

# Implementationsgrad und Wirksamkeit von Unterrichtsinterventionen zur Förderung der selbstbestimmten Lernmotivation

Dissertation zur Erlangung des akademischen Grades  
Doktor der Philosophie (Dr. phil.) im Fach Psychologie  
an der Pädagogischen Hochschule Ludwigsburg

vorgelegt von

Dipl.- Psych. Daniela Kuhn M. A.

Ludwigsburg

10.01.2018

Erstgutachter: Prof. Dr. Waldemar Mittag

Zweitgutachter: Prof. Dr. Ulf Kieschke

Datum des Abschlusses der mündlichen Prüfung: 19. Juni 2018

## Danksagung

Ganz besonders danken möchte ich Prof. Dr. Waldemar Mittag für die Betreuung bei der Anfertigung dieser Arbeit. Ebenso möchte ich Prof. Dr. Barbara Reichle danken, die mich bei meinem wissenschaftlichen Werdegang ermutigt und unterstützend begleitet hat.

Ebenfalls bedanken möchte ich mich bei Prof. Dr. Ulf Kieschke für die Begutachtung der Dissertation. Darüber hinaus haben weitere Personen einen Anteil an der Fertigstellung dieser Arbeit. So gebührt Heike Wenisch und Hans-Jürgen Schmidt ein großes Dankeschön für die Durchsicht.

Danken möchte ich auch den Lehrkräften, die Vertrauen in meine Fähigkeiten hatten und die bis dato vorhandene Freude am Lernen in mir wecken konnten. Hierzu gehören Edmund Wöllstein, Karl Herrmann Gutheil, Norbert Ebrecht und Elisabeth Wetchy.

Darüber hinaus danke ich René Schmitt, ohne dessen Ermutigung und Unterstützung ich diese Laufbahn nicht erfolgreich beschritten hätte.

Ein solches Projekt gelingt nicht ohne die Unterstützung der Familie. Herzlichen Dank an meine Eltern Rita und Dieter Kuhn für den Rückhalt und die vielseitige Unterstützung in den vergangenen Jahren.

## Zusammenfassung

Die vorliegende Arbeit untersucht, ob und inwieweit sich in Abhängigkeit von verschiedenen Maßen der Implementationsqualität unterschiedliche Programmwirkungen einer Interventionsstudie nachweisen lassen. Die Intervention beinhaltete curricular orientierte Unterrichtseinheiten auf der Grundlage der Selbstbestimmungstheorie der Motivation von Deci und Ryan (1993, 2000) sowie motivationsförderliche Maßnahmen zur Unterstützung des Erlebens von Autonomie, Kompetenz und sozialer Eingebundenheit. Die Wirksamkeitsevaluation dieser Intervention erfolgte in einem längsschnittlichen Treatment-Kontrollgruppen-Design mit insgesamt 353 Sekundarschüler\_innen aus 20 Haupt-, Real- und Gymnasialklassen der achten Jahrgangsstufe. Im Vergleich zu den Kontrollklassen zeigten sich in den Treatmentklassen positive Effekte der Unterrichtsinterventionen. Es zeigte sich allerdings auch deutliche Unterschiede in den beobachteten Interventionseffekten zwischen den Treatmentklassen (Mittag, Bieg, Hiller, Metz & Melenk, 2009). So wurde in der vorliegenden Arbeit untersucht, ob und inwieweit die Varianz zwischen den Treatmentklassen durch Unterschiede in der Implementationsqualität zu erklären ist. Als Indikatoren für die Implementationsqualität bzw. Programmintegrität werden in der Forschungsliteratur u. a. die Adhärenz (Einhaltung), Dosierung, Qualität der Programmvermittlung sowie die Responsivität der Programmteilnehmer und die Programmspezifität herangezogen (Dane & Schneider, 1998; Durlak & DuPre, 2008, Mittag, 2014). In der vorliegenden Studie wurden diese Indikatoren durch Angaben und Einschätzungen der Lehrpersonen (N = 9) erfasst und im Hinblick auf klassenbezogene Unterschiede in den Interventionseffekten untersucht. Hierzu wurden die lehrerbezogenen Angaben aus Feedback-Fragebögen im Hinblick auf Komponenten der Implementationsqualität klassifiziert und brauchbare quantitative Maße der Implementationsqualität entwickelt. Die Analysen zeigen je nach Implementationsqualität und Indikator erhebliche Wirksamkeitsunterschiede zwischen den Treatmentklassen. So zeigten sich für Klassen mit einem höheren Einsatz der Interventionsmaßnahmen (Dosierung) günstigere Entwicklungen des Autonomieerlebens und der Leistung der Schüler\_innen. Andere Indikatoren, wie z. B. die Qualität der Programmvermittlung und Responsivität der Programmteilnehmer, klären entgegen der Annahme keine Unterschiede in Bezug auf die Wirksamkeit der Interventionsmaßnahmen auf. Die Ergebnisse liefern dennoch Hinweise auf Zusammenhänge einzelner Teilaspekte der Implementationsqualität mit der Wirksamkeit der Interventionsmaßnahmen.

# Inhalt

Einleitung .....	9
<b>1 Implementationsforschung.....</b>	<b>11</b>
1.1 Indikatoren der Implementationsqualität .....	13
1.2 Methoden zur Erfassung der Indikatoren der Implementationsqualität.....	19
1.2.1 Adhärenz .....	21
1.2.2 Dosierung .....	25
1.2.3 Qualität der Programmvermittlung .....	28
1.2.4 Responsivität der Programmteilnehmer .....	30
1.2.5 Programmspezifität .....	31
1.3 Auswirkungen der Implementationsqualität auf die Programmwirksamkeit....	35
1.3.1 Studien zur Adhärenz .....	38
1.3.2 Studien zur Dosierung .....	41
1.3.3 Studien zur Responsivität der Programmteilnehmer.....	46
1.3.4 Studien zur Qualität der Programmvermittlung.....	47
1.3.5 Studien zur Programmspezifität .....	50
1.4 Die Implementation beeinflussende Faktoren.....	51
1.4.1 Adaptation von Programminhalten: Fidelitäts-Anpassungs-Debatte.....	51
1.4.2 Implementation: Merkmale von Konzeption, Vermittler und Organisation.....	54
<b>2 Ableitungen für die Intervention .....</b>	<b>66</b>
2.1 Theorien zur Lernmotivation .....	66
2.2 Die Selbstbestimmungstheorie und ihre grundlegenden Annahmen zur Förderung einer selbstbestimmten Lernmotivation .....	80
2.3 Möglichkeiten der Förderung einer selbstbestimmten Lernmotivation in Schule und Unterricht .....	87
2.4 Interventionsprogramm zur Förderung der selbstbestimmten Lernmotivation	89
<b>3 Fragestellung und Hypothesen.....</b>	<b>104</b>
<b>4 Methode.....</b>	<b>106</b>
4.1 Stichprobe .....	106
4.2 Instrumente.....	108
4.2.1 Lehrerfragebogen zur Erfassung der Implementationsqualität.....	108
4.2.2 Klassifikation und Quantifizierung der lehrerbezogenen Angaben zur Implementationsqualität.....	109
4.2.3 Schülerskalen zu Erfassung der Wirksamkeitsindikatoren.....	112
4.3 Auswertungsverfahren .....	114

<b>5</b>	<b>Ergebnisse .....</b>	<b>115</b>
5.1	Interkorrelationen der erfassten Implementationsindikatoren .....	115
5.1.1	Interkorrelationen der erfassten Implementationsindikatoren UE 1 .....	116
5.1.2	Implementationsqualität und Wirksamkeit: Interkorrelationen der erfassten Implementationsindikatoren UE 2 .....	118
5.1.3	Vergleichende Betrachtung der Zusammenhänge der Implementationsindikatoren in UE 1 und UE 2 .....	120
5.2	Korrelationen zwischen den Implementationsindikatoren und den klassenbezogenen Effektgrößen .....	121
5.2.1	Korrelationen zwischen den Implementationsindikatoren und den klassenbezogenen Effektgrößen bei den Wirksamkeitsindikatoren UE 1 .....	121
5.2.2	Korrelationen zwischen den Implementationsindikatoren und den klassenbezogenen Effektgrößen bei den Wirksamkeitsindikatoren UE 2 .....	124
5.2.3	Vergleichende Betrachtung der Zusammenhänge zwischen den Implementationsindikatoren und den klassenbezogenen Wirksamkeitsindikatoren in UE 1 u. UE 2 .....	127
5.3	Mehrebenenanalysen .....	130
5.3.1	Leistungsentwicklung UE 1 .....	131
5.3.2	Autonomieentwicklung UE 1 .....	134
5.3.3	Entwicklung der Lernfreude UE 1 .....	137
5.3.4	Leistungsentwicklung UE2 .....	139
<b>6</b>	<b>Zusammenfassende Diskussion .....</b>	<b>143</b>
	<b>Literatur .....</b>	<b>156</b>
	<b>Anhang .....</b>	<b>169</b>

## Tabellenverzeichnis

Tabelle 1 Studien zu Zusammenhängen zwischen der Adhärenz und den betrachteten Zielkriterien .....	39
Tabelle 2 Studien zu Zusammenhängen zwischen Dosierung und den betrachteten Zielkriterien .....	42
Tabelle 3 Studien zu Zusammenhängen zwischen mehreren Indikatoren und den betrachteten Zielkriterien .....	47
Tabelle 4 Rücklauf Feedback-Fragebögen.....	107
Tabelle 5 Indikatoren zur Bewertung der Implementationsqualität der Unterrichtsinterventionen.....	111
Tabelle 6 Interkorrelationen der erfassten Implementationsindikatoren UE 1 .....	116
Tabelle 7 Interkorrelationen der erfassten Implementationsindikatoren der UE 2 .....	118
Tabelle 8 Korrelationen zwischen den Implementationsindikatoren der UE 1 und den klassenbezogenen Effektgrößen bei den Wirksamkeitsindikatoren .....	122
Tabelle 9 Korrelationen zwischen den Implementationsindikatoren der UE 1 und den klassenbezogenen Effektgrößen bei dem Wirksamkeitsindikator Leistung.....	123
Tabelle 10 Korrelationen zwischen den Implementationsindikatoren der UE 2 und den klassenbezogenen Effektgrößen bei den Wirksamkeitsindikatoren .....	124
Tabelle 11 Korrelationen zwischen den Implementationsindikatoren UE2 und den klassenbezogenen Effektgrößen bei dem Wirksamkeitsindikator Leistung.....	126
Tabelle 12 Korrelationen zwischen den Implementationsindikatoren für die Unterrichtseinheiten 1 und 2 und den klassenbezogenen Effektgrößen bei den Wirksamkeitsindikatoren.....	127
Tabelle 13 Korrelationen zwischen den Implementationsindikatoren für die Unterrichtseinheiten 1 und 2 und den klassenbezogenen Effektgrößen bei der Leistung....	129
Tabelle 14 Leistungsentwicklung UE 1 .....	132
Tabelle 15 Autonomieentwicklung UE1 .....	136
Tabelle 16 Entwicklung der Lernfreude UE1 .....	139
Tabelle 17 Leistungsentwicklung UE 2 .....	141

## Abbildungsverzeichnis

<i>Abbildung 1.</i> Aspekte der Programmintegrität bzw. Implementationsqualität. ....	14
<i>Abbildung 2.</i> Häufigkeit der in den Studien betrachteten Indikatoren zur Implementationsqualität .....	37
<i>Abbildung 3.</i> Bedeutung der Programmcharakteristik und der Kontextbedingungen für die Implementationsqualität.....	55
<i>Abbildung 4.</i> Konzeptuelles Rahmenmodell der Implementationsqualität. ....	58
<i>Abbildung 5.</i> Eine Taxonomie menschlicher Motivation.....	82
<i>Abbildung 6.</i> Förderung der selbstbestimmten Lernmotivation.....	90
<i>Abbildung 7.</i> Erhebungsdesign.....	95
<i>Abbildung 8.</i> Effektstärken in Bezug auf Schülerskalen und Schulform. ....	97
<i>Abbildung 9.</i> Autonomieerleben nach Schulform. ....	98
<i>Abbildung 10.</i> Wahrgenommene Lehrerfürsorglichkeit.....	99
<i>Abbildung 11.</i> Lernfreude.....	100
<i>Abbildung 12.</i> Selbstbestimmte Lernmotivation.....	101
<i>Abbildung 13.</i> Leistungsentwicklung.....	102
<i>Abbildung 14.</i> Leistungsentwicklung UE1.....	134
<i>Abbildung 15.</i> Autonomieentwicklung UE1 .....	137
<i>Abbildung 16.</i> Leistungsentwicklung UE2.....	142



---

## Einleitung

Der Schulalltag konfrontiert Lehrkräfte und Schüler\_innen mit unterschiedlichsten Problemstellungen und Anforderungen. Deren Bewältigung wirft die Frage nach adäquaten präventiven und interventiven Maßnahmen auf, mit deren Entwicklung und Evaluation sich die Interventionsforschung beschäftigt. Die pädagogische und psychologische Forschung beschäftigt sich mit der Entwicklung und Evaluation solcher Maßnahmen. Diese Präventions- bzw. Interventionsmaßnahmen werden auf Basis theoretischer Grundlagen entwickelt, in umsetzbare Handlungselemente überführt und in ihrer Durchführung, Wirkung und ihren Folgen untersucht. Weitgehend außer Acht gelassen wurden aber bisher im Bereich der Unterrichts- und Lehr-Lern-Forschung das tatsächliche Ausmaß der Umsetzung des Programmkonzepts sowie die Folgen einer unzureichenden bzw. mangelhaften Umsetzung der Programmaktivitäten in die Praxis.

Das Konstrukt *Implementationsqualität* bezeichnet die getreue Umsetzung der geplanten Maßnahmen und Programmaktivitäten. Bisherige Befunde und der Diskurs in der wissenschaftlichen Forschung zur Bedeutung der Implementationsqualität deuten darauf hin, dass ein nicht unerheblicher Anteil der Effektivitätsvarianz durch Unterschiede in der Implementationsqualität von Unterrichtsinterventionen zu begründen ist. Allerdings wurde über die Art der Erfassung der Implementationsqualität bislang kein Konsens erzielt.

Die vorliegende Arbeit fokussiert in ihrer Hauptzielsetzung auf die Beziehungen zwischen der Implementationsqualität bei der Umsetzung von unterrichtlichen Maßnahmen zur Förderung der selbstbestimmten Lernmotivation und der Wirksamkeit dieser Maßnahmen bei den Schüler\_innen. Hierzu werden der aktuelle Forschungsstand zur Implementationsforschung aufgearbeitet und verschiedene theoretische Herangehensweisen untersucht. Die daraus abgeleiteten Erkenntnisse sind für die vorliegende Arbeit maßgeblich und forschungsleitend.

Zu Beginn werden die Hintergründe zur Implementationsforschung näher erläutert. Dann werden die für die Fragestellung zentralen Indikatoren der Implementationsqualität eingeführt und Methoden zur Erfassung der Indikatoren aus bisherigen Studien in den Abschnitten 1.1 und 1.2 ausgeführt. Dieser Begriffsklärung schließt sich in den Abschnitten zu 1.3 ein Überblick der Befunde zu Zusammenhängen der Implementationsqualität mit der Wirksamkeit von

Interventionen an Faktoren, die ein Gelingen der Implementation von Interventionsmaßnahmen beeinflussen, werden in Abschnitt 1.4 vorgestellt. In Kapitel 2 werden die Theorien zur Lernmotivation, insbesondere die Selbstbestimmungstheorie und ihre grundlegenden Annahmen zur Förderung einer selbstbestimmten Lernmotivation, dargelegt. Diesen Ausführungen schließt sich die Darstellung der Möglichkeiten zur Förderung einer selbstbestimmten Lernmotivation in Schule und Unterricht an. Im Weiteren wird das dieser Arbeit zu Grunde liegende Interventionsprogramm zur Förderung der selbstbestimmten Lernmotivation vorgestellt.

In Kapitel 3 werden die Fragestellung und die inhaltlichen Hypothesen erläutert. Im folgenden Kapitel 4 schließt sich die Beschreibung der Stichprobe, des eingesetzten Erhebungsinstrumente zur Erfassung der Implementationsqualität (Feedback-Fragebogen) und der Schülermerkmale an. Darüber hinaus die Beschreibung der Vorgehensweise bei der Klassifikation und Quantifizierung der lehrerbezogenen Angaben zur Implementationsqualität. Die Ergebnisse zu Zusammenhängen innerhalb der Implementationsindikatoren, zu den Wirksamkeitsindikatoren sowie die mehrebenenanalytischen Betrachtungen des Einflusses der Implementationsindikatoren auf die Entwicklung bei den Schüler\_innen sind Gegenstand von Kapitel 5. Abschließend erfolgt eine zusammenfassende Diskussion der Studie und der zentralen Ergebnisse.

# 1 Implementationsforschung

Unterrichtliche Maßnahmen können in Abhängigkeit von einer intendierten Zielsetzung ebenso wie Fördermaßnahmen unter den Begriff der Intervention gefasst werden. So ist jeder Eingriff in ein bestehendes System, der eine Zielsetzung verfolgt und/oder gleichzeitig eine Veränderung des Systems bewirken möchte, als Intervention zu betrachten (Leutner, 2012, 18).

Der Begriff der *Interventionsforschung* umfasst in seiner ursprünglichen Bedeutung die „Entwicklung von Interventionen auf der Grundlage technologischer Überlegungen und Theorien“ (Bortz & Döring, 2006, S. 102). Gleichwohl kann die pädagogische Interventionsforschung auch als ein umfassender Prozess im zeitlichen Verlauf verstanden werden, „bei dem sowohl die Entwicklung und Konzeption pädagogischer Interventionen als auch ihre Umsetzung und Implementation sowie die Wirksamkeit der Interventionsmaßnahmen mit empirisch wissenschaftlichen Methoden zu bewerten (evaluieren) sind“ (Mittag & Bieg, 2010, S. 31). Entsprechend erscheint eine begriffliche Trennung zwischen Evaluations- und Interventionsforschung kaum sinnvoll (Mittag et al., 2010, S. 31), da sich die Evaluationsforschung nach Rossi und Freeman (1993; Rossi, Freeman & Hofmann, 1988) auf drei Arten der Evaluation beziehen kann. Dies sind (1) die Evaluation der Programmkonzeption, (2) die Evaluation der Programmdurchführung und (3) die Evaluation der Programmwirksamkeit. Während sich die Evaluation der Programmkonzeption im engeren Sinn auf den Kernbereich der Interventionsforschung bezieht (Problembestimmung und Entscheidung über den Bereich der Intervention, eine Zielbestimmung, Konzeption und Gestaltung des Programms, Auswahl geeigneter diagnostischer Methoden und Verfahren zur Erfassung des angestrebten Programmerfolgs bzw. zur Erfassung der spezifischen Zielgruppe, Bewertung der Programmkonzeption), umfasst die Evaluation der Programmdurchführung die Supervision und Überwachung der Programmausführung und die Bewertung der Durchführbarkeit in der Praxis sowie die Prüfung der Programmreichweite. Über die Prüfung der Programmreichweite kann eruiert werden, ob die intendierte Zielgruppe an dem Programm teilnimmt und in welchem Umfang die Teilnahme erfolgt. Hierbei wird erfasst, ob Verzerrungen der Programmreichweite z. B. durch Selbstselektionen, Ausfälle oder unzureichende Definition der Zielgruppe vorliegen. Verzerrungen könnten in Folge die Einschätzung der Wirksamkeit bzw. Effizienz des Programms negativ beeinflussen (Mittag & Hager, 2000, S. 114 f.).

Unter den *Implementationsbegriff* fällt grundlegend „die Umsetzung von wissenschaftlichen Erkenntnissen in die gesellschaftliche Praxis“ (Euler & Sloane, 1998, S. 312). Eine erweiterte Definition sollte einschließen, wie theoretische Inhalte in praktisches Handeln überführt werden können. Die Erforschung der (qualitativen) Umsetzung der Maßnahmen in die Praxis kennzeichnet den Bereich der Implementationsforschung. Dabei ist sowohl die wissenschaftliche Bewertung von Maßnahmen als auch die Durchführung der Maßnahmen im Unterricht zu betrachten (Mittag, 2014). Im Hinblick auf die wissenschaftliche Bewertung der Implementation beschreibt Mittag (2014, S. 52) die Implementationsforschung als auf die *Durchführung und Umsetzung (Implementation) von Präventions- und Interventionskonzepten* ausgerichtet. Eine zentrale Fragestellung der Implementationsforschung beschäftigt sich demnach mit Merkmalen einer adäquaten Durchführung und Umsetzung der Maßnahmen und Konzepte. Eine weitere Forschungsfrage lautet, wie die Implementation in qualitativer Weise gefördert werden kann und deren Gelingen zu bewerten ist. Es ist also zu klären, durch welche Merkmale eine gelungene Implementation zu erfassen und die Implementation in der Praxis zu gewährleisten ist. Im Mittelpunkt stehen dabei die Erfassung und Sicherung der *Implementationsqualität* (implementation fidelity) und *Programmintegrität* (Mittag, 2014, S. 52 f.). Betrachtet man das Konstrukt der Implementationsqualität, so bezeichnet diese die *getreue* Umsetzung der geplanten Programmaktivitäten im Hinblick auf die Programmkonzeption in der Praxis (Mihalic, 2004). Forschungsbefunde weisen darauf hin, dass ausbleibende Interventionserfolge auch auf eine unzureichende Implementierung bzw. mangelhafte Implementationsqualität zurückgeführt werden können (Snyder, Bolin & Zumwalt, 1992).

In der Forschungsliteratur werden zur Bezeichnung der Implementationsqualität verschiedene Begriffe verwendet. In englischsprachigen Fachartikeln finden sich z. B. die Begriffe *treatment integrity* (Programmintegrität) oder *implementation fidelity* (Implementationsgüte, Implementationsqualität, Implementationstreue oder Implementationsgenauigkeit). Zudem gibt es unterschiedliche Auffassungen darüber, welche Aspekte bzw. Indikatoren unter den Begriff der Implementationsqualität einzuordnen sind. So finden sich z. B. Unterschiede in den Operationalisierungen der Indikatoren zur Erfassung der Implementationsqualität. Hinzu kommt, dass in den jeweiligen Studien meist nur einzelne Indikatoren erfasst und/oder die Indikatoren nicht einheitlich definiert und gemessen werden. „Definitions and measures of fidelity were found not to be consistent across studies, and new definitions are proposed“ (Dusenbury et al., 2003, S. 237). Neben der Überwachung der Implementationsqualität ist

demnach auch eine Vereinheitlichung bei der Definition und Erfassung der Implementationsindikatoren notwendig. Die Beachtung der Implementationsqualität und die Betrachtung der verschiedenen Indikatoren sind insofern von Bedeutung, als deren Messung Auskunft über das Gelingen der Implementation geben kann und somit auch Rückschlüsse auf ausbleibende Effekte der Intervention erlaubt. Nach Sichtung zahlreicher Studien gehen Dusenbury et al. (2003) davon aus, dass nicht alle Implementationsinhalte wie gewünscht umgesetzt werden und selbst ein Training der umsetzenden Personen nicht ausreicht, um eine zufriedenstellende Implementationsqualität zu gewährleisten. „While there has been limited research on fidelity of implementation in the social sciences, research in drug abuse prevention provides evidence that poor implementation is likely to result in a loss of program effectiveness. Studies indicate that most teachers do not cover everything in a curriculum, they are likely to teach less over time and training alone is not sufficient to ensure fidelity of implementation” (Dusenbury et al., 2003, S. 237). Seine Ausführungen weisen zudem auf die Notwendigkeit der Überprüfung, Dokumentation und zusätzlich flankierender, unterstützender Maßnahmen zur adäquaten Implementation hin. Die zentrale Fragestellung der vorliegenden Arbeit beschäftigt sich im Sinne der Implementationsforschung mit den Merkmalen einer adäquaten Durchführung und Umsetzung der Maßnahmen und Konzepte. Im Folgenden soll erörtert werden, anhand welcher Indikatoren eine gelungene Implementation erfasst werden kann.

## **1.1 Indikatoren der Implementationsqualität**

Von verschiedenen Autoren werden voneinander abgrenzbare Teilaspekte genannt, die das Konstrukt der Implementationsqualität als mehrdimensional kennzeichnen. In den einschlägigen Meta-Analysen und Überblicksartikeln (Berkel, Mauricio, Schoenfelder & Sandler, 2011; Dane & Schneider, 1998; Durlak & DuPre, 2008; Dusenbury et al., 2003) finden sich bis zu acht Teilaspekte der Implementationsqualität. In der vorliegenden Arbeit wird vor allem Bezug genommen auf die von Dane und Schneider (1998, S. 44 f.) herausgearbeiteten Aspekte, die auch in der Forschungsliteratur einen breiten Konsens gefunden haben. Dane und Schneider (1998) schlagen fünf Aspekte bzw. Dimensionen zur Erfassung der Implementationsqualität vor. Dies sind (1) die Adhärenz, (2) die Dosierung, (3) die Qualität der Programmvermittlung, (4) die Responsivität der Programmteilnehmenden und (5) die Programmdifferenzierung/Programmspezifität (s. *Abbildung 1*).

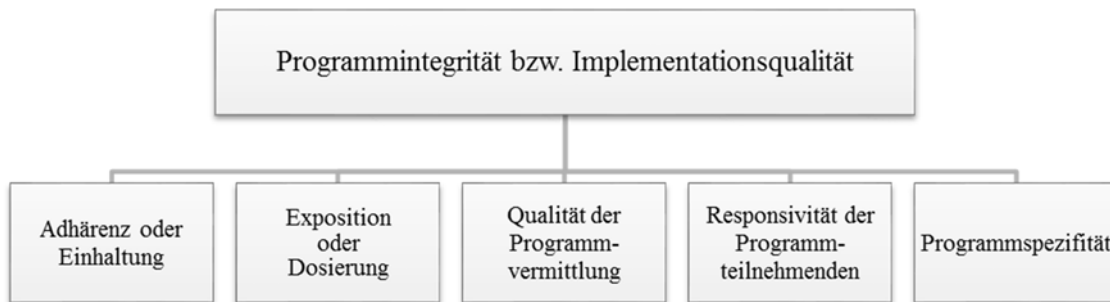


Abbildung 1. Aspekte der Programmintegrität bzw. Implementationsqualität. Aus Qualitätssicherung von Präventions- und Interventionskonzepten aus Sicht der Interventions-, Implementations- und Evaluationsforschung von W. Mittag (2014). *Wirtschaftspsychologie*, 16(2), 50–58.

Die *Adhärenz* oder *Einhaltung* (adherence) gibt an, in welchem Ausmaß ein Programm seiner Konzeption entsprechend durchgeführt wird. Es soll auch geprüft werden, inwieweit die Komponenten bzw. Bestandteile eines Programmes wie geplant umgesetzt werden. Die Einhaltung der vorgegebenen Interventionsmaßnahmen in ihrer Ganzheit entsprechend der Programmkonzeption ist ein Hauptelement der Implementationsqualität. „Fidelity may be defined to the extent to which delivery of an intervention adheres to the protocol or program model originally developed” (Mowbray, Holter, Teague & Bybee, 2003, S. 315). In vielen Studien schließt der Begriff der Einhaltung aber auch die Umsetzung des vollständigen Inhalts, die Häufigkeit der umgesetzten Maßnahmen, die Durchführungsdauer und den Durchführungsumfang (coverage; dose) ein. „Adherence includes the subcategories of content, frequency, duration and coverage (i.e., dose). The degree to which the intended content or frequency of an intervention is implemented is the degree of implementation fidelity achieved for that intervention” (Carroll et al., 2007, S. 3). Carroll et al. (2007) verweisen darauf, dass aber nicht generell die Frage nach vollständiger Einhaltung maßgeblich ist, sondern die Beachtung essentieller Faktoren ausreichend sein kann. „However, adherence may not require every single component of an intervention to be implemented. An intervention may also be implemented successfully, and meaningfully, if only the "essential" components of the model are implemented“ (Carroll et al., 2007, S. 5). Um die essentiellen Faktoren zu identifizieren, ist nach Auffassung der Autoren eine Komponentenanalyse erforderlich. Hierzu empfehlen sie die mehrmalige Interventionsdurchführung und entsprechende Evaluationen, um essentielle Wirkfaktoren bzw. die Kombinationen von wirksamen Interventionsbestandteilen zu eruieren. Ist dies aber nicht möglich, so ist eine Einhaltung aller Interventionsbestandteile not-

wendig (Carroll et al., 2007, S. 5). Soll die Adhärenz in Bezug auf die Implementationsgenauigkeit erfasst werden, so muss die geplante Intervention mit der durchgeführten Implementation verglichen werden. Wurde die Intervention wie von den Entwicklern geplant umgesetzt, so würde man von einer hohen Adhärenz und hohen Implementationsgenauigkeit sprechen (Carroll et al., 2007, S. 4). Ob eine hohe Einhaltung gewährleistet werden kann, scheint von weiteren auch die anderen Implementationsindikatoren betreffenden Faktoren abhängig zu sein. Dies sind z. B. moderierende Einflussfaktoren wie die Komplexität der Intervention und die Qualität der Programmvermittlung (s. Kapitel 1.4).

Die *Exposition* oder *Dosierung* (exposure oder auch dosage) bezieht sich in Abgrenzung zur Adhärenz ausschließlich auf den zeitlichen Umfang bzw. die Häufigkeit, mit der die Programminhalte tatsächlich umgesetzt werden. Die Erfassung erfolgt meist anhand der Anzahl der tatsächlich durchgeführten Sitzungen, der Länge der Sitzungen oder der Häufigkeit der verwendeten Programmmaßnahmen und -techniken. In ihrem Forschungsüberblick definieren Dusenbury et al. (2003): „(...), dose is defined as the amount of program content received by participants. Most studies that have measured dose have looked at the amount of the program covered (Pentz et al., 1990), often relying on teachers' self-reports; others might argue that dose can be calculated by extrapolating from the amount of coverage per session for a sample of sessions (Botvin, Baker, Dusenbury, Tortu & Botvin, 1990). However, we would argue that, ideally, a variety of factors should be included” (Dusenbury et al., 2003, S. 241). Ennett et al. (2011) operationalisieren die Exposition über die Häufigkeit der eingesetzten Maßnahmen, aber auch als Einhaltung der Vorgaben als Gesamtheit. Letztere wäre im Sinne der bisherigen Ausführungen eher der Adhärenz zuzuordnen. Ennett et al. (2011) ordnen die Einhaltung aber dem Begriff der Dosierung zu: „Exposure is typically indicated by the number of lessons taught but can reflect combinations of the number of lessons, amount of each lesson covered and adherence to the prescribed schedule” (Ennett et al., 2011, S. 362). Auch hier wird somit die bereits angesprochene Vermischung der Indikatoren deutlich.

Als ein weiterer Aspekt der Programmintegrität und Implementationsqualität kann die *Qualität der Programmvermittlung* (program delivery) angesehen werden. Es handelt sich dabei um subjektive Einschätzungen zur Programmvermittlung. Dazu können z. B. die Motivation und Einstellungen der Programmvermittler (z. B. Lehrkräfte) berücksichtigt werden, aber auch deren Einschätzung zur Wirksamkeit der einzelnen Programmsitzungen (Dane & Schneider, 1998, S. 45).

Die Qualität der Programmvermittlung nimmt nach Carroll et al. (2007) Einfluss auf die Implementationsqualität. Findet sich eine hohe Qualität der Programmvermittlung, so begünstigt dies eine vollständige Implementation der Interventionsmaßnahmen. „Quality of delivery is an obvious potential moderator of the relationship between an intervention and the fidelity with which it is implemented. It concerns whether an intervention is delivered in a way appropriate to achieving what was intended. If the content of an intervention is delivered badly, then this may affect the degree to which full implementation is realised“ (Carroll et al. 2007, S. 4). Die Qualität der Programmvermittlung steht auch in Zusammenhang zur getreuen Umsetzung der Intervention. So fanden auch Ennett et al. (2011), dass Trainingsdurchführende mit selbstsicherer Umsetzung und Ermutigung der Studenten die Maßnahmen auch vollständig umsetzen und von aktiver Teilnahme der Schüler berichten. „Notably, however, both adherence and exposure were significantly associated with quality of delivery. Providers who reported higher quality delivery-in that they were more confident of their ability to teach their evidence-based curriculum and were more encouraging of their students’ participation-were more likely to report adhering to prescribed content and delivery strategies as well as to implement the full curriculum. These providers also were more likely to report that their students actively participated in the curriculum. Provider engagement may be central to program fidelity“ (Ennett et al., 2011, S. 369). Darüber hinaus finden sich auch Zusammenhänge zwischen der Qualität der Programmvermittlung und den quantitativen Maßen der Implementationsqualität: „The various quantitative and qualitative components of implementation fidelity were found to be closely interrelated (...)“ (Griffin, Mahadeo, Weinstein & Botvin, 2006, S. 22). Neben den Zusammenhängen mit der Implementationsqualität generell und den quantitativen Aspekten der Implementationsqualität werden auch Zusammenhänge mit der Programmwirksamkeit berichtet (Griffin et al., 2006, S. 22).

Die *Responsivität der Programmteilnehmenden* (participant responsiveness) – als ein weiterer Aspekt bzw. Indikator der Programmintegrität und Implementationsqualität – umfasst die Bereitschaft und Empfänglichkeit für die Teilnahme am Programm als Ganzes oder an bestimmten Inhalten und Aktivitäten eines Programms. Operational wird dies u. a. durch Einschätzungen und Reaktionen der Programmteilnehmer selbst, aber auch durch Fremdeinschätzungen der Programmvermittler ermittelt. Hasson (2010) erweitert die genannte Begriffsdefinition, indem sie die Responsivität einerseits auf die Teilnehmer\_innen bezieht, andererseits



aber auch die Durchführenden einbezieht. Sie nimmt an, dass ein Trainer, der durch die Maßnahme erreicht wurde und zu deren Umsetzung motiviert ist, zu einer höheren Implementationsqualität beitragen kann. Ebenso tragen responsive Teilnehmer\_innen zu einer höheren Implementationsqualität bei: „Participant responsiveness refers both to individuals receiving the intervention and individuals responsible for delivering it. Higher levels of implementation fidelity are assumed to be achieved if those responsible for delivering an intervention are enthusiastic about it. Similarly, the uptake of the intervention depends on the responsiveness of those receiving it” (Hasson, 2010, S. 2). Anzunehmen ist, dass die Responsivität der Programmteilnehmer und die Befürwortung der Maßnahmen durch die Durchführenden als moderierende Faktoren für die Implementationsqualität angesehen werden können (Hasson, 2010). Für die Befürwortung der Durchführung ist jedoch auch ein Zusammenhang zu der Qualität der Programmvermittlung anzunehmen und somit ist auch ein Einfluss auf die Responsivität der Teilnehmer zu erwarten. Dane und Schneider (1998, S. 44 f.) ordnen die oben aufgeführten Merkmale (Anzahl der tatsächlich durchgeführten Sitzungen, Länge der Sitzungen oder die Häufigkeit der verwendeten Programmmaßnahmen) allein der Exposition (exposure) zu. Die Dosierung (dosage) wird bei Dane und Schneider (1998) als gewissermaßen eigenständiger Aspekt operational als Beteiligungsgrad der Programmteilnehmenden an den Programmsitzungen definiert. Diese Zuordnung weist allerdings eine starke inhaltliche Nähe zur Responsivität der Teilnehmer auf. Während bei der Adhärenz schon Vermischungen mit der Dosierung zu finden sind, zeigt sich dies bei der Dosierung in noch stärkerem Maße mit der Responsivität der Teilnehmer (Einordnung von Dane & Schneider, 1998, S. 44 f.). Diese Vermischung ist insoweit problematisch, da zwischen dem Beteiligungsgrad der Programmteilnehmenden an den Programmsitzungen und der Responsivität der Programmteilnehmer ein positiver Zusammenhang anzunehmen ist.

Unter *Programmspezifität* (program differentiation) wird schließlich verstanden, ob und inwieweit sich die durchgeführten programmspezifischen Maßnahmen und Inhalte von konkurrierenden, aber programmfremden Maßnahmen und Inhalten abgrenzen lassen. Durch diese Unterscheidung soll sichergestellt werden, dass wirklich nur die spezifischen Maßnahmen und Inhalte des in Frage stehenden Programms die beobachteten Veränderungen bewirkt haben (Mittag et al., 2010, S. 41 f.). Daher muss geprüft werden, welche programmähnlichen und konkurrierenden Inhalte in der Kontrollgruppe zu beobachten sind (Mittag, 2014). So soll eine Vermischung der Inhalte von Interventions- und Kontrollgruppe sowie eine Unterschätzung der Interventionseffekte vermieden werden.

Über die bislang diskutierten fünf Indikatoren zur Implementationsqualität hinaus werden in der Literatur drei weitere Aspekte diskutiert: die Überwachung von Kontroll- und ggf. Vergleichsgruppen, die Erfassung der Programmreichweite und die Anpassung der Programminhalte. Dane und Schneider (1998) fokussieren in ihren Ausführungen zur Programmspezifität auf die Betrachtung der Treatmentgruppe, über die sie eine Abgrenzung zu anderen Maßnahmen anstreben. Die Überwachung (monitoring) von Kontroll- und ggf. Vergleichsgruppen schlagen Durlak und DuPre (2008) zusätzlich vor. Hier liegt der Hauptfokus auf Art und Umfang von Programmbestandteilen in den Kontroll- und Vergleichsgruppen (Mittag et al., 2010, S. 41 f.; Mittag et al., 2000, S. 117 f.). Die Programmreichweite (program reach) wird von Durlak und DuPre (2008) als weiterer Indikator für die Implementationsqualität genannt. Diese definieren sie über den Grad der Beteiligung bzw. die Teilnahmequote innerhalb der Zielgruppe. Durch Beachtung der Programmreichweite können wie bereits angesprochen Aussagen über mögliche Selektionsprozesse, Ausfälle und damit einhergehende Einschränkungen der Repräsentativität getroffen werden (Mittag, 2014, S. 54; s. auch Durlak & DuPre, 2008). Die Anpassung der Programminhalte (adaption) an die realen Bedingungen wird als weiterer Indikator in Abgrenzung zum Aspekt der Adhärenz, also der Einhaltung der Programmmaßnahmen entsprechend den Vorgaben von Durlak und DuPre (2008), aufgenommen. Mit diesem Indikator werden nicht nur Abweichungen erfasst, sondern beispielsweise auch Abänderungen, Ausweitungen oder Kürzungen von Interventionsbestandteilen (Mittag, 2014, S. 54). Durch diese differenziertere Betrachtung von konkreten Veränderungen können auch Erkenntnisse über sinnvolle Abweichungen und notwendige Veränderungen der Programmbestandteile gewonnen werden. Der Aspekt der Programmanpassung deutet zudem darauf hin, dass eine absolute Einhaltung der Programmvorgaben in der Praxis kaum möglich erscheint und oft Anpassungen an die realen Bedingungen vorzunehmen sind (McGrew, Bond, Dietzen & Salyers, 1994; Gräsel & Parchmann, 2004). Entsprechend verweisen auch Dusenbury et al. (2003) auf die Notwendigkeit von Kenntnissen zur Programmanpassung. Das Ziel ist, Elemente des Programms an die realen Bedingungen zu adaptieren, ohne damit deren Wirksamkeit zu reduzieren, zugleich aber eine erhöhte Bereitschaft zur Programmumsetzung zu erzielen (s. *Mutual-Adaption-Ansatz*; Dusenbury et al., 2003; McGrew et al., 1994; Snyder et al., 1992). Bezüglich der Identifikation des Einflusses verweisen Dusenbury et al. (2003) auf die Komponentenanalyse: „We define program differentiation as identifying unique features of different components or programs so that these components or programs can be reliably differentiated from one another. (...) In fact, the greatest value of measures of program

differentiation may be their contribution to components analyses which could be used to determine the essential elements of effective prevention strategies” (Dusenbury et al., 2003, S. 244 f.). Dadurch kann bei nicht nachweisbarer Gesamteffektivität eines Programms die Wirksamkeit einzelner Komponenten geprüft werden.

Generell lässt sich im Hinblick auf die diskutierten Indikatoren der Implementationsqualität annehmen, dass keine Unabhängigkeit zwischen diesen besteht. In den bislang vorliegenden Studien wurden allerdings meist nicht alle Indikatoren berücksichtigt, sodass keine oder nur eingeschränkte Belege über derartige Zusammenhänge vorliegen. In einer Evaluationsstudie von Botvin, Dusenbury, Baker, James-Ortiz und Kerner (1989) konnten beispielsweise Beziehungen zwischen der Adhärenz und der Qualität der Programmvermittlung gefunden werden. Durch Untersuchungen verschiedener Modelle unter Einbezug der Implementationsindikatoren als Mediatoren oder Moderatoren (Interaktionseffekte) könnten die Zusammenhänge möglicherweise näher beleuchtet und einzelne Indikatoren zusammengefasst werden (Berkel et al., 2011; Carroll et al., 2007). Ob eine hohe Implementationsqualität gewährleistet werden kann, scheint auch von weiteren Einflussfaktoren abhängig zu sein (s. Kapitel 1.4). Zukünftig sollten daher alle oder möglichst viele Indikatoren berücksichtigt werden, um damit einerseits die Prüfung der Implementationsqualität zu gewährleisten und andererseits Aussagen über mögliche Zusammenhänge machen zu können.

## **1.2 Methoden zur Erfassung der Indikatoren der Implementationsqualität**

In den bisherigen Ausführungen wurden Indikatoren der Implementationsqualität betrachtet. Im Folgenden soll deren genauere Operationalisierung vorgestellt werden, wobei deren Prüfung anhand von einzelnen Studien erfolgt.

Im Fokus stehen dabei vor allem die fünf von Dane und Schneider (1998) vorgeschlagenen Indikatoren der Implementationsqualität (s. Abschnitt 1.1). Neben den verschiedenen Erhebungsverfahren werden die in den Studien verwendeten Skalierungen und Datenformate zur quantitativen oder auch qualitativen Erfassung der Implementationsqualität erörtert, weil die Art der Skalierung und das Datenformat die Operationalisierung der (Qualitäts-) Indikatoren mitbestimmen.

Bei den Erhebungsverfahren kann zunächst zwischen Befragungs- und Beobachtungsverfahren differenziert werden. Weiterhin kann bei den Befragungsverfahren danach unterschieden werden, ob schriftliche Befragungen mittels Fragebogen zum Einsatz kommen oder mündliche Befragungen in Form von Interviews oder beide Varianten. Bedeutsam erscheint zudem, wer (schriftlich oder mündlich) befragt wurde. So gibt es neben Studien, in denen Lehrkräfte als Programmvermittler/Programmdurchführende mittels Fragebogen befragt wurden (Ennett et al., 2011; (Goldberg) Lillehoj, Griffin & Spoth, 2004; Pentz et al., 1990) auch solche, in denen Interviews mit den Programmdurchführenden und den Programmteilnehmenden durch externe Personen durchgeführt wurden. Mittels Fragebogen werden aber auch Daten bei den Programmteilnehmenden erhoben (Hasson, 2010).

Einzelne Indikatoren wie beispielsweise die Adhärenz (Goldberg) Lillehoj et al., 2004; Hasson, 2010; Resnicow et al., 1998; Rohrbach, Dent, Skara Sun & Sussman, 2007; Spoth, Gyll, Lillehoj & Greenberg, 2007; Wilson et al., 2009) oder die Dosierung (Griffin et al., 2006; Songer & Gotwals, 2005; Wilson et al., 2009) werden auch über Beobachtungsverfahren erfasst. Dabei kann danach unterschieden werden, ob Fremdbeobachtungen durch externe Beobachter durchgeführt wurden (z. B. Rohrbach et al., 2007; Mihalic, Fagan & Argamaso, 2008) oder die Ergebnisse von Selbstbeobachtungen seitens der Programmdurchführenden in Form von Selbstberichten oder Logbüchern erfasst wurden (Basen-Enquist et al., 1994; Drössler, Jerusalem & Mittag, 2007; Mihalic et al., 2008; Pentz et al., 1990; Resnicow, et al., 1998; Röder & Jerusalem, 2007; Wilson et al., 2009). Allerdings sind die eingesetzten Logbücher mitunter in Teilen schon soweit standardisiert, dass eine klare und durchgehende Abgrenzung von der schriftlichen Befragung mittels Fragebogen kaum aufrecht zu erhalten ist (Drössler et al., 2007).

Betrachtet man die in den Studien verwendeten Skalierungen und Datenformate, so finden sich dort neben Häufigkeitsangaben (aus Fremd- und Selbstbeobachtungen) auch item-bezogene Einschätzungen auf Rating- und/oder Likert-Skalen. Auf Ratingskalen können beispielsweise Einschätzungen zur Qualität der Programmvermittlung (program delivery) erfasst werden. Dies wäre z. B. dann der Fall, wenn die Frage „Wie gut ist Ihrer Meinung nach die Durchführung gelungen?“ auf einer mehrstufigen Antwortskala von *gar nicht* (0 oder 1) bis *sehr gut* (3, 5 oder 7) zu beantworten ist. Als Alternative könnte auch nach der Zufriedenheit mit der Durchführung gefragt werden oder es könnten Häufigkeitseinschätzungen zum Einsatz bestimmter Methoden oder Inhalte erhoben werden. Solche Einschätzungen, Beurteilungen,

Einstufungen usw. werden dann meist zu Kennwerten für die verschiedenen (Qualitäts-) Indikatoren verrechnet und bilden die Grundlage zu deren Operationalisierung, auf die im Folgenden bezogen auf die einzelnen Indikatoren genauer eingegangen wird.

### 1.2.1 Adhärenz

Trotz eines relativ breiten Konsens, wie der Begriff der Adhärenz zu definieren ist, finden sich in den empirischen Studien erhebliche Unterschiede bei den jeweiligen Operationalisierungen und Auswertungsstrategien.

Einige Studien gehen bei der Erfassung der Adhärenz von der **Einhaltung der Programmvorgaben** auf der Grundlage des Programmmanuals aus und berücksichtigen bei der Bildung von Adhärenz-Kennwerten (z. B. Indexwerte, Summenwerte, Ist-Soll-Diskrepanzen), ob und in welchem Umfang **Abweichungen bzw. Kürzungen** von den Programmvorgaben zu verzeichnen sind.

So haben beispielsweise Rohrbach et al. (2007) in einer Interventionsstudie zur Prävention des Substanzmissbrauchs bei 13-19-Jährigen Schüler\_innen (N = 2.331) aus 18 zufällig ausgewählten Schulen die Implementationsgenauigkeit mittels **Fremdbeobachtungen** der Unterrichtseinheiten durch Projektmitarbeiter einschätzen lassen. Der **Adhärenz-Index** wurde aus den Einschätzungen zu drei Items ermittelt, die sich darauf bezogen, ob (1) die Unterrichtseinheit gekürzt wurde, ob (2) die einführenden Programmaktivitäten und (3) die zentralen Programmaktivitäten (Kernelemente) eingehalten wurden. Kritisch ist an dieser Operationalisierung der Adhärenz anzumerken, dass der z-standardisierte Mittelwert von 0 nicht die tatsächliche Programmeinhaltung (Adhärenz) wiedergibt, sondern lediglich das über alle Interventionsklassen und Unterrichtseinheiten gemittelte Ausmaß. Ein Adhärenz-Index auf der Grundlage von Ist-Soll-Diskrepanzen wäre hier anschaulicher.

Die Implementationsgenauigkeit schulbezogener Programme zur Prävention des Substanzmissbrauchs wurde auch von Ennett et al. (2011) untersucht. Die teilnehmenden Lehrer unterrichteten eins von zehn evidenzbasierten Präventionsprogrammen zum Substanzenmissbrauch. Die Ergebnisanalysen basieren auf einer Teilstichprobe von 342 teilnehmenden Schulen aus einer größeren Zufallsstichprobe von 1.721 Mittelschulen (Sekundarschulen). Zur **Erfassung der Adhärenz** wurden zwei Merkmalsbereiche einbezogen, und zwar (1) die

Implementation von Programminhalten und (2) die Anwendung interaktiver Vermittlungsstrategien. Die Implementation der Programminhalte bezog sich dabei u.a. auf Vermittlung von Informationen, Standfestigkeitsfertigkeiten sowie personalen und sozialen Kompetenzen. Dies wurde über insgesamt zwölf Items durch **Befragung** der Programmvermittler an den Schulen erhoben (Ennett et al., 2011, S. 365 f.). Die Anwendung interaktiver Vermittlungsstrategien wurde über insgesamt acht Items erfasst. Aus den Items der jeweiligen Skalen bzw. Merkmalsbereiche wurden **Adhärenz-Kennwerte** gebildet, die den gemittelten prozentualen Anteil der vorgegebenen Inhalte und Strategien widerspiegeln sollen (Ennett et al., 2011, S. 368). Es handelt sich dabei also um einen (prozentualen) **relativen Adhärenz-Index**, der über ein Selbstberichtsverfahren erhoben wurde.

Gemittelte relative Indizes zur Prüfung der Implementationsqualität wurden auch in einer Studie zur Prozessevaluation eines schulbezogenen Interventionsprogramms zur Förderung körperlicher Aktivität verwendet (Wilson et al., 2009). In das Interventionskonzept wurden Elemente der Selbstbestimmungstheorie (Deci & Ryan, 2000) und der sozial-kognitiven Lerntheorie (Bandura, 1986) einbezogen, um die intrinsische Motivation und die verhaltensbezogenen Kompetenzen bei Jugendlichen für langfristige körperliche Aktivitäten zu verbessern. Untersucht werden sollte dabei auch, inwieweit sich durch Maßnahmen der formativen Prozessevaluation die Adhärenz, die Dosierung und die Programmreichweite verbessern lassen. Die Autoren verwenden nicht den Begriff der Adhärenz, sondern ausschließlich den Begriff *fidelity*, und nehmen eine begriffliche und operationale Unterscheidung zur Dosierung vor. Allerdings erscheint die in dieser Studie zugrunde gelegte begriffliche und operationale Unterscheidung zwischen Adhärenz und Dosierung als nicht trennscharf und daher als unzureichend. So wurde die Adhärenz über Ratingskalen erfasst, auf denen einzuschätzen war, wie häufig grundlegende Programmelemente wie z. B. die Förderung von Autonomie, Partizipation und sozialer Eingebundenheit, aber auch Maßnahmen zur Förderung verhaltensbezogener Kompetenzen zu beobachten waren. Die Häufigkeitseinschätzungen erfolgten auf vierstufigen Ratingskalen, die von *none* (1; nie) über *some* (2; manchmal) und *most* (3; meistens) bis *all* (4; immer) reichten (Wilson et al., 2009, s. *Tabelle 2*). Demgegenüber wurde die von Wilson et al. (2009) bezeichnete Dosierung über Checklisten erfasst, in denen grundlegende Programmelemente wie z. B. die Vor- und Durchführung körperlicher Aktivitätsübungen oder die Erklärung und Demonstration bestimmter Fertigkeiten aufgeführt waren. Bei jedem Program-

melement war anzugeben, ob es durchgeführt (*yes*) oder nicht durchgeführt (*no*) wurde (Wilson et al, 2009, s. *Tabelle 3*). Die Ausführungen machen deutlich, dass in beiden Fällen Häufigkeiten von durchgeführten Programmmaßnahmen erfasst wurden. Die beiden Messungen unterscheiden sich lediglich in formaler Hinsicht bezogen auf Itemstamm und Antwortformat (s. dazu Eid & Schmidt, 2014, S. 79 ff.). Konzeptuell erfassen beide Messungen hingegen die Vollständigkeit der Programmdurchführung (Wilson et al, 2009) und damit die **Adhärenz**. Im ersten Fall erfolgte dies über vierstufige Ratingskalen, für die in den abschließenden Analysen **Mittelwerte** berechnet wurden. Im zweiten Fall wurden abschließend (prozentuale) **relative Adhärenz-Kennwerte** ermittelt. Die Adhärenz-Kennwerte wurden in der Studie von Wilson et al. (2009) **durch trainierte Beobachter/Evaluatoren** erhoben.

Weitere Studien gehen bei der Erfassung der Adhärenz von der **Einhaltung der Programmvorgaben** auf der Grundlage des Programmmanuals aus und betrachten relative Adhärenzkennwerte auf Grundlage **der spezifischen vorgegebenen Programminhalte und Aktivitäten** ((Goldberg) Lillehoj et al., 2004; Mihalic et al., 2008).

Die Studie von (Goldberg) Lillehoj et al. (2004) untersuchte die Einflüsse der Einhaltung der Interventionsvorgaben (implementation adherence) auf die Wirksamkeit zweier Präventionsprogramme bei Jugendlichen (Life Skills Training (LST); Strengthening Families Program (SFP:10-14)), wobei der Fokus auf der Art der Erhebungsmethode lag. Es wurden 36 zufällig ausgewählte ländliche Mittelschulen in Bezug auf die längsschnittlich angelegten Interventionen betrachtet. Am Prätest nahmen 1.664 Schüler\_innen im Alter von durchschnittlich 12,3 Jahren teil. Die operationale Definition der **Adhärenz** bezog sich auf den **relativen Anteil der vermittelten spezifischen Programminhalte und Programmaktivitäten** durch die Klassenlehrer. Hierzu wurden spezifische Programmkomponenten identifiziert, die im Hinblick auf die Adhärenz der Intervention erfasst werden sollten. Zur Erfassung wurde für jede Stunde ein *Prozessevaluationsformular* erstellt. Damit wurde festgehalten, ob die Inhalte wie beabsichtigt implementiert wurden. Die Lehrer sollten so im Selbstbericht Angaben zur Implementation in der Unterrichtsstunde machen. Es handelte sich also um Angaben zur Abdeckung der Stunde mit den geplanten spezifischen Programminhalten und Programmaktivitäten (lesson coverage). Zusätzlich wurden die Lehrer gebeten, eine Einschätzung zur Umsetzung inhaltspezifischer Ziele und Aktivitäten vorzunehmen und eine *Ja/Nein*-Auskunft abzugeben, ob diese in der Intervention vorgesehenen Bestandteile in der Stunde abgedeckt wurden.

Kritisch ist an dieser Operationalisierung der Adhärenz, dass hier ausschließlich gemessen wurde, ob die jeweiligen Interventionsinhalte umgesetzt wurden. Unklar bleibt das Ausmaß der tatsächlichen Abweichung. Im Grunde müsste man hier eine Zuordnung zur Dosierung/Exposition vornehmen, da es sich nur um eine Häufigkeitsangabe von Interventionsbestandteilen handelt. Ein Adhärenz-Index auf der Grundlage von Ist-Soll-Diskrepanzen hätte hier ein genaueres Bild der Abweichung von der intendierten Umsetzung der Interventionsinhalte ermöglicht. Im Weiteren sollte in dieser Studie angegeben werden, wie viel Zeit in der Stunde zur Umsetzung der Intervention aufgewendet wurde. Auch hier handelt es sich eher um die Dosierung von Programmbestandteilen bzw. die Exposition in Bezug auf die Teilnehmer als um Adhärenz. Sowohl die Erfassung der aufgewendeten Zeit pro Stunde für Interventionsbestandteile als auch die Erfassung der Responsivität wurden jedoch nicht in die Messung der Implementationsgenauigkeit einbezogen. Darüber hinaus schätzten auch unabhängige Beobachter die Stunde mittels der gleichen *Fragebögen* ein (s. o.). Für jede Unterrichtsstunde wurde so für Lehrkräfte und Beobachter ein Wert basierend auf dem Anteil an den umgesetzten spezifischen Vorgaben, der Zielerreichung und der abgedeckten Aktivitäten berechnet. Zusätzlich wurde ein Durchschnittswert separat für Beobachter und Lehrer berechnet. Dieser Wert bildet den Anteil der spezifischen vermittelten Interventionsinhalte in der jeweiligen Stunde ab. Unter Einbezug beider Datenquellen wurde dann ein Durchschnittswert der Adhärenz berechnet. Tatsächlich handelt es sich hier nur um ein durchschnittliches Maß (durchschnittliche Häufigkeit) für die Dosierung. Zusätzlich wurden die Zusammenhänge zwischen Selbstbericht und Beobachtung betrachtet und die Einschätzungen zur Implementationsgenauigkeit mit den Schülermaßen in Beziehung gesetzt. Bei den Erhebungsmethoden zeigten sich Unterschiede zwischen dem Selbstbericht der Lehrkräfte und unabhängigen Beobachtern. So berichten unabhängige Beobachter weniger Implementationsaktivitäten als die Lehrpersonen.

Mihalic et al. (2008) führten eine Prozessevaluation eines dreijährigen Life Skills Training (LST) durch. Das Ziel der Studie war, die Implementierung der Trainingsmaßnahmen an 423 Schulen in 105 Ortschaften mittels qualitativer und quantitativer Methoden zu überprüfen. Hierbei sollten die Zusammenhänge verschiedener Faktoren mit der Implementationsgenauigkeit untersucht werden. Insgesamt wurden vier Dimensionen der Implementationsgenauigkeit erfasst. Dabei handelt es sich um die Adhärenz, Dosierung, Qualität der Programmvermittlung und die Responsivität der Schüler\_innen. Über **Beobachtungen** im Klassenzimmer



(1-2 schulfremde Beobachter) wurde ein Implementationswert für die Adhärenz zur Vermittlung der Trainingsinhalte durch die Lehrer ermittelt. Mittels einer **Checkliste** wurde dieser *Implementationswert für jede Stunde als Prozentsatz des unterrichteten Materials am vorgesehenen Material der Stunde erfasst* (z. B. fünf von zehn der vorgegebenen Zielvorgaben entsprechen beispielsweise einem Implementationswert von fünfzig Prozent). Es handelt sich hiermit um einen **prozentualen relativen Adhärenz-Index**. Zusätzlich erfassten die Beobachter verschiedene Instruktionstechniken im Unterricht, die Teilnahme der Schüler\_innen und auftretende Schwierigkeiten. Für jeden Interventionsort wurde basierend auf der Anzahl aller Lehrer und Umsetzungsjahre ein durchschnittlicher Implementationswert berechnet. Die Ergebnisse zeigen, dass über alle Orte hinweg 86 Prozent der beabsichtigten *Programminhalte* umgesetzt wurden. Alle geforderten *Lehrmethoden* wurden eingesetzt und 71 Prozent der Lehrer unterrichteten die Trainingsstunden gemäß den Vorgaben. Die Analysen des Zusammenhangs der Programmbewertung mit der Implementationsgenauigkeit erfolgten mittels multivariater Analysen. Es zeigte sich ein signifikanter Zusammenhang zwischen der positiven Bewertung des Programms durch die Schüler\_innen und Lehrkräfte mit einem größeren prozentualen Anteil an umgesetzten vorgesehenen Material (Adhärenz).

### 1.2.2 Dosierung

Der Begriff **Dosierung** wird im Folgenden, anders als bei Dane und Schneider (1998), synonym mit dem Begriff der Exposition verwendet. Die Dosierung wird in den betrachteten Studien mit Hilfe verschiedener, zum Teil kombinierter Erhebungsmethoden erfasst. Einen Zugang bilden Selbstberichte der Programmdurchführenden, die durch Checklisten erhoben werden. Checklisten erfassen über verschiedene Skalierungen eine Einschätzung zur Umsetzung oder erfragen die Dosierung z. B. über *Ja/Nein*-Einschätzungen. Eine weitere Variante ist die Einschätzung der Intensität über eine Likert-Skalierung ((Goldberg) Lillehoj et al., 2004). Zusätzlich kommen Logbücher (Drössler et al., 2007) oder Fragebögen zum Einsatz. Wird die Dosierung von unabhängigen Beobachtern eingeschätzt, werden zur Dokumentation der Beobachtung Protokolle, Checklisten bzw. Ratingverfahren genutzt. Die Verwendung gleicher Instrumente zur Selbst- und Fremdeinschätzung gewährleistet eine Vergleichbarkeit der Angaben. Exemplarisch sollen für die Erfassung der Dosierung im Folgenden die Studien von Drössler et al. (2007) und Röder und Jerusalem (2007) vorgestellt werden. Diese Studien betrachten die Häufigkeit der eingesetzten Maßnahmen als Dosierung der Programmbestandteile. Eine ähnliche Vorgehensweise findet sich auch bei Ennett et al. (2011).

Drössler et al. (2007) führten eine indirekte Intervention zur Förderung sozialer Kompetenzen im Unterricht durch. Die Stichprobe umfasste 421 Schüler\_innen der Interventionsgruppe, 353 Schüler\_innen aus Kontrollklassen und 97 teilnehmende LehrerInnen. Die Lehrkräfte der Interventionsklassen wurden im Rahmen einer Fortbildung in der Umsetzung kooperativer Lernstrategien geschult. Im Anschluss an die Fortbildung wurden die Implementationsaktivitäten der Lehrkräfte über Lehrerfragebögen erfasst. Zum zweiten Messzeitpunkt wurden die Lehrer gebeten, Veränderungen in ihrem Unterrichtsverhalten einzuschätzen. Dabei war unter anderem die **Häufigkeit** von kooperativen Lernmethoden zur Förderung sozialer Kompetenzen auf zwei Items und einer vierstufigen Skala von *nie* (1) bis *zu jeder Gelegenheit* (4) anzugeben. Das erste Item bezog sich auf die Häufigkeit des Einsatzes im Schuljahr vor dem Workshop, das zweite auf die Häufigkeit im Schuljahr nach dem Workshop. Es handelt sich also um eine retrospektive Einschätzung der Häufigkeit. Eine Erfassung der tatsächlichen Anzahl an eingesetzten Methoden nach jeder Unterrichtsstunde (bzw. jeder Einheit) hätte ein valideres Maß bereitgestellt. Auch würde eine zeitnahe Erfassung genauere Angaben zur Dosierung ermöglichen. Auf Basis der vorgenommenen Erfassung wurden zwischen den Erhebungszeitpunkten Differenzwerte für die Dosierung der kooperativen Lernformen erstellt. Diese Differenzwerte wurden schulweise aggregiert. Mittels Medianhalbierung wurden dann zwei Gruppen gebildet. Im Ergebnis konnten vier Schulen einer Gruppe mit geringen Zuwächsen in der Verwendung von kooperativen Lernformen zugewiesen werden und vier andere Schulen einer Gruppe mit starken Zuwächsen. Diese Gruppen wurden schließlich mittels Kovarianzanalysen im Hinblick auf die Wirksamkeit der Intervention verglichen.

Bei Röder und Jerusalem (2007) findet sich eine ähnliche, aber im Hinblick auf die Messung der Dosierung geeignetere Herangehensweise. Hier wurde untersucht, inwieweit die Wirkung der Intervention mit der Intensität der Programmumsetzung variiert. Es wurden ebenfalls im Rahmen eines Workshops Inhalte an die unterrichtenden Lehrer vermittelt. Vier Schulen mit zwölf Interventionsklassen nahmen daran teil. In Workshops zum Thema „Motiviertes Lernen“ wurden die Interventionsinhalte an die Lehrkräfte vermittelt. Untersucht wurde die Implementation einer Intervention zur Förderung der Selbstwirksamkeit von Schüler\_innen. Über ein Schuljahr hinweg wurden die Interventionsklassen der 7. und 8. Klassenstufe betrachtet. Zur Förderung des Erlebens von Selbstwirksamkeit sollten von den Lehrkräften Transparenzpapiere im Unterricht eingesetzt werden. Bei Transparenzpapieren handelt es sich um Informationsblätter für Schüler\_innen, die Aufgaben und Erwartungen in Bezug auf die

Leistungsanforderungen und Leistungsbewertungen näher erläutern und so zur Vorbereitung auf Klassenarbeiten dienen können. Die Anzahl der Transparenzpapiere diene als ein Indikator der Dosierung. Die Erfassung erfolgte über die Dokumentation in einem Logbuch. Hier sollten die Lehrer einmal im Schuljahr über die Umsetzung der Inhalte im Unterricht berichten. Erfasst wurde die Häufigkeit in jeder Klasse über Fragebögen mit geschlossenen und offenen Fragen. Zwei unabhängige Rater nahmen zudem eine Beurteilung der Angaben in den Logbüchern vor. Die Daten wurden zu einem Summenwert pro Klasse über das Schuljahr hinweg aggregiert. Positiv anzumerken ist, dass die Studie von Röder und Jerusalem (2007) die absolute Häufigkeit der eingesetzten Methoden erfasst. Validiert wurde dieses Maß durch die Zusendung der verwendeten Transparenzpapiere, was die Nachteile der retrospektiven Erhebung allerdings nur bedingt ausgleicht, da nicht von allen Lehrkräften Transparenzpapiere zurückgesendet wurden. Auch hier wäre eine direkte Erfassung der verwendeten Materialien nach jeder Stunde anzudenken gewesen. Zusätzlich wurde zur Sicherung der Validität der Logbuchangaben ein Abgleich mit den zugesendeten Transparenzpapieren vorgenommen. Abschließend erfolgten mehrebenanalytische Berechnungen zur Wirksamkeit auf die abhängigen Variablen. Auf Ebene 1 fanden sich die Messzeitpunkte (Level 1), die Schüler\_innen auf Ebene 2 (Level 2) und diese geschachtelt in Klassen auf Ebene 3 (Level 3). Als erklärende Variable wurde die *Anzahl der Transparenzpapiere* auf Klassenebene (Summenwert Klasse) einbezogen.

Auch die bereits dargestellte Studie von Ennett et al. (2011) mit einer Stichprobe von 342 teilnehmenden Schulen hat den Aspekt der Dosierung berücksichtigt (s.o.). Die Dosierung wurde **(a) über die Einordnung bei einem Item** (1 *alle Stunden wie gefordert unterrichtet* (1) bzw. *nicht* (0)) und **(b) die Häufigkeit der Stunden** (*keine* (1) bis *16 oder mehr* (17)) erfasst. Zusätzlich wurde eine Checkliste eingesetzt, die den Namen und die Beschreibung der eingesetzten Stunde festhielt. Falls verfügbar, wurden diese Checklisten zur Validitätsprüfung bei der Auszählung der tatsächlich durchgeführten Stunden hinzugezogen.

Eine häufig in der Literatur zur Implementationsforschung berichtete Studie von Pentz et al. (1990) soll hier ebenfalls in Bezug auf die Dosierung vorgestellt werden. Diese Studie wurde im Rahmen des längsschnittlich angelegten Midwestern Prevention Project (MPP), einem gemeindebezogenen Programm zur Prävention von Drogenmissbrauch bei Jugendlichen, durchgeführt. Untersucht wurde der Zusammenhang zwischen dem Ausmaß der Programm-

implementation und den Veränderungen im Drogenkonsum von Jugendlichen in fünfzig Mittel- und Junior-High-Schulen. Die Interventionsinhalte wurden durch trainierte Lehrkräfte an die Schüler\_innen vermittelt. Die Implementation wurde mittels Selbstberichtsverfahren bei den Lehrern erfasst. Zur Validierung der **Lehrerselbstberichte** wurden zudem **Beobachtungen** vorgenommen und zusätzliche Berichte durch die wissenschaftliche Betreuung erstellt, deren Einschätzung von wissenschaftlicher Seite aufgrund von Telefongesprächen und Treffen mit Lehrkräften und Schulleitern vorgenommen wurde. Der Drogenkonsum wurde durch Selbstberichte seitens der Schüler\_innen und zusätzliche Atemluftmessungen erfasst. Beim Selbstbericht sollten die Lehrkräfte zwei Klassen auswählen, die sie als durchschnittlich im Hinblick auf Teilnahme am Programm, Störungen und Leistung ansahen. Die Items bezogen sich auf Einschätzungen zur Teilnahme und zu Störungen in den Klassen sowie auf subjektive Einschätzungen der Implementation. Hierzu gehörten Items, die sich auf die Dosierung beziehen. Die durchschnittliche Dauer einer Einheit betrug 45 Minuten. Um einen Wert für die Dosierung zu erhalten, wurden die Angaben zur Anzahl der Einheiten mit der durchschnittlichen Dauer multipliziert und durch 60 dividiert. In dieser Studie wird die Dosierung in ähnlicher Weise wie die Adhärenz definiert und beide werden dementsprechend absichtlich ähnlich betrachtet.

Generell ist für diese Indikatoren ein korrelativer Zusammenhang anzunehmen. Das Vorgehen von Pentz et al., (1990) hat allerdings zur Folge, dass die Indikatoren noch weniger voneinander abgegrenzt werden können. Zudem wird die notwendige Erfassung der Adhärenz über die Bildung eines Abweichungsmaßes im Sinne eines Ist-Soll-Vergleichs der eingesetzten Maßnahmen nicht vorgenommen.

### **1.2.3 Qualität der Programmvermittlung**

Zur Erfassung der **Qualität der Programmvermittlung** (*program delivery*) wird nach Dane und Schneider (1998) die Motivation und Einstellung der Programmvermittler (z. B. Lehrkräfte) berücksichtigt, aber auch deren Einschätzungen zur Wirksamkeit der einzelnen Programmsitzungen (Dane & Schneider, 1998, S. 45). In den betrachteten Studien findet sich ein sehr heterogenes Bild an operationalen Definitionen zur Qualität der Programmvermittlung. Häufig findet sich in der Literatur der Blick auf Eigenschaften der Programmvermittler (Spoth, Gyll, Lillehoj & Greenberg, 2007; Rohrbach et al., 2007; Griffin et al., 2006; Rohrbach, Gunning, Sun & Sussman, 2010; Ennett et al., 2011) oder es werden qualitative

Aspekte der Programmkonzeption einbezogen (O'Brien, 2005; Resnick et al., 2005; Carroll et al., 2007; Elliot & Mihalic, 2004; Hitt et al., 2006; Penuel & Means, 2004; Hansen, 1996; Hasson, 2010). Wieder andere Operationalisierungen weisen eine inhaltliche Nähe zur Adhärenz (Rohrbach et al., 2007, Rohrbach et al. 2010; Connell, Turner & Mason, 1985; McGraw et al., 2000) oder zur Dosierung (Mihalic et al., 2008) auf.

Griffin et al. (2006) untersuchten an 11 Mittelschulen (Lehrkräfte N = 38; Schüler\_innen N = 1.857) den Zusammenhang zwischen der Implementationsqualität bei einem Drogenpräventionstraining und dem Drogenkonsum nach der Maßnahme. Trainierte Beobachter schätzten mittels eines Beobachtungsbogens die Umsetzung der Implementation durch die Lehrkräfte im Hinblick auf qualitative und quantitative Implementationsindikatoren ein. Bei Mehrfachmessungen wurde ein Durchschnittswert gebildet. Die **Qualität der Programmvermittlung** wurde über (1) die Implementationsquantität als prozentualer Anteil umgesetzter Programmziele oder Programmpunkte (pro Unterrichtsstunde) erfasst. Im Weiteren machten die Beobachter Angaben zur (2) Qualität der Implementation. Die Qualität der Implementation umfasste a) den prozentualen Anteil an Zeit für die umgesetzten Lehrmethoden und Angaben zu Items, die (b) qualitative Merkmale der Lehrkraft erfassen (positive Einstellung der Lehrkraft gegenüber der Intervention, gute Unterrichtsvorbereitung und angemessene Interaktion mit den Schüler\_innen). Griffin et al. (2006) ordnen der Qualität der Programmvermittlung auch quantitative Merkmale zu, die wir eher der Dosierung zuordnen würden. Die qualitativen Merkmale umfassen aus unserer Sicht aufgrund der zeitlichen Erfassung hier ebenfalls die Dosierung. Zusätzlich erfolgt über die Einschätzung der Merkmale der Lehrkräfte die Erfassung der Qualität der Programmvermittlung. Im Hinblick auf die Operationalisierung kommt es so unserer Auffassung nach zu keiner klaren Trennung der Indikatoren. Allerdings sind wir der Auffassung, dass diese Indikatoren keine Unabhängigkeit voneinander aufweisen. So sollte sich beispielsweise die Einstellung der Lehrkraft in Bezug auf die Intervention auf den Zeiteinsatz und die Zielverfolgung auswirken. Anzumerken ist hier noch das Ergebnis der von Griffin et al. (2006) durchgeführten Faktorenanalyse. Es wurde eine Zwei-Faktorenstruktur für die Qualität der Programmvermittlung identifiziert. Sieben Items zeigten eine hohe Ladung auf dem Faktor „Bereitschaft zu unterrichten“ und drei Items luden auf dem zweiten Faktor „Klassenmanagementfähigkeiten“.

Rohrbach et al. (2010) untersuchten an 65 Schulen den Einfluss der Implementation eines Drogenpräventionstrainings. Die 2.983 Schüler\_innen wurden drei experimentellen Bedingungen zugewiesen: (A) Implementationsunterstützung für die Lehrkräfte (Workshopangebot, Coaching am Standort, technische Unterstützung, webbasierte Unterstützung), (B) nur Workshopteilnahme der Lehrkräfte, (C) Standardbedingung (standard care control). Die Qualität der Programmvermittlung wurde über ein **Beobachtungsinstrument** auf einer 7-stufigen Ratingskala erfasst. Der **Index für die Qualität der Programmvermittlung** ergab sich über die durchschnittlichen Angaben der Beobachter zu den Items Lehrereнтуhusiasmus, Selbstsicherheit, Ausmaß des respektvollen Umgangs mit den Schüler\_innen. Zusätzlich wurden Schülereinschätzungen auf einer 4-stufigen Skala *definitiv ja (stimme vollkommen zu)* (1) bis *definitiv nicht (stimme überhaupt nicht zu)* (4) zur Qualität der Programmvermittlung mit 5 Items erhoben (der Lehrer ermutigte zur Teilnahme, zeigte Respekt gegenüber den Schüler\_innen, zeigte Selbstsicherheit, war vorbereitet, reagierte auf Beiträge vonseiten der Schüler\_innen). Über den Durchschnittswert dieser Items wurde ein Index *Qualität der Programmvermittlungsfähigkeiten des Lehrers* aus Schülersicht gebildet.

#### **1.2.4 Responsivität der Programmteilnehmer**

Die Responsivität der Programmteilnehmer kann u. a. durch die Einschätzungen der Programmteilnehmenden selbst ermittelt werden, z. B. über **Befragung der Schüler\_innen** (Interviews; z. B. Hasson, 2010), aber auch durch Fremdeinschätzungen in **Logbüchern** (Hasson, 2010) oder **Lehrerfragebögen** ((Goldberg) Lillehoj et al., 2004; Mihalic et al., 2008). Ebenso werden **Beobachtungsdaten** der Programmvermittler oder der wissenschaftlichen Betreuung einbezogen ((Goldberg) Lillehoj et al., 2004; Rohrbach et al., 2007; Resnicow et al., 1998; Spoth et al., 2007; Wilson et al., 2009).

In der Studie von Rohrbach et al. (2007) (s. Abschnitt 1.2.1) wurde die Responsivität über **Fremdbeobachtung** erfasst. Dazu waren 1) **das Interesse** der Schüler\_innen, 2) **die Beliebtheit der implementierenden Person**; 3) **die Klassenkontrolle** der implementierenden Person einzuschätzen. Über diese Items wurde ein Programmakzeptanz-Index gebildet (3 Items;  $\alpha = 0.76$ ). Im Rahmen der Prozessevaluation wurden zusätzlich Daten im Schülerselbstbericht zur Einschätzungen der neun positiven Adjektive (z. B. angenehm, glaubhaft, interessant) erhoben. Diese Angaben der Schüler\_innen wurden aber nicht ergänzend zu den Beobachtungsdaten zur Erfassung der Responsivität eingesetzt, sondern nur zur Beurteilung der

Programmdurchführenden. Es ist anzunehmen, dass die Aspekte Beliebtheit und Klassenkontrolle in einem Zusammenhang zur Responsivität stehen, diese aber nicht konkret abbilden. Zudem kann die Klassenkontrolle aus unserer Sicht als Indikator für die *Qualität der Programmvermittlung* angesehen werden. Ebenso muss kritisch angemerkt werden, dass die Beliebtheit der Lehrkraft valider über ein Selbstberichtsverfahren der Schüler\_innen zu erfassen wäre. Dass die Angaben der Schüler\_innen zur Einschätzung der Programmittler nicht zur Erfassung der Responsivität eingesetzt werden ist bedauerlich. Diese Maße hätten durch die Einschätzung der neun positiven Adjektive im Schülerfragebogen die Responsivität besser abbilden können.

In der bereits in Abschnitt 1.2.1 vorgestellten Studie von (Goldberg) Lillehoj et al. (2004) wurde zur **Erfassung der Responsivität** eine 5-stufige Likert-Skala eingesetzt, auf der die Lehrer von *gering* (1) bis *hoch* (5) angeben sollten, wie sehr sich die Schüler\_innen an den Maßnahmen beteiligten. Unabhängige Beobachter nutzten den gleichen Fragebogen zur Einschätzung der Stunden.

Mihalic et al. (2008) erfassten die Responsivität der Schüler\_innen (student participation) am Ende jedes Schuljahres in einem **Lehrerselbstbericht**. Hierzu war von den Lehrkräften auf die Frage „Welcher Prozentsatz an Schülern nahm an den unterrichteten Life-Skills-Trainings-Aktivitäten teil?“ der **prozentuale Anteil der aktiv teilnehmenden Schüler\_innen** einzuschätzen. Aus diesen Angaben wurde ein durchschnittlicher Implementationswert für jeden Durchführungsort über alle Angaben zur Responsivität der Programmteilnehmer gebildet. In Folge wurde über die drei Implementationsjahre hinweg ein durchschnittlicher Implementationswert gebildet (Mihalic et al., 2008, S. 7). In multivariaten Analysen fanden sich keine signifikanten Zusammenhänge zwischen den erhobenen Einflussfaktoren (Trainingsqualität, Programmcharakteristika, Programmkoordinator, Schulcharakteristika, administrative Unterstützung, Unterstützung der Intervention durch die Lehrkräfte, elterliche Bewusstheit (parent awareness), Dauer der Interventionseinheit, Verhalten der Schüler\_innen) und der Responsivität der Schüler\_innen.

### 1.2.5 Programmspezifität

Obwohl es einen breiten Konsens gibt, dass die Erfassung der Programmspezifität notwendig ist, finden sich nur wenige Studien, die diesen Aspekt in die Bewertung der Implementationsqualität einbeziehen.

In der bereits vorgestellten Studie zur Implementationsgenauigkeit von Programmen zur Prävention des Substanzmissbrauchs (s. Abschnitt 1.2.1) von Ennett et al. (2011) beziehen sich die Ergebnisanalysen auf 342 Schulen aus einer größeren Zufallsstichprobe von 1.721 Mittelschulen (Sekundarschulen). Dabei wurde die Programmspezifität definiert als Einsatz nur eines evidenzbasierten Programms und als ausschließlicher Einsatz der intendierten Interventionsbedingungen (Ennett et al., 2011, S. 362). Zur Erfassung wurde der **Lehrer selbstbericht** gewählt. Darin war für die Programmdifferenzierung bei einem Item anzugeben, ob nur ein evidenzbasiertes Programm verwendet wurde (1) oder nicht (0), d. h. mehrere evidenzbasierte Programme einbezogen wurden. Zusätzlich wurde erfasst, wie viele alternative Unterrichtsbestandteile die Lehrkräfte einsetzten und ob es sich um Bestandteile eines evidenzbasierten oder nicht evidenzbasierten Programms handelte. Die globale Selbsteinschätzung und die Angabe der Häufigkeit der Bestandteile anderer Programme erfüllt die Voraussetzungen um die Programmspezifität nach Dane und Schneider (1998) erfassen zu können. Darüber hinaus ermöglicht die genauere Dokumentation, welche Art von Ergänzung stattgefunden hat, genauere Rückschlüsse in Bezug auf die Effekte der eingesetzten Programmelemente auf Schülerseite. Aus den lehrerbezogenen Angaben wurde schließlich ein **prozentualer Häufigkeitswert oder Indexwert für den ausschließlichen Einsatz nur eines evidenzbasierten Programms und seiner Interventionsbestandteile berechnet**. Insgesamt zeigte sich in dieser Studie nur eine geringe Abgrenzung von anderen konkurrierenden Programmen (Ennett et al., 2011), sodass sich die Interventionsergebnisse auch nicht ausschließlich auf die Interventionsbestandteile eines evidenzbasierten Programms zurückführen lassen.

Von Knoche, Sheridan, Edwards & Osborn (2010) wurde eine Studie im Rahmen des „Getting Ready Project“ durchgeführt. Hierbei wurde die Implementation von Maßnahmen zur Unterstützung der Schulfähigkeit von Kindern betrachtet. So sollten Pädagogen (N = 65) im Rahmen der Intervention das elterliche Engagement von der Geburt bis zum Schuleintritt des Kindes fördern. Die Programmspezifität wurde über das Ausmaß an Einhaltung des Programms (Adhärenz), die Responsivität der Eltern und die Effektivität bei der Unterstützung des elterlichen Engagements in Treatment und Kontrollgruppe erfasst. Die Grundidee zur Erfassung der Programmspezifität in dieser Studie folgt der Annahme: Wenn die Einhaltung der Maßnahmen sowie die Effektivität und Responsivität ausreichend unterschiedlich sind, kann von einer Abgrenzung der Programme in ausreichendem Maße ausgegangen werden. Die Effektivität wurde hier im Sinne der Qualität der Programmvermittlung betrachtet



(Knoche et al., 2010, S. 302). Als Erfassungsinstrument wurde die **Videobeobachtung** in der Treatmentbedingung (auf Eltern und Kind bezogene Maßnahmen) und Kontrollbedingung (kindfokussierte Maßnahmen) gewählt. Die Einschätzung erfolgte über das „Home Visit Coding Guide“ auf einer Likert-Skala von *niedrig/gering* (1) bis *hoch* (4). Hier sollten die Rater Einschätzungen zur (1) Adhärenz über die Nutzung der individuellen „Getting-Ready-Strategien“ und die *insgesamte Strategienutzung* (Summe) der Pädagogen einschätzen. Darüber hinaus war (2) die Effektivität bei der Initiierung elterlichen Engagements und Interesses einzuschätzen (Qualität der Programmvermittlung). Als weiterer Aspekt wurde (3) die Responsivität der Programmteilnehmer erfasst über die Einschätzung des Engagements und des Kontakts der Eltern zum Kind und des Engagements und des Kontakts der Eltern in Bezug zum Pädagogen. Ergänzend wurde zum ersten Responsivitätsmaß die (4) Interaktion zwischen (a) Eltern und Kind und (b) zwischen Eltern und Pädagogen erfasst. Die Erfassung der Interaktion wurde beim Vergleich zwischen Treatment- und Kontrollgruppe als jeweils eigenständiger Wert eingesetzt. Für andere Berechnungen wurden die Einschätzungen der Interaktionen zur Berechnung eines Wertes für die Responsivität miteinbezogen. Es wurden so Werte für die jeweiligen Indikatoren berechnet. Für den Wert der Adhärenz *Individuelle Strategienutzung* wurde die Gesamtstrategienutzung geteilt durch die Gesamtlänge der Hausbesuche (in Minuten) berechnet. Für den Wert der Adhärenz *Gesamtrate der Strategienutzung* wurde ein Summenwert über die jeweilige Strategienutzung gebildet. Für das Interesse und Engagement der Eltern mit den Pädagogen und Kindern wurde ebenso jeweils ein Wert über eine Likert-Skala von *niedrig/gering* (1) bis *hoch* (4) erfasst (Responsivität 1). Für die Interaktionen (Responsivität 2) (a) *Eltern-Kind*, (b) *Eltern-Pädagoge*, (c) *Pädagoge-Kind* wurde die Anzahl an dyadischen Interaktionsintervallen durch die Besuchsdauer in Minuten geteilt. Das Effektivitätsmaß wurde über den Mittelwert der Einschätzungen eines unabhängigen Beobachters (Likert-Skala) über alle 10-minütigen Beobachtungsintervalle gebildet. Zur Berechnung der Unterschiede zwischen Treatment- und Kontrollgruppe wurden t-Tests und Effektstärken berechnet. Auf diese Weise sollten die Schlüsselemente, die das Programm von anderen Interventionen unterscheidbar machen, identifiziert werden. In dieser Studie zeigte sich allerdings, dass die Strategien des Getting-Ready-Programms nicht ausschließlich in der Treatmentgruppe umgesetzt wurden.

Durch eine Einhaltung der jeweiligen Interventionsmaßnahmen bei vorliegendem Manual, eindeutige Festlegung der Maßnahmen und adäquate Erfassung der Adhärenz kann sichergestellt werden, dass nur die beabsichtigten Interventionsbestandteile umgesetzt wurden.

Auf diese Weise kann eine Aussage über die Spezifität des Programms getroffen werden. Bei der Effektivität und Responsivität scheint dies weniger nachvollziehbar, da nicht nur die Maßnahmen selbst für Effektivität und Responsivität verantwortlich sein könnten.

### **Zusammenfassung**

Bei der Erfassung der Implementationsqualität kann zwischen Befragungs- und Beobachtungsverfahren unterschieden werden. Beim Vergleich beider Verfahren erweist sich die Beobachtung als valider (Resnicow et al., 1998). Die retrospektive Einschätzung durch die Lehrkräfte liefert nur gering valide Angaben. Es ist anzunehmen, dass diese durch Gedächtniseffekte beeinflusst sind und die Erinnerung der anfänglichen Beteiligung und der Beteiligung am Ende des Schuljahres die Einschätzung beeinträchtigt. Zudem sind so keine Rückschlüsse auf die Beteiligung in einzelnen Interventionseinheiten mehr möglich. Hier wäre eine Einschätzung nach jeder Einheit sinnvoll. Zusätzlich sind hier wie generell im Selbstbericht auch Effekte sozialer Erwünschtheit anzunehmen. Die Implementation mittels Beobachtung zu erfassen ist jedoch aufwendiger und mit einem erhöhten Ressourcenaufwand verbunden, sodass aus Ökonomiegesichtspunkten allzu häufig darauf verzichtet wird.

Zur Messung der Indikatoren der Implementationsgenauigkeit finden sich in den empirischen Studien neben Vermischungen bzw. nicht eindeutigen Abgrenzungen (insbesondere zwischen Adhärenz und Dosierung) auch erhebliche Unterschiede bei den jeweiligen Operationalisierungen (insbesondere *Qualität der Programmvermittlung*), Skalierungen und Auswertungsstrategien. Diese heterogenen Erfassungs- und Auswertungsstrategien erschweren die Interpretier- und Vergleichbarkeit der Studienergebnisse bzw. machen sie in Teilen sogar unmöglich. Im Weiteren mangelt es an einer genauen Abgrenzung der einzelnen Indikatoren bei deren Erfassung und es werden meist nicht alle relevanten Aspekte analysiert. Die Betrachtung des Indikators *Programmspezifität* wurde bisher stark vernachlässigt.

Zusätzlich hat die Auswahl des Messinstruments Einfluss auf die Genauigkeit der Erfassung. Zudem werden die Gütekriterien Validität und Reliabilität in zu geringem Maße beachtet. Eine weitere Fragestellung betrifft die Zusammenhänge zwischen den Indikatoren der Implemen-

tationsqualität. Um die Unabhängigkeit der Indikatoren zu prüfen, sollten nachfolgende Studien alle Indikatoren einbeziehen und deren Zusammenhänge prüfen. Dies schließt moderierende oder medierende Zusammenhänge zwischen den Indikatoren mit ein (Berkel et al., 2011).

### 1.3 Auswirkungen der Implementationsqualität auf die Programmwirksamkeit

Die Implementationsforschung beschäftigt sich im Hinblick auf die Wirksamkeit einer Intervention unter anderem auch mit dem Einfluss der Implementationsqualität. Deren Erfassung soll die Frage beantworten, ob die Effektivität einer Intervention ausschließlich auf die Interventionsbestandteile zurückgeführt werden kann. Während diese Frage die Therapie- und Gesundheitsforschung schon länger beschäftigt, wird heute auch im Schulkontext deren Bedeutung für Interventionen anerkannt. Unterschiedliche Programmeffekte ähnlicher Interventionen warfen in den frühen 1970er Jahren die Frage auf, welchen Einfluss eine gelungene bzw. gescheiterte Implementation hat bzw. in welchem Zusammenhang diese zur Variabilität in der Wirksamkeit steht. „Studies of large-scale educational reforms and organizational change revealed that the level and quality of implementation largely determined the achievement of desired outcomes (Berman & McLaughlin, 1978; Borman, Hewes, Overman & Brown, 2003; Datnow, Borman & Stringfield, 2000; Fullan & Stieglbauer, 1991)“ (Shapley, Sheehan, Maloney & Caranikas-Walker, 2010, S. 9).

Diese Befunde führten zu dem Bewusstsein, dass die Erfassung der Implementationsqualität der Erfassung von Ergebnissen vorgelagert sein sollte. Nur so können die Interventionseffekte ursächlich der Intervention zugeschrieben werden. Darüber hinaus entwickelte sich auch eine veränderte Herangehensweise an das Forschungsdesign von Interventionen. Es wurde deutlich, dass die Evaluation der Beziehung zwischen der Implementationsqualität und den Effekten auf Schülerseite in die Anlage der Intervention einbezogen werden muss (Pentz et al., 1990, S. 265 f.).

Neben der Notwendigkeit eines entsprechenden Forschungsdesigns stellt sich die Frage nach der Messung der Implementationsqualität. Wie berichtet, findet sich in der Forschungsliteratur ein sehr heterogenes Bild der Erfassung von Indikatoren. Ebenso herrscht keine Übereinstimmung, **welche Instrumente** zur Messung der Implementationsqualität benutzt werden sollten.

Die Studien wählen eine Vielzahl an verschiedenen Herangehensweisen, Instrumenten und Auswertungsverfahren zur Implementationsqualität bzw. der einzelnen Indikatoren. Dies erschwert sowohl die Einordnung als auch die Vergleichbarkeit. Die unterschiedlichen Messinstrumente könnten zudem Einfluss auf die **Zusammenhänge mit der Wirksamkeit** nehmen. Dies muss bei der Interpretation der Studienergebnisse beachtet werden. Berücksichtigt man die unterschiedlichen Vorgehensweisen, so wird deutlich, dass sich für die Begriffe Implementationsqualität (fidelity) und Dosierung (dosage) in ihren Beziehungen zur Wirksamkeit uneinheitliche Befunde finden lassen. Eine metaanalytische Betrachtung über 26 Studien deutet darauf hin, dass ein höheres Ausmaß an Implementationsqualität (treatment integrity) in Zusammenhang mit größeren Veränderungen in den Zielindikatoren steht (Gresham, Gansle, & Noell, 1993). In anderen Studien finden sich nur geringe Effekte oder es werden keine Effekte auf Schülerseite berichtet (Cho, Hallfors, & Sánchez, 2005; Elias et al., 1986; Spoth, Gyll, Trudeau & Goldberg-Lillehoj, 2002).

Bis dato gibt es im Bereich der Unterrichts- und Lehr-Lern-Forschung noch wenige Studien, die die Zusammenhänge zwischen Implementationsqualität und schülerseitigen Effekten betrachten. In diesen zeigt sich, dass ein bedeutsamer Teil der Wirksamkeitsvarianz zwischen Treatmentklassen unter anderem auch auf Unterschiede im Implementierungsgrad rückführbar ist bzw. geringe oder nicht vorhandene Programmeffekte auf eine unzureichende Implementationsqualität hindeuten (Benner, Nelson, Stage & Ralston, 2011; Drössler et al., 2007; Röder & Jerusalem, 2007; Mittag et al., 2009). So fassen Rohrbach et al. (2007) die Befunde anderer Autoren zusammen: „The importance of program fidelity has been demonstrated (Dusenbury et al., 2003); low fidelity results in smaller or no program effects on behavioral outcomes (Botvin, Baker, Dusenbury, Tortu & Botvin, 1990; Botvin, Baker, James-Ortiz, Botvin & Kerner, 1992) and program mediators (Resnicow et al., 1998; Rohrbach, Graham & Hansen 1993)“ (Rohrbach et al., 2007, S. 2).

Im Folgenden wird eine Auswahl an Studien vorgestellt, in denen Zusammenhänge zwischen Indikatoren der Implementationsqualität und den Zielkriterien der Intervention untersucht wurden. Diese Studien wurden zwischen 2012 und 2017 über verschiedene Quellen zusammengetragen. Einerseits wurden Literaturrecherchen über die Datenbanken Psynindex und Psychinfo vorgenommen. Zusätzlich wurden über die Suchmaschinen Google und Google Scholar weitere Recherchen durchgeführt. Verwendete Suchbegriffe waren *treatment integ-*

*ity, fidelity, implementation, adherence, exposure, dosage, responsiveness, program differentiation.* Ausgehend von den Suchbefunden wurden weitere Quellen aus den Literaturverzeichnissen vorliegender Artikel erschlossen.

In den auf diese Weise recherchierten Studien (N = 15) wird am häufigsten der Indikator Dosierung in seinen Zusammenhängen mit den Zielkriterien betrachtet, am zweithäufigsten die Adhärenz und an dritter Stelle die Qualität der Programmvermittlung. Die Responsivität der Programmteilnehmer wird in einer Studie zu den Zielkriterien in Beziehung gesetzt. Für die Programmspezifität fanden sich keine Studien, die den Zusammenhang des Indikators mit der Wirksamkeit betrachtet.

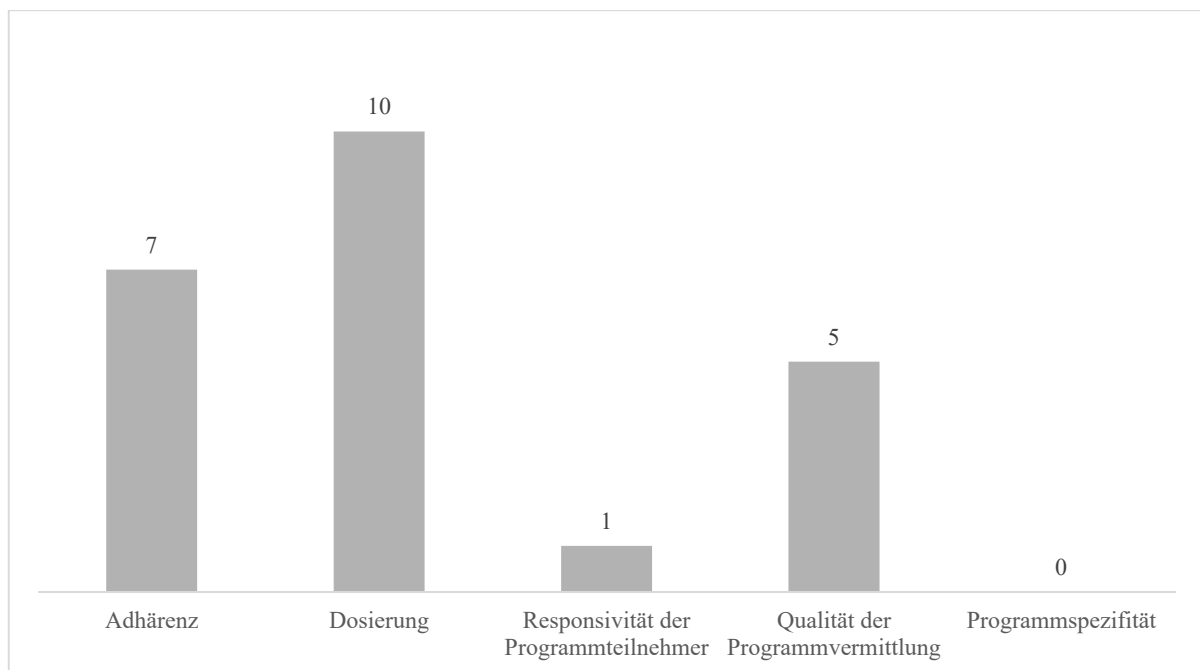


Abbildung 2. Häufigkeit der in den Studien betrachteten Indikatoren zur Implementationsqualität

Im Folgenden werden einzelne Studien und deren Zusammenhänge mit den Zielkriterien vertiefend vorgestellt. Die Auflistung verfolgt nicht den Anspruch einer Vollständigkeit oder einer Metaanalyse. Vielmehr soll ein Überblick über die gefundenen Effektstärken dargestellt werden. Um eine Vergleichbarkeit der Effektstärken zu ermöglichen, wurden die Effektstärken in Cohen's  $d$  konvertiert. Die Interpretation der Effektstärken mit  $d \leq .30$  als *kleiner Effekt*,  $d = .50$  *mittlerer Effekt*;  $d \geq .80$  „großer Effekt“ folgt der Einordnung nach Cohen (1988). Eine positive Effektstärke bestätigt die erwartete Hypothese, negative Vorzeichen widerlegen die Hypothese. Zur Berechnung von Cohen's  $d$  wurden die Berechnungsformeln

zur Umrechnung der F-Werte bei 2 Gruppen und die Überführung des standardisierten Beta in die Effektstärke  $r$  (Eisend, 2014) verwendet. Weitere Berechnungsschritte wurden über [www.psychometrica.de](http://www.psychometrica.de) durchgeführt. Hier wurden die Effektstärken  $r$  und  $\eta^2$  in Cohen's  $d$  überführt. Die Umrechnung erfolgte nach Cohen (1988) und Rosenthal (1994, S. 239) (<https://www.psychometrica.de/effektstaerke.html#transform>).

### 1.3.1 Studien zur Adhärenz

Die Bedeutung der Adhärenz für die Wirksamkeit konnte auch in größer angelegte US-Studien im Schulbereich gezeigt werden. Hier fanden sich für Interventionsgruppen mit höherer Programmeinhaltung größere Effekte auf Schülerseite bei Einstellungen und Verhaltensweisen in Bezug auf Substanzen, Substanzenmissbrauch, Sozialverhalten und Erleben des sozialen Umfelds (Battistich, Schaps, Watson, Solomon & Lewis, 2000; Battistich, Schaps & Wilson, 2004). Es scheint also ein Mindestmaß an Einhaltung der Maßnahmen erforderlich zu sein, um überhaupt eine Effektivität der Maßnahmen gewährleisten zu können (Botvin et al., 1990). Dane und Schneider (1998) weisen zudem in einer Metaanalyse darauf hin, dass größere Effekte in Teilstichproben mit größerer Programmeinhaltung zu finden sind als in der Gesamtstichprobe. Im Detail zeigten sich bei Dane und Schneider (1998) für die Adhärenz in vier von fünf Studien signifikante Effekte. Für die Exposition zeigten sich bei drei von sechs Studien signifikante Effekte, was in anderen Studien ebenso für die Dosierung (dreimal signifikant; dreimal nicht signifikant) galt.

In der Überblicksstudie von Mihalic (2004) konstatiert die Autorin zu den in ihrem Überblick betrachteten Studien: „In short, these studies show that the closer an intervention adheres to the original design, the greater the degree of behavior change“ (Mihalic, 2004, S. 100). Diese Annahme wird unter anderem gestützt von einer Metaanalyse zu Drogenpräventionsprogrammen mit 143 Studien von Tobler (1986). Hier fand sich eine mittlere Effektgröße von  $d = .34$  im Vergleich zu nicht vollständig umgesetzten Programmen. Im Vergleich zu allen betrachteten Programmen zeigte sich ein mittlerer Effekt von  $d = .30$ . Auch eine Metaanalyse von Wilson und Lipsey (2000) zu Programmen zur Prävention von Gewalt an Schulen ( $N = 196$ ) zeigte stärkere Effekte für Programme mit einer höheren Implementationsqualität. Zusammenfassend lässt sich sagen, dass Gruppen mit hoher Implementationsqualität stärkere Effekte auf die Zielkriterien bzw. eine größere Anzahl von Zielkriterien aufweisen. Teilweise zeigen sich

auch nur Effekte für die Gruppen mit hoher Implementationsqualität (Mihalic, 2004, S. 85) (s. *Tabelle 1*).

Drei Studien weisen auf bedeutsame Beziehungen der Adhärenz auf die Wirksamkeitsindikatoren hin (Battistich et al., 2000; Battistich et al., 2004; Byrnes, Miller, Aalborg, Plasencia & Keagy, 2010 (s. *Tabelle 3*); Benner et al., 2011 (s. *Tabelle 3*); Rohrbach, Graham & Hansen, 1993). Eine der Studien weist aber auch nicht signifikant gewordene Effekte der Adhärenz zu den gemessenen Kriteriumsvariablen auf (Rohrbach et al., 1993). Insgesamt zeigen diese Studien, dass die Adhärenz einen bedeutsamen Einfluss auf die Programmwirksamkeit hat. Es werden kleine bis große Effektstärken ( $d = .12$  bis  $d = 2.14$ ) berichtet. Die stärksten Effekte über die Studien hinweg finden sich in der Studie von Rohrbach et al. (1993) in Bezug auf das Zielkriterium *Programmakzeptanz*. Im Anschluss an die folgende Tabelle werden einzelne Studien und ihre zentralen Effekte umfassender vorgestellt. Tabelle 1 fasst die Effektstärken der Adhärenz bei den betrachteten Studien auf die Programmergebnisse zusammen.

Tabelle 1

*Studien zu Zusammenhängen zwischen der Adhärenz und den betrachteten Zielkriterien*

Autoren	Interventionsart	Zielgruppe	Indikator	Wirksamkeitsindikatoren	Effektstärke Cohen's $d$
<b>Battistich et al. (2000)</b>	Drogenpräventionsprogramm	Schüler_innen	Adhärenz	Reduktion Drogenkonsum Problemverhalten Einschätzung der Schulgemeinschaft	.41 bis 1.10
<b>Battistich et al. (2004)</b>	Drogenpräventionsprogramm	Schüler_innen	Adhärenz	Positive und negative Verhaltensweisen Einschätzungen der Schüler_innen durch die Lehrkräfte Schulleistungen	.12 bis .48
<b>Rohrbach et al. (1993)</b>	Drogenpräventionsprogramm	Schüler_innen	Adhärenz	Widerstandsfähigkeit Programmspezifisches Wissen Programmakzeptanz	.75 bis 2.14

Betrachtet man die Adhärenz im Sinne der Einhaltung der Interventionsbestandteile, so findet sich zum Beispiel in einer Interventionsstudie im Rahmen des längsschnittlich angelegten „Child Development Project“ (Battistich et al., 2000) mit Schüler\_innen und Lehrkräfte an 24

Schulen ( $N = 5.500$ ), dass die Einhaltung der Interventionsmaßnahmen zur Prävention von Substanzgebrauch bedeutsame Effekte auf die Wirksamkeitsindikatoren (z. B. Reduktion Drogenkonsum, Problemverhalten, Einschätzung der Schulgemeinschaft) in kleiner bis mittlerer Stärke aufweist ( $d = .14$  bis  $d = .47$ ). Die stärksten Effekte finden sich auf die „Einschätzung der Schulgemeinschaft“. Insgesamt zeigt sich, dass die Dosierung der Interventionsmaßnahmen einen günstigen Einfluss auf die Reduktion von Drogenkonsum und Involviertheit in weitere Problemverhaltensweisen nimmt.

Eine Folgestudie von Battistich et al. (2004) im „Child Development Project“ bezog 1.246 Schüler\_innen von sechs Programmschulen ein. Diese wurden in je drei Gruppen mit hoher und niedriger Adhärenz eingeteilt. Untersucht wurde der Einfluss der hohen Adhärenz in Bezug zu den Wirksamkeitsindikatoren (z. B. positive und negative Verhaltensweisen, Einschätzungen der Schüler\_innen durch die Lehrkräfte, Schulleistungen). Es zeigen sich Effekte kleiner bis mittlerer Stärke ( $d = .12$  bis  $d = .48$ ). Der stärkste Effekt zeigt sich für das Zielkriterium *Einschätzungen der Schüler\_innen durch die Lehrkräfte*. Wird der Einfluss der hohen Adhärenz dieser Studie insgesamt betrachtet, findet sich, dass diese Schüler\_innen bessere Schulleistungen, mehr prosoziale und weniger antisoziale gleichaltrige Freunde aufweisen.

Eine Interventionsstudie mit dem Ziel der Verbesserung der Lesekompetenz in der Sekundarstufe (3 Schulen) mit 281 Schüler\_innen und 21 trainierten Lehrkräften ergab (Benner et al., 2011) (s. *Tabelle 3*), dass eine hohe Implementationsqualität zu verbesserten Leseergebnissen bei Schüler\_innen mit Leseschwierigkeiten führte. Dabei erwiesen sich u. a. die Adhärenz (Einhaltung) und die Qualität der Programmvermittlung als statistisch bedeutsame Prädiktoren der Programmwirksamkeit. Insgesamt konnten 22 Prozent der Wirksamkeitsvarianz in den Effekten in den Basislesefertigkeiten und 18 Prozent der Varianz bei Verbesserungen beim Vergleichen von Textabschnitten durch Unterschiede in der Implementationsqualität (adherence; quality of delivery) erklärt werden. Es fanden sich für die Adhärenz Effektstärken von  $d = .13$  bis  $d = .24$  auf die Zielkriterien.

In einem familienbasierten Präventionsprogramm mit 381 Familien zur Prävention von Alkohol-, Tabak- und Drogenkonsum bei 11- bis 12-Jährigen von Byrnes et al. (2010) wurde die Adhärenz der Maßnahmen über Beobachtungen erfasst. Untersucht wurden die Zusammenhänge zwischen der Implementationsqualität und der Teilnahme sowie Zufriedenheit der Teilnehmer. Es zeigte sich, dass höhere Adhärenzwerte mit höheren Zufriedenheitswerten aufseiten der Schüler\_innen in Bezug auf die Interventionsmaßnahmen ( $d = 1.09$ ), aber niedrigeren



Zufriedenheitswerten der Eltern ( $d = -.82$ ) mit den Interventionsmaßnahmen einhergehen. Insgesamt zeigten sich so für die Adhärenz erwartungswidrige und erwartungskonforme Effekte in hoher Stärke. Die Autoren begründen die nicht erwartungskonformen Ergebnisse damit, dass sich eine zu starre Einhaltung ungünstig auf die Zufriedenheit der Eltern ausgewirkt haben könnte. Es fand sich kein Zusammenhang zwischen der Adhärenz und der Teilnahme am Programm.

Eine Studie von Rohrbach et al. (1993) untersuchte den Einfluss der Adhärenz auf die Zielkriterien im Rahmen eines Drogenpräventionsprogramms. Hierzu wurden zwei Gruppen jeweils mit hoher und niedriger Adhärenz gebildet. Es fanden sich nur zu Teilbereichen positive Effekte für die Gruppe mit hoher Adhärenz. Positive Effekte fanden sich für die Zielkriterien *Widerstandsfähigkeit*, *Programmspezifisches Wissen*, *Einstellung zum Substanzgebrauch* und die *Programmakzeptanz*. Für *Beurteilte Verhaltensabsichten* und *Normative Überzeugungen zum Substanzgebrauch von Freunden* fanden sich keine Effekte bei genauer Einhaltung der Programmmaßnahmen. Insgesamt fanden sich Effektstärken für die Adhärenz zwischen  $d = .75$  bis  $d = 2.14$ . Der stärkste Effekt zeigte sich auf den Wirksamkeitsindikator *Programmakzeptanz*.

### 1.3.2 Studien zur Dosierung

Zum Indikator Dosierung und den Bezügen zu den jeweiligen Zielkriterien findet sich eine größere Anzahl an Studien. Dies deckt sich mit den betrachteten Metaanalysen. Domitrovich, Gest, Jones, Gill und Sanford DeRousie (2010) berichten in ihrer metaanalytischen Auswertung, dass einige Studien auf Zusammenhänge zwischen Dosierung und Wirksamkeit auf Schülerseite hinweisen (Aber, Jones, Brown, Chaudry & Samples, 1998; August, Bloomquist, Lee, Realmuto & Hektner, 2006; Botvin et al., 1990; Spoth et al., 2002), andererseits fanden sich in anderen Studien keine Zusammenhänge (Basch, Sliepcevich, Gold, Duncan & Kolbe, 1985; (Goldberg) Lillehoj et al., 2004; Resnicow et al. 1992; Resnicow et al., 1998). Mihalic (2004) schlussfolgert in ihrem Forschungsüberblick, dass die Effektivität von Programmen durch eine nicht im vollen Umfang realisierte Dosierung insgesamt beeinträchtigt ist. Auch hier muss die uneinheitliche Erfassung der Dosierung bei der Interpretation von Forschungsergebnissen beachtet werden. Connell et al., 1985 verweisen darüber hinaus auf die Einflüsse sozialer Erwünschtheit, die selbst berichtete Angaben zur Dosierung beeinflussen und so die Ergebnisse verzerren könnten. Trotz mancher Einschränkungen und einiger uneinheitlicher

Forschungsbefunde kann aber angenommen werden, dass Programme bei zu geringer Dosierung der beabsichtigten Interventionsbestandteile weniger effektiv sind (Drössler et al., 2007).

Alle in dieser Arbeit betrachteten Studien (s. *Tabelle 2*) weisen auf bedeutsame Beziehungen der Dosierung auf die Zielkriterien hin (Gottfredson, Gottfredson & Hybl, 1993; Drössler et al., 2007; Drössler, 2010; Johnson & Johnson, 1983; Johnson, Johnson, Buckman & Richards, 1985; Olweus & Alsaker, 1991; Röder & Jerusalem, 2007). Sechs der sieben hier betrachteten Studien weisen aber auch nicht signifikant gewordene Effekte der Dosierung zu einzelnen Kriteriumsvariablen nach (Gottfredson et al., 1993; Drössler et al., 2007; Drössler, 2010; Johnson & Johnson, 1983; Johnson et al., 1985; Röder & Jerusalem, 2007). Insgesamt zeigt sich, dass die Dosierung bedeutsamen Einfluss auf die Programmwirksamkeit hat. Je nach Studie und betrachteten Zielkriterien finden sich geringe bis starke Effekte auf die Wirksamkeitsindikatoren. Die stärksten Effekte für die Dosierung werden bei Olweus und Alsaker (1991) berichtet. Im Folgenden werden in der Tabelle die Effektstärken der betrachteten Studien berichtet. Im Anschluss werden die einzelnen Studien näher vorgestellt.

Tabelle 2

*Studien zu Zusammenhängen zwischen Dosierung und den betrachteten Zielkriterien*

<b>Autoren</b>	<b>Interventionsart</b>	<b>Zielgruppe</b>	<b>Indikator</b>	<b>Wirksamkeitsindikatoren (Beispiele)</b>	<b>Effektstärke Cohen's <i>d</i></b>
<b>Drössler et al. (2007)</b>	Förderung sozialer Kompetenz	Schüler_innen/ Lehrkräfte	Dosierung	Selbstwirksamkeitserleben Hilfsbereitschaft	.16
<b>Drössler (2010)</b>	Förderung sozialer Kompetenz	Schüler_innen/ Lehrkräfte	Dosierung	Selbstwirksamkeit, Teamfähigkeit Hilfsbereitschaft in der Klasse Fürsorglichkeit der Lehrkraft	.16 bis .52
<b>Gottfredson et al. (1993)</b>	Verbesserung Problemverhalten	Schüler_innen/ Lehrkräfte	Dosierung	Klassenorganisation Störungen im Klassenraum Wahrgenommene Gerechtigkeit der Regeln	-.08 bis .74

Fortsetzung Tabelle 2

Studien zu Zusammenhängen zwischen Dosierung und den betrachteten Zielkriterien

Autoren	Interventionsart	Zielgruppe	Indikator	Wirksamkeitsindikatoren (Beispiele)	Effektstärke Cohen's <i>d</i>
<b>Johnson &amp; Johnson (1983)</b>	Kooperatives Lernen	Schüler_innen	Dosierung	Individualistisches Lernen Akad. Unterstützung durch Schüler_innen Arbeit in heterogenen Gruppen	.28 bis .34
<b>Olweus &amp; Alsaker (1991)</b>	Gewaltprävention/ Bullying	Schüler_innen	Dosierung	Klassendurchschnitt Bullying/ Opferprobleme	1.19
<b>Röder &amp; Jerusalem (2007)</b>	Förderung von Selbstwirksamkeit	Schüler_innen/ Lehrkräfte	Dosierung	Erleben schulischer Selbstwirksamkeit Erlebte subjektive Transparenz von Klassenarbeiten	.30 bis 1.01

Die Studie von Drössler et al. (2007) untersuchte eine Stichprobe mit 421 Schüler\_innen der Interventionsgruppe und 353 Schüler\_innen aus Kontrollklassen sowie 97 teilnehmenden Lehrkräften. Es wurden kooperative Lernformen in den Unterricht implementiert. Für den Indikator Dosierung wurde die Umsetzungshäufigkeit der Maßnahmen erfasst. Insgesamt zeigte sich, dass bei einer häufigeren Umsetzung der Maßnahmen tendenziell günstigere Werte bei Schüler\_innen in den Zielkriterien *Selbstwirksamkeit*, *Hilfsbereitschaft in der Klasse* auftreten ( $d = .16$ ), als für Schüler\_innen aus Schulen, die nicht die erwartete Häufigkeit in der Umsetzung leisteten. Nicht erwartungskonform konnten keine Effekte auf die *Perspektivenübernahme* der Schüler\_innen gefunden werden.

Drössler (2010) untersuchte in Folge 364 Schüler\_innen aus acht Interventionsschulen und 198 Schüler\_innen aus fünf Kontrollschulen über drei Jahre (drei Messzeitpunkte) in der Entwicklung sozialer Kompetenzen, Kompetenzüberzeugungen und Klimaerleben in der Klasse. Es wurden kooperative Lernformen sowie Maßnahmen der Klimaförderung (soziale Regeln, Klimadiagnose) im Rahmen einer indirekten Intervention in den Unterricht implementiert. Als Kennwert des Implementationsausmaßes wurde die Umsetzungshäufigkeit der Maßnahmen erfasst. Bei der Analyse der Werte auf Klassenebene zeigten sich Effekte für die Häufigkeit der Umsetzung der Interventionsmaßnahmen. Für die Umsetzungshäufigkeit der

Klassenregeln fanden sich Effekte auf die Zielkriterien (*Entwicklung der sozialen Selbstwirksamkeitserwartung*, *Wahrgenommene Lehrerfürsorglichkeit*) in einer mittleren Stärke (jeweils  $d = .52$ ). Für die Häufigkeit der Umsetzung der Maßnahmen der Klimadiagnose fand sich ein positiver Effekt auf die *Selbstwirksamkeitserwartung Teamfähigkeit* ( $d = .16$ ). Beim Zielkriterium *Erlebte Restriktivität der Lehrkraft* konnten ungünstigere Ausgangswerte der Interventionsgruppen durch den häufigen Einsatz der Klimadiagnose auf das Niveau der Kontrollgruppe (t3) abgeschwächt werden. Für die Häufigkeit der Umsetzung kooperativer Lernformen konnten aber auch in dieser Studie keine signifikanten Effekte gefunden werden. Insgesamt zeigt sich, dass Schüler\_innen aus Schulen, in denen die Maßnahmen häufiger umgesetzt wurden, günstigere Werte in den Zielkriterien *Selbstwirksamkeit Teamfähigkeit*, *Hilfsbereitschaft in der Klasse*, *Fürsorglichkeit des Lehrers* und *Erlebte Restriktivität der Lehrkraft* aufweisen als Schüler\_innen aus Schulen, die nicht über die erwartete Häufigkeit in der Umsetzung verfügten ( $d = .20$  bis  $d = .28$ ). Dies zeigte sich auch nach Kontrolle der Ausgangswerte bei den Zielkriterien (Drössler, 2010, S. 179-185).

Die Intervention von Gottfredson et al. (1993) zielte auf die Verbesserung von Problemverhaltensweisen ab. An einer Stichprobe mit Schüler\_innen und Lehrkräften von 8 Schulen (6. bis 8. Schuljahr;  $N = 5.719$ ) konnte gezeigt werden, dass die Dosierung Effekte auf die Zielkriterien aufweist (z. B. *Klassenorganisation*, *Rebellisches Verhalten*, *Wahrgenommene Gerechtigkeit der Regeln*) ( $d = -.08$  bis  $d = .74$ ). Der stärkste Effekt der Dosierung fand sich für das Zielkriterium *Klassenorganisation*. Hypothesenwidrige Ergebnisse zeigten sich bei den Schülereinschätzungen *Rebellisches Verhalten* und *Belohnung*. Hier fanden sich geringere Effekte in der *High Implementation-Gruppe* im Vergleich zur *Low Implementation-Gruppe*. Besonders anzumerken ist der Befund zum Zielkriterium *Rebellisches Verhalten*. Hier zeigte sich der stärkste Effekt für die Gruppe *Mittlere Implementation*.

Bei einer schulischen Intervention von Johnson und Johnson (1983) wurde der Einfluss der Dosierung von kooperativen Lernformen bei 859 Schüler\_innen der 5. bis 9. Jahrgangsstufe betrachtet. Hier ergaben sich bei dem Vergleich der Klassen mit hoher und niedriger Dosierung kooperativer Lernformen Effekte in der Höhe von  $d = .28$  bis  $d = .34$  auf die Zielkriterien (z. B. *Individualistisches Lernen*, *Akademische Unterstützung durch Schüler*, *Arbeit in heterogenen Gruppen*). Die stärksten Effekte der Dosierung der Interventionsmaßnahmen zeigten sich erwartungsgemäß mit  $d = .34$  in einer Reduzierung von *Individualistisches Lernen*

( $d = -.34$ ) (s. *Tabelle 2*). Insgesamt zeigte sich, dass Kooperation und die regelmäßige Teilnahme (hohe Dosierung) an kooperativen Lernformen die Wahrnehmung von Unterstützung, Hilfeverhalten und freundschaftlichem Verhalten vonseiten der Lehrkräfte und Gleichaltrigen günstig beeinflussen (Johnson & Johnson, 1983). Eine nachfolgende Studie von Johnson et al. (1985) untersuchte den Einfluss der Dosierung von Kooperativen Lernformen auf die soziale Unterstützung im Klassenraum bei 91 Schüler\_innen der 8. Klasse. Es zeigten sich Effekte in Höhe von  $d = .32$  bis  $d = .61$  auf die Zielkriterien (z. B. Kooperatives Lernen, Akademische Unterstützung durch die Lehrkraft, Wahrgenommene Fairness der Benotung). Der stärkste Effekt findet sich erwartungsgemäß mit  $d = .61$  auf das *Erreichen sozialer Zustimmung*. Die Befunde deuten insgesamt darauf hin, dass die Häufigkeit der Teilnahme und die Dauer der Teilnahme an kooperativen Lernformen (hohe Dosierung) die soziale Unterstützung im Klassenraum günstig beeinflussen (Johnson et al., 1985).

Bei Olweus und Alsaker (1991) zeigten sich Effekte der Dosierung im Rahmen der Maßnahmen zum „Bullying Prevention Program“ bezogen auf den Rückgang von Gewaltproblemen beim Vergleich von Klassen mit hoher versus niedriger Dosierung. Hier konnte durch eine hohe Dosierung der Interventionsmaßnahmen ein großer Effekt in Bezug auf Opferprobleme (Klassendurchschnitt) erzielt werden ( $d = 1.19$ ). Weitere Befunde zur Evaluation schulbezogener Programme zeigen, dass bei zunehmendem Einsatz programmspezifischer Maßnahmen (Dosierung) die interventionsbedingten Veränderungen positiver ausfallen.

In der schulischen Intervention von Röder und Jerusalem (2007) konnte der Einfluss der Dosierung beim Einsatz von Transparenzpapieren in 14 Klassen ( $N = 180$ ) der 7. und 8. Stufe über ein Schuljahr hinweg gezeigt werden. Hier fand sich, dass die Dosierung nahezu vollständig (über die erfasste Anzahl der eingesetzten Transparenzpapiere in der Klasse) das subjektive Transparenzerleben der Schüler\_innen im Hinblick auf Klassenarbeiten vorher-sagt. Ebenso konnte dadurch, dass die Dosierung der Transparenzpapiere auf Klassenebene bei der schulbezogenen Selbstwirksamkeitserwartung einbezogen wurde, eine Aufklärung der Residualvarianz von 38 Prozent erreicht werden. Insgesamt zeigten sich Effektstärken von  $d = .30$  bis  $d = 1.01$ . Der stärkste Effekt der Dosierung fand sich auf das *Erleben schulischer Selbstwirksamkeit*. Es konnte allerdings kein Einfluss der Dosierung auf die *Prüfungsängstlichkeit* nachgewiesen werden (Röder & Jerusalem, 2007, S.37 f.).

Auch im Rahmen der bereits vorgestellten Studie (s. Kapitel 1.2) von Pentz et al. (1990) fanden sich deutliche Unterschiede in den Zielkriterien (z. B. Substanzgebrauch) zwischen

Schulen mit hoher versus geringer Dosierung ( $d = .52$  m.s. bis  $d = .76$ ; s. *Tabelle 2*). So zeigte sich für die Gruppe mit hoher Dosierung ein geringerer Zuwachs an Substanzgebrauch im Vergleich zur Kontrollgruppe (Pentz et al., 1990, S. 278).

### **1.3.3 Studien zur Responsivität der Programmteilnehmer**

Die Erfassung der Responsivität der Programmteilnehmer zu den Zielkriterien wurde bisher weitestgehend vernachlässigt. In Folge kann hier nur eine Studie berichtet werden. Die Responsivität wurde hier allerdings nur in die Berechnung eines global angelegten Implementationsmaßes einbezogen. So können keine Aussagen zum Zusammenhang zwischen Responsivität und Effekten gemacht werden.

In der Studie von (Goldberg) Lillehoj et al. (2004) wurde die Responsivität der Teilnehmer in der jeweiligen Stunde erfasst. Dabei wurde auch die Übereinstimmung der Einschätzungen durch die Interventionsdurchführenden mit den Einschätzungen durch trainierte unabhängige Beobachter betrachtet. Darüber hinaus wurde der Einfluss der Implementationsqualität auf die Ergebnisse bei den Jugendlichen untersucht. Zusätzlich sollte die Beziehung zwischen Merkmalen der Programmvermittler und der Implementationsqualität ermittelt werden. Wie bereits in Kapitel 1.2 berichtet, wurde die Responsivität der Programmteilnehmer nur in die Berechnung eines global angelegten Implementationsmaßes einbezogen. Dieses globale Maß wurde zum Vergleich des Lehrerselbstberichts mit dem Bericht unabhängiger Beobachter genutzt. In Regressionsanalysen zeigten sich signifikante Zusammenhänge zwischen den Angaben zur Implementation durch die Beobachter und dem Substanzgebrauch der Jugendlichen. Für die Selbstberichtsangaben durch die Programmvermittler zeigten sich diese Zusammenhänge nicht. Für einige Subskalen konnten aber auch für die Implementationsangaben durch die Beobachter keine Effekte auf die Zielkriterien gefunden werden. Allerdings fand sich, dass einige Merkmale der Programmvermittler (z. B. vorherige Vermittlungserfahrung in Bezug auf Präventionsprogramme, Einstellung zu Tabak und Alkohol, Wissen über Marihuana, Überzeugungen im Hinblick auf normatives Peerverhalten in Bezug auf Drogen, Glaubenssätze in Bezug auf normatives Verhalten) Effekte aufseiten der Jugendlichen vorhersagen. Es finden sich Effekte in geringer bis mittlerer Stärke ( $d = .12$  bis  $d = .50$ ). Der stärkste Effekt zeigt sich für den Wirksamkeitsindikator *Glaubenssätze in Bezug auf normatives Verhalten* ( $d = .50$ ) (s. *Tabelle 3*).

### 1.3.4 Studien zur Qualität der Programmvermittlung

Tabelle 3 zeigt die Studien, die mehrere Indikatoren berücksichtigt haben. Drei Studien, die die Qualität der Programmvermittlung berücksichtigen, weisen auf bedeutsame Beziehungen der Qualität der Programmvermittlung mit den Zielkriterien hin (Byrnes et al., 2010; Benner et al., 2011; Griffin et al., 2006). Insgesamt zeigt sich, dass die Qualität der Programmvermittlung bedeutsamen Einfluss auf die Programmwirksamkeit hat. Die stärksten Effekte finden sich bei Griffin et al. (2006). Hier findet sich für die Qualität der Programmvermittlung ein starker positiver Effekt auf die Zufriedenheit mit der Intervention. Besonders anzumerken ist, dass die Qualität der Programmvermittlung bei zwei von drei Studien stärkere Effekte aufweist als die anderen gemessenen Indikatoren (Byrnes et al., 2010; Griffin et al., 2006).

Tabelle 3

*Studien zu Zusammenhängen zwischen mehreren Indikatoren und den betrachteten Zielkriterien*

Autoren	Interventionsart	Zielgruppe	Indikator	Wirksamkeitsindikatoren	Effektstärke Cohen's <i>d</i>
Benner et al. (2011)	Leseförderung	Schüler_innen	Adhärenz	Lesekompetenz	.13 bis .24
			Qualität der Programmvermittlung		.11 bis .15
Byrnes et al. (2010)	Drogenpräventionsprogramm	Familien/Adoleszente	Adhärenz	Zufriedenheit	.82 bis 1.09 <sup>a</sup>
			Qualität der Programmvermittlung		1.54 <sup>a</sup>
Griffin, et al. (2006)	Drogenpräventionsprogramm	Schüler_innen	Dosierung	Substanzgebrauch	Kein sign. Effekt
			Qualität der Programmvermittlung		.14 bis 1.96 <sup>a</sup>

Fortsetzung Tabelle 3

Studien zu Zusammenhängen zwischen mehreren Indikatoren und den betrachteten Zielkriterien

Autoren	Interventionsart	Zielgruppe	Indikator	Wirksamkeitsindikatoren	Effektstärke Cohen's <i>d</i>
(Goldberg) Lillehoj et al. (2004)	Drogenpräventionsprogramm	Schüler_innen	globales Maß Implementationsgenauigkeit (Adhärenz (Dosierung)), Responsivität, Qualität der Programmvermittlung)	Einstellung in Bezug auf Alkohol Wissen über Marihuana Glaubenssätze in Bezug auf normatives Verhalten	.12 bis .50 <sup>a</sup>
Pentz et al. (1990)	Drogenpräventionsprogramm	Schüler_innen	Adhärenz (reinvention) <b>Dosierung</b>		kein sign. Effekt <b>.52 m. s. bis .76</b>

Anmerkungen. <sup>a</sup> Hier wurde nur der Effekt der Implementation auf die Zielkriterien berechnet. Es fand kein Vergleich zu einer Kontrollgruppe statt. Das standardisierte Beta-Gewicht wurde in Cohen's *d* überführt.

Im Folgenden werden die Studien nun mit dem Fokus auf die Qualität der Programmvermittlung vorgestellt.

In der in Abschnitt 1.3.1 angesprochenen Interventionsstudie zur Steigerung der Lesekompetenz von Benner et al. (2011) fand sich für die Qualität der Vermittlung (quality of delivery) insgesamt ein positiver Effekt auf die Lesekompetenz. Die Qualität der Vermittlung wurde über die *Wiederholung der Einheiten nach Bedarf* erfasst. Für diesen Aspekt zeigte sich ein positiver Effekt auf die Basislesefertigkeiten und auf den Vergleich von Textabschnitten. Für den weiteren Aspekt der Qualität der Programmvermittlung *etablierter Umgang mit Fehlern* zeigte sich ein signifikanter Effekt in Bezug auf die Basislesefertigkeiten, der sich für die Teilaspekte des Indikators (*Lob und Rückmeldung durch die Lehrkraft, Überwachung der Reaktionen der Schüler\_innen*) nicht nachweisen ließ.

In der Studie von Byrnes et al. (2010) mit 381 Familien mit Kindern im Alter von 11 bis 12 Jahren ergab sich, dass höhere Werte der Qualität der Programmvermittlung bei dem Programm „Strengthen Families Program: For Parents and Youth 10-14 (SFP)“ höhere Zufriedenheitswerte der Eltern vorhersagen. Die Zufriedenheit der Jugendlichen konnte jedoch nicht vorhergesagt werden. Darüber hinaus zeigten sich bei einem Vergleichsprogramm



keine Zusammenhänge zwischen der Qualität der Programmvermittlung und den Zielindikatoren *Partizipation*, *Zufriedenheit der Eltern* und *Zufriedenheit der Jugendlichen*.

Die Interventionsstudie von Griffin et al. (2006) zur Prävention des Substanzenmissbrauchs wurde an 11 New Yorker Mittelschulen mit 38 Lehrern und 1.857 Schüler\_innen durchgeführt. Die Ergebnisse der Studie weisen auf die Bedeutung der qualitativen Aspekte der Programmvermittlung hin. Es wurden multiple Regressionen zur Prüfung der Vorhersagequalität der Implementationsmaße Qualität und Quantität mit dem Substanzenkonsum der Jugendlichen als Kriteriumsvariable durchgeführt. Um die Unterschiede des Substanzenkonsums der Jugendlichen in Abhängigkeit den jeweiligen Lehrkräften zu testen, wurden zudem t-Tests berechnet. Es zeigte sich, dass die Lehrmethoden und die Lehrqualität (teacher quality) einen bedeutsamen Einfluss auf die Ergebnisse der Schüler\_innen nehmen ( $d = .14$  bis  $d = 1.96$ ). Darüber hinaus konnte auch ein Zusammenhang zwischen den quantitativen und qualitativen Aspekten der Implementation gefunden werden: Die Interventionsdurchführenden, die einen *Unterrichtsvortrag halten* (lecturing), decken zwar viele Inhalte des Curriculums ab, zeigen aber einen geringeren *Einsatz effektiver Lehrstrategien* (Diskussionen) und wichtiger Lehrqualitäten wie die *Bereitschaft zu unterrichten*. Darüber hinaus fand sich, dass bei dem Einsatz von Diskussionen als Lehrmethode weniger Punkte des Programms umgesetzt wurden (Griffin, 2006, S.15). Hinsichtlich der Zusammenhänge zwischen Implementation und Substanzgebrauch zeigten sich für die quantitativen Aspekte (Prozentsatz an implementierten Interventionsbestandteilen) (Dosierung) in multivariaten Analysen nicht die erwarteten Veränderungen auf Schülerseite. Nur die qualitativen Aspekte (Einsatz von Diskussionen als Lehrmethode bzw. Grad der Lehrbefähigung) erwiesen sich als signifikante Prädiktoren eines verringerten *Substanzgebrauchs der Schüler\_innen* (Griffin et al., 2006, S. 22). Die Faktoren *Bereitschaft zu unterrichten* und *Klassenmanagementfähigkeiten* sagten die Abnahme im Substanzgebrauch für das Rauchen und Trinken ein Jahr später vorher. Es zeigen sich für die Qualität der Programmvermittlung geringe bis starke Effekte ( $d = .14$  bis  $d = 1.96$ ) auf den Substanzgebrauch, während sich kein Einfluss der Dosierung der Interventionsmaßnahmen findet. Die Befunde weisen generell darauf hin, dass entsprechende Trainingsmaßnahmen für Lehrkräfte für eine hohe Qualität der Programmvermittlung und somit auch auf für das erwünschte Zielverhalten von Bedeutung sind.

### 1.3.5 Studien zur Programmspezifität

Der Aspekt der Programmspezifität wird von Dusenbury et al. (2003) als zentraler Aspekt der Implementationsqualität angesehen. „Measuring program differentiation can be the key to assessing aspects of fidelity that are related to immediate outcomes” (Dusenbury et al., 2003, S. 244). Bedauerlicherweise findet die Abgrenzung eines Programms zu anderen Komponenten oder Programmen bisher überwiegend nur in der theoretischen Betrachtung zur Erfassung der Implementationsqualität Beachtung.

#### Zusammenfassung

In diesem Kapitel wurden Studien betrachtet, die sich mit den Effekten von Implementationsqualität auf die Zielkriterien beschäftigen. Am häufigsten werden die Zusammenhänge zwischen der Dosierung und den Zielkriterien untersucht. Dies lässt sich durch die einfache Erfassung des Indikators Dosierung erklären. Als zweithäufigster Indikator wird die Adhärenz berücksichtigt. An dritter Stelle findet sich die Qualität der Programmvermittlung, die zumeist gemeinsam mit weiteren Indikatoren betrachtet wurde. Die Responsivität der Programmteilnehmer wurde nur in einer von 15 Studien erfasst. Für die Programmspezifität konnte keine Studie gefunden werden. Dies deutet auf einen erhöhten Bedarf an zukünftiger Forschung zu den Indikatoren Responsivität und Programmspezifität und deren Effekten auf die Zielkriterien hin. Betrachtet man die Anzahl der Indikatoren sowie deren Effekte auf die Zielkriterien, zeigt sich, dass in zehn von fünfzehn Studien nur ein Indikator geprüft wurde, drei der fünfzehn Studien untersuchten zwei Indikatoren. Mehr als drei Indikatoren wurden nur in zwei von fünfzehn Studien einbezogen. Insgesamt zeigt sich, dass für den Indikator Adhärenz Effekte in großer Stärke auf die Teilaspekte der betrachteten Zielkriterien gefunden wurden ( $d = .12$  bis  $d = 2.14$ ). Im Weiteren zeigt sich für die Dosierung ein ähnliches Bild. Hier werden Effektstärken in Höhe von  $d = -.08$  bis  $d = 1.01$  berichtet. Für einzelne Zielindikatoren werden aber auch keine Effekte einer hohen Dosierung gefunden (Drössler et al., 2007; Drössler, 2010; Gottfredson et al., 1993; Griffin et al., 2006; Johnson & Johnson, 1983; Johnson et al., 1985; Röder & Jerusalem, 2007). Abschließend wurden noch Studien betrachtet, die mehrere Indikatoren erfasst und zu Zielindikatoren in Beziehung gesetzt haben. Hier zeigt sich, dass die stärksten Effekte für die Qualität der Programmvermittlung zu beobachten sind, nämlich Effekte in kleiner bis großer Stärke ( $d = .11$  bis  $d = 1.96$ ). Allerdings finden sich keine Zusammenhänge mit Teilaspekten oder auch nur Effekte für Teilgruppen (Byrnes, 2010), wobei zu beachten ist, dass die Qualität der Programmvermittlung in den hier berichteten Studien

mit unterschiedlichen Indikatoren erfasst wurde. Für die Responsivität kann keine Aussage zu Effekten gemacht werden, da sie bei (Goldberg) Lillehoj et al. (2004) ausschließlich als ein globales Implementationsmaß einbezogen wurde. Zu Effekten der Programmspezifität kann hier ebenso keine Aussage gemacht werden, da dieser Aspekt in den Studien nicht berücksichtigt wurde.

Insgesamt zeigt sich, dass der Indikator Adhärenz die stärksten Effekte auf die Zielkriterien aufweist. Beachtet werden muss bei der Interpretation dieses Befundes allerdings, dass sich die Zielindikatoren je nach Studie qualitativ bzw. inhaltlich unterscheiden. Annähernd können nur die gefundenen Effekte der Implementationsindikatoren bei Drogenpräventionsprogrammen verglichen werden.

## **1.4 Die Implementation beeinflussende Faktoren**

Im Rahmen der Implementationsforschung wird eine Anpassung von Maßnahmen an die jeweiligen Bedingungen und Gegebenheiten in der Praxis dahingehend diskutiert, ob solche Anpassungen sinnvoll sind und evtl. zu einer Steigerung der Effektivität beziehungsweise im negativen Fall zu deren Minderung führen.

Damit eine Intervention adäquat in die Praxis implementiert werden kann, besteht generell die Notwendigkeit, günstige Bedingungen bereitzustellen. Hierzu zählen eine adäquate Programmkonzeption, die Unterstützung der Maßnahme durch die Lehrkraft (Merkmale der Programmvermittler), Responsivität der Teilnehmer\_innen und Merkmale des Kontexts wie die Rahmenbedingungen in der Schule, die Unterstützung durch Schulleiter, Kollegium sowie eine unterstützende Haltung der Eltern. Im Folgenden werden auszugsweise die Merkmale der Programmkonzeption, der Programmvermittler und des Organisationskontextes, hier der Schule, in den Blick genommen. Es sollen sowohl Fragen zur Notwendigkeit der Anpassung als auch zu den weiteren Bedingungen und deren Zusammenhänge mit der Implementationsqualität betrachtet werden.

### **1.4.1 Adaptation von Programminhalten: Fidelitäts-Anpassungs-Debatte**

In der Interventionsforschung kristallisieren sich zwei gegensätzliche Positionen zur Frage der Adaption von Programminhalten heraus, die als Anpassungsperspektive und Fidelitätsperspektive beschrieben werden können. Einige Autoren wie z. B. Berman und McLaughlin

(1976) äußern sich zugunsten der (1) **Anpassungsperspektive**. Sie verweisen auf die Notwendigkeit von Modifizierungen, um auf unterschiedliche Kontextbedingungen angemessen reagieren zu können und so eine höhere Effektivität der Intervention zu erzielen. Ein weiteres Argument der Vertreter dieser Perspektive ist, dass so ein stärkerer Zuspruch seitens der Interventionsdurchführenden zu erzielen sei und damit eher eine längerfristige Umsetzungsbereitschaft erzeugt werden könne. Nachdrückliche Verfechter dieses Gedankens proklamieren die generelle Unterstützung und insbesondere Entwicklung von Interventionen in den jeweiligen Kontexten, um individuellen Bedürfnissen gerecht zu werden (Blakely et al. 1987).

Befunde stützen die Sichtweise der Vertreter der Anpassungsperspektive (Blakely et al., 1987; McGraw et al. 1996; Kerr, Kent & Lam, 1985) und zeigen, dass die Programmanpassung die Programmwirksamkeit begünstigen kann. Berman und McLaughlin (1976) fanden, dass Interventionen am erfolgreichsten waren, die mit einer Anpassung an die Bedingungen umgesetzt und eingeführt wurden.

Befürworter einer (2) **Fidelitätsperspektive** betonen die Notwendigkeit einer Einhaltung der spezifischen Interventionsbestandteile und somit der Orientierung am ursprünglichen Interventionsmodell: „They have argued that validated innovative programs should be adopted with close correspondence (fidelity) to the original model (Boruch & Gomez, 1977; Calsyn, Tornatzky, & Dittmar, 1977)“ (Blakely et al. 1987, S. 255). Blakely et al. (1987) fanden, dass bei Gruppen Interventionen mit hoher Implementationsqualität eine höhere Effektivität aufweisen als solche mit niedriger Implementationsqualität. Weitere Autoren der high-fidelity-Perspektive schlussfolgern: „Jede Abweichung ist nicht mehr theoriekonform und führt potentiell zu einer Minderung der Wirksamkeit“ (Gollwitzer & Jäger, 2007, S. 129).

So findet sich in der Debatte einerseits die Forderung nach Programmeinhaltung (adherence oder fidelity) und andererseits die Forderung nach Programmanpassung (adaptation); (Dane und Schneider, 1998, S. 25 f.; Durlak & DuPre, 2008, S. 341 f.; Berkel et al., 2011, S. 26 f.). Mittag (2014) versucht beide Aspekte als dienlich für die Effektivität von Interventionsmaßnahmen zu betrachten und konstatiert: „(...), dass die Implementationsqualität sowohl durch die Programmeinhaltung als auch durch erfolgreiche Programmanpassungen an unterschiedliche Rahmenbedingungen gekennzeichnet sein kann“ (Mittag, 2014, S. 54).

Berman (1981) führt an, dass eine Anpassung bzw. die hohe Fidelität **je nach Bedingungen** gewählt werden sollte und es keinen generellen und idealen Weg für die Implementation einer

Intervention gibt. Hierbei wird unterschieden zwischen sehr strukturierten speziellen Maßnahmen, die mit einer hohen Genauigkeit umgesetzt werden sollten, und weniger strukturierten Maßnahmen, die eine Anpassung an die Bedingungen erforderlich und möglich machen. Es werden allerdings keine genaueren Angaben zur Einordnung vorgeschlagen. Darüber hinaus zeigte sich aber auch, dass Ergänzungen die Passung und die Effektivität erhöhen können. Bei Modifizierungen war dies nicht der Fall. In der Literatur finden sich weiterhin Hinweise darauf, dass Anpassung an spezifische Gegebenheiten keine negativen Auswirkungen auf die Effektivität erwarten lässt, wenn **essenzielle Interventionsbestandteile** beibehalten werden (Bauman, Stein & Ireys, 1991; Meyer, Miller & Herman, 1993). So ist man in Komponentenanalysen der Frage nachgegangen, welche Interventionsbestandteile für den jeweiligen Interventionserfolg essenziell sind. Darüber hinaus wurde untersucht, welche Interventionsbestandteile auch im Sinne einer besseren Passung veränderbar sind, ohne die Effektivität zu beeinträchtigen bzw. diese im günstigsten Fall sogar zu steigern (Domitrovich & Greenberg, 2000). O'Connor, Small und Cooney (2007) betrachten in einem Überblicksartikel, welche Aspekte der Anpassung ohne negative Folgen auf die Programmwirksamkeit bleiben und welche Aspekte der Veränderung sich ungünstig auf den Interventionserfolg auswirken können. Sie beschreiben, dass „akzeptierbare“ bzw. auch günstige Anpassungen an die jeweilige Zielgruppe durch eine Veränderung der Sprache bzw. die Überführung der Sprache, der Bildmaterialien und der kulturellen Bezüge erfolgen können. Solche Anpassungen können darüber hinaus Veränderungen verschiedener Aktivitäten (z. B. Körperkontakt) sein und das Hinzufügen relevanter evidenzbasierter Inhalte, die die Attraktivität für die Teilnehmer erhöhen können.

Als riskant im Hinblick auf die Programmwirksamkeit beurteilen sie hingegen die Reduzierung von Anzahl bzw. Länge der Einheiten. Gleiches gilt für eine Reduzierung der Involviertheit bzw. des Engagements der Teilnehmer. Auch die Entfernung von Schlüsselementen und zu vermittelnden Schlüsselkompetenzen (die zuvor identifiziert werden müssten) sowie das Entfernen von Themen bzw. Veränderungen des theoretischen Zugangs werden als negativ für die Effektivität von Interventionen beschrieben. Gleichmaßen kann auch die Anpassung äußerer Bedingungen ungünstig sein. So stehen sowohl eine unzureichende Schulung der Trainingsdurchführenden als auch unzureichende und von der Konzeption abweichende Zuordnung von Mitarbeitern in einem negativen Zusammenhang mit der Programmwirksamkeit (O'Connor et al., 2007, S. 2).

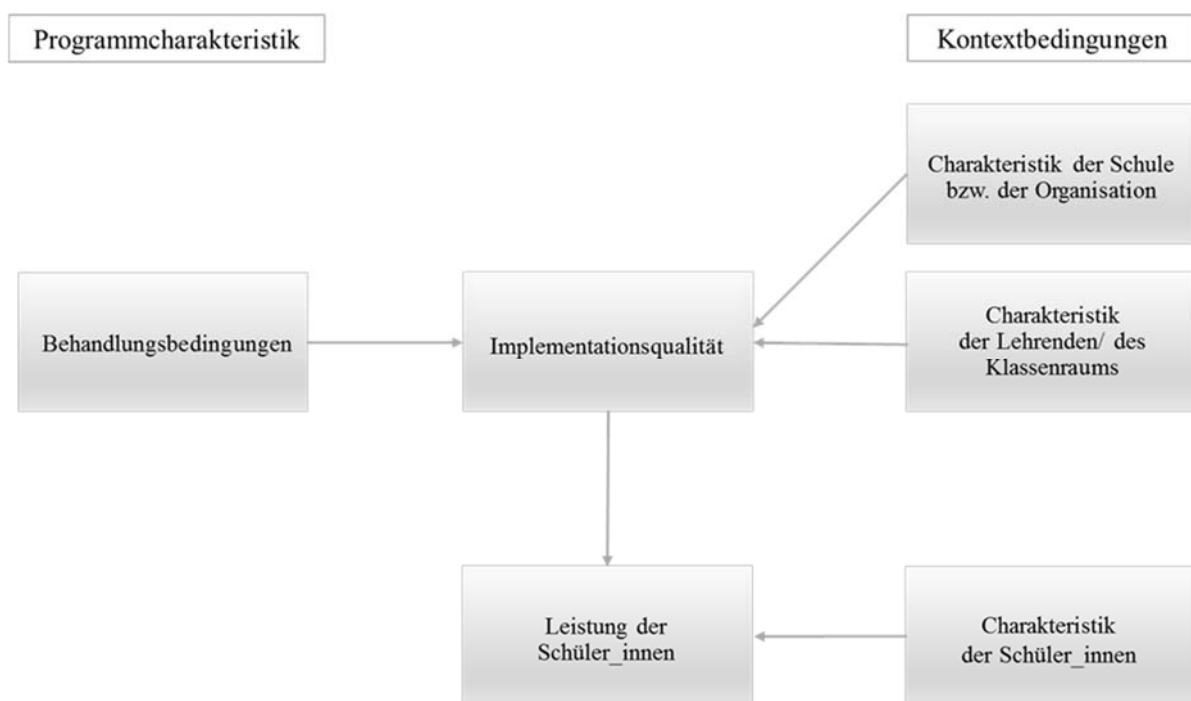
Wenn sich die Frage nach notwendigen und dienlichen Anpassungen der Programmbestandteile stellt, ist auch zu beachten, welche Maßnahmen bzw. Anpassung vonseiten der Praxis geleistet werden können bzw. müssen, um einen günstigen Einfluss auf die Implementation der Programmbestandteile nehmen zu können. Hierzu soll im Folgenden ein Blick auf die Konzeption, auf Aspekte der Vermittler und auf den Kontext in Bezug auf das Gelingen der Implementation geworfen werden.

### **1.4.2 Implementation: Merkmale von Konzeption, Vermittler und Organisation**

Berman und McLaughlin (1976) beschreiben drei verschiedene Implementationsmuster, die sich im Grad der Anpassung an die jeweiligen Bedingungen unterscheiden. Im ersten Fall handelt es sich um die Anpassung der Intervention an das System (cooptation). Hierbei findet eine Anpassung des Interventionsprogramms ohne Veränderungen im Verhalten innerhalb der Organisation statt. Bei der zweiten Form findet eine Anpassung des Interventionsprogramms mit begleitenden Veränderungen im organisatorischen Verhalten statt. Es handelt sich somit um eine einvernehmliche Anpassung der Intervention und des Systems (mutual adaption). Misslingt die Einführung und Implementation der Intervention, sprechen die Autoren von einer nicht umgesetzten Implementation (nonimplementation) (Blakely et al., 1987).

Die Anpassung äußerer Bedingungen für die Umsetzung einer Innovation kann insofern notwendig sein, als sich bestehende Strukturen gegebenenfalls verändern müssen. Von den Einzelpersonen und der Institution wird dabei in der Regel eine hohe Bereitschaft gefordert, bestehende Strukturen zu hinterfragen, Neuerungen zuzulassen, flexibel zu reagieren und bisherige Denk- und Handlungsmuster zu reflektieren. Letztlich wird so auch das professionelle Selbstkonzept der beteiligten Individuen als auch das der Institution zumindest in Teilen infrage gestellt. Nicht immer geschieht dies ohne Widerstände. So sind Merkmale der Programmvermittler und deren Einfluss auf die Implementationsqualität zu untersuchen. Die Implementation einer Intervention bedeutet zudem auch einen Eingriff in ein bestehendes System. Auch hier scheint es sinnvoll, die Implementation beeinflussende Merkmale zu identifizieren und begünstigende Bedingungen zu fördern. Andererseits stellt sich die Frage nach Merkmalen der Programmkonzeption, die entweder eine adäquate Anpassung erlauben oder eine Implementation mit hoher Fidelität ermöglichen.

In der Forschungsliteratur werden zwei Hauptdimensionen postuliert, die Einfluss auf die Implementationsqualität nehmen. Diese umfassen die Programmcharakteristik sowie die Kontextbedingungen, in denen das Programm eingeführt wird (Stein et al. 2008) (s. *Abbildung 3*). “Educational interventions are implemented within a social context that includes, among other things, the varying conditions of school organizations, teachers and classrooms. These conditions also influence the degree of implementation fidelity, which in turn affects student achievement outcomes.” (Stein et al., 2008). *Abbildung 3* zeigt die Bedeutung der Programmcharakteristik und der Kontextbedingungen für die Implementationsqualität.



*Abbildung 3.* Bedeutung der Programmcharakteristik und der Kontextbedingungen für die Implementationsqualität. In M. L. Stein et al. (2008): Scaling up an early reading program: Relationships among teacher support, fidelity of implementation, and student performance across different sites and years. *Educational Evaluation & Policy Analysis*, 30, 368-388.

In ihrer empirischen Studie zur Prüfung des Modells, bei der die Intervention zufällig ausgewählten Lehrkräften zugewiesen wurde, konnten die Autoren zeigen, dass die Leistungsergebnisse der Schüler\_innen sowohl durch die Implementationsqualität als auch durch Charakteristika der Schüler\_innen beeinflusst werden. Im Hinblick auf die Modellbildung ist kritisch anzumerken, dass Charakteristika der Schüler\_innen vermutlich auch im Sinne der Responsi-

vität einen Einfluss auf die Implementationsqualität nehmen könnten. So könnte die Implementationsqualität durch geringere Sprachkenntnisse der Schüler\_innen beeinträchtigt sein (Stein et al., 2008).

#### **1.4.2.1 Merkmale der Programmkonzeption**

Günstige Bedingungen für die Implementationen von Trainingsmaßnahmen können durch eine gelungene Programmkonzeption geschaffen werden. Im Hinblick auf die Programmkonzeption sind verschiedene Herangehensweisen zu unterscheiden. Es unterscheiden sich Interventionen u. a. darin, ob sie über (1) *direkte Interventionen*, also über direkte Maßnahmen das Zielverhalten beim Individuum zu verändern beabsichtigen, oder über (2) *indirekte Interventionen* z. B. durch Einwirkung auf die Personen im Umfeld, u.a. Eltern und Lehrer – oder auch durch eine Veränderung der Lernumgebung Einfluss nehmen möchten. So geschieht dies bei der Veränderung von Lehrplänen bzw. der Einführung von Bildungsstandards und Schulprogrammen.

**(1) Direkte Interventionen** weisen bestimmte Vorteile auf. So ist nachvollziehbar, ob die Intervention tatsächlich implementiert wurde und Störfaktoren minimiert werden können. Eine Erfassung der Effekte kann in die Intervention miteinbezogen werden. Bei direkten Interventionen können die Effekte direkt beim Individuum gemessen werden. Ein Nachteil direkter Interventionen ist die relativ ressourcenaufwendige Umsetzung.

In der vorliegenden Arbeit sollte einerseits über die Veränderung der Personen im Umfeld, also durch Schulungsmaßnahmen bei den Lehrkräften Einfluss auf die Lernmotivation der Schüler\_innen genommen werden. Im Weiteren sollte eine Veränderung der Lernumgebung z. B. durch vorgegebene Medien und Arbeitsformen, Einfluss auf die Lernmotivation der Schüler\_innen nehmen. Es handelt sich somit um eine indirekte Intervention. Die methodischen Hintergründe sollen hierzu im Folgenden ausführlicher ausgeführt werden.

**(2) Indirekte Interventionen** haben den Vorteil, dass durch die Gestaltung der (Lern)- Umgebung eine große Zielpopulation erreicht werden kann. So kann ebenso über geschulte Multiplikatoren eine große Anzahl an Zielpersonen erreicht werden. Darüber hinaus können durch die Veränderungen der Lernumgebung bzw. der damit involvierten Personen schon bei sehr jungen Kindern Effekte erzielt werden. Als nachteiliger Aspekt wird die Notwendigkeit einer ausreichenden Implementation zur Messung von Effekten auf Schülerseite genannt. Auch erfordert die erhöhte Komplexität bei indirekten Interventionen komplexere Evaluationspläne



(Otto, Kistner, Perels, Schmitz & Büttner, 2009). Zur Evaluation bedarf es somit einer Berücksichtigung der Implementation und deren Erfassung. Es ist einerseits zu prüfen, ob die beteiligten Individuen die vermittelten Wissensbestandteile, Strategien und Kompetenzen anwenden, also tatsächlich eine Implementation der angedachten Inhalte stattfindet. Andererseits muss auch die Wirkung auf die Zielgruppe erfasst werden. Wenn sich auf Schülerseite keine Effekte nachweisen lassen, können eine unzureichende Implementierung, Störfaktoren, ein ungeeignetes Design oder Messinstrument verantwortlich zu machen sein (Otto et al., 2009).

Der Aspekt Design schließt ein, dass eine Programmkonzeption auch eine Entscheidung bezüglich des Interventionszeitraums erforderlich macht. Hier kann zwischen **(1) Kurzzeitinterventionen** und **(2) Langzeitinterventionen** unterschieden werden. Generell ist die Dauer einer Intervention im Hinblick auf Interventionsziele festzulegen. Es ist ebenso miteinzubeziehen, wie viel Zeit eine Veränderung beansprucht und wie diese Veränderung messbar gemacht werden kann (Petermann, 2006). Dies bestimmt mit, ob ein Interventionseffekt überhaupt aufgedeckt werden kann.

Im Hinblick auf die Implementation stellt sich die Frage, wann sich eine zu lange Dauer einer Intervention ungünstig sowohl auf die Implementationsqualität als auch auf die Wirksamkeitsmessung auswirkt. Generell besteht die Gefahr von Ermüdungseffekten. Kurzfristig positive Effekte im Interventionsverlauf lassen sich in der Postmessung nicht mehr abbilden. Vorhandene Effekte schwächen sich ab oder verschwinden. Die Folge einer zu langen Dauer kann auch eine geringere Responsivität sein, was z. B. durch einen Stichproben-Drop-out sichtbar wird. Dies kann wiederum zur Ergebnisverzerrung führen. Darüber hinaus sind bei längeren Interventionen die Betreuung der Probanden und die Sicherung der Implementation durch entsprechende Monitoringmaßnahmen und durch Unterstützung/Betreuung von wissenschaftlicher Seite mit hohem Aufwand verbunden.

Allerdings finden sich auch Belege dafür, dass längerfristige Maßnahmen stärkere Effekte aufweisen. So fand sich ein Anstieg der Wahrscheinlichkeit des Auftretens längerfristiger Effekte (zeitlicher Transfer) bei einer mehrjährigen Interventionsdauer in einer Metanalyse von Greenberg, Domitrovich und Bumbarger (2001). Ursächlich hierfür könnte die Dauer sein, die benötigt wird, um Neuerungen in den Unterricht zu implementieren. Bestehende Handlungsroutinen können wahrscheinlicher über langfristige Maßnahmen verändert werden und in Folge Veränderungen bei der Zielgruppe erreichen (Gräsel & Parchmann, 2004; Anson

et al., 1991). Der Einfluss von verschiedenen Merkmalen der Programmkonzeption auf die Umsetzungsgenauigkeit der Programmeffekte einschließlich moderierender Variablen wird im Modell von Carroll et al. (2007) dargestellt. Im Modell wird die Implementationsqualität gleichgesetzt mit der Einhaltung (Adhärenz) der Programmmaßnahmen. Als Subkategorien der Adhärenz werden der Inhalt, die Häufigkeit, die Dauer und die Abdeckung (Dosierung) aufgenommen. In diesem Modell werden die potentiellen Moderatoren der Implementationsqualität abgebildet. Hierbei handelt es sich um den Umfang (Komplexität) der Methodenbeschreibung (Comprehensiveness of policy discription) (in Bezug auf die Beschreibung und auf die generelle Intervention), unterstützende Strategien (Manuale, Richtlinien, Trainingsmaßnahmen, Feedback), die Qualität der Programmvermittlung und die Responsivität der Teilnehmer (Carroll et al., 2007).

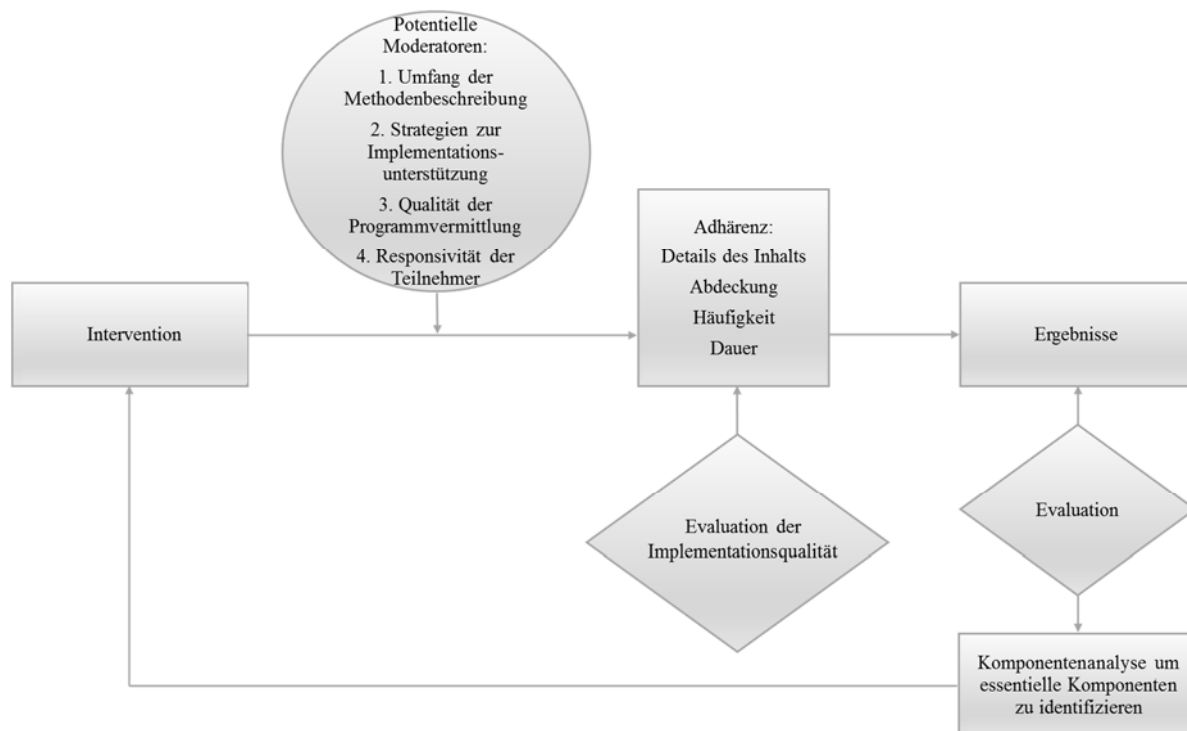


Abbildung 4. Konzeptuelles Rahmenmodell der Implementationsqualität. Aus A conceptual framework for implementation fidelity von C. Carroll et al. (2007). *Implementation Science*, 40(2).

Zum Aspekt der Komplexität findet sich einerseits, dass die **Komplexität** eine zusätzliche Belastung für die Implementation darstellen kann, andererseits wird sie im Zusammenhang mit dem Aspekt der Struktur als Erleichterung der Implementation durch vereinfachende und handlungsleitende Vorgaben genannt. Greenhalgh, Robert, Macfarlane, Bate und Kyriakidou (2004) fanden, dass bei Interventionen mit einer hohen Komplexität schwieriger eine hohe

Implementationsgenauigkeit zu erreichen ist als in einfach gestalteten Interventionsmaßnahmen. Auch Forschungsbefunde von Olweus et al. (1991) weisen auf die Komplexität als potenziellen Moderator hin. Sie schlagen vor, dass Organisationen Interventionen wählen sollten, die sich durch ein Minimum an Komplexität, aber ein Maximum an Flexibilität auszeichnen (Olweus & Alsaker, 1991).

Neben der Komplexität zeigt sich auch der Aspekt der **Strukturiertheit** als weiterer Moderator (Mihalic et al., 2008). So werden Interventionen, die genauere strukturelle Vorgaben haben, mit einer höheren Implementationsqualität umgesetzt werden (Hasson, 2010). Mihalic et al. (2008) fanden, dass komplexe und stark strukturierte Interventionen eher weniger Abweichungen von den Vorgaben zeigen und so eher eine hohe Programmeinhaltung resultiert: „Program complexity and structure have been associated with successful delivery; programs with clear goals and procedures are easier to implement and less likely to result in deviation [40,49,52,53].“ (Mihalic et al., 2008, S. 3).

Zusätzlich zu den genannten Aspekten haben **unterstützende Strategien** einen bedeutsamen Einfluss auf die Implementation von Interventionsmaßnahmen. Einen besonderen Stellenwert hat die Schulung der beteiligten Lehrkräfte. Hier ist die Bereitstellung geeigneter Materialien (standardisierte Trainingsmanuale) (Payne, Gottfredson & Gottfredson, 2006) notwendig. Eine verständlich konzipierte Schulung der beteiligten Lehrkräfte durch geschulte Trainer im Hinblick auf die Wissensinhalte, Bedeutung, Nützlichkeit der Maßnahme (auch für die Problembewältigung im Schulalltag), den Umgang mit Materialien und ein zeitnaher Interventionsbeginn sind unerlässlich, um den Transfer auf die eigentliche Zielgruppe zu begünstigen. Diese Maßnahmen ermöglichen eine subjektive Überzeugung der Realisierbarkeit (fühlen sich gut vorbereitet) und Sinnhaftigkeit für die Lehrkräfte und begünstigen so das Gelingen der Implementation. Ebenso scheinen selbst ausgewählte (Payne et al., 2006) einfach in den Schulalltag zu implementierende Maßnahmen eine Implementation zu begünstigen (Gager & Elias, 1997). Neben diesen Bestandteilen der Programmkonzeption kann auch eine externe Betreuung und Beratung in der Konzeption angelegt sein. Dies fördert das Verbindlichkeitsgefühl der Lehrkräfte und kann so die getreue Umsetzung der Implementation begünstigen (Peterson, Mori, Selby, & Rosen, 1988; Wilson, Lipsey & Derzon, 2003). Bei Langzeitinterventionen ist dies allerdings meist nicht vollständig zu leisten. Dennoch scheint die Häufigkeit unterstützender Strategien nicht zwangsläufig die Implementationsqualität günstig zu beeinflussen. Hier scheint es einen starken Zusammenhang zur Komplexität der Intervention zu

geben (Hasson, 2010). So sind bei der Planung der Programmkonzeption unterstützende Strategien nach qualitativen Gesichtspunkten auszuwählen, da eine Vielzahl an Stützstrategien die Komplexität einer Intervention erhöhen kann.

#### **1.4.2.2 Merkmale der Programmvermittler**

Bei der Durchführung von Interventionen sind in besonderem Maße die Kooperation und Unterstützung der Maßnahme durch die ausführenden Personen notwendig. Offene oder verdeckte Widerstände können die Umsetzung einer Maßnahme erschweren. Auch Stein et al., (2008) stellen fest: „Without the dedication of teachers who embrace the educational intervention, no reform will be enacted, no matter how effective it may be (Berends, Bodilly, & Kirby, 2002; Fullan, 2001)“ (S. 373). So ist auch im Rahmen der Implementationsforschung die Frage zu beantworten, welche Merkmale der Programmvermittler eine Implementation begünstigen bzw. behindern können. Darüber hinaus ist von Interesse, welche flankierenden Maßnahmen günstig auf Lehrkräfte einwirken, sodass diese die von Forscherseite entwickelte Intervention umfassend und erfolgreich in die Praxis implementieren.

In einer Studie von Pant, Vock, Pöhlmann und Köller (2008) wurden prototypische Profile der Auseinandersetzung mit Innovationen identifiziert. Die überwiegende Mehrheit der Lehrkräfte wies ein Profil auf, das die Autoren als *Kooperierer* im Hinblick auf die Innovation bezeichnen (Pant et al., 2008).

Rohrbach et al. (1993) fanden, dass ein hohes Ausmaß an Implementation durch Personen geleistet wurde, die weniger Lehrerfahrung hatten und ein hohes Maß an implementationsbezogene Selbstwirksamkeit und Programmakzeptanz aufwiesen ((Goldberg) Lillehoj et al., 2004, S. 254). Ursächlich für die höhere Implementationsaktivität der jüngeren Lehrkräfte könnte deren Suche nach unterstützenden Methoden und Materialien für die eigene Tätigkeit sein. Sie weisen so mitunter eine größere Offenheit für neue Herangehensweisen auf (Huberman, 1995). Andererseits finden sich auch Befunde, die darauf verweisen, dass ältere, aufgeschlossene, experimentierfreudige und analytisch orientierte Programmausführende ein höheres Implementationsausmaß als jüngere Lehrkräfte zeigen ((Goldberg) Lillehoj et al., 2004, S. 254). Allerdings ist anzunehmen, dass ein langjährig stabilisiertes Lehrerselbstkonzept dazu beiträgt, bestehende als erfolgreich bewertete Strategien beizubehalten, und diese Lehrkräfte somit weniger Veränderungsbereitschaft zeigen (Coburn, 2004). So scheint also nicht die eigentliche Berufsdauer die Zusammenhänge vollständig zu erklären, es können auch

weitere Aspekte eine Implementation hemmen oder begünstigen. Weitere Befunde deuten darauf hin, dass insbesondere das **Selbstwirksamkeitserleben** der Lehrer in Zusammenhang mit der Implementation von Innovationen steht (Dusenbury et al., 2003; Moncher & Prinz, 1991; Ruiz-Primo, 2005; Smylie, 1988; Vaughn, Klingner, & Hughes, 2000). Auch in der Studie von Olweus et al. (1991) zum Bullying Prevention Program wurde der Einfluss der Lehrkräfte auf die Implementationsqualität betrachtet. Sie fanden, dass Lehrkräfte eine Schlüsselrolle bei der gelungenen Implementation spielen und ein großer Anteil an Varianz der Implementation durch Unterschiede zwischen den Lehrkräften erklärt werden konnte. Als relevante Variablen zeigten sich auch hier sowohl die wahrgenommene Bedeutsamkeit der Lehrkraft (perceived staff importance) als auch die Menge an (mehr) gelesenem Programmamaterial und die affektive Involviertheit der Lehrkraft (empathische Identifikation mit Opfern von Bullying) (Olweus & Limber, 2010). Darüber hinaus führen ((Goldberg) Lillehoj et al. (2004) an, dass ein hohes Implementationsmaß in Zusammenhang steht mit einer hohen Passung zum eigenen Lehrstil und den eigenen Lehrmethoden ((Goldberg) Lillehoj et al. 2004, S. 254).

#### 1.4.2.3 Organisatorische Merkmale

Neben den eben dargestellten Merkmalen der Programmvermittler sind organisatorische und soziale Kontextfaktoren für die Umsetzung einer Intervention durch Lehrkräfte bedeutsam (Dusenbury et al., 2003; Ruiz-Primo, 2005; Smylie, 1988). In der Studie von Olweus und Alsaker (1991) konnten Schulvariablen die Implementationsqualität vorhersagen. Als bedeutsame Variablen zeigten sich die Offenheit in der Kommunikation im Kollegium (positive Einstellung zur Veränderung) und die Aufmerksamkeit der Schule in Bezug auf die in der Intervention thematisierte Problemstellung (bullying problems) (Olweus & Limber, 2010).

Als soziale Einflussfaktoren für die Umsetzung einer Intervention werden die **Führung durch die Schulleiter, das Schulklima, die Gewissenhaftigkeit (moral) des Kollegiums und die Kommunikation in der Schule** genannt (Dusenbury et al., 2003; Fullan & Pomfret, 1977; Gottfredson, 1984). „Both of these characteristics - leadership and community - are important influences on the willingness and ability of teachers to implement new educational programs. Research has consistently shown that the principal strongly influences the likelihood of program implementation and change (Berends et al., 2002; Fullan, 2001; Berman & McLaughlin, 1978)“ (Stein, et al., 2008, S. 373). Wenn von Forscherseite entwickelte Inter-

ventionen im Schulkontext eingeführt bzw. über *top down-Prozesse* von außen in ein Schulsystem oder in bestehende Strukturen einer Schule implementiert werden, bedarf es auch der Unterstützung der jeweiligen Institution als Gesamtsystem mit ihren beteiligten Funktionsträgern. So hat die Schule die Möglichkeit, durch entsprechende Maßnahmen Strukturen zu schaffen. Dies kann z. B. die Raumgestaltung betreffen oder die Freistellung, um sich in Interventionsinhalte einzuarbeiten. Das kann die Einführung einer Intervention im günstigsten Fall unterstützen, aber bei mangelhafter Kooperation von Seiten der Schule auch behindern. Hier scheint insbesondere eine Führungsfähigkeit des Schulleiters von Bedeutung zu sein, der die Lehrkräfte ermutigt, die Interventionsmaßnahmen einzuführen, und ihre Bemühungen durch ausreichende Ressourcen unterstützt (Newmann, Smith, Allensworth & Bryk, 2001). So schlagen Mihalic et al. (2008) in Folge ihrer Studienergebnisse zur Verbesserung der Implementation vor, dass Schulen Lehrkräften generell **Fortbildungen und technische Unterstützung** zur Verbesserung der Vermittlungsqualität (quality of delivery) bereitstellen sollten. Dies gilt insbesondere für den Bereich Klassenmanagement. Mit einem adäquaten Klassenmanagement kann das Schülerverhalten reguliert und so erst die Grundlage geschaffen werden, eine Intervention wie beabsichtigt umsetzen zu können. Darüber hinaus scheint die Verbindung mit dem weiteren Lehrstoff und Programmen förderlich für die Implementation zu sein (Newmann, Smith, Allensworth & Bryk, 2001). Schließlich ist auch die Schule aufgefordert, die Einführung der **Maßnahmen gegenüber den Eltern angemessen zu rechtfertigen, diese zu informieren** und bei diesen um deren Unterstützung im häuslichen Umfeld zu werben.

Die Aspekte **Kommunikation/Kooperation** sind ebenso von Bedeutung. Sie sollten innerhalb der Programmkonzeption beabsichtigt sein, erfordern aber auch die Bereitstellungen entsprechender Rahmenbedingungen innerhalb der Schule. Es finden sich Befunde dafür, dass die Einführung von Neuerungen dann genauer umgesetzt wird, wenn Personen in Teams zusammenarbeiten und sich unterstützen (Drössler, 2010; Ishler, Johnson & Johnson, 1998). Im Weiteren kann ein motivationsförderliches Rückmeldeverhalten (Rückmeldung von Stärken und Schwächen, Bekräftigung) innerhalb des Kollegiums positiv auf die Implementation wirken (Drössler, 2010). Ebenso scheint eine generelle Zusammenarbeit der Lehrkräfte in einem positiven Zusammenhang mit der Implementationsqualität zu stehen. „Implementing educational interventions may also result in greater fidelity of implementation if the program is . . . implemented by a community of teachers who work together frequently outside of their

---

classrooms to focus on the academic needs of students (Louis & Marks, 1998)” (Stein et al., 2008, S. 373).

### **Zusammenfassung**

Neben den Befunden, die auf die Notwendigkeit einer genauen Einhaltung der Interventionsvorgaben verweisen, zeigen weitere Befunde, dass auch die Bedeutung der Anpassung der Interventionen an die Bedingungen in der Praxis beachtenswert ist. Dies deutet darauf hin, dass sowohl Programmeinhaltung als auch Programmanpassungen im Sinne der Implementationsqualität betrachtet werden können. Das Gelingen der Implementation einer Maßnahme wird durch verschiedene weitere Aspekte beeinflusst. So sind neben der Programmkonzeption als solches auch Rahmenbedingungen wie Charakteristika der Personen und Merkmale der Organisation zu beachten und entsprechend zu berücksichtigen. Bei einer Programmkonzeption gilt es, sich auch immer für Dauer und Umsetzung in direkter oder indirekter Form zu entscheiden, indem die Schüler\_innen entweder direkt mit Interventionsmaßnahmen konfrontiert werden oder dies über die Vermittlung von Wissensinhalten an die Lehrkräfte geschieht. Das Format der indirekten Intervention weist den Hauptvorteil auf, eine große Anzahl an Zielpersonen erreichen zu können. Dies erfordert aber eine besondere Überwachung der Implementation und intensive Schulung der beteiligten Lehrkräfte. Je nach betrachtetem Merkmal und der benötigten Zeit der Veränderung ist eine Entscheidung über die Dauer einer Intervention zu treffen. Langzeitinterventionen z. B. über ein Schuljahr weisen den Nachteil auf, dass sie zu Ermüdung der Beteiligten und zur Vernachlässigung der beabsichtigten Implementationsqualität beitragen können. Allerdings bedarf die Veränderung der Zielkriterien möglicherweise eines längeren Zeitraumes, um Effekte der Interventionsmaßnahmen messbar zu machen. Ein weiterer Aspekt der Anlage der Intervention ist deren Komplexität. Zu komplexe Interventionen können sich nachteilig auf die Implementationsqualität auswirken. Anders verhält es sich mit der Strukturiertheit. Hier kann ein hohes Maß struktureller Vorgaben günstig für ein Gelingen der Implementation sein.

Bei der Programmkonzeption gilt es, Aspekte zu bedenken und zu planen, die mit der Einführung der Maßnahme zusammenhängen. Hierzu zählen sowohl die Konzeption der Schulung der beteiligten Personen als auch die Betreuung der Maßnahme durch Unterstützung bei Fragen und Umsetzungsschwierigkeiten. Eine gelungene Programmkonzeption kann sowohl die Qualität der Programmvermittlung als auch die Responsivität der Teilnehmer günstig beeinflussen. Zahlreiche Befunde zeigen relevante Merkmale der Lehrkraft auf. Es finden sich

zum Einfluss der Lehrerfahrung kontroverse Befunde. Eine Vielzahl von Autoren verweisen auf die Bedeutung des Selbstwirksamkeitserlebens als Prädiktor einer guten Implementationsqualität. Darüber hinaus wird in Anlehnung an Hall und Lord auch die Programmakzeptanz genannt und eine Involviertheit vonseiten der Lehrkraft, die sich durch ein erhöhtes Engagement und erhöhte affektive Beteiligung auszeichnet. Ebenso scheint die Passung zum eigenen Lehrstil günstig. Ein förderliches Merkmal der Organisation ist ein allgemein positives Schulklima. Spezielle Aspekte, für die hier ein Zusammenhang zur Implementationsqualität gezeigt wurde, umfassen die Führung durch den Schulleiter, die Gewissenhaftigkeit des Kollegiums und eine positive Kommunikation und Arbeitsbeziehung der Lehrkräfte. Grundlegend positiv wirkt sich eine hohe Aufmerksamkeit der Schule in Bezug auf die in der Intervention bearbeitete Problemstellung aus. Günstige Unterstützungsstrukturen auf der Leitungsebene umfassen die Ermöglichung von Fortbildungen, Freistellungen der in die Intervention eingebundenen Lehrkräfte, funktionale Raumgestaltung bzw. Raumvergabe und Information bzw. Einbindung der Eltern.

### **Fazit**

Die bisherige Forschung zur Erfassung der Implementationsqualität ist gekennzeichnet durch Heterogenität der theoretischen Definitionen, Operationalisierungen und Verwendung von Messinstrumenten. Zusätzlich gibt es eine uneinheitliche Auffassung darüber, welche Indikatoren zur Messung der Implementationsqualität herangezogen werden sollten. Blickt man auf die ausgewählten Studien, scheint zumindest ein Konsens in Bezug auf die Mehrdimensionalität des Konstrukts vorzuliegen. Über eine Vielzahl von Studien hinweg werden in Metaanalysen bis zu acht Teilaspekte der Implementationsgenauigkeit genannt. Es ist anzunehmen, dass keine Unabhängigkeit zwischen den Indikatoren vorliegt. Dies sollte allerdings durch die Berücksichtigung aller Indikatoren in den jeweiligen Studien geprüft werden. Die bisherige Forschungslage deutet darauf hin, dass eine Kombination der Erhebungsmethoden, Fremd- und Selbstbericht, deren jeweilige Nachteile ausgleichen können und so validere Ergebnisse möglich sind. Dies ist allerdings meist ressourcenaufwendig und nicht immer umsetzbar. Zusätzlich sollten weitere Forschungsbemühungen zur Bewertung der Erhebungsinstrumente in Bezug auf die Erfassung der jeweiligen Indikatoren geprüft werden, um Aussagen darüber treffen zu können, welches Instrument für welchen Indikator zu validen Daten beiträgt.

Neben Studien, die Einflussfaktoren auf die Implementation von Maßnahmen betrachten, bezieht sich eine Forschungsrichtung auf die Aufdeckung der Zusammenhänge zwischen



---

Implementationsqualität und Programmwirksamkeit in Bezug auf die Zielindikatoren. Insbesondere im Bereich der Unterrichts- und Lehr-Lern-Forschung stehen hierzu noch Studien aus. Die bisherige Forschungslage weist auf eine Variation der Wirksamkeit zwischen Klassen mit hoher Implementationsqualität und niedriger Implementationsqualität in den Interventionsklassen hin. Größtenteils werden die Indikatoren Adhärenz und Dosierung und deren Beziehungen mit der Programmwirkung betrachtet.

Zur Adhärenz finden sich gemischte Befunde, die darauf hindeuten, dass einerseits eine Einhaltung der angedachten Maßnahmen zu größeren Effekten auf Schülerseite führt. Allerdings kann sich auch eine zu starre Einhaltung ungünstig auswirken oder keinen Einfluss zum Beispiel auf die Teilnahmebereitschaft der Schüler\_innen haben (Byrnes et al., 2010). Für die Dosierung kann insgesamt angenommen werden, dass Programme dann weniger effektiv sind, wenn die Zielgruppe nicht die intendierte *Dosierung* (dosage/exposure) erhält. Die Befunde zur Wirksamkeit deuten darauf hin, dass sowohl Programmeinhaltung als auch Programmanpassungen im Sinne der Implementationsqualität betrachtet werden können. Neben der evtl. notwendigen Anpassung des Programms an die vorgefundenen Bedingungen finden sich in den Studien zur Erforschung der Rahmenbedingungen einer gelungenen Implementation Hinweise darauf, dass in der Programmkonzeption unterstützende Maßnahmen angedacht werden müssen.

## 2 Ableitungen für die Intervention

Die bisherigen Ausführungen haben sich mit den Aspekten des Implementationsbegriffs, den Implementationsindikatoren, den Bedingungen für eine gelingende Implementation und der Bedeutung der Implementation für die Wirksamkeit von Interventionen beschäftigt. Im folgenden Kapitel sollen nun die theoretischen Grundlagen der durchgeführten Unterrichtsin-tervention zur Förderung der selbstbestimmten Lernmotivation vorgestellt werden. In einem ersten Schritt werden Motivationstheorien vorgestellt, die bereits in Interventionskonzepte zur Förderung der Lernmotivation eingeflossen sind und Bezüge zur Selbstbestimmungstheorie aufweisen. In einem zweiten Schritt wird die Selbstbestimmungstheorie von Deci und Ryan (Deci & Ryan, 1985; Deci & Ryan, 2000; Ryan & Deci, 2002) näher ausgeführt. Sie liefert die theoretischen Hintergründe zur Förderung der selbstbestimmten Lernmotivation sowie zur Entwicklung handlungsleitender Konzepte und bietet die Basis für die dieser Arbeit zu Grunde liegende Intervention. Der letzte Teil des Kapitels beschäftigt sich mit den handlungsleitenden Konzepten der hier vorliegenden Intervention und ersten Analysen zu deren Wirksamkeit.

### 2.1 Theorien zur Lernmotivation

In spezifischen Situationen können individuelle Motive ihre Wirkung entfalten und sich in beobachtbaren Verhaltensweisen zeigen. McClelland (1987) definiert den Begriff des Motivs folgendermaßen: „Motive sind zeitlich überdauernde und interindividuell unterschiedliche Präferenzen für bestimmte Verhaltensklassen und die mit diesen Verhaltensklassen einhergehenden Anreize, insbesondere das Erleben emotionaler Befriedigung“ (McClelland, 1987 nach Dresel & Lämmle, 2017, S. 94).

Beim Beispiel des Leistungsmotivs suchen Individuen bevorzugt Leistungssituationen auf und orientieren sich an einem verbindlich erachteten Gütemaßstab. Sie streben nach der Erfüllung dieses Gütemaßstabs und dem daraus folgenden Erleben positiver Emotionen bei der Zielerreichung. So orientieren sich Schüler\_innen mit einem ausgeprägten Erfolgsmotiv (Hoffnung auf Erfolg) an Aufgaben, die das Demonstrieren ihrer Leistungsfähigkeit ermöglichen, und sie zeigen positive leistungsbegleitende Emotionen bei Erfolg wie Freude und/oder Stolz über die erbrachte Leistung. Dabei geht es nicht generell um den Wert und die erlebte

Nützlichkeit, sondern um den Prozess und das positive Erleben. „Für das angeregte Leistungsmotiv sind jedoch der Zustand, auf ein noch nicht erreichtes Ziel hocheffizient hinzuarbeiten, sowie der Stolz, es dann mit eigener Tüchtigkeit erreicht zu haben, ausschlaggebend“ (Rheinberg & Fries, 2001, S. 354).

Rheinberg (2002) definiert den Begriff der Motivation als die „aktivierende Ausrichtung des momentanen Lebensvollzugs auf einen positiv bewerteten Zielzustand“ (S. 17) Rheinberg (2004) ergänzt: „Der ‚positiv bewertete Zielzustand‘ kann gelegentlich in der Vermeidung oder der Abwehr unerwünschter Ereignisse bestehen. (...) Der Zielzustand ist das, was Anreiz besitzt, und die zielführende Aktivität ist lediglich das Instrument, das diesen Zielzustand herbeiführt“ (S. 2). Rheinberg (2004) erweitert die eigene Definition dahingehend, dass „(...) bei klar zielgerichteten Lebensvollzügen Anreize nicht nur in den angestrebten Ergebnisfolgen, sondern auch im Vollzug der Tätigkeit liegen können. Dabei können die Anreize von Tätigkeiten und ihren Ergebnisfolgen das gleiche Vorzeichen haben (z. B. attraktive Tätigkeiten führen zu erwünschten Folgen), sie können aber auch verschiedene Vorzeichen haben (aversive Tätigkeiten führen zu erwünschten Folgen und umgekehrt)“ (S. 5). Anreize, die im Vollzug der Tätigkeit liegen, können so als intrinsisch, also innen verortet, beschrieben werden. Um extrinsische Anreize handelt es sich bei anreizbesetzten Ereignissen bzw. Veränderungen von „außen“, nach erfolgreicher Ausführung der Tätigkeit (Rheinberg, 2004, S. 6). Neben der Motivationsform der intrinsischen Motivation, bei der Handlungen bzw. Tätigkeiten um ihrer selbst willen ausgeführt werden und die motivationalen Anreize in der Handlung selbst liegen, wird so die Motivationsform der extrinsischen Motivation beschrieben. Bei dem Vorliegen einer extrinsischen Motivation werden Handlungen auf Grund von Handlungsanreizen, die außerhalb der Handlung liegen, ausgeführt. Hier sollen positive Konsequenzen erreicht oder negative Konsequenzen vermieden werden.

Es stellt sich die Frage, wie die Aspekte intrinsische vs. extrinsische Motivation im Kontext von schulischem Lernen zu bewerten sind und welche Faktoren günstig auf die Lernmotivation der Schüler\_innen wirken. Die Lernmotivation der Schüler\_innen zeigt sich generell in der Initiierung eigener Lernhandlungen, in der Anstrengung und Ausdauer beim Lernen, dem Setzen adäquater Unterziele und sinnvoller Planung der Lernhandlungen. Der Lernende stellt günstige Bedingungen zum Lernen her und erlebt dabei positive handlungsleitende Emotionen und weist wenig handlungsirrelevante Kognitionen auf. All diese einzelnen Handlungen, die das Verhalten während des Lernens kennzeichnen und als Indikatoren Rückschlüsse auf

die Motivation erlauben, haben einen großen Einfluss auf schulisches Lernen und die Schülerleistung (Helmke & Schrader, 2006). Befunde zeigen, dass sich schon in der Grundschule eine abnehmende Lernmotivation, die sich bis in die Sekundarstufe fortsetzt, bei den Schüler\_innen zeigt (Lewalter & Bieg, 2009; Helmke, 1993; Wild, Rammert, & Siegmund, 2006). In der Praxis im Kontext Schule wird häufig versucht, externe Anreize (Lob, gute Noten ...) einzusetzen. Die empirischen Befunde legen ein *Primat der intrinsischen Motivation* nahe. So finden sich positive Zusammenhänge intrinsischer Motivation mit dem Kompetenzerwerb von Schüler\_innen (Ryan & Deci, 2002). Dennoch sind auch extrinsische Motivationsquellen (z. B. das Setzen von Belohnungen nach dem Erreichen von Teilerfolgen im Sinne der operanten Konditionierung) notwendig, um über längere Lernphasen ein Lernziel zu verfolgen (Fries, 2010).

Generell stellt sich die Frage, woran ein motiviertes Verhalten von Schüler\_innen im Unterricht sichtbar wird und welche Bedeutsamkeit die Motivation für das Lernen der Schüler\_innen hat. Die Motivation der Schüler\_innen steht in Zusammenhang mit der Auswahl, der Persistenz und der Intensität von Handlungen bzw. deren Ausführung und dem Erleben bei der Handlungsausführung (Fries, 2010). So zeigt sich Motivation im Lernprozess durch die Initiierung von Handlungen, adäquates Setzen von Subzielen und Planen von Lernhandlungen und dem Herstellen günstiger Bedingungen. Zudem wird sie sichtbar in der gezeigten Anstrengung und Ausdauer beim Lernprozess. Mit ihr einhergehen günstige handlungsbegleitende Emotionen (z. B. Lernfreude) und wenig handlungsirrelevante Kognitionen. Dies hat somit große Bedeutung für schulisches Lernen und Leistung (Helmke & Schrader, 2006). So verweisen auch Otto et al. (2009) auf die Relevanz der Lernmotivation für alle Phasen des Lernens: „Häufig wird angenommen, dass die Lernmotivation nur am Anfang eines Lernprozesses von Bedeutung ist, damit das Lernen begonnen wird. Sie kann jedoch in allen Phasen des Lernens (vor, während, nach) günstige Auswirkungen auf das Lernverhalten und damit auf das Lernergebnis nehmen“ (S. 288).

Eine Vielzahl an Theorien beschäftigt sich somit mit motivationalen Vorbedingungen für erfolgreiches Lernen in Lehr-Lern-Settings. Die Selbstbestimmungstheorie (Deci & Ryan, 1985, 2000) und die Person-Gegenstandstheorie des Interesses (Krapp, 1996, 1998, 2002) beschäftigen sich mit intrinsisch motiviertem Verhalten bzw. selbstbestimmtem Lernen und dem Erreichen dieses Zustands. Die Selbstbestimmungstheorie befasst sich auch damit, wie ein

hohes Maß an Identifikation, eine identifizierte Motivation im Hinblick auf externale Anforderungen entstehen und ein hoher Grad an Selbstbestimmung erlebt wird. In der Selbstbestimmungstheorie wird angenommen, dass dies über die Erfüllung der drei psychologischen Grundbedürfnisse gelingen kann. Dies umfasst das eigene Kompetenzerleben, d. h. positive Rückmeldung über die eigenen Fähigkeiten zu erhalten, das Autonomieerleben, also möglichst selbstbestimmt handeln zu können, und insbesondere das Erleben von sozialer Eingebundenheit. Das dem Menschen inhärente Bedürfnis der sozialen Eingebundenheit, der Orientierung an dem sozialen Gefüge und den sozialen Beziehungen, trägt dazu bei, dass ursprünglich externale Ziele verinnerlicht werden. Die Bedeutsamkeit einer Handlung sollte so über die Verwirklichung der drei psychologischen Grundbedürfnisse nach Autonomie, Kompetenzerleben und sozialer Eingebundenheit erreicht werden (Nähere Ausführungen zur Selbstbestimmungstheorie hierzu in Kapitel 2.2).

Die Person-Gegenstandstheorie des Interesses (Krapp, 1996, 1998, 2002) ergänzt diese Annahmen durch die besondere Betonung der subjektiven Bedeutsamkeit des Lerngegenstands. Die Übernahme des externen Ziels geschieht durch die Herstellung eines lebensweltlichen Bezugs (Krapp, 2002). Es ergeben sich eine subjektive Bedeutsamkeit und Identifikation mit dem Gegenstand und in Folge Interesse am Lerngegenstand. Die intrinsische Motivation wird als Interesse bezeichnet, wenn der Anreiz mit dem Gegenstand verknüpft ist. Eine Lernaktivität ist im Sinne der Person-Gegenstandstheorie des Interesses „intrinsisch motiviert“, wenn sie als selbstbestimmt bzw. selbstintentional erlebt wird, weil es eine Identifikation mit dem Lerngegenstand gibt. Dies kann auch der Fall sein, wenn die Aufgabe von außen vorgegeben ist. Die entscheidenden Aspekte sind die Selbstbestimmtheit und die persönliche Bedeutsamkeit des Gegenstands (Rheinberg, 2004). „Liegt der Anreizschwerpunkt dagegen auf der Tätigkeit [...], so haben wir es nach Schiefele (1996) mit einer intrinsischen Motivation zu tun, die von tätigkeitsspezifischen Vollzugsanreizen im Sinne Rheinbergs (1989, 1993) getragen wird“ (Rheinberg, 2004, S. 8). Interesse sollte sich, so auch Fries (2010), in Anlehnung an die Selbstbestimmungstheorie in Handlungskontexten einstellen, wenn die drei psychologischen Grundbedürfnisse erfüllt werden (Fries, 2010, S. 151). Die Selbstbestimmungstheorie und die Pädagogische Interessentheorie weisen somit Ähnlichkeiten auf, unterscheiden sich aber hinsichtlich der Einschätzung/Einordnung der Wahrnehmung der subjektiven Bedeutsamkeit für den Internalisierungsprozess und der generell notwendigen Bedingung der Internalisierung externaler Anforderungen für die selbstbestimmte Lernmotivation (Rakoczy,

Klieme, & Pauli, 2008). So betont die Interessentheorie das Erleben der **subjektiven Bedeutsamkeit** zur Internalisierung externer Anforderungen. Die Selbstbestimmungstheorie legt einen Schwerpunkt auf die verschiedenen Integrationsformen und den damit verbundenen **Grad an erlebter Selbstbestimmung**.

In der Zielorientierungstheorie (Harackiewicz, Barron, Pintrich, Elliot, & Trash, 2002; Köller & Schiefele, 2006) werden Aussagen darüber gemacht, an welchen Zielen Individuen sich in leistungsthematischen Situationen orientieren. Hier wird unterschieden zwischen einer Lern- und Leistungszielorientierung. Die Theorie nimmt an, dass Personen mit einer Lernzielorientierung ihre Lernziele aufgrund des Interesses am Lerngegenstand verfolgen. Erfolgserleben und positive Emotionen stellen sich im Zusammenhang persönlicher Lernzuwächse ein (Fries, 2010). Individuen mit einer Leistungszielorientierung führen hingegen Handlungen zum Erreichen eines Lernziels aus, um sich kompetent zu zeigen (Annäherungs-Leistungszielorientierung), oder sie wollen vermeiden, dass mangelhafte Kompetenz (Vermeidungs-Leistungszielorientierung) sichtbar wird. Diese Konzepte weisen eine hohe inhaltliche Nähe zu den Konstrukten intrinsische und extrinsische Motivation auf. So wird die Lernzielorientierung der intrinsischen Lernmotivation und die Leistungszielorientierung der extrinsischen Lernmotivation zugeordnet (Rheinberg, 2004, S. 16).

Die Erwartung-mal-Wert-Theorie (Wigfield & Eccles, 2000) stellt eine multiplikatorische Verknüpfung zwischen der Erwartung der Zielerreichung und dem persönlichen Wert des Ziels her, die die Intensität der Motivation beschreibt. Je nach Erwartung der Zielerreichung und dem persönlichen Wert des Ziels sollten Lernsituationen aufgesucht bzw. gemieden werden. Der Wert einer Aufgabe setzt sich aus verschiedenen Komponenten zusammen, nämlich der Relevanz für das eigene Selbstkonzept (attainment value), dem Nutzen (utility value), dem intrinsischen Anreiz (intrinsic value) abzüglich der entstehenden Kosten (z. B. Anstrengung) bei der Aufgabenausführung (Fries, 2010). Aus individuellen Kausalattributionen und Interpretationen in Bezug auf vorangegangenen Leistungssituationen resultieren subjektive motivationale Überzeugungen. Motivationale Überzeugungen umfassen das Interesse, die Zielorientierung, das Selbstkonzept und die Selbstwirksamkeit. Diese nehmen Einfluss auf die Erwartung der Zielerreichung und den Aufgabenwert (Eccles & Wigfield, 2002).

Während die genannten Theorien im Interventionsbereich bisher weniger Beachtung fanden, hat die Leistungsmotivationsforschung einen erheblichen Beitrag für Maßnahmen in Interventionen bereitgestellt. Erste Beiträge für Interventionen wurden durch theoretische Ansätze im

Rahmen der Leistungsmotivationsforschung von McClelland (1965) vorgenommen. Bei leistungsmotiviertem Verhalten geht es um die Demonstration und Bewertung der eigenen Kompetenz als Verhaltensanreiz. Diese Bewertung kann sowohl vom Individuum selbst als auch von anderen Personen vorgenommen werden. Ein Bewertungsmaßstab ergibt sich aus internalisierten Werten und Normen. Bei der Leistungsmotivation handelt es sich so um eine besondere Form der Zielverfolgung. Diese Zielverfolgung zeigt sich durch „Handlungen oder Handlungsergebnisse, (die) auf einen Tüchtigkeitsmaßstab bezogen werden, den man für verbindlich hält, sodass am Ende letztlich Erfolg oder Misserfolg steht“ (Heckhausen, 1974, S. 170).

Auch Heckhausen, Schmalt & Schneider (1985) haben sich in diesem Rahmen mit dem Aufsuchen und Vermeiden von Lernsituationen beschäftigt. Durch das Selbstbewertungsmodell konnte die theoretische Grundlage für weitere Interventionsmaßnahmen geschaffen werden. Das Modell unterscheidet zwischen erfolgsmotivierten und misserfolgsmeidenden Personen. Erfolgsmotivierte und misserfolgsmeidende Personen unterscheiden sich in der Ziel- und Anspruchsniveau-Setzung, in der Ursachenzuschreibung von Leistungsergebnissen und der Selbstbewertung in Folge der Aufgabenbewältigung. So wird das Leistungsmotiv über diese drei Prozesskomponenten dargestellt. Heckhausen nimmt eine Wechselwirkung und wechselseitige Stabilisierung zwischen den Prozesskomponenten Ziel- und Anspruchsniveausetzung, Ursachenzuschreibung und Selbstbewertung an. In Bezug auf die Ziel- und Anspruchsniveausetzung sollten erfolgszuversichtliche Personen eher realistische Ziele wählen und mittelschwere Aufgaben bevorzugen, die herausfordernd, aber noch erreichbar sind. Misserfolgsängstliche hingegen setzen sich zu hohe oder zu niedrige Ziele und bevorzugen zu schwere oder zu leichte Aufgaben. Im Hinblick auf die Ursachenzuschreibung sehen erfolgszuversichtliche Personen Ursachen für Erfolg internal verortet (hohe Fähigkeit, hohe Anstrengung) und Misserfolge werden variabel attribuiert (z. B. auf mangelnde Anstrengung zurückgeführt). Misserfolgsängstliche verorten Ursachen für Erfolg eher external (Zufall, Glück). In Folge dieser Wahl von Aufgaben und deren Bewertung haben erfolgszuversichtliche Personen mehr Anlass zu einer positiven Selbstbewertung. Die Selbstbewertung in Folge von Erfolg ist Freude über die eigene Tüchtigkeit. Bei Misserfolg stellt sich nur in geringem Maße Ärger ein. Im Modell werden zwei Varianten der Misserfolgsangst beschrieben. Variante 1 zeigt sich darin, dass niedrige Ziele gesetzt werden. Bei Erfolg geschieht die Ursachenzuschreibung external, z. B. wird Glück für die Bewältigung der Aufgabe als ursächlich betrachtet. Im Falle

eines Misserfolgs wird dieser internalen und stabilen Faktoren (mangelnde Fähigkeit) zugeschrieben. Die Selbstbewertung fällt also ungünstig aus, da Erfolg keine Freude über die eigene Leistung nach sich zieht und bei Misserfolg eine große Enttäuschung erlebt wird. Ein Misserfolg trotz geringer Aufgabenschwierigkeit legt eine negative Selbstbewertung als Folge einer Ursachenzuschreibung im Hinblick auf mangelnde Fähigkeiten nahe. Um weitere Misserfolge so gering wie möglich zu halten und eine Bedrohung des Selbstwerts zu minimieren, kommt es zur Verfestigung der Aufgabenwahl *leichte Aufgabe* und der Stabilisierung dieser Strategie. Bei der 2. Variante der Misserfolgsangst werden zu hohe Ziele gewählt. Erfolge werden external (Glück, Zufall) attribuiert und Misserfolg external und stabil (z. B. zu hohe Anforderung) erklärt. Auch hier stellt sich eine negative Selbstbewertungsbilanz ein: Aus Erfolg resultiert kein Stolz über die eigene Tüchtigkeit, aus Misserfolg allerdings auch weniger Ärger über das Scheitern. Generell werden von misserfolgsmotivierten Personen Anforderungssituationen so eher nicht aufgesucht und insbesondere mittelschwere Aufgaben gemieden, die Rückschlüsse auf die eigene Verantwortung, also die Attribution auf interne Faktoren wie Fähigkeit oder Anstrengung, zulassen.

Die Einordnung der Leistung als Erfolg oder Misserfolg ergibt sich durch die Orientierung an einer akzeptierten Bezugsnorm. So stehen die individuelle und die soziale Bezugsnorm in Zusammenhang mit der Leistungsmotivation (Rheinberg, 1987). Es sind drei Arten von Bezugsnormen zu unterscheiden: erstens die soziale Bezugsnorm, die die Leistung des Individuums im Vergleich zu seiner sozialen Gruppe betrachtet, zweitens die individuelle Bezugsnorm, die die Leistung des Schülers bzw. der Schülerin zu den vorangegangenen eigenen Leistungen in Beziehung setzt, und drittens die sachliche bzw. kriteriumsbezogene Bezugsnorm, die ein festgelegtes Kriterium wie z. B. den Bildungsplan als Ausgangspunkt zur Leistungsbeurteilung vorsieht. Bei der sachlichen Bezugsnorm werden an den Aufgaben orientierte Kriterien herangezogen und bedingen somit ein als falsch oder richtig bewertetes Handlungsergebnis.

Eine soziale Bezugsnorm wird meist zwecks einer Vergleichbarkeit von Leistungen im Rahmen einer Schulklasse eingesetzt. Dabei wird außer Acht gelassen, welches absolute Leistungsniveau die Klasse aufweist. So wird ein durchschnittlicher Schüler bzw. eine Schülerin in einer Klasse mit geringem Leistungsniveau und leistungsschwächeren Mitschüler\_innen im sozialen Vergleich eine bessere Leistungsbeurteilung erhalten als in einer Klasse auf einem hohen Leistungsniveau (*big-fish-little-pond-effect*). So erfolgt die Leistungsbewertung des



Schülers bzw. der Schülerin auch immer in Relation zur Vergleichsgruppe: ein Schüler bzw. eine Schülerin macht so möglicherweise die Erfahrung, trotz vermehrter Anstrengung und individueller Leistungssteigerung im sozialen Vergleich stets schlecht abzuschneiden. Dies hat in Folge einen negativen motivationalen Einfluss. Die Erfahrung, trotz vermehrter Anstrengung schlechte Ergebnisse im sozialen Vergleich zu erzielen, hat Konsequenzen für die Kausalattribution. Misserfolge werden eher auf mangelnde Fähigkeiten (internal, stabil) zurückgeführt. Leistungssituationen werden so im Hinblick auf diese Ursachenzuschreibung eher gemieden bzw. entsprechend zu leichte oder zu schwere Anforderungen gewählt (Misserfolgsmeidende Variante 1 oder 2). Zu leichte Aufgaben reduzieren die Misserfolgswahrscheinlichkeit. Misserfolge nach zu schwierigen Aufgaben können auf den zu hohen Schwierigkeitsgrad der Aufgabe (external) attribuiert werden. Eine Attribution auf internale Faktoren kann so vermieden und der eigene Selbstwert geschützt werden.

Leistungsstarke Schüler\_innen machen durch die Verwendung einer sozialen Bezugsnorm die Erfahrung, bessere Leistungen als Schüler ihrer Vergleichsgruppe zu erbringen. Erfolge können somit auf die eigene Fähigkeit attribuiert werden, Misserfolge auf external-variable Faktoren wie Zufall bzw. Pech oder den internal-variablen Faktor Anstrengung. Diese Schüler werden in Folge eher erfolgszuversichtlich, was sich wiederum günstig auf die Leistungsmotivation der Schüler auswirken sollte, da diese Schüler bestrebt sind, das Leistungsniveau beizubehalten. Um eine Misserfolgsmeidung bzw. Misserfolgsängstlichkeit der Schüler zu reduzieren, bietet sich die Verwendung der individuellen Bezugsnorm an. Durch die Betrachtung des eigenen Lernfortschritts werden für die Schüler\_innen eigene Lernzuwächse deutlich, und es kann eine Verbindung von eigener Anstrengung und Leistung erreicht werden.

Die individuelle Bezugsnorm fördert eine Kausalattribution auf die eigene Anstrengung (internal-variabler Ursachenfaktor). Durch die individuelle Bewertung wird ein Zuwachs deutlich und dies legt nahe, sich auch zukünftig vermehrt anzustrengen, um die eigenen Handlungsergebnisse zu steigern. In Folge können bessere Leistungen erbracht werden, was die Hoffnung auf Erfolg erhöht und somit günstig auf die Leistung und das Aufsuchen von Leistungssituationen wirkt. Ebenso wird die Kausalattribution auf internale Faktoren (Fähigkeit) gefördert. So weisen Schüler\_innen, die von einer Lehrkraft mit individueller Bezugsnorm unterrichtet werden, mehr Hoffnung auf Erfolg und im geringeren Ausmaß Furcht vor Misserfolg auf (Holodynski & Oerter, 2008).

Die Bezugsnormorientierung kann in Hinblick auf die Leistungsbeurteilung je nach Verwendung förderlichen oder hemmenden Einfluss auf die Lernmotivation nehmen. Ist also die Anstrengungsbereitschaft der Schüler zu fördern bzw. zu erhalten, ist auch eine unterschiedliche Verwendung der Bezugsnorm anzudenken. Ein Nahelegen des Ursachenfaktors Anstrengung verweist auf die eigene Beeinflussbarkeit der Zielerreichung. Misserfolgsmeidende profitieren durch eine Verwendung der individuellen Bezugsnorm insbesondere dadurch, dass ihre Zuwächse auf die eigene Anstrengung zurückgeführt werden können. So kann Misserfolgsangst reduziert werden. Dieser Zusammenhang konnte vielfach aufgezeigt werden. Dieses für die Lehrkräfte recht einfach umsetzbare Verhalten bei Feedback und Leistungsbeurwertungsprozessen konnte vielfach erfolgreich in Interventionsmaßnahmen umgesetzt werden (Rheinberg & Krug, 1999).

Die unterschiedlichen Theorien können Beiträge zur Konstruktion von Interventionsmaßnahmen liefern. In Folge soll dargelegt werden, welche Empfehlungen aus den einzelnen Theorien abgeleitet werden können.

Die Selbstbestimmungstheorie von Deci und Ryan (1985, 2000), die auch die Basis für die hier betrachtete Intervention darstellt, soll an dieser Stelle nur kurz ausgeführt werden. Die Selbstbestimmungstheorie bietet mit ihrer Prämisse einer Verwirklichung der drei psychologischen Grundbedürfnisse nach Kompetenzerleben, Autonomieerleben und dem Erleben von sozialer Eingebundenheit den Rahmen, den Lehr- und Lernkontext so zu beeinflussen, dass er diesen Anforderungen gerecht wird. Dies kann durch Veränderungen der Unterrichtsmaßnahmen und Materialien, aber auch der Schulung der Lehrkräfte mit Wissensinhalten zur Selbstbestimmungstheorie versucht werden. Schiefele (2004) schlägt insbesondere zur *Förderung der Autonomie* folgende Punkte vor: Schüler\_innen sollten an der Festlegung von Lernstoff und Lernziel mitwirken können, Lehrkräfte mittels autonomieförderlichen Lernmethoden Handlungsspielräume schaffen (z. B. offener Unterricht), Verhaltensregeln gemeinsam ausgearbeitet werden, Maßnahmen zur angemessenen Selbstbewertung und Ankopplung an übergeordnete Ziele fokussiert werden. *Kompetenzerleben* kann sowohl durch ein individuelles Feedback vonseiten der Lehrkraft als auch in Peer-Teaching-Situationen bzw. durch kooperative Lernformen unterstützt werden. Diese Lernformen können ebenso die *soziale Eingebundenheit* fördern. Darüber hinaus dient ein individuelles Rückmeldeverhalten der Förderung der Lehrer-Schüler Beziehung und somit dem Erleben sozialer Eingebundenheit. Vertiefend wird dies im folgenden Abschnitt 2.2 ausgeführt. Generell findet sich zu den Inhalten

der Selbstbestimmungstheorie, dass für autonomieförderliche Kontexte eine motivationsfördernde Wirkung gefunden werden konnte (Rakoczy et al., 2008). Schüler\_innen zeigten sich in einer Untersuchung von Rakoczy et al. (2008) umso motivierter, je autonomieförderlicher der Unterricht gestaltet wurde. Darüber hinaus konnte gezeigt werden, dass Unterweisungen und ein Selbststudium angehender Lehrer in Bezug auf autonomieförderliche Lehrverhalten eine positive Wirkung auf die Motivation der unterrichteten Schüler hatten (Reeve, 1998; Reeve, Jang, Carell, Jeon und Barch, 2004 nach Fries, 2010, S. 153). Die Interessentheorie möchte zusätzlich die Bedeutsamkeit des Lerngegenstandes erhöhen. So kann im Zuge der Interessentheorie folgende Interventionsregel abgeleitet werden. Lehr- Lernsettings sind so zu gestalten, dass die drei psychologischen Grundbedürfnisse zu erfüllen sind. Darüber hinaus ist die subjektive Bedeutsamkeit des Lerngegenstands zu erhöhen (Schiefele, 2004). „Eine subjektive Bedeutsamkeit (kann) dadurch erreicht werden, dass der Bezug des Lerngegenstands zur Lebenswelt der Schülerin bzw. des Schülers aufgezeigt wird“ (Assor, Kaplan & Roth, 2002 nach Rakoczy et al., 2008, S. 27). Zur Selbstbestimmungstheorie bzw. zu Inhalten der Interessentheorie findet sich die evaluierte Intervention LEKOLEMO (Programm zur Förderung von Lesekompetenz und Lesemotivation) (Streblow, Holodynski, & Schiefele, 2007). Hier sollte neben den didaktischen Elementen, die die Förderung des Leseverstehens betreffen, auch die intrinsische Motivation durch das Erleben von Kompetenz, Autonomie und der Sozialen Eingebundenheit gefördert werden. Generell handelt es sich um ein Sachtexttraining, bei dem Informationen aus dem Text ermittelt, interpretiert, reflektiert und bewertet werden sollen. Darüber hinaus geht es um die Vermittlung von metakognitivem Wissen über Lern- und Leseprozesse. Durch Bearbeitung der Texte in Kleingruppen, das Klären von Verständnisschwierigkeiten und den Austausch in der Gesamtgruppe sollte das Erleben sozialer Eingebundenheit gefördert werden. Eine freie Wahl der Themen sollte Autonomieerleben ermöglichen. Reciprocal Teaching (wechselseitiges Lernen und Lehren) sollte das Kompetenzerleben der Schüler fördern, und in Anlehnung an die Interessentheorie von Krapp sollte der Bezug des Textmaterials zur Lebenswelt die erlebte Nützlichkeit erhöhen. In einer Evaluationsstudie zu LEKOLEMO von Meyer (2009) mit insgesamt 18 Schulklassen in einem Prä-Post-Treatment-Kontrollgruppen-Design zeigten sich insgesamt nur geringe Effektstärken. Eine Steigerung der Lesekompetenz zeigte sich nur für die Trainingsgruppe, in der die Implementierung des Trainings (erfasst durch hohe Trainingsakzeptanz seitens des Kollegiums, separate Durchführungsräume für die Trainingsgruppe und Sitzungen außerhalb der 5. und 6. Schulstunde) erfolgte.

Die Zielorientierungstheorie schlägt vor, Maßnahmen zu entwickeln, um die Lernzielorientierung der Schüler zu fördern. In mehreren Interventionen konnte dies über entsprechende Hinweise in der Aufgabeninstruktion erreicht werden. Hier konnte auch ein positiver Zusammenhang mit der Schülerleistung gefunden werden. Leider liegen noch keine umfassenden Interventionsmaßnahmen vor, die eine längerfristige Veränderung der Zielorientierungen nachweisen können. Meist werden diese o. a. Maßnahmen als ein Interventionsbestandteil in Kombination mit den in anderen Motivationstheorien vorgeschlagenen Maßnahmen eingesetzt (Guthrie, Wigfield & Perencevich, 2004). Kurzfristige Veränderungen der Lernzielorientierung konnten durch entsprechende Hinweise in der Aufgabenstellung erreicht werden. In Folge zeigten sich bessere Leistungen der Schüler. Eine langfristige Veränderung scheint auch möglich, wurde aber noch nicht durch wiederholte Erhebungen bestätigt. Das Training Concept-Orientated Reading Instruction (CORI; Guthrie, et al., 2004) verknüpft Annahmen der Zielorientierungstheorie mit Elementen zur Förderung des Leseengagements, von Lese- und Schreibkompetenzen sowie naturwissenschaftlichem Verständnis. Diese Intervention wurde über einen Zeitraum von drei bis zwölf Monaten täglich ca. eine Stunde bei Schüler\_innen der dritten bis fünften Klasse durchgeführt. Es wurden dabei Übungen zum Einsatz von Lesestrategien und motivationsfördernde Maßnahmen eingesetzt. Auch hier finden sich neben dem Fokus auf die Zielorientierungstheorie durch den Aspekt der Autonomieförderung Bezüge zur Selbstbestimmungstheorie. Generell wird der Versuch unternommen, mit autonomieförderlichen Maßnahmen, z. B. einer Leseübung in Einzel- oder Gruppenarbeit, zu arbeiten. Ergebnisse werden in kooperativen Arbeitsformen festgehalten und so auch Kompetenzerleben und soziale Eingebundenheit ermöglicht. Ebenso finden sich durch naturwissenschaftliche Exkursionen und Experimente zur Bearbeitung von Sachtexten Bezüge zur Interessentheorie. Eine Förderung der Motivation soll so über die ganzheitliche Auseinandersetzung in Exkursionen zu den Sachthemen (Interessentheorie) erreicht werden. Insgesamt sollte durch diese Maßnahmen die Lernzielorientierung der Schüler\_innen günstig beeinflusst werden (Zielorientierungstheorie). Die Verknüpfung dieser Elemente hat sich in Evaluationsstudien als effektiv erwiesen. In Evaluationen des Programms zeigen sich insbesondere positive Effekte auf die Lesekompetenz sowie die intrinsische und extrinsische Motivation. Auch bei Guthrie, McRae und Klauda (2007) konnte die Motivationssteigerung durch CORI gezeigt werden und eine Förderung der Neugier, Zielorientierung und der erlebten Selbstwirksamkeit der Schüler\_innen (Guthrie et al., 2007).

Das Selbstbewertungsmodell von Heckhausen (1975) hat einen erheblichen Beitrag für die Entwicklung von Interventionen geleistet. Die Einordnung in erfolgswahrscheinliche und misserfolgsmeidende Personen und in Folge deren Aufsuchen bzw. Vermeiden von Leistungssituationen, deren spezifische Attribuierung von Leistung und die unterschiedlichen Folgen für die Selbstbewertung bieten vielfältige Anknüpfungspunkte für Interventionsmaßnahmen. Da es sich um ein sich selbst stabilisierendes System handelt und keine Veränderungen durch einzelne Erfahrungen zu erwarten sind, müssen Interventionsmaßnahmen an allen drei Prozesskomponenten ansetzen. So gilt es generell, die Hoffnung auf Erfolg zu erhöhen und die Furcht vor Misserfolg zu reduzieren. So sollten Fördermaßnahmen eine realistische Anspruchsniveausetzung und eine erfolgswahrscheinliche Attributionsweise fördern womit sie zu einer positiven Selbstbewertungsbilanz beitragen würden. Es kann auch über entsprechende Aufgabengestaltung eine Leistungssituation mit einem individuell an die Schüler\_innen angepassten mittleren Schwierigkeitsgrad bereitgestellt werden. Solche Aufgaben sind sowohl herausfordernd als auch bewältigbar und legen darüber hinaus eine Attribution auf die eigene Anstrengung und ein Selbstwirksamkeitserleben nahe. Ein entsprechendes Feedbackverhalten der Lehrkraft kann eine günstige Kausalattribution nahelegen. Hierbei sollte der Schüler auch zur Reflexion ermutigt werden, welche Ursachen alternativ für den Erfolg bzw. Misserfolg verantwortlich zu machen sein könnten. Erfolge sollten internal, stabil und global attribuiert werden und so eine positive Selbstbewertung ermöglichen. Misserfolge sollten möglichst external, variabel und lokal attribuiert werden. So ist die Annahme wahrscheinlich, dass solcherart eine günstige Selbstbewertung erfolgt und günstige Entwicklungen der Erfolgswahrscheinlichkeit und eine Reduktion der Misserfolgsangst zu erwarten sind. Ergeben sollte sich daraus auch ein günstiger Einfluss auf die Leistungsmotivation. Ansatzpunkte für Interventionsmaßnahmen bietet das Modell über seine drei Komponenten Aufgabewahl- bzw. Anspruchsniveausetzung, Ursachenzuschreibung und Selbstbewertung nach leistungsbezogenen Aufgaben. Hier sollte an allen drei Elementen des Prozessmodells angesetzt werden. Bisherige Ansätze zeigen den Erfolg unterrichtsintegrierter Maßnahmen durch geschulte Lehrkräfte (Rheinberg et. al., 2001). Einfacher in den Unterricht integrierbar scheinen Veränderungen der Bezugsnorm der Lehrkraft. So haben Anregungen zur Verwendung einer individuellen Bezugsnorm breiten Einfluss auf Motivationsförderprogramme mit der Orientierung am Prozessmodell genommen. Zur Umsetzung des Selbstbewertungsmodells findet sich in der Literatur häufig eine Beschreibung einer frühen Intervention von (Krug & Hanel, 1976).

Hier sollten in einem 16-stündigen Training in der 4. Klasse (Schüler mit ungünstiger Motivationsprägung und schwache Schüler) die drei Komponenten des Modells über Geschicklichkeitsspiele gefördert werden. Die 30 ausgewählten Schüler wurden einer Interventionsgruppe, Messwiederholungsgruppe und Erwartungskontrollgruppe zugeordnet. Die Erhebung erfolgte in einem Prä-Post-Kontrollgruppen-Design mit Messungen bei allen Gruppen. In einem ersten Schritt wurden über die Spiele eine günstige Zielsetzung (Aufgabenwahl), Ursachenzuschreibung und in Folge Selbstbewertung vermittelt. In einer weiteren Trainingsphase wurden Rechen- und Rechtschreibaufgaben bearbeitet. Hier sollten dieselben Schritte wie bei den Geschicklichkeitsspielen im Sinne des Modells geleistet werden. Vor der Aufgabenbearbeitung sollten auf Arbeitsblättern bestimmte Zielsetzungen formuliert werden. Im Anschluss konnte so eine Ist-Soll-Bewertung vorgenommen und die Zielerreichung bewertet werden. Darüber hinaus sollten Ursachenerklärungen vorgenommen und die eigene Zufriedenheit mit dem Ergebnis formuliert werden. Die Arbeitsblätter konnten anschließend eingesammelt und rückblickend eine Entwicklungsbeurteilung im Sinne der individuellen Bezugsnorm vorgenommen werden. Positive Effekte zeigten sich für die realistische Zielsetzung der Schüler\_innen und eine geringere Furcht vor Misserfolg und mehr Hoffnung auf Erfolg. Die Schüler attribuierten Misserfolge stärker auf den variablen Faktor Anstrengung als auf mangelnde Fähigkeiten. Darüber hinaus zeigten sich für die Interventionsgruppe die stärksten Effekte auf einen Intelligenz- und einen Schulleistungstest. Aufgrund der geringen Stichprobengröße und des fehlenden Nachweises längerfristiger Effekte mittels Folgerhebungen ist der Befund aber nur als Hinweis für die Wirksamkeit der Intervention zu betrachten (Fries, 2010).

Die Erwartung-mal-Wert-Theorie nimmt an, dass eine Erhöhung der Erfolgserwartung und des Aufgabenwerts zu einer Steigerung der Lernmotivation beiträgt. So scheint es durchaus denkbar, dass sich Maßnahmen entwickeln lassen, die bedeutsame Aspekte der einzelnen Theorien verbinden: die Relevanz für das eigene Selbstkonzept, der subjektive Nutzen, der intrinsische Anreiz und eine Reduktion von Kosten bei der Aufgabenausführung (z. B. der Anstrengung). Ebenso ist es denkbar, dass subjektive motivationale Überzeugungen durch entsprechende unterrichtliche Maßnahme beeinflussbar sind. Auch ist vorstellbar, z. B. durch entsprechende Feedbackverfahren eine motivationsförderliche Kompetenzrückmeldung bereitzustellen (indiv. Bezugsnorm) und somit die Erfolgserwartung der Schüler zu verbessern.

Leider sind spezifische Interventionen im Rahmen dieser Theorie bislang unterblieben. Allerdings sind die einzelnen Elemente auch in anderen Grundlagentheorien zu finden und so können Maßnahmen, die auf diesen basieren, die Erfolgserwartung und den Aufgabenwert ebenso steigern. So schreibt auch Fries (2010): „Das Modell verweist zwar auf vielfältige Größen, die der Erfolgserwartung und dem Aufgabenwert vorgeordnet sind, doch lassen sich aus dem Modell keine Empfehlungen ableiten, was durch eine Intervention geändert werden sollte und wie“ (S. 154).

Betrachtet man die Theorien zur Lernmotivation und ihre Interventionsansätze, so scheinen sie vielversprechende Wege zur Verbesserung motivationaler Faktoren für Schüler aufzuzeigen. Über alle Theorien hinweg lassen sich recht einfach in den Unterrichtsalltag integrierbare Elemente finden. Sei es durch die Verwirklichung der drei psychologischen Grundbedürfnisse, die Förderung des erlebten Nutzens, die Erhöhung des Aufgabenwerts oder die Verwendung einer individualisierten Bezugsnorm im Rückmeldeverhalten von Lehrkräften. Neben diesen Maßnahmen, die ganz generell in den Unterricht und in das Lehrerhandeln einfließen können, gibt es die Möglichkeit der Entwicklung von Manualen, die zur systematischen Förderung der Lernmotivation von Schülern beitragen können. Diese Manuale können handlungsleitend und unterstützend zur Integration in den Unterricht konzipiert sein bzw. durch Schulungen in den Basistheorien der umsetzenden Lehrkräfte ergänzt werden. Hier gibt es im Bereich der Motivationsförderung noch Nachholbedarf. Die motivationsförderlichen Interventionsprogramme sind noch nicht so weit offen zugänglich, wie das in anderen Teilbereichen der Psychologie der Fall ist. So findet sich ein breiteres Angebot von evaluierten Interventionsmanualen beispielsweise zu Trainingsmaßnahmen zur sozialen Kompetenz, der jeweiligen Fachkompetenzbereiche oder kognitiver Trainingsprogramme. Darüber hinaus stellt sich die Frage nach der Implementation bzw. deren Dokumentation bei allen Interventionsbemühungen im Unterricht. Es stellt sich auch im Sinne der Implementation die Frage, wie Interventionen ressourcensparend in Lehr-Lernsettings umgesetzt werden können, sei es durch die Praxisnähe der Maßnahmen und somit deren erleichterte Umsetzbarkeit oder auch durch die Art der Intervention. So ist durch indirekte Interventionen, also durch Schulung der Lehrkräfte und die Umsetzung in mehreren Klassen, eine große Zahl an Schülern zu erreichen. Dieser Ansatz wurde auch bei der hier dargestellten Intervention verfolgt. Im Sinne einer indirekten Intervention wurden die Lehrkräfte zu Inhalten der Selbstbestimmungstheorie nach Deci und Ryan (1985; Ryan & Deci, 2000, 2002) geschult und konnten über Manuale durch

Umstellungen in den Arbeitsformen und den verwendeten Materialien die Zielsetzung Kompetenzerleben, Autonomieerleben und soziale Eingebundenheit verfolgen.

## **2.2 Die Selbstbestimmungstheorie und ihre grundlegenden Annahmen zur Förderung einer selbstbestimmten Lernmotivation**

Die Selbstbestimmungstheorie der Motivation von Deci und Ryan (1985, 2000) beschäftigt sich mit der Frage, wie eine selbstbestimmte Lernmotivation gefördert werden kann. Sie sieht die Befriedigung der psychologischen Grundbedürfnisse als Grundlage einer selbstbestimmten Lernmotivation und nimmt an, dass jedes Individuum einen inneren Antrieb (intrinsische Motivation, Neugier, psychologische Grundbedürfnisse) zu Wachstum als motivationale Grundlage für eine hochqualitative Involviertheit bei der Lernaktivität und Schulanpassung hat (Deci & Ryan, 1985, 2000; Reeve, Deci & Ryan, 2004; Ryan & Deci, 2000, 2002; Vansteenkiste, Niemiec & Soenens, 2010). Jüngere Befunde legen nahe, dass die aktive Involviertheit beim Lernen, wie sie bei einer intrinsischen Motivation zu beobachten ist, positive Effekte auf Lernen und Leistung hat. Veränderungen in der aktiven Involviertheit verändern die Motivation und das Lernumfeld. Darüber hinaus beeinflusst der Kontext auch das Engagement des Schülers (Reeve, 2012).

Neben der Annahme, dass ein Individuum die genannten Wachstumsaspekte schon mitbringt, beschäftigt sich die Selbstbestimmungstheorie damit, wie sich durch die Erfüllung von angeborenen psychologischen Grundbedürfnissen eine intrinsische Motivation einstellen kann und wie ursprünglich fremdbestimmte, extrinsisch motivierte Handlungen schrittweise als selbstbestimmte Handlungen internalisiert werden können. So kann laut der Selbstbestimmungstheorie von Deci und Ryan (1985, 2000) eine intrinsische Motivation über die Erfüllung von angeborenen psychologischen Grundbedürfnissen nach Autonomie- und Kompetenzerleben entstehen. Über die Erfüllung des Bedürfnisses nach Sozialer Eingebundenheit können nach Deci und Ryan aber auch fremdbestimmte Handlungen (*externale Regulation*) über bestimmte Entwicklungsstadien der *introjizierten*, der *identifizierten* und der *integrierten Regulation* in das Selbst integriert und als selbstbestimmt erlebt werden (Deci & Ryan, 2000).

Die **Selbstbestimmungstheorie der Motivation** von Deci und Ryan (1993, 2000; Ryan & Deci, 2000b) umfasst sechs Teiltheorien:



1. Cognitive Evaluation Theory (CET)
2. Organismic Integration Theory (OIT)
3. Causality Orientations Theory (COT)
4. Basic Psychological Needs Theory (BPNT)
5. Goal Contents Theory (GCT)
6. Relationships Motivation Theory (RMT)

Die **Cognitive Evaluation Theory** spezifiziert die Bedingungen, die intrinsische Motivationsprozesse der Lernenden fördern und begünstigen können. Sie erläutert zudem, welche Bedingungen ihnen entgegenwirken und die intrinsische Motivation unterminieren, wobei sie davon ausgeht, dass Ereignisse, die Einfluss auf das Erleben von Autonomie und Kompetenz haben, auch die intrinsische Motivation beeinflussen. So haben alle externalen Ereignisse einen Kontroll- und Informationsaspekt für das Individuum (Deci & Ryan, 1985). Der Effekt auf die intrinsische Motivation wird durch die relative Salienz des kontrollierenden versus informativen Aspekts bestimmt.

So kann ein externes Ereignis (z. B. Feedback, Noten, Tests) eine Kompetenzrückmeldung bereitstellen und deshalb einen *Informationsaspekt* für den Lernenden aufweisen. Aber auch dies kann als Kontrolle erlebt werden. Eine Verstärkung sollte deshalb möglichst eine Rückmeldung im Hinblick auf Kompetenz und verbesserte Arbeitsweise beinhalten und somit nach der erbrachten Leistung eingesetzt werden. Eine in Aussicht gestellte Belohnung für erwünschtes Verhalten hat allerdings für den Schüler einen deutlicheren *Kontrollaspekt*, da er damit zu einer bestimmten Verhaltensweise bewegt werden soll. Solch gestaltete externale Ereignisse schwächen die intrinsische Motivation (*Korruptionseffekt*), während nicht kontrollierende, informative Rückmeldungen bzw. Verstärker die intrinsische Motivation sichern. Der informative Aspekt einer Rückmeldung kann auch als Förderung des wahrgenommenen Autonomieerlebens betrachtet werden, da der Schüler nicht in eine bestimmte Richtung des Verhaltens gedrängt wird. Das Ausmaß an Autonomieunterstützung bzw. an erlebter Kontrolle kann als Kontext für das jeweilige externe Ereignis fungieren und so je nach Ausrichtung Einfluss auf die intrinsische Motivation nehmen.

Schulisches Lernen ist meist durch extrinsische Zielsetzungen beeinflusst. Die **Organismic Integration Theory** erläutert, unter welchen Gegebenheiten Lernende extrinsische Ziele internalisieren (bzw. nicht internalisieren) und extrinsische motivationale Prozesse in das

Selbstmotivationssystem integrieren (Reeve, 2012). Individuen sind bestrebt, soziale Normen, Regeln und Verhaltensweisen zu internalisieren. Zur Steigerung der eigenen Kompetenz und zur Intensivierung sozialer Beziehungen suchen sie aktiv hierzu geeignete Möglichkeiten. So nimmt die Organismic Integration Theory an, dass Individuen von Natur aus willentlich Anteile ihres sozialen Umfelds internalisieren und einige Werte als angeeignete Motivationen in das Selbstmotivationssystem integrieren. Je mehr es gelingt, extrinsische Ziele zu internalisieren, zu integrieren, desto mehr Autonomieerleben und in Folge positive Leistungsfähigkeit im betreffenden Bereich sind möglich (Ryan, 1993; Ryan & Connell, 1989). Verschiedene Arten extrinsischer Motivation sind mit einem unterschiedlichen Ausmaß an wahrgenommenem Autonomieerleben verbunden und Lernende können sich so in unterschiedlichem Grad als Verursacher ihrer Verhaltensweisen und Überzeugungen erleben.

Nach Ryan & Deci (2000a) können die verschiedenen Arten der Motivation auf einem unipolaren Kontinuum dargestellt werden:

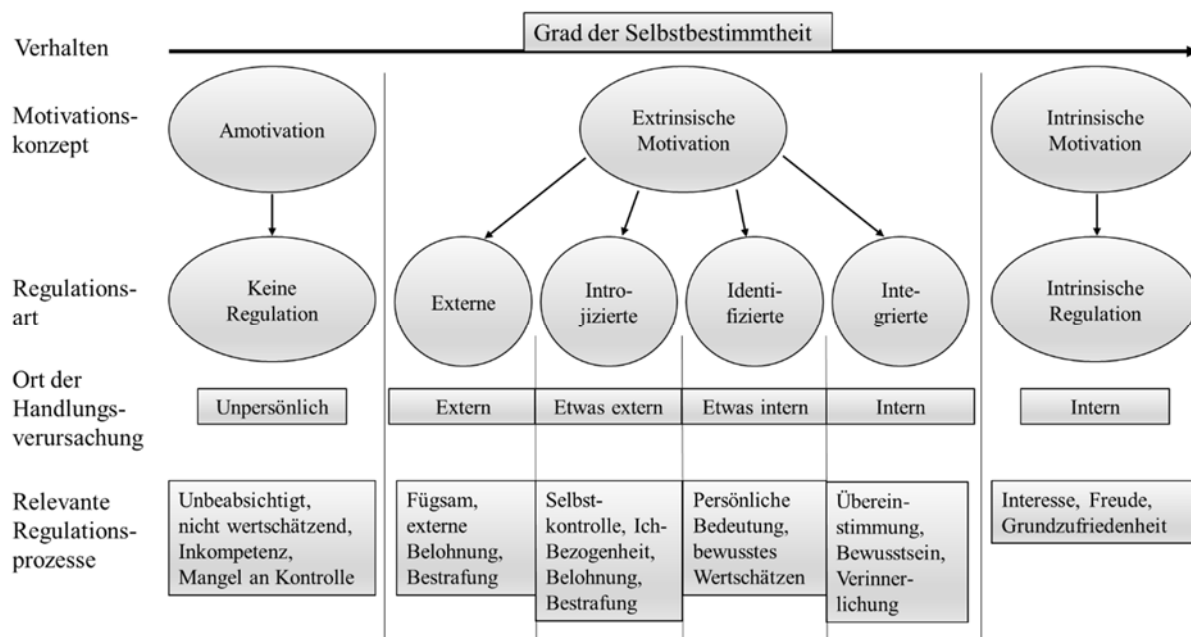


Abbildung 5. Eine Taxonomie menschlicher Motivation. Aus Intrinsic and extrinsic motivations: Classic definitions and new directions von Ryan und Deci (2000a). *Contemporary Educational Psychology*, 25(1), 54–67.

Amotivation ist ein Zustand ohne eigene Intention zur Handlung. Eine Handlung erfolgt somit entweder unbeabsichtigt und ohne Wertschätzung oder die Überzeugung, kompetent zur Aufgabenbewältigung zu sein, oder mit der Überzeugung, das beabsichtigte Ziel nicht erreichen zu können.

Bei der externalen Motivation handelt es sich um die am geringsten autonome Form der Motivation da sich das Individuum bei externaler Regulation den Anforderungen nur fügt, um eine Belohnung zu erhalten oder eine Strafe zu umgehen. Die persönliche Wertschätzung des Verhaltens ist sehr gering.

Eine weitere Art der kontrollierten Motivation ist die Form der introjizierten Regulation. Der Schüler möchte externe Anforderungen erfüllen, um Schaden am eigenen Selbstwert abzuwenden bzw. um diesen in den Augen von anderen zu bestätigen. Dadurch soll Schuld oder Scham vermieden werden.

Um eine annähernd selbstbestimmte Form der Motivation handelt es sich bei der identifizierten Regulation. Bei dieser werden willentlich externaler Ziele transformiert und es findet eine Identifikation mit diesen Zielen statt, da persönliche Bedeutung, Wertschätzung, ein Gefühl der Wahl und ein persönliches Einverständnis vorliegen.

Die integrierte Regulation ist die am meisten selbstbestimmte Form der extrinsischen Motivation. Die identifizierte und die intrinsische Motivation haben eine interne Perspektive und der Ort der Verursachung liegt in der Person selbst (deCharms, 1968). Dennoch unterscheidet sich die integrierte Motivation von der intrinsischen Motivation insofern, als dass sie auf Relevanz der Aufgabe basiert und ein Bewusstsein und Reflexion erfordert, während die intrinsische Motivation allein aus der Aufgabe und an dem Interesse daran entsteht (Ryan & Deci, 2000a, S. 61 ff.).

Die **Causality Orientations Theory** beschäftigt sich mit interindividuell unterschiedlichen Orientierungen, die eine Initiierung von Aktivitäten im Klassenraum beeinflussen. So unterscheiden sich Schüler\_innen im Hinblick darauf, ob sie sich dabei überwiegend von dem Interesse, befürworteten Werten und persönlichen Zielen (autonome selbstbestimmte Führung) oder vielmehr sozialen Vorschriften, verinnerlichten Anforderungen oder Anreizen aus der Umwelt leiten lassen (kontrollierende Führung). Im Gegensatz zu Persönlichkeitseigenschaften handelt es bei einer kontrollierenden bzw. einer selbstbestimmten Kausalorientierungsdisposition um oberflächliche individuelle Unterschiede, die von der Sozialisationserfahrung beeinflusst und veränderbar sind (Vansteenkiste et al., 2010). Darüber hinaus nimmt die Causality Orientations Theory an, dass beide Dispositionen im Individuum vorhanden sind und sich nur in der relativen Ausprägung unterscheiden. Eine hohe Ausprägung einer selbstbestimmten Kausalorientierung hat zur Folge, dass Lernende auf intrinsische Motivation,

integrierte Regulation und identifizierte Regulation vertrauen, während Lernende mit einer überwiegend kontrollierenden Kausalorientierung eher extrinsische, introjizierte Motivationsquellen bevorzugen. Lernende mit einer selbstbestimmten Kausalorientierung sind in der Tendenz selbstbewusster, weisen eine reifere Selbstbildentwicklung und weniger Selbstbeeinträchtigung auf. Dominieren kontrollierende Kausalorientierungen, tendieren Lernende zu mehr Erleben von täglichem Stress, Defensivität und ein auf äußere Einflüsse bezogenes Selbstbewusstsein (Reeve, 2012).

Nach der **Basic Psychological Needs Theory** ist die Erfüllung der drei psychologischen Grundbedürfnisse des Individuums nach Autonomieerleben, sozialer Eingebundenheit und Kompetenzerleben die Grundlage der intrinsischen Motivation und autonomen Selbstregulation (Bieg & Mittag, 2011) und begünstigt psychisches und soziales Wohlbefinden. Sie beschreibt das **Bedürfnis nach Autonomie** als das Verlangen, das eigene Verhalten als von sich selbst ausgehend und von sich selbst unterstützt zu erleben, das **Bedürfnis nach sozialer Eingebundenheit** als den Wunsch enge soziale emotionale Beziehungen zu etablieren und Beziehungen zu anderen zu sichern. Bei dem **Bedürfnis nach Kompetenzerleben** geht es darum, im Rahmen eigener Ziele und Interaktionen mit der Umwelt effektiv zu sein.

Die Befriedigung der psychologischen Grundbedürfnisse unterstützt eine aktive Involviertheit des Lernenden und das Versagen dieser Bedürfnisbefriedigung sagt verschiedene Anzeichen von Unzufriedenheit vorher (Deci & Ryan, 2000; Reis, Sheldon, Gable, Roscoe & Ryan, 2000). Die Basic Psychological Needs Theory bietet so eine Erklärung dafür, warum sich Schüler\_innen aktiv beim Lernen engagieren bzw. dies bei anderer Gelegenheit nicht tun. Die Theorie liefert Hinweise darauf, welche Bestandteile des Unterrichts und der Klassenumgebung sich günstig bzw. unterminierend auf die Befriedigung der Grundbedürfnisse und somit auf das Engagement der Schüler\_innen auswirken. Die Theorie nimmt an, dass die Befriedigung der drei Grundbedürfnisse nach Autonomie, sozialer Eingebundenheit und Kompetenzerleben zu einem besseren Gelingen von Lernprozessen beiträgt und eine selbstbestimmte Lernmotivation begünstigt. Verhalten kann dann als autonom oder selbstbestimmt bezeichnet werden, wenn Schüler\_innen frei bei der Aufgabenwahl sind, ihre Handlungen als selbstgesteuert erleben, das Verhalten aus inneren Ressourcen angetrieben wird und man so auch von volitionalen Handlungen sprechen kann (Mittag et al., 2009). Die Förderung des Autonomieerlebens durch unterstützende Maßnahmen der Lehrkraft hat, so zeigen Längsschnittstudien,

einen günstigen Einfluss auf das Kompetenzerleben der Schüler und deren intrinsische Motivation (Guay, Boggiano & Vallerand, 2001; Skinner & Belmont, 1993; Vallerand, Fortier & Guay, 1997). Sich als kompetent und autonom erlebende Schüler\_innen weisen darüber hinaus mehr positive Emotionen in Lernsituationen auf, sind neugieriger, engagierter und aktiver im Lernprozess. Kontrollierende Bedingungen und Begrenzungen bei Wahl- und Beteiligungsmöglichkeiten sind hingegen oft mit Verschlechterung in den Lernleistungen assoziiert (Boggiano, Flink, Shields, Seelbach & Barrett, 1993). Um sich als kompetent zu erleben, müssen Bedingungen geschaffen werden, die eine Passung von Anforderungen und Fähigkeiten ermöglichen und dem Lernenden so die Gelegenheit geben, sich als selbstwirksam zu erleben, bei der Handlungsausführung Vertrauen in die eigenen Fähigkeiten zu entwickeln und deren Erweiterung ermöglichen. Die Bewältigung von Lernaufgaben erzeugt positive Rückmeldungen im Hinblick auf die eigene Leistungsfähigkeit und begünstigt eine intrinsische Motivation (Mittag et al., 2009). Hierzu gehört neben der Passung von Anforderung und Fähigkeit auch eine individuelle und informative Leistungsrückmeldung im Lernprozess (Deci, 1975). Menschen haben das Bedürfnis nach emotionaler Stabilität und fürsorglichen Beziehungen. Sichere Bindungen und positive Beziehung in der Klasse tragen so zur Befriedigung des psychologischen Grundbedürfnisses nach sozialer Eingebundenheit bei (Baumeister & Leary, 1995; Deci & Ryan, 1991). Günstig wirkt sich ein hohes Maß an Fürsorglichkeit (Skinner & Belmont, 1993), Empathie und Wertschätzung (Reis et al., 2000) seitens der Lehrkraft aus (Bieg & Mittag, 2010). Durch das Erleben sozialer Eingebundenheit können Werte beziehungsweise Handlungsziele anderer Personen verinnerlicht und die Identifikation mit dem Lehrgegenstand erhöht werden, so dass die Entwicklung einer selbstbestimmten Lernmotivation unterstützt wird. Neben der Lehrer-Schüler-Beziehung kommt auch dem Beziehungsgeflecht innerhalb der Lerngruppe eine hohe Bedeutung zu, die durch kooperative Lern- und Arbeitsformen unterstützt werden kann und das Erleben von sozialer Eingebundenheit in der Klasse begünstigt.

Ob sich Wohlbefinden und eine Befriedigung der psychologischen Grundbedürfnisse durch das Erreichen eines Ziels einstellen, ist abhängig von der Art des Ziels. So sollte sich beim Erreichen intrinsischer Ziele eher Wohlbefinden einstellen als bei Erreichen extrinsischer Zielsetzungen. Wie intrinsische und extrinsische Ziele die Motivation und das Wohlbefinden beeinflussen, damit beschäftigt sich die **Goal Contents Theory** als Teiltheorie der Selbstbestimmungstheorie. In der Theorie wird dargelegt, dass das Verfolgen intrinsischer Ziele die

inneren motivationalen Ressourcen aktiviert und begünstigt und so im Gegensatz zur Verfolgung extrinsischer Ziele ein vertiefendes Lernen, eine bessere Performanz, größeres psychisches Wohlbefinden und auch die Ausdauer bei einer Leistungsanforderung günstig beeinflusst (Vansteenkiste, Simons, Lens, Sheldon & Deci, 2004; Vansteenkiste, Lens & Deci, 2006). Die Orientierung an extrinsischen Zielen hingegen trägt nicht zur Befriedigung der psychologischen Grundbedürfnisse bei, begünstigt Unwohlsein und unterminiert das Lernen – und das gilt selbst dann, wenn extrinsische Ziele erreicht werden (Reeve, 2012).

Bei der **Relationships Motivation Theory (RMT)** wird angenommen, dass soziale Verbundenheit über die Zugehörigkeit zu engen Freunden, Partnern oder sozialen Gruppen und ein bestimmtes Maß an sozialer Interaktion innerhalb dieser Gefüge, essenziell für die Anpassung und das Wohlbefinden des Individuums sind. Entscheidend für das Erleben sozialer Eingebundenheit und einer hochqualitativen Beziehung ist hierbei, dass die Interaktionspartner in der Beziehung Autonomie erleben und autonomieunterstützend agieren. Kontrollerleben, Objektivierung und beständige Aufmerksamkeit wirken nicht nur dem Autonomieerleben entgegen, sondern auch dem Erleben sozialer Eingebundenheit. Dies resultiert in einer geringeren erlebten Beziehungsqualität. So kann Autonomieerleben auch in einer Arbeitsumgebung das Wohlbefinden und die Leistungsfähigkeit unterstützen. „Further studies have shown that, in current situations such as school and workplaces, when the environment is autonomy supportive, people report higher levels of psychological need satisfaction, which has positive links to engagement, performance, and psychological well-being (Baard, Deci, & Ryan, 2004; Deci, Ryan, Gagné, Leone, Usunov, & Kornazheva, 2001)“ (Deci & Ryan, 2014, S. 60). Dies zeigt sich beispielsweise bei Untersuchungen zur freiwilligen Partnerwahl bei der Aufgabenbewältigung. Diese Arbeitsgruppen zeichnen sich durch eine höhere Effektivität und Nähe aus (Weinstein, Hodgins & Ryan, 2010).

Generell liefert insbesondere die Basic Psychological Needs Theory eine Erklärung für motiviertes Lernen und für bessere Lernleistungen, indem sie diese auf eine Steigerung der intrinsischen Motivation durch die Befriedigung der drei psychologischen Grundbedürfnisse zurückführen. Darüber hinaus liefert die Organismic Integration Theory Hinweise darauf, welche Maßnahmen eine Integration extrinsischer Ziele begünstigen. Beide Theorien können so als Grundlage für die Entwicklung von Interventionsmaßnahmen dienen, die eine Förderung der selbstbestimmten Lernmotivation beabsichtigen.

Im Folgenden soll nun anhand der durchgeführten Intervention dargestellt werden, inwieweit sich selbstbestimmte Lernmotivation in Schule und Unterricht fördern lässt. Dabei sollen die Maßnahmen näher betrachtet und begründet sowie deren Umsetzung erläutert werden.

### **2.3 Möglichkeiten der Förderung einer selbstbestimmten Lernmotivation in Schule und Unterricht**

In Schule und Unterricht stellt sich insbesondere in Lernfeldern die für den individuellen Schüler nicht durch eine hohe intrinsische Motivation geprägt sind, immer wieder die Frage nach motivationsförderlichen Bedingungen. Darüber hinaus gilt es Voraussetzungen zu schaffen, die die Schülermotivation über einen längeren Zeitraum stabilisieren.

Die Selbstbestimmungstheorie von Deci und Ryan (1985, 2000) nimmt an, dass das Individuum proaktiv Möglichkeiten aufsucht, um seine Kompetenzen im sozialen Umfeld zu steigern und intensivere Beziehung zu anderen herzustellen. Dieses Bedürfnis kann durch ein unterstützendes soziales Umfeld begünstigt werden.

Sie nimmt darüber hinaus an, dass Individuen willentlich Aspekte ihres sozialen Umfelds internalisieren und einige Werte und Verhaltensweisen als angeeignete Motivationen integrieren. Eine weitere zentrale Annahme der Selbstbestimmungstheorie ist, dass Entwicklungs- und Lernprozesse umso besser gelingen, je mehr das Bedürfnis nach Autonomie, Kompetenzerleben und sozialer Eingebundenheit berücksichtigt und befriedigt wird (Deci & Ryan, 1993).

Dies gilt es in der praktischen Unterrichtsgestaltung umzusetzen. Ermöglicht werden soll das Erleben von **Autonomie** über den Einbezug in Entscheidungsprozesse im Klassenalltag, aber auch bei der Gestaltung von Lernzielen, Lerngegenständen und Lernaktivitäten mit Handlungsspielräumen, Unterstützung von selbstreguliertem Lernen (self-direction) (Reeve, Nix & Hamm, 2003) und der Bereitstellung von Materialien und Bedingungen, die eine Selbstbewertung und Fehlerkorrektur ermöglichen. Die Mitwirkung bei der Auswahl von Lernaufgaben mit Alltagsbezug mit Wahl- (Katz & Assor, 2007) und Partizipationsmöglichkeiten im Unterricht (u.a. auch das gemeinsame Aushandeln von Klassenregeln), begründete Erklärungen (explanatory rationales) (Reeve, Jang, Hadre & Omura, 2002), Wahrnehmung von Gefühlen (acknowledgement of feelings) (Koestner, Ryan, Bernier & Holt, 1984), Ermuti-

gung (Reeve & Jang, 2006) sowie positives Feedback sind Maßnahmen, um das Autonomieerleben der Schüler im Unterricht zu begünstigen. Als Autonomiehindernisse im Klassenzimmer werden Kontrolle und Überwachung (Leppner & Greene, 1975), Deadlines (Amabile, DeJong & Lepper, 1976), aufgezwungene Regeln und Grenzen (Koestner et al., 1984), aufgezwungene Ziele (Mossholder, 1980), Anweisungen und Kommandos (Reeve & Jang, 2006), Wettbewerb (Deci, Betley, Kahle, Abrams, & Porac, 1981) und Bewertungen (Ryan, 1982) genannt (Reeve, 2012, S. 156).

Als weiteres psychologisches Grundbedürfnis, dessen Befriedigung im Unterricht verwirklicht werden sollte, führen Deci und Ryan (1993) das **Kompetenzerleben** an. Dies ist das Bedürfnis, im Rahmen eigener Ziele und Interaktionen mit der Umwelt effektiv zu sein, eigene Fähigkeiten bzw. Ressourcen zu erweitern, zu üben und umweltbezogene Anforderungen aufzusuchen und zu bewältigen (Deci, 1975). Als unterstützende Maßnahmen werden individualisierte und informierende Rückmeldungen im Sinne einer Verwendung der individuellen und sachlichen Bezugsnorm (Mittag et al., 2009) genannt. Weitere genannte Maßnahmen sind häufiges und positives Feedback, eine differenzierte Aufgabenstellung (Deci, 1975) mit verständlichen und strukturierten Anleitungen, individuell an das jeweilige Kompetenzniveau angepasste Aufgabenschwierigkeit (Individualisierung der Lern- und Leistungsanforderungen, Aufgaben mit subjektiv optimalem Anforderungsniveau, Unterstützung bei schwieriger Problemlösung, Aufgabenstellungen, die über die fachspezifischen Lerninhalte hinausgehen und das Zeigen vielfältiger Fähigkeiten ermöglichen (z. B. die Koordination der Arbeitsgruppe). Hierzu können kooperative Lerngelegenheiten und Möglichkeiten zum wechselseitigen Lehren- und Lernen (Peer-Teaching) geschaffen werden. Im Weiteren kann eine generelle Verbesserung der Transparenz von Leistungsanforderungen (z. B. durch Transparenzpapiere) und der Bewertung erbrachter Leistung zu einem erhöhten Kompetenzerleben des Schülers beitragen. Hierzu werden Maßnahmen zur Trennung von Lern- und Leistungsräumen genannt. Dies kann über die Einführung spezieller Lernmöglichkeiten erfolgen, in denen Schüler ohne Leistungsbeurteilung ihr bestehendes Wissen erweitern können. Um mehr Transparenz der Leistungsanforderungen zu erreichen, können beispielsweise Transparenzpapiere, die eine genaue Beschreibung der Anforderungen und Benotungen beinhalten, eingesetzt werden (Bieg & Mittag, 2010).



Die Schule ist ein Raum emotionaler und sozialer Beziehungen. Die Qualität sozialer und emotionaler Beziehungen im Sinne einer Befriedigung des psychologischen Grundbedürfnisses nach **sozialer Eingebundenheit** gilt es zu fördern bzw. zu erhalten. Dies gilt über die Schüler-Schüler-Beziehung hinaus auch für die positiv zu gestaltende Lehrer-Schüler-Beziehung. So Bieg und Mittag (2010): „Positive Lehrer-Schüler-Beziehungen finden sich insbesondere bei fürsorglichen Lehrkräften, die von den Schüler\_innen als empathische Ansprechpartner wahrgenommen werden. Durch die damit einhergehenden positiven Emotionen identifizieren sich die Schüler\_innen mit dem Lerngegenstand und entwickeln so eine eher selbstbestimmte Lernmotivation“ (S. 188 f.).

So nehmen Deci und Ryan an, dass die Internalisierung fremdbestimmter Handlungen durch das Bedürfnis nach sozialer Eingebundenheit ermöglicht wird. Dieses Bedürfnis reflektiert das Bedürfnis, emotional mit anderen verbunden und interpersonell in warme, fürsorgliche und responsive Beziehungen eingebunden zu sein (Deci & Ryan, 1991). Schüler erfahren Befriedigung dieses Bedürfnisses, indem sie mit anderen in authentischer, fürsorglicher und wechselseitiger Weise verbunden sind (Ryan, 1993). Die Ermöglichung einer solchen Erfahrung kann durch kooperative Lern- und Arbeitsformen, die die Schüler-Schüler-Beziehung fördern, begünstigt werden. Durch differenzierte, individuelle Aufgabenbetreuung und Beachtung kann vom Schüler eine wertschätzende Lehrerfürsorglichkeit erlebt und somit das Gefühl der sozialen Eingebundenheit unterstützt werden (Reeve, 2012).

Die genannten Maßnahmen sollen durch den Einbezug in Schule und Unterricht die selbstbestimmte Lernmotivation des Schülers fördern. Es stellt sich die Frage, ob und inwieweit eine strukturierte und konzeptualisierte Umsetzung dieser Maßnahme im täglichen Unterrichtsgeschehen möglich ist. Im Hinblick auf diese Fragestellung und die daraus ableitbare Zielsetzung wurde von Mittag et al. (2009) eine Programmkonzeption vorgenommen.

## **2.4 Interventionsprogramm zur Förderung der selbstbestimmten Lernmotivation**

Das vorzustellende Interventionsprogramm von Mittag et al. (2009) zielt darauf ab, die selbstbestimmte Lernmotivation der Schüler\_innen zu steigern. Es basiert auf der Selbstbestimmungstheorie von Deci und Ryan (1993), insbesondere der Teiltheorie *Basic Psychological Needs Theory*, und bezieht zusätzlich Ansätze aus der *Cognitive Evaluation Theory* mit ein.

Die Intervention folgt einem theoriegeleiteten Modell, aus dem Unterrichtshandeln und Methoden für die Praxis ableitbar sind (Bieg & Mittag, 2010, S. 189). Abbildung 6 zeigt die Ebenen der Programmkonzeption auf.

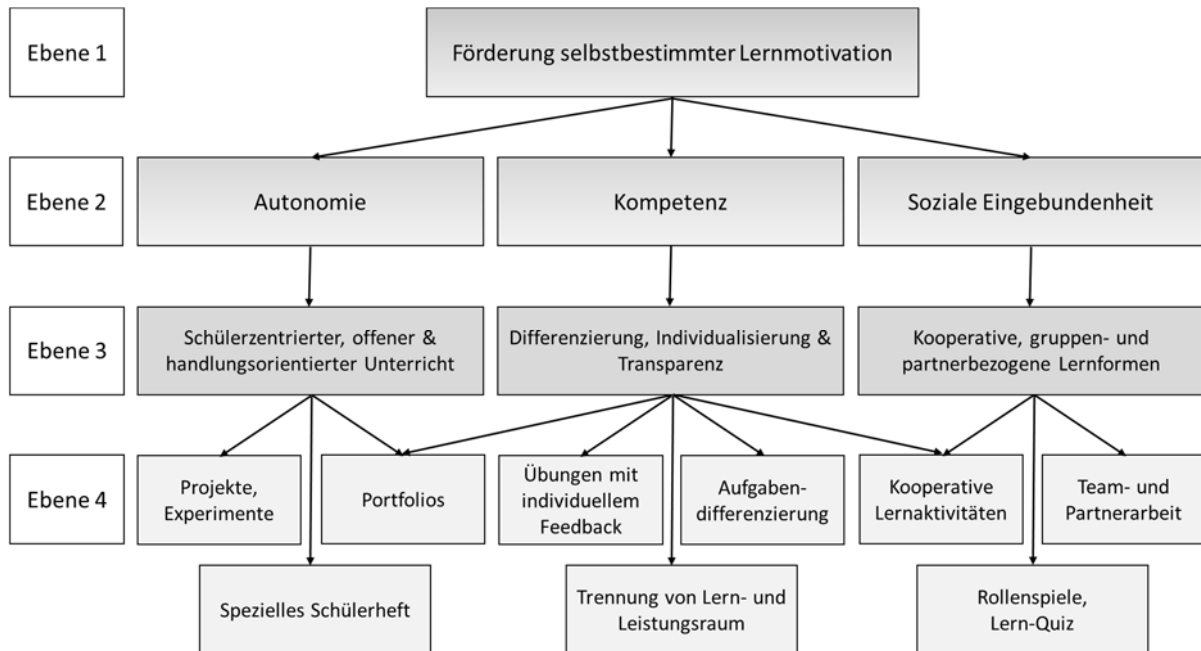


Abbildung 6. Förderung der selbstbestimmten Lernmotivation. Aus Zur Bedeutung der Lehrer-Schüler-Beziehung für die Wahrnehmung der Unterrichtsqualität und die Lernmotivation von W. Mittag (2011, Mai). *Unveröffentlichte Präsentation an der Universität Erfurt.*

Durch die Verwirklichung von Autonomie, Kompetenzerleben und sozialer Eingebundenheit (Ebene 2) soll das Interventionsziel der Steigerung einer selbstbestimmten Lernmotivation (Ebene 1) erreicht werden.

Aus den drei Grundbedürfnissen leiten sich verschiedene methodisch-didaktische Konzepte und Lehrstrategien ab (Ebene 3). Auf einer untersten Ebene finden sich konkrete unterrichtliche Maßnahmen, die im Hinblick auf die Zielerreichung förderlich sein sollen (Bieg & Mittag, 2010).

Unterstützende Bedingungen in Schule und Unterricht, insbesondere die positive Schüler-Lehrer-Beziehung, aber auch Schüler-Eltern-Beziehung, können günstig auf die Schülermotivation und das Schülerengagement wirken. So weisen Bieg und Mittag (2010) darauf hin, „[...] dass durch die Unterstützung der Grundbedürfnisse intrinsisch motiviertes und nachhaltiges Lernen gefördert werden kann (z. B. Black & Deci, 2000; Reeve & Jang, 2006)“ (S. 189). Dem entsprechend betont Reeve (2012), dass Lehrkräfte bzw. der Klassenkontext

die Motivation bzw. das Engagement so unterstützen sollten, dass eine hochqualitative Motivation bzw. ein hochqualitatives Engagement entstehen kann.

Eine hochqualitative Motivation wird angenommen, wenn Schüler\_innen möglichst selbstbestimmt und intrinsisch motiviert ihre Lernaktivitäten aufnehmen. In diesem Sinne war das allgemeine Interventionsziel die Unterstützung einer selbstbestimmten Lernmotivation. Die sollte im weitesten Sinne über methodisch-didaktische Lehrstrategien zur Unterstützung der drei Grundbedürfnisse im Unterricht und Medien und Arbeitsformen erreicht werden (Bieg & Mittag, 2010).

Für das Fach Deutsch der 8. Klasse wurden in Zusammenarbeit zwischen der Fachdidaktik Deutsch und Psychologie zwei manualisierte Unterrichtskonzepte entworfen, die sich thematisch an dem Lehrplan der Jahrgangsstufe orientieren. In einem ersten Schritt wurden interventionsbezogene Unterrichtsmethoden bzw. Unterrichtsmedien im Hinblick auf die Vorgaben des theoretischen Modells, aber auch die Bedingungen in der Praxis ausgewählt. In einem zweiten Schritt wurden eine Pilotierung in der Praxis und eine formative Evaluation vorgenommen. Die entwickelten Interventionsmaßnahmen wurden im Hinblick auf ihre Umsetzbarkeit, das anvisierte Interventionsziel und ihre Wirksamkeit beurteilt (Mittag et al., 2009).

Die auf den Grundlagen der Selbstbestimmungstheorie basierende Intervention wurde im Rahmen einer indirekten Intervention durch die Schulung der Lehrkräfte in die gewohnte Lernumgebung der Schüler eingebracht. Die Schulung der Lehrkräfte erfolgte im Hinblick auf die im theoretischen Grundmodell explizierten Ziel- und Förderbereiche. Diese werden unter Bezugnahme auf die psychologischen Grundbedürfnisse näher ausgeführt bzw. die Umsetzung und angestrebte Verwirklichung durch verschiedene konkrete Handlungen im Unterricht näher dargelegt.

### **Folgerungen von Unterrichtsempfehlungen im Hinblick auf die Interventionsziele**

Im Hinblick auf die verschiedenen Förderbereiche wurden mithilfe eines Lehrermanuals Materialien und Arbeitsformen in die Unterrichtseinheiten eingebracht. Dies sollte den Schülern die Möglichkeit zu geben, die behandelten Themen bzw. Aufgaben/Experimente selbstbestimmt bearbeiten zu können, wobei ihnen die Wahl der Inhalte und Art der Aufgaben, des Schwierigkeitsgrades und die Zeitplanung überlassen wurden. Zudem wurde eine Entscheidung zwischen Einzel-, Partner- oder Gruppenarbeit freigestellt. Ebenso sollte die Möglich-

keit einer selbstständigen Zielsetzung und Einteilung des Lernstoffs gegeben sein. Die Lehrkraft hatte nur in einer beratenden Funktion und im Sinne der Unterstützung des Erlebens sozialer Eingebundenheit zu agieren, konnte aber auf Verlangen der Schüler individuelle Rückmeldungen anbieten. Die Aufgaben sollen einen konkreten Alltagsbezug bzw. Verweise zum Alltag aufweisen. Daneben waren neben verbindlichen Aufgaben auch Aufgaben und Experimente mit freier Wahlmöglichkeit enthalten. So sollte eine autonome Gestaltung des eigenen Lernwegs für die Schüler eröffnet werden. Für den Förderbereich Autonomieerleben wurden für jede Unterrichtseinheit und jedes Fach spezielle Schülerarbeitsmappen eingesetzt, die Informationen zur selbstständigen Bearbeitung der Aufgaben und Übungen durch die Schüler sowie Basiswissen zu den Lerninhalten enthielten. Zusätzlich wurden Aufgaben und Anleitungen zur Umsetzung kooperativer Lernformen im Sinne von Partner- und Gruppenarbeit umgesetzt. Die Themen konnten so vom Schüler\_innen wahlweise in Einzel- oder Gruppenarbeit selbstständig erschlossen werden, auf ein vom Lehrer kontrolliertes und lehrerzentriertes Klassengespräch wurde verzichtet. Falls erforderlich erhielten die Schüler\_innen Lösungshefte, die ein Prüfen der eigenen Lernergebnisse ermöglichten. Zusätzlich wurde mit Portfolios gearbeitet. Hier konnten Schüler eigenständig ihre individuellen Lernfortschritte, die Bearbeitung freiwilliger Zusatzaufgaben und informierende Rückmeldung der Lehrkräfte dokumentieren.

Neben den kooperativen Lernformen der Partner- und Gruppenarbeit wurde die Methode des Lernzirkels zur Förderung des Autonomieerlebens eingesetzt. Hier können Schüler\_innen themenbezogen ihren individuellen Beginn im Lernzirkel festlegen und Zusatzstationen selbst wählen. Die Verwendung der Schülerarbeitsmappen verfolgt darüber hinaus die Zielsetzung, das Kompetenzerleben der Schüler\_innen zu unterstützen. Aufgaben mit unterschiedlichem Schwierigkeitsgrad ermöglichen den Schüler\_innen ihr optimales Anforderungsniveau zu finden, dieses zu bearbeiten bzw. zu anderen Anforderungsniveaus zu wechseln. Durch erfolgreiche Aufgabenbewältigung wird die Erfahrung von Lernerfolg ermöglicht. Leistungsschwächere Schüler\_innen als auch leistungsstärkere Schüler\_innen können entsprechend der ausgewiesenen Aufgabenschwierigkeit herausfordernde, aber bewältigbare Aufgaben wählen. Diese informieren über die eigene Tüchtigkeit und können so die Leistungsmotivation der Schüler\_innen fördern (Rheinberg, 2004, S. 69 ff.).

Generell sollte der Unterricht so gestaltet werden, dass er durch Differenzierung und Aufgabenpassung gekennzeichnet ist. Schüler\_innen sollten kooperative Übungsphasen mit

Aufgaben entsprechend ihrem individuellen Anforderungsniveau als *Lernraum* nutzen. Die Aufgaben im Lernraum sollten ohne Lösungsvorgaben und kontrollierende Lehrerrückmeldungen gestaltet sein. Werden von der Lehrkraft Rückmeldungen gegeben, sollten diese an der individuellen und sachlichen Bezugsnorm orientiert sein und zu vermehrter Transparenz der Anforderungen beitragen. Es bietet sich der Lehrkraft in einer beratenden Funktion zudem die Möglichkeit, eine positive Lehrer-Schüler-Beziehung und ein positives Lern- und Unterrichtsklima aufzubauen und individueller auf den einzelnen Schüler einzugehen. Durch die Wahlmöglichkeit der Arbeitsform kann neben der Erfahrung der sozialen Eingebundenheit die selbstständige Organisation der eigenen Rolle in der Gruppe vorgenommen werden. Zudem können kooperative Arbeitsformen ermöglichen, dass man die eigene Kompetenz erlebt, wenn man die Ergebnisse der Arbeitsgruppe als Experte präsentiert (Mittag et al., 2009; Bieg & Mittag, 2010).

Die genannten Maßnahmen sollten von den Lehrkräften in den jeweiligen Unterrichtseinheiten umgesetzt werden. Um den Lehrkräften theoretisches Wissen zur Selbstbestimmungstheorie und der Förderung der selbstbestimmten Lernmotivation zu vermitteln, wurden die Lehrkräfte der Treatmentklassen zu einer zweitägigen Schulung eingeladen. In dieser wurden am ersten Tag Inhalte zu den Grundlagen der Selbstbestimmungstheorie erarbeitet. Es wurde auf förderliche Merkmale des Unterrichts, günstige Verhaltensweisen zur Förderung von Lernmotivation und leistungsbezogener Verhaltensweisen fokussiert. Am zweiten Tag der Fortbildung wurden die Lehrkräfte an das Manual und das zentrale Element der Schülerarbeitsmappe herangeführt und es wurden fachliche und organisatorische Fragen besprochen. Die Lehrkräfte der Kontrollgruppe wurden in einem eintägigen Workshop über die Unterrichtsziele und Unterrichtsmethoden der Unterrichtseinheiten informiert. Sie erhielten aber weder Informationen zur Selbstbestimmungstheorie und den pädagogischen Unterrichtsimplicationen noch Arbeitsmappen und manualisierte Anleitungen zum Unterrichtsverlauf.

Die primäre Zielsetzung der konzipierten Intervention war eine Förderung der selbstbestimmten Lernmotivation und eine Leistungsverbesserung der Teilnehmer\_innen. Folgende Forschungsfragen leiteten die Evaluation der Programmwirksamkeit: Gibt es einen Einfluss der Interventionsbestandteile auf die Wahrnehmung des Unterrichts durch die Schüler, insbesondere was das Autonomieerleben und die Fürsorglichkeit der Lehrkraft betrifft? Weiterhin sollte geprüft werden, ob und in welchem Ausmaß sich Interventionseffekte auf die affektiven

und motivationalen Lernhaltungen wie die selbstbestimmte Lernmotivation und die Lernfreude zeigen und ob sich Auswirkungen auf die Fachleistungen finden. Ein weiteres Forschungsinteresse lag darin, zu erfassen, ob je nach Schulform bzw. dem Leistungsniveau der Schüler in Gymnasial-, Real- oder Hauptschulklassen unterschiedliche Effekte der Intervention zu beobachten sind. Um zu erfassen, welche Inhalte von den Lehrkräften umgesetzt wurden, sollten die vermittelten Inhalte mithilfe eines von den Lehrkräften nach jeder Unterrichtsstunde ausgefüllten Feedback-Fragebogens erfasst werden. In diesem sollten sie genaue Angaben zu den verwendeten Materialien, Arbeitsformen, Einschätzungen zu Schülervariablen, Abweichungen und Schwierigkeiten bei der Umsetzung machen.

Die Studie wurde in der Sekundarstufe in der achten Jahrgangsstufe mit 353 Schülern durchgeführt. Diese besuchten sieben Hauptschulklassen, fünf Realschulklassen und acht Gymnasialklassen in Baden-Württemberg. Die quasi-experimentelle Studie wurde mittels eines Treatment-Kontrollgruppen-Designs mit drei Messzeitpunkten (s. *Abbildung 7*) durchgeführt. Die Rekrutierung der Lehrer erfolgte über Ausschreibungen und direkte Kontaktaufnahme. Neun Klassen wurden den Treatmentklassen (N = 165) und 11 Klassen (N = 188) der Kontrollgruppe zugewiesen. Die Schüler\_innen waren im Mittel 13.6 Jahre alt (Range 12-17 Jahre; 50 % weiblich) (Mittag et al., 2009).

Im Fach Deutsch wurden zwei Unterrichtseinheiten *Grammatik und Kommasetzung* und das Thema *Sachtexte erschließen* als Basisthemen zur Umsetzung der Interventionsbestandteile einbezogen. Diese Unterrichtseinheiten wurden in 12 bzw. 13 Unterrichtsstunden innerhalb von 4 Wochen umgesetzt.

Vor Beginn der Unterrichtseinheit 1 wurden im ersten Schritt mittels Fragebogen Daten zu den Zielkriterien Lernfreude, Lernmotivation und die wahrgenommenen Unterrichtsmerkmale erhoben. Der Prätest 1 umfasste die Erfassung des themenbezogenen Vorwissens zu *Grammatik und Kommasetzung*. Auf die Unterrichtseinheit 1 folgten ein erneuter Leistungstest und sechs Wochen später eine Follow-up Erhebung des Leistungstests. Der Messzeitpunkt 2, der vor der Unterrichtseinheit 2 lag, erfasste erneut die Lernfreude, Lernmotivation und die wahrgenommenen Unterrichtsmerkmale.

Auch hier erfolgte vor der Unterrichtseinheit 2 erneut ein Leistungstest zur Erfassung des Vorwissens zum Thema *Sachtexte erschließen*. Im Anschluss wurden Veränderungen in der Leistung gemessen und sechs Wochen später erneut eine Follow-up Erhebung zur Leistung

durchgeführt. Die Erhebung wurde mit der Fragebogenerhebung bei den Schülern abgeschlossen. Die Durchführung der Intervention und der Leistungstest oblagen den Lehrkräften und erfolgten im regulären Lernsetting. Die Fragebogenerhebung wurde von der wissenschaftlichen Begleitung der Intervention durchgeführt. Es wurden nur Teilnehmer\_innen einbezogen, bei denen zuvor die Eltern die Zustimmung zur Teilnahme gegeben hatten.

<b>t1</b>	<b>Prätest I</b>	<b>Unterrichtseinheit I</b>	<b>Posttest I</b>	<b>Follow-up I</b>	
Fragebogen (Schüler & Lehrer)	Leistungstest (Vorwissen)	Grammatik & Kommasetzung (13 Stunden)	Leistungstest	Leistungstest	
<b>t2</b>	<b>Prätest I</b>	<b>Unterrichtseinheit II</b>	<b>Posttest II</b>	<b>Follow-up II</b>	<b>t3</b>
Fragebogen (Schüler & Lehrer)	Leistungstest (Vorwissen)	Sachtexte / Texterschließung (13 Stunden)	Leistungstest	Leistungstest	Fragebogen (Schüler & Lehrer)

Abbildung 7. Erhebungsdesign. Aus Zur Bedeutung der Lehrer-Schüler-Beziehung für die Wahrnehmung der Unterrichtsqualität und die Lernmotivation von W. Mittag (2011, Mai). Unveröffentlichte Präsentation an der Universität Erfurt.

Zur Erfassung der Programmwirksamkeit wurden neben den Leistungstests Fragebögen zur wahrgenommenen Autonomieunterstützung, zur erlebten Lehrerfürsorglichkeit, der Lernfreude und der motivationalen Selbstregulation eingesetzt. Verwendet wurde die Skala *Autonomieunterstützung im Unterricht* (modifizierte Skala nach Röder & Kleine, 2003) mit fünf Items. Erfasst wurde mit fünf Items die Skala *Fürsorglichkeit* (gekürzte Version nach v. Saldern & Littig, 1987). Mit vier Items wurde die Skala *Lernfreude* (Laukenmann et al., 2003) genutzt. Vier Subskalen des Fragebogens zur *Motivationalen Selbstregulation* (MoS-J 16, Bieg & Mittag, 2005) wurden eingesetzt (Zur genaueren Beschreibung der Instrumente zur Erfassung der Schülervariablen s. Kapitel 5.2). Zur Erhebung der Leistung zum jeweiligen Messzeitpunkt wurden in einer Pilotstudie getestete, an den curricularen Inhalten orientierte und in gleichförmiger Strukturiertheit konzipierte Leistungstests eingesetzt.

Neben der Evaluation der Programmwirksamkeit sollte auch die Programmdurchführung, also die Implementation der Intervention, erfasst werden. Dies geschah mit Hilfe von nach den

Unterrichtsstunden ausgefüllten Feedback-Fragebögen (s. Anhang), in denen die Lehrer im Rahmen eines geschlossenen Antwortformats angeben sollten, welche Methoden und welche Arbeitsformen sie einsetzen hatten. Zudem sollte von den Lehrkräften eine Einschätzung der Schüler\_innen zu den Variablen Interesse, Lernfreude, Mitarbeit, Selbstbestimmung, Kompetenzerleben, soziales Miteinander, Verständnisschwierigkeiten und Unterrichtsstörungen vorgenommen werden. Darüber hinaus wurde eine Teilstichprobe von Lehrkräften mithilfe von einem offenen Antwortformat befragt, wie sie die Unterrichtsstunde insgesamt einschätzen, welche Schwierigkeiten auftraten und welche Veränderungen vorgenommen wurden. Ziel dieser Erhebung war es, zu erfassen, inwieweit die Intervention entsprechend der im Manual und in den theoretischen Hintergrundinformationen gegebenen Anleitungen im Unterschied zur Kontrollgruppe in den Unterricht implementiert wurde. Ebenso sollte erfasst werden, ob und inwiefern die Effekte der Intervention in Abhängigkeit vom der Implementationsqualität variieren (Bieg & Mittag, 2010). Die Erfassung der Implementationsgenauigkeit und somit die Evaluation der Programmdurchführung ist Ziel dieser Arbeit und wird in den weiteren Ausführungen näher beschrieben.

Um standardisierte und vergleichbare Maße zur Verfügung zu haben, wurden die empirischen Effektgrößen (*demp*) (Netto-Effekte), also die an der mittleren Binnenstreuung standardisierten Mittelwertsdifferenzen zwischen Treatment und Kontrollgruppe für kurzfristige Effekte  $t_1$  zu  $t_2$  und für langfristige Effekte  $t_1$  zu  $t_3$ , berechnet. Die gebildeten empirischen Effektgrößen (*demp*) weisen so bei negativen Vorzeichen auch auf negative Effekte der Intervention hin (Mittag et al. 2009; Bieg & Mittag, 2010). Die folgenden Ergebnisse der varianzanalytischen Auswertung beziehen sich auf die Angaben in Mittag et al. (2009). Dort finden sich ebenso differenziertere Angaben zu den Analysen für die jeweiligen Messzeitpunkte.



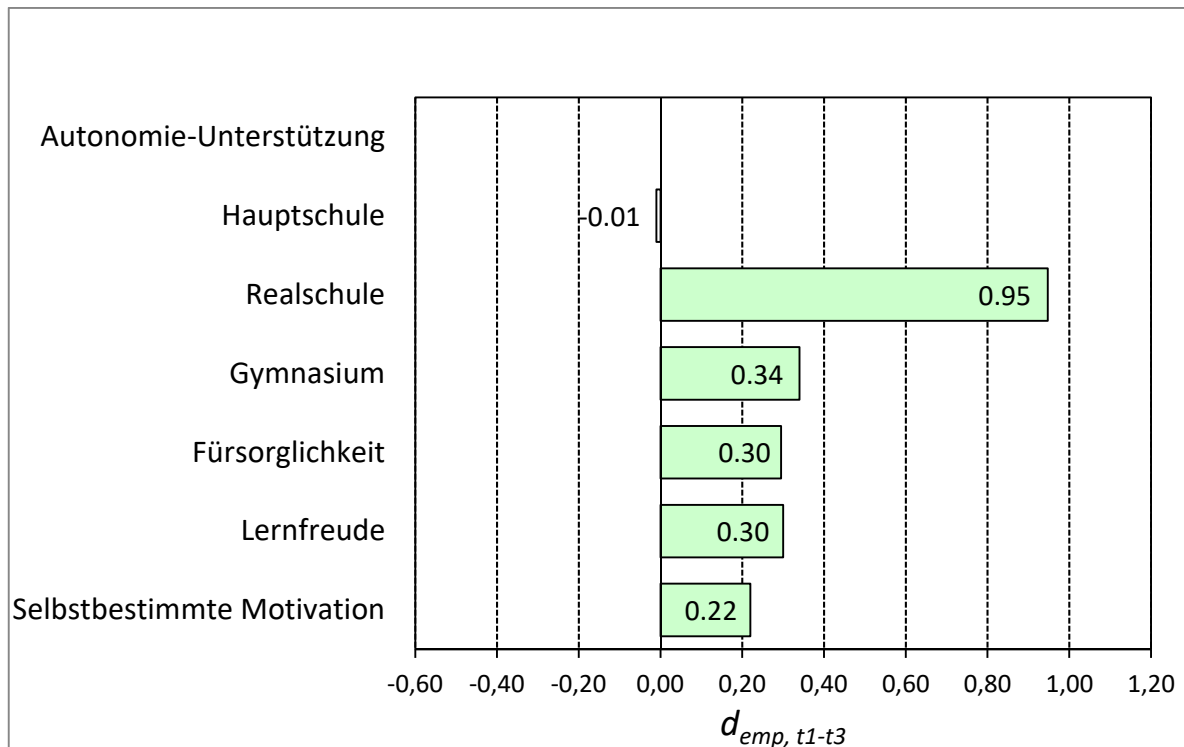


Abbildung 8. Effektstärken in Bezug auf Schülerskalen und Schulform. Aus Zur Bedeutung der Lehrer-Schüler-Beziehung für die Wahrnehmung der Unterrichtsqualität und die Lernmotivation von W. Mittag (2011, Mai). Unveröffentlichte Präsentation an der Universität Erfurt.

Für die wahrgenommene **Autonomieunterstützung** fanden sich keine positiven Interventionseffekte für Hauptschüler, obwohl zu allen drei Messzeitpunkten im Mittel ein höheres Autonomieerleben berichtet wurde als bei den anderen Schulformen. Für die Realschulen ergibt sich der signifikante Interventionseffekt durch das stetige Absinken der wahrgenommenen Autonomieunterstützung über die Messzeitpunkte in der Kontrollgruppe.

Die folgende Abbildung zeigt die Unterschiede der wahrgenommenen **Autonomieunterstützung** nach Schulform. Für die Hauptschulen finden sich keine signifikanten Interventionseffekte. Hier bestehen auch schon zum ersten und zweiten Messzeitpunkt geringere Werte für die wahrgenommene Autonomieunterstützung in der Treatmentgruppe. Ein signifikanter Interventionseffekt zeigt sich für die Gymnasien und die Realschulen. Hier findet sich ein Anstieg in der Treatmentgruppe, während in den Kontrollgruppen kein Anstieg zu beobachten ist. Betrachtet man die Netto-Effekte für die Schulform Gymnasium, so werden diese auch nur zwischen den Messzeitpunkten t1-t3 mit  $d_{emp} = .34$  signifikant. Für die Realschule findet sich ebenso ein signifikanter Interventionseffekt zwischen t1 und t3 mit  $d_{emp} = .95$ . Insgesamt sind die Effekte nach Cohen im kleinen bis hohen Bereich einzuordnen (s. Abbildung 9).

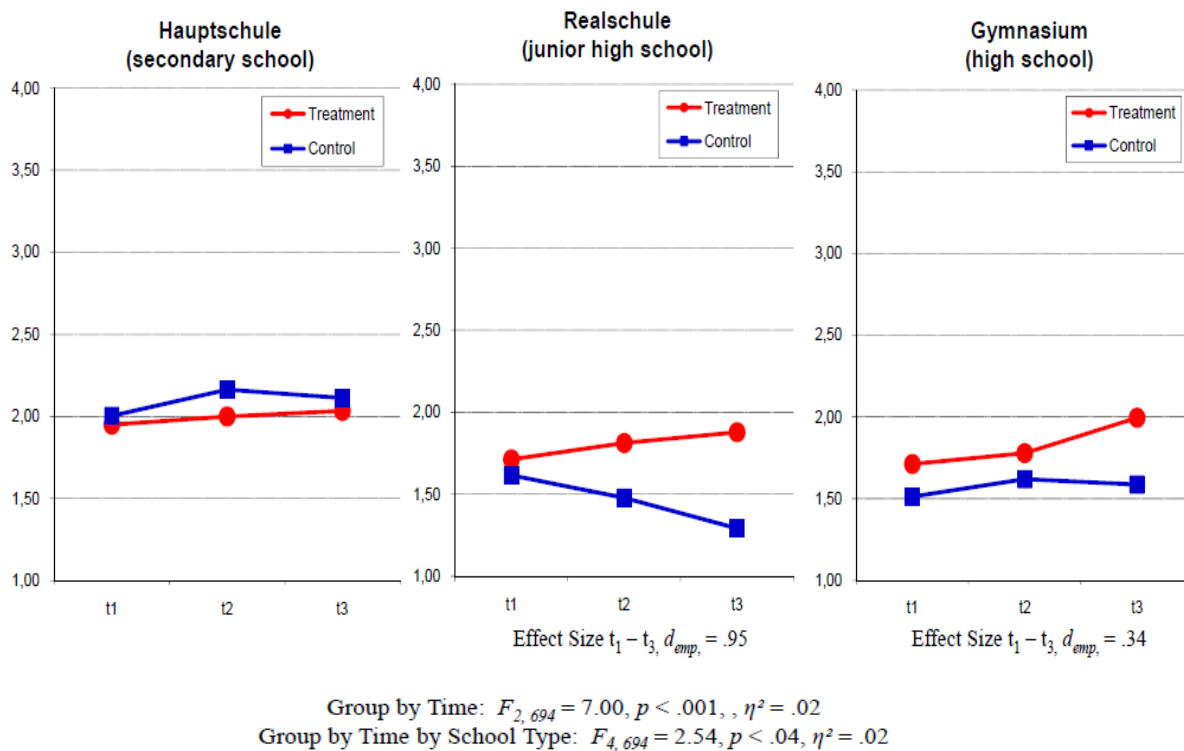
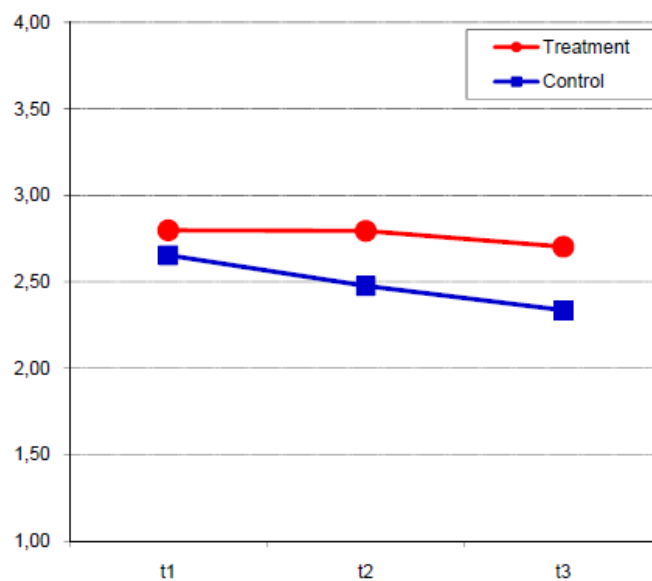


Abbildung 9. Autonomieerleben nach Schulform. Aus Zur Bedeutung der Lehrer-Schüler-Beziehung für die Wahrnehmung der Unterrichtsqualität und die Lernmotivation von W. Mittag (2011, Mai). Unveröffentlichte Präsentation an der Universität Erfurt.

Auch für die **Lehrerfürsorglichkeit** zeigt sich, dass die erlebte Lehrerfürsorglichkeit in den Kontrollgruppen über die Messzeitpunkte hinweg kontinuierlich abnimmt, während in den Treatmentgruppen dies nur von t2 zu t3 der Fall ist. Betrachtet man die Netto-Effekte, findet sich für die Betrachtung der Veränderung von t1 zu t3 ein  $d_{emp} = .30$ . Insgesamt sind die Effekte nach Cohen im kleinen bis mittleren Bereich einzuordnen.

Die folgende Abbildung 10 zeigt die Unterschiede der wahrgenommenen Lehrerfürsorglichkeit nach Gruppenzuordnung.



Effect Size  $t_1 - t_3$ ,  $d_{emp} = .30$

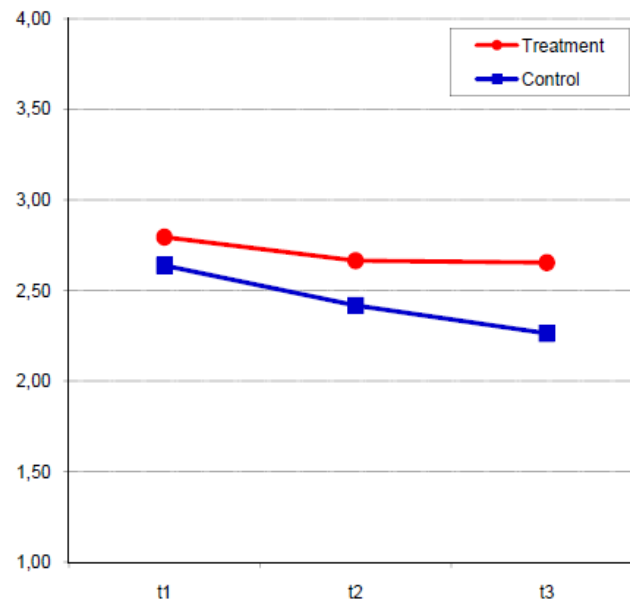
Group by Time:  $F_{2, 694} = 4.75, p < .01, \eta^2 = .01$

Abbildung 10. Wahrgenommene Lehrerfürsorglichkeit. Aus Zur Bedeutung der Lehrer-Schüler-Beziehung für die Wahrnehmung der Unterrichtsqualität und die Lernmotivation von W. Mittag (2011, Mai). Unveröffentlichte Präsentation an der Universität Erfurt.

Für die **Lernfreude** kann gezeigt werden, dass sich Treatment- und Kontrollgruppe zu den Messzeitpunkten t2 und t3 signifikant unterscheiden.

Für die Kontrollgruppe zeigt sich, wie auch bei der Lehrerfürsorglichkeit, eine signifikante Abnahme über alle Messzeitpunkte. Auch für die Treatmentgruppe kommt es erst zu einer Abnahme (t1-t2), der Wert stabilisiert sich aber wieder bis zum dritten Messzeitpunkt. Statistisch signifikant wird nur der Nettoeffekt für t1-t3 in geringer Größe mit  $d_{emp} = .30$ .

Die folgende Abbildung zeigt die Unterschiede in der Lernfreude nach Gruppenzuordnung.

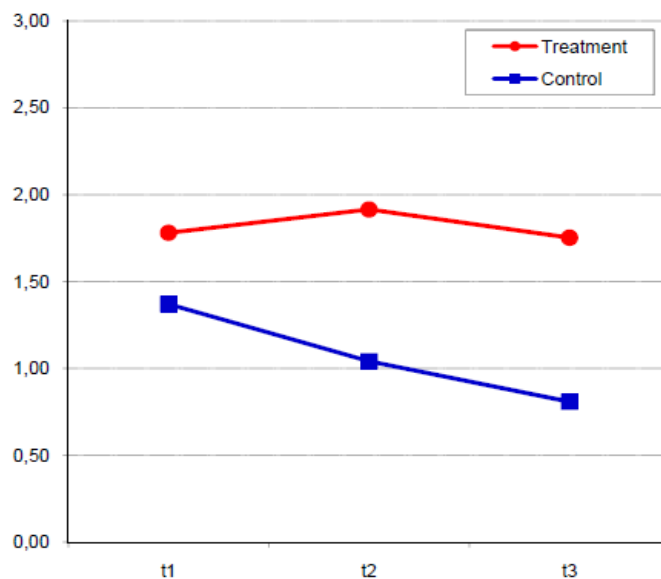


Effect Size  $t_1 - t_3$ ,  $d_{emp} = .30$

Group by Time:  $F_{2, 694} = 3.95$ ,  $p < .02$ ,  $\eta^2 = .01$

*Abbildung 11.* Lernfreude. Aus Zur Bedeutung der Lehrer-Schüler-Beziehung für die Wahrnehmung der Unterrichtsqualität und die Lernmotivation von W. Mittag (2011, Mai). *Unveröffentlichte Präsentation an der Universität Erfurt.*

Für die **selbstbestimmte Lernmotivation** (s. *Abbildung 12*) finden sich Unterschiede zwischen Treatment- und Kontrollgruppe und diese werden zum zweiten und dritten Messzeitpunkt signifikant. Ein kleiner signifikanter Netto-Effekt ergibt sich für die Unterschiede zum ersten und dritten Messzeitpunkt ( $d_{emp} = .22$ ). Dieser Effekt ist in der bedeutsamen Abnahme der selbstbestimmten Lernmotivation in der Kontrollgruppe begründet, während in der Interventionsgruppe keine signifikante Abnahme zu beobachten ist.



Effect Size  $t_1 - t_3$ ,  $d_{emp} = .22$

Group by Time:  $F_{2, 694} = 2.98$ ,  $p < .06$ ,  $\eta^2 = .01$

Abbildung 12. Selbstbestimmte Lernmotivation. Aus Zur Bedeutung der Lehrer-Schüler-Beziehung für die Wahrnehmung der Unterrichtsqualität und die Lernmotivation von W. Mittag (2011, Mai). Unveröffentlichte Präsentation an der Universität Erfurt.

In Bezug auf die **Lernleistung** konnten sich nur die Schüler der Realschule und der Gymnasien verbessern. Für die Hauptschule zeigen sich günstigere Werte für die Kontrollgruppe (s. Abbildung 11).

Betrachtet man die Werte von t1 zu t2, so findet sich kein bedeutsamer Unterschied zwischen Treatment- und Kontrollgruppe. Von t2 bis t3 sinken die Werte für die Lernleistung in der Treatmentgruppe allerdings ab. So kann für die Hauptschule nur ein negativer Interventionseffekt ( $d_{emp} = -.43$ ) berichtet werden. Für die Schulformen Realschule und Gymnasium finden sich über alle Messzeitpunkte höhere Mittelwerte für die Lernleistung der Treatmentgruppen als in den Kontrollgruppen. Für Realschule und Gymnasium zeigen sich zum ersten Messzeitpunkte keine signifikanten Gruppenunterschiede. Es können folgende positive mittlere bis hohe Netto-Effekte berichtet werden: Realschule (t1-t3)  $d_{emp} = 1.17$ ; Gymnasium (t1-t3)  $d_{emp} = 0.61$ . Für die zweite Unterrichtseinheit finden sich bezogen auf das Zielkriterium Lernleistung für die Hauptschule keine signifikanten Unterschiede zwischen den Gruppen zu allen Messzeitpunkten. Für die Betrachtung der Werte von t1 zu t3 muss für die Realschule ein negativer Interventionseffekt berichtet werden ( $d_{emp} = -0.50$ ).

Die Treatment- und Kontrollgruppen im Gymnasium unterscheiden sich zum ersten und zweiten Messzeitpunkt nicht signifikant. Erwartungsgemäß zeigt sich hier für die Interventionsgruppe ein positiver Lernzuwachs von t1 zu t3 und ein mittlerer Netto-Effekt von  $d_{emp} = .51$ .

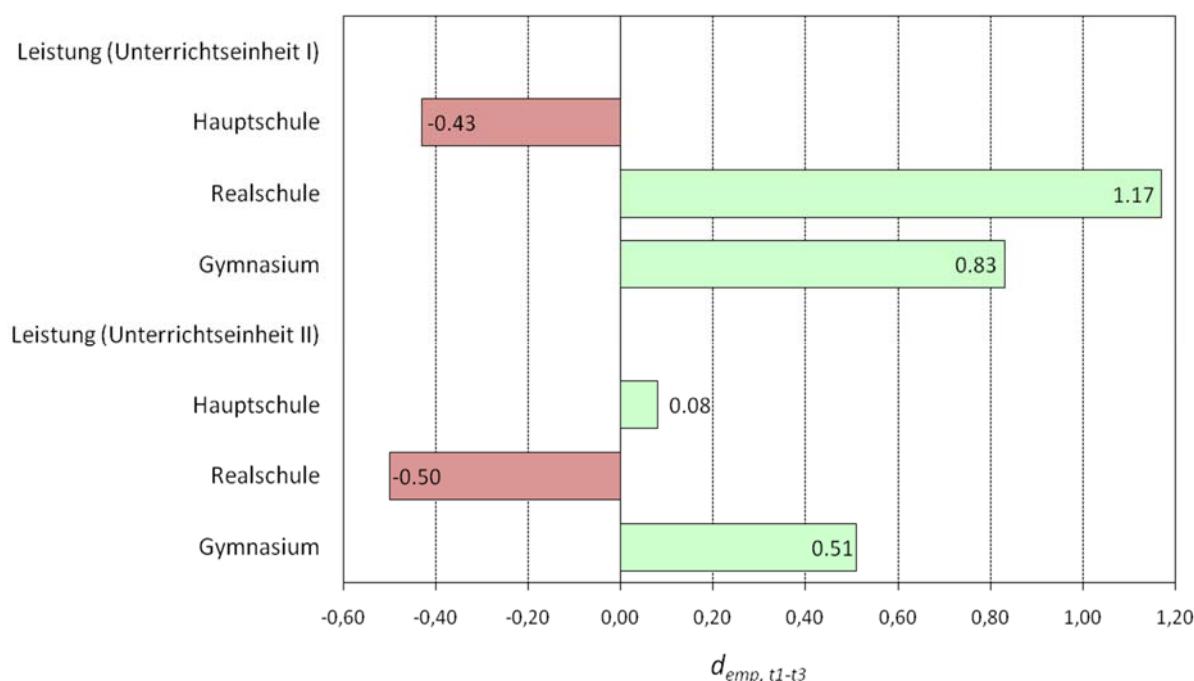


Abbildung 13. Leistungsentwicklung. Aus Zur Bedeutung der Lehrer-Schüler-Beziehung für die Wahrnehmung der Unterrichtsqualität und die Lernmotivation von W. Mittag (2011, Mai). Unveröffentlichte Präsentation an der Universität Erfurt.

Insgesamt zeigen sich für alle Schulformen signifikante Interventionseffekte, aber auch Unterschiede in Abhängigkeit von der Schulform. Es finden sich nur teilweise erwartungskonforme Ergebnisse und auch differenzielle Interventionseffekte zuungunsten der Klassen, die von der Intervention profitieren sollten. Dies fand sich für die Hauptschule nach Durchlaufen der ersten Unterrichtseinheit. Bei der zweiten Unterrichtseinheit zeigten sich ungünstige Effekte für die Realschule.

Es finden sich für die Messung t1-t2 Effekte für die Lehrerfürsorglichkeit und die selbstbestimmte Lernmotivation bei allen Schulformen. Darüber hinaus ergeben sich über alle Schulformen hinweg höhere Netto-Effekte bei der Messung t1-t3 als in der Messung von t1-t2, was darauf hindeutet, dass Veränderungen durch eine höhere Interventionszeit begünstigt werden (Streblov & Hasselhorn, 2001, nach Mittag et al. 2009). Die insgesamt zufriedenstellenden Effekte auf die Zielindikatoren können auch durch die deutlicheren Rückgänge in den Kontrollklassen im Vergleich zu den weniger deutlichen Rückgängen in den gemessenen

Zielindikatoren bei den Treatmentklassen zurückgeführt werden. Nur bei den Gymnasien konnte ein signifikanter Anstieg der wahrgenommenen Autonomieunterstützung beobachtet werden. So konnte ein in der Forschungsliteratur bekannter Rückgang in den affektiven und motivationalen Lernhaltungen (Fend, 1997; Pekrun, 1993; Saldern & Littag, 1987, nach Mittag et al., 2009) im Schuljahresverlauf durch die Intervention günstig beeinflusst werden.

Bei der Betrachtung differenzieller Effekte auf die Lernleistung finden sich für die erste Unterrichtseinheit positive Effekte der Intervention für Realschule und Gymnasien. Nicht erwartungskonform zeigen sich aber auch aufgrund eines Abfalls von Posttest zu Follow-up Effekte zuungunsten der Treatmentklassen für die Hauptschulen in Bezug auf die erste Unterrichtseinheit und für Realschulen bei Unterrichtseinheit 2. Erwartungskonform zeigten sich mittlere bis große Effekte für die erste Unterrichtseinheit für Realschulen und Gymnasien in der Leistungsentwicklung zum Posttest und zur Follow-up Erhebung. In der zweiten Unterrichtseinheit zeigen sich positive Interventionseffekte auf die Leistung zum Follow-up nur für Gymnasien (Mittag et al., 2009). Zu den Ergebnissen muss angemerkt werden, dass die negativen Interventionseffekte bei den Haupt- und Realschulklassen möglicherweise darauf zurückzuführen sind, dass die Unterrichtsmaterialien nicht an den Leistungsstand der schwächeren Schüler angepasst waren. Einschränkend muss im Weiteren angeführt werden, dass relative niedrige Reliabilitäten für die Messung bei der zweiten Unterrichtseinheit vorliegen. Die zeitliche Abfolge der aufeinanderfolgenden Unterrichtseinheiten könnte darüber hinaus zu Reihenfolgeeffekten geführt haben, was bei der Einschätzung der Ergebnisse ebenso beachtet werden muss.

Die Ergebnisse stützen insgesamt die Annahmen der Selbstbestimmungstheorie. Zunehmende oder stabile Erfahrungen von Autonomieunterstützung und sozialer Eingebundenheit (Fürsorglichkeit) sowie das Erleben von Lernfreude korrespondieren mit eher positiven Entwicklungen in der selbstbestimmten Lernmotivation und den schulischen Leistungen. Insbesondere die ausgebliebenen bzw. negativen Interventionseffekte bei den leistungsschwächeren Lernern zeigen die Notwendigkeit, die Lernmaterialien stärker an das Leistungsniveau der Lernenden anzupassen.

Letztlich stehen bei diesen Ergebnissen die Prüfung der Implementationsqualität und von deren Zusammenhängen mit der Wirksamkeit der Unterrichtsintervention noch aus. Diese Zusammenhänge sollen im Folgenden näher untersucht werden.

### 3 Fragestellung und Hypothesen

Das Hauptanliegen der vorliegenden Arbeit besteht darin, die Implementation von Unterrichtseinheiten zur Förderung der selbstbestimmten Lernmotivation zu erfassen. Es sollen Aussagen darüber getroffen werden, in welcher Beziehung die Implementationsqualität zur Wirksamkeit der Unterrichtsinterventionen steht. Die Angaben der Lehrkräfte zur Umsetzung der Interventionsmaßnahmen müssen klassifiziert und quantifiziert werden und in Implementationsindikatoren überführt werden. Auf der Basis subjektiver Schülereinschätzungen des Unterrichts und ihrer Lernmotivation sowie objektiver Fachleistungsdaten der Schüler\_innen sollen differenzierte Aussagen über Zusammenhänge zwischen Implementationsindikatoren und Wirksamkeit der Intervention getroffen werden. Ein solches Anliegen setzt voraus, dass das gewählte Erhebungsverfahren zur Erfassung von Indikatoren der Implementationsqualität (Feedback-Fragebögen) geeignet ist und die gebildeten Indikatoren geringe korrelative Zusammenhänge und somit eine hohe diskriminative Validität aufweisen. Die Teilaspekte der Indikatoren, sollten jeweils dasselbe Konstrukt messen und somit hohe korrelative Zusammenhänge aufweisen (konvergente Validität). Dies gilt es zu zeigen.

Die zentralen Forschungsfragen lauten daher:

- 1) Zeigen sich Zusammenhänge zwischen den verschiedenen Teilaspekten der Indikatoren der Implementationsqualität?
- 2) Welche Zusammenhänge weisen die einzelnen Indikatoren der Implementationsqualität zur Entwicklung des Autonomieerlebens, der erlebten Lehrerfürsorglichkeit, der Lernfreude und der selbstbestimmten Lernmotivation auf?
- 3) Welche Zusammenhänge weisen die einzelnen Indikatoren der Implementationsqualität mit der Leistungsentwicklung der Schüler\_innen auf?

Die Klärung dieser Forschungsfragen erfolgt nach den in Abschnitt 4.2.4 aufgeführten Arbeitsschritten.



## Hypothesen

Ausgehend von dem in den vorherigen Kapiteln berichteten Forschungsstand werden im Hinblick auf die Forschungsfragen konkrete inhaltliche Hypothesen formuliert. Diese sollen mit den Projektdaten der Studie überprüft werden.

Hypothese 1: Teilaspekte eines Indikators weisen vergleichsweise hohe Zusammenhänge innerhalb der Indikatorengruppe auf. Bei Indikatoren, die eine geringere Ähnlichkeit in der Operationalisierung aufweisen, zeigen sich geringe Zusammenhänge. Für einzelne Indikatoren der Implementationsqualität zeigen sich aufgrund der ähnlichen Operationalisierung hohe Zusammenhänge:

Hypothese 1a: Die Teilaspekte eines Indikators weisen vergleichsweise hohe Zusammenhänge auf (positive Intra-Indikatoren-Zusammenhänge).

Hypothese 1b: Die Indikatoren (bzw. Teilaspekte der Indikatoren) weisen vergleichsweise niedrige Zusammenhänge mit anderen Indikatoren (bzw. Teilaspekten anderer Indikatoren) auf (relative Inter-Indikatoren-Unabhängigkeit). Eine Ausnahme bilden die Indikatoren Adhärenz und Dosierung. Aufgrund der Ähnlichkeit ihrer Operationalisierungen werden positive Zusammenhänge zwischen beiden Indikatoren (bzw. Teilaspekten beider Indikatoren) vermutet.

Hypothese 2: Eine hohe Implementationsqualität sollte sich auf Schülerseite in positiven Veränderungen des Autonomieerlebens, der erlebten Lehrerfürsorglichkeit, der Lernfreude und der selbstbestimmten Lernmotivation niederschlagen:

Hypothese 2a: Es gibt einen positiven Zusammenhang zwischen Implementationsqualität und den Unterrichtsmerkmalen Autonomieerleben und erlebte Lehrerfürsorglichkeit (aus Schülersicht).

Hypothese 2b: Es gibt einen positiven Zusammenhang zwischen Implementationsqualität und emotional-motivationalen Selbsteinschätzungen zu Lernfreude und selbstbestimmter Lernmotivation (aus Schülersicht).

Hypothese 3: Es gibt einen positiven Zusammenhang zwischen Implementationsqualität und Veränderungen in der Lernleistung.

## 4 Methode

Nachdem im Kapitel 2.4 die Programmkonzeption und die Ergebnisse der Wirksamkeitsanalyse vorgestellt wurden, soll im Folgenden auf die vorliegende Untersuchung und die Auswertung der Angaben zur Implementation fokussiert werden.

In Kapitel 4.1 wird die Stichprobe beschrieben. In Kapitel 4.2 werden die eingesetzten Instrumente sowie die Klassifikation und Quantifizierung der lehrerbezogenen Angaben zur Implementationsqualität dargelegt.

### 4.1 Stichprobe

Entsprechend der Selbstbestimmungstheorie der Motivation von Deci und Ryan (1993, 2000) sowie unter Berücksichtigung fachdidaktischer Überlegungen wurden im FuN-Kolleg „Förderung selbstbestimmten Lernens im Unterricht“ je zwei am Lehrplan orientierte Unterrichtseinheiten für das Fach Deutsch entwickelt.

Die Unterrichtseinheiten wurden zunächst in einer Pilotstudie formativ evaluiert. Die so auf ihre pädagogisch-praktische Anwendbarkeit optimierten Unterrichtseinheiten wurden anschließend in einem längsschnittlich angelegten Treatment-Kontrollgruppen-Design mit Treatment- und Kontrollklassen der 8. Jahrgangsstufe summativ auf ihre Wirksamkeit untersucht. Die teilnehmenden Lehrkräfte wurden mittels öffentlicher Ausschreibungen im Großraum Stuttgart und im Umland von Stuttgart rekrutiert. Die Lehrkräfte wurden anschließend den Treatmentklassen (mit Unterrichtsmaterialien zur Förderung der selbstbestimmten Lernmotivation) und Kontrollklassen (mit regulärem Deutschunterricht) zugeordnet. Beide Gruppen hatten dieselbe curriculare Zielsetzung. Die beiden Unterrichtseinheiten wurden in 9 Treatmentklassen mit 165 Schüler\_innen sowie 11 Kontrollklassen mit 188 Schüler\_innen in der ersten Hälfte des Schuljahres 2005/06 durchgeführt. Zunächst wurde in allen Projektlassen die Unterrichtseinheit *Grammatik und Kommasetzung* durchgeführt. In der zweiten Schuljahreshälfte zu Beginn des Jahres 2006 wurde die Unterrichtseinheit *Sachtexte erschließen* eingesetzt. Zu beiden Unterrichtseinheiten wurden für die 14 Unterrichtsstunden bei den Lehrkräften über Feedback-Fragebögen Daten zur Umsetzung erhoben. Beim Rücklauf der Feedback-Fragebögen handelt es sich um Angaben von 8 Lehrkräften der 8. Klasse (davon

7 weiblich) zu der Unterrichtseinheit 1. Für die Unterrichtseinheit 2 liegen Angaben von 9 Lehrkräften vor (davon 8 weiblich). Zwei Lehrkräfte waren an einer Hauptschule tätig, drei der Lehrkräfte unterrichteten an Realschulen und vier Lehrkräfte an Gymnasien.

Tabelle 4  
*Rücklauf Feedback-Fragebögen*

Lehrkraft	Geschlecht	Schulform	Rücklauf Feedback- Fragebögen UE 1	Rücklauf Feedback- Fragebögen UE 2
<b>Lehrkraft 1</b>	weiblich	Hauptschule	15	16
<b>Lehrkraft 2</b>	weiblich	Hauptschule	10	16
<b>Lehrkraft 3</b>	weiblich	Realschule	16	20
<b>Lehrkraft 4</b>	weiblich	Realschule	13	15
<b>Lehrkraft 5</b>	männlich	Realschule	14	20
<b>Lehrkraft 6</b>	weiblich	Gymnasium	/	15
<b>Lehrkraft 7</b>	weiblich	Gymnasium	16	15
<b>Lehrkraft 8</b>	weiblich	Gymnasium	15	15
<b>Lehrkraft 9</b>	weiblich	Gymnasium	11	10

Die Tabelle 4 zeigt Klassen mit Rücklauf der Feedback-Fragebögen nach den durchgeführten Unterrichtseinheiten. Bei der Unterrichtseinheit 1 haben acht von neun Lehrkräften regelmäßig die Durchführung über Feedback-Fragebögen dokumentiert. Drei Lehrkräfte haben nur 10, 11 bzw. 13 der vorgesehenen 14 Bögen ausgefüllt. Zwei der Lehrkräfte haben Wiederholungsstunden durchgeführt und diese mit jeweils einem zusätzlichen Bogen dokumentiert. Drei Lehrkräfte haben zusätzliche inhaltliche Stunden durchgeführt und auch zu diesen einen Feedback-Fragebogen eingereicht.

Für die Unterrichtseinheit 2 liegt insgesamt eine höhere Dosierung bezogen auf die Anzahl der Unterrichtsstunden vor: Hier hat nur eine Lehrkraft mit 10 Bögen weniger als das Soll von 14 Feedbackbögen eingereicht. Vier der neun Lehrkräfte haben das beabsichtigte Soll von 14

Unterrichtsstunden und zusätzliche mit Feedbackbögen dokumentierte Wiederholungsstunden eingesetzt. Vier Lehrkräfte haben zusätzliche Stunden mit neuen Inhalten durchgeführt.

## 4.2 Instrumente

### 4.2.1 Lehrerfragebogen zur Erfassung der Implementationsqualität

In dem Feedback-Fragebogen (s. Anhang) war von den Lehrkräften für jede der 14 Unterrichtsstunden anzugeben, welche Medien und Arbeitsformen zum Einsatz kamen, wie aus Sicht der Lehrkraft die Unterrichtsstunde von den Schüler\_innen angenommen wurde, wie sie selbst die Durchführung der Unterrichtsstunde qualitativ einschätzen, welche Schwierigkeiten sie bemerkt hatten und welche Konsequenzen und Veränderungen im Unterrichtsverlauf daraus resultierten.

Der Feedback-Fragebogen ist in vier Bereiche eingeteilt. Im ersten werden Angaben zur Lehrperson, zur Klasse, zum Unterrichtsthema und zur Zielsetzung erfragt. Im zweiten Abschnitt war anzugeben, welche Medien bzw. Arbeitsformen eingesetzt wurden: In einem halboffenen Antwortformat sollten die Lehrkräfte einerseits aus vorgegebenen Medien und Arbeitsformen auswählen, konnten darüber hinaus unter der Kategorie *Sonstiges* aber auch andere Medien oder Arbeitsformen ergänzen. Im dritten Bereich werden intervallskalierte Daten erhoben. Mittels eines geschlossenen Antwortformats sind Gesamteinschätzungen aller Schüler\_innen jeder Klasse auf einer 5-stufigen Skala von *hoch/gut* (1) über *mittel* (3) bis *niedrig/schlecht* (5) in der Unterrichtsstunde vorzunehmen. Einzuschätzen sind dabei das Interesse, die Lernfreude, Mitarbeit, Selbstbestimmung, das Kompetenzerleben und soziale Miteinander sowie Verständnisschwierigkeiten und Unterrichtsstörungen. Ein weiteres Item erfasst die Einschätzung der Lehrkräfte zur Umsetzungsqualität der Unterrichtsvorhaben. Im letzten Abschnitt sind in einem offenen Antwortformat anzugeben, ob und ggf. welche Schwierigkeiten während der Unterrichtsstunden auftraten, zu welchen Konsequenzen dies führte und inwieweit aus von den Lehrkräften anzugebenden Gründen Veränderungen durchgeführt werden sollten.

#### 4.2.2 Klassifikation und Quantifizierung der lehrerbezogenen Angaben zur Implementationsqualität

Nach Sichtung und Auswertung theoretischer und empirischer Vorarbeiten wurde auf der Grundlage der in Abbildung 1 dargestellten Indikatoren zur Programmintegrität bzw. Implementationsqualität ein Klassifikationsschema entwickelt, in das die Angaben aus dem Feedback-Fragebogen eingeordnet wurden. Auf der Grundlage dieser Daten wurden für jede Unterrichtsstunde drei Indikatoren für die Adhärenz, vier Indikatoren für die Dosierung, zwei Indikatoren für die Qualität der Programmvermittlung und ein Indikator für die Responsivität der am Programm teilnehmenden Schüler\_innen gebildet.

Wie in Abschnitt 1.2.1 dargelegt, soll durch die **Adhärenz** oder Einhaltung erfasst werden, in welchem Ausmaß ein Programm als Ganzes entsprechend seiner Konzeption durchgeführt wird. Aufgrund der Angaben im Feedback-Fragebogen wurden *drei Indikatoren* für Adhärenz gebildet: (1) ein Indikator bezogen auf die in jeder Unterrichtsstunde eingesetzten Medien (s. Anlage, Item 3); (2) ein weiterer Indikator bezogen auf die in jeder Unterrichtsstunde eingesetzten Arbeitsformen (s. Anlage, Item 4); (3) und ein dritter Indikator schließlich für das *spezielle Schülerheft* (s. Anlage, Item 3). Diesem dritten Indikator liegt die Überlegung zugrunde, dass den speziell für die beiden Unterrichtsinterventionen entwickelten Schülerarbeitsheften eine zentrale Bedeutung bei der Programmkonzeption zur Autonomie- und Kompetenzunterstützung zugeschrieben wird (Mittag et al., 2009).

Zur Quantifizierung der drei Indikatoren waren zunächst die Soll-Werte für jede Unterrichtsstunde entsprechend den Programmvorgaben zu bestimmen. Diese erfolgte anhand der Lehrerhefte, die den Lehrkräften als Programmmanual zur Durchführung der Unterrichtsinterventionen vorlagen und neben fachlichen und organisatorischen Aspekten zur Unterrichtsdurchführung auch die genaue Verlaufsplanung zu jeder Unterrichtsstunde sowie Handlungsvarianten enthielten. Die jeweiligen Soll-Werte wurden von zwei unabhängigen Ratern aufgrund der dort aufgeführten Angaben bestimmt. Aus dem Vergleich mit den Angaben der Lehrkräfte im Feedback-Fragebogen ergeben sich dann Werte als *Ist-Soll-Diskrepanzen* für jede Unterrichtsstunde. Diese Werte wurden anschließend über alle gehaltenen Unterrichtsstunden der jeweiligen Unterrichtseinheit aufsummiert. Negative und positive Werte indizieren je nach Ausprägung mehr oder weniger starke Abweichungen von den Vorgaben im Lehrerheft (s. *Tabelle 5*). Ein Wert von *Null* zeigt hingegen an, dass die Programmvorgaben eingehalten wurden und somit Adhärenz vorliegt.

In Abgrenzung dazu wurde zur Quantifizierung der **Dosierung** ausschließlich die Einsatzhäufigkeit herangezogen, d. h. je häufiger Medien, Arbeitsformen und das *spezielle Schülerheft* eingesetzt wurden, desto höher der entsprechende Häufigkeitsindex für die Dosierung (s. *Tabelle 5*). Die Bildung von reinen Häufigkeitsindizes begründet sich dadurch, dass in zahlreichen Anwendungsbereichen die Häufigkeit der eingesetzten Maßnahmen sowie die Dauer der Intervention eine positive Beziehung zur Wirksamkeit aufweisen. Da bei der Umsetzung der Unterrichtseinheiten zwischen den Klassen zudem Unterschiede in der Anzahl der durchgeführten Unterrichtsstunden zu beobachten waren, wurden drei Häufigkeitsindizes für *Medien* (1) *Arbeitsformen* (2) und das *spezielle Schülerheft* (3) gebildet.

Als Maß für die **Qualität der Programmversorgung** können nach Dane und Schneider (1998, S. 45) u.a. auch Einschätzungen der Programmvermittler zur Wirksamkeit der einzelnen Programmsitzungen herangezogen werden. Dazu hatten die Lehrkräfte im Feedback-Fragebogen anzugeben, wie aus ihrer Sicht jede Unterrichtsstunde von den Schüler\_innen im Hinblick auf Interesse, Lernfreude, Selbstbestimmung, Kompetenzerleben und soziales Miteinander angenommen wurde (s. Anhang, Item 6). Selbstbestimmung, Kompetenzerleben und soziales Miteinander wie auch Interesse und Lernfreude repräsentieren zentrale Elemente der Selbstbestimmungstheorie und sind daher als zentrale Zielvariablen der Unterrichtsinterventionen anzusehen. Dementsprechend wurden die lehrkraftbezogenen Einschätzungen zu diesen fünf Zielvariablen zu einem Skalenmittelwert über alle durchgeführten Unterrichtsstunden verrechnet und als Indikator zur Bewertung für die Qualität der Programmversorgung in weiteren Analysen verwendet. Daneben wurden auch die Einschätzungen der Lehrkräfte zur ersten Aussage in Item 6 („Wie gut ist Ihrer Meinung nach die Durchführung der Unterrichtsstunde gelungen?“) zu einem Skalenmittelwert über alle Unterrichtsstunden verrechnet und als weiterer Indikator einbezogen.

Die **Responsivität der Programmteilnehmenden** wird von Dane und Schneider (1998) als Empfänglichkeit und Bereitschaft thematisiert, sich nicht nur passiv, sondern vor allem aktiv mit den Inhalten und Maßnahmen eines Programmes während der Durchführung auseinander zu setzen und mitzuarbeiten. Aus unserer Sicht können dazu die Einschätzungen der Lehrkräfte im Feedback-Fragebogen zur schülerseitigen Mitarbeit und zu Unterrichtsstörungen herangezogen werden (s. Anlage, Item 5). Für die weiteren Analysen wurden die lehrkraftbezogenen Angaben nach Umpolung der Einschätzungen zu den Unterrichtsstörungen zu einem Skalenmittelwert verrechnet und als Indikator für die Responsivität der Schüler\_innen

verwendet. Die Items in den Feedback-Fragebögen wurden von zwei unabhängigen Personen dahingehend eingeschätzt, welche Materialien und Arbeitsformen laut Manual zu verwenden waren. Um die Übereinstimmungen zwischen den Beurteilern zu prüfen, wurde das Übereinstimmungsmaß Kappa (Cohen, 1960) berechnet. Der Wertebereich von Cohen's Kappa liegt zwischen 0.00 (keine Übereinstimmung) bis 1.00 (perfekte Übereinstimmung) Werte von  $\geq .60$  weisen auf eine ausreichend hohe Übereinstimmung hin (Altman, 1991). Es fand sich für alle einbezogenen Daten der Unterrichtseinheit 1 ( $\kappa = .71$ ) und der Unterrichtseinheit 2 ( $\kappa = .85$ ) eine gute bis sehr gute Übereinstimmung zwischen den Ratern. In der folgenden Tabelle 5 finden sich die jeweiligen Items zur Bildung der einzelnen Indikatoren der Implementationsqualität und die Vorgehensweise bei der Skalenbildung.

Tabelle 5

*Indikatoren zur Bewertung der Implementationsqualität der Unterrichtsinterventionen*

<b>Items</b>	<b>Aspekte der Implementationsqualität</b>	<b>Skalenbildung</b>
Welche <b>Medien</b> haben Sie eingesetzt? (a) insgesamt (b) nur <b>Schülerheft</b>	Adhärenz	<b>Differenzwerte:</b> Summe der <b>Ist-Werte</b> (Feedback-Fragebögen) - Summe der <b>Soll-Werte</b> (Lehrermanual)
	Dosierung	<b>Summe der Medien</b> über alle U-Stunden
Welche <b>Arbeitsformen</b> kamen in der Stunde vor?	Adhärenz	<b>Differenzwerte:</b> Summe der <b>Ist-Werte</b> (Feedback-Fragebögen) - Summe der <b>Soll-Werte</b> (Lehrermanual)
	Dosierung	<b>Summe der Arbeitsformen</b> über alle U-Stunden

Fortsetzung Tabelle 5

*Indikatoren zur Bewertung der Implementationsqualität der Unterrichtsinterventionen*

<b>Items</b>	<b>Aspekte der Implementationsqualität</b>	<b>Skalenbildung</b>
Wie haben die Schüler_innen diese Unterrichtsstunde wahrgenommen?	Qualität der Programmvermittlung 1	<b>Mittelwert</b> über alle U-Stunden: <b>Interesse, Lernfreude, Selbstbestimmung, Kompetenzerleben und soziales Miteinander</b>
	Responsivität 1	<b>Mittelwert</b> über alle U-Stunden: Einschätzungen der Lehrkräfte zur <b>Mitarbeit</b> der Schüler_innen
	Responsivität 2	<b>Mittelwert</b> über alle U-Stunden: Einschätzungen der Lehrkräfte zu <b>Unterrichtsstörungen</b>
Ihre Einschätzung zur Unterrichtsstunde insgesamt?	Qualität der Programmvermittlung 2	<b>Mittelwert</b> über alle U-Stunden

### 4.2.3 Schülerskalen zu Erfassung der Wirksamkeitsindikatoren

Die zur Erfassung der Schülervariablen eingesetzten Instrumente waren für alle Messzeitpunkte für Treatment- und Kontrollklassen identisch. Es wurden getrennt von den Leistungstests standardisierte Fragebögen eingesetzt, die während einer Unterrichtsstunde von den Schüler\_innen ausgefüllt wurden. Dies erfolgte vor der ersten Unterrichtseinheit, unmittelbar nach dieser Unterrichtseinheit und direkt nach der zweiten Unterrichtseinheit. Es wurden die Kriteriumsvariablen wahrgenommene Autonomieunterstützung, erlebte Lehrerfürsorglichkeit, Lernfreude und motivationale Selbstregulation mittels Selbstberichtsfragebogen erhoben. Die Fragen, die bezogen auf den Deutschunterricht beantwortet werden sollten, waren auf einer vierstufigen Antwortskala von *trifft gar nicht zu* (0) bis *trifft völlig zu* (3) beantwortet.



Verwendet wurde die Skala *Autonomieunterstützung im Unterricht* (modifizierte Skala nach Röder & Jerusalem, 2007) mit fünf Items. Ein Beispielitem zu dieser Skala lautet: „In meinem Unterricht können wir oft unter verschiedenen Themen auswählen.“ Die Skala weist eine interne Konsistenz von  $\alpha = .77$  auf.

Um zu erfassen, in welchem Umfang die Schüler die Lehrkräfte als unterstützend und kooperationsbereit wahrnehmen, wurde die fünf Items umfassende Skala *Fürsorglichkeit* (gekürzte Version nach v. Saldern et al., 1987) (z. B. „Unser Lehrer kümmert sich um die Probleme der Schüler.“,  $\alpha = .84$ ) verwendet.

Die Skala *Lernfreude* (Laukenmann et al., 2003) erfasst mit vier Items die Freude am Lernen und die Zufriedenheit mit dem Deutschunterricht ( $\alpha = .89$ ). Ein Beispielitem der Skala ist: „Der Unterricht im Fach Deutsch macht mir Freude.“

Mithilfe der vier Subskalen des Fragebogens zur *Motivationalen Selbstregulation* (MoS-J 16, Bieg & Mittag, 2005) wurden die vier Stufen (external, introjiziert, identifiziert und intrinsisch) der motivationalen Regulation erfasst. Der Fragebogen umfasst drei typische Lernsituationen mit je vier Items. Die Werte der vier Subskalen können zusätzlich zur Bildung eines Selbstbestimmungsindex herangezogen werden (Levesque, Zuehlke, Stanek & Ryan, 2004; Vallerand, 1997 nach Mittag et al., 2009). Der Selbstbestimmungsindex wird über die gewichteten Skalenmittelwerte berechnet (Formel:  $SDI = [2x \text{ intrinsisch}] + [\text{identifiziert}] - [\text{introjiziert}] - [2x \text{ external}]$ ). Die Skalen weisen eine interne Konsistenz im Bereich ( $\alpha = .75$ ) für die introjizierte Regulation und ( $\alpha = .87$ ) für die intrinsische Regulation auf (Mittag et al., 2009).

Zusätzlich zu den Selbstberichtsdaten der Schüler\_innen wurden Leistungsdaten erhoben. Für die Unterrichtseinheit *Grammatik und Kommasetzung* wurden Testaufgaben aus schülersprachigen, von den Schülern selbst konzipierten Textvorlagen verwendet. Für die Unterrichtseinheit *Sachtexte erschließen* wurden den Schülern unbekannte Beispieltex te vorgelegt. Mithilfe der Retest reliabilitäten wurden die Tests auf Zuverlässigkeit geprüft. Die Korrelation des jeweiligen Posttest mit dem jeweiligen Prätest betrug  $r = .72$  für die erste Unterrichtseinheit und  $r = .50$  für die zweite Unterrichtseinheit (jeweils  $p < .01$ ). Die Korrelation zwischen den Posttests und den Follow-up Werten betrug für die erste Unterrichtseinheit  $r = .81$  und für die zweite Unterrichtseinheit  $r = .54$  (jeweils  $p < .01$ ).

### 4.3 Auswertungsverfahren

Die Auswertung der Daten erfolgt in mehreren Schritten. Um zu überprüfen, ob zwischen den einzelnen Indikatoren bzw. Teilindikatoren der Implementationsqualität bedeutsame Zusammenhänge bestehen, werden in einem ersten Schritt Interkorrelationen berechnet. In einem zweiten Schritt geht es um die Klärung der Frage, ob und inwieweit sich auch zwischen den Indikatoren zur Implementationsqualität und schülerseitigen Wirksamkeitsindikatoren bedeutsame Zusammenhänge nachweisen lassen. Dazu werden zunächst die klassenbezogenen Implementationsindikatoren mit den klassenbezogenen Effektgrößen (s. *Tabellen 8 - 11*) korreliert. Auf der Basis dieser Analysen werden für die zuvor als signifikant erwiesenen Zusammenhänge längsschnittliche Wachstumsmodelle modelliert. Als statistisches Verfahren bietet sich hierzu die Mehrebenenanalyse an. Mehrebenenanalysen erlauben die Schätzung individueller Wachstumsmodelle. Dabei kann neben einer linearen auch eine quadratische Wachstumskomponente geprüft werden. Im vorliegenden Datensatz gibt es längsschnittliche Daten zu den jeweiligen Messzeitpunkten auf Schüler\_innenebene (Ebene 1). Zudem gibt es Lehrerdaten zur Implementationsgüte auf Klassenebene (Ebene 2). Auf der ersten Ebene werden die zeitlich aufeinander folgenden Messzeitpunkte modelliert (z. B. die Leistungsdaten zum Prätest, Posttest und Follow-up). In diesem Fall werden die Messzeitpunkte innerhalb der Person geschachtelt. Auf der zweiten Ebene werden dann die Lehrerdaten zur Implementationsgüte einbezogen und im Hinblick auf ihre Effekte auf die längsschnittlichen Veränderungen geprüft. Aufgrund des relativ geringen Stichprobenumfangs auf Ebene 2 (Klassenebene) haben die mehrebenenanalytischen Ergebnisse hier allerdings eher einen explorativen Status.

## 5 Ergebnisse

Ein Anliegen dieser Arbeit war die Einordnung der Lehrerangaben zur Umsetzung der Unterrichtseinheiten. Hierzu wurden in einem ersten Schritt Indikatoren bzw. Teilindikatoren zur Implementationsqualität gebildet.

Für die mehrbenenanalytischen Berechnungen wurden zur besseren Interpretierbarkeit der Ergebnisse die Kennwerte der jeweils einbezogenen Implementationsindikatoren an ihrem Gesamtmittel standardisiert (Mittelwerte liegen damit jeweils bei null). Die Zusammenhangsanalysen wurden mit dem Statistikprogramm SPSS 23.0 durchgeführt. Die Mehrebenenanalysen wurden mit der Software HLM, Version 7 (Raudenbush, Bryk, Cheong & Congdon, 2011) berechnet.

### 5.1 Interkorrelationen der erfassten Implementationsindikatoren

Im Folgenden werden die Interkorrelationen der erfassten Implementationsindikatoren der Unterrichtseinheit 1 (Abschnitt 5.1.1) und der Unterrichtseinheit 2 (Abschnitt 5.2.1) vorgestellt.

Als Kennwerte der Implementationsqualität wurden die folgenden Indikatoren bzw. Teilindikatoren einbezogen: Adhärenz Medien, Adhärenz Arbeitsformen, Adhärenz spezielles Schülerheft, Dosierung Medien, Dosierung Arbeitsformen, Dosierung spezielles Schülerheft, Qualität der Programmvermittlung 1, Qualität der Programmvermittlung 2, Responsivität 1 und Responsivität 2.

Die Teilaspekte eines Indikators sollten vergleichsweise hohe Zusammenhänge aufweisen (positive Intra-Indikatoren-Zusammenhänge) (Hypothese 1a). Die Indikatoren (bzw. Teilaspekte der Indikatoren) sollten vergleichsweise niedrige Zusammenhänge mit anderen Indikatoren (bez. Teilaspekten anderer Indikatoren) aufweisen (relative Inter-Indikatoren-Unabhängigkeit). Eine Ausnahme bilden die Indikatoren Adhärenz und Dosierung. Aufgrund der Ähnlichkeit ihrer Operationalisierungen werden positive Zusammenhänge zwischen beiden Indikatoren (bzw. Teilaspekten beider Indikatoren) vermutet. So erwarten wir für die Teilaspekte der Adhärenz (erfasst über Differenzwerte: Summe der Ist-Werte (Lehrerfeedbackbögen) -

Summe der Soll-Werte (Lehrermanual) und die Teilaspekte der Dosierung (Summe der Medien/bzw. Arbeitsformen über alle Unterrichtsstunden) bedeutsame korrelative Zusammenhänge (Hypothese 1b).

### 5.1.1 Interkorrelationen der erfassten Implementationsindikatoren UE 1

Tabelle 6 zeigt die Interkorrelationen der erfassten Implementationsindikatoren zur Unterrichtseinheit 1:

Tabelle 6  
Interkorrelationen der erfassten Implementationsindikatoren UE 1

Implementationsindikatoren	Adhärenz			Dosierung			Qual. Programmvermittlung		Responsivität	
	Medien	Arbeitsf.	S-Heft	Medien	Arbeitsf.	S-Heft	1	2	1	2
Adhärenz Medien	<b>.71<sup>t</sup></b>	.58		.56	.25	.23	-.52	.08	-.07	.42
Adhärenz Arbeitsformen		.48		<b>.84**</b>	<b>.66<sup>t</sup></b>	-.12	-.29	.51	.05	.16
Adhärenz Schülerheft				.31	.10	.20	-.59	.31	.03	.47
Dosierung Medien					<b>.93***</b>	.03	-.09	.55	.33	.23
Dosierung Arbeitsformen						.10	.08	.49	.28	-.03
Dosierung Schülerheft							-.34	-.52	-.54	-.33
Qual. Programmvermittlung 1								.15	.13	.05
Qual. Programmvermittlung 2									.52	.46
Responsivität 1										<b>.62<sup>t</sup></b>
Responsivität 2										

Anmerkungen.  $N = 8$ . Pearson-Korrelationen. \*\*\*  $p < .001$ , \*\*  $p < .01$ , \*  $p < .05$ , <sup>t</sup>  $p < .10$

### Intra-Indikatoren-Zusammenhänge

Die Ergebnisse bivariater Korrelationen weisen bei den Indikatoren Adhärenz, Dosierung und Responsivität der Programmteilnehmer\_innen erwartungsgemäß vergleichsweise positive Intra-Indikatoren-Zusammenhänge auf (Hypothese 1a). Beim Indikator Adhärenz korrelieren die Teilaspekte *Medien* und *Arbeitsformen* tendenziell positiv ( $r = .71, p < .10$ ). Beim dritten Teilaspekt *Schülerheft* werden die Zusammenhänge nicht signifikant. Innerhalb des Indikators Dosierung korrelieren die Teilaspekte *Medien* und *Arbeitsformen* positiv ( $r = .93, p < .001$ ), nicht jedoch wiederum der Teilaspekt *Schülerheft*. Tendenziell signifikant wird die Korrelation der beiden Teilaspekte des Indikators Responsivität der Programmteilnehmer (*Responsivität 1*: Mittelwert über aller U-Stunden: Einschätzungen der Lehrkräfte zur Mitarbeit der Schüler\_innen, *Responsivität 2*: Mittelwert über alle U-Stunden: Einschätzungen der Lehrkräfte zu Unterrichtsstörungen) ( $r = .62; p < .10$ ). Erwartungswidrig weisen die beiden Teilaspekte des Indikators Qualität der Programmvermittlung keinen Zusammenhang auf ( $r = .15, n.s.$ ).

### Inter-Indikatoren-Unabhängigkeit

Erwartungskonform verpassen die Zusammenhänge zwischen den Indikatoren Qualität der Programmvermittlung und Responsivität der Programmteilnehmer (bzw. deren Teilaspekte) die Vertrauensgrenzen (Hypothese 1b). Ebenfalls zeigen sich keine signifikanten Zusammenhänge zwischen diesen beiden Indikatoren und den Indikatoren Adhärenz und Dosierung.

Abweichend von dieser relativen Unabhängigkeit der Indikatoren zeigen sich allerdings erwartungskonform Zusammenhänge zwischen Teilaspekten der Indikatoren Adhärenz und Dosierung. Eine signifikante Beziehung findet sich für die *Adhärenz der Arbeitsformen* zur *Dosierung der Medien* ( $r = .84; p < .01$ ). Tendenziell signifikant wird der korrelative Zusammenhang der *Adhärenz der Arbeitsformen* mit *Dosierung der Arbeitsformen* ( $r = .66; p < .10$ ). Keine Zusammenhänge zeigen sich für den Aspekt der *Schülerhefte*.

### 5.1.2 Implementationsqualität und Wirksamkeit: Interkorrelationen der erfassten Implementationsindikatoren UE 2

Tabelle 7 zeigt die Interkorrelationen der erfassten Implementationsindikatoren der Unterrichtseinheit 2:

Tabelle 7

*Interkorrelationen der erfassten Implementationsindikatoren der UE 2*

Implementationsindikatoren	Adhärenz			Dosierung			Qual. Programmvermittlung		Responsivität	
	Medien	Arbeitsf.	S-Heft	Medien	Arbeitsf.	S-Heft	1	2	1	2
Adhärenz Medien		.40	<b>.79*</b>	<b>.91**</b>	<b>.69*</b>	<b>.74*</b>	-.18	-.32	.10	-.36
Adhärenz Arbeitsformen			.37	.33	<b>.77*</b>	.40	-.33	-.33	-.19	-.28
Adhärenz Schülerheft				.51	.43	<b>.99***</b>	-.23	-.57	-.42	-.29
Dosierung Medien					<b>.72*</b>	.46	-.05	-.02	.46	-.29
Dosierung Arbeitsformen						.43	-.28	-.28	.09	-.57
Dosierung Schülerheft							-.20	-.58	-.46	-.30
Qual. Programmvermittlung 1								<b>.77*</b>	.45	-.06
Qual. Programmvermittlung 2									<b>.81**</b>	.13
Responsivität 1										.10
Responsivität 2										

*Anmerkungen. N = 9. Pearson-Korrelationen. \*\*\*  $p < .001$ , \*\*  $p < .01$ , \*  $p < .05$ , †  $p < .10$*

### Intra-Indikatoren-Zusammenhänge

Die Ergebnisse bivariater Korrelationen weisen bei den Indikatoren Adhärenz, Dosierung und Qualität der Programmvermittlung erwartungsgemäß vergleichsweise positive Intra-Indikatoren-Zusammenhänge auf (Hypothese 1a). Beim Indikator Adhärenz korrelieren die Teilaspekte *Medien* und *Schülerheft* tendenziell positiv ( $r = .79, p < .05$ ). Beim dritten Teilaspekt *Arbeitsformen* werden die Zusammenhänge nicht signifikant. Innerhalb des Indikators Dosierung korrelieren die Teilaspekte *Medien* und *Arbeitsformen* positiv ( $r = .72, p < .05$ ), nicht jedoch wiederum der Teilaspekt *Schülerheft*. Signifikant wird die Korrelation der beiden Teilaspekte des Indikators Qualität der Programmvermittlung (*Qualität der Programmvermittlung 1*: Mittelwert über alle Unterrichtsstunden für Interesse, Lernfreude, Selbstbestimmung, Kompetenzerleben und soziales Miteinander; *Qualität der Programmvermittlung 2*: Mittelwert über alle Unterrichtsstunden aus den Einschätzungen der Lehrkräfte zur Unterrichtsstunde insgesamt) ( $r = .77; p < .05$ ). Erwartungswidrig weisen die beiden Teilaspekte des Indikators Responsivität der Programmteilnehmer keinen Zusammenhang auf ( $r = .10, n. s.$ ).

### Inter-Indikatoren-Unabhängigkeit

Erwartungsgemäß verfehlen die Zusammenhänge der Adhärenz, der Dosierung, der Qualität der Programmvermittlung und der Responsivität der Programmteilnehmer die Vertrauensgrenzen.

Abweichend von dieser relativen Unabhängigkeit der Indikatoren zeigen sich allerdings erwartungskonform Zusammenhänge zwischen Teilaspekten der Indikatoren Adhärenz und Dosierung. Eine signifikante Beziehung findet sich für die *Adhärenz der Medien* zur *Dosierung der Medien* ( $r = .91; p < .01$ ), der *Dosierung der Arbeitsformen* ( $r = .69; p < .05$ ) und der *Dosierung des speziellen Schülerheftes* ( $r = .74; p < .05$ ). Signifikant wird der korrelative Zusammenhang der *Adhärenz der Arbeitsformen* mit *Dosierung der Arbeitsformen* ( $r = .77; p < .05$ ). Hochsignifikant wird der Zusammenhang der *Adhärenz des Schülerheftes* mit der *Dosierung des Schülerheftes* ( $r = .99; p < .001$ ).

Nicht erwartungskonform zeigen sich signifikant positive Zusammenhänge zwischen den Indikatoren Qualität der Programmvermittlung und Responsivität der Programmteilnehmer (bzw. deren Teilaspekten) ( $r = .81; p < .01$ ) (Hypothese 1b).

### 5.1.3 Vergleichende Betrachtung der Zusammenhänge der Implementationsindikatoren in UE 1 und UE 2

Bisher wurden die Zusammenhänge zwischen den Teilaspekten einer Indikatorengruppe und die Zusammenhänge dieser Teilaspekte mit den anderen Indikatorengruppen getrennt nach den Unterrichtseinheiten betrachtet. Darüber hinaus ist die Frage relevant, ob sich auch Unterschiede für diese Zusammenhänge je nach Unterrichtseinheit zeigen. Die Zusammenhänge werden nun vergleichend betrachtet.

Die Teilaspekte eines Indikators sollten höhere Zusammenhänge miteinander aufweisen als mit den Teilaspekten der anderen Indikatoren. Hohe Zusammenhänge innerhalb einer Indikatorengruppe konnten bei beiden Unterrichtseinheiten nur teilweise beobachtet werden. Für die Indikatorengruppe der Adhärenz fand sich in der Unterrichtseinheit 1 nur für den Teilaspekt *Adhärenz der Medien* eine signifikant positive Korrelation mit der *Adhärenz der Arbeitsformen*. In der Unterrichtseinheit 2 zeigte sich eine signifikant positive Korrelation für die *Adhärenz der Medien* mit der *Adhärenz des Schülerheftes*. Für die Indikatorengruppe der Dosierung findet sich für den Teilaspekt *Dosierung der Medien* bei beiden Unterrichtseinheiten ein signifikant positiver Zusammenhang mit der *Dosierung der Arbeitsformen*. Die *Qualität der Programmvermittlung 1* steht in der Unterrichtseinheit 1 erwartungsgemäß in einer signifikant positiven Beziehung mit *Qualität der Programmvermittlung 2*. Für die Teilaspekte der Implementationsindikatoren *Qualität der Programmvermittlung* und *Responsivität der Programmteilnehmer* wurden niedrige Zusammenhänge erwartet. Bei der Unterrichtseinheit 1 konnte erwartungskonform kein signifikanter Zusammenhang zwischen der Qualität der Programmvermittlung und der Responsivität der Programmvermittler gefunden werden. Bei der Unterrichtseinheit 2 zeigt sich entgegen der Erwartung, dass eine korrelative Beziehung zwischen der *Qualität der Programmvermittlung 2* und der *Responsivität 1* vorliegt.

Es wurde erwartet, dass die Indikatoren der Adhärenz und der Dosierung eine Ausnahme bilden und hohe Zusammenhänge aufweisen. Während in der Unterrichtseinheit 1 die Adhärenz der Medien keine Zusammenhänge zu den Teilaspekten der Dosierung aufwies, zeigten sich in der Unterrichtseinheit 2 signifikante positive Zusammenhänge zu allen Teilaspekten der Dosierung. Setzen die Lehrkräfte die Medien entsprechend den Vorgaben ein, geht dies sowohl mit einer hohen Dosierung im Hinblick auf den Medieneinsatz und die Verwendung der Arbeitsformen als auch die Häufigkeit des Einsatzes des speziellen Schülerheftes einher. Bei der Adhärenz der Arbeitsformen findet sich für beide Unterrichtseinheiten ein signifikant



positiver Zusammenhang mit der *Dosierung der Arbeitsformen*. Die *Adhärenz des Schülerheftes* steht ausschließlich bei der Unterrichtseinheit 2 in einer hochsignifikant positiven Beziehung zur *Dosierung des Schülerheftes*. Die höchste korrelative Beziehung findet sich so über beide Unterrichtseinheiten hinweg nicht innerhalb einer Indikatorengruppe, sondern für *Adhärenz des Schülerhefts* mit der *Dosierung des Schülerheftes*.

## **5.2 Korrelationen zwischen den Implementationsindikatoren und den klassenbezogenen Effektgrößen**

Nachfolgend werden die Korrelationen zwischen den Implementationsindikatoren und klassenbezogenen Effektgrößen vorgestellt, um die Wirksamkeit der Unterrichtseinheiten 1 und 2 zu bewerten. Um die Veränderungen zwischen den einzelnen Messzeitpunkten zu dokumentieren, wurden zunächst die klassenbezogenen Effektgrößen in Bezug auf die Wirksamkeitsindikatoren Autonomie, Lehrerfürsorglichkeit, Lernfreude, Lernmotivation ermittelt und so getrennt nach Unterrichtseinheiten betrachtet (Unterrichtseinheit 1: t1 bis t2; Unterrichtseinheit 2: t2 bis t3). Zur Berechnung der Effektgrößen für die Leistung der jeweiligen Unterrichtseinheit wurden die Prätest, Posttest und Follow-up Werte herangezogen (Unterrichtseinheit 1: Prätest I bis Posttest I, Posttest I bis Follow-up I, Prätest I bis Follow-up I; Unterrichtseinheit 2: Prätest II bis Posttest II, Posttest II bis Follow-up II, Prätest II bis Follow-up II; s. dazu auch *Abbildung 7*).

### **5.2.1 Korrelationen zwischen den Implementationsindikatoren und den klassenbezogenen Effektgrößen bei den Wirksamkeitsindikatoren UE 1**

Die Tabelle 8 zeigt die Korrelationen zwischen den Implementationsindikatoren und den klassenbezogenen Effektgrößen bei den Wirksamkeitsindikatoren der Unterrichtseinheit 1.

Tabelle 8  
 Korrelationen zwischen den Implementationsindikatoren der UE 1 und den klassenbezogenen Effektgrößen bei den Wirksamkeitsindikatoren

Implementationsindikatoren	Wirksamkeitsindikatoren			
	Autonomie t1 bis t2	Fürsorglichkeit t1 bis t2	Lernfreude t1 bis t2	Lernmotivation t1 bis t2
Adhärenz Medien	-.20	.11	-.29	-.05
Adhärenz Arbeitsformen	-.37	-.22	<b>-.78*</b>	.22
Adhärenz Schülerheft	-.10	-.45	-.38	-.08
Dosierung Medien	-.14	.10	<b>-.69<sup>t</sup></b>	-.21
Dosierung Arbeitsformen	.06	.20	<b>-.70<sup>t</sup></b>	-.29
Dosierung Schülerheft	<b>.78*</b>	.61	-.01	-.60
Qual. Programmvermittlung 1	.04	.27	.30	.22
Qual. Programmvermittlung 2	<b>-.68<sup>t</sup></b>	-.54	-.01	.30
Responsivität 1	-.37	-.22	-.01	-.32
Responsivität 2	-.46	-.19	.25	-.05

Anmerkungen.  $N = 8$ . Pearson-Korrelationen \*  $p < .05$ , <sup>t</sup>  $p < .10$

Erwartet wurden positive Zusammenhänge zwischen den Implementationsindikatoren und den Wirksamkeitsindikatoren zur Einschätzung des Unterrichts (Autonomie, Fürsorglichkeit) (Hypothese 2 a) und der Motivation (Lernfreude und Lernmotivation) (Schülerselbstbericht) (Hypothese 2b). Die Ergebnisse weisen überwiegend auf nicht belegbare Zusammenhänge hin. Erwartungskonform korreliert lediglich die *Dosierung des Schülerhefts* signifikant positiv mit der klassenbezogenen Effektgröße zum *Autonomieerleben* ( $r = .78^*$ ;  $p < .05$ ) vom ersten zum zweiten Messzeitpunkt. Nicht erwartungskonform zeigen sich hingegen eine signifikant negative Korrelation zur *Lernfreude* für die *Adhärenz der Arbeitsformen* ( $r = -.78^*$ ;  $p < .05$ ) sowie tendenziell signifikant negative Zusammenhänge für die *Dosierung der Medien* ( $r = -.69^t$ ;  $p < .10$ ) und die *Dosierung der Arbeitsformen* ( $r = -.70^t$ ;  $p < .10$ ).

Erwartungswidrig ist auch die tendenziell negative Beziehung zwischen der *Qualität der Programmvermittlung 2* (Einschätzung der Unterrichtsstunde insgesamt) und dem *Autonomieerleben der Schüler* ( $r = -.68^t$ ;  $p < .10$ ). Dies bedeutet, dass mit zunehmend positiven Einschätzungen der Unterrichtsstunden seitens der Lehrkräfte eher ungünstigere (negative) Entwicklungen im Autonomieerleben der Schüler\_innen einhergehen.

Tabelle 9

*Korrelationen zwischen den Implementationsindikatoren der UE 1 und den klassenbezogenen Effektgrößen bei dem Wirksamkeitsindikator Leistung*

Implementations- indikatoren UE 1	Wirksamkeitsindikator		
	Prätest I bis Posttest I	Leistung Posttest I bis Follow-up I	Prätest I bis Follow-up I
Adhärenz Medien	-.19	.10	.02
Adhärenz Arbeitsformen	.45	<b>.74<sup>t</sup></b>	.66
Adhärenz Schülerheft	-.22	.11	.07
Dosierung Medien	.61	.60	<b>.69<sup>t</sup></b>
Dosierung Arbeitsformen	<b>.80*</b>	.66	<b>.80*</b>
Dosierung Schülerheft	-.09	-.17	-.01
Qual. Programmverm. 1	.33	.27	.23
Qual. Programmverm. 2	.30	.57	.43
Responsivität 1	.14	-.23	-.09
Responsivität 2	-.42	-.27	-.36

*Anmerkungen. N = 7. Pearson-Korrelationen \*  $p < .05$ , <sup>t</sup>  $p < .10$*

Erwartet wurden positive Zusammenhänge zwischen den Implementationsindikatoren und den Veränderungen in der Leistung (Hypothese 3). Diese fand sich signifikant nur für drei der betrachteten Beziehungen.

Zwischen der Effektgröße der *Leistung* findet sich vom Posttest bis zum Follow-up eine tendenziell signifikant positive Beziehung zur *Adhärenz der Arbeitsformen* ( $r = .74^t$ ;  $p < .10$ ). Tendenzuell positiv zeigt sich auch der Zusammenhang zwischen *Dosierung der Medien* mit den klassenbezogenen Effektgrößen zur *Leistungsentwicklung* zwischen Prätest und Follow-

up ( $r = .69^t$ ;  $p < .10$ ). Die *Dosierung der Arbeitsformen* korreliert signifikant positiv mit den klassenbezogenen Effektgrößen zur *Leistungsentwicklung* zwischen Prätest und Posttest ( $r = .80^*$ ;  $p < .05$ ) sowie vom Prätest bis zum Follow-up ( $r = .80^*$ ;  $p < .05$ ). Werden die verschiedenen Arbeitsformen häufiger eingesetzt, zeigt sich eine positivere Leistungsentwicklung der Schüler bis zum Abschluss der Unterrichtseinheit und darüber hinaus bis zum Follow-up. Die Veränderungen zwischen Posttest und Follow-up korrelieren hingegen nicht mit der *Dosierung der Arbeitsformen*. Auch die anderen Implementationsindikatoren weisen keine signifikanten Zusammenhänge mit den klassenbezogenen Effektgrößen in der *Leistung* auf.

### 5.2.2 Korrelationen zwischen den Implementationsindikatoren und den klassenbezogenen Effektgrößen bei den Wirksamkeitsindikatoren UE 2

Tabelle 10 zeigt die korrelativen Zusammenhänge der Implementationsindikatoren mit den klassenbezogenen Effektgrößen bei den Wirksamkeitsindikatoren für die Unterrichtseinheit 2.

Tabelle 10

*Korrelationen zwischen den Implementationsindikatoren der UE 2 und den klassenbezogenen Effektgrößen bei den Wirksamkeitsindikatoren*

Implementationsindikatoren UE 2	Wirksamkeitsindikatoren			
	Autonomie	Fürsorglichkeit	Lernfreude	Lernmotivation
	t2 bis t3	t2 bis t3	t2 bis t3	t2 bis t3
Adhärenz Medien	.45	.14	.28	.29
Adhärenz Arbeitsformen	.17	-.17	-.31	-.41
Adhärenz Schülerheft	<b>.63<sup>t</sup></b>	.25	.40	.38
Dosierung Medien	.08	.05	.20	.13
Dosierung Arbeitsformen	.19	-.16	.05	-.26
Dosierung Schülerheft	<b>.65<sup>t</sup></b>	.19	.42	.30

Fortsetzung Tabelle 10

*Korrelationen zwischen Implementationsindikatoren der UE 2 und den klassenbezogenen Effektgrößen bei den Wirksamkeitsindikatoren*

Implementationsindikatoren UE 2	Wirksamkeitsindikatoren			
	Autonomie	Fürsorglichkeit	Lernfreude	Lernmotivation
	t2 bis t3	t2 bis t3	t2 bis t3	t2 bis t3
Qual. Programmvermittlung 1	.04	.17	.54	-.39
Qual. Programmvermittlung 2	-.53	.03	.08	-.34
Responsivität 1	<b>-.64<sup>t</sup></b>	-.14	-.06	-.21
Responsivität 2	.06	-.07	-.26	-.07

*Anmerkungen. N = 9. Pearson-Korrelationen<sup>t</sup> p < .10*

Auch für die Unterrichtseinheit 2 wurden positive Zusammenhänge zwischen den Implementationsindikatoren und den Wirksamkeitsindikatoren zur Einschätzung des Unterrichts (Autonomie, Fürsorglichkeit) (Hypothese 2a) und der Motivation (Lernfreude und Lernmotivation) (Schülerselbstbericht) (Hypothese 2b) erwartet. Dies konnte nur für drei der Implementationsindikatoren und bei der Entwicklung im Autonomieerleben gefunden werden.

Für die *Adhärenz des Schülerheftes* zeigt sich eine tendenziell positive Korrelation zu den klassenbezogenen Effektgrößen im *Autonomieerleben* (t2 bis t3;  $r = .63^t$ ;  $p < .10$ ). Die *Dosierung des Schülerheftes* korreliert ebenfalls tendenziell positiv mit den Veränderungen im *Autonomieerleben* zwischen zweitem und drittem Messzeitpunkt ( $r = .65^t$ ;  $p < .10$ ). Darüber hinaus zeigt sich eine tendenziell negative Beziehung zwischen der *Responsivität 1* (Mitarbeit) und den Veränderungen im Autonomieerleben vom zweiten bis zum dritten Messzeitpunkt ( $r = -.64^t$ ;  $p < .10$ ). Tabelle 11 zeigt die korrelativen Zusammenhänge der Implementationsindikatoren mit den klassenbezogenen Effektgrößen der Leistung der Schüler\_innen für die Unterrichtseinheit 2.

Tabelle 11  
 Korrelationen zwischen den Implementationsindikatoren UE2 und den klassenbezogenen Effektgrößen bei dem Wirksamkeitsindikator Leistung

Implementations- indikatoren UE 2	Wirksamkeitsindikator		
	Leistung		
	Prätest II bis Posttest II	Posttest II bis Follow-up II	Prätest II bis Follow-up II
Adhärenz Medien	-.49	<b>-.59<sup>t</sup></b>	.28
Adhärenz Arbeitsformen	.06	-.06	.01
Adhärenz Schülerheft	-.56	.36	.05
Dosierung Medien	-.42	<b>.69*</b>	.42
Dosierung Arbeitsformen	-.30	.35	.16
Dosierung Schülerheft	<b>-.59<sup>t</sup></b>	.28	-.06
Qual. Programmvermittlung 1	.11	-.41	-.30
Qual. Programmvermittlung 2	.41	-.13	-.17
Responsivität 1	.15	.31	.40
Responsivität 2	.06	-.02	.15

Anmerkungen.  $N = 9$ . Pearson-Korrelationen \*  $p < .05$ , <sup>t</sup>  $p < .10$

Es wurde ein positiver Zusammenhang der Implementationsindikatoren mit der Leistungsentwicklung der Schüler\_innen erwartet. Die gefundenen Zusammenhänge entsprechen nur bei einem Teilaspekt der Adhärenz den erwarteten Beziehungen.

Für die Effektgrößen beim Wirksamkeitsindikator Leistung findet sich ein signifikant positiver Zusammenhang der *Dosierung der Medien* mit der *Leistungsentwicklung* zwischen Posttest und Follow-up ( $r = .69^*$ ;  $p < .10$ ). Darüber hinaus ergeben sich in der Tendenz signifikant negative Beziehungen zwischen der *Adhärenz der Medien* und der *Leistungsentwicklung* vom Posttest zum Follow-up ( $r = -.59^t$ ;  $p < .10$ ) sowie zwischen der *Dosierung des Schülerhefts* und der *Leistungsentwicklung* vom Prätest zum Posttest ( $r = -.59^t$ ;  $p < .10$ ). Die Zusammenhänge für die Unterrichtseinheit 2 deuten tendenziell an, dass ein „übermäßiger“ Einsatz der Schülerhefte mit geringeren Lernzuwächsen einhergehen kann. Erklärungsbedürftig ist zudem der gegenläufige Befund in Bezug auf die Adhärenz der Medien und der Leistungsentwick-

lung vom Posttest zum Follow-up. Hier scheint sich die starre Einhaltung tendenziell ungünstig auf die Leistungsentwicklung auszuwirken, wobei allerdings eine hohe Dosierung von Medien einen positiven Zusammenhang zur Leistungsentwicklung aufweist.

### 5.2.3 Vergleichende Betrachtung der Zusammenhänge zwischen den Implementationsindikatoren und den klassenbezogenen Wirksamkeitsindikatoren in UE 1 und UE 2

Bisher wurden die Zusammenhänge zwischen den Implementationsindikatoren und den klassenbezogenen Wirksamkeitsindikatoren getrennt nach Unterrichtseinheiten betrachtet. Darüber hinaus ist hier auch die Frage relevant, ob sich hier Unterschiede je nach Unterrichtseinheit zeigen. Im Folgenden werden deshalb die Zusammenhänge für beide Unterrichtseinheiten zusammengefasst dargestellt.

Tabelle 12

*Korrelationen zwischen den Implementationsindikatoren für die Unterrichtseinheiten 1 und 2 und den klassenbezogenen Effektgrößen bei den Wirksamkeitsindikatoren*

Implementationsindikatoren	Wirksamkeitsindikatoren							
	Autonomie		Fürsorglichkeit		Lernfreude		Lernmotivation	
	UE 1 t1-t2	UE 2 t2-t3	UE 1 t1-t2	UE 2 t2-t3	UE 1 t1-t2	UE 2 t2-t3	UE 1 t1-t2	UE 2 t2-t3
<b>Adhärenz</b>								
Medien	-.20	.45	.11	.14	-.29	.28	-.05	.29
Arbeitsformen	-.37	.17	-.22	-.17	<b>-.78*</b>	-.31	.22	-.41
Schülerheft	-.10	<b>.63<sup>t</sup></b>	-.45	.25	-.38	.40	-.08	.38
<b>Dosierung</b>								
Medien	-.14	.08	.10	.05	<b>-.69<sup>t</sup></b>	.20	-.21	.13
Arbeitsformen	.06	.19	.20	-.16	<b>-.70<sup>t</sup></b>	.05	-.29	-.26
Schülerheft	<b>.78*</b>	<b>.65<sup>t</sup></b>	.61	.19	-.01	.42	-.60	.30

Fortsetzung Tabelle 12

Korrelationen zwischen den Implementationsindikatoren für die Unterrichtseinheiten 1 und 2 und den klassenbezogenen Effektgrößen bei den Wirksamkeitsindikatoren

Implementationsindikatoren	Wirksamkeitsindikatoren							
	Autonomie		Fürsorglichkeit		Lernfreude		Lernmotivation	
	UE 1	UE 2	UE 1	UE 2	UE 1	UE 2	UE 1	UE 2
	t1-t2	t2-t3	t1-t2	t2-t3	t1-t2	t2-t3	t1-t2	t2-t3
<b>Qualität der Programmvermittlung</b>								
Qual. Programmverm. 1	.04	.04	.27	.17	.30	.54	.22	-.39
Qual. Programmverm. 2	<b>-.68<sup>t</sup></b>	-.53	-.54	.03	-.01	.08	.30	-.34
<b>Responsivität</b>								
Responsivität 1	-.37	<b>-.64<sup>t</sup></b>	-.22	-.14	-.01	-.06	-.32	-.21
Responsivität 2	-.46	.06	-.19	-.07	.25	-.26	-.05	-.07

Anmerkungen.  $N_{UE1} = 8$ ,  $N_{UE2} = 9$ . Pearson-Korrelationen \*  $p < .05$ , <sup>t</sup>  $p < .10$

Über beide Unterrichtseinheiten konnten positive Zusammenhänge zwischen der *Dosierung des speziellen Schülerheftes* und den Veränderungen im *Autonomieerleben* der Schüler\_innen beobachtet werden. Die *Adhärenz des speziellen Schülerheftes* wirkt sich tendenziell nur in der Unterrichtseinheit 2 positiv auf das *Autonomieerleben* aus. Die *Adhärenz der Arbeitsformen*, die *Dosierung der Medien* als auch die *Dosierung der Arbeitsformen* scheinen in der Unterrichtseinheit 1 tendenziell negative Zusammenhänge mit der *Lernfreude* aufzuweisen, während dies bei der zweiten Unterrichtseinheit nicht der Fall ist. Die *Qualität der Programmvermittlung 2* hat nur in der ersten Unterrichtseinheit eine tendenziell negative Beziehung zu dem *Autonomieerleben* der Schüler\_innen.



Tabelle 13

Korrelationen zwischen den Implementationsindikatoren für die Unterrichtseinheiten 1 und 2 und den klassenbezogenen Effektgrößen bei der Leistung

Implementations- indikatoren	Wirksamkeitsindikator					
	Leistung					
	UE 1	UE 2	UE 1	UE 2	UE 1	UE 2
	Prätest I bis Post- test I	Prätest II bis Post- test II	Posttest I bis Follow- up I	Posttest II bis Follow- up II	Prätest I bis Follow- up I	Prätest II bis Follow- up II
Adhärenz Medien	-.19	-.49	.10	<b>-.59<sup>t</sup></b>	.02	.28
Adhärenz Arbeitsformen	.45	.06	<b>.74<sup>t</sup></b>	-.06	.66	.01
Adhärenz Schülerheft	-.22	-.56	.11	.36	.07	.05
Dosierung Medien	.61	-.42	.60	<b>.69*</b>	<b>.69<sup>t</sup></b>	.42
Dosierung Arbeitsformen	<b>.80*</b>	-.30	.66	.35	<b>.80*</b>	.16
Dosierung Schülerheft	-.09	<b>-.59<sup>t</sup></b>	-.17	.28	-.01	-.06
Qual. Programmvermittlung 1	.33	.11	.27	-.41	.23	-.30
Qual. Programmvermittlung 2	.30	.41	.57	-.13	.43	-.17
Responsivität 1	.14	.15	-.23	.31	-.09	.40
Responsivität 2	-.42	.06	-.27	-.02	-.36	.15

Anmerkungen.  $N_{UE1} = 7$ ,  $N_{UE2} = 9$ . Pearson-Korrelationen \*  $p < .05$ , <sup>t</sup>  $p < .10$

In Bezug auf die Effektgrößen der *Leistung* zeigen sich in der Unterrichtseinheit 1 für die *Dosierung der Arbeitsformen* signifikant positive Zusammenhänge (Posttest bis Follow-up; Prätest bis Follow-up). Tendenziell signifikant werden hier die positiven Zusammenhänge der *Adhärenz der Arbeitsformen* mit der *Leistungsentwicklung* (Posttest I bis Follow-up I) und der *Dosierung der Medien* mit der *Leistungsentwicklung* (Prätest I bis Follow-up I).

Bei der zweiten Unterrichtseinheit weist ebenso die *Dosierung der Medien* einen signifikant positiven Zusammenhang zur *Leistungsentwicklung* auf. Tendenziell ist die *Adhärenz der Medien* in der zweiten Unterrichtseinheit negativ mit der *Leistungsentwicklung* assoziiert.

Ebenso zeigt sich tendenziell eine negative Beziehung der *Dosierung des speziellen Schülerheftes* mit der *Leistung* (Prätest II bis Posttest II). Dies ist bei der Unterrichtseinheit 1 nicht der Fall.

### 5.3 Mehrebenenanalysen

Zur Durchführung der Mehrebenenanalysen wurden die Daten der Schüler\_innen auf der ersten Ebene und die Daten der Lehrkräfte auf der zweiten Ebene über drei Messzeitpunkte in die Analysen einbezogen. Das Vorgehen für die Mehrebenenanalyse orientiert sich an den Empfehlungen von Hox (2010). In einem ersten Schritt wird ein Nullmodell, d.h. ein intercept-only Modell ohne Hinzunahme weiterer Prädiktoren, berechnet. Das Nullmodell kann als Referenzmodell für weitere Modelle unter Hinzunahme von Prädiktoren genutzt werden. Im nächsten Schritt werden dann weitere Prädiktoren in das Modell aufgenommen. So soll schrittweise zusätzliche Varianz in der Kriteriumsvariablen aufgeklärt und Fehlervarianz reduziert werden. Auf der ersten Ebene wurde die Zeit (z. B. Prätest, Posttest und Follow-up bei den Leistungstests), also die Veränderungen über die drei Messzeitpunkte, aufgenommen. So kann mit dem reinen Messwiederholungsmodell ohne Unterschiede in den Implementationsindikatoren eine Aussage zu den Veränderungen bei den schülerseitigen Kriteriumsmaßen gemacht werden. Um individuelle Wachstumsmodelle bestimmen zu können, wurde die Mehrebenenanalyse trotz des geringen Stichprobenumfangs auf Ebene 2 als Auswertungsmethode gewählt. Die Schätzung wurde mittels Restricted Maximum Likelihood (REML) durchgeführt. Dies ermöglicht Analysen auch mit wenigen Gruppen auf Ebene 2 (Browne und Draper, 2000; Maas und Hox, 2004). Die Daten wurden unzentriert in die Analysen einbezogen. Berichtet werden somit die unstandardisierten Beta-Gewichte.

Zur Schätzung der individuellen Wachstumsmodelle wurden eine lineare und in Teilanalysen eine quadratische Wachstumskomponente einbezogen. Eine quadratische Wachstumskomponente ist insbesondere bei der Leistungsentwicklung zu erwarten, da nach Ende der Intervention mit einer Stagnation bzw. zum Follow-up mit einem Leistungsabfall zu rechnen ist. Begründet werden kann dies durch zu erwartende Interferenz- bzw. Überlagerungseffekte nachfolgender Lerninhalte (Hasselhorn & Gold, 2006, S. 56 f.).

Zur Modellierung der linearen Wachstumskomponente wurden der Prätest mit 0, der Posttest mit 1 und das Follow-up mit 2 kodiert. Die Kodierung der quadratischen Wachstumskomponente erfolgte mit 0, 1 und 4. Die Wachstumskomponenten wurden bei der Betrachtung der Leistungsentwicklung additiv in das Modell aufgenommen. Um den Einfluss der Indikatoren auf die Implementationsqualität zu prüfen, wurden diese in einem dritten Schritt auf der zweiten Ebene als Prädiktoren in das jeweilige Modell aufgenommen. Die Prädiktorvariablen wurden für die Analysen nach den Empfehlungen von Hox (2010) am *grand-mean* zentriert. Der Intercept entspricht so dem Wert, den die Kriteriumsvariable annimmt, wenn alle anderen Prädiktoren gleich Null sind. Es wurde zusätzlich geprüft, wie viel Varianz in der Kriteriumsvariablen durch die Varianz der Prädiktorvariablen erklärt werden kann.

Für die Unterrichtseinheit 1 wurde in einem ersten Schritt die Entwicklung des Autonomieerlebens durch die Dosierung des Schülerheftes geschätzt und geprüft. Im Weiteren wurde für die erste Unterrichtseinheit die Leistungsentwicklung durch das Ausmaß an Dosierung der Arbeitsformen geschätzt und geprüft. Schließlich wurde dann mehrebenenanalytisch betrachtet, welchen Einfluss die Adhärenz der Arbeitsformen auf die Entwicklung der Lernfreude bei der Unterrichtseinheit 1 hat. Für die Unterrichtseinheit 2 wurde die Leistungsentwicklung durch das Ausmaß der Dosierung der Medien geschätzt und geprüft. Im Folgenden werden die jeweiligen Modellgleichungen und die Ergebnisse der mehrebenenanalytischen Analysen berichtet. Die Notationen entsprechen den Angaben in HLM 7 Hierarchical Linear and Nonlinear Modeling (Raudenbush, Bryk, & Congdon, 2011).

### 5.3.1 Leistungsentwicklung UE 1

Die zu erwartenden Unterschiede (Varianzen) in der Leistungsentwicklung bei Unterrichtseinheit 1 sollen durch die Dosierung der Arbeitsformen auf der zweiten Ebene aufgeklärt werden. Hierzu wurde zunächst ein Nullmodell berechnet. Anschließend wurde ein Messwiederholungsmodell mit einem linearen Trend und quadratischen Trend analysiert. In einem dritten Schritt wurde dann der Prädiktor Dosierung der Arbeitsformen in das Modell aufgenommen.

#### **Nullmodell:**

#### **Ebene-1 Modell**

$$LEISTUNG\ UE\ I_{it} = \pi_{0i} + e_{it}$$

**Ebene-2 Modell**

$$\pi_{0i} = \beta_{00} + r_{0i}$$

**Gemischtes Modell**

$$LEISTUNG\ UE\ I_{ii} = \beta_{00} + r_{0i} + e_{ii}$$

**Modell 1: Leitungsentwicklung****Ebene-1 Modell**

$$LEISTUNG\ UE\ I_{ii} = \pi_{0i} + \pi_{1i} * (ZEIT_{ii}) + \pi_{2i} * (QUADRIERTE\ ZEIT_{ii}) + e_{ii}$$

**Ebene-2 Modell**

$$\pi_{0i} = \beta_{00} + r_{0i}$$

$$\pi_{1i} = \beta_{10} + \beta_{11} * (DOSIERUNG\ ARBEITSFORMEN\ UE\ I_i)$$

$$\pi_{2i} = \beta_{20}$$

**Gemischtes Modell**

$$LEISTUNG\ UE\ I_{ii} = \beta_{00} + \beta_{10} * ZEIT_{ii} + \beta_{11} * DOSIERUNG\ ARBEITSFORMEN\ UE\ I_i * ZEIT_{ii} + \beta_{20} * QUADRIERTE\ ZEIT_{ii} + r_{0i} + e_{ii}$$

Die Tabelle 14 zeigt die Ergebnisse des berechneten Mehrebenenmodells zur Leistungsentwicklung in der Unterrichteinheit 1.

Tabelle 14  
Leistungsentwicklung UE 1

	<b>Beta</b>	<b>SE</b>
<b>Ebene 1</b>		
Mittelwert Prätest I (Ausgangsniveau)	32.62***	4.61
Lineare Wachstumskomponente	55.89***	1.16
Quadratische Wachstumskomponente	-20.13***	1.33

Fortsetzung Tabelle 14  
*Leistungsentwicklung UE 1*

<b>Ebene 2</b>		
Dosierung Arbeitsformen UE 1	1.22***	0.21
<b>Varianzaufklärung</b>		
R-Quadrat (Ebene 1)	49.1 %	
R-Quadrat (Ebene 2)	53.6 %	
<i>Anmerkung. Restricted Maximum Likelihood. *** <math>p &lt; .001</math>, ** <math>p &lt; .01</math>, * <math>p &lt; .05</math>, ' <math>p &lt; .10</math></i>		

Bei der Parameterschätzung auf der Messwiederholungsebene erweisen sich sowohl die *lineare* als auch die *quadratische Wachstumskomponente* als signifikante Prädiktoren der *Leistungsentwicklung* ( $\gamma_{10} = 55.89$ ,  $p < .001$ ;  $\gamma_{20} = -20.13$ ,  $p < .001$ ). Die negative quadratische Wachstumskomponente steht so für einen Leistungsabfall vom Posttest zum Follow-up.

Der Wert  $\gamma_{00} = 32.62$  ( $p < .001$ ) repräsentiert den mittleren Wert für das Ausgangsleistungsniveau auf Individualebene, soweit alle übrigen Prädiktoren den Wert Null annehmen (Nullmodell). Das Regressionsgewicht von  $\gamma_{11} = 1.22$  ( $p < .001$ ) für die *Dosierung der Arbeitsformen* besagt, dass der lineare Anstieg in der *Leistung* auf der Personenebene um 1.22 Einheiten höher ausfällt, wenn die Dosierung der Arbeitsformen um eine Einheit ansteigt. Wenn man die Standardabweichung einbezieht, dann beträgt der Anstieg 1.22 mal die Standardabweichung bei der Dosierung ( $SD = 5.21$ ) (s. Eid, Gollwitzer & Schmitt, 2015, S. 742 f). Die Schüler\_innen profitieren somit in ihrer Leistungsentwicklung von einer höheren Dosierung der Arbeitsformen.

Durch Hinzunahme der *linearen* und *quadratischen Wachstumskomponente* werden 49.1 % der Varianz auf Personenebene aufgeklärt. Auf der zweiten Ebene werden 53.6 % der Unterschiede der personenspezifischen Leistungsentwicklung durch Unterschiede in der *Dosierung der Arbeitsformen* aufgeklärt.

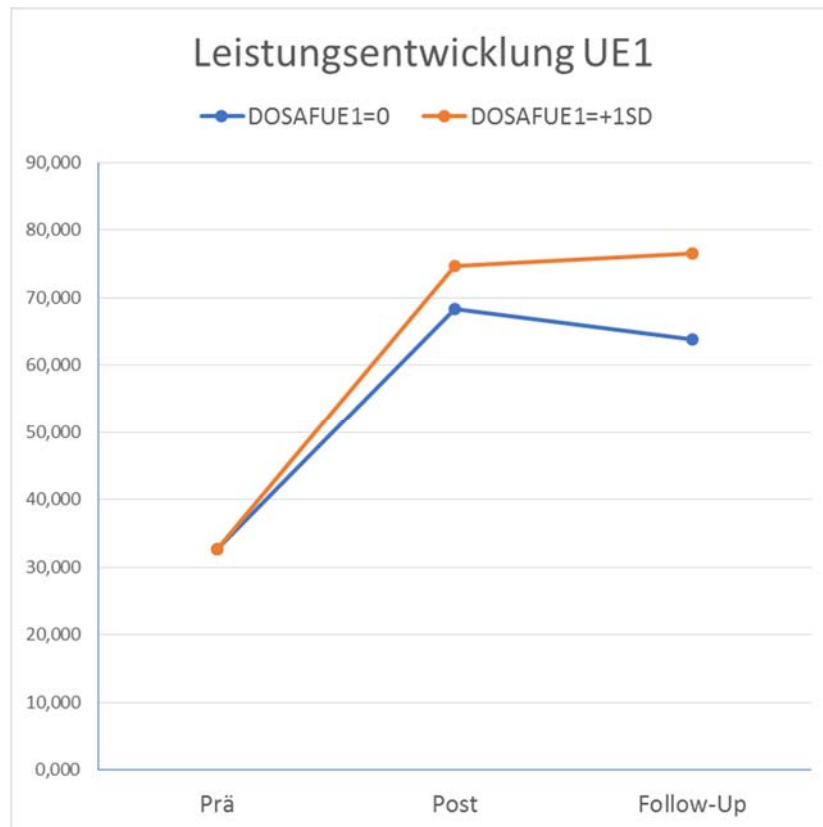


Abbildung 14. Leistungsentwicklung UE1 (DOSAFU1= Dosierung der Arbeitsformen in der Unterrichtseinheit 1)

### 5.3.2 Autonomieentwicklung UE 1

Die zu erwartenden Unterschiede (Varianzen) in der Entwicklung des Autonomieerlebens sollen durch den Einfluss der Dosierung des speziellen Schülerheftes in der Unterrichtseinheit 1 auf der zweiten Ebene aufgeklärt werden. Hierzu wurde erneut ein Nullmodell berechnet. Da hier nur zwei Messzeitpunkte (t1 und t2) einbezogen wurden (s. *Abbildung 7*), wurde in einem zweiten Schritt ein Messwiederholungsmodell mit einem linearen Trend berechnet. In einem dritten Schritt wurde dann der Prädiktor Dosierung des Schülerheftes in das Modell aufgenommen.

**Nullmodell:****Ebene-1 Modell**

$$AUTONOMIEERLEBEN_{ii} = \pi_{0i} + e_{ii}$$

**Ebene-2 Modell**

$$\pi_{0i} = \beta_{00} + r_{0i}$$

**Gemischtes Modell**

$$AUTONOMIEERLEBEN_{ii} = \beta_{00} + r_{0i} + e_{ii}$$

**Modell 2: Autonomieentwicklung****Ebene-1 Modell**

$$AUTONOMIEERLEBEN_{ii} = \pi_{0i} + \pi_{1i} * (ZEIT_{ii}) + e_{ii}$$

**Ebene-2 Modell**

$$\pi_{0i} = \beta_{00} + r_{0i}$$

$$\pi_{1i} = \beta_{10} + \beta_{11} * (DOSIERUNG SCHÜLERHEFT UE1_i)$$

**Gemischtes Modell**

$$AUTONOMIEERLEBEN_{ii} = \beta_{00} + \beta_{10} * ZEIT_{ii} + \beta_{11} * DOSIERUNG SCHÜLERHEFT UE1_i * ZEIT_{ii} + r_{0i} + e_{ii}$$

Die Tabelle 15 zeigt die mehrebenenanalytischen Ergebnisse zum Einfluss des Implementationsindikators Dosierung des speziellen Schülerheftes auf die Entwicklung des Autonomieerlebens der Schüler.

Tabelle 15  
*Autonomieentwicklung UE1*

	Beta	SE
<b>Ebene 1</b>		
Mittelwert zu t1 (Ausgangsniveau)	1.74***	0.08
Lineare Wachstumskomponente	0.08***	0.02
<b>Ebene 2</b>		
Dosierung Schülerheft UE1	0.06***	0.01
<b>Varianzaufklärung</b>		
R-Quadrat (Ebene 1)	2.5 %	
R-Quadrat (Ebene 2)	1.3 %	

*Anmerkung. Restricted Maximum Likelihood. \*\*\*  $p < .001$ , \*\*  $p < .01$ , \*  $p < .05$ , †  $p < .10$*

Es zeigte sich ein Einfluss der *Dosierung des Schülerheftes* auf die Entwicklung des *Autonomieerlebens* der Schüler ( $\gamma_{11} = 0.06, p < .001$ ). Bei der Parameterschätzung auf der Messwiederholungsebene erweist sich zudem die *lineare Wachstumskomponente* als bedeutsam für die Vorhersage der *Autonomieentwicklung*  $\gamma_{10} = 0.08 (p < .001)$ .

Bei dem Wert  $\gamma_{00} = 1.74 (p < .001)$  handelt es sich um das Ausgangsniveau auf Individual-ebene, soweit alle anderen Prädiktoren den Wert Null annehmen (Nullmodell). Das Regressionsgewicht für die *Dosierung des Schülerheftes* ( $\gamma_{11} = 0.06, p < .001$ ) besagt, dass der lineare Anstieg im *Autonomieerleben* um 0.06 Einheiten höher ausfällt, wenn die Dosierung um eine Einheit ansteigt. Es zeigt sich, dass Schüler\_innen aus Klassen, in denen das Schülerheft häufiger eingesetzt wurde, eine positivere Entwicklung des Autonomieerlebens berichten.

Auf der Messwiederholungsebene werden 2.5 % der Varianz durch die *lineare Wachstumskomponente* aufgeklärt. Auf der zweiten Ebene sind es 1.3 % der Varianz in der personenspezifischen Autonomieentwicklung, die durch die Unterschiede in der *Dosierung des Schülerheftes* aufgeklärt werden können.



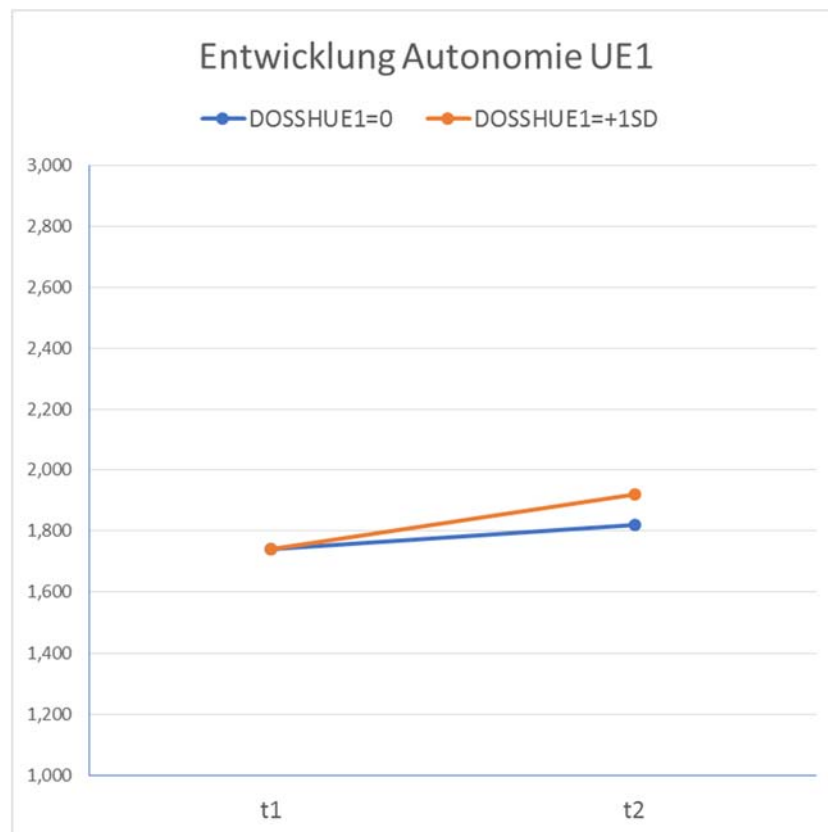


Abbildung 15. Autonomieentwicklung UE1 (DOSSHU1= Dosierung des speziellen Schülerheftes Unterrichteinheit 1)

### 5.3.3 Entwicklung der Lernfreude UE 1

Auch zur Lernfreude wurden für jede Person individuelle Wachstumsmodelle geschätzt, bei denen die Lernfreude über die Zeit betrachtet wird. Da hier nur zwei Messzeitpunkte (t1 und t2) einbezogen wurden (s. *Abbildung 7*), wurde nach dem Nullmodell in einem zweiten Schritt ein Messwiederholungsmodell mit einem linearen Trend berechnet. Unterschiede (Varianz) in der Entwicklung der Lernfreude sollen dann durch den Einfluss der Adhärenz der Arbeitsformen in der Unterrichteinheit 1 auf der zweiten Ebene aufgeklärt werden.

**Nullmodell:****Ebene-1 Modell**

$$LERNFREUDE_{ij} = \beta_{0j} + r_{ij}$$

**Ebene-2 Modell**

$$\beta_{0j} = \gamma_{00} + u_{0j}$$

**Gemischtes Modell**

$$LERNFREUDE_{ij} = \gamma_{00} + u_{0j} + r_{ij}$$

**Modell 3: Lernfreude****Ebene-1 Modell**

$$LERNFREUDE_{ij} = \beta_{0j} + \beta_{1j} * (Zeit_{ij}) + r_{ij}$$

**Ebene-2 Modell**

$$\beta_{0j} = \gamma_{00} + u_{0j}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11} * (ADHÄRENZ\_ARBEITSFORMEN\_UE1_j)$$

**Gemischtes Modell**

$$LERNFREUDE_{ij} = \gamma_{00} + \gamma_{10} * ZEIT_{ij} + \gamma_{11} * ADHÄRENZ\_ARBEITSFORMEN\_UE1_j * ZEIT_{ij} + u_{0j} + r_{ij}$$

In der folgenden Tabelle 16 sind die mehrebenenanalytischen Ergebnisse zum Einfluss des Implementationsindikators Adhärenz der Arbeitsformen auf die Entwicklung der Lernfreude der Schüler\_innen aufgeführt.

Tabelle 16  
*Entwicklung der Lernfreude UE1*

	<b>Beta</b>	<b>SE</b>
<b>Ebene 1</b>		
Mittelwert zu t1 (Ausgangsniveau)	2.68***	0.15
Lineare Wachstumskomponente	-0.09**	0.04
<b>Ebene 2</b>		
Adhärenz Arbeitsformen UE1	-0.03 n. s.	0.02
<b>Varianzaufklärung</b>		
R-Quadrat (Ebene 1)	0.9 %	
<i>Anmerkung. Restricted Maximum Likelihood. *** <math>p &lt; .001</math>, ** <math>p &lt; .01</math>, * <math>p &lt; .05</math>, ' <math>p &lt; .10</math></i>		

Die gefundene negative Korrelation zwischen *Adhärenz der Arbeitsformen* (s. Tabelle 14) bleibt in der mehrbenenanalytischen Betrachtung statistisch unbedeutend ( $\gamma_{11} = -0.03$ ,  $p > .10$ ). Ausschließlich die *lineare Wachstumskomponente* erweist sich als statistisch bedeutsamer Prädiktor für die Entwicklung der *Lernfreude*  $\gamma_{10} = -0.09$  ( $p < .01$ ). Die Lernfreude der Schüler\_innen nimmt zwischen dem ersten und zweiten Messzeitpunkt ab.

Bei dem Wert  $\gamma_{00} = 2.68$  ( $p < .001$ ) handelt es sich um den mittleren Wert für das Ausgangsniveau auf Individualebene, soweit alle Prädiktoren den Wert Null annehmen (Nullmodell). Auf der Messwiederholungsebene konnten 0.9 % der Varianz durch die *lineare Wachstumskomponente* aufgeklärt werden.

### 5.3.4 Leistungsentwicklung UE2

Zur Vorhersage der Leistungsentwicklung in der Unterrichtseinheit 2 wurde erneut ein hierarchisches Modell mit zwei Ebenen gewählt. Die dabei zu erwartenden Unterschiede (Varianzen) in der Leistungsentwicklung sollen durch den Einfluss der Dosierung der Medien auf der zweiten Ebene aufgeklärt werden.

**Nullmodell:****Ebene-1 Modell**

$$LEISTUNG_{UE2ij} = \beta_{0j} + r_{ij}$$

**Ebene-2 Modell**

$$\beta_{0j} = \gamma_{00} + u_{0j}$$

**Gemischtes Modell**

$$LEISTUNG_{UE2ij} = \gamma_{00} + u_{0j} + r_{ij}$$

**Modellbildung: Leistungsentwicklung UE2****Ebene-1 Modell**

$$LEISTUNG_{UE2ij} = \beta_{0j} + \beta_{1j} * (ZEIT_{ij}) + \beta_{2j} * (QUADRIERTE ZEIT_{ij}) + r_{ij}$$

**Ebene-2 Modell**

$$\beta_{0j} = \gamma_{00} + u_{0j}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11} * (DOSIERUNG MEDIEN UE2_j)$$

$$\beta_{2j} = \gamma_{20} + \gamma_{21} * (DOSIERUNG MEDIEN UE2_j)$$

**Gemischtes Modell**

$$LEISTUNG_{UE2ij} = \gamma_{00} + \gamma_{10} * ZEIT_{ij} + \gamma_{11} * DOSIERUNG MEDIEN UE2_j * ZEIT_{ij} + \gamma_{20} * QUADRIERTE ZEIT_{ij} + \gamma_{21} * DOSIERUNG MEDIEN UE2_j * QUADRIERTE ZEIT_{ij} + u_{0j} + r_{ij}$$

Die folgende Tabelle 17 zeigt die Ergebnisse der mehrbenenanalytischen Auswertung zum Einfluss des Implementationsindikators Dosierung der Medien auf die Leistungsentwicklung der Schüler\_innen.

Tabelle 17  
*Leistungsentwicklung UE 2*

	Beta	SE
<b>Ebene 1</b>		
Mittelwert Prätest II (Ausgangsniveau)	43.50***	4.08
Lineare Wachstumskomponente	39.03***	2.44
Quadratische Wachstumskomponente	-15.23***	1.18
<b>Ebene 2</b>		
Linearer Trend Dosierung Medien UE2	-1.09***	0.33
Quadrierter Trend Dosierung Medien UE2	0.68***	0.16
<b>Varianzaufklärung</b>		
R-Quadrat (Ebene 1)	33.4 %	
R-Quadrat (Ebene 2)	35.9%	

*Anmerkung. Restricted Maximum Likelihood. \*\*\*  $p < .001$ , \*\*  $p < .01$ , \*  $p < .05$ , '  $p < .10$*

Bei der Parameterschätzung auf der Messwiederholungsebene erwiesen sich sowohl die *lineare Wachstumskomponente* als auch die *quadratische Wachstumskomponente* als statistisch bedeutsam für die Vorhersage der *Leistungsentwicklung* ( $\gamma_{10} = 39.03$ ;  $p < .001$ ;  $\gamma_{20} = -15.23$ ;  $p < .001$ ).

Der Wert  $\gamma_{00} = 43.50$  ( $p < .001$ ) ist der mittlere Wert für das Ausgangsleistungsniveau auf Individualebene, soweit alle Prädiktoren den Wert Null annehmen (Nullmodell). Die negative *quadratische Wachstumskomponente* auf der Ebene 1 ohne Hinzunahme des Prädiktors der Implementationsqualität verweist auf einen Leistungsabfall vom Posttest zum Follow-up.

Auf der zweiten Ebene zeigt das Regressionsgewicht von  $\gamma_{11} = -1.09$  ( $p < .001$ ) für die *Dosierung der Medien* auf dem linearen Trend, dass für Personen mit hoher Dosierung der Medien der lineare Anstieg in der *Leistungsentwicklung* um 1.09 Einheiten vom Prä- zum Posttest geringer ausfällt, wenn die *Dosierung der Medien* um eine Einheit ansteigt. Das Regressionsgewicht von  $\gamma_{21} = 0.68$  ( $p < .001$ ) für die Hinzunahme der *Dosierung der Medien*

zur *quadratischen Wachstumskomponente* zeigt allerdings auch, dass die *Leistungsentwicklung* vom Posttest zum Follow-up um 0.68 Einheiten höher ausfällt, wenn die *Dosierung der Medien* um eine Einheit ansteigt. Langfristig profitieren die Schüler\_innen also von einer hohen Dosierung der Medien.

Auf der Messwiederholungsebene werden 33.4 % der Varianz auf Personenebene durch Hinzunahme der *linearen und quadratischen Wachstumskomponente* aufgeklärt. Auf der zweiten Ebene sind es 35.9 % der Varianz in der personenspezifischen *Leistungsentwicklung*, die durch Unterschiede in der *Dosierung der Medien* aufgeklärt werden können.

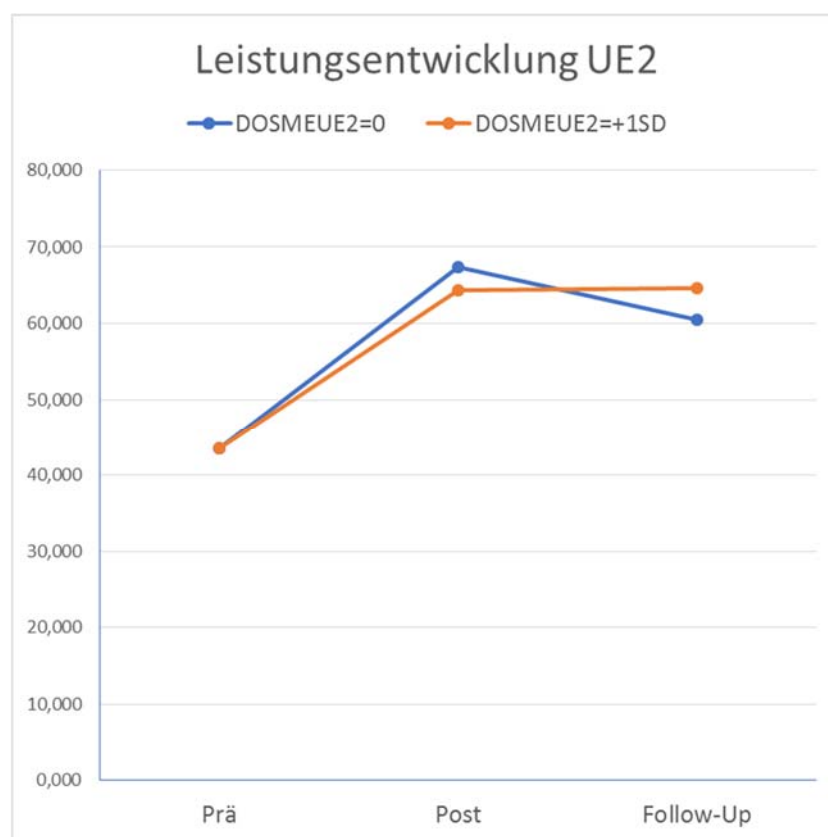


Abbildung 16. Leistungsentwicklung UE2 (DOSMEDU1= Dosierung der Medien Unterrichteinheit 2)

## 6 Zusammenfassende Diskussion

Die vorliegende Arbeit befasst sich damit, Angaben der Lehrkräfte zur Implementation der Interventionsmaßnahmen und ihrer Effekte auf die gemessenen Schülermaße einzuordnen. Dazu wurden die Angaben der Lehrkräfte aus den Feedback-Fragebögen zu Komponenten der Implementationsqualität klassifiziert und brauchbare quantitative Maße der Implementationsqualität entwickelt. In einem ersten Schritt wurden die Zusammenhänge zwischen Teilaspekten der Indikatoren und zu den anderen Indikatoren der Implementationsqualität untersucht. Im zweiten Schritt wurden die Zusammenhänge einzelner Indikatoren mit der Programmwirksamkeit in Bezug auf die Entwicklung des Autonomieerlebens, der erlebten Lehrerfürsorglichkeit, der Lernfreude und der selbstbestimmten Lernmotivation betrachtet. Darüber hinausgehend sollte die Forschungsfrage nach Zusammenhängen der Implementationsqualität mit der Leistungsentwicklung der Schüler\_innen beantwortet werden.

Ausgehend von der Forschungsfragestellung wurden vor dem in Kapitel 1 dargestellten theoretischen Hintergrund zentrale Forschungsbefunde zur Erfassung der Implementationsqualität dargestellt (Kapitel 2). Vor diesem wurde weiterhin in Kapitel 3 ein Klassifikationsschema zur Einordnung der lehrerbezogenen Angaben aus dem Feedback-Fragebogen entwickelt. Auf der Grundlage der so ermittelten Daten wurden für jede Unterrichtsstunde drei Indikatoren für die Adhärenz, drei für die Dosierung, zwei für die Qualität der Programmvermittlung und zwei für die Responsivität der am Programm teilnehmenden Schüler\_innen gebildet. Diese Indikatoren bildeten die Grundlage für die weiteren Analysen. Diese zeigten je nach Implementationsqualität und Indikator erhebliche Wirksamkeitsunterschiede zwischen den Treatmentklassen. So ergaben sich für Klassen mit einem höheren Einsatz der Interventionsmaßnahmen (Dosierung) günstigere Entwicklungen des Autonomieerlebens und der Leistung der Schüler\_innen.

Die dieser Arbeit zugrundeliegenden Daten wurden im Rahmen des FuN-Kollegs „Förderung selbstbestimmten Lernens im Unterricht“ erhoben. Zwei Unterrichtseinheiten im Fach Deutsch wurden in 9 Treatmentklassen (11 Kontrollklassen) der 7. und 8. Klassenstufe in der ersten Hälfte des Schuljahres 2005/06 eingesetzt. Zu beiden Unterrichtseinheiten wurden für 14 Unterrichtsstunden über Feedback-Fragebögen bei den Lehrkräften Daten zur Umsetzung erhoben. Die inhaltlichen Auswertungen stützen sich auf die Angaben von insgesamt 9 Lehrkräften zur Implementation der Interventionsmaßnahmen und den erhobenen Schülermaßen

zum Autonomieerleben, zur erlebten Lehrerfürsorglichkeit, der Lernfreude, der selbstbestimmten Lernmotivation und der Leistung.

### **Diskussion des Erhebungsinstruments**

Bei den Verfahren, mit denen die Indikatoren der Implementationsqualität erhoben werden, kann in der Forschungsliteratur zwischen Befragung und Beobachtung unterschieden werden. Bei der Datenerhebung der vorliegenden Arbeit, wurde eine schriftliche Befragung der Lehrkräfte mittels eines Feedback-Fragebogens nach jeder Unterrichtsstunde bei den Lehrkräften vorgenommen. In diesem war für jede der 14 Unterrichtsstunden anzugeben, welche Medien und Arbeitsformen zum Einsatz kamen, wie die Unterrichtsstunde von den Schüler\_innen angenommen wurde, wie sie selbst die Durchführung qualitativ einschätzten, welche Schwierigkeiten und Veränderungen sich ergaben und welche Konsequenzen dies für den Unterrichtsverlauf hatte. Über Fragebögen lässt sich die Implementationsaktivität der Lehrkräfte retrospektiv, kontinuierlich und ökonomisch über die gesamte Interventionsdauer erfassen. Für die Lehrkräfte ist es jedoch eine zusätzliche Belastung, den Verlauf der Implementation, die Einschätzung zu den Schüler\_innen sowie das Gelingen der Stunde zu dokumentieren. Somit ist eine entsprechende Disziplin bei der Dokumentation erforderlich. Betrachtet man den Rücklauf der Feedbackbögen (s. *Tabelle 4*), war diese mit einer Ausnahme in der Unterrichtseinheit 1 gegeben.

Dennoch muss kritisch betrachtet werden, wie valide und reliabel diese Form der Selbsteinschätzung sein kann: Einerseits kann die Dokumentation in der Praxis als belastend empfunden werden, andererseits liegen auch keine Erkenntnisse darüber vor, unter welchen Bedingungen die Dokumentationsbögen ausgefüllt wurden. So können beispielsweise Zeitmangel, Störungen oder die retrospektive Dokumentation zu Ungenauigkeiten bzw. Fehlangaben beitragen. Diese Urteile können somit durch eine Reihe verschiedener Faktoren beeinflusst und verzerrt sein. Zudem ist im Rahmen von nicht anonymisierten Selbstberichten auch mit einem erwünschten Antwortverhalten (*overreporting*) zu rechnen. So berichten unabhängige Beobachter weniger Implementationsaktivitäten als die Lehrpersonen (Mc Graw et al., 2000; (Goldberg) Lillehoj et al., 2004). Zusätzlich muss das subjektive Fremdurteil der Lehrkräfte zu den Schülervariablen kritisch betrachtet werden. Andere Formen der Datenerhebung wie z. B. flankierende Beobachtungen mit mehreren Beobachtern bzw. eine Videodokumentation der Unterrichtssituationen wären weniger ökonomisch aber nicht retrospektiv und würden so validere Daten bieten. Allerdings wären hierzu, um valide Implementationsdaten zu erhalten,



mehrere Unterrichtsbeobachtungen einzuplanen. Ergänzend könnten auch Interviewverfahren eingesetzt werden. Diese sind wegen ihres zeitlichen Aufwandes zwar weniger ökonomisch, ermöglichen aber ein flexibleres Vorgehen als der verwendete Feedback-Fragebogen. Die qualitative Auswertung könnte über die Fragen im Feedback-Fragebogen hinaus Einblick in Bezug auf Implementationshindernisse liefern und so evtl. auch schon im Rahmen einer Prozessevaluation eingesetzt werden, um von wissenschaftlicher Seite Unterstützung bei der Implementation bereitzustellen.

### **Validitätsdiskussion der Indikatoren**

Im Hinblick auf die Ergebnisse der Interkorrelationen der Indikatoren bzw. deren Teilaspekte ist die Frage zu stellen, ob sowohl die vorgenommene Konzeptualisierung als auch die vorgenommene Operationalisierung geeignet war, um Indikatoren der Implementationsqualität zu bilden und mit diesen eine valide Messung zu gewährleisten. Die Indikatoren Adhärenz und Dosierung weisen in ihren Teilaspekten erhebliche Zusammenhänge auf. Dies ist auf die ähnliche Herangehensweise bei deren Operationalisierung zurückzuführen und ergibt, wie in den bivariaten Zusammenhängen sichtbar wird, eine nicht hinreichende diskriminative Validität bei einer Messung der Konstrukte: Mit der Adhärenz sollte ein Ist-Soll Vergleich der Programmvorgaben ermöglicht werden. Mit der Dosierung wird eine Häufigkeit von Maßnahmen erfasst, die über das laut Manual gewünschte Ausmaß hinausgehen kann. Es stellt sich die Frage, ob diese beiden Konstrukte bei der Messung der Implementationsqualität getrennt erfasst werden sollten. Letztlich zeigen die korrelativen Zusammenhänge, dass eine Einhaltung der Programmvorgaben mit einer hohen Dosierung einhergeht. In Bezug auf die tatsächliche Umsetzung scheinen diese Befunde auch inhaltlich nachvollziehbar. Halten sich die Programmdurchführenden an die Programmvorgaben, dann setzen sie die Programmbestandteile auch in einer hohen Anzahl um. Da bei der Betrachtung der Zusammenhänge mit der Leistung auch eine negative Beziehung zur Dosierung des Schülerheftes gefunden wurden, also ein Übermaß an Maßnahmen negative Wirkung entfalten kann (tendenziell signifikant), scheint aus unserer Sicht demzufolge eine getrennte Erfassung von Adhärenz und Dosierung sinnvoll, um Erkenntnisse darüber zu gewinnen, ob eine strikte Einhaltung der Interventionsmaßnahmen notwendig ist, um unerwünschte Effekte einer Überdosierung zu vermeiden (s. *Fidelitäts-Anpassungs-Debatte*; Dane & Schneider, 1998, S. 25 f.; Durlak & DuPre, 2008, S. 341 f.; Berkel et al., 2011, S. 26 f.). Hier wäre es darüber hinaus sinnvoll, im Rahmen einer Untersuchung mit einem größeren Stichprobenumfang differenzierter zu betrachten, ab wann

bzw. unter welchen Bedingungen die Dosierung der Maßnahmen als ein Übermaß zu bewerten ist und welche Folgen damit einhergehen.

### **Diskussion der bivariaten Befunde**

Die gebildeten Indikatoren sollten geringe korrelative Zusammenhänge und somit eine hohe diskriminative Validität aufweisen. Es wurde erwartet, dass die Adhärenz und die Dosierung aufgrund der ähnlichen Operationalisierung eine Ausnahme bilden. Die Teilaspekte der Indikatoren sollten jeweils dasselbe Konstrukt messen, dementsprechend wurde angenommen, dass diese hohe Zusammenhänge aufweisen (konvergente Validität). Bei der Betrachtung der bivariaten Beziehungen der **Unterrichtseinheit 1** zeigt sich erwartungsgemäß, dass eine hohe Einhaltung der Manualvorgaben in Bezug auf den Medieneinsatz mit einer hohen Einhaltung in Bezug auf die Arbeitsformen einhergeht. Gleiches zeigt sich auch für den Vergleich zwischen der Dosierung der Medien mit der Dosierung der Arbeitsformen: Setzen die Lehrkräfte viele Medien im Unterricht ein, so setzen sie auch eine Vielzahl an Arbeitsformen um. Schätzen die Lehrkräfte die Mitarbeit der Schüler\_innen als hoch ein, geben sie weniger Unterrichtsstörungen an. Nicht erwartungsgemäß trägt die Einhaltung bei Medien und Arbeitsformen in der Unterrichtseinheit 1 nicht zu einer Einhaltung bzgl. der Vorgaben zur Verwendung des Schülerheftes bei. Ein häufiger Einsatz von Medien und Arbeitsformen begünstigt zudem hingegen der Erwartung nicht den häufigen Einsatz des Schülerheftes.

Wie erwartet zeigt sich die Unabhängigkeit der Indikatoren *Qualität der Programmvermittlung* und *Responsivität der Programmteilnehmer*. Diese Indikatoren weisen ebenso Unabhängigkeit von Adhärenz und Dosierung auf. Erwartungskonform zeigt sich keine vollständige Unabhängigkeit der Adhärenz von der Dosierung. Werden die Vorgaben im Manual in Bezug auf die Arbeitsformen in hohem Maße eingehalten, so setzen die Lehrkräfte häufiger Medien und Arbeitsformen ein. Keine Zusammenhänge zeigen sich hingegen mit dem Einsatz der speziellen Schülerhefte.

Für die **Unterrichtseinheit 2** zeigt sich erwartungskonform bezogen auf die Zusammenhänge innerhalb der Teilaspekte der Indikatoren, dass eine hohe Einhaltung der vorgegebenen Medien mit einer Einhaltung der Vorgaben zum speziellen Schülerheft einhergeht. Setzen die Lehrkräfte viele Medien im Unterricht ein, so entspricht dem eine Vielzahl an Arbeitsformen im Unterricht.

Entgegen der Erwartung gibt es keine signifikante Beziehung der anderen Teilaspekte der Adhärenz mit der Adhärenz der Arbeitsformen. Erwartungswidrig weisen die beiden Teilaspekte des Indikators *Responsivität der Programmteilnehmer* keinen Zusammenhang auf.

Im Hinblick auf die Unabhängigkeit der Indikatoren zeigt sich wie erwartet eine relative Unabhängigkeit der Indikatoren Adhärenz, der Dosierung, der Qualität der Programmvermittlung und der Responsivität der Programmteilnehmer. Nur einzelne Teilaspekte der Adhärenz und der Dosierung bilden eine Ausnahme.

Bei einer genaueren Einhaltung der Vorgaben bezüglich der Medien werden auch häufiger Medien, Arbeitsformen und das spezielle Schülerheft eingesetzt. Halten sich die Lehrkräfte bezüglich der Arbeitsformen in hohem Maße an die Vorgaben im Manual, setzten sie auch häufiger verschiedene Arbeitsformen ein. Gleiches gilt für die Einhaltung bezüglich des Schülerheftes. Auch hier geht eine Einhaltung der Vorgaben zur Verwendung des Schülerheftes mit einem häufigeren Einsatz des speziellen Schülerheftes einher. Nicht erwartungskonform zeigt sich, dass die positive Einschätzung der Unterrichtswahrnehmung der Schüler durch die Lehrkräfte in Zusammenhang steht mit einer besseren Bewertung der eigenen Unterrichtsdurchführung.

Beim **Vergleich beider Unterrichtseinheiten** zeigen sich je nach Unterrichtseinheit unterschiedliche Zusammenhänge. Übereinstimmend zeigen allerdings die Analysen, dass die Adhärenz der angedachten Interventionsvorgaben in der Regel auch mit einer umfänglicheren Dosierung der Implementationsbestandteile einhergeht. Zwischen dem Indikator Qualität der Programmvermittlung und dem Indikator Responsivität der Teilnehmer besteht entgegen der Erwartung in Unterrichtseinheit 1 ein signifikanter Zusammenhang. Dieser konnte aber in der Unterrichtseinheit 2 nicht gefunden werden. Ebenso zeigen die Analysen sowohl bedeutsame Zusammenhänge innerhalb der Teilaspekte der Adhärenz als auch innerhalb der Teilaspekte der Dosierung. Es finden sich aber insgesamt häufiger signifikante Zusammenhänge bei Teilaspekten der Adhärenz mit den Teilaspekten der Dosierung, so dass die erwarteten hohen Korrelationen der Teilaspekte innerhalb eines Indikators einerseits und die Unabhängigkeit zwischen den Implementationsindikatoren andererseits in den Unterrichtseinheiten nicht für alle Teilaspekte gezeigt werden konnten.

Im Hinblick auf die **bivariaten Beziehungen zu den Schülermaßen in der Unterrichtseinheit 1** zeigt sich erwartungskonform, dass bis zum Ende der ersten Unterrichtseinheit ein häufiger Einsatz des Schülerheftes mit einer positiven Veränderung im Autonomieerleben der Schüler\_innen einhergeht. Entgegen der Erwartung zeigt sich, dass die Adhärenz der Arbeitsformen einen negativen Einfluss auf die Entwicklung in der Lernfreude aufweist. Ebenso scheint ein sehr häufiger Einsatz von Medien und Arbeitsformen einen tendenziell ungünstigen Einfluss auf die Entwicklung der Lernfreude zu nehmen. Für die Qualität der Programmvermittlung zeigt sich, dass je positiver die Lehrkräfte die Unterrichtsstunde insgesamt einschätzten, desto weniger Entwicklung zeigte sich im Autonomieerleben der Schüler. Die Einhaltung der Manualvorgaben bzgl. der Arbeitsformen entfaltet nach dem Ende der Unterrichtseinheit ihre nachhaltige positive Wirkung auf die Leistungsentwicklung. Werden die verschiedenen Arbeitsformen häufiger eingesetzt (Dosierung), zeigt sich eine langfristige und nachhaltige positive Leistungsentwicklung der Schüler\_innen von Beginn der Intervention bis zur Folgerhebung (Posttest bis Follow-up; Prätest bis Follow-up) Entgegen der Erwartung besteht zu dieser bei den anderen Implementationsindikatoren kein signifikanter Zusammenhang.

Bei der **zweiten Unterrichtseinheit** zeigt sich eine umso größere positive Entwicklung des Autonomieerlebens, je höher die Adhärenz in Bezug auf das spezielle Schülerheft ist. Auch dessen hohe Dosierung trägt zu einer positiven Entwicklung im Autonomieerleben bei. Je höher die Lehrkräfte allerdings die Mitarbeit der Schüler\_innen einschätzen, desto weniger Entwicklung im Autonomieerleben wird sichtbar.

Wie erwartet geht eine hohe Dosierung der Medien mit der positiven Leistungsentwicklung nach dem Abschluss der Unterrichtseinheit bis zur Folgerhebung einher. Erklärungsbedürftig ist, dass tendenziell ein übermäßiger Einsatz der Schülerhefte vom Beginn der Unterrichtseinheit bis zur Folgerhebung mit geringeren Lernzuwächsen einhergeht. Nicht erwartungskonform zeigt sich zudem für eine starre Einhaltung der Medien ein negativer Zusammenhang zur Leistungsentwicklung (Posttest bis Follow-up). So scheint sich eine Abweichung der Manualvorgaben in Bezug den Medieneinsatz günstig auszuwirken. Arbeiten die Schüler\_innen aber (zu) häufig eigenständig mit dem speziellen Schülerheft zeigen sich geringere Leistungszuwächse. Hier könnte allerdings eine mangelhafte Betreuung durch die Lehrkräfte ursächlich sein. Wird die Arbeit mit diesem Medium nicht wie beabsichtigt von der Lehrkraft begleitet

und erhält der Schüler nicht die entsprechenden Rückmeldungen vonseiten der Lehrkraft, kann das spezielle Schülerheft nicht die beabsichtigte Wirkung entfalten.

Beim **Vergleich der Unterrichtseinheiten** zeigt sich, dass jeweils andere Aspekte der Implementationsqualität von Bedeutung sind oder aber die Beziehungen wider Erwarten nicht signifikant werden. Es wurde allerdings angenommen, dass spezielle Indikatoren unabhängig von der Unterrichtseinheit bzw. deren Inhalten sind. So steht die Adhärenz des speziellen Schülerheftes in Unterrichtseinheit 2 auch nur tendenziell in einem positiven Zusammenhang mit der Entwicklung der erlebten Autonomie. Seine Dosierung weist hingegen für beide Unterrichtseinheiten eine signifikante Beziehung zur Entwicklung des Autonomieerlebens auf. Andererseits findet sich ausschließlich in der Unterrichtseinheit 2 ein positiver Zusammenhang zwischen dessen Dosierung mit der Entwicklung der erlebten Lehrerfürsorglichkeit.

Nicht erwartungskonform finden sich für einzelne Indikatoren der Unterrichtseinheit 1 negative Zusammenhänge mit der Lernfreude was bei der zweiten Unterrichtseinheit nicht der Fall ist. Auch für den Indikator Qualität der Programmvermittlung 2 zeigt sich nur in der ersten Unterrichtseinheit eine tendenziell negative Beziehung zum Autonomieerleben. Auch stehen in den Unterrichtseinheiten jeweils andere Indikatoren mit der Leistungsentwicklung der Schüler\_innen in Beziehung. In der zweiten Unterrichtseinheit findet sich anders als in der Unterrichtseinheit 1 für die Adhärenz der Medien ein negativer Zusammenhang mit der Leistungsentwicklung.

Für die Qualität der Programmvermittlung und die Responsivität der Programmteilnehmer zeigt sich insgesamt nicht die erwartete Häufigkeit der Zusammenhänge zu den Schülermaßen. Dies könnte auf eine ungeeignete Operationalisierung der Indikatoren hindeuten. Für beide Unterrichtseinheiten zeigen sich keine signifikanten Zusammenhänge der Indikatoren mit der erlebten Lehrerfürsorglichkeit und der Lernmotivation. Insgesamt weisen die Ergebnisse über die genannten signifikanten Zusammenhänge hinaus teilweise starke Korrelationen auf, für diese zeigt sich aber aufgrund der geringen Stichprobengröße keine Signifikanz.

Zudem wäre es im Rahmen einer größeren Stichprobe von Interesse zu prüfen, ob die Indikatoren *Qualität der Programmvermittlung* und *Responsivität der Programmteilnehmer* moderierenden bzw. mediierenden Einfluss auf die anderen Indikatoren und ihre Beziehung zur Wirksamkeit nehmen. Hierzu finden sich Hinweise in der Forschungsliteratur (Berkel, 2011). Zudem könnten weitere Moderatoreffekte vorliegen, die die gefundenen Zusammenhänge

beeinflussen (z. B. Merkmale der Programmvermittler; Abschnitt 1.4.2). Dies könnte im Rahmen einer größeren Stichprobe geprüft werden.

### **Diskussion der Ergebnisse der Mehrebenenanalyse**

In einem zweiten Schritt wurden die gefundenen Beziehungen zusätzlich mehrebenenanalytisch überprüft. Zur Durchführung der Mehrebenenanalysen wurden Daten der Schüler\_innen auf der ersten Ebene und die Daten der Lehrkräfte auf der zweiten Ebene über drei Messzeitpunkte in die Analysen einbezogen. Das Vorgehen für die Mehrebenenanalyse orientiert sich an den Empfehlungen von Hox (2010). Die Schätzung erfolgte mittels Restricted Maximum Likelihood (REML). Da nur Daten von 9 Lehrkräften auf der zweiten Ebene vorliegen, müssen die Folgen dieses geringen Stichprobenumfangs für die Validität der Parameterschätzungen angesprochen werden. Bei schlanken Modellen mit wenigen Variablen, ist generell eine geringe Fallzahl auf der zweiten Ebene als weniger problematisch zu bewerten. Wenig beeinflusst von einer geringen Fallzahl auf Ebene 2 sind die Koeffizienten der fixen Effekte und deren Standardfehler (Hox, 2010, S. 233 f.). Analysen mit weniger als zehn Einheiten auf der zweiten Ebene können dennoch nur erste Hinweise auf mögliche Zusammenhänge liefern (Stegmueller, 2013, S. 753).

Entsprechende Hinweise zu Zusammenhängen der Indikatoren mit den Veränderungen auf Schülerseite zeigen sich für die Leistungsentwicklung in der Unterrichtseinheit 1. Hier konnte neben den signifikanten Prädiktoren *lineare* und *quadratische Wachstumskomponente* durch die Hinzunahme des Prädiktors *Dosierung der Arbeitsformen* zusätzliche Varianz in der Leistungsentwicklung aufgeklärt werden. So konnten die Schüler\_innen neben der positiven Veränderung in der Leistung im Interventionszeitraum erwartungsgemäß von einer höheren Dosierung der Arbeitsformen profitieren.

In der Unterrichtseinheit 1 zeigt sich die *lineare Wachstumskomponente* als bedeutsam zur Vorhersage der *Autonomieentwicklung*. Im Weiteren zeigt sich erwartungskonform der bedeutsame Einfluss der *Dosierung des Schülerheftes* auf die Entwicklung des *Autonomieerlebens* der Schüler\_innen. So fällt der lineare Anstieg des Autonomieerlebens im Interventionszeitraum bei einer hohen Dosierung des Schülerheftes günstiger aus. Es zeigt sich insgesamt, dass für Schüler\_innen aus Klassen, in denen das Schülerheft häufiger eingesetzt wurde, eine positivere Entwicklung des Autonomieerlebens berichtet werden kann.

Für die Entwicklung der *Lernfreude* in der Unterrichtseinheit 1 erweist sich die *lineare Wachstumskomponente* als statistisch bedeutsamer Prädiktor. So zeigt sich eine ungünstige Entwicklung der Lernfreude zwischen dem ersten und zweiten Messzeitpunkt. Hier wurde ein möglicher Zusammenhang der *Adhärenz der Arbeitsformen* mit der Entwicklung der Lernfreude betrachtet, dieser verfehlt aber trotz Signifikanz in der Betrachtung der korrelativen Zusammenhänge die statistische Vertrauensgrenze in der Mehrebenenanalyse.

Auch bei der Unterrichtseinheit 2 konnten die Schüler\_innen im Interventionszeitraum erwartungsgemäß von einer höheren *Dosierung der Medien* profitieren. So konnten neben den signifikanten Prädiktoren *lineare Wachstumskomponente* und *quadratische Wachstumskomponente* durch die Hinzunahme des Prädiktors *Dosierung der Medien* zusätzliche Varianz in der Leistungsentwicklung aufgeklärt werden. Die signifikante negative quadratische Wachstumskomponente auf der Ebene 1 ohne Hinzunahme des Prädiktors der Implementationsqualität verweist zwar auf einen Leistungsabfall nach der Unterrichtseinheit bis zur Folgerhebung (Posttest bis zum Follow-up), dieser kann aber durch eine hohe Dosierung der Medien abgefedert werden. Langfristig profitieren die Schüler\_innen somit von einer hohen Dosierung der Medien.

Die bisher dargestellten Ergebnisse unterstützen die Annahmen, wonach ein häufiger Einsatz der Interventionsmaßnahmen in einer positiven Beziehung zur Entwicklung Autonomieerleben (*Dosierung des Schülerheftes*) und zur Leistungsentwicklung (*Dosierung der Arbeitsformen*) der Schüler\_innen. Eine zu starre Einhaltung scheint sich ungünstig auf die Lernfreude auszuwirken. Diese Beziehung deutet auf die Notwendigkeit der Flexibilisierung der Interventionsmaßnahmen hin. Allerdings wurde der Befund aufgrund der geringen Stichprobengröße in der Mehrebenenanalyse nicht signifikant und sollte insbesondere wegen der praktischen Bedeutsamkeit erneut geprüft werden.

## **Fazit und Ausblick**

Das zentrale Anliegen der vorliegenden Arbeit bestand darin, die Implementationsqualität zu erfassen, geeignete Indikatoren zu deren Messung zu bilden und die Zusammenhänge zwischen der Implementationsqualität und der Wirksamkeit der Interventionsmaßnahmen zu bestimmen. Die Analysen zur Unterrichtseinheit 1 zeigen einerseits erwartungskonform, dass das Ausmaß an Dosierung der Interventionsmaßnahmen zur Erklärung von Leistungsunterschieden herangezogen werden kann. Übereinstimmend mit der Selbstbestimmungstheorie zeigt sich zudem, dass der Einsatz unterrichtsbegleitender Schülerhefte mit erhöhtem Autonomieerleben einhergeht. Andererseits zeigt sich für die Unterrichtseinheit 2 tendenziell aber auch, dass ein übermäßiger Einsatz der Schülerhefte mit geringeren Lernzuwächsen einhergehen kann. Erwartungswidrig und erklärungsbedürftig sind die negativen Beziehungen zwischen einigen Adhärenz- und Dosierungsmaßen und den Veränderungen in der Lernfreude. Hier deutet sich an, dass stärkere Abweichungen von den Programmvorgaben mit positiveren Effektgrößen in der Lernfreude in Zusammenhang stehen. Dies könnte auf positiv zu bewertende Programmanpassungen hinweisen.

Eine Schwäche der Studie zeigt sich in der Erhebungsmethode. Zur Erfassung von Indikatoren der Implementationsqualität scheint die schriftliche Befragung der Lehrkräfte nur bedingt geeignet. Interviews und/oder Unterrichtsbeobachtungen führen vermutlich zu valideren Schätzungen der Implementationsqualität. Hier sollten Beobachtungsdaten mehrerer Beobachter erhoben und so eine Erfassung der Übereinstimmung der Angaben ermöglicht werden. Im Hinblick auf die Messinstrumente sollten in Folgeerhebungen hochinferente Schätz- und Beurteilungsskalen vermieden werden.

In Bezug auf die gebildeten Indikatoren sind weitere Untersuchungen erforderlich. Nicht für alle Indikatoren konnten die Zusammenhänge der Teilindikatoren gezeigt werden. Die Adhärenz und die Dosierung sind ähnliche Maße, die einen hohen Zusammenhang aufweisen und so nicht eindeutig voneinander abgrenzbar sind.

Ein weiterer Kritikpunkt ergibt sich daraus, dass die Angaben der Kontrollgruppe vernachlässigt wurden. Zur Überprüfung der Kausalbeziehungen sind prospektive Längsschnittstudien mit mehreren Erhebungswellen mit einem Treatment-Kontrollgruppen-Design anzustreben. Auf diese Weise könnten Störeinflüsse kontrolliert, Veränderungen in den zentralen Kriteriumsvariablen und deren Kausalbeziehungen aufgezeigt werden.



Die mehrebenenanalytischen Auswertungen konnten erste Hinweise zu den erwarteten Zusammenhängen zwischen den Indikatoren der Implementationsqualität und insbesondere der Leistungsentwicklung und der Entwicklung des Autonomieerlebens aufzeigen. Einschränkend lässt sich anmerken, dass ein zu kleiner Stichprobenumfang vorliegt, der die Aussagekraft der Ergebnisse einschränkt. Da es sich um eine geringe Anzahl an Analyseeinheiten auf Ebene 2 handelt, können die Ergebnisse nur erste Hinweise für weitere Forschungsbemühungen in diesem Feld liefern. Hier wäre in Folgeuntersuchungen eine höhere Anzahl an teilnehmenden Klassen bzw. Lehrkräften wünschenswert. In der vorliegenden Arbeit blieb zudem der Aspekt der Programmspezifität unberücksichtigt. Es wurde zwar ein Item formuliert, mit dem eine Abgrenzung gegenüber anderen Programmen bestimmt werden sollte. Dies wurde allerdings wegen der geringen Aussagekraft nicht zur Bestimmung der Programmspezifität eingesetzt. Dieser Teilaspekt der Implementationsqualität blieb auch in der bisherigen Implementationsforschung weitgehend unberücksichtigt. Dies liegt insbesondere daran, dass dieser Aspekt bei der Erhebung der Daten besonders berücksichtigt werden muss und hierzu ein gesondertes Erfassungsinstrument eingesetzt werden sollte, das die Erfassung anderer Programmmaßnahmen ermöglicht. Abschließend muss angemerkt werden, dass innerhalb der Forschungsgemeinschaft ein stärkerer Konsens bzgl. der Indikatoren der Implementationsqualität angestrebt und eine konsequentere Erfassung der Implementationsqualität erfolgen sollte. Gelingt dies, könnten die Vergleichbarkeit der Indikatoren und die Vergleichbarkeit der Ergebnisse zur Implementationsqualität ebenso wie die Vergleichbarkeit in Bezug auf die Wirksamkeit von Interventionsmaßnahmen erhöht werden.

## Literatur

- Aber, J. L., Jones, S. M., Brown, J. L., Chaudry, N. & Samples, F. (1998). Resolving conflict creatively: evaluating the developmental effects of a school-based violence prevention program in neighborhood and classroom context. *Development and Psychopathology*, 10(2), 187–213.
- Altman, D. G. (1991). *Practical statistics for medical research*. London: Chapman and Hall.
- Amabile, T. M., DeJong, W. & Lepper, M. R. (1976). Effects of externally-imposed deadlines on subsequent intrinsic motivation. *Journal of Personality and Social Psychology*, 34, 92–98.
- Anson, A. R., Cook, T. D., Habib, F., Grady, M. K., Haynes, N. & Comer, J. P. (1991). The Comer school development program: A theoretical analysis. *Urban Education*, 26, 56–82.
- August, G. J., Bloomquist, M. L., Lee, S. S., Realmuto, G. M. & Hektner, J. M. (2006). Can evidence-based prevention programs be sustained in community practice settings? The Early Risers' Advanced-Stage Effectiveness Trial. *Prevention Science*, 7(2), 151–65.
- Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Englewood, NJ: Prentice Hall.
- Basch, C. E., Sliepcevich, E. M., Gold, R. S., Duncan, D. F. & Kolbe, L. J. (1985). Avoiding type III errors in health education program evaluations: a case study. *Health Education Quarterly*, 12(4), 315–331.
- Basen-Enquist, K., O'Hara-Tompkins, N., Lovato, C. Y., Lewis, M. J., Parcel, G. S. & Gingiss, P. (1994). The effect of two types of teacher training on implementation of Smart Choices: a tobacco prevention curriculum. *Journal of School Health*, 64, 334–339.
- Battistich, V., Schaps, E., Watson, M., Solomon, D. & Lewis, C. (2000). Effects of the Child Development Project on students' drug use and other problem behaviors. *The Journal of Primary Prevention*, 21(1), 75–99. <http://doi.org/10.1023/a:1007057414994>
- Battistich, V., Schaps, E. & Wilson, N. (2004). Effects of an elementary school intervention on students' "connectedness" to school and social adjustment during middle school. *Journal of Primary Prevention*, 24(3), 243–262. <http://doi.org/10.1023/B:JOPP.0000018048.38517.cd>
- Bauman, L. J., Stein, R. E. & Ireys, H. T. (1991). Reinventing fidelity: The transfer of social technology among settings. *American Journal of Community Psychology*, 19(4), 619–639.
- Baumeister, R. F. & Leary, M. R. (1995). The need to belong: Desire for interpersonal attachments as a fundamental human motivation. *Psychological Bulletin*, 117(3), 497–529.
- Benner, J. G., Nelson, R. J., Stage, S. A. & Ralston, N. C. (2011). The influence of fidelity of implementation on the reading outcomes of middle school students experiencing reading difficulties. *Remedial and Special Education: RASE*, 32(1), 79–88.

- Berkel, C., Mauricio, A., Schoenfelder, E. & Sandler, I. N. (2011). Putting the pieces together: an integrated model of program implementation. *Prevention Science, 12*(1), 23–33.
- Berman, P. (1981). Educational change: An implementation paradigm. In R. Lehming & M. Kane (Eds.), *Improving schools: Using what we know* (pp. 253–286). Beverly Hills: Sage.
- Berman, P. & McLaughlin, M. W. (1976). Implementation of educational innovation. *The Educational Forum, 40*(3), 345–370.
- Bieg, S. & Mittag, W. (2005). *Fragebogen zur Motivationalen Selbstregulation bei Jugendlichen (MoS-J)*. Univeröffentlichtes Manuskript. Pädagogische Hochschule Ludwigsburg.
- Bieg, S. & Mittag, W. (2010). Selbstbestimmte Lernmotivation. In T. Hascher & B. Schmitz (Hrsg.), *Pädagogische Interventionsforschung. Theoretische Grundlagen und empirisches Handlungswissen* (S. 188–198). Weinheim: Juventa.
- Bieg, S. & Mittag, W. (2011). Leistungsverbesserung durch Förderung der selbstbestimmten Lernmotivation. In M. Dresel & L. Lämmle (Hrsg.), *Motivation, Selbstregulation und Leistungsexzellenz* (S. 219–236).
- Blakely, C. H., Mayer, J. P., Gottschalk, R. G., Schmitt, N., Davidson, W. S., Roitman, D. B. & Emshoff, J. G. (1987). The Fidelity-Adaptation Debate: Implications for the implementation of public sector social programs. *American Journal of Community Psychology, 15*(3), 253–268.
- Boggiano, A. K., Flink, C., Shields, A., Seelbach, A. & Barrett, M. (1993). Use of techniques promoting students' self-determination: Effects of students' analytic problem-solving skills. *Motivation and Emotion, 17*, 319–336.
- Bortz, J., & Döring, N. (2006). *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler*. Berlin, Heidelberg: Springer.
- Botvin, G. J., Baker, E., Dusenbury, L., Tortu, S. & Botvin, E. M. (1990). Preventing adolescent drug abuse through a multimodal cognitive-behavioral approach: Results of a 3-year study. *Journal of Consulting and Clinical Psychology, 58*, 437–446.
- Botvin, G. J., Dusenbury, L., Baker, E., James-Ortiz, S. & Kerner, J. (1989). A skills training approach to smoking prevention among hispanic youth. *Journal of Behavioral Medicine, 12*(3), 279–296.
- Byrnes, H. F., Miller, B. A., Aalborg, A. E., Plasencia, A. V., & Keagy, C. D. (2010). Implementation fidelity in adolescent family-based prevention programs: relationship to family engagement. *Health Education Research, 25*(4), 531–541.
- Carroll, C., Patterson, M., Wood, S., Booth, A., Rick, J. & Balain, S. (2007). A conceptual framework for implementation fidelity. *Implementation Science IS, 40*(2).
- Cho, H., Hallfors, D. D. & Sánchez, V. (2005). Evaluation of a high school peer group intervention for at-risk youth. *Journal of Abnormal Child Psychology, 33*(3), 363–74.
- Coburn, C. E. (2004). Beyond decoupling: Rethinking the relationship between the institutional environment and the classroom. *Sociology of Education, 77*(3), 211–244.

- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20, 37–46.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Connell, D. B., Turner, R. R. & Mason, E. F. (1985). Summary of findings of the school health education evaluation: Health promotion effectiveness, implementation, and costs. *Journal of School Health*, 55(8), 316–321.
- Dane, A. V. & Schneider, B. (1998). Program integrity in primary and early secondary prevention: are implementation effects out of control? *Clinical Psychology Review*, 18(1), S. 23–45.
- DeCharms, R. (1968). *Personal causation*. New York: Academic Press.
- Deci, E. L. (1975). *Intrinsic motivation*. New York: Plenum Press.
- Deci, E. L. Betley, G., Kahle, J., Abrams, L., & Porac, J. (1981). When trying to win: Competition and intrinsic motivation. *Personality and Social Psychology Bulletin*, 7, 79–83.
- Deci, E. L. & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior*. New York: Plenum Press.
- Deci, E. L. & Ryan, R. M. (1991). A motivational approach to self: Integration in personality. In R. Dienstbier (Ed.), *Nebraska symposium on motivation: Vol. 38, Perspectives on motivation* (pp. 237–288). Lincoln, NE: Nebraska University Press.
- Deci, E. L. & Ryan, R. M. (1993). Die Selbstbestimmungstheorie der Motivation und ihre Bedeutung für die Pädagogik. *Zeitschrift für Pädagogik*, 39, 223 – 237.
- Deci, E. L. & Ryan, R. M. (2000). The “what” and “why” of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry*, 11, 227–268.
- Deci, E. L. & Ryan, R. M. (2014). Autonomy and need satisfaction in close relationships: Relationships motivation theory. In N. Weinstein (Ed.), *Human motivation and interpersonal relationships: Theory, research, and applications* (pp. 53–73). Dordrecht: Springer.
- Domitrovich, C. E. Gest, S. D., Jones, D., Gill, S., & Sanford DeRousie, R. M. (2010). Implementation quality: Lessons learned in the context of Head Start REDI trial. *Early Childhood Research Quarterly*, 25, 284–298.
- Domitrovich, C. E. & Greenberg, M. T. (2000). The study of implementation: Current findings from effective programs that prevent mental disorders in school-aged children. *Journal of Educational & Psychological Consultation*, 11(2), 193–221.
- Dresel, M. & Lämmle, L. (2017). Emotion, Motivation und selbstreguliertes Lernen. In T. Götz, A. C. Frenzel, M. Dresel, & R. Pekrun (Hrsg.), *Emotion, Motivation und selbstreguliertes Lernen* (S. 80–142).
- Drössler, S., Jerusalem, M. & Mittag, W. (2007). Förderung sozialer Kompetenzen im Unterricht. Implementation eines Lehrerfortbildungsprojektes. *Zeitschrift Für Pädagogische Psychologie*, 21(2), 157–168.

- Durlak, J. A. & DuPre, E. P. (2008). Implementation matters: a review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology*, 41(3-4), 327–50. <http://doi.org/10.1007/s10464-008-9165-0>
- Dusenbury, L., Brannigan, R., Falco, M. & Hansen, W. B. (2003). A review of research on fidelity of implementation: implications for drug abuse prevention in school settings. *Health Education Research*, 18(2), 237–56.
- Eccles, J. S. & Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annual Reviews Psychology*, 53, 109–132.
- Eid, M., Gollwitzer, M., & Schmitt, M. (2015). *Statistik und Forschungsmethoden* (4. Aufl.). Weinheim: Beltz.
- Eid, M., & Schmidt, K. (2014). *Testtheorie und Testkonstruktion*. Göttingen, Bern, Wien: Hogrefe.
- Eisend, M. (2014). *Metaanalyse* (1. Aufl.). Mering: Hampp.
- Elias, M. J., Gara, M., Ubriaco, M., Rothbaum, P. A., Clabby, J. F. & Schuyler, T. (1986). Impact of a preventive social problem solving intervention on children's coping with middle-school stressors. *American Journal of Community Psychology*, 14, 259–275.
- Elliot, D., & Mihalic, S. (2004). Issues in disseminating and replicating effective prevention programs. *Prevention Science*, 5, 47–53.
- Ennett, S. T., Haws, S., Ringwalt, C. L., Vincus, A. A., Hanley, S., Bowling, J. M. & Rohrbach, L. A. (2011). Evidence-based practice in school substance use prevention: fidelity of implementation under real-world conditions. *Health Education Research*, 26(2), 361–371.
- Euler, D. & Sloane, P. F. E. (1998). Implementation als Problem der Modellversuchsforschung. *Unterrichtswissenschaft*, 26(4), S. 312–326.
- Fries, S. (2010). Motivation. In T. Hascher & B. Schmitz (Hrsg.), *Handbuch Pädagogische Interventionsforschung* (S. 149–161).
- Fullan, M. & Pomfret, A. (1977). Research on curriculum and instruction implementation. *Review of Educational Research*, 47, 335–397.
- Gager, P. J. & Elias, M. J. (1997). Implementing prevention programs in high-risk environments: Application of the resiliency paradigm. *American Journal of Orthopsychiatry*, 67(3), 363–373.
- Gollwitzer, M., & Jäger, R. S. (2007). *Evaluation-Workbook*. Weinheim: Beltz PVU.
- Gottfredson, D. C., Gottfredson, G. D. & Hybl, L. G. (1993). Managing adolescent behavior: A multiyear, multischool study. *American Educational Research Journal*, 30(1), 179–215.
- Gottfredson, G. (1984). A theory-ridden approach to program evaluation: A method for stimulating researcher-implementer collaboration. *American Psychologist*, 39, 1101–1112.

- Gräsel, C. & Parchmann, I. (2004). Implementationsforschung-oder: der steinige Weg, Unterricht zu verändern. *Unterrichtswissenschaft*, 32(3), 196–214.
- Greenberg, M. T., Domitrovich, C. E. & Bumbarger, B. (2001). The prevention of mental disorders in school-aged children: Current state of the field. *Prevention and Treatment*, 4(1), 1–59.
- Greenhalgh, T., Robert, G., Macfarlane, F., Bate, P. & Kyriakidou, O. (2004). Diffusion of innovations in service organizations: systematic review and recommendations. *The Milbank Quarterly*, 82, 581–629.
- Gresham, F. M., Gansle, K. A. & Noell, G. H. (1993). Treatment integrity in applied behavior analysis with children. *Journal of Applied Behavior Analysis*, 26(2), 257–263.
- Griffin, K. W., Mahadeo, M., Weinstein, J. & Botvin, G. J. (2006). Program implementation fidelity and substance use outcomes among middle school students in a drug abuse prevention program. *Health and Addictions*, 6(1), 9–28.
- (Goldberg) Lillehoj, C. J., Griffin, K. W. & Spoth, R. (2004). Program provider and observer ratings of school-based preventive intervention implementation: agreement and relation to youth outcomes. *Health Education & Behavior: The Official Publication of the Society for Public Health Education*, 31(2), 242–57. <http://doi.org/10.1177/1090198103260514>
- Guay, F., Boggiano, A. K. & Vallerand, R. J. (2001). Autonomy support, intrinsic motivation, and perceived competence: Conceptual and empirical linkages. *Personality and Social Psychology Bulletin*, 27, 643 – 650.
- Guthrie, J. T., McRae, A. & Klauda, S. L. (2007). Contributions of concept-orientated reading instruction to knowledge about interventions for motivations in reading. *Educational Psychologist*, 42, 237–250.
- Hansen, W. B. (1996). Pilot test results comparing the All Stars program with seventh grade D.A.R.E.: program integrity and mediating variable analysis. *Substance Use and Misuse*, 31, 1359–1377.
- Harackiewicz, J. M., Barron, K. E., Pintrich, P. R., Elliot, A. J., & Trash, T. (2002). Revision of Achievement Goal Theory: Necessary and illuminating. *Journal of Educational Psychology* 2002, 94(3), 638–645.
- Hasson, H. (2010). Systematic evaluation of implementation fidelity of complex interventions in health and social care. *Implementation Science IS*, 5(1), 67.
- Heckhausen, H. (1975). Fear of failure as a self-reinforcing motive system. In I. G. Sarason & C. Spielberger (Eds.), *Stress and anxiety* (pp. 117-128). Washington, D. C.: Hemisphere.
- Heckhausen, H., Schmalt, H.-D. & Schneider, K. (1985). *Achievement motivation in perspective*. Orlando, FL: Academic Press. Orlando, FL.
- Helmke, A. (1993). Die Entwicklung der Lernfreude vom Kindergarten bis zur 5. Klassenstufe. *Zeitschrift für Pädagogische Psychologie*, 7, 77–86.
- Helmke, A. & Schrader, F.-W. (2006). Determinanten der Schulleistung. In D. H. Rost (Hrsg.), *Handwörterbuch Pädagogische Psychologie* (3 Aufl., S. 83–94). Weinheim: Beltz Psychologie Verlags Union.

- Hitt, J., Robbins, A., Galbraith, J., Todd, J., Patel-Larson, A., McFarlane, J., Spikes, P. et al. (2006). Adaptation and implementation of a evidence-based prevention counselling intervention in Texas. *AIDS Education and Prevention*, 18 (SA): 108-118.
- Holodyski, M. & Oerter, R. (2008). Tätigkeitsregulation und die Entwicklung von Motivation, Emotion, Volition. In R. Oerter & L. Montada (Hrsg.), *Entwicklungspsychologie* (6 Aufl., S. 535–571). Weinheim: Beltz PVU.
- Hox, J. J. (2010). *Multilevel analysis: techniques and applications*. New York: Routledge.
- Huberman, M. (1995). Professional careers and professional development: Some intersections. In T. R. Guskey & M. Huberman (Eds.), *Professional development in education: New paradigms and practices* (pp. 193–224). New York: Teachers College Press, Columbia University.
- Ishler, A. L., Johnson, R. T., & Johnson, D. W. (1998). Long-term effectiveness of a statewide staff development program on cooperative learning. *Teaching and Teacher Education*, 14, 273–281.
- Johnson, D. W. & Johnson, R. T. (1983). Social interdependence and perceived academic and personal support in the classroom. *The Journal of Social Psychology*, 120, 77–82.
- Johnson, D. W., Johnson, R. T., Buckman, L. A. & Richards, S. P. (1985). The effect of prologed implementation of cooperative learning on social support within the classroom. *The Journal of Psychology*, 119(5), 405–411.
- Katz, I. & Assor, A. (2007). When choice motivates an when it does not. *Educational Psychology Review*, 19, 429–442.
- Kerr, D. M., Kent, L. & Lam, T. C. M. (1985). Measuring Program Implementation with a Classroom Observation Instrument. The Interactive Teaching Map. *Evaluation Review*, 9(4), 461–482.
- Knoche, L. L., Sheridan, S. M., Edwards, C. P. & Osborn, A. Q. (2010). Implementation of a relationship-based school readiness intervention: A multidimensional approach to fidelity measurement for early childhood. *Early Child Research Quaterly*, 25(3), 299–313. DOI:10.1016/j.ecresq.2009.05.003
- Koestner, R., Ryan, R. M., Bernier, F. & Holt, K. (1984). Setting limits on children`s behavior: The differential effects of controlling versus informational styles on intrinsic motivation and creativity. *Journal of Personality*, 52, 233–248.
- Köller, O., & Schiefele, U. (2006). Zielorientierung. In D. H. Rost (Hrsg.), *Handwörterbuch Pädagogische Psychologie* (S. 880–886). Weinheim: Beltz PVU.
- Krapp, A. (1996). Die Bedeutung von Interesse und intrinsischer Motivation für den Erfolg und die Steuerung schulischen Lernens. In G. W. Schnaitmann (Hrsg.), *Theorie und Praxis der Unterrichtsforschung. Methodologische und praktische Ansätze zur Erforschung von Lernprozessen* (S. 87–110). Donauwörth: Auer.
- Krapp, A. (1998). Entwicklung und Förderung von Interessen im Unterricht. *Psychologie in Erziehung und Unterricht*, 45, 186–203.

- Krapp, A. (2002). An educational-psychological theory of interest and its relation to self-determination theory. In E. Deci, R. M. Ryan (ed.), *The handbook of self-determination research* (pp. 405–427). Rochester: University of Rochester Press.
- Krug, S. & Hanel, J. (1976). Motivänderung. Erprobung eines theoriegeleiteten Trainingsprogramms. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 8, 274–287.
- Laukenmann, M., Bleicher, M., Fuß, S., Gläser-Zikuda, M., Mayring, P. & v. Rhöneck, C. (2003). An investigation on the influence of emotions on learning in physics. *International Journal of Science Education*, 25, 489–507.
- Lenhard, W. & Lenhard, A. (2015). *Normwertrechner*. verfügbar unter: <https://www.psychometrica.de/normwertrechner.html>. Bibergau: Psychometrica. DOI: 10.13140/RG.2.1.4592.5363
- Leppner, M. R. & Greene, D. (1975). Turning play into work: Effects of adult surveillance and extrinsic rewards on children's intrinsic motivation. *Journal of Personality and Social Psychology*, 31, 479–486.
- Leutner, D. (2012). Perspektiven pädagogischer Interventionsforschung. In E. Severin (Hrsg.), *Qualitätsentwicklung in der Berufsbildungsforschung* (S. 17–28). Bielefeld: Bertelsmann. <http://doi.org/10.3278/111-054w017>
- Lewalter, D. & Bieg, S. (2009). Editorial zum Themenheft: Lernmotivation – Bedingungen erkunden, Förderungsmöglichkeiten entwickeln. *Psychologie in Erziehung und Unterricht*, 56(4), 241–242.
- McClelland. (1965). Toward a theory of motivation acquisition. *American Psychologist*, 20, 321–332.
- McGraw, S. A., Sellers, D., Stone, E., Resnicow, K. A., Kuester, S., Fridinger, F., & Wechsler, H. (2000). Measuring implementation of school programs and policies to promote healthy eating and physical activity among youth. *Preventive Medicine*, 31(2), 86–S97. <http://doi.org/10.1006/pmed.2000.0648>
- McGrew, J. H., Bond, G. R., Dietzen, L., & Salyers, M. (1994). Measuring the fidelity of implementation of a mental health program model. *American Psychological Association*, 62(4), 670–678.
- Meyer, A., Miller, S. & Herman, M. (1993). Balancing the priorities of evaluation with the priorities of the setting: A focus on positive youth development programs in school settings. *Journal of Primary Prevention*, 14(2), 95–113.
- Meyer, S. (2009). Entwicklung und Evaluation eines Trainings zur Förderung der Lesekompetenz und Lesemotivation (LekoLemo) für die Sekundarstufe I. *Universität Bielefeld*. Verfügbar unter <http://pub.uni-bielefeld.de/luur/download?func=downloadFile&recordId=2305609&fileId=2305612> [09.01.2018]
- Mihalic, S. (2004). The importance of implementation fidelity. *Emotional & Behavioral Disorders in Youth*, 4(4), S. 83–86.
- Mihalic, S. F., Fagan, A. A. & Argamaso, S. (2008). Implementing the life skills training drug



- prevention program: factors related to implementation fidelity. *Implementation Science IS*, 3, 5.
- Mittag, W. (2011). Zur Bedeutung der Lehrer-Schüler-Beziehung für die Wahrnehmung der Unterrichtsqualität und die Lernmotivation. *Unveröffentlichte Präsentation an der Universität Erfurt*.
- Mittag, W. (2014). Qualitätssicherung von Präventions- und Interventionskonzepten aus Sicht der Interventions-, Implementations- und Evaluationsforschung. *Wirtschaftspsychologie*, 16(2), 50–58.
- Mittag, W., & Bieg, S. (2010). Die Bedeutung und Funktion pädagogischer Interventionsforschung und deren grundlegende Qualitätskriterien. In T. Hascher & B. Schmitz (Hrsg.), *Pädagogische Interventionsforschung. Theoretische Grundlagen und empirisches Handlungswissen* (S. 31–47). Weinheim: Juventa.
- Mittag, W., Bieg, S., Hiller, F., Metz, K. & Melenk, H. (2009). Förderung selbstbestimmter Lernmotivation im Deutschunterricht. *Psychologie in Erziehung und Unterricht*, 56, 271–286.
- Mittag, W. & Hager, W. (2000). Ein Rahmenkonzept zur Evaluation psychologischer Interventionsmaßnahmen. In W. Hager, J.-L. Patry, & H. Brezing (Hrsg.), *Evaluation psychologischer Interventionsmaßnahmen. Standards und Kriterien: Ein Handbuch* (S. 102–128). Göttingen: Hans Huber.
- Moncher, F. J. & Prinz, R. J. (1991). Treatment studies in outcome studies. *Clinical Psychology*, 11, 247–266.
- Mossholder, K. W. (1980). Effects of externally mediated goal setting on intrinsic motivation: A laboratory experiment. *Journal of Applied Psychology*, 65, 202–210.
- Mowbray, C. T., Holter, M. C., Teague, G. B. & Bybee, D. (2003). Fidelity Criteria: Development, measurement, and validation. *The American Journal of Evaluation*, 24(3), 315–340. [http://doi.org/10.1016/S1098-2140\(03\)00057-2](http://doi.org/10.1016/S1098-2140(03)00057-2)
- Newmann, F. M., Smith, B. A., Allensworth, E. & Bryk, A. S. (2001). Instructional program coherence: What it is and why it should guide school improvement policy. *Educational Evaluation and Policy Analysis*, 23(4).
- O'Connor, C., Small, S. A. & Cooney, S. M. (2007). Program fidelity and adaptation: Meeting local needs without compromising program effectiveness. *What Works, Wisconsin – Research to Practice Series*, 4, 1–6.
- O'Brien, R. (2005). Translating a research intervention into community practice: The nurse-family partnership. *The Journal of Primary Prevention*, 26, 241–257.
- Olweus, D. & Alsaker, F. D. (1991). Assessing change in a cohort longitudinal study with hierarchical data. In D. Magnusson, L. R. Bergman, G. Rudinger, & B. Torestad (Eds.), *Problems and methods in longitudinal research* (pp. 107–132). New York: Cambridge University Press.

- Olweus, D. & Limber, S. (2010). *The Olweus bullying prevention program: Implementation and evaluation over two decades*. In S. R. Jimerson, S. M. Swearer, & D. L. Espelage (Eds.), *The Handbook of Bullying in Schools: An International Perspective*. New York: Routledge.
- Otto, B., Kistner, S., Perels, F., Schmitz, B. & Büttner, G. (2009). Effekte direkter und indirekter Interventionen auf die Lernmotivation. *Psychologie in Erziehung und Unterricht*, 56, 287–302.
- Pant, H. A., Vock, M., Pöhlmann, C. & Köller, O. (2008). Offenheit für Innovationen. Befunde aus einer Studie zur Rezeption der Bildungsstandards bei Lehrkräften und Zusammenhänge mit Schülerleistungen. *Zeitschrift für Pädagogik*, 54(6), 827–845.
- Payne, A. A., Gottfredson, D. C. & Gottfredson, G. D. (2006). School Predictors of the intensity of implementation of school-based prevention programs: results from a National Study. *Prevention Science*, 7(2), 225–237.
- Pentz, M. A., Trebow, E. A., Hansen, W. B., MacKinnon, D. P., Dwyer, J. H., Johnson, C. A. et al. (1990). Effects of program implementation on adolescent drug use behavior: The Midwestern Prevention Project (MPP). *Evaluation Review*, 14(3), 264–289. <http://doi.org/10.1177/0193841X9001400303>
- Penuel, W. & Means, B. (2004). Implementation variation and fidelity in an Inquiry Science Program: Analysis of GLOBE data reporting patterns. *Journal of Research in Science Teaching*, 41, 294–315.
- Petermann, F. (2006). Veränderungsmessung. In H. R. Rost (Ed.), *Handwörterbuch Pädagogische Psychologie* (3 Aufl., S. 821–827). Weinheim: Beltz PVU.
- Peterson, L., Mori, L., Selby, V. & Rosen, B. N. (1988). Community interventions in children's injury prevention: Differing costs and differing benefits. *Journal of Community Psychology*, 16, 188–204.
- Rakoczy, K., Klieme, E. & Pauli, C. (2008). Die Bedeutung der wahrgenommenen Unterstützung motivationsrelevanter Bedürfnisse und des Alltagsbezugs im Mathematikunterricht für die selbstbestimmte Motivation. *Zeitschrift für Pädagogische Psychologie*, 22(1), 25–35.
- Raudenbush, S., Bryk, A., Cheong, Y. & Congdon, R. (2011). HLM 7-Hierarchical linear and nonlinear modeling. Lincolnwood: Scientific Software International.
- Reeve, J. (2012). A self-determination theory perspective on student engagement. In S. L. Christenson, A. L. Reschly, & C. Wylie (Eds.), *Handbook of Research on Student Engagement* (pp. 149–172). New York: Springer.
- Reeve, J., Deci, E. L. & Ryan, R. M. (2004). Self-determination theory: A dialectic framework for understanding the sociocultural influences on student motivation. In D. McInerney & S. Van Etten (Eds.), *Research on sociocultural influences on student motivation and learning: Big theories revisited* (4th ed., pp. 31–59). Greenwich, CT: Information Age.
- Reeve, J., & Jang, H. (2006). What teachers say and do to support students' autonomy during learning activities. *Journal of Educational Psychology*, 98, 209–218.

- Reeve, J., Jang, H., Hadre, P. M. & Omura, M. (2002). Providing a rationale in an autonomy-supportive way as a strategy to motivate others during an uninteresting activity. *Motivation and Emotion*, 26, 183–207.
- Reeve, J., Nix, G., & Hamm, D. (2003). Testing models of the experience of self-determination in intrinsic motivation and the conundrum of choice. *Journal of Educational Psychology*, 95, 375–392.
- Reis, H. T., Sheldon, K. M., Gable, S. L., Roscoe, J. & Ryan, R. M. (2000). Daily well-being: The role of autonomy, competence, and relatedness. *Personality and Social Psychology Bulletin*, 419–435.
- Resnick, B., Bellg, A. J., Borrelli, B., Defrancesco, C., Breger, R., Hecht, J. et al. (2005). Examples of implementation and evaluation of treatment fidelity in the BCC studies: Where we are and where we need to go. *Annals of Behavioral Medicine a Publication of the Society of Behavioral Medicine*, 29 Suppl(7), 46–54.
- Resnicow, K. A., Cohn, L., Reinhardt, J., Cross, D., Futterman, R., Kirschner, E. et al. (1992). A three-year evaluation of the Know Your Body Program in inner-city schoolchildren. *Health Education Quarterly*, 16(4), 463–480.
- Resnicow, K., Davis, M., Smith, M., Lazarus-Yaroch, A., Baranowski, T., Baranowski, J. et al. (1998). How best to measure implementation of school health curricula: A comparison of three measures. *Health Education Research*, 13(2), 239–50.
- Rheinberg, F. (1987). Soziale vs. individuelle Leistungsvergleiche und ihre motivationalen Folgen in Lehr- und Lernsituationen. In R. Olechowski & E. Persy (Hrsg.), *Fördernde Leistungsbeurteilung* (S. 80–115). Jugend und Volk.
- Rheinberg, F. (2002). *Motivation*. Stuttgart: Kohlhammer.
- Rheinberg, F. (2004). *Motivation*. Stuttgart: Kohlhammer.
- Rheinberg, F. & Fries, S. (2001). Motivationstraining. In K. J. Klauer (Hrsg.), *Handbuch Kognitives Training* (S. 349–371). Göttingen: Hogrefe.
- Rheinberg, F. & Krug, S. (1999). *Motivationsförderung im Schulalltag* (2. Aufl.). Göttingen: Hogrefe.
- Röder, B. & Jerusalem, M. (2007). Implementationsgrad und Wirkungen eines Programms zur Förderung von Selbstwirksamkeit. *Schweizerische Zeitschrift für Psychologie und ihre Anwendungen*, 54, 30 – 46.
- Röder, B. & Kleine, D. (2003). Selbstbestimmung. In M. Jerusalem (Hrsg.), *Skalendokumentation zum Forschungsprojekt "Selbstwirksamkeit und Selbstbestimmung im Unterricht"*. Berlin: Humboldt-Universität zu Berlin.
- Rohrbach, L. A., Dent, C. W., Skara, S., Sun, P. & Sussman, S. (2007). Fidelity of implementation in Project Towards No Drug Abuse (TND): a comparison of classroom teachers and program specialists. *Prevention Science the Official Journal of the Society for Prevention Research*, 8(2), 125–132. <http://doi.org/10.1007/s11121-006-0056-z>.

- Rohrbach, L. A., Graham, J. W. & Hansen, W. B. (1993). Diffusion of a school-based substance abuse prevention program: predictors of program implementation. *Preventive Medicine, 22*, 237–260.
- Rohrbach, L. A., Gunning, M., Sun, P. & Sussman, S. (2010). The Project Towards No Drug Abuse (TND) dissemination trial: Implementation fidelity and immediate outcomes. *Prevention Science the Official Journal of the Society for Prevention Research, 11*(1), 1–17. <http://doi.org/10.1007/s11121-009-0151-z>.The
- Rosenthal, R. (1994). Parametric measures of effect size. In H. Cooper & L. V. Hedges (Eds.), *The Handbook of Research Synthesis* (pp. 231–244). New York, NY: Sage.
- Rossi, P. H. & Freeman, H. E. (1993). *Evaluation: A Systematic Approach* (5th ed.). Thousand Oaks: SAGE Publications.
- Rossi, P. H., Freeman, H. E. & Hofmann, G. (1988). *Programm-Evaluation. Einführung in die Methoden angewandter Sozialforschung*. Stuttgart: Enke.
- Ruiz-Primo, M. A. (2005). A multi-method and multi-source approach for studying fidelity of implementation. In *The annual meeting of the American Education Research Association*. Montreal, Canada.
- Ryan, R. M. (1982). Control and information in the intrapersonal sphere: An extension of cognitive evaluation theory. *The Journal of Personality and Social Psychology, 43*, 450–461.
- Ryan, R. M. (1993). Agency and organization: Intrinsic motivation, autonomy and the self in psychological development. In J. Jacobs (Ed.), *Nebraska symposium on motivation: Developmental perspectives on motivation* (Vol. 40, pp. 1–56). Lincoln, NE: University of Nebraska Press.
- Ryan, R. M. & Connell, J. P. (1989). Perceived locus of causality and internalization: Examining reasons for acting in two domains: *Journal of Personality and Social Psychology, 57*, 749–761.
- Ryan, R. M. & Deci, E. L. (2000a). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology, 25*(1), 54–67. <http://doi.org/10.1006/ceps.1999.1020>
- Ryan, R. M. & Deci, E. L. (2000b). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist, 55*(1), 68–78.
- Ryan, R. M. & Deci, E. L. (2002). Overview of self-determination theory: An organismic dialectical perspective. In R. M. Ryan & E. L. Deci (Eds.), *Handbook of self-determination research* (pp. 3–33). Rochester, NY: The University of Rochester Press.
- Schiefele, U. (2004). Förderung von Interessen. In G. W. Lauth, M. Grünke, & J. C. Brunstein (Hrsg.), *Interventionen bei Lernstörungen* (S. 134–144). Göttingen: Hogrefe.
- Shapley, K. S., Sheehan, D., Maloney, C. & Caranikas-Walker, F. (2010). Evaluating the implementation fidelity of technology immersion and its relationship with student achievement. *Journal of Technology, 9*(4).

- Skinner, E. A. & Belmont, M. J. (1993). Motivation in the classroom: Reciprocal effects of teacher behavior and student engagement across the school year. *Journal of Educational Psychology, 85*, 571 – 581.
- Smylie, M. A. (1988). The enhancement function of staff development: Organizational and psychological antecedents to individual teacher change. *American Educational Research Journal, 25*, 1–30.
- Snyder, J., Bolin, F. & Zumwalt, K. (1992). Curriculum implementation. In P. W. Jackson (Ed.), *Handbook of research on curriculum* (pp. 402–435). New York: Macmillan Publishing Company.
- Songer, N. B. & Gotwals, A. W. (2005). *Fidelity of implementation in three sequential curricular units. Paper presented at the Annual Meeting of the American Education Research Association, Montreal, Canada.* Verfügbar unter <http://www.biokids.umich.edu/papers/songergotwals.fidelity05.pdf> [09.01.2018]
- Spoth, R., Gyll, M., Lillehoj, C. & Greenberg, M. (2007). Prosper study of evidence-based intervention implementation quality by community-university partnerships. *Community Psychology, 35*(8), 981–999.
- Spoth, R., Gyll, M., Trudeau, L. & Goldberg-Lillehoj, C. (2002). Two studies of proximal outcomes and implementation quality of universal preventive interventions in a community university collaboration context. *Journal of Community Psychology, 30*, 499–518.
- Stegmueller, D. (2013). How many countries for Multilevel Modeling? A comparison of Frequentist and Bayesian approaches. *American Journal of Political Science, 57*(3), 748–761.
- Streblov, L., Holodyski, M. & Schiefele, U. (2007). Entwicklung eines Lesekompetenz- und Lesemotivationsstrainings für die siebte Klassenstufe. Bericht über zwei Evaluationsstudien. *Psychologie in Erziehung und Unterricht, 54*(4), 287–297.
- Tobler, N. K. (1986). Meta-analysis of 143 adolescent drug prevention programs: Quantitative outcome results of program participants compared to a control or comparison group. *Journal of Drug Issues, 16*, 537–567.
- Vallerand, R. J., Fortier, M. S., & Guay, F. (1997). Self-determination and persistence in a real-life setting: Toward a motivational model of high school dropout. *Journal of Personality and Social Psychology, 72*, 1161 – 1176.
- Vansteenkiste, M., Lens, W. & Deci, E. L. (2006). Intrinsic versus extrinsic goal contents in self-determination theory: Another look at the quality of academic motivation. *Educational Psychologist, 41*, 19–31.
- Vansteenkiste, M., Niemiec, C. P. & Soenens, B. (2010). The development of the five mini-theories of self-determination theory: an historical overview, emerging trends, and future directions. *Advances in Motivation and Achievement: The Decade Ahead: Theoretical Perspectives on Motivation and Achievement, 16A*, 105–167.
- Vansteenkiste, M., Simons, J., Lens, W., Sheldon, K. M. & Deci, E. L. (2004). Motivating learning, performance, and persistence: The synergistic effects of intrinsic goal contents and autonomy-supportive contexts. *Journal of Personality and Social Psychology, 87*(2),

246–260.

- Vaughn, S., Klingner, J. & Hughes, M. (2000). Sustainability of research-based practices. *Exceptional Children*, 66(2), 163–171.
- von Saldern, M. & Littig, K. E. (1987). *Landauer Skalen zum Sozialklima (Lasso 4-13)*. Weinheim: Beltz.
- Weinstein, N., Hodgins, H. S. & Ryan, R. M. (2010). Autonomy and control in dyads: effects on interaction quality and joint creative performance. *Personality and Social Psychology Bulletin*, 36, 1603–1617. <http://doi.org/doi: 10.1177/0146167210386385>
- Wigfield, A. & Eccles, J. (2000). Expectancy-value theory of achievement motivation. *Contemporary Educational Psychology*, 25(1), 68–81. <http://doi.org/10.1006/ceps.1999.1015>
- Wild, E., Rammert, R. & Siegmund, A. (2006). Die Förderung selbstbestimmter Formen der Lernmotivation in Elternhaus und Schule. In M. Prenzel & L. Allolio-Näcke (Hrsg.), *Untersuchungen zur Bildungsqualität von Schule* (S. 370–397). Münster: Waxmann.
- Wilson, D. K., Griffin, S., Saunders, R. P., Kitzman-Ulrich, H., Meyers, D. C. & Mansard, L. (2009). Using process evaluation for program improvement in dose, fidelity and reach: the ACT trial experience. *The International Journal of Behavioral Nutrition and Physical Activity*, 6, 79. <http://doi.org/10.1186/1479-5868-6-79>
- Wilson, S. J. & Lipsey, M. W. (2000). Effects of school violence prevention programs on aggressive and disruptive behavior: a meta-analysis of outcome evaluations. *Nashville: Center for Evaluation Research and Methodology, Vanderbilt Institute for Public Policy*.
- Wilson, S. J., Lipsey, M. W. & Derzon, J. H. (2003). The effects of school-based intervention programs on aggressive behavior: a meta-analysis. *Journal of Consulting and Clinical Psychology*, 71, 136–149.

## Anhang

<b>Feedback zur Unterrichtsstunde</b>		
<i>Bitte beantworten Sie folgende Fragen nach jeder Unterrichtsstunde.</i>		
Name:	Klasse:	Fach: Deutsch
Unterrichtsstunde Nr.		Datum:

<b>1. Thema der heutigen Stunde:</b>
<b>2. Unterrichtsziele:</b>

<b>3. Welche Medien haben Sie eingesetzt? Zusätzliches bitte eintragen.</b>			
Tafel	<input type="radio"/>	Schulbuch	<input type="radio"/>
Abbildung / Grafik / (Wand-)Plakat	<input type="radio"/>	weitere Bücher / Lexika	<input type="radio"/>
Internet / Lernsoftware / Computer	<input type="radio"/>	reguläres Schülerheft	<input type="radio"/>
Lösungskarten / -heft	<input type="radio"/>	Spezielles Schülerheft	<input type="radio"/>
Arbeitsblatt / -blätter	<input type="radio"/>	Film / Dia / Video	<input type="radio"/>
OHP	<input type="radio"/>	Sonstiges:	

<b>4. Welche Arbeitsformen kamen in der Stunde vor?</b>			
Arbeit im Klassenverband	<input type="radio"/>	Gruppenarbeit (arbeitsgleich)	<input type="radio"/>
Einzelarbeit	<input type="radio"/>	Gruppenarbeit (arbeitsteilig)	<input type="radio"/>
Partnerarbeit	<input type="radio"/>	Arbeit an Stationen	<input type="radio"/>
Lernspiel / Quiz	<input type="radio"/>	Lerntheke / Werkstatt	<input type="radio"/>
Schüler-Präsentation / Schüler-Referat / Expertenrunde	<input type="radio"/>	Sonstiges:	

<b>5. Wie haben die Schüler diese Unterrichtsstunde wahrgenommen? Schätzen Sie bitte die Schüler bzgl. folgender Aspekte ein:</b>					
	hoch / gut		mittel		niedrig / schlecht
Interesse	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Lernfreude	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Mitarbeit	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Selbstbestimmung	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Kompetenzerleben	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Soziales Miteinander	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Verständnisschwierigkeiten	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Unterrichtsstörungen	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>



6. Ihre Einschätzung zur Unterrichtsstunde insgesamt:					
	sehr gut / sehr positiv		mittel	sehr schlecht / sehr negativ	
Wie gut ist Ihrer Meinung nach die Durchführung der Unterrichtsstunde gelungen?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Wie haben Sie diese Unterrichtsstunde emotional erlebt?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Gründe aus Ihrer Sicht:					

7. Traten bei der Stunde (so wie sie vorgesehen war) Schwierigkeiten auf?	
Art der Schwierigkeit	Konsequenz (z.B. Abbruch, Kürzung etc.)
_____	_____
_____	_____
_____	_____
_____	_____
_____	_____
_____	_____

8. Haben Sie in dieser Unterrichtsstunde weitere Veränderungen vorgenommen?	
Vorgenommene Veränderung	Grund
_____	_____
_____	_____
_____	_____
_____	_____
_____	_____
_____	_____

9. Haben Sie kritische Anmerkungen und Verbesserungsvorschläge?
_____
_____
_____
_____
_____
_____
_____

**Vielen Dank für Ihre Mitarbeit!**