



universität
wien

Diplomarbeit

Titel der Diplomarbeit

Durchführung des Schulprojekts *Reflect* in der Praxis

Implementierungstreue und ihre Relevanz für die

Programmwirksamkeit

Verfasserin

Michaela Pichler, BA

Angestrebter akademischer Grad

Magistra der Naturwissenschaften (Mag. rer. nat.)

Wien, 2014

Studienkennzahl: A 298

Studienrichtung: Psychologie

Betreuerin: Univ.-Prof. Dipl.-Psych. Dr. Barbara Schober

Zunächst möchte ich Barbara Schober für die Betreuung dieser Diplomarbeit danken. Danke für die vielen Anregungen und interessanten Diskussionen!

Des Weiteren möchte ich mich besonders bei Marie-Therese Schultes für ihre engagierte Mitbetreuung der Diplomarbeit bedanken: dafür, dass sie mein Interesse immer wieder durch spannende Gespräche geschürt und mich fachlich und emotional im Entstehungsprozess unterstützt hat.

Vielen herzlichen Dank auch an Monika Finsterwald und Gregor Jöstl für die ExpertInnenratings, ohne die meine Diplomarbeit nicht funktioniert hätte, und Gregor Jöstl darüber hinaus noch für die gemeinsame Konsens-Suche, die trotz aller Mühen auch viel Spaß gemacht hat.

Des Weiteren liegt es mir am Herzen, mich bei meinen StudienkollegInnen zu bedanken, die durch unzählige Denkanstöße und ihre Unterstützung auf verschiedensten Ebenen zum Gelingen dieser Arbeit beigetragen haben. Vor allem meine KollegInnen aus dem Projektstudium und aus dem Forschungsseminar haben mich immer wieder angeregt und motiviert.

Nicht nur während meines Studiums wichtig waren und sind meine Familie und meine FreundInnen. Danke für alles!

Einleitung.....	1
1. Implementierungsforschung: Ein Überblick.....	3
2. Implementierungstreue	8
2.1. Definition	8
2.2. Implementierungstreue und -adaption – ein Gegensatz?	9
2.3. Zusammenhang mit Trainingseffekten.....	10
2.4. Förderung von Implementierungstreue in der Praxis	13
2.5. Erfassung von Implementierungstreue	14
2.5.1. Identifikation und Operationalisierung der Core Components	15
2.5.2. Datenerhebung	17
2.5.3. Validierung und Ebene der Analyse	19
3. Das Schulprojekt <i>Reflect</i>	21
3.1. Theoretische Grundlagen	21
3.1.1. Die Selbstbestimmungstheorie und Förderung von Motivation	22
3.2. Umsetzung.....	26
3.2.1. Vermittelte Umsetzungsmöglichkeiten der Selbstbestimmungstheorie	26
3.2.2. Die Klassenprojekte und Dokumentation der Umsetzung	27
4. Forschungsanliegen.....	28
5. Methode.....	32
5.1. Datenerhebung.....	32
5.2. Stichprobe.....	33
5.3. Instrumente	33
5.4. Statistische Auswertung	39
5.4.1. Mehrebenenanalyse	40
6. Ergebnisse.....	44
6.1. Reliabilitäten der Skalen und Interrater-Übereinstimmungen	44

6.2. Beurteilungen der Implementierungstreue von SchülerInnen, LehrerInnen und ExpertInnen	45
6.3. Unterschiede zwischen den verschiedenen Beurteilungen der Implementierungstreue	47
6.4. Zusammenhang zwischen Implementierungstreue und Wirksamkeit der Klassenprojekte	48
7. Diskussion	50
7.1. Zusammenfassung und Interpretation der Ergebnisse	50
7.1.1. Beurteilungen der Implementierungstreue von SchülerInnen, LehrerInnen und ExpertInnen.....	50
7.1.2. Unterschiede zwischen den verschiedenen Beurteilungen der Implementierungstreue	50
7.1.3. Zusammenhang zwischen Implementierungstreue und Wirksamkeit der Klassenprojekte.....	52
7.2. Forschungsausblick.....	55
8. Zusammenfassung.....	59
Literaturverzeichnis.....	62
Abbildungsverzeichnis.....	70
Tabellenverzeichnis	70
Anhang.....	71
I. Verwendete Skalen.....	71
Skala Umgang mit Unterschiedlichkeit (SchülerInnen)	71
Skalen zur Beurteilung der implementierungstreuen Umsetzung der Selbstbestimmungstheorie (SchülerInnen)	71
Skalen zur Beurteilung der implementierungstreuen Umsetzung der Selbstbestimmungstheorie (LehrerInnen).....	72
Beurteilungsbogen des ExpertInnenratings.....	73
II. Lebenslauf.....	74

Einleitung

Gerade im Bildungs- und Gesundheitsbereich ist man seit einiger Zeit bestrebt, Präventions- und Interventionsprogramme anzubieten, die auf aktuellsten Forschungserkenntnissen beruhen und deren Effektivität im Vorfeld mittels wissenschaftlich fundierter Evaluationen bestätigt werden konnte. Doch der Transfer von Forschungserkenntnissen in die Praxis ist kein leichtes Unterfangen und selbst die unter strengen Forschungsbedingungen effektivsten Programme können bei der Umsetzung im Feld an Effektivität einbüßen oder sogar gänzlich ineffektiv werden. Lange Zeit hat man sich bei der Evaluation von Interventionen nur die Trainingseffekte angesehen und die Programme lediglich danach beurteilt, wie groß die erzielten Veränderungen in der Stichprobe waren. Den Prozess der Durchführung (Implementierung) hat man gänzlich aus der Untersuchung ausgeklammert und wie eine Art „Black-Box“ behandelt (Nelson, Cordray, Hulleman, Darrow, & Sommer, 2012). Doch die Art der Durchführung ist einer der wesentlichsten Faktoren für die Wirksamkeit von Interventionen (Fixsen, Blase, Naoom, & Wallace, 2009). Relevante Merkmale der Durchführung können aufgrund verschiedener TrainerInnen, TeilnehmerInnen, Kontextfaktoren etc. in der Praxis oft stark variieren (Dusenbury, Brannigan, Falco, & Hansen, 2003). Aufgrund von Mangel an Ressourcen, guter Ausbildung etc. werden Interventionen oftmals verkürzt oder inhaltlich inadäquat umgesetzt, was sich direkt auf die Effektivität der Interventionen auswirken kann (Lane, Bocian, MacMillan, & Gresham, 2004).

Die Implementierungsforschung ist nun darum bemüht, die genauen Wirkmechanismen, welche bei der Implementierung von Interventionen beteiligt sind, zu analysieren und zu optimieren. Dabei werden verschiedene Aspekte der Durchführung betrachtet: Merkmale der TrainerInnen, der TeilnehmerInnen, Rahmenbedingungen etc. (Mowbray, Holter, Teague, & Bybee, 2003). Eines der wichtigsten Merkmale guter Implementierung ist die implementierungstreue Umsetzung einer Intervention (Umsetzung gemäß dem ursprünglichen Interventionsplan). Je genauer sich TrainerInnen bei der Durchführung an den Originalplan halten, desto größer sind die durch die Intervention erzielten Effekte (Berkel, Mauricio, Schoenfelder, & Sandler, 2011). Zur Erhebung von Implementierungstreue kann man verschiedene Personengruppen (TrainerInnen, TeilnehmerInnen, ExpertInnen) mittels unterschiedlicher Methoden befragen. Selten werden bei Programmevaluationen jedoch überhaupt Angaben zur Implementierungstreue gemacht und wenn Daten

vorhanden sind, so wurden diese zumeist durch die Befragung von nur einer Personengruppe ermittelt (Durlak & DuPre, 2008). Falls mehrere verschiedene Gruppen befragt werden, so werden die Perspektiven kaum einmal miteinander verglichen (Durlak & DuPre, 2008; Dusenbury et al., 2003). Außerdem werden Daten zur Implementierungstreue zu selten mit den erzielten Trainingseffekten in Zusammenhang gebracht (Berkel et al., 2011), sodass es bis jetzt nur wenige Erkenntnisse über die Validität verschiedener Perspektiven und Methoden gibt.

Ziel dieser Studie ist es deshalb, verschiedene Perspektiven auf die Implementierungstreue des Schulprojekts *Reflect* zu vergleichen und die Zusammenhänge der Beurteilungen mit einem Trainingseffekt zu analysieren. *Reflect* war ein Projekt der Universität Wien zur nachhaltigen Verankerung des Kompetenzerwerbes zur reflexiven Koedukation in der LehrerInnenbildung (Finsterwald et al., 2012). Im Zuge dieses Trainings führten die Lehrkräfte supervidiert durch das Universitätspersonal in ihren Klassen Projekte zum Thema „Gender und Beruf“ durch. Diese Arbeit vergleicht drei Beurteilungen der implementierungstreuen Umsetzung dieser Klassenprojekte: Beurteilungen aus Sicht der SchülerInnen, der LehrerInnen, sowie von ExpertInnen der Universität Wien. Basierend auf diesem Vergleich werden Vorteile und Nachteile der verschiedenen Beurteilungsperspektiven diskutiert mit dem Ziel, Informationen zu sammeln, die bei der Planung von Implementierungsevaluationen nützlich sein können: Wie sehen die verschiedenen Beurteilungen aus? Nehmen TrainerInnen, TeilnehmerInnen und ExpertInnen die implementierungstreue Umsetzung ähnlich oder gänzlich unterschiedlich wahr? Welche der Perspektiven kann die Wirksamkeit eines Programms am besten vorhersagen?

1. Implementierungsforschung: Ein Überblick

In den letzten Jahrzehnten ist man im Gesundheits- und Bildungsbereich immer mehr darum bemüht, Interventionen gemäß dem aktuellen Stand der Forschung zu entwickeln und zu evaluieren, um die Nutzung evidenzbasierter Interventionen zu verbreiten (Fixsen et al., 2009). Gerade im anglo-amerikanischen Raum sind diese Bestrebungen bereits weit verbreitet (Spiel & Strohmeier, 2012). Eine Intervention gilt dann als „evidenzbasiert“, wenn nach strengen wissenschaftlichen Kriterien überprüft wurde, ob die Intervention die angestrebten Programmziele tatsächlich erfüllen kann (Esbensen, Matsuda, Taylor, & Peterson, 2011). Mittlerweile gibt es ein breites Angebot an solchen Programmen, welche unter kontrollierten Bedingungen gute Effekte erzielen – jedoch fehlt es teilweise noch an Anhaltspunkten, wie diese erfolgreich in der Praxis umgesetzt werden können (Fagan, Hanson, Hawkins, & Arthur, 2008). Der Transfer von Forschungserkenntnissen in die Praxis ist kein leichtes Unterfangen. Nicht selten scheitern Programme an der konkreten Umsetzung im Feld und werden dadurch ineffektiv oder in ihrer Wirkung sogar kontraproduktiv (Durlak & DuPre, 2008; Fixsen et al., 2009). Selbst die effektivsten Programme können dadurch nutzlos werden (Elliott & Mihalic, 2004).

Um ein evidenzbasiertes Interventionsprogramm erfolgreich und längerfristig in der Praxis zu etablieren, müssen verschiedene Schritte realisiert werden (Durlak & DuPre, 2008; Durlak, 1998): Informationen über das Programm und seine Nützlichkeit müssen an InteressentInnen herangetragen werden (Dissemination) und die Organisationen müssen sich dazu entschließen, das Programm tatsächlich einzuführen (Adoption). Erst dann kann die Phase der Implementierung (Implementation, Durchführung) beginnen, welche idealerweise in die Phase der Nachhaltigkeit (Sustainability, Maintainance) mündet, in der das Programm auch nach einer Testphase weitergeführt wird. Nur wenn ein Programm auch in der Praxis die erwünschten Effekte erzielen kann, besteht die Chance, dass es längerfristig weitergeführt wird. Die meisten breit angelegten Programme, welche beispielsweise landesweit in mehreren Schulen implementiert werden, werden allerdings von Personen durchgeführt, welche nicht in deren Entwicklung eingebunden waren (Durlak, 1998). So ist es wenig verwunderlich, dass die Qualität der Implementierung über verschiedene TrainerInnen hinweg meist stark variiert (Dusenbury et al., 2003). Beispielsweise konnten Gottfredson und Gottfredson (2002) in einer Meta-Analyse von schulbasierten

Präventionsprogrammen zeigen, dass die Hälfte aller untersuchten Programme in der Praxis nur unregelmäßig oder unvollständig durchgeführt wurden. Darüber hinaus waren die verwendeten Methoden und Instruktionen oftmals nicht passend und stimmten nicht mit den Empfehlungen der ProgrammentwicklerInnen überein. Hallfors und Godette (2002) fanden ähnliche Ergebnisse: Die Umsetzungen von Präventionsprogrammen in US-amerikanischen Schulen variierten in ihrer Qualität sehr stark. Oftmals wurden die Präventionsprogramme von unausgebildeten TrainerInnen abgehalten, es fehlten relevante Materialien zur Umsetzung oder die Programme wurden bei Zielgruppen implementiert, für die die Methoden und Inhalte (z.B. aufgrund des Alters) nicht geeignet oder irrelevant waren. Trotz der aufgezeigten Defizite in diesem Bereich wurde die Phase der Implementierung in der Evaluationsforschung lange Zeit erheblich vernachlässigt (Durlak & DuPre, 2008). Es wurde stets mehr Wert auf die Überprüfung von Trainingseffekten als auf die Überwachung des Prozesses der Implementierung gelegt (Elliott & Mihalic, 2004; Mihalic & Irwin, 2003). In einer klassischen Outcome-Evaluation mit Versuchs- und Kontrollgruppendesign interessiert nur die randomisierte Zuteilung zu den beiden Gruppen – den Unterschied bildet lediglich der Erhalt der Intervention. Die gemessenen Unterschiede in den interessierenden Outcomevariablen werden allein darauf zurückgeführt. Alles, was während der Durchführungsphase in der Interventions- aber auch in der Kontrollgruppe passiert, wird klassischer Weise nicht gemessen. Diese Phase bildet eine Art „Black Box“ in der Evaluation (Nelson et al., 2012).

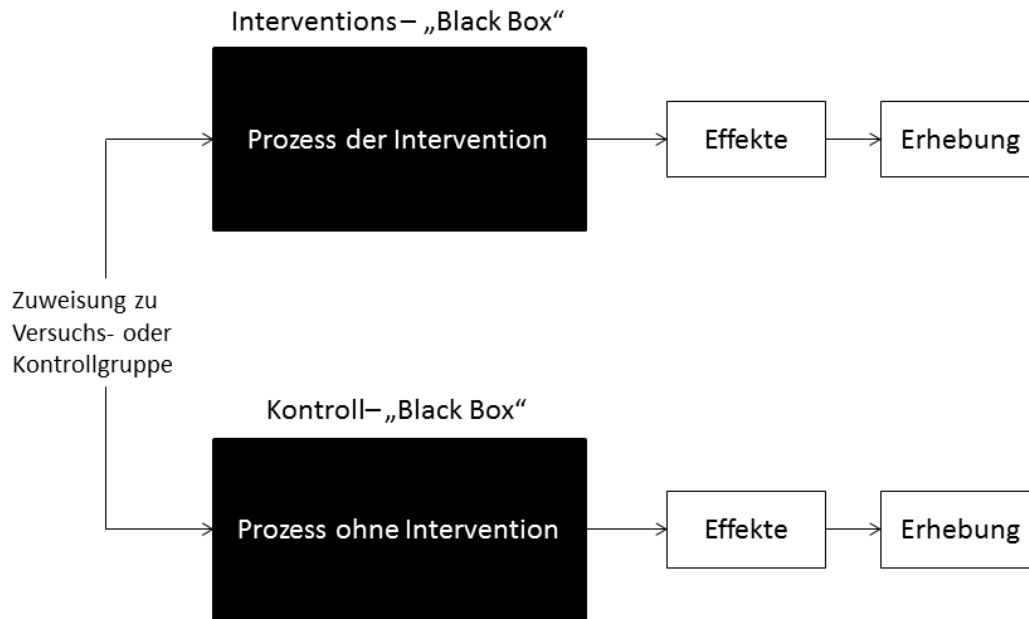


Abbildung 1: klassisches Evaluationsmodell mit Versuchs- und Kontrollgruppe, Grafik angelehnt an Nelson et al. (2012), S. 376

Die Implementierungsforschung setzt genau hier an und untersucht systematisch, wie gut Interventionen in der Praxis umgesetzt werden und welchen Modifizierungen sie unterliegen. Man hat sich vom „Black-Box Modell“ der Intervention verabschiedet, und versucht, die genauen Wirkmechanismen, die während der Durchführung einer Intervention bedeutsam werden, zu verstehen (Mowbray et al., 2003). Ein weiteres Interesse der Implementierungsforschung besteht darin, wie man den Transfer von Erkenntnissen aus der Forschung in die Praxis möglichst effektiv gestalten kann, um damit dem Replizieren von evidenzbasierten Programmen in Schulen, Gemeinden etc. eine bessere wissenschaftliche Grundlage zu geben und die Qualität der Umsetzung zu erhöhen (Eccles & Mittman, 2006; Elliott & Mihalic, 2004).

Dadurch, dass dies ein relativ junger Forschungszweig ist, gibt es teilweise keine einheitlichen Begriffsdefinitionen, Modelle und Messstandards (Berkel, Mauricio, Schoenfelder, & Sandler, 2011; Dane & Schneider, 1998; Durlak & DuPre, 2008; Nelson et al., 2012). Außerdem finden sich oftmals Begriffsüberschneidungen, welche einen Diskurs erschweren. Angaben zur Implementierung sind immer noch eher eine Seltenheit in Evaluationsstudien: Dane und Schneider (1998) stellten bei ihrer Meta-Analyse von

Präventionsliteratur aus den Jahren zwischen 1980 und 1994 fest, dass in nur 39 der 162 genauer analysierten Studien Daten zur Implementierung der Interventionen erhoben worden waren (24%). Obwohl die Anzahl der Programme mit Implementierungsevaluation über die Jahre leicht gestiegen ist, waren diese am Ende der 90er Jahre noch immer die Minderheit. Wilson et al. (2003) fanden in ihrer Meta-Analyse von Studien zu Aggressionspräventionsprogrammen ebenfalls kaum Angaben zur Implementierung, obwohl anhand der Trainingseffekte klar erkennbar war, dass die Effektivität der Programme beim Transfer in die Praxis gelitten hatte. Zwei Drittel aller Studien gaben entweder an, dass es keine Probleme bei der Implementierung gegeben hätte oder machten gar keine Angaben zum Thema Implementierung.

Implementierung ist ein multidimensionales Konstrukt (Berkel et al., 2011). Die Aspekte, die in den verschiedenen Modellen von Implementierung beschrieben werden, variieren bezüglich Anzahl und Benennungen (verschiedene Rahmenmodelle bei Berkel et al., 2011; Carroll et al., 2007; Crawford, Carpenter, Wilson, Schmeister, & McDonald, 2012; Dane & Schneider, 1998; Durlak & DuPre, 2008; Dusenbury et al., 2003). Im Folgenden werden acht verschiedene Implementierungsaspekte nach dem Modell von Durlak & DuPre (2008) näher erläutert. Dieses basiert auf einer Meta-Analyse verschiedener Rahmenmodelle und vereint sie zu einem umfassenden Modell (alternative Benennungen aus anderen Rahmenmodellen in Klammern):

Fidelity (Implementierungstreue) beschreibt, ob eine Intervention so umgesetzt wird, wie geplant (adherence, integrity, completeness, compliance).

Quality (of Delivery) (Implementierungsqualität) beschreibt, wie gut die durchführenden Personen das Programm hinsichtlich der Methodik und des Inhalts implementieren können (competence).

Dosage (Implementierungsdosierung) beschreibt, wie viel des ursprünglichen Interventionsplans (zeitlich oder nach Interventionseinheiten) durchgeführt wird (exposure, quantity, intervention strength).

Participant Responsiveness (Aktive Teilnahme der Teilnehmenden) beschreibt, wie gut die Teilnehmenden auf die Umsetzung des Programms ansprechen und wie sehr sie in den Prozess der Implementierung involviert sind.

Adaptation (Implementierungsadaption) beschreibt, welche Veränderungen am ursprünglichen Interventionsplan bei der Umsetzung vorgenommen werden (reinvention, modification).

Program Reach (Implementierungreichweite) beschreibt, wie repräsentativ die Interventionsstichprobe für die angedachte Zielpopulation der Intervention ist.

Monitoring of Control Conditions (Monitoring der Kontrollgruppe) ist notwendig, um zu dokumentieren, welche Interventionen die Kontrollgruppe möglicherweise parallel zur Evaluation erhalten hat, die die Ergebnisse verzerren könnten.

Program Differentiation (Ausdifferenzierung) beschreibt, wie sehr sich die einzelnen Komponenten einer Intervention von den Bedingungen in der Kontrollgruppe oder anderen vergleichbaren Programmen unterscheiden (program uniqueness).

Es gibt keine Belege dafür, dass alle Teilaspekte von Implementierung hundertprozentig kontrolliert sein müssen, um mit einem Programm in der Praxis gute Erfolge erzielen zu können (Dusenbury et al., 2003), jedoch hat sich gezeigt, dass Interventionen, die die verschiedenen Aspekte besser kontrollieren, bis zu zwei- oder dreimal größere Effekte erzielen als weniger sorgfältig implementierte Programme (Durlak & DuPre, 2008; Mihalic & Irwin, 2003). Die verschiedenen Aspekte von Implementierung hängen zusammen, können aber (zu Erhebungszwecken) auch getrennt voneinander betrachtet werden – am meisten verbreitet ist die Erhebung von Implementierungstreue (Durlak & DuPre, 2008).

2. Implementierungstreue

Lange Zeit wurde in der evidenzbasierten Praxis davon ausgegangen, dass Interventionen so implementiert werden, wie sie ursprünglich von den ProgrammentwicklerInnen geplant wurden (Carroll et al., 2007; Dumas, Lynch, Laughlin, Phillips Smith, & Prinz, 2001). Diese Annahme wurde erstmals in den 1970ern hinterfragt, als man begann, sich über die tatsächliche Umsetzung durch unterschiedliche TrainerInnen in unterschiedlichen Organisationen und all die möglichen Einflüsse darauf Gedanken zu machen (Dusenbury et al., 2003). Eine wichtige Rolle bei diesem Umdenken in der Evaluationsforschung spielte der *Rand Report on the Implementation of Educational Innovations* von Berman und McLaughlin (1976; zitiert nach Dusenbury et al., 2003). Diese hatten staatliche Programme aus dem Bildungssektor untersucht und waren zu der Erkenntnis gelangt, dass die tatsächliche Umsetzung der Interventionen stark von den durchführenden Lehrpersonen abhing. Oftmals wurden im Vergleich zum Interventionsplan Kompromisse eingegangen und die Implementierungstreue war in vielen Fällen sehr eingeschränkt. Diese Ergebnisse wurden seither vielfach bestätigt (Durlak, 1998; Dusenbury et al., 2003; Elliott & Mihalic, 2004). LehrerInnen oder TrainerInnen befinden sich oftmals in einem Spannungsfeld zwischen dem Versuch, Interventionen an den Kontext und die Teilnehmenden anzupassen und dem Anspruch, sich dennoch möglichst genau an den zugrundeliegenden Interventionsplan zu halten (Bopp, Saunders, & Lattimore, 2013).

2.1. Definition

Implementierungstreue wird in der Literatur unter anderem definiert als „der Grad, zu dem eine Intervention wie geplant umgesetzt wird“ (Berkel u. a., 2011; Breitenstein u. a., 2010; Dane & Schneider, 1998; Durlak & DuPre, 2008; Lane, Bocian, MacMillan, & Gresham, 2004), als „der Grad, zu dem Programme so implementiert werden, wie es die ProgrammentwicklerInnen angedacht hatten“ (Dusenbury et al., 2003), als „der Grad, zu dem die Komponenten einer Intervention so wie vorgeschrieben und im Unterschied zur Kontrollgruppe vermittelt werden“ (Nelson et al., 2012) oder als der Grad zu dem „die einzelnen Komponenten allen Teilnehmenden in einer vergleichbaren Art und Weise vermittelt werden, welche sich an der Theorie und den Zielen aus der Forschung orientiert“ (Dumas et al., 2001).

Uneinigkeit besteht vor allem in Bezug darauf, wie man diese Definitionen operationalisieren kann und welche Aspekte von Implementierung eine implementierungstreue Umsetzung ausmachen (Nelson et al., 2012). Oftmals wird in der Literatur zwischen Content-Fidelity (Structure-Fidelity) und Process-Fidelity unterschieden (Carroll et al., 2007; Crawford et al., 2012; Dumas et al., 2001; Mowbray et al., 2003; Power et al., 2005). Die Treue bezüglich der Struktur beschreibt hierbei die Vollständigkeit der Vermittlung in Bezug auf zeitliche und inhaltliche Aspekte. Die Treue bezüglich des Prozesses bezieht sich auf die Qualität dieser Vermittlung und somit auf Merkmale des Trainingspersonals und deren Interaktion mit den Teilnehmenden.

Folgende zwei Teilaspekte bilden zusammen die Treue bezüglich der Struktur einer Intervention (Reinke, Herman, Stormont, Newcomer, & David, 2013):

(1) die Vollständigkeit, mit der alle zentralen Inhalte eines Trainings (auch Core Components genannt) gemäß der Theorie und der Vorgaben der ProgrammentwicklerInnen vermittelt werden (Adherence)

(2) wie viel von der geplanten Intervention die Teilnehmenden vermittelt bekommen (Dosage oder Exposure)

Die Treue bezüglich des Prozesses wird einerseits durch die Qualität der Vermittlung definiert und bezieht sich somit auf die Bereitschaft und Fähigkeit des Trainingspersonals, die Methoden und Inhalte theoriekonform und wie geplant anzuwenden und zu vermitteln. Andererseits beschreibt diese Dimension aber auch die aktive Teilnahme, Beteiligung und Aufmerksamkeit der Teilnehmenden und die Interaktion zwischen TrainerInnen und Teilnehmenden (Reinke et al., 2013).

2.2. Implementierungstreue und -adaption – ein Gegensatz?

Zwischen zwei Merkmalen guter Implementierung liegt eine gewisse Spannung vor: Sowohl Implementierungstreue als auch Implementierungsadaption gelten als wichtige Aspekte gelungener Implementierung (Durlak & DuPre, 2008). Dies hat zu einer theoretischen Debatte geführt, in welcher zwei Lager prominent vertreten sind: VertreterInnen der „**Pro-Fidelity**“ Perspektive sehen den Schlüssel für gelungene Interventionen zur Gänze in der strikten Einhaltung methodischer und inhaltlicher Vorgaben. Sie befürchten, dass

Änderungen dazu führen könnten, dass wichtige Komponenten von Interventionen nicht durchgeführt oder theoretische Überlegungen aus der Planungsphase zunichte gemacht werden und das Programm in der Praxis somit an Effektivität einbüßt (Elliott & Mihalic, 2004).

AnhängerInnen der „**Pro-Adaptation**“ Perspektive hingegen sind der Meinung, dass ein blindes Festhalten am Interventionsplan dazu führen kann, dass man nicht auf die lokalen Bedürfnisse eingeht, was somit ebenfalls zu einer Verminderung der Effektivität des Programms führen könnte (Blakely et al., 1987).

Es gibt aber auch die Ansicht, dass **sowohl** Implementierungstreue **als auch** Implementierungsadaption wichtig für effektive Trainings sind (Bopp et al., 2013; Castro, Barrera, & Martinez, 2004; Kumpfer, Alvarado, Smith, & Bellamy, 2002; O’Connor, Small, & Cooney, 2007) – eine Perspektive, welche auch immer wieder durch empirische Evidenz bestärkt wird (Berkel et al., 2011; Blakely et al., 1987; Larsen & Samdal, 2007; Skolits & Richards, 2010). AutorInnen, die diese Ansicht vertreten, sind der Meinung, dass gewisse Modifizierungen bei der konkreten Durchführung notwendig und akzeptabel sind, solange die zentralen Inhalte einer Intervention erhalten bleiben und adäquat umgesetzt werden. So gesehen ist es wichtig, diese kritischen Elemente von Interventionen zu identifizieren und den TrainerInnen klar als solche zu vermitteln bzw. sie in Manualen deutlich zu kennzeichnen (Berkel et al., 2011; Dusenbury et al., 2003). Adaptierbare Teilaspekte von Interventionen können so besser von Schlüsselementen unterschieden werden, welche auf alle Fälle nach Plan durchgeführt werden müssen. Somit kann ein effektives Zusammenspiel zwischen Implementierungstreue und einer Adaption an konkrete Bedingungen funktionieren. Adaptionen am Interventionsplan sollten allerdings stets sorgfältig dokumentiert werden, um sie und ihre Wirkung auf Trainingseffekte untersuchen zu können (Berkel et al., 2011).

2.3. Zusammenhang mit Trainingseffekten

Studien zeigen immer wieder, dass implementierungstreue umgesetzte Interventionen bessere Effekte erzielen (Berkel et al., 2011; Crawford et al., 2012; Durlak & DuPre, 2008; Fagan et al., 2008; Fixsen et al., 2009; Fixsen, Naoom, Blase, Friedman, & Wallace, 2005; Kalafat, Illback, & Sanders, 2007; Mihalic & Irwin, 2003; Wilson & Lipsey, 2007). Oder es

können überhaupt nur bei den Trainingsgruppen Effekte gefunden werden, bei denen die Intervention implementierungstreu umgesetzt wurde (Durlak, 1998). Beispielsweise konnten Salmivalli et al. (2005) bei ihrer Studie zu einem Anti-Bullying Programm in Schulen zeigen, dass Lehrkräfte, die das Programm mit hoher Implementierungstreue umsetzten, bei ihren SchülerInnen wesentlich größere Veränderungen in den Einstellungen bezüglich Bullying erzielen konnten als KollegInnen, die wenig implementierungstreu arbeiteten. Signifikante Veränderungen im relevanten Verhalten konnten (mit wenigen Ausnahmen) überhaupt nur in Klassen mit hoher Implementierungstreue erzielt werden (Salmivalli et al., 2005). Burke et al. (2011) zeigten in ihrer Evaluation eines Schulprogramms zur Verbesserung des SchülerInnenverhaltens im Klassenzimmer, dass die implementierungstreue Umsetzung durch das Lehrpersonal ein signifikanter Prädiktor für mehr Mitarbeit im Unterricht war ($\beta = .75$, $p = .000$) und 57% der Varianz im Trainingseffekt erklären konnte. In einer Studie von Lillehoj et al. (2004) konnten ExpertInnenbeurteilungen der Implementierungstreue von schulbasierten Präventionstrainings die Einstellungsveränderungen von SchülerInnen gegenüber Alkohol und Marihuana signifikant vorhersagen.

Zu geringe Implementierungstreue kann dafür verantwortlich sein, dass Interventionen in der Praxis nicht dieselben Effekte erzielen, wie unter stärker kontrollierten Forschungsbedingungen (Breitenstein, Robbins, & Muennich Cowell, 2012). Eine implementierungstreue Umsetzung ist in der Praxis jedoch durch verschiedene Umstände oft nicht gewährleistet: limitierte Ressourcen, unzureichende Vorbereitung und wenig Commitment von Seiten der Institution, Zeitmangel, Probleme mit dem Curriculum und Personalwechsel können die Implementierungstreue negativ beeinflussen (Dusenbury et al., 2003; Elliott & Mihalic, 2004). Gut geplante und evidenzbasierte Programme können dadurch beim Transfer in die Praxis an Effektivität einbüßen oder in ihrer Wirkung sogar verzerrt werden (Crawford et al., 2012). Deshalb sind die Effekte, welche in der breiten Implementierung durch verschiedenes Trainingspersonal erzielt werden, zumeist niedriger als in einer ersten Projektphase, in der Interventionen unter kontrollierten Bedingungen, mit mehr Ressourcen und zumeist noch von den ProgrammentwicklerInnen selbst durchgeführt werden (Wilson & Lipsey, 2007). Wilson et al. (2007) konnten in ihrer Meta-Analyse von Studien zu schulbasierten Projekten zeigen, dass in 87% der Studien nur Daten aus der Pilotphase von Programmen veröffentlicht werden, was Annahmen bezüglich der

tatsächlichen Effektivität bei einer breit angelegten Implementierung im Feld erheblich erschwert (Wilson & Lipsey, 2007).

Die Erfassung von Daten zur Implementierungstreue eines Programms ist demnach aus verschiedenen Gründen essenziell. Zunächst wird durch solche Daten die Interpretierbarkeit von Trainingseffekten erhöht (Dane & Schneider, 1998; Dumas et al., 2001): Ohne Daten zur Implementierungstreue ist es nicht möglich zu sagen, wie genau signifikante Trainingseffekte zustande gekommen sind, beziehungsweise ob schlechte Ergebnisse auf das Programmkonzept oder auf die unzulängliche Umsetzung in der Praxis zurückzuführen sind (interne Validität). Nur mit Angaben zur Implementierungstreue der Umsetzung können klare Aussagen über die Entstehung der Trainingseffekte gemacht werden. Beurteilungen der Implementierungstreue können Trainingseffekte vorhersagen und fungieren als Moderatorvariablen, welche Varianzen in den Outcomes erklären können (Mowbray et al., 2003). Somit besteht durch die Missachtung der Implementierungsebene bei der Evaluation die Gefahr, dass Interventionen, welche aufgrund von unzureichender Implementierungstreue in der Praxis keine zufriedenstellenden Effekte erzielen, fälschlicherweise als ineffektiv beurteilt und verworfen werden (Type III Error) (Durlak, 1998; Dusenbury et al., 2003). Fixsen et al. (2009) schreiben dazu (S. 533): „Ineffective programs can be implemented well and effective programs can be implemented poorly. Neither option is desirable. Desirable outcomes are achieved only when effective programs are implemented well.“ Nur Interventionen, welche mit hoher Implementierungstreue umgesetzt werden und dennoch schlechte Effekte erzielen, müssen in ihrer Konzeption neu überdacht werden (Dusenbury et al., 2003). Auch die Beurteilung der Wirksamkeit und Übertragbarkeit auf andere Kontexte (externe Validität) ist von Daten zur Implementierungstreue abhängig (Dumas et al., 2001). Fehlen entsprechende Daten, so kann es sein, dass effektive Trainings nicht angewandt bzw. Interventionen angewandt werden, welche in der Praxis aufgrund von Ressourcenmangel oder bestimmter Kontextfaktoren nicht adäquat repliziert werden können und somit bei einer breiten Implementierung an Effektivität einbüßen (Dumas et al., 2001).

Somit sind die Validität und Interpretierbarkeit von Outcome-Evaluationen ohne Implementierungsevaluation stark eingeschränkt (Durlak, 1998; Lane et al., 2004). Darüber hinaus können Daten zur Implementierungstreue auch die statistische Power von Analysen

erhöhen (Dumas et al., 2001). Werden die Ergebnisse von Erhebungen zur Implementierungstreue laufend rückgemeldet, so kann die Implementierung noch während der Umsetzung angepasst und verbessert werden, was wiederum die Effektivität von Interventionen mit sofortiger Wirkung erhöht (Fagan et al., 2008). Daher ist die Erhebung von Daten zur Implementierungstreue essenziell für eine aussagekräftige Evaluation von Interventionen.

2.4. Förderung von Implementierungstreue in der Praxis

Die vorhandenen Befunde zur implementierungstreuen Umsetzung von Interventionen im Feld stimmen eher pessimistisch. Oft werden bei der Durchführung willkürliche Änderungen am Programm vorgenommen oder Teile der Intervention werden gar gänzlich gestrichen (Dusenbury et al., 2003). Ein Grund dafür scheint sicherlich Ressourcenknappheit zu sein. Denn obwohl man zumeist hohe Erwartungen an Interventionen und ihre Wirkungen stellt, sehen sich TrainerInnen in der Praxis zu oft mit limitierten Ressourcen konfrontiert, die das Arbeiten erschweren (Esbensen et al., 2011).

Es hat sich gezeigt, dass vor allem die Komplexität des Interventionsplans, die benötigten Ressourcen (Zeit, Materialien und Personal), die wahrgenommene Effektivität des Programms und die Motivation der LehrerInnen (TrainerInnen) dafür verantwortlich sind, ob eine Intervention in der Praxis so umgesetzt wird, wie es geplant war (Lane et al., 2004). In der Literatur findet man Überlegungen und Studien dazu, wie man Implementierungstreue in der Praxis fördern kann: Da die Komplexität von Programmen die implementierungstreue Umsetzung stark einschränken kann, ist es wichtig, Trainingsmanuale möglichst übersichtlich und verständlich zu gestalten. Die zentralen Komponenten der Intervention, welche auf jeden Fall durchgeführt werden müssen (Core Components), sollten explizit dargelegt werden (Dusenbury et al., 2003). Diese klare Aufschlüsselung von essenziellen und weniger wichtigen Teilen eines Programms hilft auch bei der Erhebung von Daten zur Implementierungstreue (siehe Kapitel 2.5. *Erfassung von Implementierungstreue*). Eine möglichst spezifische Beschreibung der Kontexte, in denen die Intervention nach Meinung der ProgrammentwicklerInnen angewendet werden kann, ist ebenfalls nützlich (Bickman et al., 2009). Wünschenswert sind außerdem Umsetzungsbeispiele in den Manualen, die genauer erläutern, wie man erlernte Methoden und Konzepte konkret in Instruktionen und

Übungen einbauen kann und wie man am besten auf mögliche Hürden und Probleme reagiert (Dusenbury et al., 2003). Genaue Anleitungen für Aktivitäten, die als relevant erachtet werden und die einfach durchzuführen sind, können die Umsetzung für das Trainingspersonal erheblich erleichtern (Mihalic & Irwin, 2003).

Essenziell scheint vor allem auch ein gutes, vorbereitendes Training für das Interventionspersonal zu sein: Eine umfangreiche Einführung mit Follow-Up Treffen resultiert zumeist in besseren Werten in Bezug auf die implementierungstreue Umsetzung (Dusenbury et al., 2003). Es hat sich außerdem gezeigt, dass Lehrpersonen (bzw. TrainerInnen), welche sich sicher und qualifiziert fühlen, eine Intervention gut umsetzen zu können und die von der Wirksamkeit des Programms überzeugt sind, implementierungstreuer arbeiten (Dusenbury et al., 2003; Larsen & Samdal, 2007). Trainings sollten daher einerseits das Wissen und die nötigen Fertigkeiten zur Umsetzung vermitteln, aber andererseits das Personal auch motivieren und es von der Nützlichkeit des Programms überzeugen (Larsen & Samdal, 2007). Interventionspersonal, welches ein solch gründliches Training erhält, ist besser vorbereitet und eher dazu bereit, die Intervention gänzlich und adäquat umzusetzen (Mihalic & Irwin, 2003). Zusätzlich zu einem gut konzipierten Training hat sich begleitende Supervision durch die ProgrammentwicklerInnen (telefonisch oder mittels persönlicher Treffen) als förderlich für Implementierungstreue erwiesen (Fagan et al., 2008). Darüber hinaus kann die Nutzung formativer Evaluationsmethoden die Qualität der Implementierung und somit die Effektivität von Interventionen erheblich verbessern. Mittels Beobachtung und Feedback durch BeraterInnen kann die Implementierungstreue noch während der Umsetzung optimiert werden (Lane et al., 2004; Smith, Daunic, & Taylor, 2007).

2.5. Erfassung von Implementierungstreue

„In most cases [...] authors [...] only say that implementation was effectively achieved without supplying any data.“ (Durlak & DuPre 2008, S. 344)

Wie bereits dargelegt, sind Datenangaben zur Implementierungstreue noch immer eher eine Seltenheit in Evaluationsstudien (siehe Kapitel 1. *Implementierungsforschung: Ein Überblick*). In Anbetracht der Bedeutung von Implementierungstreue für die Effektivität von Trainingsprogrammen stellt dies ein enormes Defizit in Bezug auf die Validität und

Interpretierbarkeit von summativen Outcome-Evaluationen dar. Es besteht eine große Varianz an Erhebungsmöglichkeiten im Bereich der Implementierungsevaluation und es gibt keine verbindlichen Standards für die Erhebung (Nelson et al., 2012). Die meisten Studien stützen sich nur auf eine Erhebungsmethode oder -perspektive, ohne die Vorgangsweise genauer zu begründen (Durlak & DuPre, 2008; Esbensen et al., 2011). Darüber hinaus findet man in den wenigsten Studien überhaupt genauere Angaben zu den Erhebungs- und Auswertungsmethoden (Lillehoj, Griffin, & Spoth, 2004). Es soll hier deshalb nun ein Überblick über gängige Vorgangsweisen bei der Erhebung von Implementierungstreue sowie wünschenswerte Entwicklungen in diesem Bereich gegeben werden.

Mowbray et al. (2003) haben eine ideale Vorgangsweise bei der Erfassung von Implementierungstreue vorgeschlagen:

- (1) Identifikation und Operationalisierung der Core Components
- (2) Datenerhebung
- (3) Validierung

2.5.1. Identifikation und Operationalisierung der Core Components

An erster Stelle im Prozess steht die Identifikation und Operationalisierung der zentralen Inhalte einer Intervention. Um diese erkennen zu können, muss man das Wirkmodell hinter der Intervention kennen und verstehen (Kalafat et al., 2007; Nelson et al., 2012). Ein solches Wirkmodell beschreibt die Interventionskomponenten (Übungen, Methoden etc.) und die durch die Umsetzung erhofften Veränderungen bei den Teilnehmenden. Es verdeutlicht die angenommenen kausalen Wirkmechanismen eines Programms und bildet die Grundlage von theoriegeleiteten Programmevaluationen (Nelson et al., 2012). Idealerweise wird das Wirkmodell von den ProgrammentwicklerInnen im Manual beschrieben (Bickman et al., 2009). Wenn dies nicht der Fall ist, kann man ExpertInnen zur Wirkung der Intervention befragen oder Meinungen von PraktikerInnen einholen und diese qualitativ analysieren. Eine weitere Möglichkeit besteht darin, Pilotstudien durchzuführen, um zu ermitteln, welche Komponenten unbedingt durchgeführt werden müssen, um die gewünschten Effekte erzielen zu können (Blase & Fixsen, 2013; Mowbray et al., 2003).

Wenn man die zentralen Inhalte einer Intervention ermittelt hat, werden Indikatoren und Operationalisierungen für die Beurteilung entwickelt. Zu diesem Zweck werden die Core Components in kleinere Aspekte aufgeschlüsselt, die die Handlungen und Ressourcen beschreiben, welche für die Durchführung einer bestimmten Komponente notwendig sind. Für diese differenzierteren Beschreibungen von Subkomponenten, welche bereits konkrete Handlungsbeschreibungen beinhalten, können in weiterer Folge Indikatoren zur Beurteilung entwickelt werden (Nelson et al., 2012). Wünschenswert ist hierbei auch die Festlegung von Ankerwerten, welche eine besonders gute oder besonders niedrige Implementierungstreue in Bezug auf die verschiedenen Aspekte signalisieren (Mowbray et al., 2003). Zumeist gibt es keine solchen Anhaltspunkte, ab wann ein bestimmter Wert akzeptable oder gute Implementierungstreue bedeutet (Smith et al., 2007), weshalb die Interpretation von Ergebnissen oftmals schwer fällt.

Je nachdem welchen Aspekt von Implementierungstreue man beurteilen möchte, sehen die Operationalisierungen unterschiedlich aus (Esbensen et al., 2011). Gängige Indikatoren sind:

- (1) gemäß der Theorie und der Vorgaben der ProgrammentwicklerInnen adäquat durchgeführte Core Components (nach Interventionseinheiten aufgeschlüsselt) (Adherence)
- (2) aufgewendete Zeit pro Interventionskomponente (nach Interventionseinheiten aufgeschlüsselt oder global) (Dosage)
- (3) globale Beurteilung der Involviertheit der Teilnehmenden
- (4) globale Beurteilung der Qualität der Vermittlung

Die Erhebung von Implementierungstreue bezüglich des Prozesses ist wesentlich aufwändiger und beruht zumeist auf sehr subjektiven Maßen. Deshalb sind solche Daten oftmals weit weniger reliabel. Daten bezüglich der Struktur und des Inhalts (Dosage und Adherence) lassen sich im Vergleich dazu wesentlich einfacher operationalisieren und standardisierter erheben (Mowbray et al., 2003). Die gängigste Form, in der Implementierungstreue in Studien angegeben wird, ist eine Prozentangabe aller in der Einheit vollständig und adäquat vermittelten Interventionskomponenten (Adherence) (Smith et al., 2007). Wenn zum Beispiel sieben der neun geplanten Übungen entsprechend in der Einheit durchgeführt wurden, so würde die Prozentangabe für die Einheit 78% betragen. Daraus lässt sich ein Gesamtprozentwert für die Adherence der Intervention berechnen.

2.5.2. Datenerhebung

Im nächsten Schritt werden auf Basis der Operationalisierungen Daten zur Implementierungstreue gesammelt. Je nachdem welche BeurteilerInnengruppe herangezogen wird, müssen zu diesem Zweck unterschiedliche Verfahren verwendet werden (Lane et al., 2004; Lillehoj et al., 2004; Nelson et al., 2012).

Die beiden gängigsten Erhebungsformen sind (Durlak & DuPre, 2008):

(1) **ExpertInnenratings** (in Form von Interviews, direkten Beobachtungen, Beurteilungen von Videoaufnahmen, Programmdokumentationen etc.)

(2) **Fragebogen oder Interviews von TrainerInnen** (Selbstberichte)

Beide Perspektiven bieten Vor- aber auch Nachteile: TrainerInnen geben zumeist an, dass sie die Intervention so wie geplant durchgeführt haben. So sind Werte aus Selbstberichten oft durch sozial erwünschte Antworten verzerrt und liegen somit meist höher als beispielsweise die Einschätzungen von objektiven BeobachterInnen (Lane et al., 2004; Lillehoj et al., 2004). Auf der anderen Seite können gerade Selbstberichte aber auch einen guten Einblick in die tatsächliche Umsetzung und die Beweggründe für etwaige Adaptionen bieten (Dusenbury et al., 2003; Lillehoj et al., 2004).

Direkte Beobachtungen durch ExpertInnen können zwar zumeist am besten Trainingseffekte vorhersagen und gelten als vergleichsweise objektiv und valide (Durlak & DuPre, 2008; Dusenbury et al., 2003), sind in der Praxis aufgrund des enorm hohen Aufwands aber oftmals nicht durchführbar (Lane et al., 2004). Darüber hinaus besteht die Möglichkeit, dass durch die Anwesenheit von BeobachterInnen das Verhalten der Durchführenden aber auch der Teilnehmenden verzerrt wird (Nelson et al., 2012). Als Alternative bieten sich daher beispielsweise ExpertInnenratings auf Grundlage von Portfolios und Dokumentationen des Implementierungsverlaufs oder von Produkten der Intervention an. Dies hat den Vorteil, dass die Umsetzung nicht gestört wird und die Datenerhebung nicht direkt in den Prozess der Implementierung eingreift (Lane et al., 2004; Salmivalli et al., 2005). Ein Nachteil besteht allerdings darin, dass die tatsächliche Qualität der Umsetzung dadurch nicht erfasst werden kann. Es kann lediglich beurteilt werden, ob die relevanten Komponenten durchgeführt wurden (Smith et al., 2007). So ist mittels solch indirekter Maße nur eine Beurteilung der Structure- bzw. Content-Fidelity möglich (siehe Kapitel 2.1. *Definition*).

Die Befragung von Teilnehmenden ist bisher nur sehr wenig verbreitet (Mowbray et al., 2003). Dies mag unter anderem daran liegen, dass Teilnehmende dazu neigen, in Projektevaluationen ihren persönlichen Befindlichkeiten gegenüber dem Trainingspersonal und ihrer generellen Ablehnung oder Euphorie gegenüber dem Projekt Ausdruck zu verleihen. Oftmals kann man daher nicht zur Gänze nachvollziehen, wie die Beurteilungen zustande gekommen sind und welche Faktoren diese möglicherweise zusätzlich beeinflusst haben (Nelson et al., 2012). Die Perspektive der Teilnehmenden kann jedoch ein umfassendes und unvoreingenommenes Bild der Umsetzung vermitteln, da diese genau wie die TrainerInnen über die gesamte Intervention hinweg anwesend sind, aber keine Informationen bezüglich des Interventionsmodells haben. Einige wenige Studien haben Teilnehmende befragt und dabei vor allem Fragebogen bzw. Checklists zur Beurteilung der Implementierungstreue verwendet (Mowbray et al., 2003).

Generell wird sehr selten mehr als eine Erhebungsperspektive zur Erfassung von Implementierungstreue herangezogen (Durlak & DuPre, 2008). Noch seltener sind Vergleiche zwischen verschiedenen Beurteilungsperspektiven (Durlak & DuPre, 2008; Cho, Hallfors, & Sánchez, 2005). In den wenigen Studien, in denen verschiedene Perspektiven auf die Implementierungstreue eines Trainings verglichen wurden, konnten zumeist signifikante und hohe Korrelationen zwischen den verschiedenen Beurteilungen gefunden werden (Lillehoj et al., 2004). Zumeist werden hierbei objektive ExpertInnenratings mit Selbstberichten von TrainerInnen verglichen – die wenigsten dieser Studien beziehen sich jedoch auf Projekte im Feld. Somit fehlen praxisnahe Angaben von TrainerInnen, die tatsächlich längerfristig in Schulen oder Gemeinden Projekte implementieren (z.B. Lehrpersonen) (Lillehoj et al., 2004).

Da sowohl die Verwendung direkter als auch indirekter Methoden (Beobachtungen auf der einen Seite; Selbstberichte, Ratings von Dokumentationen und Produkten auf der anderen Seite) Grenzen aufweisen, bietet es sich an, diesen Limitationen durch die ergänzende Verwendung verschiedener Beurteilungsperspektiven und -methoden entgegenzuwirken, um dadurch die Informationsbreite zu erweitern (Lane et al., 2004; Lillehoj et al., 2004). Es ist somit bei der Erhebung von Implementierungsdaten ein multi-method und multi-informant Ansatz erstrebenswert (Bickman et al., 2009; Esbensen et al., 2011; Lillehoj et al., 2004; Mowbray et al., 2003). Ein daraus resultierender kombinierter Index für

Implementierungstreue kann Verzerrungen, die die verschiedenen Erhebungsmethoden mit sich bringen, verringern (Nelson et al., 2012). Eine solche Vorgangsweise ist bisher jedoch noch sehr selten (Esbensen et al., 2011). Da Projektressourcen es oftmals nicht ermöglichen, Implementierungstreue ergänzend aus mehreren verschiedenen Perspektiven zu erheben, ist es notwendig, mehr über unterschiedliche Perspektiven und ihre Vor- und Nachteile zu erfahren. Relevante Informationen für die passende Auswahl von Beurteilungsperspektiven und Methoden zur Erhebung von Implementierungstreue eines Programms können durch Vergleiche zwischen verschiedenen Erhebungszugängen gewonnen werden.

2.5.3. Validierung und Ebene der Analyse

Angaben zur Reliabilität von verwendeten Skalen und die Validierung der festgelegten Indikatoren und Messinstrumente zur Erhebung von Implementierungstreue sind essenziell, um die Nützlichkeit der Angaben für Forschung und Praxis zu garantieren (Mowbray et al., 2003). Zunächst müssen verwendete Skalen auf ihre Reliabilität überprüft werden. Wenn Ratingskalen verwendet werden, ist außerdem auch die Angabe von Interrater-Übereinstimmungen obligat (Mowbray et al., 2003).

Wenn das einer Intervention zugrundeliegende Wirkmodell tatsächlich die gewünschten Trainingseffekte erzielen kann, seine zentralen Komponenten gut operationalisiert wurden und die Erfassung von Implementierungstreue mittels reliabler und valider Messinstrumente erfolgt ist, so können die ermittelten Implementierungsdaten Trainingseffekte vorhersagen (Mowbray et al., 2003). Solche positiven Zusammenhänge zwischen Indikatoren von Implementierungstreue und Trainingseffekten können als Validierung der Messung von Implementierungstreue aber auch des Wirkmodells der Intervention gesehen werden (Mowbray et al., 2003). Daher ist es wünschenswert, dass Beurteilungen der implementierungstreuen Umsetzung einer Intervention mit Trainingsoutcomes in Verbindung gebracht werden (Mowbray et al., 2003; Nelson et al., 2012). Bisher ist eine solche Vorgangsweise leider noch kaum verbreitet. Durlak und DuPre (2008) fanden in ihrer Meta-Analyse von beinahe 500 Studien heraus, dass nur 59 den Zusammenhang zwischen Implementierungstreue und Trainingseffekten ermittelt hatten.

Bei der Analyse von Zusammenhängen zwischen Daten zur Implementierungstreue, welche zumeist auf Projektebene vorliegen, und Trainingseffekten auf Individualebene stößt man

auf Probleme bei der statistischen Auswertung, da die Daten auf verschiedenen Ebenen vorliegen (Mowbray et al., 2003). Eine Möglichkeit, damit umzugehen, ist das Aggregieren der Outcome-Daten auf Projektebene, wodurch jedoch eine Menge Informationen auf Individualebene verloren gehen. Ein Rechnen auf Individualebene wiederum ignoriert die hierarchische Struktur der Daten gänzlich und lässt außer Acht, dass sich Individuen, welche gemeinsam in einer Interventionsgruppe (Klasse) sind, zumeist mehr ähneln (Mowbray et al., 2003). Eine gute statistische Möglichkeit, mit dieser methodischen Herausforderung umzugehen, ist das Rechnen eine Mehrebenenanalyse (Mowbray et al., 2003). Nähere Erläuterungen finden sich im Kapitel 5.4.1. *Mehrebenenanalyse*.

3. Das Schulprojekt *Reflect*

Das Projekt *Reflect – Genderkompetenz durch Reflexive Koedukation* der Universität Wien war ein Train-the-Trainer Programm zur nachhaltigen Verankerung des Kompetenzerwerbes zur reflexiven Koedukation in der LehrerInnenbildung (Finsterwald et al., 2012). Im Gegensatz zum monoedukativen Unterricht beschreibt reflexive Koedukation eine Form des Unterrichts, in der Mädchen und Burschen gemeinsam unterrichtet werden und bei der keine Einschränkungen durch Geschlechtsstereotype entstehen sollen. Die SchülerInnen sollen sich gemeinsam in einer Klasse ihrer individuellen Potenziale bewusst werden (Finsterwald et al., 2012). In einem von wissenschaftlichen MitarbeiterInnen des Arbeitsbereichs Bildungspsychologie und Evaluation durchgeführten LehrerInnentrainings wurden die teilnehmenden Lehrkräfte zum Thema reflexive Koedukation inhaltlich und methodisch geschult. In einer anschließenden Supervisionsphase führten die Lehrkräfte supervidiert durch das Universitätspersonal in ihren Klassen Projekte zum Thema Berufswahl und Geschlecht durch.

3.1. Theoretische Grundlagen

Reflect baut auf das für die Erklärung und Veränderung von geschlechtsbezogenen Leistungsunterschieden etablierte Modell des Aktiotops (Ziegler, Heller, Schober, & Dresel, 2006) auf. Alle Inhalte und Methoden wurden in Bezug auf das Aktiotopmodell geplant und im LehrerInnentraining vermittelt. Die Klassenprojekte wurden von den Lehrpersonen gemeinsam mit den ExpertInnen der Universität entwickelt und mit diesen laufend abgestimmt, um eine theoriekonforme Umsetzung zu garantieren. Das LehrerInnentraining umfasste folgende Inhalte (siehe Tabelle 1):

Tabelle 1: Inhaltliche Module des LehrerInnentrainings von *Reflect*

Modul 1	Einführung in die Thematik <i>Reflexive Koedukation</i>
Modul 2	Hintergrund zur Entstehung und Möglichkeiten zur Verminderung von Geschlechtsunterschieden im Unterricht
Modul 3	Unterrichtsentwicklung I - Gendersensibilität durch Förderung von Selbstwert und Interesse
Modul 4	Unterrichtsentwicklung II - Gendersensibilität durch adäquates Feedback

Neben der Fokussierung genderthematischer Inhalte wurde besonders in den Modulen 3 und 4 auch Wert darauf gelegt, den Lehrpersonen Kompetenzen zur motivations- und selbstwertförderlichen Unterrichtsgestaltung zu vermitteln – eine Grundvoraussetzung für gendersensiblen Unterricht (Finsterwald et al., 2012). Im Zuge dieses Methodenschwerpunkts war die Selbstbestimmungstheorie von Deci und Ryan (2002) ein zentrales Thema.

3.1.1. Die Selbstbestimmungstheorie und Förderung von Motivation

Zumeist wird in der Wissenschaft ein Gegensatz zwischen intrinsischer und extrinsischer Motivation propagiert. In der Selbstbestimmungstheorie von Deci & Ryan (2002) stellen sie hingegen Phasen auf einem „Selbstbestimmungskontinuum“ dar (siehe Abbildung 2). Je mehr wir uns mit einer Aufgabe identifizieren und diese in unsere Werthaltungen integrieren, desto selbstbestimmter und -regulierter arbeiten wir.

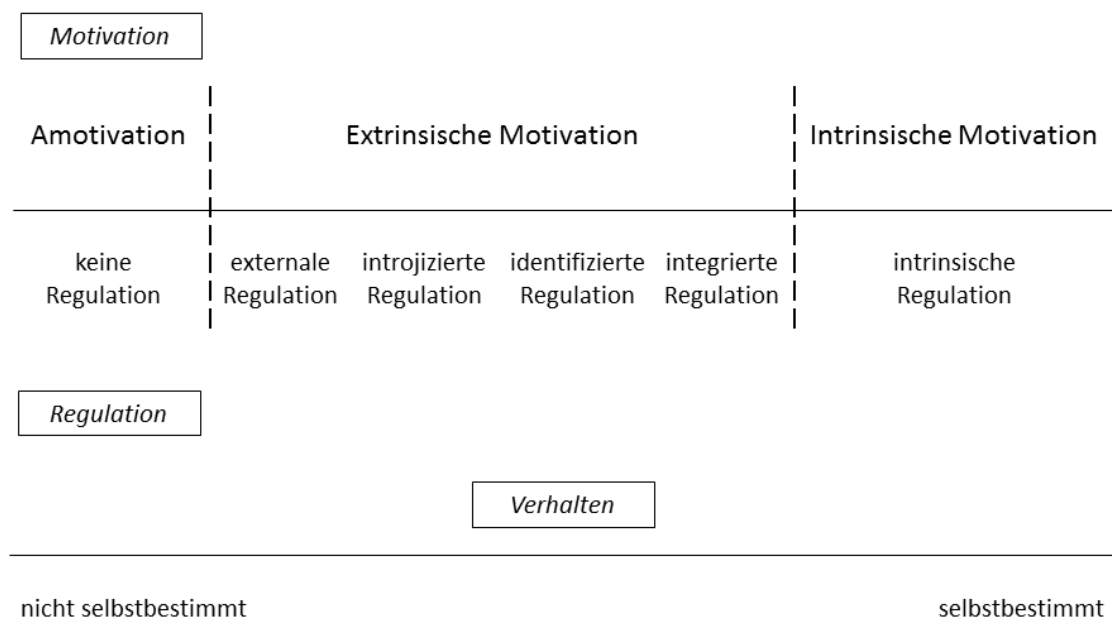


Abbildung 2: Selbstbestimmungskontinuum mit den verschiedenen Formen der Motivation und Regulation nach Deci & Ryan (2002, S. 16)

Ganz links steht die **Amotivation** - ein Zustand, in dem wir entweder gar nicht handeln oder nur sehr passiv. Ein solcher Motivationsstatus resultiert zumeist aus dem Empfinden, die geforderten Ziele nicht erreichen zu können oder aus einer Geringschätzung der vorgegebenen Ziele und Aufgaben.

Die Phase der **extrinsischen Motivation** wird in der Selbstbestimmungstheorie in verschiedene kleinere Phasen unterteilt, welche sich durch verschiedene Regulationsformen unterscheiden:

Externale Regulation bildet die am stärksten von außen gesteuerte Form der extrinsischen Motivation und steht im Kontinuum gleich nach der Phase der Amotivation. Wenn wir external reguliert handeln, so tun wir dies, um eine Belohnung zu erhalten oder um einer Bestrafung zu entgehen. Diese Form der extrinsischen Motivation steht im stärksten Kontrast zur intrinsischen Motivation. Die Ursachenzuschreibung für die eigenen Handlungen erfolgt external (außerhalb der eigenen Person).

Introjierte Regulation beschreibt eine Form der extrinsischen Motivation, bei der die externale Regulation schon bis zu einem gewissen Grad verinnerlicht wurde. Allerdings stellt auch sie noch immer eine sehr regulierte Form der Motivation dar, da man Handlungen vor allem setzt, um Gefühle von Scham oder Schuld zu vermeiden, sowie um das eigene Selbstwertgefühl zu verbessern.

Regulation aufgrund von Identifikation impliziert eine bewusste Wertschätzung bestimmten Verhaltens. Wenn man sich mit dem eigenen Verhalten identifiziert, nimmt man eine gewisse Autonomie des eigenen Handelns wahr und die Ursachenzuschreibung erfolgt bereits ziemlich internal (der eigenen Person zugehörig). Diese Wertschätzung ist aber noch nicht gänzlich in die eigenen Werte und Vorstellungen integriert, was in manchen Situationen zu Konflikten innerhalb der Person führen kann.

Integrierte Regulation ist die autonomste Form der extrinsischen Motivation. Die Identifikation mit einem bestimmten Verhalten ist mit den eigenen Werten und Vorstellungen im Einklang. Doch obwohl ein solches Verhalten bereits volitional gesetzt wird, beschreibt diese Form der Regulation noch einen Teil der extrinsischen Motivation, da man die Handlungen zum Erreichen der gewünschten Ergebnisse setzt und nicht wegen der Freude oder dem Interesse an der Handlung selbst.

Ganz rechts steht die **intrinsische Motivation**, welche dafür verantwortlich ist, dass wir Handlungen aus eigenem Interesse und aufgrund der in der Aufgabe selbst gefundenen Befriedigung ausführen. Dieser Zustand bildet sozusagen den Prototyp autonomen und selbstbestimmten Handelns.

Die Selbstbestimmungstheorie fußt auf dem Grundgedanken, dass alle Menschen sowohl nach persönlichem Wachstum als auch nach sozialer Anerkennung und Integration streben (Krapp, 2005). Soziale Kontexte können diese Bestrebungen entweder fördern oder aber einschränken (Deci & Ryan, 2002). Im Bildungsbereich sehen sich PädagogInnen und TrainerInnen oftmals damit konfrontiert, Verhalten fördern zu wollen, welches nicht an sich intrinsisch motiviert ist. Die Qualität der Motivation trägt jedoch wesentlich dazu bei, dass SchülerInnen gute Leistungen erbringen, gerne in die Schule gehen, und ihrem Leistungsniveau entsprechende Herausforderungen suchen (Reeve, Deci, & Ryan, 2002). Die Befriedigung dreier grundlegender psychischer Bedürfnisse kann hier zu einem Fortschreiten auf dem Motivations- und Selbstbestimmungskontinuum beitragen und zu einem höheren Grad an Selbstregulation bei den SchülerInnen führen. Diese grundlegenden psychischen Bedürfnisse sind die Bedürfnisse nach Autonomie, Kompetenzerleben und sozialer Eingebundenheit (Deci & Ryan, 2002):

Autonomie

Wir wollen uns als ursächlich und verantwortlich für unser eigenes Verhalten wahrnehmen. Eines unserer grundlegenden psychischen Bedürfnisse ist es, aus unserem eigenen Interesse und unseren eigenen Werthaltungen heraus zu handeln.

Kompetenzerleben

Das Bedürfnis nach Kompetenzerleben lässt uns immer wieder nach Herausforderungen suchen, anhand derer wir unsere Leistungsfähigkeit unter Beweis stellen können. Aus diesem Grund suchen wir vor allem nach Aufgaben, die unseren Fähigkeiten optimal entsprechen, um uns als effektiv erleben zu können.

soziale Eingebundenheit

Das dritte grundlegende psychische Bedürfnis ist das Bedürfnis nach der Verbundenheit mit anderen, nach einem Gefühl der Zugehörigkeit, der Sorge um andere und dem Bedürfnis, auch selbst umsorgt zu sein.

Die Befriedigung dieser grundlegenden psychischen Bedürfnisse wirkt sich auf den Grad der Motivation aus, mit dem wir bestimmte Aufgaben erledigen. Als schädlich für die Motivation haben sich Androhungen von Strafen, Deadlines, auferlegte Ziele, Kontrolle, permanenter Vergleich mit anderen und negatives Feedback herausgestellt, da diese Maßnahmen als kontrollierend wahrgenommen werden und die Bedürfnisse nach Autonomie- und Kompetenzerleben einschränken (Deci & Ryan, 2002). Förderlich für die Motivation hingegen sind das Vorgeben interessanter Aufgaben, Wahlmöglichkeiten in Bezug darauf, ob, wie und wann gearbeitet wird, sowie eine Ausrichtung des Feedbacks und der Benotung auf die Erledigung der Aufgaben und nicht auf die Person an sich (Reeve et al., 2002), da dies Autonomie- und Kompetenzerleben fördert und Druck aus dem Lernprozess nimmt. Auch soziale Eingebundenheit spielt eine wichtige Rolle bei der Internalisierung von Regulation, also dem Fortschreiten auf dem Selbstbestimmungskontinuum. Wenn wir von Personen, die wichtig für uns sind, Aufgaben gestellt bekommen oder wenn wir durch sie Interesse oder Wertschätzung für bestimmte Handlungen vermittelt bekommen, so werden wir dieses Verhalten eher zeigen, um implizite oder explizite Anerkennung zu erlangen (Deci & Ryan, 2002). Die Befriedigung dieser Bedürfnisse begünstigt die Entwicklung von Motivation und selbstbestimmtem Verhalten. Wenn sich SchülerInnen kompetent fühlen, eine bestimmte Aufgabe erfüllen zu können und sie sich dadurch Anerkennung durch ihnen wichtige Personen versprechen, so kommt es zu externaler Regulation. Durch die Förderung von sozialer Eingebundenheit und Kompetenzerleben wird introjizierte Regulation erzielt und diese durch die Förderung von Autonomieerleben immer weiter in die eigene Person integriert (Deci & Ryan, 2002) – die Voraussetzung für selbstbestimmtes und motiviertes Handeln.

3.2. Umsetzung

Reflect wurde von Oktober 2010 bis August 2012 durchgeführt. 38 Lehrkräfte aus der Sekundarstufe wurden von ExpertInnen der Universität Wien in einer Intensivphase zum Thema reflexive Koedukation geschult. In der Supervisionsphase (Jänner - April 2012) führten anschließend 30 Lehrkräfte supervidiert durch das Universitätspersonal selbst Klassenprojekte durch (Finsterwald et al., 2012). Im Modulblock 3 des LehrerInnentrainings wurde die Bedeutung von Autonomie- und Kompetenzerleben sowie sozialer Eingebundenheit für die Lernmotivation und das Interesse der SchülerInnen thematisiert. Die Teilnehmenden bekamen eine theoretische Einführung zum Thema Selbstbestimmungstheorie (Deci & Ryan, 2002) und es wurden konkrete Umsetzungsmöglichkeiten besprochen, welche sodann von den Lehrkräften in den Klassenprojekten umgesetzt werden sollten.

3.2.1. Vermittelte Umsetzungsmöglichkeiten der Selbstbestimmungstheorie

Als Beispiele zur **Förderung von Autonomieerleben** wurden Rückmeldungen und Wahlmöglichkeiten bei der Bearbeitung von Aufgaben besprochen. Den Lehrkräften wurde vermittelt, dass Rückmeldungen nicht kontrollierend sein sollten und dass es wichtig ist, dass sie Hilfe zur Selbsthilfe bieten. Wahlmöglichkeiten bei der Bearbeitung von Aufgaben können die Reihenfolge, den Inhalt, die Sozialform aber auch das Bearbeitungstempo betreffen. Die SchülerInnen sollten die Möglichkeit haben, selbstständig den eigenen Lernprozess zu verwalten.

Zur **Förderung von Kompetenzerleben** sollten die Aufgaben einen optimalen, an die SchülerInnen individuell angepassten, Schwierigkeitsgrad aufweisen. Dies kann beispielsweise erzielt werden, indem man zum selben Thema mehrere Aufgaben vorgibt, die je nach Schwierigkeit gekennzeichnet sind. Die SchülerInnen können selbst entscheiden, welcher Herausforderung sie sich gewachsen fühlen. Rückmeldungen sollten stets individuell erfolgen und ausschließlich sachorientiertes Feedback über den eigenen Lernfortschritt beinhalten. Den SchülerInnen soll so signalisiert werden, dass man von Fehlern profitieren und aus ihnen lernen kann. Die Lehrkräfte sollten sich stets bemühen, den SchülerInnen Zutrauen in ihre Fähigkeiten zu signalisieren.

Um die **soziale Eingebundenheit** zu fördern, ist es vor allem wichtig, auf einen wertschätzenden und respektvollen Umgang untereinander zu achten und prosoziales Verhalten in der Klasse zu fördern. Die Lehrkräfte sollten sich dabei stets bemühen, authentisch zu bleiben. Außerdem ist es wichtig, die SchülerInnen gut zu beobachten und ihnen stets aktiv zuzuhören. Die LehrerInnen wurden dazu angehalten, den SchülerInnen im unterstützenden Dialog zu helfen.

3.2.2. Die Klassenprojekte und Dokumentation der Umsetzung

Die Klassenprojekte aus der Supervisionsphase sollten sich thematisch mit dem Thema Berufswahl und Gender befassen und dabei folgende Themen beinhalten (siehe Tabelle 2):

Tabelle 2: zentrale Themen der Klassenprojekte von *Reflect*

Zentrale Themen

Aufklärung: keine Unterschiede in den Begabungen zwischen Mädchen und Buben

Reflexion über Unterschiede in den Berufswahlen

Reflexion über Gleichberechtigung von Männern und Frauen

Eine weitere Vorgabe war die Umsetzung der Selbstbestimmungstheorie in den Klassenprojekten zur Förderung der Motivation und des Interesses der SchülerInnen an der Thematik. Diese Vorgabe bildete eine zentrale Core Component der Klassenprojekte von *Reflect* (mehr zum Thema Core Components siehe Kapitel 2.1. *Definition* und 2.5.1. *Identifikation und Operationalisierung der Core Components*).

Die Umsetzung der Klassenprojekte wurde von den Lehrpersonen in Portfolios dokumentiert. Sie wurden dazu angehalten, einen detaillierten Projektplan inklusive Planung der Einheiten abzugeben, verwendete Materialien und Produkte der SchülerInnen (beispielhaft) beizulegen und die Durchführung in einem Logbuch zu dokumentieren. Ein zentraler Bestandteil des Portfolios war auch die abschließende, schriftliche Reflexion über die Umsetzung der zentralen Inhalte und Core Components sowie mögliche Hürden dabei.

4. Forschungsanliegen

Es hat sich wiederholt gezeigt, dass Implementierungstreue ein wichtiger Faktor für die erfolgreiche Umsetzung evidenzbasierter Interventionen in der Praxis ist. Dies manifestiert sich in besseren Trainingseffekten von Projekten mit guter Implementierungstreue im Vergleich zu Interventionen mit geringer oder gar keiner Implementierungstreue (Berkel et al., 2011; Durlak & DuPre, 2008; Fixsen et al., 2005). Daten zur Implementierungstreue sind essenziell, um differenzierte Aussagen über Trainingseffekte machen zu können und erklären zu können, wie diese entstanden sind. Ohne solche Daten ist es kaum möglich zu beurteilen, ob Trainingseffekte tatsächlich auf die Wirksamkeit der Intervention zurückzuführen sind, oder ob andere Variablen bei der Umsetzung konfundierend beteiligt waren (Dumas et al., 2001; Mowbray et al., 2003). Daten zur Implementierungstreue können die interne und externe Validität von Outcome-Evaluationen erhöhen (Dusenbury et al., 2003; Lane et al., 2004) und können bereits während der Implementierungsphase wichtige Informationen zur Verbesserung der Umsetzung liefern (Fagan et al., 2008).

Die Erhebung von Daten zur Implementierungstreue ist in der Evaluationspraxis dennoch noch immer eher eine Seltenheit (Dane & Schneider, 1998; Wilson et al., 2003). Bisher gibt es außerdem keine verbindlichen Standards in Bezug auf die Herangehensweise an die Erhebung (Dusenbury et al., 2003) und es finden sich viele unterschiedliche Erhebungsmöglichkeiten, welche allesamt teils erhebliche Limitationen aufweisen. In den meisten Studien wird Implementierungstreue nur aus der Perspektive einer Personengruppe (zumeist ExpertInnen oder TrainerInnen) beurteilt (Esbensen et al., 2011), obwohl es naheliegend erscheint, dass ExpertInnen, TrainerInnen und Teilnehmende die Umsetzung von Projekten unterschiedlich wahrnehmen. Ein zentrales Anliegen dieser Studie ist die Einbeziehung mehrerer Perspektiven in die Beurteilung der Implementierungstreue von *Reflect*. Um Informationen über verschiedene Beurteilungsperspektiven zu gewinnen, werden folgende Personen befragt: LehrerInnen, SchülerInnen und ExpertInnen.

(1) Wie beurteilen SchülerInnen, LehrerInnen und ExpertInnen die implementierungstreue Umsetzung der Klassenprojekte?

Die implementierungstreue Umsetzung von Projekten kann auf zwei Ebenen betrachtet werden: auf der Inhaltsebene oder auf der Ebene des Prozesses (siehe Kapitel 2.1.

Definition). Da die Umsetzung der Core Components zur Verinnerlichung der Inhalte besonders wichtig ist (siehe Kapitel 2.5.1. *Identifikation und Operationalisierung der Core Components*), wird in dieser Studie die theoriekonforme Umsetzung in Bezug auf eine Core Component gemessen. Dafür wurde die Umsetzung gemäß der Selbstbestimmungstheorie (Deci & Ryan, 2002) ausgewählt. Die implementierungstreue Umsetzung in Bezug auf *Kompetenzerleben, Autonomie* und *soziale Eingebundenheit* wird aus drei verschiedenen Perspektiven beurteilt: von SchülerInnen, LehrerInnen und ExpertInnen (zur Selbstbestimmungstheorie siehe Kapitel 3.1.1. *Die Selbstbestimmungstheorie und Förderung von Motivation* und Kapitel 3.2.1. *Vermittelte Umsetzungsmöglichkeiten der Selbstbestimmungstheorie*).

Die Beurteilungen aus SchülerInnen-, LehrerInnen- und ExpertInnenperspektive bieten verschiedene Vorteile: Die LehrerInnenbeurteilung kann einen guten Einblick in die tatsächliche Umsetzung und die Beweggründe für etwaige Adaptionen bieten (Dusenbury et al., 2003; Lillehoj et al., 2004). Beurteilungen durch ExpertInnen gelten als vergleichsweise objektiv und valide (Durlak & DuPre, 2008; Dusenbury et al., 2003). Eine Befragung von Teilnehmenden ist bisher noch kaum verbreitet (Mowbray et al., 2003) und stellt einen innovativen Zugang zur Erhebung von Implementierungstreue dar.

Die gängigsten Erhebungsformen bei der Beurteilung von Implementierungstreue sind Fragebogen für TrainerInnen, Interviews mit TrainerInnen oder ExpertInnen und direkte Beobachtungen oder Beurteilung von Videoaufnahmen, Programmdokumentationen etc. durch ExpertInnen (Durlak & DuPre, 2008). Fragebogen und Interviews mit TrainerInnen bieten einen guten Einblick in die tatsächliche Umsetzung von Interventionen, sind aber anfällig für Verzerrungen aufgrund von sozial erwünschter Antworten (Lillehoj et al., 2004). Direkte Beobachtungen durch ExpertInnen gelten als vergleichsweise objektiv (Durlak & DuPre, 2008) – sie sind in der Praxis aufgrund des enormen Aufwands aber kaum einsetzbar (Lane et al., 2004). Als Alternative bietet sich daher die Beurteilung mittels Dokumentationen der Implementierung an (z.B. Aufzeichnungen, Portfolios etc.) (Lillehoj et al., 2004). Ein weiterer Vorteil dieser Vorgangsweise ist, dass dadurch nicht direkt in den Prozess der Implementierung eingegriffen wird, was ebenfalls zu Verzerrungen führen könnte (Nelson et al., 2012).

(2) Unterscheiden sich die verschiedenen Beurteilungen der Implementierungstreue voneinander?

Es gibt kaum Studien, die verschiedene Beurteilungsperspektiven miteinander vergleichen (Durlak & DuPre, 2008; Dusenbury et al., 2003). In den wenigen Studien, in denen bisher mehrere Beurteilungsperspektiven verglichen wurden, wurden signifikante und hohe Korrelationen zwischen Selbstberichten und ExpertInnenratings gefunden, obwohl Selbstberichte zumeist besser ausfielen als ExpertInnenratings (Lillehoj et al., 2004). Diese Studien beziehen sich meistens auf Pilotprojekte und es fehlen Daten von TrainerInnen, die tatsächlich in Schulen oder Gemeinden längerfristig Projekte implementieren (z.B. Lehrpersonen) (Lillehoj et al., 2004). Außerdem gibt es kaum Studien, welche auch die Teilnehmendenperspektive in die Analyse miteinbeziehen (Mowbray et al., 2003), obwohl diese ebenfalls ein sehr umfassendes und darüber hinaus unvoreingenommenes Bild der Umsetzung liefern können. Diese Studie soll daher auch Informationen über diesen noch wenig genutzten Zugang liefern. Ziel des Vergleichs ist es, mehr über die Charakteristika verschiedener BeurteilerInnengruppen sowie über die Unterschiede zwischen den Perspektiven zu erfahren. Solche Informationen können in zukünftigen Implementierungsevaluationen bei der Auswahl von BeurteilerInnengruppen und bei der Interpretation von Daten aus den verschiedenen Perspektiven behilflich sein. In Anbetracht bisheriger Studien wird davon ausgegangen, dass die LehrerInnenbeurteilungen aufgrund von sozial erwünschten Antworten tendenziell höher liegen werden als die ExpertInnenbeurteilungen. Darüber hinaus wird vermutet, dass die verschiedenen Beurteilungen hoch miteinander korrelieren (Lillehoj et al., 2004).

(3) Welche Beurteilung kann am besten die *Veränderung des Umgangs mit Unterschiedlichkeit* vorhersagen?

Wenn das einer Intervention zugrundeliegende Wirkmodell tatsächlich die gewünschten Trainingseffekte erzielen kann, seine zentralen Komponenten gut operationalisiert wurden und die Erfassung von Implementierungstreue mittels reliabler und valider Messinstrumente erfolgt ist, so können die ermittelten Implementierungsdaten Trainingseffekte vorhersagen (Mowbray et al., 2003). Um dies zu überprüfen, müssen Zusammenhänge zwischen Daten zur Implementierungstreue und Outcome-Daten berechnet werden (Durlak & DuPre, 2008). Ein solches In-Verbindung-Bringen von Daten zur Implementierungstreue mit Outcome-

Daten hat zwei verschiedene Bedeutungen: Auf der einen Seite lässt sich so das Zustandekommen von Trainingseffekten besser erklären, auf der anderen Seite kann man damit die Validität von verschiedenen Beurteilungen der Implementierungstreue überprüfen (Mowbray et al., 2003). Doch obwohl eine solche Analyse essenziell ist, gibt es kaum Studien, die Angaben zum Zusammenhang mit Trainingseffekten machen (Durlak & DuPre, 2008).

Im Fall von *Reflect* war es ein zentrales Ziel, den Umgang mit Unterschiedlichkeit der SchülerInnen zu verbessern. Deshalb wird in dieser Studie die Vorhersagekraft der verschiedenen Beurteilungen für den Trainingseffekt *Veränderung des Umgangs mit Unterschiedlichkeit* berechnet. Die aus der Analyse gewonnenen Informationen können im Zusammenhang mit den Ergebnissen der zweiten Fragestellung Auskunft über die Güte der verschiedenen Erhebungszugänge liefern, was bei der Planung künftiger Implementierungsevaluationen nützlich sein kann. Bisher haben sich im Gegensatz zu Selbstberichten der TrainerInnen zumeist nur direkte Beobachtungen von ExpertInnen als signifikante Prädiktoren von Trainingseffekten erwiesen (Durlak & DuPre, 2008; Lillehoj et al., 2004). Zur Vorhersagekraft von Teilnehmendenbeurteilungen ist im Bereich von schulbasierten Projekten noch nichts bekannt.

5. Methode

5.1. Datenerhebung

Die teilnehmenden Lehrpersonen bearbeiteten online Evaluationsfragebogen zu drei Messzeitpunkten: vor Beginn des Lehrkräfte-Trainings zur Erhebung der Baseline (Messzeitpunkt 1 = MZP1), am Ende des Trainings (MZP2) und nach der Supervisionsphase (MZP3). Auch die SchülerInnen, die an den Klassenprojekten teilnahmen, wurden zu Beginn des Programms (MZP1) und nach Beendigung der Klassenprojekte (MZP3) mittels einer Online-Erhebung befragt (eine Übersicht zum Evaluationsdesign siehe Abbildung 3).

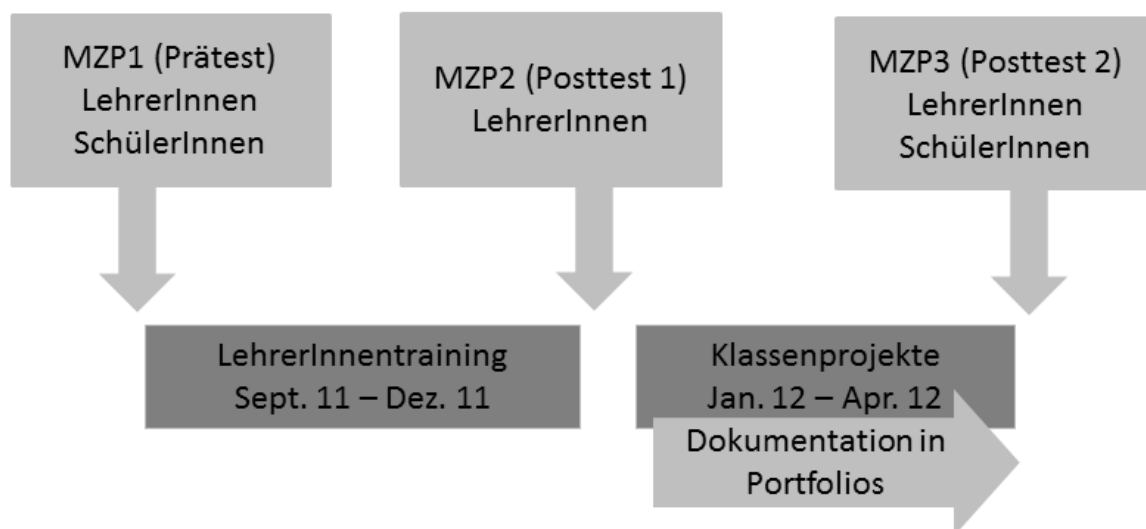


Abbildung 3: Evaluationsdesign der begleitenden Evaluation von *Reflect*

Die Daten zur *Veränderung des Umgangs mit Unterschiedlichkeit* bei den SchülerInnen stammen aus dem Vergleich der SchülerInnendaten von MZP1 und MZP3. Die implementierungstreue Umsetzung der Klassenprojekte wurde im Zuge der begleitenden Evaluation von den LehrerInnen und SchülerInnen beim letzten Messzeitpunkt (MZP3) beurteilt. Die Beurteilungen auf ExpertInnenebene wurden anhand der von den LehrerInnen erstellten Portfolios nach Beendigung von *Reflect* vorgenommen. Drei RaterInnen beurteilten die Portfolios zunächst unabhängig voneinander. Im zweiten Schritt wurde ein Konsens erzielt.

Alle drei Beurteilungsgruppen beurteilten die Implementierungstreue in Bezug auf drei Variablen: *Kompetenzerleben*, *Autonomie* und *soziale Eingebundenheit*.

5.2. Stichprobe

Die Stichprobe bestand aus 30 Lehrpersonen (24 Lehrerinnen und 6 Lehrer) und 569 SchülerInnen (durchschnittlich 19,62 SchülerInnen pro Klasse, m = 288, w = 258, keine Angabe = 23; durchschnittliches Alter = 12,75 Jahre). Zehn Lehrkräfte führten die Klassenprojekte in Teams durch (zwei Dreier- und zwei Zweiergruppen). Für diese Team-Projekte wurde jeweils nur ein Portfolio gemeinsam erstellt. Drei wissenschaftliche MitarbeiterInnen des Arbeitsbereichs Bildungspsychologie und Evaluation der Universität Wien beurteilten die insgesamt 24 abgegebenen Portfolios.

5.3. Instrumente

Zur Beurteilung der *Veränderung des Umgangs mit Unterschiedlichkeit* wurde den SchülerInnen an beiden Messzeitpunkten eine Skala von Finsterwald et al. (Finsterwald, Schober, Wagner, & Spiel, 2010) zum Umgang mit Unterschiedlichkeit in ihrer Klasse vorgegeben. Die Skala misst den subjektiv wahrgenommenen Umgang mit Diversität in der Klasse allgemein und genderbezogen (siehe Items Tabelle 3). Die Veränderung in Bezug auf den Trainingseffekt wurde mittels der Differenz zwischen MZP3 und MZP1 berechnet.

Tabelle 3: Items der Skala *Umgang mit Unterschiedlichkeit*

Einleitung

In unserer Klasse...

Items

...werden Schüler/Schülerinnen ausgeschlossen, die anders sind.

...wird jede/jeder so angenommen, wie er/sie ist.

...wird jemand leicht ausgeschlossen, wenn er/sie eine andere Meinung hat als die anderen.

...ist es egal, welche Muttersprache jemand hat.

...mögen wir auch Mädchen, die sich nicht wie „typische Mädchen“ verhalten.

...mögen wir auch Buben, die sich nicht wie „typische Buben“ verhalten.

...ist es eine besondere Stärke, dass wir alle verschiedene Begabungen und Fähigkeiten haben.

Antwortformat

Schieberegler; 1 (stimme gar nicht zu) - 101 (stimme voll zu); markierte Mitte bei 51

Die implementierungstreue Umsetzung der Klassenprojekte wurde bei den SchülerInnen mittels Items aus dem Erhebungsinstrument „Unterrichtsqualität und mathematisches Verständnis“ (Lipowsky, Rakoczy, Buff, & Klieme, 2005; Rakoczy, Klieme, & Pauli, 2008) zur wahrgenommenen *Autonomie*, *Kompetenz* und *sozialen Eingebundenheit* abgefragt, welche sprachlich leicht angepasst wurden, um den Bezug zum Projekt herzustellen (Beispielitems siehe Tabelle 4).

Tabelle 4: Beispiele für SchülerInnenitems zur Beurteilung von Autonomie¹, Kompetenzerleben² und sozialer Eingebundenheit³

Einleitung

Im Projekt...

Items

...konnte ich selber entscheiden, wie ich arbeiten will.¹

...konnte ich wählen, wann ich welche Aufgabe bearbeite.¹

...fanden meine Leistungen Anerkennung.²

...sagte mir der Lehrer/die Lehrerin, was ich noch verbessern könnte.²

...hatte ich das Gefühl, dass mir die anderen in der Klasse helfen würden, wenn es nötig wäre.³

...fühlte ich mich von den anderen in der Klasse verstanden.³

Antwortformat

Schieberegler; 1 (stimme gar nicht zu) - 101 (stimme voll zu); markierte Mitte bei 51

Die LehrerInnen beurteilten die implementierungstreue Umsetzung der Klassenprojekte anhand verschiedener Skalen: Ob die SchülerInnen in ihrem Klassenprojekt autonom arbeiten konnten, wurde mittels Items von Röder und Kleine (2007; zitiert nach Jerusalem et al., 2009) erfragt. Als Äquivalent zur Förderung von *Kompetenzerleben* wurden sie nach der *Bezugsnorm* für die Beurteilung gefragt (Fensterwald et al., 2010). Förderung von Kompetenzerleben beinhaltet unter anderem eine Ausrichtung von Feedback und Beurteilung auf informationelle Aspekte und das Eindämmen von Wettbewerb zwischen den SchülerInnen (Deci & Ryan, 2002). Somit kann eine soziale Bezugsnorm als schädigend für das Kompetenzerleben gewertet werden, wohingegen eine sachliche und individuelle Beurteilung förderlich für das Kompetenzerleben sind. Wie sehr sie es generell schaffen, die SchülerInnen sozial einzubinden, beurteilten die LehrerInnen (nicht direkt in Bezug auf das Projekt) mittels der Skala *Fürsorglichkeit* (Jerusalem et al., 2009). Diese Skala misst das wertschätzende und fürsorgliche Verhalten von Lehrpersonen in Bezug auf persönliche Interessen und Probleme ihrer SchülerInnen und stellt somit einen wichtigen Aspekt der

Förderung *sozialer Eingebundenheit* aus LehrerInnenperspektive dar (Deci & Ryan, 2002)
(Beispielitems siehe Tabelle 5).

Tabelle 5: Beispiele für LehrerInnenitems zur Beurteilung der Autonomie¹, der Bezugsnorm² und der Fürsorglichkeit³

Einleitung

Im Projekt der Reflect-Supervisionsphase...

Items

...konnten die SchülerInnen zwischen unterschiedlich schwierigen Aufgaben wählen.¹

...konnten die Schülerinnen mitentscheiden, wann und wie lange sie sich mit einer bestimmten Aufgabe beschäftigen.¹

Einleitung

Ich habe den Erfolg/Misserfolg eines/einer SchülerIn im Projekt daran gemessen, ob...

Items

...die Leistung des Schülers/der Schülerin besser/schlechter ist als die der anderen SchülerInnen.²

...die Leistung des Schülers/der Schülerin besser/schlechter ist als die vorherige Leistung desselben Schülers.²

...der Schüler/die Schülerin ein vorher festgesetztes Ziel erreicht hat oder nicht.²

Items

Die persönlichen Beziehungen zu meinen SchülerInnen haben Vorrang vor schnellem Vorankommen im Stoff.³

Für persönliche und soziale Angelegenheiten nehme ich mir auch im Unterricht Zeit.³

Antwortformat

Schieberegler; 1 (stimme gar nicht zu) - 101 (stimme voll zu); markierte Mitte bei 51

Die ExpertInnen beurteilten die implementierungstreue Umsetzung in Bezug auf *Autonomie*, *Kompetenzerleben* und *sozialer Eingebundenheit* mittels eines an der Universität entwickelten Ratingbogens (Schultes et al., in Vorbereitung). Dieser orientierte sich an der Checkliste, anhand derer die Lehrkräfte die Klassenprojekte umsetzen sollten. Durch die

Vergabe von 0 bis 3 Punkten (0 – nicht vorhanden, 1 – angesprochen, 2 – weitgehend vorhanden, 3 – vollständig vorhanden) wurden die Portfolios nach folgenden Kriterien beurteilt (siehe Tabelle 6):

Tabelle 6: Beurteilungsgrundlage für das ExpertInnenrating zu den Portfolios

Wird durch die Konzeption des Projekts die Autonomie der SchülerInnen gefördert?

z.B.: Den SchülerInnen stehen Wahlangebote und Gestaltungsfreiräume zur Verfügung (Themen, Aufgaben, Arbeitsformen, Präsentationsformen, Bearbeitungszeit etc.);

Wird durch die Konzeption des Projekts das Kompetenzerleben der SchülerInnen gefördert?

z.B.: Die SchülerInnen haben Möglichkeiten sich selbstwirksam zu erleben; die LehrerInnen geben konstruktives Feedback auf SchülerInnenarbeiten und ermöglichen Verbesserungen; individuelle Lernfortschritte von SchülerInnen werden honoriert;

Wird durch die Gestaltung der Einheiten die soziale Eingebundenheit bzw. soziale Zugehörigkeit der SchülerInnen gefördert?

z.B.: Es werden gemeinschaftliche Aktivitäten initiiert, die über den herkömmlichen Unterricht hinausgehen. Auf persönliche Anliegen der SchülerInnen wird eingegangen. Kooperative Lernformen werden eingesetzt. Reflexions- und Feedbackmöglichkeiten werden angeboten. Die individuellen Fähigkeiten von SchülerInnen werden anerkannt.

Zur besseren Nachvollziehbarkeit der ExpertInnenbeurteilung werden im Folgenden beispielhaft Auszüge aus den Portfolios dargestellt, welche mit einer hohen Punktezahl beurteilt wurden.

Autonomie

Beschreibung einer Einheit: *„Wichtig ist mir, verschiedene Begabungen entsprechend zu belohnen, indem die SchülerInnen aus einer Vielfalt von Aufgaben auswählen können und so selbst bestimmen, was sie tun und wie sie die Arbeit präsentieren.“*

Beschreibung einer Übung: *„Selbständige Auswahl, Auseinandersetzung und Ausarbeitung eines Bereichs zum Themenbereich „Beruf und Gender im Wandel von Vergangenheit und Gegenwart“ (im Projekt hieß das Thema „(Zeit-)Reise durch die Berufswelt“) aus einem vorgegebenen Themenkatalog. Die Schülerinnen und Schüler hatten aber auch die Möglichkeit andere als die von der Lehrkraft vorgeschlagenen Themen einzubringen – sofern sie inhaltlich kompatibel waren.“*

Reflexion: *„Bei der Reflexion gaben die SchülerInnen mehrheitlich positive Rückmeldungen zum Projekt: freie Wahl der PartnerInnen, Arbeit im Team nach Interesse, [...].“*

Kompetenzerleben

Reflexion: *„Da der Unterricht in einer heterogenen Klasse stattfand, wurden den SchülerInnen Differenzierungsmöglichkeiten angeboten, die in der Checklist mit einem Stern und einem schwarzen Kreis gekennzeichnet sind. Die Aufgabenstellung von Übungen mit dem Kreis war eher leicht, die von Übungen mit dem Stern eher anspruchsvoll. Die SchülerInnen hatten je nach Leistungsstand die Möglichkeit sich für eine der beiden Varianten individuell zu entscheiden.“*

Beschreibung einer Übung: *„Ich wähle die jeweiligen Arbeiten aus der Checklist aus, gebe sie zur Korrektur ab. Diese Erstfassung wird korrigiert und mit Vorschlägen und Kommentaren versehen. Dann überarbeite ich sie und bringe sie in Reinform.“*

Beschreibung einer Übung: *„Schreibe deinen Namen untereinander in farbigen Druckbuchstaben auf. Notiere nun neben jedem Buchstaben eine deiner Eigenschaften /Fähigkeiten, die mit diesem Buchstaben beginnen.“*

Notiz zu einer Einheit: *„Das Augenmerk liegt auf dem individuellen Lernzuwachs – Einzelgespräche mit den SchülerInnen während ihrer Arbeit (Was ist dir wichtig, was hast du dazugelernt, siehst du das Thema jetzt anders?).“*

soziale Eingebundenheit

Reflexion: *„Die SchülerInnen hatten während des gesamten Projekts die Möglichkeit mich auch außerhalb der Unterrichtszeit per E-Mail zu kontaktieren und mir ihre Arbeiten zu schicken. Dies ermöglichte häufige, individuelle und konstruktive Rückmeldungen über ihr Tun und den Stand ihrer Fähigkeiten. Sie fühlten sich dadurch ernst genommen, unterstützt und begleitet. Ich konnte den SchülerInnen dadurch mein Interesse und die Bedeutung des Themas „Gender“ deutlich machen und sie vielleicht ein bisschen anstecken. Sie waren stolz, mit mir am Nachmittag kommunizieren zu können und haben diese Möglichkeit wirklich genutzt.“*

Beschreibung einer Übung: *„Die SchülerInnen erhalten ein Blatt mit den Feldern „Mein Name ist/ Meine Klasse ist...“, „Das kann ich gut...“, „Das interessiert mich...“, „Meine liebsten Fächer...“, „Meine Hobbies...“. Wenn sie fertig sind, sollen sie sich mit anderen MitschülerInnen besprechen. Seid ihr voneinander überrascht?“*

Notiz zu einer Einheit: *„Nun ging es weiter mit der Erforschung der eigenen Stärken. Hier war es für mich schön zu beobachten, welche Wertschätzung die SchülerInnen durch die anderen MitschülerInnen erfuhren. Viele konnten es gar nicht glauben, welche Stärken man als „Außenstehende/r“ in ihnen entdeckte.“*

5.4. Statistische Auswertung

Zur Beantwortung der Forschungsfrage **(1) Wie beurteilen SchülerInnen, LehrerInnen und ExpertInnen die implementierungstreue Umsetzung der Klassenprojekte?** wurden die deskriptiven Daten näher betrachtet.

Zur Beantwortung der Forschungsfrage **(2) Unterscheiden sich die verschiedenen Beurteilungen der Implementierungstreue voneinander?** wurden eine Korrelationsanalyse und ein T-Test durchgeführt.

Zur Beantwortung der Forschungsfrage **(3) Welche Beurteilung kann am besten die Veränderung des Umgangs mit Unterschiedlichkeit vorhersagen?** wurde eine Mehrebenenanalyse durchgeführt. Da diese Analyse sehr komplex ist und ein Verständnis

für die Vorgangsweise bei der Auswertung die Nachvollziehbarkeit der Ergebnisse beeinflusst, wird die Mehrebenenanalyse im Folgenden näher erläutert.

Alle Analysen wurden mit SPSS 20 durchgeführt.

5.4.1. Mehrebenenanalyse

Daten, welche bei der Evaluation von Schulprojekten erhoben werden, sind beinahe immer auf zumindest zwei Ebenen verschachtelt: die SchülerInnen gehören zu verschiedenen Klassen, die die Inhalte von unterschiedlichen Lehrpersonen vermittelt bekommen haben. Daten zum Projekt liegen somit zumeist auf Klassenebene vor, während sich die Outcomes auf SchülerInnenebene befinden (Lüdtke & Köller, 2010). Abbildung 4 soll diese Datenstruktur veranschaulichen.

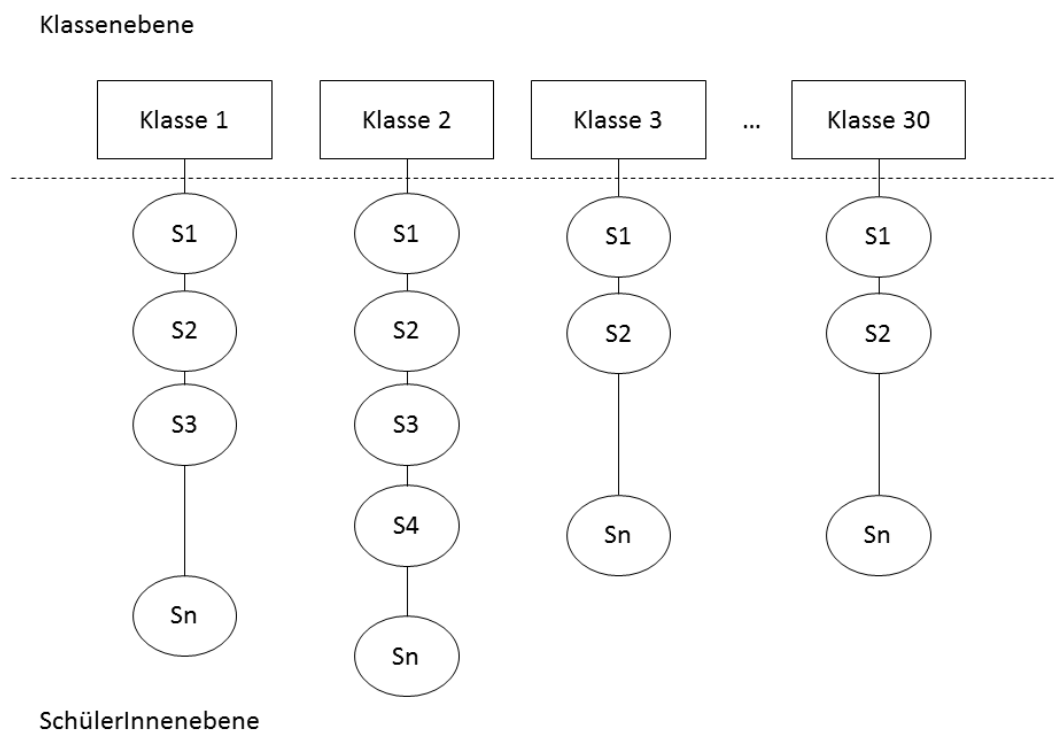


Abbildung 4: Veranschaulichung der hierarchischen Datenstruktur in den Klassenprojekten

Wie bereits an anderer Stelle beschrieben (siehe Kapitel 2.5.3. *Validierung und Ebene der Analyse*) gibt es mehrere Möglichkeiten, um mit einer solchen Datenstruktur umzugehen.

Während jedoch durch das Aggregieren der Daten auf die Klassenebene oder das simple Rechnen auf Individualebene wesentliche Informationen verloren gehen, bietet eine Mehrebenenanalyse die Möglichkeit, Daten auf verschiedenen Ebenen zuzulassen und somit alle vorhandenen Informationen miteinzubeziehen (Field, 2009). Die Gefahr, welche beim Ignorieren von hierarchischen Datenstrukturen entsteht, ist, dass man außer Acht lässt, dass ein Teil der Varianz in der abhängigen Variable durch Kontextvariablen (z.B. Klassenzugehörigkeit) entstanden sein könnte (Field, 2009). Dies lässt sich mittels der Berechnung einer Intraklassenkorrelation (ICC) überprüfen. Dieses Maß gibt an, wie viel der Varianz des Outcomes auf Variablen der Klassenebene zurückzuführen ist. Ab einem ICC von 0,05 (entspricht 5% der Gesamtvarianz) sollte man eine Mehrebenenanalyse rechnen (Heck, Thomas, & Tabata, 2010).

Die verschiedenen Beurteilungen der implementierungstreuen Umsetzung wurden in dieser Studie mit dem Trainingseffekt *Veränderung des Umgangs mit Unterschiedlichkeit* in Verbindung gebracht. Die Berechnung eines Nullmodells (ein Modell ohne Prädiktoren) für diesen Trainingsoutcome ergab eine Intraklassenkorrelation (ICC) von 0,05. Dies bedeutet, dass 5% der Varianz des Trainingseffekts durch Variablen der Klassenebene erklärt werden kann und somit eine Mehrebenenanalyse zu rechnen ist. Deshalb wurde ein Random-Intercept-Modell gerechnet, bei dem der Mittelwert der abhängigen Variable in den verschiedenen Klassen unterschiedliche Werte annehmen kann. Der Unterschied zu einer linearen Regression besteht darin, dass verschiedene mittlere Ausprägungen der Veränderung in den Klassen einkalkuliert werden können (Field, 2009; Heck et al., 2010). Das hypothetisierte Modell zum Zusammenhang zwischen den verschiedenen Beurteilungen der implementierungstreuen Umsetzung der Klassenprojekte bezüglich der Selbstbestimmungstheorie und dem Trainingsoutcome *Veränderung des Umgangs mit Unterschiedlichkeit* sieht demnach wie folgt aus (siehe Abbildung 5).

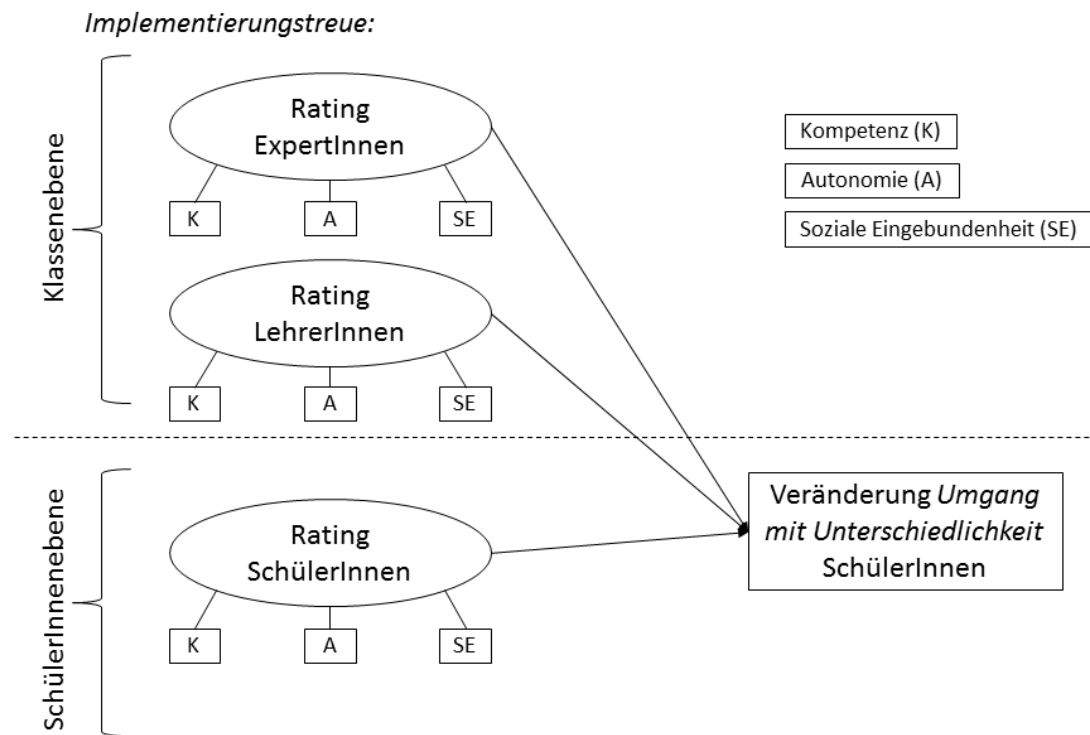


Abbildung 5: hypothetisches Modell des Zusammenhangs zwischen den verschiedenen Beurteilungen der implementierungstreuen Umsetzung bezüglich Selbstbestimmungstheorie und dem Trainingsoutcome *Veränderung des Umgangs mit Unterschiedlichkeit*

Die SchülerInnenbeurteilungen der implementierungstreuen Umsetzung der Klassenprojekte sowie die Daten zum Trainingseffekt liegen auf der Individualebene. Die LehrerInnen- und ExpertInnenbeurteilungen der implementierungstreuen Umsetzung der Klassenprojekte befinden sich auf der Klassenebene. Das 2-Level Random-Intercept-Model wurde mittels SPSS 20 berechnet. Die daraus resultierenden signifikanten β -Werte wurden zur besseren Vergleichbarkeit nach Hox (2010) mit folgender Formel manuell standardisiert:

$$\frac{(\beta_{\text{unstand.}} \times SD_{UV})}{SD_{AV}}$$

$\beta_{\text{unstand.}}$ = unstandardisierter Regressionskoeffizient, SD_{UV} = Standardabweichung der unabhängigen Variable, SD_{AV} = Standardabweichung der abhängigen Variable

Um sicherstellen zu können, dass das Modell die vorliegenden Zusammenhänge tatsächlich adäquat abbilden kann, gibt es einige Parameter, die Auskunft über die Güte des berechneten Modells geben können. Die Parameter beruhen allesamt auf der Berechnung der -2Log-Likelihood – ein Maß dafür, wie viel Information durch ein bestimmtes Modell unerklärt bleibt (Field, 2009). Das bedeutet: je kleiner der angegebene Wert, desto besser das Modell. Ein weiterer Parameter für die Güte ist Akaike's information criterion (AIC). Dieses bezieht die Komplexität des Modells (Anzahl der Prädiktoren) mit ein und ist eine korrigierte Form der Log-Likelihood. Der Parameter ist per se nicht interpretierbar, sondern nur im Verhältnis: beim Vergleich zweier Modelle ist jenes mit dem kleineren AIC das bessere Modell (Field, 2009). Um Modelle vergleichen zu können, bietet sich außerdem die Berechnung des χ^2 -Wertes der Veränderung zwischen den beiden Modellen und dessen Überprüfung auf Signifikanz an. Folgende Formel wird dafür verwendet (Field, 2009):

$$\chi^2_{Change} = (-2\text{Log} - \text{Likelihood}_{Old}) - (-2\text{Log} - \text{Likelihood}_{New})$$

$$df_{Change} = \text{Number of Parameters}_{Old} - \text{Number of Parameters}_{New}$$

Der resultierende χ^2 -Wert wird unter Berücksichtigung der Differenz der Freiheitsgrade (df_{Change}) mittels einer χ^2 -Verteilungstabelle auf Signifikanz überprüft.

6. Ergebnisse

6.1. Reliabilitäten der Skalen und Interrater-Übereinstimmungen

Die Reliabilitäten der Skalen zur Beurteilung der Förderung von *Autonomie*, *Kompetenzerleben* und *sozialer Eingebundenheit* von SchülerInnen und LehrerInnen waren allesamt sehr zufriedenstellend (siehe Tabelle 7).

Tabelle 7: Reliabilitäten der Beurteilungsskalen von SchülerInnen und LehrerInnen

Skala	Cronbach α
SchülerInnen Autonomie	.80
SchülerInnen Kompetenz	.73
SchülerInnen soziale Eingebundenheit	.83
LehrerInnen Autonomie	.72
LehrerInnen Fürsorglichkeit (soziale Eingebundenheit)	.91

Die drei Items der LehrerInnenbeurteilung der *Bezugsnorm* (Äquivalent zu *Kompetenzerleben*) stellen im eigentlichen Sinne keine Skala dar, da sie drei verschiedene Aspekte der Bezugsnormorientierung abdecken: soziale, sachliche und individuelle Orientierung der Lehrperson bei der Beurteilung ihrer SchülerInnen. Verrechnet wurden die Items analog zu einer Skala. Der Mittelwert stellt die Güte der Bezugsnorm in Hinblick auf die Förderung des Kompetenzerlebens dar (zum Zusammenhang zwischen Bezugsnormorientierung und Kompetenzerleben siehe Kapitel 5.3. *Instrumente*; mehr Informationen zum Thema Bezugsnormorientierung nachzulesen bei Rheinberg & Fries, 2010).

Die Interrater-Übereinstimmungen bei den ExpertInnenbeurteilungen wurden mittels des statistischen Kennwertes Kendall-W berechnet. Dieser misst den Grad der Übereinstimmung zwischen verschiedenen BeurteilerInnen und kann von 0 (keine Übereinstimmung) bis 1 (komplette Übereinstimmung) variieren (Bühl, 2008). Die Übereinstimmungswerte der ExpertInnenbeurteilungen waren .65 ($p < .01$) für die Beurteilung der Förderung von *Autonomie*, .64 ($p < .01$) für die Beurteilung der Förderung von *Kompetenzerleben* und .55 (p

< .05) für die Beurteilung der Förderung *sozialer Eingebundenheit*. Auch diese Werte können als ausreichend hoch bewertet werden (Bühl, 2008).

6.2. Beurteilungen der Implementierungstreue von SchülerInnen, LehrerInnen und ExpertInnen

SchülerInnen und LehrerInnen beurteilten die implementierungstreue Umsetzung der Selbstbestimmungstheorie auf einer Skala von 1 – 101. Die Mittelwerte der Beurteilungen sahen wie folgt aus (siehe Tabelle 8):

Tabelle 8: Mittelwerte (M) und Standardabweichungen (SD) der SchülerInnen- und LehrerInnenbeurteilungen der implementierungstreuen Umsetzung in Bezug auf die Umsetzung der Selbstbestimmungstheorie

	M	SD
SchülerInnen		
Autonomie	64.10	21.59
Kompetenzerleben	62.46	23.55
soziale Eingebundenheit	72.34	23.85
LehrerInnen		
Autonomie	82.10	15.59
Kompetenz	76.90	15.52
soziale Eingebundenheit	89.13	11.63

Die ExpertInnenbeurteilungen wurden auf einer Skala von 0 – 3 angegeben (Mittelwerte und Standardabweichung siehe Tabelle 9):

Tabelle 9: Mittelwerte (M) und Standardabweichungen (SD) der ExpertInnenbeurteilungen

ExpertInnen	M	SD
Autonomie	2.08	0.93
Kompetenzerleben	1.90	0.98
soziale Eingebundenheit	2.02	0.77

Bei genauerer Betrachtung der Rohdaten erkennt man, dass SchülerInnen und ExpertInnen die gesamte Beurteilungsspanne der Skalen genutzt haben, während sich die LehrerInnenbeurteilungen eher im oberen Teil der Skala befinden (siehe Tabelle 10):

Tabelle 10: Minimum- und Maximumwerte der SchülerInnen-, LehrerInnen- und ExpertInnenbeurteilungen

	Minimum	Maximum
SchülerInnen		
Autonomie	1.00	101.00
Kompetenzerleben	1.00	101.00
soziale Eingebundenheit	1.00	101.00
LehrerInnen		
Autonomie	36.00	101.00
Kompetenzerleben	41.33	101.00
soziale Eingebundenheit	50.50	101.00
ExpertInnen		
Autonomie	0.00	3.00
Kompetenzerleben	0.00	3.00
soziale Eingebundenheit	0.00	3.00

6.3. Unterschiede zwischen den verschiedenen Beurteilungen der Implementierungstreue

Die Korrelationsanalyse der drei Beurteilungen ergab folgendes Bild (Tabelle 11): Die Korrelationen zwischen den Beurteilungen sind zwar fast durchgängig hoch signifikant aber eher niedrig.

Tabelle 11: Korrelationen zwischen den drei verschiedenen Beurteilungen der Implementierungstreue bezüglich der Umsetzung der Selbstbestimmungstheorie in den Klassenprojekten (** $p < .01$, * $p < .05$)

Autonomie	SchülerInnen	LehrerInnen	ExpertInnen
SchülerInnen		.272**	.160**
LehrerInnen			.436**
ExpertInnen			
Kompetenzerleben	SchülerInnen	LehrerInnen	ExpertInnen
SchülerInnen		.139**	.147**
LehrerInnen			.110**
ExpertInnen			
Soziale Eingebundenheit	SchülerInnen	LehrerInnen	ExpertInnen
SchülerInnen		.059	.139**
LehrerInnen			.212**
ExpertInnen			

Die höchste Korrelation besteht zwischen der LehrerInnen- und der ExpertInnenbeurteilung von *Autonomie* (.44, $p < .01$). Aber auch die niedrigste signifikante Korrelation findet sich zwischen einer LehrerInnen- und einer ExpertInnenbeurteilung: der Zusammenhang zwischen den Beurteilungen von *Kompetenzerleben* liegt nur bei .11 ($p < .01$). Zwischen der SchülerInnen- und LehrerInnenbeurteilungen der *sozialen Eingebundenheit* besteht kein signifikanter Zusammenhang (.06, $p = .17$). Ein T-Test zeigt, dass die LehrerInnenbeurteilungen von *Autonomie*, *Kompetenzerleben* und *sozialer Eingebundenheit*

über alle Klassen hinweg signifikant ($p < .01$) höher liegen als die Beurteilungen der SchülerInnen.

6.4. Zusammenhang zwischen Implementierungstreue und Wirksamkeit der Klassenprojekte

Die Mehrebenenanalyse zum Vergleich des Zusammenhangs zwischen den Beurteilungen der Implementierungstreue und dem Trainingseffekt *Veränderung des Umgangs mit Unterschiedlichkeit* ergab folgendes Modell (siehe Abbildung 6). In der Darstellung sind nur (marginal) signifikante Parameter enthalten.

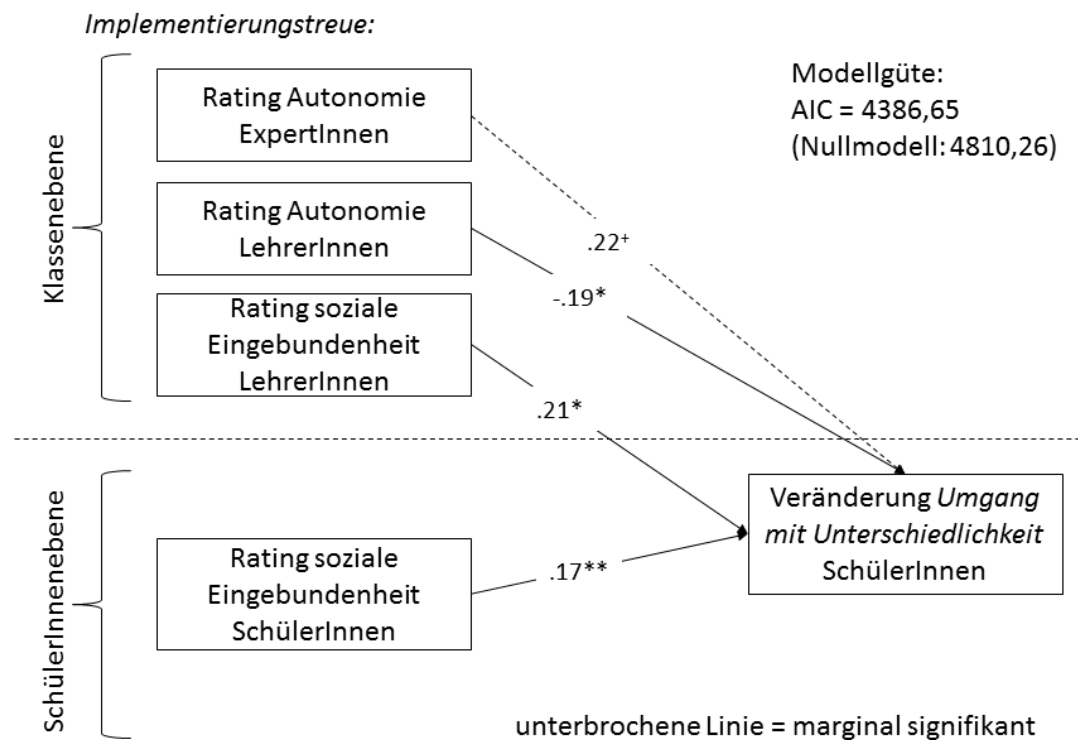


Abbildung 6: Modell des Zusammenhangs zwischen den verschiedenen Beurteilungen der implementierungstreuen Umsetzung bezüglich Selbstbestimmungstheorie und dem Trainingsoutcome *Veränderung des Umgangs mit Unterschiedlichkeit* (* $p < .05$, ** $p < .01$; + marginal signifikant)

Das berechnete Modell wurde mit dem Nullmodell verglichen. Die Güteparameter weisen auf eine gute Passung des Modells zu den Daten hin ($AIC_{\text{Modell}} = 4386.65$, $AIC_{\text{Nullmodell}} =$

4810.26; $\chi^2(8) = 423.61$, $p < .01$). Die eingetragenen β -Werte sind bereits standardisiert (siehe Kapitel 5.4.1. *Mehrebenenanalyse*). Die Ergebnisse zeigen, dass die SchülerInnenbeurteilung der *sozialen Eingebundenheit* in signifikant positivem Zusammenhang mit der *Veränderung im Umgang mit Unterschiedlichkeit* steht ($\beta = .17$, $p < .01$). Auch die LehrerInnenbeurteilung der *sozialen Eingebundenheit* steht in signifikant positivem Zusammenhang mit dem Outcome ($\beta = .21$, $p < .05$). Die LehrerInnenbeurteilung der *Autonomie* jedoch steht in signifikant negativem Zusammenhang mit dem Outcome ($\beta = -.19$, $p < .05$). Die ExpertInnenbeurteilung der *Autonomie* steht in marginal signifikantem, positivem Zusammenhang mit dem Trainingseffekt ($\beta = .22$, $p = .07$).

7. Diskussion

7.1. Zusammenfassung und Interpretation der Ergebnisse

7.1.1. Beurteilungen der Implementierungstreue von SchülerInnen, LehrerInnen und ExpertInnen

Die SchülerInnen beurteilten die implementierungstreue Umsetzung in Bezug auf *soziale Eingebundenheit* am höchsten ($M = 72.34$, $SD = 23.85$) und in Bezug auf *Kompetenzerleben* am niedrigsten ($M = 62.64$, $SD = 23.55$). Bei der Beurteilung nutzten die SchülerInnen die gesamte Beurteilungsspanne der Skala (1 – 101), was für eine differenzierte Beurteilung der Implementierungstreue spricht.

Die LehrerInnen beurteilten die implementierungstreue Umsetzung in Bezug auf alle drei Bereiche der Selbstbestimmungstheorie sehr hoch. Am besten fiel ebenfalls die Beurteilung der *sozialen Eingebundenheit* aus ($M = 89.13$, $SD = 11.63$) und am niedrigsten jene des *Kompetenzerlebens* ($M = 76.90$, $SD = 15.52$). Außerdem ist die Beurteilungsspanne vergleichsweise eng: Keine der Beurteilungen lag unter 36.00. Bei der Beurteilung der *sozialen Eingebundenheit* ist die niedrigste Beurteilung gar 50.50. Dies könnte darauf hindeuten, dass die Lehrkräfte sozial erwünscht geantwortet haben – ein Ergebnis, welches sich mit den Ergebnissen anderer Studien deckt (Lane et al., 2004; Lillehoj et al., 2004).

Die ExpertInnen beurteilten die implementierungstreue Umsetzung in Bezug auf alle drei Bereiche ebenfalls eher hoch. Auf einer Skala von 0 – 3 lagen die durchschnittlichen Beurteilungen zwischen $M = 2.08$ (*Autonomie*, $SD = 0.93$) und $M = 1.90$ (*Kompetenzerleben*, $SD = 0.98$). Wie die SchülerInnen nutzten die ExpertInnen die gesamte Spannweite der Beurteilungsskala (0 – 3).

7.1.2. Unterschiede zwischen den verschiedenen Beurteilungen der Implementierungstreue

Die Analyse der Daten hat gezeigt, dass sich die Beurteilungen der Implementierungstreue durch die drei verschiedenen Gruppen nur in geringem Maße decken. Die Werte korrelieren zwar alle bis auf die Beurteilung der *sozialen Eingebundenheit* durch SchülerInnen und

LehrerInnen signifikant, jedoch deutlich niedriger als erwartet. Vor allem bei den Beurteilungen der LehrerInnen und SchülerInnen ist dies überraschend, da beide Gruppen direkt in die Intervention involviert waren. Die Beurteilung der *Autonomie* basiert sogar auf derselben Skala, doch auch hier liegt die Korrelation nur bei .27 ($p < .01$). Ein T-Test hat gezeigt, dass die Beurteilungen der LehrerInnen im Mittel über alle drei Bereiche und alle Klassen hinweg signifikant ($p < .01$) höher lagen als die der SchülerInnen. Dies deckt sich auch mit der genaueren Betrachtung der deskriptiven Daten, die gezeigt hat, dass die Lehrpersonen die implementierungstreue Umsetzung in Bezug auf alle drei Variablen tendenziell in der oberen Hälfte der Skala beurteilt haben. Bereits in früheren Studien konnte immer wieder gezeigt werden, dass Werte aus Selbstberichten tendenziell höher liegen als jene aus objektiveren ExpertInnenbeurteilungen (Lane et al., 2004) – der signifikante Unterschied zwischen LehrerInnen- und SchülerInnenbeurteilungen bestätigt diese Tendenz durch einen zusätzlichen Vergleich mit einer bisher kaum untersuchten Beurteilungsgruppe.

Auch der tendenziell niedrige Zusammenhang zwischen den ExpertInnen- und den LehrerInnenbeurteilungen ist auffallend. Zwar liegt die Korrelation bei der Beurteilung der *Autonomie* vergleichsweise hoch (.44, $p < .01$), die Korrelation zwischen den Beurteilungen der *sozialen Eingebundenheit* (.21, $p < .01$) und der Förderung von *Kompetenzerleben* (.11, $p < .01$) sind aber wiederum sehr niedrig. In den wenigen anderen vergleichbaren Studien wurden zumeist höhere Korrelationen zwischen den ExpertInnenbeurteilungen und Selbstberichten gefunden (Lillehoj et al., 2004). Die SchülerInnen- und ExpertInnenbeurteilungen korrelieren allesamt sehr niedrig, was jedoch weniger verwundert, da die Ratings der ExpertInnen auf den Portfolios der Lehrpersonen beruhen und die Beurteilungen der SchülerInnen- und LehrerInnen nur in geringem Maße korrelieren.

Es zeigt sich somit, dass die drei verschiedenen Beurteilungsgruppen sich nur bis zu einem gewissen Grad einig über die implementierungstreue Umsetzung in Bezug auf die Selbstbestimmungstheorie waren und dass die Lehrpersonen diese über alle Klassen und Variablen hinweg signifikant höher beurteilt haben als die SchülerInnen. Die niedrige Korrelation zwischen SchülerInnen und LehrerInnen könnte demnach unter anderem durch die systematisch bessere Beurteilung durch die LehrerInnen und die geringere Spannweite der LehrerInnenbeurteilungen bedingt sein. Die niedrigen Korrelationen zwischen

SchülerInnen- und ExpertInnenbeurteilungen könnten dadurch bedingt sein, dass die Beurteilungen der ExpertInnen auf den Unterlagen der Lehrpersonen beruhen. Zusammengefasst zeigen die Ergebnisse, dass durch die Betrachtung von nur einer Perspektive bei der Beurteilung von Implementierungstreue mit Verzerrungen gerechnet werden sollte. Gerade wenn nur Selbstberichte von Lehrpersonen als Quelle verwendet werden, muss man davon ausgehen, dass diese die Implementierungstreue aufgrund von sozial erwünschten Antworten besonders gut beurteilen.

7.1.3. Zusammenhang zwischen Implementierungstreue und Wirksamkeit der Klassenprojekte

Das Mehrebenenmodell des Zusammenhangs zwischen den Beurteilungen der Implementierungstreue und dem Trainingseffekt *Veränderung des Umgangs mit Unterschiedlichkeit* beinhaltet vier verschiedene Prädiktoren: Insgesamt gibt es drei signifikante Prädiktoren (die LehrerInnenbeurteilungen der implementierungstreuen Umsetzung von *Autonomie* und *sozialer Eingebundenheit*, sowie die SchülerInnenbeurteilung der implementierungstreuen Umsetzung *sozialer Eingebundenheit*) und einen marginal signifikanten Prädiktor (die ExpertInnenbeurteilung der implementierungstreuen Umsetzung von *Autonomie*). Je besser die LehrerInnen und die SchülerInnen die theoriekonforme Förderung der *sozialen Eingebundenheit* im Projekt einschätzten, desto größer war auch die durch das Projekt erzielte Veränderung des *Umgangs mit Unterschiedlichkeit* (LehrerInnenbeurteilung $\beta = .21$, $p < .05$; SchülerInnenbeurteilung $\beta = .17$, $p < .01$). Ebenso verhält es sich mit der Beurteilung der *Autonomie* durch die ExpertInnen, wenngleich dieser Zusammenhang nur marginal signifikant ausfällt ($\beta = .22$, $p = .07$).

Somit gibt es aus allen drei Perspektiven eine Beurteilung der implementierungstreuen Umsetzung, welche den Trainingseffekt (zumindest marginal) signifikant vorhersagen kann. Auch der noch neuere Zugang der SchülerInnenbefragung hat sich als durchaus vielversprechend erwiesen. Dass die ExpertInnenbeurteilung nur einen marginal signifikanten Prädiktor für den Trainingseffekt aufweist, kann darauf hinweisen, dass ExpertInnenbeurteilungen auf Grundlage von Projektdokumentationen noch einer Verbesserung in Hinblick auf Standardisierung und Validität bedürfen. Anders als in der Literatur bisher zumeist berichtet, kann auch eine Beurteilung aus der

LehrerInnenperspektive (Selbstberichte) den Trainingseffekt signifikant vorhersagen. Es findet sich jedoch auch ein überraschendes Ergebnis, welches man etwas genauer betrachten muss: Je höher die Beurteilung der *Autonomie* durch die Lehrpersonen ausfiel, desto geringer war die erzielte Veränderung im Projekt in Bezug auf den *Umgang mit Unterschiedlichkeit* ($\beta = -.19$, $p < .05$). Diese Beurteilung der Lehrpersonen steht somit anders als erwartet in negativem Zusammenhang mit dem Trainingseffekt. Eine mögliche Erklärung für dieses kontraintuitive Ergebnis könnten durch soziale Erwünschtheit verzerrte Antworten der Lehrpersonen sein. Dies würde bedeuten, dass die *Autonomie* im Projekt tatsächlich nicht (oder nicht in dem Maße) wie durch die LehrerInnen angegeben stattgefunden hat, und somit eine wesentliche Komponente der Intervention nicht umgesetzt wurde. Dies könnte im Endeffekt zu schlechteren Trainingseffekten in den besagten Projekten geführt haben. Eine weitere Erklärung könnte sein, dass die Lehrpersonen inadäquate „Laissez-faire“-Prinzipien als *Autonomie* gewertet haben, welche jedoch zu schlechteren Ergebnissen geführt haben, da die SchülerInnen nicht ausreichend für die autonome Arbeit vorbereitet und bei ihren Tätigkeiten unterstützt wurden (dazu vgl. Kapitel 3.1.1. *Die Selbstbestimmungstheorie und Förderung von Motivation*). Um diesen Erklärungsansatz überprüfen zu können, müsste man sich zusätzlich zur implementierungstreuen Umsetzung auch noch Maße der Process-Fidelity wie die Qualität der Vermittlung oder die aktive Teilnahme der SchülerInnen ansehen, um erkennen zu können, wie qualifiziert die Lehrpersonen die SchülerInnen zu autonomem Arbeiten angeleitet haben und wie gut die SchülerInnen darauf angesprochen haben. Ebenfalls wichtig zu beachten ist hierbei, dass die Beurteilung der *Autonomie* durch die ExpertInnen in (zwar nur marginal signifikant aber) positivem Zusammenhang mit dem Trainingseffekt steht ($\beta = .22$, $p = .07$), obwohl die Ratings der ExpertInnen auf Unterlagen der Lehrpersonen beruhen. Auch dies könnte ein Hinweis darauf sein, dass die Lehrpersonen zwar theoretisch wussten, wie sie die Autonomie der SchülerInnen fördern können und in den Portfolios auch passende Strategien angegeben haben, bei der tatsächlichen Interaktion in der Klasse aber möglicherweise inadäquat gehandelt haben. Das Ergebnis zeigt somit einerseits die Nachteile von Selbstberichten auf. Andererseits kann es aber auch ein Hinweis darauf sein, dass die Lehrpersonen Schwierigkeiten hatten, die Autonomie der SchülerInnen adäquat zu unterstützen – eine Information, mit der man bei einer formativen Evaluation und entsprechenden Interventionen die Implementierungstreue der Klassenprojekte bereits während der

Umsetzung verbessern könnte. Des Weiteren fällt auf, dass gleich zwei Beurteilungen der *sozialen Eingebundenheit* signifikante Prädiktoren des Trainingseffekts darstellen. Dies lässt sich vermutlich durch die inhaltliche Nähe zur Variable *Umgang mit Unterschiedlichkeit* erklären. Im Sinne der *sozialen Eingebundenheit* ist es zentral, auf einen wertschätzenden und respektvollen Umgang untereinander zu achten und prosoziales Verhalten in der Klasse zu fördern. Die implementierungstreue Umsetzung dieser Komponente beeinflusst klarerweise den *Umgang mit Unterschiedlichkeit* auch auf einer inhaltlichen Ebene.

Die Beantwortung der zweiten Fragenstellung stellt sich somit etwas schwieriger dar: Es hat sich gezeigt, dass alle drei Perspektiven zur genaueren Erklärung des Zustandekommens der Veränderung des *Umgangs mit Unterschiedlichkeit* beitragen können. Je eine Beurteilungsvariable pro Perspektive entspricht einem (zumindest marginal) signifikanten Prädiktor im Modell. Anders als in der Literatur bisher zumeist berichtet, gibt es auch zwischen einer Selbstberichte-Variable und dem Trainingseffekt einen signifikanten Zusammenhang (Beurteilung der *sozialen Eingebundenheit* durch LehrerInnen: $\beta=.21$, $p<.05$). Das kontraintuitive Ergebnis zur LehrerInnenbeurteilung der implementierungstreuen Umsetzung in Bezug auf *Autonomie* und deren negativer Zusammenhang mit dem Trainingseffekt kann durch die gemeinsame Betrachtung mit den ExpertInnen- und SchülerInnenbeurteilungen besser interpretiert werden. Durch die gemeinsame Verwendung dreier Beurteilungsperspektiven hat sich außerdem gezeigt, dass der *sozialen Eingebundenheit* eine besondere Bedeutung bei der Veränderung des *Umgangs mit Unterschiedlichkeit* zukommt. Alle drei Beurteilungsperspektiven tragen zu einem besseren Verständnis der Durchführung in der Klasse und der implementierungstreuen Umsetzung der Core Component bei und liefern unterschiedliche Informationen zum Zustandekommen des Trainingseffekts. Es zeigen sich auch in dieser Studie die verschiedenen Limitationen der unterschiedlichen Perspektiven sowie die Vorteile und der Informationsgewinn durch die gemeinsame Verwendung verschiedener Erhebungsansätze. Es kann somit keine Entscheidung für eine Perspektive fallen, welche am besten die Trainingseffekte voraussagen kann, sondern es muss für einen multi-method und multi-informant Ansatz bei Implementierungsevaluationen plädiert werden. Erst die Befragung verschiedener Personen (LehrerInnen, SchülerInnen und ExpertInnen) kann ein umfassendes Bild der Implementierung und der Wirkmechanismen bei der Entstehung von Trainingseffekten bieten.

7.2. Forschungsausblick

Die vorliegende Studie hatte zum Ziel, verschiedene Perspektiven auf die implementierungstreue Umsetzung der Klassenprojekte von *Reflect* zu vergleichen und ihre Relevanz für die Wirksamkeit der Intervention näher zu analysieren. Die Übereinstimmungen zwischen den verschiedenen Beurteilungen erwiesen sich als signifikant aber vergleichsweise niedrig. Die Beurteilungen der LehrerInnen lagen dabei im Schnitt über alle Klassen hinweg höher als die der SchülerInnen und der ExpertInnen. Es muss zwar angemerkt werden, dass aufgrund von Mangel an guten Skalen für die Beurteilungen der Lehrpersonen nur Teilaspekte der Umsetzung (*Bezugsnorm* und *Fürsorglichkeit*) abgefragt werden konnten, was die Vergleichbarkeit der Beurteilungen mit den beiden anderen Perspektiven etwas einschränken könnte. Jedoch auch in Bezug auf die Beurteilung der *Autonomie* waren die Korrelationen mit den SchülerInnen sehr niedrig – obwohl in diesem Fall die Skalen sogar gänzlich ident waren.

Eine generelle Problematik dieser Studie ist, dass sich die verschiedenen Perspektiven auf die implementierungstreue Umsetzung nur mittels unterschiedlicher Erhebungsinstrumente erfassen ließen. Daraus folgte eine Konfundierung der Ebenen *Messinstrument* und *Beurteilungsperspektive*. Bei weiteren Forschungen in diesem Bereich sollte darauf Wert gelegt werden, dass man diese Ebenen zur besseren Interpretierbarkeit methodisch angleicht. Dies wird nicht zur Gänze möglich sein, da die Befragung unterschiedlicher Personengruppen eine Verwendung unterschiedlicher Erhebungsmethoden notwendig macht. Jedoch sollte man versuchen, möglichst ähnliche Skalen und Formulierungen von Items zu verwenden. Eine weitere Schwierigkeit ergibt sich durch die Vermischung der implementierungstreuen Umsetzung mit inhaltlichen Aspekten der Selbstbestimmungstheorie. Da die drei Bereiche *Kompetenzerleben*, *Autonomie* und *soziale Eingebundenheit* getrennt voneinander beurteilt werden mussten, kann es sein, dass die Bedeutsamkeit der Beurteilungen als Prädiktoren für den Trainingseffekt durch die inhaltliche Relevanz mitbeeinflusst wird. So war keine der Beurteilungen der implementierungstreuen Umsetzung in Bezug auf *Kompetenzerleben* signifikant, wohingegen aber gleich zwei Beurteilungen von *Förderung sozialer Eingebundenheit* als signifikante Prädiktoren vertreten waren, was durch die inhaltliche Nähe zum Trainingseffekt *Umgang mit Unterschiedlichkeit* bedingt sein könnte. Auch solche Verzerrungen sollten in künftigen

Studien beachtet und wenn möglich eliminiert werden. Die Klassenprojekte von *Reflect* waren zwar nicht gänzlich standardisiert, die Projekte hatten aber gemeinsame Core Components und die Planung und Durchführung wurde von ExpertInnen supervidiert. Durch diese systematische Planung der Projekte fiel die Operationalisierung der implementierungstreuen Umsetzung in Bezug auf die Core Component *Umsetzung der Selbstbestimmungstheorie* nicht schwer. Somit hat sich auch in dieser Arbeit herausgestellt, wie wichtig es ist, das Wirkmodell hinter den zu evaluierenden Interventionen gut zu kennen (siehe Kapitel 2.5.1. *Identifikation und Operationalisierung der Core Components*).

Die Ratings der ExpertInnen beruhten auf den Portfolios der Lehrpersonen, was eine objektive Beurteilung erheblich erschwerte. Alternativ wären fallweise Besuche in den Klassen und Beurteilungen vor Ort möglich gewesen, was jedoch ressourcentechnisch nicht möglich war. Außerdem bringt auch ein solcher Zugang Probleme mit sich: Wie in Kapitel 2.5.2. *Datenerhebung* erläutert, kann die Anwesenheit von BeurteilerInnen das Verhalten von TrainerInnen und Teilnehmenden erheblich verzerren. Außerdem können einzelne Visiten nur einen sehr kleinen Ausschnitt der gesamten Umsetzung vermitteln. Die Beurteilung auf Grundlage von Portfolios oder anderen Dokumentationen der Umsetzung ist zwar weniger objektiv als mittels direkter Beobachtungen während des Trainings. Dennoch ergeben sich durch diese weniger eingreifende Form der ExpertInnenbeurteilung auch einige Vorteile, die in der Praxis durchaus Relevanz haben. Es werden weniger Ressourcen benötigt und die TrainerInnen fühlen sich dadurch wesentlich weniger kontrolliert. Deshalb wird diese Form der ExpertInnenbeurteilung auch weiterhin angezeigt sein. In zukünftigen Studien sollte man Fokus auf die methodische Optimierung dieser Form der Beurteilungen legen und sich Möglichkeiten überlegen, wie man diese reliabler und valider gestalten kann. Die Portfolios in dieser Studie waren sehr unstandardisiert. Dies bringt Vorteile aber auch Nachteile mit sich: Durch die freie Gestaltung ist die Gefahr, nur sozial erwünschte Antworten von den TrainerInnen zu erhalten, geringer als bei standardisierteren, fragebogenähnlichen Checklisten. Außerdem bietet sich den ExpertInnen durch die offenere Form mehr Interpretationsspielraum und es werden nicht zwingenderweise einfach die LehrerInnenangaben repliziert. Auf der anderen Seite fallen die Ratings auf Grundlage unterschiedlichster Unterlagen natürlich aber sehr schwer. Teilweise waren notwendige Informationen in den Portfolios nicht enthalten oder man konnte die Ausführungen der Lehrpersonen nur sehr schwer deuten. Eine gewisse Standardisierung von Portfolios würde

in künftigen Arbeiten ein solches Rating sicherlich erleichtern. Man könnte zum Beispiel gezieltere Fragen stellen, welche in der Reflexion beantwortet werden sollen. So wäre es denkbar, dass die Umsetzung nach Core Components aufgeschlüsselt reflektiert wird, was garantieren könnte, dass alle für eine Beurteilung notwendigen Informationen auch tatsächlich vorhanden sind. Eine weitere Schwierigkeit bei den ExpertInnenratings ergab sich durch den gemeinsamen Konsens, der gefunden werden musste. Dadurch ergab sich für die Bewertung eine starke Tendenz zur Mitte, was sich auch in den tendenziell weniger signifikanten Ergebnissen widerspiegelt. Es ist wichtig, dies bei der Interpretation der Ergebnisse zu bedenken.

Für zukünftige, umfangreichere Forschungsarbeiten ist außerdem unbedingt zu überlegen, auch Prozess-Variablen in die Beurteilung der Implementierungstreue miteinzubeziehen, damit man Ergebnisse wie bei der LehrerInnenbeurteilung von *Autonomie* besser interpretieren kann. Obwohl die Erhebung solcher Daten wesentlich subjektivere Maße erfordert und auch schwieriger durchzuführen ist als die Erhebung von Inhalts- bzw. Strukturmerkmalen der Implementierungstreue (siehe Kapitel 2.5.1. *Identifikation und Operationalisierung der Core Components*), so lassen sich dadurch dennoch weitere sehr wichtige Informationen zu den Prozessen sammeln, welche bei der Umsetzung von Interventionen in der Praxis relevant sind.

Der hier ausgewählte Trainingseffekt *Veränderung des Umgang mit Unterschiedlichkeit* ist zwar einer der zentralen Trainingseffekte von *Reflect*, es ist aber auch ein schwer zu interpretierender Trainingseffekt, da die Skala sehr allgemein ist und für das Zustandekommen der Veränderung viele verschiedene Faktoren relevant sind. In zukünftigen, umfangreicheren Studien wäre es wünschenswert, dass man die verschiedenen Beurteilungen von Implementierungstreue mit mehreren Trainingseffekten in Verbindung bringt. Dies könnte zu einem differenzierteren Bild in Bezug auf die Umsetzung und die Wirksamkeit von Interventionen wesentlich beitragen.

Die meisten Studien erheben Implementierungstreue nur aus einer Perspektive. In Anbetracht der Ergebnisse dieser Studie stellt sich diese Vorgangsweise als sehr problematisch dar, da sich gezeigt hat, dass sich die verschiedenen Perspektiven doch stark unterscheiden und die verschiedenen Beurteilungen unterschiedliche Zusammenhänge mit dem Trainingseffekt aufweisen. Wenn aus ressourcentechnischen Gründen nicht mehrere

verschiedene Beurteilungsperspektiven erhoben werden können, so sollte man sich genau über die Vorteile und Nachteile der einzelnen Perspektiven informieren, bevor man sich für die Befragung einer Personengruppe entscheidet. Diese Studie konnte durch den Vergleich dreier verschiedener Perspektiven einen Beitrag zu mehr Wissen über die verschiedenen Zugänge leisten (siehe Kapitel 7.1. *Zusammenfassung und Interpretation der Ergebnisse*). Man sollte auch weiterhin Untersuchungen durchführen, in denen Implementierungstreue aus verschiedenen Perspektiven erhoben wird, da man noch mehr Erkenntnisse für die Entwicklung von Messstandards und validen Instrumenten zur Erhebung von Implementierungsdaten braucht. Außerdem wäre es erstrebenswert, mehrere verschiedene Aspekte von Implementierungstreue (auch in Bezug auf Process-Fidelity) parallel zu betrachten, um das Konstrukt Implementierungstreue noch besser definieren zu können. So sollte sich in weiterer Folge auch herausstellen, welche Operationalisierungen und Erhebungsinstrumente für den jeweiligen Teilaspekt von Implementierungstreue am besten funktionieren.

Alle drei Perspektiven können zur Erklärung der Wirksamkeit von *Reflect* beitragen. Es wurden aber auch verschiedene Einschränkungen evident: LehrerInnen scheinen die Implementierungstreue aufgrund von sozial erwünschten Antworten systematisch zu überschätzen. Die Beurteilungen von ExpertInnen auf der Basis von Portfolios bedürfen weiterer methodischer Optimierungen zur Sicherung der Validität. Die SchülerInnenperspektive hat sich in dieser Studie als vielversprechend erwiesen. Bisher ist die Teilnehmendenperspektive wenig beachtet worden. In zukünftigen Studien sollte man sich auch auf die Entwicklung und Verbesserung geeigneter Instrumente für diese Personengruppe konzentrieren. Alles in allem ist auch bei zukünftigen Evaluationen ein multi-method und multi-informant Ansatz anzustreben, da bei der Befragung von nur einer Gruppe von Personen die dadurch erzielten Ergebnisse nur sehr schwer verallgemeinerbar sind und durch verschiedene Limitationen nur wenig Aussagekraft besitzen. Nur durch die ergänzende Beurteilung aus verschiedenen Perspektiven kann ein umfassendes und aussagekräftiges Bild der Umsetzung gewonnen werden und der Beitrag von implementierungstreuer Umsetzung für das Zustandekommen von Trainingseffekten zufriedenstellend erklärt werden.

8. Zusammenfassung

Die Erhebung von Implementierungstreue ist für die Erklärung des Zustandekommens und die Interpretation von Trainingseffekten von zentraler Bedeutung (Dane & Schneider, 1998). Dennoch wird in der klassischen Programmevaluation stets mehr Wert auf die Überprüfung von Trainingseffekten als auf die Überwachung des Prozesses der Implementierung gelegt (Elliott & Mihalic, 2004). Angaben zur Implementierung sind immer noch eine Seltenheit in Evaluationsstudien (Dane & Schneider, 1998). Falls überhaupt Daten zur Implementierungstreue erhoben werden, so stammen diese zumeist nur von einer Personengruppe und es werden kaum einmal verschiedene Perspektiven miteinander verglichen (Durlak & DuPre, 2008).

Daher war es das zentrale Anliegen dieser Studie, die Implementierungstreue des schulbasierten Gendertrainings *Reflect* aus drei verschiedenen Perspektiven zu betrachten (LehrerInnen, SchülerInnen und ExpertInnen). Eine zentrale Komponente des Projekts war die Umsetzung gemäß der Selbstbestimmungstheorie (Deci & Ryan, 2002). Die Lehrkräfte sollten ihre SchülerInnen in Bezug auf *Autonomie*, *Kompetenzerleben* und *soziale Eingebundenheit* unterstützen. Die SchülerInnen und LehrerInnen beurteilten die theoriekonforme Umsetzung mittels eines Onlinefragebogens, drei ExpertInnen anhand von Projektportfolios. Die drei Perspektiven wurden mittels einer Korrelationsanalyse und eines T-Tests verglichen. Um die Relevanz der implementierungstreuen Umsetzung für die Wirksamkeit des Projekts zu ermitteln, wurde eine Mehrebenenanalyse durchgeführt, die die verschiedenen Beurteilungen mit einem Trainingseffekt in Zusammenhang bringt (*Veränderung des Umgangs mit Unterschiedlichkeit* der SchülerInnen).

Die Korrelationsanalyse hat gezeigt, dass die verschiedenen Perspektiven durchwegs signifikant aber niedrig korrelieren. Die Beurteilungen der Lehrpersonen fielen signifikant ($p < .01$) höher aus als die Beurteilungen der SchülerInnen. Die SchülerInnen- und LehrerInnenbeurteilungen *sozialer Eingebundenheit* stellten sich als signifikante Prädiktoren des Trainingseffekts heraus (SchülerInnen $\beta = .17$, $p < .01$, LehrerInnen $\beta = .21$, $p < .05$). Die ExpertInnenbeurteilung der *Autonomie* war ein marginal signifikanter Prädiktor ($\beta = .22$, $p = .07$), die LehrerInnenbeurteilung der *Autonomie* war ein signifikanter aber negativer Prädiktor ($\beta = -.19$, $p < .05$). Die Ergebnisse zeigen, dass alle drei Beurteilungsperspektiven zur Erklärung der Wirksamkeit des Projekts beitragen können. In Anbetracht des

Informationsgewinns durch die ergänzende Verwendung muss für einen multi-method multi-informant Ansatz bei zukünftigen Implementierungsevaluationen plädiert werden.

Schlüsselbegriffe: Erfassung von Implementierungstreue, verschiedene Perspektiven, Zusammenhang mit Trainingseffekt, Implementierungsevaluation, schulbasierte Projekte

Abstract

Measuring Implementation fidelity is crucial for interpreting intervention outcomes and explaining their emergence. Despite its importance for the validity of evaluations and for replicating evidence-based programs effectively in practice (Dane & Schneider, 1998; Dumas et al., 2001; Mowbray et al., 2003) there is still a lack of implementation analysis in most evaluation studies (Dane & Schneider, 1998). If implementation fidelity is measured at all mostly only one perspective (coaches, experts or participants) is factored in and perspectives are hardly ever compared to each other (Durlak & DuPre, 2008). Therefore there is only little information about differences and agreements between the perspectives.

Accordingly this study's aim was to include three perspectives into the assessment of intervention fidelity of a school-based gender training called *Reflect*. One of the program's core components was to implement it in regard to Self-Determination Theory (Deci & Ryan, 2002). Teachers were supposed to support students' perceptions of competence, autonomy and relatedness. Students and teachers were asked to rate the degree to which teachers' implementation had adhered to this core component using an online questionnaire and three experts rated the adherence based on intervention-portfolios provided by the teachers. The three perspectives were then compared through correlational analyses and a t-test. A multilevel-analysis was conducted to assess the different predictive values for a training outcome (students' change in handling diversity).

Correlation analyses showed that the three perspectives correlated significantly but low. Teacher self-reports provided significantly higher ratings for each variable compared to student ratings ($p < .01$). Student rating and teacher rating of adherence to support of relatedness were highly significant predictors of the outcome (students $\beta = .17$, $p < .01$, teachers $\beta = .21$, $p < .05$). Expert rating of adherence to autonomy support was a marginal significant predictor of the outcome ($\beta = .22$, $p = .07$), Teacher rating of adherence to autonomy support was a significant but negative predictor of the outcome ($\beta = -.19$, $p < .05$). Based on these results the advantages and disadvantages of the three different perspectives are discussed and the benefits of a multi-method and multi-informant approach in measuring implementation fidelity are pointed out.

Keywords: implementation fidelity, adherence, measurement, different perspectives, outcome

Literaturverzeichnis

- Berkel, C., Mauricio, A. M., Schoenfelder, E., & Sandler, I. N. (2011). Putting the Pieces Together: An Integrated Model of Program Implementation. *Prevention Science, 12*(1), 23–33. doi:10.1007/s11121-010-0186-1
- Bickman, L., Riemer, M., Brown, J., Jones, S., Flay, B., Li, K.-K., DuBois, D., Pelham, W. & Massetti, G. (2009). Approaches to Measuring Implementation Fidelity in School-Based Program Evaluations. *Journal of Research in Character Education, 7*(2), 75 – 101.
- 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. doi:10.1007/BF00922697
- Blase, K., & Fixsen, D. (2013). *Core Intervention Components: Identifying and Operationalizing What Makes Programs Work* (Research Brief). University of North Carolina at Chapel Hill: National Implementation Research Network. Retrieved from <http://evidence-innovation.findyouthinfo.gov/node/21076>
- Bopp, M., Saunders, R. P., & Lattimore, D. (2013). The Tug-of-War: Fidelity Versus Adaptation Throughout the Health Promotion Program Life Cycle. *The Journal of Primary Prevention, 34*(3), 193–207. doi:10.1007/s10935-013-0299-y
- Breitenstein, S. M., Gross, D., Garvey, C. A., Hill, C., Fogg, L., & Resnick, B. (2010). Implementation fidelity in community-based interventions. *Research in Nursing & Health, 33*(2), 164–173. doi:10.1002/nur.20373

- Breitenstein, S., Robbins, L., & Muennich Cowell, J. (2012). Attention to Fidelity: Why Is It Important. *The Journal of School Nursing, 28*(6), 407–408.
doi:10.1177/1059840512465408
- Bühl, A. (2008). *SPSS 16: Einführung in die moderne Datenanalyse*. München: Pearson Deutschland GmbH.
- Burke, R. V., Oats, R. G., Ringle, J. L., Fichtner, L. O., & DelGaudio, M. B. (2011). Implementation of a Classroom Management Program with Urban Elementary Schools in Low-Income Neighborhoods: Does Program Fidelity Affect Student Behavior and Academic Outcomes? *Journal of Education for Students Placed at Risk (JESPAR), 16*(3), 201–218. doi:10.1080/10824669.2011.585944
- Carroll, C., Patterson, M., Wood, S., Booth, A., Rick, J., & Balain, S. (2007). A conceptual framework for implementation fidelity. *Implementation Science, 2*(1), 40.
doi:10.1186/1748-5908-2-40
- Castro, F. G., Barrera, M., Jr, & Martinez, C. R., Jr. (2004). The cultural adaptation of prevention interventions: resolving tensions between fidelity and fit. *Prevention Science: The Official Journal of the Society for Prevention Research, 5*(1), 41–45.
- 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), 363–374.
doi:10.1007/s10802-005-3574-4
- Crawford, L., Carpenter, D. M., Wilson, M. T., Schmeister, M., & McDonald, M. (2012). Testing the Relation Between Fidelity of Implementation and Student Outcomes in Math. *Assessment for Effective Intervention, 37*(4), 224–235.
doi:10.1177/1534508411436111

- Dane, A. V., & Schneider, B. H. (1998). Program integrity in primary and early secondary prevention: are implementation effects out of control? *Clinical Psychology Review*, *18*(1), 23–45.
- Deci, E. L., & Ryan, R. M. (2002). An Overview of Self-Determination Theory. In E.L. Deci & R.M. Ryan (Hrsg.), *Handbook of self-determination research* (S. 3 – 35). Rochester, NY [u.a.]: University Rochester Press.
- Dumas, J. E., Lynch, A. M., Laughlin, J. E., Phillips Smith, E., & Prinz, R. J. (2001). Promoting intervention fidelity: Conceptual issues, methods, and preliminary results from the EARLY ALLIANCE prevention trial. *American Journal of Preventive Medicine*, *20*(1), 38–47. doi:10.1016/S0749-3797(00)00272-5
- Durlak, J. A. (1998). Why Program Implementation is Important. *Journal of Prevention & Intervention in the Community*, *17*(2), 5–18. doi:10.1300/J005v17n02_02
- 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), 327–350. doi: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), 237–256. doi:10.1093/her/18.2.237
- Eccles, M. P., & Mittman, B. S. (2006). Welcome to Implementation Science. *Implementation Science*, *1*(1), 1. doi:10.1186/1748-5908-1-1
- Elliott, D. S., & Mihalic, S. (2004). Issues in Disseminating and Replicating Effective Prevention Programs. *Prevention Science*, *5*(1), 47–53. doi:10.1023/B:PREV.0000013981.28071.52

- Esbensen, F.-A., Matsuda, K. N., Taylor, T. J., & Peterson, D. (2011). Multimethod Strategy for Assessing Program Fidelity: The National Evaluation of the Revised G.R.E.A.T. Program. *Evaluation Review*, 35(1), 14–39. doi:10.1177/0193841X10388126
- Fagan, A. A., Hanson, K., Hawkins, J. D., & Arthur, M. W. (2008). Bridging Science to Practice: Achieving Prevention Program Implementation Fidelity in the Community Youth Development Study. *American Journal of Community Psychology*, 41(3-4), 235–249. doi:10.1007/s10464-008-9176-x
- Field, A. (2009). *Discovering Statistics Using SPSS*. London: SAGE Publications.
- Finsterwald, M., Jöstl, G., Popper, V., Hesse, N., Spiel, C., & Schober, B. (2012). *Abschlussbericht Reflect: Genderkompetenz durch Reflexive Koedukation*. Unpublished manuscript, Universität Wien.
- Finsterwald, M., Schober, B., Wagner, P., & Spiel, C. (2010). *TALK – Skalen zur Erfassung von LLL-Kompetenzen*. Unpublished manuscript, Universität Wien.
- Fixsen, D. L., Blase, K. A., Naoom, S. F., & Wallace, F. (2009). Core Implementation Components. *Research on Social Work Practice*, 19(5), 531–540. doi:10.1177/1049731509335549
- Fixsen, D. L., Naoom, S. F., Blase, K. A., Friedman, R. M., & Wallace, F. (2005). *Implementation Research: A Synthesis of the Literature*. Tampa, FL: University of South Florida.
- Gottfredson, D. C., & Gottfredson, G. D. (2002). Quality of School-Based Prevention Programs: Results from a National Survey. *Journal of Research in Crime and Delinquency*, 39(1), 3–35. doi:10.1177/002242780203900101

- Hallfors, D., & Godette, D. (2002). Will the 'Principles of Effectiveness' improve prevention practice? Early findings from a diffusion study. *Health Education Research, 17*(4), 461–470. doi:10.1093/her/17.4.461
- Heck, R. H., Thomas, S. L., & Tabata, L. N. (2010). *Multilevel and Longitudinal Modeling with IBM SPSS*. New York: Taylor & Francis.
- Hox, J. J. (2010). *Multilevel analysis: Techniques and applications*. New York: Routledge.
- Jerusalem, M., Drössler, S., Kleine, D., Klein-Heßling, J., Mittag, W., & Röder, B. (2009). *Skalenbuch: Förderung von Selbstwirksamkeit und Selbstbestimmung im Unterricht. Skalen zur Erfassung von Lehrer- und Schülermerkmalen*. Berlin: Humboldt-Universität Berlin.
- Kalafat, J., Illback, R. J., & Sanders, D. (2007). The relationship between implementation fidelity and educational outcomes in a school-based family support program: Development of a model for evaluating multidimensional full-service programs. *Evaluation and Program Planning, 30*(2), 136–148. doi:10.1016/j.evalprogplan.2007.01.004
- Krapp, A. (2005). Das Konzept der grundlegenden psychologischen Bedürfnisse. Ein Erklärungsansatz für die positiven Effekte von Wohlbefinden und intrinsischer Motivation im Lehr-Lerngeschehen. *Zeitschrift Für Pädagogik, 51*(5), 626–641.
- Kumpfer, K. L., Alvarado, R., Smith, P., & Bellamy, N. (2002). Cultural sensitivity and adaptation in family-based prevention interventions. *Prevention Science: The Official Journal of the Society for Prevention Research, 3*(3), 241–246.
- Lane, K. L., Bocian, K. M., MacMillan, D. L., & Gresham, F. M. (2004). Treatment Integrity: An Essential—But Often Forgotten—Component of School-Based Interventions.

- Preventing School Failure: Alternative Education for Children and Youth*, 48(3), 36–43.
doi:10.3200/PSFL.48.3.36-43
- Larsen, T., & Samdal, O. (2007). Implementing Second Step: Balancing Fidelity and Program Adaptation. *Journal of Educational and Psychological Consultation*, 17(1), 1–29.
doi:10.1080/10474410709336588
- 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*, 31(2), 242–257.
doi:10.1177/1090198103260514
- Lipowsky, F., Rakoczy, K., Buff, A., & Klieme, E. (2005). *Dokumentation der Erhebungs- und Auswertungsinstrumente zur schweizerisch-deutschen Videostudie "Unterrichtsqualität, Lernverhalten und mathematisches Verständnis"*. Frankfurt am Main: DIPF.
- Lüdtke, O., & Köller, O. (2010). Mehrebenenanalyse. In D.H. Rost (Hrsg.) *Handwörterbuch der Pädagogischen Psychologie* (S. 530 – 534). Weinheim, Basel: Beltz.
- Mihalic, S. F., & Irwin, K. (2003). Blueprints for Violence Prevention From Research to Real-World Settings—Factors Influencing the Successful Replication of Model Programs. *Youth Violence and Juvenile Justice*, 1(4), 307–329. doi:10.1177/1541204003255841
- Mowbray, C. T., Holter, M. C., Teague, G. B., & Bybee, D. (2003). Fidelity Criteria: Development, Measurement, and Validation. *American Journal of Evaluation*, 24(3), 315–340. doi:10.1177/109821400302400303
- Nelson, M. C., Cordray, D. S., Hulleman, C. S., Darrow, C. L., & Sommer, E. C. (2012). A Procedure for Assessing Intervention Fidelity in Experiments Testing Educational and

- Behavioral Interventions. *The Journal of Behavioral Health Services & Research*, 39(4), 374–96. doi:<http://dx.doi.org/10.1007/s11414-012-9295-x>
- O'Connor, C., Small, S. A., & Cooney, S. M. (2007). Program fidelity and adaptation: Meeting local needs without compromising program effectiveness. In *What Works, Wisconsin Research to Practice Series* (Vol. 4). Madison: University of Wisconsin–Madison.
- Power, T. J., Blom-Hoffman, J., Clarke, A. T., Riley-Tillman, T. C., Kelleher, C., & Manz, P. H. (2005). Reconceptualizing intervention integrity: A partnership-based framework for linking research with practice. *Psychology in the Schools*, 42(5), 495–507.
- 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.
- Reeve, J., Deci, E. L., & Ryan, R. M. (2002). Self-determination theory applied to educational settings. In E.L. Deci & R.M. Ryan, *Handbook of self-determination research* (S.183–204). Rochester, NY [u.a.]: University Rochester Press.
- Reinke, W. M., Herman, K. C., Stormont, M., Newcomer, L., & David, K. (2013). Illustrating the Multiple Facets and Levels of Fidelity of Implementation to a Teacher Classroom Management Intervention. *Administration and Policy in Mental Health and Mental Health Services Research*, 40(6), 494–506. doi:[10.1007/s10488-013-0496-2](https://doi.org/10.1007/s10488-013-0496-2)
- Rheinberg, F., & Fries, S. (2010). Bezugsnormorientierung. In D.H. Rost (Hrsg.), *Handwörterbuch Pädagogische Psychologie* (S. 61 – 67). Weinheim, Basel: Beltz.
- Salmivalli, C., Kaukiainen, A., & Voeten, M. (2005). Anti-bullying intervention: Implementation and outcome. *British Journal of Educational Psychology*, 75(3), 465–487. doi:[10.1348/000709905X26011](https://doi.org/10.1348/000709905X26011)

Schultes, M.-T., Joestl, G., Finsterwald, M., Schober, B., & Spiel, C. (in Vorbereitung).

Measuring program adherence from different perspectives with multiple methods: Comparing experts' and students' ratings on the implementation of the "Reflect" program in their relations with outcomes.

Skolits, G. J., & Richards, J. (2010). Designing and applying project fidelity assessment for a teacher implemented middle school instructional improvement pilot intervention.

The Canadian Journal of Program Evaluation, 24(1), 133–156.

Smith, S. W., Daunic, A. P., & Taylor, G. G. (2007). Treatment Fidelity in Applied Educational Research: Expanding the Adoption and Application of Measures to Ensure Evidence-Based Practice. *Education and Treatment of Children*, 30(4), 121–134.

doi:10.1353/etc.2007.0033

Spiel, C., & Strohmeier, D. (2012). Evidence-based practice and policy: When researchers, policy makers, and practitioners learn how to work together. *European Journal of Developmental Psychology*, 9(1), 150–162. doi:10.1080/17405629.2011.616776

Wilson, S. J., & Lipsey, M. W. (2007). School-Based Interventions for Aggressive and Disruptive Behavior: Update of a Meta-Analysis. *American Journal of Preventive Medicine*, 33(2), 130–143. doi:10.1016/j.amepre.2007.04.011

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(1), 136–149.

Ziegler, A., Heller, K., Schober, B., & Dresel, M. (2006). The Actiotope: A heuristic model for a research program designed to examine and reduce adverse motivational conditions influencing scholastic achievement. In D. Frey, H. Mandl & L. v. Rosenstiel (Hrsg.), *Knowledge and action* (S.143–173), Seattle: Hogrefe.

Abbildungsverzeichnis

Abbildung 1: klassisches Evaluationsmodell mit Versuchs- und Kontrollgruppe, Grafik angelehnt an Nelson et al. (2012), S. 376	5
Abbildung 2: Selbstbestimmungskontinuum mit den verschiedenen Formen der Motivation und Regulation nach Deci & Ryan (2002, S. 16)	22
Abbildung 3: Evaluationsdesign der begleitenden Evaluation von <i>Reflect</i>	32
Abbildung 4: Veranschaulichung der hierarchischen Datenstruktur in den Klassenprojekten	40
Abbildung 5: hypothetisches Modell des Zusammenhangs zwischen den verschiedenen Beurteilungen der implementierungstreuen Umsetzung bezüglich Selbstbestimmungstheorie und dem Trainingsoutcome <i>Veränderung des Umgangs mit Unterschiedlichkeit</i>	42
Abbildung 6: Modell des Zusammenhangs zwischen den verschiedenen Beurteilungen der implementierungstreuen Umsetzung bezüglich Selbstbestimmungstheorie und dem Trainingsoutcome <i>Veränderung des Umgangs mit Unterschiedlichkeit</i> (* $p < .05$, ** $p < .01$; + marginal signifikant).....	48

Tabellenverzeichnis

Tabelle 1: Inhaltliche Module des LehrerInnentrainings von <i>Reflect</i>	21
Tabelle 2: zentrale Themen der Klassenprojekte von <i>Reflect</i>	27
Tabelle 3: Items der Skala <i>Umgang mit Unterschiedlichkeit</i>	34
Tabelle 4: Beispiele für SchülerInnenitems zur Beurteilung von Autonomie ¹ , Kompetenzerleben ² und sozialer Eingebundenheit ³	35
Tabelle 5: Beispiele für LehrerInnenitems zur Beurteilung der Autonomie ¹ , der Bezugsnorm ² und der Fürsorglichkeit ³	36
Tabelle 6: Beurteilungsgrundlage für das ExpertInnenrating zu den Portfolios.....	37
Tabelle 7: Reliabilitäten der Beurteilungsskalen von SchülerInnen und LehrerInnen.....	44
Tabelle 8: Mittelwerte (M) und Standardabweichungen (SD) der SchülerInnen- und LehrerInnenbeurteilungen der implementierungstreuen Umsetzung in Bezug auf die Umsetzung der Selbstbestimmungstheorie.....	45
Tabelle 9: Mittelwerte (M) und Standardabweichungen (SD) der ExpertInnenbeurteilungen	46
Tabelle 10: Minimum- und Maximumwerte der SchülerInnen-, LehrerInnen- und ExpertInnenbeurteilungen	46
Tabelle 11: Korrelationen zwischen den drei verschiedenen Beurteilungen der Implementierungstreue bezüglich der Umsetzung der Selbstbestimmungstheorie in den Klassenprojekten (** $p < .01$, * $p < .05$).....	47

Anhang

I. Verwendete Skalen

Skala Umgang mit Unterschiedlichkeit (SchülerInnen)

Umgang mit Unterschiedlichkeit

Beim nächsten Thema geht es darum, wie du eure Klasse wahrnimmst.

Auch dabei gibt es keine richtigen oder falschen Antworten, es geht um deine persönliche Meinung. Klicke dazu auf den Balken, und platziere dann den erscheinenden Knopf je nachdem wie sehr du der Aussage zustimmst oder nicht. Die Mitte bedeutet, dass du weder zustimmst noch ablehnst.

Wie ist das bei euch in der Klasse?

In unserer Klasse...

- ...werden Schüler/Schülerinnen ausgeschlossen, die anders sind.
- ...wird jede/jeder so angenommen, wie er/sie ist.
- ...wird jemand leicht ausgeschlossen, wenn er/sie eine andere Meinung hat als die anderen.
- ...ist es egal, welche Muttersprache jemand hat.
- ...mögen wir auch Mädchen, die sich nicht wie „typische Mädchen“ verhalten.
- ...mögen wir auch Buben, die sich nicht wie „typische Buben“ verhalten.
- ...ist es eine besondere Stärke, dass wir alle verschiedene Begabungen und Fähigkeiten haben.

Skalen zur Beurteilung der implementierungstreuen Umsetzung der Selbstbestimmungstheorie (SchülerInnen)

Beurteilung der implementierungstreuen Umsetzung der Selbstbestimmungstheorie

Nun geht es darum, wie du euer Projekt erlebt hast. Wie du schon weißt, geht es wieder um deine Meinung, es gibt keine richtigen oder falschen Antworten.

Im Projekt...

Selbstbestimmungstheorie: Autonomie

- ...konnte ich selber entscheiden, wie ich arbeiten will.
- ...hatte ich die Gelegenheit, mich mit interessanten Aufgaben oder Inhalten genauer zu beschäftigen.
- ...hatte ich die Möglichkeit, neue Themen selbstständig zu erkunden.
- ...wurde ich zu selbstständigem Arbeiten ermuntert.
- ...konnte ich zwischen verschiedenen Aufgaben wählen.
- ...konnte ich wählen, wann ich welche Aufgabe bearbeite.

Selbstbestimmungstheorie: Kompetenzerleben

- ...fanden meine Leistungen Anerkennung.
- ...sagte mir der Lehrer/die Lehrerin, was ich noch verbessern könnte.
- ...wurde ich für gute Leistungen gelobt.
- ...informierte mich der Lehrer/die Lehrerin über meine Fortschritte.

Selbstbestimmungstheorie: soziale Eingebundenheit

- ...hatte ich das Gefühl, dazuzugehören.
- ...fühlte ich mich von den anderen in der Klasse verstanden.
- ...hatte ich das Gefühl, dass mir die anderen in der Klasse helfen würden, wenn es nötig wäre.
- ...wurde ich von den anderen in der Klasse als Kollege/Kollegin behandelt.

Skalen zur Beurteilung der implementierungstreuen Umsetzung der Selbstbestimmungstheorie (LehrerInnen)

Beurteilung der Autonomie

Im Projekt der Reflect-Supervisionsphase

- ...konnten die SchülerInnen unter verschiedenen Themen auswählen.
- ...konnten die SchülerInnen entscheiden, ob sie allein oder in Gruppen arbeiten.
- ...konnten die SchülerInnen entscheiden, wie sie ein Thema behandeln (z.B. Lehrbuch, Video, Gruppendiskussion oder LehrerInnenvortrag).
- ...konnten die Schülerinnen mitentscheiden, wann und wie lange sie sich mit einer bestimmten Aufgabe beschäftigen.
- ...konnten die SchülerInnen zwischen unterschiedlich schwierigen Aufgaben wählen.

Bezugsnorm (Aspekt von Kompetenzerleben)

Ich habe den Erfolg/Misserfolg eines/einer SchülerIn im Projekt daran gemessen, ob...

- ...die Leistung des Schülers/der Schülerin besser/schlechter ist als die der anderen SchülerInnen.
- ...die Leistung des Schülers/der Schülerin besser/schlechter ist als die vorherige Leistung desselben Schülers.
- ...der Schüler/die Schülerin ein vorher festgesetztes Ziel erreicht hat oder nicht.

Beurteilung der Fürsorglichkeit (Aspekt von sozialer Eingebundenheit)

In meinem Unterricht...

- Wenn ein/e SchülerIn ein persönliches Anliegen hat, gehe ich auch im Unterricht darauf ein.
- Ich bespreche mit den SchülerInnen allgemeine oder aktuelle Themen, auch wenn der Fachunterricht dann ein wenig zurücktritt.
- Die persönlichen Beziehungen zu meinen SchülerInnen haben Vorrang vor schnellem Vorankommen im Stoff.
- Für persönliche und soziale Angelegenheiten nehme ich mir auch im Unterricht Zeit.

Beurteilungsbogen des ExpertInnenratings

2. CORE FRAMEWORK I: Self Determination Theory	Projektkonzept	Bsp.Einheit (1. UE): 0/1 Punkte	Alle Einheiten	Materialien/ Produkte der S.	Reflexion	Logbuch	sonstiges:	Summe:
<p>Wird durch die Konzeption des Projekts das Kompetenzerleben der SchülerInnen gefördert?</p> <p><i>z.B.: Die SchülerInnen haben Möglichkeiten sich selbstwirksam zu erleben; die LehrerInnen geben konstruktives Feedback auf SchülerInnenarbeiten und ermöglichen Verbesserungen; individuelle Lernfortschritte von SchülerInnen werden honoriert;</i></p>								
<p>Wird durch die Konzeption des Projekts die Autonomie der SchülerInnen gefördert?</p> <p><i>z.B.: Den SchülerInnen stehen Wahlangebote und Gestaltungsfreiräume zur Verfügung (Themen, Aufgaben, Arbeitsformen, Präsentationsformen, Bearbeitungszeit etc);</i></p>								
<p>Wird durch die Gestaltung der Einheiten die soziale Eingebundenheit bzw. soziale Zugehörigkeit der SchülerInnen gefördert?</p> <p><i>z.B.: Es werden gemeinschaftliche Aktivitäten initiiert, die über den herkömmlichen Unterricht hinausgehen. Auf persönliche Anliegen der SchülerInnen wird eingegangen. Kooperative Lernformen werden eingesetzt. Reflexions- und Feedbackmöglichkeiten werden angeboten. Die individuellen Fähigkeiten von SchülerInnen werden anerkannt.</i></p>								

II. Lebenslauf

CURRICULUM VITAE

Persönliche Angaben

Name: Michaela Pichler, BA
Email-Adresse: michaela.pichler@univie.ac.at

Ausbildung

10/2007- : **Universität Wien**

10/2008- : Studium der Psychologie

Spezialisierungen: Bildungspsychologie, Evaluation und Training; Angewandte Kinder- und Jugendpsychologie

10/2007-03/2012: Studium der Theater-, Film- und Medienwissenschaften, abgeschlossen am 22.03.2012

09/2010-06/2012: **PH Linz**

09/2011-06/2012: Theaterpädagogik Lehrgang II

09/2010-06/2011: Theaterpädagogik Lehrgang I

09/1999-06/2007: **Khevenhüllergymnasium Linz**

Vertiefung: Englisch und Psychologie; Abschlusschwerpunkte: Englisch, Musik und Physik

Berufliche Erfahrung

10/2012-: **Forschungsprojekt Pro-SRL, Universität Wien**
studentische Mitarbeiterin

01/2013-03/2013: **Universität Wien, Institut für Angewandte Psychologie, Arbeitsbereich „Bildungspsychologie und Evaluation“**
psychologisches Pflichtpraktikum

10/2011-10/2012: **Caritas Oberösterreich, St. Isidor bei Leonding**

regelmäßige freizeitpädagogische Betreuung von Kindern mit
Beeinträchtigung

02/2011-10/2012: **Kinderfreunde Wien, Bereich Veranstaltungsservice**
Animation und Betreuung von Kindern bei Veranstaltungen

02/2009-10/2011: **Komödie am Kai**
Ticketservice

07/2008-09/2008: **Linz2009 Kulturhauptstadt Europas, Projektentwicklung**
Praktikum

weitere Kenntnisse und Fähigkeiten

Sprachkenntnisse: Deutsch (Muttersprache)
 Englisch (ausgezeichnet)
 Spanisch (Maturaniveau)
 Schwedisch (Grundkenntnisse)

EDV-Kenntnisse: MS-Office, SPSS, MaxQDA

„Hiermit bestätige ich, dass die vorliegende Arbeit in allen relevanten Teilen selbstständig durchgeführt wurde.“

Wien, 11.02.2014

Michaela Pichler