiner bisherigen Arbeit zusammengefasst und teilweise neu bewertet. Schon in seinem 1977er Buch über das Schulklima - Fend steht auch für die Verbindung von Schulklima- und Schuleffektivitätsforschung - richtete er die Aufmerksamkeit auf das Innenleben der Schule. Er nutzte den Begriff des Schulklimas, um Prozessmerkmale von Bildungssystemen zu beschreiben, die auf kulturelle Gestaltungskräfte zurückgehen. Die Prozessmerkmale, die aus der gemeinsamen Deutung der Lehrer- und Schülerschaft hervorgehen, veranlassten Fend, die

Wahrnehmung aller Betroffenen zu erfassen. Er erhob die Wahrnehmung von Eltern, Schülerinnen und Schülern sowie Lehrerinnen und Lehrern. An den Untersuchungen in den 1970er und 1980er Jahren waren 183 Schulen aus mehreren Bundesländern beteiligt (Fend, 1998). Fend untersuchte in diesen Schulen eine Vielzahl von Merkmalen. Zu ihnen gehörten die Architektur und Ausstattung, die organisatorische Bewältigung des Schulbetriebes und die sozialen Belastungen von Schülerschaft, Lehrpersonal und Eltern. Ein dritter Merkmalskomplex bildet die Art der Entscheidungsfindung, wobei die Partizipationsspielräume der Schüler- und Lehrerschaft berücksichtigt wurden. Sehr umfassend wurde das Schulklima erhoben. Dabei wurden verschiedene Aspekte des Schulklimas in die Betrachtung einbezogen: der Leistungsdruck, die Disziplin, die Kontrolle, die Mitbestimmung, das Lehrerengagement und das persönliche Vertrauen. Fend klassifizierte die Schulen und gelangte so zu einem Bündel von Qualitätskriterien.

Auf der Grundlage seiner Untersuchungen versuchte er, die Frage „Was ist eine gute Schule?“ zu beantworten (Fend, 1989) und fand dabei folgende Qualitätskriterien: hohe Arbeitszufriedenheit, anregungsreiches Schulleben, hoher Integrationsgrad des Lehrerkollegiums, konstruktive Konfliktbewältigung im Kollegium, keine Gleichgültigkeit und Anonymität im Verhältnis zu den Schülerinnen und Schülern sowie eine hohe Qualität der Schulleitung (Fend, 1998, S. 103). Diese Ebene der Schulqualität ergänzte er durch die Ebene der Systemqualität.

#### *Internationales Netzwerk Innovativer Schulen (INIS)*

Im Jahre 1996 wurden die Schulsysteme von sechs Ländern (Kanada, Niederlande, Norwegen, Schottland, Schweiz, Ungarn) für den Carl-Bertelsmann-Preis als besonders „innovative Schulsysteme im internationalen Vergleich“ nominiert. Die internationale Recherche im Vorfeld der Preisverleihung hatte gezeigt, dass jedes der nominierten Länder bereits über ein beträchtliches Know-how zum Thema „Innovation in Schulen und Schulsystemen“ verfügte. Zugleich wurde ein erheblicher Bedarf an zusätzlichen Informationen in Hinblick auf die Erforschung und Erprobung neuer Ideen festgestellt. Daraufhin gründeten Vertreterinnen und Vertreter aus Schulministerien und Schulen der nominierten Länder auf Initiative der Bertelsmann-Stiftung 1997 das Internationale Netzwerk innovativer Schulsysteme (INIS). Das internationale Netzwerk sah es als seine Aufgabe an, den internationalen Erfahrungsaustausch systematisch mit dem Ziel auszubauen, aus den jeweiligen Stärken der Schulsysteme zu lernen und daraus einen Beitrag zur nachhaltiger Verbesserung der Arbeit aller Beteiligten vor Ort zu leisten. INIS verfolgt dementsprechend drei zentrale Anliegen:

- die Stärkung der Innovationskraft der Schulsysteme,
- die Stärkung des internationalen Know-how-Transfers zwischen den Schulen und Schulsystemen durch hierarchieübergreifenden Dialog und
- die Entwicklung und Verbreitung beispielhafter Lösungen zu drängenden Problemen der Schulsystem-Reform.

Die erste Arbeitsphase, auch als INIS I bezeichnet, umfasste den Zeitraum 1997-2000. Im Mittelpunkt des Erfahrungsaustauschs der Teilnehmerländer standen in dieser Phase die Themen „Selbstevaluation von Schulen“ („Quality and Standards“) und „Instrumente der Rechenschaftslegung“ („Accountability“). Beispiele für gelungene Ansätze wurden zusammengetragen und systematisch ausgewertet (Bertelsmann-Stiftung, 2004).

Im Januar 2001 begann mit INIS II die zweite Projektphase (Laufzeit 2001-2005). Bei INIS II ist der Fokus auf den Qualitätsvergleich gerichtet. Im Gegensatz zu der stärker wissenschaftlich orientierten Bildungsforschung geht es bei INIS II explizit um praktische Hilfestellungen für Schulen. Ausgehend von der Erkenntnis, dass Entwicklungsprozesse



maßgeblich verbessert werden können, wenn sie sich auf empirische Vergleichsdaten und nicht allein auf pädagogische Theorien, Traditionen, individuelle Beobachtungen oder Intuition stützen, wurde aus der internationalen Praxis des Netzwerks ein Steuerungsinstrument für den Schulentwicklungsprozess von Einzelschulen erarbeitet. Vor dem Hintergrund eines internationalen Trends zur Stärkung der Autonomie der Einzelschule war es eine wichtige Motivation, ein Steuerungsinstrument speziell für einzelne Schulen zur Verfügung zu stellen (Stern, Mahlmann & Vaccaro, 2003). Dies impliziert zugleich eine stärkere Eigenverantwortung für die Schulqualität. Die besondere Herausforderung an die einzelnen Schulen besteht in der Qualitätsentwicklung innerhalb der gesetzlich vorgeschriebenen Rahmenbedingungen, um ein dem besonderen Profil der jeweiligen Schule entsprechendes individuelles Angebot zu entwickeln. Zur realistischen Einschätzung der eigenen Stärken und Schwächen ist jedoch die Entwicklung geeigneter Steuerungsinstrumente notwendig.

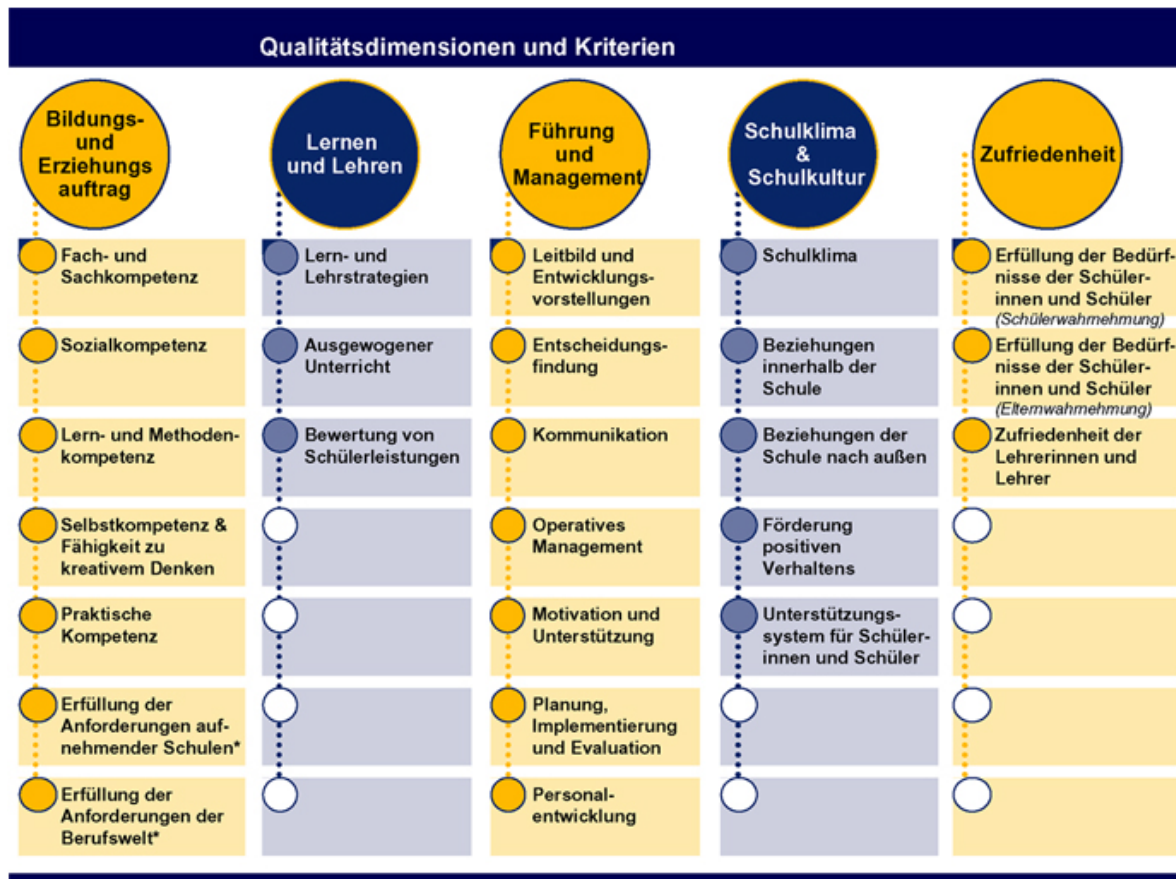
Seit Anfang 2001 arbeiten 41 Schulen aus den Teilnehmerländern der ersten Studie sowie den deutschen Bundesländern Bayern, Hessen, Niedersachsen, Nordrhein-Westfalen und Thüringen gemeinsam mit Schulentwicklungsberatern und Vertretern der jeweiligen Bildungsministerien im internationalen Netzwerk innovativer Schulsysteme (INIS II) an dieser Entwicklungsaufgabe. Die gemeinsame Zielsetzung ist die Entwicklung eines international tragfähigen Steuerungsinstruments, „...mit Hilfe dessen Schulleitungen und Kollegien in die Lage versetzt werden sollen, ihren Entwicklungsprozess systemischer, effizienter, effektiver und nachhaltiger zu steuern.“ (Stern, Mahlmann & Vaccaro, 2003, S. 2). In den Jahren 2001 bis 2003 ist das von dem internationalen Team erarbeitete Erhebungsinstrument zur Selbstevaluation mehreren Probeläufen und statistischen Analysen unterzogen worden. Hervorzuheben ist dabei der Anspruch, ein Instrument zu entwickeln, dass in verschiedenen nationalen und kulturellen Kontexten zum Einsatz kommen kann. In einem ersten Probelauf der Instrumente im Jahre 2002 wurden 10-jährige und 15- bis 17-jährige Schülerinnen und Schüler und ihre Eltern befragt. Außerdem kamen „Checklisten“ für die Schüler- und Lehrerschaft zum Einsatz, mit deren Hilfe mehrere Perspektiven erfasst werden konnten. Diese Checklisten wurden später in die Erhebungsinstrumente integriert, um den Befragungsablauf zu optimieren. Seit 2004 liegt es in einer wissenschaftlich überprüften und in der internationalen Praxis erprobten Fassung vor. Elemente dieses Steuerungsinstrumentes sind: (a) ein international tragfähiges Qualitätsverständnis von guter Schule, (b) wissenschaftlich geprüfte und praxiserprobte Erhebungsinstrumente, (c) ein handhabbares Berichtswesen und (d) Unterstützungsmaterial zur Durchführung von Evaluationen, Datenanalyse und Maßnahmenplanung.

Grundvoraussetzung für den Qualitätsvergleich ist ein gemeinsames Verständnis aller Beteiligten von guter Schule. Fünf zentrale Qualitätsdimensionen wurden entwickelt. Die fünf Dimensionen sind: (1) Bildungs- und Erziehungsauftrag, (2) Lernen und Lehren, (3) Führung und Management, (4) Schulklima und Schulkultur sowie (5) Zufriedenheit. Ihnen wurden 25 spezifische Kriterien zugeordnet und ein Indikatorenkatalog wurde erstellt. Abbildung 1.2 zeigt die Verbindung diese Kriterien mit den fünf Qualitätsdimensionen. Die fünf Qualitätsdimensionen wurden durch die Zuordnung von Kriterien und die Bestimmung geeigneter Indikatoren operationalisiert.

Die empirische Untersuchung selbst ist mehrperspektivisch angelegt. Es wurden für zwei Altersgruppen von Schülerinnen und Schülern Fragebögen entwickelt. Damit fand die Befragung in der 4., 9. und 11. Klasse statt. Erfasst werden auch die Sichtweisen der Lehrerinnen und Lehrer und Mitarbeiterinnen bzw. Mitarbeiter der Schule und die der Eltern. Um die Befragungsdaten in einen Kontext einordnen zu können, werden sie durch Angaben der Schulleitung ergänzt, die in einem Formular abgefragt werden. Die Befragungsinstrumente umfassen die Qualitätsdimensionen sehr umfangreich. So enthält der

Fragebogen für die 4. Klasse rund 100 Items. Dabei werden nicht nur erlernte Kompetenzen und Fähigkeiten abgefragt, sondern auch die Lernmethodik des Unterrichts. Kaum untersucht werden die Klassenbeziehungen und die Attraktivität des Schullebens. Es fällt auf, dass alle Aussagen positiv formuliert sind und manche Fragen sehr detailliert gestellt sind. Dafür ein Beispiel: „Ich kann vieles von dem, was ich in der Schule lerne, auch außerhalb der Schule gebrauchen (ich kann z.B. einen Zugfahrplan lesen, Aufgaben im Haushalt übernehmen, einen Brief schreiben, Anweisungen folgen, ein Problem lösen, auf etwas sparen etc.).“

**Abbildung 2 Qualitätsdimensionen von INIS**



Quelle: Bertelsmann-Stiftung (2004)

## 2.6 Integrative Ansätze

Die im Folgenden vorgestellten Studien und Programme integrieren mehrere Forschungsperspektiven. So verbindet die neuere Schulleistungsforschung in Studien wie TIMS und PISA die Leistungsmessung mit der Erhebung des Klassen- und Schulklimas. Schuleffektivitätsforschung wird durch den Vergleich von Schulen untereinander möglich, insbesondere der Vergleich von Schulen unterschiedlicher Schulformen kann in diesem Sinne interpretiert werden. Von den in den vorhergehenden Abschnitten beschriebenen Ansätzen findet allerdings die Lehr-Lernforschung die stärkste Anwendung. Die Hinwendung zum Produkt „Leistung“ ist ein inhärenter Bestandteil dieser Tradition. Auch neuere Qualitätssicherungssysteme nutzen mehrere Perspektiven von Schulqualität. Abschließend sollen deshalb einige Beispiele für die Integration dieser einzelnen Forschungsansätze betrachtet werden.

### 2.6.1 Nationale und internationale Schulleistungsuntersuchungen

Die Leistungsevaluation hat im Bildungswesen eine lange Tradition. Eine empirisch fundierte Leistungsforschung gibt es allerdings erst seit der Mitte des letzten Jahrhunderts: Erst mit den Mitteln der Sozialforschung und der psychologischen Testentwicklung, deren Voraussetzungen sich in der ersten Hälfte des letzten Jahrhunderts dynamisch entwickelten (Methoden, Statistik, Datenverarbeitung) wurde es möglich, Schulleistungen systematisch zu vergleichen. Schulleistungsstudien verfolgen unterschiedliche Ziele – entsprechend groß ist ihre Varianz. Allen gemeinsam ist die Erfassung der individuellen Schülerleistung, in der Regel mit Hilfe eines speziell für diese Untersuchung entwickelten Leistungstests. Ziel solcher Studien kann die Prüfung sein, ob angestrebte Bildungs- oder Lehrziele erreicht werden. Es ist aber auch möglich, den eigenen Standort im (inter)nationalen Vergleich zu bestimmen. Weiterhin können sie dazu dienen, spezielle Maßnahmen zu prüfen oder das schulische Qualitätsmanagement zu sichern. Wichtig ist auch ihr Beitrag zur Grundlagenforschung, wenn man z.B. der Frage nachgeht, wie Schülerleistungen und Leistungsunterschiede zustande kommen.

In den 1950er Jahren begannen internationale Organisationen wie die UNESCO, das „International Bureau of Education“ (IBE) (1925 als NGO gegründet; seit 1969 gehört es zur UNESCO) und die OECD damit, Informationen über die Schulsysteme unterschiedlicher Länder zu sammeln und deren Auswirkungen auf die sozioökonomische Entwicklung zu interpretieren (Bos & Postlethwaite, 2002). Dabei stellte man fest, dass sich die Ergebnisse einzelner Bildungsabschlüsse von Land zu Land erheblich unterschieden. Die Notwendigkeit von Leistungsvergleichen wurde formuliert. 1958 begann man mit den Vorarbeiten zur „Pilot Twelve-Country Study“ (1959-1962). Zielgruppe waren 13-jährige Schüler und Schülerinnen, untersucht wurden die Leistungen in den Fächern Mathematik, Lesen, Geographie, Naturwissenschaft und nonverbale Fertigkeiten. An dieser Studie waren neben Westdeutschland und anderen westeuropäischen Ländern auch Polen, Jugoslawien und die USA beteiligt. Schon damals zeigte sich eine erhebliche Varianz innerhalb der beteiligten Länder, die deutlich geringer war als zwischen den Ländern.

In der „First International Mathematics Study“ (FIMS) wurden 1964 die mathematischen Kompetenzen von Schülerinnen und Schüler in zwölf Ländern gemessen. Aus der Bundesrepublik Deutschland beteiligten sich lediglich zwei Bundesländer an der Erhebung. Es folgten weitere Studien, in denen die Schülerleistungen in Naturwissenschaft, Englisch und Leseverständnis erhoben wurden. Diese Studien wurden von der International Association for the Evaluation of Educational Achievement (IEA)<sup>1</sup> durchgeführt. Die letzte Studie unter deutscher Beteiligung war die Six Subject Study (1970-1971). Die Bundesrepublik beteiligte sich an Teilstudien zu Schulleistungen in den Naturwissenschaften, Englisch und zur politischen Bildung (Baumert, 2001). Im Mittelpunkt des Interesses zu dieser Zeit stand die Frage nach der besten Schulform (s. Untersuchungen von Fend). Leistungsmessung galt als pädagogisch nutzlos oder gar gefährlich (Helmke & Schrader, 2001). Kritiker der Leistungsmessung sahen Leistungsanforderungen und die Messung von Leistung als kinderfeindliche, antireformpädagogische und ökonomisch instrumentalisierbarer Kontrollmechanismen zur Disziplinierung von Lehrerschaft und Schülerschaft an (Weinert, 2001). Auch die DDR beteiligte sich nicht an internationalen Vergleichstests.

Erst in den 1990er Jahren beteiligte sich Deutschland wieder an internationalen Vergleichsuntersuchungen. Sowohl an der IEA „Reading Literacy Study“ (1990/1991 durchgeführt) als auch an der „CIVIC Education Study“ (eine Studie, die das historisch-

---

<sup>1</sup> Die IEA ist ein gemeinnütziger Verband, der internationale vergleichende Schulforschung durchführt. Zu ihr gehören Bildungsministerien, Forschungsinstitute und Universitäten aus 39 Mitgliedsländern.

politische Bewusstsein von 14-jährigen Schülerinnen und Schüler in 28 Ländern misst) nahm Deutschland teil. Von besonderer Bedeutung war die „Third International Mathematics and Science Study“ (TIMSS), an der sich 46 Länder beteiligten. Die TIMS-Studie wurde Mitte der 90er Jahre von der IEA durchgeführt und besteht aus mehreren Untersuchungen. Deutschland nahm nur an TIMSS 1995 teil, nicht an der Untersuchung im Jahr 1999. Im Verlauf von TIMSS wurde das Leistungsvermögen von Schülerinnen und Schülern der Grundschulen und der Sekundarstufe I und II erfasst (Baumert & Lehmann, 1997). In Deutschland wurden über 7000 Jugendliche der Klassenstufen 7 und 8 in die Untersuchung einbezogen (TIMSS/II) sowie mehr als 5000 Befragte am Ende der gymnasialen Oberstufe (TIMSS/III). Getestet wurden die Fächer Mathematik und Naturwissenschaft. In der Bundesrepublik Deutschland gab es im Rahmen von TIMSS keine Grundschuluntersuchung (TIMSS/I). Sie wurde später durch die Internationale Grundschul-Leseleistung (IGLU) „nachgeholt“. Neben Leistungstests wurde ein Schülerfragebogen eingesetzt, der eine Reihe von Fragen zur Unterrichtsqualität enthielt. Ergänzt wurde die Untersuchung durch eine Videostudie. Sie erfasste didaktische Abläufe des Unterrichts (Skripts) in Deutschland, den USA und Japan.

Die TIMSS-Resultate lösten in Deutschland Bestürzung aus, da die deutschen Schülerinnen und Schüler dabei nur mittelmäßig abschnitten. In der Folge wurde in Deutschland eine öffentliche Diskussion ausgelöst. Leistungsüberprüfungen und Leistungsmessungen, die jahrzehntelang tabuisiert worden waren, wurden nicht länger als nutzlos angesehen. Am Ende dieses Diskussionsprozesses stand ein ganz neues Bildungsverständnis, das bundesweiten Leistungsstandards und deren Evaluation eine neue Bedeutung einräumte. Kurz, man kann diesen Prozess als einen Wechsel von der Inputsteuerung zur Outputsteuerung charakterisieren. Nicht mehr gut ausgebildete Lehrerinnen und Lehrer und akribisch ausformulierte Lehrpläne wurden als Garanten für eine gute Schule abgesehen, sondern das Ergebnis.

Als Reaktion auf das mäßige Abschneiden der deutschen Schülerinnen und Schüler in der TIMSS wurden von drei Bundesländern große, flächendeckende Leistungstests durchgeführt: Die Lern-Ausgangslagen-Untersuchung (LAU) in Hamburg ist eine Längsschnitterhebung, um den Lernfortschritt in den Fächern Deutsch, Mathematik, in der ersten Fremdsprache und in fächerübergreifenden Kompetenzen zu erfassen. Sie begann 1996 in der 5. Jahrgangsstufe und wird im zweijährigen Abstand durchgeführt (z.B. Lehmann et al., 1997 bzw. 2002). Eine weitere Untersuchung ist die Qualitätsuntersuchung an Schulen zum Unterricht in Mathematik (QuaSUM) in Brandenburg. QuaSUM wurde 1999 vom Ministerium für Bildung, Jugend und Sport des Landes Brandenburg in Auftrag gegeben, um den Stand der Aneignung mathematischer Kenntnisse am Ende der 5. und 9. Klassenstufe zu bilanzieren (Lehmann et al., 2000). Um die Lernstände in Mathematik zu untersuchen, wurden in einer repräsentativen Studie 12500 Schülerinnen und Schüler befragt. Neben dem Lerntest, der zum Teil TIMSS-Aufgaben nutzte, wurden die Jugendlichen, Mathematiklehrerinnen und Mathematiklehrer, Eltern und Schulleiterinnen bzw. Schulleiter befragt. In der Schülerbefragung ging es darum, schul- und unterrichtsbezogene Einstellungen zu messen. Dazu wurden 85 Items entwickelt, welche die Schulzufriedenheit, das leistungsbezogene Selbstkonzept, das Mathematikinteresse, die Anstrengung und die Unterrichtsqualität erfassen. Der Elternfragebogen bezieht sich auf die Bildungsnähe des Elternhauses, den sozioökonomischen Hintergrund der Schülerschaft und die Alltagsgestaltung der Jugendlichen. Die Lehrkräfte in Mathematik wurden gebeten, die Unterrichtssituation einzuschätzen, zudem wurde ihre Bildungsbiographie erfasst. Nur die Schulleiter und Lehrkräfte schätzen das Schulklima ein.

Eine weitere bundeslandspezifische Untersuchung nach TIMSS ist die Mathematik-Gesamterhebung: Kompetenzen, Unterrichtsmerkmale, Schulkontext (MARKUS) in Rheinland-Pfalz. Hier handelt es sich wie bei LAU um eine Vollerhebung. Diese Untersuchung dient ebenfalls der Erfassung von Mathematikleistungen. Sie wurde vom

Kultusministerium Rheinland-Pfalz in Auftrag gegeben und im Mai 2000 durchgeführt. Neben der Erfassung der Mathematiktestleistung der Schülerinnen und Schüler in allen 8. Klassen wurden Schülerfeedbacks zum Unterricht eingeholt (Helmke & Jäger, 2002). Ihr Ziel war es, die Lernvoraussetzungen, Unterrichtsmerkmale und Schülermerkmale (wie das Geschlecht, die Nationalität, die Schulabschlüsse der Eltern, das Vorhandensein eines Computers zu Hause) zu erfassen und in Beziehung zu den Schulleistungen zu setzen. Dabei wurden auch eine Reihe von Merkmalen des Unterrichts und des Mathematiklehrers bzw. der Lehrerin erhoben. So wurde erfragt, wie motivierend und unterstützend die Lehrkraft ist, wie strukturiert der Unterricht ist, welche Relevanz das Gelernte hat. Wichtige Aspekte waren auch die Schülerorientierung, Binnendifferenzierung, Kleingruppenarbeit und die Klassenführung. Diese Ergebnisse der Einzelschule wurden an die Schulen klassen- bzw. kursbezogen rückgemeldet.

Die OECD entwickelte parallel und in Konkurrenz zur IEA ein langfristiges Forschungsprogramm zur Erfassung von Schülerleistungen (Baumert, 2001). Das „Programme for International Student Assessment“ (PISA) will den Teilnehmerländern Prozess- und Ertragsindikatoren zur Verfügung stellen, die für politisch-administrative Entscheidungen zur Verbesserung der nationalen Bildungssysteme brauchbar sind (Baumert et al., 2002). Zurzeit liegen die Ergebnisse von PISA 2000 und PISA-E 2000 vor (E steht für „Ergänzend“). An beiden Untersuchungen nahmen über 50.000 Schülerinnen und Schüler aus 1466 deutschen Schulen teil (Baumert, Stanat & Demmrich, 2001).

Im Zentrum dieser Schulleistungsstudie steht die Analyse schulisch erworbener Kompetenzen. Will man Leistungsunterschiede zwischen und in den einzelnen Ländern erklären, muss man die Kontextbedingungen berücksichtigen, unter denen Schülerinnen und Schüler lernen, und ein theoretisches Modell entwickeln, wie diese die Schülerleistung beeinflussen. Dies geschieht in der PISA-Untersuchung. Neben den familiären Lebensverhältnissen, der sozialen Herkunft und dem Freizeitverhalten werden die Rahmenbedingungen des Lernens erfasst, die vom Einschulungsalter bis zu den schulischen Ressourcen reichen. Zu den Kontextbedingungen gehört auch das Unterrichts- und Schulklima; beides wird in PISA durch Befragungen erhoben. PISA integriert also verschiedene Ansätze der Schulqualitätsforschung und nutzt dazu unterschiedliche Methoden. Als ein wesentliches Qualitätsmerkmal von Schulen wird dabei das Schulklima angesehen. Das folgende Zitat zeigt, von welchem Klimaverständnis die Autoren ausgehen: „Beim Schulklima handelt es sich um ein zentrales Qualitätsmerkmal des pädagogischen Prozesses in der Schule, das sowohl das didaktisch-methodische Arrangement des Fachunterrichts als auch die schulische Lebenswelt durchdringt. Mit der Erfassung von Merkmalen des Schulklimas werden subjektiv wahrgenommene Umwelten konstruiert.“ (Baumert, Stanat & Demmrich, 2001, S. 47). Zum Schulklima gehören die allgemeine Schulzufriedenheit, die Schüler-Schüler- und Schüler-Lehrer-Beziehung, physische Gewalt, Leistungs- und Disziplindruck und die Kommunikation im Kollegium. Schon die Dimensionen zeigen, dass man hier unterschiedliche Perspektiven berücksichtigen muss. So wurde ein Schüler-, Lehrer- und Elternfragebogen entwickelt. Das deutsche PISA-Konsortium modifizierte und erweiterte die internationalen Fragebögen. Allerdings stand man nun vor dem Problem, diese Dimensionen (die jede für sich genommen schon Gegenstand einer Untersuchung sein könnten und auch schon waren) so zu operationalisieren, dass sie in 20-30 Minuten abgefragt werden konnten. Man tat dies unter Rückgriff auf etablierte Skalen. So enthielt der internationale Schülerfragebogen fünf Items zur Schüler-Lehrer-Beziehung, die das persönliche Verhältnis, das Engagement der Lehrkräfte und die Unterstützung und Behandlung durch die Lehrerschaft erfragen (Kunter, Schümer, Artelt et al., 2001). Problematisch erscheint die Erfassung physischer Gewalt. Hier wird nur eigenes vandalisches und gewalttätiges Verhalten der letzten zwölf Monate erfasst (Prügeleien, Belästigungen,

Bedrohungen und Beschädigungen). Obwohl dies alles sehr stark miteinander zusammenhängt (Cronbachs Alpha = 0.9), ist das eigene aggressive Verhalten nur bedingt zur Beschreibung des Gewaltklimas an der Schule geeignet.

Neuere nationale Schulleistungsuntersuchungen wie PISA und IGLU geben eine Rückmeldung über die Untersuchungsergebnisse an die Schulen. Es werden also nicht mehr nur aggregierte Daten auf der Basis einer repräsentativen Stichprobe untersucht, sondern diese können auch zur Evaluation von Einzelschulen eingesetzt werden. Dabei wird auch die Zufriedenheit mit der Schule rückgemeldet, wie sie im Schüler- und Elternfragebogen erhoben wurde. Erfasst werden die Schüler-Lehrer-Beziehung, die Schüler-Schüler-Beziehung und die allgemeine Zufriedenheit der Schülerschaft mit der Schule. Zur Einordnung der eigenen Ergebnisse dienen Vergleichswerte von anderen Schulen. Um den Schulen diesen Vergleichsmaßstab anzubieten, werden sie bei ausreichender Fallzahl mit Schulen gleichen Schultyps im gleichen Bundesland verglichen.

Positiv hervorzuheben ist bei der PISA-Untersuchung die Entwicklung eines Modells der Wirkung der Kontextmerkmale auf die Schulleistung, wobei diese Kontextmerkmale aus unterschiedlicher Perspektive erhoben werden. Neben der Schulleiterbefragung werden auch die Sichtweisen der Schülerinnen und Schüler sowie der Eltern erfragt. Da der Schwerpunkt von PISA auf der Leistungsmessung liegt, können jedoch nur wenige Merkmale der Schule erfasst werden. So mussten einzelne Items aus Platzgründen aufgegeben werden, obwohl damit die Skalengüte schlechter wird (Kunter et al., 2001). Wichtig für die Qualitätsentwicklung der Einzelschulen wären Längsschnittuntersuchungen, um individuelle Kompetenzzuwächse durch den Schulbesuch zu erfassen (dies würde auch Unterschiede durch unterschiedliche Schulprofile nivellieren) und um Kausalanalysen zu ermöglichen. Die Beschränkung auf die 9. Jahrgangsstufe und auf spezielle Wissensgebiete mindert die Aussagekraft der Ergebnisse, sodass keine umfassende Beschreibung der Fähigkeiten vorliegt. Für das Gelingen von Schulentwicklungsprozessen ist ein Input in Form einer Evaluation zwar nützlich, keinesfalls aber eine hinreichende Bedingung. PISA liefert die Diagnostik für einen selektiven Aspekt der Schülerleistung, eine Begleitung von Schulentwicklungsprozessen kann und will diese Untersuchung jedoch nicht leisten. Die Zufallsauswahl einzelner Schüler und Schülerinnen ermöglicht zudem keine Aussagen auf der Ebene von Schulklassen. Ziel von PISA ist die Bildungsberichterstattung, Schulevaluation ist hier ein „Nebenprodukt“. Im nächsten Abschnitt sollen deshalb Verfahren beschrieben werden, die explizit der Schulevaluation dienen.

## **2.6.2 Schulinspektion**

In den Medien hat sich für die externe Qualitätsuntersuchung der Ausdruck „Schul-TÜV“ durchgesetzt (vgl. zum Beispiel: Der Spiegel 20/2004). Mehrere Bundesländer planen derzeit die Einführung solcher Evaluationssysteme. Sie sind integrativ im dem Sinne, dass alle bisher diskutierten Aspekte von Schulqualität auf die eine oder andere Weise mit berücksichtigt werden. Diese Qualitätssysteme sind nicht neu. Das Office for Standards in Education (Ofsted) in England oder die Schulinspektionen der Niederlande nutzen schon geraume Zeit diese Form der Schulevaluation. Am Beispiel von England, Niedersachsen und Schleswig-Holstein werden nachfolgend Verfahren der Qualitätskontrolle vorgestellt. Anschließend wird der Blick auf unsere Nachbarländer Österreich und die Schweiz (Kanton Luzern) gerichtet.

### *Schulevaluation in England*

Auf der Basis der Ofsted-Homepage soll kurz die Vorgehensweise bei englischen Schulinspektionen vorgestellt werden. Bis Anfang der 1990er Jahre waren staatliche

Schulinspektoren („Her Majesty’s Inspectors“) für die Evaluation von englischen<sup>2</sup> Schulen zuständig. Bis dahin war die systematische Überprüfung von Schulen eher die Ausnahme: Im Durchschnitt wurde jede Schule nur einmal in zwanzig Jahren überprüft, und Schulberichte wurden grundsätzlich nicht der Öffentlichkeit zugänglich gemacht (Sammons et al., 2004). Das änderte sich schlagartig mit der Gründung des Office for Standards in Education, kurz „Ofsted“, im Jahre 1992. Ofsteds erklärtes Ziel ist die Verbesserung der Schulleistungsstandards und der Bildungsqualität durch regelmäßige unabhängige Inspektionen, öffentliche Berichterstattung und die Bereitstellung professioneller Hilfestellungen für Lehrerinnen und Lehrer sowie Schulverwaltungen. Mit diesen Evaluationen sollen die Eltern einen externen und unabhängigen Blick darauf erhalten, was an der Schule ihrer Kinder gut und was weniger gut ist. Da diese Berichte öffentlich zugänglich sind, dienen sie Eltern auch als Entscheidungshilfe für die Schulwahl. Auf Grundlage dieser Berichte und der erreichten Punktezahlen in Abschlussprüfungen werden Rankings erstellt. Die von Ofsted durchgeführte Evaluation der Schulqualität und der Leistung der Schulverwaltungen betrifft vier Kernbereiche: die von der Schule vermittelte Bildungsqualität, die von der Schule erreichten Bildungsstandards, die geistige, moralische, soziale und kulturelle Entwicklung der Schülerinnen und Schüler und die Verwaltung und Effizienz der Schule. Schulinspektionen finden wenigstens alle sechs Jahre statt, weniger erfolgreiche Schulen werden häufiger evaluiert. Schulinspektionen sollen den Schulen helfen, ihre Stärken und Schwächen zu identifizieren.

Ofsted legt das Evaluationsprogramm für jedes Schuljahr fest. Die Aufträge zur Inspektion werden ausgeschrieben. So können verschiedene Dienstleister<sup>3</sup> um die Verträge konkurrieren. Die Mitglieder des Inspektionsteams sind unabhängig und werden von den verschiedenen Dienstleistern bereitgestellt, die regional verteilt sind. Die Inspektionsteams werden von staatlich anerkannten Inspektoren geleitet. Die Kontrolle dieser Teams wird durch „Her Majesty’s Inspectors“ (HMI) wahrgenommen, die bei Ofsted festangestellt sind und an allen Inspektionen teilnehmen können. An jeder Inspektion muss wenigstens ein Laieninspektor teilnehmen, der kein Pädagoge ist und auch nicht in der Schulverwaltung tätig war oder ist.

Bevor eine Evaluation beginnt, werden separate Vorgespräche mit der Schulleitung und den Eltern geführt. Diese Treffen sind gesetzlich vorgeschrieben. Inspektionen dauern gewöhnlich eine Woche. Im Mittelpunkt der eigentlichen Schulevaluation stehen Hospitationen. Es wird angestrebt, bei den meisten Lehrerinnen und Lehrern zwei- bis dreimal zu hospitieren. Dies führt zu einer sehr großen Anzahl von Unterrichtsstunden, die beobachtet werden. Die Lehrqualität wird auf der Basis einer 7-Punkt-Skala beurteilt. Die Lehrerinnen und Lehrer erhalten ein Feedback zu ihrer Leistung. Der Schulleitung werden zudem vertrauliche Informationen über die Beurteilung zur Verfügung gestellt. Außerdem wird ein Gesamtbericht zur Lehrqualität der Schule angefertigt.

Am Beispiel der St. Michael’s RC School (in London, rund 600 Schülerinnen und Schüler von 11-16 Jahren, Modern School) sei kurz das Vorgehen beschrieben. Im Mai 2003 evaluierten 15 Inspektoren eine Woche lang die Schule. Dabei wurden 128 Unterrichtsstunden beobachtet. Diese Unterrichtsstunden wurden wie folgt bewertet (Roberts, 2003):

---

<sup>2</sup> Das englische Schulsystem wird separat von jenen in den anderen Teilen des Vereinigten Königreichs Schottland, Wales und Nordirland verwaltet.

<sup>3</sup> Inspektionsdienstleister sind überwiegend kommerzielle Organisationen oder Local Education Authorities (Ofsted, 2003).

	Excellent	Very good	Good	Satisfactory	Unsatisfactory	Poor	Very poor
Number	4	24	51	41	5	1	1
Percentage	3	19	40	32	4	1	1

*The table gives the number and percentage of lessons observed in each of the seven categories used to make judgements about teaching.*

Schularbeiten, Hospitationen, Daten von nationalen Leistungsüberprüfungen und öffentliche Prüfungen werden zu Rate gezogen, um die Bildungsstandards an der betreffenden Schule zu beurteilen. Schuldaten und nationale Datensätze werden genutzt, um Informationen über die Leistungen der Schülerinnen und Schüler im nationalen Vergleich zu erhalten. Darüber hinaus werden Kinder und Jugendliche befragt sowie Schularbeiten herangezogen, um die Fertigkeiten und Lernfortschritte der Schülerinnen und Schüler, ihre Anwesenheit und Einstellungen sowie ihr Verhalten und ihre persönliche Entwicklung zu evaluieren. Die Inspektoren müssen jedes Unterrichtsfach des nationalen Lehrplans („National Curriculum“) und den Religionsunterricht evaluieren und darüber Bericht erstatten. Separate Berichte müssen für die Pflichtfächer Englisch, Mathematik und Naturwissenschaften angefertigt werden.

Die Inspektoren beurteilen außerdem die Leitungsfähigkeit und das Management der Schulleitung und Schulverwaltung. Auch die Selbstevaluationsmechanismen der Schule sind Gegenstand der Inspektion. Daneben wird auch die materielle und personelle Ausstattung mit Ressourcen (wie z.B. das Fehlen von Fachlehrern und Fachlehrerinnen, Vertretungslehrern und Vertretungslehrerinnen, Schwierigkeiten bei der Anwerbung neuen Lehrpersonals) und ihre Auswirkungen auf die Qualitätsstandards überprüft.

Zur Inspektion gehören strukturierte Interviews mit führenden und mittleren Verwaltungsangestellten sowie eine schriftliche Befragung der Eltern. Der Elternfragebogen umfasst 15 Items zu verschiedenen Aspekten des Schullebens aus der Elternperspektive (z.B. das Wohlbefinden der Kinder, die Qualität des Unterrichts sowie die Kompetenz des Lehrkörpers und der Schulverwaltung). Antworten erfolgen mittels einer 4-Punkt-Skala. Bemerkenswert ist, dass die Elternfragebögen außer in Englisch in 24 weiteren Sprachen vorliegen (auch auf Deutsch). Dies führt aber nicht notwendigerweise zu einer hohen Beteiligung der Eltern an der Befragung (an der schon erwähnten St. Michael's School betrug die Rücklaufquote gerade einmal 28%). Von den Schülerinnen und Schülern, die ebenfalls einen Kurzfragebogen zu ihrer Schulsituation erhalten, wird hingegen erwartet, dass sie diese Fragen auf Englisch beantworten können. Es gibt entsprechend des Alters der Befragten drei Versionen des Schülerfragebogens (Grundschule, Sekundarstufe I und II). Grund- und Sekundarschüler beantworten 12, Abiturienten 16 Items mittels einer 4-Punkt-Skala.

Nach Abschluss der Evaluation wird ein Bericht erstellt, der innerhalb von sieben Wochen der Schule und Ofsted vorliegen muss. Er enthält u. a. einen Überblick über die erreichten Schülerleistungen sowie Einschätzungen der Unterrichtsqualität. Weiterhin gibt er Auskunft über die Spezifika der Schule, das Schulmanagement und die Einstellungen der Schülerinnen und Schüler zum Lernen. Der Bericht gibt Empfehlungen, auf dessen Grundlage die Schule einen Handlungsplan erarbeitet. Um zu illustrieren, welche Empfehlungen gegeben werden, wird wieder auf das Beispiel der St. Michael School zurückgegriffen:



### **What could be improved**

- Results in mathematics and science and those of girls in English and a number of other subjects
- The effectiveness of the school's improvement planning, with the headteacher's long-term vision at the forefront, and the coordination of the school's key policies to ensure consistent approaches to improvement
- The use of the data on pupils' performance which the school already collects to secure further improvement in their achievements
- Provision of information and communication technology (ICT) lessons for all older pupils
- The rigour with which governors carry out some of their responsibilities, including those relating to health and safety

Quelle: Roberts, 2003

Die Zusammenfassung des Berichts wird an die Eltern verteilt. Ofsted stellt die Berichte ins Internet. Die Ergebnisse werden so der Öffentlichkeit zugänglich gemacht. Neben individuellen Ergebnissen der Schulen werden leistungsbezogene Rankings publiziert, in denen gute und schlechte Schulen jeweils unter Berücksichtigung des sozialen Umfelds herausgestellt werden (Schwippert & Bos, 2003).

### *Schulinspektion in Niedersachsen*

Im Rahmen des Projekts „Qualitätsentwicklung in Netzwerken“ wurde ein Orientierungsrahmen zur Schulqualität entwickelt, der Modellcharakter trägt und mit Hilfe der holländischen Inspectie van het Onderwijs auch praktisch erprobt wurde (Niedersächsisches Kultusministerium, 2003a). Das Projekt „Von anderen Nationen lernen - Beurteilung schulischer Qualität“ wurde im Laufe der Jahre 2002-2003 im Rahmen der schon länger bestehenden Zusammenarbeit zwischen der niederländischen Inspectie van het Onderwijs und dem Niedersächsischen Kultusministerium durchgeführt. Im Projekt beurteilten beide Partner gemeinsam die Qualität der Schulen und des Unterrichts in einer Reihe ausgewählter Grundschulen und Schulen des Sekundarbereichs. Die folgende Darstellung basiert auf dem Abschlussbericht des Projektes: „Von anderen Nationen lernen'. Beurteilung schulischer Qualität in Niedersachsen“ (Niedersächsisches Kultusministerium, 2003b).

Das Anliegen des Projekts war vielschichtig. Im Mittelpunkt stand die Frage, wie die niederländischen Schulinspektoren die Schulqualität in Niedersachsen beurteilen, um so Qualitätsverbesserungen zu erzielen. Zu Beginn musste deshalb geklärt werden, ob das System der niederländischen Schulinspektion auf Niedersachsen übertragbar ist. Da diese Voraussetzung gegeben war, wurde anschließend analysiert, welche Gemeinsamkeiten und Unterschiede sich beim Vergleich der Schulqualität in den Niederlanden und Niedersachsen ergeben. Wichtig war aber auch, den einzelnen Schulen einen Spiegel vorzuhalten, wie es um ihre Schulqualität bestellt ist. Dafür wurden Schulen des Projektes „Qualitätsentwicklung in Netzwerken“ gewonnen. Es sollte Aufgabe der Evaluationssachverständigen sein, mit transparenten, deutlich strukturierten Methoden die Stärken und Entwicklungsmöglichkeiten der einzelnen Schulen zu identifizieren. Ein weiteres Anliegen war die Qualifizierung der niedersächsischen Evaluatoren. Hauptziel der Inspektoren war aber die Erfassung der Schulqualität an ausgewählten Schulen. Im Mittelpunkt steht dabei: „Die Qualität des Unterrichts einer Schule muss sich vor allem aus dem unterrichtlichen Lernprozess in Relation zu den realisierten Ergebnissen ergeben. Zum Lernprozess gehören das Unterrichtsangebot, die Lehrzeit, das pädagogische Klima, das didaktische Handeln der Lehrerinnen und Lehrer und die Schülerbetreuung.“ (Niedersächsisches Kultusministerium,

2003b, S. 26). Künftig sollen an allen Schulen in Niedersachsen Schulinspektoren die Qualität der schulischen Arbeit analysieren. Dazu soll ab 2005 eine Schulinspektion nach niederländischem Vorbild aufgebaut werden.

Wichtig für das Gelingen des Vorhabens waren transparente Qualitätsmerkmale und Indikatoren. Diese wurden auf deutsche Verhältnisse adaptiert. Insgesamt wurden 13 Qualitätsmerkmale erfasst. Den Qualitätsmerkmalen wurden fünf bis acht Indikatoren zugeordnet. Die Bewertung erfolgte dreistufig. Diese Einstufung bildete die Basis für die Bewertung des jeweiligen Qualitätsniveaus („Überwiegend schwach“, „Eher schwach als stark“, „Eher stark als schwach“ bis „Überwiegend stark“). Die Schulinspektion wurde von drei Evaluatoren durchgeführt, von denen zwei aus Niedersachsen kamen und einer aus den Niederlanden. Die deutschen Inspektoren wurden im Vorfeld geschult und nahmen an Schulinspektionen in den Niederlanden teil.

Es werden an zwei Tagen Unterrichtshospitationen durchgeführt (Dauer: 20-25 Minuten im Sekundarschulbereich, 45 Minuten in Grundschulen). Zusätzlich finden mündliche Interviews und Gespräche mit der Schulleitung (Leitfadeninterviews) statt. Weiterhin gibt es Gruppendiskussionen mit Eltern, Lehrerinnen und Lehrern sowie Schülerinnen und Schülern. Komplettiert wird diese Vorgehensweise durch Beobachtung und Dokumentenanalyse. Im Abschlussbericht werden die verschiedenen Aspekte und Fragen behandelt. Zudem werden im letzten Abschnitt Schlussfolgerungen und Empfehlungen formuliert. Das Modellprojekt wurde abschließend von der Schulleitung und Evaluationssachverständigen evaluiert.

Am Beispiel Schulklima kann man erkennen, dass nur sehr wenige Indikatoren in die Berichtslegung eingingen. Das übergreifende Qualitätsmerkmal heißt: Die Schule sorgt für ein sicheres und motivierendes Schulklima. Dabei wurden folgende Aspekte berücksichtigt:

- In der Schule gehen die Beschäftigten positiv mit den Schülern um.
- Die Schüler schätzen die Schule.
- Die Schule macht einen gepflegten Eindruck und wirkt einladend auf die Schüler.
- Die Schule organisiert Aktivitäten, durch die die Identifikation der Schüler mit der Schule erhöht wird.
- Die Schule wacht über die Sicherheit und Eigentum der Schüler.
- Die Schule besitzt klar formulierte Regeln für den Umgang miteinander.
- Das Personal wendet die vereinbarten Umgangsregeln abgewogen an.

(Norm Fünf von sieben Indikatoren)

Positiv hervorzuheben ist, dass versucht wird, neben der Unterrichtsqualität und der Betrachtung des Abschneidens bei Vergleichsarbeiten auch die Prozessqualität des Lernens mit zu berücksichtigen. Die Bewertung der Schulevaluatoren ist trotz transparenter, nachvollziehbarer und fundierter Bewertungskriterien immer noch durch große Ermessensspielräume geprägt und intensiver Schulung eine Sichtweise mit Ermessensspielraum. Kernpunkt der Evaluation ist der Unterricht. In den beschriebenen Beispielen wird an mehr als 20 Unterrichtsstunden teilgenommen.

### *Schul-Inspektion in Schleswig-Holstein*

Schleswig-Holstein führte als erstes deutsches Bundesland landesweite Qualitätstests durch. Jährlich sollen 160 Schulen einem Qualitätscheck unterzogen werden. „Externe schulische Evaluation im Team“ (EVIT) heißt dieser Qualitätstest. Für die folgende Darstellung wurde überwiegend auf das Qualitätshandbuch zu EVIT zurückgegriffen (Ministerium für Bildung, Forschung und Kultur des Landes Schleswig-Holstein, 2004). Das Anliegen dieses Tests ist

es, verlässliche Aussagen über die Qualität einer Schule zu gewinnen. EVIT formuliert Qualitätsmaßstäbe für Grundschulen, weiterführende allgemein bildende Schulen und Sonderschulen. Da nicht alle Aktivitäten einer Schule evaluiert werden können, beschränkt sich EVIT auf die Bereiche schulischer Qualität, die für den Erziehungs- und Bildungsauftrag als von zentraler Bedeutung angesehen werden. „Für EVIT gilt der Grundsatz, dass eine Schule nur für den Gestaltungsspielraum, den sie tatsächlich verantwortet, Rückmeldung zum Erfolg ihrer Arbeit erhalten kann. Deshalb werden bei EVIT die Lern- und Arbeitsbedingungen einer Schule als wichtige Voraussetzung schulischer Arbeit mit erhoben und entsprechend berücksichtigt. Hinzu kommen die rechtlichen Vorgaben, die soziale Umgebung und die institutionellen Voraussetzungen der jeweiligen Schullart...“ (Ministerium für Bildung, Forschung und Kultur des Landes Schleswig-Holstein, 2004).

Dazu wurden 42 Qualitätsindikatoren entwickelt, die vom Abschneiden bei Vergleichsarbeiten bis hin zum Vorhandensein von Kooperationsbeziehungen reichen. Ziel ist es, für jede Schule ein Stärken-Schwächen-Profil zu erarbeiten, eine Rückmeldung über den Erfolg der eigenen Arbeit zu erhalten und neue Impulse zu geben. Dieses Projekt ist nicht freiwillig und bei schlechten Evaluationsergebnissen mit einer verpflichtenden Komponente versehen. Weiterhin soll der Austausch zwischen den Schulen gefördert werden und die Schule im Schulentwicklungsprozess beraten werden. Alle sechs Jahre (also ähnlich wie in England) sollen die Schulen eine „abgesicherte und faire Außensicht“ erhalten. Die jährlich zu evaluierenden Schulen werden zufällig ausgewählt, eine freiwillige Beteiligung ist möglich. Die Instrumente von EVIT können aber auch zu einer schulinternen Evaluation eingesetzt werden. Um diese Zielsetzung umzusetzen, wurden über 40 Expertenteams gebildet. Sie setzen sich aus drei Vertretern folgender Institutionen zusammen: der Schulaufsicht, des Instituts für Qualitätsentwicklung an Schulen Schleswig-Holstein (IQSH)<sup>4</sup> und ein Schulleiter bzw. eine Schulleiterin einer vergleichbaren Schule der Region. Die ersten beiden Mitglieder werden von der Schulaufsicht benannt. Das dritte Mitglied wird von der zu evaluierenden Schule vorgeschlagen. Alle drei Mitglieder des Teams sind gleichberechtigt.

Die Überprüfung der sechs Teilbereiche mit ihren 42 Indikatoren erfolgt auf fünf Niveaustufen von „Trifft voll und ganz zu“ bis „Trifft überhaupt nicht zu“. Um diese Indikatoren einzuschätzen, wird auf eine Reihe verschiedener Methoden zurückgegriffen. Es werden Eltern-, Lehrer- und Schülerfragebogen im Vorfeld des Schulbesuchs eingesetzt, Dokumente werden eingesehen (z.B. Schulprogramme, Protokolle von Konferenzen), Daten der Schulstatistik werden ebenso einbezogen wie die Ergebnisse der Vergleichsarbeiten in den Fächern Mathematik und Deutsch. Dies wird ergänzt durch eine Unterrichtsbeobachtung, Gespräche und einen Schulrundgang. Der Evaluationsbesuch findet an einem oder zwei Tagen statt und wird protokolliert. Hierzu werden Fragebögen und Beobachtungsschemata verwendet. Alle Ergebnisse werden zusammengefasst und nach der Evaluation in einem Abschlussgespräch besprochen. Anschließend wird ein Bericht erstellt. Er gibt Auskunft über die Erfolge und Probleme der Schule und schlägt Maßnahmen zur Qualitätsverbesserung vor. Diese sind aber nicht bindend, sondern sollten vielmehr als Input für Maßnahmen zur Schulentwicklung verstanden werden. Sollten diese aus Sicht der Schulaufsicht nicht ausreichen, vereinbart sie mit der betreffenden Schule überprüfbare Schritte zur Lösung der Probleme. Das IQSH unterstützt die Schulen durch Moderatoren und Fortbildung.

---

<sup>4</sup> Das Landesinstitut Schleswig-Holstein für Praxis und Theorie der Schule (IPTS) besteht seit 1971. Es wurde am 3.2.2003 umbenannt in „Institut für Qualitätsentwicklung an Schulen, Schleswig-Holstein (IQSH). Das IQSH ist eine nicht rechtsfähige Anstalt des öffentlichen Rechts im Bereich des Ministeriums für Bildung, Wissenschaft, Forschung und Kultur des Landes Schleswig-Holstein. Im IQSH sind die zweite und dritte Phase der Lehrerbildung (Berufseinführung, Fort- und Weiterbildung sowie Unterrichtsfachberatung) unter einem Dach vereinigt.

Von besonderer Bedeutung für das Evaluationsvorhaben sind die eingesetzten Fragebögen. Deshalb soll zum Abschluss auf Art und Umfang der Erhebung eingegangen werden. Die Fragebögen sind für alle Schulen gleich, bieten also keine Möglichkeit, schon im Vorfeld des Besuchs Besonderheiten der Schule zu berücksichtigen. Spätestens drei Wochen vor dem Schulbesuch werden die Fragebögen ausgefüllt. In ein- und zweizügigen Schulen erfolgt eine Vollerhebung, ab dreistufigen Jahrgängen können Stichproben gezogen werden (auf Klassenbasis). Befragt werden alle Eltern, Lehrpersonen sowie Schülerinnen und Schüler „spätestens ab Klassenstufe 4“. Die Befragung ist anonym. Die Ergebnisauswertung auf Klassenebene ist möglich, wenn die entsprechende Lehrkraft dies wünscht. Der Schülerfragebogen umfasst 27 Items aus unterschiedlichen Bereichen. Viele Fragen in den drei Fragebögen beziehen sich auf die Zusammenarbeit zwischen den am Schulleben beteiligten Personen.

Wenn man die Items der Schülerbefragung liest, wird deutlich, dass besonders viel Wert auf die Schüler-Lehrer-Beziehung und die Unterrichtsgestaltung gelegt wurde. Es wird auch erfasst, inwieweit die Kinder und Jugendlichen etwas über Drogen und die friedliche Lösung von Konflikten lernen. Wünschenswert wäre aber auch die Erfassung des Umfangs des Drogenkonsums bzw. der Existenz von Schulgewalt gewesen. Positiv ist die Einbeziehung mehrerer Perspektiven in den Evaluationsprozess. Die repräsentative bzw. vollständige Erhebung der Lehrer-, Schüler- und Elternsicht sowie die Mehrsprachigkeit der Elternfragebögen sprechen für das Verfahren. Hervorzuheben ist weiterhin die Transparenz der Vorgehensweise. Auch die Anlage der Untersuchung als Onlineerhebung überzeugt, weil sie schnell ist, Kosten spart und die Auswertung von den Schulen selbst vom Server herunter geladen werden kann (s. Kapitel 4).

Da aber noch nicht alle Eltern, Lehrerinnen und Lehrer sowie Schülerinnen und Schüler die Möglichkeit haben, den Fragebogen im Internet auszufüllen, wird auch eine Papierversion angeboten. Im Qualitätshandbuch wird die Erwartung formuliert, dass die Eltern und Schulen die schriftliche Befragung selbst auswerten. Nur dann, wenn ihnen dies nicht gelingt, unterstützt das IQSH. Die Auswertung von mehreren hundert Fragebögen durch die Eltern erscheint unrealistisch. Auch die Probleme der Aggregation, die bei Verwendung zweier verschiedener Befragungsformen entstehen, werden nicht thematisiert. Da bisher kaum systematische Untersuchungen zu den methodischen Effekten beider Untersuchungsformen vorliegen, besteht hier weiterer Forschungsbedarf. Auch die Auswertung und Darstellung der Ergebnisse wird nicht beschrieben: Werden Norm- bzw. Vergleichswerte in Abhängigkeit von Schultypen, Region und sozialstrukturellen Gegebenheiten berechnet wie z.B. in den Niederlanden? Werden aus Einzelaussagen Skalen gebildet? Ein weiterer offener Punkt der Untersuchungsanlage ist die Teilnahmequote. Das Ausfüllen der Schülerfragebögen außerhalb der regulären Unterrichtszeit wird die Beteiligungsquote reduzieren. Die Erfahrungen von Ofsted zeigen, dass eine hohe Beteiligung der Eltern keineswegs selbstverständlich ist, insbesondere wenn dabei für die Eltern Kosten in Form von Internetgebühren anfallen. Es wird also davon abhängen, inwieweit den Eltern, Schülerinnen und Schülern und Lehrkräften die Nützlichkeit des Vorgehens verständlich gemacht werden kann.

EVIT überzeugt durch seine Vielschichtigkeit in Kombination mit der Nutzung moderner Erhebungsverfahren. Es bleibt aber ein Evaluationsinstrument. Die angestrebte Qualitätsverbesserung im Schulsystem von Schleswig-Holstein kann vermutlich aber nur gelingen, wenn Unterstützungs- und Hilffsysteme zur Verfügung gestellt werden, die individuell auf die Schulsituation zugeschnitten sind. Kurz, das Projekt steht und fällt mit den Schulentwicklungsmoderatoren des IQSH, den angebotenen Fortbildungen und den materiellen Rahmenbedingungen.

*Q.I.S. in Österreich*

„Qualität in Schulen“ (Q.I.S.) wurde vom Bundesministerium für Bildung, Wissenschaft und Kultur des Landes Österreich initiiert, um die Qualitätsdiskussion an den Schulen anzuregen (Bundesministerium für Bildung, Wissenschaft und Kultur, 2004). Schulqualität wird dabei nicht normativ vorgegeben, sondern es werden verschiedene Perspektiven und unterschiedliche Maßstäbe zugelassen. Es gibt dabei zwar allgemeine Hinweise in Gesetzen und Verordnungen zu beachten, die konkrete Umsetzung obliegt aber der Einzelschule. Deshalb wird bei Q.I.S. betont, dass die Schulqualität nicht eindimensional ist und ein Spannungsfeld unterschiedlicher Ziele umfasst (Bundesministerium für Bildung, Wissenschaft und Kultur, 2004: 3). Dieses Projekt soll dazu dienen, über die Schulqualität und das Leitbild der eigenen Schule nachzudenken und auf dieser Basis ein Schulprogramm zu entwickeln. Dazu bedarf es zu Beginn einer Bestandsaufnahme. Empfohlen wird die Erfassung von (allen) fünf Qualitätsbereichen: „Lehren und Lernen“, „Lebensraum Klasse und Schule“, „Schulpartnerschaft und Außenbeziehungen“, „Schulmanagement, Professionalisierung und Personalentwicklung“. Eine Arbeitsgruppe wählt und modifiziert diese Merkmale und schlägt Indikatoren zu ihrer Messung vor. Der Q.I.S.-Leitfaden empfiehlt zwar die Durchführung anonymer Befragungen, aber es sind auch Dokumentenanalysen, Beobachtungen und Interviews möglich.

Betrachtet man die vorgeschlagenen Erhebungsinstrumente, so fällt auf, dass neben den offenen Formen der Datenerhebung eine Reihe von standardisierten Fragebögen angeboten werden. Diese erfassen die Perspektive der Eltern, der Lehrerschaft und der Schülerschaft. Für Schülerinnen und Schüler wurden zwei Fragebögen entwickelt: Zum einen wird ein Unterrichtsfeedbackbogen für eine einzelne Lehrerin bzw. einen einzelnen Lehrer angeboten, zum anderen können mit einem Schülerfragebogen verschiedene Aspekte des Schullebens erfasst werden (Unterricht, Mitschülerinnen und Mitschüler, Schule allgemein). In den Q.I.S.-Materialien findet sich kein Hinweis darauf, ob die Schule bei der Auswertung unterstützt wird. Die Übertragung der Fragebögen von Vollerhebungen und Stichproben mittels der mitgelieferten Auswertungsschablone in eine entsprechende Tabelle ist sehr zeitaufwändig. Auch die Eingabe der Fragebögen in ein Computerprogramm wie beispielsweise Excel erfordert viel Zeit zur Erstellung der Datenmaske und der Dateneingabe. Den Schulen entstehen also beträchtliche Kosten. Ohne zusätzliche Ressourcen in Form von Abminderungsstunden oder zusätzlichem Personal ist das nicht zu leisten. Bei den Musterfragebögen besteht zudem noch Optimierungsbedarf. Der Wechsel des Antwortmodells, die Hervorhebung von negativ formulierten Aussagen (sodass sie kaum als Kontrollitems taugen), die Mehrdimensionalität und die Kompliziertheit sprachlicher Wendungen fallen negativ auf (Zum Beispiel Item 49 aus dem Schülerfragebogen zur Qualitätsentwicklung: „Unser Direktor/unsere Direktorin bemüht sich, auch uns Schüler/innen zu verstehen, und sucht immer eine faire Lösung, wenn wir einmal mit einem Problem nicht zurecht kommen.“ Verwirrend ist außerdem die Anordnung und Anzahl der „Smiley“-Skalen bzw. der dazugehörigen Abbildungen (fünf Smileys für eine 4-Punkt-Skala).

Die Entwicklung eines Schulprogramms und die Erhebung des Ist-Zustandes zur Schulqualität liefern zweifellos wichtige Beiträge zur Schulentwicklung und Qualitätssicherung. Nach Durchsicht der bisher vorliegenden Dokumente drängt sich aber der Eindruck auf, dass das Q.I.S.-System die Schulen zur Selbstevaluation befähigen soll, ohne alle dafür notwendigen Hilfestellungen bereitzustellen.

#### *Schulevaluation im Kanton Luzern*

In der Schweiz wurden in den letzten Jahren in den einzelnen Kantonen eine Reihe von Schulevaluations- und Qualitätssicherungsinstrumenten entwickelt. Im Folgenden soll am Beispiel von Luzern, einem Kanton mit 350.000 Einwohnern und 250 Volksschulen, die dabei gewählte Vorgehensweise beschrieben werden. Schon zu Beginn der 1990er Jahre dachte man in Luzern über eine Reform der kantonalen Volksschulen nach. Das Projekt „Schulen mit

Profil“ wurde begonnen. An diesem Projekt beteiligten sich alle kantonalen Volksschulen. Es ist beispielhaft für die zurzeit laufenden Schulreformvorhaben in der Schweiz und wurde sehr detailliert begleitet sowie von der Universität Zürich evaluiert (Büeler, 2004). Zu Beginn des Jahres 2000 wurden im Kanton Luzern die Aufgaben des Schulinspektorates für Volksschulen im Rahmen des Reformprojektes sukzessive durch ein ganzheitliches Qualitätsmanagement ersetzt, erweitert und verschiedenen Instanzen übertragen. Die kantonale Fachstelle für Schulevaluation wurde gebildet und führt die externe Evaluation der einzelnen Schulen und des gesamten Volksschulsystems durch. Gymnasien werden zentral von der Universität Zürich evaluiert, Volksschulen von der Schulaufsicht der einzelnen Kantone. Sie erstattet den Schulen, den Schulpflegern und dem zuständigen Departement periodisch Bericht. Hervorzuheben ist am Modell „Schulen mit Profil“, dass die externe Evaluation eingebettet ist in ein Qualitätssicherungsprogramm an den Einzelschulen. Hierzu gehören beispielsweise die Bildung einer Q-Gruppe, Selbstevaluationen, Lehrerbeurteilungen und Weiterbildungen. Basierend auf dem Vortrag von Kramis (2004) soll die Vorgehensweise der externen Evaluation skizziert werden. Die Abbildung 1.3 zeigt die Qualitätsbereiche, die in den Schulen erfasst werden.

Abbildung 3 Qualitätsbereiche in der Schulevaluation des Kantons Luzern

<b>Was evaluieren wir? Q-Bereiche, Q-Ansprüche FSE LU</b> 			
<small>Basierend auf: Q-Ansprüchen von FSE LU, INIS, NisKM, QZE u.a. Zusammenstellung von: Jo Kramis, FSE LU. Stand: 20.9.2004</small>			
<b>Fachstelle für Schulevaluation</b>	<b>1. Schul- und Unterrichtsklima</b> <ul style="list-style-type: none"> <li>☐ Die Schüler/innen fühlen sich in Schule ...</li> <li>☐ wohl, gut aufgenommen, integriert</li> <li>☐ sicher vor Gewalt, gerecht behandelt</li> <li>☐ ernst genommen, unterstützt</li> <li>☐ ihren Fähigkeiten gemäss gefördert</li> <li>☐ Lehrpersonen:</li> <li>☐ Wohlbefinden in Schule und Unterricht</li> <li>☐ Entfaltung beruflich und menschlich</li> <li>☐ Identifikation mit Schule</li> <li>☐ Verhältnis Belastung / Befriedigung</li> <li>☐ Arbeitszufriedenheit, Berufszufriedenheit</li> </ul>	<b>2. Erfüllung Bildungs-, Erz'auftrag</b> <ul style="list-style-type: none"> <li>☐ Schule und Unterricht fördern ausgewogen kognitive, affektive, soziale, instrumentelle Kenntnisse, Fähigkeiten, Fertigkeiten:</li> <li>☐ Persönlichkeitsbildung</li> <li>☐ Fachkompetenz</li> <li>☐ Sozialkompetenz, Selbstkompetenz</li> <li>☐ Lern- und Methodenkompetenz</li> <li>☐ Praktische Kompetenzen</li> <li>☐ Schlüsselkompetenzen, Lernen fürs Leben</li> <li>☐ d. Schüler/innen erreichen die Ergebnisse, die ihren persönlichen Möglichkeiten entsprechen (Schulerfolg, Bildungsweg)</li> </ul>	<b>3. Qualität von Lehren u. Lernen</b> <ul style="list-style-type: none"> <li>☐ Bedeutsamkeit, Wirksamkeit, Effizienz</li> <li>☐ bedeutsame Ziele und Inhalte</li> <li>☐ klare Orientierung an Zielen</li> <li>☐ Methodenvielfalt, Schüleraktivierung, Selbsttätigkeit, Medieneinsatz</li> <li>☐ Nutzung der Zeit, time on task</li> <li>☐ Prüfen und beurteilen</li> <li>☐ angemessenes Fördern und Fordern</li> <li>☐ Individualisierung</li> <li>☐ Fächerverbindung, Hausaufgaben</li> <li>☐ Klassenführung</li> </ul>
	<b>4. Führung und Management (+QM)</b> <ul style="list-style-type: none"> <li>☐ Führung fachlich, pädagogisch, menschlich</li> <li>☐ Gemeinsame Werte, Ziele, tragfäh. Konsens</li> <li>☐ Regelung von Aufgaben und Kompetenzen</li> <li>☐ Organisation, Administration</li> <li>☐ Personalführung, Entscheidungsprozesse</li> <li>☐ Einbezug der Beteiligten u. Betroffenen</li> <li>☐ Informationsfluss intern und extern</li> <li>☐ Qualitätsmanagement:</li> <li>☐ Steuerung der Q-Prozesse d. Schulleitung</li> <li>☐ Individualfeedback u. ind. Q-Entwicklung</li> <li>☐ Praxis der Selbstevaluation, Professionalität</li> <li>☐ Grav. Mängel erkannt, angegangen, beseitigt</li> <li>☐ Schule pflegt Stärken, behebt Schwächen</li> </ul>	<b>5. Schulkultur u. Schulgemeinschaft</b> <ul style="list-style-type: none"> <li>☐ Kommunikationskultur: Wertschätzung, Vertrauen, Offenheit, Konfliktregelung, Feedbackkultur, Streitkultur, bei Entscheiden beste Lösung im Zentrum</li> <li>☐ Schulgemeinschaft</li> <li>☐ gegenseitige Unterstützung</li> <li>☐ gute Beziehungen zwischen L-S, S-S, L-L</li> <li>☐ gute Zusammenarbeit, Kooperation</li> <li>☐ gemeinsame Aktivitäten, Projekte, Anlässe</li> <li>☐ häufiges Lob für Einsatz und Leistungen</li> <li>☐ klare Verhaltensregeln, konsequent eing.</li> <li>☐ Schule - ein einladender, freundlicher Ort</li> </ul>	<b>6. Inputqualitäten (Rahmenbedingungen)</b> <ul style="list-style-type: none"> <li>☐ Vorgaben von Gemeinde und Schule:</li> <li>☐ Leitbild, Leistungsauftrag</li> <li>☐ Qualitätskonzept, Schulprogramm</li> <li>☐ Schulorganisation, Std'plan, Zeitgefässe</li> <li>☐ Stütz- u. Förderangebote, Zusatzang.</li> <li>☐ Voraussetzungen:</li> <li>☐ Zusammensetzung des Kollegiums</li> <li>☐ Zusammensetzung, Erstsprache, Kulturkreis, Bildungshgr. Lernende/Elt</li> <li>☐ Ressourcen: Schulausstattung, Infrastruktur, Hilfsmittel, Medien, Informatik</li> <li>☐ Finanzen, Mittelverwendung, -nutzung</li> </ul>
	<small>© JoK, FSE LU 9/2004</small>		

Quelle: Kramis, 2004

Die bisherigen Erfahrungen basieren auf der Evaluation von Schulen, die sich freiwillig gemeldet haben. Bevor eine Evaluation beginnen kann, gibt es im Kanton eine Reihe von Vorabsprachen mit den Schulen. Sie reichen Materialien (Portfolios) und eine Stärken-Schwächen-Analyse ein. Erst danach werden Instrumente ausgewählt oder entwickelt, die in der Evaluation zum Einsatz kommen. Insgesamt dauert eine Schulevaluation sechs bis acht

Wochen. Die verwendeten Methoden reichen von der Dokumentenanalyse des Leitbildes und den Konzepten für Qualitätsmanagement und Öffentlichkeitsarbeit der Schule, einem Schulrundgang bis hin zu Befragungen und Unterrichtsbeobachtungen. Der Zeitaufwand für die beteiligten Lehrerinnen und Lehrer wird mit acht Stunden angegeben (für Stärken-Schwächen-Analysen, Interviews/Fragebögen und Berichtsvorstellung). Alle Lehrkräfte einer Schule werden schriftlich mit einem standardisierten Fragebogen befragt. Zusätzlich werden Gruppendiskussionen durchgeführt. Im Fragebogen werden beispielsweise die Freude am Unterrichten, das Wohlbefinden, die Entfaltung, die Identifikation mit der Schule, die Arbeitsbelastungen und die Berufszufriedenheit abgefragt. Andere Themen sind die „Schulkultur“ und die „Schulgemeinschaft“. Es wird ein siebenstufiges Antwortmodell vorgegeben. Hinzu kommt, dass bei jeder Aussage eingeschätzt werden soll, wie nötig eine Veränderung in diesem Bereich ist. Aus den Angaben der Lehrkräfte werden zu jeder Aussage Verteilungen, Mittelwerte und eine Qualitätsstufe (Sterne) berechnet. Aus den Einzelitems werden semantische Profile berechnet. Analog dazu wird eine Schüler- und Elternbefragung durchgeführt. Schülerinnen und Schüler werden ab der 4. Klasse einbezogen. Es werden mehrsprachige Fragebögen eingesetzt. Die bisherigen Rücklaufquoten sind mit 85-100% als sehr hoch anzusehen. Nach Abschluss aller Erhebungen wird ein Bericht erstellt, der auch an den Schulen vorgestellt wird. Dabei entscheidet die Schule, wer zur Berichtspräsentation eingeladen wird. Pro Qualitätsbereich werden fünf bis sieben Empfehlungen gegeben, von denen zwei Empfehlungen umgesetzt werden müssen.

## 2.7 Zusammenfassung

Wenn Schulen in einer Gesellschaft Anerkennung erringen wollen, dann müssen sie ihre spezifischen Leistungen nachweisen. Am besten geschieht dies mit überzeugenden Daten (Oelkers, 2000). Hinzu kommt aber noch ein zweiter Gesichtspunkt. Begriffe wie Schulevaluation und Qualitätsmanagement sind populär geworden. Sie haben in bildungspolitische Debatten genauso Einzug gehalten wie in die Schulgesetze vieler Bundesländer. Doch auch die besten politischen Ziele und Normen nutzen wenig, wenn man sie nicht mit Leben erfüllt. Wie aber sieht der Alltag an deutschen Schulen aus? Auch um diese Frage beantworten zu können, braucht man Daten, die in regelmäßigen Abständen erhoben werden und die die erreichten Leistungen der Schulen transparent machen sowie ihre Stärken und Schwächen aufzeigen.

In den letzten 50 Jahren wurden eine Reihe von Verfahren entwickelt, um die Qualität einer Schule zu bestimmen. Die bisherigen Forschungsarbeiten legen nahe, verschiedene Ebenen von Schulqualität zu unterscheiden (Fend, 1998; Holtappels, 2003). Die kleinste Einheit ist die Schulklasse oder Lerngruppe. In ihr stehen die Unterrichtsqualität und das Klassenklima im Mittelpunkt. Die Unterrichtsqualität zu messen gelingt am besten durch eine Kombination von Schülerfeedbacks, Lehrerselbsteinschätzungen und Beobachtungen. Während Hospitationen auf eine lange Tradition im deutschen Schulwesen zurückblicken können, wurden die Schülermeinungen lange Zeit nicht ernstgenommen. Schülerwahrnehmungen können jedoch ein wichtiger Ansatz zur Verbesserung der Unterrichtsqualität sein, wie neuere Forschungen belegen (Clausen, 2001; Ditton, 2002a). Zur Analyse des Klassenklimas gibt es Fragebögen, welche die Sichtweise der Schülerinnen und Schüler erfassen. Arbeiten, die das Klassenklima analysieren, sind relativ zahlreich. Im Zuge dieser Forschung wurden Instrumente entwickelt, die mittlerweile gut bewährt sind. Dabei wurde die Wechselwirkung zwischen Klassenklima, Schülerleistung und Devianz untersucht. Während es einen klaren Zusammenhang zwischen Schulklima und deviantem Verhalten gibt, ist jener zwischen Schülerleistung und Klima eher schwach ausgeprägt.

Schulen stellen nicht nur die Gebäude und Räume zur Verfügung, in denen unterrichtet wird, sondern Schulen haben auch ein individuelles Profil. Schulen haben eine eigene



Leistungskultur oder Schwerpunktsetzung: Hier ein musikalisches Gymnasium, dort eine berufsvorbereitende Hauptschule. In beiden sind die Leistungsanforderungen ganz unterschiedlich, aber klassenübergreifend definiert. Deshalb wird eine zweite Qualitätsebene eingeführt – die der Schule. Eine Schule ist aber nicht nur dann eine gute Schule, wenn sie eine gute „Wissensvermittlungsanstalt“ ist, sondern wenn sie den Schülerinnen und Schüler Handlungsspielräume im Rahmen eines pädagogisch gestalteten Schullebens bietet. Die Schuleffektivitätsforschung hat eine Reihe von Bedingungen identifiziert, die eine gute Schule ausmachen. Auch wenn sich hier im Detail Unterschiede ergeben, gibt es doch in den Grundannahmen eine relativ hohe Übereinstimmung. Von besonderer Bedeutung sind in diesem Zusammenhang Organisations-, Arbeits- und Schulklima und die Möglichkeiten zur Partizipation. Den Mikrokosmos Schule in allen Facetten zu analysieren wird kaum gelingen. Wohl aber kann man durch Befragungen verschiedener Personengruppen – von der Schülerschaft bis hin zum Hausmeister – Indikatoren der Schulqualität erfassen. Sie geben Aufschluss über die Befindlichkeiten (Ist-Zustand) und die Wünsche (Soll-Zustand) der am Schulprozess Beteiligten sowie deren Bedeutung.

Die dritte Ebene, die Fend als Systemqualität bezeichnet, umfasst die Rahmenbedingungen des Lernens an Schulen. Dazu gehören die Schulgesetze, die Bildungsziele definieren, die Bildungsstandards, die Schulaufsicht, die Lehrerausbildung und -beschäftigung, die Schuldauer, die Schulabschlüsse etc. All dies sind Faktoren, auf die Schulen kaum Einfluss haben, die aber maßgeblich auf ihre Arbeit einwirken. Wird zum Beispiel die Autonomie der Einzelschule erhöht, so zieht das die Notwendigkeit einer stärkeren externen Kontrolle nach sich. In diese Richtung geht die Etablierung des sogenannten Schul-TÜVs. Besonders in dem beschriebenen Beispiel aus Schleswig-Holstein konnte gezeigt werden, dass neben den Leistungen (Vergleichsarbeiten in den Fächern Mathematik und Deutsch) die Zufriedenheit der Schülerschaft, der Eltern und Lehrerinnen und Lehrer zu einem zentralen Qualitätsindikator werden.

Wenn von Schulqualität gesprochen wird, müssen also alle drei Ebenen berücksichtigt werden. Es scheint allerdings, dass die für den Bildungssektor verantwortlichen Politikerinnen und Politiker in der Evaluation von Schülerleistungen den Königsweg der Qualitätssicherung von Schulen sehen (Melzer, 1997). Wenn Schulqualität auf die Fachleistungen in den Kernfächern Mathematik, Deutsch und in den Naturwissenschaften reduziert wird, verliert man die Prozessmerkmale des Lernens aus dem Blick. Das Lernumfeld kann mittlerweile valide erfasst werden und wird in neuere integrative Verfahren einbezogen. Ein Beispiel dafür ist PISA. Hier werden gleichzeitig Schülerleistungen und Kontextfaktoren empirisch erhoben.

Integrative Ansätze wie PISA wären also ein geeignetes Instrument, um Schulqualität zu analysieren. Die Verbindung von Leistungsmessung mit sozialen Komponenten der Schulqualität ergibt ein Gesamtbild von Schule. Dies erfordert allerdings ein hohes Maß an Ressourcen. Für die PISA-Datenerhebungen wurden an jeder teilnehmenden Schule zwei Unterrichtstage benötigt. Die Gesamtkosten für Datenerhebung und -analyse für 1.500 Schulen gingen in die Millionen. Repräsentative Stichproben sind zwar für die Bildungsberichterstattung wichtig, sie helfen aber nicht den anderen 45.000 Schulen, die nicht an der PISA-Studie teilgenommen haben: Nur wenn jede einzelne Schule eine genaue Rückmeldung über ihre Stärken und Schwächen erhält, kann sie ihre Schulqualität verbessern. Die Lösung dieses Dilemmas liegt darin, die Leistungen der Schülerinnen und Schüler in landesweiten Vergleichsarbeiten zu erfassen (im Sinne einer Selbstevaluation) und diese in Bezug zu setzen zu den Kontextbedingungen des Lernens. In allen Qualitätskonzepten, wie sie zum Beispiel in der Schuleffektivitätsforschung entwickelt wurden, spielt die Selbstevaluation eine zentrale Rolle. In Deutschland, wo Schulen, die sich selbst evaluieren und ein umfassendes System des Qualitätsmanagement entwickelt haben, eher die Ausnahme als die Regel sind, erscheinen hierzu Hilfestellungen notwendig. So bietet beispielsweise das



baden-württembergische EIS (Evaluationsinstrumente für Schulen) neben Lernstandsdiagnostikinstrumenten auch Fragebögen zum Unterricht, zum Schulklima, zur Schulführung an. Ob diese Selbstevaluationen aber gelingen, hängt jedoch (zumindest am Anfang) ganz entscheidend von der Unterstützung der Einzelschulen ab.

Auch wenn der Effekt der verschiedenen Dimensionen von Schulqualität auf das Lernverhalten und die Leistung vergleichsweise beschränkt ist, ist es doch die einzige Einflussmöglichkeit, die die Schule selbst hat. Weder die kognitiven Eingangsvoraussetzungen noch die familialen Bedingungen können durch die Schule signifikant beeinflusst werden. Zahlreiche Forschungen belegen, dass Schulqualität einen wichtigen Beitrag zur Sozialisation von jungen Menschen leistet. Kinder und Jugendliche sollen sich an Schulen wohlfühlen und dort soziale Kompetenzen erwerben. So kann Schulqualität zur Prävention delinquenten Verhaltens beitragen. Bildungsstandards werden vielleicht in absehbarer Zeit das Abschneiden von deutschen Kindern und Jugendlichen in internationalen Schulleistungstudien verbessern, aber sie werden nichts zum Abbau sozialer Ungleichheit in der deutschen Gesellschaft beitragen.

## 2.8 Literatur

- Aikin, W. M. (1942). *The story of the Eight-Year-Study*. New York/London: Progressive Education Association Publications: Adventure in American Education.
- Arnold, K.-H. (2002). Schulentwicklung durch Rückmeldung der Lernwirksamkeit an der Einzelschule: Möglichkeiten und Grenzen der Schuleffizienzforschung. *Zeitschrift für Pädagogik*, 48, 741-764.
- Aurin, K. (1990). *Gute Schulen – worauf beruht ihre Wirksamkeit?* Bad Heilbrunn: Klinkhardt.
- Anderson, G. J. (1973). *The assessment of learning environments: A manual for the learning environment inventory and my class inventory*. Halifax: Atlantic Institute of Education.
- Anderson, G. J. & Walberg, H. J. (1974). *Learning environments*. In H. J. Walberg (ed.), *Evaluating educational performance*. Berkeley: McCutchan Publishing Corporation.
- Anderson, L. W., Ryan, D. W. & Shapiro, B. J. (eds.). (1989). *The IEA classroom environmental study*. Oxford: Pergamon Press.
- Baumert, J. (2001). Internationale Schulleistungsvergleiche. In D.H. Rost (Hrsg.), *Handwörterbuch Pädagogische Psychologie* (2. Auflage) (S. 294- 303). Weinheim: Beltz.
- Baumert, J. & Lehmann, R. et al. (1997). *TIMSS – Mathematisch-naturwissenschaftlicher Unterricht im internationalen Vergleich. Deskriptive Befunde*. Opladen: Leske + Budrich.
- Baumert, J., Stanat, P. & Demmrich, A. (2001). *PISA 2000: Untersuchungsgegenstand, theoretische Grundlagen und Durchführung der Studie*. In Deutsches PISA-Konsortium (Hrsg.), *PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich* (S. 15-68). Opladen: Leske + Budrich.
- Baumert, J. et al. (2002). *Untersuchungsgegenstand, Fragestellung und technische Grundlagen der Studie*. In: Deutsches PISA-Konsortium (Hrsg.). *PISA 2000 – Die Länder der Bundesrepublik Deutschland im Vergleich*. Opladen: Leske + Budrich.
- Berner, B. (2002). Die 'Fragen über den Schul-Unterricht' von 1771 oder Die erste Evaluation des Zürcher Schulwesens. *Zeitschrift für pädagogische Historiographie*, 8 (2), 128–136.
- Bertelsmann-Stiftung (2004). *Einführung: Netzwerke innovativer Schulforschung in Deutschland*. Gütersloh: Bertelsmann-Stiftung. [www.bertelsmann-stiftung.de](http://www.bertelsmann-stiftung.de) (17.10.2004).
- Bessoth, R. (1989). *Die Verbesserung des Unterrichtsklimas*. Neuwied: Luchterhand.
- Binet, A. & Simon, Th. *Methodes nouvelles pour le diagnostique du niveau interllectuel des anormaux*. *Annee Psychologique* 1905, 11, 191-244.
- Bloom, B. S. (1971). *Individual differences in school achievement*. Annual Meeting of American Educational Research Association. New York.

- Bos, W. & Postlethwaite, T. N. (2002). Möglichkeiten, Grenzen und Perspektiven internationaler Schulleistungsforschung. In R. Tippelt (Hrsg.), *Handbuch Bildungsforschung* (S. 241-261). Opladen: Leske + Budrich.
- Büeler, X. (2004). Chancen und Grenzen von „Schulen mit Profil“. *FS&S aktuell*. Ausgabe 6/04. Zürich: Forschungsbereich Schulqualität & Schulentwicklung.
- Bundesministerium für Bildung, Wissenschaft und Kultur (2004). *Q.I.S. Qualität in Schulen. Leitfaden*. Wien.
- Carroll, J. B. (1963). A model of school learning. *Teachers College Record*, 64, 723-733.
- Central Advisory Council for Education (1967). *Children and their Primary Schools ('The Plowden Report')*, London: HMSO.
- Clausen, M. (2001). *Unterrichtsqualität: Eine Frage der Perspektive?* Münster: Waxmann.
- Clausen, M., Schnabel, K. & Schröder, S. (2002). Konstrukte der Unterrichtsqualität im Expertenurteil. *Unterrichtswissenschaft*, 30, 245-260.
- Coleman, J. S., Campbell, E. Q., Hobson, C. J. et al. (1966). *Equality of educational opportunity*. Washington: US Government Printing Office.
- Der Spiegel (2004). Die Rache der Igel. 20, 73.
- Ditton, H. (2002a). Lehrkräfte und Unterricht aus Schülersicht. Ergebnisse einer Untersuchung im Fach Mathematik. *Zeitschrift für Pädagogik*, 48, 262-286.
- Ditton, H. (2002b). Unterrichtsqualität – Konzeptionen, methodische Überlegungen und Perspektiven. *Unterrichtswissenschaft*, 30, 197-212.
- Dreesmann, H. R. (1979). *Das Unterrichtsklima als situative Bedingung für kognitive Prozesse und das Leistungsverhalten von Schülern*. Dissertation im Fach Psychologie. Wirtschafts- und Sozialwissenschaftliche Fakultät der Universität Heidelberg.
- Dreesmann, H. R. (1982). *Unterrichtsklima. Wie Schüler den Unterricht wahrnehmen*. Weinheim: Beltz.
- Edelstein, W. & Fauser, P. (2001). *Demokratie lernen und leben. Gutachten zum Programm. (Materialien zur Bildungsplanung und Forschungsförderung, Heft 96)*. Bonn: Bund-Länder-Kommission für Bildungsplanung und Forschungsförderung (BLK).
- Eder, F. (1995). Das Befinden von Schülerinnen und Schülern in den öffentlichen Schulen - Ergebnisse der Repräsentativerhebung. In F. Eder (Hrsg.), *Das Befinden von Kindern und Jugendlichen in der Schule* (S. 24-168). Innsbruck: StudienVerlag..
- Eder, F. (1996). *Schul- und Klassenklima. Ausprägung, Determinanten und Wirkungen des Klimas an weiterführenden Schulen*. Innsbruck: StudienVerlag.
- Eder, F. (1998). *Linzer Fragebogen zum Schul- und Klassenklima für die 8.-13. Klasse*. Göttingen: Hogrefe.
- Eder, F. (2001). Schul- und Klassenklima. In D. H. Rost (Hrsg.), *Handwörterbuch Pädagogische Psychologie (2. Auflage)* ( S. 578- 586 ). Weinheim: Beltz.
- Eder, F. & Mayr, J. (2000). *Linzer Fragebogen zum Schul- und Klassenklima für die 4.-8. Klasse*. Göttingen: Westermann.
- Eder, F. (2002). Unterrichtsklima und Unterrichtsqualität. *Unterrichtswissenschaft*, 30, 213-229.
- Eikenbusch, G. (2001). Erfahrungen mit der Schülerrückmeldung in der Oberstufe. *Pädagogik*, 53 (5), 18-22.
- Einsiedler, W. (2002). Das Konzept "Unterrichtsqualität". *Unterrichtswissenschaft*, 30 (3), 194-196.
- Fend, H. (1977). *Schulklima*. Weinheim: Beltz.
- Fend, H. (1982). *Gesamtschule im Vergleich. Bilanz der Ergebnisse des Gesamtschulversuchs*. Weinheim: Beltz
- Fend, H. (1986). „Gute Schulen - schlechte Schulen“. *Die einzelne Schule als pädagogische Handlungseinheit. Die deutsche Schule*, 79, 275-293.
- Fend, H. (1989). Was ist eine gute Schule? In K.-J. Tillmann (Hrsg.), *Was ist eine gute Schule?* Hamburg: Bergmann + Helbig.

- Fend, H. (1998). *Qualität im Bildungswesen. Schulforschung zu Systembedingungen, Schulprofilen und Lehrerleistungen*. Weinheim: Juventa.
- Fiala-Thier, I. (2001). *Qualität – die Chance für Schulen: ein Handbuch*. Wien: Manz.
- Fisher, D. L. & Waldrup, B. G. (1997). A culturally sensitive learning environment questionnaire for use in science classrooms. *Research in Science Education*, 27, 41-49.
- Fraser, B. J. (1985). *Individualized classroom environment questionnaire*. Melbourne: Australian Council for Educational Research.
- Fraser, B. J. & Fisher, D. L. (1997). Using short forms of classroom climate instruments to assess and improve classroom psychosocial environment. *Journal of Research in Science Teaching*, 23, 387-413.
- Fraser, B. J., Giddings, G. G. & McRobbie, C. J. (1995). Evolution and validation of a personal form of an instrument for assessing science laboratory classroom environments. *Journal of Research in Science Teaching*, 32, 399-422.
- Fraser, B. J. (1999). Using learning environment assessments to improve classroom and school climates. In H. J. Freiberg (ed.), *School climate. Measuring, improving and sustaining healthy learning environments* (S. 65-83). London: Falmer Press.
- Freiberg, H. J. (1999). Introduction. In Freiberg, H. J. (ed.), *School climate: Measuring, improving and sustaining healthy learning environments* (S. 1-10). London: Falmer Press.
- Friedenburg, L.v. (1989). *Bildungsreform in Deutschland*. Frankfurt a.M.: Suhrkamp.
- Gage, N. L. (1963). *Handbook of research on teaching*. Chicago: Rand McNally.
- Goh, S. C. & Fraser, B. J. (1996). Validation of an elementary school version of the questionnaire on teacher interaction, *Psychological Reports*, 64, 29-40.
- Goltz, E. & Landua, D. (2003) *Entwicklung der „Sozialen Schulqualität“*. Ergebnisse eine Zeitreihenuntersuchung an 10 Brandenburger Schulen im Rahmen des Projekts „Schulqualität als Präventionsstrategie“. Abschlussbericht.
- Good, T. L., Biddle, B. J. & Brophy, J. E. (1975). *Teachers make a difference*. New York: Holt, Rinehart & Winston.
- Gruehn, S. (2000). *Unterricht und schulisches Lernen. Schüler als Quellen der Unterrichtsbeschreibung*. Münster: Waxmann.
- Heller, K. A. & Hany, E. A. (2001). Standardisierte Schulleistungsmessungen. In F. E. Weinert (Hrsg.), *Leistungsmessungen in Schulen* (S. 87-101). Weinheim: Beltz.
- Helmke, A. (2002). Kommentar: Unterrichtsqualität und Unterrichtsklima: Perspektiven und Sackgassen. *Unterrichtswissenschaft*, 30, 261-277.
- Helmke, A. (2003). Unterrichtsqualität erfassen, bewerten, verbessern. Seelze: Kallmeyer.
- Helmke, A. & Jäger, R. S. (2002). *Die Studie MARKUS - Mathematik-Gesamterhebung Rheinland-Pfalz: Kompetenzen, Unterrichtsmerkmale, Schulkontext*. Landau: Verlag Empirische Pädagogik.
- Helmke, A., Schneider, W. & Weinert, F. E. (1986). Quality of instruction and classroom learning outcomes – Results of the German contribution to the classroom environment study of the IEA. *Teaching and Teacher Education*, 2, 1-18.
- Helmke, A. & Schrader, F.-W. (1998). Determinanten der Schulleistung. In D. H. Rost (Hrsg.), *Handbuch Pädagogische Psychologie* (2. Auflage) (S. 183-187). Weinheim: Beltz..
- Helmke, A. & Schrader, F.-W. (2001). Jenseits von TIMSS: Messungen sprachlicher Kompetenzen, komplexe Längsschnittstudien und kulturvergleichende Analysen. Ergebnisse und Perspektiven ausgewählter Längsschnittstudien. In F. E. Weinert (Hrsg.), *Leistungsmessungen in Schulen* (S. 237-250). Weinheim: Beltz.
- Helmke, A. & Weinert, F. E. (1997). Bedingungsfaktoren schulischer Leistungen. In F. E. Weinert (Hrsg.), *Psychologie des Unterrichts und der Schule. Enzyklopädie der Psychologie, Serie Pädagogische Psychologie*, Bd. 3 (S. 71-176). Göttingen: Hogrefe.
- Holtappels, H.-G. (2003). *Schulqualität durch Schulentwicklung und Evaluation. Konzepte, Forschungsbefunde, Instrumente*. München: Luchterhand.

- Horstkemper, M. & Tillmann, K.-J. (2004). Schulformvergleiche und Studien zu Einzelschulen In: W. Helsper & J. Böhme (Hrsg.), *Handbuch der Schulforschung* (S. 287-323). Wiesbaden: Verlag für Sozialwissenschaften.
- Institut für Schulentwicklungsforschung (1999). IFS Schulbarometer. Ein mehrperspektivisches Instrument zur Erfassung der Schulwirklichkeit. Dortmund: IFS-Verlag.
- Ingenkamp, K. & Saldern, M. v. (1983). Zwischenbericht über das DFG-Projekt. In 22/10-1 ‚Schulische Umwelt und Verhalten von Schülern‘. Landau: EWH.
- Jencks, C., Smith, M., Acland, H. et al. (1972). *Inequality: A reassessment of the effect of family and schooling in America*. New York: Basic Books.
- Kahl, T. N. (1977). *Unterrichtsforschung. Probleme, Methoden und Ergebnisse der empirischen Untersuchung unterrichtlicher Lernsituationen*. Kronberg/Ts.: Scriptor.
- Kempfert, G. & Rolff, H.-G. (2002). *Pädagogische Qualitätsentwicklung. Ein Arbeitsbuch für Schule und Unterricht* (3. Aufl.). Weinheim: Beltz.
- Klieme, E., Abs, H.-J. & Diedrich, M. (2004). Evaluation im BLK-Modellprogramm Demokratie lernen und leben: Musterrückmeldungen und erste deskriptive Befunde aus der Eingangserhebung 2003. Frankfurt a.M.: Deutsches Institut für Internationale Pädagogische Forschung.
- Kramis, J. (2004). Externe Evaluation des Kts. Luzern als Teil eines ganzheitlichen Q-Management. Vortrag auf der Herbsttagung der KBBB am 4./5.10.2004 in Dortmund.
- Kunter, M., Schümer, G., Artelt, C. et al. (2001). PISA 2000. Dokumentation der Erhebungsinstrumente. Berlin: Max-Planck-Institut für Bildungsforschung.
- Lehmann, R. H. & Peek, R. (1997). Aspekte der Lernausgangslage von Schülerinnen und Schülern der fünften Klassen an Hamburger Schulen. Bericht über die Untersuchung im September 1996. Hamburg.
- Lehmann, R. H., Peek, R., Gänsfuß, R. & Husfeldt, V. (2002). Aspekte der Lernausgangslage und der Lernentwicklung – Klassenstufe (LAU 9). Ergebnisse der Längsschnittuntersuchung in Hamburg. Hamburg: Behörde für Bildung und Sport, Amt für Schule.
- Lehmann, R. H. et al. (2000). QuaSUM. Qualitätsuntersuchung an Schulen zum Unterricht in Mathematik. Ergebnisse einer repräsentativen Untersuchung im Land Brandenburg (Reihe Schulforschung in Brandenburg, Heft 1) Potsdam.
- Leiske, M., Sturzbecher, D. & Keil, J.-G. (2001). Soziale Schulqualität aus der Sicht von Jugendlichen in Brandenburg. In D. Sturzbecher (Hrsg.), *Jugend in Ostdeutschland - Lebenssituationen und Delinquenz* (S. 210 – 248). Opladen: Leske + Budrich.
- Lüders, M. & Rauin, U. (2004). Unterrichts- und Lehr-Lern-Forschung. In W. Helsper & J. Böhme (Hrsg.), *Handbuch der Schulforschung* (S. 691-719). Wiesbaden: Verlag für Sozialwissenschaften.
- Mayer, J. K. & Nickolaus, R. (2003). Der Unterrichts-Beurteilungsbogen zur Bewertung von Unterricht durch Schüler. <http://www.uni-stuttgart.de/bwt> (20.9.2004).
- McRobbie, C. J. & Fraser, B. J. (1993). Associations between student outcomes and psychosocial science laboratory environments. *Journal of Educational Research*, 87, 78-85.
- Melzer, W. (1997). Der Beitrag der Schulforschung zur Qualitätssicherung und Entwicklung von Schule. *Erziehungswissenschaft*, 8, 16-21.
- Ministerium für Bildung, Forschung und Kultur des Landes Schleswig-Holstein (2004). Externe Evaluation im Team - EVIT -. *Qualitätshandbuch für die Arbeit an Schulen in Schleswig-Holstein*. Kiel.
- Moos, R. H. (1979). *Evaluating educational environments: Procedures, measures, findings, and policy implications*. San Francisco: Jossey-Bass.
- Moos, R. H. & Trickett, E. H. (1987). *Classroom environment scale manual* (2nd ed.). Palo Alto, CA: Consulting Psychologists Press.
- Murray, H. A. (1938). *Explorations in personality*. New York: Oxford University Press.
- National Center for Education Statistics (2000). *Monitoring school quality: An indicators Report*. Washington, D.C.: U.S. Department of Education

- Niedersächsisches Kultusministerium (2003a). Orientierungsrahmen Schulqualität in Niedersachsen. Qualitätsbereiche und Qualitätsmerkmale guter Schulen. Hannover.
- Niedersächsisches Kultusministerium (2003b). Von anderen Nationen lernen – Beurteilung schulischer Qualität in Niedersachsen. Hannover.
- OECD (1989). Schools and quality. An international report. Paris: OECD.
- Oelkers, J. (2000). Welche Effizienz braucht Schule? Vortrag auf dem 8. Forum Gymnasium der Gewerkschaft Erziehung und Wissenschaft (GEW) Nordrhein-Westfalen am 1. September 2000 in Unna.
- Ofsted (2003). Inspecting schools. Framework for inspecting schools. London: Ofsted.
- Ofsted (2004). <http://www.ofsted.gov.uk> (24.10.2004).
- Pine, G. J. & Boy, A. V. (1977). Learner-centred teaching: A humanistic view. Denver, CO: Love.
- Renolds, D. & Teddlie, C (2000). An introduction to school effectiveness. In C. Teddlie & D. Renolds (eds.), International handbook of school effectiveness research (S. 3-25). London: Falmer Press.
- Rentoul, A. J. & Fraser, B. J. (1983). Development of a school-level environment questionnaire. *Journal of Educational Administration*, 21, 21-39.
- Roberts, D.H. (2003). Inspection Report. St. Michael's RC School. London.
- Rosenthal, R. & Jacobson, L. (1968). *Pygmalion in the classroom*. New York: Holt, Rinehart and Winston.
- Roßbach, H.-G. (2002). Unterrichtsqualität im 2. Schuljahr – Ergebnisse einer empirischen Untersuchung. *Unterrichtswissenschaft*, 30, 230-245.
- Rutter, M., Maughan, B., Mortimore, P. & Ouston, J. (1979). Fifteen thousand hours. Secondary schools and their effects on children. Reissued in 1994. London: Paul Chapman.
- Saldern, M. von (1987). Schulklima von Schulklassen. Überlegungen und mehrbenenanalytische Untersuchungen zur subjektiven Wahrnehmung von Lernumwelten. Frankfurt/Main: Peter Lang.
- Saldern, M. v. & Littig, K. E. (1987). Die Konstruktion der Landauer Skalen zum Sozialklima (Lasso). *Zeitschrift für Entwicklungspsychologie und pädagogische Psychologie*, 17, 138-149.
- Sammons, P., Elliot, K., Welcomme, W., Taggart, B. & Levacic, R. (2004). England. In H. Döbert, E. Klieme & W. Sroka (eds.), *Conditions of School Performance in Seven Countries. A quest for understanding the international variation of PISA results* (S. 65-149) Münster: Waxmann..
- Schnabel, K. U. (2001): Schuleffekte. In D. H. Rost (Hrsg.), *Handwörterbuch Pädagogische Psychologie* (2. Auflage.) (S. 586- 591). Weinheim: Beltz.
- Schwippert, K. & Bos, W. (2003). Systemmonitoring. In Arbeitsgruppe Internationale Vergleichsstudie. Vertiefender Vergleich der Schulsysteme ausgewählter PISA-Teilnehmer-staaten (S. 135-148). Bonn: Bundesministerium für Bildung und Forschung.
- Slavin, R. E. (1990). Mastery learning re-reconsidered. *Review of Educational Research*, 60, 300-302.
- Slavin, R. E. (1996). *Education for all*. Lisse: Swets & Zeitlinger.
- Standop, J. (2002). Emotionen und kognitives schulisches Lernen aus interdisziplinärer Perspektive. Frankfurt a.M.: Peter Lang.
- Stern, C., Mahlmann, J. & Vaccaro, E. (2003). Qualitätsentwicklung braucht Steuerung. Erfahrungen mit der Bestimmung, Messung und dem Vergleich von Qualität in einem internationalen Kontext. *Pädagogische Rundschau*, 57, 473-481.
- Steffens, U. & Bargel, T. (1993). *Erkundungen zur Qualität von Schule*. Neuwied: Luchterhand.
- Sturzbecher, D. & Hess, M. (2002). Soziale Schulqualität aus Schülersicht. In D. Sturzbecher (Hrsg.), *Jugendtrends in Ostdeutschland: Bildung, Freizeit, Politik, Risiken* (S. 155-181). Opladen: Leske + Budrich.
- Taylor, P. C. S., Fraser, B. J. & Fisher, D. J. (1997). Monitoring constructivist classroom learning environment. *International Journal of Educational Research*, 27 (4), 293-302.
- Tillmann, K.-J., Holler-Nowitzki, B., Holtappels, H. G., Meier, U. & Popp, U. (1999). Gewalt als Schulproblem. Verursachende Bedingungen, Erscheinungsformen und pädagogische Handlungsperspektiven. Weinheim: Juventa.

- Tillmann, K.-J. (2003). Schulqualität und Schulentwicklung. Erfahrungen, Perspektiven, Fallstricke. Vortrag auf der Jubiläumsveranstaltung des "Forums Eltern und Schule" am 28.11.2003 in Wuppertal.
- Tillmann, K.-J. & Meier, U. (2001): Schule, Familie und Freunde – Erfahrungen von Schülerinnen und Schülern. In Deutsches PISA-Konsortium (Hrsg.), PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich (S. 468-509). Opladen: Leske + Budrich.
- Walberg, H. J. & Anderson G. J. (1968). Classroom climate and individual learning. *Journal of Educational Psychology*, 59, 414-419.
- Weinert, F. E. (2001). Vergleichende Leistungsmessung in Schulen – eine umstrittene Selbstverständlichkeit. In F. E. Weinert (Hrsg.), *Leistungsmessungen in Schulen* (S.17-31). Weinheim: Beltz..
- Weinert, F. E. & Helmke, A. (1997). *Entwicklung im Grundschulalter*. Weinheim: Psychologie Verlags Union.
- Weinert, F. E., Schrader, F. W. & Helmke, A. (1989). Quality of instruction and achievement outcomes. *International Journal of Educational Research*, 13, 895-914.
- Weiß, M. & Steinert, B. (2001). Institutionelle Vorgaben und ihre aktive Ausgestaltung – Die Perspektive der deutschen Schulleitungen. In Deutsches PISA-Konsortium (Hrsg.), PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich (S. 427-454). Opladen: Leske + Budrich.
- Wenzel, H. (2004). Studien zur Organisations- und Schulkulturentwicklung. In W. Helsper & J. Böhme (Hrsg.), *Handbuch der Schulforschung* (S. 391-415). Wiesbaden: Verlag für Sozialwissenschaften..

## 3 Evaluation sozialer Schulqualität – eine Serviceleistung für Schulen

D. Landua

### 3.1 Methodische Anmerkungen zur Evaluation von Schulen

#### 3.1.1 Zur Geschichte der Evaluationsforschung

Die Evaluationsforschung in den USA kann auf eine lange methodologische Tradition zurückblicken. Erste (kleine) Feldprojekte gab es im späten 19. und frühen 20. Jahrhundert beispielsweise in der Psychologie, in der Arbeitswissenschaft und in der Industriesoziologie (Mertens, 2000). Ein bedeutender Entstehungszusammenhang der Evaluationsforschung liegt auch in der staatsphilosophischen und politikwissenschaftlichen Diskussion über die Wirkungsweise und Wirksamkeit des Staates begründet. Diese eher staatstheoretische und verwaltungswissenschaftliche Diskussion ist mit der Entwicklung und Ausdifferenzierung einer empirischen Sozialforschung und ihrer zunehmenden „Indienststellung“ zugunsten einer pragmatischen und praktischen Nutzung der Evaluation weitgehend verloren gegangen. In den 50er Jahren des 20. Jahrhunderts verlagerte sich der Schwerpunkt in der Nutzung und Entwicklung auf die Evaluation von internationalen Entwicklungshilfeprojekten. So war es die UNESCO, die Ende der 50er Jahre das erste Handbuch zur Entwicklungsforschung herausgab. Doch erst mit dem Anstieg staatlicher Reformen, Aufgaben und Ausgaben in den 60er Jahren beginnt der gezielte Einsatz der Evaluationsforschung für staatliche Aufgaben und Programme. Die ersten „professionellen“ Evaluationen wurden in den 60er Jahren des 20. Jahrhunderts im Rahmen der Great-Society Gesetze durchgeführt (Stockmann, 2002).

In Deutschland setzt die Evaluationsforschung mit etwa 10 Jahren Verspätung Ende der 60er, Anfang der 70er Jahre ein, ebenfalls im Zuge politischer Reformprogramme. Reformiert werden sollte beispielsweise das Bildungswesen, Städtebau, Gesundheitswesen, Sozialwesen. Weiterhin sollten Maßnahmen zur Humanisierung der Arbeitswelt durchgeführt werden.

Prinzipiell lassen sich fünf Richtungen von Evaluation als Analyse- und Kontrollverfahren politisch-administrativen Handelns unterscheiden:

1. Evaluation als "Rückmeldeschleife" in einem Planungs- und Managementsystem,
2. Evaluation als Mittel der Kostenreduzierung,
3. Evaluation als Analysemittel zum Abbau von "Bürokratismus",
4. Evaluation als Hilfsmittel des Parlaments zur Kräftigung seiner Kontrollfunktion und
5. Evaluation als Verfahren systematischen Lernens im Rahmen von "Reformpolitik".

Vor allem Punkt 5 kommt an dieser Stelle eine besondere Bedeutung zu. Ende der 60er Jahre wurde in Deutschland, nicht zuletzt unter sozialdemokratischen Regierungen, eine Politik der inneren Reformen eingeleitet. Es entwickelte sich eine Bereitschaft, durch "Sozialexperimente" Erkenntnisse zu gewinnen, um dadurch rationale politische Entscheidungen treffen zu können. So evaluierte man bildungspolitische Experimente wie die Einführung der ersten Ganztagschulen. Dabei standen die Forscher im Spannungsfeld zwischen "Aktionsforschung" und "distanzierter Kontrollforschung", konnten einerseits experimentell-innovativ mit dem Ziel systematischen Lernens vorgehen, mussten andererseits aber auch Innovationen in einem komplizierten Handlungsfeld entwickeln und durchsetzen. Im Zuge der ersten bildungspolitischen Evaluationen kam es auch zum Paradigmenwechsel innerhalb pädagogischer Modellversuche: Das traditionelle Paradigma, die erkenntnistheoretische Position des kritischen Rationalismus, wurde weitgehend zugunsten

eines Paradigmas der Handlungsforschung aufgegeben, die nicht nur die empirische Kontrolle fordert, sondern auch Konstruktion, Optimierung und Legitimierung der Modellmaßnahme. Wesentliches methodologisches Charakteristikum war dabei die Integration von Analyse-, Konstruktions-, Kontroll- und Bewertungsleistungen im Begriff der handlungsbezogenen Kommunikation.

### **3.1.2 Schulqualität und Schulevaluation**

Regelmäßige Leistungsbilanzen, wie sie im Bereich von Industrie, Wirtschaft und Finanzen obligatorisch eingeführt sind, um den Fortbestand der betreffenden Einrichtung zu sichern, kennen öffentliche Bildungseinrichtungen in Deutschland bislang kaum. Ihre Leistungsfähigkeit wird bis heute vorwiegend im Rahmen eines administrativen Inspektionssystems, durch Erlass von Richtlinien sowie Verordnungen für die Aus- und Weiterbildung der Lehrerschaft und der Schulleitungen oder die Beteiligung der Eltern und der Schülerschaft am Schulgeschehen, vor allem aber durch eine hierarchisch gegliederte Schulaufsicht zu gewährleisten versucht. Konzepte, eine relative Schulautonomie einzuführen, den Schulen mehr Gestaltungsspielräume zuzubilligen, Elemente der Selbstevaluation zu implementieren und eine "Schulentwicklung von unten" anzustoßen, sind noch relativ selten bzw. stehen erst am Anfang ihrer Implementierung. Eine konsequente Reform des gegenwärtigen Schul- und Schulaufsichtssystems, wie sie in einigen Nachbarländern verwirklicht wurde, ist in Deutschland erst im Entstehen. Die Möglichkeiten, Evaluationen zum Instrument der Qualitätssicherung von Schulen zu machen, sind hier noch lange nicht ausgeschöpft.

Bei der exakten Evaluation von Programmen und Projekten als systematischer Untersuchung und Überprüfung ihres Wertes und Nutzens sind verschiedene Evaluationsstandards zu beachten; im Einzelnen handelt es sich dabei um (Joint Committee, 1994):

- Nützlichkeitsstandards: Diese sollen dazu beitragen, dass sich die Evaluation an den Informationsbedürfnissen der „Nutzer“ orientiert.
- Durchführbarkeitsstandards: Diese sollen gewährleisten, dass der Evaluationsprozess geplant, praxisverträglich und kosteneffizient durchgeführt wird.
- Korrektheitsstandards: Diese Standards dienen dazu, dass die Evaluation rechtlich und ethisch korrekt verläuft.
- Genauigkeitsstandards: Diese Standards sollen sicherstellen, dass dem Evaluationsprozess wissenschaftliche Erkenntnisse und Gütekriterien zugrunde liegen.

Bezogen auf den Bildungssektor beinhalten Evaluationen Verfahren zur systematischen Prüfung und Verbesserung der Qualität des Schulsystems und der Schulen. Grundsätzlich werden Überprüfungen von Schulen seit langem von der Schulaufsicht durchgeführt, und auch in den meisten Lehrerkollegien ist es nicht ungebräuchlich, dass über pädagogische Arbeitsabläufe und -ergebnisse reflektiert wird. Oft fehlen solchen Reflexionen und Überprüfungen jedoch die für Evaluationen notwendige Systematik und verlässliche Methoden.

Das wachsende öffentliche Interesse an der Qualität im Schulwesen und damit zusammenhängend an der Evaluation von Schulqualität hat sich in den letzten Jahren teilweise stark auf technische Fragen der Qualitätsmessung konzentriert. Die Frage, was eigentlich unter Schulqualität verstanden werden soll, ist dabei etwas im Hintergrund geblieben. Wie die vorangegangenen Kapitel gezeigt haben hat diese Zurückhaltung gute Gründe, weil es nämlich auf diese Frage verschiedene und einander teilweise widersprechende Antworten gibt. Worin unterscheiden sich gute Schulen von weniger guten?



An welchen Merkmalen kann man gute Schulen erkennen? Die folgende Übersicht orientiert sich an den von Peter Posch und Herbert Altrichter (1999) vorgestellten Kriterien.

*Orientierung an hohen, allen bekannten fachlichen und überfachlichen Leistungsstandards: positive Leistungserwartung und intellektuelle Herausforderung*

Gute Schulen erwarten gute Leistungen von den Schülerinnen und Schülern: Sie glauben, dass die Lernenden 'etwas zustande bringen' werden und lassen sie das auch spüren: 'Fördern und Fordern', intellektuelle Anforderungen und didaktische und individuelle Hilfestellung stehen in einer guten Balance. (Aurin, 1991). Vor allem zwei Aspekte sind dabei interessant: die Standards sollten allen bekannt sein und sowohl fachlicher als auch überfachlicher Art sein. "Allen bekannt" bedeutet, dass die Schülerinnen und Schüler im Unterricht eine möglichst klare Vorstellung von den Ansprüchen gewinnen, die in einem Fach gestellt werden. "Sowohl fachlich als auch überfachlich" bedeutet, dass neben disziplinären Standards auch das Bemühen um fächerverbindende Problemstellungen, die Schülerinnen und Schüler eine Integration des Wissens aus verschiedenen Disziplinen ermöglichen, einen hohen Stellenwert haben.

*Hohe Wertschätzung von Wissen und Kompetenz*

Je mehr an einer Schule den Schülerinnen und Schülern verdeutlicht wird, wie wichtig Wissen ist, desto eher wird offensichtlich auch den Jugendlichen die Auseinandersetzung mit Wissen als selbstverständliche Tätigkeit nahe gebracht. Allerdings sind der Schule hier Grenzen gesetzt. Die Analyse der Ergebnisse der TIMS-Studie hat z.B. gezeigt, dass Unterschiede in der gesellschaftlichen Wertschätzung von Wissen und Kompetenz in erheblichem Maße zur Erklärung der Leistungsunterschiede zwischen Schülerinnen und Schülern in Europa (auch den USA) und jenen Japans herangezogen werden können. Wissen, vor allem mathematisch-naturwissenschaftliches Wissen, wird in Japan und anderen fernöstlichen Staaten außerhalb der Schule wesentlich höher bewertet als in Europa. In der Folge erhält auch das Angebot der Schule einen höheren Stellenwert, was u. a. wieder darin zum Ausdruck kommt, dass der Anteil an unterrichtsbezogener Zeit in diesen Wissensbereichen an der insgesamt verfügbaren Zeit höher ist (Baumert & Lehmann, 1997).

*Mitsprache und Verantwortungsübernahme durch Schülerinnen und Schüler*

Die Schülerschaft ist nicht nur und auch nicht in erster Linie Kunde der Schule oder Konsument schulischer Leistungen, sondern Prozess und Output der Schule hängen in hohem Maße von der Beteiligung und „Mitarbeit“ der Schülerinnen und Schüler ab. In guten Schulen werden Schülerinnen und Schüler explizit als „Mitproduzenten" schulischer Leistungen angesehen. Aus dieser Haltung heraus wird der Schülerschaft Mitverantwortung für die Gestaltung des Lebens an der Schule zugewiesen.

*Wertschätzende Beziehungen zwischen Leitung, Lehrerkollegium und Schülerschaft*

In guten Schulen werden die Schülerinnen und Schüler respektvoll behandelt und als Personen ernst genommen. Es herrschen ein höflicher Umgangston und eine reversible Anrede. In ähnlicher Weise sind auch die Beziehungen innerhalb des Kollegiums und zwischen Lehrenden und Leitung von gegenseitiger Achtung geprägt. Dies dürfte eines der wichtigsten Qualitätsmerkmale guter Schulen sein. Der psychologische Hintergrund besteht darin, dass mit dem Respekt auch erhebliche Verantwortung für das Verhalten dem Gegenüber zugewiesen wird. Diesem fällt es dadurch schwerer, Regeln zu verletzen, weil er etwas (nämlich Anerkennung) zu verlieren hat. Ein Aspekt dieser Wertschätzung ist auch die Erfolgserwartung im Hinblick auf die Leistungsfähigkeit der Schülerschaft (und der Lehrkräfte).

### *Konsequente Handhabung von Regeln: Berechenbarkeit des Verhaltens*

In guten Schulen werden Regeln mit den Schülerinnen und Schülern vereinbart bzw. in den Fällen, in denen dies nicht möglich ist (z.B. weil Regeln vorgegeben sind), auf verständliche Weise begründet. Es wird darauf geachtet, dass Regeln für alle gelten. Eng damit verbunden ist auch die Praxis, Regelverstöße sofort zu thematisieren. Es zeigt sich, dass eine Ahndung in vielen Fällen gar nicht nötig ist, wenn ein Regelverstoß sofort als solcher hingestellt und deutlich gemacht wird, dass dieses Verhalten nicht akzeptiert wird. Gute Schulen haben "... viele kunstvolle Einrichtungen geschaffen, um zu Vereinbarungen zu kommen und ihre Einhaltung zu sichern. Sie sind sich des Dauercharakters solcher Bemühungen sehr bewusst geworden. Demokratische Prozesse der Entscheidungsfindung, Komitees zur Konfliktbearbeitung, 'Kummerlöser' usw. gehören zu diesen Schulen." (Fend 1998)

### *Reichhaltiges Schulleben und vielfältige Entfaltungsmöglichkeiten für Lehrerkollegium und Schülerschaft*

In guten Schulen gilt: "Vieles ist möglich". Auf Initiativen, die aus der Schülerschaft kommen, wird großen Wert gelegt und die Schülerinnen und Schüler werden bei der Realisierung unterstützt, wenn diese begründet und mit den Aufgaben der Schule in Zusammenhang gebracht werden können. Ähnliches gilt für Lehrkräfte: Ihre Initiativen werden von der Leitung grundsätzlich gefördert, aber es wird verlangt, dass sie beobachtet und evaluiert werden und dass aus Fehlern gelernt wird.

### *Eine kooperative, aber deutlich wahrgenommene und zielbewusste Schulleitung*

Zwei fast widersprüchlich erscheinende Merkmale der Schulleitung zeichnen gute Schulen aus: Eine Leitung, die zuhören kann, sich in die Situation von Lehrkräften, Schülerschaft und Eltern hineinversetzen kann, die um Konsens bemüht ist, die aber gleichzeitig die Schule "leitet". Sie nimmt pointiert Führungsaufgaben wahr und stellt dabei pädagogische über organisatorische Interessen.

### *Zusammenarbeit und Konsens im Kollegium*

Sehr häufig wird berichtet, dass „gute Schulen“ durch eine intensive kollegiale Zusammenarbeit im Lehrkörper gekennzeichnet sind. Für eine Schule ist es günstig, "...sich nach innen und nach außen auf wichtige pädagogische Programmpunkte zu einigen, die sich zu einem charakteristischen 'Schulprofil' verbinden lassen". Es sollte einen "... Grundkonsens, zumindest einen Minimalkonsens, in wichtigen fachlichen Fragen, bei didaktisch-methodischen Problemen und Fragen der curricularen Gestaltung des Unterrichtes geben, möglichst auch in Fragen der Leistungsbeurteilung." (Hurrelmann 1991, S. 341 und 338). Gute Schulen zeichnen sich durch eine hohe Arbeitszufriedenheit der Lehrerschaft aus.

### *Einbeziehung der Eltern*

„Gute Schulen“ sind weiterhin dadurch gekennzeichnet, dass sie Raum für Engagement von Eltern geben und solches auch stimulieren können. Gute Schulen schotten sich offenbar nicht von ihrem sozialen Umfeld ab, sondern verstehen sich als Partner. Sie tun aber auch nicht alles und jedes, was von ihnen gefordert wird, sondern halten eine Balance zwischen Offenheit zum Gemeinwesen und Konzentration auf die eigenen Anliegen.

### *Schulinterne Lehrerfortbildung*

In einer Reihe von Studien über die Qualität von Schulen (z.B. OECD 1989) wird hervorgehoben, dass „gute Schulen“ eine Strategie für die Fortbildung und Weiterentwicklung der Lehrerinnen und Lehrer haben, die auf die Bedürfnisse des Schulprofils abgestimmt ist. Das Bewusstsein einer Schule über diese Bedürfnisse, über ihre Stärken und Schwächen und

die Bereitschaft, diese offen zu besprechen und gemeinschaftlich an einer Weiterentwicklung zu arbeiten, scheinen durchaus Kennzeichen für die Güte einer Schule zu sein.

### **3.1.3 Schulinterne und schulexterne Evaluation**

Eine schulinterne Evaluation kann als ein systematischer Prozess zur Qualitätsverbesserung von Schulen verstanden werden, der von der Schule selbst geplant und gesteuert wird. Die diesem Prozess zugrunde liegenden Bestandsaufnahmen und Bewertungen sind unter bestimmten Umständen durchaus geeignet, aus der Eigenperspektive die Stärken und Schwächen im Schulsystem aufzuzeigen und damit die Basis für eine Verbesserung der Schulqualität bereit zu stellen. Das Gelingen eines solchen Prozesses ist jedoch an mehrere Voraussetzungen gebunden, die im Schulalltag nicht immer leicht umzusetzen sind (Burkhard & Eikenbusch, 2000):

1. Ein internes Evaluationsvorhaben benötigt Zielklarheit hinsichtlich des Zwecks und der Richtung der Evaluation; Schwächen sollten zur ihrer Beseitigung offen gelegt und Stärken der Schule zu ihrer Bekräftigung betont werden.
2. Eine schulinterne Evaluation muss eine breitere Schulöffentlichkeit hinreichend beteiligen (Lehrerkollegium, Schulleitung, Schülerschaft und Eltern).
3. Eine schulinterne Evaluation setzt die Akzeptanz der schulbeteiligten Gruppen voraus und benötigt klare und offene Vereinbarungen über Regeln und Verantwortlichkeiten des Gesamtprozesses.
4. Das Vorhaben muss für die Beteiligten überschaubar und bearbeitbar bleiben; eine mögliche Überforderung, beispielsweise in methodischer Hinsicht, muss vermieden werden.
5. Die Ziele und die Themen der internen Evaluation sollten dennoch grundsätzlich alle Gestaltungsbereiche von Schulen berücksichtigen. Dies kann zu durchaus komplexen Zusammenhangsanalysen führen.
6. Die einer internen Evaluation zugrunde liegenden Analysen benötigen ausreichend Zeit für eine substanzielle und inhaltliche Auseinandersetzung mit den ermittelten Befunden. Trotzdem sollte eine möglichst kurzfristige Rückmeldung der Ergebnisse an alle Betroffenen erfolgen; die Diskussion über die Bedeutung bzw. über Folgerungen aus den Resultaten der Evaluation sollte dabei weitgehend öffentlich erfolgen.
7. Die kontinuierliche Anwendung von internen Evaluationen sollte eine fortwährende Prüfung der gewählten Verfahren hinsichtlich ihrer Aussagekraft, ihrer Gültigkeit und ihrer Effizienz beinhalten. Besonders hinsichtlich methodischer Alternativen sollte der interne Evaluationsprozess offen bleiben für Anregungen und Entwicklungsimpulse von außen.

Gerade die zuletzt genannten Voraussetzungen deuten an, dass in der Praxis viele deutsche Schulen mit der (alleinigen) Durchführung von internen Evaluationen bislang noch überfordert sein dürften und vielleicht deshalb entsprechende Maßnahmen noch weitgehend scheuen. Neben den nicht unerheblichen zeitlichen Aufwendungen zur Durchführung einer solchen internen Evaluation spielen hier wahrscheinlich auch fehlende methodische Kenntnisse eine Rolle. Entsprechende Beratungs- und Unterstützungssysteme für Schulen zur Selbstevaluation werden zurzeit zwar in mehreren Bundesländern aufgebaut (als ein Beispiel für Brandenburg: <http://www.lisum.brandenburg.de/sesus/>), ihre tatsächliche Effizienz bei der Implementierung schulinterner Evaluationsprogramme muss sich jedoch in der Praxis erst noch erweisen. Es ist – aus unserer Sicht – davon auszugehen, dass für eine flächendeckende interne Evaluation von Schulen noch auf längere Zeit ein ergänzendes Angebot von (für die Schulen kostenpflichtigen) externen Evaluationsprogrammen bereitgestellt werden sollte.

Auf externe Schulevaluationen wird auch aus anderen Gründen jedoch kaum zu verzichten sein. So sind beispielsweise die methodische Qualität und die Vergleichbarkeit der Ergebnisse

von Schulevaluationen weitaus eher durch zentral gesteuerte externe Evaluationen zu sichern als durch eine Vielzahl selbst gesteuerter interner Evaluationen. Weiterhin können die Ergebnisse von schulinternen Evaluationen letztlich nur vor dem Hintergrund der Ergebnisse anderer Vergleichsschulen verstanden und gedeutet werden; eine reine Selbstevaluation ist stets dem Risiko der „Betriebsblindheit“ ausgesetzt. Die Einbeziehung von Vergleichsmaßstäben, die eine Außenperspektive in die schulinterne Diskussion der eigenen Befunde einbringt, fehlt dem Konzept der internen Evaluation. Eine externe Schulevaluation kann auch die Aufgabe übernehmen, Einzelschulen mit alternativen Konzepten zu konfrontieren und Ergebnisdeutungen auf der Basis einer neutralen Außensicht zu geben.

Die Basis der Evaluationsstudien des Projekts „Unsere Schule...“ bilden die Ergebnisse der Befragung von Schülerinnen und Schülern, denn die Beurteilung von Unterricht, Lehrerverhalten, Klassenklima und Schulklima durch die „Kunden“ des Schulsystems stellt eine gute Möglichkeit dar, wichtige Qualitätsaspekte auf Schulebene zu erfassen und zu beschreiben. Schüler betrachten ihre Schule aus anderen Perspektiven als ihre Lehrkräfte und greifen dabei auch auf andere Beurteilungskriterien zurück. Beurteilungsdifferenzen zwischen Lehrkräften und Schülerschaft sind beim Thema Schulqualität deshalb nahezu vorbestimmt. Dies bedeutet aber nicht, dass die Schülerperspektive weniger zutreffend oder wichtig wäre als die Lehrerperspektive: Wie in Kapitel 1 bereits dargestellt wurde, verfügen bereits Kleinkinder über generalisierte stabile Vorstellungen über die erlebte Betreuungsqualität, und ab dem Vorschulalter können solche Erfahrungen mit Erziehungspersonen zuverlässig und gültig erfragt werden, wenn man altersangemessene Befragungsmethoden verwendet (Sturzbecher, 2001a).

Aggregiert man Schülereinschätzungen auf Klassen- oder Schulebene zu Gruppenurteilen, so erhält man Rückmeldungen dazu, wie es aus Schülersicht um die Qualität des Unterrichts und des Schulklimas bestellt ist. Kein ernsthafter Pädagoge sollte ein solches Feedback gering schätzen – im Gegenteil: Die Auseinandersetzung mit solchen Bewertungen auf Klassen- und Schulebene kann dem standortbezogenen Qualitätsmanagement von Schulen wertvolle Impulse verleihen. Dagegen würde das Ignorieren der Schülersicht auf Schulqualität bedeuten, eine gute Chance zur Förderung von Schulentwicklungsprozessen zu verschenken.

### 3.2 Ziele und Organisation des Projekts „Unsere Schule...“

Unser Bild über die Jugend wird weitgehend durch beunruhigende Medienberichte geprägt. Die Jugendforschung bietet hier oft ein anderes Bild, wie beispielsweise die Studien des Instituts für angewandte Familien-, Kindheits- und Jugendforschung an der Universität Potsdam (IFK) zur Lebenssituation Jugendlicher zeigen (Sturzbecher, 2001b und 2002). Aufbauend auf diesen Jugendstudien wurde vom IFK zusammen mit dem Institut für berufliche Bildung und Weiterbildung Göttingen (ibbw) im Jahr 2000 beim Bundesministerium für Bildung und Forschung (BMBF) der Antrag zur Förderung des Modellprojekts „Unsere Schule...“ gestellt. Das Projekt schloss zwei Projektbereiche bzw. Angebote für Schulen ein, nämlich eine schulinterne Evaluation<sup>5</sup> von sozialer Schulqualität und abweichendem Schülerverhalten sowie ein Fortbildungsprogramm für Lehrerinnen und Lehrer.

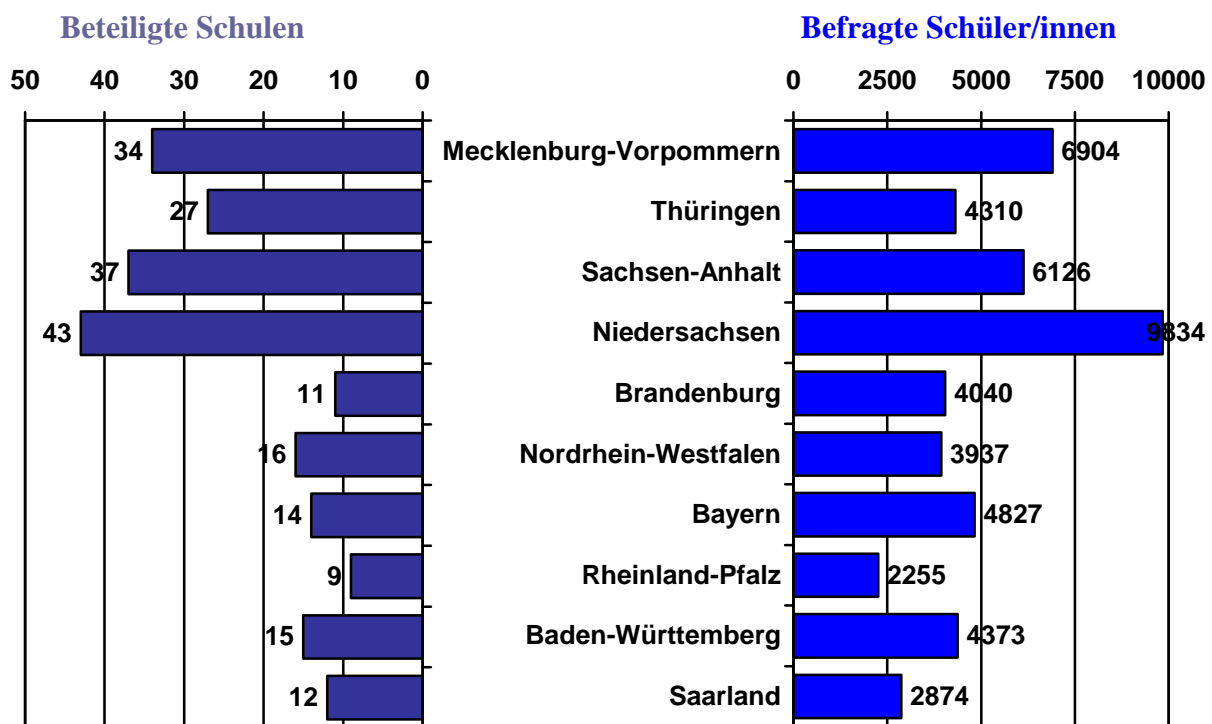
Das Hauptziel des Projekts lag in der Förderung der sozialen Schulqualität, also in der Verbesserung von Umfeldbedingungen des Sozialraums Schule. Deshalb wurden im ersten Teilprojekt in zwei Erhebungswellen an 218 Schulen fast 50.000 Schülerinnen und Schüler

---

<sup>5</sup> Kurz formuliert kann unter „Evaluation“ die „systematische Untersuchung und Überprüfung des Wertes und Nutzens eines Gegenstandes“ verstanden werden (Joint Committee 1994, S.3).

zum Schwerpunkt „Schulqualität“ befragt (vgl. Abbildung 1.4). Weiterhin wurden Informationen zum Themenbereich „Jugenddelinquenz“ erhoben. Schulleitung, Lehrerinnen und Lehrer, Eltern sowie Schülerinnen und Schüler sollten so die Möglichkeit erhalten, sich präzise Informationen über die Bedingungen an „ihrer“ Schule zu verschaffen. Gleichzeitig wurden die Ergebnisse der Einzelschule mit den Ergebnissen anderer Schulen (gleicher Schultyp, gleiches Bundesland) verglichen. Auf der Grundlage dieser Befragungsergebnisse sollte jede Schule ihr spezifisches Qualitäts- und Problemprofil erarbeiten können und in die schulinterne Diskussion über die Weiterentwicklung der Schulkonzeption und Schulqualität einbringen. Die Ergebnisse der schulinternen Evaluation wurden für die jeweilige Schule analytisch aufbereitet und anhand eines schriftlichen Datenreports sowie durch anschließende mündliche Präsentationsveranstaltungen in den einzelnen Schulen vorgestellt. Diese Maßnahmen sollten für alle Beteiligten sowohl zur Ermutigung als auch als empirische Basis dienen, um Problemlagen offensiv angehen und präventive Vorhaben in die Tat umsetzen zu können. Um der Schulentwicklung und Veränderungen an den teilnehmenden Schulen nachgehen zu können, wurden die Schülerbefragungen nach etwa zwei Jahren wiederholt und die Ergebnisse dieser Befragungen erneut an den Schulen vorgestellt.

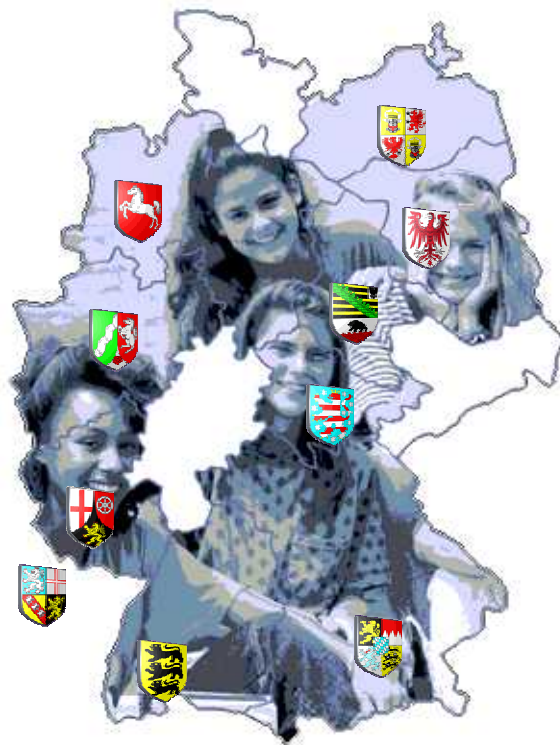
**Abbildung 4 Zum Umfang der Datenerhebungen im Projekt „Unsere Schule...“**



Das zweite Teilprojekt beschäftigte sich mit der Entwicklung und Erprobung eines modularen Fernlehrgangs, der flexibel und berufsübergreifend die schulinterne und schulexterne Lehrerfortbildung ergänzt. Verknüpft mit den Befragungsergebnissen und der Vorstellung der Ergebnisse aus den Schulreports wurden für die beteiligten Schulen Konzepte zur Förderung der Schulqualität aber auch für eine effiziente, zielgruppenorientierte Prävention gegen Gewalt, politischen Extremismus und Ausländerfeindlichkeit entwickelt. Jede der beteiligten Schulen konnte für sich aus einem Pool von 25 Lehreinheiten insgesamt bis zu sieben für die eigene Schulentwicklung besonders bedeutsame Module auswählen. Diese wurden dann in einer Gruppe von drei bis vier Lehrerinnen und Lehrern, einschließlich Vertretern der Schulöffentlichkeit, im Zeitraum von einem Schuljahr gemeinsam bearbeitet. Alle Teilnehmer am Fernlehrgang wurden dabei kontinuierlich von Tutoren betreut. Die teilnehmenden

Schulen hatten weiterhin die Möglichkeit, im Rahmen der schulinternen Lehrerfortbildung für einen Seminartag zu einem von der Schule als besonders wichtig definierten Thema beim ibbw einen Fachdozenten zu buchen, der dann in der einzelnen Schule für die Lehrerschaft und die Schulöffentlichkeit eine Fortbildung durchführte. Angestrebt wurde eine enge Kooperation mit den Lehrerfortbildungsinstitutionen der jeweiligen Bundesländer, um eine effektive Durchführung sowie die Nachhaltigkeit der angestrebten Entwicklungen an den einzelnen Schulen zu sichern. In Abstimmung mit der jeweiligen staatlichen Lehrerfortbildung wurden deshalb regionale und überregionale Seminare organisiert, die einerseits dem Austausch der beteiligten Projektschulen über Erfahrungen ihrer schulischen Entwicklung dienen sollten, so dass Netzwerke zwischen Schulen entstanden, die sich gegenseitig auch über das Projektende hinaus unterstützen können; andererseits wurden in diesen Seminaren Inhalte vermittelt, die die Fortbildungsteilnehmer qualifiziert, als Multiplikatoren in ihrem jeweiligen Kollegium aufzutreten.

**Abbildung 5 Beteiligte Bundesländer**



Das Projekt startete Ende 2000 in den sechs Bundesländern Brandenburg, Mecklenburg-Vorpommern, Niedersachsen, Nordrhein-Westfalen, Sachsen-Anhalt und Thüringen (erster Länderblock). Ende 2001, nach der Bewilligung eines Aufstockungsantrages durch das BMBF, konnten durch die Einbeziehung der Länder Baden-Württemberg, Bayern, Rheinland-Pfalz und Saarland (zweiter Länderblock) zusätzliche Schulen aus diesen Bundesländern an

dem Projekt teilnehmen. Das Jahr 2000 wurde vor allem zur Klärung organisatorischer Fragen und zur Entwicklung des Erhebungsinstruments genutzt. Seit Anfang 2001 konnte das Projekt „Unsere Schule...“ als Forschungs- und Dienstleistungsangebot für Bildungseinrichtungen durchgeführt werden.

Das Modellprojekt "Unsere Schule..." wurde während seiner gesamten Laufzeit von einem Lenkungsbeirat begleitet, der sich in regelmäßigen Abständen traf und den Projektträgern bei der Steuerung des Projektablaufs beratend zur Seite stand. Dem Lenkungsbeirat gehörten Vertreterinnen und Vertreter des BMBF, der verantwortlichen Projektträger (ibbw und IFK) und der beteiligten Länder (Ministerien und Lehrerfortbildung) an. Bedarfsweise wurden weitere Experten oder andere Gäste zu den Sitzungen hinzugezogen.

Die Internethomepage ([http://www.ibbw.de/projekte/unsere\\_schule/](http://www.ibbw.de/projekte/unsere_schule/)) des Projekts diente zur Information über das Projekt. Dort wurden der Ablauf, die Fortbildung und die Lehrbriefe ausführlich beschrieben. Die Homepage sollte auch ein Forum zum Austausch der Schulen im Projekt sein. Um eine offene Diskussion zu ermöglichen, wurde das Forum der Öffentlichkeit nicht zugänglich gemacht.

Die Träger des Projekts „Unsere Schule...“ waren das „Institut für angewandte Familien-, Kindheits- und Jugendforschung“ e.V. an der Universität Potsdam (IFK) sowie das „Institut für berufliche Bildung und Weiterbildung“ e.V. (ibbw) in Göttingen. Gefördert wurde das Projekt durch das Bundesministerium für Bildung und Forschung (BMBF). Das Projekt hatte eine Laufzeit von Oktober 2000 bis Ende September 2004 (IFK) bzw. bis Ende Dezember 2004 (ibbw).

### 3.3 Zur Projektentwicklung in den einzelnen Bundesländern

Der Grundsatz „Bildung ist Ländersache“ galt auch für die Durchführung des Projekts „Unsere Schule...“. Der offizielle Start des Projekts auf Landesebene kann mit der ersten Sitzung des Lenkungsbeirats auf den 9. Februar 2001 in Göttingen datiert werden. Herr Dr. Eisfeld wies als anwesender Vertreter des Bundesministeriums für Bildung und Forschung darauf hin, dass die Umsetzung und Verstetigung des Modellprojekts vor allem in die Verantwortlichkeit der beteiligten Länder falle. Bereits auf dieser ersten Sitzung wurde deutlich, dass die Vertreterinnen und Vertreter der einzelnen Landesministerien teilweise recht unterschiedliche Vorstellungen hinsichtlich der konkreten Umsetzung der von IFK und ibbw geplanten Projektstrukturen hatten. So wurde für Brandenburg die Zahl der teilnehmenden Schulen auf 10 Einrichtungen begrenzt – dafür wurde eine intensivere Betreuung der teilnehmenden Schulen sowie ein verbindlicher Ablaufplan für einzelne Projektphasen vereinbart. Aus Niedersachsen wurde der Wunsch geäußert, mit Hilfe der Schülerbefragungen auch zu landesrepräsentativen Befunden zu den Forschungsthemen „Jugenddelinquenz“ und „Schulqualität“ zu gelangen. In Thüringen wurde nur eine Schulform (staatliche Regelschulen) für die Projektteilnahme vorgesehen. In Sachsen-Anhalt sollten nicht nur allgemein bildende Schulen sondern auch Berufsbildende Schulen in das Projekt einbezogen werden. In Mecklenburg-Vorpommern kam eine Sondervereinbarung zustande, die die Projektteilnahme auch für Förderschulen des Landes vorsah. Vergleichbare Vereinbarungen, die sich an den jeweiligen Interessenlagen einzelner Bundesländer orientierten, wurden auch mit den Teilnehmern des zweiten Länderblocks getroffen. Im Folgenden sollen die wichtigsten Entwicklungen in den teilnehmenden Bundesländern kurz zusammengefasst werden.

### **3.3.1 Zur Entwicklung in Niedersachsen**

Da es zunächst ein Ziel des Niedersächsischen Kultusministeriums war, mit den Schülerbefragungen des Projekts auch landesrepräsentative Informationen zu den Themen „Jugenddelinquenz“ und „Soziale Schulqualität“ zu erhalten, erarbeitete das IFK in diesem Bundesland ein entsprechendes Auswahlverfahren. Hierzu wurden über das Statistische Landesamt Daten über die Gesamtheit aller allgemein bildenden Schulen des Landes Niedersachsen für das Jahr 2000 eingeholt. Diese Daten enthielten Angaben über Schulart, Schulgröße und Anschrift der Schulen. Ziel des Auswahlverfahrens war es, aus der Grundgesamtheit aller Schulen eine Zufallsauswahl von 100 Schulen zu ziehen. Hierzu wurden anhand der amtlichen Daten die Schulen zunächst nach einzelnen Schultypen geschichtet und innerhalb jeder Schicht die genaue Zahl von Schulen ermittelt, die dem tatsächlichen Anteil des Schultyps in der Grundgesamtheit entsprach. Bei den anschließend ermittelten Zufallsstichproben innerhalb einzelner Schichten wurde die unterschiedliche Anzahl der Schüler an den jeweiligen Schulen berücksichtigt. Die angestrebte Gleichwahrscheinlichkeit der Auswahl von Schülern an Schulen mit unterschiedlicher Größe konnte dabei über die Anwendung eines PPS-Verfahrens („Probability Proportional to Size“) gewährleistet werden. Die so ermittelten 100 Schulen wurden anschließend durch das IFK sowie durch das Niedersächsische Kultusministerium angeschrieben und um die Teilnahme am Projekt gebeten. Die angestrebte Ausschöpfung der Stichprobe für eine landesrepräsentative Auswahl wurde bei mindestens 40 Prozent (40 Schulen) angesetzt. Leider erwies sich die Teilnahmebereitschaft deutlich niedriger als erwartet: Letztlich konnten nur 10 der 100 kontaktierten Schulen für eine Projektbeteiligung gewonnen werden. Für eine landesrepräsentative Umfrage war diese Ausschöpfungsquote bei weitem zu gering.

Da die Vorbereitung und Durchführung eines zweiten Auswahlverfahrens zu zeitaufwändig gewesen wäre, wurde das Ziel einer landesrepräsentativen Studie in Niedersachsen notgedrungen aufgegeben. Die Einbeziehung weiterer Schulen erfolgte in Niedersachsen ab diesem Zeitpunkt eher über informelle Kanäle wie Verteiler auf der Ebene von Bezirksregierungen oder die „Mundpropaganda“ bereits teilnehmender Schulen. Der Aufnahmeprozess von Schulen erfolgte in Niedersachsen demnach weitgehend in offener Form und wurde erst Ende 2002 beendet. Letztlich nahmen 43 allgemein bildende und Berufsbildende Einrichtungen aus Niedersachsen am Projekt „Unsere Schule...“ teil. Vertreten waren alle Schulformen mit einer weiträumigen geografischen Standortverteilung. Befragt wurden in Niedersachsen in der ersten Erhebungswelle 4425 und in der zweiten Erhebungswelle 5409 Jugendliche.

### **3.3.2 Zur Entwicklung in Mecklenburg-Vorpommern**

Bereits im Frühjahr 2001 kam es in Mecklenburg-Vorpommern zu einer offiziellen Auftaktveranstaltung für interessierte Schulleiterinnen und Schulleiter. Diese Veranstaltung erfolgte auf Einladung des Ministeriums für Bildung, Wissenschaft und Kultur des Landes Mecklenburg-Vorpommern sowie durch das Landesinstitut für Schule und Ausbildung (LISA) in Schwerin. Die konkrete Auswahl der Schulen wurde von den Schulämtern übernommen. Zu den eingeladenen Schulen gehörten auch mehrere Fördereinrichtungen. Die überwiegende Mehrheit der vertretenen Schulen konnte im Rahmen dieser Veranstaltung für eine Projektbeteiligung gewonnen werden. Zweifellos trug hierzu auch der Umstand bei, dass für die Projektteilnahme in Mecklenburg-Vorpommern zusätzliche Abminderungsstunden gewährt wurden. Die so bereits in einer frühen Projektphase ermittelten Schulen repräsentierten alle Schultypen des Landes Mecklenburg-Vorpommern und verteilten sich über einen weiträumigen geografischen Einzugsbereich. Von den insgesamt 34 teilnehmenden Schulen nahm die Mehrheit bis zum Abschluss des Projekts teil. Ein wichtiger Ausfallgrund



lag in Mecklenburg-Vorpommern in den parallel zum Projekt ablaufenden, demografisch bedingten Schulschließungsprozessen. Einige der einbezogenen Schulen in Mecklenburg-Vorpommern nahmen das Angebot des IFK wahr, die Schülerbefragungen zur Schulqualität durch eine Lehrerbefragung zu ergänzen. Befragt wurden in Mecklenburg-Vorpommern in der ersten Erhebungswelle 2636 und in der zweiten Erhebungswelle 4268 Jugendliche.

### **3.3.3 Zur Entwicklung in Sachsen-Anhalt**

Auch in Sachsen-Anhalt startete das Projekt im Rahmen einer zentralen Auftaktveranstaltung für interessierte Schulleiterinnen und Schulleiter. Entsprechend den Vorstellungen zur Projektrealisierung des Kultusministeriums des Landes Sachsen-Anhalt waren hierbei neben allgemein bildenden auch mehrere Berufsbildende Einrichtungen vertreten. Auch in Sachsen-Anhalt konnten den teilnehmenden Schulen für die Projektteilnahme zusätzliche Abminderungsstunden gewährt werden. Die Auftaktveranstaltung fand im Mai 2001 statt, anschließend begann das Projekt in Sachsen-Anhalt mit 17 Schulen. Aufgrund der positiven Projektentwicklung wurde auf Beschluss des zuständigen Kultusministeriums der Teilnehmerkreis der Schulen im Jahr 2002 um zusätzliche 20 Einrichtungen erweitert. Aufgrund des relativ späten Projektbeitritts konnte die Mehrzahl dieser Nachzügler leider nicht mehr in die zweite Befragungswelle einbezogen werden. Auch mehrere Schulen in Sachsen-Anhalt sahen eine Möglichkeit zur Förderung der eigenen Schulqualität darin, die Schülerbefragung durch eine Befragung der eigenen Lehrkräfte zu ergänzen. Befragt wurden in Sachsen-Anhalt in der ersten Erhebungswelle 4163 und in der zweiten Erhebungswelle noch 1963 Jugendliche.

### **3.3.4 Zur Entwicklung in Brandenburg**

Öffentliche Aushänge an den Schulämtern informierten die Schulleiterinnen und Schulleiter in Brandenburg im Frühjahr 2001 über die Möglichkeit einer Projektbeteiligung. Das Interesse an einer solchen Beteiligung war unerwartet hoch. Allerdings wurde aufgrund besonderer Vereinbarungen zur Projektumsetzung sowie durch das begrenzte Kontingent an verfügbaren Abminderungsstunden die Zahl der teilnehmenden Schulen auf zehn begrenzt. Der Projektverlauf in Brandenburg unterschied sich von dem in den anderen Bundesländern weiterhin auch dadurch, dass seitens des zuständigen Ministeriums für Bildung, Jugend und Sport (MBJS) genaue Vorstellungen hinsichtlich der Maßnahmen bestanden, mit denen das Ziel der Förderung sozialer Schulqualität an den teilnehmenden Schulen erreicht werden sollte. Zu diesen Maßnahmen zählte insbesondere die Bildung fester Steuerungsgruppen an den Schulen, die – mit Betreuung durch das IFK – den Prozess der Evaluation und der Weiterbildung möglichst eigenständig tragen und schulintern verbreitern sollten. Deshalb wurden noch vor Beginn der Datenerhebungen in Brandenburg die Mitglieder der Steuerungsgruppen im Rahmen einer eintägigen Fortbildungsveranstaltung durch einen Mitarbeiter des IFK auf ihre Aufgaben vorbereitet. Ziel dieser Veranstaltung war es, die Teilnehmer im Umgang mit Befragungsdaten und deren Präsentation soweit vertraut zu machen, dass die angestrebte eigenständige Auswertung und Diskussion der Ergebnisse des schriftlichen Schulreports über die Steuerungsgruppen möglich war.

Obwohl die feste Absicht hierzu bei den meisten Schulen gegeben war, zeigte sich im weiteren Verlauf der Projektentwicklung schnell, dass eine effektive – d.h. insbesondere Prozessauslösende – Wirkung von Datenpräsentationen und -diskussionen über ausschließlich schulinterne Aktivitäten nur schwer zu erreichen ist. Viele Schulen versprachen sich (auch) in Brandenburg einen stärkeren Effekt durch den Auftritt einer außen stehenden neutralen Person. Auf ausdrücklichen Wunsch der meisten teilnehmenden Schulen in Brandenburg

wurde deshalb die Auswertung der Daten sowie die hieraus abzuleitenden Schlussfolgerungen durch das IFK-Team übernommen. Einige der Schulen, die dies zunächst eigenständig versuchten, baten später um eine ergänzende Datenpräsentation durch das IFK.

Als eine weitere wichtige Maßnahme zur Optimierung der Projektentwicklung waren regelmäßige Treffen der Steuerungsgruppen in Brandenburg vorgesehen. Diese sollten dem Erfahrungsaustausch und zur Diskussion der Projektentwicklung an den Einzelschulen dienen. Leider verlor die anfänglich feste Terminstruktur dieser Treffen mit fortschreitender Projektlaufzeit ihre Bindung – ein Umstand, der von vielen Mitgliedern der Steuerungsgruppen ausdrücklich bedauert wurde.

In Brandenburg nahm zwar die Mehrzahl der teilnehmenden Schulen das zusätzliche Angebot des IFK für eine Lehrerbefragung an; an zwei Schulen musste das geplante Vorhaben jedoch aufgrund der geringen Teilnahmebereitschaft der Lehrkräfte abgesagt werden. Eine Schule erfuhr erst 2003 von dem Projekt und nahm die Möglichkeit einer verspäteten Beteiligung – auch ohne die Gewährung von Abminderungsstunden – gern an. Befragt wurden in Brandenburg in der ersten Erhebungswelle 2009 und in der zweiten Erhebungswelle 2031 Jugendliche.

### **3.3.5 Zur Entwicklung in Thüringen**

Im Mai 2001 fand in den Räumen des Thüringer Kultusministeriums eine Informationsveranstaltung für interessierte Schulleiterinnen und Schulleiter von staatlichen Regelschulen statt. Hierzu wurden im Vorfeld pro Schulamt zwei Schulen ausgewählt. Die Einbeziehung anderer Schulformen war zunächst nicht geplant. Mit der Konzentration auf eine Schulform sowie mit einer weiträumigen geografischen Streuung der Standorte verband sich die Hoffnung, dass die zusammen gefassten Angaben Rückschlüsse auf die landesweiten Gegebenheiten zumindest innerhalb eines wichtigen Schultyps in Thüringen gestatten könnten. Da bereits 2001 (ohne Gewährung von Abminderungsstunden) 21 staatliche Regelschulen in Thüringen für eine Projektbeteiligung gewonnen werden konnten, erwies sich diese Hoffnung durchaus als begründet. Die Begrenzung der Teilnahmemöglichkeit auf staatliche Regelschulen wurde während des Jahres 2002 aufgehoben. Hierdurch nahmen als Nachzügler auch Gymnasien, Gesamtschulen und Berufsbildende Einrichtungen am Projekt teil. Die Gesamtzahl der teilnehmenden Bildungseinrichtungen erhöhte sich in Thüringen dadurch auf 29 (zwei Schulen beteiligten sich ausschließlich an der Weiterbildung des ibbw). Allerdings gab es mehrere Ausfälle in Folge von Schulschließungen und -zusammenlegungen. Auch in Thüringen wurden die Schulevaluationen vereinzelt durch Ergebnisse aus der Befragung von Lehrkräften ergänzt. Befragt wurden in Thüringen in der ersten Erhebungswelle 2483 und in der zweiten Erhebungswelle 1827 Jugendliche.

### **3.3.6 Zur Entwicklung in Nordrhein-Westfalen**

Obwohl Nordrhein-Westfalen (NRW) zum ersten Länderblock gehörte, begannen konkrete Maßnahmen der Projektrealisierung hier relativ spät und nahezu parallel mit dem Projektstart innerhalb des zweiten Länderblocks, d.h. etwa gegen Ende 2001. Eine Hauptursache dieser Verzögerung lag in dem langen Verfahren der Projektbewilligung. Die Prüfung und Bewilligung des Erhebungsinstrumentariums durch die Datenschutzbeauftragte des Landes Nordrhein-Westfalen nahm deutlich mehr Zeit in Anspruch als ursprünglich geplant. Eine zentrale Auftaktveranstaltung war für NRW nicht vorgesehen. Die Schulen erhielten hier auf höchst unterschiedlichen Wegen Informationen über die Möglichkeit einer Projektbeteiligung. Hierzu zählten auch mehrere Informationsveranstaltungen durch Mitarbeiter des ibbw und des IFK im Rahmen von Treffen regionaler Schulbehörden des Landes NRW. Mit Unterstützung

durch das Ministerium für Schule, Wissenschaft und Forschung des Landes Nordrhein-Westfalen konnte die Zahl der teilnehmenden Schulen bis Ende 2002 auf insgesamt 16 erhöht werden. Befragt wurden in NRW in der ersten Erhebungswelle 2961 und in der zweiten Erhebungswelle 976 Jugendliche.

### **3.3.7 Zur Entwicklung in Bayern**

Von zentraler Bedeutung für die Entwicklung des Projekts in Bayern war die enge Anbindung der Projektkoordination an die Akademie für Lehrerfortbildung und Personalführung in Dillingen. Mit Beginn des Schuljahres 2001/2002 wurden die bayerischen Schulen über das Projekt informiert. Im Januar 2002 fand eine erste Informations- und Auftaktveranstaltung mit Vertreterinnen und Vertretern von 15 Schulen in Dillingen statt. Im Freistaat Bayern nahmen letztlich 14 Schulen aus fünf Regierungsbezirken am Projekt teil. Lediglich eine Bildungseinrichtung schied vorzeitig aus dem Projekt aus. Befragt wurden in Bayern in der ersten Erhebungswelle 2116 und in der zweiten Erhebungswelle 2711 Jugendliche.

### **3.3.8 Zur Entwicklung in Baden-Württemberg**

Das Ministerium für Kultus, Jugend und Sport Baden-Württemberg konnte bis Ende 2001 12 Schulen für eine Projektteilnahme interessieren. Im November 2001 fand für diese Bildungseinrichtungen in Stuttgart eine gemeinsam von IFK und ibbw gestaltete Einführungsveranstaltung statt. Auch die staatliche Lehrerfortbildung wurde zu dieser Veranstaltung eingeladen. Im Verlauf des Jahres 2002 konnte die Zahl der teilnehmenden Schulen in Baden-Württemberg auf 15 Einrichtungen erweitert werden. Für die Teilnahme an dem Projekt wurde in Baden-Württemberg eine zusätzliche Abminderungsstunde eingeplant. Befragt wurden in Baden-Württemberg in der ersten Erhebungswelle 2750 und in der zweiten Erhebungswelle 1623 Jugendliche.

### **3.3.9 Zur Entwicklung in Rheinland-Pfalz**

Als recht schwierig erwies es sich, in Rheinland-Pfalz Schulen für eine Projektbeteiligung zu gewinnen. Erst am 1. März 2002 fand eine Informationsveranstaltung für Schulleiterinnen und Schulleiter in den Räumen des Ministeriums für Bildung, Frauen und Jugend in Mainz statt. Zwar zeigten sich die anwesenden Vertreterinnen und Vertreter von etwa 20 Schulen an den Zielen des Projekts durchaus interessiert; zu einer positiven Entscheidung für die Projektteilnahme kam es jedoch nur in wenigen Fällen. Die Gründe für das erkennbar niedrige Interesse waren sicherlich vielschichtig; eine Rolle spielte dabei zweifellos der Umstand, dass den Schulen keine Entlastungsstunden angeboten werden konnten. Trotz nachfolgender Bemühungen nahmen in Rheinland-Pfalz lediglich 9 Schulen am Projekt „Unsere Schule...“ teil. Befragt wurden in Rheinland-Pfalz in der ersten Erhebungswelle 1538 und in der zweiten Erhebungswelle 717 Jugendliche.

### **3.3.10 Zur Entwicklung im Saarland**

Startschwierigkeiten bei der Projektrealisierung ergaben sich im Saarland zunächst durch Besonderheiten des saarländischen Schulgesetzes, demzufolge die Nutzung von so genannten „Selbstbeichtigungsfragen“ (z.B.: „Wie oft nimmst du Drogen?“) im Rahmen von Schülerbefragungen unzulässig ist. Da der Standardfragebogen des Projekts „Unsere Schule...“ im Bereich des Themas „Jugenddelinquenz“ entsprechende Fragen enthielt, wurde das Erhebungsverfahren des Projekts zunächst nicht genehmigt. Erst nach einer Überarbeitung

des Fragebogens und der Genehmigung des Erhebungsverfahrens konnte das Projekt im Saarland starten.

Für die Anwerbung von Schulen wurden mit Unterstützung des Ministeriums für Bildung, Kultur und Wissenschaft des Saarlands die turnusmäßig stattfindenden Schulleitertreffen von Gesamtschulen Ende April 2002 sowie von Erweiterten Realschulen im September 2002 als Informations- und Auftaktveranstaltungen erfolgreich genutzt. Interessierte Schulen setzten sich im direkten Anschluss an diese Veranstaltungen mit dem IFK in Verbindung, um noch offene Fragen zu klären und ggfs. die Einzelheiten für die erste Datenerhebung abzusprechen. Auch ohne die Gewährung von zusätzlichen Entlastungsstunden konnten auf diese Weise im Saarland insgesamt zunächst 11 Schulen für die Teilnahme am Projekt gewonnen werden. Leider hatte der verzögerte Projektstart unvermeidliche Auswirkungen auf die weitere Projektentwicklung an den teilnehmenden Schulen. So konnte die zweite Erhebungswelle im Saarland nur über eine kurze Zeitdistanz zur letzten Befragung realisiert werden. Auch für die Durchführung des Fortbildungsprogramms des ibbw verblieb weniger Zeit als ursprünglich vorgesehen. Ein Gymnasium schloss sich noch Anfang 2003 dem Projekt an. Befragt wurden im Saarland in der ersten Erhebungswelle 2038 und in der zweiten Erhebungswelle 836 Jugendliche.

### **3.3.11 Abschließende Hinweise zur Projektentwicklung**

Die Schulauswahl basierte im Rahmen des Projekts nicht auf zufallsgesteuerten oder quotenbestimmten Verfahren der Stichprobenziehung, sondern im wesentlichen auf dem Interesse und dem freiwilligen Mitwirken von interessierten Bildungseinrichtungen. Dennoch zeigte der Vergleich mit den amtlichen Daten der Schulstatistik, dass in mehreren Bundesländern die Struktur der befragten Schülerinnen und Schüler nach den Merkmalen „Geschlecht“, „Klassenstufe“ und „Schultyp“ weitgehend den landesweiten Verteilungen entsprachen. Auch die geografische Gliederung von Flächenstaaten wie Sachsen-Anhalt, Niedersachsen und Mecklenburg-Vorpommern konnte über die Streuung der jeweiligen Schulstandorte angemessen abgebildet werden. Die aus den zusammengefassten Befunden der Schülerbefragungen gebildeten „Länderreports“ gehen in ihrer Bedeutung demzufolge über eine Darstellung der Befunde einiger projektbeteiligter Schulen hinaus und lassen – mit Einschränkungen – durchaus auch gewisse Rückschlüsse auf landesweite Gegebenheiten zu.

Als Grundgesamtheit für die Befragungen wurden zu Beginn der Projektplanungen zunächst Schülerinnen und Schüler der Jahrgänge sieben bis 13 an verschiedenen allgemein bildenden Schulformen definiert, wie sie in den jeweiligen Bildungseinrichtungen der einzelnen Bundesländer vertreten waren. Aufgrund der zu erwartenden Mehraufwendungen bei der Durchführung standardisierter Befragungen waren Fördereinrichtungen und Berufsbildende Schulen demnach ursprünglich für eine Projektteilnahme nicht vorgesehen. Schon bald zeigte sich im Rahmen der Informationsveranstaltungen in den einzelnen Bundesländern, dass auch seitens dieser Einrichtungen ein großes Interesse an einer Projektteilnahme bestand. Die Projektträger erweiterten deshalb kurzfristig den Kreis der einzubeziehenden Schulformen. Dies hatte allerdings zur Folge, dass weder ein einheitliches Erhebungsinstrument noch ein einheitliches Auswahlverfahren zum Einsatz kamen. Vielmehr mussten die eingesetzten Materialien und Verfahren flexibel an die jeweiligen Besonderheiten einzelner Bildungseinrichtungen angepasst werden.

In einigen Bundesländern wurde von den zuständigen Ministerien die Zahl der einzubeziehenden Schulen begrenzt. In anderen Ländern gab es solche Obergrenzen nicht. Dies hatte zur Folge, dass die durchschnittliche Zahl der teilnehmenden Schulen pro Bundesland erheblich variierte (s. Abbildung 1.4 in diesem Kapitel). Die Auswahl der Schulen wurde in der Regel in einem kurzen Zeitraum abgeschlossen. Lediglich in

Niedersachsen, Sachsen-Anhalt und in Thüringen wurde aufgrund des anhaltenden Interesses an einer Projektteilnahme eine größere Zahl von Nachzögler\*innen in das bereits angelaufene Projekt aufgenommen. Ein endgültiger Aufnahmestopp wurde erst Anfang 2003 gesetzt. So erfreulich das anhaltende Interesse der Schulen in den Bundesländern war, der außerplanmäßig späte Projektbeitritt hatte meist eine Begrenzung der Projektleistungen zur Folge. So war beispielsweise an vielen Nachzügler\*erschulen eine zweite Datenerhebung und -auswertung im noch verbleibenden Projektzeitraum nicht mehr durchführbar.

## 3.4 Die Entwicklung des Erhebungsinstruments

### 3.4.1 Allgemeine Anmerkungen zur Fragebogenkonstruktion

Für die Durchführung einer schriftlichen Befragung unter Schülerinnen und Schülern benötigt man ein entsprechendes Erhebungsinstrument – einen Fragebogen. Im Rahmen der quantitativen Sozialforschung handelt es sich dabei vor allem um so genannte standardisierte Erhebungsinstrumente, bei denen darauf geachtet wird, dass die Stimuli – d.h. die Fragen des Fragebogens – die auf die Befragten wirken, über alle Probanden hinweg unverändert bleiben. Durch dieses Vorgehen wird angestrebt, die abgegebenen Antwortreaktionen miteinander vergleichbar zu halten. Die Regeln zur Konstruktion von Fragebögen beziehen sich dabei meist auf zwei unterschiedliche Aspekte: die inhaltlich-methodische Gestaltung und die optische Gestaltung des Fragebogens.

Wenden wir uns zunächst der inhaltlich-methodischen Gestaltung des Fragebogens zu. Leider führen die zahlreichen Vorgaben und Richtlinien zur Frageformulierung in vielen Fällen nicht zu den angestrebten Resultaten. Dies hat verschiedene Ursachen. Ein hierbei zu beachtender Faktor ist, dass sich die Richtlinien zur Frageformulierung meist nur auf die inhaltliche Gestaltung von Einzelfragen beziehen und bei ihrer Anwendung übersehen wird, dass die Einzelfragen eines Fragebogens oft in einem Sinnzusammenhang stehen. Dadurch besteht die Gefahr, dass die Beantwortung einer bestimmten Frage die Beantwortung nachfolgender Fragen beeinflusst („Halo-Effekt“). Da die genaue Wirkung solcher Effekte nur durch gezielte Tests mit verschiedenen Frageanordnungen überprüft werden kann, griff das Team des Projekts „Unsere Schule...“ bei der Konstruktion des Fragebogens auf Erfahrungen aus früheren Jugendstudien zurück, bei denen die Reihenfolge einzelner Themenbereiche im Fragebogen variiert wurde. Dabei zeigten sich niemals auffällige Abweichungen, die auf eine beachtenswerte Wirkung von Reihungs- bzw. Ausstrahlungseffekten hätte schließen lassen.

Bei der Konstruktion eines schriftlichen Fragebogens ist weiterhin darauf zu achten, dass der Einführung in die Handhabung des Fragebogens eine besondere Bedeutung zukommt: Die Instruktion entscheidet zum einen nicht selten über die Motivation des Befragten zur Fragebogenbearbeitung und zum anderen über die Zuverlässigkeit der erhaltenen Antworten (Landua, 1995). Weiterhin wurden bei der Konstruktion des Standardfragebogens im Projekt „Unsere Schule...“ folgende allgemeine Richtlinien berücksichtigt:

- Zu komplexen Themenbereichen wurden stets mehrere Fragen gestellt (Konzept der „Multiplen Indikatoren“).
- Fragen, die einem übergeordneten Themenbereich zuzuordnen waren, wurden stets nacheinander abgefragt. Dabei wurden allerdings – als Kontrollfragen – auch invers formulierte Items eingesetzt.
- Neue Themenbereiche wurden stets mit Überleitungssätzen eingeleitet.
- Themenbereiche, die nicht für alle Befragten von Belang waren, wurden durch den Einsatz von Filterfragen auf den relevanten Probandenkreis beschränkt.

- Sensible oder „heikle“ Fragen wurden – soweit dies der thematische Aufbau des Fragebogens zuließ – an das Ende des Fragebogens gestellt, um das Abbruchrisiko bei der Bearbeitung zu mindern.

Für das Design eines Fragebogens sind nicht nur ästhetische Maßstäbe geltend zu machen. Vielmehr ist die äußere Gestaltung stets so auszurichten, dass der Befragte keine formalen Schwierigkeiten mit der Handhabung des Erhebungsmaterials hat. Dies betrifft zwar grundsätzlich alle Bestandteile des Fragebogens, trifft jedoch insbesondere dort zu, wo unvermeidbare Zusatzelemente des Designs – z.B. die Steueranweisungen von Filterfragen – die Handhabung des Fragebogens erschweren. Aus diesem Grund wird für einen Fragebogen z.B. in der Regel stets nur eine Art der Filterführung benutzt.

**Abbildung 6 Titelblatt des Fragebogens zum Projekt „Unsere Schule...“**



Die Frage, wie lang ein Fragebogen sein darf, ist nicht einfach zu beantworten. Zwar haben Forschungsergebnisse zu diesem Thema gezeigt, dass eine Befragungsdauer von 60 bis 90 Minuten keineswegs als problematisch gelten muss (Fowler, 1984); allerdings beziehen sich die Ergebnisse solcher Untersuchungen meist auf die Befragung von Erwachsenen. Erfahrungen aus früheren Jugendstudien des IFK zeigen, dass – auch bei anfänglich hoher Teilnahmebereitschaft – die Aufmerksamkeit insbesondere bei Schülern der siebten und achten Klassen nach spätestens einer Schulstunde merklich nachließ. Weiterhin fanden sich Unterschiede zwischen der Schülerschaft verschiedener Schultypen.

Die zumutbare Länge eines Fragebogens ist jedoch auch über die Form des Layouts beeinflussbar. So kann eine ansprechende Fragebogengestaltung gerade bei Jugendlichen dazu beitragen, dass die Zeit beim Ausfüllen „vergessen“ wird. Deshalb ist es

empfehlenswert, ein Fragebogenformat großzügig anzulegen und nicht zu versuchen, möglichst viele Fragen auf einer Seite unterzubringen. Die schnelle Abarbeitung von Seiten und die Sichtbarkeit dieses Fortschritts hat auf die Befragten meist einen motivationsfördernden Einfluss. Weiterhin wurde bei der Gestaltung des Standardfragebogens des Projekts „Unsere Schule...“ auch auf den Einsatz comicartiger Bilder von „Strichmännchen“ gesetzt. Die Auswahl und Platzierung dieser Figuren orientierte sich – in loser Form – an den Befragungsthemen der jeweiligen Seite. Maßnahmen dieser Art sind in der Umfrageforschung nicht unumstritten, da sie vom eigentlichen Zweck des Fragebogens ablenken und von ihnen ein zusätzlicher, schwer kalkulierbarer Stimulus ausgeht. Dennoch wurde davon ausgegangen, dass die motivationsfördernde Wirkung solcher Figuren die mit ihrem Einsatz verbundenen Risiken mehr als aufwog.

Pretests und der praktische Einsatz des Fragebogens an den Schulen bestätigten die dargestellten bzw. berücksichtigten Konstruktionsgrundsätze. Obwohl vor allem in den siebten Klassen sowie unter den Befragten an Hauptschulen das Ausfüllen des Fragebogens in Einzelfällen bis zu zwei Schulstunden (90 Minuten) andauerte, kamen Abbrüche oder monotone Antwortmuster auch gegen Ende des Fragebogens nur selten vor.

Auf der folgenden Seite wird das Konzept des Fragebogenlayouts anhand der Einführungsseite dargestellt. Erläutert werden die Handhabung von allen im Fragebogen vorkommenden Fragetypen, einschließlich des Umgangs mit Filterführungen und der Anweisung, wie nicht zutreffende Angaben korrigiert werden können. Für die einzelnen Elemente jedes Themenkomplexes – die eigentliche Frage, die Antwortvorgaben, zusätzliche Hinweise und Erläuterungen – wurden unterschiedliche Schrifttypen benutzt. Die Gestaltung des Fragebogens orientierte sich an einigen Stellen zwar auch an den technischen Voraussetzungen für die anschließende maschinelle Datenerfassung, diese beeinflussten das Gesamtlayout jedoch nur unwesentlich. Die Anredeform „Sie“ wurde unter Berücksichtigung der Befragten aus der Sekundarstufe II gewählt und stieß auch bei jüngeren Jahrgängen nicht auf prinzipielle Ablehnung.

Abbildung 7 Auszug aus dem Fragebogen

Liebe Jugendliche: Wenn Sie mitmachen wollen und Sie Ihre Einverständniserklärung unterschrieben haben, können Sie jetzt loslegen!

Wir möchten Ihnen das Ausfüllen unseres Fragebogens erleichtern. Deshalb geben wir Ihnen auf dieser Seite wichtige Hinweise zur Beantwortung der Fragen anhand von Beispielen.

Es gibt vier verschiedene Typen von Fragen.

Beispiel 1: Fragen, bei denen eine Zahl eingetragen werden soll.	
B1	<p>Wie viel Zeit werden Sie für die Beantwortung des Fragebogens benötigen?</p> <p style="text-align: right;"><b>23</b> Minuten (DIE ZEIT BITTE HIER EINTRAGEN.)</p>

Beispiel 2: Fragen, bei denen alle zutreffenden Antworten anzukreuzen sind.	
B2	<p>Geben Sie bitte Dinge an, die Sie mit auf eine einsame Insel nehmen würden.</p> <p>(KREUZEN SIE BITTE ALLES ZUTREFFENDE AN!)</p> <ul style="list-style-type: none"> <li><input checked="" type="checkbox"/> Mein Robinson Crusoe-Buch</li> <li><input type="checkbox"/> Meine Musik</li> <li><input type="checkbox"/> Sonnencreme</li> <li><input checked="" type="checkbox"/> Schuhcreme</li> <li><input type="checkbox"/> Kondome</li> </ul>

Beispiel 3: Fragen, bei denen nur eine Antwort anzukreuzen ist. (Das sind die meisten Fragen...)	
B3	<p>Wie oft putzen Sie Ihre Zähne?</p> <p>(KREUZEN SIE NUR EIN KÄSTCHEN AN!)</p> <ul style="list-style-type: none"> <li><input type="checkbox"/> Täglich</li> <li><input type="checkbox"/> Nur wochentags</li> <li><input type="checkbox"/> Nur am Wochenende</li> <li><input checked="" type="checkbox"/> Ganz unregelmäßig</li> <li><input type="checkbox"/> Nur im Urlaub</li> </ul> <p>(GLEICH ZU FRAGE 99)</p>

Achtung:

Wenn Sie eine Antwort angekreuzt haben und dahinter steht dieses Hinweiszeichen, dann können Sie gleich bei der Frage weitermachen, deren Nummer angegeben ist. Im Beispiel 3 könnten Sie also gleich bei Frage 99 fortfahren.

Beispiel 4: Änderungsbeispiel: Wenn Sie bei den „Ankreuz-Fragen“ eine bereits angekreuzte Antwort korrigieren möchten, schwärzen Sie das Kästchen mit der „falschen“ Antwort bitte vollkommen aus. Danach kreuzen Sie dann die Antwort an, die Ihrer Meinung besser entspricht.					
B4	Wie zufrieden sind Sie insgesamt mit den Möglichkeiten zur Freizeitgestaltung?	Zufrieden	Eher zufrieden	Eher unzufrieden	Unzufrieden
		■	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>

Bitte benutzen Sie zum Ausfüllen des Fragebogens nur Schreibgeräte mit schwarzer oder dunkelblauer Farbe!

So - hier endet die Einleitung, und jetzt geht es richtig los!





### 3.4.2 Einige Details zur Entwicklung des Erhebungsinstruments

Das für die Schülerbefragungen des Projekts „Unsere Schule...“ genutzte Erhebungsinstrument setzt sich aus einer Vielzahl – teils neu entwickelter – Fragen zusammen, die vor ihrem Einsatz umfangreichen Gütetests und insbesondere Validierungsuntersuchungen unterzogen wurden. Die ursprüngliche Fragebogenfassung wurde auf diesem Wege mehrfach überarbeitet und gekürzt. Bei der Überarbeitung wurden die Wünsche und Anregungen der Landesvertretungen einzelner Bundesländer mitberücksichtigt. Praktische Umsetzungsprobleme ergaben sich zum einen daraus, dass der Fragebogen auch in Klassen mit einem hohen Anteil an ausländischen Schülerinnen und Schülern einsetzbar sein sollte. Zum anderen wurden in einigen Bundesländern auch Sonder- bzw. Förderschulen in das Projekt integriert. Das erste Problem konnte durch relativ einfache Maßnahmen bei der Gestaltung des Standardfragebogens gelöst werden; das zweite Problem bedurfte der Entwicklung eines separaten Erhebungsinstruments. Dies geschah in enger Zusammenarbeit mit den betreffenden Schulen. Da die übliche Erhebungsmethode (standardisierte, schriftliche Befragungen im Klassenverband) unter den besonderen Gegebenheiten an Förderschulen in der Regel nicht möglich war, wurden die Schülerbefragungen hier auf der Basis eines neu entwickelten Fragebogens und unter Einsatz speziell geschulter Sonderpädagogikstudierender erfolgreich durchgeführt. Der Fragebogen wurde hierzu sprachlich vereinfacht und verkürzt. Weiterhin wurde das Erhebungsinstrument so gestaltet, dass es nicht nur als schriftliche Fragebogenfassung, sondern auch im Rahmen von mündlichen „face-to-face“-Interviews eingesetzt werden konnte.

Eine vollständige Abbildung der Inhalte des Standardfragebogens muss hier aus Platzgründen unterbleiben. Die inhaltlichen Schwerpunkte der Schülerbefragungen können jedoch der folgenden Übersicht entnommen werden. Dabei ist zu beachten, dass die meisten der angeführten Inhalte nicht mit Einzelfragen, sondern mit neu gebildeten Skalen erfasst wurden; die diesbezüglichen Skalenwerte ergeben sich also stets aus den zusammengefassten Antworten von mehreren Einzelitems des Fragebogens. Genauere Hinweise zum angewendeten Verfahren der Skalenbildung werden nachfolgend noch dargestellt.

Die im Fragebogen des Projekts enthaltenen Formulierungen sind überwiegend keine Neuentwicklungen, sondern beruhen auf Items, die in früheren Jugendstudien bzw. Schülerbefragungen bereits erfolgreich eingesetzt wurden. Hierzu gehören vor allem die IFK-Jugendstudien (Sturzbecher, 1997 und 2001). Andere Einzelitems sind folgenden Publikationen bzw. Studien entnommen: dem Linzer Fragebogen zum Schul- und Klassenklima (Eder, 1998), den Landauer Skalen zum Sozialklima für Klasse 4-13 (v. Saldern & Littig, 1987), dem Schülerfragebogen des Teilprojekts B8 „Gewalt in der Schule“ (Holtappels & Tillmann et. al., 1999. Sonderforschungsbereich Prävention und Intervention im Kindes- und Jugendalter der Universität Bielefeld), den Instrumenten des Quasum-Projekts (Ministerium für Bildung, Jugend und Sport des Landes Brandenburg, 2000) und des PISA-Projekts (Schulleiterfragebogen) sowie dem Schülerfragebogen aus der Internationalen Studie zur Politischen Bildung der International Association for the Evaluation of Educational Achievement.

## Abbildung 8 Übersicht über Inhalte der Schülerbefragungen des Projekts „Unsere Schule...“

### *Personenbezogene Merkmale*

- Allgemeine Lebensziele, Berufsbezogener Zukunftsoptimismus
- Selbstvertrauen, Externale Kontrollüberzeugungen, Quietismus

### *Schwerpunktbereich „Jugenddelinquenz“*

- Ausländerfeindlichkeit, Rechtsextremismus, Antisemitische Vorurteile
- Schulgewalt, Gewaltakzeptanz, Häufigkeit von Gewalthandlungen
- Motive von Gewalttätern, Opfer von Jugendgewalt
- Bewaffnung von Schülerinnen und Schülern
- Schulvandalismus, Drogenkonsum, Fahren ohne Führerschein

### *Schwerpunktbereich „Schulqualität“*

- Durchschnittliche Noten der Schülerinnen und Schüler
- Zufriedenheit mit einzelnen Schulbereichen
- Schulspaß, Schulunlust, Schulstress/ Schulangst, Umfang und Gründe von Schulschwänzen
- Fachliche Lehrqualität, Soziale Lehrqualität, Zufriedenheit mit der Lehrerschaft
- Klassenzusammenhalt, Klassenzerrüttung
- Schulattraktivität, Schulverbundenheit
- Schülermitgestaltung, Schuldemokratie, Soziale Verantwortung

### 3.4.3 Maßnahmen im Vorfeld der Datenerhebungen

Die Datenerhebungen im Projekt „Unsere Schule...“ erfolgten über standardisierte schriftliche Schülerbefragungen im Klassenverband. Die Erhebungen wurden in zwei Wellen (2001 und 2003) durchgeführt. Die Befragungsergebnisse wurden den Projektschulen in Form von schriftlichen Schulreports sowie auf anschließenden Präsentationsveranstaltungen vorgestellt. Aufbauend auf diesen Ergebnissen schloss sich ein Fort- und Weiterbildungsprogramm an, das zur Qualifizierung von Steuerungsgruppen unter den Lehrkräften diente. Um Entwicklungstrends an den beteiligten Schulen im Projektzeitraum abschätzen zu können, wurde schließlich eine zweite Schülerbefragung realisiert.

Die Befragungen mussten zunächst von den zuständigen Ministerien und Datenschutzbeauftragten der 10 beteiligten Länder genehmigt werden. Nach der Genehmigung des Fragebogens und der Verfahren zur Datenerhebung und -speicherung wurden in allen teilnehmenden Ländern Informationsveranstaltungen abgehalten, die interessierten Schulen einen Einblick in das geplante Vorhaben geben sollten. Diese Auftaktveranstaltungen wurden von den beiden Projektträgern gemeinsam durchgeführt. Zeigten die Schulleiterinnen und Schulleiter im Anschluss an die Auftaktveranstaltung Interesse an einer Projektbeteiligung, so wurde ihnen schriftliches Informationsmaterial über die Ziele und Methoden sowie über den zeitlichen Projekttablauf zugestellt. Projektflyer, Genehmigungsbescheide von Landesbehörden, Fragebögen zur Einsicht und von den Projektträgern vorbereitete Aushänge informierten Schülerinnen und Schüler, Lehrerinnen und Lehrer, Eltern und Elternvertretungen über das Anliegen des Projekts; offene Fragen wurden in einer konstruktiven Zusammenarbeit geklärt. Die Entscheidungsfindung an den Schulen für oder gegen eine Projektteilnahme vollzog sich meist schrittweise und in

unterschiedlichen Gremien: War die Schulleitung von den Vorteilen einer Projektbeteiligung überzeugt, so wurde das Vorhaben in der Regel anschließend im Lehrerkollegium diskutiert. Formell wurde schließlich über die Projektteilnahme im Rahmen von Gesamtkonferenzen der Schulen entschieden.

Bei der Entscheidungsfindung waren sich die Schulen durchaus des zusätzlichen Arbeitsaufwands im Schulalltag bewusst, der sich vor allem durch die Teilnahme am Weiterbildungsprogramm des ibbw ergeben würde. In vielen Bundesländern konnten leider keine zusätzlichen Abminderungsstunden für die teilnehmenden Schulen bewilligt werden. Die Projektbeteiligung war also für die einbezogenen Lehrkräfte mit einem erheblichen Mehraufwand an Arbeit verbunden. Dennoch entschieden sich letztlich 218 Bildungseinrichtungen für eine aktive Mitwirkung im Projekt „Unsere Schule...“. Auf der Grundlage dieser Ausgangssituation konnte die erste Befragungswelle für das erste Halbjahr 2001 vorbereitet werden, die zweite Erhebungswelle wurde auf den Zeitraum Frühjahr/Sommer 2003 terminiert.

### **3.4.4 Die Befragungszeiträume der beiden Erhebungswellen**

Die Datenerhebungen an den Projektschulen wurden mit Beginn des Projekts im Frühjahr 2001 in Zusammenarbeit mit den Schulen des ersten Länderblocks vorbereitet. Am 15. Mai 2001 wurde in Mecklenburg-Vorpommern die erste Befragung durchgeführt. Es folgten bis zum Beginn der Sommerferien 2001 weitere Datenerhebungen in den Bundesländern Thüringen, Sachsen-Anhalt, Brandenburg und Niedersachsen. Mit Beginn des Schuljahres 2001/2002 wurden die Datenerhebungen in diesen Bundesländern fortgesetzt und kurz nach den Herbstferien 2001 abgeschlossen. Aufgrund eines längeren Projektbewilligungsverfahrens in Nordrhein-Westfalen konnten die dortigen Befragungen erst parallel zu den Erhebungen des zweiten Länderblocks beginnen.

Nach der Bewilligung des Erweiterungsantrags und dem Vorliegen der entsprechenden Genehmigungen auf Länderebene, wurden Ende der Herbstferien 2001 die Erhebungen für die Länder Baden-Württemberg, Bayern, Rheinland-Pfalz, Saarland (zweiter Länderblock) sowie für Nordrhein-Westfalen vorbereitet. Mit Beginn des Jahres 2002 fanden die ersten Datenerhebungen in diesen Bundesländern statt.

Bis zu den Sommerferien 2002 konnten auch die Befragungen zur ersten Erhebungswelle in den genannten Ländern weitgehend abgeschlossen werden. Für das Saarland ergab sich ein längerer Vorbereitungszeitraum, der mit einem schwierigeren Genehmigungsverfahren und der hieraus resultierenden Überarbeitung des Erhebungsmaterials in Verbindung stand. Deshalb konnten im Saarland die Befragungen erst im Frühherbst 2002 beginnen, die letzten Erhebungen zogen sich bis in das Kalenderjahr 2003 hin. Dies hatte zur Folge, dass sich die Vorbereitungen für die allgemeine zweite Erhebungswelle mit der Endphase der ersten Erhebungswelle im Saarland zeitlich überlagerten.

Der Abstand von der Erhebung (erste Welle) bis zur Übergabe der Ergebnisdarstellung an die Projektschulen (schriftlicher Schulreport) betrug im Durchschnitt vier Monate.

Die zweite Erhebungswelle konnte wie geplant im Frühjahr 2003 gestartet werden, jedoch nicht – wie vorgesehen – im Sommer 2003, sondern erst zum Jahresende 2003 abgeschlossen werden. Für die Projektschulen in Mecklenburg-Vorpommern, Brandenburg, Thüringen, Sachsen-Anhalt und Niedersachsen wurde dabei der im Rahmen der Projektkonzeption vorgesehene Zeitabstand von zwei Jahren zwischen den beiden Erhebungen fast vollständig eingehalten. Ausnahmen bildeten in diesem Zusammenhang 19 Schulen in Sachsen-Anhalt, sechs Schulen in Thüringen, acht Schulen in Niedersachsen, eine Schule in Mecklenburg-Vorpommern und eine Schule in Brandenburg. Einige der Nachzügler konnten insbesondere

in Niedersachsen, Nordrhein-Westfalen und im Saarland trotz des späten Projektbeitritts auch bei der zweiten Welle einbezogen werden. Hier ergaben sich hinsichtlich der geplanten Distanz zwischen den beiden Erhebungen (zwei Jahre) durch die Verzögerungen beim Projektbeitritt einige Probleme. Auch in diesen Fällen konnte jedoch gesichert werden, dass die Distanz zur ersten Datenerhebung mindestens ein Jahr betrug.

### **3.4.5 Auswahl und Schulung von Honorarkräften zur Datenerhebung**

Ein wichtiges Ziel bei der Anwerbung von Honorarkräften zur Datenerhebung war der Aufbau eines dezentralen Netzes von „Feldabteilungen“ an Standorten in mehreren Bundesländern. Dieser aufwändige Vorbereitungsprozess diente nicht nur zur Reduzierung der Reisekosten im Zuge der Datenerhebungen. Vielmehr war diese Maßnahme auch eine Voraussetzung zur Einhaltung der Termine der Schülerbefragungen an Schulen: Da die Befragungen meist unmittelbar nach Schulbeginn starteten, hätten mit einer zentralen Feldabteilung an entfernt gelegenen Schulen die Termine nur durch Anreisen am Vortag (und Übernachtungen) eingehalten werden können. Dies hätte zusätzliches Personal erfordert. In Mannheim, Leipzig, München, Göttingen und Potsdam wurden insgesamt fünf Feldabteilungen aufgebaut. Die Feldabteilung in Mannheim rekrutierte sich aus wissenschaftlichen Hilfskräften des Zentrums für Umfragen, Methoden und Analysen (ZUMA); in Leipzig wurden über persönliche Kontakte und Aushänge an der Universität studentische Honorarkräfte angeworben; in München konnte das IFK auf die Unterstützung einer ehemaligen Mitarbeiterin zurückgreifen; in Göttingen bildete sich das Erhebungsteam aus dem personellen Netzwerk des ibbw. Der größte Interviewerstab wurde in geografischer Nähe zum IFK aufgebaut. Alle Erhebungsteams wurden vor ihrem Einsatz anhand einer schriftlichen Anleitung eingehend geschult. Insgesamt waren bundesweit bis zu 30 Honorarkräfte bei der Datenerhebung tätig.

Eine besondere Herausforderung ergab sich durch die Einbeziehung von Fördereinrichtungen in das Projekt. Vor allem an Schulen mit verhaltensauffälligen Kindern und Jugendlichen war eine Durchführung der Befragungen im Klassenverband völlig ausgeschlossen. Um diesen Schulen dennoch eine Projektteilnahme zu ermöglichen, wurden über die Universität Rostock fortgeschrittene Studierende der Sonderpädagogik angeworben. Diese führten mit den Jugendlichen die Befragungen teils in Kleingruppen, teils in individuellen „face-to-face“-Interviews durch. Der Standardfragebogen wurde hierzu überarbeitet und gekürzt. Durch das große Engagement der Erhebungsteams und Lehrkräfte sowie nicht zuletzt auch aufgrund des erstaunlich hohen Interesses der Schülerinnen und Schüler der Fördereinrichtungen konnte das Projekt auch an diesen Schulen erfolgreich umgesetzt werden.

### **3.4.6 Die Auswahl der Klassen**

Als Grundgesamtheit für die Befragungen an den Projektschulen wurden zu Beginn der Projektplanungen zunächst Schülerinnen und Schüler der Jahrgänge sieben bis 13 an verschiedenen allgemein bildenden Schulformen definiert, wie sie in den jeweiligen Bildungseinrichtungen der einzelnen Bundesländer vertreten sind. Im Rahmen der Datenerhebungen an den Einzelschulen wurden aus Kostengründen jedoch nicht alle Klassen in die Befragung einbezogen. Um dennoch ein repräsentatives Abbild der Schülermeinungen an den Schulen zu erhalten, wurde die Auswahl der Klassen nicht den Schulen überlassen, sondern unterlag einem vom IFK vorgegebenen und von der Schulleitung nicht beeinflussbaren systematischen Auswahlsschlüssel. Bei der Auswahl der zu befragenden Klassen wurde in jeder Schule im Rahmen der ersten Erhebungswelle zunächst eine Klasse pro Jahrgang (ab Klassenstufe 7) für die Befragung ausgewählt. Existierten innerhalb eines

Jahrgangs mehrere Parallelklassen, so wurden in allen „ungeraden“ Klassenstufen (7, 9, 11, 13) jeweils die erste Klasse (z.B. die „7a“), in allen „geraden“ Klassenstufen (8, 10, 12) jeweils die höchste Parallelklasse befragt. Aufgrund der relativ hohen Ausfallquote wurde ab November 2001 die Zahl der zu befragenden Klassen pro Schule auf jeweils zwei Klassen pro Jahrgang (die namentlich erste sowie die namentlich letzte Klasse) erhöht. Dieses veränderte Auswahlverfahren betraf im Zeitraum der ersten Befragungswelle insbesondere die Projektschulen in Baden-Württemberg, Bayern, Rheinland-Pfalz, Saarland und Nordrhein-Westfalen. Für die zweite Befragungswelle wurde die Auswahl von zwei Klassen pro Jahrgang vollständig umgesetzt.

### **3.4.7 Weitere Maßnahmen im Vorfeld der Datenerhebungen**

Waren die in die Befragung einzubeziehenden Klassen ermittelt, galt es in einem nächsten Schritt, die Zustimmung der Eltern zur Teilnahme ihrer (minderjährigen) Kinder an der Befragung einzuholen. Hierzu wurden der Schulleitung Elternbriefe und schriftliche Einwilligungserklärungen für die Eltern zur Verfügung gestellt. Diese Materialien wurden in den ermittelten Klassen ausgeteilt. Die Elterninformationen enthielten Anschrift und Telefonnummern von Ansprechpartnern am IFK. Telefonische Rückfragen zu dem geplanten Vorhaben kamen seitens der Eltern nur selten; im Rahmen der Telefongespräche wurden vorhandene Fragen geklärt und Bedenken ausgeräumt. Allerdings machten einige wenige Eltern auch von ihrem Recht Gebrauch, den Datensatz ihres Kindes nachträglich löschen zu lassen. Das eigens aufgebaute Angebot für die Eltern, über die in den Schulsekretariaten ausliegenden Fragebögen Einsicht in die konkreten Inhalte der Befragung zu nehmen, wurde so gut wie gar nicht genutzt. Im Durchschnitt brachten etwa 60 bis 70 Prozent der Schülerinnen und Schüler die unterschriebenen Einverständniserklärungen ihrer Eltern in die Schule zurück. Nach dem endgültigen Rücklauf der schriftlichen Elternzustimmung wurde der Termin für die Befragung mit der Projektschule festgelegt und der Befragungsablauf abgestimmt. Parallel zu den Absprachen mit den Projektschulen wurden mit den Datenerhebungsteams die Befragungsabläufe an den Einzelschulen organisiert. Dabei wurde gesichert, dass die Befragung aller Klassen an einem einzigen Schultag abgeschlossen werden konnte.

### **3.4.8 Durchführung und Ablauf der Datenerhebungen**

Für die Befragungen standen an den Projektschulen unterschiedliche räumliche Bedingungen zur Verfügung. Vorzugsweise wurden größere Räume (beispielsweise Aula, Hörsaal, Musikzimmer, Speisesaal) für die Durchführung der Befragung genutzt. In anderen Schulen erfolgte die Befragung (Klasse für Klasse) in den jeweiligen Klassenräumen. Bevor die Schülerinnen und Schüler in den Befragungsablauf eingewiesen wurden, stellten sich die Mitarbeiter der Erhebungsteams den Jugendlichen vor und gaben ihnen einen kurzen Projektüberblick. In diesem Zusammenhang verwiesen sie auch auf die an der Schule ausliegenden Informationsmaterialien zum Projekt. Daran anschließend wurden die Erhebungsmaterialien erläutert und Instruktionen zum Ausfüllen des Fragebogens gegeben. Ein wichtiger Bestandteil dieser Einführung bildete stets der ausdrückliche Hinweis auf die Anonymität der Angaben sowie auf die Freiwilligkeit der Teilnahme an der Befragung. Dabei wurde sowohl auf die bereits unterschriebene Einwilligungserklärung der Eltern (von Minderjährigen) als auch auf die Relevanz der eigenen Einwilligung als Voraussetzung zur Teilnahme an der Befragung verwiesen: Der Fragebogen wurde nur von Schülerinnen und Schülern ausgefüllt, bei denen eine unterschriebene Erklärung der Eltern (bei Minderjährigen) vorlag und die ein separates Einwilligungsschreiben persönlich unterschrieben hatten.

Nachdem die Voraussetzungen für die Teilnahme an der Befragung geklärt und abgesichert waren, wurde das Ausfüllen des Fragebogens erläutert. Dazu wurden insbesondere die Beispiele auf der ersten Seite des Fragebogens genutzt. Anhand dieser Beispiele wurden die im Fragebogen enthaltenen Fragentypen erläutert. Jüngeren Probanden wurden die Beispiele zusätzlich an der Tafel vorgeführt. Bevor die Schülerinnen und Schüler mit dem Ausfüllen des Fragebogens begannen, wurde ihnen vom Befragungsleiter die jeweilige Schulnummer genannt, die auf der ersten Seite des Fragebogens einzutragen war. Diese Kennziffer war eine zentrale Voraussetzung für die Zuordnung der Schülerdaten zu einer bestimmten Schule. In der Regel wurde diese Nummer deshalb an die Tafel geschrieben oder in einer anderen Form (Plakat, Karton) den Schülerinnen und Schülern „narrensicher“ mitgeteilt. Erst danach begann das Ausfüllen des Fragebogens. Traten dabei Rückfragen oder erkennbare Unsicherheiten auf, wurden die Probanden individuell unterstützt.

Eine wichtige Funktion der Befragungsleiter bestand während des Ausfüllens der Fragebögen gelegentlich auch in der „Beaufsichtigung“ von anwesenden Lehrkräften. Vor allem die bei einigen Lehrerinnen und Lehrern vorhandene Angewohnheit, schreibenden Schülerinnen und Schülern über die Schulter zu schauen, wurde stets im Ansatz unterbunden; es wurde also stets darauf geachtet, dass weder anwesende Lehrkräfte noch andere Personen während der Befragung Einsicht in die Erhebungsunterlagen erhielten. Die Notwendigkeit dieser Maßnahme geboten nicht nur die Bestimmungen des Datenschutzes: Die Gültigkeit der Schülereinschätzungen ist gerade bei „sensiblen“ Themen vor allem von der Überzeugung der Befragten abhängig, dass die versprochene Anonymität der Angaben auch tatsächlich eingehalten wird. Auch während der Abgabe der Fragebögen wurde erkennbar gewährleistet, dass andere Personen (Schulleitung, Lehrer oder Eltern) keinerlei Einblick in die ausgefüllten Fragebögen erhielten. Abschließend wurde den Probanden für die Teilnahme an der Befragung ausdrücklich gedankt; Anfragen zum weiteren Projektverlauf wurden beantwortet.

Von den Befragungsleitern wurde für jede Klasse ein schriftliches Erhebungsprotokoll erstellt. Hier wurden Details aufgenommen, die für eine Einschätzung des Datenerhebungsprozesses von Bedeutung sein konnten (z.B. Anzahl der Befragten, Gründe für Ausfälle, Auffälligkeiten während der Erhebung, Inhalte von Rückfragen etc.).

### **3.4.9 Rückmeldungen nach der Befragung**

Unmittelbar nach Abschluss der Datenerhebung übermittelten die Befragungsleiter eine erste Einschätzung zum Befragungsablauf an das IFK. Darin wurden auch Schwierigkeiten und Probleme während der Befragung benannt, in der Mehrzahl wurden jedoch positive Rückmeldungen zum Befragungsablauf gegeben. Auf der Basis dieser Kurzinformationen wurde durch das IFK am folgenden Tag, gemeinsam mit den Projektschulen, die Befragung ausgewertet. Hinweise, Probleme, Erfahrungswerte zum Befragungsablauf wurden dabei dokumentiert. Diese Auswertungen konnten dann bei der Vorbereitung und Durchführung der zweiten Befragungswelle berücksichtigt werden.

Die Schülerinnen und Schüler, als wichtige Partner in dem Projekt, zeigten sich in allen Bundesländern der Befragung gegenüber meist aufgeschlossen. Ausnahmen konnten vor allem altersspezifisch (besonders in den 9. und 10. Klassen) und schulspezifisch (besonders in Haupt- und Berufsschulen) beobachtet werden. An der Befragung eher desinteressiert und ohne Motivation zeigte sich die Schülerschaft jedoch nur an wenigen Bildungseinrichtungen. Dies betraf von 218 teilnehmenden Projektschulen nur etwa 15 Fälle. Gelegentlich artikulierten die Jugendlichen ihre Meinung zur Befragung auf dem Fragebogen oder am Ende der Erhebung; dabei wurden vorrangig positive Bewertungen geäußert.

Das Feedback der Schulleitungen zur Vorbereitung und Durchführung der Befragungen war bis auf wenige Ausnahmen ebenfalls positiv. Viele Projektschulen bedankten sich schriftlich für die gute Zusammenarbeit und die reibungslose Durchführung der Datenerhebung.

#### **3.4.10 Einige Erfahrungsberichte aus den Erhebungsprotokollen**

In den neuen Bundesländern wurden seitens der Befragungsleiter teilweise andere Problemfelder in den Erhebungsprotokollen festgehalten als in den alten Ländern. So wurde vor allem in den neuen Bundesländern in einigen Schulformen eine Verbindung zwischen dem hohen Stand der Arbeitslosigkeit der Eltern und dem Verhalten der Schülerinnen und Schüler erwähnt; die betroffenen Schülerinnen und Schüler zeigten sich während der Befragung vergleichsweise unmotiviert bzw. unkonzentriert. Darüber hinaus war das beobachtbare Verhalten zu den Lehrerinnen und Lehrern relativ häufig unangemessen, was sich negativ auf die gesamte Befragungssituation auswirkte. An einer Berufsbildenden Schule mussten beispielsweise pro Klasse zwei bis drei Lehrkräfte bei der Befragung anwesend sein, um eine annehmbare Befragungssituation gewährleisten zu können. Allerdings wurde selbst unter diesen Bedingungen die Respektlosigkeit der Auszubildenden gegenüber ihren Lehrkräften offensichtlich.

Weiterhin berichteten die Befragungsleiter in den neuen Ländern häufiger von schlechten allgemeinen Rahmenbedingungen (Schulgebäude, Klassenräume und Schulhof machten zum Teil einen verwahrlosten Eindruck), wohingegen Lehrkräfte und Schulleitungen trotz widriger Arbeitsbedingungen oft recht kompetent und motiviert wirkten. An vielen Projektschulen wurde deshalb der Eindruck bestätigt, dass unzureichende Rahmenbedingungen keineswegs nur Resignation und Desinteresse hervorrufen müssen. Gerade einige der Projektschulen mit unzureichenden Bedingungen zeigten während der Feldphase ein großes Interesse an der Weiterentwicklung der Schulqualität an ihrer Schule. In solchen Fällen waren es eher die Auswirkungen von Schulreformen, ständigen Umstrukturierungen und Unsicherheiten im Bildungssystem, von denen Schulleitungen und Lehrkräfte unseren Befragungsleitern als negative Einflussfaktoren auf die Schulentwicklung berichteten.

An einigen Projektschulen in den alten Bundesländern ließ sich anhand der Erhebungsprotokolle eine besondere Problematik bei der Integration von Kindern aus Aussiedlerfamilien ausmachen. Laut Aussagen der Lehrkräfte bildeten diese häufig geschlossene Cliquen, von denen teilweise ein hohes Gewaltpotenzial ausgeht. Vereinzelt wurden bei der Vorbereitung des Befragungstermins unzureichende Lesefertigkeiten dieser Schülerinnen und Schüler erwähnt. Bei der Befragung selbst richteten sich die Befragungsleiter stets auf die von den Projektschulen vorab benannten Probleme und Schwierigkeiten ein. In der Regel verliefen deshalb die Befragungen auch an diesen Schulen weitgehend reibungslos.

Beim Verständnis des Fragebogens zeigten sich insgesamt keine hervorhebendwerten Unterschiede zwischen den neuen und alten Ländern. Verständnisprobleme zu bestimmten Inhalten des Fragebogens traten vorwiegend unter den Schülerinnen und Schülern an Hauptschulen sowie an einigen Berufsbildenden Einrichtungen auf. Auch mit dem Aufbau des Fragebogens (z.B. „Filterfragen“) waren die Probanden an diesen Schulformen gelegentlich überfordert; allerdings waren solche Schwierigkeiten eher in den unteren Klassen zu beobachten. Alles in allem erwies sich der eingesetzte Standardfragebogen für die meisten Befragten als verständlich.

Gelegentlich bezogen sich Nachfragen der Schülerinnen und Schüler auf die Themen „Datenschutz“ und „Ergebnispräsentation“. Die der Projektdurchführung zugrunde liegenden Richtlinien hinsichtlich des Datenschutzes wurden von den Schülerinnen und Schülern

teilweise sehr genau hinterfragt. Rückfragen dieser Art kamen natürlich nicht überraschend. Die Befragungsleiter waren im Hinblick auf Anfragen zum Datenschutz genau instruiert und konnten die Anonymität der Befragung deshalb plausibel und glaubwürdig vermitteln. Ebenso versuchten die Befragungsleiter, die auf die Befragung folgenden Schritte (Auswertung, Schulreport, Datenpräsentation, Fortbildung) verständlich darzulegen und damit den Nutzen der Projektteilnahme auch für die Schülerinnen und Schüler zu verdeutlichen.

### 3.5 Die Datenerfassung und -auswertung

#### 3.5.1 Zur Methode der maschinellen Erfassung der Fragebögen

Für eine schnelle und präzise Datenerfassung von ca. 50.000 Fragebögen mit einem Umfang von jeweils 16 Seiten kam eine manuelle Dateneingabe von Anfang an nicht in Betracht. Die Auslagerung dieser Aufgabe an eine externe Firma hatte jedoch unmittelbare Auswirkungen auf das Layout des Fragebogens, da technische und optische Voraussetzungen zum Einscannen der Einzelblätter erfüllt werden mussten. Allerdings haben Fortschritte sowohl bei der Hardware als auch bei der Software in diesem Bereich die Ansprüche an das notwendige Layout der gedruckten Fragebögen mittlerweile deutlich herabgesetzt.

Für die Datenerfassung der Fragebögen musste ein zuverlässiger Partner gefunden werden, der auf die oft kurzfristig anfallenden Projekttermine flexibel reagiert. Da für die Schülerbefragungen keine einheitliche Fragebogenfassung vorlag, sondern in Fördereinrichtungen, an einigen Berufsbildenden Schulen sowie an den teilnehmenden Schulen aus dem Saarland jeweils eigene Fragebogenfassungen zum Einsatz kamen, ergaben sich demzufolge auch hohe Anforderungen an eine zuverlässige Erfassung unterschiedlicher Vorlagen. Da die genaue Anzahl der notwendigen Fragebogenfassungen im Vorfeld der Datenerhebungen nie genau bekannt war und der Druck überzähliger Fragebögen vermieden werden sollte, wurde der Druck der Fragebögen stets kurzfristig und in mehreren Schüben in Auftrag gegeben. Folgende Aufgaben mussten demzufolge erfüllt werden:

- der individualisierte Druck und Versand der Fragebögen für einzelne Schulen auf Anforderung durch das IFK,
- die maschinelle Erfassung der anonymisierten Befragungsbögen und
- die Aufbereitung und Kontrolle der Daten zur Auswertung im IFK.

Die Firma DATEN.WERK GmbH ([www.daten-werk.de](http://www.daten-werk.de)) übernahm diese Aufgaben. Sie erwies sich als ein zuverlässiger Partner und war stets in der Lage, alle Anforderungen im eigenen Hause und mit der geforderten Qualität auszuführen. Die Erfassung der Fragebögen erfolgte mit Hochleistungsscannern und der Software FORMS der Firma Readsoft (Schweden) in Verbindung mit einer MS-SQL-Datenbank. Das Scannen der Dokumente erfolgte im Duplexverfahren mit einer Auflösung von 200 dpi. Alle Bilder wurden zum Zweck der späteren Archivierung gespeichert. Dazu wurden die Einzelbilder eines Fragebogens wieder zu einer Acrobat-Datei zusammengeführt. Nach der Interpretation der Markierungen (Ankreuzungen), Zahlen und Buchstaben wurden alle handgeschriebenen Buchstaben und Zahlen, wie auch die nicht eindeutig zu interpretierenden Markierungen manuell durch DATEN.WERK verifiziert. Nach dem Export der Daten aus der Datenbank wurden diese in das Softwarepaket SPSS eingelesen und weiterverarbeitet.

In einer ersten Kontrolle wurden die Fragebogennummern von allen Einzelseiten gelesen und verglichen. Bei Nichtübereinstimmung wurde nach manuellem Abgleich eine Zuordnung festgelegt. Der zweite Kontrolllauf identifizierte doppelte Nummern und eliminierte entsprechende Fälle. Nach einigen standardisierten Transaktionen zur Aufbereitung der Daten



für die Analyse wurden einige kritische Variablen, wie einzeln stehende Zahlen, einem Bereichstest unterzogen. Dabei wurden alle außerhalb plausibler Bereiche liegende Werte gelistet, an Hand der archivierten Bilder geprüft und ggf. korrigiert. Alle Transaktionen über die Rohdaten bis hin zur übergebenen SPSS-Datendatei wurden reproduzierbar und kontrollierbar über Jobs ausgeführt. Stichprobenweise Tests ergaben dabei eine Fehlerrate von unter einem Prozent. Nach Ablauf von drei Monaten nach der Datenübergabe wurden die Datensätze sowie die Sicherungen der Abbilder aller Fragebögen bei der Firma DATEN.WERK gelöscht.

### **3.5.2 Datenauswertung**

Die Grundlage für die im Projekt „Unsere Schule...“ durchgeführten Schulevaluationen bildeten die Ergebnisse der Schülerbefragungen an den Einzelschulen. Für den sich hierauf aufbauenden Evaluationsprozess wurden die vom Joint Committee on Standards for Educational Evaluation formulierten Standards für Evaluationen zugrunde gelegt (s. o.). Der Umsetzung von Nützlichkeitsstandards, Durchführbarkeitsstandards, Korrektheitsstandards und Genauigkeitsstandards liegen dabei allerdings keine unveränderlichen Richtlinien zugrunde; sie sind in der Praxis stets auf die Zielgruppe auszurichten. Diese Zielgruppe bildeten im Rahmen des Projekts „Unsere Schule...“ zwar an erster Stelle die beteiligten Schulpädagogen. Es war jedoch zu berücksichtigen, dass auch interessierte Eltern sowie Schülerinnen und Schüler in den Evaluationsprozess einbezogen werden konnten. Deshalb zeichnete sich im Vorfeld der Planungen zu den einzusetzenden Methoden der Datenauswertungen ein gewisses Spannungsfeld zwischen den Kriterien der Nützlichkeit (z.B. die Informationsbedürfnisse der Nutzer) und der Genauigkeit (z.B. die Einhaltung wissenschaftlicher Standards) ab.

Unter Berücksichtigung des Umstands, dass der Nutzerkreis der Evaluationsergebnisse fast ausschließlich aus statistischen Laien bestand, konnte der Versuch einer weitestgehenden Einhaltung aller Evaluationsstandards in der Praxis nur durch ein mehrstufiges Verfahren gelingen. So wurde im Zuge der ersten Datenpräsentation, d.h. bei der Ausarbeitung des schriftlichen Schulreports, zunächst vollständig auf die Anwendung komplexerer (z.B. multivariater) Datenanalyseverfahren verzichtet. Die Ergebnisse der Schulreports wurden ausschließlich in Form einfacher statistischer Maßzahlen dargestellt, deren (Selbst-)Verständnis auch bei einem Laienpublikum keine größeren Schwierigkeiten erwarten ließ. Dennoch sollten die Möglichkeiten einer deskriptiven Datenanalyse selbstverständlich so weit wie möglich ausgeschöpft werden: Anhand von Mittelwerten wurden alle relevanten Informationen zu einem Thema in einem einzigen Zahlenwert zusammengefasst. Da durch solche Zusammenfassungen meist auch Detailinformationen verloren gehen, wurden stets auch prozentuale Verteilungen und Streuungswerte angegeben; weiterhin wurden die Ergebnisse von Teilgruppenanalysen dargestellt. Zum besseren Verständnis der Ergebnisdarstellung wurden auch die Methode und die Vorteile der Skalenbildung im Schulreport erläutert. Schließlich enthielt der Anhang der Schulreports nähere Erläuterungen zu allen eingesetzten Methoden und zur Ergebnisdarstellung.

Der schriftliche Schulreport stellte allerdings nur den ersten Schritt der Ergebnispräsentation dar. Die Einzelschulen konnten sich anhand dieser Vorlage einen Überblick über die Meinungen der Schülerinnen und Schüler zu einer Vielzahl von Themen verschaffen, meist blieb aber eine Reihe von Detailfragen durch den Report unbeantwortet. Die Klärung solcher Zusatzfragen sowie eine stärker auf die spezifische Einzelsituation der Schule hin ausgerichtete Ergebnisauswertung der Daten sollte im Rahmen einer mündlichen Datenpräsentation erfolgen. Der entsprechende Vortrag wurde deshalb im Vorfeld mit der Schule inhaltlich genau abgestimmt. In diesem Zusammenhang wurden seitens der Schulen

oft Fragen aufgeworfen, deren Beantwortung über eine rein deskriptive Datenanalyse nicht möglich war. Im Rahmen der mündlichen Datenpräsentation wurden deshalb bei Bedarf auch die Ergebnisse zusätzlicher multivariater Analysen dargestellt.

## 3.6 Die Datenpräsentation

### 3.6.1 Die Zielgruppen der Datenpräsentationen

Das Hauptanliegen des Projekts war die Förderung Sozialer Schulqualität durch die Kombination einer schulinternen Evaluation mit einem an den Ergebnissen dieser Evaluation orientierten flexiblen Weiterbildungsprogramm für Lehrkräfte. Dieses Vorhaben konnte nur durch den Aufbau eines Evaluationsangebots gelingen, das sowohl dem speziellen Informationsbedarf von Schulen als auch bestimmten Merkmalen der betreffenden Zielgruppen Rechnung trug. Diesbezügliche Überlegungen führten zu dem bereits skizzierten zweistufigen Präsentationsmodell, bei dem den beteiligten Schulen die Ergebnisse der Schülerbefragungen zunächst in Form eines schriftlichen Schulreports zur Verfügung gestellt wurden. Anschließend wurden die Ergebnisse durch mündliche Datenpräsentationen an den Einzelschulen erläutert und diskutiert.

Die Beeinflussbarkeit von Schulentwicklungsprozessen durch die Präsentation von Evaluationsergebnissen wächst mit dem Kreis der Personen, die Zugang zu diesen Ergebnissen erhalten. Bei der Projektplanung wurde deshalb eine Präsentationskonzeption ausgearbeitet, die als mögliche Zielgruppen nicht nur die Schulleitungen und Lehrkräfte sondern auch interessierte Schülerinnen und Schüler sowie Eltern vorsah. Weiterhin musste die Konzeption berücksichtigen, dass bei einem Großteil der relevanten Zielgruppen nicht von nennenswerten Vorkenntnissen im Bereich der Datenanalyse und Statistik auszugehen war. Bei der Präsentation der Daten musste deshalb stets eine Gratwanderung gelingen, die sich zwischen den Vorgaben wissenschaftlicher Methodik und dem Gebot leichter Verständlichkeit bewegte. Dabei lag die eigentliche Herausforderung darin, die Befunde in einer leicht verständlichen und möglichst unterhaltsamen Form zu vermitteln sowie das Publikum schnell in den Präsentationsprozess einzubinden. Besonders im Falle schlechter Ergebnisse war resignativen Tendenzen im Lehrerkollegium durch einen motivierenden Präsentations- und Diskussionsstil entgegenzuwirken.

Weiterhin war im Falle schlechter Resultate vor allem seitens der Schulleitung und der Lehrkräfte sowohl mit Zweifeln an der Aussagefähigkeit von Statistiken im Allgemeinen als auch an den Ergebnissen von Schülerbefragungen im Speziellen zu rechnen. Um Einwände dieser Art ausräumen zu können, musste die Datenpräsentation nicht nur in methodischer Hinsicht überzeugen. Die Ergebnisse der Schülerbefragungen wurden zusätzlich bereits im Vorfeld der Datenpräsentationen auf mögliche Auffälligkeiten im Antwortverhalten der Befragten hin untersucht. Eine Häufung von sich inhaltlich widersprechenden Meinungen eines Probanden oder ein konstant monotones Antwortverhalten führten letztlich zum Ausschluss des entsprechenden Fragebogens aus der Auswertung. Allerdings machte der Anteil unbrauchbarer Fragebögen je Schule selten mehr als zwei bis drei Prozent aus.

Die mündliche Datenpräsentation beschränkte sich auf ausgewählte Evaluationsthemen, um einer Überforderung der Zielgruppen durch ein Überangebot an Ergebnissen vorzubeugen. Darüber hinaus war davon auszugehen, dass für eine exakte Ausarbeitung der individuellen Qualitäts- bzw. Problemprofile an den meisten Einzelschulen letztlich nur eine relativ kleine Auswahl von Themenbereichen notwendig sein würde. Von Bedeutung bei der Auswahl der Präsentationsthemen waren die unterschiedlichen Interessen einzelner Zielgruppen. So zeigte

sich, dass Schulleitung und Lehrkräfte vor allem Interesse an Themen formulierten, die in unmittelbarem Zusammenhang mit ihrer Arbeit standen (Unterrichtsqualität, Schüler-Lehrerverhältnis, Schulverweigerung usw.). Die Eltern- und Schülerschaft, sofern sie in die Datenpräsentation einbezogen wurden, richteten ihre Aufmerksamkeit hingegen auch auf die Ergebnisse zum Schwerpunktbereich „Jugenddelinquenz“ (Gewalt, Extremismus usw.).

### **3.6.2 Der schriftliche Schulreport**

In einem ersten Schritt wurden die Ergebnisse der Schülerbefragung den Einzelschulen in Form eines schulbezogenen Schulreports zur Verfügung gestellt: Jede Schule erhielt drei gedruckte Exemplare sowie den gesamten Report als PDF-Datei. Letztlich verfügten die Schulen über diese Reports zwar stets nach eigenem Ermessen; geplant war in der Projektkonzeption allerdings, dass ein Exemplar der Schulleitung, ein Exemplar dem Lehrerkollegium und das dritte Exemplar insbesondere interessierten Schülerinnen und Schüler sowie deren Eltern zur Verfügung stehen sollte. Bei Bedarf konnten die Schulen darüber hinaus die PDF-Datei als Basis zur Herstellung weiterer Arbeitsmaterialien (z.B. für den Druck von Folien oder für zusätzliche Belegexemplare) nutzen.

Die Aufgabe des schriftlichen Schulreports bestand vor allem darin, der Schule einen ersten, eigenständigen Zugang zu den Ergebnissen der Schülerbefragung zu ermöglichen. Im Einzelfall reichten bereits das Studium des Reports und die anschließende schulinterne Diskussion über die Bedeutung der Einzelresultate sowie über die hieraus abzuleitenden Konsequenzen aus, um Impulse für einen Schulentwicklungsprozess zu geben. In solchen Fällen nahm die anschließende mündliche Datenpräsentation durch das IFK meist nur eine ergänzende und unterstützende Nebenrolle im Rahmen einer sich selbst entfaltenden Entwicklung ein. Dies war aber wie erwartet nicht der Regelfall.

Um Schwierigkeiten bei der eigenständigen Aufarbeitung der Inhalte des schriftlichen Schulreports vorzubeugen, wurden bereits bei der Projektplanung Erarbeitungskriterien festgelegt. Dazu gehörten:

- eine verständliche Darstellungsform der Datenpräsentation,
- eine hohe aber überschaubare Informationsdichte,
- eine theoriebasierte aber auch praxisorientierte Interpretation der Daten,
- eine nachvollziehbare Inhaltsstruktur,
- eine wissenschaftlich fundierte Einführung zu jedem Themenkomplex,
- eine auch für statistische Laien selbsterklärende Form der Textgestaltung und
- ein optisch ansprechendes Layout.

Die Entwicklung der Struktur des Schulreports wurde bis September 2001 abgeschlossen. Die Handhabung des Instruments erfolgte jedoch nicht in allen Bundesländern einheitlich. So wurden beispielsweise, gemäß einer Übereinkunft mit dem Ministerium für Bildung, Jugend und Sport (MBS) in Brandenburg, Vertreterinnen und Vertreter der beteiligten Schulen in einer speziellen Weiterbildungsveranstaltung durch das IFK zum Inhalt der Reports sowie in der Handhabung statistischer Analyseverfahren geschult. Leider erwiesen sich solche vorbereitenden Maßnahmen sowohl aus zeitlichen als auch aus organisatorischen Gründen nicht in allen Bundesländern als durchführbar. Der Inhalt des schriftlichen Schulreports kann der folgenden Übersicht 2 entnommen werden. Es ist dabei unschwer zu erkennen, dass der thematische Schwerpunkt des Schulreports auf den Bereich „Schulqualität“ gelegt wurde.

## Abbildung 9 Übersicht zum Inhalt des schriftlichen Schulreports

1. Schwerpunkt: „Jugenddelinquenz“		
„Ausländerfeindlichkeit“	}	<ul style="list-style-type: none"> <li>• Themen zur Darstellung des Umfangs fremdenfeindlicher und extremistischer Einstellungen.</li> <li>• Verbreitung von Gewalt akzeptierenden Meinungen sowie von Gewalthandlungen.</li> <li>• Bedeutung der Gewalttätigkeiten vor dem Hintergrund eines möglichen Waffeneinsatzes.</li> </ul>
„Rechtsextremismus“		
„Antisemitismus“		
„Gewaltakzeptanz“		
„Gewalthandeln“		
Bewaffnung von Schülern		
2. Schwerpunkt: „Schulqualität“		
Zufriedenheit mit einzelnen Schulbereichen	}	<ul style="list-style-type: none"> <li>• Übersicht zum Zufriedenheitsniveau der Schülerinnen und Schüler mit zentralen Schulbereichen.</li> </ul>
„Schulspaß“	}	<ul style="list-style-type: none"> <li>• Gefühle von Schülerinnen und Schüler in Verbindung mit dem sozialen Umfeld der Schule.</li> <li>• Mögliche Folgen von negativen oder belastenden Empfindungen: Schulverweigerung</li> </ul>
„Schulunlust“		
„Schulstress/ Schulangst“		
Schulschwänzen		
„Fachliche Lehrqualität“	}	<ul style="list-style-type: none"> <li>• Einschätzung der Unterrichtsqualität aus der Perspektive von Schülerinnen und Schüler.</li> </ul>
„Soziale Lehrqualität“		
„Zufriedenheit mit der Lehrerschaft“	}	<ul style="list-style-type: none"> <li>• Wahrnehmung der Qualität der sozialen Beziehungen zwischen Lehrerschaft und Schülerschaft sowie innerhalb der Schülerschaft.</li> </ul>
„Klassenzusammenhalt“		
„Klassenzerrüttung“		
„Schulattraktivität“	}	<ul style="list-style-type: none"> <li>• Wahrnehmung und Bewertung institutioneller Rahmenbedingungen von Schule.</li> <li>• Folgen von günstigen/schlechten Einschätzungen dieser Schulqualitätsmerkmale.</li> </ul>
„Schülermitgestaltung“		
„Schuldemokratie“		
„Schulverbundenheit“		
„Soziale Verantwortung“		
„Schulvandalismus“	}	<ul style="list-style-type: none"> <li>• Wahrgenommener Umfang von Vandalismus und Gewalt innerhalb des letzten Schuljahres.</li> </ul>
„Schulgewalt“		

Eine besondere Zielstellung bei der Konzeption des Schulreports lag darin, mit diesem Arbeitsmaterial auch statistischen Laien einen weitgehend selbsterklärenden Zu- und Umgang mit dem Datenmaterial zu ermöglichen. Die beiden Schwerpunkte „Delinquenz“ und „Schulqualität“ wurden deshalb jeweils mit einer Einleitung begonnen, die eine kurze punktuelle Aufarbeitung des aktuellen Forschungsstands zum Inhalt hatten. Durch diese Maßnahme sollte das zum Verständnis und zur Einordnung bestimmter Detailbefunde notwendige Hintergrundwissen vermittelt werden. Auch die Unterkapitel zu den Einzelthemen begannen mit einer knappen Einführung, die unter anderem methodische Erläuterungen zur Bildung des Untersuchungsmerkmals und/oder Verweise auf die Ergebnisse anderer Studien enthielten.

Methodische Erklärungen zu den in den Reports genutzten statistischen Angaben und Grafiken waren Gegenstand einer Beispielseite, die zum besseren Verständnis des Schulreports beitragen sollte. Auf dieser Beispielseite wurden die für Laien am schwersten zu verstehenden statistischen Angaben und Abbildungen des Reports dargestellt; ergänzende detaillierte Erläuterungen zu den verwendeten statistischen Verfahren fanden interessierte Leserinnen und Leser im Anhang des Schulreports.

Der grafische Aufbau der Ergebnisdarstellung wurde für fast alle Einzelthemen beibehalten und folgte dem Aufbau der Beispielseite. Der auf diese Weise eintretende Lerneffekt sollte den Umgang mit dem Datenmaterial erleichtern und zu einem besseren Verständnis von zentralen Ergebnissen beitragen. Das allgemeine Prinzip der Ergebnisdarstellung wird anhand des Beispiels „Ausländerfeindlichkeit“ auf der folgenden Seite verdeutlicht.

Abbildung 10 Ausschnitt aus dem Schulreport

Skala:

## Ausländerfeindlichkeit

7 Einzelaussagen – Bsp.:

„Deutschland den Deutschen – Ausländer raus.“

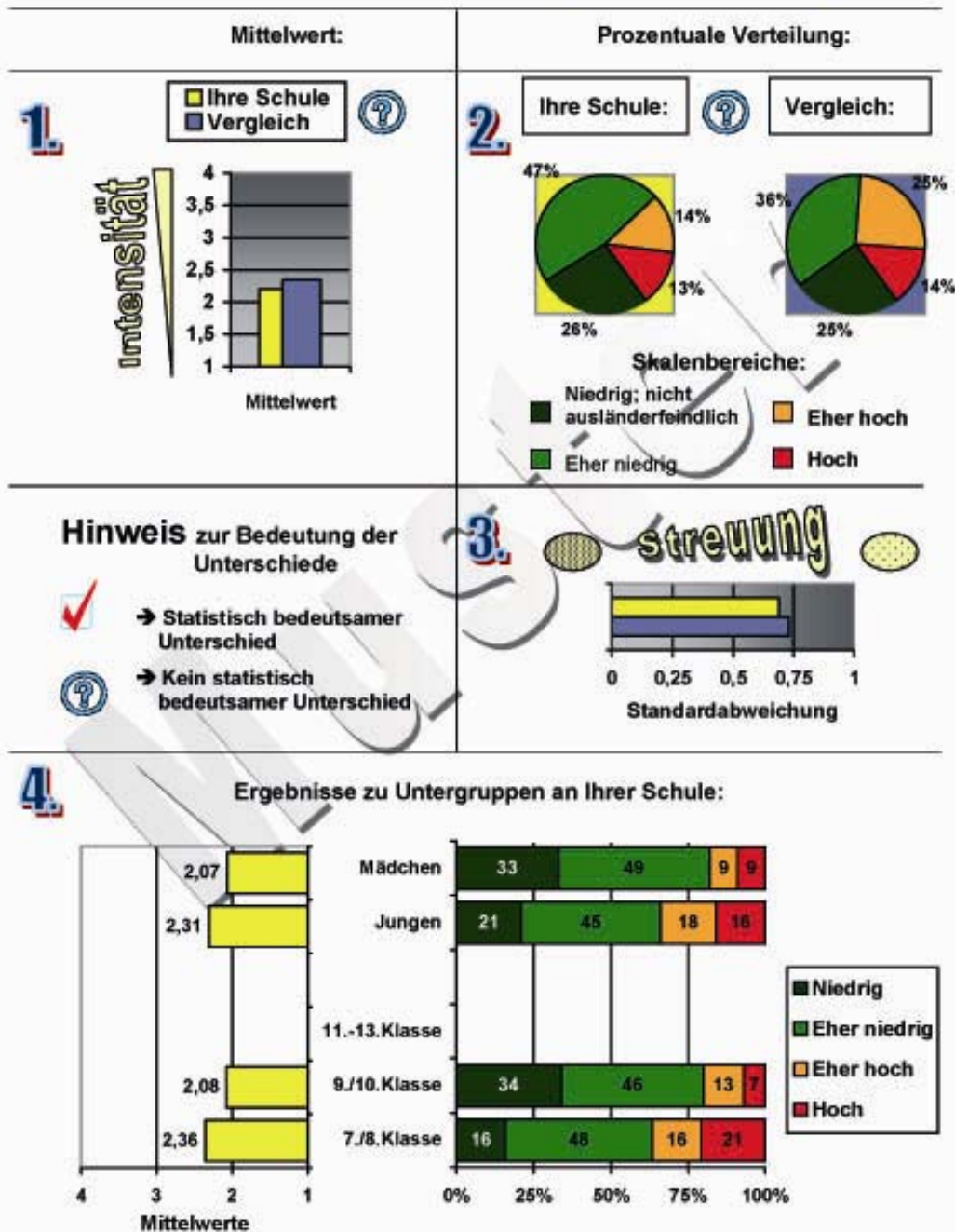
Antwortvorgaben:

4 – „Stimmt völlig“

3 – „Stimmt ziemlich“

2 – „Stimmt wenig“

1 – „Stimmt gar nicht“



In der ersten Zeile weist die Angabe „Skala“ darauf hin, dass es sich bei den Ergebnissen dieser Seite nicht um die Antworten auf Einzelfragen aus dem Fragebogen handelt, sondern dass das Merkmal „Ausländerfeindlichkeit“ eine neu gebildete Variable darstellt, die sich aus den zusammengefassten Angaben zu sieben Einzelaussagen des Fragebogens zusammensetzt. Dabei wird stets eine der hierbei zugrunde liegenden Einzelaussagen als Beispiel ausgewiesen. Schließlich werden auch noch die Antwortvorgaben angeführt, die den Jugendlichen im Fragebogen zur Verfügung standen, um ihre Einstellung zu den jeweiligen Einzelaussagen zum Ausdruck zu bringen.

Die Grafiken in den Feldern 1 bis 4 beinhalten die eigentlichen Ergebnisse zu den jeweiligen Einzelthemen. Neben statistischen Lagemaßen sollten in die Ergebnisdarstellung auch Verteilungsinformationen einfließen. Die genutzten statistischen Informationen (Mittelwerte, Prozentverteilungen, Streuungsmaße) erwiesen sich in der Regel sowohl von ihrer inhaltlichen Bedeutung als auch in der genutzten grafischen Darstellung (Balken- und Kreisdiagramme) für die Leser des Reports als weitgehend selbsterklärende Informationen. Auch die farbige Darstellung der Ergebnisse wurde von fast allen Nutzern als Unterstützung beim Umgang mit dem Arbeitsmaterial bezeichnet.

Feld 1 stellt den Mittelwert einer Skala für die betreffende Schule (gelber Balken) und für die Vergleichsschulen des jeweiligen Bundeslandes gleichen Schultyps dar (blauer Balken). Der Mittelwert ist eine der wichtigsten statistischen Maßzahlen und leicht zu interpretieren: Je länger die Balken sind bzw. je näher ein Mittelwert an den Skalen-Maximalwert „4“ reicht, desto stärker ist der erfasste Sachverhalt an den Schulen ausgeprägt. Feld 2 stellt die prozentuale Verteilung einer Skala für die Einzelschule und für die Vergleichsschulen dar. Hierzu wurde der Wertebereich der Skalen in vier gleichgroße Abschnitte unterteilt: Das unterste Viertel wurde mit „Niedrig“, das höchste Viertel mit „Hoch“ benannt. Dargestellt werden die Anteile der Befragten in jedem der vier Skalenbereiche. Das „Hinweisfeld“ enthält eine Erklärung zur Bedeutung der Unterschiede zwischen der Einzelschule und den Vergleichsschulen. Nicht alle Unterschiede, die augenscheinlich erkennbar sind, haben auch unter der „Lupe“ statistischer Prüfverfahren Bestand. Rote Häkchen bzw. blaue Fragezeichen in den Feldern 1 und 2 signalisierten die Ergebnisse dieser Tests.

Feld 3 stellt die Streuung (Standardabweichung) des Antwortverhaltens dar. Dieses Maß drückt aus, wie stark die Antworten der Befragten „streuen“; also ob die Befragten eher ähnlich geantwortet haben oder ob die Meinungen in der Schülerschaft weit auseinander liegen. Diese Information sollte vor allem Aufschluss darüber geben, ob innerhalb der befragten Schülerschaft substanzielle Meinungsunterschiede auszumachen sind. War dies der Fall, so lieferte die nach Teilgruppen differenzierte Darstellung in Feld 4 meist interessante Informationen zu den Hintergründen dieser Ergebnisunterschiede. Die gelben Balken auf der linken Seite stellen die Mittelwerte der Skala für jede der Untergruppen dar. Auf der rechten Seite der Abbildung werden die prozentualen Anteile der vier Skalenausprägungen („Niedrig“ bis „Hoch“) für jede Untergruppe angeführt. Die Angaben für alle Untergruppen sind also zeilenweise zu lesen.

Ein Blättern von Seiten sollte bei der Lektüre der Ergebnisse vermieden werden. Die (beidseitig bedruckten) Schulreports wurden deshalb so gegliedert, dass die Ergebnisseite zu einem Thema stets auf einer Rückseite und die Erläuterungen zu dieser Ergebnisseite auf der gegenüberliegenden Vorderseite platziert war. Ergebnisse und Erläuterungen zu einem Thema waren für die Leser also stets auf einen Blick zu erfassen; auch dies sollte zur Erleichterung des Umgangs mit dem umfangreichen Datenmaterial beitragen.

### **3.6.3 Die mündliche Datenpräsentation**

Nach der Zustellung des schriftlichen Schulreports hatten die betreffenden Schulen jeweils ein bis zwei Monate zur Verfügung, sich eigenständig mit den Ergebnissen der Schülerbefragung

vertraut zu machen. Anschließend wurde mit dem IFK ein Termin für eine mündliche Datenpräsentation vereinbart. Meldeten sich die Schulen nicht aus eigenem Antrieb (was durchaus vorkam), so vollzogen sich die Kontaktaufnahme und die Terminabsprache auf Initiative des IFK. Die Durchführung dieser Veranstaltung verfolgte mehrere Zwecke. Zum einen sollte den Schulen die Möglichkeit gegeben werden, eigene Interpretationen und Schlussfolgerungen aus den Daten ihres Schulreports mit der Perspektive eines Außenstehenden zu vergleichen. Zum anderen sollten mit dieser Präsentationsveranstaltung auch möglicherweise vorhandene Zweifel hinsichtlich der Zuverlässigkeit der Daten ausgeräumt und ein (insbesondere in Verbindung mit schlechten Befunden) motivierender Impuls auf das Lehrerkollegium gegeben werden. Schließlich diente die Veranstaltung noch dazu, die Projektphase der ersten Datenauswertung durch das IFK vorerst abzuschließen und die teilnehmenden Schulen in die nun anschließende Weiterbildungsphase des Projekts zu begleiten. Diese Aufgabe wurde am gleichen Tag, im Anschluss an die mündliche Datenpräsentation, durch eine/n ebenfalls anwesende/n Vertreter/in des ibbw übernommen.

Die Durchführung der Präsentationsveranstaltung wurde nicht nur terminlich, sondern auch inhaltlich mit den einzelnen Schulen abgestimmt. Dabei standen bei der Vorbereitung der Veranstaltung vor allem die jeweiligen Interessen der Schulen im Vordergrund. Viele der untersuchten Themen der Schülerbefragungen betrafen höchst sensible Schulbereiche, weshalb die Schulen die organisatorischen Rahmenbedingungen der Veranstaltung selbst festlegten. Zwar wurde den Schulen stets nahe gelegt, die Präsentation unter Einschluss einer möglichst breiten Schulöffentlichkeit – d.h. vor allem unter Einbeziehung von Eltern und der Schülerschaft – durchzuführen. Dennoch traf diese öffentliche Form der Datenpräsentation nur für eine Minderzahl der Veranstaltungen zu. Dies hatte eine Reihe von Ursachen: In vielen Fällen äußerte die Schulleitung den direkten Wunsch, die Ergebnisse vorerst nur innerhalb des Lehrerkollegiums zu diskutieren, um anschließend im Rahmen einer zweiten Präsentationsveranstaltung die Eltern und die Schülerschaft mit einzubeziehen. Doch auch die ausdrücklich gewünschte Einbeziehung von Eltern- und Schülerschaft führte nicht immer zu einer größeren Verbreitung der Ergebnisse der Befragungen: Oft war das Interesse der Eltern sowie der Schülerinnen und Schüler an den Ergebnissen einfach nicht vorhanden, trotz offizieller und fristgerechter Einladung durch die Schulleitung und trotz der Terminlegung auf Abendstunden oder auf Wochenenden. Selbst Eltern- und Schülervvertretungen waren an diesen Terminen meist unvollständig zugegen.

In der Regel gingen die teilnehmenden Schulen bei der Verbreitung der Ergebnisse ihrer Schülerbefragungen eher zurückhaltend vor. Vor allem in den ostdeutschen Bundesländern versuchten jedoch einige Schulen mit guten Befragungsergebnissen, diese als Mittel gegen geplante Schulschließungen zu nutzen. Eine Reihe von Präsentationsveranstaltungen fand deshalb in Anwesenheit der (meist örtlichen) Presse statt.

Die meisten mündlichen Präsentationsveranstaltungen fanden im Rahmen von Schulkonferenzen statt oder wurden als Bestandteil von „SchiLf-Tagen“ (Schulinterne Lehrerfortbildung, Pädagogische Tage) eingeplant. Um Ermüdungserscheinungen vorzubeugen, wurde für die eigentliche Datenpräsentation ein Zeitrahmen von maximal 45 Minuten festgesetzt. Für die Diskussion und für die Beratung der Schulen waren weitere 45 Minuten vorgesehen; dieser Zeitrahmen wurde in der Regel nicht überschritten. Im Anschluss an die mündliche Datenpräsentation stellte ein Vertreter des ibbw die Inhalte, die Struktur und den Ablauf des folgenden Weiterbildungsprogramms vor. Die Schulen hatten auf diese Weise die Möglichkeit, bei der späteren Auswahl der Fortbildungsmodule sowohl die Ergebnisse der Schülerbefragungen zur Schwerpunktsetzung zu nutzen als auch die Themen der Fortbildung eher nach den jeweils vorhandenen Interessenlagen innerhalb des Lehrerkollegiums zu bestimmen.



Im Rahmen der ersten Welle der Datenerhebung wurden in den 10 Bundesländern für die teilnehmenden Schulen 218 Schulreports angefertigt und den betreffenden Schulen zugestellt. In fast allen Fällen erfolgte anschließend eine Präsentationsveranstaltung unter Beteiligung des IFK; nur wenige Schulen verzichteten auf das weiterführende Informationsangebot einer mündlichen Datenpräsentation. Vielmehr wurde an vielen Schulen die Datenpräsentation vor dem Lehrerkollegium, den Eltern und/oder vor den Schülerinnen und Schülern mehrfach durchgeführt. Bei mehr als 40 Schulen kam es weiterhin vor, dass die Datenauswertungen eine Reihe von neuen Fragen aufwarfen, die nur durch detaillierte Zusatzauswertungen beantwortet werden konnten. Der Aufwand solcher Zusatzauswertungen lohnte sich jedoch in fast allen Fällen, da die betreffenden Schulen die Ergebnisse der Schülerbefragungen zur Verbesserung der Schulqualität intensiv nutzten.

### 3.7 Ergänzende Maßnahmen und Leistungsangebote des Projekts

Ein Projekt, das das Ziel verfolgt, soziale Schulqualität zu fördern und dabei ein Dienstleistungsanbieter für Schulen zu sein, kann über einen Zeitraum von mehreren Jahren seine Strukturen kaum unverändert beibehalten. Vielmehr wird eine gute Projektkonzeption durch die Feststellung des jeweiligen Einzelbedarfs und den sich (teilweise unterschiedlich) entwickelnden Interessen von Kooperationspartnern stets einen eigenen Entwicklungsprozess durchlaufen. Dies traf auch für das Projekt „Unsere Schule...“ zu. Einige Details dieser Entwicklung wurden bereits an anderer Stelle erwähnt: So wurde beispielsweise der ursprüngliche Plan, nur allgemein bildende Schulen in das Projekt einzubeziehen, schnell aufgegeben und auch Berufsbildenden Einrichtungen sowie Förder- und Sonderschulen eine Projektbeteiligung ermöglicht. Die hierfür notwendigen Mehraufwendungen (Entwicklung neuer Erhebungsmaterialien, Schulung geeigneter Befragungsleiter usw.) mussten allerdings ohne zusätzliche Projektmittel umgesetzt werden. Als unerwartet hoch erwies sich auch der Auswertungsaufwand für viele der teilnehmenden Schulen: Nicht selten reichten weder der schriftliche Schulreport noch die mündliche Datenpräsentation aus, um den Informationsbedarf der Schulen zu decken. Erst Detailanalysen zu bestimmten Fragestellungen führten dann zu zufrieden stellenden Ergebnissen. Da der Wunsch nach solchen Einzelauswertungen jedoch gleichzeitig den hohen Stellenwert zum Ausdruck brachte, den manche Schulen den Ergebnissen der Schülerbefragungen einräumten, wurde entsprechenden Anfragen seitens der Schulen stets gern entsprochen. Im den folgenden Abschnitten sollen einige andere, weiterführende Maßnahmen und Leistungsangebote des Projekts vorgestellt werden, deren Umsetzung vor allem aus der praktischen Projektarbeit und -entwicklung resultierte.

#### 3.7.1 Die Ausarbeitung von „Landesreports“

Das anfangs nur von einigen Bundesländern formulierte Interesse an der schriftlichen Ausarbeitung von so genannten „Landesreports“ wurde schnell zu einem allgemeinen Anliegen. Zwar konnte keine der 10 Einzelstichproben nach auswahltheoretischen Maßstäben als repräsentativ für die Grundgesamtheit aller Bildungseinrichtungen dieser Bundesländer gelten; nachträgliche Vergleiche mit amtlichen Schuldaten mehrerer Bundesländer zeigten jedoch, dass die Merkmalsverteilungen der Stichproben (nach Geschlecht, Alter, Schultyp und geografischer Verteilung der Schulstandorte) weitgehend den realen Gegebenheiten dieser Länder entsprachen. Die Ergebnisse der zusammengefassten Schülerbefragungen stellten also nicht nur die Befunde an mehreren projektteilnehmenden Schulen dar, sondern gestatteten für einige Bundesländer durchaus auch gewisse Rückschlüsse auf landesweite Gegebenheiten. Die tatsächliche Nutzung der Landesreports gestaltete sich unterschiedlich: Manche

Länderministerien nutzten die Informationen eher intern bzw. als länderbezogenen Abschlussbericht; in anderen Bundesländern wurden die Ergebnisse weiter verbreitet und dienten als Arbeitsmaterialien für mehrere Landesbehörden und öffentliche Einrichtungen.

### **3.7.2 Die Durchführung von Lehrerbefragungen**

Bereits in einigen Auftaktveranstaltungen wurde von Schulen Kritik an der Projektkonzeption formuliert: Schülerbefragungen würden die soziale Schulqualität nur einseitig erfassen; für eine effektive Förderung von Schulentwicklungsprozessen fehle die Lehrerperspektive. Deshalb wurde ein Lehrerfragebogen entwickelt, dessen Inhalt sich auf die Erfassung der Arbeitsbedingungen, der Kollegenbeziehungen, der Schüler-Lehrerbeziehungen und des allgemeinen Schulklimas konzentrierte. Weiterhin hatten die Lehrkräfte die Möglichkeit, ihre Meinungen zu bestimmten Themen offen im Fragebogen zu vermerken. Die Lehrkräfte konnten den Fragebogen zu Hause ausfüllen und postalisch an das IFK zurückschicken. Die Präsentation der Ergebnisse erfolgte dann wie bei den Schülerbefragungen in zwei Schritten durch einen schriftlichen Report und eine anschließende mündliche Datenpräsentation. Insgesamt wurden auf diese Weise an 16 Schulen fast 400 Lehrkräfte befragt.

### **3.7.3 Die Durchführung eines Schülerwettbewerbs**

Im Rahmen des Projekts „Unsere Schule...“ sollten Schülerinnen und Schüler nicht nur bei den Datenerhebungen eine Rolle spielen, sondern als Projektbeteiligte auch aktiv zur Förderung der sozialen Schulqualität beitragen. Um projektbezogene Schüleraktivitäten anzuregen, führte das IFK im Jahr 2003 einen Ausschreibungswettbewerb durch. Schülerinnen und Schüler aller projektbeteiligten Schulen sollten eigene Vorstellungen und Wünsche zur Verbesserung der sozialen Schulqualität an ihrer Schule formulieren und vorstellen. Die Jugendlichen sollten hierzu eine Internetseite zum Thema „Schulqualität“ gestalten und ihre Vorstellungen von einer guten Schule in Form von Bildern, Texten, Liedern, Videoclips usw. umsetzen. Besonders kreative und wirksame Entwürfe wurden am Ende des Wettbewerbs veröffentlicht und mit Preisen in Höhe von insgesamt 1.500,- Euro prämiert.

### **3.7.4 Eine Veröffentlichung zu innovativen Schulbeispielen**

Als Ergebnis des direkten Kontakts zu über 200 Bildungseinrichtungen entstand die Idee, einige dieser Schulen im Rahmen eines Buchprojekts vorzustellen. Diese Maßnahme entwickelte sich vor dem Hintergrund einer negativ geprägten Bildungsdebatte in Deutschland: Man konnte den Eindruck haben, dass die Themen „Schule“ und „Schulqualität“ hier überwiegend von schlechten Nachrichten geprägt waren. Der Schock über das Abschneiden deutscher Schülerinnen und Schüler bei PISA saß tief und fast täglich – so schien es – hörte man Meldungen über Gewalt und Mobbing an Schulen, über Unterrichtsausfall und vermeintlich hilflose Lehrerinnen und Lehrer. Unsere Kontakte zu den Einzelschulen in 10 Bundesländern zeigten ein differenzierteres Bild von der deutschen Schullandschaft. Natürlich gibt es sie: innovative Schulen, in denen sich Lehrer, Eltern und Schüler (lange vor PISA) mit viel Engagement gemeinsam auf den Weg zu mehr sozialer Schulqualität gemacht und sichtbare Erfolge vorzuweisen haben. Einige solcher Schulen wollten wir in dem Buch „Innovative Beispiele auf dem Weg zu mehr sozialer Schulqualität“ vorstellen. Es ging uns bei der Planung dieses Buches also nicht darum, ein repräsentatives Bild von der deutschen Schullandschaft zu produzieren. Ebenfalls stand bei der konkreten Auswahl der Schulen keineswegs nur die Idee im Vordergrund, so genannte „Leuchttürme“ aus dem deutschen Bildungssystem vorzustellen und damit von vorhandenen Problemen abzulenken. Bei der Auswahl der Schulen und damit bei der Festlegung des Buchinhalts ging

es uns vor allem darum aufzuzeigen, wie kreativ und engagiert einzelne Schulen auf besondere Anforderungen reagieren oder gegen zum Teil gravierende Probleme vorgehen.

An allen Schulen gewährten die Schulleitungen, Lehrerinnen und Lehrer sowie Schülerinnen und Schüler gleichermaßen aufgeschlossen und bereitwillig Einblick in ihren Schulalltag. Ausgehend von der Situation und den Verhältnissen vor Ort entwickelten sie im Gespräch mit dem Autor und freien Journalisten Kai Nusser ihre persönlichen Vorstellungen einer guten Schule. Diese Gespräche bildeten die Grundlage der zehn Schulprofile.

### **3.7.5 Entwicklung innovativer Befragungstechniken**

Bis heute werden in der sozialwissenschaftlichen Umfrageforschung Daten vor allem unter Nutzung schriftlicher Fragebögen erhoben. Im Rahmen des Projekts „Unsere Schule...“ sollten jedoch auch die Voraussetzungen erarbeitet und erprobt werden, Schulbefragungen via Internet durchzuführen. Dies geschah unter der Vorannahme, dass die Ausstattung der Schulen mit (vernetzten) Computerterminals in einzelnen Bundesländern in naher Zukunft abgeschlossen ist. Zwar wurden die Datenerhebungen in den Projektschulen grundsätzlich noch mittels schriftlicher Fragebögen geplant und durchgeführt; für die gleichzeitige Erprobung dieser neuen Befragungsmethode sprachen jedoch mehrere Gründe:

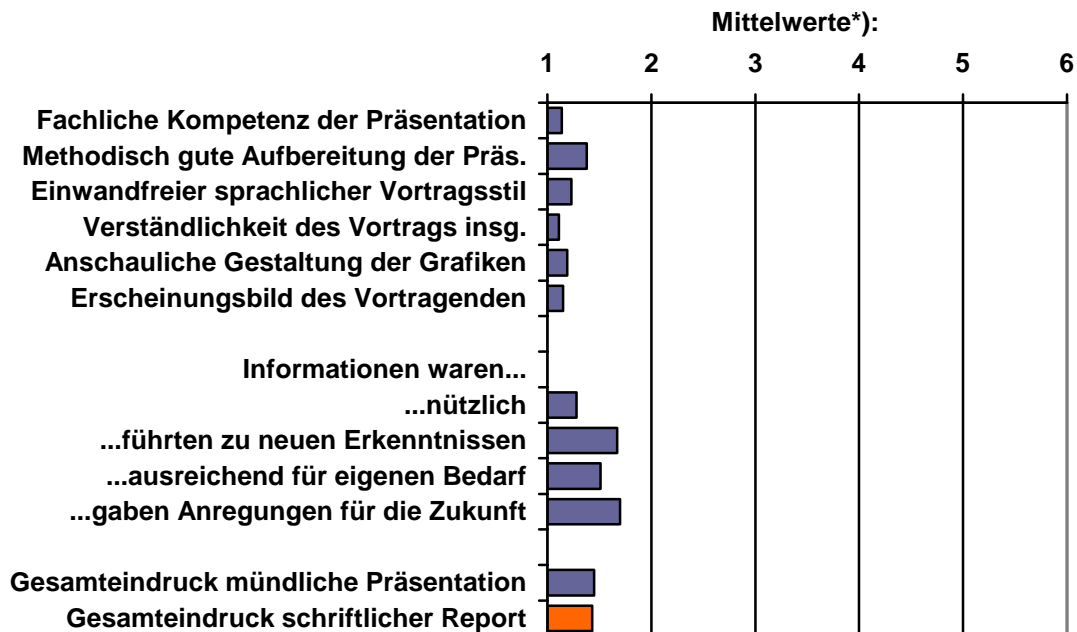
- Im Gegensatz zu Datenerhebungen mit Hilfe schriftlicher Fragebögen entfallen bei einer Internetbefragung die Druckkosten für Fragebögen und die Kosten für die Übertragung der Fragebogendaten auf EDV-lesbare Datenträger. Außerdem reduzieren sich die Fahrtkosten erheblich.
- Die Daten stehen weitaus schneller für Auswertungen zur Verfügung.
- Im Zusammenhang mit jeder Befragungsmethode ist zu sichern, dass die Ausfallrate in der Stichprobe möglichst gering gehalten wird. In diesem Zusammenhang ist u. a. für eine hohe Antwortmotivation der zu Befragenden zu sorgen. Es war zu erwarten, dass sich das Ausfüllen eines „Bildschirm-Fragebogens“ für Jugendliche spannender gestaltet und beispielsweise über softwaregesteuerte Animationssequenzen zusätzlich aufgelockert werden kann. Nähere Einzelheiten zum Verfahren und zu den Ergebnissen der Online-Befragung an einer teilnehmenden Schule des Projekts werden separat dargestellt.

### **3.7.6 Steuerung durch kontinuierliche Selbstevaluation**

Um die Qualität der Datenpräsentationsleistungen einzuschätzen und zu optimieren, wurde unmittelbar im Anschluss an die mündliche Datenpräsentation jeder Schule ein Fragebogen zugestellt, der anonymisiert an das IFK zurückgeschickt und ausgewertet wurde. Die Schulen konnten zur Bewertung der Präsentationsleistungen eine mehrstufige Ratingskala nutzen, deren Anwendung den Lehrkräften vertraut sein sollte: Der Wert „1“ stellte die beste Note und der Wert „6“ die schlechteste Note dar. Zum Projektende zeichnete sich im Rahmen dieser Maßnahme folgendes Evaluationsergebnis ab (Abbildung 1.11).

Die Evaluationsergebnisse zeigen, dass die Datenpräsentationen durch die Mitarbeiterinnen und Mitarbeiter des IFK von den projektbeteiligten Schulen alles in allem sehr positiv bewertet wurden. Hohen Anteil an diesem Ergebnis hatten zweifellos die intensive Vorbereitung und die Konzeption des schriftlichen Schulreports, der sich in seinem Aufbau und in seiner Struktur speziell an dem Informationsbedarf der Schulen orientierte. Den mündlichen Datenpräsentationen ging eine ausgiebige Schulung und Einarbeitung der beteiligten Mitarbeiterinnen und Mitarbeiter voraus. Die gesammelten Erfahrungen wurden kontinuierlich ausgetauscht und trugen so zu einer anhaltenden Verbesserung der mündlichen Datenpräsentationen bei. Begleitet wurde dieser Entwicklungsprozess durch Weiterbildungen beispielsweise im Rahmen von Rhetorik-Seminaren.

**Abbildung 11 Ergebnisse der Selbstevaluation der Datenpräsentationsleistungen im Rahmen des Projekts „Unsere Schule...“; (N = 176)**



\*) Bewertung der Präsentationsleistungen durch eine mehrstufige Ratingskala; der Wert „1“ stellt die beste Note und der Wert „6“ die schlechteste Note dar.

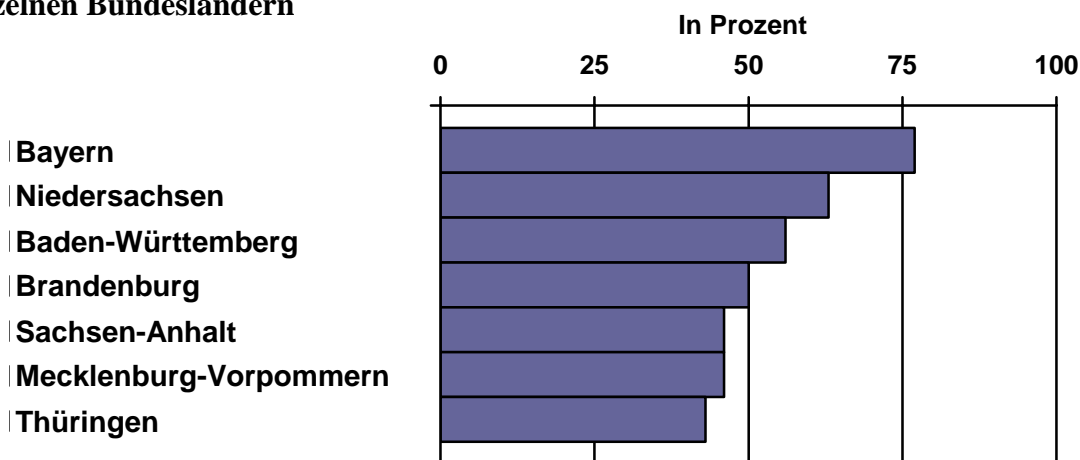
### 3.8 Zusammenfassung und Ausblick

#### 3.8.1 Zur Entwicklung der teilnehmenden Schulen in einzelnen Bundesländern

Das Projekt „Unsere Schule...“ stand in erster Linie den beteiligten Einzelschulen für eine Laufzeit von vier Jahren als Dienstleistungsanbieter zur Förderung der eigenen Schulqualität zur Seite. Ob bzw. inwieweit das Ziel, die soziale Schulqualität der teilnehmenden Schulen zu fördern erreicht werden konnte, wird – außer von den einzelnen Schulen selbst – letztlich schwer zu beantworten sein. Die Bildungslandschaft in Deutschland hat sich während der Projektlaufzeit durch eine Vielzahl unterschiedlicher Einflussfaktoren weiter entwickelt und verändert.

Durch das Fehlen einer Kontrollgruppe wird es deshalb schwer zu entscheiden sein, welche der feststellbaren Veränderungen an unseren teilnehmenden Schulen tatsächlich auf die Projektbeteiligung und welche eher auf allgemeine Entwicklungen zurückzuführen sind. Dennoch waren wir selbstverständlich sehr an der Frage interessiert, wie sich die am Projekt „Unsere Schule...“ teilnehmenden Schulen – aus Sicht der Schülerinnen und Schüler – im Untersuchungszeitraum tatsächlich entwickelt haben. Die Antwort auf diese Frage kann den Ergebnissen aus Abbildung 1.12 entnommen werden.

**Abbildung 12** Anteile von Schulen mit verbessertem „Schulqualitätsindex“(\*) in einzelnen Bundesländern



(\*) Summenindex aus aggregierten Schuldaten der Skalen „Schulattraktivität“, „Schülermitgestaltung“, „Fachliche Lehrqualität“, „Klassenzusammenhalt“, „Zufriedenheit mit der Lehrerschaft“, „Schulspaß“ und „Schulverbundenheit“; Vergleich der Indexwerte aus Welle 1 und Welle 2 (ausgewiesen sind nur Bundesländer mit 10 und mehr Schulen in beiden Erhebungen).

Der Vergleich aus den beiden Schülerbefragungen zeigt, dass in Bayern, Niedersachsen und Baden-Württemberg eine Mehrheit der teilnehmenden Schulen im Zeitraum der Projektteilnahme eine positive Entwicklung durchlaufen hat. In den östlichen Bundesländern verlief die Entwicklung der projektbeteiligten Schulen hingegen etwas schlechter. Die genauen Ursachen dieser Unterschiede bleiben unbestimmt. Es muss deshalb hervorgehoben werden, dass diese Angaben weder schlüssige Hinweise auf die Wirkung landesweiter Prozesse noch auf konkrete Einflüsse der Projektbeteiligung enthalten. Allerdings kommen aus unserer Sicht die oben vorgestellten Abweichungen zwischen den Entwicklungen in einzelnen Bundesländern nicht allzu überraschend: Vor allem in den neuen Bundesländern liefen parallel zum Projekt „Unsere Schule...“ eine Vielzahl – leider oft eher belastender – Veränderungen im Bildungsbereich ab. Zwar waren von deren Folgen (Schulschließungen, Schulzusammenlegungen, Änderungskündigungen beim Personal, Arbeitszeitkürzungen usw.) unmittelbar und überwiegend die beteiligten Lehrkräfte betroffen, jedoch ist davon auszugehen, dass eine Beeinträchtigung der Arbeitsbedingungen sowie des Arbeitsklimas für Lehrkräfte mittelbar und wirkungsvoll auch die Schulqualität der Schülerschaft negativ beeinflusst. Wir wurden im Rahmen unserer vielen persönlichen Schulkontakte „vor Ort“ immer wieder mit den Folgen dieser Veränderungen konfrontiert – leider ohne den teilnehmenden Schulen bei diesen Problemen wirkungsvoll helfen zu können.

### 3.8.2 Folgerungen aus den gemachten Erfahrungen der Projektstätigkeit

Ein Rückblick des Projektteams nach vier Jahren intensiver Arbeit im Projekt „Unsere Schule...“ verbindet sich durchaus mit einem Gefühl der Zufriedenheit: Die Mitarbeiterinnen und Mitarbeiter des Projekts haben mit großem Engagement versucht, die Entwicklung von Schulen zu fördern; die direkten Kontakte zu den Menschen, die sichtbaren Folgen der eigenen Arbeit und auch die Vielschichtigkeit der Herausforderungen an den Einzelschulen haben zu einer hohen Arbeitsmotivation beigetragen. Die teilnehmenden Schulen haben dies durchaus wahrgenommen und in positiven Projektbewertungen wiedergegeben (s. o.).

Trotz des positiven Gesamteindrucks muss eine zweite, sowohl nahe liegende als auch selbstkritische Frage lauten: Was würde man nach den gesammelten Erfahrungen heute

anders gestalten? Die Antwort auf diese Frage ergibt sich dabei keineswegs nur aus den Eindrücken des Projektteams, sondern spiegelt auch die Rückmeldungen und Verbesserungsvorschläge der Lehrkräfte aus unseren Projektschulen wider. Die nachfolgend aufgeführten Erfahrungen und Schlussfolgerungen wären aus unserer Sicht bei der Planung künftiger Projekte einzubeziehen.

#### *Weniger Schulen, mehr individuelle Betreuung.*

Der Wunsch, möglichst viele Schulen in ein Projekt einzubeziehen, liegt auf der Hand: Zum einen resultiert aus einer großen Zahl teilnehmender Schulen eine weiträumigere, direkte Wirkung des Projekts an Einzelschulen. Zum anderen verbindet sich mit einer großen Projekteinahme stets auch die Hoffnung, dass die teilnehmenden Schulen die positiven Erfahrungen aus dem Projekt an andere Bildungseinrichtungen weitergeben und durch diesen „Schneeballeffekt“ zu einer Verbreitung und Verstetigung bestimmter Projektinhalte beitragen. In der Praxis ergeben sich aus einer großen Zahl teilnehmender Schulen jedoch leider auch negative Auswirkungen auf die angestrebten Projektziele, denn je mehr Schulen man in einem Projekt betreut, desto geringer fällt zwangsläufig die Intensität der Betreuung der Einzelschulen aus. Viele Schulen waren (und sind) einfach nicht in der Lage, die Situation ihrer Bildungseinrichtung aus eigener Kraft zu verbessern. Gerade zur Förderung von Schulentwicklungsprozessen an solchen „schwierigen“ Schulen hätte die prozessbegleitende Betreuung in regelmäßigen Abständen und von den Projektträgern durch verbindliche Vorgaben stärker strukturiert erfolgen müssen. Aber auch an Schulen, die die Ergebnisse ihrer Schulevaluation interessiert aufnahmen und sich mit großem Engagement und hoher Motivation „auf den Weg“ machten, hätten die Fortschritte dieser Entwicklungsprozesse durch eine intensivere Betreuung seitens der Projektträger zusätzlich an Kraft und Dynamik gewinnen können. Eine solche, stärker an den Besonderheiten der Einzelschulen orientierte Betreuung, setzt jedoch eine längerfristige Zusammenarbeit zwischen Schule und Projektträger voraus.

#### *Aktivere Einbeziehung von Eltern und Schülerinnen und Schüler.*

Obwohl vor allem die Einbeziehung von Schulleitungen und Lehrkräften in die Umsetzung des Projekts eine unverzichtbare Voraussetzung für den Anstoß von Schulentwicklungsprozessen darstellt, sah die Projektkonzeption für den Erfolg und zur Verstetigung dieser Prozesse stets auch eine starke Einbeziehung anderer Beteiligengruppen in einzelne Projektabläufe vor. Insbesondere im Zuge der vorbereitenden Maßnahmen wurde, nicht zuletzt um den partnerschaftlichen Charakter der folgenden Zusammenarbeit zu unterstreichen, seitens der Projektträger den Einzelschulen allerdings wenig Vorschriften hinsichtlich der konkreten Projektumsetzung gemacht. Dies schloss auch die Frage ein, ob bzw. inwieweit interessierte Schülerinnen und Schüler sowie Eltern in die Leistungen und in die Ergebnisse des Projekts (Information über Ergebnisse der Schulevaluation, Beteiligung an der Fortbildung) einbezogen werden sollten. Da sich an einigen Schulen die persönlichen Kontakte zwischen Schulleitung und Lehrkräften auf der einen Seite sowie Schülerinnen und Schüler sowie Eltern auf der anderen Seite nicht immer positiv gestalteten, drangen Einzelheiten der Projektteilnahme manchmal nicht über das Lehrerkollegium hinaus. Dies resultierte jedoch keineswegs nur aus Versäumnissen oder einer Blockadehaltung seitens der Schulleitung, sondern war zuweilen auch dem fehlenden Interesse von Schüler- und Elternschaft geschuldet. Zwar hätte das in manchen Fällen demonstrative Desinteresse dieser Gruppen durch Projektmaßnahmen wohl kaum nennenswert beseitigt werden können; verbindlichere Vorgaben an die Schulleitungen zur Information oder hinsichtlich der Möglichkeiten einer Beteiligung von Schüler- und Elternschaft wären jedoch machbar und zweckdienlich gewesen.

### *Mehr Hilfen zur Selbstevaluation.*

Die Vorteile von externen Evaluationen liegen für die betreffenden Schulen geradezu auf der Hand: Erstens erfordern sie seitens der Schule kaum eigene Aufwendungen, die eigentliche Arbeit wird von Außenstehenden übernommen. Zweitens fehlt es an fast allen Schulen sowohl an der fachlichen Kompetenz als auch an den Sachmitteln (z.B. geeignete Statistiksoftware), um eine solch komplexe Maßnahme effizient durchführen zu können. Drittens erhält die Präsentation der Daten durch Außenstehende vor der – meist „befangenen“ oder „betriebsblinden“ – Schulöffentlichkeit ein höheres Gewicht, die Evaluation wirkt nicht zuletzt wegen der Einbeziehung von Vergleichswerten anderer Schulen glaubwürdiger und nachhaltiger. Viertens sind vor allem im Zusammenhang mit sensiblen Themen bestimmte Voraussetzungen für die Durchführung von Evaluationen (z.B. Datenschutzbestimmungen und damit auch die Zuverlässigkeit und Gültigkeit der Daten) meist leichter durch neutrale Außenstehende zu gewährleisten, insbesondere aus wissenschaftlichen Einrichtungen.

Die Nachteile externer Evaluationen sind jedoch ebenfalls unübersehbar. Erstens erfordert die Beauftragung von externen Institutionen bzw. Personen zur Durchführung von Evaluationen entweder (meist knappe) Schulmittel oder die externe Evaluation bleibt auf das Angebot öffentlich finanzierter Projekte beschränkt. Zweitens bleiben Schulen in der Handhabung von Evaluationen auf diese Weise weitgehend unflexibel, da sie durch die Übertragung der Aufgaben nur wenig eigene Kompetenzen erwerben, Evaluationen jederzeit und zu aktuellen Themen selbst durchzuführen. Drittens bleiben Evaluationen als Instrumente zur Förderung von Schulqualität durch ausschließlich externe Träger meist auf wenige interessierte Schulen begrenzt. Zu ihrer Verbreitung müssen solche Maßnahmen aber vielmehr in die Schulen hineingetragen werden. Dies geschieht aktuell in vielen Bundesländern über den Aufbau von „Schulevaluationsberatern“, die interessierten Schulen bei der eigenen Durchführung eines solchen Projekts als Unterstützung zur Seite stehen. Das Projekt „Unsere Schule...“ hat mit der Durchführung einer begleiteten Online-Befragung versucht, die Möglichkeiten zur Förderung der Kompetenz von Schulen bei der Durchführung von Selbstevaluationen zu erproben. Spezielle, durch das Projekt getragene Praktika für Lehrkräfte hätten diese Kompetenzen an den teilnehmenden Schulen zusätzlich fördern können.

### *Ein praktisches Projekt zur Förderung Sozialer Schulqualität.*

Eine weitere Möglichkeit der Nutzenoptimierung durch die Beteiligung an dem Projekt „Unsere Schule...“ hätte wohl darin bestanden, den Prozess der Evaluation ganzheitlich zu verstehen und – ausgehend von den Befragungsergebnissen – gemeinsam mit den Schulen zwischen den beiden Befragungen jeweils ein konkretes Schulprojekt zur Förderung von Schulqualität zu entwickeln und in die Tat umzusetzen. Viele Schulen äußerten die Überzeugung, dass sich die gezielte Umsetzung einer speziellen Maßnahme zur Verbesserung der sozialen Schulqualität in den Ergebnissen der zweiten Befragung gezeigt hätte. Ein solcher positiver Befund hätte sicher ein Gefühl von Selbstwirksamkeit und die Motivation für weitere Aktivitäten vermitteln können. Wir möchten diesen Gedanken – wenn auch nur nachträglich – voll unterstützen.

### 3.9 Literatur

- Aurin, K. (Hrsg.). (1991). Gute Schulen – Worauf beruht ihre Wirksamkeit? Bad Heilbrunn: Klinkhardt.
- Baumert, J., Lehmann, R. et al. (1997). TIMSS – Mathematisch-naturwissenschaftlicher Unterricht im internationalen Vergleich. Deskriptive Befunde. Opladen: Leske und Budrich.
- Burkhard, Ch., Eikenbusch, G. (2000). Praxishandbuch Evaluation in der Schule. Berlin.
- Eder, F. (1998). Linzer Fragebogen zum Schul- und Klassenklima, LFSK. Göttingen: Hogrefe.
- Fend, H. (1998). Qualität im Bildungswesen. Weinheim: Juventa.
- Fowler, F.J. (1984). Survey Research Methods: Beverly Hills.
- Holtappels, H.G., Heitmeyer, W., Melzer, W. & Tillmann, K.-J. (Hrsg.). (1999). Forschung über Gewalt an Schulen: Erscheinungsformen und Ursachen, Konzepte und Prävention (2. Aufl.) Weinheim, München: Juventa.
- Hurrelmann, K. (1991). Was ist eine „gute Schule?“. Erziehungskunst 4, 336-347.
- Joint Committee on Standards for educational Evaluation (1994). The program evaluation standards (2nd Edition). Thousand Oaks: Sage.
- Landua, D. (1995). Einführung in die Fragebogenkonstruktion. Berlin: Manuskript.
- Mertens, D.M. (2000). Institutionalizing evaluation in the United States of America. In R. Stockmann (Hrsg.), Evaluationsforschung (S. 41-57). Opladen: Leske und Budrich.
- OECD (1989). Schools and quality. An international report. Paris: OECD.
- Ministerium für Bildung, Jugend und Sport des Landes Brandenburg (2000). Quasum (Qualitätsuntersuchung an Schulen zum Unterricht in Mathematik) Potsdam.
- Posch, P. & Altrichter, H. (1999). Merkmale guter Schulen. <http://www.qis.at> BMUK: Wien.
- Saldern, M. v. & Littig, K.-E (1987). LASSO (Landauer Skalen zum Sozialklima für Klasse 4-13). Weinheim: Beltz.
- Stockmann, R. (2002). Qualitätsmanagement und Evaluation – Konkurrierende oder sich ergänzende Konzepte? Centrum für Evaluation. Saarbrücken, CEval-Arbeitspapiere 3.
- Sturzbecher, D. (Hrsg.). (1997). Jugend und Gewalt in Ostdeutschland - Lebenserfahrungen in Schule, Freizeit und Familie. Göttingen: Verlag für Angewandte Psychologie.
- Sturzbecher, D. (2001a). Methodische Lösungsansätze zur Befragung jüngerer Kinder. In: D. Sturzbecher (Hrsg.). Spielbasierte Befragungstechniken. Göttingen: Hogrefe, Seite 51-62.
- Sturzbecher, D. (2001b) (Hrsg.). Jugend in Ostdeutschland. Lebenssituationen und Delinquenz. Opladen: Leske und Budrich.
- Sturzbecher, D. (2002) (Hrsg.) Jugendtrends in Ostdeutschland: Bildung, Freizeit, Politik, Risiken. Opladen: Leske und Budrich.



## 4 Erhebung sozialer Schulqualität – computerbasierte Befragungstechniken

C. Gerbich<sup>6</sup>

### 4.1 Methoden zur Schulevaluation

Nicht zuletzt die Ergebnisse international vergleichender Studien wie beispielsweise TIMSS oder PISA (Baumert et al., 1997; Baumert, 2001) haben auf einen Reformbedarf hingewiesen und anhaltende Diskussionen über die Qualität und Leistungsfähigkeit des deutschen Bildungssystems angestoßen. Eine Reaktion auf diese Debatte ist das „Forum Bildung“, das von der Bund-Länder-Kommission für Bildungsplanung und Forschungsförderung ins Leben gerufen wurde. Dieses Forum beschäftigte sich auch mit der Qualitätsentwicklung und Qualitätssicherung des deutschen Bildungswesens. Zu den zwölf Empfehlungen, die das Forum 2001 veröffentlichte, gehörte auch die einer „Förderung des Lernens durch Evaluationen“:

„Zunehmende Komplexität der Anforderungen an Bildung und Qualifizierung erfordert mehr Eigenverantwortung der Bildungseinrichtungen. [...] Mehr Eigenverantwortung setzt die Bereitschaft zur Rechenschaftslegung voraus. Interne und externe Evaluationen sind zugleich wichtige Instrumente der Selbststeuerung von Bildungseinrichtungen. Das Lernen aus Evaluationen muss zu einer Selbstverständlichkeit werden.“ (Arbeitsstab Forum Bildung, November 2001, S. 25).

Hinter dieser Empfehlung steht der Gedanke, dass zur Entwicklung und Verbesserung von Bildungsqualität zunächst der Ist-Zustand von Bildungseinrichtungen systematisch erfasst werden muss, bevor die Wirkung von Projekten oder Programmen hinsichtlich ihrer Effizienz und Effektivität beurteilt werden kann. Für den Bereich der schulischen Bildung bedeutet dies konkret, dass zunächst Informationen über den Zustand von Schulqualität vorliegen müssen, bevor die Verständigung über den konkreten Entwicklungsbedarf erfolgen kann (Kohlmeyer, 2002, S.7). Dabei ist Evaluation nicht mit einer bloßen Bestandsaufnahme gleichzusetzen, sondern geht über das Identifizieren von Fehlern im Sinne der Qualitätssicherung hinaus: Erst wenn Problemursachen bekannt sind, besteht die Möglichkeit, ihrer Entstehung vorzubeugen.

Die Qualität einer Schule stellt keine statische Größe dar, sondern ist abhängig von sich verändernden Bedingungen. Aus diesem Grund setzt ein systematisches Qualitätsmanagement für Schulen die Durchführung regelmäßiger Schulevaluationen voraus. Schulentwicklungsprozesse können so gemessen und Erfolge bzw. Ursachen von Problemen identifiziert werden<sup>7</sup>.

Es stellt sich nun die Frage, welche Methode angewandt wird, um diese Evaluationen durchzuführen. Im schulischen Alltag können beispielsweise schon einfache Feedbackverfahren wie „Stimmungsbarometer“ oder „Blitzlichter“ dazu dienen, Reaktionen auf neue Unterrichtsmethoden zu bekommen. Für eine differenzierte Analyse des Ist-Zustands einer Schule sind derartige Verfahren jedoch nicht ausreichend. Qualitative Verfahren wie Beobachtungen oder Leitfadeninterviews ermöglichen eine differenziertere Sicht auf verschiedene schulische Aspekte, denn schon durch Interviews mit wenigen

---

<sup>6</sup> Dieses Kapitel basiert in Teilen auf einer Recherche zu computergestützten Datenerhebungen, die von Reinhard Schrul durchgeführt und von Karen Kammler bearbeitet wurde.

<sup>7</sup> Allerdings sollten derartige Evaluationen in angemessenen Abständen durchgeführt werden, da sonst die Gefahr besteht, dass sie aufgrund des hohen Diskussionsbedarfs eher kontraproduktiv wirken.

Schülerinnen und Schülern können Problemschwerpunkte deutlich werden. Solche Verfahren sind jedoch ergebnisoffen und erlauben keine Verallgemeinerungen.

Quantitative Verfahren, wie sie im Rahmen des Projekts „Unsere Schule...“ zum Einsatz kamen ermöglichen es hingegen, sich auf der Basis einer zufällig ausgewählten Stichprobe von Schülerinnen und Schülern einer Schule ein Gesamtbild der Schule aus Schülersicht zu machen. Im Projekt „Unsere Schule...“ wurde hierfür ein standardisierter Fragebogen eingesetzt, mit dem die Schülermeinungen zu verschiedenen, für die soziale Schulqualität relevanten, Themen gesammelt wurden. Die Befragungen erfolgten schriftlich, was mit einem nicht unbeträchtlichen zeitlichen und finanziellen Aufwand verbunden war. Aus diesem Grund kam bereits bei den Vorbereitungen zum Projekt die Idee auf, neue Verfahren der computergestützten Datenerhebung im Rahmen des Projekts zu erproben.

Im nächsten Abschnitt wird nun kurz das Verfahren von so genannten „Papier-und-Bleistift“-Erhebungen beschrieben, um im Anschluss daran auf die Vorteile computergestützter Datenerhebungen einzugehen.

#### **4.1.1 „Papier-und-Bleistift“-Erhebungen**

Das Institut für angewandte Familien-, Kindheits- und Jugendforschung (IFK) hat bereits vor dem Projekt „Unsere Schule...“ Erfahrungen mit der Evaluation sozialer Schulqualität gesammelt (Goltz & Landua, 2003). Im Rahmen des Projekts wurden 218 Schulen in zehn Bundesländern evaluiert, in dem Schülerbefragungen durchgeführt wurden. Zu den kostenintensivsten Bestandteilen des Projekts gehörten die Datenerhebung, die Erfassung der Daten sowie die Datenaufbereitung. Der Ablauf dieser drei Prozesse, die den Standards wissenschaftlicher Datenerhebungen genügen, soll im Folgenden kurz beschrieben werden. Das ausführliche Verfahren wurde in Kapitel 3 beschrieben

Vor der eigentlichen Datenerhebung sind einige organisatorischen Aufgaben zur Vorbereitung zu erfüllen. Hierzu gehört bei Papier-und-Bleistift-Erhebungen neben dem Druck der Fragebögen auch die Rekrutierung und Schulung des Erhebungspersonals. Die Datenerheber müssen über den Inhalt und die Funktionsweise des Fragebogens informiert sein, um vor Ort Fragen von Schülerseite beantworten zu können. Zudem müssen sie ein Verständnis des Untersuchungsverlaufs haben, um Fehler beim Ausfüllen des Fragebogens verhindern zu können.

Am Tag der Datenerhebung fährt das Erhebungspersonal mit den gedruckten Fragebögen an die Schulen. Die zu befragenden Klassen werden nach einem durch das IFK festgelegten, von den Schulen nicht zu beeinflussenden Schlüssel ausgewählt. Das Erhebungspersonal beaufsichtigt die Klassen, während die Schülerinnen und Schüler den Fragebogen selbständig ausfüllen. Im Anschluss an die Befragung werden die ausgefüllten Fragebögen klassenweise verpackt und von den Erheberinnen und Erhebern zum IFK zurücktransportiert. Da die manuelle Eingabe der Daten einen zu hohen Kosten- und Zeitaufwand bedeuten würde, wird zur Datenerfassung ein externer Anbieter hinzugezogen, der die Fragebögen mittels Scantechnik für die elektronische Datenverarbeitung aufbereitet.

Nach der Datenerfassung wird der Datensatz zunächst grob von Fehlern bereinigt und an das Institut zurückgeschickt, wo in einem nächsten Schritt eine genauere Prüfung und Bereinigung der Daten vorgenommen wird. Diese zweite Fehlerprüfung umfasst beispielsweise die Korrektur der Antworten so genannter Filterfragen. Diese werden eingesetzt, um das Verhalten einer bestimmten Gruppe von Schülern (z.B. Schulschwänzer) genauer zu untersuchen (Gründe für Schulschwänzen). Hier kommt es bei Papier-und-Bleistift-Befragung nicht selten zu Fehlern, weil die Filteranweisungen von den Schülerinnen und Schülern übersehen werden. Bei der Datenbereinigung werden jedoch unlogische

Angaben beseitigt. Als Beispiel kann hier die Angabe einer 18-jährigen Oberstufenschülerin dienen, die angab, eine 16-jährige Mutter zu haben.

Dieser kurze Abriss verdeutlicht bereits den hohen zeitlichen, personellen und finanziellen Aufwand beim Einsatz der Papier-und-Bleistift-Methode. Wie anfangs ausgeführt, verlangt die Etablierung eines Qualitätsmanagements jedoch regelmäßige Evaluationen. Um dies zu realisieren, ist es dringend notwendig, den Aufwand für Datenerhebungen, -erfassungen und -bereinigungen zu senken. Hier bieten computergestützte Verfahren der Datenerhebung eine Alternative.

#### **4.1.2 Computergestützte Erhebungsverfahren**

Durch die technologische Entwicklung im Hardwarebereich in den letzten Jahren haben sich die Möglichkeiten computergestützter Datenerhebungen verändert. Neben Desktop-PCs, Laptops und Notebooks sind unter anderem jetzt auch Tablet PCs und Taschencomputer, so genannte „persönliche digitale Assistenten“ (kurz: PDA), erhältlich, die den Interviewerinnen und Interviewern eine größere Mobilität einräumen (Schneid, 2001).

Für den Einsatz von Computern in Befragungssituationen sprechen viele Gründe. Zunächst sind hier die geringeren Erhebungskosten zu nennen: Sowohl die Kosten für Druck und Transport der Fragebögen können eingespart werden als auch der recht hohe finanzielle Aufwand für das Einscannen der Fragebögen. Es entsteht aber auch eine Zeitersparnis dadurch, dass der organisatorische Aufwand, der allein schon für die Aufbereitung der Fragebögen zum Einscannen anfällt, geringer ist. Die Daten sind so schneller verfügbar. Die Technisierung erlaubt zudem eine Vielzahl von Gestaltungsmöglichkeiten für das Layout des Fragebogens. Diese beziehen sich nicht nur auf ästhetische Gesichtspunkte, sondern sind auch rein praktischer Natur; beispielsweise können während der Befragung Informationen zur Verfügung gestellt werden, die das Ausfüllen des Fragebogens erleichtern (z.B. die Definition unklarer Begriffe).

Darüber hinaus entsteht ein zusätzlicher Informationsgewinn dadurch, dass Computerprogramme auch Metadaten wie die Antwortlatenz, den Zeitpunkt der Beantwortung und eventuelle Unterbrechungen der Bearbeitung erfassen können. Vor allem bedeutet eine Technisierung der Befragung jedoch auch eine deutliche Verbesserung der Gütekriterien des Erhebungsinstruments. Über die gesteigerte Durchführungsobjektivität bei der Anwendung von computerisierten Verfahren besteht allgemein Konsens (s. bspw. Finger & Ones, 1999). Der Computer gibt die Items vor, und die Instruktion, die über den PC gegeben wird, ist vollständig standardisiert. Es ist daher davon auszugehen, dass der Befragungsablauf um ein Vielfaches einheitlicher gestaltet ist, als dies ein menschlicher Versuchsleiter ermöglichen könnte. Ein deutlicher Vorteil der technisierten Befragung ist auch die Reduktion von Fehlern im Datenerhebungs- und -verarbeitungsprozess. Zum einen werden Fehler beim Ausfüllen der Fragebögen vermieden, die bei Anwendung der Papierform durch das Überspringen von Seiten, das Missachten von Filtern, unvollständige Antworten oder fälschliche Mehrfachantworten entstehen. All diese Probleme werden bei der technisierten Form durch den automatischen Wechsel zur nächsten Frage unter Beachtung der Filterführung ausgeschlossen. Fehler bei der Zuordnung der Fragebögen beispielsweise zu den jeweiligen Klassen bzw. Schulen, die durch die Eingabe einer falschen Identifikationsnummer entstehen, werden aufgrund der Vergabe eines automatischen Codes durch den Befragungsleiter minimiert. Auch Fehler bei der Datenerfassung im Sinne von Eingabefehlern bei der manuellen oder elektronischen Dateneingabe können nicht mehr auftreten. All dies führt dazu, dass der Aufwand für die Datenbereinigung verringert wird und die Daten so schneller für die Auswertung zur Verfügung stehen.

Während dieser standardisierte Ablauf also zum einen die Durchführungsobjektivität der Befragung maximiert, nimmt Jäger (1990) zum anderen den gleichen Effekt für die

Veränderung der Reliabilität an. Es verstehe sich von selbst, „daß ein Programmsystem zuverlässiger ‚funktioniert‘ und daher reliabler ist, als ein Experte“ (Jäger, 1990, S. 107). Mit einer erhöhten Zuverlässigkeit und Objektivität ist im Allgemeinen auch eine erhöhte Validität gewährleistet. Diese kann weiter gesteigert werden, da der Computer die Möglichkeit bietet, Reihungseffekte durch die Randomisierung der Skalen bzw. Items auszuschließen.

Nicht übersehen werden darf jedoch die Gefahr des Verlusts der Konstruktvalidität. Durch eine Überlagerung mit Nebeneffekten, die aufgrund der Verwendung von Computertechnologien entstehen können, treten möglicherweise Verzerrungen der Ergebnisse auf, die darauf hindeuten, dass die PC-Version ein gegenüber der zugrunde liegenden Papier-Bleistift-Version verändertes Konstrukt misst. So untersuchen bspw. verschiedene Autoren die Evozierung von Angst durch die Bearbeitung von Testverfahren am Computer (z.B. Bringsjord, 2001; Powers, 2001). Es wird angenommen, dass einerseits das Medium an sich für manche Probanden Angst auslösend wirkt (Stichwort „Technophobie“), während andererseits bestimmte Merkmale der Testsituation, wie bspw. die Unmöglichkeit, eingegebene Antworten im Nachhinein zu korrigieren, Angst hervorrufen können. Spray, Ackerman, Reckase und Carlson (1989) erlangten in ihrer Übersicht zu Äquivalenzuntersuchen die Erkenntnis, dass vergleichbare Testergebnisse bei Papier- und Bleistift- und Computerversionen von Tests dann entstehen, wenn die Software die gleiche Flexibilität erlaubt wie das Papier-Bleistift-Verfahren, wenn der Proband also die Möglichkeit hat, zu bereits beantworteten Items zurückzukehren und diese zu korrigieren. Die Ergebnisse derartiger Untersuchungen müssen bei der Erstellung der technisierten Form des Fragebogens berücksichtigt werden. So weisen die Forschungsergebnisse auch auf die Notwendigkeit hin, zu Beginn der Befragung eine Einführung in die Benutzung des Erhebungsmediums zu geben. Im Folgenden werden nun verschiedene Möglichkeiten zur computergestützten Datenerhebung aus technischer Sicht miteinander verglichen.

## 4.2 Technische Möglichkeiten computergestützter Datenerhebungen

### 4.2.1 Überblick

In diesem Teil werden zunächst einleitend zwei Möglichkeiten computergestützter Datenerhebungen vorgestellt und die Minimalanforderungen erläutert, die diese Optionen erfüllen müssen. Im Anschluss daran werden ausführlich die Möglichkeiten ihrer Umsetzung aus technischer Sicht erläutert und gegeneinander abgewogen.

Bei den zwei möglichen Zugängen handelt es sich erstens um Datenerhebungen, die mit Hilfe mobiler Geräte durchgeführt werden. Die Möglichkeiten der mobilen Befragung sind vielfältig. Aus organisatorischen und ökonomischen Gesichtspunkten erscheint für unsere Zwecke jedoch die Datenerhebung mittels PDA (Personal Digital Assistant) bzw. Tablet-PC am besten geeignet zu sein. Bei mobilen Befragungen werden Schülerinnen und Schüler mit Hilfe dieser mobilen Geräte vor Ort (in einem Klassenzimmer der Schule) befragt.

Zweitens sollen internetbasierte Befragungstechniken betrachtet werden. Hier werden die von den Schülerinnen und Schülern an schuleigenen Computern eingegebenen Daten direkt an das Evaluationsinstitut oder einen externen Server gesendet. Dabei sind wiederum zwei Vorgehensweisen möglich: Online-Befragungen können autark durchgeführt werden, d.h. der Fragebogen wird von der die Befragung durchführenden Institution selbst erstellt, mit einem speziellen Layout versehen und ist mit einem (institutseigenen) Server verbunden. Die Schülerinnen und Schüler füllen den Fragebogen online in der Schule aus und die Daten werden an den Server geschickt und dort institutsintern verwaltet. Die zweite Möglichkeit besteht darin, über einen externen Anbieter einen so genannten „Host“ zu mieten, der sowohl

das erforderliche Programm zur Generierung und Gestaltung des Fragebogens zur Verfügung stellt als auch die eingehenden Daten speichert und verwaltet

Tabelle 1: Möglichkeiten der technisierten Datenerhebung

Online-Befragungen		Mobile Befragungen	
<b>Autark mit Desktop-PC</b>	<b>Dienstleister stellt Desktop-PC</b>	<b>Einsatz von Tablet-PC</b>	<b>Einsatz von Personal Digital Assistants (PDA)</b>

Um Risiken frühzeitig zu erkennen und unangenehme Überraschungen bei der Einführung eines neuen Datenerhebungs- und -verarbeitungssystems zu vermeiden, ist es hilfreich, die Anforderungen an die Methoden klar und eindeutig zu definieren. Schwierigkeiten, die erst später im Entwicklungsprozess entdeckt werden, sind kostenintensiv und machen aufwändige Korrekturen nötig. Aus diesem Grund wird im Folgenden eine Reihe von Kriterien formuliert, die Minimalanforderungen darstellen und deren Erfüllung für eine Einbeziehung in den Vergleich unverzichtbar oder zumindest stark zu favorisieren ist.

Aus der Recherche im Bereich des internetbasierten Zugangs werden diejenigen Systeme vom Vergleich ausgeschlossen, die nicht allen Anforderungen wissenschaftlicher Datenerhebungen genügen. Eine Software für die Eigenerstellung von Befragungsinstrumenten muss beispielsweise alle datenschutzrechtlichen Ansprüche erfüllen und die Programmierung der gängigen Fragetypen (Einfach- und Mehrfachauswahl, offene und halboffene Fragen, Ranking, Matrix, Filterfragen etc.) erlauben<sup>8</sup>.

Innerhalb der Methode der mobilen Datenerhebung sollen vorrangig diejenigen Systeme Gegenstand der vorliegenden Analyse sein, die marktbeherrschend sind. Die bewertete Technik sollte demnach nicht aus Bereichen stammen, die hinsichtlich ihrer technischen Umsetzung zu individuellen Technologien mit geringer Marktakzeptanz neigen. Außerdem wird Technik, deren Anschaffungskosten deutlich über dem in dieser Gruppe bewerteten Durchschnittspreis liegen (mehr als 50 Prozent), im Rahmen dieser Analyse nicht berücksichtigt.

Darüber hinaus werden solche Systeme aus der Betrachtung ausgeschlossen, die bei der Betrachtung der im Folgenden dargestellten Abwägungskriterien auffallend schlechte Werte aufweisen. So werden bspw. PDA mit Graustufen-Display nicht berücksichtigt, da die Qualität der Darstellung ungenügend ist. In gleicher Weise werden auch solche Methoden und technische Geräte, die den zu speichernden sicherheitsrelevanten Daten keinen Schutz durch entsprechende Soft- bzw. Hardwarelösungen bieten können, hier nicht bewertet. Zudem muss zur Durchsetzung der Integrität eines Systems der Betriebssystemkern stabil und unempfindlich gegen Ausfälle sein. Ist dieser Aspekt auf der Basis von Nutzungs- und Testuntersuchungen in Hinblick auf ein Datenerhebungssystem nicht genügend berücksichtigt, wird das jeweilige System hier ebenfalls nicht vorgestellt und bewertet.

Probleme bei der technisierten Erfassung entstehen durch die zu erfassende Datenmenge. Diese ist das Produkt aus der Anzahl der befragten Schüler und der Anzahl der im Fragebogen enthaltenen Items. Die Anzahl der Befragten kann bei Datenerhebungen stark variieren. Einmal sind hierfür die Forderungen des Auftraggebers ausschlaggebend (Bund, Land,

---

<sup>8</sup> Aufgrund der in den letzten Jahren gestiegenen Popularität von Online-Befragungen hat sich eine Fülle von Softwareanbietern auf dem Markt etabliert. Eine gute Übersicht über Softwareangebote, die sich für wissenschaftliche Erhebungen eignen, findet man unter:  
[www.gesis.org/Methodenberatung/Datenerhebung/Online/Informationsquellen/Links/software.htm](http://www.gesis.org/Methodenberatung/Datenerhebung/Online/Informationsquellen/Links/software.htm)

Schulamt, etc.), und zum anderen sind wissenschaftliche Ziele zu beachten. So kann beispielsweise die Einzelschule im Fokus des Interesses stehen oder ein Ländervergleich angestrebt werden. Neben der Anzahl der Befragten variieren Untersuchungen jedoch auch hinsichtlich der Anzahl der Items, also der „Fragen“, die den Probanden zur Beantwortung vorgelegt werden. Die Menge der Items kann durch die Differenzierung der Zielstellung ebenfalls stark schwanken.

Die in diesem Gutachten vorgestellten Systeme sollen sich hinsichtlich ihrer Leistungsgrenze an der maximal zu erwartenden Datenmenge orientieren. Für die vorliegende Untersuchung wird daher von der größten zu erwartenden Itemzahl ausgegangen, die Schülern zugemutet werden kann, d. h. mit einer Itemzahl von 150 sollten die Systeme kapazitiv den Anforderungen genügen. Des Weiteren soll von einer maximalen Zahl von 5000 Befragten ausgegangen werden. Es ist anzunehmen, dass technischen Systeme, die so leistungsstark sind, dass sie 1.000 Datensätze von Befragten speichern können, auch die Erfassung von größeren Datenmengen möglich ist, ohne dass der Aufwand für ihre Anpassung deutlich gesteigert werden müsste.

Im Folgenden werden nun Kriterien aufgeführt, die die Bewertungsgrundlage für den Vergleich der verfügbaren Systeme zur technisierten Befragung von Schülerinnen und Schülern bilden. Dabei ist jedoch zu beachten, dass nicht alle Kriterien für alle Varianten von Erhebungsmethoden relevant oder zutreffend sind. Die folgende Auflistung stellt vielmehr eine systematische Darstellung aller möglichen Kriterienbereiche dar, die als Überblick zu verstehen ist.

## **4.2.2 Vergleich und Abwägung der Erhebungsmethoden**

### *4.2.2.1 Erforderliche Leistungsparameter der Hardware*

Online-Befragungen und mobile Erhebungen unterscheiden sich klar im Hinblick auf die Hardware-Ausstattung. Aus diesem Grund soll hier nur abgewogen werden, ob die beurteilte Technik über das erforderliche Leistungsspektrum verfügt, um Befragungen internetbasiert bzw. mobil durchzuführen. Hier sind die Rechenleistung, die Ausstattung der Geräte mit Speicher, die Möglichkeiten der Speichererweiterung sowie das Handling der Datenentnahme von Bedeutung.

#### (a) Rechenleistung

Deutlich unterscheiden sich die Rechenleistungen der in den verschiedenen Verfahren eingesetzten Hardware. Arbeitet der Desktop-PC, der bei Online-Befragungen zum Einsatz kommt, heute mit einer durchschnittlichen Taktfrequenz von 2,5 GHz (Gigahertz), so ist das bei den Tablet-PCs mehrheitlich 1 GHz. Im PDA-Feld liegt die Rechenleistung bei 250 bzw. 400 MHz (Megahertz). Die Leistung der Hardware ist bei diesen PDAs allerdings abgestimmt mit dem jeweiligen Betriebssystem und der eingesetzten Anwendungssoftware, sodass bei einem Vergleich von PDAs für die Anschaffung explizit der Verwendungszweck berücksichtigt werden muss<sup>9</sup>. Auch die Zielgruppe der Befragung sollte dabei berücksichtigt werden, die möglicherweise den aufwändigen Einsatz multimedialer Elemente zur Steigerung der Motivation sinnvoll erscheinen lässt.

#### (b) Speicher-Ausstattung

Desktop-PC und Tablet-PC haben heute eine durchschnittliche Hauptspeichergröße von 512 Megabyte. Es ist anzunehmen, dass diese Speicherkapazität noch zunehmen wird. Der Festplattenspeicher bewegt sich im Größenbereich von 60 bis 120 GByte und ist ebenfalls mehr als ausreichend. Beim PDA-Speicher wird konzeptgebunden bei der

---

<sup>9</sup> Unterschiede gibt es beispielsweise bei Palm OS und Pocket PC.

Ressourcenverwendung großzügig bis sparsam ausgerüstet. Auch hier ist also ein Vergleich der Geräte notwendig, wobei auch die Trendentwicklungen auf dem PDA-Markt zu berücksichtigen sind.

#### (c) Möglichkeiten der Speichererweiterung

Während bei Desktop-PCs für die Durchführung von Online-Datenerhebungen und für Tablet-PCs bei mobilen Befragungen ausreichend Leistungsreserven zur Verfügung stehen, steht auf PDAs nicht genügend freier Speicher zur Verfügung, um größere Datenmengen aufzunehmen. Aus diesem Grund müssen Steckplätze für Speicherkarten implementiert werden. Die Schnittstellen für diese Speicherkarten unterscheiden sich allerdings je nach Gerätetyp bzw. Hersteller voneinander<sup>10</sup>, sodass auch hier wieder die Relevanz einer genauen Bedarfsanalyse im Fall der Anschaffung der Taschencomputer deutlich wird. Allerdings ist zu erwarten, dass sich sowohl die Leistungsfähigkeit als auch das Preis-Leistungsverhältnis in Bezug auf PDAs in den kommenden Jahren vermutlich erheblich verbessern wird.

#### (d) Handling der Speichererweiterung (Datenentnahme)

Dieser Aspekt kann bei Online-Befragungen unberücksichtigt bleiben, da die Daten hier via Internet an den Server zurückgeschickt und dort bearbeitet werden. Bei der mobilen Methode müssen hingegen die auf Tablet-PC bzw. PDA gespeicherten Datensätze auf ein anderes Gerät mit größerem Speichervolumen (z.B. Laptop) übertragen werden. Diese Datenentnahme funktioniert über so genannte Schnittstellen am Gerät. Hierzu werden beispielsweise USB-Sticks oder Bluetooth-Module benötigt, mit deren Hilfe die Daten auf einem Laptop gesammelt werden können.

#### 4.2.2.2 Ressourcenbedarf und Wartung der Hardware

Für die Erhaltung ihrer Betriebsbereitschaft ist sowohl bei Desktop-PCs als auch bei Tablet-PCs bzw. PDAs ein gewisser Wartungsaufwand erforderlich. In der Befragungssituation ist weiterhin die Frage nach dem Strombedarf von Bedeutung.

##### (a) Wartungsaufwand

Betriebssystem, Hardware, Befragungssoftware sowie notwendige Internet- und Sicherheitssoftware müssen aufeinander abgestimmt sein und optimal miteinander funktionieren. Der Aufwand hierfür gestaltet sich für fast alle betrachteten Erhebungsmethoden ähnlich. Bei autarken Online-Befragungen wird hier kompetentes Personal benötigt. Dies gilt auch für beide Möglichkeiten der mobilen Befragung, allerdings ist hier der Abstimmungsaufwand etwas geringer. Lediglich bei gehosteten Online-Befragungen wird die Wartung vom Hostanbieter übernommen.

Der Wartungsaufwand der Hardware für das Forschungsinstitut ist bei mobilen Befragungen, bedingt durch den Transport der Geräte und die mehrfache Nutzung der Geräte durch mehrere Befragte, größer als bei Online-Erhebungen. Beispielsweise können die Geräte beim Transport oder in der Befragungssituation beschädigt werden. Die Robustheit der PDAs bzw. Tablet-PCs ist daher bei der Anschaffung zu berücksichtigen.

##### (b) Strombedarf und Stromversorgung

Im Falle einer Online-Erhebung werden sowohl die schuleigenen PCs als auch der Server durch das reguläre Stromnetz gespeist, sodass keine weiteren Vorkehrungen zur Stromversorgung getroffen werden müssen. Bei mobilen Befragungen gestaltet sich die Stromversorgung aufwändiger. Eine Verkabelung der Geräte vor Ort erscheint nicht sinnvoll,

---

<sup>10</sup> Die wichtigsten Speicherkartentypen sind CompactFlash (CF), Secure Digital (SD), MultiMediaCard (MMC), SmartMedia (SM).

u. a. weil sich die räumlichen Gegebenheiten von Schule zu Schule unterscheiden, für den Aufbau der Geräte ein zu großer Aufwand erforderlich wäre und auch eine nicht unerhebliche Unfallgefahr besteht. Aus diesem Grund muss auf Akkus zurückgegriffen werden. Dies ist jedoch nicht ganz unproblematisch. Die Laufzeit eines PDA-Akkus sollte mindestens 90 Minuten betragen, also etwa zwei Schulstunden, die für die Beantwortung von etwa 150 Items notwendig sind. Wird ein vorzeitiger Akkuwechsel notwendig, so ist dies nicht problemlos möglich, u. a. weil die Befragungsdaten dann zunächst abgespeichert werden müssen, was eine Störung im Befragungsablauf zwischen zwei Befragungen bedeuten würde. Des Weiteren ist zu beachten, dass u. U. ein zweiter Akku benötigt wird, wenn nämlich Befragungen in mehreren Klassen hintereinander stattfinden sollen. Um die Akkus neu aufzuladen, sind für die PDA-Lösung spezielle Mehrfachladegeräte anzuschaffen. PDA und Tablet-PC unterscheiden sich hier insofern voneinander, als der Stromverbrauch bei PDAs geringer ist als bei Tablet-PCs, diese dafür aber leistungsfähigere Akkus besitzen. Die Probleme beim Akkuwechsel und dem Handling der Ladeeinrichtungen unterscheiden sich hingegen kaum voneinander.

#### *4.2.2.3 Darstellungsmöglichkeiten*

Im Bereich der Visualität stellen die Größe des Displays, gemessen über die Diagonale, die Bildauflösung (angegeben in Anzahl dargestellter Pixel), die Anzahl möglicher Farben, die Helligkeit und der Kontrast wichtige Kriterien zur Einschätzung der Benutzerfreundlichkeit dar. Ziel ist es, durch die Art der Darstellung eine übersichtliche Anzeige der Fragestellung zu gewährleisten. Als Abwägungskriterien fungieren demnach bei mobilen Befragungen vor allem das Vermögen der Hardware, in Verbindung mit der verwendeten Software die Displayverkleinerung zu kompensieren, sowie ausreichende Helligkeits- und Kontrastwerte des Displays. Bei der Online-Lösung benutzen die Probanden ausschließlich Desktop-PCs, die mit herkömmlichen Monitoren ausgerüstet sind. Bei diesen sind vergleichsweise wenig Einschränkungen hinsichtlich der Darstellungsmöglichkeiten zu konstatieren, weil der Bildschirm über eine ausreichende Größe verfügt. Wir beziehen uns hier daher ausschließlich auf die Darstellungsmöglichkeiten bei mobilen Verfahren.

Sowohl bei PDAs als auch bei Tablet-PC variieren die Darstellungsmöglichkeiten je nach Anbieter enorm. Während bei Tablet-PCs eine gute Auflösung (1280x1024) bei einer Bildschirmdiagonalen von 10-12'' (25-30cm) bzw. 15-19'' (38-48 cm) möglich ist, stößt man bei der Verwendung von PDAs auf das Problem, dass es bei einer Verkleinerung der Schrift zur Verbesserung der Auflösung zu einer größeren Fehlerrate kommen kann, da die stiftorientierte Eingabe eine gewisse Praxisgewöhnung und einen höheren Grad an Konzentration erfordert. Nur PDA-Displays mit genügend Helligkeit und ausreichend starkem Kontrast eignen sich für die Darstellung eines ganzen Itemblocks, sodass auch dieser Aspekt bei der Anschaffung der Geräte mit berücksichtigt werden muss. So spiegeln transreflektive Displays einen Teil des einfallenden Lichtes wider und lassen gleichzeitig das Licht der Hintergrundbeleuchtung passieren, so dass die Anzeige bei Auflicht eine deutlich höhere Bildschärfe und Lichtstärke als herkömmliche reflektive LCD-Displays ermöglicht.

#### *4.2.2.4 Bedienbarkeit*

Die Anforderungen an die Dateneingabe und Gerätesteuerung bei der Beantwortung des Fragebogens sind im Vergleich mit anderen Standardanwendungen aus dem Office- oder Grafikbereich relativ gering, denn bei Fragebogenerhebungen wird vorrangig mit geschlossenen Antwortvorgaben gearbeitet, die nur markiert werden müssen. Der Desktop-PC für die internetbasierten Datenerhebungen wird prinzipiell mit Tastatur und Maus gesteuert bzw. bedient. Bei Geräten, die bei mobilen Erhebungsverfahren zum Einsatz kommen, ist hingegen mehrheitlich keine Tastatur vorhanden bzw. sie ist in einer deutlich kleineren Bauform ausgeführt. Um die mobilen Geräte funktionell zu steuern bzw. Daten einzugeben,



wurde eine Reihe neuer Bedienelemente entwickelt. Dazu gehören beispielsweise externe (faltbare) Tastaturen oder Stifte, die die Funktionalität von Tastatur und Maus übertreffen können.

Zur Bewertung der Benutzerfreundlichkeit der Bedienelemente muss abgewogen werden, inwiefern die Bedienungs- und Funktionselemente eines Gerätes dazu geeignet sind, alle notwendigen Eingabe- und Steuerungshandlungen einfach und eventuell auch intuitiv zu handhaben. Da die Häufigkeit der Nutzung von Desktop-PCs bei Jugendlichen immer stärker zunimmt, ist anzunehmen, dass die meisten von ihnen mit der Benutzung von Keyboard und Maus vertraut sind. Über die Benutzerfreundlichkeit anderer Bedienelemente kann an dieser Stelle nichts ausgesagt werden, da in der Bundesrepublik z.B. PDAs unter Jugendlichen nicht besonders weit verbreitet sind<sup>11</sup>.

#### 4.2.2.5 Anforderungen an die Software

Im Bereich der Software ist einerseits zwischen den Eigenschaften von Betriebssystemen zu unterscheiden, die u. a. für die Verwaltung der Speicher und der Programme notwendig sind, also die Schnittstelle zwischen Hard- und Software darstellen. Andererseits müssen die Eigenschaften der Software betrachtet werden, die für die Erstellung und das Layouten des Fragebogens, das Portieren<sup>12</sup> der Programme auf die jeweilige Hardware sowie die Datenweiterleitung bzw. die Datensicherung verwendet werden.

##### (a) Eigenschaften von Betriebssystemen

Bei Online-Befragungen erfolgt die Programmierung des Fragebogens in der Regel entweder online oder lokal auf Desktop-PCs, die in der Regel ein Standardbetriebssystem (z.B. Windows XP, Linux) nutzen.

Als Betriebssysteme für PDA sind „Windows CE/Pocket PC“ sowie „PalmOS“ am weitesten verbreitet. Beides sind „embedded Betriebssysteme“, d. h. Hard- und Softwarebausteine sind in ein gemeinsames System „eingebettet“. Das Problem dieser Betriebssysteme ist, möglichst wenige Ressourcen zu verwenden, aber gleichzeitig für verschiedene Aufgaben tauglich zu sein. Funktionen der Betriebssysteme, welche in der Lage sind, die Befragungssoftware funktionell komplett zu bedienen, müssen daher im Nachhinein nachgeladen werden können. Auch wenn sich hier in relativ kurzer Zeit enorme Veränderungen ergeben werden, dürften in der Regel immer Geräte mit relativ großen Ressourcen im Vorteil sein. Entsprechend ihrer individuellen Marktausrichtungen erweisen sich auch die derzeitig marktbeherrschenden Betriebssysteme für PDA und Tablet-PCs teilweise als völlig unterschiedlich. Natürlich ist ein Vergleich zwischen zwei Produkten möglich, aber die Spezialisierung auf bestimmte Anwendungsfunktionen erschwert diese Bewertung erheblich.

In der folgenden Tabelle sind die Bewertungskriterien für Betriebssysteme verschiedener Geräte tabellarisch dargestellt.

---

<sup>11</sup> Ganz anders sieht dies beispielsweise in Kanada aus, wo derartige Techniken bereits von Jugendlichen im schulischen Alltag genutzt werden.

<sup>12</sup> Portieren: Programme, die für ein bestimmtes Betriebssystem oder eine Plattform geschrieben wurden so abändern, dass sie auch auf einem anderen Betriebssystem laufen.

Tabelle 2: Bewertungskriterien für Betriebssysteme verschiedener Geräte

	Online-Befragungen		Mobile Befragungen	
	Autark mit Desktop-PC	Gehostet mit Desktop-PC	Einsatz von Tablet-PC	Einsatz von Personal Digital Assistants (PDA)
<b>Ausstattung der Geräte mit Basissoftware</b>	Ja	Ja	Ja	Ja
<b>Möglichkeiten der Erweiterbarkeit der Systeme mit Speicher und Schnittstellen</b>	Ja	Ja	Ja	Ja
<b>Connectivity (Möglichkeit des Anschlusses an das Internet)</b>	Ja	Ja	Ja	Ja
<b>Multimediale Funktionalität</b>	Ja	Ja	Ja	Ja
<b>Kompatibilität mit anderen Betriebssystemen / anderer Software</b>	Ja	Ja	Ja	Nein
<b>Möglicher Einsatz von Programmiersprachen</b>	Ja	Ja	Ja	Ja
<b>Auswirkung der Ressourcenverteilung auf die Funktionalität der Geräte<sup>13</sup></b>	Nein	Nein	Nein	Ja

(b) Eigenschaften der Software für Datenerhebung

Bereits vorab muss darauf hingewiesen werden, dass die Konzepte der Produkthersteller große Unterschiede aufweisen, was primär auf die unterschiedlichen Lösungsansätze und Zielstellungen hinsichtlich wichtiger Produktmerkmale zurückzuführen ist. Dadurch wird u. U. die Entscheidung für eine Datenerhebungsmethode erschwert. Ein Vergleich gestaltet sich auch dadurch schwierig, dass die technischen Parameter einer Gerätekonzeption einander bedingen und in einem Wechselverhältnis stehen. So fordern Leistungsmerkmale wie eine hohe Rechengeschwindigkeit, eine breite funktionale Anwendungsbasis sowie eine hohe Connectivity und Visualität ihren Tribut hinsichtlich der Stromversorgung. Sie verlangen auch eine bessere Ausrüstung mit Speicherkapazität und verursachen damit letztlich höhere Kosten. Es gilt demnach abzuwägen, welche technischen Parameter erforderlich sind, um Datenerhebungen durchzuführen, und welche Konzepte dazu die günstigsten Voraussetzungen bilden.

---

<sup>13</sup> Widersprüchliche Entwicklungskonzepte beeinflussen die Funktionalität: Eine bessere Videodarstellung benötigt mehr Energie und verkürzt damit die Arbeitszeit der autonomen Stromversorgung; eine höhere Funktionalität der Betriebssysteme erfordert einen größeren Speicherbedarf und fordert deshalb mehr zusätzlichen Speicher; externe Tastaturen verbessern die Möglichkeiten der Dateneingabe, erschweren aber gleichzeitig das Handling der Technik.

Dargestellt werden soll hier in erster Linie plattformunabhängige Software, also Software, die auf Programmierstandards wie JSP<sup>14</sup>, XML<sup>15</sup> oder SQL<sup>16</sup> setzt und auch in heterogenen Systemen zum Einsatz kommen kann. Wird auf unterschiedlicher Hardware gearbeitet oder werden Hardware bzw. Betriebssysteme ausgetauscht, kann mit plattformunabhängiger Software problemlos weitergearbeitet werden, so dass keine zusätzlichen finanziellen Mittel oder Bildungsressourcen notwendig werden. Tabelle 3 stellt die Abwägungskriterien für die Software zur Durchführung von Datenerhebungen für die verschiedenen Erhebungsmethoden dar.

#### (c) Systemerhaltung

Der Aufwand für die Regulierung und Abstimmung der Hard- und Softwareinstallation (s. 1.2.2) und periodisch durchzuführende Maßnahmen der Softwareaktualisierung gestaltet sich für die hier betrachteten Erhebungsmethoden ähnlich.

### 4.2.3 Datentechnische Organisation

#### 4.2.3.1 Datenvolumen

Die Erfahrungen im Rahmen des Projekts „Unsere Schule...“ zeigen, dass die Anzahl von Variablen in der für die Auswertung entwickelten Datenmaske zwei- bis dreimal so groß ist wie die Anzahl der Items im eingesetzten Fragebogen. Dies ist auf die verwendeten Fragetypen wie bspw. Fragen mit der Möglichkeit von Mehrfachantworten und deren Codierungsmöglichkeiten zurückzuführen. Das Datenvolumen eines Datensatzes wird vom Dateninhalt der Variablen bestimmt. Eine Texteingabe trägt dazu natürlich mehr als ein numerischer Wert bei. Die Auswertung von etwa 5.000 Fallbeispielen an Rohdaten mit jeweils 120 Items ergab ein durchschnittliches Datenvolumen von 335 Byte je Datensatz (alle Responder-Typen). Im Folgenden wird abgewogen, inwiefern dieses Datenvolumen durch die verschiedenen Systeme erfasst werden kann.

*4.2.3.2 Beim Einsatz von PDA kommt dem verfügbaren Speicher bzw. seinen Erweiterungsmöglichkeiten eine besondere Bedeutung zu, da hier bei einem größeren Datenvolumen eine Speichererweiterung notwendig wird, die sich momentan aufgrund des uneinheitlichen Angebots noch etwas schwierig gestaltet. Tablet-PCs bzw. Desktop-PCs sind hingegen in der Regel für die Aufnahme eines größeren Datenvolumens geeignet. Datenaustausch*

Bei computergestützten Datenerhebungen sind Datenaustauschprozesse von erheblicher Relevanz, wie im Folgenden kurz skizziert werden soll. Nach der Programmierung und dem Layouten des Fragebogens wird dieser bei der Online-Befragung auf einen Server exportiert. Der Fragebogen wird dann durch die Befragten vom Server hochgeladen und am PC von den Befragten ausgefüllt. Die Antworten werden schließlich wieder auf dem Server gespeichert,

---

<sup>14</sup> Java Server Page

<sup>15</sup> Extensible/Extended Markup Language – Metasprache zur Beschreibung von Datenstrukturen

<sup>16</sup> Structured Query Language – Abfragesprache für die Datenbankprogrammierung

Tabelle 3: Abwägungskriterien für die Software zur Durchführung von Datenerhebungen

	Online-Befragungen		Mobile Befragungen	
	Autark mit Desktop-PC	Gehostet mit Desktop-PC	Einsatz von Tablet-PC	Einsatz von PDA
<b>Befragungssoftware ist erhältlich</b>	Ja	Ja	Ja	Nein
<b>Webbasierte und plattformunabhängige Software ist verfügbar<sup>17</sup></b>	Ja	Ja	Ja	Ja
<b>Administration und Steuerung geschieht über einen Browser</b>	Ja	Ja	Ja	Ja
<b>Benötigte Fragetypen für wissenschaftliche Untersuchungen (z.B. Einfach- und Mehrfachauswahl, Ratingskalen, offene Fragen usw.) werden durch Software bereitgestellt</b>	Ja	Ja	Ja	Nein
<b>Die Layoutgestaltung des Fragebogens kann flexibel erfolgen</b>	Ja	Ja	Ja	Nein
<b>Möglichkeit der Einbindung von Multimediaelementen</b>	Ja	Ja	Ja	Nein
<b>Komplexe Filterführung möglich</b>	Ja	Ja	Ja	Nein
<b>Rücksprünge bei der Beantwortung des Fragebogens möglich</b>	Ja	Ja	Ja	Nein
<b>Software prüft Aufbau des Fragebogens bei Erstellung auf Plausibilität</b>	Ja	Ja	Ja	Nein
<b>In der Software sind Validierungs- und Plausibilitätsprüfungen bei der Auswertung integriert</b>	Ja	Ja	Ja	Nein
<b>Befragte lassen sich in Gruppen (z.B. Schüler) zusammenfügen und als Gruppen verwalten</b>	Ja	Ja	Ja	Nein

wo sich neben der Befragungssoftware die Steuerungsprogramme sowie die benötigten Datenbanken befinden. Bei mobilen Verfahren wird der Fragebogen auf einem Desktop-PC bzw. Laptop programmiert und layoutet, dann auf PDA oder Tablet-PC geladen und dort von den Probanden ausgefüllt. Anschließend werden die Datensätze wieder auf den Desktop-PC bzw. den Laptop exportiert und dort gespeichert sowie weiter verarbeitet.

Für die dargestellten Datenaustauschprozesse sind Merkmale der Datenweiterleitung, der Datensicherung und des Datenschutzes von Bedeutung, die im Folgenden für die verschiedenen hier betrachteten Methoden der Datenerhebung vorgestellt und gegeneinander abgewogen werden sollen.

Für die Datenweiterleitung stehen drei Optionen zur Verfügung: Die Daten können über das Internet an den Server übertragen, kabelgebunden<sup>18</sup> bzw. kabellos<sup>19</sup> an einen Rechner

---

17 Technologie mit Standards wie Java und der Verwendung von HTML, XML und HTTP können unabhängig vom jeweiligen Betriebssystem und Rechnertechnologie eingesetzt werden.

<sup>18</sup> Serielle Schnittstelle RS 232

<sup>19</sup> FireWire, Bluetooth

geschickt oder über ein Speicherkartenlesegerät<sup>20</sup> in ein Notebook eingelesen werden. Beim Online-Verfahren erfolgt die Übertragung über das Internet, während bei mobilen Befragungen die letzten beiden Optionen in Frage kommen.

Der technische Aufwand für die Weiterleitung der Daten ist demnach beim Online-Verfahren am geringsten, da die Übertragung am PC über das Internet erfolgt, ohne dass hier zusätzliche Hardware erforderlich wird, wie das bei Befragungen mit PDA oder Tablet-PC der Fall ist. Allerdings ist die Bedienerfreundlichkeit dieser Weiterleitungstechniken bei der mobilen Methode recht gut, sodass hier keine größeren Schwierigkeiten zu erwarten sind.

#### 4.2.3.3 Datensicherheit und Datenschutz

Datensätze einer Datenerhebung sind einmalig und in ihrer originären Form nicht reproduzierbar. Schwachpunkt des Datenhandlings ist die Zeit vom Beginn der Dateneingabe bis zur Übergabe der Daten an einen Online-Server bzw. an einen anderen PC. Aus diesem Grund kommt der Datensicherheit eine besondere Bedeutung zu, also der „Sachlage, bei der Daten unmittelbar oder mittelbar so weit wie möglich vor Beeinträchtigung oder Missbrauch bewahrt sind. Beeinträchtigung von Daten umfasst dabei u. a. den Verlust, die Zerstörung oder die Verfälschung.“<sup>21</sup>

Als Vergleichskriterium der betrachteten Erhebungsmethoden gilt hier u. a. das Vorhandensein eines Backups bzw. einer Datensicherung und deren Funktionalität bzw. Arbeitsweise. Die Sicherung der Daten sollte automatisch sowie ohne Zutun des Nutzers bzw. EDV-Personals erfolgen; die Rückführung der Daten sollte zu jedem Zeitpunkt in vollständigem Zustand möglich sein.

Tabelle 4: Übersicht über Kriterien der Datensicherung für verschiedene Verfahren

	Online-Befragungen		Mobile Befragungen	
	Autark mit Desktop-PC	Gehostet mit Desktop-PC	Tablet-PC	PDA
<b>Vorhandensein eines automatischen Backups</b>	Ja	Ja	Ja	Ja
<b>Datensicherung</b>	Ja	Ja	Ja	Ja
<b>Schutzmaßnahmen gegen Viren bei Datenübertragung über das Internet</b>	Ja	Ja	Ja	Ja

Datensicherheit ist nicht mit Datenschutz zu verwechseln. Beim Datenschutz handelt es sich um die „Sachlage, bei der die schutzwürdigen Belange Betroffener vor Beeinträchtigung, die von der Verarbeitung von Daten ausgeht, bewahrt sind.“<sup>22</sup> Kennzeichen für sichere Systeme im Datenschutz sind Integrität (die Daten werden nicht durch das Programm verfälscht, sind also konsistent und entsprechen der Realität), Vertraulichkeit (die Daten sind nur Berechtigten zugänglich), Verbindlichkeit bzw. Authentizität (die Herkunft der Daten ist nachvollziehbar) und Verfügbarkeit (Berechtigte haben immer Zugriff auf die Daten bzw. Rechnerressourcen) (Ott, 2004).

Integrität ist dann gegeben, wenn bei einem Datenbestand oder einer Menge von übertragenen Daten die Daten konsistent und korrekt sind. Konsistenz liegt dann vor, wenn die Daten sich

<sup>20</sup> Über USB-Schnittstelle mit PC verbundenes Gerät, welches die verschiedenen Speicherkartentypen, die bei PDA verwendet werden, auslesen kann.

<sup>21</sup> Quelle: DIN 44 300

<sup>22</sup> Quelle: DIN 44 300

selbst und der Realität nicht widersprechen. Die Konsistenz von Daten automatisch (d.h. durch den Computer) zu überprüfen bzw. zu gewährleisten ist relativ einfach möglich. Beispielsweise kann man durch Regeln oder Trigger bei der Datenspeicherung oder durch Prüfsummen bzw. Signaturen bei der Datenübertragung Inkonsistenzgefährdungen entdecken. Des Weiteren kann durch Normalisierung von relationalen Datenbanksystemen verhindert werden, dass bestimmte Inkonsistenzgefährdungen überhaupt eintreten können. Die Korrektheit von Daten zu prüfen bzw. zu gewährleisten, ist allerdings sehr schwer automatisierbar. Es können im Prinzip nur Plausibilitätsbedingungen angegeben werden. Die Integrität mobil erhobener Daten kann überprüft werden, wenn die Daten zur Weiterverarbeitung auf einem Desktop-PC gespeichert sind.

Vertraulichkeit ist dann gewährleistet, wenn Daten nur für Berechtigte zugänglich sind. Berechtigter ist bei Daten grundsätzlich der Eigner („Owner“) der Daten, d.h. derjenige, der sie erstellt bzw. übermittelt hat bzw. (beim Datenschutz) auf den sie sich beziehen. Der Owner kann jeweils das Zugriffsrecht weitergeben bzw. wieder entziehen; die Weitergabe kann auch die Autorisierung zur Vergabe der Rechte mit einschließen. Berechtigungen können an bestimmte Personen (identifiziert durch den User-Namen) oder an Personengruppen verteilt werden, wobei dann Personen, die einer Gruppe zugeteilt sind, die der Gruppe zugeteilten Rechte besitzen. Übliche Gruppen sind „Administratoren“ (meist alle Rechte im System) und „Public“ (keine oder nur Leserechte). Vor unberechtigtem Zugriff können Daten wirksam nur durch Verschlüsselung geschützt werden. Daraus folgt in der Praxis, dass die verschlüsselten Daten zwar prinzipiell für jeden zugänglich gemacht werden, dass aber sichergestellt wird, dass nur die Zugriffs-Berechtigten die Möglichkeit zum Entschlüsseln erhalten. Bei mobilen Verfahren reicht eine Verschlüsselung der Daten nicht aus, die Geräte müssen unter Verschluss gehalten werden.

Die Verbindlichkeit von Transaktionen, die Daten generieren, transportieren oder ändern, ist dann gegeben, wenn Daten über eine definierte Zeit erhalten bleiben (Persistenz) und der Nachweis erbracht werden kann, wer zu welchem Zeitpunkt und auf welche Weise die Daten gespeichert bzw. übertragen hat (Authentizität). Persistenz wird durch Verwendung von Datenbanksystemen gewährleistet. Authentizität wird durch Identifikation des Benutzers und Protokollierung der Vorgänge im Computer (Log-File) erreicht<sup>23</sup>.

Verfügbarkeit bedeutet, dass Berechtigte zu den im Rahmen ihrer Berechtigungen erlaubten Zeiten immer Zugriff auf Daten bzw. Rechnerressourcen haben. Für diese Verfügbarkeit kann zwischen Nutzer und Provider auch eine Wahrscheinlichkeit, ausgedrückt durch einen prozentualen Anteil, vereinbart werden. Eine 94-prozentige-Verfügbarkeit besagt dann, dass beispielsweise der Zugriff auf eine Website im vereinbarten Zeitraum bei 94 von 100 Zugriffen erfolgreich ist.

#### 4.2.3.4 Technische Vorbereitung der Datenerhebung

Neben Kriterien, die bei der Auswahl der Hard- und Software zu berücksichtigen sind, soll bei der Entscheidung für ein bestimmtes Datenerhebungsverfahren auch der Aufwand bei der technischen Vorbereitung der Datenerhebung abgewogen werden. Die Unterschiede zwischen den Verfahren sind hier beträchtlich. Der Einsatz verschiedener Hardware- bzw. Softwarelösungen führt wegen seiner unterschiedlich definierten Schnittstellen zu deutlichen

---

<sup>23</sup> Im Electronic Commerce und dabei vor allem beim Online-Shopping und Online-Banking ist dieses Kriterium auf der einen Seite sehr bedeutend ("Fake-Orders"), auf der anderen Seite jedoch derzeit noch sehr schwer zu gewährleisten. Erst digitale Signaturen bzw. darauf basierende PKI-Infrastrukturen bieten eine wirksame Möglichkeit hierfür.

Schwankungen hinsichtlich der benötigten Zeit und der sonstigen Ressourcen. Deshalb sollten diese Ressourcen bei der Bewertung abgewogen werden.

Bei autark durchgeführten Online-Erhebungen muss zunächst ein Server aufgebaut und installiert werden. Anschließend müssen verschiedene Programmierleistungen (Datenbank, Datenhandling, Fragebogen etc.) durch qualifiziertes Personal erbracht werden. Die dafür zu verwendende Zeit ist abhängig von der Komplexität des Fragebogens bzw. des gesamten Projektvorhabens. Bereits an dieser Stelle wird deutlich, dass sich dieses Verfahren nicht für kurzfristige bzw. einmalige Projekte eignet. Die komplette Vorbereitung erfolgt hier am Ort des durchführenden Instituts, es ist kein Transportaufwand erforderlich, weil die Daten via Internet weitergeleitet werden.

Bei gehosteten Verfahren erfolgt die Eingabe und das Layouten des Fragebogens sowie die Datensammlung und Datenauswertung ebenfalls vom regulären Arbeitsplatz aus. Zusätzliche technische Vorbereitungen sind hier nicht nötig, da sämtliche Funktionen für die Datenerhebung im Angebot des Hosts integriert sind. Es entsteht also lediglich ein zeitlicher Aufwand für die Einarbeitung in das jeweilige Programm.

Zur Vorbereitung mobil durchgeführter Datenerhebungen mittels PDA bzw. Tablet-PC muss zunächst die Hardware auf Funktionstüchtigkeit geprüft werden. Weiterhin müssen eventuelle Schäden behoben sowie der gerätinterne Akku und die Zusatzakkus geladen werden. Im Anschluss daran muss die Befragungssoftware auf den Geräten installiert und geprüft werden. Der Transport der Geräte erfordert ebenfalls einen gewissen organisatorischen Aufwand, denn neben geeignetem Verpackungsmaterial wird auch ein Transportfahrzeug benötigt. Vor Ort müssen die Geräte vom Erhebungspersonal aufgebaut und ggf. getestet werden.

#### 4.2.3.5 Fragebogengenerierung und Portierung

Bevor mit der Datenerhebung begonnen werden kann, muss der Fragebogen in Dateiform auf den Server bzw. Client gebracht und getestet werden. Die damit in Zusammenhang stehenden Arbeitsschritte der Entwicklung bzw. Anpassung des Fragebogens und der Gestaltung des Layouts decken sich für die verschiedenen Erhebungsmethoden weitgehend. Die Rahmenbedingungen dieser Arbeiten differieren jedoch, wodurch auch der organisatorische Aufwand dieser Arbeitsschritte unterschiedlich groß ist. Wesentlich für die Abwägung sind die Einfachheit und Verständlichkeit der Bedienung des Fragebogengenerators, die Leistungsparameter der Layoutmöglichkeiten sowie der Zeitaufwand für die Erstellung des Fragebogens einschließlich Layout. Diese Aspekte werden in Tabelle 5 für die verschiedenen Verfahren miteinander verglichen.

Tabelle 5: Aspekte der Fragebogengenerierung für verschiedene Verfahren

	Online-Befragungen		Mobile Befragungen	
	Autark mit Desktop-PC	Gehostet mit Desktop-PC	Tablet-PC	PDA
<b>Einfachheit und Verständlichkeit der Bedienung des Fragebogengenerators</b>	Gering, geschultes Personal erforderlich	Hoch	Hoch	Fragebogengenerator noch nicht vorhanden? <sup>24</sup>
<b>Leistungsparameter der Layoutmöglichkeiten</b>	Hoch	Variiert mit Anbieter	Gering	s. o.
<b>Zeitaufwand für Erstellung einschließlich Layout</b>	Für geschultes Personal gering	gering	Gering	s. o.

<sup>24</sup>Artverwandte Software existiert, konnte aber nicht getestet werden.

#### *4.2.3.6 Nicht-technische Aspekte*

Neben den in den vorangegangenen Abschnitten ausführlich beschriebenen technischen Kriterien spielen für den Vergleich zwischen Online-Befragungen und mobilen Verfahren noch weitere Aspekte eine Rolle, auf die an dieser Stelle nun kurz eingegangen wird. Zu nennen ist hier beispielsweise das für die Durchführung der Datenerhebung notwendige Know-how. Während für die Online-Befragungen, bei denen auf einen Hostanbieter zurückgegriffen wird, nur wenige technische Kenntnisse erforderlich sind bzw. diese in relativ kurzer Zeit angeeignet werden können, ist für die autarke Lösung gut geschultes Personal erforderlich, das nicht nur im Umgang mit der Hardware versiert ist, sondern auch Programmierfähigkeiten mitbringt. Weiterhin ist zu berücksichtigen, dass bei der Online-Methode nicht zwingend Personal für die Datenerhebung erforderlich ist, bei der mobilen Lösung hingegen geschultes Personal anwesend sein muss, das im Umgang und mit der Wartung der Geräte vertraut ist. Weiterhin sind die Anforderungen an die räumlichen Voraussetzungen zu beachten. Während bei Online-Befragungen die schuleigenen Computerpools genutzt werden, müssen bei mobilen Befragungen geeignete Räume durch die Schule bereitgestellt werden, in denen ausreichend Steckdosen zur möglichen Wiederaufladung der Akkus vorhanden sind und die ggf. sicher verschlossen werden können, falls beispielsweise eine Pause im Befragungsablauf auftritt oder zweitägige Befragungen durchgeführt werden sollen. Des Weiteren muss bei allen computergestützten Verfahren eine Schulung der Probanden erfolgen. Während die Nutzung von Desktop-PCs unter den Jugendlichen inzwischen recht weit verbreitet ist, könnten bei der Anwendung von PDAs bzw. Tablet-PCs gewisse Schwierigkeiten auftreten.

#### *4.2.3.7 Kostenbetrachtung*

Abschließend werden im Folgenden die Investitions- und Betriebskosten der verschiedenen Verfahren miteinander verglichen.



Tabelle 6: Investitions- und Betriebskosten der verschiedenen Verfahren im Vergleich (in Euro)

	Online-Befragungen		Mobile Befragungen	
	Autark	Gehostet	Tablet-PC	PDA
<b>Investitionskosten</b>				
Server + Peripherie <sup>25</sup> Mindestbedarf	5.000			
Durchschnittspreis Software (Betriebssystem, Standardsoftware, Programmiersoftware)	2.400			7.000 <sup>26</sup>
Durchschnittspreis Hardware PDA / Tablet-PC <sup>27</sup>			57.690	13.500
Geeignete Transportbehälter			5.000	2.500
Speichererweiterung 64 MByte				1.050
Ladeeinrichtungen (handelsüblich)			2.500	2.500
Gesamt Investitionskosten	7.400		65.190	19.550
<b>Betriebskosten</b>				
ISDN-Kanäle / Internet (6 Leitungen, zeitliche Befristung)	1.000			
Programmierung des Datenbankmodells <sup>28</sup>	7.650			
Einrichtung und Unterhaltung von Server / Datentechnik etc <sup>29</sup>	11.750			
Hosting (3000 Datensätze, 3 Monate)		500		
Hosting (bis 50.000 Datensätze / Jahr <sup>30</sup> )		6.000		
Leasingkosten Transportmittel (3 Monate)			1.500	1.500
Betriebskosten Transportmittel (3.000 km)			300	300
Transportversicherung			1.500	500
Servicekosten EDV-Technik <sup>31</sup>	Noch unbekannt	Noch unbekannt	Noch unbekannt	Noch unbekannt
Gesamt Betriebskosten	20.400		3.300	2.300
<b>Summe Betrieb- und Investitionskosten</b>	<b>27.800</b>	<b>500 - 6000</b>	<b>68.490</b>	<b>21.850</b>

#### 4.2.4 Fazit

In den vorangegangenen Abschnitten wurden verschiedene Möglichkeiten computergestützter Datenerhebungsverfahren nach verschiedenen Kriterien verglichen. Für den Vergleich wurde eine Fragebogenerhebung an Schulen mit einem Fragebogen zugrunde gelegt, der etwa 150

25 Unterbrechungsfreie Stromversorgung, Datensicherung, Störschutz, etc.

21 Eine Lizenz für Desktop-PC, 30 Lizenzen PDA (durchschnittlich 150€)

27 Preise am 09.09.2003 (Quelle: www.guenstiger.de)

28 Dauer: ca. 40 WT= 1/6 Stelle 4a (47.000€ AG-Brutto)

29 Dauer: ca. 55 WT = ¼ Stelle 4a (47.000€ AG-Brutto); Befragungs-Standzeit, Zeit DB-Einrichtung und DB-Testlauf, Datentransfer. Keine technisch bedingten Zeitrressourcen.

30 Auch befristet möglich, dann 500 € je Monat

31 Servicekosten sind bisher nicht berechnet worden. Keine Erfahrungen bisher. Hier sind zwei Varianten möglich: Entweder eigenes Personal kümmert sich um die technischen Aspekte der Datenerhebung oder der Support wird durch eine EDV-Firma vorgenommen.

Items enthält. Verglichen wurden Online-Erhebungen und mobile Befragungen, wobei bei ersteren zwischen einer autarken Variante mit eigenem Server und selbst programmierter Software einerseits und einer „gehosteten“ Variante andererseits unterschieden wurde, die die Dienstleistung eines Hostanbieters in Anspruch nimmt, der aller notwendigen Programme zur Durchführung der Datenerhebung zur Verfügung stellt. Beim mobilen Verfahren wurden zwei Optionen betrachtet: zum einen die Datenerhebung mittels Personal Digital Assistants (PDA), d.h. Taschencomputer, und zum anderen die Datenerhebung mittels Tablet-PC. Für den Vergleich wurden die Leistungsparameter der Hardware, der Ressourcenbedarf und die Wartung der Hardware, die visuellen Darstellungsmöglichkeiten, die Bedienbarkeit der Geräte, Anforderungen an die Software, die datentechnische Organisation, Aspekte der technischen Vorbereitung der Datenerhebung, Verfahren der Fragebogengenerierung und Portierung sowie eine Kostenbetrachtung berücksichtigt. Des Weiteren wurden einige nicht-technische Aspekte computergestützter Verfahren betrachtet. Der Einsatz mobiler Techniken zur Datenerhebung wurde in Erwägung gezogen, weil damit auch solche Schulen befragt werden können, die noch unzureichend mit Computern ausgestattet sind. Des Weiteren können derartige Geräte auch neue Forschungsmethoden eröffnen, indem beispielsweise bestimmte Unterrichtsmethoden sofort nach Anwendung durch die Schüler evaluiert werden.

Die Abwägung nach technischen Kriterien zeigte, dass der Markt für PDA und Tablet-PCs momentan noch stark in Bewegung ist und sich das Angebot und die Funktionsweise von Geräten je nach Anbieter enorm unterscheiden. Speziell für Datenerhebungen entwickelte Geräte sind noch nicht auf dem Markt. Aus diesem Grund müssen beim Produktvergleich sowohl die äußere Beschaffenheit der einzelnen Modelle (z.B. Robustheit), die Darstellungsmöglichkeiten als auch die Funktionsweise genauestens verglichen und gegeneinander abgewogen werden. Dies gilt auch für die benötigte Software. Bisher ist auf dem Markt kein Produkt zu finden, das sich für die Durchführung mobiler wissenschaftlicher Untersuchungen eignet. Der Einsatz derartiger Technologien würde also einen nicht unerheblichen Programmieraufwand erfordern. Des Weiteren ist bei der mobilen Methode der vergleichsweise hohe Kostenaufwand zu berücksichtigen, der sich u. a. aus Investitionskosten in die Geräte, den Transportkosten sowie den Kosten für die Pflege und Wartung der Geräte zusammensetzt. Insgesamt erscheint die Nutzung mobiler PDAs bzw. Tablet-PCs für die Durchführung von Datenerhebungen zumindest momentan noch als wenig sinnvoll. Es bleibt allerdings abzuwarten, wie sich der Markt entwickelt und ob derartige Techniken sich in Zukunft stärker verbreiten und ein größeres Spektrum an Funktionen anbieten.

Die beiden hier berücksichtigten Optionen von Online-Befragungen scheinen hinsichtlich der Verfügbarkeit von Software und Darstellungsmöglichkeiten für die Datenerhebung an Schulen aus technischer Sicht besser geeignet, da sie sich als weniger aufwändig und problemloser verfügbar erweisen. Die autarke Variante, bei der die Untersuchung über einen eigenen Server verwaltet wird, erweist sich für kurzfristige und einmalige Projekte allein aus Kostengründen jedoch wenig geeignet. Für langfristige Projekte wie beispielsweise Panel-Befragungen erscheint die Anschaffung der Hard- und Software jedoch als sinnvoll, da dadurch Kosten eingespart und spezielle Layoutvarianten oder Fragestellungen berücksichtigt werden können. Dies ist zukünftig zu berücksichtigen, denn Online-Befragungen werden in der wissenschaftlichen Forschung vermutlich immer häufiger Verwendung finden. Es ist dann möglich, dass Kosten für gute Dienstleistungsangebote, die einen Host zur Verfügung stellen, die Kosten für die einmalige Anschaffung eines Servers übersteigen.

Für kurz- und mittelfristige Projekte erscheint die Möglichkeit der gehosteten Online-Befragungen zum jetzigen Zeitpunkt jedoch am sinnvollsten. Dies rechtfertigen schon allein die geringen Investitionskosten. Hinzu kommt, dass kein zusätzlicher Qualifikationsaufwand notwendig ist, da die Datenverwaltung, die Datenaufbereitung und Datenweiterleitung durch den Dienstleistungsanbieter vorgenommen wird.

## 4.3 Methodische Aspekte computergestützter Befragungen

### 4.3.1 Formen computergestützter Befragungsmethoden

Zu Beginn dieses Kapitels wurde auf die Notwendigkeit von Evaluationen im Bildungsbereich eingegangen. Der Vergleich verschiedener Erhebungsmethoden aus technischer Sicht ergab, dass Online-Befragungen, die mit Hilfe eines gemieteten Hosts durchgeführt werden zum jetzigen Zeitpunkt eine kostengünstige Option für einmalige oder in unregelmäßigen Abständen durchgeführte Evaluationsprojekte darstellen. In diesem Abschnitt soll nun gezeigt werden, dass Online-Befragungen sich nicht zuletzt auch aus methodischer Sicht besonders gut zur Durchführungen von Schulevaluationen eignen. Dazu wird zunächst die finanzielle und technische Ausstattung von Schulen erörtert, um im Anschluss daran zu diskutieren, welche Relevanz methodische Probleme für den Kontext „Schule“ haben, die mit dem Einsatz von Online-Verfahren im Allgemeinen verbunden sind.

Vorab soll jedoch das Online-Verfahren genauer beschrieben werden, das hier von Interesse ist. Der Begriff „Online“ weist ja zunächst nur darauf hin, dass die Befragung internetbasiert erfolgt. Das konkrete Verfahren wird damit jedoch nicht bezeichnet. So unterscheidet man zwischen nicht-reaktiven Verfahren, die eine Art Beobachtung im Netz darstellen. Dazu gehören beispielsweise Logfile-Analysen, mit deren Hilfe das Verhalten im Internet nachgezeichnet werden kann (z.B. Häufigkeit des Aufrufs bestimmter Seiten) (Bandilla, 2004). An dieser Stelle interessieren hingegen lediglich reaktive-Verfahren (Bandilla et al., 1999), die sich des Internets bedienen, um Fragebogenerhebungen, Online-Interviews oder Experimente durchzuführen. Bei Fragebogenuntersuchungen stehen im Wesentlichen drei Optionen zur Verfügung (Hauptmanns, 1999): Erstens Email-Surveys, bei denen den Probanden eine Email mit Fragebogen im Anhang zugesendet wird, der dann ausgefüllt via Email zurückgeschickt werden soll; zweitens Befragungen in Newsgroups und drittens Befragungen im WWW, auf die im Folgenden ausführlicher eingegangen werden soll. Bei diesem Verfahren werden die Probanden aufgefordert, eine Seite im Internet zu öffnen, auf der sich der Fragebogen befindet, diesen auszufüllen und durch Mausclick an den Host abzusenden. Diese Form der Befragung steht im Folgenden im Fokus unseres Interesses.

### 4.3.2 Technische Ausstattung von Schulen

Eine wesentliche Voraussetzung für die Durchführung von Online-Befragungen an Schulen stellt die Ausstattung der Schulen mit Computerarbeitsplätzen und Internetanschlüssen dar. Es ist festzustellen, dass sich die IT-Ausstattung der Schulen in den letzten Jahren wesentlich verbessert hat. So hat die vom Bundesministerium für Bildung und Forschung in Auftrag gegebene Bestandsaufnahme zur „IT-Ausstattung der allgemein bildenden und berufsbildenden Schulen in Deutschland“ ergeben, dass sich die Relation „Schüler/innen pro Computer“ in den letzten Jahren deutlich verbessert hat. Sie lag 2003 bei 14 Schülerinnen und Schülern pro Computer in den Sekundarschulen I und II und bei 11 Schülerinnen und Schülern pro Computer in den berufsbildenden Schulen<sup>32</sup> (BMBF, 2003). Im Durchschnitt befinden sich in den Sekundarschulen I und II 31 Computer, in den berufsbildenden Einrichtungen sind es 94 Geräte. Insgesamt sind in den Sekundarschulen I und II 63 Prozent dieser Geräte multimediafähig, in den berufsbildenden Schulen sind es 65 Prozent. Ist ein Server vorhanden, so ist dieser am häufigsten mit Windows-Software ausgestattet.

Interessant ist in diesem Zusammenhang, wie schnell sich die IT-Ausstattung der Schulen sowie die Möglichkeiten des Zugangs zum Internet verbessert haben. Wie die Untersuchung

---

<sup>32</sup> Im Rahmen des Aktionsplans „eLearning“ der Europäischen Kommission wurde ein Verhältnis von 15 Schülerinnen und Schülern angestrebt, das damit erreicht wird.

des BMBF zeigt, waren 2001 nur 22 Prozent der Schulen mit dem Internet verbunden, im Jahr 2003 war dieser Anteil bereits auf 62 Prozent gestiegen. Im Durchschnitt sind jetzt in jeder allgemein bildenden Schule 21 Rechner mit dem Internet verbunden, an den berufsbildenden Einrichtungen sind es 67 Rechner. Es ist anzunehmen, dass sich die Ausstattung der Schulen in den nächsten Jahren noch weiter verbessern wird, da nicht nur verstärkt multimediale Lernsoftware eingesetzt wird, sondern auch der Umgang mit dem Internet einen größeren Stellenwert im Unterricht einnimmt.

Diese Ergebnisse zeigen, dass Online-Evaluationen zur Zeit zwar nicht in jedem Fall das geeignete Verfahren zur Evaluation bundesdeutscher Schulen darstellen; in den nächsten Jahren ist jedoch aufgrund verschiedenster politischer Initiativen („Schulen ans Netz“, „Neue Medien in der Bildung“) mit einer weiteren Verbesserung der IT- Ausstattung zu rechnen, die derartige Verfahren in den meisten Schulen ermöglichen wird.

### **4.3.3 Kosten**

Eine exakte Kostenberechnung für die Durchführung von Online-Befragung an Schulen kann hier nicht erfolgen. Der Preis variiert, je nachdem, ob die Befragung intern (von der Schule selbst) oder extern (durch ein Unternehmen) durchgeführt wird. Bei externen Erhebungen fallen Honorar- und Reisekosten an, während bei intern durchgeführten Untersuchungen enorme arbeitszeitliche Ressourcen benötigt werden. Des Weiteren sind hier die preislichen Unterschiede der Hostanbieter zu berücksichtigen. Während das von der Bundeszentrale für politische Bildung angebotene Programm „Grafstat“ bereits für 4 € erworben werden kann und Schulen bei Verwendung dieses Programms die Möglichkeit haben, einen Server kostenlos zu nutzen, können die Miet- und Verwaltungskosten bei anderen Dienstleistern bis zu 1000 € betragen. Allerdings ergeben sich deutliche Unterschiede zwischen den verschiedenen Softwareangeboten. Diese beziehen sich vor allem auf die Layoutmöglichkeiten des Fragebogens, die angebotenen Frageformen sowie die Möglichkeiten zur Datenanalyse<sup>33</sup>.

### **4.3.4 Bedeutung allgemeiner methodischer Probleme für schulinterne Online-Evaluationen**

Die meisten wissenschaftlichen Erhebungen, die mit quantitativen Methoden durchgeführt werden, basieren auf Zufallsstichproben, die den Vorteil haben, dass man von dieser zufällig ausgewählten Stichprobe Rückschlüsse auf die Gesamtpopulation (z.B. alle Schülerinnen und Schüler einer Schule) ziehen kann. Gleichgültig welches Erhebungsverfahren angewendet wird, sind im Wesentlichen vier mögliche Fehlertypen zu berücksichtigen, die die Qualität der Stichprobendaten beeinträchtigen können (Dillman & Bowker, 2001). Diese vier Fehlertypen werden im Folgenden im Hinblick auf ihre Relevanz für die Durchführung von Online-Befragungen an Schulen skizziert.

#### **(a) Unbekannte Grundgesamtheit (Coverage Error)**

Bei Internet-Befragungen im WWW stellt sich zunächst die Frage, wer durch die Befragung erreicht wird. Beispielsweise waren zwar 2003 nur 51 Prozent der deutschen Haushalte mit einem Internetanschluss ausgestattet (Statistisches Bundesamt 2004, S.9). Beispielsweise nutzen 27 Prozent der „Onliner“ das Internet vom Arbeitsplatz aus (Statistisches Bundesamt 2004, S. 24). Da sich hier Überschneidungen ergeben können, kann die Grundgesamtheit<sup>34</sup> der Internet-Nutzer nicht genau definiert werden. Wenn man nun beispielsweise auf der Basis

---

<sup>33</sup> So sind mit „Grafstat“ die Programmierung von Filterfragen sowie die Anwendung multivariater Verfahren nicht möglich.

<sup>34</sup> Grundgesamtheit: Menge von Personen oder Objekten, für die die Aussagen einer Untersuchung gelten sollen (Schnell et al., 1993, 279).

einer Online-Befragung den Ausgang der nächsten Bundestagswahl vorhersagen wollte, wäre dies kaum möglich, da sich der Teil der Bevölkerung, der das Internet nutzt, nicht mit dem Teil der Deutschen „deckt“ (englisch: „to cover“), die wahlberechtigt sind (vgl. dazu auch Faas, 2003). Es kommt deshalb bei Online-Befragungen häufig zum so genannten „Coverage Error“, was man frei mit „Deckungsungleichheit“ übersetzen kann. Aus diesem Grund muss vor der Durchführung einer Online-Befragung genau überprüft werden, wie stark die Vernetzung innerhalb der Zielpopulation ist. Bei Online-Befragungen, die an Schulen durchgeführt werden, stellt sich dieses Problem nicht, denn die Grundgesamtheit (alle Schülerinnen und Schüler der Schule) ist nicht nur bekannt, sondern alle Schülerinnen und Schüler haben in dem selbem Maße Zugang zu den schuleigenen Computern. Insofern spielt das Problem des Coverage Error bei der Befragung von Schülern für die schulinterne Evaluation eine untergeordnete Rolle. Eine Befragung der Eltern sollte aufgrund möglicher Ausfälle durch nicht vorhandene Technik eher über eine Papier- und Bleistift-Befragung erfolgen.

#### (b) Stichprobenfehler (Sampling Error)

Stichprobenfehler (Sampling Error) kommen dadurch zustande, dass bei einer Zufallsauswahl nicht alle Personen der Grundgesamtheit dieselbe Wahrscheinlichkeit besitzen, an der Befragung teilzunehmen. Bei „offenen“ Online-Befragungen, die im WWW durchgeführt werden, gestaltet sich die Rekrutierung von Teilnehmern, das so genannte „Sampling“<sup>35</sup>, schwierig. Zwar existieren vielfältige Verfahren, angefangen von Teilnahmeaufrufen auf Internetseiten über gekaufte Email-Adressen bis hin zu gezielten Ansprachen via Email oder Telefon, um Probanden für eine Befragung zu finden. Dennoch ist es schwierig, Stichprobenfehler zu vermeiden, weil sich bei offenen Befragungen im Internet die Teilnehmerschaft meist selbst rekrutiert und so die Teilnahme einiger Gruppen eher unwahrscheinlich ist (Dillman & Bowker, 2001, S.162). Darüber hinaus ist festzustellen, dass nicht alle Personen, die einen Internetzugang besitzen, diesen auch regelmäßig nutzen. Nur ein Viertel der Bundesdeutschen ist täglich im Internet (Statistisches Bundesamt, 2004, S. 17). Verzerrungen können auch dadurch auftreten, dass (für Bevölkerungsumfragen) bestimmte Gruppen unterrepräsentiert sind. So nutzten 2003 zwar 97 Prozent der Studierenden das Internet, aber nur 42 Prozent der Arbeitslosen und 16 Prozent der Rentner (Statistisches Bundesamt, 2004, S. 22). Stichprobenfehler können bei Schulevaluationen auf ein Minimum reduziert werden, wenn die Jugendlichen während der Schulzeit klassenweise befragt werden. Während es bei offenen Befragungen, die im Internet durchgeführt werden, zudem nur schwer möglich ist, Mehrfachteilnahmen auszuschließen, kann bei Internetbefragungen im schulischen Kontext durch die Vergabe von Zugangscodes die mehrfache Teilnahme ausgeschlossen werden.

#### (c) Ausfälle (Nonresponse Error)

Ausfälle bei Befragungen können durch zwei Arten zustande kommen: Zum einen ist es möglich, dass nicht alle Personen, die befragt werden sollen, auch tatsächlich an der Befragung teilnehmen. Man spricht dann von „Unit-Nonresponse“. Geschieht diese Nichtteilnahme zufällig, beispielsweise weil jemand am Tag der Umfrage einen Schnupfen hat, dann stellt sie kein größeres Problem dar. Nehmen Personen jedoch bewusst nicht an der Befragung teil, dann kann es zu systematischen Verzerrungen kommen, wenn nämlich die Verweigerer bestimmte Gemeinsamkeiten (z.B. sozialer Hintergrund) aufweisen (Schnell et al., 1993). Für diese Art von Ausfällen sind häufig eine schlechte Erreichbarkeit bzw. die mangelnde Motivation der Probanden verantwortlich.

---

<sup>35</sup> „Sampling“ von „sample“ = Stichprobe.

Ausfälle können weiterhin dadurch zustande kommen, dass die Probanden zwar an der Befragung teilnehmen, aber den Fragebogen nicht bis zum Ende ausfüllen. Dieser so genannte „Item-Nonresponse“ kann viele Ursachen haben. Online-Befragungen unterscheiden sich hier nicht wesentlich von anderen Erhebungsformen wie telefonischen oder schriftlichen Befragungen, bei denen die Gründe für die Nichtbeantwortung von Fragen mangelnde Motivation, Verweigerung, Interviewereffekte usw. sein können.

Sowohl für den Item-Nonresponse als auch für den Unit-Nonresponse spielen bei Online-Befragungen aber zudem auch unzureichende Fähigkeiten im Umgang mit Computer und Internet, Designeffekte (räumliche Anordnung von Fragen und Eingabeoptionen) sowie technische Probleme eine Rolle (s. dazu ausführlicher die Anmerkungen im nächsten Abschnitt). Vor allem systematischen Ausfällen sollte generell vorgebeugt werden, um die Qualität der Stichprobe zu gewährleisten. Hier können gezielte Maßnahmen eingesetzt werden. Beispielsweise ist es sinnvoll, den Geldgeber der Studie sowie den genauen Verwendungszusammenhang der Studie anzugeben („für die Verbesserung des Klimas an Eurer Schule...“); die Ansprache der Probanden zu personalisieren („Wir möchten dich bitten...“), die tatsächliche Dauer der Untersuchung anzukündigen, Hinweise zum Datenschutz zu geben und auf die Freiwilligkeit der Teilnahme hinzuweisen sowie bestimmte Anreize (Stichwort „Mitbestimmung“) bereitzustellen.

Auch bei schulinternen Evaluationen wird das Problem des Nonresponse nicht vollständig vermieden werden können. Als Vorteil ist jedoch zu sehen, dass die Zielgruppe und damit auch die wirkungsvollsten Anreize bekannt sind. Des Weiteren können Ausfälle (Krankheit, Motivationsdefizite) besser erfasst werden. Zudem hat sich die Befragung im Klassenverband im Rahmen des Projekts „Unsere Schule...“ als probates Mittel erwiesen, um zumindest den Unit-Nonresponse zu minimieren. Ein großer Vorteil von Befragungen mittels Computer oder ähnlicher Medien ist zudem in der großen Attraktivität der Medien für junge Menschen zu vermuten<sup>36</sup>. Dies lässt eine Motivationssteigerung erwarten, die sich ihrerseits vermutlich positiv auf die Teilnahmebereitschaft und eine vollständige Bearbeitung des Fragebogens auswirken wird.

#### (d) Messfehler (Measurement Error)

Messfehler (Measurement Error) tauchen beispielsweise dann auf, wenn die Antworten, die die Probanden auf eine Frage geben, von den eigentlichen „wahren“ Antworten abweichen (Schnell et al., 1993). Beispielsweise ist es möglich, dass Jugendliche bei bestimmten Fragen nicht ehrlich antworten, weil sie Konsequenzen erwarten (z.B. bei Fragen nach dem eigenen Drogenkonsum). Beim Online-Verfahren kann hingegen größere Anonymität gewährleistet werden, da (bei internen Evaluationen) die Lehrer den Fragebogen zu keinem Zeitpunkt in den Händen halten. Im Fall von Online-Befragungen, die zeit- und ortsunabhängig durchgeführt werden, sind zwar im Gegensatz zu Befragungen in der Schule Rückfragen nicht möglich. Allerdings kann man unklare Begriffe oder konkrete Hilfestellungen im Fragebogen selbst (z.B. durch Anklicken) zur Verfügung zu stellen.

Speziell bei Online-Erhebungen stellt sich die Frage, in wie weit Form, Design und Inhalt von Surveys die Antworten der Befragten beeinflussen. Es ist daher von großer Wichtigkeit, dass der Fragebogen selbst erklärend ist. Negative Effekte durch ein bestimmtes optisches Design können durch ein gutes Survey-Design vermieden werden.

---

<sup>36</sup> Über 95 Prozent der 10-24-Jährigen ist im Umgang mit dem Computer vertraut (Statistisches Bundesamt, 2004, S. 15).

## 4.4 Zum Ablauf von Schulevaluationen mit Hilfe des Online-Verfahrens

### 4.4.1 Überblick

Nachdem im letzten Abschnitt einige methodische Aspekte von Online-Befragungen behandelt wurden, soll nun der Ablauf von Schulevaluationen mit Hilfe von Online-Erhebungen grob umrissen werden. Abbildung 1.13 zeigt schematisch den Ablauf eines Evaluationsprojekts. Unterschiede zu mündlichen oder schriftlichen Befragungen ergeben sich beim Ablauf in erster Linie hinsichtlich der technischen Voraussetzungen, die an den Schulen gegeben sein müssen und bei der Vorbereitung der Fragebögen. Diese Punkte werden im Folgenden ausführlich dargestellt, während die Aspekte, die generell für die Durchführung von schulinternen Evaluationen berücksichtigt werden müssen, nur kurz umrissen werden. Sie werden an anderem Ort (vgl. Kapitel 3) ausführlicher erörtert.

### 4.4.2 Zieldefinition, Analyse der Ausgangslage und Recherche des Evaluationsgegenstands

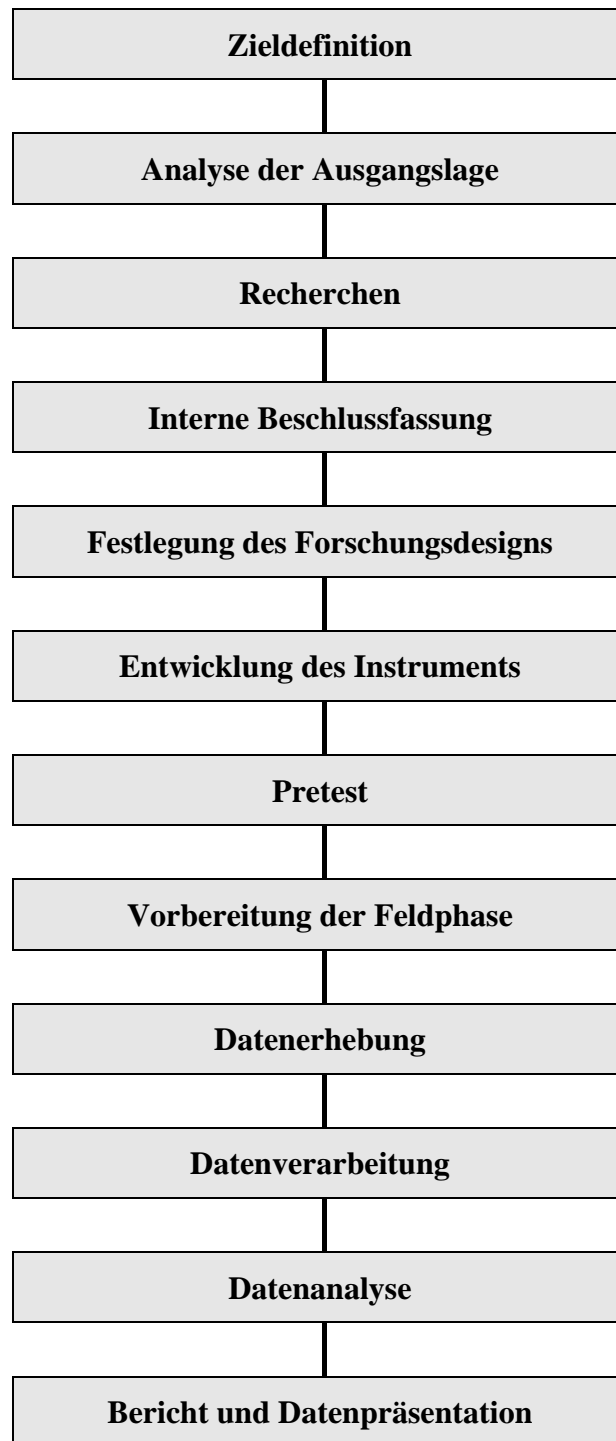
Was soll evaluiert werden? Der Formulierung der Ziele einer Evaluation kommt eine wesentliche Bedeutung für das Gelingen des Evaluationsvorhabens zu. Neben der Tatsache, dass sich durch eine exakte Zieldefinition spätere Interpretationsschwierigkeiten vermeiden lassen, erhöht sie auch die Bereitschaft der Probanden, an der Befragung teilzunehmen. Auf die Zieldefinition folgt dann die Analyse der Ausgangslage. Diese Analyse sollte bereits Kriterien zur Überprüfung der Zielstellung enthalten. Zu Beginn empfiehlt sich eine Recherche, die zum einen den thematischen Aspekt der Evaluation zum Inhalt hat und zum anderen Erfahrungen mit verschiedenen Methoden zur Erfassung des Evaluationsgegenstands beinhalten sollte. So ist es beispielsweise möglich von den Erfahrungen anderer Schulen profitieren oder sich wissenschaftliche Erkenntnisse für die eigene Befragung zunutze zu machen.

### 4.4.3 Interne Beschlussfassung

Zu Anfang dieses Kapitels wurde ausführlich auf die Notwendigkeit und den Nutzen von Evaluationen von Schulen für den Schulentwicklungsprozess hingewiesen: Evaluationen dienen dazu, mit Hilfe wissenschaftlicher Methoden systematisch Informationen über das Verhältnis zwischen den angestrebten Zielen und den tatsächlich erzielten Erfolgen eines Projekts oder Programms zu gewinnen. Selbst wenn der Sinn und Zweck schulinterner Evaluationen bekannt ist, wird eine Evaluation aber keineswegs von den Schulen in jedem Fall begrüßt. Diese Skepsis kann vielerlei Ursachen haben. Nicht zuletzt spielen hier Vorbehalte gegenüber der Qualität der Ergebnisse von Fragebogenuntersuchungen eine Rolle, aber auch Misstrauen gegenüber denjenigen, die die Daten erheben. Aus diesem Grund ist es sinnvoll, vor einer Untersuchung zunächst das Einverständnis aller Beteiligten – Schülerinnen und Schüler, Eltern, Schulleitung – im Rahmen der Schulkonferenz einzuholen.

Dieses Einverständnis vorausgesetzt, kann die Arbeit beginnen. Hierzu ist es sinnvoll, wenn sich eine Steuerungsgruppe von ca. 5-6 Personen zusammenfindet, die bereits im Vorfeld eine grobe Terminplanung vornimmt und sich in regelmäßigen Abständen trifft.

**Abbildung 13** Ablauf einer Online-Erhebung



#### **4.4.4 Festlegung des Forschungsdesigns**

Sobald Klarheit über die inhaltlichen Ziele besteht, muss geklärt werden, mit welcher Methode die formulierten Ziele erreicht werden sollen (qualitative vs. quantitative Forschungsmethoden; Form der Datenerhebung usw.). Zudem ist die Zielpopulation festzustellen. Bei quantitativen Fragebogenerhebungen in Schulen ist beispielsweise zu klären, ob sich die Befragung ausschließlich auf die Schülerinnen und Schüler einer Schule beschränkt oder ob auch das Lehrerkollegium und die Eltern in die Befragung einbezogen



werden sollen. Des Weiteren sollte diskutiert werden, ob eine Vollerhebung durchgeführt werden soll oder ob es sinnvoller ist, nur eine Stichprobe der Schülerschaft befragen.

Vor der Entscheidung für die Durchführung einer Online-Erhebung ist es zudem zwingend notwendig, die Ausstattung der Schule mit Computern zu berücksichtigen. Für die Durchführung von Online-Befragungen ist ein Klassensatz funktionstüchtiger Computer mit Internetanschluss notwendig. Die meisten Computerpools an Schulen sind zudem mit einem Beamer ausgestattet, der es den Lehrerinnen und Lehrer bzw. dem Erhebungspersonal ermöglicht, zu Beginn und während der Befragung auf Fragen seitens der Schülerinnen und Schüler einzugehen.

#### **4.4.5 Entwicklung des Fragebogens**

Mit der Entwicklung des Fragebogens beginnen die konkreten Vorbereitungen der Evaluation. Hier ist zu berücksichtigen, dass die Konstruktion eines guten Fragebogens, der sich durch kurze, eindeutige Fragen sowie eine klare und logische Struktur auszeichnet, sehr zeitaufwändig ist. Zunächst müssen die Forschungsfragen in konkret messbare Sachverhalte umgesetzt werden. Dies ist bei personenbezogenen Fragen (Alter, Geschlecht) einfach, sobald aber Einstellungen, Meinungen und dergleichen abgefragt werden sollen (z.B. Gewaltakzeptanz), erfordert die Formulierung der Fragestellungen mehrfache ausführliche Diskussionen im Rahmen der Steuerungsgruppe. Aufgrund der Komplexität von Evaluationsprojekten empfiehlt es sich daher fast immer, auf bereits erprobte Instrumente zurückzugreifen. Eine übersichtliche Einführung in die Konstruktion von Fragebögen findet sich bei Schnell et al. (1993).

#### **4.4.6 Klärung datenschutzrechtlicher Aspekte**

Wie bereits weiter oben ausgeführt wurde, müssen bei Online-Befragungen Datensicherheit und Datenschutz technisch gewährleistet werden. Zweck des Bundesdatenschutzgesetzes ist es „den einzelnen davor zu schützen, dass er durch den Umgang mit seinen personenbezogenen Daten in seinem Persönlichkeitsrecht beeinträchtigt wird“ (§1 Abs. 1 BDSG; zitiert nach Janetzko, 2000, S. 198). Demnach liegt eine datenschutzrechtlich relevante Situation dann vor, wenn Daten auf konkrete Personen bezogen oder beziehbar sind. Vor Durchführung der Befragung empfiehlt es sich, die Befragung vom Landesdatenschutzbeauftragten genehmigen zu lassen.

In der Regel wird der Datenschutz durch den Hostanbieter, beispielsweise durch die Vergabe von Passwörtern gewährleistet. Auch von Seiten der Schule muss dafür Sorge getragen werden, dass die Erhebung anonym stattfindet und personenbezogene Daten nicht an Dritte weitergegeben werden bzw. nicht für Dritte zugänglich sind. Es ist zu gewährleisten, Transaktionsnummern oder Passwörter getrennt von den Daten aufzubewahren. Bei Befragung Minderjähriger ist zudem zu beachten, ob die Einwilligung der Erziehungsberechtigten vorliegt. Es empfiehlt sich auch, die für die Datenerhebung zuständigen Personen (z.B. Lehrerinnen und Lehrer) auf Ihre Verpflichtung zum Datenschutz nochmals explizit hinzuweisen.

Die Gewährleistung von Datenschutz und Datensicherheit ist übrigens nicht nur aus rechtlicher sondern auch aus methodischer Sicht (gerade bei Schulevaluationen) von Bedeutung. Beispielsweise wird das Misstrauen potenzieller Probanden gegenüber der Befragung geringer sein, wenn glaubhaft vermittelt werden kann, dass die Prinzipien des Datenschutzes eingehalten werden. Auch die Qualität der Daten ist davon abhängig, denn sobald bei Schülerinnen und Schülern der Eindruck entsteht, dass Lehrerinnen und Lehrer oder Erziehungsberechtigte Einsicht von ihren Angaben Rückschlüsse auf die eigene Person ziehen können, kann dies zu Messfehlern führen. Beispielsweise könnten Angaben über

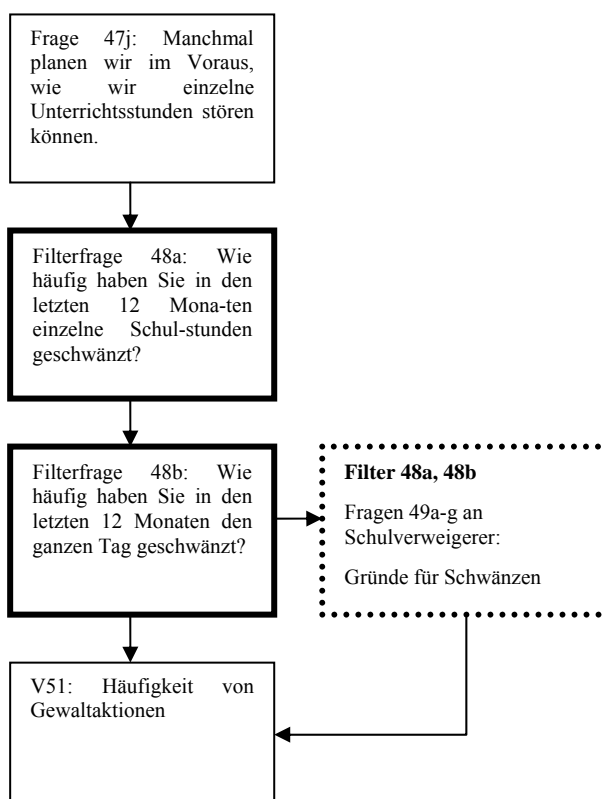
eigenes Fehlverhalten (Schwänzen, Alkohol- und Drogenkonsum) durch ein sozial erwünschtes Antwortverhalten verfälscht werden.

#### 4.4.7 Programmierung des Fragebogens

Die Programmierung des Fragebogens sollte erst dann erfolgen, wenn sich die Verantwortlichen auf eine endgültige Version geeinigt haben und die Befragung aus datenschutzrechtlicher Sicht unbedenklich ist. Für die Programmierung des Fragebogens empfiehlt es sich, zunächst auf Papier einen so genannten Codeplan zu erstellen, der jedem Item eine eigene Nummer zuweist. Um bei der Programmierung die Übersicht zu behalten, empfiehlt Kaczmirek (2004), ein so genanntes Flussdiagramm zu erstellen, in dem beispielsweise Filter vermerkt werden können. Ein Ausschnitt eines solchen Flussdiagramms ist in Abbildung 1.14 dargestellt.

Für das Layout des Fragebogens steht bei computergestützten Befragungen eine Vielzahl von Gestaltungsoptionen zur Verfügung. Hier ist darauf zu achten, dass die Gestaltung des Fragebogens „crossbrowserkompatibel“ vorgenommen wird, also von jedem Browser gelesen werden kann. Ist dies nicht der Fall, so besteht die Gefahr, dass der Fragebogen je nach Betriebssystem und Browser anders dargestellt wird, u. U. also der Fragebogen auf dem Schulcomputer vollkommen anders erscheint als auf dem Computer, der zur Programmierung des Fragebogens verwendet wurde.

**Abbildung 14** Ausschnitt aus einem Flussdiagramm



Beim Layout selbst ist zu berücksichtigen, dass ein gutes Web Design noch lange kein gutes Survey Design sein muss. So ist es bei Online-Befragungen zwar durchaus möglich, mit Elementen wie Abbildungen, bewegten Bildern, Animationen, unterschiedlichen Farben usw. zu arbeiten. Diese Elemente sollten jedoch nur dann zum Einsatz kommen, wenn sie einen

klar definierten Zweck erfüllen. Andernfalls besteht schnell die Gefahr der Ablenkung und der Unübersichtlichkeit. Durch die Wahl der Farben kann bei sparsam eingesetzten Kontrasten die Aufmerksamkeit der Probanden gesteuert werden. Vorsicht ist bei der Wahl der Farbkombinationen geboten. So können bestimmte Farbkombinationen wie Blau/Gelb, Rot/Grün, Rot/Blau oder Grün/Blau Nachbilder und Interferenzen erzeugen.

Auch bei der Wahl der Eingabeoptionen ist eine Reihe von Aspekten zu berücksichtigen. Zunächst ist die Entscheidung zwischen Multiple-Item-Screens und Single-Item-Screens zu treffen. Obwohl die zweite Variante übersichtlicher erscheint, hat sich hier gezeigt, dass die Befragungsdauer dadurch steigt und sich der Item-Nonresponse erhöht. Bei den meisten Items handelt es sich um geschlossene Fragen, d.h. die Probanden werden aufgefordert, zwischen vorgegebenen Antwortoptionen auszuwählen. Hier besteht die Möglichkeit entweder Entry Boxen zu verwenden, bei denen die Antworten mit der Tastatur eingegeben werden müssen, oder sich so genannter Radio Buttons zu bedienen, die durch einen Mausklick aktiviert werden. Hier sind eindeutig Radio Buttons zu bevorzugen, da sie für die Probanden den Antwortaufwand minimieren. Beim Einsatz so genannter Dropdown Boxes konnte eine Dominanz der sichtbaren Antwortalternativen festgestellt werden. Bei offenen Fragen, bei denen die Probanden einen freien Text eingeben können, hat sich gezeigt, dass die Größe des Testfelds als Hinweis für die erwartete Textmenge verstanden wird. Schließlich kann es sinnvoll sein, den Fortschritt des Befragungsprozesses anzuzeigen. Hier hat sich bisher gezeigt, dass diese Fortschrittsbalken zwar motivierend bei kurzen Befragungen wirken, bei längeren Untersuchungen jedoch eher das Gegenteil zur Folge haben (zum Design von Online-Fragebögen siehe Bandilla, 2004; Tipps für ein gutes Web-Design finden sich bei Pearrow, 2000 und Wirth, 2002). Der Vergleich verschiedener Anbieter ergab, dass die meisten Angebote, die sich für wissenschaftliche Untersuchungen eignen, auch geeignete Layout-Vorlagen zur Verfügung stellen.

#### **4.4.8 Vorbereitung der Feldphase**

Erfahrungsgemäß sollte für die Vorbereitung der Feldphase ausreichend Zeit kalkuliert werden, damit ein reibungsloser Ablauf der Befragung gewährleistet werden kann. Im Folgenden wird kurz auf die wesentlichsten Aspekte eingegangen, die bei der Planung zu berücksichtigen sind. Hierzu gehört eine sinnvolle Auswahl der zu befragenden Klassen, eine realistische Planung des zeitlichen Ablaufs, die auch Unvorhergesehenes berücksichtigt, die rechtszeitige Information der Erziehungsberechtigten, ein ausführlicher inhaltlicher und technischer Test des Fragebogens sowie eine gründliche Schulung des Erhebungspersonals.

##### **(a) Auswahl der zu befragenden Klassen**

Wie bereits deutlich wurde, ist mit computergestützten Verfahren der Datenerhebung eine Reihe von Vorteilen verbunden. Für Datenerhebungen, die online durchgeführt werden, gilt eine prinzipielle Unabhängigkeit von Zeitpunkt und Ort der Erhebung, d. h. theoretisch wäre es möglich, den Schülerinnen und Schülern eine Internetadresse zu nennen, die sie dann von einem beliebigen Ort zu einem beliebigen Zeitpunkt besuchen können, um den Fragebogen auszufüllen. Aus verschiedenen Gründen (s. Abschnitt 3) ist es jedoch gerade bei Schülerbefragungen sinnvoll, die Befragungen im Klassenverband durchzuführen: Nicht alle Schülerinnen und Schüler verfügen über einen Computer, Fragen können so direkt beantwortet werden, die Rücklaufquote wird höher sein, weil niemand „vergessen“ kann, den Fragebogen auszufüllen.

Die Schülerzahlen der meisten Schulen sind zu groß, um eine Vollerhebung durchzuführen. Um später dennoch auf repräsentative Ergebnisse zurückgreifen zu können, ist es notwendig, eine zufällige Auswahl aus der Gesamtschülerschaft zu treffen. Hier sind mehrere Verfahren möglich; im Rahmen des Projekts „Unsere Schule...“ hat sich hier ein Verfahren als praktikabel erwiesen, bei dem jeweils zwei Klassen eines Jahrgangs nach einem bestimmten

Schema ausgewählt werden. Es können dies jeweils die entsprechend der Klassenbezeichnung ersten und letzten Klassen sein (z.B. a und d); an manchen Schulen kann es hier zu systematischen Fehlern kommen, wenn sich beispielsweise hinter einer Klassenbezeichnung immer dasselbe Schülerklientel verbirgt (Schülerinnen und Schüler mit mangelnden Sprachkenntnissen, Pendler), sodass es hier u. U. zu Verzerrungen kommen kann. Schließlich ist bei der Auswahl der Klassen darauf zu achten, dass die Klassengröße nicht die Anzahl der verfügbaren Computer überschreitet und dass die Schülerinnen und Schüler über ausreichend Erfahrungen im Umgang mit dem Computer verfügen.

#### (b) Planung des zeitlichen Ablaufs

Die regelmäßige Durchführung schulinterner Evaluationen zur Überprüfung des Schulentwicklungsprozesses ist von großer Bedeutung. Ist eine Wiederholung der Evaluation geplant, so empfiehlt es sich, aus Gründen der Vergleichbarkeit der Ergebnisse einen „neutralen“ Befragungszeitpunkt auszuwählen. So erscheint eine Befragung direkt nach den Sommerferien oder in Prüfungszeiten nicht sinnvoll, da die Einschätzungen durch die besonderen Gegebenheiten (große An- bzw. Entspannung) verzerrt sein könnten. Wählt man jedoch einen solchen Zeitpunkt, so sollte dieser auch bei Folgeerhebungen beibehalten werden.

Der zeitliche Ablauf der aktuellen Befragung sollte cirka ein bis zwei Wochen vor dem gewählten Termin in der Schule bekannt gegeben werden, um die Unterrichtsplanung so wenig wie möglich zu beeinträchtigen.

#### (c) Schriftliches Einverständnis einholen

Für Befragungen Minderjähriger ist das schriftliche Einverständnis der Erziehungsberechtigten eine Voraussetzung. Es empfiehlt sich, dieses Einverständnis ebenfalls etwa ein bis zwei Wochen vor der Befragung einzuholen, um Erinnerungslücken zum eigentlichen Befragungszeitpunkt vorzubeugen. Das Schreiben an die Eltern sollte Informationen darüber enthalten, wer die Befragung durchführt, was das Ziel der Befragung ist und wie die Befragung abläuft (zeitlicher Rahmen). Darüber hinaus sollte auf die Anonymität der Untersuchung sowie auf das Widerrufsrecht der Jugendlichen hingewiesen werden. Das Schreiben sollte möglichst einfach und klar formuliert sein und einen Ansprechpartner für Rückfragen benennen. Um die Teilnahmebereitschaft zu steigern, empfiehlt es sich zudem, auch das schriftliche Einverständnis der Schülerinnen und Schüler einzufordern. Beispiele für Einverständniserklärungen sind im Anhang enthalten.

#### (d) Fragebogen ins Netz stellen und testen

Bevor der Fragebogen bei der eigentlichen Befragung eingesetzt wird, muss er noch einmal durch seine Autoren auf Konsistenz und Verständlichkeit überprüft werden. Im Anschluss daran wird er auf der Homepage veröffentlicht. Der Fragebogen muss dazu als HTML-Formular aufbereitet sein. Zudem muss die Befragung auf dem Datensammelpunkt des Dienstleisters angemeldet werden. Von dort sind die Daten jederzeit über das Internet abrufbar und können in die Auswertung einbezogen werden.

Um einen ungestörten Befragungsablauf zu gewährleisten und letzte Fehlerquellen auszuschließen, empfiehlt es sich zudem, den Fragebogen an den Schülerinnen und Schülern selbst zu erproben. Auch die ungefähre Dauer der Befragung kann so ermittelt werden. Bei diesem Pretest sollte möglichst detailliert das Verhalten der Schülerinnen und Schüler während des Tests, ihre Fragen sowie ihre Kritik dokumentiert und ggf. im Anschluss berücksichtigt werden. Für den Test sollten Schülergruppen ausgewählt werden, bei denen Schwierigkeiten im Befragungsverlauf antizipiert werden, um einen ungestörten Befragungsablauf in der eigentlichen Erhebung besser gewährleisten zu können.

#### (e) Schulung des Erhebungspersonals

Ein letzter wichtiger Punkt stellt die Schulung des Erhebungspersonals dar. Bei extern durchgeführten Befragungen ist dies Sache des Instituts, bei selbst initiierten Befragungen Aufgabe der Schule. Die Datenerheber (Eltern, Lehrer) müssen, um auf Rückfragen während der Befragungssituation antworten zu können, mit Inhalt und Aufbau des Fragebogens vertraut gemacht werden und über den Ablauf der Befragung genau informiert sein. Wesentlich bei Online-Befragungen ist jedoch, dass sich die Verantwortlichen auch mit den technischen Aspekten der Befragung vertraut machen. Hierzu gehören einerseits Informationen darüber, wie der Fragebogen aufgerufen und ausgefüllt wird, wie der Beamer funktioniert, was im Falle eines Ausfalls zu tun ist und mit was sich die Schülerinnen und Schüler nach der Befragung beschäftigen sollen. Diese Informationen sollten den Datenerhebern auch in schriftlicher Form vorliegen. Des Weiteren ist es sinnvoll, u. a. für die spätere Analyse des Nonresponse, ein Erhebungsprotokoll zu führen, in dem Klasse, Teilnehmer, Name des Datenerhebers, Rückfragen während der Erhebung sowie die Gründe für eine Nichtteilnahme aufgeführt sind. Diese Erhebungsprotokolle (s. Anhang) sind nicht nur nützlich, um den Verlauf der aktuellen Untersuchung im Nachhinein zu analysieren, sondern eignen sich auch dazu, Fehlerquellen zukünftiger Befragungen auszuschließen.

#### **4.4.9 Datenerhebung**

Die Datenerhebung selbst erfolgt im schuleigenen Computerpool. Um Zeit zu sparen, sollten die Computer bereits vor Beginn der Erhebung hochgefahren werden und die Internetadresse des Fragebogens unter „Favoriten“ abrufbar sein. Zunächst ist sicher zu stellen, dass jeder Schülerin und jedem Schüler ein eigenes Gerät zur Verfügung steht und von jedem Jugendlichen die eigene Einverständniserklärung und die der Eltern vorliegen. Die Jugendlichen laden dann zusammen mit dem Datenheber den Fragebogen hoch, wobei sich der Datenerheber zu Vorführungszwecken eines Beamers bedient. Sind alle Fragebögen geladen, kann die Befragung beginnen. Am Ende der Befragung werden die Daten dann durch einen Mausklick auf dem Datensammelpunkt gespeichert.

### 4.5 Praxisbericht

#### **4.5.1 Überblick**

In diesem Abschnitt werden die Erfahrungen einer verbundenen Haupt- und Realschule aus Niedersachsen beschrieben, die sich im Rahmen der Fortbildung des Projekts „Unsere Schule...“ mit den beiden Lehrbriefen „Evaluation als Instrument von Schulentwicklung“ (Kohlmeyer, 2002) sowie „Schulmeinungsforschung zur Verbesserung sozialer Schulqualität“ (Keil & Schroeter, 2002) des ibbw beschäftigte.

Aus methodischer Sicht besonders interessant war hier die Tatsache, dass die Schule eine eigene Evaluation durchführen wollte. Zwar wurde mit den Teilnehmern der Fortbildung vereinbart, für Fragen zur Verfügung zu stehen, was besonders für die Entwicklung des Fragebogens relevant war; bei der Erhebung selbst nahm das IFK jedoch lediglich eine beobachtende Rolle ein. Dieses Vorgehen wurde als sinnvoll erachtet, um Probleme bei Selbstevaluationen von Schulen zu identifizieren.

## **4.5.2 Planung und Organisation der Selbstevaluation**

### *Ziel*

Als Ziel der Befragung wurde die Evaluation sozialer Schulqualität definiert.

### *Beschlussfassung*

Mit der Durchführung der Erhebung wurde eine Steuergruppe von fünf Lehrerinnen und Lehrern betraut, die bereits die relevanten Lehrbriefe bearbeitet hatte. Das Evaluationsvorhaben wurde von Seiten der Schulleitung, den Eltern sowie den Kolleginnen und Kollegen unterstützt. Ein Mitglied der Schulleitung war in der Steuergruppe vertreten.

### *Ausstattung der Schule*

Die Schule ist gut mit Computern ausgerüstet. Es sind insgesamt zwei Computerpools mit insgesamt 32 internetfähigen Rechnern vorhanden. Die Räume liegen nebeneinander, sodass bei großen Klassen problemlos eine Teilung in zwei verschiedene Räume vorgenommen werden kann bzw. bei technischen Problemen der Wechsel zu einem anderen Arbeitsplatz möglich ist. In jedem Raum ist zudem ein Beamer vorhanden.

### *Zeitlicher Rahmen der Untersuchung*

Die Selbstevaluation erfolgte erstmalig mit beratender Unterstützung des IFK und ist als Auftakt einer kontinuierlichen Selbstevaluation gedacht, die nun etwa alle zwei Jahre an dieser Schule stattfinden und den Schulentwicklungsprozess unterstützen soll. Diese soll jeweils nach den Herbstferien stattfinden, da zu diesem Zeitpunkt in der Regel keine besonderen Termine (Prüfungen, Klassenfahrten u. ä.) anstehen. Alle Befragungen sollen im Zeitraum von einer Woche durchgeführt werden, wobei zum Ausfüllen des Fragebogens für jede Klasse eine Schulstunde zur Verfügung gestellt wird.

### *Entwicklung des Instruments*

Zunächst wurde von der Steuergruppe ein Fragebogen entworfen und dieser in einem gemeinsamen Treffen mit einer Mitarbeiterin des IFK überarbeitet. Die Grundidee bei der Konstruktion des Fragebogens war es, einen unveränderlichen Fragenkomplex zusammenzustellen, der auch bei zukünftigen Befragungen zum Einsatz kommen soll. Dieser wird durch einen modularen Fragenteil ergänzt, der sich mit aktuellen Themen befasst. Die Fragen des Basisteils orientieren sich im Wesentlichen an dem Fragenkatalog des IFK. Da bereits im Vorfeld zwei Evaluationen im Rahmen des Projekts „Unsere Schule...“ zum Thema soziale Schulqualität durchgeführt wurden, wird so ein Vergleich zwischen verschiedenen Zeitpunkten möglich. Im modularen Teil des Fragebogens wurden Fragen zur aktuellen Situation in der Schule gestellt. Hier wurden beispielsweise Fragen zur Schülerbücherei, zu Klassenfahrten und zu Aktivitäten außerhalb des Unterrichts formuliert. Insgesamt bestand der Fragebogen aus 112 geschlossenen Items und einer offenen Frage.

### *Klärung datenschutzrechtlicher Aspekte*

Zur Klärung der datenschutzrechtlichen Aspekte wurde der Fragebogen dem Landesdatenschutzbeauftragten vorgelegt. Zudem wurde er über die genaue Vorgehensweise der Datenerhebung und Datenverwaltung informiert. Vor Beginn der Befragung wurde zunächst das Einverständnis der Eltern und Schülerinnen und Schüler eingeholt. Auf der Einverständniserklärung der Jugendlichen befand sich ein Abschnitt, auf dem eine Transaktionsnummer (TAN) vermerkt war, die zu Beginn der Befragung eingegeben werden musste. Der Abschnitt verblieb nach der Befragung bei den Jugendlichen.

### *Wahl der Software, Programmierung des Fragebogens*

Die Wahl der Software orientierte sich an zwei Kriterien: Zum einen sollten die Kosten möglichst niedrig gehalten werden, zum anderen sollte sie aber auch alle Anforderungen wissenschaftlicher Erhebungen genügen. Die Steuergruppe entschied sich für das Programm GrafStat, das über die Bundeszentrale für politische Bildung für vier Euro bezogen werden kann<sup>37</sup>. Das Programm wurde explizit für Datenerhebungen im Bildungsbereich entwickelt und bedarf keiner langen Einarbeitungszeiten. Es gibt Hilfestellung bei der Entwicklung von Items, ermöglicht die selbständige Programmierung des Fragebogens, die Durchführung der Untersuchung sowie die Auswertung der Daten auf deskriptivem Niveau. Im Vorfeld der Untersuchung wurde beschlossen, dass ein Teil der Steuergruppe sich in das Programm einarbeiten sollte, um den Programmieraufwand für den Einzelnen gering zu halten und ggf. das Know-How an neue Steuergruppenmitglieder weitergeben zu können.

### *Auswahl der zu befragenden Klassen*

An der ersten Befragung sollten alle Schülerinnen und Schüler der Schule teilnehmen. Dabei handelte es sich um die Klassenstufen 6 bis 10 des Haupt- und Realschulzweigs der Schule. Für zukünftige Erhebungen ist geplant, jeweils nur eine Klasse aus jedem Schulzweig zu befragen.

### *Pretest*

Der Fragebogen sowie die Erhebungsmethode wurde eine Woche vor dem eigentlichen Erhebungstermin getestet. Für den Pretest wurde eine 6. Klasse ausgewählt, da vermutet wurde, dass diese Schüler sowohl inhaltliche als auch technische Schwierigkeiten bei der Fragenbeantwortung haben könnten. Beim Pretest ergaben sich weniger Probleme als erwartet. Bereits im Vorfeld war von den Lehrerinnen und Lehrern vermutet worden, dass die meisten Schülerinnen und Schüler dieser Klassenstufe Erfahrungen im Umgang mit Computern gesammelt haben.

### **4.5.3 Durchführung der Evaluation**

Die Untersuchung fand im Zeitraum von einer Woche statt. Insgesamt wurden 20 Klassen befragt, denen eine Schulstunde zur Verfügung stand, um den Fragebogen online zu beantworten. Es wurden nur die Jugendlichen befragt, deren Eltern ihr Einverständnis erklärt hatten. Die Befragung wurde von den Lehrerinnen und Lehrern geleitet, die in der jeweiligen Stunde Unterricht hatten. Sie bekamen vom Befragungsleiter ein Merkblatt zum Befragungsablauf, das auch technische Informationen enthielt und wurden gebeten, ein Erhebungsprotokoll auszufüllen. Auch die Gründe, warum jemand nicht an der Befragung teilnahm, sollten in Form von kurzen Notizen festgehalten werden (z. B. Krankheit, Verweigerung, Angst vor Datenunsicherheit).

Vor Beginn der Befragung wurden die Schülerinnen und Schüler nochmals kurz über das Thema der Befragung informiert. Sie wurden dann gebeten, die Einverständniserklärung durchzulesen und zu unterschreiben. Der Abschnitt mit der TAN wurde abgetrennt und verblieb bei den Schülerinnen und Schülern. Die Hinleitung zum Fragebogen erfolgte über durch die Lehrerin bzw. den Lehrer mit Hilfe des Beamers.

---

<sup>37</sup> Das Programm kann per Email unter [www.bpb.de](http://www.bpb.de) bestellt werden.

### Abbildung 15 Beispielhafter Ablauf einer Online-Befragung (Pretest)

8:00 Uhr:	<ul style="list-style-type: none"><li>• Die Schüler schalten den Computer ein.</li><li>• Die Einverständniserklärungen werden ausgeteilt und ausgefüllt.</li><li>• Der Lehrer zeigt, wie die Schüler zum Fragebogen gelangen.</li></ul>
8:10 Uhr:	<ul style="list-style-type: none"><li>• Die Befragung beginnt.</li></ul>
8:38 Uhr:	<ul style="list-style-type: none"><li>• Der erste Fragebogen wird an den Server geschickt.</li></ul>
8:46 Uhr:	<ul style="list-style-type: none"><li>• Der letzte Fragebogen wird an den Server geschickt.</li></ul>

#### 4.5.4 1.1 Probleme bei der Durchführung der Evaluation

##### *Technische Probleme*

Insgesamt traten im Befragungszeitraum nur relativ wenige technische Probleme auf. So stürzte zu Beginn der Befragung ein Computer aufgrund eines Gerätefehlers (mangelnde Energiezufuhr) ab. Bei einem weiteren Gerät konnte der Zugang zum Server nicht hergestellt werden.

##### *Probleme bei der Bedienung des Computers*

Bei der Bedienung des Computers gab es größere Probleme, die in erster Linie damit zusammen hingen, dass die zuständigen Lehrerinnen und Lehrer erst kurz vor Beginn der Befragung das Merkblatt zum Befragungsablauf durchlasen. Einigen von ihnen war auch die Bedienung des Beamers (Anschalten, Ausschalten, Standby) unklar. Beim Einloggen gab es Probleme, weil die Zugangspasswörter nicht bekannt waren oder falsch geschrieben wurden. Entsprechend den Erwartungen gab es von Seiten der Jugendlichen fast keine Probleme bei der Bedienung der Computer. Dies ist u. a. darauf zurückzuführen, dass sich alle befragten Klassen bereits vor der Befragung im Rahmen des Unterrichts mit den Computern vertraut gemacht hatten. Daher hatten nur wenige Schülerinnen und Schüler Probleme damit, den Fragebogen im Internet zu finden oder die Maus zu bedienen (Klick auf rechte statt linke Maustaste).

##### *Probleme mit Software*

Probleme mit der Software ergaben sich häufig bei der Eingabe der TAN-Nummern. Wurde die TAN-Nummer mit der Eingabetaste bestätigt, dann wurde der Fragebogen unausgefüllt an den Server geschickt und ein neues Passwort musste vergeben werden, da das Programm bei doppelt vergebenen TAN-Nummern automatisch die zweite Fragebogenversion unberücksichtigt lässt. Des Weiteren hat es sich als problematisch erwiesen, dass das Passwort nicht durch eine nochmalige Eingabe bestätigt wird, da es so zu falschen Eingaben kommen kann und der ausgefüllte Fragebogen mit der falschen TAN möglicherweise nicht gespeichert wird. Weitere Schwierigkeiten ergaben sich dadurch, dass einige Jugendliche den Fragebogen nach Eingabe der TAN zunächst durchscrollten und dann auf „Abschicken“ klickten. Dies führte dazu, dass ein leerer Fragebogen versendet wurde. Vereinzelt kam es auch vor, dass die Schülerinnen und Schüler statt auf „Abschicken“ auf „alle Angaben löschen“ klickten. Leider war dieser Befehl nicht mehr rückgängig zu machen. Da die Software die Programmierung von Filterfragen nicht erlaubt, mussten durch ein spezielles Layout eingebaute Filterfragen im Nachhinein bereinigt werden.

##### *Verständnisprobleme*

Es gab nur vereinzelte Verständnisprobleme (In meiner Klasse fühle ich mich wohl). Teilweise lasen die Schülerinnen und Schüler die einleitenden Texte nicht oder übersahen die durch das Layout hervorgehobenen Filteranweisungen.



Zeitliche Probleme ergaben sich nur, wenn dem Ausfüllen des Fragebogens ein technisches Problem vorangegangen war. Ansonsten schafften es alle Schülerinnen und Schüler, die 114 Fragen in einer Schulstunde zu beantworten.

#### *Sonstiges*

Es ist überaus sinnvoll, den Jugendlichen eine unterrichtsbezogene Aufgabe zu stellen (Recherche, Erstellung von Powerpoint-Präsentationen), die sie bearbeiten können, wenn sie mit dem Ausfüllen des Fragebogens fertig sind. Bei der Befragung an unserer Beispielschule zeigte sich, dass die Schülerinnen und Schüler mit großem Enthusiasmus an einer Präsentation zu einem Gegenstand ihres Interesses arbeiteten. In Klassen, in denen auf eine solche Aufgabe verzichtet wurde, kam es zu Störungen am Ende der Befragung.

Es hat sich bestätigt, dass die Schulung des Erhebungspersonals – und hier vor allem der Hinweis auf datenschutzrechtliche Aspekte - von größter Wichtigkeit ist. Zwar sind Lehrerinnen und Lehrer dienstrechtlich zu Vertraulichkeit verpflichtet, es ist aber nicht auszuschließen, dass sich einzelne Schülerinnen und Schüler durch die Anwesenheit des Lehrpersonals während der Befragung verunsichert fühlten und die Ergebnisse durch Interviewereffekte oder sozial erwünschtes Antwortverhalten verzerrt wurden. Durch eine ausführliche Schulung können zudem technische Unsicherheiten und Probleme im Vorhinein geklärt und behoben werden.

#### **4.5.5 Fazit aus Lehrersicht**

Nach Abschluss des Projekts wurden die Lehrerinnen und Lehrer, die in der „Projektgruppe Evaluation“ mitarbeiteten, gebeten, ein Feedback abzugeben. Insgesamt stellten die Lehrerinnen und Lehrer fest, dass die gesamte Durchführung einer Befragung (Fragebogenerstellung, Durchführung und Auswertung der Befragung) neben der Vollzeitarbeit als Lehrpersonen nur mit viel Idealismus und Einsatzfreude möglich ist. Besonders gilt dies für die Datenauswertung, die von vier Kolleginnen und Kollegen durchgeführt wurde. Dieses Team benötigte ungefähr fünf Stunden, um im Rahmen einer Gruppendiskussion die wichtigsten Ergebnisse zu diskutieren. Anschließend beschäftigte sich ein Mitglied des Teams noch etwa weitere vier Stunden mit der Datenaufbereitung. Berücksichtigt man dazu noch die Zeit, die für die Fragebogenentwicklung (drei Nachmittage), die Programmierung des Fragebogens (ein Abend) sowie die Organisation und Durchführung der Befragung (ca. eineinhalb Wochen) benötigt wurde, so wird deutlich, dass die Selbstevaluation tatsächlich mit einem großen Arbeitsaufwand verbunden war. Die Ergebnisse der Befragung wurden im Rahmen der Gesamtkonferenz durch eine Powerpointpräsentation dem Kollegium vorgestellt. Sie wurden dann durch den Beratungslehrer an die Schülerinnen und Schüler, die dem Schülerrat angehören vermittelt. Es wurde zudem diskutiert, die Befragungsergebnisse im Intranet bzw. auf den Internetseiten der Schule zu präsentieren. Die hauptsächliche organisatorische Schwierigkeit bei der Planung und Durchführung der Befragung lag für die Lehrerinnen und Lehrer darin, dass es innerhalb der Gruppe zu viele Kollegen im Team gab, die nicht hinreichend in die Bedienung des Programms Grafstat eingewiesen waren. Daraus ergibt sich die Forderung, dass zu Beginn der Planung einer Evaluation möglichst viele Personen über den konkreten Ablauf der Evaluation informiert sowie in die technischen Aspekte der Datenerhebung eingewiesen werden sollten. So lässt sich eine Konzentration der Arbeitsaufgaben auf eine einzige Person verhindern.

## **4.6 Zusammenfassung und Ausblick**

Von bildungspolitischer Seite werden Evaluationen als ein unverzichtbares Instrument zur Förderung von Qualität im Bildungsbereich anerkannt. In einigen deutschen Bundesländern

wurden bereits Verfahren etabliert, bei denen die schulinterne Qualitätssicherung und Qualitätsentwicklung durch externe Beobachterinnen und Beobachter eingeschätzt wird. So wird beispielsweise in Niedersachsen eine regelmäßige systematische Überprüfung und Bewertung der Unterrichts- und Erziehungsarbeit an Schulen vorgenommen; dabei werden die Schulentwicklungsprozesse beurteilt. Gleichzeitig ist ein steigendes Interesse einzelner Schulen am Thema „Selbstevaluation“ festzustellen.

Es herrscht weitgehend Einigkeit darüber, dass Schulevaluationen mit einer gewissen Regelmäßigkeit durchgeführt werden müssen, um zu einem besseren Gelingen von Qualitätsentwicklungsprozessen beizutragen. Dies gilt besonders für Schulen, die sich an neuartige Voraussetzungen anpassen müssen. Stellvertretend seien an dieser Stelle konzeptionelle Veränderungen wie die Entwicklung zu einer Ganztagschule oder strukturelle Veränderungen wie sinkende Schülerzahlen genannt. Es lohnt sich also, darüber nachzudenken, wie die Evaluation von Schulen möglichst effektiv vonstatten gehen kann.

Regelmäßige schriftliche Fragebogenuntersuchungen, die mittels Papier- und Bleistift-Verfahren durchgeführt werden, sind auf Dauer mit einem hohen finanziellen Aufwand verbunden. Als Alternative dazu wurden im vorliegenden Kapitel verschiedene computergestützte Datenerhebungsverfahren diskutiert, für die eine Reihe von Vorteilen sprechen: Neben geringeren Kosten sind dies die Zeitersparnis bei der Datenaufbereitung, die schnellere Verfügbarkeit der Daten, die durch den Einsatz von digitalen Medien bedingte Vielfalt von Gestaltungsoptionen, die mögliche Erhebung von Metadaten sowie die Reduktion von Fehlern im Datenerhebungs- und Datenverarbeitungsprozess. Um herauszufinden, welches Verfahren beim Einsatz von computerbasierten Systemen am sinnvollsten ist, wurden mobile Datenerhebungstechniken – der Einsatz von PDAs bzw. der Einsatz von Tablet PCs – mit Verfahren verglichen, die sich des Internets bedienen. Bei diesen internetbasierten Befragungen wurde wiederum unterschieden zwischen einer „autarken“ Befragung, bei der die technische Abwicklung von den Evaluatoren selbst übernommen wird, und einer nicht-autarken Befragung, die sich die Hilfe externer Anbieter zunutze macht.

Der Vergleich ergab, dass Online-Befragungen mobilen Datenerhebungsverfahren vor allem aus technischer Sicht in vieler Hinsicht (noch) überlegen sind: Dies gilt für die Ausstattung der Hardware (z.B. Rechenleistung, Speicherausstattung) genauso wie für den Wartungsbedarf und die Stromversorgung (z.B. die Anzahl der Netzanschlüsse zum Aufladen von Akkus). Hinsichtlich der visuellen Darstellungsmöglichkeiten waren zwischen Desktop-PC und Tablet-PC geringe Unterschiede festzustellen; zumindest zum momentanen Zeitpunkt sind die Darstellungsmöglichkeiten der PDAs noch nicht ausreichend. Ähnliches gilt für die Bedienbarkeit der Taschencomputer sowie die Software, die für die Geräte verfügbar ist. Was die datentechnische Organisation (z.B. Datenaustausch, Datensicherheit) angeht, erwiesen sich die Online-Verfahren als praktikabler. Und schließlich weisen auch der Vergleich der Kalkulationen sowie die organisatorischen Voraussetzungen darauf hin, dass Online-Verfahren gegenwärtig als Alternative zu Papier- und Bleistift-Befragungen am besten geeignet sind. Dafür spricht auch, dass dieses zeit- und ortonabhängige Verfahren für die Befragung des Lehrpersonals und der Eltern eingesetzt werden kann und so vergleichsweise einfach verschiedene Perspektiven des schulischen Alltags erfasst werden können.

Die Durchführung von Befragungen mittels mobiler Techniken erweist sich derzeit noch als wenig praktikabel. Gründe hierfür sind nicht nur der Preis, der technische Entwicklungsstand der Geräte und der Mangel an einer geeigneten Software, sondern auch die relativ geringe Verbreitung der Geräte. Entwicklungen in Kanada und den USA zeigen jedoch, dass PDAs dort nicht nur im universitären Alltag von vielen Studierenden zunehmend für die Arbeit genutzt werden, sondern auch im Rahmen von Feldarbeiten erprobt werden. Perspektivisch ist davon auszugehen, dass es auch in Europa zu einer noch stärkeren Ausbreitung der Geräte kommen wird.

Neben technischen Merkmalen wurden auch methodische Erwägungen in unsere Überlegungen und Abwägungen einbezogen. Zusätzlich zu der Frage nach der Ausstattung der Schulen mit Computerarbeitsplätzen wurden vor allem allgemeine methodische Probleme für Online-Evaluationen diskutiert. Es zeigte sich, dass methodische Schwierigkeiten, die aufgrund einer fehlenden Grundgesamtheit mit Online-Befragungen im Allgemeinen verbunden werden, bei einer schulinternen Befragung wegfallen; des Weiteren können Stichprobenfehler, vor allem wenn die Befragungen klassenweise durchgeführt werden, vermieden und der Anteil der Ausfälle durch ein attraktives Design verringert werden. Zwar treten auch bei Online-Befragungen Probleme auf; dies zeigen die praktischen Erfahrungen mit der Selbstevaluation einer Schule des Projekts „Unsere Schule“, von denen in Abschnitt 5 berichtet wurde. Diese Fehlerquellen können jedoch bereits durch einen ausführlichen Test der eingesetzten Software und eine gute Schulung des Erhebungspersonals minimiert werden.

Die Befürchtung, Schülerbefragungen könnten an den mangelnden technischen Fertigkeiten der Schülerinnen und Schülern scheitern, wurde im Rahmen dieser Befragung interessanterweise nicht bestätigt. Dies kann als Hinweis darauf verstanden werden, dass der Umgang mit Computern bereits zu einer Kulturfertigkeit der jüngeren Generation geworden ist. Es ist anzunehmen, dass viele Kinder in Zukunft bereits im Kindergartenalter Gelegenheit haben, sich mit den neuen Techniken vertraut zu machen. Spätestens jedoch mit dem Eintritt ins Schulalter wird die Nutzung von Computern immer selbstverständlicher. Hierzu trägt nicht zuletzt die sich stetig verbessernde Ausstattung der Schulen mit Computern bei, die den Einsatz neuer Lehr- und Lernmethoden ermöglicht und den Umgang mit dem Internet und gängigen Softwareanwendungen selbstverständlich macht.

Es ist also für Schulen, die ihren Schulentwicklungsprozess durch kontinuierliche Selbstevaluationen begleiten wollen, sinnvoll, sich mit den Möglichkeiten und den Angeboten internetbasierter Befragungsmethoden zu beschäftigen. Wurde die Selbstevaluation bereits in Angriff genommen, so ist es ratsam, das Expertenwissen über die technischen Details, die Entstehung des Fragebogens sowie die Modalitäten der Auswertung im Evaluationsteam breit zu streuen. Dies hilft dabei, die Kontinuität der Evaluationen zu gewährleisten.

Es wurde bereits darauf hingewiesen, dass im angloamerikanischen Raum vermehrt mobile Befragungsmethoden für Evaluationsaufgaben und die wissenschaftliche Arbeit genutzt werden. Obwohl sich im Verlauf des vorliegenden Kapitels gezeigt hat, dass für schulinterne Befragungen internetbasierte Befragungstechniken mobilen Verfahren überlegen sind, scheint es aus wissenschaftlicher Sicht zukünftig unverzichtbar, sich auch mit der technischen Entwicklung im Bereich der mobilen Befragungstechniken sowie den methodischen Besonderheiten beim Einsatz dieser Techniken zu beschäftigen.

## 4.7 Literatur

- Arbeitsstab Forum Bildung (2001). Empfehlungen des Forum Bildung. [www.forumbildung.de](http://www.forumbildung.de).
- Bandilla, W., Bosnjak, M., Schneid, M & Stiegler, A. (1999). Interaktive Medien als Instrument und Gegenstand der empirischen Sozialforschung. In Berghaus, M. (Hrsg.), *Interaktive Medien – interdisziplinär vernetzt* (S. 129-148). Opladen. Wiesbaden: Westdeutscher Verlag.
- Bandilla, W. (2004). Materialien zum Workshop „Online Surveys“ am 18./19. März 2004 im Zentrum für Umfragen, Methoden und Analysen (ZUMA). Unveröffentlichtes Manuskript.

- Baumert, J., Lehmann, R. (1997). TIMSS – Mathematisch-naturwissenschaftlicher Unterricht im internationalen Vergleich. Deskriptive Befunde. Opladen: Leske + Budrich.
- Baumert, J. (Hrsg.) (2001). PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich. Opladen: Leske + Budrich.
- BMBF (Hrsg.) (2003). IT-Ausstattung der allgemein bildenden und berufsbildenden Schulen in Deutschland. Reihe BMBF Publik. [www.bmbf.de](http://www.bmbf.de).
- Bringsjörd, E. L. (2001). Computerized-adaptive versus paper-and-pencil testing environments: An experimental analysis of examinee experience. Dissertation Abstracts International Section A: Humanities & Social Sciences, 62 (5-A), S. 1717.
- Dillman, D., Bowker, D. (2001). The web questionnaire challenge to survey methodologists. In Reips, U., Bosnjak, M. (Hrsg.), Dimensions of internet science (S. 159-177). Lengerich: Pabst Science Publishers.
- Faas, T. (2003). Umfragen im Umfeld der Bundestagswahl 2002: Offline und Online im Vergleich. ZA-Information 52, 120-135.
- Finger, M. S. & Ones, D. S. (1999). Psychometric equivalence of the computer and booklet forms of the MMPI: A meta-Analysis. Psychological Assessment, 11 (1), 58-66.
- Goltz, E. & Landua, D. (2003). Entwicklung der „Sozialen Schulqualität“. Ergebnisse einer Zeitreihenuntersuchung an 10 Brandenburger Schulen im Rahmen des Projekts „Schulqualität als Präventionsstrategie“. Unveröffentlichter Abschlussbericht.
- Hauptmanns, P. (1999). Grenzen und Chancen von quantitativen Befragungen mit Hilfe des Internet. In Batinic, B. (Hrsg.), Online-Research. Methoden, Anwendungen und Ergebnisse (S. 21). Göttingen: Hogrefe.
- Jäger, S. (1990). Computerdiagnostik – ein Überblick. Diagnostics, 36 (2), 96-114.
- Janetzko, D. (1999). Statistische Anwendungen im Internet. Daten in Netzwerkumgebungen erheben, auswerten, präsentieren. München, Reading (MA): Addison-Wesley.
- Kaczmirek, L. (2004). Umfragepraxis. Von der Idee zum Ergebnis. Materialien zum Workshop „Online Surveys“ am 18./19. März 2004 im Zentrum für Umfragen, Methoden und Analysen (ZUMA).
- Keil, J.-G. & Schroeter, K. (2002). Schulmeinungsforschung zur Verbesserung sozialer Schulqualität. Institut für berufliche Bildung und Weiterbildung (Hrsg.), Lehrbrief zum Fort- und Weiterbildungsmodell „Unsere Schule...“. Soziale Schulqualität – Schulinterne Evaluation, Fort- und Weiterbildung.
- Kohlmeier, K. (2002). Evaluation als Instrument von Schulentwicklung. Institut für berufliche Bildung und Weiterbildung (Hrsg.), Lehrbrief zum Fort- und Weiterbildungsmodell „Unsere Schule...“. Soziale Schulqualität – Schulinterne Evaluation, Fort- und Weiterbildung.
- Ott, H. J. (2004). eBusiness-Kompendium. [www.kecos.de/script](http://www.kecos.de/script) (29.09.Stand 2004).
- Pearrow, M. (2000). Web site usability handbook. Rockland, MA: Charles River Media.
- Powers, D.E. (2001). Test anxiety and test performance : Comparing paper –based and computer-adaptive versions of the Graduate Record Examination (GRE ©) General Test. Journal of Educational Computing Research, 24 (3), 249-273.
- Schneid, M. (2001). Mobile computergestützte Befragungen mit PDAs. ZUMA-Nachrichten 50, 121-142.
- Schnell, R., Hill, P. & Esser, E. (1993). Methoden der empirischen Sozialforschung. München, Wien : Oldenbourg.

Spray, J.A., Ackermann, T.A., Rechase, M.D. & Carlson, J.E. (1989). Effect of medium of item presentation on examinee performance and item characteristics. *Journal of Educational Measurement*, 25, 261-271.

Statistisches Bundesamt (Hrsg.). (2004). *Informationstechnologien in Haushalten. Ergebnisse einer Pilotstudie für das Jahr 2003*. Wiesbaden: Statistisches Bundesamt.

Wirth, T. (2002). *Missing Links. Über gutes Web Design*. Hauser: München, Wien.

## 5 Anhang

### *Anhang 1: Brief an den Landesdatenschutzbeauftragten den Landes Niedersachsen*

#### **Erläuterungen zur Durchführung der Schülerbefragung an der Beispielschule in Beispielort**

##### **1. Zur Befragung**

Die Befragung findet an einer allgemein bildenden Schule ab Klassenstufe 5 statt. Die Befragung erfolgt online, im Klassenverband und unter Nutzung eines standardisierten Fragebogens. Der Fragebogen wird mit dem über die Bundeszentrale für politische Bildung erhältlichen Programm GrafStat programmiert. In die Untersuchung werden mehrere Klassen einbezogen, die über einen nicht beeinflussbaren Auswahlsschlüssel ermittelt werden. Alle Eltern und jede Schülerin bzw. jeder Schüler erhält zwei Wochen vor der Befragung Informationsmaterial und Erhebungsunterlagen zur Einsicht.

##### **2. Verwaltung der Daten**

Schulen, die mit Hilfe des Programms GrafStat Online-Befragungen durchführen, haben die Möglichkeit, ihre Daten auf verschiedenen öffentlichen „Datensammelpunkten“ abzulegen, die auf folgenden Servern kostenlos zur Verfügung gestellt werden:

<b>Server</b>	<b>Adresse</b>
Learn:line	<a href="http://php.learn-line.de/GrafStat">http://php.learn-line.de/GrafStat</a>
Schule.de	<a href="http://www.nw.schule.de/GrafStat">www.nw.schule.de/GrafStat</a>
AMbis	<a href="http://www.know-how-netzwerk.de/GrafStat/">http://www.know-how-netzwerk.de/GrafStat/</a>
IDEE Haefele KEG	<a href="http://www.qualifizierung.com/GrafStat">http://www.qualifizierung.com/GrafStat</a>
Studienseminare	<a href="http://www.studienseminare.nrw.de/GrafStat">www.studienseminare.nrw.de/GrafStat</a>
Ralph-Erich Hildebrandt	<a href="http://www.rehs.de/GrafStat/">www.rehs.de/GrafStat/</a>

Am Ende der Befragung werden die Daten (Arbeitsgruppe „Soziale Schulqualität“) vom Server abgerufen. Hierzu sind zwei spezielle Passwörter notwendig, die nur den für die Befragung Verantwortlichen bekannt sind: Ein Serverpasswort und ein Passwort für die Befragung.

Die Beispielschule beabsichtigt, die Befragungsdaten auf dem Learn-Line-Server zu sammeln. Da jedoch nicht immer alle Server mit Datensammelpunkten am Netz sind und sich von Zeit zu Zeit die Adressen ändern, behält sich die Schule vor, im Notfall auf einen der oben aufgeführten Server zurückzugreifen.

##### **3. Anonymität und Widerrufsrecht**

Die Befragung erfolgt anonym. Es werden keine Daten erhoben, gespeichert oder der Schulöffentlichkeit präsentiert, die Rückschlüsse auf die Person des Befragten erlauben.

Alle Schüler erhalten zu Beginn der Befragung eine so genannte Transaktionsnummer, die nur ihnen bekannt ist und die vor der Beantwortung des Fragebogens eingegeben werden muss. Diese Nummer soll dafür sorgen, dass jeder Schüler nur einmal an der Befragung teilnehmen kann. Die Nummern verbleiben auch nach der Befragung bei den Schülern. Möchte ein Schüler seine Angaben widerrufen, so teilt er den verantwortlichen Lehrern seine Transaktionsnummer mit, die den Datensatz dann umgehend löschen.

##### **4. Verbleib der Einwilligungserklärungen**

Es ist vorgesehen, dass die unterschriebenen Einwilligungserklärungen von der Beispielschule entgegengenommen und bis ein Jahr nach der Befragung verwahrt werden. Es wird sichergestellt, dass personenbezogene Angaben der Einverständniserklärungen und die Befragungsdaten räumlich und personell getrennt aufbewahrt werden.

## **Anhang 2: Einwilligungserklärung von Jugendlichen zur freiwilligen Teilnahme an einer schriftlichen Befragung**

Liebe Schülerin, lieber Schüler,

wir möchten gerne wissen, wie zufrieden unsere Schülerinnen und Schüler mit dem Leben und dem Lernumfeld an unserer Schule sind. Damit wir einen guten Überblick darüber erhalten, was ihr über die Situation an der Beispielschule denkt, ist es für uns sehr wichtig, dass sich möglichst viele Schülerinnen und Schüler beteiligen. Jede einzelne Meinung ist uns wichtig! Deshalb möchten wir dich herzlich darum bitten, unseren Fragebogen ausfüllen.

Der Fragebogen wird am Computer ausgefüllt. Das ist ganz einfach: Neben den Antworten, die man auf eine Frage geben möchte, ist ein Kästchen, das man einfach mit der Maus anklicken kann.

Die Beantwortung der Fragen ist freiwillig. „Freiwillig“, das heißt: Wenn du nicht mitmachen willst, kannst du die Teilnahme an der Befragung (oder die Beantwortung einzelner Fragen) verweigern – und zwar auch dann, wenn deine Eltern der Befragung bereits zugestimmt haben! Ob du nun mitmachst oder nicht - es entstehen dir keinerlei Nachteile. Des Weiteren werden deine Angaben namenlos erfasst. „Namenlos“, das heißt: Es kann vom Fragebogen nicht auf die Person geschlossen werden, die ihn ausgefüllt hat. Weder deine Lehrerinnen und Lehrer noch deine Eltern oder andere Personen können herausfinden, welches dein Fragebogen ist!

Praktisch geht's wie folgt weiter:

1. Wenn du mitmachen möchtest, fülle bitte den unteren Abschnitt dieser Einwilligungserklärung aus.
2. Gebe den Abschnitt deiner Lehrerin oder deinem Lehrer.
3. Gebe folgende Codenummer für den Datensatz am Bildschirm ein:

1234567 (zufällig generierte Nummer)

4. Fülle den Fragebogen aus.

**Vielen Dank für deine Mitarbeit!**

**Unterschrift!**

Bitte eintragen: .....

(dein Vor- und Zuname)

.....

(deine Klasse)

Ich bin mit der Teilnahme an der Befragung einverstanden: Ja  Nein

(Zutreffendes bitte ankreuzen)

.....  
Datum und Unterschrift

### Anhang 3: Einwilligungserklärung Eltern

#### Einwilligungserklärung zur freiwilligen Teilnahme Ihres Sohnes/ Ihrer Tochter an der Schülerbefragung der Beispielschule

Liebe Eltern,

uns interessiert, wie zufrieden unsere Schülerinnen und Schüler mit dem Leben und dem Lernumfeld an der Beispielschule sind. Aus diesem Grunde beabsichtigen wir, vom 10. bis zum 14. Mai 2004 erstmals eine eigene Befragung unter unseren Schülerinnen und Schülern durchzuführen.

Der Fragebogen wird am Computer ausgefüllt. Die Befragung erfolgt anonym, das heißt mit dem Fragebogen werden keine Angaben erfasst, mit denen Rückschlüsse auf die Identität der befragten Person möglich sind. Die Daten werden dann in anonymisierter Form auf einem Datensammelpunkt außerhalb der Schule gespeichert und entsprechend den Datenschutzbestimmungen des Landes Niedersachsen vor dem Zugriff Unbefugter geschützt. Die Ergebnisse der Befragungen werden anschließend durch die Schulleitung vorgestellt und mit Schülern, Eltern und Lehrern diskutiert.

Die Schulleitung, der Gesamtelternbeirat sowie das Lehrerkollegium haben sich bereits mit der Durchführung der Befragung einverstanden erklärt. Da Ihr Kind eine der zu befragenden Klassen besucht, bitten wir nun auch Sie um Ihre Einwilligung, dass Ihr Kind an der oben beschriebenen Untersuchung teilnehmen darf. Jede einzelne Meinung ist uns wichtig, denn wir möchten einen guten Überblick darüber bekommen, was die Schülerinnen und Schüler über die Beispielschule denken.

Die Teilnahme an der Untersuchung ist freiwillig und jederzeit widerrufbar. Ihr Kind ist trotz Ihrer Zustimmung nicht verpflichtet, an der Untersuchung teilzunehmen, es kann aus eigener Entscheidung die Teilnahme verweigern bzw. widerrufen. Falls Sie die Erlaubnis zur Teilnahme zu einem späteren Zeitpunkt noch widerrufen möchten, wenden Sie sich einfach an die Schulleitung (Telefonnummer). Wenn Ihre Tochter bzw. Ihr Sohn nicht an der Befragung teilnimmt, hat das keinerlei Nachteile für Ihr Kind.

Zum weiteren Ablauf:

1. Füllen Sie bitte zunächst den unteren Abschnitt der Einwilligungserklärung aus.
2. Trennen Sie ihn vom oberen Informationsteil ab.
3. Geben Sie Ihrem Kind den unteren Abschnitt mit zurück in die Schule.

Es können nur (minderjährige) Jugendliche an der Untersuchung teilnehmen, deren Eltern der Teilnahme an der Untersuchung zugestimmt haben und die selbst ebenfalls bereit sind, an der Befragung mitzuwirken.

Vielen Dank für Ihre Mitarbeit!

(Unterschrift)

✂-----✂-----✂-----✂-----✂-----✂-----

Wir stimmen der Teilnahme unserer Tochter/ unseres

Sohnes..... an der Schülerbefragung der Beispielschule zu:

Ja                       Nein

(Zutreffendes bitte ankreuzen)

\_\_\_\_\_  
Datum und Unterschrift der Eltern



**Anhang 4: Teilnahmeliste**

**Schülerbefragung der Don-Bosco-Schule 2004**

**Teilnahmeliste**

<b>Datum:</b>	
<b>In welcher Stunde findet Befragung statt?</b>	
<b>Klasse:</b>	
<b>Anzahl der Schüler in der Klasse:</b>	
<b>Anzahl der teilnehmenden Schüler/innen:</b>	
<b>Anzahl der nicht teilnehmenden Schüler/innen:</b>	

Name, Vorname	Einwilligung Eltern		Einwilligung Schüler/in		Krankheit	Sonstige Gründe für Nichtteilnahme (bitte benennen)
	Ja	Nein	Ja	Nein		
1.						
2.						
3.						
4.						
5.						
6.....						

**Anhang 5: Protokoll der Datenerhebung**

Datum:	
Klasse:	
Anzahl teilnehmende Schüler:	
Aufsicht:	
Technische Probleme	
Probleme bei der Bedienung des Computers	
Verständnisprobleme, die sich auf die Inhalte des Fragebogens beziehen	
Zeitliche Probleme	
Sonstige Bemerkungen zum Verlauf der Befragung	

### **Anhang 6: Abbildungsverzeichnis**

Abbildung 1	Schulqualitätsmerkmale aus den USA .....	39
Abbildung 2	Qualitätsdimensionen von INIS .....	42
Abbildung 3	Qualitätsbereiche in der Schulevaluation des Kantons Luzern .....	54
Abbildung 4	Zum Umfang der Datenerhebungen im Projekt „Unsere Schule...“ .....	69
Abbildung 5	Beteiligte Bundesländer .....	70
Abbildung 6	Titelblatt des Fragebogens zum Projekt „Unsere Schule...“ .....	78
Abbildung 7	Auszug aus dem Fragebogen.....	80
Abbildung 8	Übersicht über Inhalte der Schülerbefragungen des Projekts „Unsere Schule...“	82
Abbildung 9	Übersicht zum Inhalt des schriftlichen Schulreports.....	92
Abbildung 10	Ausschnitt aus dem Schulreport.....	94
Abbildung 11	Ergebnisse der Selbstevaluation der Datenpräsentationsleistungen im Rahmen des Projekts „Unsere Schule...“; (N = 176) .....	100
Abbildung 12	Anteile von Schulen mit verbessertem „Schulqualitätsindex“(*) in einzelnen Bundesländern	101
Abbildung 13	Ablauf einer Online-Erhebung .....	128
Abbildung 14	Ausschnitt aus einem Flussdiagramm .....	130
Abbildung 15	Beispielhafter Ablauf einer Online-Befragung (Pretest).....	136

### **Anhang 7: Tabellenverzeichnis**

Tabelle 1: Möglichkeiten der technisierten Datenerhebung .....	109
Tabelle 2: Bewertungskriterien für Betriebssysteme verschiedener Geräte .....	114
Tabelle 3: Abwägungskriterien für die Software zur Durchführung von Datenerhebungen .	116
Tabelle 4: Übersicht über Kriterien der Datensicherung für verschiedene Verfahren.....	117
Tabelle 5: Aspekte der Fragebogengenerierung für verschiedene Verfahren.....	119
Tabelle 6: Investitions- und Betriebskosten der verschiedenen Verfahren im Vergleich (in Euro).....	121