

***Wirksamkeit präventiver Interventionen zur sozio-emotionalen  
Förderung bei Kindern und Jugendlichen mit  
Migrationshintergrund:  
Eine multinationale Meta-Analyse***

**Dissertation**

**zur Erlangung des akademischen Grades**

**doctor philosophiae (Dr. phil.)**

vorgelegt dem Rat der Fakultät für Sozial- und Verhaltenswissenschaften

der Friedrich-Schiller-Universität Jena

von Dipl.-Psych. Sebastian Schulz

geboren am 16.10.1985 in Lauchhammer

### **Gutachter**

1. Prof. Dr. Andreas Beelmann
2. Prof. Dr. Daniel Haun
3. Prof. Dr. Peter Noack

**Tag der mündlichen Prüfung: 10.06.2016**

## INHALTSVERZEICHNIS

Tabellenverzeichnis .....	4
Abbildungsverzeichnis .....	5
1. Einleitung .....	6
2. Theoretischer Hintergrund .....	8
2.1 Migration und Migrationshintergrund .....	8
2.2 Migrationshintergrund als Risikofaktor .....	11
2.3 Prävention im Kindes- und Jugendalter .....	17
2.4 Prävention bei Kindern und Jugendlichen mit Migrationshintergrund .....	23
2.5 Ziele der Arbeit .....	31
3. Methoden .....	33
3.1 Forschungssynthese mittels Meta-Analyse .....	33
3.2 Selektionskriterien .....	34
3.3 Suchstrategie zum Sichten relevanter Primärstudien .....	38
3.4 Suchergebnisse .....	41
3.5 Kodierung relevanter Primärstudien .....	42
3.5.1 Dokumentenebene .....	44
3.5.2 Vergleichsebene .....	44
3.5.3 Effektstärkenebene .....	46
3.6 Statistische Auswertung .....	47
3.6.1 Bestimmung der Roheffektstärken .....	47
3.6.2 Integration der Roheffektstärken .....	52
3.6.3 Gewichtung der Effektstärken und Schätzung der Populationseffektstärke .....	54
3.6.4 Moderatoranalysen .....	58
3.6.5 Publikationsverzerrung .....	59
4. Ergebnisse .....	64
4.1 Deskriptive Auswertungen .....	64
4.1.1 Forschungsberichte .....	64
4.1.2 Vergleiche .....	65
4.2 Allgemeine Effektivität .....	73
4.3 Heterogenitätsanalyse .....	74

4.4	Stabilität der Effekte über die Zeit .....	75
4.5	Wirksamkeit nach Erfolgskriterien .....	77
4.6	Publikationsverzerrung.....	80
4.7	Moderatoranalysen .....	81
4.7.1	Publikationsbezogene Merkmale.....	81
4.7.2	Methodische Merkmale .....	83
4.7.3	Merkmale der Präventionsprogramme .....	86
4.7.4	Stichprobenmerkmale.....	94
4.7.5	Weiterführende Analysen .....	96
5.	Diskussion.....	99
5.1	Zusammenfassende Betrachtung der Ergebnisse .....	99
5.1.1	Allgemeine Wirksamkeit der präventiven Interventionen.....	100
5.1.2	Stabilität der Effekte über die Zeit .....	103
5.1.3	Differenzielle Wirksamkeit hinsichtlich verschiedener Erfolgskriterien.....	103
5.1.4	Publikationsbezogene Merkmale.....	104
5.1.5	Methodische Merkmale .....	105
5.1.6	Merkmale der präventiven Interventionen .....	108
5.1.7	Merkmale der Stichprobe .....	112
5.2	Einschränkungen der Arbeit.....	114
5.3	Implikationen für die Praxis .....	116
5.4	Weiterer Forschungsbedarf .....	118
6.	Zusammenfassung .....	120
7.	Literaturverzeichnis .....	122
	Anhänge.....	150
	Anhang A – Suchbefehle für die Online-Suche in den einzelnen Datenbanken .....	150
	Anhang B – Kodierschema.....	163
	Ehrenwörtliche Erklärung.....	215
	Danksagung .....	216

## TABELLENVERZEICHNIS

Tabelle 1	Beschreibung der 68 Forschungsberichte in absoluten und relativen Häufigkeiten.....	65
Tabelle 2	Beschreibung der 93 Vergleiche in absoluten und relativen Häufigkeiten.....	69
Tabelle 3	Effektstärken für die verschiedenen Zeiträume in Abhängigkeit von für den jeweiligen Vergleich vorhandenen Messzeiträumen.....	76
Tabelle 4	Wirksamkeit der präventiven Interventionen nach verschiedenen Erfolgskriterien auf Effektstärken- und Vergleichsebene .....	79
Tabelle 5	Moderatoranalysen für publikationsbezogene Merkmale.....	82
Tabelle 6	Moderatoranalysen für methodische Merkmale .....	84
Tabelle 7	Wirksamkeit in Abhängigkeit von verschiedenen Programminhalten .....	87
Tabelle 8	Wirksamkeit in Abhängigkeit von verschiedenen sozio-emotionalen Programmzielen .....	89
Tabelle 9	Moderatoranalysen für Implementationsvariablen .....	91
Tabelle 10	Moderatoranalysen für weitere Programmmerkmale .....	92
Tabelle 11	Moderatoranalysen für Stichprobenmerkmale.....	95
Tabelle 12	Meta-Regressionen für weiterführende Fragestellungen.....	97

**ABBILDUNGSVERZEICHNIS**

Abbildung 1	Selektionsprozess zur Identifikation relevanter Forschungsberichte .....	43
Abbildung 2	Darstellung des Fixed-Effects-Models im Vergleich zum Random-Effects-Model .....	57
Abbildung 3	Stabilität der gewichteten mittleren Effektstärken ( $d_+$ ) über die Zeit (mit 95 %-Konfidenzintervallen) .....	77

## 1. EINLEITUNG

Ob Thilo Sarrazins Buch „Deutschland schafft sich ab“, medial geschürte Ängste vor Armutsmigration, das Unwort des Jahres 2013 „Sozialtourismus“ oder die aktuelle „Flüchtlingskrise“ - das Thema Migration ist im gesellschaftlichen Diskurs seit längerem präsent wie kaum ein anderes. In der Bildungsforschung sorgte die PISA-Studie für Aufsehen mit der empirisch fundierten Schlussfolgerung, Schüler mit Migrationshintergrund schneiden in der Schule in fast allen Ländern schlechter ab, als Schüler ohne Migrationshintergrund (OECD, 2007, 2011). Doch auch wenn die Bildung von Migranten verstärkt im Fokus der öffentlichen Wahrnehmung steht, so gibt es ebenso klare Risiken in Bezug auf die psychische Gesundheit und soziale Entwicklung, die mit einem Migrationshintergrund einhergehen. Betrachtet man den Anstieg rechtsextremistischer Straftaten seit der „Flüchtlingskrise“ in Deutschland, die Grenzschießung in Ungarn oder die immer wieder aufkommende Rassismusdebatte in den USA, so kann in vielen Kontexten von einer „Willkommenskultur“ keine Rede sein. Doch selbst wenn sich Personen mit Migrationshintergrund nicht direkt mit Vorurteilen und Diskriminierung konfrontiert sehen, so existieren dennoch verschiedene psychische und soziale Belastungsfaktoren, die dem Migrationsprozess inhärent sind. Dieser beinhaltet nicht nur das Verlassen einer gewohnten Umgebung mit einem häufig damit einhergehendem Wegfall des sozialen Netzes, sondern auch die Notwendigkeit der Integration in bzw. Anpassung an eine neue Gesellschaft, deren Sprache und Kultur. Kinder von Migranten wachsen zudem häufig im Spannungsfeld zweier Kulturen auf, was sowohl Einfluss auf ihre Identitätsentwicklung als auch auf ihre Beziehung zu den Eltern haben kann. Dass Kinder und Jugendliche mit Migrationshintergrund daher einem erhöhten Risiko im Hinblick auf ihre soziale und emotionale Entwicklung ausgesetzt sind, ist nicht neu und wurde bereits in vielen Studien bestätigt (Aronowitz, 1984). Dementsprechend gibt es international mittlerweile auch viele Präventionsprogramme, die darauf abzielen, diesen negativen Folgen von Migration vorzubeugen und davon betroffene Kinder und Jugendliche sozio-emotional zu fördern. Deren Wirksamkeit wurde ebenfalls bereits in zahlreichen Untersuchungen überprüft. Jedoch fehlt bisher eine systematische Aufbereitung dieser Befunde, die es ermöglichen würde, zu beurteilen ob, wie und unter welchen Bedingungen welche Programme am erfolgversprechendsten sind. Einige Arbeiten fassten bereits einen Teil der Befunde zu dieser Fragestellung zusammen, jedoch

schlossen sie in der Regel entweder nur bestimmte Migrantengruppen (z. B. Lateinamerikaner in den USA, siehe z. B. Castro et al.; 2006) oder eng umrissene Programmtypen ein (z. B. Kong, Singh & Krishnan-Sarin, 2012). Diese Lücke soll mit der vorliegenden Arbeit geschlossen werden. Ziel ist es demnach, eine Meta-Analyse durchzuführen, welche die existierenden Befunde zur Wirksamkeit präventiver Interventionen zur sozio-emotionalen Förderung bei Kindern und Jugendlichen mit Migrationshintergrund umfassend integriert.



## 2. THEORETISCHER HINTERGRUND

### 2.1 Migration und Migrationshintergrund

Migration kann als eine „auf einen längerfristigen Aufenthalt angelegte räumliche Verlagerung des Lebensmittelpunktes“ (Oltmer, 2013, S. 31) definiert werden (siehe auch Treibel, 2008). Hierbei ist zu unterscheiden zwischen der Binnenmigration, welche lediglich innerhalb eines Landes stattfindet und der internationalen Migration, bei welcher Ländergrenzen überschritten werden (Treibel, 1999). Verursacht schon eine Binnenmigration häufig eine Herauslösung aus dem sozialen Umfeld (Aksel, Gün, Irmak & Cengelci, 2007; Pyakuryal, Tausig, S. Subedi & J. Subedi, 2011), so hat eine grenzüberschreitende Migration noch weitaus größere soziale, finanzielle und rechtliche Folgen für die Betroffenen (Berry, 1997). Die Auswirkungen einer solchen internationalen Migration können zwar sehr unterschiedlich sein, jedoch geht sie in der Regel mit einer Vielzahl von Stressoren und Belastungsfaktoren einher (Belhadj Kouider, Koglin & Petermann, 2014; Bhugra, 2004; Carta, Bernal, Hardoy & Haro-Abad, 2005; Quintana, Vogel & Ybarra, 1991). Berry (1997) spricht in diesem Zusammenhang von sogenanntem *Akkulturationsstress*; dieser ist definiert als eine durch den allgemeinen Prozess des interkulturellen Kontaktes und die hierbei notwendige Anpassungsleistung (*Akkulturation*) bedingte Reduktion des (psychischen, somatischen oder sozialen) Gesundheitszustands (Berry, 1997; Berry, Kim, Mindle & Mok, 1987). Berry (1997) unterscheidet dabei vier Akkulturationsstrategien. Er spricht von *Integration*, wenn die eingewanderte Person zugleich versucht ihre kulturelle Identität zu erhalten und außerdem häufig mit Angehörigen der Kultur des Einwanderungslandes zu interagieren. Ist die Person bestrebt ihre ursprüngliche kulturelle Identität zu erhalten, meidet jedoch Kontakt zu anderen Kulturen, so spricht man von *Segregation*. *Adaptation* bezeichnet eine Strategie, bei der versucht wird, mit den Angehörigen der Kultur des Einwanderungslandes zu interagieren, die eigene kulturelle Identität dabei jedoch verworfen wird. Wird sich schließlich sowohl von der ursprünglichen Kultur distanziert als auch die Interaktion mit anderen Kulturen vermieden, so handelt es sich um *Marginalisierung*. Migranten sind hierbei hinsichtlich der Wahl ihrer Akkulturationsstrategie keineswegs vollkommen frei. Stattdessen hat die Gesellschaft, in welche die Person einwandert, bedeutsamen Einfluss darauf, welche Akkulturationsstrategie gewählt wird bzw. welche

überhaupt gewählt werden kann. Von der Akkulturationsstrategie hängt wiederum auch ab, wie gut die Anpassung des Migranten gelingt, wobei Integration in der Regel mit besserer, Marginalisierung typischer Weise mit schlechterer Anpassung einhergeht (Berry, 1997; Berry & Sam, 1997).

Hierbei kann unterschieden werden zwischen *psychologischer Anpassung*, welche sich eher auf die persönliche und kulturelle Identität, mentale Gesundheit und persönliche Zufriedenheit bezieht und *soziokultureller Anpassung*, mit der vor allem die Fähigkeit mit täglichen Problemen in Lebensbereichen wie der Familie, im Beruf oder in der Schule umzugehen, gemeint ist (Berry, 1997; Searle & Ward, 1990). In Bezug auf den zeitlichen Verlauf der Anpassung von Migranten unterscheidet Sluzki (2010) fünf Stadien des Migrationsprozesses: (a) die Vorbereitungsphase, (b) den Migrationsakt, (c) die Phase der Überkompensierung, (d) die Phase der Dekompensation und (e) die Phase der generationsübergreifenden Anpassungsprozesse. Diesen Verlauf sieht Sluzki (2010) als universell und kulturübergreifend an. Kennzeichnend ist hierbei, dass die Anpassungsfähigkeit von Migranten in den ersten Wochen und Monaten nach der Ankunft (Phase der Überkompensation) zunächst hoch erscheint. In dieser Zeit treten Konflikte zwischen den Erwartungen der Migranten und der sie umgebenden Realität eher in den Hintergrund oder werden verleugnet. Erst in der darauffolgenden Phase der Dekompensation treten solche Konflikte offen zu Tage und die über die Zeit kumulierten Belastungen offenbaren ihre Wirkung. Es folgt ein Anpassungsprozess, welcher sich über mehrere Generationen fortsetzen kann.

Ein Großteil des entstehenden Akkulturationsstresses ist direkt oder indirekt auf mangelnde Kenntnisse der Sprache des Einwanderungslandes zurückzuführen (Gil, Wagner & Vega, 2000; Grünigen, Perren, Nägele & Alsaker, 2010; Titzmann, Silbereisen, Mesch & Schmitt-Rodermund, 2011; Yakhnich, 2008). Während Kinder der ersten Migrantengeneration sowie solche mit zumindest einem immigrierten Elternteil innerhalb der Familie noch vorwiegend die Sprache des Heimatlandes sprechen, spricht bereits die dritte Generation, deren Eltern in der Regel vollständig die Schulbildung des Einwanderungslandes durchlaufen haben, zu Hause hauptsächlich die Sprache des Einwanderungslandes (Alba, 1999). Dies spricht dafür, dass der durch mangelnde Sprachkenntnisse auftretende Akkulturationsstress in der Regel für die ersten beiden Generationen nach der Einwanderung größer sein sollte. Auch können Migrantenkinder der ersten und zweiten Migrantengeneration als besonders gefährdet angesehen werden,

da sie zugleich durch ihre Ursprungsfamilie, aber auch durch die Kultur des Landes, in dem sie leben, sozialisiert werden, was zu Generationskonflikten und vor allem im Jugendalter zu Problemen in der Identitätsentwicklung führen kann (Belhadj Kouider et al., 2014; Jonas, Stroebe & Hewstone, 2014). Es kann folglich davon ausgegangen werden, dass insbesondere die ersten beiden Generationen von den Folgen der Migration betroffen sind. Auch wenn es in der Literatur bisher keinen Konsens über die angemessenste Operationalisierung eines *Migrationshintergrundes* gibt (Kemper, 2010; Sattelmeyer & Erbe, 2010), soll ein solcher daher in dieser Arbeit definiert werden als die Abstammung von mindestens einem Elternteil, welches in einem anderen Land als dem gegenwärtigen Wohnort der Familie geboren wurde (Nauck, Clauß & Richter, 2008; Schenk et al., 2008).

Eng mit dem Begriff des Migrationshintergrundes verknüpft ist das Konzept der *ethnischen Minorität*. Der Begriff *Ethnizität* bezieht sich hierbei auf eine Gruppe, die Merkmale wie „Sprache, eine gemeinsame Geschichte oder Tradition, eine gemeinsame Abstammung oder eine gemeinsame geografische Herkunft“ (Europäische Kommission, 2010; S. 23) miteinander teilt. Minorität kann allgemein verstanden werden als distinkte, gegenüber einer Majorität kleinere Gruppe (Erb & Bohner, 2006; Fuchs-Heinritz et al., 2011). Eine ethnische Minorität unterscheidet sich folglich durch Merkmale der Ethnizität von der Mehrheitsgesellschaft. Dieser Unterschied kann durch Migration, jedoch auch durch fortgesetzte Segregation bedingt sein. Entscheidend ist, dass sich die Mitglieder einer ethnischen Minderheit selbst als verschieden von anderen Menschen in der Gesellschaft wahrnehmen, aber auch, dass die Gruppe von anderen Menschen ebenfalls als verschieden wahrgenommen wird (Bös, 2008). Dies bedeutet wiederum auch, dass die Kategorisierung verschiedener ethnischer Minoritäten nicht nur abhängig von der Herkunft und den Merkmalen der Minderheitengruppe selbst ist, sondern auch von der Gesellschaft, in der sie sich befindet. So ist in den USA beispielsweise die Kategorisierung *Hispanics*, welche die spanischsprachige Bevölkerung sowohl aus Lateinamerika als auch aus Europa bezeichnet, sehr gebräuchlich. In Deutschland ist eine Unterscheidung in Spanier und Lateinamerikaner hingegen üblich. Einige ethnische Minoritäten umfassen vorwiegend Personen mit Migrationshintergrund (z. B. *Latino/Hispanics* oder *Asian Americans*, siehe Grieco, 2010; Lee, 1998), andere jedoch überwiegend Personen ohne Migrationshintergrund (z. B. *African Americans* oder *Native Americans*). Beide Gruppen nehmen sich, entsprechend der von Bös (2008) formulierten

Definition, selbst hinsichtlich ihrer Ethnizität als verschieden von der Mehrheitsgesellschaft wahr und werden von dieser ebenfalls als unterschiedlich angesehen. Wodurch dieser Unterschied jedoch bedingt ist, variiert. In der ersten Gruppe sind diese Unterschiede vermutlich vor allem durch die aus dem Emigrationsland mitgebrachte Kultur bedingt (z. B. Sprache, Tradition, Religion). Im zweiten Fall können verschiedenste Faktoren eine Rolle spielen, wie äußere Merkmale (z. B. Hautfarbe) oder eine Kultur, die aufgrund von Aus- oder Abgrenzung lange Zeit neben der Mehrheitsgesellschaft besteht, ohne dass eine Integration oder Anpassung erfolgte.

Da die Sprache zum einen ein Definitionskriterium ethnischer Minderheiten darstellt, zum anderen mangelnde Kenntnisse der im Einwanderungsland hauptsächlich gesprochenen Sprache ein Hauptstressor bei Personen mit Migrationshintergrund darstellt (Titzmann et al., 2011; Yakhnich, 2008), werden Minderheiten häufig anhand ihrer Muttersprache kategorisiert (z. B. Spanisch sprechende Bevölkerung in den USA oder Türkisch sprechende Bevölkerung in Deutschland; Sáenz, L. S. Fuchs & D. Fuchs, 2005; Stanat, Becker, Baumert, Lüdtke & Eckhardt, 2012, Vadasy & Sanders, 2010). Gruppen mit mangelnden Sprachkenntnissen hinsichtlich der im Einwanderungsland hauptsächlich gesprochenen Sprachen bestehen bis auf wenige Ausnahmen (vor allem indigene Gruppen) in der Regel mehrheitlich aus Personen mit Migrationshintergrund (Capps et al., 2005).

Zusammenfassend kann festgestellt werden, dass ein Migrationshintergrund durch eine internationale Wanderungsbewegung einer Person oder deren Eltern gekennzeichnet ist und trotz verschiedener Migrationsbedingungen sowie kultureller Besonderheiten häufig mit einem ähnlichen Muster von Belastungen und Stressoren einhergeht. Zudem ist er zwar eng mit der Sprache einer Person sowie deren Ethnizität verbunden, kann jedoch nicht ausschließlich anhand dieser Merkmale definiert werden.

## **2.2 Migrationshintergrund als Risikofaktor**

Wie bereits beschrieben, geht eine internationale Migration mit einer Vielzahl unterschiedlicher Belastungen und Stressoren sowie einem Wegfall bestimmter Ressourcen einher. Das Ausmaß der unmittelbaren Migrationsfolgen und deren Auswirkungen auf Kinder und Jugendliche mit Migrationshintergrund können jedoch je nach Einwanderungsland, Migrantengruppe und individuellen Faktoren sehr unter-

schiedlich sein (Berry & Sam, 1997). So unterscheidet Berry (1997) vier verschiedene Akkulturationsstrategien, die einen Einfluss darauf haben, ob eine positive Anpassung an die neue Umgebung gelingt oder nicht. Welche dieser Akkulturationsstrategien gewählt wird bzw. gewählt werden kann, ist wiederum nicht nur von individuellen Präferenzen der betreffenden Person abhängig, sondern ebenso von der Gesellschaft und Politik des Einwanderungslandes sowie den Akkulturationspräferenzen der Mitglieder der eigenen Kultur in diesem Land. Hovey (2000) konnte dabei zeigen, dass die Höhe des Akkulturationsstresses unter anderem vom Einkommen im Einwanderungsland abhängt sowie davon, ob eine geografische Trennung von der Familie existiert. Zudem fanden Belhadj Kouider et al. (2014) in ihrem systematischen Review zur psychischen Gesundheit von Kindern und Jugendlichen mit Migrationshintergrund zwar konsistente Befunde hinsichtlich einer höheren Prävalenz internalisierender Probleme, allerdings waren die Befunde bezüglich externalisierender Probleme weniger eindeutig und es konnte demnach nicht übereinstimmend gezeigt werden, dass ein Migrationshintergrund häufiger mit derartigen Problemen einhergeht. Sam, Vedder, Ward und Horenczyk (2006) konnten in einer multinationalen Studie zeigen, dass Jugendliche mit Migrationshintergrund verglichen mit einheimischen Jugendlichen aus denselben Nachbarschaften und Schulen eine genauso gute psychologische Anpassung und zum Teil sogar eine bessere soziokulturelle Anpassung aufweisen können.

Andererseits geht eine Migration nahezu zwangsläufig mit einem zumindest teilweisen Verlust des sozialen Netzwerks einher (Aksel et al., 2007; Carta et al., 2005; Steinbach & Nauck, 2004). Auch haben Personen mit Migrationshintergrund ein erhöhtes Risiko, die Sprache sowie die Struktur von Institutionen wie etwa dem Bildungssystem des Einwanderungslandes nur unzureichend zu kennen und ein geringeres Einkommen zu erzielen (Capps et al., 2005; Grabka & Frick, 2010; U.S. Census Bureau, 2011, Table 713). Diese Faktoren stellen wiederum nicht nur ein Risiko für die bildungsbezogene Entwicklung von Kindern und Jugendlichen dar, sondern sind auch mit sozialen und emotionalen Defiziten assoziiert (Arnold & Doctoroff, 2003; Bradley & Corwyn, 2002; Chu, Saucier & Hafner, 2010; Herbert-Myers, Guttentag, Swank, Smith & Landry, 2006; Vedder, 2005). Auf theoretischer Ebene unterscheidet Bourdieu (1983) in seinem Modell zur Aufrechterhaltung sozialer Ungleichheit drei Wege, über die der sozioökonomische Status der Eltern an die Kinder weitergegeben werden kann: *ökonomisches Kapital* (z.B. Geld und Eigentum), *kulturelles Kapital* (z.B. Wissen und Bildung) und *soziales Kapital*

(z.B. soziale Unterstützung und Anerkennung). Diese sind nach der Theorie von Bourdieu (1983) nicht nur ineinander überführbar, sondern bestimmen zugleich auch die Bildungs- und Entwicklungschancen der Kinder. Folgt man den oben genannten Argumenten, kann geschlussfolgert werden, dass Personen mit Migrationshintergrund in allen drei Kapitalformen weniger Ressourcen zur Verfügung haben und folglich auch ein höheres Risiko für Bildungs- und Entwicklungsprobleme aufweisen sollten. Für Migrantenkinder können auch durch schwach ausgeprägte Kenntnisse der Sprache des Einwanderungslandes die Interaktion mit Gleichaltrigen sowie der Erwerb sozialer Kompetenzen erschwert sein. Ebenso besteht aufgrund des ethnischen Minderheitenstatus ein erhöhtes Risiko von Diskriminierung. Die Migration selbst sowie der damit einhergehende Wechsel des kulturellen Kontextes kann traumatischen Stress erzeugen und zudem Konflikte innerhalb der Familie, insbesondere zwischen den Generationen, hervorrufen und stellt auch zusätzliche Anforderungen an die Identitätsentwicklung jugendlicher Migranten (Belhadj Kouider et al., 2014; Sluzki, 1979). All diese Faktoren können für Kinder und Jugendliche mit Migrationshintergrund ebenfalls ein Risiko hinsichtlich ihrer sozio-emotionalen Entwicklung darstellen.

Nach der Definition von Kraemer et al. (1997) handelt es sich bei einem Merkmal bereits dann um einen Risikofaktor, wenn nachgewiesen werden kann, dass es mit einem bestimmten negativen Ereignis oder Verhalten korreliert und diesem vorausgeht, unabhängig davon, ob es selbst ein direkter kausaler Wirkfaktor ist oder ob es verändert werden kann. Im Folgenden soll anhand von Untersuchungen aus verschiedenen Ländern gezeigt werden, dass es sich bei einem Migrationshintergrund um einen Risikofaktor nach dieser Definition handelt und mit welchen konkreten Bildungs- und Entwicklungsrisiken ein solcher demnach verbunden ist. Dabei soll der Schwerpunkt auf den sozialen und emotionalen Folgen der Migration liegen, jedoch auch auf Risiken in Bezug auf das häusliche Milieu sowie die kognitive und bildungsbezogene Entwicklung eingegangen werden, da es zwischen diesen Parametern beträchtliche Zusammenhänge gibt (Aronowitz, 1984; Belhadj Kouider et al., 2014; Manongdo & Ramírez García, 2007; McKee, Colletti, Rakow, Jones & Forehand, 2008; Moilanen, Shaw & Maxwell, 2010; Stevens, Vollebergh, Pels & Crijnen, 2007).

Hinsichtlich der bildungsbezogenen Entwicklung zeigte die PISA-Studie über nahezu alle Länder hinweg, dass Schüler der ersten genauso wie die der zweiten Migrantengeneration im Vergleich zu einheimischen Schülern nicht nur schlechtere Leseleistungen,

sondern auch schlechtere Mathematikleistungen aufweisen (OECD, 2007, 2011). In Studien aus den USA konnten ähnliche Unterschiede für die durchschnittlichen Schulnoten sowie für den Bildungsweg aufgezeigt werden (Kuperminc, Blatt, Shahar, Henrich & Leadbeater, 2004; National Science Foundation, Division of Science Resources Statistics, 2003; Pew Hispanic Center, 2004). Auch in den Niederlanden brechen ethnische Minderheiten häufiger die Schule ab als einheimische (Bosker, 1994), was sich letztlich auch im beruflichen Werdegang von Personen mit Migrationshintergrund widerspiegelt (De Vries & Wolbers, 2004). Eine Studie aus Deutschland konnte wiederum belegen, dass bereits beim Schulbeginn die kognitiven Fähigkeiten von Schülern mit Migrationshintergrund geringer waren als jene von Schülern ohne Migrationshintergrund (Becker & Biedinger, 2006). Es kann demnach davon ausgegangen werden, dass Kinder und Jugendliche mit Migrationshintergrund ein erhöhtes Entwicklungsrisiko für kognitive und bildungsbezogene Defizite aufweisen.

Auch in Bezug auf ihr häusliches Milieu sind Kinder und Jugendliche mit Migrationshintergrund häufig schlechteren Bedingungen ausgesetzt als gleichaltrige Einheimische. Aufgrund des häufig niedrigeren sozioökonomischen Status und der mangelnden sozialen und kulturellen Ressourcen der Eltern sowie der durch die Migration bedingten Belastungen, die auf das Familiensystem einwirken, kann einer destruktiven, wenig förderlichen Erziehung sowie einer schlechten Eltern-Kind-Beziehung Vorschub geleistet werden (Belhadj Kouider et al., 2014, Sluzki, 1979). So konnte gezeigt werden, dass lateinamerikanische Kinder in den USA zu Hause weniger kognitiv gefördert werden und ihre Eltern eine geringere Responsivität aufweisen als jene von nichtlateinamerikanischen Kindern (Bradley, Corwyn, McAdoo & Coll, 2011). Kinder und Jugendliche mit Migrationshintergrund in der Schweiz sind zudem stärkerer elterlicher Kontrolle ausgesetzt und erleben weniger Wärme und Akzeptanz von ihren Müttern sowie mehr Zurückweisung durch ihre Eltern als Kinder und Jugendliche ohne Migrationshintergrund (Steinhausen, Bearth-Carrari & Metzke, 2009). Auch in Israel erleben jugendliche Migranten aus dem Gebiet der ehemaligen Sowjetunion weniger positive Erziehung (Dwairy & Dor, 2009; Finzi-Dottan, Bilu & Golubchik, 2011). Ein weiterer möglicher Stressor kann emotionale Parentifizierung sein, welche beispielsweise bei jugendlichen Migranten aus dem Gebiet der ehemaligen Sowjetunion in Deutschland häufiger vorkommt als bei einheimischen (Titzmann, 2012). In diesem Zusammenhang ist der Begriff der *Akkulturationslücke* von Bedeutung, welche dadurch entsteht, dass sich

Kinder und Jugendliche schneller an eine neue Sprache und Kultur anpassen, als ihre Eltern (Martinez, 2006; Szapocznik & Kurtines, 1993). Diese geht nicht nur mit einer ineffizienteren Erziehung und einer damit verbundenen erhöhten Wahrscheinlichkeit von jugendlichem Substanzmissbrauch einher, sondern führt häufig dazu, dass die Jugendlichen für ihre Eltern übersetzen (Martinez, 2006; Martinez, McClure & Eddy, 2009; Tse, 1995). Es konnte gezeigt werden, dass gerade für jüngere Jugendliche diese Übersetzungsaufgaben, welche sie für ihre Eltern übernehmen, mit mehr internalisierenden Problemen einhergehen (Schulz, Titzmann & Michel, 2013). Folglich scheinen Kinder und Jugendliche mit Migrationshintergrund auch hinsichtlich ihres häuslichen Milieus ungünstigeren Entwicklungsbedingungen ausgesetzt zu sein.

Bezüglich der sozialen und emotionalen Entwicklung liegen schließlich ebenfalls Studien aus verschiedenen Ländern vor, die auf ein erhöhtes Risiko für Kinder und Jugendliche mit Migrationshintergrund hindeuten. Hierzu führte bereits Aronowitz (1984) in einer frühen Übersichtsarbeit aus, dass Migrantenkinder häufig mit einer Vielzahl unterschiedlicher sozialer und emotionaler Anpassungsprobleme zu kämpfen haben. Quintana et al. (1991) konnten in einer Meta-Analyse zeigen, dass Collegestudenten lateinamerikanischer Herkunft in den USA in verschiedenen Lebensbereichen höhere Stresslevel aufwiesen als Collegestudenten nichtlateinamerikanischer Herkunft. In ähnlicher Weise wurden auch bei kolumbianischen Migranten in Spanien höhere Stresswerte nachgewiesen (Patiño & Kirchner, 2009). Probleme in der Identitätsentwicklung sind bei jugendlichen Migranten, aufgrund des Aufwachsens in zwei Kulturen, ebenfalls nicht selten und können dann zu weiteren internalisierenden Problemen führen (Belhadj Kouider et al., 2014). So konnte bei Lateinamerikanern in den USA genauso wie in Israel für jugendliche Migranten aus dem Gebiet der ehemaligen Sowjetunion eine höhere Prävalenz von Identitätskonflikten nachgewiesen werden (Huynh, Devos & Smalarz, 2011; Finzi-Dottan et al., 2011). Auch zeigte sich im Vergleich zur einheimischen ethnischen Majorität ein niedrigerer Selbstwert für Lateinamerikaner in den USA sowie für Kinder und Jugendliche mit Migrationshintergrund in der Schweiz (Grossman, Wirt & Davids, 1985; Steinhausen et al., 2009). Letztere weisen zudem eine stärkere Präferenz für vermeidende Stressbewältigungsstrategien auf als gleichaltrige Einheimische (Steinhausen et al., 2009). Weiterhin gibt es Hinweise darauf, dass Personen mit Migrationshintergrund in Deutschland verstärkt Angstsymptome aufweisen (Merbach, Wittig & Brähler, 2008). In Schweden geht ein



Migrationshintergrund mit mehr Angst vor Viktimisierung einher (Martens, 2000). Kinder mit koreanischem Migrationshintergrund in den USA neigen zudem stärker zur Zurückgezogenheit als gleichaltrige ohne Migrationshintergrund (Jung & Stinnett, 2005). Auch finden sich Belege dafür, dass in den USA Kinder und Jugendliche der ersten bis dritten Migrantengeneration ein niedrigeres Wohlbefinden sowie mehr depressive Symptome aufweisen und mit höherer Wahrscheinlichkeit einen Selbstmordversuch unternehmen (Centers for Disease Control and Prevention, 2006; Knight, Viridin, Ocampo & Roosa, 1994; Maynard & Harding, 2010; Vega, Zimmerman, Warheit, Apospori & Gil, 1993; Wang, 2008). Diese Befunde sprechen allgemein für ein erhöhtes Risiko internalisierender Probleme, welches für Kinder und Jugendliche mit Migrationshintergrund in der Schweiz und für lateinamerikanische Jugendliche in den USA auch nachgewiesen werden konnte (Kuperminc et al., 2004; Steinhausen et al., 2009) und in einem europaweiten Review ebenfalls bestätigt wurde (Belhadj Kouider et al., 2014).

Auch wenn Belhadj Kouider et al. (2014) darauf hinwiesen, dass die Befundlage in diesem Bereich für Europa nicht eindeutig ist, so gibt es doch eine Vielzahl von Studien, die ein erhöhtes Risiko für externalisierende Probleme bei Kindern und Jugendlichen mit Migrationshintergrund nahelegen (Steinhausen et al., 2009). Lateinamerikanische Jugendliche in den USA zeigen mehr delinquentes Verhalten und ein weniger gut angepasstes Verhalten als nichtlateinamerikanische Gleichaltrige (Grossman et al., 1985; Vega et al., 1993). Auch weisen Kinder mit Migrationshintergrund in der Schweiz mehr aggressives Verhalten auf (Steinhausen et al., 2009). In den USA haben hispanische Achtklässler höhere Prävalenzraten für den Konsum legaler und illegaler Drogen (Johnston, Malley, Bachman & Schulenberg, 2011). Dieser Trend zeigt sich jedoch nicht für andere Altersklassen und es muss dabei auch festgestellt werden, dass die erste Generation geringere Prävalenzraten aufweist, als in den USA geborene Hispanics (Prado et al., 2009). Lateinamerikanische Jugendliche haben außerdem häufiger Geschlechtsverkehr vor ihrem dreizehnten Lebensjahr und verwenden bei ihrer sexuellen Aktivität seltener Kondome als nichtlateinamerikanische Jugendliche (Centers for Disease Control and Prevention, 2006). Kinder und Jugendliche mit Migrationshintergrund in der Schweiz weisen zudem schlechtere Sozialbeziehungen zu ihren Klassenkameraden auf und berichten eine erhöhte Zahl negativer Lebensereignisse (Grünigen et al., 2010; Steinhausen et al., 2009). In der Schweiz und in Schweden erleben Personen mit

Migrationshintergrund darüber hinaus häufiger eine Viktimisierung (Grünigen et al., 2010; Martens, 2000).

Dass schließlich Personen mit Migrationshintergrund, wie durch multinationale Forschungsbefunde belegt, im Erwachsenenalter ein höheres Risiko für psychische Störungen wie Schizophrenie oder affektive Störungen aufweisen, zeigt, dass die schlechtere psychologische Anpassung von Personen mit Migrationshintergrund kein auf die Kindheit oder das Jugendalter beschränktes Phänomen ist (Cantor-Graae & Selten, 2005; Porter & Haslam, 2005; Swinnen & Selten, 2007).

Abschließend kann gesagt werden, dass Kinder und Jugendliche mit Migrationshintergrund hinsichtlich verschiedenster Kriterien einem erhöhten Entwicklungsrisiko ausgesetzt sind. Auch wenn die Befundlage nicht in allen Bereichen gleichermaßen eindeutig ist, kann doch allgemein von einer erhöhten Belastung für diese Gruppe ausgegangen werden, zumal die verschiedenen beschriebenen Parameter untereinander korreliert sind (Belhadj Kouider et al., 2014; Manongdo & Ramírez García, 2007; McKee et al., 2008; Moilanen et al., 2010; Reitz, Dekovic & Meijer, 2005; Stevens et al., 2007).

Wenn nun, wie gezeigt, ein Migrationshintergrund für Kinder und Jugendliche einen Risikofaktor darstellt, wirft dies die Frage auf, wie den genannten Entwicklungsrisiken präventiv oder kompensatorisch entgegengewirkt werden kann.

### **2.3 Prävention im Kindes- und Jugendalter**

Unter *präventiven Interventionen* kann der theoretisch und empirisch begründete Versuch verstanden werden, unerwünschten psychischen und physischen Störungen vorzubeugen bzw. diese zu verhindern (Beelmann & Raabe, 2007; Coie et al., 1993; Perrez, 1994, 1998). Von psychologischer Prävention spricht man dann, wenn hierfür psychologische Mittel verwendet werden. Psychologische Mittel sind dabei all jene, die im Erleben und Verhalten ansetzen und sich in der sozialen Interaktion zwischen dem Durchführenden der Intervention und den Teilnehmern vollziehen (Perrez & Baumann, 2011). Hinsichtlich der Klassifikation psychologischer Prävention lassen sich verschiedene Modelle unterscheiden. Im Folgenden sollen einige solcher Modelle Erwähnung finden, wobei aufgrund der Vielfältigkeit verschiedener Programme und der Möglichkeiten diese einzuteilen, die nachstehende Aufzählung unvollständig bleiben muss.

Eines der ersten solcher Klassifikationssysteme stammt von Caplan (1964). Er unterteilte Prävention je nachdem, ob die jeweilige Maßnahme vor, während oder nach Auftreten einer Störung beginnt, in *primäre, sekundäre* bzw. *tertiäre Prävention*. Die hierin enthaltene Präventionsdefinition ist jedoch sehr weit gefasst und ermöglicht es nicht, den Begriff trennscharf von anderen Interventionsformen zu unterscheiden. Aus diesem Grund wurde sie von vielen Autoren kritisiert, welche ihrerseits eine engere Definition vorschlugen (Gordon, 1983; Munoz, Mrazek & Haggerty, 1996; Perrez, 1998). Der Begriff Prävention soll demnach von den Begriffen *Therapie* und *Rehabilitation* abgegrenzt und nur für jene Interventionen genutzt werden, welche bereits ansetzen, bevor sich eine diagnostizierbare Störung manifestiert (Gordon, 1983; Munoz et al., 1996, Perrez, 1994). Munoz et al. (1996) empfehlen als Klassifizierungssystem die, von Gordon (1983) vorgeschlagene, Unterscheidung in *universelle, selektive* und *indizierte Prävention*. Universelle Prävention bezeichnet dabei Interventionen, welche für eine Gesamtpopulation vorgesehen sind, die kein erhöhtes Risiko für Verhaltensprobleme aufweist. Selektive Prävention wendet sich hingegen an eine Gruppe, welche zwar ein erhöhtes Risiko für Verhaltensprobleme bzw. eine Störung aufweist, jedoch bis dahin noch keine Auffälligkeit zeigt. Indizierte Prävention schließlich adressiert Personen, welche bereits erste Auffälligkeiten hinsichtlich des zu vermeidenden Problems bzw. der Störung zeigen, die jedoch noch nicht das Ausmaß einer diagnostizierbaren psychischen Störung erreichen (Gordon, 1983; Munoz et al., 1996). Eine ähnliche Kategorisierung wählten Offord, Kraemer, Kazdin, Jensen und Harrington (1998), indem sie zwischen *universeller* und *gezielter (targeted) Prävention* unterschieden. Letzteres kann als Oberkategorie verstanden werden, welche sowohl selektive als auch indizierte Präventionsmaßnahmen einschließt (Beelmann & Raabe, 2007). Universelle Präventionsprogramme haben den Vorteil, dass alle potentiell gefährdeten Personen angesprochen werden und die Eingriffsintensität in der Regel relativ gering ist. Allerdings ist es schwer sicherzustellen, dass tatsächlich alle Mitglieder einer Population erreicht werden und die große Zahl an Personen steigert zum einen die Kosten und führt zum anderen dazu, dass viele von denen, welche die Intervention bekommen, sie eigentlich gar nicht bräuchten, da sie sich ohnehin positiv entwickeln würden. Gezielte Präventionsstrategien ermöglichen spezifischere Hilfen, die auch auf die Bedürfnisse der Zielgruppe besser zugeschnitten sind. Sie erfordern allerdings differenzierte Erkenntnisse über Risikofaktoren und Entstehungsbedingungen der zu vermeidenden Probleme sowie zuverlässige

Screeningverfahren, um diese zu erkennen und bergen zudem das Risiko von Stigmatisierungseffekten (Beelmann & Raabe, 2007; Offord, 2000; Offord et al., 1998). Weiterhin können Programme bezüglich des Adressaten der Intervention kategorisiert werden. Perrez (1994) unterscheidet dabei Prävention bei Kindern und Jugendlichen, Programme zur Förderung der Sozialisationskompetenzen von Eltern sowie umgebungsbezogene Ziele.

Ein weiteres mögliches Unterscheidungsmerkmal ist die Spezifität einer Intervention. Soll eine bestimmte Störung verhindert werden, so spricht man von einer *spezifischen Prävention*. Daneben gibt es jedoch auch Interventionen mit dem Ziel einer generellen Verbesserung der biopsychosozialen Lebensbedingungen, auch als *unspezifische Prävention* bezeichnet (Perrez, 1998). Hinsichtlich der Zielstellung kann zudem zwischen präventiven Interventionen, die eine Reduktion von Risikofaktoren anstreben und solchen, die darauf abzielen, *protektive Faktoren* (Faktoren, welche die negativen Auswirkungen von Risikofaktoren abschwächen) zu fördern (Munoz et al., 1996, Perrez, 1998), unterschieden werden. Interventionen, welche auf Letzteres abzielen, werden häufig auch als *Gesundheitsförderung* bezeichnet, sollen aber im Folgenden mit unter den Begriff der Prävention subsumiert werden, da gerade bei unspezifischen, universellen Präventionsprogrammen eine trennscharfe Differenzierung kaum möglich ist. Perrez (1994) schlägt schließlich noch eine weitere Grobkategorisierung der Ziele präventiver Interventionen für Kinder und Jugendliche vor. Er unterscheidet dabei zwischen psychomotorischen und kognitiven Förderungs- und Präventionszielen, sozio-emotionalen Förderungs- und Präventionszielen sowie Förderungs- und Präventionszielen im Gesundheitsbereich. Erstere umfassen z.B. Sprach- und Leseförderung, Intelligenzförderung sowie kognitive Förderung nach Piaget. Sozio-emotionale Förderungs- und Präventionsziele betonen hingegen die emotionale und soziale Entwicklung, wobei beispielsweise Gefühle, Einstellungen und Werte aktiver wahrgenommen, akzeptiert und in ihrer Bedeutung für die soziale Interaktion erkannt oder soziale Kompetenzen trainiert werden sollen. Präventionsziele im Gesundheitsbereich fokussieren auf Körperwissen, -wahrnehmung und -bewertung sowie Gesundheitsverhalten wie z.B. die Verhütung von HIV-Infektionen. Die Bildung weiterer Unterkategorien hinsichtlich der genauen Zielstellung ist hierbei jeweils möglich.

Bezüglich der Wirksamkeit primärpräventiver Programme für Kinder und Jugendliche veröffentlichten Durlak und Wells (1997) eine Meta-Analyse, in der alle bis 1991

veröffentlichten kontrollierten Studien zum Thema eingeschlossen wurden. Ausgenommen waren dabei lediglich Interventionen, die sich ausschließlich auf schulleistungsbezogene Förderung oder Drogenprävention bezogen. Hierbei zeigten sich je nach Programmtyp Effektstärken zwischen  $d = 0.16$  und  $d = 0.93$ , wobei Elterntrainings die niedrigsten Effekte hatten und als einziger Programmtyp keine signifikante Wirksamkeit aufwiesen.

Beelmann (2006) gibt einen Überblick über den Stand der integrativen Forschung im Bereich der Wirksamkeit präventiver Interventionen für Kinder und Jugendliche. Aus seiner Übersichtsarbeit wird insbesondere deutlich, dass bereits eine Vielzahl von Studien und integrativen Arbeiten zur Prävention dissozialer bzw. externalisierender Verhaltensprobleme, aber auch zur Substanzmissbrauchsprävention existieren. In diesen Bereichen zeigten sich kleine bis mittlere Effekte. Lösel und Beelmann (2003) ermittelten beispielsweise in ihrer Meta-Analyse zur Wirksamkeit sozialer Trainingsprogramme zur Prävention dissozialen Verhaltens eine Effektstärke von  $d = 0.38$  als besten Schätzer. Tobler et al. (2000) berichteten in ihrer Meta-Analyse zur Suchtprävention je nach Suchtmittel eine gewichtete mittlere Effektstärke von 0.08 bis 0.16. Meta-Analysen existieren außerdem zur Wirksamkeit von Programmen zur Prävention jugendlicher Schwangerschaft sowie zu Elterntrainings, insbesondere bei bereits bestehenden Verhaltensproblemen des Kindes. Franklin, Grant, Corcoran, Miller und Bultman (1997) berechneten in ihrer Meta-Analyse zur Wirksamkeit von Programmen zur Vermeidung jugendlicher Schwangerschaften für sexuelle Aktivitäten eine nicht von Null signifikante Effektstärke von  $d = 0.01$ , für Verhütungsmittelgebrauch einen Effekt von  $d = 0.27$  und für Schwangerschaftsraten eine Effektstärke von  $d = 0.15$ . Lundahl, Risser und Lovejoy (2006) berichteten in einer Meta-Analyse zur Wirksamkeit von Elterntrainings bei externalisierenden Störungen Effekte von  $d = 0.42$  auf kindliches Problemverhalten,  $d = 0.47$  auf elterliches Verhalten und  $d = 0.53$  auf elterliche Wahrnehmung. Der Effekt auf kindliches Problemverhalten war hierbei für die 14 Studien, in denen die kindlichen Symptome vor dem Elterntraining kein klinisches Level erreichten, deutlich geringer, lag aber immerhin noch bei  $d = 0.31$ . Pinquart und Silbereisen (2014) geben ebenfalls eine Übersicht über Meta-Analysen zu Präventionsprogrammen mit Jugendlichen. In dieser werden auch Meta-Analysen zur Wirksamkeit von Programmen zur Prävention spezifischer psychischer Störungen aufgeführt. Teubert und Pinquart (2011) untersuchten in ihrer Meta-Analyse die Wirksamkeit von Programmen zur Prävention von Angst-

symptomen bei Kindern und Jugendlichen und kamen dabei zu Effektstärken von  $d = 0.22$  für Messungen direkt nach der Intervention und  $d = 0.19$  für Follow-up-Messungen. Stice, Shaw, Bohon, Marti und Rohde (2009) berichteten in ihrer Meta-Analyse zur Wirksamkeit von Depressionspräventionsprogrammen für Kinder und Jugendliche einen mittleren post-Test Effekt von  $r = .15$  (entspricht  $d = 0.30$ ) und einen mittleren Follow-up-Effekt von  $r = .11$  ( $d = 0.22$ ). Röhrle (2008) fasste die Ergebnisse von 113 Meta-Analysen zur Prävention psychischer Störungen und Förderung psychischer Gesundheit zusammen. Für die in den Meta-Analysen enthaltenen 1799 Studien ergab sich eine mittlere Effektstärke von  $d = 0.37$ . Hinsichtlich der Prävention psychischer Störungen zeigte sich hierbei, dass Programme zur Prävention von Depression mit  $d = 0.39$  am effektivsten waren, wohingegen Drogenmissbrauchsprogramme mit  $d = .20$  eine deutlich niedrigere Wirksamkeit hatten. Im Mittelfeld lagen Programme zur Prävention von Verhaltensstörungen bzw. Delinquenz mit  $d = 0.30$ . Höhere Effekte zeigten sich für Elterntrainingsprogramme mit  $d = 0.53$ .

Neben dem Programmtyp gibt es noch eine Reihe anderer Faktoren, welche allgemein die Wirksamkeit präventiver Interventionen beeinflussen können. Ein solcher Faktor ist die Implementation, also die Umsetzung eines Präventionsprogrammes. Diese wird in Übersichtsarbeiten häufig als sehr bedeutsam, zugleich aber empirisch noch zu schwach beforscht, beurteilt (Beelmann, 2006; Beelmann & Karing, 2014). S. J. Wilson, Lipsey und Derzon (2003) konnten für schulbasierte Interventionsprogramme zur Aggressionsprävention zeigen, dass die Implementationsqualität, welche anhand der berichteten Probleme bei der vollständigen Durchführung der beabsichtigten Intervention bei der Zielgruppe beurteilt wurde, positiv mit der Wirksamkeit assoziiert ist. Beelmann und Karing (2014) diskutieren außerdem als mögliche Ansätze zur Förderung individueller Implementationsfaktoren die Ausbildung und Supervision der Programmadministratoren sowie teilnehmerorientierte Maßnahmen zur Steigerung der Inanspruchnahme und der Compliance wie z.B. Kinderbetreuung, Transportservice zum Trainingsort oder monetäre Anreize für die Teilnahme. Auch der berufliche Hintergrund der Programmadministratoren kann den Erfolg einer präventiven Intervention mitbestimmen. So konnten S. J. Wilson, Lipsey & Derzon (2003) zeigen, dass die Wirkung schulbasierter Interventionen auf aggressives Verhalten stärker ausfällt, wenn es sich bei den Administratoren um Lehrer handelt, Programme, die von Laien durchgeführt werden dagegen weniger effektiv sind. Hingegen ergaben sich in der Meta-Analyse von Teubert

und Pinquart (2011) höhere Effekte für Administratoren, die beruflich im Bereich der psychischen Gesundheit arbeiteten als für solche, die als Lehrer tätig waren. Weiterhin ist von Belang, in welchem Format eine Intervention stattfindet. Während Lösel und Beelmann (2003) im Bereich sozialer Trainingsprogramme zur Prävention dissozialen Verhaltens bei Kindern und Jugendlichen keinen Unterschied zwischen Programmen, die in der Gruppe und solchen, die einzeln durchgeführt wurden, fanden, konnten Lundahl et al. (2006) für Elterntrainings zeigen, dass diese wirksamer waren, wenn sie einzeln durchgeführt wurden. Letztere fanden dabei auch, dass dieser Effekt bei Familien mit niedrigem sozioökonomischem Status besonders stark ausgeprägt war. Neben diesen eher qualitativen Merkmalen sollte schließlich auch die Intensität eine Rolle spielen, mit der eine präventive Intervention durchgeführt wird. Allerdings ist hier die Befundlage teilweise widersprüchlich (Beelmann, 2006). Während Tobler et al. (2000) sowie S. J. Wilson, Lipsey und Derzon (2003) einen positiven Effekt zwischen Programmintensität und Effektstärke zeigen konnten, fanden Stice et al. (2009) für depressionspräventive Programme, dass solche mit geringerer Dauer (weniger als 12 Stunden) höhere Effekte aufwiesen, als jene mit größerer Dauer.

Auch die Präventionsart hinsichtlich der Zielgruppe der Maßnahme kann einen Einfluss auf die Effektstärke haben, welche in Studien ermittelt wird. Beelmann (2006) weist darauf hin, dass in Meta-Analysen in konsistenter Weise geringere Effekte für universelle im Vergleich zu gezielten Präventionsmaßnahmen gefunden wurden, was vor allem mit geringeren Problemprävalenzen und daraus resultierenden Boden- bzw. Deckeneffekten begründet wird. Eine Ausnahme stellt hierbei die Meta-Analyse von Lundahl et al. (2006) dar, in der gezeigt werden konnte, dass Familien mit niedrigem sozioökonomischem Status insgesamt deutlich weniger von Elterntrainingsprogrammen profitieren als ökonomisch weniger benachteiligte Familien. Ein weiteres im Hinblick auf die Stichprobe bedeutsames Merkmal kann das Alter der Kinder bzw. Jugendlichen sein. Tobler et al. (2000) konnten dabei eine mit dem Alter bzw. der damit einhergehenden Schulform ansteigende Wirksamkeit von Drogenpräventionsprogrammen nachweisen.

Die in Studien gefundenen Effekte sind jedoch auch von forschungsmethodischen Faktoren abhängig. Als erstes ist hierbei die Art der Kontrollgruppe zu nennen. Erhält eine Kontrollgruppe eine minimale oder unspezifische Intervention (z.B. Aufmerksamkeitskontrollgruppe), so kann auch das einen positiven Einfluss auf die dieser Gruppe zugeordneten Teilnehmer haben. Entsprechend konnte auch gezeigt werden, dass Studien,

die derartige aktive Kontrollgruppen verwendeten, tendenziell kleinere Effekte aufweisen als gänzlich unbehandelte Kontrollgruppen oder Wartelistenkontrollgruppen (Teubert & Pinquart, 2011). Bezüglich der internen Validität konnte gezeigt werden, dass Studien, in denen die Vergleichbarkeit zwischen der Interventions- und Kontrollgruppe gefährdet ist, größere Effekte aufweisen (Beelmann, Pfost & Schmitt, 2014). Hingegen konnten S. J. Wilson, Lipsey & Derzon (2003) zeigen, dass eine höhere Dropout-Rate mit einer niedrigeren Effektstärke einhergehen kann. Hinsichtlich des Umgangs mit Dropouts und Nichtteilnahme an Sitzungen bieten *Intention-to-treat*-Analysen die konservativste Schätzung, wogegen sich bei alternativen Auswertungsmethoden höhere Effekte zeigen (Gupta, 2011; Abraha et al., 2015). Für randomisiert-kontrollierte Studien zeigte sich, dass Studien mit *Cluster-Randomisierung* (z.B. auf Basis von Klassen oder Schulen), die häufig auf größeren Stichproben beruhen, niedrigere Effekte aufweisen als solche mit individuell randomisierter Zuweisung (Lösel & Beelmann, 2003).

Insgesamt gibt es auch Hinweise darauf, dass das Publikationsjahr einen Zusammenhang sowohl mit den in Studien ermittelten als auch mit den in Meta-Analysen integrierten Wirksamkeitsparametern aufweist. Später publizierte Arbeiten ergeben hierbei tendenziell niedrigere Effekte (Beelmann 2006).

## **2.4 Prävention bei Kindern und Jugendlichen mit Migrationshintergrund**

Da, wie in Kapitel 2.2 erläutert, ein Migrationshintergrund einen Risikofaktor für die Entwicklung von sozialen und emotionalen Defiziten darstellt, handelt es sich bei präventiven Interventionen zur sozio-emotionalen Förderung, die speziell für Kinder und Jugendliche mit Migrationshintergrund entwickelt wurden und mit diesen durchgeführt werden, um gezielte Prävention nach der Kategorisierung von Offord et al. (1998). Wenn die Kinder und Jugendlichen mit Migrationshintergrund, welche mit der Präventionsstrategie angesprochen werden sollen, bereits erste Auffälligkeiten hinsichtlich des zu vermeidenden Problems bzw. der Störung zeigen, so handelt es sich um indizierte Prävention. Andernfalls kann man, der Kategorisierung von Gordon (1983) folgend, von einer selektiven Präventionsmaßnahme sprechen. Des Weiteren werden Kinder und Jugendliche mit Migrationshintergrund jedoch auch neben einheimischen Kindern und Jugendlichen in universelle Präventionsmaßnahmen einbezogen. Die hierbei erzielte Wirkung auf die Gruppe der Kinder und Jugendlichen mit Migrationshintergrund kann



zum einen von der Effektivität des entsprechenden Präventionsprogrammes für Teilnehmer ohne Migrationshintergrund abweichen und sich zum anderen auch von der Wirkung unterscheiden, welche für die Zielgruppe in gezielten Präventionsmaßnahmen zu erzielen wäre. Generell sollte, entsprechend der Argumentation von Offord et al. (1998), für Kinder und Jugendliche mit Migrationshintergrund die Wirksamkeit präventiver Interventionen größer sein als für einheimische Kinder und Jugendliche, da es sich bei erstgenannter Gruppe, wie bereits erläutert, um eine Risikogruppe handelt, in der es folglich auch weniger Personen gibt, die sich auch ohne Intervention positiv entwickeln würden. Um Kinder und Jugendliche mit Migrationshintergrund sozio-emotional zu fördern, kann es unter Umständen sinnvoll sein, auf universelle Prävention zurückzugreifen, da ethnische Minderheiten ohnehin häufig von Vorurteilen und Diskriminierung betroffen sind und somit bei gezielten Präventionsstrategien für diese Gruppe das Risiko einer Stigmatisierung besteht (Titzmann et al., 2011; T. C. Wilson, 1996). Gezielte Präventionsstrategien bieten jedoch den Vorteil, dass sie besser auf die Bedürfnisse der Zielpopulation zugeschnitten werden können. Dies kann insbesondere für Kinder und Jugendliche mit Migrationshintergrund von Bedeutung sein (Castro, Cota & Vega, 1999). So sind universelle Präventionsprogramme häufig an der in dem Einwanderungsland vorherrschenden Kultur orientiert und legen damit auch deren Risiko- und Schutzfaktoren zugrunde. Überdies werden Einstellungen und Werte angesprochen, welche möglicherweise in dieser Konstellation bei der Zielpopulation kulturbedingt gar nicht vorliegen müssen (Dent, Sussman, Ellickson, Brown & Richardson, 1996; Johnson, et al., 2005). Diese Fehlpassung sollte gerade dann nur schwer zu kompensieren sein, wenn die kulturelle Minderheit, wie es bei universellen Präventionsmaßnahmen der Fall ist, zusammen mit Mitgliedern der Mehrheitsgesellschaft trainiert wird. Dent et al. (1996) liefern jedoch für den Bereich der Drogenprävention auch theoretische Gründe, weswegen eine kulturelle Anpassung von Programmen *nicht* unbedingt notwendig ist. Zum einen gibt es zwar für viele verbreitete Drogenpräventionsprogramme Belege dafür, dass sie allgemein wirksam sind, jedoch keine klaren Hinweise darauf, dass sie es für ethnische Minderheiten nicht sind. Außerdem haben in interaktiven Programmen, in denen die Kinder bzw. Jugendlichen eine aktive Rolle einnehmen, diese hierüber ebenfalls die Möglichkeit ihre kulturell bedingten Einstellungen und Werte einfließen zu lassen. Darüber hinaus gibt es auch eine substantielle Überschneidung in den Risikofaktoren für Substanzmissbrauch zwischen unterschiedlichen ethnischen Gruppen in den

USA. Nicht zuletzt ist es auch kostenintensiv für die unterschiedlichen Subpopulationen jeweils unterschiedliche Programme zu entwickeln und zu implementieren und die Betonung der Unterschiede zwischen den Gruppen könnte Stereotypisierungsprozesse weiter verstärken (Dent et al., 1996).

Auch in empirischen Arbeiten ist der Nutzen einer kulturellen Anpassung von Programmen keineswegs unumstritten. So konnte zwar von Johnson et al. (2005) gezeigt werden, dass für eine mehrheitlich lateinamerikanische und asiatischstämmige Stichprobe in den USA ein multikulturelles Substanzmissbrauchsprogramm, welches ein kollektivistisches Selbst anspricht, eine höhere Wirksamkeit erzielen kann als ein Standardcurriculum. Bledsoe (2002) konnte hingegen in seiner Meta-Analyse einen Vorteil für kulturell angepasste gegenüber unangepassten Programmen nicht bestätigen. Ebenso konnten S. J. Wilson, Lipsey und Soydan (2003) zeigen, dass kulturell nicht angepasste universelle Präventionsprogramme zur Delinquenz- bzw. Dissozialitätsprävention für ethnische Minoritäten genauso wirksam sind wie für die ethnische Majorität in den USA. Jedoch kann es sowohl aus ethischen wie aus statistischen Gründen angemessen sein, von Programmen eine höhere Wirksamkeit für Kinder und Jugendliche mit Migrationshintergrund zu erwarten als für solche ohne Migrationshintergrund. Da sich bei Kindern und Jugendlichen mit Migrationshintergrund häufiger Probleme herausbilden, ist die Zahl an problematischen Verläufen, die durch ein Präventionsprogramm verhindert werden könnten, bei dieser Gruppe demnach größer als bei Kindern und Jugendlichen ohne Migrationshintergrund. Dies sollte auch dazu führen, dass in Präventionsstudien Boden- und Deckeneffekte bei Kindern und Jugendlichen mit Migrationshintergrund geringer ausfallen. Aus ethischen Gründen erscheint es zudem erstrebenswert, soziale Ungleichheit zu reduzieren, um Chancengerechtigkeit und Teilhabe zu gewährleisten. Dies gelingt jedoch nicht, wenn benachteiligte oder besonders gefährdete Gruppen nur genauso viel und nicht mehr von einer Intervention profitieren, da dies die relative Ungleichheit nicht verändert. Des Weiteren kann es als ein Ziel präventiven Handelns gelten, denjenigen, die besonders gefährdet sind, auch am meisten und wirksamsten zu helfen. Meta-Analysen haben gezeigt, dass universelle Präventionsprogramme im Allgemeinen niedrigere Effektstärken aufweisen als gezielte Präventionsprogramme (Beelmann, 2006). Schon aus diesem Grund erscheint es fragwürdig, gerade bei der Risikogruppe der Kinder und Jugendlichen mit Migrationshintergrund Effektgrößen für angemessen zu erachten, welche denen universeller Präventionsprogramme

gleichen. Folglich kann auch der Befund, dass kulturell nicht angepasste, universelle Präventionsprogramme für ethnische Minderheiten und die Mehrheitsbevölkerung eine ähnlich hohe Wirksamkeit besitzen, nicht als Hinweis darauf gewertet werden, dass keine Notwendigkeit gezielter Prävention bzw. kultureller Anpassung einer Intervention an ethnische Minderheiten bestünde. Zudem gibt die Meta-Analyse von S. J. Wilson, Lipsey und Soydan (2003) keinerlei Auskunft darüber, ob eine kulturelle Anpassung die Wirksamkeit präventiver Interventionen für Kinder und Jugendliche mit Migrationshintergrund nicht noch steigern würde.

Ein weiterer wichtiger Faktor, der speziell bei präventiven Interventionen für Kinder und Jugendliche mit Migrationshintergrund zu beachten ist, ist deren Familiensystem bzw. häusliches Milieu. Auf der einen Seite ist dieses gerade bezüglich der sozialen und emotionalen Entwicklung von Kindern und Jugendlichen mit Migrationshintergrund besonders bedeutsam (Belhadj Kouider et al., 2014; Birman & Taylor-Ritzler, 2007; Choi, He & Harachi, 2008; Stevens et al., 2007). Auf der anderen Seite ist das Familiensystem jedoch gerade bei dieser Population besonders anfällig für Störungen und Konflikte und stellt daher einen Risikofaktor dar (Belhadj Kouider et al., 2014; Sluzki, 1979). Folglich ist das Familiensystem bzw. das Elternverhalten auch ein bedeutsamer Ansatzpunkt, wenn es um die sozio-emotionale Förderung von Kindern und Jugendlichen mit Migrationshintergrund geht.

Zur Wirksamkeit von Präventionsprogrammen für Kinder und Jugendliche gibt es bereits viele Reviews und Meta-Analysen, von denen einige auch ethnische Minderheiten oder Personen mit Migrationshintergrund berücksichtigen. Diese betrachten jedoch lediglich als Teilpopulation bestimmte ethnische Minderheiten oder schließen nur ganz spezifische Stichproben oder Präventionsansätze ein (Bledsoe; 2002; Cardoza, Documet, Fryer, Gold & Butler, 2012; Castro et al., 2006; Huey & Polo, 2008; Jackson, 2009; Jani, Ortiz & Aranda, 2009; Kong, et al., 2012; S. J. Wilson, Lipsey & Soydan, 2003; Yuen, 2004).

Die Meta-Analyse von Yuen (2004) untersuchte die Wirksamkeit von kulturell angepassten Programmen bei ethnischen Minderheiten. Die Suche wurde auf ethnische Minderheiten in den USA und zuvor definierte traditionelle Minderheiten in anderen Ländern beschränkt und beinhaltete zudem nur Studien, die bis einschließlich 2003 veröffentlicht wurden. 99 % der Studien in dieser Meta-Analyse kamen aus den USA und neben der überwiegenden Zahl an Studien die zu über 90 % afroamerikanische

Stichproben oder amerikanische Ureinwohner untersuchten, wurden dabei auch einige Studien mit Lateinamerikanern und Amerikanern asiatischer Herkunft eingeschlossen. Yuen (2004) fand dabei für Lateinamerikaner in den USA eine mittlere Effektstärke von  $d = 0.14$  und für Amerikaner asiatischer Herkunft eine mittlere Effektstärke von  $d = 0.32$ . Die Wirksamkeit präventiver Interventionen bei Kindern und Jugendlichen mit Migrationshintergrund ist jedoch mit dieser Studie nicht umfassend genug untersucht, da die Einschränkung auf bestimmte ethnische Minderheiten genauso wie der Ausschluss von kulturell nicht angepassten Programmen bedeutsame Befunde unberücksichtigt lässt. Ebenfalls ist diese Meta-Analyse bereits über 10 Jahre alt und enthält dadurch viele aktuelle Forschungsarbeiten nicht.

Huey & Polo (2008) fassten die Befunde zur psychosozialen Behandlung ethnischer Minderheiten unter 18 Jahren mit emotionalen oder Verhaltensauffälligkeiten zusammen. Eingeschlossen wurden unter anderem vier Studien mit lateinamerikanischen Stichproben, für die eine mittlere Effektstärke von  $d = 0.47$  errechnet wurde. Von diesen testete jedoch lediglich eine Studie eine präventive Intervention anhand einer nicht-klinischen Stichprobe. Da in diesem Review außerdem nur Studien eingeschlossen wurden, die ausschließlich Kinder und Jugendliche behandelten, welche bereits emotionale oder Verhaltensauffälligkeiten aufwiesen, waren sämtliche universellen und selektiven Präventionsprogramme ausgeschlossen. Damit wird ein Großteil relevanter Programme nicht berücksichtigt.

Eine Meta-Analyse von S. J. Wilson, Lipsey & Soydan (2003) untersuchte die Wirkung etablierter Programme gegen jugendliche Delinquenz bei ethnischen Minderheiten im Vergleich zur Mehrheit. Es wurden veröffentlichte und unveröffentlichte Studien eingeschlossen, die zwischen 1950 und 1996 durchgeführt wurden. Ein weiteres Einschlusskriterium war, dass die Jugendlichen zwischen 12 und 21 Jahre sein und in den USA, Kanada, Großbritannien, Neuseeland oder Australien wohnen mussten. Sie sollten zudem als delinquent oder dissozial beschrieben oder es musste delinquentes oder dissoziales Verhalten als primäre Erfolgsvariable erfasst worden sein. Von den 141 eingeschlossenen Studien, die eine Stichprobe untersuchten, welche hauptsächlich (>60 %) ethnische Minderheiten enthielt, waren jedoch nur 19 Studien mit hauptsächlich hispanoamerikanischer Stichprobe. Weiterhin wurde für diese Studien keine separate mittlere Effektstärke berichtet, sondern lediglich für Studien, in denen selbst nach Ethnie getrennte Ergebnisse berichtet wurden. Sowohl für den Vergleich hispanoamerikanischer

Teilnehmer mit afroamerikanischen als auch für den mit weißen US-Amerikanern existierten 12 Studien und es konnte für beide Vergleiche kein signifikanter Unterschied in den Effekten festgestellt werden. Allerdings waren die in dieser Meta-Analyse gefundenen Effektstärken insgesamt sehr klein, berücksichtigt man die Tatsache, dass die eingeschlossenen Stichproben mehrheitlich bereits zu Interventionsbeginn delinquent waren und es sich somit um indizierte Präventionsmaßnahmen handelte. Dies traf insbesondere auf jene Studien zu, in denen nach Ethnie getrennte Ergebnisse angegeben wurden. Bei diesen Analysen ergaben sich für die lateinischen Teilstichproben jeweils nicht signifikant von null verschiedene Effektstärken, von  $d = -0.01$  für Studien in denen ein Vergleich mit Afroamerikanern erfolgte und  $d = 0.04$  für Studien, in denen auch Ergebnisse für weiße US-Amerikaner berichtet wurden. Da die Meta-Analyse jedoch eine Reihe anderer Migrationsländer (z. B. Niederlande, Norwegen, Deutschland) ausschloss und die vorhandenen Daten zu hispanoamerikanischen Jugendlichen nicht vollständig in Analysen berichtet, kann die Studie keine angemessene Auskunft über die Wirksamkeit der Programme für Jugendliche mit Migrationshintergrund liefern. Auch fand eine Einschränkung auf Dissozialitäts-/Delinquenzpräventionsprogramme statt, wodurch viele Programme beispielsweise mit dem Ziel der Erhöhung des Selbstwertes, der Reduktion von Akkulturationsstress und generell der Vermeidung internalisierender Probleme ausgeschlossen wurden. Da zudem der Suchzeitraum dieser Meta-Analyse über 15 Jahre zurück liegt, kann sie den aktuellen Forschungsstand nicht annähernd abdecken.

Jani et al. (2009) gaben einen Überblick über die Befunde von Studien, in denen Interventionen aus den Bereichen Gesundheit, geistiger Gesundheit und Substanzmissbrauch anhand von Stichproben evaluiert wurden, welche unter anderem Lateinamerikaner enthielten. Die Studien mussten zwischen 1999 und 2005 in einer Fachzeitschrift veröffentlicht worden sein und Prä- und Postinterventionswerte liefern, um eingeschlossen zu werden. Außerdem sollten Lateinamerikaner vom Festland der USA enthalten und die Intervention kulturell an diese angepasst sein. Es wurden zwar keine Einschränkungen hinsichtlich des Alters der Stichprobe gemacht, jedoch befassten sich von den fünf eingeschlossenen Substanzmissbrauchsprogrammen alle mit Kindern und Jugendlichen bzw. ihren Eltern. Effektstärken werden nicht angegeben, allerdings werden für drei der vier randomisiert-kontrollierten Studien zu Substanzmissbrauchsinterventionen Ergebnisse für den Vergleich zwischen Interventions- und Kontrollgruppe berichtet. Zwei der Studien fanden hierbei positive Effekte im Vergleich zur Kontroll-

gruppe, in einer dritten wurde jedoch festgestellt, dass spanischsprachige mexikanischstämmige Jugendliche keinen Interventionseffekt aufweisen. In einer Studie wurde zudem gezeigt, dass eine Intervention, in die lateinamerikanische Werte einbezogen wurden, keine höheren Effekte hatte als eine in der dies nicht geschah. Hinsichtlich des Bereiches der Substanzmissbrauchsprogramme fassten Jani et al. (2009) schließlich zusammen, dass die Präventionsmethoden bildungsbasiert waren, in der Sprache vorgebracht wurden, in der die Teilnehmer am sichersten waren, sowie die Familie des Kindes einschlossen und dass diese Merkmalskombination mit positiven Effekten zusammenhängen. Die Wirksamkeit von Substanzmissbrauchsprogrammen für Kinder und Jugendliche mit Migrationshintergrund wurde in dieser Studie jedoch nicht erschöpfend untersucht. Die Beschränkung auf Lateinamerikaner in den USA stellt erneut eine selektive Suchstrategie dar und auch der Suchzeitraum war mit sechs Jahren relativ eng. Mögliche Publikationsverzerrungen wurden durch die Beschränkung auf Fachzeitschriften nicht berücksichtigt genauso wie nicht kultursensitive Interventionen mit Migrantenstichproben keine Beachtung fanden.

Auch Castro et al. (2006) fassten den Forschungsstand zur Substanzmissbrauchsprävention für Lateinamerikaner zusammen. Dabei schlossen sie jedoch auch korrelative Studien und Stichproben ein, die nur zum kleineren Teil aus Lateinamerikanern in den USA bestanden. Es wurde sich auf Publikationen in Fachzeitschriften zwischen 1978 und 2003 beschränkt und Studien die ausschließlich auf Alkohol und Tabakkonsum abzielten, wurden explizit ausgeschlossen. Genauere Einschlusskriterien wurden jedoch nicht formuliert. Ebenso wurden keine Effektstärken angegeben, jedoch insgesamt von einer Wirksamkeit der Programme gesprochen. Auch hier wurde aufgrund der selektiven Stichprobenauswahl, des etwa 10 Jahre zurück liegenden Suchzeitraums, der Beschränkung auf publizierte Fachzeitschriftenbeiträge und des Ausschlusses von Studien zur Wirksamkeit von Programmen, die auf die Prävention bzw. Reduktion von Alkohol- und Tabakmissbrauch abzielten, der gegenwärtige Forschungsstand zur Wirksamkeit von Substanzmissbrauchsprogrammen bei Kindern und Jugendlichen mit Migrationshintergrund nicht angemessen abgedeckt.

Kong et al. (2012) fassten in einem Review Studien zu kulturell angepassten Rauchpräventionsprogrammen für Jugendliche in den USA zusammen, die zwischen 1990 und 2011 veröffentlicht wurden. Dabei sollten jeweils mindestens 50 % der Jugendlichen einer ethnischen Minderheit angehören; als Wirksamkeitsvariablen musste

entweder das Anfangen mit dem Rauchen oder dessen Reduktion oder Beendigung erfasst worden sein. Einige der Studien untersuchten hauptsächlich hispanoamerikanische Jugendliche. Mittlere Effektstärken wurden nicht angegeben, jedoch festgestellt, dass das Initiieren von Rauchen allgemein wirksam verhindert werden konnte, die kulturell sensitiven Programme jedoch in der Regel nicht dazu führten, dass mehr Jugendliche mit dem Rauchen aufhörten. Durch die Beschränkung auf kultursensitive Programme und den Fokus auf Tabakkonsum deckt jedoch auch dieses Review nur einen kleinen Bereich des Forschungsstandes zur Substanzmissbrauchsprävention ab.

In der Meta-Analyse von Bledsoe (2002) wurde die Wirksamkeit von Substanzmissbrauchsprogrammen bei ethnischen Minderheiten in den USA untersucht. Dabei wurden quasi-experimentelle sowie experimentelle Studien eingeschlossen, die zu mindestens 50 % ethnische Minderheiten enthielten und auf Jugendliche von der fünften bis zur 12. Klasse abzielten. Es wurde unter anderem untersucht, ob Präventionsprogramme mit kulturellen Komponenten höhere Effekte erzielen als solche ohne kulturelle Elemente. Dabei zeigte sich kein Unterschied in der mittleren gewichteten Effektstärke von kulturell angepassten ( $d = .12$ ) und nicht angepassten ( $d = .13$ ) Programmen. Für einzelne Subanalysen wurden auch separate Ergebnisse für Lateinamerikaner angegeben, die z. B. für Studien mit zugrunde gelegtem theoretischen Modell zu einer Effektstärke von  $d = 0.18$  ( $k = 15$ ) und für solche ohne ein solches Modell zu einem Effekt von  $d = 0.23$  ( $k = 5$ ) führen. Da in dieser Meta-Analyse jedoch nur ethnische Minderheiten in den USA betrachtet und nur für Lateinamerikaner separate Ergebnisse angegeben wurden, ermöglicht sie keinerlei Aussage für andere Migrantengruppen. Da die Meta-Analyse zudem 2003 veröffentlicht wurde und somit schon annähernd 10 Jahre alt ist, deckt sie den gegenwärtigen Forschungsstand zu Substanzmissbrauchsprogrammen nur unzureichend ab.

Jackson (2009) fasste Befunde zur Wirksamkeit von kultursensitiven Interventionen bei ethnischen Minderheiten in den USA zusammen, die in Fachzeitschriften veröffentlicht wurden. Unter den integrierten Studien befanden sich auch eine experimentelle Studie mit mexikanischstämmigen Amerikanern sowie je eine quasi-experimentelle Studie mit Lateinamerikanern und chinesischstämmigen Amerikanern. Effektstärken wurden nicht berichtet, jedoch wurden marginal- bis signifikante Effekte für Alkohol- und Tabakkonsum sowie die Sicherheit beim Zurückweisen von Drogen hervorgehoben. Auch dieses Review untersuchte die Wirksamkeit von Substanzmissbrauchsprogrammen

bei Kindern und Jugendlichen mit Migrationshintergrund keinesfalls erschöpfend. So wurden zum einen unveröffentlichte Studien nicht berücksichtigt und zum anderen eine Einschränkung auf Populationen in den USA vorgenommen. Auch ermöglicht dieses Review keine Rückschlüsse auf Substanzmissbrauchsprogramme, die zwar nicht kultursensitiv gestaltet wurden, deren Wirksamkeit jedoch auch bei Kindern und Jugendlichen mit Migrationshintergrund untersucht wurde.

Im Bereich der Sexualprävention fassten Cardoza et al. (2012) in einem systematischen Review Studien zusammen, deren Zielpopulation hauptsächlich Lateinamerikaner in den USA zwischen 11 und 21 Jahren waren. Dabei wurden nur Artikel aus Fachzeitschriften eingeschlossen, die zwischen 1993 und 2011 veröffentlicht wurden. Von den 15 im Review eingeschlossenen Studien hatten fünf ein randomisiert-kontrolliertes Design genutzt. Effektstärken wurden in dem Review nicht berichtet, allerdings zeigten die meisten Studien eine Wirksamkeit hinsichtlich der Reduktion sexuellen Risikoverhaltens sowie Veränderungen in Einstellungen, Überzeugungen und Wahrnehmungen bezüglich sexueller Gesundheit und sexuell übertragbarer Krankheiten. Die Befundlage zur Wirksamkeit von Sexualpräventionsprogrammen bei Kindern und Jugendlichen mit Migrationshintergrund wurde in diesem Review jedoch ebenfalls nicht vollständig erfasst, da zum einen die Zielpopulation, zum anderen aber auch Publikationsform und -zeitraum eingeschränkt waren.

Abschließend kann damit festgestellt werden, dass bislang keine forschungssynthetische Arbeit existiert, die den gesamten aktuellen internationalen Forschungsstand zur Wirksamkeit präventiver Interventionen zur sozio-emotionalen Förderung von Kindern und Jugendlichen mit Migrationshintergrund integriert und systematisch aufarbeitet.

## **2.5 Ziele der Arbeit**

Aufgrund des beschriebenen bisherigen Mangels integrativer Arbeiten in diesem Forschungsbereich, soll es das Ziel der vorliegenden Arbeit sein, die bisherigen Befunde zur Wirksamkeit präventiver Interventionen zur sozio-emotionalen Förderung bei Kindern und Jugendlichen mit Migrationshintergrund mit meta-analytischen Methoden zu integrieren. Dabei sollen insbesondere folgende Fragestellungen untersucht werden:



1. Wie hoch ist die allgemeine Wirksamkeit präventiver Interventionen zur sozio-emotionalen Förderung bei Kindern und Jugendlichen mit Migrationshintergrund?
2. Wie stabil sind die erzielten Effekte der präventiven Interventionen über die Zeit?
3. Wie ist die differenzielle Wirksamkeit hinsichtlich verschiedener Erfolgskriterien?
4. Inwiefern unterscheiden sich die in den Studien ermittelten Effekte in Abhängigkeit von publikationsbezogenen Charakteristika?
5. Welchen Zusammenhang gibt es zwischen der Höhe der gemessenen Effekte und verschiedenen methodischen Merkmalen?
6. Welchen Einfluss haben verschiedene Interventionscharakteristika auf die Wirksamkeit der Programme?
7. Inwieweit unterscheidet sich die Wirksamkeit der präventiven Interventionen in Abhängigkeit von verschiedenen Merkmalen der Stichprobenzusammensetzung?

### 3. METHODEN

#### 3.1 Forschungssynthese mittels Meta-Analyse

Der Begriff Meta-Analyse bezeichnet im Allgemeinen eine Integration der Ergebnisse einer größeren Anzahl von einzelnen empirischen Primärstudien mithilfe statistischer Analysemethoden (Glass, 1976). Eine Meta-Analyse kann damit auch als eine Art wissenschaftliche Studie verstanden werden, in der statt Personen, ganze Forschungsberichte statistisch erfasst werden (Lipsey & D. B. Wilson, 2001).

Eine solche forschungssynthetische Arbeit kann gerade in viel beforschten Forschungsfeldern sinnvoll sein. Wenn zu einem Thema viele einzelne unterschiedliche Forschungsbefunde existieren, die zu verschiedenen, teils widersprüchlichen Schlussfolgerungen kommen, so kann es nicht nur für Anwender der Forschung, sondern auch für Forscher selbst schwierig werden den Forschungsstand zu überblicken, ihn angemessen zu bewerten und zu einem klaren Verständnis der Befundlage zu gelangen (Glass, 1977). Hierfür kann eine strukturierte Aufarbeitung und Integration der einzelnen Forschungsbefunde notwendig sein. Auch haben Ergebnisse aus forschungssynthetischen Arbeiten in der Regel eine höhere Generalisierbarkeit als einzelne empirische Primärstudien, da Effekte über verschiedene Settings und Stichproben hinweg getestet werden können (Matt, 2003). Ein weiterer Vorteil von Meta-Analysen ist, dass auch Fragestellungen bearbeitet werden können, die entweder erst durch Integration mehrerer verschiedener Studie sinnvoll untersucht werden können oder aus den Einzelergebnissen von Primärstudien bisher nicht erschließbar sind (Cooper & Lemke, 1991). Beispiele sind die Fragen, ob publizierte Studien höhere Effekte zeigen als unpublizierte oder welchen Einfluss die Dropout-Rate auf die Wirksamkeit einer Intervention hat.

Neben der Meta-Analyse existieren auch noch andere forschungsintegrative Verfahren wie narrative Reviews und das Auszählen von Signifikanzen (*Box-Score Methode*, *vote counting* oder *voting method* genannt; Glass, 1977; Hedges & Olkin, 1985; Lipsey & D. B. Wilson, 2001). Ersteres ist jedoch bei einer sehr hohen Zahl an Forschungsbefunden zu dem Thema unpraktikabel und letzteres hängt zum einen in seinem Ergebnis wesentlich von der Größe der Stichproben in den Primärstudien ab und liefert zum anderen keine Information darüber, wie groß der gemessene Effekt tatsächlich ist (Glass, 1976, 1977; Hedges & Olkin, 1985; Lipsey & D. B. Wilson, 2001). Eine Meta-Analyse hingegen

integriert die Befunde *statistisch* und kann folglich quantitative Aussagen über die Stärke eines Zusammenhanges bzw. die Größe eines Effektes treffen.

Bei der Durchführung einer Meta-Analyse werden typischerweise 6 Phasen unterschieden (Beelmann & Bliesener, 1994; Cooper, Hedges & Valentine, 2009):

1. Formulierung und Operationalisierung einer Fragestellung, die bereits empirisch geprüft wurde,
2. Systematische Literatursuche und Studienausswahl,
3. Extraktion relevanter inhaltlicher und methodischer Merkmale der Studien,
4. Aufbereitung und Analyse der aus den Studien extrahierten Daten,
5. Interpretation der Ergebnisse,
6. Dokumentation und Veröffentlichung der Meta-Analyse.

In den folgenden Unterkapiteln soll entsprechend dokumentiert werden, welche Einschlusskriterien zugrunde gelegt (Operationalisierung der Fragestellung) und welche Suchstrategie zur Identifikation relevanter Primärstudien genutzt wurden, wie der Selektionsprozess verlief, wie die in die Meta-Analyse eingeschlossenen Studien kodiert und wie die Daten zur weiteren Analyse aufbereitet wurden.

### **3.2 Selektionskriterien**

In die vorliegende Meta-Analyse wurden ausschließlich Studien aufgenommen, welche bis Ende 2013 publiziert oder anderweitig zugänglich gemacht wurden und zudem die folgenden Selektionskriterien erfüllten:

1. Psychologische präventive Intervention mit dem Ziel der sozio-emotionalen Förderung,
2. Mehrheitlich nicht-klinische Stichprobe von Kindern bzw. Jugendlichen mit Migrationshintergrund,
3. Randomisiert-kontrolliertes Design,
4. Stichprobengröße von mindestens  $N = 25$  pro Gruppe,
5. Berechenbare oder schätzbare Effektstärken.

Diese Kriterien sollen im Folgenden näher erläutert werden:

1. Die Studie musste die Wirksamkeit einer präventiven Intervention (siehe Kapitel 2.3) untersuchen, welche mithilfe psychologischer Mittel sozio-emotionale Förderungs-

und Präventionsziele verfolgt. Mit psychologischen Mitteln sind jegliche Methoden gemeint, welche am inneren menschlichen Erleben und äußeren Verhalten ansetzen, wobei eine soziale Interaktion zwischen Durchführenden der Intervention und Teilnehmenden stattfindet (Perrez & Baumann, 2011). Beispielhaft können hier Schulungen, Beratungen, Verhaltenstrainings, kognitive Trainings, Sprach-, Problemlöse- oder soziale Kompetenztrainings, Durchführung von Übungsaufgaben und Psychoedukation genannt werden. Ausgeschlossen wurden im Gegensatz dazu Maßnahmen, die lediglich aus einem reinen Geldtransfer oder medizinischen Maßnahmen wie Medikamentengabe, chirurgischen oder physikalischen Eingriffen bestanden. Ziele der präventiven Intervention sollten angelehnt an Perrez's (1994) Klassifikationssystem die emotionale und soziale Entwicklung sein. Hierzu zählten insbesondere Programme zur Förderung der sozialen bzw. sozial-kognitiven Kompetenzen, des Selbstwerts, der ethnischen Identität oder der Emotionsregulation sowie jene zur Prävention internalisierender und externalisierender Problemverhaltensweisen. Des Weiteren sollten ergänzend zur Klassifikation von Perrez (1994) auch Maßnahmen zur Stressbewältigung, zum Umgang mit Drogen sowie Sexualpräventionsprogramme mit eingeschlossen werden, insofern sie nicht primär auf die Prävention körperlicher Krankheiten abzielten. Hierfür sind folgende Gründe zu nennen: Stress hängt mit der Entstehung von internalisierenden Problemen eng zusammen und eine Stressreduktion kann vor allem auch das emotionale Wohlbefinden steigern (Brock & Lawrence, 2014; Nykliček & Kuijpers, 2008; Sheidow, Henry, Tolan & Strachan, 2014). Substanzmissbrauch und riskantes Sexualverhalten hängen wiederum eng mit externalisierenden Problemverhaltensweisen zusammen (Iacono, Malone, McGue, 2003; Timmermans, van Lier & Koot, 2008). Auch sind soziale Kompetenzdefizite wie die mangelnde Fähigkeit *nein* zu sagen (sogenannte *refusal skills*) ein Risikofaktor für Substanzmissbrauch und riskantes Sexualverhalten (Griffin, Epstein, Botvin & Spoth, 2001; Warzak, Grow, Poler & Walburn, 1995). Die Hauptmotivation für die Nutzung von Kondomen ist zudem nicht die Prävention von HIV, sondern die Schwangerschaftsverhütung (Sheeran, Abraham & Orbell, 1999). Ausgeschlossen wurden hingegen Studien, in denen Programme untersucht wurden, die eine rein kognitive, sprachliche oder schulische Förderung zum Ziel hatten oder primär darauf abzielten, körperliche Krankheiten oder Risikofaktoren zu vermeiden, wie beispielsweise Programme zur Prävention von Über- oder Untergewicht oder zur Erhöhung der Impfquote. Auch Elterntrainingsprogramme sollten nur dann eingeschlossen

werden, wenn explizit dargelegt wurde, dass diese das Ziel der sozio-emotionalen Förderung der Kinder bzw. Jugendlichen verfolgen.

2. Bei der Stichprobe, die durch die Maßnahme erreicht werden sollte, musste es sich um Kinder oder Jugendliche handeln, die keine psychische Störung aufwiesen und zum überwiegenden Teil einen Migrationshintergrund hatten. Dass die Stichprobe keine klinisch-psychiatrische Diagnose aufweisen sollte, geht bereits aus der Definition von Prävention (und somit aus Kriterium 1) hervor. Hierbei wurde eine Studie nicht nur dann ausgeschlossen, wenn sie ausschließlich und explizit auf Kinder und Jugendliche mit psychischer Störung abzielte, sondern auch dann, wenn über die Hälfte der Stichprobe eine solche Störung aufwies oder die Untersuchung in einem Setting stattfand, welches auf eine derartige Prävalenz schließen ließ (z.B. psychiatrische Klinik). Das Alter der Stichprobe sollte im Mittel unter 18 Jahre sein. Ausschlaggebend für den Ein- oder Ausschluss war jedoch das Entwicklungsalter bzw. die Abhängigkeit oder Autonomie vom Elternhaus. Dies bedeutet, dass beispielsweise eine Studie, die Jugendliche untersuchte, welche zwar im Mittel knapp über 18 Jahren waren, jedoch (z. B. aufgrund Sitzenbleibens) noch zur Schule gingen und bei ihren Eltern wohnten die Einschlusskriterien erfüllte, eine Studie mit 17-jährigen, die bereits eine eigene Familie gegründet hatten, arbeitstätig oder Studenten waren und nicht mehr im Haushalt der Eltern wohnten, jedoch nicht. Weiterhin mussten mehr als 50 % der Kinder bzw. Jugendlichen einen Migrationshintergrund haben (Definition siehe Kapitel 2.1). Neben direkten Angaben über den Migrationsstatus der Stichprobe bzw. deren Eltern wurden dabei auch indirekte Hinweise auf einen Migrationshintergrund der Stichprobe berücksichtigt. So kann davon ausgegangen werden, dass bestimmte ethnische Minderheiten oder Gruppen, welche eine Muttersprache haben, die in der einheimischen Bevölkerung nicht verbreitet ist, zum überwiegenden Teil aus Personen mit Migrationshintergrund bestehen. Dies trifft insbesondere auf Lateinamerikaner, Amerikaner asiatischer Herkunft und nicht-englische Muttersprachler in den USA zu (Capps et al., 2005; Grieco, 2010; Lee, 1998; Pew Research Center, 2013). Stichproben, die zu über 90 % aus solchen Gruppen bestanden, erfüllten daher ebenfalls das Einschlusskriterium. Wurden in der Studie für die beschriebene Stichprobe separate Ergebnisse berichtet, so wurde die Studie auch dann eingeschlossen, wenn die Gesamtstichprobe der Studie die Einschlusskriterien nicht erfüllte. In diesem Fall wurden nur die separaten Ergebnisse für die relevante Stichprobe berücksichtigt. Eltern- und Lehrertrainingsprogramme wurden dann eingeschlossen, wenn

klar ersichtlich war, dass durch die Maßnahme primär die beschriebene Stichprobe von Kindern und Jugendlichen mit Migrationshintergrund erreicht werden sollte. Dies war der Fall, wenn die Erziehungskompetenzen der Eltern von Kindern und Jugendlichen mit Migrationshintergrund gestärkt werden sollten, nicht jedoch, wenn der inhaltliche Fokus, wie beispielsweise im Falle einer Paarberatung dieser Eltern, nicht auf der genannten Zielgruppe lag.

3. Die Evaluationsstudie musste ein randomisiert-kontrolliertes Studiendesign aufweisen, welches nach Einteilung der Maryland Skala dem methodischen Level 5 entspricht (Farrington, 2003). Dies bedeutet zum einen, dass neben der Gruppe, welche die zu untersuchende präventive Intervention erhielt, eine Kontrollgruppe existieren musste, welche die Intervention nicht bekam und am Ende mit der Interventionsgruppe verglichen werden konnte. Eingeschlossen waren dabei neben unbehandelten und Wartelisten-Kontrollgruppen auch Aufmerksamkeitskontrollgruppen sowie Gruppen, die nur eine minimale Intervention oder ein in dem untersuchten Setting populationsübergreifend angebotenes Standardprogramm erhielten. Studien, in denen die Vergleichsgruppe jedoch eine alternative Intervention bekam oder sich lediglich durch die Administratoren oder Anwendungsmodi (z.B. Einsatz unterschiedlicher Medien oder Vergleich von Ganztags- mit Halbtagschulen) von der Interventionsgruppe unterschieden, wurden hingegen ausgeschlossen. Weiterhin musste die Zuordnung der Teilnehmer auf die beiden Gruppen zufällig erfolgen. Nur durch die randomisierte Zuweisung zu den Gruppen ist es möglich, sowohl für gemessene als auch für unbekannt konfundierende Variablen zu kontrollieren und somit eine mögliche Verzerrung der gemessenen Interventionseffekte zu vermeiden (Baytop, 2006; Farrington, 2003). Um eine möglichst hohe Vergleichbarkeit der Gruppen zu gewährleisten, wurden Studien, in denen eine Cluster-Randomisierung (zufällige Zuweisung z.B. von Schulen oder Klassen) vorgenommen wurde, nur dann eingeschlossen, wenn jeder Gruppe mindestens 5 Cluster zugewiesen wurden. Insgesamt mussten mindestens 50% der in der Post-Test-Messung erfassten Stichprobe zufällig auf die Untersuchungsbedingungen zugewiesen worden sein. Demnach wurden auch Studien, in denen mehr als die Hälfte der Stichprobe erst nach der randomisierten Zuweisung nicht zufällig in eine der Gruppen aufgenommen wurde, ausgeschlossen, da eine Vergleichbarkeit der daraus resultierenden Gruppen dann nicht mehr gewährleistet war. Bei einer randomisierten Zuweisung zur Interventions- und Kontrollgruppe sind keine systematischen, bereits vor der Intervention bestehenden,

Gruppenunterschiede zu erwarten, weshalb eine Vortestmessung nicht notwendig ist (Campbell & Stanley, 1963). Daher wurden sowohl experimentelle Designs mit als auch ohne Vortestmessung eingeschlossen.

4. Die Stichprobe musste eine Größe von mindestens  $N=25$  in jeder Gruppe aufweisen. Kleine Stichproben liefern aufgrund des erhöhten Standardfehlers eine ungenauere Effektstärkenschatzung als Studien mit größerer Stichprobe (Carrasco, Volkmar & Bloch, 2012; Viechtbauer, 2007). Weiterhin sind kleinere Studien anfälliger für eine Publikationsverzerrung, da nicht signifikante Ergebnisse aus kleineren Stichproben häufig nicht veröffentlicht werden, wohingegen größere Studien in der Regel unabhängig von ihrem Ergebnis publiziert werden (Dickersin, 1990; Gilbody, Song, Eastwood & Sutton, 2000; Newcombe, 1987; Thornton & Lee, 2000). Die Stichprobengröße von insgesamt 50 Teilnehmern empfiehlt sich, da diese zum einen bei einem Mittelwertvergleich zu einer Normalverteilung der Prüfgröße führt (Bortz, 2005, S. 141) und zum anderen durch sie ein Mindestmaß an Teststärke erreicht wird (Cuijpers, van Straten, Bohlmeijer, Hollon & Andersson, 2010).

5. Die Studie musste Ergebnisse in einem Format präsentieren, aus dem einheitliche Effektstärken reliabel berechnet oder geschätzt werden konnten, die einen Unterschied zwischen Interventions- und Kontrollgruppe nach der Intervention ausdrückten und sich direkt oder indirekt auf die unter 2. beschriebene Stichprobe bezogen. Ausgeschlossen waren damit Studien, die lediglich eine Effektstärkenberechnung für den Vergleich zwischen Prä- und Post-Werten einer der beiden Gruppen oder zwischen den Gruppen vor der Intervention ermöglichten. Bezüglich der Stichprobe waren auch jene Effekte eingeschlossen, die sich beispielsweise auf die elterliche Erziehungskompetenz bezogen, wohingegen solche ausgeschlossen wurden, welche die Eltern selbst oder die elterliche Beziehung zum Gegenstand hatten, ohne dass ein klar dargelegter Bezug zu den Kindern bzw. Jugendlichen bestand. Wie die Effektstärken geschätzt oder berechnet wurden, soll in Kapitel 3.6.1 näher erläutert werden.

### **3.3 Suchstrategie zum Sichten relevanter Primärstudien**

Die Sichtung relevanter Primärstudien erfolgte zwischen Februar 2013 und September 2014. Da die vorliegende Meta-Analyse im Rahmen eines größeren, vom Bundesministerium für Bildung und Forschung (BMBF) finanzierten und von Prof. Dr. Andreas

Beelmann geleiteten Projektes mit dem Titel „Wirksamkeit von Bildungs- und Interventionsprogrammen zur Prävention und Kompensation von Armuts- und Migrationsfolgen bei Kindern und Jugendlichen. Ein Forschungsüberblick und eine Meta-Analyse internationaler Evaluationsforschung“ (Förderkennzeichen: 01JC1105) durchgeführt wurde, gestaltete sich die Literatursuche entsprechend umfangreicher (Schulz, Maichrowitz, Arnold & Beelmann, 2014). Folgende Suchstrategien wurden hierbei verwendet:

1. Am 09.02.2013 wurden Suchanfragen an verschiedene psychologische, pädagogische, medizinische und soziologische Datenbanken gestellt und die Suchergebnisse dabei, zunächst auf Arbeiten begrenzt, welche bis Ende 2012 publiziert oder anderweitig zugänglich gemacht wurden. Die durchsuchten Datenbanken waren PsycINFO, PSYINDEX, PsycARTICLES, ProQuest Dissertations & Theses Full Text, ERIC, Web of Science, PubMed, The Cochrane Library, Sociological Abstracts und Social Services Abstracts. Gesucht wurde mit einem komplexen Kombinationssystem englischsprachiger Suchbegriffe, wobei, angelehnt an die Einschlusskriterien, jeweils mindestens ein Suchbegriff innerhalb von vier Gruppen von Suchbegriffen im Titel, in Schlüsselwörtern, Themen oder Abstract vorkommen musste. Die erste Suchbegriffsgruppe sollte dazu dienen, die Suchergebnisse auf Studien zu beschränken, die Personen mit Migrationshintergrund enthielten und umfasste verschiedene Begriffe, die einen Bezug zu Migration bzw. zu Migranten haben, sowie einige Bezeichnungen von häufig in der Literatur untersuchten Migrantengruppen und relevanten Minderheiten (z. B. *Latino* und *Asian American*). In der zweiten Gruppe lag der Fokus auf der unabhängigen Variable, weswegen Synonyme für *Intervention* verwendet wurden (z. B. *training* und *treatment*). Die dritte Gruppe von Suchbegriffen bezog sich auf das Alter der Probanden und enthielt Wörter, die mit Minderjährigen und ihren Familien assoziiert sind (z. B. *adolescents* und *parent*). Mit der letzten Suchbegriffsgruppe sollte eine erste methodische Vorselektion erreicht werden, indem Wörter verwendet wurden, welche in randomisiert-kontrollierten Studien häufig vorkommen, um das Studiendesign zu beschreiben (z. B. *experimental* und *control condition*). Diese vier Gruppen von Suchbegriffen wurden in jeder der Datenbanken verwendet. Jedoch gab es große Unterschiede zwischen den Datenbanken hinsichtlich ihrer Thesauri sowie der Suchmöglichkeiten und -beschränkungen, sodass für jede der Datenbanken ein individuell angepasster Suchbefehl verwendet werden musste. Eine detaillierte Auflistung dieser Suchbefehle ist im Anhang



A zu finden. Am 08.08.2014 erfolgte zudem eine Aktualisierung der Online-Suche. Hierbei wurden im Wesentlichen dieselben Datenbanken durchsucht, wie bei der ersten Suche und auch die gleichen Suchbefehle verwendet. Lediglich die Datenbanken ProQuest Dissertations & Theses Full Text sowie Sociological Abstracts und Social Services Abstracts wurden nicht erneut durchsucht. ProQuest Dissertations & Theses Full Text stand zu diesem Suchzeitpunkt aufgrund eingeschränkter Universitätslizenzen nicht mehr zur Verfügung. Die letztgenannten zwei Datenbanken konnten durch eine zwischenzeitlich veränderte Suchmaske nicht erneut in der selben Differenziertheit wie im ersten Suchdurchgang durchsucht werden und da sich zu diesem Zeitpunkt bereits herausstellte, dass die beiden Datenbanken kaum relevante Suchergebnisse lieferten, wurde auch hier auf eine weitere Verwendung verzichtet. Bei diesem zweiten Durchgang der Online-Suche wurde der Suchzeitraum auf Studien eingeschränkt, die entweder 2012 oder 2013 publiziert oder anderweitig zugänglich gemacht wurden. Mit der unteren Grenze sollten möglicherweise verspätet in den Datenbanken gelistete und dadurch in der ersten Suche nicht gefundene Studien erfasst werden. Die obere Grenze entspricht dem zeitlichen Einschlusskriterium.

2. Die Literaturverweise folgender zentraler Überblicksarbeiten und Meta-Analysen wurden nach relevanten Studien durchsucht: Adesope, Lavin, Thompson und Ungerleider (2011), Bledsoe (2002), Cardoza et al. (2012), Castro et al. (2006), Foxcroft und Tsertsvadze (2011), Gersten und Baker (2000), Huey und Polo (2008), Jackson (2009), Jani et al. (2009), Kong et al. (2012), Miranda et al. (2005), Perez-Johnson und Maynard (2007), Shanahan und Beck (2006), Sherman und Primack (2009), Slavin und Cheung (2005), Szapocznik und Williams (2000), Waldron und Turner (2008), Yuen (2004).

3. Relevante Primärstudien aus einer im Rahmen des Projektes parallel durchgeführten Meta-Analyse zur Prävention von Armutfolgen bei Kindern und Jugendlichen wurden ebenfalls auf Relevanz für die vorliegende Meta-Analyse geprüft (Maichrowitz, Arnold, Schulz, Beelmann, 2014).

4. Wurden in einer eingeschlossenen oder ausschließlich aufgrund (bis dahin) fehlender Post-Messung ausgeschlossenen Studie spätere möglicherweise relevante Publikationen (z.B. Follow-up-Untersuchungen) angekündigt, so wurden diese nachträglich über die Autoren sowie den Programmnamen in den oben genannten Datenbanken gesucht.

5. Abschließend wurden die Literaturverzeichnisse aller in den vorangegangenen Schritten eingeschlossenen Primärstudien durchsucht.

### **3.4 Suchergebnisse**

Die Literatursuche ergab insgesamt 9 661 Forschungsberichte. Von diesen stammten 8 933 aus elektronischen Datenbanken, 310 aus den Literaturverweisen zentraler Überblicksarbeiten und Meta-Analysen, 12 aus der Meta-Analyse zur Wirksamkeit von Prävention von Armutfolgen bei Kindern und Jugendlichen, neun wurden aufgrund von Ankündigungen in vorangegangenen Studien, über die Suche nach Autoren- und Programmnamen gefunden und 397 aus den Literaturverzeichnissen relevanter Studien entnommen. Von den insgesamt 9 661 Forschungsberichten wurden 8 075 bereits anhand ihres Titels bzw. Abstracts ausgeschlossen. 1 586 wurden anhand der Prüfung von Titel und Abstract als möglicherweise relevant eingestuft und von diesen konnten 1 480 im Volltext geprüft werden (106 waren nicht beschaffbar). Von diesen erfüllten 152 nicht das Interventionskriterium (Selektionskriterium 1; 8 davon bedienten sich keiner psychologischer Mittel, 144 verfolgten nicht das Ziel der sozio-emotionalen Förderung). Weitere 602 erfüllten die Stichprobenkriterien nicht (Selektionskriterium 2; 542 enthielten nicht mehrheitlich Personen mit Migrationshintergrund, 26 untersuchten keine Kinder oder Jugendlichen, 34 enthielten eine klinische Stichprobe). 493 der Studien mussten aufgrund methodischer Mängel ausgeschlossen werden (Selektionskriterium 3; 24 waren keine Evaluationsstudie, 10 waren Überblicksarbeiten, 68 hatten keine Kontrollgruppe, 360 wiesen die Gruppen nicht zufällig den Bedingungen zu, 31 verwendeten eine Cluster-randomisierung mit weniger als 5 Clustern pro Bedingung). Außerdem wurden 107 Forschungsberichte ausgeschlossen, da die Stichprobe weniger als 25 Teilnehmer pro Gruppe umfasste (Selektionskriterium 4) und ein Forschungsbericht enthielt keine relevanten Effektmaße (Selektionskriterium 5). Schließlich befanden sich noch 57 Doppelpublikationen unter den überprüften Volltexten. Davon wurden 14 als partielle oder vollständige Mehrfachpublikationen ausgeschlossen, da sie keine zusätzlichen relevanten Informationen enthielten und 43 waren ergänzende Doppelpublikationen, welche bei späteren Kodierungen berücksichtigt wurden. In die Meta-Analyse eingeschlossen wurden schließlich 68 Forschungsberichte, die 71 Studien mit 93 Vergleichen enthielten. Insgesamt wurden darin 16 365 Interventionsgruppenteilnehmer

und 16 345 Kontrollgruppenteilnehmer untersucht. Der komplette Selektionsprozess zur Identifikation relevanter Forschungsberichte ist in Abbildung 1 dargestellt.

### 3.5 Kodierung relevanter Primärstudien

Um die Ergebnisse der in die Meta-Analyse eingeschlossenen Forschungsberichte statistisch, also quantitativ analysieren zu können, sind Variablen notwendig, die ausgewertet und miteinander in Beziehung gesetzt werden können. Hierfür müssen verschiedene Merkmale eines Forschungsberichtes und der in ihm beschriebenen Forschungsmethoden, Stichproben und Messungen extrahiert werden. Dies geschah anhand eines Kodierschemas, welches im Rahmen des vom BMBF geförderten Projektes vor Beginn der Datenextraktion erstellt und während einer ersten Testphase wiederholt überarbeitet wurde. Das Kodierschema definierte Variablen sowie Werte, welche diese annehmen konnten und gab Bedingungen an, die festlegten welcher Wert einer Variable in einem konkreten Fall zugeordnet werden sollte. Die eigentliche Kodierung erfolgte mithilfe des Programmes *FileMaker Pro*, in welches das Kodierschema eingearbeitet wurde. Dies ermöglichte es, für jeden Forschungsbericht den Variablen aus dem Kodierschema entsprechend der aus dem Forschungsbericht zu extrahierenden Merkmale und der im Kodierschema definierten Zuordnungsregeln die passenden Werte zuzuweisen und diese Zuweisung zu speichern.

Die Kodierung der Forschungsberichte erfolgte in dem Projekt durch drei Doktoranden und drei wissenschaftliche Hilfskräfte (im Folgenden *Kodierende* genannt). Hierbei gab es zunächst ein *Kodiertraining*, in dem insgesamt 10 Primärstudien durch jeden der sechs Kodierenden kodiert wurden. Nach der Kodierung eines jeden Forschungsberichtes wurden die Kodierungen der sechs Kodierenden verglichen, Abweichungen diskutiert und die Interraterreliabilität mittels Cohens Kappa berechnet (Cohen, 1960). Ziel dieses Verfahrens war es, das Kodierschema zu optimieren, Unklarheiten zu beseitigen und eine möglichst hohe Übereinstimmung hinsichtlich der Kodierungen zu erreichen. Am Ende des Trainings konnte für jeden Vergleich zwischen Kodierenden eine Interraterreliabilität von  $\kappa > 0.8$  erzielt werden, was von Landis und Koch (1977) als *fast perfekte (almost perfect)* Übereinstimmung beurteilt wird.

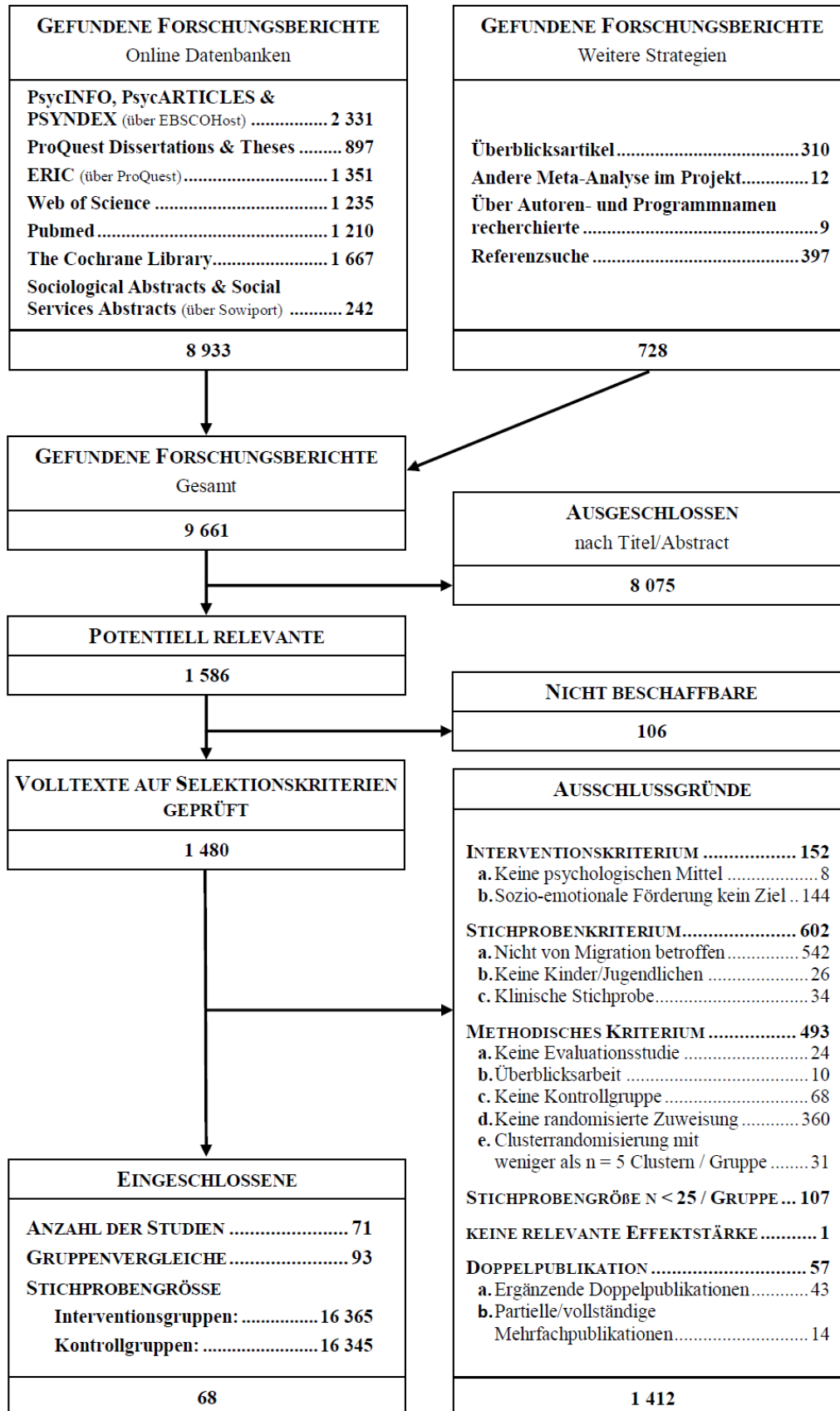


Abbildung 1. Selektionsprozess zur Identifikation relevanter Forschungsberichte.

Anschließend erfolgte die eigentliche Kodierphase, in der jeder Forschungsbericht jeweils nur noch durch einen der Kodierenden kodiert wurde. Das Kodierschema, welches im Anhang B aufgeführt ist, wurde für das gesamte Projekt verwendet und umfasste mehr als 250 verschiedene Variablen, die jedoch nicht alle in der vorliegenden Meta-Analyse Verwendung fanden. Es ist, hierarchisch absteigend, in Variablen der Dokumenten-, Vergleichs- und Effektstärkenebene unterteilt. Die eigentliche Kodierung über den *FileMaker Pro* enthielt zusätzlich eine Studienebene, die sich hierarchisch zwischen der Dokumenten- und Vergleichsebene befand. Da auf dieser Ebene jedoch keine zusätzlichen Variablen kodiert wurden, ist sie im Kodierschema nicht aufgeführt und soll hier auch nicht weiter berücksichtigt werden. Alle anderen Ebenen werden im Folgenden näher erläutert:

### *3.5.1 Dokumentenebene*

Auf der hierarchisch obersten Ebene des Kodierschemas wurden jene Variablen kodiert, welche sich auf den gesamten Forschungsbericht beziehen. Dies betraf neben dem Publikationsjahr, Titel und Autor unter anderem auch die Publikationssprache sowie einen Identifikationscode, der eine eindeutige Zuordnung innerhalb der Datenbank ermöglichte. Aber auch das Publikationsformat (z. B. Fachzeitschrift oder Dissertation) und der Ort, in der die Untersuchung stattfand, wurden auf dieser Ebene kodiert.

### *3.5.2 Vergleichsebene*

Innerhalb eines Forschungsberichtes bzw. einer Studie kann es mehrere verschiedene Interventions- und Kontrollgruppen-Kontraste geben, die untersucht wurden. Diese können sich nicht nur hinsichtlich der Art der Intervention und der Zusammensetzung der Stichprobe unterscheiden, sondern auch bezüglich methodischer Variablen. Jeder dieser Gruppenvergleiche wurde daher, insofern er die Einschlusskriterien erfüllte, jeweils separat auf der Vergleichsebene kodiert. Hierbei wurde keine Interventionsgruppe mehrfach kodiert, das heißt, dass auch beim Vorliegen mehrerer Kontrollgruppen jeweils nur der Vergleich zwischen der bzw. den Interventionsgruppe(n) und einer der Kontrollgruppen (in der Regel der am wenigsten behandelten bzw. trainierten) kodiert wurde. Die Vergleichsebene enthielt unter anderem den Identifikationscode aus der Dokumentenebene, eine Kennziffer für die Studie sowie eine individuelle Identifikations-

nummer für den jeweils betrachteten Kontrast. Damit wurde eine genaue Zuordnung ermöglicht.

Als programmbezogene Variablen wurden unter anderem der Programmtyp sowie die Programminhalte und -ziele kodiert. Ebenfalls wurde festgehalten, wer das Programm durchführte (Administrator) und in welchem Setting es stattfand. Außerdem wurde das Interventionsformat (Selbsthilfe, Gruppentraining, Einzeltraining oder Gruppen- und Einzeltraining) erfasst. Bezüglich der Programmintensität wurde kodiert, welchen Zeitraum und wie viele Sitzungen mit welcher durchschnittlichen Dauer das Programm umfasste. Hinsichtlich der kulturellen Angepasstheit wurde zudem erfasst, ob der Autor das Programm als an die Kultur der untersuchten Migrantengruppe bzw. Minderheit angepasst oder zugeschnitten bezeichnete bzw. in einer Weise beschrieb, die darauf schließen ließ oder ob keine Hinweise auf eine solche Passung vorlagen. In Bezug auf die Implementation wurden Angaben über das Training und die Supervision des Administrators, teilnehmerorientierte Angebote zur Sicherung der Compliance (z.B. finanzielle Belohnung für Teilnahme an Sitzungen, Fahrdienste zum Ort der Interventionsdurchführung oder Kinderbetreuung während des Elterntrainings) sowie die berichteten Implementationsprobleme erfasst. Dabei wurde jeweils kodiert, ob der Autor zu dem jeweiligen Merkmal keine Angabe machte, eine geringe Ausprägung oder eine große Intensität beschrieb.

Hinsichtlich der Stichprobe wurden unter anderem die Stichprobengrößen der beiden Gruppen zum Zeitpunkt der Zuweisung, Mittelwert und Spannweite des Alters sowie die Geschlechtsverteilung der Kinder bzw. Jugendlichen in den Gruppen erfasst. Kodiert wurden als Stichprobenvariablen außerdem vorhandene Risikofaktoren und ob die Stichprobe aus einer urbanen oder ländlichen Gegend stammte. Hinsichtlich des Migrationshintergrundes wurde erfasst, ob es sich um eine ethnische Minderheit oder um eine Gruppe handelte, die eine in dem Untersuchungsland für Migranten typische Muttersprache sprachen oder ob explizit angegeben wurde, dass die Stichprobe mehrheitlich aus Personen mit Migrationshintergrund oder Migranten der ersten Generation stammte. Falls genaue Angaben dazu existierten, wurde auch kodiert, wie viel Prozent der Stichprobe der ersten Migrantengeneration angehörten und wie viele einen Migrationshintergrund hatten. Auch wurde erfasst, ob die kodierte, mehrheitlich aus Personen mit Migrationshintergrund bestehende Stichprobe die Gesamtheit der in der Studie untersuchten Stichprobe umfasste oder hiervon nur eine Teilmenge ausmachte, für

die separate Ergebnisse berichtet wurden. Und schließlich wurden vorhandene Angaben zur Muttersprache, der ethnischen Gruppe sowie zum Herkunftsland der Kinder und Jugendlichen bzw. deren Eltern festgehalten und kodiert, ob in der Region, in der die Stichprobe rekrutiert wurde, die Mehrzahl der Bevölkerung keinen Migrationshintergrund hatte, einen anderen oder den gleichen Migrationshintergrund hatte wie die untersuchte Stichprobe.

Als methodische Variablen wurden die Art der Randomisierung (individuell oder Cluster-Randomisierung), die Art der Kontrollgruppe und die Dropout-Rate kodiert. Auch wurde festgehalten, ob die Vergleichbarkeit der Gruppen vor der Intervention getestet wurde und wenn ja, ob sie als vergleichbar oder nicht vergleichbar angesehen werden konnten. Im Falle von Dropouts wurde außerdem kodiert, ob diese in die Analysen eingeschlossen wurden (z.B. durch Imputationsverfahren; Intention-to-treat), ob nur jene analysiert wurden, die an der Post-Testung teilnahmen (abgeschwächte Form des *Per-protocol*) oder ob nur diejenigen in den Analysen berücksichtigt wurden, die zudem ein Mindestmaß an Intervention erhalten hatten (stärkere Form des *Per-protocol*).

### 3.5.3 Effektstärkenebene

Für jeden der Interventions-Kontrollgruppen-Kontraste konnten verschiedene Effektmaße erfasst worden sein. Jedes entsprechend der Einschlusskriterien relevante Effektmaß wurde daher auf der hierarchisch untersten Ebene des Kodierschemas separat kodiert, wobei wiederum Identifikationsvariablen zur eindeutigen Zuordnung zu einem Interventions-Kontrollgruppen-Kontrast, einer Studie und dem entsprechenden Forschungsbericht vorhanden waren. Auf der Effektstärkenebene wurden die Effektstärkenmaße zum einen in Konstrukte und Konstruktkategorien eingeordnet, zum anderen aber auch Informationen zur Standardisierung, Messmethodik sowie zur konzeptionellen Nähe des Maßes zu der durchgeführten präventiven Intervention kodiert. Die übergeordneten Konstruktkategorien umfassten hierbei neben sozialen und emotionalen Parametern auch die kognitive Entwicklung (Leistungsparameter, wie z.B. Intelligenz, Lese- oder Matheleistung), bildungsbezogene Maße (z.B. Lernverhalten, Schulnoten oder Schulabbrüche), den schulischen Kontext (z.B. Lehrer-Schüler-Beziehung, Lehrerverhalten und -fähigkeiten) sowie den familiären Kontext (z.B. Erziehungseinstellungen und -verhalten der Eltern, Eltern-Kind-Beziehung und Anzahl der Bücher im Haus). Soziale und emotionale Konstrukte waren internalisierende Probleme, externalisierende

Probleme, selbstbezogene Reflektion, prosoziale Kognitionen und Verhaltensweisen, Sexualverhalten, Suchtmittelkonsum, allgemeine Verhaltensprobleme sowie sonstige soziale und emotionale Maße. Festgehalten wurde zudem in welchem zeitlichen Abstand zur Intervention die Messung durchgeführt wurde. Zu den wohl wichtigsten Variablen gehörten die Informationen zur Höhe der Effektstärke, welche je nach Angaben in der Studie entweder über Mittelwerte und Standardabweichungen oder aber direkt als fertig berechnete Effektstärke  $d$  mit zusätzlicher Information darüber, aus welchen statistischen Parametern die Effektstärke rückgerechnet bzw. geschätzt wurde, in die Datenbank eingegeben werden konnten. In jedem Fall wurde jedoch, jeweils für die Interventions- und die Kontrollgruppe getrennt, die Stichprobengröße kodiert, welche bei der Messung erfasst wurde bzw. auf welche sich die Effektstärke bezog, was für die spätere Berechnung des Standardschätzfehlers der Effektstärke unabdingbar war. Auch wurde erfasst, ob hohe Werte des Effektmaßes als positiv bzw. wünschenswert oder negativ bzw. unvorteilhaft zu bewerten waren. Zur Beurteilung wurden dabei bevorzugt explizite oder implizite Hinweise in der Studie selbst zugrunde gelegt. Weitere Informationen zur Ermittlung und Aufbereitung der Effektstärken sind im folgenden Kapitel zu finden.

### **3.6 Statistische Auswertung**

#### *3.6.1 Bestimmung der Roheffektstärken*

Wie bereits beschrieben, konnten für jede der Interventions-Kontrollgruppen-Kontraste mehrere Effektmaße verwendet worden sein. Eine Effektstärke, die sich auf ein einzelnes solches Erfolgsmaß bezieht, soll im folgenden Roheffektstärke genannt werden.

Die in den Studien angegebenen statistischen Parameter oder getroffenen Aussagen, die Auskunft über die Wirksamkeit einer präventiven Intervention bezüglich eines Erfolgsmaßes geben sollen, können sehr unterschiedliche sein bzw. verschiedene Formate annehmen. So können beispielsweise in einer Studie Mittelwerte und Standardabweichungen für Kontroll- und Interventionsgruppe zu den Zeitpunkten vor (*Prä*) und nach (*Post*) der Intervention angegeben sein, in der nächsten ein  $t$ -Wert für den Mittelwertunterschied zum post-Zeitpunkt berichtet werden und in wieder einer anderen lediglich erwähnt werden, dass alle Effekte nicht signifikant waren. Auch einzelne Erfolgsmaße innerhalb einer Studie können in dieser Art unterschiedlich ausgewertet bzw. berichtet werden. Um die Effekte integrieren zu können, ob nun jene von einzelnen



Erfolgsmaßen innerhalb einer Studie bzw. eines Gruppenvergleichs oder zur Berechnung eines mittleren Effektes über alle Studien hinweg, ist es jedoch notwendig, dass für jedes einzelne Erfolgsmaß der gleiche statistische Parameter verwendet wird, sodass die zu integrierenden Werte dieselbe Metrik aufweisen. Einen möglichen solchen Parameter stellt die *standardisierte Mittelwertdifferenz* dar (im Englischen *standardized mean difference*; Lipsey & D. B. Wilson, 2001). Die Terminologie ist an vielen Stellen uneinheitlich (vergleiche z. B. Cohen, 1988, Hedges & Olkin, 1985 sowie Lipsey & D. B. Wilson, 2001). Jedoch beruhen alle Berechnungen auf folgender grundlegender Formel:

$$\delta = \frac{\mu_{IG} - \mu_{KG}}{\sigma} \quad (1)$$

Hierbei bezeichnet  $\delta$  die Effektstärke,  $\mu_{IG}$  und  $\mu_{KG}$  jeweils die Populationsmittelwerte der Interventions- und Kontrollgruppe und  $\sigma$  die Standardabweichung der beiden Populationen. Folglich gibt die Effektstärke den Mittelwertunterschied zwischen beiden Gruppen in Relation zur Standardabweichung an und liefert damit eine standardisierte Größe, welche einen Vergleich zweier Gruppen, unabhängig von der Stichprobe, erlaubt. Hierbei können Effektstärken von  $\delta = 0.2$  als *klein*,  $\delta = 0.5$  als *mittel* und  $\delta = 0.8$  als *groß* gelten (Cohen, 1988). Während die Populationsmittelwerte der Gruppen typischerweise über die Stichprobenmittelwerte geschätzt werden, existieren hinsichtlich der Populationsstandardabweichung verschiedene Empfehlungen. Glass, McGaw und Smith (1981) schlugen hierbei vor, die Standardabweichung der Kontrollgruppe zu verwenden, da die Standardabweichung in der Interventionsgruppe durch die Intervention selbst beeinflusst sein kann. Dies führt in der Regel jedoch zu einer Verdopplung des Standardfehlers der Standardabweichung im Vergleich zu Verfahren, in denen die Standardabweichung beider Gruppen herangezogen wurde, da die Stichprobengröße der Gesamtstichprobe größer ist als die der Kontrollgruppe allein (Hunter & Schmidt, 1995). Hedges und Olkin (1985) verwenden daher als Schätzer für die Populationsstandardabweichung die gepoolte Standardabweichung der Interventions- und Kontrollgruppe. Hieraus ergaben sich folgende Formeln zur Berechnung der standardisierten Mittelwertdifferenz (siehe Lipsey & D. B. Wilson, 2001):

$$ES_{sm} = \frac{\bar{X}_{IG} - \bar{X}_{KG}}{SD_{pooled}} \quad (2)$$

$$SD_{pooled} = \sqrt{\frac{(n_{IG} - 1) \times SD_{IG}^2 + (n_{KG} - 1) \times SD_{KG}^2}{n_{IG} + n_{KG} - 2}} \quad (3)$$

$ES_{sm}$  bezeichnet hierbei die standardisierte Mittelwertdifferenz,  $\bar{X}_{IG}$  und  $\bar{X}_{KG}$  die Stichprobenmittelwerte der Interventions- und Kontrollgruppe,  $SD_{pooled}$  die gepoolte Standardabweichung,  $n_{IG}$  und  $n_{KG}$  die Stichprobengrößen der Interventions- und Kontrollgruppe und  $SD_{IG}^2$  und  $SD_{KG}^2$  die Varianzen (quadrierte Standardabweichungen) innerhalb der Interventions- und Kontrollgruppe.

Hedges (1981) konnte jedoch zeigen, dass die daraus folgende Effektstärke besonders für kleinere Stichproben nach oben verzerrt ist. Daher wird empfohlen, diese verzerrte standardisierte Mittelwertdifferenz mit einem Korrekturfaktor  $J$  zu multiplizieren, um einen unverzerrten Schätzer zu erhalten (Hedges & Olkin, 1985; Lipsey & D. B. Wilson, 2001). Dieser Korrekturfaktor errechnet sich folgendermaßen:

$$J = \left(1 - \frac{3}{4N - 9}\right) \quad (4)$$

$N$  steht dabei für die Gesamtstichprobengröße von Interventions- und Kontrollgruppe.

Der daraus resultierende unverzerrte Schätzer der standardisierten Mittelwertdifferenz, soll Hedges und Olkins (1985) folgend mit  $d$  bezeichnet werden und hat folgende Formel:

$$d = \left(1 - \frac{3}{4N - 9}\right) \times \left[ \frac{\bar{X}_{IG} - \bar{X}_{KG}}{\sqrt{\frac{(n_{IG} - 1) \times SD_{IG}^2 + (n_{KG} - 1) \times SD_{KG}^2}{n_{IG} + n_{KG} - 2}}} \right] \quad (5)$$

Diese Formel wurde in dieser Meta-Analyse für alle Effektstärken zugrunde gelegt, die ausschließlich auf dem Vergleich zweier Gruppen zu einem Zeitpunkt beruhen. In Interventionsstudien wird jedoch häufig auch ein Prä-Post-Kontrollgruppendesign verwendet, in dem nicht nur nach der Intervention Messungen in der Interventions- und Kontrollgruppe durchgeführt werden (Post-Test-Messung), sondern das selbe Maß auch vor Beginn der Evaluation in beiden Gruppen erfasst wird (Prä-Test-Messung). In diesen Fällen sollten auch die Prä-Test-Messungen bei der Effektstärkenberechnung berücksichtigt werden, da somit möglicherweise zufällig entstandene Gruppenunterschiede, die bereits vor der Intervention existierten, besser ausgeglichen werden können. Auch hier existieren verschiedene Möglichkeiten, die vier zur Verfügung stehenden Standardabweichungen zur Berechnung der Effektstärke heranzuziehen. Um das passende Vorgehen herauszufinden, verglich Morris (2008) drei unterschiedliche Methoden zur Schätzung der Effektstärke für dieses Design hinsichtlich Robustheit,

Heterogenität, Genauigkeit und möglicher Verzerrungen. Er kam dabei zu dem Schluss, dass die Verwendung der gepoolten Standardabweichung der Prä-Test-Messungen den besten Schätzer darstellt. Die Formel zur Berechnung der Effektstärke in einem Prä-Post-Kontrollgruppendesign lautet demnach:

$$d = \left(1 - \frac{3}{4N - 9}\right) \times \left[ \frac{(\bar{X}_{post,IG} - \bar{X}_{prä,IG}) - (\bar{X}_{post,KG} - \bar{X}_{prä,KG})}{\sqrt{\frac{(n_{IG} - 1) \times SD_{prä,IG}^2 + (n_{KG} - 1) \times SD_{prä,KG}^2}{n_{IG} + n_{KG} - 2}}} \right] \quad (6)$$

$\bar{X}_{post,IG}$ ,  $\bar{X}_{prä,IG}$ ,  $\bar{X}_{post,KG}$  und  $\bar{X}_{prä,KG}$  bezeichnen hier jeweils die Mittelwerte der Interventionsgruppe bei Post-Test-Messung und Prä-Test-Messung bzw. der Kontrollgruppe zu den beiden Messzeitpunkten.  $SD_{prä,IG}^2$  und  $SD_{prä,KG}^2$  stehen für die Varianzen der Interventions- und Kontrollgruppe zur Prä-Test-Messung.

Die Verwendung der gepoolten Prä-Standardabweichung hat mit Hinblick auf die oben erwähnte Diskussion den doppelten Vorteil, dass durch die Verwendung der Standardabweichungen beider Gruppen die Gesamtstichprobe zur Schätzung herangezogen wird, wodurch der Standardfehler nicht unnötig erhöht wird und zudem die Standardabweichung der Interventionsgruppe bei der Prä-Test-Messung noch nicht durch die Intervention selbst beeinflusst sein kann. Diese Formel wurde daher in Prä-Post-Kontrollgruppendesigns bei der Berechnung oder Schätzung der Effektstärke zugrunde gelegt.

Wann immer Mittelwerte und Standardabweichungen angegeben waren, wurden diese zur Berechnung der Roheffektstärken genutzt. In Fällen, wo dies nicht der Fall war, aber andere statistische Parameter berichtet wurden, die eine Rückrechnung ermöglichten, wurden diese entsprechend auf den unverzerrten Schätzer der standardisierten Mittelwertdifferenz rückgerechnet (Lypsey & D. B. Wilson, 2001). Dies war beispielsweise für  $t$ -Werte und  $F$ -Werte aus einfaktoriellen Varianzanalysen jeweils in Kombination mit den Stichprobengrößen der Fall, aber auch für Korrelationskoeffizienten sowie für Häufigkeitsangaben bei dichotomen Daten. Während solche Rückrechnungen relativ präzise möglich sind, gilt dies nicht für Fälle, in denen Drittvariablen in die Berechnung der statistischen Parameter eingeflossen sind. Dies trifft beispielsweise auf multiple Regressionen zu, in denen die Gruppenzugehörigkeit nur eine von mehreren Prädiktorvariablen ist, aber auch auf Kovarianzanalysen. In Fällen, in denen ausschließlich

statistische Parameter dieser Art für ein Erfolgsmaß berichtet wurden, sollten die Effektstärken jedoch dennoch geschätzt werden. Hierzu wurde, falls dies von den Autoren nicht anders berichtet wurde, davon ausgegangen, dass sämtliche Variablen, welche vor der Intervention erfasst wurden, mit der dichotomen Gruppenvariable vollkommen unkorreliert waren ( $r = 0$ ). Diese Grundannahme erschien angemessen, da die Zuweisung der Teilnehmer zu den Gruppen in allen eingeschlossenen Studien zufällig erfolgte. Mit dieser Grundannahme konnten standardisierte Regressionskoeffizienten in multiplen Regressionen als Korrelationskoeffizienten betrachtet und als solche rückgerechnet werden (siehe Wirtz & Nachtigall, 1998, S. 177). In Fällen, in denen nur der unstandardisierte Regressionskoeffizient angegeben war oder nur Kovarianzanalysen für ein Erfolgsmaß berechnet wurden, wurde die Effektstärke über die Verwendung des  $t$ - bzw.  $F$ -Wertes geschätzt. Effektstärken für Erfolgsmaße, für welche keiner dieser Parameter verfügbar war, wurden anderweitig geschätzt. Wurden Signifikanzniveaus berichtet, so wurde, falls in der Studie nicht anders dargestellt, von einem  $t$ -Test ausgegangen und die Effektstärke konservativ geschätzt. Dies bedeutet, dass für nicht signifikante Ergebnisse die Effektstärke gleich Null gesetzt und beispielsweise für Ergebnisse, welche bei einem Signifikanzniveau von  $\alpha = .05$  als signifikant berichtet wurden, ein  $p$ -Wert von  $.05$  als Grundlage für die Rückrechnung genutzt wurde. Zwar kann dies zu einer Unterschätzung der mittleren Effektstärke sowie deren Varianz führen, jedoch könnte ein Nichtkodieren solcher Effektstärken zu einer Überschätzung der mittleren Effektstärke führen, da nicht davon auszugehen ist, dass die Art, in der Effekte berichtet werden, zufällig erfolgt (Pigott, 2001; Weisburd, Lum & Yang, 2003).

Wenn für dieselbe Wirksamkeitsvariable mehrere statistische Parameter angegeben wurden, so wurde zur Berechnung bzw. Schätzung der Effektstärke in absteigender Reihenfolge nach folgender Präferenz vorgegangen: (a) Mittelwerte und Standardabweichungen zum Prä- und Post-Messzeitpunkt in beiden Gruppen; (b) Mittelwerte und Standardabweichungen zum Post-Messzeitpunkt in beiden Gruppen; (c) vom Autor berichtete Effektstärke; (d)  $t$ -Wert oder  $F$ -Wert aus einer einfaktoriellen Varianzanalyse für den Vergleich der Interventions- mit der Kontrollgruppe zum Post-Zeitpunkt bzw. ein genauer  $p$ -Wert, der sich auf einen solchen Vergleich bezieht; (e) Korrelationskoeffizient oder Regressionskoeffizient aus einer einfachen linearen Regression, der den Zusammenhang zwischen Gruppenzugehörigkeit (Interventions- oder Kontrollgruppe) und dem Effektmaß angibt; (f) Odds Ratio oder Häufigkeitsangaben für Interventions- und

Kontrollgruppe bei dichotomen Variablen; (g) standardisierter Regressionskoeffizient in einer multiplen Regression; (h) unstandardisierter Regressionskoeffizient mit dazugehörigem Standardfehler in einer multiplen Regression oder  $F$ -Wert in einer mehrfaktoriellen Varianzanalyse bzw. einer Kovarianzanalyse, der sich auf den Vergleich der Interventions- mit der Kontrollgruppe bezieht; (i) Aussage, dass die Intervention keinen signifikanten Effekt auf das Effektmaß hat.

Jede Effektstärke wurde zunächst so kodiert, dass eine positive Effektstärke höhere Variablenwerte in der Interventions-, verglichen mit der Kontrollgruppe bedeutete. Zusätzlich existierte eine Variable, die, je nachdem ob hohe Variablenwerte auf dem Effektmaß als positiv oder negativ zu beurteilen waren, mit 1 bzw. -1 kodiert war. War unklar, ob ein Effekt als positiv oder negativ zu beurteilen war und waren auch im Text keinerlei Aussagen diesbezüglich zu finden, so wurde das Effektmaß von jeder weiteren Analyse ausgeschlossen. Die verbleibenden Effektstärken wurden mit der genannten Variable multipliziert. Die resultierende Effektstärke, welche schließlich für alle weiteren Analysen genutzt wurde, nahm folglich immer dann einen positiven Wert an, wenn die Intervention im Vergleich zur Kontrollgruppe zu einer Verbesserung hinsichtlich des Erfolgskriteriums führte und war jeweils dann negativ, wenn die Interventionsgruppe schlechtere Werte hinsichtlich dieses Maßes aufwies.

### *3.6.2 Integration der Roheffektstärken*

Wie bereits erwähnt, kann ein Forschungsbericht bzw. eine Studie mehrere Interventions-Kontrollgruppen-Kontraste enthalten und für jeden dieser Kontraste können wiederum mehrere Roheffektstärken für unterschiedliche Erfolgsmaße berichtet werden. Da der Vergleich zwischen einer Interventions- und einer Kontrollgruppe im Zentrum der Untersuchung steht, ist die Vergleichsebene die Hauptanalyseebene in dieser Meta-Analyse. Daher mussten die Roheffektstärken innerhalb eines Interventions-Kontrollgruppen-Kontrastes für die weiteren Analysen zunächst aggregiert werden. Da jedoch die Wirksamkeit einer Intervention in unterschiedlichen Bereichen von Erfolgsmaßen unterschiedlich ausfallen kann und auch eine Abhängigkeit der Wirkung vom zeitlichen Abstand zwischen der Intervention und der Erfolgsmessung in Betracht gezogen werden muss, wurde hierbei folgendermaßen vorgegangen: Die Erfolgsmaße wurden zunächst anhand der kodierten Konstrukte und Konstruktkategorien zusammengefasst. Da in dieser Arbeit der Fokus auf der sozio-emotionalen Förderung der Kinder

und Jugendlichen lag, wurde hinsichtlich der Konstruktkategorie der sozialen und emotionalen Parameter die detailliertere Unterscheidung in die, in Kapitel 3.5.3 beschriebenen, acht Konstrukte verwendet. Die anderen vier Konstruktkategorien, kognitive Entwicklung, bildungsbezogene Maße, schulischer Kontext und häuslicher Kontext, wurden aufgrund der Korrelation zwischen den verschiedenen Entwicklungsbereichen und Kontexten zwar in die Analysen einbezogen, jedoch im Hinblick auf den Schwerpunkt der vorliegenden Arbeit nicht hinsichtlich des genauen Konstruktes unterschieden. Für diese Arbeit ergaben sich demnach 12 Kategorien von Erfolgskriterien (vier Konstruktkategorien und acht soziale oder emotionale Konstrukte). Neben dieser Kategorisierung der Erfolgskriterien wurden die Messungen zusätzlich hinsichtlich der Zeit, die zwischen der Intervention und der Erfassung des Effektmaßes lag, unterschieden. Da in den verschiedenen Studien unterschiedliche Messzeitpunkte gewählt wurden, für eine Integration von Befunden jedoch ein gewisser Grad an Standardisierung erforderlich ist, wurden in der vorliegenden Meta-Analyse drei Messzeiträume festgelegt. Der erste Messzeitraum umfasste Messungen, die spätestens drei Monate nach Ende der Intervention durchgeführt wurden und stellt somit Kurzeffekte dar. Der zweite Messzeitraum enthielt Messungen, die länger als 3 Monate, aber höchstens 12 Monate nach Interventionsende erfasst wurden und spiegelt mittelfristige Effekte wieder. Der dritte und letzte Messzeitraum stellt Langzeiteffekte dar und umfasste jene Messungen, die länger als 12 Monate nach dem Ende der Intervention durchgeführt wurden. Unabhängig von der Bezeichnung in den Primärstudien soll im Folgenden der erste Messzeitraum auch als Postmessung, der zweite als erste Follow-up-Messung und der dritte als zweite Follow-up-Messung bezeichnet werden. Zur Integration der Daten wurde zunächst innerhalb jedes Kontrastes, jeweils getrennt nach den drei Zeiträumen und den zwölf Kategorien von Erfolgskriterien aggregiert, indem die Roheffektstärken gemittelt wurden. Hierdurch konnten jedem Kontrast bis zu 36 ( $3 \text{ Zeiträume} \times 12 \text{ Kategorien}$ ) aggregierte Effektstärken zugeordnet werden. Diese wurden anschließend wiederum für jeden Kontrast innerhalb der Zeiträume aggregiert, indem die Effektstärken der bis zu 12 Kategorien jeweils gemittelt wurden. So entstanden für jeden Interventions-Kontrollgruppen-Kontrast bis zu drei aggregierte Effektstärken (für jeden Messzeitraum eine).

### 3.6.3 Gewichtung der Effektstärken und Schätzung der Populationseffektstärke

Um die Ergebnisse der verschiedenen Studien meta-analytisch integrieren zu können, war es notwendig, die auf Vergleichsebene aggregierten Effektstärken rechnerisch zusammenzufassen. Lipsey und D. B. Wilson (2001) schlugen hierfür vor, die Effektstärken zunächst zu gewichten, bevor sie gemittelt bzw. aus ihnen eine Populationseffektstärke geschätzt wird. Die Effektstärken sollten dabei so gewichtet werden, dass jene Ergebnisse stärker in die Berechnung einfließen, die eine genauere Schätzung für die Populationseffektstärke liefern, also einen möglichst kleinen Stichprobenfehler beinhalten. Folglich wurde von Lipsey und D. B. Wilson (2001) als Gewichtungsfaktor ( $w_i$ ) die Inverse des quadrierten Standardfehlers empfohlen:

$$w_i = \frac{1}{SE_{d_i}^2} \quad (7)$$

Der Standardfehler ( $SE_{d_i}$ ) errechnet sich dabei über folgende Formel:

$$SE_d = \sqrt{\frac{n_{IG} + n_{KG}}{n_{IG}n_{KG}} + \frac{d^2}{2(n_{IG} + n_{KG})}} \quad (8)$$

Der Standardfehler ist somit insbesondere bei kleineren Effektstärken im Wesentlichen eine Funktion der Stichprobengröße. Der Gewichtungsfaktor  $w_i$ , der einem Kontrast zugewiesen wird, ist folglich umso größer, je größer die Stichprobe und je kleiner daher der Standardfehler ist.

Die Schätzung der Populationseffektstärke erfolgte mithilfe von SPSS-Macros, deren Syntaxen von Lipsey und D. B. Wilson (2001) in Textform veröffentlicht und von D. B. Wilson (2010) als Dateien frei zugänglich gemacht wurden. Da die Effektstärken in der vorliegenden Meta-Analyse auf Ebene der Vergleiche aggregiert wurden, handelte es sich hierbei entsprechend um die primäre Analyseebene, auf welcher die SPSS-Macros Anwendung fanden.

Zur Schätzung einer Populationseffektstärke gibt es zunächst zwei grundlegende Modelle, welche hierfür in Frage kommen: Das *Fixed-Effects-Model* und das *Random-Effects-Model*. Ersteres geht von der Grundannahme aus, dass alle zu integrierenden Effektstärken auf eine gemeinsame Populationseffektstärke zurückzuführen sind und nur auf Grund des Stichprobenfehlers um diesen herum schwanken. Demzufolge wird von einer grundsätzlichen Homogenität der Effektstärken ausgegangen und jegliche Abweichung von der Populationseffektstärke auf den Prozess der Stichprobenziehung

zurückgeführt. Dies bedeutet, dass die Varianz der Effektstärken durch keine anderen Variablen erklärbar ist als durch den Standardfehler. Daraus folgt auch, dass der beste Schätzer für die Populationseffektstärke die gewichtete mittlere Effektstärke ( $d_+$ ) ist, die sich aus den Effektstärken der einzelnen Vergleiche ( $d_i$ ) sowie den dazugehörigen Gewichtungsfaktoren ( $w_i$ ) wie folgt errechnet (Lipsey & D. B. Wilson, 2001):

$$d_+ = \frac{\sum_{i=1}^k (w_i \times d_i)}{\sum_{i=1}^k w_i} \quad (9)$$

Hierbei bezeichnet  $k$  die Gesamtzahl an Vergleichen.

Die Gültigkeit bzw. Anwendbarkeit dieses statistischen Modells ist jedoch von der Homogenitätsannahme abhängig. Die Homogenität der Effektstärken und somit die Passung des Fixed-Effects-Modells kann mithilfe der Q-Statistik ermittelt werden. Diese errechnet sich folgendermaßen:

$$Q = \sum_{i=1}^k w_i (d_i - d_+)^2 \quad (10)$$

Die Teststatistik  $Q$  ist Chi<sup>2</sup>-verteilt. Überschreitet  $Q$  den kritischen Chi<sup>2</sup> Wert bei  $k - 1$  Freiheitsgraden, so muss die Null-Hypothese der Homogenität verworfen werden. Eine derartige signifikante Heterogenität würde gegen die Annahmen des Fixed-Effects-Modells verstoßen, weswegen dieses in einem solchen Fall nicht angewendet und stattdessen das Random-Effects-Modell zugrunde gelegt werden sollte (Lipsey & D. B. Wilson, 2001).

Neben dieser Möglichkeit der Signifikanztestung, mithilfe derer beurteilt werden kann, ob Heterogenität vorliegt, schlugen Higgins und Thompson (2002) außerdem eine Verwendung des Parameters  $I^2$  vor. Dieser gibt den relativen Anteil der Heterogenität, also jener Varianz, die nicht auf Stichprobenfehler zurückführbar ist, an der Gesamtvarianz an. Der Parameter  $I^2$  hat den Vorteil, dass er, im Gegensatz zum Signifikanztest, unabhängig von der Anzahl der Vergleiche ist und zudem zusätzlich eine quantitative Aussage darüber erlaubt, wie groß die Heterogenität ist. Ist  $Q \leq k - 1$ , so gilt  $I^2 = 0$ ; in allen anderen Fällen errechnet sich  $I^2$  mit folgender Formel:

$$I^2 = \frac{Q - (k - 1)}{Q} \times 100\% \quad (11)$$



Als Interpretationshilfen geben Huedo-Medina, Sánchez-Meca, Marín-Martínez und Botella (2006) an, dass die Heterogenität bei einem  $I^2$  von 25 % als *klein*, 50 % als *mittel* und 75 % als *groß* eingeschätzt werden kann.

Das Random-Effects-Model geht davon aus, dass die Variabilität in den Vergleichseffektstärken nicht ausschließlich auf Stichprobenfehler zurückzuführen ist, sondern zumindest ein Teil der Varianz auf zufällige Unterschiede zwischen den Gruppenvergleichen zurückgeht. Grundlage dieser Annahme ist dabei, dass nicht nur innerhalb jedes Gruppenvergleiches eine Zufallsstichprobe von möglichen Teilnehmern gezogen wird, deren Stichprobeneffektstärke um die Populationseffektstärke des Gruppenvergleiches schwanken kann (Stichprobenfehler), sondern, dass auch die erfassten Vergleiche eine Zufallsstichprobe aus einer größeren Population möglicher Gruppenvergleiche darstellen, deren Vergleichseffektstärken ebenfalls variieren können (Cooper et al., 2009; Lipsey & D. B. Wilson, 2001). Die Varianz in den wahren Vergleichseffektstärken werden in dem Random-Effects-Model jedoch zunächst nicht durch spezifische Merkmale der Gruppenvergleiche erklärt, sondern es wird lediglich eine zusätzliche Varianzkomponente zum Modell hinzugefügt. Da zur Berechnung des Gewichtungsfaktors jedoch die komplette Varianz, die mit der Verteilung der Effektstärke einhergeht, berücksichtigt werden soll, errechnet sich dieser nicht einfach wie im Fixed-Effects-Model als inverse des Stichprobenfehlers entsprechend der Formel aus (7). Stattdessen geht im Random-Effects-Model die Varianz in den wahren Vergleichseffektstärken in die Berechnung des Gewichtungsfaktors ein. Es ergibt sich folgende Formel:

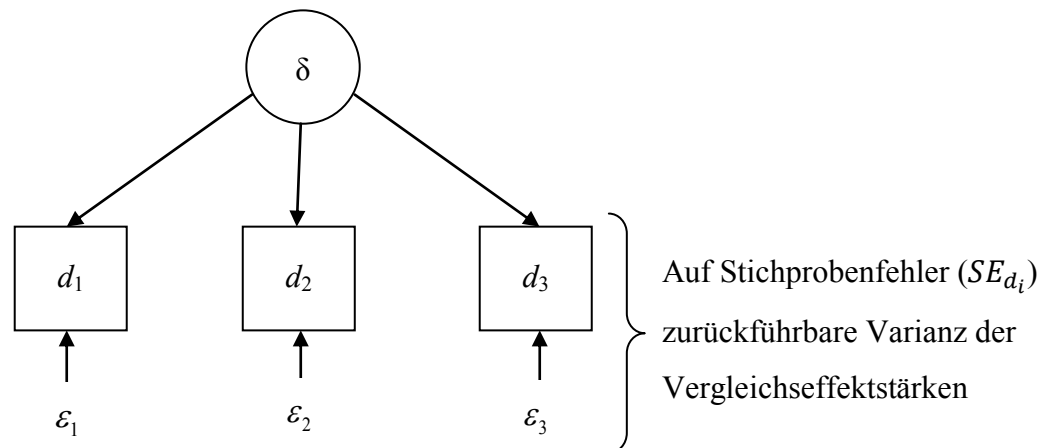
$$w_i' = \frac{1}{SE_{d_i}^2 + v_\theta} \quad (12)$$

Dabei stellt  $v_\theta$  die Varianz der wahren Vergleichseffektstärken dar. Sie errechnet sich aus der  $Q$ -Statistik und dem Gewichtungsfaktor aus dem Fixed-Effects-Model (Lipsey & D. B. Wilson, 2001):

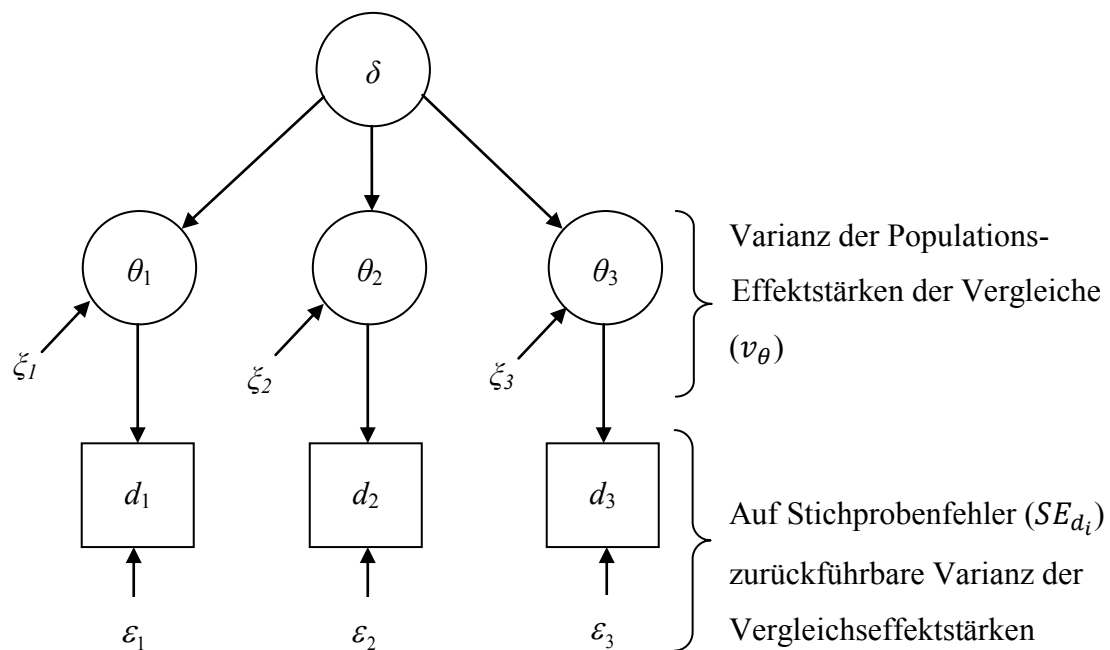
$$v_\theta = \frac{Q - (k - 1)}{\sum_{i=1}^k w_i - \left( \sum_{i=1}^k w_i^2 / \sum_{i=1}^k w_i \right)} \quad (13)$$

Im Random-Effects-Model wird die gewichtete mittlere Effektstärke nun anhand des neu berechneten Gewichtungsfaktors  $w_i'$  entsprechend der Formel (9) errechnet. Abbildung 2 gibt einen Überblick über beide Modelle.

## Fixed-Effects-Model



## Random-Effects-Model



Legende:

$\delta$  = (mittlere) Populationseffektstärke;  $\theta_i$  = Populationseffektstärke des Vergleichs  $i$ ;  
 $\zeta_i$  = alle vergleichsspezifischen Einflüsse auf die Effektstärke;  $d_i$  = gemessene  
 Vergleichseffektstärke der Studie  $i$ ;  $\varepsilon_i$  = Stichprobenfehler des Vergleichs  $i$

Abbildung 2. Darstellung des Fixed-Effects-Modells im Vergleich zum Random-Effects-  
 Model.

### 3.6.4 Moderatoranalysen

Insbesondere bei einer hohen Heterogenität der Ergebnisse kann es sinnvoll sein, Moderatoranalysen durchzuführen. Diese stellen in gewisser Weise eine Alternative zum Random-Effects-Model dar. Hierbei wird versucht, mögliche Unterschiede in den Populationseffektstärken der Gruppenvergleiche nicht nur wie im Random-Effects-Model als zufällig verteilt zu modellieren, sondern diese stattdessen anhand konkreter Merkmale der Gruppenvergleiche, ihrer Methodik, der in Ihnen untersuchten Stichproben oder der verwendeten Operationalisierungen für die unabhängigen und abhängigen Variablen zu erklären. Je höher die Heterogenität der Effektstärken zwischen den Gruppenvergleichen, desto höher ist der Anteil der tatsächlich zu erklärenden Varianz in den Populations-effektstärken an der Gesamtvarianz der Vergleichseffektstärken und umso eher sollten sich auch Moderatoren identifizieren lassen, die diese Unterschiede erklären. Allerdings sollte ein signifikanter Heterogenitätstest ausdrücklich nicht als eine Vorbedingung für Moderatoranalysen angesehen werden, insbesondere wenn die Moderatoren a-priori spezifiziert wurden (Bosnjak & Viechtbauer, 2009).

Im Rahmen einer Meta-Analyse gibt es zunächst zwei mögliche Varianten von Moderatoranalysen. Die erste Möglichkeit ist eine meta-analytische Varianzanalyse (Hedges, 1982). Durch diese kann geprüft werden, ob eine kategoriale Variable einen Einfluss auf die Höhe der Effektstärke hat und folglich, je nach Ausprägung der Variable, die mittleren Effektstärken signifikant unterschiedlich groß sind. Hierzu werden zunächst angelehnt an Formel (9) die gewichteten mittleren Effektstärken für die unterschiedlichen Ausprägungen der Variable errechnet:

$$d_{j+} = \frac{\sum_{i=1}^k (w_{ji} \times d_{ji})}{\sum_{i=1}^k w_{ji}} \quad (14)$$

Hierbei stellt  $j$  die jeweilige Ausprägung der kategorialen Variable mit insgesamt  $m$  möglichen Ausprägungen dar. Die Summe aus den  $Q$ -Werten innerhalb der Kategorien ( $Q_W$ ) ergibt:

$$Q_W = \sum_{j=1}^m \sum_{i=1}^k w_{ji} (d_{ji} - d_{j+})^2 \quad (15)$$

Der  $Q$ -Wert zwischen den Kategorien ( $Q_B$ ) errechnet sich sodann folgendermaßen:

$$Q_B = \sum_{j=1}^m \sum_{i=1}^k w_{ji} (d_{j+} - d_{+})^2 \quad (16)$$

Dabei gilt folgender Zusammenhang zwischen den entsprechenden  $Q$ -Werten:

$$Q = Q_B + Q_W \quad (17)$$

Die Unterschiede zwischen den Kategorien sind signifikant, wenn  $Q_B$  den kritischen  $\text{Chi}^2$  Wert bei  $m - 1$  Freiheitsgraden überschreitet. Ist dies der Fall, so klärt die kategoriale Moderatorvariable einen signifikanten Anteil an der Gesamtheterogenität der Effektstärken auf. Überschreitet  $Q_W$  bei  $k - m$  Freiheitsgraden den kritischen  $\text{Chi}^2$  Wert, dann muss von einer signifikanten Heterogenität innerhalb der Kategorien ausgegangen werden. In diesem Fall sollte ein sogenanntes Mixed-Model verwendet werden, welches ähnlich dem Random-Effects-Model bei der Berechnung des Gewichtungsfaktors entsprechend Formel (11) dem Stichprobenfehler eine zusätzliche Varianzkomponente hinzufügt. Diese muss jedoch mit einer Matrix-Algebra geschätzt werden und soll an dieser Stelle nicht näher erläutert werden (Lipsey & D. B. Wilson, 2001).

Die zweite Möglichkeit der Moderatoranalyse in einer Meta-Analyse ist eine gewichtete Regressionsanalyse (Hedges & Olkin, 1985). Mit dieser können zum einen auch kontinuierliche Variablen auf ihren Einfluss auf die Effektstärke getestet und zum anderen auch der Einfluss mehrerer Variablen gleichzeitig modelliert werden. Dies geschieht mit einer angepassten Variante der Methode der gewichteten kleinsten Quadrate. Hierbei wird auch die aufgeklärte Heterogenität ( $Q_R$ ) der unaufgeklärten bzw. Residualheterogenität ( $Q_E$ ) gegenübergestellt, die addiert die Gesamtheterogenität ergeben. Überschreitet  $Q_R$  den kritischen  $\text{Chi}^2$  Wert bei  $p$  Freiheitsgraden ( $p$  entspricht hier der Anzahl enthaltener Prädiktoren), klärt das Modell bzw. die darin enthaltenen Prädiktoren einen signifikanten Anteil der Heterogenität auf. Ist  $Q_E$  bei  $k - p - 1$  Freiheitsgraden signifikant, so ist ein signifikanter Anteil der Heterogenität trotz der Prädiktoren noch unaufgeklärt. In diesem Fall sollte ähnlich wie bei der meta-analytischen Varianzanalyse auf ein Mixed-Model zurück gegriffen werden (Lipsey & D. B. Wilson, 2001).

### 3.6.5 Publikationsverzerrung

Ein grundsätzliches methodisches Problem von Meta-Analysen, welches in dieser Arbeit ebenfalls explizit Beachtung finden soll, ist das der *Publikationsverzerrung* (in der

englischsprachigen Literatur als *publication bias*, *sampling bias* oder *availability bias* bekannt). Hierunter ist eine mögliche Verzerrung der Ergebnisse einer Meta-Analyse in Richtung höherer Effektstärken zu verstehen, die dadurch zustande kommen kann, dass statistisch signifikante Befunde und somit solche mit höheren Effektstärken mit größerer Wahrscheinlichkeit publiziert werden, zugleich aber jene publizierten Befunde auch einfacher zu finden und zu beschaffen und hierdurch in Meta-Analysen meist überrepräsentiert sind (Lipsey & D. B. Wilson, 2001). So konnten Lipsey und D. B. Wilson (1993) in ihrer Meta-Analyse zur Wirksamkeit psychologischer, edukativer und verhaltensbezogener Interventionen zeigen, dass publizierte Studien im Mittel höhere Effekte haben als unpublizierte Studien. Dieser Effekt ist besonders bei Studien mit kleinerer Stichprobengröße ausgeprägt, da diese zum einen erst bei größeren Effektstärken signifikante Ergebnisse erbringen und zum anderen nicht signifikante Ergebnisse bei kleiner Stichprobengröße seltener veröffentlicht werden (Dickersin, 1990; Gilbody, et al., 2000; Thornton & Lee, 2000).

Um mit dieser bekannten Validitätsgefährdung meta-analytischer Ergebnisse umzugehen, wurden verschiedene Vorgehensweisen vorgeschlagen. Rosenthal (1979) schlug hierfür die Berechnung des so genannten *Fail-Safe N* vor (Orwin, 1983). *Fail-Safe N* gibt allgemein die Anzahl an Studien mit einer durchschnittlichen Effektstärke von Null an, die notwendig wären, um den in einer Meta-Analyse gefundenen signifikanten Effekt in einen nicht mehr signifikanten oder zumindest kleineren Effekt umzuwandeln. Es errechnet sich folgendermaßen:

$$X = \left[ \left( \sum_{i=1}^k Z_i \right)^2 / (Z_{\alpha/2})^2 \right] - k \quad (18)$$

Wobei  $X$  die Anzahl der Studien mit einer Effektstärke von Null widerspiegelt, die notwendig wären, um den signifikanten Effekt in einen nicht mehr signifikanten umzuwandeln,  $k$  der Zahl der eingeschlossenen Studien entspricht und  $Z_{\alpha/2}$  den kritischen  $Z$ -Wert bei einem zweiseitigen Test darstellt, für den getestet werden soll (bei einem Signifikanzniveau von  $\alpha = .05$  ist dieser gleich 1.96). Die Variable  $Z_i$  entspricht dem  $Z$ -Wert der Effektstärke der Studie  $i$  und lässt sich mittels folgender Formel schätzen:

$$Z = \left[ d^2 / \sqrt{(d^2 + 4)} \right] \times \sqrt{N} \quad (19)$$

$d$  entspricht dabei der Effektstärke der Studie und  $N$  deren Gesamtstichprobengröße.

Ausgangspunkt von Rosenthal (1979) war hierbei das *File-Drawer* Problem. Dieses besteht darin, dass egal wie viel publizierte Studien gefunden und in die Meta-Analyse eingeschlossen wurden, man dennoch nicht weiß, wie viele Studien mit nicht signifikanten Ergebnissen nicht publiziert wurden und dadurch für die Meta-Analyse nicht auffindbar waren. Im Extremfall sind demnach entsprechend des Signifikanzlevels nur lediglich die 5 % aller durchgeführten Studien, welche zufällig das Signifikanzniveau erreichten in der publizierten Literatur zu finden, wohingegen die restlichen 95 %, welche keine signifikanten Ergebnisse zeigen konnten, nicht publiziert wurden und hierdurch in einer Meta-Analyse nicht auftauchen (Rosenthal, 1979). Mithilfe des Fail-Safe  $N$ 's ist es nun zumindest möglich, die Stabilität des gefundenen Effekts angesichts möglicher nicht veröffentlichter Studien mit Null-Effekten, abzusichern. Ist das Fail-Safe  $N$  sehr groß, so dass es unwahrscheinlich erscheint, dass so viele Studien zu dieser Fragestellung durchgeführt aber nicht veröffentlicht wurden, so kann davon ausgegangen werden, dass der in der Meta-Analyse gefundene Effekt zumindest nicht vollständig auf das File-Drawer Problem zurückführbar ist. Als mögliche Grenze hierfür schlägt Rosenthal (1979) das Fünffache der bereits gefundenen Studien plus 10 weitere Studien vor.

Eine weitere Methode der Aufdeckung einer Publikationsverzerrung in den meta-analytischen Ergebnissen geht von der Grundannahme aus, dass, wenn Publikationsverzerrungen zugunsten höherer Effektstärken auftreten, diese bei kleineren Stichproben stärker ausgeprägt sein sollten als bei größeren Stichproben. Kleinere Stichproben ergeben zum einen erst bei größeren Effektstärken signifikante Ergebnisse, zum anderen werden nicht signifikante Ergebnisse bei kleiner Stichprobengröße seltener veröffentlicht als bei größerer (Dickersin, 1990; Gilbody, et al., 2000; Thornton & Lee, 2000). Daraus folgt, dass, wenn es eine Publikationsverzerrung gibt, sich in der Meta-Analyse auch ein negativer Zusammenhang zwischen der Studieneffektstärke und der Teilnehmerzahl zeigen sollte. Dieser Zusammenhang wird häufig grafisch mittels Funnel-Plot betrachtet, bei dem entweder die Stichprobengrößen der Studien oder deren Varianz in Abhängigkeit von deren Effektstärke abgetragen werden. Die Interpretation solcher grafischer Darstellungen enthält jedoch stets ein subjektives Element, da die Schlussfolgerungen, die aus der Abbildung gezogen werden, je nach Betrachter unterschiedlich ausfallen können. Aus diesem Grund wurden statistische Methoden entwickelt, mithilfe derer das Vorhandensein einer Publikationsverzerrung anhand eines Signifikanztests überprüft werden kann. Kromrey und Rendina-Gobioff (2006) konnten zeigen, dass von diesen

Tests *Begg's Rangkorrelationsmethode* (Begg & Mazumdar, 1994) unter Verwendung der Stichprobengröße am besten den alpha-Fehler kontrollierte und zudem bei ausreichend hoher Anzahl an Studien eine mindestens genauso hohe Teststärke aufwies wie alternative Methoden der Signifikanztestung auf Publikationsverzerrungen. Nach *Begg's Rangkorrelationsmethode* wird zunächst die standardisierte Effektstärke jeder Studie berechnet und sodann deren Zusammenhang mit der Stichprobengröße mithilfe des Rangkorrelationskoeffizienten *Kendall's tau* ermittelt. Bei Vorliegen einer Publikationsverzerrung sollte sich ein signifikant negativer Zusammenhang zeigen. Allerdings muss an dieser Stelle beachtet werden, dass in der vorliegenden Meta-Analyse nur Studien eingeschlossen wurden, die eine Mindeststichprobengröße von 25 Teilnehmern pro Gruppe, also insgesamt 50 aufwiesen, wodurch die mögliche Varianz in der Stichprobengröße eingeschränkt und gerade Studien mit sehr kleinen Stichproben, für die die Publikationsverzerrung am größten sein sollte, folglich von vornherein ausgeschlossen waren. Dies begrenzt auch die Möglichkeit, mit dieser Methode eine Publikationsverzerrung aufzudecken.

Eine dritte Methode, eine Publikationsverzerrung in einer Meta-Analyse mithilfe der gegebenen Daten sichtbar zu machen, ist, publizierte und folglich leicht verfügbare Studien mit jenen Studien zu vergleichen, die entweder öffentlich weniger gut zugänglich sind oder aufgrund der Art wie sie verfügbar gemacht wurden, eine geringere Publikationsverzerrung erwarten lassen. Lipsey & D. B. Wilson (1993) nennen als publizierte und leicht verfügbare Studien insbesondere solche, welche in Zeitschriften oder Monographien veröffentlicht wurden. Im Gegensatz dazu sollten Dissertationen weniger einer Publikationsverzerrung unterworfen sein, da ihre Verfügbarkeit in der Regel nicht von der Höhe gefundener Effekte abhängig ist. Auch bieten sich Dissertationen zur Testung der Publikationsverzerrung im Besonderen an, da sie typischerweise eine ähnlich hohe methodische Qualität aufweisen wie publizierte Studien (Beelmann, 2006).

In dieser Meta-Analyse sollten alle diese drei Methoden genutzt werden, um zu prüfen, inwiefern die Validität der Ergebnisse aufgrund einer Publikationsverzerrung in Frage gestellt werden muss. Da in der vorliegenden Meta-Analyse die Vergleichsebene die primäre Analyseebene war, hinsichtlich derer auch die Effektstärken aggregiert wurden, sollten hierbei auch alle Auswertungen bezüglich der Publikationsverzerrung auf

dieser hierarchischen Ebene durchgeführt werden. Daraus folgte unter anderem auch, dass das Fail-Safe  $N$  in der dieser Arbeit eine Anzahl von Gruppenvergleichen darstellt.



## 4. ERGEBNISSE

### 4.1 Deskriptive Auswertungen

Bevor auf die Wirksamkeit der präventiven Interventionen eingegangen wird und weitergehende Analysen dargestellt werden, sollen in diesem Kapitel zunächst die gefundenen Forschungsberichte und die in ihnen enthaltenen Gruppenvergleiche im Detail beschrieben werden. Dies dient hauptsächlich dazu, die Generalisierbarkeit und Aussagekraft nachfolgend dargestellter Befunde besser beurteilen zu können, liefert aber zugleich einen Überblick über den aktuellen Forschungsstand, indem Aussagen darüber ermöglicht werden, in welchen Kontexten und unter welchen Rahmenbedingungen die Fragestellung bereits untersucht wurde.

#### 4.1.1 Forschungsberichte

Von den insgesamt 68 eingeschlossenen Forschungsberichten wurden 51 in Fachzeitschriften publiziert. Diese Forschungsberichte verteilten sich dabei auf 31 verschiedene Fachzeitschriften. Des Weiteren waren 13 der Forschungsberichte Dissertationen, zwei wurden als Forschungsberichte von Ministerien veröffentlicht, einer als Monographie und bei ebenfalls einem Forschungsbericht handelte es sich um eine Masterarbeit. Die 68 Forschungsberichte wurden zwischen 1967 und 2013 publiziert bzw. zugänglich gemacht, eine davon 1967, zwei in den 1970er Jahren, vier in den 1980er Jahren, 12 in den 1990er Jahren, 28 in den 2000er Jahren und 21 von 2010 bis einschließlich 2013. Von den 68 Forschungsberichten wurden nahezu alle in englischer Sprache verfasst. Lediglich ein Forschungsbericht wurde in deutscher Sprache publiziert. Hinsichtlich des Publikationslandes stammten die Forschungsberichte in großer Mehrheit aus den USA (62; 91.2 %). Lediglich sechs Forschungsberichte stammten aus anderen Ländern, darunter zwei aus Kanada und jeweils ein Forschungsbericht aus den Niederlanden, Deutschland, Norwegen und Israel. Eine detaillierte Auflistung der deskriptiven Merkmale der Forschungsberichte ist in Tabelle 1 zu finden.

Tabelle 1

*Beschreibung der 68 Forschungsberichte in absoluten und relativen Häufigkeiten*

Merkmal	<i>n</i>	%
<b>Publikationstyp</b>		
Zeitschriftenartikel	51	75.0
Dissertation	13	19.1
Forschungsbericht eines Ministeriums	2	2.9
Monographie	1	1.5
Masterarbeit	1	1.5
<b>Publikationsjahr</b>		
1960–1969	1	1.5
1970–1979	2	2.9
1980–1989	4	5.9
1990–1999	12	17.6
2000–2009	28	41.2
2010–2013	21	30.9
<b>Publikationssprache</b>		
Englisch	67	98.5
Deutsch	1	1.5
<b>Erhebungsland</b>		
USA	62	91.2
Kanada	2	2.9
Niederlande	1	1.5
Deutschland	1	1.5
Norwegen	1	1.5
Israel	1	1.5

*Anmerkung.* *n* = Anzahl der Forschungsberichte.

#### 4.1.2 Vergleiche

Die in den 68 Forschungsberichten enthaltenen Vergleiche können zusätzlich hinsichtlich der Merkmale der durchgeführten präventiven Intervention und der untersuchten Stichprobe sowie bezüglich methodischer Charakteristika unterschieden werden. Von den insgesamt 93 enthaltenen Vergleichen bezogen sich die meisten auf Präventions-

programme, die sich ausschließlich an Kinder bzw. Jugendliche richteten (68.8 %). Lediglich 11 Programme bezogen zusätzlich zu den Kindern bzw. Jugendlichen die dazugehörigen Eltern mit ein, 15 arbeiteten ausschließlich mit Eltern und drei der Interventionen waren Lehrertrainingsprogramme. Hinsichtlich der Programmlänge erstreckten sich 28 % über höchstens einen Monat, 12.9 % dauerten 1 bis 2 Monate, 17.2 % 2 bis 3 Monate, 16.1 % 3 bis 6 Monate, 15.1 % 6 Monate bis zu 1 Jahr und immerhin noch 10.8 % waren sogar länger als ein Jahr. Bezogen auf die Kontakthäufigkeit in dieser Zeit zeigte sich, dass jeweils etwa ein Viertel der Programme auf solche mit bis zu fünf Sitzungen (25.8 %), 6 bis 10 Sitzungen (26.9 %) und 11 bis 20 Sitzungen (26.9 %) entfiel. Lediglich 11.8 % der Interventionen enthielten 21 bis 40 Sitzungen und 8.6 % hatten sogar mehr als 100 Sitzungen. Dabei wurden die meisten Programme ausschließlich als Gruppentraining durchgeführt (75.3 %). 14 % der präventiven Interventionen trainierten die Teilnehmer sowohl in der Gruppe als auch in Einzelsitzungen und 7.5 % arbeiteten ausschließlich einzeln mit den Teilnehmern. Nur zwei der Programme basierten auf Selbsthilfe und in einem Fall gab es keine Angabe zum Interventionsformat. Die präventiven Interventionen wurden in den meisten Fällen in Schulen, Vorschulen oder Kindergärten bzw. Kinderkrippen durchgeführt (65.6 %). In Gemeinschaftszentren bzw. Begegnungsstätten fanden 4.3 % der Programme statt, jeweils 2.2 % wurden zu Hause, in Wohlfahrtseinrichtungen sowie in Krankenhäusern oder Sonderschulen durchgeführt und 1.1 % in Universitäten oder Forschungseinrichtungen. Häufiger kamen noch Programme vor, die an mehreren verschiedenen Orten durchgeführt wurden (14 %). Für 8.6 % der Vergleiche gab es keine Angabe zum Ort der Interventionsdurchführung. In den meisten Fällen wurden die Programme von psychosozialen Fachkräften (29 %), Lehrpersonal (23.7 %) oder Hochschulpersonal (20.4 %) durchgeführt. In 5.4 % der Vergleiche gab es keinen Administrator und 4.3 % der Interventionen wurden von Laien oder Freiwilligen geleitet. Ein substantieller Anteil von 12.9 % wurde schließlich durch mehrere verschiedene der zuvor genannten Personengruppen durchgeführt. Lediglich für 4.3 % der Gruppenvergleiche wurden keine Angaben zum Administrator der Intervention gemacht.

Die Stichproben umfassten bei 26.9 % der Vergleiche zwischen 50 und 100 Teilnehmer, bei 33.3 % 101 bis 200 Teilnehmer, 201 bis 500 Teilnehmer groß waren 14 % der untersuchten Stichproben und jeweils 12.9 % der Vergleiche enthielten 501 bis 1 000 Teilnehmer und mehr als 1 000 Teilnehmer. Die meisten Stichproben hatten

mehrheitlich einen spanischsprachigen bzw. lateinamerikanischen Migrationshintergrund, davon 44 Gruppenvergleiche ohne genauere Angabe hinsichtlich der Herkunft, 25 aus Mexiko, fünf aus Puerto Rico und eine aus Kuba. Asiatische Teilnehmer unspezifischer Herkunft und türkischstämmige Teilnehmer wurden in jeweils drei der Vergleiche untersucht und jeweils eine Stichprobe war chinesischen, haitianischen und afrikanischen Ursprungs. Insgesamt 8 der 93 Stichproben enthielten jeweils mehrere verschiedene Migrantengruppen und für einen Vergleich gab es keine Angabe bezüglich der Herkunft der Eltern. Hinsichtlich des Geburtslandes der untersuchten Kinder bzw. Jugendlichen gab es nur für etwa ein Viertel der Vergleiche überhaupt eine Angabe in den Forschungsberichten. Lediglich 4 der 93 Stichproben enthielten mehrheitlich Kinder und Jugendliche, die in dem Herkunftsland ihrer Eltern geboren wurden und folglich selbst migriert sind, wohingegen für 21 Vergleiche bekannt ist, dass mindestens 50 % der Stichprobe im Einwanderungsland geboren wurde. Eine Hochrisikostichprobe mit bereits vorhandenen kognitiven, Lern- oder Verhaltensauffälligkeiten untersuchten 15.1 % der Vergleiche, wohingegen 84.9 % der Stichproben keine derartigen Auffälligkeiten zeigten. Die teilnehmenden Kinder bzw. Jugendlichen waren bei allen Vergleichen im Mittel jünger als 17 Jahre. In 4.3 % der Fälle lag das mittlere Alter bei bis zu 3 Jahren, in 8.6 % zwischen 3 und 6 Jahren, 21.5 % der Stichproben waren im Mittel zwischen 6 und 10 Jahre, 46.2 % zwischen 10 und 14 Jahre und 19.4 % zwischen 14 und 17 Jahre alt. In den meisten Fällen waren beide Geschlechter in der Stichprobe etwa gleich häufig vertreten (61.3 %). In 22 der 93 Vergleiche waren nur 20 bis 44.9 % der Teilnehmer Mädchen, auf der anderen Seite waren bei 10 Gruppenvergleichen 55.1 bis 80 % der Teilnehmer Mädchen und drei Vergleiche umfassten ausschließlich weibliche Teilnehmerinnen. Für einen Vergleich war keine Aussage bezüglich der Geschlechtsverteilung möglich. Die meisten Untersuchungen fanden in einer urbanen Umgebung statt (64.5 %). Lediglich 2.2 % wurden in ländlichen oder suburbanen Regionen durchgeführt, 22.6 % fanden sowohl in ländlicher als auch in urbaner Umgebung statt und für 10.8 % der Gruppenvergleiche wurde hierzu keine Angabe gemacht. Von den Stichproben kamen die meisten aus Regionen, in denen hauptsächlich die nicht-migrierten Angehörige der ethnischen Majorität des Landes lebten (52.7 %), 32.3 % aus Gebieten, in denen die untersuchte Migrantengruppe selbst eine Mehrheit darstellten und 9.7 % aus Regionen, in denen eine andere Migrantengruppe oder ethnische Minderheit den größten Teil der Bevölkerung darstellten. Für 5.4 % der Stichproben waren hierzu keine Aussagen möglich.

Auch innerhalb der relativ restriktiven Einschlusskriterien gab es noch methodische Unterschiede zwischen den verschiedenen Vergleichen. Die Randomisierung fand in 62.4 % der Fälle auf individueller Ebene statt, in 37.6 % wurden ganze Cluster, wie beispielsweise Klassen oder Schulen, randomisiert zugewiesen. Die mittlere Dropout-Rate lag bei 20.4 % der Gruppenvergleiche unter 5 %, bei 17.2 % zwischen 5 und 10 %, bei 18.3 % der Vergleiche zwischen 10 und 20 % und bei 22.6 % zwischen 20 und 30 %. Lediglich 7.5 % hatten Dropout-Raten von 30 bis 40 %, 3.2 % Dropout-Raten von 40 bis 50 % und bei 2.2 % der Vergleiche lag die Dropout-Rate sogar über 50 %. In 8.6 % der Fälle war keine Angabe zur Dropout-Rate vorhanden. Interventions- und Kontrollgruppe wurden in den meisten Fällen vor der Intervention als äquivalent getestet (55.9 %), 25.8 % waren hingegen schon vorher nicht vergleichbar. Bei 18.3 % der Vergleiche wurde die Äquivalenz der Gruppen nicht getestet. Die Datenauswertung fand schließlich in einer überwiegenden Mehrheit Per-protocol statt, ohne dass Teilnehmer, die eine bestimmte Sitzungszahl nicht mit gemacht hatten explizit ausgeschlossen wurden (72.0 %). Bei 10.8 % der Vergleiche fand sogar ein solcher Ausschluss statt und weitere 10.8 % bezogen alle Personen in die Analysen ein, die randomisiert auf die beiden Gruppen verteilt wurden (Intention-to-treat). Für 6.5 % machte eine Unterscheidung nach der Datenauswertungsmethode keinen Sinn, da alle Teilnehmer an der Post-Testung teilnahmen. Die Kontrollgruppen der Untersuchungen bekamen mehrheitlich zumindest ein Standardprogramm, eine minimale, oder eine unspezifische Intervention (58.1 %). Etwas weniger Kontrollgruppen waren passiv, erhielten folglich während der Studie keinerlei Intervention (41.9 %). Alle deskriptiven Analysen auf Vergleichsebene sind in Tabelle 2 zu finden.

Tabelle 2

*Beschreibung der 93 Vergleiche in absoluten und relativen Häufigkeiten*

Merkmal	<i>k</i>	%
Angaben zum Präventionsprogramm		
Adressat(en)		
Kinder/Jugendliche	64	68.8
Eltern	15	16.1
Lehrer	3	3.2
Kinder/Jugendliche und Eltern	11	11.8
Länge des Programmes		
Bis zu 1 Monat	26	28.0
1–2 Monate	12	12.9
2–3 Monate	16	17.2
3–6 Monate	15	16.1
6 Monate bis zu 1 Jahr	14	15.1
Länger als 1 Jahr	10	10.8
Anzahl der Sitzungen		
≤ 5	24	25.8
6–10	25	26.9
11–20	25	26.9
21–40	11	11.8
> 100	8	8.6
Interventionsformat		
Selbsthilfe	2	2.2
Individuell	7	7.5
Gruppe	70	75.3
Individuell und Gruppe	13	14.0
Keine Angabe	1	1.1

Merkmal	<i>k</i>	%
<b>Ort der Interventionsdurchführung</b>		
Schule, Vorschule oder Kindergarten/-krippe	61	65.6
Universität oder Forschungseinrichtung	1	1.1
Zu Hause	2	2.2
Gemeinschaftszentrum / Begegnungsstätte	4	4.3
Krankenhaus oder Sonderschule	2	2.2
Wohlfahrtseinrichtungen	2	2.2
Mehrere verschiedene Orte	13	14.0
Keine Angabe	8	8.6
<b>Administrator der Intervention</b>		
Keine Person	5	5.4
Lehrpersonal	22	23.7
Psychosoziale Fachkräfte	27	29.0
Laien oder Freiwillige	4	4.3
Hochschulpersonal	19	20.4
Mehrere verschiedene Administratoren	12	12.9
Keine Angabe	4	4.3
<b>Angaben zur Stichprobe</b>		
<b>Stichprobengröße</b>		
50–100	25	26.9
101–200	31	33.3
201–500	13	14.0
501–1 000	12	12.9
> 1 000	12	12.9

Merkmal	<i>k</i>	%
Herkunft		
Hispano-/lateinamerikanisch ohne genauere Angabe	44	47.3
Mexikanisch	25	26.9
Kubanisch	1	1.1
Puerto-ricanisch	5	5.4
Asiatisch ohne genauere Angabe	3	3.2
Chinesisch	1	1.1
Haitianisch	1	1.1
Afrikanisch ohne genauere Angabe	1	1.1
Türkisch	3	3.2
Mehrere verschiedene Migrantengruppen	8	8.6
Keine Angabe	1	1.1
Geburtsland der Kinder/Jugendlichen		
≤ 50 % Herkunftsland der Eltern	21	22.6
> 50 % Herkunftsland der Eltern	4	4.3
Keine Angabe	68	73.1
Hochrisikostichprobe <sup>a</sup>		
Bereits erste Auffälligkeiten	14	15.1
Keine bestehenden Auffälligkeiten	79	84.9
Mittleres Alter der Kinder/Jugendlichen in Jahren		
≤ 3	4	4.3
3–6	8	8.6
6–10	20	21.5
10–14	43	46.2
14–17	18	19.4
Geschlecht (% Mädchen)		
20–44.9	22	23.7
45–55	57	61.3
55.1–80	10	10.8
100	3	3.2
Keine Angabe	1	1.1



Merkmal	<i>k</i>	%
Urbanität		
Ländlich bis suburban	2	2.2
Urban	60	64.5
Gemischt	21	22.6
Keine Angabe	10	10.8
Minoritätenstatus in der untersuchten Region		
Hauptsächlich Nicht-Migranten der ethnischen Majorität	49	52.7
Hauptsächlich andere ethnische Minorität	9	9.7
Untersuchte Gruppe stellt in der Region Mehrheit dar	30	32.3
Keine Angabe / unsicher	5	5.4
Methodische Angaben		
Randomisierung		
Auf individueller Ebene	58	62.4
Auf Ebene von Clustern (z.B. Klassen oder Schulen)	35	37.6
Mittlere Dropout-Rate (in % der Gesamtstichprobe)		
≤ 5	19	20.4
5–10	16	17.2
10–20	17	18.3
20–30	21	22.6
30–40	7	7.5
40–50	3	3.2
> 50	2	2.2
Keine Angabe	8	8.6
Äquivalenz der Gruppen vor der Intervention		
Äquivalent	52	55.9
Nicht äquivalent	24	25.8
Nicht getestet	17	18.3
Datenauswertung		
Per-protocol mit Mindestteilnahme an Intervention	10	10.8
Per-protocol ohne Mindestteilnahme an Intervention	67	72.0
Intention-to-treat	10	10.8
Keine Angabe (keine Dropouts)	6	6.5

Merkmal	<i>k</i>	%
Kontrollgruppe		
Passiv	39	41.9
Aktiv	54	58.1

*Anmerkungen.* *k* = Anzahl der Vergleiche.

<sup>a</sup> Als Hochrisikostichprobe galt eine Stichprobe dann, wenn mindestens 50% bereits vor Interventionsbeginn erste kognitive, Lern- oder Verhaltensauffälligkeiten zeigte.

\*  $p < .05$ ; \*\*  $p < .01$ ; \*\*\*  $p < .001$ .

## 4.2 Allgemeine Effektivität

Von den insgesamt 93 Vergleichen berichteten 76 Post-Effektstärken, 27 Vergleiche Effektstärken im Messzeitraum zwischen 3 und 12 Monaten nach der Intervention und 17 Vergleiche Effektstärken, die mehr als 12 Monate nach der Intervention erhoben wurden. Von den insgesamt 850 berichteten Effektstärken, entfielen dabei 529 auf den Post-Zeitraum, 212 auf den ersten Follow-up-Zeitraum und 109 auf den zweiten Follow-up-Zeitraum.

Die Effektstärken im ersten Messzeitraum schwankten auf Effektstärkenebene zwischen  $d_{es} = -0.56$  und  $d_{es} = 1.90$  ( $M_{des} = 0.19$ ,  $SD_{des} = 0.30$ ). Da Effektstärken in dieser Spannweite durchaus realistisch erschienen und es keinen Hinweis darauf gab, dass die Extremwerte auf Messfehler zurückzuführen waren, wurden diese weder ausgeschlossen noch umkodiert. Nach Aggregation der Post-Effektstärken ergaben sich daraus auf der Ebene der Vergleiche Effektstärken im Bereich zwischen  $d = -0.17$  und  $d = 0.82$  (ungewichtete  $M_d = 0.20$ ,  $SD_d = 0.19$ ). Als gewichtete mittlere Effektstärken ergaben sich hieraus jeweils von Null verschiedene Effektstärken von  $d_+ = 0.17$  (95 %-CI [0.14, 0.19]) für das Fixed-Effects-Model und  $d_+ = 0.18$  (95 %-CI [0.14, 0.21]) für das Random-Effects-Model.

Im zweiten Messzeitraum ergaben sich auf Effektstärkenebene Effekte von  $d_{es} = -0.55$  bis  $d_{es} = 1.39$  ( $M_{des} = 0.15$ ,  $SD_{des} = 0.27$ ). Da auch hier die Extremwerte plausibel erschienen und es wiederum keinen Anhaltspunkt dafür gab, dass sie auf Messfehlern beruhten, wurden diese ebenfalls unbearbeitet in der vorliegenden Form aggregiert. Auf Vergleichsebene ergaben sich für diesen Messzeitraum Effektstärken von  $d = -0.18$  bis  $d = 0.63$  (ungewichtete  $M_d = 0.19$ ,  $SD_d = 0.20$ ). Die gewichteten mittleren Effektstärken

waren schließlich  $d_+ = 0.12$  (95 %-CI [0.07, 0.16]) für das Fixed-Effects-Model und  $d_+ = 0.16$  (95 %-CI [0.08, 0.23]) für das Random-Effects-Model. Diese waren beide signifikant von Null verschieden.

Die Effektstärken für die Langzeit-Follow-up-Messungen schwankten auf Effektstärkenebene zwischen  $d_{es} = -0.78$  bis  $d_{es} = 2.66$  ( $M_{des} = 0.16$ ,  $SD_{des} = 0.37$ ). Da auch in diesem Fall die Extremwerte nicht ungewöhnlich erschienen und es keinen Hinweis auf Messfehler gab, wurden die Effektstärken ohne Korrektur aggregiert. Auf Ebene der Vergleiche schwankten die Effektstärken für diesen Messzeitraum zwischen  $d = -0.05$  bis  $d = 0.42$  (ungewichtete  $M_d = 0.16$ ,  $SD_d = 0.15$ ). Als gewichtete mittlere Effektstärken ergaben sich jeweils signifikante Effekte von  $d_+ = 0.11$  (95 %-CI [0.08, 0.15]) für das Fixed-Effects-Model und  $d_+ = 0.12$  (95 %-CI [0.06, 0.19]) für das Random-Effects-Model.

### 4.3 Heterogenitätsanalyse

Um die Varianz der gewichteten Effektstärken besser bewerten und die Passung der beiden Modelle beurteilen zu können, ist es notwendig, die Heterogenität der Effektstärken zu betrachten.

Die Homogenitätsanalyse ergab für die Post-Messungen keine signifikante Heterogenität ( $Q = 96.17$ ,  $df = 75$ ,  $p = .05$ ). Es ergab sich ein Anteil der Heterogenität, also der nicht auf Stichprobenfehler zurückführbaren Varianz, an der Gesamtvarianz von  $I^2 = 22.0$  %. Das  $I^2$  lag somit sogar noch knapp unter der von Huedo-Medina et al. (2006) als *klein* eingestuften Heterogenität. Sowohl die Ergebnisse des Signifikanztests aus der  $Q$ -Statistik als auch die deskriptiven, von der Anzahl der Gruppenvergleiche unabhängigen, Befunde sprachen folglich dafür, im ersten Messzeitraum das Fixed-Effects-Model beizubehalten und das Random-Effects-Model nicht zu verwenden.

Für den ersten Follow-up-Zeitraum ergab sich eine signifikante Heterogenität der Effektstärken ( $Q = 62.77$ ,  $df = 26$ ,  $p < .001$ ). Die Heterogenität der Vergleiche hatte folglich einen Anteil an der Gesamtvarianz von  $I^2 = 58.6$  %, was einer mittleren bis hohen Heterogenität entspricht. Die Ergebnisse der Heterogenitätsanalysen zeigten für diesen Messzeitraum demnach eine Nichtpassung des Fixed-Effects-Modells an, was bedeutet, dass das Random-Effects-Model hier Anwendung finden sollte.

Die Langzeit-Follow-up-Messungen wiesen ebenfalls eine signifikante Heterogenität auf ( $Q = 45.13$ ,  $df = 16$ ,  $p < .001$ ). Betrachtet man das quantitative Ausmaß der Heterogenität, so zeigte sich, dass die nicht auf Stichprobenfehler zurückführbare Varianz an der Gesamtvarianz einen Anteil von  $I^2 = 64.5\%$  hatte. Dies entspricht einer mittleren bis hohen Heterogenität. Demzufolge sprachen auch hier sowohl die Ergebnisse des Signifikanztests aus der  $Q$ -Statistik als auch die deskriptive Betrachtung des prozentualen Anteils der Heterogenität an der Gesamtvarianz gegen die Annahmen des Fixed-Effects-Models. Folglich sollte auch für diesen Messzeitraum stattdessen das Random-Effects-Model genutzt werden.

#### **4.4 Stabilität der Effekte über die Zeit**

Um Unterschiede in der Programmwirksamkeit zwischen dem Post-Zeitraum sowie den zwei Follow-up-Zeiträumen und damit die Stabilität der Effekte beurteilen zu können, ist es notwendig, sich auch die Ergebnisse für jene Vergleiche anzuschauen, für die auch Messungen zu jedem der zu vergleichenden Zeiträume vorgenommen wurden.

Eine Übersicht über die Effektstärken für die einzelnen Zeiträume in Abhängigkeit von in den jeweiligen Vergleichen ebenfalls erfassten Messzeiträumen gibt Tabelle 3. Es zeigte sich, dass von den 27 Vergleichen, für die Ergebnisse für den ersten Follow-up-Zeitraum angegeben waren, lediglich 15 auch Post-Effektstärken beinhalteten. Von den insgesamt 17 Vergleichen mit Langzeit-Follow-up-Ergebnissen enthielten nur 9 auch Effektstärken für den Post-Zeitraum und 7 Erfolgsmessungen für den ersten Follow-up-Zeitraum. Lediglich bei vier Gruppenvergleichen waren Effekte für alle drei Zeiträume angegeben, was einen Vergleich zwischen den Zeiträumen erschwerte.

Tabelle 3

*Effektstärken für die verschiedenen Zeiträume in Abhängigkeit von für den jeweiligen Vergleich vorhandenen Messzeiträumen*

Vergleiche		Post	Follow-up 1	Follow-up 2
Mit Post-Messung(en)	$M_d (SD)$	0.20 (0.19)	0.16 (0.20)	0.09 (0.15)
	FEM $d_+$	0.17***	0.05	0.04
	REM $d_+$	0.18***	0.07	0.04
	$Q (df)$	96.17 (75)	21.23 (14)	6.97 (8)
	$k$	76	15	9
Mit Messung(en) im Follow-up-Zeitraum 1	$M_d (SD)$	0.22 (0.19)	0.19 (0.20)	0.20 (0.18)
	FEM $d_+$	0.21***	0.12***	0.06
	REM $d_+$	0.20***	0.16***	0.12
	$Q (df)$	27.48 (14)*	62.77 (26)***	10.82 (6)
	$k$	15	27	7
Mit Messung(en) im Follow-up-Zeitraum 2	$M_d (SD)$	0.16 (0.14)	0.18 (0.20)	0.16 (0.15)
	FEM $d_+$	0.09***	0.04	0.11***
	REM $d_+$	0.09***	0.10	0.12***
	$Q (df)$	8.10 (8)	10.69 (6)	45.13 (16)***
	$k$	9	7	17
Mit Messung(en) in allen drei Zeiträumen	$M_d (SD)$	0.17 (0.12)	0.17 (0.23)	0.15 (0.22)
	FEM $d_+$	0.09	0.01	0.02
	REM $d_+$	0.09	0.07	0.07
	$Q (df)$	1.83 (3)	6.75 (3)	6.20 (3)
	$k$	4	4	4

*Anmerkungen.* Die Angaben beziehen sich jeweils auf die Effektstärken des in der Spalte angegebenen Zeitraums für Vergleiche, in denen auch Messungen für den in der Zeile angegebenen Zeitraum vorhanden waren.  $M_d$  = ungewichtete mittlere Effektstärke. FEM  $d_+$  = gewichtete mittlere Effektstärke nach dem Fixed-Effects-Model. REM  $d_+$  = gewichtete mittlere Effektstärke nach dem Random-Effects-Model.  $Q$  = Heterogenität.  $k$  = Anzahl der Vergleiche.

\*  $p < .05$ ; \*\*  $p < .01$ ; \*\*\*  $p < .001$ .

Betrachtete man nun die Effektstärken innerhalb von Gruppenvergleichen, so zeigten sich sowohl für Messungen zum Post- im Vergleich zum ersten Follow-up-Zeitraum ( $d = 0.20$  vs.  $0.07$ ; Random-Effects-Model) als auch für Post- und Langzeit-Follow-up-Messungen tendenziell geringer werdende Effektstärken ( $d = 0.09$  vs.  $0.04$ ; Fixed-Effects-Model). Die Effektstärken für die Follow-up-Zeiträume waren dabei für diese Vergleiche, die auch Post-Messungen beinhalteten, nicht mehr signifikant von Null verschieden ( $p = .10$  bzw.  $p = .09$ ). Für Vergleiche mit Messungen zu beiden Follow-up Zeiträumen zeigte sich tendenziell jedoch eher eine zunehmende Effektivität von der ersten zur zweiten Follow-up-Messung ( $d = 0.04$  vs.  $0.06$ ; Fixed-Effects-Model). Auch diese waren jedoch jeweils nicht signifikant von Null verschieden ( $p = .35$  bzw.  $p = .19$ ). In Abbildung 3 ist die Stabilität von der post-Messung jeweils zum ersten und zweiten Follow-up-Zeitpunkt sowie von der ersten zur zweiten Follow-up-Messung dargestellt.

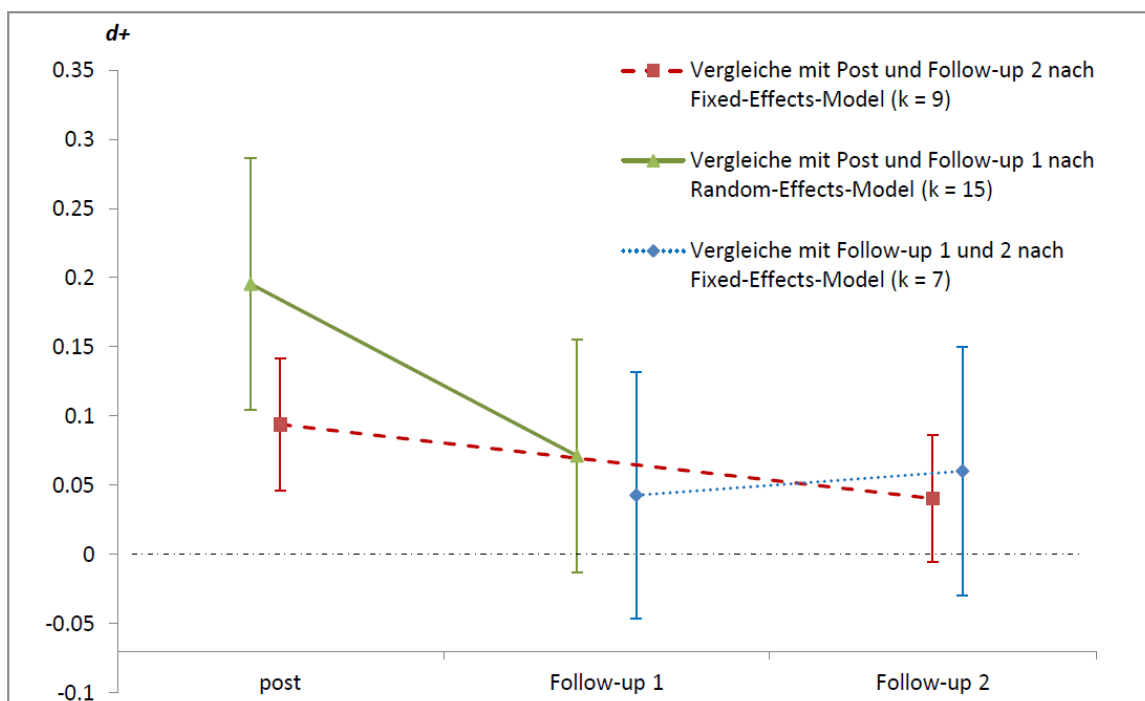


Abbildung 3. Stabilität der gewichteten mittleren Effektstärken ( $d_+$ ) über die Zeit (mit 95 %-Konfidenzintervallen).

#### 4.5 Wirksamkeit nach Erfolgskriterien

Die Wirksamkeit präventiver Interventionen zur sozio-emotionalen Förderung bei Kindern und Jugendlichen mit Migrationshintergrund kann für verschiedene Erfolgs-

kriterien unterschiedlich sein. Zugleich ist es von wissenschaftlicher, aber auch praktischer Relevanz, welche Probleme, Kompetenzen oder Verhaltensweisen durch die Interventionen besser beeinflusst werden können und welche weniger gut modifizierbar sind. Daher soll im Folgenden auf die Wirksamkeit der Programme zum einen auf die verschiedenen Erfolgskriterien im Bereich der sozialen und emotionalen Parameter, zum anderen auf weitere Kategorien von Erfolgskriterien eingegangen werden. Die ausführliche Darstellung der entsprechenden Effektstärken ist in Tabelle 4 zu finden.

Hierbei zeigte sich zunächst, dass von den 76 Vergleichen mit berichteten Post-Effektstärken 67 Vergleiche in mindestens einem der sozialen und emotionalen Maße Erfolgsmessungen vornahmen und diese insgesamt 309 Einzeleffektstärken umfassten. Außerhalb sozialer und emotionaler Parameter waren Messungen im Bereich des familiären Kontexts mit 105 Einzeleffekten in 18 Vergleichen und im Bereich der kognitiven Entwicklung mit 55 Einzeleffekten in 20 Vergleichen am häufigsten vertreten. Bis auf den Bereich des Schulkontextes mit nur 4 Vergleichen, das Sexualverhalten mit 6 Vergleichen und sonstige soziale und emotionale Maße mit 9 Vergleichen kamen alle Erfolgskriterien in mindestens 10 der 76 Vergleiche vor.

Die gewichtete mittlere Effektstärke für alle Messungen im Bereich der sozialen und emotionalen Entwicklung lag mit  $d_+ = 0.16$  nur geringfügig unter der über alle Erfolgskriterien hinweg ermittelten Gesamteffektstärke. Die höchsten gewichteten mittleren Effektstärken in diesem Bereich zeigten sich für internalisierende Probleme mit  $d_+ = 0.31$ . Am niedrigsten waren die Effekte für Suchtmittelkonsum mit  $d_+ = 0.09$  und externalisierende Probleme mit  $d_+ = 0.12$ . Außerhalb der Konstruktkategorie der sozialen und emotionalen Parameter fanden sich zum Teil sogar noch höhere Effekte, insbesondere für bildungsbezogene Kriterien mit  $d_+ = 0.36$ . Lediglich Erfolgsmessungen im Bereich des Schulkontextes wiesen mit  $d_+ = 0.08$  eine unterdurchschnittliche (nicht signifikant von Null verschiedene) gewichtete mittlere Effektstärke auf, wobei hier neben der geringen Zahl an Vergleichen auch die große Streuung in Abhängigkeit vom Berechnungsverfahren zu beachten ist.

Tabelle 4

*Wirksamkeit der präventiven Interventionen nach verschiedenen Erfolgskriterien auf Effektstärken- und Vergleichsebene*

Erfolgskriterien	Effektstärkenebene				Vergleichsebene			
	$n_{es}$	$M_{des}$	$SD_{des}$	$k$	$M_d$	$SD_d$	$d_+ \pm CI^a$	$Q (df)$
Soziale und emotionale Parameter	309	0.16	0.26	67	0.18	0.21	$0.16 \pm 0.04$	90.73 (66)*
Internalisierende Probleme	31	0.30	0.29	18	0.33	0.31	$0.31 \pm 0.14$	34.80 (17)**
Externalisierende Probleme	51	0.15	0.25	26	0.15	0.23	$0.12 \pm 0.09$	41.67 (25)*
Selbstbezogene Reflektion	45	0.19	0.34	28	0.24	0.38	$0.22 \pm 0.13$	72.56 (27)***
Prosoziale(s) Kognitionen/Verhalten	34	0.14	0.18	17	0.15	0.20	$0.20 \pm 0.06$	24.11 (16)
Sexualverhalten	22	0.24	0.24	6	0.16	0.18	$0.15 \pm 0.10$	5.34 (5)
Suchtmittelkonsum	78	0.12	0.19	18	0.11	0.13	$0.09 \pm 0.04$	9.43 (17)
Allgemeine Verhaltensprobleme	16	0.16	0.30	10	0.13	0.27	$0.15 \pm 0.16$	21.96 (9)**
Sonstige	32	0.07	0.28	9	0.16	0.31	$0.20 \pm 0.23$	36.79 (8)***
Kognitive Entwicklung	55	0.27	0.35	20	0.27	0.27	$0.25 \pm 0.11$	53.47 (19)***
Bildungsbezogene Kriterien	32	0.26	0.36	16	0.34	0.45	$0.36 \pm 0.24$	126.86 (15)***
Schulkontext	28	0.45	0.60	4	0.24	0.36	$0.08 \pm 0.18$	2.66 (3)
Familiärer Kontext	105	0.16	0.22	18	0.21	0.17	$0.17 \pm 0.07$	22.89 (17)

*Anmerkungen.*  $n_{es}$  = Anzahl der Effektstärken.  $M_{des}$  = Mittlere Effektstärke auf Effektstärkenebene.  $k$  = Anzahl der Vergleiche.  $M_d$  = ungewichtete mittlere Effektstärke auf Vergleichsebene.  $Q$  = Heterogenität.

<sup>a</sup> gewichtete mittlere Effektstärke mit 95 %-Konfidenzintervall; bei signifikanter Heterogenität wurde das Random-Effects-Model verwendet.

\*  $p < .05$ ; \*\*  $p < .01$ ; \*\*\*  $p < .001$ .



#### 4.6 Publikationsverzerrung

Um die gewichtete mittlere Effektstärke gegen den Einfluss von Publikationsverzerrungen abzusichern, wurden verschiedene Ansätze verwendet.

Zunächst wurde mithilfe des Fail-Safe N ermittelt, wie viele Vergleiche aus unveröffentlichten Studien mit Effektstärken von Null zum gefundenen Datenpool hinzukommen müssten, damit der Gesamteffekt nicht mehr signifikant wäre. Die Berechnung ergab eine Zahl von 2 428 Vergleichen, die sich entsprechend des File-Drawer Problems in Studien befinden müssten, welche bisher nicht publiziert wurden. Dies ist mehr als das 30-fache der Anzahl der in der Meta-Analyse enthaltenen Vergleiche und liegt damit deutlich über der von Rosenthal (1979) vorgeschlagenen Grenze, welche für die Größe der vorliegenden Meta-Analyse bei 390 Vergleichen läge. Somit ist es sehr unwahrscheinlich, dass die gefundene signifikante Effektstärke allein auf das File-Drawer Problem zurückführbar war.

Um die Ergebnisse dieser Meta-Analyse zur allgemeinen Wirksamkeit der Programme auf Hinweise auf eine Publikationsverzerrung zu untersuchen, wurde außerdem mithilfe von Begg's Rangkorrelationsmethode der Zusammenhang der Stichprobengröße mit den standardisierten Effektstärken der Vergleiche untersucht (Begg & Mazumdar, 1994). Die Analysen ergaben für diesen Zusammenhang einen nicht signifikanten Rangkorrelationskoeffizienten Kendall's tau von  $\tau = -.10$  ( $p = .20$ ). Folglich ergibt sich auch aus diesen Analysen kein Hinweis auf eine Publikationsverzerrung.

Eine dritte und letzte Möglichkeit die meta-analytisch integrierten Daten auf eine Publikationsverzerrung hin zu überprüfen, bietet der direkte Vergleich von Ergebnissen aus in Zeitschriften oder Monographien publizierten Studien mit denen aus Dissertationen. Hier zeigte sich, dass die Ergebnisse von Dissertationen nicht etwa, wie man es bei Vorliegen einer Publikationsverzerrung annehmen müsste, eine kleinere gewichtete mittlere Effektstärke aufwiesen als publizierte bzw. leicht zugängliche Studien, sondern dass diese sogar tendenziell, wenn auch nicht signifikant größer waren ( $Q_b = 2.96$ ,  $df = 1$ ,  $p = .09$ ). Die gewichtete mittlere Effektstärke nach dem Fixed-Effects-Model lag dabei für die insgesamt immerhin 15 Vergleiche innerhalb von Dissertationen bei  $d = 0.24$ , für Vergleiche innerhalb publizierter Arbeiten jedoch lediglich bei  $d = 0.16$ . Demzufolge kann auch in diesem direkten Vergleich kein Hinweis darauf gefunden

werden, dass die gefundenen Ergebnisse aufgrund von Publikationsverzerrungen positiv nach oben hin verfälscht sind.

#### **4.7 Moderatoranalysen**

Im Folgenden soll dargestellt werden, welchen Einfluss verschiedene Moderatorvariablen auf die berichtete Wirksamkeit hatten. Hierzu soll betrachtet werden, ob die auf Vergleichsebene aggregierten Post-Effektstärken in Abhängigkeit von publikationsbezogenen Charakteristika, methodischen Merkmalen, Stichprobenmerkmalen und Merkmalen der Präventionsprogramme schwankten. Da sich in den vorangegangenen Analysen keine signifikante Heterogenität der Effektstärken des Post-Zeitraums zeigte, sollte auch in den folgenden Moderatoranalysen das Fixed-Effects-Model als Grundlage dienen.

##### **4.7.1 Publikationsbezogene Merkmale**

Als erstes galt es zu untersuchen, ob publikationsbezogene Merkmale, also Variablen, die sich auf den gesamten Forschungsbericht beziehen und folglich auf Dokumentenebene kodiert wurden, einen Zusammenhang mit der Höhe der ermittelten Effektstärken aufweisen. Da im Kapitel zur Publikationsverzerrung bereits auf den Publikationstyp eingegangen wurde, sollen dabei das Publikationsjahr sowie das Land, in dem die Untersuchung stattfand, näher betrachtet werden (siehe Tabelle 5).

Tabelle 5

*Moderatoranalysen für publikationsbezogene Merkmale*

Merkmal	$d_+ \pm CI$	$k$	$Q_b (df)$
Publikationsjahr			12.95 (5)*
1960 – 1969	0.39 ± 0.20	1	
1970 – 1979	0.19 ± 0.21	3	
1980 – 1989	0.22 ± 0.16	6	
1990 – 1999	0.25 ± 0.11	15	
2000 – 2009	0.19 ± 0.04	32	
2010 – 2013	0.12 ± 0.04	19	
Erhebungsland			5.67 (5)
USA	0.17 ± 0.03	65	
Kanada	0.17 ± 0.26	2	
Niederlande	0.28 ± 0.31	1	
Deutschland	0.02 ± 0.19	6	
Norwegen	-0.01 ± 0.40	1	
Israel	0.45 ± 0.40	1	

*Anmerkungen.*  $d_+ \pm CI$  = gewichtete mittlere Effektstärke nach dem Fixed-Effects-Model mit 95 %-Konfidenzintervall.  $k$  = Anzahl der Vergleiche.  $Q_b$  = Heterogenität zwischen den Gruppen.

\*  $p < .05$ ; \*\*  $p < .01$ ; \*\*\*  $p < .001$ .

Hinsichtlich des Jahres der Veröffentlichung zeigten sich signifikante Unterschiede in den Effektstärken ( $Q_b = 12.95$ ,  $df = 5$ ,  $p < .05$ ). Die niedrigste Effektstärke ergab sich für Studien, die zwischen 2010 und 2013 veröffentlicht wurden ( $d_+ = 0.12$ ), die höchste hingegen für Publikationsjahre vor 1970 ( $d_+ = 0.39$ ). Da letztere sich jedoch nur auf einen einzigen Vergleich bezog und aus den gewählten Kategorien keine eindeutige lineare Tendenz ablesbar war, wurden zusätzlich alle Vergleichseffektstärken aus Studien, welche vor dem Jahr 2000 veröffentlicht wurden, mit all jenen verglichen, deren Publikation ab 2000 erfolgte und zudem eine Meta-Regression mit der kontinuierlichen Variable des Publikationsjahres berechnet. Bezüglich ersterer Analyse ergab sich eine signifikant größere Effektstärke für die älteren, vor dem Jahr 2000 veröffentlichten Studien im Vergleich zu den später publizierten Arbeiten ( $d_+ = 0.25$  vs.  $0.15$ ;  $Q_b = 5.62$ ,  $df = 1$ ,  $p < .05$ ). Auch für die Meta-Regression ergab sich ein Abfall der Effektstärken mit steigendem Publikationsjahr (standardisiertes  $\beta = -0.22$ ,  $Q_R = 4.85$ ,  $df = 1$ ,  $p < .05$ ).

Keine signifikanten Unterschiede im Hinblick auf die Effektstärken zeigten sich in Abhängigkeit davon, in welchem Land die Untersuchung stattfand ( $Q_b = 5.67$ ,  $df = 5$ ,  $p = .34$ ). Jedoch war hierbei die Zellenbesetzung zum Teil sehr gering, weswegen die Teststärke in dieser Analyse begrenzt war.

#### 4.7.2 Methodische Merkmale

In den nächsten Analysen wurde der Fragestellung nachgegangen, ob sich Unterschiede in den berichteten Wirksamkeitsparametern der Gruppenvergleiche in Abhängigkeit von verschiedenen methodischen Merkmalen zeigen (siehe Tabelle 6). Da methodische Schwächen nicht nur die mittlere Höhe ermittelter Effekte beeinflussen können, sondern auch zu einer größeren Ungenauigkeit und somit Streuung von Forschungsergebnissen beitragen können, soll dabei auch die Heterogenität innerhalb der verschiedenen Ausprägungen der methodischen Variablen Beachtung finden.

In den Analysen zeigte sich, dass bei Vergleichen, in denen die Intervention mit einer Kontrollgruppe kontrastiert wurde, welche keinerlei Behandlung erfuhr, signifikant größere Effekte auftraten als bei solchen, in denen eine Kontrollgruppe genutzt wurde, die ein unspezifisches, minimales oder ein Standardtraining erhielt ( $d_+ = 0.21$  vs.  $0.15$ ;  $Q_b = 4.74$ ,  $df = 1$ ,  $p < .05$ ). Während sich für Vergleiche mit unbehandelter Kontrollgruppe keine signifikante Heterogenität zeigte ( $Q = 27.86$ ,  $df = 34$ ,  $p = .76$ ;  $I^2 = 0\%$ ), ergab sich für Vergleiche mit aktiver Kontrollgruppe eine signifikante Heterogenität ( $Q = 63.56$ ,  $df = 40$ ,  $p < .05$ ;  $I^2 = 37.1\%$ ).

Weiterhin ergaben sich Unterschiede hinsichtlich der Art der Randomisierung. Wurde diese auf individueller Ebene vorgenommen, also einzelne Teilnehmer zufällig zugewiesen, so zeigte sich eine größere Wirksamkeit der untersuchten Interventionen, als wenn stattdessen ganze Cluster von Teilnehmern, also beispielsweise Schulen oder Klassen randomisiert auf die Untersuchungsbedingungen verteilt wurden ( $d_+ = 0.21$  vs.  $0.14$ ;  $Q_b = 5.88$ ,  $df = 1$ ,  $p < .05$ ). Dabei ergaben sich zwar innerhalb der Kategorien keine signifikanten Heterogenitäten, jedoch waren deskriptiv betrachtet Vergleiche, bei denen die Randomisierung auf der Ebene von Clustern durchgeführt wurde, deutlich heterogener als solche mit Zufallszuweisung auf individueller Ebene ( $Q = 35.08$ ,  $df = 23$ ,  $p = .05$ ;  $I^2 = 34.4\%$  bzw.  $Q = 55.21$ ,  $df = 51$ ,  $p = .32$ ;  $I^2 = 7.6\%$ ).

Tabelle 6

*Moderatoranalysen für methodische Merkmale*

Merkmal	$d_+ \pm CI$	$k$	$Q (df)$	$Q_b (df)$
Kontrollgruppe				4.74 (1)*
Passiv	0.21 ± 0.05	35	27.86 (34)	
Aktiv	0.15 ± 0.03	41	63.56 (40)*	
Randomisierung				5.88 (1)*
Auf individueller Ebene	0.21 ± 0.05	52	55.21 (51)	
Auf Ebene von Clustern	0.14 ± 0.03	24	35.08 (23)	
Äquivalenz der Gruppen vor Intervention				18.27 (2)***
Äquivalent	0.18 ± 0.05	43	45.17 (42)	
Nicht äquivalent	0.26 ± 0.06	16	20.80 (15)	
Nicht getestet bzw. nicht berichtet	0.11 ± 0.04	17	11.93 (16)	
Mittlere Dropout-Rate (in %)				5.11 (6)
≤ 5	0.24 ± 0.09	17	11.18 (16)	
5 – 10	0.22 ± 0.10	15	15.71 (14)	
10 – 20	0.15 ± 0.06	14	23.35 (13)*	
20 – 30	0.19 ± 0.09	12	6.97 (11)	
30 – 40	0.16 ± 0.05	6	20.09 (5)**	
40 – 50	0.06 ± 0.20	2	0.33 (1)	
> 50	0.16 ± 0.14	2	5.16 (1)*	
Datenanalysemethode <sup>a</sup>				6.98 (2)*
Per-protocol mit aktivem Ausschluss	0.26 ± 0.13	10	12.09 (9)	
Per-protocol ohne aktiven Ausschluss	0.18 ± 0.04	51	53.22 (50)	
Intention-to-treat	0.11 ± 0.05	9	16.03 (8)*	

*Anmerkungen.*  $d_+ \pm CI$  = gewichtete mittlere Effektstärke nach dem Fixed-Effects-Model mit 95 %-Konfidenzintervall.  $k$  = Anzahl der Vergleiche.  $Q$  = Heterogenität innerhalb der einzelnen Gruppen.  $Q_b$  = Heterogenität zwischen den Gruppen.

<sup>a</sup> bei der Methode Per-protocol wurde unterschieden zwischen Analysen die auch Teilnehmer ausschlossen, die eine bestimmte Mindestteilnahmerate an den Sitzungen der Intervention nicht erreichten (mit aktivem Ausschluss) und Analysen, bei denen lediglich jene Teilnehmer nicht berücksichtigt wurden, die nicht an der Post-Testung teilnahmen (ohne aktiven Ausschluss).

\*  $p < .05$ ; \*\*  $p < .01$ ; \*\*\*  $p < .001$ .

Eine weitere bedeutsame Moderatorvariable war die Äquivalenz der beiden Gruppen vor Beginn der Intervention ( $Q_b = 18.27$ ,  $df = 2$ ,  $p < .001$ ). Hierbei zeigte sich für Vergleiche, bei denen diese nicht getestet wurde, die niedrigste Effektstärke ( $d_+ = 0.11$ ). Wurde die Äquivalenz der Gruppen getestet, so ergaben sich höhere Effekte in Fällen, in denen die Gruppen nicht vergleichbar waren ( $d_+ = 0.26$ ) als bei Kontrasten, für die eine anfängliche Äquivalenz bestand ( $d_+ = 0.18$ ). Dieser Unterschied erwies sich auch bei separater Betrachtung der Vergleiche, in denen die Äquivalenz getestet wurde, als signifikant ( $Q_b = 4.10$ ,  $df = 1$ ,  $p < .05$ ). Kontrastierte man zudem alle Gruppenvergleiche, für welche die Äquivalenz überprüft wurde ( $d_+ = 0.21$ ), mit jenen für die keine Testung stattfand, so ergab sich auch hier ein signifikanter Unterschied ( $Q_b = 14.17$ ,  $df = 1$ ,  $p < .001$ ). Für keine der drei Kategorien waren die Effektstärken signifikant heterogen, jedoch ergab sich prozentual zumindest eine *kleine* Heterogenität für Kontraste, bei denen die Gruppen nicht vergleichbar waren ( $Q = 20.80$ ,  $df = 15$ ,  $p = .14$ ;  $I^2 = 27.9\%$ ). Für Vergleiche, bei denen eine Gruppenäquivalenz positiv getestet wurde ( $Q = 45.17$ ,  $df = 42$ ,  $p = .34$ ;  $I^2 = 7.0\%$ ) sowie für solche, bei denen keine Testung auf Äquivalenz stattfand, war die Heterogenität geringer ( $Q = 11.93$ ,  $df = 16$ ,  $p = .75$ ;  $I^2 = 0\%$ ).

Die mittlere Dropout-Rate hing nicht signifikant mit der Effektstärke zusammen ( $Q_b = 5.11$ ,  $df = 6$ ,  $p = .53$ ). Auch in einer Regression mit der kontinuierlich gemessenen Dropout-Rate ergab sich kein signifikanter Zusammenhang ( $\beta = -0.15$ ,  $Q_R = 2.00$ ,  $df = 1$ ,  $p = .16$ ). Während sich für Vergleiche mit einem Dropout von 0 bis 5 % ( $Q = 11.18$ ,  $df = 16$ ,  $p = .80$ ;  $I^2 = 0\%$ ) und zwischen 5 und 10 % ( $Q = 15.71$ ,  $df = 14$ ,  $p = .33$ ;  $I^2 = 10.9\%$ ) keine signifikante Heterogenität zeigte, war die Heterogenität für Vergleiche mit über 50 %-iger Dropout-Rate am größten ( $Q = 5.16$ ,  $df = 1$ ,  $p < .05$ ;  $I^2 = 80.62\%$ ). Kontrastierte man all jene Vergleiche für die eine Dropout-Rate von bis zu 10 % vorlag mit solchen, die eine Dropout-Rate über 10 % aufwiesen, so zeigte sich, dass für erstere die Heterogenität deutlich niedriger und im Gegensatz zu Gruppenvergleichen mit höherer Dropout-Rate nicht signifikant ausfiel ( $Q = 26.94$ ,  $df = 31$ ,  $p = .68$ ;  $I^2 = 0\%$  bzw.  $Q = 57.40$ ,  $df = 35$ ,  $p < .01$ ;  $I^2 = 39.0\%$ ).

Als letzte methodische Moderatorvariable wurde die Datenanalysemethode untersucht. Auch diese wies einen signifikanten Zusammenhang mit der Effektstärke auf ( $Q_b = 6.98$ ,  $df = 2$ ,  $p < .05$ ). Untersuchungen, bei denen von den Autoren explizit Personen von der Analyse ausgeschlossen wurden, die ein Mindestteilnahmekriterium nicht erfüllten, zeigten dabei den höchsten Effekt ( $d_+ = 0.26$ ). Vergleiche, bei denen

Personen zwar nicht explizit aufgrund einer geringen Teilnehmerate ausgeschlossen wurden, jedoch auch nur diejenigen in die Analysen einbezogen wurden, die auch an der Post-Messung teilgenommen hatten, ergaben etwas geringere Effekte ( $d_+ = 0.18$ ). Die niedrigste Effektstärke ergab sich jedoch für Intention-to-treat-Analysen ( $d_+ = 0.11$ ). Eine signifikante Heterogenität ergab sich lediglich für die Intention-to-treat-Analysen ( $Q = 16.04$ ,  $df = 8$ ,  $p < .05$ ;  $I^2 = 50.1\%$ ). Vergleiche, bei denen ein teilnahmebezogener Ausschluss vorgenommen wurde, wiesen zumindest deskriptiv betrachtet eine größere Heterogenität auf ( $Q = 12.09$ ,  $df = 9$ ,  $p = .21$ ;  $I^2 = 25.6\%$ ) als Vergleiche, bei denen alle eingeschlossen wurden, die an der Post-Testung teilnahmen ( $Q = 53.22$ ,  $df = 50$ ,  $p = .35$ ;  $I^2 = 6.1\%$ ).

#### *4.7.3 Merkmale der Präventionsprogramme*

Im Folgenden soll dargelegt werden, inwieweit Merkmale der untersuchten präventiven Interventionen im Zusammenhang mit den ermittelten Wirksamkeitsparametern standen.

Hierbei sollten als erstes die Inhalte der Programme einer Analyse unterzogen werden. Da sehr viele Interventionen mehr als nur ein konkretes Thema bearbeiteten und einige Inhalte aufgrund konzeptioneller Nähe hauptsächlich in Kombination auftraten (z.B. Sprach und Leseprogramme in Kombination mit kognitiven oder Lernprogrammen oder Substanzmissbrauchsprogramme in Kombination mit sozialen Kompetenztrainings) wurde hierfür zunächst eine Grobkategorisierung vorgenommen. Programme, die im weitesten Sinne Sprache, Kognition oder Lernen trainierten, wurden zusammengefasst. Ebenso wurden alle Programme, die Substanzmissbrauch oder Sexualprävention thematisierten unter soziale und emotionale Trainingsprogramme subsumiert. Die Tabelle 7 gibt einen Überblick über die Wirksamkeit in Abhängigkeit von den verschiedenen Programminhalten.

Tabelle 7

*Wirksamkeit in Abhängigkeit von verschiedenen Programminhalten*

Inhalte	$d_+ \pm CI$	$k$
Fokussierte Programme	$0.15 \pm 0.03$	57
Elternt trainingsprogramme	$0.21 \pm 0.11$	10
Lehrertrainingsprogramme	$0.24 \pm 0.17$	3
Sprach-/Kognitions-/Lernprogramme	$0.14 \pm 0.09$	3
Soziale und emotionale Trainingsprogramme	$0.14 \pm 0.03$	41
mit Komponente Substanzmissbrauch	$0.10 \pm 0.04$	13
mit Komponente Sexualprävention	$0.18 \pm 0.11$	4
Kombinierte Programme	$0.29 \pm 0.09$	16
Sprach-/Kognitions-/Lernprogramme in Kombination mit sozialen und emotionalen Trainingsprogrammen	$0.20 \pm 0.15$	7
Elternt trainingsprogramme in Kombination mit sozialen und emotionalen Trainingsprogrammen	$0.34 \pm 0.12$	8
Mehrkomponentenprogramme	$0.26 \pm 0.37$	1
Sonstige Programme	$0.16 \pm 0.17$	3

*Anmerkungen.*  $d_+ \pm CI$  = gewichtete mittlere Effektstärke nach dem Fixed-Effects-Model mit 95 %-Konfidenzintervall.  $k$  = Anzahl der Vergleiche.

Aus dieser ist zu erkennen, dass die meisten Programme auf einzelne Inhaltsbereiche fokussierten ( $k = 57$ ), jedoch auch eine substantielle Zahl an Vergleichen kombinierte Programme zum Gegenstand hatten ( $k = 16$ ). Letztere hatten dabei eine höhere Effektstärke ( $d = 0.29$ ) als fokussierte Programme ( $d = 0.15$ ), wobei ein direkter Vergleich dieser Oberkategorien ergab, dass es sich hierbei um einen signifikanten Unterschied handelte ( $Q_b = 8.04$ ,  $df = 1$ ,  $p < .01$ ). Lediglich drei Programme waren nicht eindeutig zu kategorisieren. Da diese jedoch eine mittlere Effektstärke aufwiesen, welche sich nicht wesentlich von den Effektstärken anderer Programme unterschieden ( $d_+ = 0.16$ ) und innerhalb dieser Kategorie sonstiger Programme keine signifikante Heterogenität bestand ( $Q = 3.05$ ,  $df = 2$ ,  $p = .15$ ), wurde sie nicht weiter differenziert. Im Bereich fokussierter Programme waren am häufigsten solche mit sozialem oder emotionalem Training vertreten ( $k = 41$ ). Gleichzeitig ergaben sich für diese jedoch nur unterdurchschnittliche Effekte ( $d_+ = 0.14$ ). Innerhalb der sozialen und emotionalen Trainings-



programme zeigte sich auch, dass Programme, die unter anderem auch Substanzmissbrauch thematisierten, signifikant kleinere Effekte aufwiesen als Programme ohne diesen Inhalt ( $d_+ = 0.10$  vs.  $0.21$ ;  $Q_b = 11.73$ ,  $df = 1$ ,  $p < .001$ ). Am wirksamsten von allen Programmen waren solche, die sowohl ein soziales oder emotionales Trainingsprogramm für Kinder bzw. Jugendliche enthielten als auch deren Eltern trainierten ( $d_+ = 0.34$ ). Diese waren auch im direkten Vergleich signifikant wirksamer als Programme, die nur ein soziales oder emotionales Trainingsprogramm umfassten ohne zusätzlich die Eltern zu trainieren ( $Q_b = 10.52$ ,  $df = 1$ ,  $p < .01$ ). Auch waren Programme, die sich nur an die Kinder bzw. Jugendlichen wendeten, allgemein weniger wirksam als solche, welche die Eltern entweder allein oder in Kombination mit ihren Kindern einbezogen ( $d_+ = 0.15$  vs.  $0.27$ ;  $Q_b = 8.31$ ,  $df = 1$ ,  $p < .01$ ).

Über die Wirksamkeit der Programme in Abhängigkeit von verschiedenen sozio-emotionalen Zielen gibt Tabelle 8 einen Überblick. Da die wenigsten Programme ausschließlich ein spezifisches Ziel verfolgten, wurde auch hier, um die Vielzahl möglicher Kombinationen zu begrenzen, eine Grobkategorisierung vorgenommen und zwischen der sozio-emotionalen Gesundheitsförderung, in der die Förderung protektiver Faktoren im Vordergrund stand, Prävention internalisierender Probleme sowie Vermeidung externalisierender Probleme im weitesten Sinne unterschieden. Unter die Oberkategorie der Prävention externalisierender Probleme wurden dabei auch Programme zur Sexualprävention sowie zur Substanzmissbrauchsprävention subsumiert, da sich zwischen den entsprechenden Problemverhaltensweisen ein starker Zusammenhang gezeigt hat (Iacono et al., 2003; Timmermans et al., 2008).

Tabelle 8

*Wirksamkeit in Abhängigkeit von verschiedenen sozio-emotionalen Programmzielen*

Ziele	$d_+ \pm CI$	$k$
Sozio-emotionale Gesundheitsförderung	$0.21 \pm 0.05$	26
mit Förderung sozialer / sozialkognitiver Kompetenz	$0.22 \pm 0.05$	19
mit Förderung der Emotionsregulation	$0.08 \pm 0.09$	3
mit Förderung des Selbstwerts	$0.20 \pm 0.10$	12
mit Förderung der ethnischen Identität	$0.20 \pm 0.12$	11
Prävention internalisierender Probleme	$0.25 \pm 0.54$	1
Prävention externalisierender Probleme <sup>a</sup>	$0.13 \pm 0.06$	20
mit Prävention dissozialen Verhaltens	$0.14 \pm 0.10$	10
mit Prävention sexuellen Risikoverhaltens	$0.15 \pm 0.10$	5
mit Prävention von Substanzmissbrauch	$0.09 \pm 0.09$	12
Sozio-emotionale Gesundheitsförderung und Prävention internalisierender Probleme	$0.34 \pm 0.28$	3
Sozio-emotionale Gesundheitsförderung und Prävention externalisierender Probleme <sup>a</sup>	$0.11 \pm 0.04$	15
Prävention internalisierender und externalisierender Probleme <sup>a</sup>	$0.40 \pm 0.14$	4
Sozio-emotionale Gesundheitsförderung und Prävention internalisierender und externalisierender Probleme	$0.32 \pm 0.14$	7

*Anmerkungen.*  $d_+ \pm CI$  = gewichtete mittlere Effektstärke nach dem Fixed-Effects-Model mit 95 %-Konfidenzintervall.  $k$  = Anzahl der Vergleiche.

<sup>a</sup> Unter die Oberkategorie der Prävention externalisierender Probleme wurden hier nicht nur alle Programme zur Prävention dissozialen Verhaltens subsumiert, sondern auch jene, die auf sexuelles Risikoverhalten oder Substanzmissbrauch abzielten.

Es zeigte sich, dass die meisten Programme eine sozio-emotionale Gesundheitsförderung bezweckten ( $k = 26$ ) und hierbei insbesondere soziale bzw. sozial-kognitive Kompetenzen fördern sollten ( $k = 19$ ). Für diese Programme zeigten sich leicht überdurchschnittliche Effekte ( $d_+ = 0.21$  bzw.  $0.22$ ). Leicht unterdurchschnittliche Effekte zeigten sich hingegen für Programme, die der Prävention externalisierender Probleme dienten und innerhalb dieser insbesondere solche zur Prävention von Substanzmissbrauch ( $d_+ = 0.13$  und  $0.09$ ). Verfolgte das Programm zusätzlich das Ziel, auch internalisierenden Problemen

vorzubeugen bzw. diese zu reduzieren ( $d_+ = 0.40$ ), so zeigte sich ein signifikant höherer Effekt als in Fällen, bei denen nur die Prävention externalisierender Probleme angestrebt wurde ( $Q_b = 11.68$ ,  $df = 1$ ,  $p < .001$ ). Zwar waren Programme, die sowohl internalisierenden als auch externalisierenden Problemen präventiv entgegenwirken sollten, relativ selten ( $k = 4$ ), was die Interpretierbarkeit dieses Ergebnisses einschränken kann, jedoch ergab sich ein ähnliches Muster für Programme in denen zusätzlich das Ziel der sozio-emotionalen Gesundheitsförderung verfolgt wurde. Auch hier ergab sich für Programme, die der Prävention externalisierender Probleme sowie der sozio-emotionalen Gesundheitsförderung dienen sollten, geringere Effekte ( $d_+ = 0.11$ ) als für Programme, die neben diesen beiden Zielen auch die Prävention internalisierender Probleme bezweckten ( $d_+ = 0.32$ ;  $Q_b = 7.82$ ,  $df = 1$ ,  $p < .01$ ). Eine weitere Analyse ergab zudem, dass Programme, die unter anderem auf internalisierende Probleme abzielten, auch höhere Effekte auf externalisierende Probleme aufwiesen als Programme, welche sich die Prävention internalisierender Probleme nicht zum Ziel gesetzt hatten ( $d_+ = 0.28$ ,  $k = 7$  vs.  $d_+ = 0.06$ ,  $k = 19$ ;  $Q_b = 7.38$ ,  $df = 1$ ,  $p < .01$ ).

Bezüglich der Implementation der präventiven Interventionen zeigten sich ebenfalls einige Unterschiede im Hinblick auf deren Wirksamkeit (siehe Tabelle 9). Hinsichtlich des Trainings der Administratoren ergaben sich zwar für Programme, bei denen ein solches entweder nicht stattfand oder von den Autoren nicht berichtet wurde, überdurchschnittliche Effekte ( $d_+ = 0.22$ ), jedoch gab es für diese Moderatorvariable keine signifikanten Unterschiede hinsichtlich der Wirksamkeit ( $Q_b = 4.47$ ,  $df = 2$ ,  $p = .11$ ). Signifikante Heterogenität gab es jedoch in Abhängigkeit von der Supervision des Administrators ( $Q_b = 14.96$ ,  $df = 2$ ,  $p < .001$ ). Hierbei gab es keinen signifikanten Wirksamkeitsunterschied zwischen Programmen bei denen keine Supervision berichtet wurde ( $d_+ = 0.17$ ) und solchen für die eine Supervision geringer Intensität beschrieben wurde ( $d_+ = 0.11$ ;  $Q_b = 3.79$ ,  $df = 1$ ,  $p = .05$ ). Programme mit sehr intensiver Supervision wiesen jedoch eine signifikant höhere Effektivität auf ( $d_+ = 0.29$ ), sowohl im Vergleich zu Programmen, welche keine Supervision berichteten ( $Q_b = 6.50$ ,  $df = 1$ ,  $p < .05$ ), als auch zu solchen, in denen nur eine geringfügige Supervision stattfand ( $Q_b = 14.63$ ,  $df = 1$ ,  $p < .001$ ). Auch bezüglich teilnehmerorientierter Maßnahmen zur Implementationsförderung zeigte sich eine mit dem Ausmaß der Implementation steigende Wirksamkeit der Programme ( $d_+ = 0.15$  vs.  $0.22$  vs.  $0.33$ ;  $Q_b = 11.66$ ,  $df = 2$ ,  $p < .01$ ). Signifikante Unterschiede gab es zudem hinsichtlich berichteter Implementationsprobleme

( $Q_b = 11.74$ ,  $df = 2$ ,  $p < .01$ ). Hierbei zeigten sich jedoch für Vergleiche mit geringen Implementationsproblemen ( $d_+ = 0.27$ ) nicht nur höhere Effekte als für Gruppenvergleiche mit starken Implementationsproblemen ( $d_+ = 0.14$ ;  $Q_b = 8.58$ ,  $df = 1$ ,  $p < .01$ ), sondern auch höhere Effektstärken als für Vergleiche ohne Implementationsprobleme ( $d_+ = 0.14$ ;  $Q_b = 10.59$ ,  $df = 1$ ,  $p < .01$ ).

Tabelle 9

*Moderatoranalysen für Implementationsvariablen*

Merkmal	$d_+ \pm CI$	$k$	$Q_b (df)$
Training der Administratoren <sup>a</sup>			4.47 (2)
Keines bzw. keine Angabe	0.22 ± 0.06	21	
Wenig Training	0.15 ± 0.08	10	
Viel Training	0.15 ± 0.03	42	
Supervision der Administratoren <sup>a</sup>			14.96 (2)***
Keine bzw. keine Angabe	0.17 ± 0.04	47	
Wenig Supervision	0.11 ± 0.04	12	
Viel Supervision	0.29 ± 0.08	14	
Teilnehmerorientierte Maßnahmen			11.66 (2)**
Keine bzw. keine Angabe	0.15 ± 0.03	56	
Wenig teilnehmerorientierte Maßnahmen	0.22 ± 0.09	10	
Viele teilnehmerorientierte Maßnahmen	0.33 ± 0.11	10	
Implementationsprobleme			11.74 (2)**
Keine bzw. keine Angabe	0.14 ± 0.04	50	
Wenig Implementationsprobleme	0.27 ± 0.06	17	
Viele Implementationsprobleme	0.14 ± 0.06	9	

*Anmerkungen.*  $d_+ \pm CI$  = gewichtete mittlere Effektstärke nach dem Fixed-Effects-Model mit 95 %-Konfidenzintervall.  $k$  = Anzahl der Vergleiche.  $Q_b$  = Heterogenität zwischen den Gruppen.

<sup>a</sup> da es bei einigen Programmen keinen Administrator gab addiert sich die Anzahl der Vergleiche nicht auf 76 auf.

\*  $p < .05$ ; \*\*  $p < .01$ ; \*\*\*  $p < .001$ .

Als weitere Programmmerkmale soll im Folgenden auf die Programmintensität, das Interventionsformat, die Administratoren des Programmes und die kulturelle Angepasstheit eingegangen werden (siehe Tabelle 10).

Tabelle 10

*Moderatoranalysen für weitere Programmmerkmale*

Merkmal	$d_+ \pm CI$	$k$	$Q_b (df)$
Gesamtinterventionsdauer in Stunden			12.65 (4)*
$\leq 5$	0.23 $\pm$ 0.06	20	
5–10	0.10 $\pm$ 0.05	15	
10–20	0.15 $\pm$ 0.07	13	
20–50	0.18 $\pm$ 0.06	19	
$> 50$	0.20 $\pm$ 0.07	9	
Interventionsformat <sup>a</sup>			5.64 (3)
Selbsthilfe	0.16 $\pm$ 0.46	1	
Individuell	0.23 $\pm$ 0.15	6	
Gruppe	0.16 $\pm$ 0.03	61	
Individuell und Gruppe	0.28 $\pm$ 0.10	7	
Administratoren der Intervention <sup>a</sup>			21.47 (5)***
Keine Person	0.19 $\pm$ 0.17	3	
Lehrpersonal	0.11 $\pm$ 0.04	21	
Psychosoziale Fachkräfte	0.19 $\pm$ 0.08	19	
Laien oder Freiwillige	0.28 $\pm$ 0.08	3	
Hochschulpersonal	0.18 $\pm$ 0.09	18	
Mehrere verschiedene Administratoren	0.28 $\pm$ 0.10	9	
Kulturelle Angepasstheit			2.33 (1)
Kulturell zugeschnitten/angepasst	0.18 $\pm$ 0.04	42	
Nicht kulturell zugeschnitten/angepasst	0.14 $\pm$ 0.04	34	

*Anmerkungen.*  $d_+ \pm CI$  = gewichtete mittlere Effektstärke nach dem Fixed-Effects-Model mit 95 %-Konfidenzintervall.  $k$  = Anzahl der Vergleiche.  $Q_b$  = Heterogenität zwischen den Gruppen.

<sup>a</sup> nicht alle Studien gaben Auskunft über das Programmmerkmal, weswegen sich die Zahl der Vergleiche nicht auf 76 aufaddiert.

\*  $p < .05$ ; \*\*  $p < .01$ ; \*\*\*  $p < .001$ .

Um den Zusammenhang zwischen der Programmintensität und der Effektstärke zu ermitteln wurde ein Gesamtmaß der Interventionsdauer errechnet, welches sich als Produkt der Anzahl der Sitzungen und der durchschnittlichen Dauer der Sitzungen in Stunden ergab. Für die meta-analytische Varianzanalyse wurde diese Gesamtinterventionsdauer zunächst kategorial betrachtet in den Stufen bis 5 Stunden, 5 bis 10, 10 bis 20, 20 bis 50 und mehr als 50 Stunden. Dabei zeigte sich ein signifikanter Unterschied hinsichtlich dieser Kategorisierung ( $Q_b = 12.64$ ,  $df = 4$ ,  $p < .05$ ). Allerdings wies dieser nicht linear in eine bestimmte Richtung. Die höchsten Effekte ergaben sich demnach für Programme mit einer Dauer bis zu fünf Stunden ( $d_+ = 0.23$ ), die niedrigsten jedoch für die nächstfolgende Kategorie von Programmen mit 5-10-stündiger Dauer ( $d_+ = 0.10$ ). Für länger andauernde Programme gab es dann wiederum einen Anstieg der Effektstärke mit zunehmender Dauer, sodass sich der zweithöchste Effekt für Programme ergab, die mehr als 50 Stunden umfassten ( $d_+ = 0.20$ ). Fasste man alle 5- bis 20-stündigen Programme und alle Programme über 20 Stunden zusammen, so zeigten sich für erstere zum einen signifikant niedrigere Effektstärken als für letztere ( $d_+ = 0.11$  vs.  $0.19$ ;  $Q_b = 4.96$ ,  $df = 1$ ,  $p < .05$ ), zum anderen aber auch geringere Effekte als für maximal 5-stündige Programme ( $d_+ = 0.11$  vs.  $0.23$ ;  $Q_b = 9.95$ ,  $df = 1$ ,  $p < .01$ ). Ein signifikanter Unterschied zwischen den bis zu 5-stündigen und den über 20-stündigen Programmen zeigte sich hierbei nicht ( $d_+ = 0.23$  vs.  $0.19$ ;  $Q_b = 1.43$ ,  $df = 1$ ,  $p = .23$ ). Auch ein linearer Zusammenhang zwischen der Effektstärke und dem kontinuierlichen Maß der Gesamtinterventionsdauer ergab sich innerhalb einer Meta-Regression nicht ( $\beta = -0.03$ ,  $Q_R = 0.10$ ,  $df = 1$ ,  $p = .75$ ).

Hinsichtlich des Interventionsformats zeigte sich kein signifikanter Unterschied, wobei zu beachten ist, dass die Zellenbesetzung zum Teil recht klein war ( $k < 8$  für drei der vier Kategorien;  $Q_b = 5.64$ ,  $df = 3$ ,  $p = .13$ ). Allerdings schienen Programme, die zumindest teilweise individuell durchgeführt wurden, tendenziell höhere Effektstärken zu zeigen. Fasste man Programme, die zum Teil individuell, zum Teil in Gruppen durchgeführt wurden mit ausschließlich individuell durchgeführten Programmen zusammen, so zeigte sich für diese ein signifikant höherer Effekt als für ausschließlich in Gruppen durchgeführte Programme ( $d_+ = 0.26$  vs.  $0.16$ ;  $Q_b = 5.34$ ,  $df = 1$ ,  $p < .05$ ).

Die Effektstärken wiesen eine signifikante Heterogenität in Abhängigkeit von den Administratoren der präventiven Intervention auf ( $Q_b = 21.47$ ,  $df = 5$ ,  $p < .001$ ). Die niedrigste Effektstärke zeigte sich dabei für Lehrpersonal ( $d_+ = 0.11$ ). Die höchsten Effekte ergaben sich für Laien oder Freiwillige ( $d_+ = 0.28$ ), für die jedoch nur eine

Datengrundlage von drei Vergleichen existierte, sowie für Programme mit mehreren verschiedenen Administratoren ( $d_+ = 0.28$ ). Kontrastierte man Programme, die von Lehrpersonal durchgeführt wurden mit allen anderen Programmen ( $d_+ = 0.23$ ), so zeigte sich, dass erstere niedrigere Effektstärken aufwiesen ( $Q_b = 16.51$ ,  $df = 1$ ,  $p < .001$ ).

Hinsichtlich der kulturellen Angepasstheit der Intervention zeigten sich keine signifikanten Unterschiede. Zwar wiesen Programme, welche kulturell zugeschnitten bzw. angepasst waren, tendenziell höhere Effekte auf als Programme, die dies nicht waren, jedoch war dieser Unterschied statistisch nicht signifikant ( $d_+ = 0.18$  vs.  $0.14$ ;  $Q_b = 2.33$ ,  $df = 1$ ,  $p = .13$ ).

#### 4.7.4 Stichprobenmerkmale

Im Folgenden soll der Zusammenhang der Vergleichseffektstärken mit verschiedenen Stichprobenmerkmalen betrachtet werden. Alle Analysen hierzu sind Tabelle 11 zu entnehmen.

Hinsichtlich der Herkunft der Kinder und Jugendlichen mit Migrationshintergrund zeigte sich kein signifikanter Zusammenhang mit den ermittelten Effektstärken ( $Q_b = 6.57$ ,  $df = 9$ ,  $p = .68$ ). Hierbei ist jedoch zu beachten, dass lediglich hispano-amerikanische Gruppen (insgesamt  $k = 60$ ) sowie Stichproben mit verschiedenen unterschiedlichen Migrantengruppen ( $k = 8$ ) in ausreichender Zahl vorhanden waren, wodurch die Sensibilität der Analyse für Gruppenunterschiede eingeschränkt ist.

Auch hinsichtlich der Risikobelastung der Stichprobe ergaben sich keine signifikanten Unterschiede ( $Q_b = 0.28$ ,  $df = 1$ ,  $p = .60$ ). Vergleiche, in denen die Stichprobe der Kinder und Jugendlichen mit Migrationshintergrund überwiegend bereits vor Interventionsbeginn eine kognitive, Lern- oder Verhaltensauffälligkeit aufwies (indizierte Prävention), unterschieden sich folglich hinsichtlich der Effektstärke nicht von Vergleichen, bei denen dies nicht der Fall war ( $d_+ = 0.19$  vs.  $0.16$ ).

Für das mittlere Alter der Stichprobe zeigte sich ebenfalls kein Unterschied im kategorialen Vergleich ( $Q_b = 8.10$ ,  $df = 4$ ,  $p = .09$ ). Auch eine Regressionsanalyse ergab keinen signifikanten Zusammenhang zwischen der kontinuierlichen Variable des Alters und der Effektstärke ( $\beta = 0.06$ ,  $Q_R = 0.37$ ,  $df = 1$ ,  $p = .54$ ).

Tabelle 11  
*Moderatoranalysen für Stichprobenmerkmale*

Merkmal	$d_+ \pm CI$	$k$	$Q_b (df)$
Herkunft <sup>a</sup>			6.57 (9)
Hispano-/lateinamerikanisch <sup>b</sup>	0.16 ± 0.03	29	
Mexikanisch	0.22 ± 0.07	25	
Kubanisch	0.14 ± 0.45	1	
Puerto-ricanisch	0.15 ± 0.19	5	
Asiatisch <sup>b</sup>	0.06 ± 0.22	1	
Chinesisch	0.27 ± 0.52	1	
Haitianisch	0.27 ± 0.44	1	
Afrikanisch <sup>b</sup>	0.45 ± 0.40	1	
Türkisch	0.09 ± 0.25	3	
Mehrere verschiedene Migrantengruppen	0.15 ± 0.13	8	
Hochrisikostichprobe <sup>c</sup>			0.28 (1)
Bereits erste Auffälligkeiten	0.19 ± 0.10	13	
Keine bestehenden Auffälligkeiten	0.16 ± 0.03	63	
Mittleres Alter der Kinder/Jugendlichen in Jahren			8.10 (4)
≤ 3	0.14 ± 0.11	4	
3–6	0.16 ± 0.09	8	
6–10	0.18 ± 0.08	20	
10–14	0.14 ± 0.04	29	
14–17	0.24 ± 0.06	15	
Teilstichprobe der Studie			11.07 (1)***
Nur Teilstichprobe der Studie eingeschlossen	0.12 ± 0.04	25	
Gesamte Stichprobe der Studie eingeschlossen	0.21 ± 0.04	51	

*Anmerkungen.*  $d_+ \pm CI$  = gewichtete mittlere Effektstärke nach dem Fixed-Effects-Model mit 95 %-Konfidenzintervall.  $k$  = Anzahl der Vergleiche.  $Q_b$  = Heterogenität zwischen den Gruppen.

<sup>a</sup> In einem Fall existierten keinerlei Angaben bezüglich der Herkunft, weswegen sich die Zahl der Vergleiche nicht auf 76 aufaddiert. <sup>b</sup> Ohne genauere Angabe der Herkunft. <sup>c</sup> Als Hochrisikostichprobe galt eine Stichprobe dann, wenn mindestens 50% bereits vor Interventionsbeginn erste kognitive, Lern- oder Verhaltensauffälligkeiten zeigte.

\*  $p < .05$ ; \*\*  $p < .01$ ; \*\*\*  $p < .001$ .



Es zeigte sich ein signifikanter Unterschied in den Effektstärken in Abhängigkeit davon, ob in der Studie der Fokus auf einer Stichprobe lag, die zum überwiegenden Teil einen Migrationshintergrund besaß oder ob in der Studie eine weniger selektive Stichprobe untersucht wurde (universelle Prävention) und lediglich separate Ergebnisse für eine Teilstichprobe berichtet wurden, die mehrheitlich einen Migrationshintergrund aufwies ( $Q_b = 11.07$ ,  $df = 1$ ,  $p < .001$ ). Jene Vergleiche, welche aus einer Studie stammten, deren Gesamtstichprobe mehrheitlich einen Migrationshintergrund besaß, wiesen hierbei eine größere Effektstärke auf als solche, für die im Rahmen einer größeren Studie separate Ergebnisse berichtet wurden ( $d_+ = 0.21$  vs.  $0.12$ ).

#### 4.7.5 Weiterführende Analysen

Aus den vorangegangenen Analysen ergaben sich Ansatzpunkte für weitere Fragestellungen, weswegen im Folgenden einige weiterführende Analysen dargestellt werden sollen.

Als erstes wurde dabei versucht, den gefundenen Zusammenhang zwischen dem Publikationsjahr und der Effektstärke näher zu explorieren. Dabei sollte überprüft werden, inwieweit dieser durch den Zusammenhang zwischen der Effektstärke und der Art der Kontrollgruppe erklärt werden kann bzw. ob die Verknüpfung zwischen Publikationsjahr und Effektstärke von der Art der Kontrollgruppe abhängig ist. In einer zweifachen gewichteten Regression von der Effektstärke auf das Publikationsjahr und die Art der Kontrollgruppe war die durch die Gesamtregression erklärte Heterogenität zwar signifikant, jedoch erreichte keiner der beiden Regressoren das Signifikanzniveau ( $Q_R = 8.12$ ,  $df = 2$ ,  $p < 0.05$ ;  $\beta = -0.19$ ,  $p = .07$  bzw.  $\beta = 0.19$ ,  $p = .07$ ). Betrachtete man den Zusammenhang zwischen Effektstärke und Publikationsjahr in Abhängigkeit von der Art der Kontrollgruppe, so zeigte sich für passive Kontrollgruppen kein signifikanter Regressionskoeffizient des Publikationsjahres ( $\beta = 0.06$ ;  $Q_R = 0.10$ ,  $df = 1$ ,  $p = .76$ ), während für aktive Kontrollgruppen ein signifikanter negativer Zusammenhang zwischen Publikationsjahr und Effektstärke zu finden war ( $\beta = -0.31$ ;  $Q_R = 5.93$ ,  $df = 1$ ,  $p < .05$ ).

Als nächstes wurde der Zusammenhang zwischen der Art der Randomisierung mit der Effektstärke weitergehend untersucht. Da Studien mit Cluster-Randomisierung häufig an Schulen und die darin getesteten Programme daher mit höherer Wahrscheinlichkeit von Lehrern durchgeführt werden, welche auch niedrigere Effekte erzielten, sollten Lehrer als Administratoren und die Art der Randomisierung gleichzeitig als Regressoren verwendet

werden. Es zeigte sich dabei, dass Lehrer als Administratoren auch in dieser zweifachen Meta-Regression mit signifikant niedrigeren Effekten einhergingen ( $\beta = -0.37, p < .001$ ), die Art der Randomisierung jedoch in dieser gewichteten zweifachen Regression keinen signifikanten Einfluss auf die Effektstärke hatte ( $\beta = -0.16, p = .14; Q_R = 18.66, df = 2, p < .001$ ).

Ebenso fand bei Studien, in denen die eingeschlossene Stichprobe lediglich einen Teil der mehrheitlich nicht aus Personen mit Migrationshintergrund bestehenden Gesamtstichprobe der Studie ausmachten, die Randomisierung häufiger auf Ebene von Clustern statt, weswegen auch diese beiden Variablen gemeinsam in einer zweifachen gewichteten Regression analysiert wurden. Dabei zeigten sich signifikant niedrigere Effektstärken, wenn die eingeschlossene Stichprobe nur eine Teilstichprobe der Studie war ( $\beta = -0.29, p < .05$ ), wohingegen sich die Art der Randomisierung nicht als signifikanter Prädiktor erwies ( $\beta = -0.09, p = .47; Q_R = 11.59, df = 2, p < .01$ ). Eine Übersicht über die durchgeführten zweifachen Regressionsanalysen ist in Tabelle 12 dargestellt.

Tabelle 12  
*Meta-Regressionen für weiterführende Fragestellungen*

Merkmale	$\beta$	$Q_R (df)$	$Q_E (df)$
Regression 1		8.12 (2)*	88.05 (73)
Publikationsjahr	-0.19		
Passive Kontrollgruppe	0.19		
Regression 2		18.66 (2)***	75.50 (70)
Cluster-Randomisierung	-0.16		
Lehrer als Administrator	-0.37***		
Regression 3		11.59 (2)**	84.58 (73)
Cluster-Randomisierung	-0.09		
Nur Teilstichprobe der Studie eingeschlossen	-0.29*		

*Anmerkungen.*  $\beta$  = standardisierter Regressionskoeffizient.  $Q_R$  = durch das Regressionsmodell erklärte Heterogenität.  $Q_E$  = durch das Regressionsmodell nicht erklärte Heterogenität.

\*  $p < .05$ ; \*\*  $p < .01$ ; \*\*\*  $p < .001$ .

Als Letztes sollte untersucht werden, ob der Befund, wonach Studien, in denen die eingeschlossene Stichprobe lediglich einen Teil der Gesamtstichprobe ausmachte, niedrigere Effektstärken aufwiesen, vom Interventionsformat abhängig war. Es zeigte sich, dass ausschließlich bei reinen Gruppenformaten Studien, in denen die eingeschlossene Stichprobe lediglich einen Teil der Gesamtstichprobe ausmachte, signifikant niedrigere Effektstärken aufwiesen als jene, in denen die gesamte Studienstichprobe die Einschlusskriterien erfüllte ( $d_+ = 0.12$  vs.  $0.20$ ;  $Q_b = 7.80$ ,  $df = 1$ ,  $p < .01$ ), wohingegen dieser Effekt nicht bei Einzel- oder gemischten Formaten auftrat ( $d_+ = 0.23$  vs.  $0.27$ ;  $Q_b = 0.14$ ,  $df = 1$ ,  $p = .70$ ).

## 5. DISKUSSION

Ziel dieser Arbeit war es, die bisherigen Befunde zur Wirksamkeit präventiver Interventionen zur sozio-emotionalen Förderung bei Kindern und Jugendlichen mit Migrationshintergrund mit meta-analytischen Methoden zu integrieren. Dabei stellte sich heraus, dass die zu integrierenden Befunde hauptsächlich aus den USA stammten und in Englisch verfasst wurden. Der eingeschlossene Pool von 68 qualitativ besonders hochwertigen und methodisch stringenten Publikationen sowie den darin enthaltenen 93 Vergleichen bot in diesem Rahmen eine angemessene Grundlage, um die eingangs formulierten Fragestellungen in Bezug auf die allgemeine sowie bedingte Effektivität präventiver Interventionen für diesen Bereich zu bearbeiten. In diesem Kapitel soll nun zunächst eine Zusammenfassung der Ergebnisse im Kontext bisheriger Forschung erfolgen, anschließend auf Implikationen für Forschung und Praxis eingegangen und schließlich die Einschränkungen der vorliegenden Arbeit diskutiert werden.

### 5.1 Zusammenfassende Betrachtung der Ergebnisse

Zunächst sollen die Ergebnisse der vorliegenden Arbeit zusammenfassend betrachtet und mit anderen Forschungsarbeiten in Beziehung gesetzt werden. Dabei soll beschrieben werden, welche Antworten sich auf die folgenden zentralen Fragestellungen dieser Meta-Analyse ableiten lassen:

1. Wie hoch ist die allgemeine Wirksamkeit präventiver Interventionen zur sozio-emotionalen Förderung bei Kindern und Jugendlichen mit Migrationshintergrund?
2. Wie stabil sind die erzielten Effekte der präventiven Interventionen über die Zeit?
3. Wie ist die differenzielle Wirksamkeit hinsichtlich verschiedener Erfolgskriterien?
4. Inwiefern unterscheiden sich die in den Studien ermittelten Effekte in Abhängigkeit von publikationsbezogenen Charakteristika?
5. Welchen Zusammenhang gibt es zwischen der Höhe der gemessenen Effekte und verschiedenen methodischen Merkmalen?
6. Welchen Einfluss haben verschiedene Interventionscharakteristika auf die Wirksamkeit der Programme?
7. Inwieweit unterscheidet sich die Wirksamkeit der präventiven Interventionen in Abhängigkeit von verschiedenen Merkmalen der Stichprobensammensetzung?

### 5.1.1 Allgemeine Wirksamkeit der präventiven Interventionen

Über alle Vergleiche hinweg zeigte sich für präventive Interventionen zur sozio-emotionalen Förderung bei Kindern und Jugendlichen mit Migrationshintergrund zur Post-Messung eine gewichtete mittlere Effektstärke von  $d = 0.18$ , welche keine signifikante Heterogenität aufwies. Auch wenn sich diese als signifikant erwies, muss berücksichtigt werden, dass sie damit noch unter der von Cohen (1988) als *klein* eingestuften Effektstärke von  $d = 0.2$  lag. Vergleicht man dies mit Ergebnissen aus anderen Meta-Analysen zur allgemeinen Wirksamkeit präventiver Interventionen, so zeigt sich zudem, dass diese in der Regel geringfügig höhere Effekte aufweisen als die vorliegende Arbeit (siehe z.B. Durlak & Wells, 1997 oder die Übersichten von Beelmann, 2006 und Piquart & Silbereisen, 2014). Ebendies steht zunächst scheinbar im Widerspruch zu dem Befund, dass gezielte Präventionsprogramme eine höhere Wirksamkeit aufweisen als universelle Präventionsmaßnahmen (Beelmann, 2006). Wenn es sich bei einem Migrationshintergrund um einen Risikofaktor handelt, so müsste nach der Argumentation von Offord et al. (1998) die Wirksamkeit präventiver Interventionen für eine von Migration betroffene Stichprobe tendenziell sogar eher höher ausfallen als für eine bevölkerungsrepräsentative Stichprobe, da sich in letztgenannter Gruppe mehr Personen befinden sollten, welche das Präventionsprogramm gar nicht bräuchten bzw. wenig Verbesserungspotential in Bezug auf die Erfolgsmaße besitzen. Die vorliegende Meta-Analyse, welche die Wirksamkeit für Kinder und Jugendliche mit Migrationshintergrund untersuchte, hätte demnach eher höhere als niedrigere Effekte aufweisen müssen als Meta-Analysen, in denen der Fokus nicht auf dieser Risikogruppe lag.

Was dabei jedoch beachtet werden muss, ist, dass die Vergleichbarkeit der vorliegenden Arbeit mit vorangegangenen Meta-Analysen in Frage gestellt werden kann. Die Selektionskriterien der hier durchgeführten Meta-Analyse waren vergleichsweise restriktiv. So wurden beispielsweise nur Studien eingeschlossen, in welchen die Zuweisung der Teilnehmer zur Interventions- bzw. Kontrollgruppe randomisiert erfolgte. Es gibt einige Hinweise dafür, dass randomisiert-kontrollierte Studien eine geringere Heterogenität aufweisen und zum anderen auch tendenziell mit niedrigeren Effekten einhergehen (Baytop, 2006; Beelmann, 2006). Die niedrigere Effektstärke in der vorliegenden Meta-Analyse könnte folglich auf diesen Unterschied zurückführbar sein.

Weiterhin ist zu beachten, dass in dieser Arbeit lediglich Studien eingeschlossen wurden, welche eine Stichprobengröße von mindestens 25 Personen pro Gruppe

aufwiesen. Die Stichprobengröße ist jedoch wiederum typischerweise negativ mit der Effektstärke assoziiert, was in der Regel zumindest teilweise auf Publikationsverzerrungen zurückgeführt wird (Beelmann, 2006; Dickersin, 1990; Gilbody, et al., 2000; Lösel & Beelmann, 2003). Auch hierin könnte demnach eine Ursache dafür liegen, dass sich in dieser Meta-Analyse vergleichsweise niedrige Effektstärken fanden, zudem in früheren Meta-Analysen im Gegensatz zur vorliegenden Arbeit häufig das Random-Effects-Model verwendet wurde, bei welchem der Stichprobenfehler relativ betrachtet weniger stark in die Berechnung des Gewichtungsfaktors eingeht und somit auch kleinere Studien stärker in die Gesamteffektstärke einfließen als im Fixed-Effects-Model.

Des Weiteren ist die vorliegende Meta-Analyse aktueller als die zum Vergleich herangezogenen Arbeiten und schloss somit auch Studien ein, welche ihrerseits wiederum jüngeren Ursprungs sind. Es wurde bereits darauf hingewiesen, dass nicht nur ein negativer Zusammenhang zwischen dem Publikationsjahr einer Primärstudie und der Wirksamkeit präventiver Interventionen existiert, sondern auch Meta-Analysen älteren Ursprungs größere Effekte aufzeigen (Beelmann, 2006). Dementsprechend könnte auch in dem Publikationsjahr ein Grund dafür zu finden sein, dass die in der vorliegenden Meta-Analyse ermittelte Effektstärke niedriger ausfiel als in vorangegangenen integrativen Arbeiten zur Wirksamkeit präventiver Interventionen.

Sollten diese methodischen Unterschiede nicht allein für das erwartungskonträre Ergebnis, dass sich in der vorliegende Arbeit im Vergleich mit Meta-Analysen, die nicht auf die Risikogruppe von Kindern und Jugendlichen mit Migrationshintergrund abzielten, eher kleinere als größere Effekte zeigten, verantwortlich sein, gibt es zudem mögliche inhaltliche Gründe, die zumindest erklären könnten, warum die gefundenen Effektstärken trotz Fokus auf eine selektive Stichprobe nicht größer waren.

Zum einen kann argumentiert werden, dass ein Migrationshintergrund nicht unter allen Umständen und bezüglich jedes sozio-emotionalen Entwicklungsparameters bzw. jeder Art von Verhaltensproblem einen Risikofaktor darstellen muss. So legen Berry und Sam (1997) dar, dass der Prozess der Akkulturation nicht nur je nach Einwanderungsland und Migrantengruppe unterschiedlich sein kann, sondern es auch individuelle Faktoren gibt, die das Maß an Akkulturationsstress sowie den Erfolg der psychologischen und soziokulturellen Anpassung mitbestimmen. Dementsprechend fanden auch Sam et al. (2006) in einer multinationalen Studie, dass Jugendliche mit Migrationshintergrund verglichen mit einheimischen Jugendlichen aus denselben Nachbarschaften und Schulen

eine genauso gute psychologische Anpassung und zum Teil sogar eine bessere sozio-kulturelle Anpassung aufweisen können. Belhadj Kouider et al. (2014) schlussfolgerten in ihrem systematischen Review, dass Kinder und Jugendliche mit Migrationshintergrund zwar eine höhere Prävalenz für internalisierende Probleme aufweisen, es in Bezug auf externalisierende Probleme aber keine konsistente Befundlage gäbe.

Auf der anderen Seite stellt sich, auch wenn man eine Risikobelastung von Kindern und Jugendlichen mit Migrationshintergrund als gegeben voraussetzt, die Frage, ob sich dadurch die zu erwartende Wirksamkeit von Präventionsprogrammen tatsächlich erhöht. Zwar weisen gezielte Präventionsmaßnahmen im Allgemeinen höhere Effekte auf (Beelmann, 2006), jedoch gibt es Hinweise darauf, dass dieser Unterschied insbesondere auf besonders hohe Effekte für indizierte Präventionsprogramme zurückführbar ist (Lösel & Beelmann, 2003). Auch konnten Lundahl et al. (2006) zeigen, dass Elterntrainings für Familien mit niedrigem sozioökonomischem Status sogar eine geringere Wirksamkeit aufweisen als für weniger benachteiligte Familien. Da ein niedriger sozioökonomischer Status ebenfalls einen Risikofaktor darstellt (Bradley & Corwyn, 2002), der zudem häufiger bei Personen mit Migrationshintergrund auftritt (Grabka & Frick, 2010; U.S. Census Bureau, 2011, Table 713), spricht dies dafür, dass die Wirkung präventiver Interventionen für Risikogruppen nicht in jedem Fall größer ausfallen muss. Als Ursache für diesen Effekt diskutieren Lundahl et al. (2006) die Möglichkeit, dass die schwierigeren Familienverhältnisse die Implementation von Präventionsprogrammen erschweren. Aufgrund der vielfältigen migrationsbedingten Belastungen, die auf das Familiensystem einwirken können, könnte ein ähnlicher Mechanismus auch bei Personen mit Migrationshintergrund wirksam sein.

Eine abschließende Bewertung wodurch die vergleichsweise niedrige Effektstärke in der vorliegenden Meta-Analyse bedingt ist, kann anhand der gegebenen Befunde jedoch nicht getroffen werden. Allerdings kann aufgrund der angemessenen Anzahl und hohen methodischen Qualität der hier eingeschlossenen Studien, die sich auch in der besonders geringen Heterogenität der Befunde widerspiegelt, davon ausgegangen werden, dass die ermittelte Effektstärke einen guten Schätzer für die Wirksamkeit präventiver Interventionen zur sozio-emotionalen Förderung bei Kindern und Jugendlichen mit Migrationshintergrund liefert und von hoher interner Validität zeugt.

### 5.1.2 Stabilität der Effekte über die Zeit

Wie in der Präventionsforschung häufig angemerkt, gab es auch in dieser Meta-Analyse nur wenige Studien, die eine Schätzung der Stabilität der Effekte über die Zeit ermöglichten (Beelmann, 2006). Die separaten Analysen über die drei festgelegten Zeiträume sprechen zwar dafür, dass es möglich ist sowohl kurz als auch mittel und langfristig kleine, aber bedeutsame Effekte zu erzielen, jedoch zeigten die Analysen innerhalb von Gruppenvergleichen, für die über mehrere Messzeiträume hinweg gemessen wurde, auch, dass der Effekt über die Zeit hinweg abnahm, sodass bei diesen Vergleichen schließlich auch keine längerfristigen Effekte mehr nachweisbar waren. Generelle Aussagen über die längerfristige Wirksamkeit präventiver Interventionen zur sozio-emotionalen Förderung bei Kindern und Jugendlichen mit Migrationshintergrund sind daher kaum möglich. Zudem erschienen aufgrund der geringen Zahl an Vergleichen, in denen die Stabilität abgebildet wurde, Moderatoranalysen zu Faktoren, welche die längerfristigen Effekte beeinflussen, wenig sinnvoll.

### 5.1.3 Differenzielle Wirksamkeit hinsichtlich verschiedener Erfolgskriterien

Insgesamt können präventive Interventionen zur sozio-emotionalen Förderung bei Kindern und Jugendlichen mit Migrationshintergrund für eine breite Palette von Erfolgskriterien als wirksam erachtet werden. Dabei kann festgestellt werden, dass die untersuchten präventiven Interventionen neben ihrer Wirksamkeit in Bereichen der sozialen und emotionalen Entwicklung auch andere Erfolgskriterien positiv beeinflussten. Insbesondere sind dabei der familiäre Kontext zu nennen sowie die kognitive Entwicklung und bildungsbezogene Kriterien, für welche die Wirksamkeit sogar noch höher ausfiel als für soziale und emotionale Parameter als primäre Maße. Dabei ist zu beachten, dass in der vorliegenden Meta-Analyse Programme, die neben der sozio-emotionalen Förderung noch andere Ziele verfolgten, nicht ausgeschlossen wurden, weswegen diese Erfolgskriterien aus Sicht der jeweiligen Studien nicht ohne weiteres als *sekundär* bezeichnet werden können. Innerhalb des Bereiches sozialer und emotionaler Parameter waren die untersuchten Programme am wirksamsten in Bezug auf internalisierende Probleme, wogegen sich hinsichtlich des Suchtmittelkonsums sowie externalisierender Probleme die kleinsten Effekte zeigten. Dies korrespondiert mit dem Befund, dass Meta-Analysen zur Prävention von Depressionen höhere Wirksamkeitsparameter erzielen als Meta-Analysen zur Prävention von Verhaltensstörungen bzw. Delinquenz und



insbesondere zur Prävention von Drogenmissbrauch (Röhrle, 2008). Möglicherweise lassen sich auch bei Kindern und Jugendlichen mit Migrationshintergrund Probleme, welche nach innen gerichtet sind, trotz ihrer unter Umständen geringeren Auffälligkeit für das soziale Umfeld, leichter korrigieren bzw. vorbeugen als Verhaltensprobleme, die sich nach außen hin manifestieren, zumal letztere häufig in einen sozialen Kontext eingeflochten sind und allgemein eine höhere Stabilität aufweisen (Fergusson, Swain-Campbell & Horwood, 2002; Laucht, Esser & Schmidt, 2000). Insgesamt waren die Unterschiede zwischen den verschiedenen Erfolgsmaßen im Hinblick auf die Wirksamkeit jedoch relativ gering.

#### *5.1.4 Publikationsbezogene Merkmale*

Eine Publikationsverzerrung wird in Meta-Analysen häufig berichtet (Beelmann, 2006). In der vorliegenden Arbeit ergab sich hierauf jedoch kein Hinweis. Weder gab es einen signifikanten Zusammenhang zwischen Stichprobengröße und Effektstärke noch zeigten Dissertationen eine geringere Effektivität und auch das ermittelte Fail-Safe N überschreitet die von Rosenthal (1979) als sicher angesehene Grenze um ein Vielfaches. Grund hierfür könnten wiederum die verwendeten Selektionskriterien sein. Diese begrenzten die Varianz der Stichprobengröße durch den Ausschluss von Studien, in denen entweder die Interventions- oder die Kontrollgruppe weniger als 25 Teilnehmer enthielten, was natürlich Auswirkungen auf die sich aus Begg's Rangkorrelationsmethode (Begg & Mazumdar, 1994) ergebenden Auswertungen hat. Jedoch spricht der Befund, dass unter den eingeschlossenen Studien die Dissertationen keine geringere Effektstärke aufwiesen als in Monographien oder Zeitschriften publizierte Forschungsberichte, dafür, dass durch die Selektionskriterien tatsächlich der Einfluss von Publikationsverzerrungen auf die forschungssynthetischen Befunde wirkungsvoll reduziert, wenn nicht gänzlich vermieden werden konnte. Dies gilt umso mehr, als dass sich in der vorliegenden Arbeit zum einen eine genügend große Zahl an eingeschlossenen Dissertationen befand, um Analysen mit angemessener Teststärke durchzuführen und zum anderen die Ergebnisse auch deskriptiv tendenziell sogar auf höhere Effekte für Dissertationen hindeuteten. Dies unterstreicht wiederum die Validität der ermittelten Effektstärken.

Der gefundene negative Zusammenhang zwischen Publikationsjahr und Effektstärke stützt die Vermutung, dass die vergleichsweise niedrigen Effekte, welche in der vorliegenden Arbeit gefunden wurden, zumindest teilweise auch auf die Aktualität der

Meta-Analyse zurückzuführen sind. Auch ist erwähnenswert, dass sich insgesamt nur wenige ältere Studien in der Meta-Analyse fanden. Nur 4.4 % der Arbeiten wurden vor 1980 veröffentlicht, was im Vergleich zu anderen Meta-Analysen ein verschwindend geringer Teil ist (z.B. wurden bei Lösen & Beelmann, 2003, 22.6 % der Studien bis 1980 veröffentlicht). Dies mag teilweise auf die restriktiven Selektionskriterien zurückzuführen sein, weißt aber möglicherweise auch auf die Aktualität der Migrationsforschung im wissenschaftlichen Diskurs bzw. die erst spät einsetzende Entwicklung dieses Forschungsfeldes hin.

Der Befund, wonach die Effektstärke nicht signifikant mit dem Land, in dem die Studie durchgeführt wurde, zusammenhängt, könnte als Hinweis auf eine Generalisierbarkeit der Befunde über verschiedene Einwanderungsländer hinweg interpretiert werden, wären nicht Studien zu diesem Thema außerhalb der USA nur sehr spärlich vorhanden. Aufgrund der Tatsache jedoch, dass über 90 % der eingeschlossenen Forschungsberichte ihre Daten aus den USA bezogen und nur insgesamt 11 der Vergleiche sich auf Stichproben aus fünf verschiedenen anderen Ländern bezogen, kann dies nicht abschließend beurteilt werden. Da allerdings andere multinationale Meta-Analysen zu einer ähnlich schiefen Verteilung hinsichtlich des Erhebungslandes kommen (Lösel & Beelmann, 2003; Yuen, 2004), kann nicht davon ausgegangen werden, dass dieser Befund allein auf die methodischen oder stichprobenbezogenen Eigenschaften dieser Arbeit zurückzuführen ist. Es handelt sich eher um eine typische Konzentrierung der Präventionsforschung im angloamerikanischen Raum, die jedoch fundierte Aussagen über die Generalisierbarkeit von Befunden erschwert.

#### *5.1.5 Methodische Merkmale*

Trotz der vergleichsweise restriktiven methodischen Selektionskriterien der vorliegenden Arbeit konnten bedeutsame Effektstärkenunterschiede in Abhängigkeit von methodischen Variablen beobachtet werden.

Gruppenvergleiche, bei denen die Kontrollgruppe keinerlei Behandlung erfuhr, wiesen insgesamt größere Effekte auf. Dies ist vermutlich damit zu begründen, dass die minimalen oder unspezifischen Interventionen, denen die aktiven Kontrollgruppen ausgesetzt waren, die Erfolgskriterien, wenn auch weniger deutlich als die eigentlich zu testenden präventiven Interventionen, positiv beeinflussten. Die höhere Heterogenität aktiver Kontrollgruppen könnte dann auch auf die unterschiedliche Wirksamkeit eben

dieser minimalen oder unspezifischen Interventionen zurückzuführen sein. Beelmann (2006) sieht in dem zunehmenden Mangel an unbehandelten Kontrollgruppen in der neueren Forschung einen möglichen Grund für den negativen Zusammenhang zwischen Publikationsjahr und Effektstärke. In der vorliegenden Meta-Analyse fanden sich hierfür Hinweise. Zwar sagte in einer zweifachen gewichteten Regression, vermutlich aufgrund der Kollinearität, neben dem Publikationsjahr auch die Art der Kontrollgruppe nicht mehr signifikant die Effektstärke vorher, jedoch war das Publikationsjahr ausschließlich für Vergleiche mit aktiver Kontrollgruppe ein bedeutsamer Prädiktor der Wirksamkeit. Für Gruppenvergleiche mit passiver Kontrollgruppe hatte das Publikationsjahr hingegen keinerlei Einfluss auf die Effektstärke. Bezüglich des Zusammenhanges für aktive Kontrollgruppen stellt sich dabei auch die Frage, ob dieser dadurch bedingt ist, dass die in jüngerer Zeit durchgeführten unspezifischen oder minimalen Interventionen wirksamer sind als die, welche in früheren Jahren Anwendung fanden. Dies ist mit den vorliegenden Daten jedoch nicht abschließend zu klären.

Ähnlich wie in der Meta-Analyse von Lösel und Beelmann (2003) wiesen auch in der vorliegenden Arbeit Gruppenvergleiche, bei denen die Teilnehmer individuell den Untersuchungsbedingungen randomisiert zugewiesen wurden, höhere Effekte auf als Vergleiche, bei denen die Randomisierung auf Ebene von Clustern erfolgte. Zwar beruhen Vergleiche, bei denen ganze Schulen oder Klassen randomisiert zugewiesen werden, in der Regel auch auf größeren Stichproben, insbesondere wenn dabei, wie in den angewendeten Selektionskriterien festgelegt, mindestens 10 solcher Cluster vorhanden sein müssen, jedoch hing die Stichprobengröße, wie sich in Begg's Rangkorrelationsmethode zeigte, in dieser Arbeit nicht signifikant mit der Effektstärke zusammen. Daher ist auch in einer größeren Stichprobe keine Ursache für den gefundenen Zusammenhang zwischen der Art der Randomisierung und der Effektstärke zu suchen. Die größere Heterogenität von Vergleichen, in denen ganze Cluster randomisiert zugewiesen wurden, zeigt, dass auch bei solchen Untersuchungen höhere Effekte erzielt werden können. Jedoch spricht diese größere Streuung der Ergebnisse dafür, dass eine solche Randomisierung nicht unbedingt auch auf Individualebene zu vergleichbaren Gruppen führen muss, wodurch die interne Validität im Einzelfall reduziert sein kann.

Dass die Vergleichbarkeit von Interventions- und Kontrollgruppe Auswirkungen auf die ermittelten Effektstärken haben kann, wurde bereits in früheren Meta-Analysen gezeigt (Beelmann et al., 2014). Auch in der vorliegenden Arbeit wiesen Vergleiche, bei

denen die beiden Gruppen vergleichbar waren, kleinere Effekte auf als Vergleiche, bei denen eine solche Äquivalenz vor der Intervention nicht gegeben war. Dies ging einher mit einer größeren Heterogenität bei Nichtvergleichbarkeit der Gruppen. Während die höhere Heterogenität vermutlich auf die unterschiedliche Art bzw. Richtung, in der sich die Gruppen jeweils unterschieden, zurückzuführen war, deutet der Unterschied in der gewichteten mittleren Effektstärke darauf hin, dass die Unterschiede vor der Intervention tendenziell häufiger zu Lasten der Kontrollgruppe gingen. Es ist jedoch auch hervorzuheben, dass sich für die Kontraste, in denen die Vergleichbarkeit der Gruppen sichergestellt war, mit  $d_+ = 0.18$  eine Effektstärke ergab, die sich nicht wesentlich von der ermittelten Gesamteffektstärke für alle Vergleiche unterschied. Folglich kann nicht davon ausgegangen werden, dass die ermittelte Gesamteffektstärke aufgrund methodischer Mängel einiger Studien bedeutsam nach oben hin verzerrt wurde. Kontraste, bei denen die Vergleichbarkeit vor der Intervention nicht getestet wurde, wiesen die kleinsten Effekte auf. Jedoch ist unklar, worauf dieser Unterschied zurückzuführen ist, da das Testen der Vergleichbarkeit der Gruppen wesentlich durch Entscheidungen beeinflusst ist, welche die Autoren während der Auswertung und des Verfassens ihrer Studie treffen.

Die mittlere Dropout-Rate hatte zwar keinen Einfluss auf die gewichtete mittlere Effektstärke, jedoch trug sie zur Heterogenität der Vergleichseffektstärken bei. Dies spricht dafür, dass sie zumindest in dieser Meta-Analyse keine systematische Verzerrung der Ergebnisse in eine bestimmte Richtung erzeugte. Auf Ebene der einzelnen Gruppenvergleiche beeinträchtigte sie jedoch vermutlich die interne Validität, wodurch sich aus Vergleichen mit hoher Dropout-Rate ungenauere Effektschätzungen ergeben können. Diese fallen, je nachdem welche Teilnehmer aus welcher Gruppe die Studie verlassen, entsprechend höher oder niedriger aus.

Die für einen Vergleich verwendete Analyseverfahren hing sowohl mit der gewichteten mittleren Effektstärke als auch mit der Heterogenität der Ergebnisse zusammen. Die konservativste Schätzung über Intention-to-treat-Analysen ergab jedoch auch die heterogensten Ergebnisse. Da diese Analyseverfahren auch Teilnehmer einschließt, die überhaupt nicht oder nur an wenigen Sitzungen der Intervention teilnahmen und die folglich auch weniger von der Intervention profitieren konnten, scheint diese niedrige Effektstärke sehr plausibel. Die höhere Heterogenität kann zum einen darauf zurückzuführen sein, dass der Anteil der Teilnehmer, welche durch mangelnde Teilnahme wenig von der Intervention profitierten, von Vergleich zu Vergleich sehr unterschiedlich sein

kann und zum anderen darauf, dass fehlende Daten in Bezug auf das Erfolgskriterium mittels Imputationsverfahren geschätzt wurden und diese Schätzungen nicht so exakt sein können wie eine tatsächliche Messung. Die höchste mittlere Effektivität wurde ermittelt, wenn in den Analysen explizit Teilnehmer der Interventionsgruppe ausgeschlossen wurden, die eine für den Vergleich definierte Mindestanzahl von Sitzungen nicht besuchten. Mit dieser Methode wurden jene Personen aus der Interventionsgruppe ausgeschlossen, welche durch mangelnde Teilnahme weniger von der Präventionsmaßnahme profitieren konnten und für die entsprechend geringere Verbesserungen bzw. schlechtere Ergebnisse im Hinblick auf die Erfolgskriterien zu erwarten waren. Bezüglich der Streuung der Effektstärken führten jedoch jene Analysen zu den homogensten Ergebnissen, welche zwar nur jene Teilnehmer in die Auswertungen einbezogen, die an der Post-Test-Messung teilnahmen, allerdings niemanden aufgrund der Teilnahmerate explizit ausschloss. Dieses Vorgehen vermeidet zum einen die mangelnde Genauigkeit von Imputationsverfahren, zum anderen aber auch die stets subjektive Festlegung von Mindestteilnahmeraten, die ebenfalls zu Heterogenität beitragen könnte. Dass sich für diese Vergleiche eine dem Durchschnitt entsprechende mittlere Effektstärke zeigte, ist vermutlich darauf zurückzuführen, dass zwar nicht direkt anhand der Teilnahmerate selektiert wurde, jedoch jene Personen, die zur Post-Test-Messung nicht erschienen sind mit höherer Wahrscheinlichkeit auch an der präventiven Intervention seltener teilnahmen.

#### *5.1.6 Merkmale der präventiven Interventionen*

Die in den Studien ermittelte Wirksamkeit der präventiven Interventionen hing auch wesentlich mit Merkmalen der getesteten Präventionsprogramme zusammen. Insbesondere waren hierbei Inhalte, Ziele, Implementation, Intensität und die Administratoren einer präventiven Intervention relevant, aber auch ob das Programm Einzelsitzungen enthielt.

Hinsichtlich der Inhalte der Programme erwiesen sich jene Programme als am wenigsten wirksam, die Substanzmissbrauch thematisierten. Dies korrespondiert mit Befunden aus anderen Meta-Analysen und Übersichtsarbeiten, in denen für Sucht- bzw. Drogenprävention ebenfalls vergleichsweise niedrige Effekte beschrieben wurden (Röhrle, 2008; Tobler et al., 2000) und hängt vermutlich auch mit den allgemein niedrigeren Effekten für Suchtmittelkonsum zusammen.

Kombinierte Programme mit mehreren unterschiedlichen Inhalten waren allgemein wirksamer als Programme, die sich nur auf einen Inhaltsbereich fokussierten. Vermutlich sind gerade aufgrund der multiplen Risiken, die mit einem Migrationshintergrund einhergehen, solche umfangreicheren Programme für diese Kinder und Jugendlichen erfolgversprechender, zumal sich diese unterschiedlichen Probleme, denen präventiv entgegengewirkt werden soll, auch gegenseitig beeinflussen (Belhadj Kouider et al., 2014; Manongdo & Ramírez García, 2007; Reitz et al., 2005; Stevens et al., 2007). In diesem Zusammenhang ist auch der Befund zusehen, dass Präventionsprogramme, bei denen die Eltern der Kinder und Jugendlichen mit Migrationshintergrund einbezogen wurden, wirksamer waren. Die Bedeutung des Familiensystems für die soziale und emotionale Entwicklung ist gerade bei Kindern und Jugendlichen mit Migrationshintergrund kaum zu überschätzen (Belhadj Kouider et al., 2014; Birman & Taylor-Ritzler, 2007; Choi et al., 2008).

Bezüglich sozio-emotionaler Ziele der präventiven Interventionen erwiesen sich Programme, welche ausschließlich auf die Prävention externalisierender Probleme abzielten, als weniger wirksam, was wiederum mit den Befunden von Röhrle (2008) korrespondiert. Möglicherweise ist dies auch darauf zurückzuführen, dass die Frage, ob Kinder und Jugendliche mit Migrationshintergrund ein höheres Risiko für externalisierende Probleme aufweisen, nicht eindeutig geklärt ist (Belhadj Kouider et al., 2014). Allerdings zeigte sich in der vorliegenden Arbeit auch, dass Programme, die neben diesem Ziel auch auf die Prävention internalisierender Probleme abzielten, am wirksamsten waren und dass sich diese höhere Wirksamkeit auch hinsichtlich externalisierender Probleme zeigte. Womöglich können externalisierende Probleme bei Kindern und Jugendlichen mit Migrationshintergrund auch dadurch wirksam verhindert oder reduziert werden, dass internalisierende Probleme abgebaut werden, zumal letztere bei dieser Risikogruppe überdurchschnittlich häufig vorkommen und sich beide Problemformen gegenseitig verstärken (Belhadj Kouider et al., 2014; Stevens et al., 2007).

Auch die Implementation hatte einen Einfluss auf die Wirksamkeit der präventiven Interventionen, allerdings nicht für alle betrachteten Variablen in konsistenter Weise. Erschwert ist die Interpretation der Ergebnisse insbesondere dadurch, dass die Kodierung besonders für Implementationsvariablen verstärkt von der Art der Beschreibung durch die Autoren bzw. von der deskriptiven Validität der Studie abhängig ist. Das Bild der Implementation als ein empirisch erst schwach beforschter, aber dennoch möglicherweise

sehr bedeutsamer Wirkfaktor (Beelmann, 2006; Beelmann & Karing, 2014) bestätigte sich auch in der vorliegenden Arbeit. So gab es in vielen der eingeschlossenen Forschungsberichte keinen Hinweis auf verschiedene Implementationsfaktoren. Als bedeutsam erwies sich dennoch die Supervision der Administratoren. Interventionen, bei denen häufig eine Supervision durchgeführt wurde, waren dabei deutlich wirksamer als solche, bei denen nur geringfügig supervidiert wurde. Dies deutet darauf hin, dass intensive Supervision zur Qualitätssicherung einer Präventionsmaßnahme beitragen kann. Unklar ist jedoch, welche Bedeutung eine weniger umfangreiche Supervision für die Wirksamkeit einer präventiven Intervention hat. Für diese ergab sich kein signifikanter Unterschied zu Interventionen, für welche die Autoren keine Angaben bezüglich einer Supervision machten. Diese letztgenannte Gruppe von Vergleichen, für die keine Supervision Erwähnung fand, enthielt jedoch auch Interventionen, die nur eine einzige Sitzung umfassten. Es kann auch nicht ausgeschlossen werden, dass Autoren eine stattfindende Supervision unerwähnt ließen. Ähnlich unklar ist, weshalb ein Training der Administratoren keinen Einfluss auf die Wirksamkeit einer präventiven Intervention hatte. Allgemein ist hinsichtlich der Durchführung implementationsfördernder Maßnahmen und deren Erwähnung in Studien davon auszugehen, dass diese auch mit anderen relevanten Bedingungen, wie der Zahl der Sitzungen, dem Setting, der Manualisierung sowie den Administratoren einer Präventionsmaßnahme zusammenhängen. Deutlicher war das Ergebnis im Hinblick auf teilnehmerorientierte Maßnahmen zur Implementationsförderung, wie beispielsweise monetäre Anreize zur Teilnahme an den Sitzungen der Intervention, Fahrdienste zum Erreichen des Interventionsortes oder Kinderbetreuung bei Elterntrainings. Je mehr solche zusätzlichen Angebote gemacht wurden, desto wirksamer war die präventive Intervention. Dies betont die Bedeutung solcher Implementationsfaktoren auch im Vergleich zum Training und zur Supervision der Administratoren. Der Einfluss von Implementationsproblemen auf die Wirksamkeit der präventiven Interventionen fiel in der vorliegenden Arbeit weniger eindeutig aus, als man ihn erwarten sollte. Auch hier gilt jedoch, dass in Studien, welche keine Implementationsprobleme berichteten, nicht mit Sicherheit gesagt werden kann, ob es auch tatsächlich keine gab. Dass präventive Interventionen mit vielen Implementationsproblemen weniger wirksam waren als solche, für die nur geringfügige Probleme berichtet wurden, entspricht hingegen der Erwartung, dass eine Präventionsmaßnahme seine Wirkung umso besser entfalten kann je besser sie umgesetzt wird.

Auch zwischen der Interventionsintensität und der Wirksamkeit der Präventionsmaßnahmen zeigte sich ein Zusammenhang. Dieser war jedoch nicht linear. Während die Wirksamkeit ansonsten tendenziell mit der Gesamtinterventionsdauer zunahm, erwiesen sich die kürzesten Programme mit bis zu fünf-stündiger Dauer sogar am wirksamsten. Möglicherweise sind besonders kurze Interventionen gut dazu geeignet, Anstöße für weitergehende Veränderungsprozesse zu geben, während eine direkte Bearbeitung und Veränderung von Problemen innerhalb von Interventionssitzungen durchaus sehr viel Zeit erfordert. Ein solcher kurvilinearere Zusammenhang könnte auch für die divergierenden Befunde aus anderen Meta-Analysen eine Erklärung bieten (vergleiche Stice et al., 2009, Tobler et al., 2000, und S. J. Wilson, Lipsey & Derzon, 2003). Der von Stice et al. (2009) gefundene negative Zusammenhang zwischen Programmintensität und Effektstärke, könnte demnach damit zusammenhängen, dass die mittlere Dauer der Präventionsmaßnahmen in ihrer Meta-Analyse mit 12.84 Stunden vergleichsweise gering war.

Weiterhin zeigte sich, dass Programme, in denen ausschließlich mit Gruppen gearbeitet wurde, weniger wirksam waren als Präventionsmaßnahmen, die auch Einzelsitzungen beinhalteten. Dies korrespondiert mit dem Befund von Lundahl et al. (2006), welche einen ähnlichen Effekt insbesondere bei der Risikogruppe der Familien mit niedrigem sozioökonomischem Status zeigen konnten, zumal es sich bei einem Migrationshintergrund ebenfalls um einen Risikofaktor handelt, welcher unter anderem auch mit einem niedrigeren sozioökonomischem Status einhergeht (Grabka & Frick, 2010; U.S. Census Bureau, 2011, Table 713). Möglicherweise ist gerade bei soziodemografischen Risikogruppen ein sehr individuelles Vorgehen erforderlich, um auf die verschiedenen Problemlagen und Besonderheiten der Kinder bzw. Jugendlichen eingehen zu können und eine gute Implementation zu gewährleisten.

Außerdem waren die Administratoren einer Intervention für deren Wirksamkeit von Bedeutung. Insbesondere konnten hier Lehrpersonen nur geringe Effekte erzielen. Dies entspricht zumindest teilweise den Ergebnissen von Teubert und Piquart (2011), wonach Lehrer als Administratoren weniger effektiv sind als Personen, die beruflich im Bereich der psychischen Gesundheit arbeiten. Fraglich ist hierbei, ob Lehrer Interventionen schlechter umsetzen, die typischerweise auf vorwiegend kognitivem Wissensaustausch basierende Lehrer-Schüler-Beziehung eine wirksame sozio-emotionale Förderung erschwert oder ob Schüler ihrem Lehrer weniger Aufmerksamkeit schenken als einer ihnen unbekanntem Person. Der Zusammenhang erwies sich darüber hinaus auch bei



Kontrolle für die Art der Randomisierung als relevant. Da sich diese wiederum als nicht mehr bedeutsam für die Vorhersage der Effektstärke erwies, wenn die Lehrer als Administratoren als Moderatorvariable in die Regression eingeschlossen wurden, ist in dieser Konfundierung auch eine mögliche Erklärung dafür zu finden, dass cluster-randomisierte Studien niedrigere Effekte zeigten.

Die kulturelle Angepasstheit spielte hingegen keine Rolle für die Wirksamkeit der präventiven Interventionen. Zwar bietet eine kulturelle Anpassung von Programmen die Möglichkeit spezifische Risiko- und Schutzfaktoren sowie Werte und Einstellungen anzusprechen, die für die trainierte Migrantengruppe von besonderer Relevanz sind, jedoch ist unklar wie sehr sich diese von denen Einheimischer tatsächlich unterscheidet. Zudem ist zu bedenken, dass auf diese spezifischen Risiko- und Schutzfaktoren sowie Werte und Einstellungen innerhalb von kulturell nicht angepassten Programmen zumindest in Einzelsitzungen oder bei überschaubarer Gruppengröße auch individuell eingegangen werden kann. Dies scheint ohnehin wichtig angesichts individueller Faktoren, welche den Akkulturationsprozess beeinflussen und der damit einhergehenden Heterogenität auch innerhalb von Migrantengruppen (Berry, 1997). Des Weiteren ist es möglich, dass bereits etablierte Programme, die mehrfach getestet und optimiert wurden, durch ihre bessere empirische Fundierung effektiver sind als viele neu entwickelte kulturell zugeschnittenen Programme, deren Entwicklung noch nicht so weit fortgeschritten ist. All dies relativiert vermutlich die möglichen Vorteile, die sich aus kulturell angepassten Programmen ergeben könnten.

#### *5.1.7 Merkmale der Stichprobe*

Stichprobenmerkmale hatten in der vorliegenden Arbeit bis auf eine Ausnahme kaum Einfluss auf die Wirksamkeit der präventiven Interventionen. Insbesondere ergaben sich keine Unterschiede in Abhängigkeit von der untersuchten Migrantengruppe. Ähnlich wie die Ergebnisse hinsichtlich des Erhebungslandes, kann dies jedoch nicht unbedingt als Hinweis auf eine Generalisierbarkeit der gewichteten mittleren Effektstärke für alle Migrantengruppen gewertet werden. Nahezu dreiviertel der Studien untersuchten entweder Hispanics ohne Konkretisierung bezüglich des Herkunftslandes oder mexikanischstämmige Teilnehmer, wobei ein Vergleich dieser beiden Gruppen nicht sinnvoll war, da sich auch in der Gruppe der Hispanics jeweils ein mehr oder weniger großer Anteil an Teilnehmern mexikanischer Herkunft befunden haben kann. Um eine

Generalisierbarkeit der Ergebnisse wirklich statistisch absichern zu können, hätte es einer ausreichend großen Zahl an Studien bedurft, die spezifisch eine Migrantengruppe untersuchten, welche nicht unter die Oberkategorie der Hispanics subsumiert werden kann. Diese lag jedoch nicht vor, weswegen Wirksamkeitsunterschiede in Abhängigkeit von der Herkunft der Kinder und Jugendlichen mit Migrationshintergrund kaum sinnvoll zu untersuchen waren.

Auch machte es in Bezug auf die Wirksamkeit der Präventionsmaßnahmen keinen Unterschied, ob die Stichprobe bereits erste kognitive, Lern- oder Verhaltensauffälligkeiten zeigte oder nicht. Da es sich bei Kindern und Jugendlichen mit Migrationshintergrund bereits um eine Risikogruppe handelt ist es denkbar, dass die kumulierten Probleme in solchen Hochrisikogruppen eher die Implementation erschweren. Auch könnte der Fokus auf eine bestimmte bereits vorhandene Auffälligkeit dazu führen, dass andere Risikofaktoren, welche mit einem Migrationshintergrund einhergehen und ebenfalls einen Einfluss auf die Erfolgsmaße haben, vernachlässigt werden.

Ebenfalls hatte das Alter keinen Einfluss auf die Wirksamkeit präventiver Interventionen. Es gab gleichermaßen wirksame Programme für Kleinkinder, Schulkinder und Jugendliche mit Migrationshintergrund. Den in anderen Meta-Analysen gefundenen positiven Zusammenhang zwischen dem Alter und der Wirksamkeit präventiver Interventionen führten Beelmann et al. (2014) auf eine Konfundierung mit der Präventionsart (universell vs. gezielt) zurück. Da jedoch in der vorliegenden Meta-Analyse alle eingeschlossenen Stichproben mehrheitlich einen Migrationshintergrund aufweisen mussten und indizierte Präventionsmaßnahmen keine höheren Effekte zeigten, war demnach auch kein Alterseffekt zu erwarten.

Von Bedeutung war allerdings, ob lediglich ein kleinerer Teil der Stichprobe der Studie einen Migrationshintergrund besaß und dieser gesondert ausgewertet wurde oder ob die Untersuchung als gezielte Prävention speziell auf Kinder und Jugendliche mit Migrationshintergrund bzw. eine entsprechende ethnische Minderheit abzielte. Bedenkt man, dass es sich im ersteren Falle um eine universelle Präventionsstrategie handelt, da die Interventionsprogramme in solchen Fällen mit Teilnehmern durchgeführt wurden, welche nicht hinsichtlich des Migrationshintergrundes selektiert waren, so entsprechen die Ergebnisse der vorliegenden Arbeit dem häufig berichteten Befund, dass gezielte Präventionsmaßnahmen höhere Effekte aufweisen als universelle (Beelmann, 2006). Dieser Unterschied ist offensichtlich nicht ausschließlich auf geringere Prävalenzraten

und damit einhergehende Boden- bzw. Deckeneffekte bei universellen Programmen zurückzuführen, da die Risikobelastung der in der vorliegenden Arbeit bei der Auswertung berücksichtigten Stichprobe in beiden Fällen vergleichbar war. Möglicherweise können die Präventionsmaßnahmen bei einer Beschränkung auf eine bestimmte Risikogruppe besser implementiert und an die individuellen Bedürfnisse der Teilnehmer angepasst werden. Dies scheint insbesondere bei Programmen von Bedeutung zu sein, welche ausschließlich im Gruppenformat angeboten wurden, wohingegen es in Einzelsitzungen, unabhängig von der Konzeption als gezielte Maßnahme für Kinder bzw. Jugendliche mit Migrationshintergrund, möglich war auf die individuellen Bedürfnisse der Teilnehmer einzugehen. Der Effekt der Konzeption als gezielte Maßnahme für Kinder bzw. Jugendliche mit Migrationshintergrund zeigte sich auch unabhängig von der Art der Randomisierung, welche wiederum in der zweifachen gewichteten Regression keinen Einfluss mehr auf die Effektstärke hatte. Auch hierin wäre eine mögliche Erklärung für den bivariaten Zusammenhang zwischen der Art der Randomisierung und der Effektstärke zu finden.

## **5.2 Einschränkungen der Arbeit**

Auch wenn die in dieser Arbeit eingeschlossenen Studien hohe methodische Qualitätsstandards erfüllen mussten, die gewichteten mittleren Effektstärken in den Analysen vergleichsweise homogen sind und es keinen Hinweis auf Publikationsverzerrungen gibt, müssen bei der Interpretation der Ergebnisse der vorliegenden Arbeit einige Limitationen beachtet werden.

Zunächst ist darauf hinzuweisen, dass sämtliche Moderatoranalysen auf Grundlage des Fixed-Effects-Modells berechnet wurden. Dieses ist generell weniger konservativ als das Mixed-Effects-Modell. Jedoch lag bei Analyse der Post-Effektstärken keine signifikante Heterogenität vor, was für die Verwendung des Fixed-Effects-Modells spricht (Lipsey & D. B. Wilson, 2001). Auch ist zu erwähnen, dass die Teststärke von Moderatoranalysen unter Verwendung des Mixed-Effects-Modells herabgesetzt ist, was insbesondere bei Analysen mit geringer Zellenbesetzung ein hohes Risiko für einen Beta-Fehler bedeutet.

Ein weiteres Problem ist der kumulierte Alpha-Fehler. Da in der vorliegenden Arbeit über 20 verschiedene Moderatoranalysen berechnet wurden, ist bei einer Irrtums-

wahrscheinlichkeit von 5 % von mindestens einem Zusammenhang auszugehen, der nur zufällig das Signifikanzniveau erreicht hat. Allerdings ergaben sich auch für über die Hälfte dieser Analysen signifikante Effekte, die meisten davon mit  $p < .01$  oder sogar  $p < .001$ . Dies spricht eher gegen reine Zufallsbefunde, wobei ein Restrisiko diesbezüglich nicht ausgeschlossen werden kann. Es ist zudem zu beachten, dass viele Analysen auch gering besetzte Zellen enthielten und bei Anwendung einer Bonferoni-Korrektur die Teststärke noch zusätzlich verringert worden wäre (Bortz, 2010, S. 232).

Des Weiteren sind in Meta-Analysen gefundene Zusammenhänge nie eindeutig kausal zu interpretieren. Da die Merkmale der Forschungsberichte und der in ihnen untersuchten Programme, enthaltenen Stichproben sowie verwendeten Methoden nicht zufällig den Vergleichen zugewiesen werden können, sind die berichteten Analysen korrelativ zu verstehen. Es kann davon ausgegangen werden, dass alle Moderatorvariablen mit anderen Merkmalen konfundiert sind, weswegen man keine reinen Effekte messen kann. Aus diesem Grund sollte eine Meta-Analyse auch nicht primär als hypothesentestendes Verfahren betrachtet werden und darin gefundene Ergebnisse auch stets in gut kontrollierten Primärstudien abgesichert werden (Charlton, 1996; Cooper & Lemke, 1991). Dies gilt insbesondere für die Analysen zur Implementation, da zu dieser in vielen Studien keine Angaben gemacht wurden, was inhaltlich kaum interpretierbar ist.

Weiterhin ist zu erwähnen, dass die Teststärke für einige Analysen aufgrund der geringen Zellbesetzung nur sehr klein war. Dies traf insbesondere auf Auswertungen hinsichtlich des Erhebungslandes sowie der Migrantengruppe zu. Der deutliche Überschuss an Publikationen aus den USA ist zwar ein durchaus üblicher Befund metaanalytischer Arbeiten, jedoch ergeben sich hieraus Einschränkungen hinsichtlich der Generalisierbarkeit der Ergebnisse. Ob diese auch für nicht-lateinamerikanische innerhalb oder Migranten außerhalb der USA Gültigkeit besitzen, kann daher nicht mit Sicherheit gesagt werden. Auch können Unterschiede in Abhängigkeit vom Herkunftsland auch für hispanoamerikanische Einwanderer nicht gänzlich ausgeschlossen werden, da dieses in etwa der Hälfte der Studien gar nicht näher spezifiziert wurde. In ähnlicher Weise ist unklar, ob die Ergebnisse auch auf Kinder und Jugendliche der ersten Migrantengeneration anwendbar sind. In den eingeschlossenen Studien gab es nur selten Angaben zum Geburtsland und in Fällen, in denen es solche gab, wurde nur in einer verschwindend kleinen Zahl von Vergleichen die erste Migrantengeneration untersucht. Hinsichtlich der Generalisierbarkeit ist auch der negative Zusammenhang zwischen Publikationsjahr und

Effektstärke von Belang. Dieser verdeutlicht die Kontextgebundenheit der Ergebnisse und gibt Anlass zu der Annahme, dass zukünftige Studien und Meta-Analysen zu diesem und anderen Themen zu niedrigeren Wirksamkeitsschätzungen führen könnten – zumindest insofern sich diese nicht auf unbehandelte Kontrollgruppen beschränken.

Und schließlich muss festgestellt werden, dass auch in dieser Arbeit das Vorliegen einer Publikationsverzerrung nicht gänzlich ausgeschlossen werden kann. Zwar ergaben sich in keiner der Analysen Hinweise auf eine solche, jedoch ist zu beachten, dass für die vorliegende Meta-Analyse keine Autoren kontaktiert wurden, weshalb wirklich unveröffentlichte Arbeiten nicht eingeschlossen werden konnten. Allerdings gibt es auch bei den über ProQuest Dissertations & Theses Full Text zugänglich gemachten Dissertationen, die in der vorliegenden Arbeit zum Vergleich herangezogen wurden, keinen Anhaltspunkt dafür, weshalb sie einer Publikationsverzerrung unterliegen sollten. Vielmehr wird die Vermutung eines Publikationsbias häufig gerade auf das Vorliegen geringerer Effekte für Dissertationen gestützt (Beelmann, 2006).

Insgesamt kann jedoch aufgrund der hohen methodischen Qualität der eingeschlossenen Arbeiten, der geringen Heterogenität der Effektstärken sowie des Fehlens sämtlicher Indizien für eine Publikationsverzerrung von einer hohen Validität der berichteten Ergebnisse ausgegangen werden, weswegen aus diesen im folgenden Kapitel Implikationen für die Praxis abgeleitet werden sollen.

### **5.3 Implikationen für die Praxis**

Trotz der genannten Einschränkungen kann Beachtung finden, dass die vorliegende Arbeit alle bisher verfügbaren methodisch besonders hochwertigen und damit aussagekräftigen Befunde und somit den aktuell verfügbaren Wissenstand zu diesem Thema integriert, weshalb auch die Ableitung einiger Implikationen für die Praxis geboten erscheint.

Zunächst muss festgestellt werden, dass mit präventiven Interventionen zur sozio-emotionalen Förderung bei Kindern und Jugendlichen mit Migrationshintergrund im Durchschnitt nur kleine Effekte erzielt werden. Allerdings gibt es auch gute Gründe, nicht auf derartige Präventionsmaßnahmen zu verzichten. Zunächst ist zu sagen, dass die in der vorliegenden Meta-Analyse ermittelte gewichtete mittlere Effektstärke von 0.18 einer Korrelation von 0.09 und damit immerhin einer Erhöhung der Erfolgsrate von 45.5 % auf

54.5 % entsprechen würde (Rosenthal & Rubin, 1982). Auch ist zu erwähnen, dass in vielen Studien keine wirklich unbehandelten Kontrollgruppen verwendet wurden, sodass dies die tatsächliche Wirksamkeit der Interventionen vermutlich unterschätzt. Ein weiterer Grund für die Notwendigkeit von Prävention bei dieser Zielgruppe sind die im Kapitel 2.2 beschriebenen erhöhten Entwicklungsrisiken, mit denen ein Migrationshintergrund einhergeht. Hierbei sind auch die hohe Stabilität internalisierender und externalisierender Probleme sowie das höhere Risiko für Personen mit Migrationshintergrund im Erwachsenenalter an einer psychischen Störung zu erkranken, zu berücksichtigen (Cantor-Graae & Selten, 2005; Porter & Haslam, 2005; Reitz et al., 2005; Swinnen & Selten, 2007). Dass indizierte Präventionsmaßnahmen, trotzdem hier kleinere Decken- und Bodeneffekte zu erwarten wären, keine höhere Wirkung erzielen, spricht zudem dafür, nicht erst zu warten, bis sich erste Auffälligkeiten zeigen oder gar nur den am schlimmsten Betroffenen Hilfestellungen anzubieten, sondern im Sinne einer selektiven Prävention auf Kinder und Jugendliche mit Migrationshintergrund einzugehen. Auch aus diesen Gründen sollte in diesem Fall in einer auf Chancengerechtigkeit und Teilhabe bedachten Gesellschaft der Fokus weniger auf dem *ob* als vielmehr auf dem *wie* der Prävention liegen, insbesondere da sich gezeigt hat, dass unter den richtigen Voraussetzungen durchaus Effekte von  $d > 0.3$  zu erreichen sind. Auch zur Beantwortung dieser Frage liefert die vorliegende Arbeit einige Hinweise.

Als erstes erscheint es wichtig, bei präventiven Interventionen für Kinder und Jugendliche mit Migrationshintergrund sehr individuell vorzugehen. Hierbei sind insbesondere Einzelsitzungen zu empfehlen. Wenn dies nicht möglich ist, so sollte doch zumindest eine selektive Präventionsstrategie gewählt werden, in der speziell Kinder und Jugendliche mit Migrationshintergrund die Zielgruppe sind. Dabei sollte besonders auch auf die Belastungen und Probleme eingegangen werden, welche äußerlich nicht sofort sichtbar sind, wie z.B. Depressivität oder Ängstlichkeit, selbst dann, wenn es primär darum geht, externalisierenden Problemen entgegenzuwirken.

Allgemein sind auch komplexere Programme wirksamer. Es ist also zur sozio-emotionalen Förderung von Kindern und Jugendlichen mit Migrationshintergrund von Vorteil, mehrere verschiedene Komponenten in ein Präventionsprogramm einzubeziehen. Insbesondere ist hierbei die Einbeziehung der Eltern in die präventive Intervention von Bedeutung, durch welche die Wirksamkeit der Maßnahme erhöht werden kann.

Hinsichtlich der Implementation kann eine sehr intensive Supervision Vorteile bringen. Auch kann es lohnenswert sein, verschiedene teilnehmerorientierte Maßnahmen zur Implementationssicherung zu ergreifen, wie beispielsweise einen kleinen finanziellen Anreiz für die Programmteilnahme zu offerieren, Fahrdienste zum Interventionsort anzubieten oder während Elterntrainings Kinderbetreuung zur Verfügung zu stellen. Bezüglich der Intensität ist festzustellen, dass auch sehr kurze Programme eine vergleichsweise hohe Wirksamkeit haben können. Falls dennoch intensivere Präventionsmaßnahmen geplant werden, sollte sichergestellt sein, dass ausreichend Ressourcen zur Verfügung stehen, um die präventive Intervention in sehr großem Umfang umzusetzen.

Als Administrator der Präventionsprogrammes sollte nach Möglichkeit nicht auf Lehrer, sondern eher auf Personen zurückgegriffen werden, welche keinen alltäglichen Kontakt zu den Kindern und Jugendlichen haben. Das Alter der Zielgruppe hat hingegen kaum Relevanz. Präventive Interventionen scheinen demnach für von Migration betroffene Kinder und Jugendliche jeglichen Alters wirksam und somit empfehlenswert zu sein.

#### **5.4 Weiterer Forschungsbedarf**

In der vorliegenden Arbeit hat sich auch gezeigt, dass es trotz der insgesamt recht großen Zahl methodisch hochwertiger Studien in einigen Bereichen noch Forschungsbedarf gibt. Aus diesem Grund sollen einige Empfehlungen hinsichtlich zukünftiger Forschungsarbeit gegeben werden.

Als erstes ist die bisher nur geringe Zahl an Studien außerhalb der USA zu beachten sowie der bisherige Fokus auf lateinamerikanische Migranten. Hier sollte die Wirksamkeit präventiver Interventionen auch für andere Migrantengruppen und außerhalb der USA verstärkt untersucht werden. Dies ist insbesondere relevant, da die Auswirkungen einer internationalen Migration auch abhängig von dem Herkunfts- und Einwanderungsland sind, weswegen auch geklärt werden sollte, ob präventive Interventionen über verschiedene Herkunfts- und Einwanderungsländer hinweg wirksam sind (Berry & Sam, 1997; Möllering, Schiefer, Knafo & Boehnke, 2014). Dabei sollten dann stets auch explizite Angaben darüber gemacht werden, wie viele Teilnehmer aus welchem Herkunftsland kommen und in welchem Land die Kinder bzw. Jugendlichen selbst geboren wurden.

Weiterhin wäre eine experimentelle Absicherung einiger der in der vorliegenden Arbeit gefundenen Zusammenhänge in randomisiert kontrollierten Studien wünschenswert, da in meta-analytischen Moderatoranalysen stets mit verschiedenen konfundierenden Variablen zu rechnen ist. Dies trifft insbesondere auf die Implementation zu, aber auch beispielsweise auf die Einbeziehung der Eltern in die Intervention sowie die Administratoren des Präventionsprogrammes. Die Effekte könnten dabei durch den direkten Vergleich zweier Interventionsgruppen untersucht werden, indem z.B. die Administratoren oder die Supervision experimentell variiert werden.

Im Hinblick auf den Programmtyp könnten zukünftige Studien auch verstärkt Präventionsprogramme in den Fokus rücken, bei denen Lehrer trainiert werden. Die Wirksamkeit solcher Programme ist insbesondere deswegen interessant, weil viele soziale und emotionale Schwierigkeiten von Kindern und Jugendlichen mit Migrationshintergrund bevorzugt im Schulkontext auftreten (Aronowitz, 1984). Im Gegensatz zu den Befunden über Lehrer als Administratoren scheinen hier auch die wenigen ersten Untersuchungen vielversprechend. Es sind jedoch mehr Studien in diesem Bereich notwendig, um das Potential dieses Ansatzes ausreichend abschätzen zu können. Außerdem sollten verstärkt Mehrkomponentenprogramme untersucht werden, da sich auch hier eine hohe Wirksamkeit andeutet, welche jedoch bisher nur auf einem einzigen Gruppenvergleich basiert. Weiterhin gab es nur wenige Programme, die primär auf die Prävention internalisierender Probleme abzielte und das, obwohl gerade hier das migrationsbedingte Risiko am besten nachgewiesen ist (Belhadj Kouider et al., 2014). Auch hier könnte die Forschung intensiviert werden.

Schließlich sollte in zukünftigen Untersuchungen eine längerfristige Nacherhebung stattfinden. Bestenfalls sollten hierbei mehrere Messzeitpunkte eingeplant werden, um den Verlauf der Erfolgskriterien nach der Intervention über längere Zeit hinweg nachverfolgen zu können. Dies ist insbesondere deswegen wünschenswert, weil die Stabilität der Effekte, auch aufgrund einer bisher sehr begrenzten Anzahl entsprechender Studien, bislang keineswegs als gesichert angesehen werden kann.



## 6. ZUSAMMENFASSUNG

Ziel dieser Arbeit war es, die bestehende Forschung zur Wirksamkeit präventiver Interventionen zur sozio-emotionalen Förderung bei Kindern und Jugendlichen mit Migrationshintergrund mit Hilfe der Methode der Meta-Analyse zu integrieren. Dabei wurde zunächst der Begriff *Migrationshintergrund* definiert, Beziehungen zu Sprache und Ethnizität aufgezeigt und Risiken benannt, mit denen ein solcher einhergeht. Es wurde das Konzept der präventiven Intervention erläutert, Besonderheiten der Prävention bei Kindern und Jugendlichen mit Migrationshintergrund hervorgehoben und auf bestehende synthetische Arbeiten sowohl zur allgemeinen Wirksamkeit präventiver Interventionen als auch zur Wirksamkeit für bestimmte ethnische Minderheiten eingegangen. Dabei zeigte sich im Hinblick auf die Wirksamkeitsforschung für Kinder und Jugendliche mit Migrationshintergrund ein Mangel umfassender integrativer Arbeiten, welcher durch die vorliegende Meta-Analyse behoben werden sollte.

In die Meta-Analyse wurden hierbei nur Studien eingeschlossen, welche bis 2013 veröffentlicht bzw. zugänglich gemacht wurden, die Wirksamkeit einer präventiven psychologischen Intervention zur sozio-emotionalen Förderung mit Hilfe eines randomisiert-kontrollierten Designs untersuchten und dabei für eine nicht-klinische Stichprobe, die mehrheitlich aus Kindern und Jugendlichen mit Migrationshintergrund bestand, die Berechnung oder Schätzung einheitlicher Erfolgsparameter ermöglichten. Die Stichprobengröße musste dabei mindestens  $N = 25$  pro Gruppe betragen und bei Cluster-Randomisierung sollten jeder Gruppe mindestens fünf Cluster zugeordnet worden sein. Mittels einer systematischen Suche über eine Vielzahl pädagogischer, psychologischer, medizinischer oder soziologischer elektronischer Datenbanken sowie die Literaturverweise verschiedener Überblicksarbeiten und bereits eingeschlossener Studien wurden dabei insgesamt 69 relevante Forschungsberichte mit 93 Gruppenvergleichen gesichtet.

Die Meta-Analyse ergab eine kleine, aber signifikante Effektstärke von  $d_+ = 0.18$ , die sich als homogen erwies, deren Stabilität über die Zeit jedoch nicht abgesichert werden konnte. Die Wirksamkeit erstreckte sich über eine breite Palette an Erfolgskriterien, wobei internalisierende Probleme am wirksamsten reduziert werden konnten, die Wirksamkeit für Suchtmittelkonsum und externalisierende Probleme jedoch etwas

geringer ausfiel. Hinweise auf eine Publikationsverzerrung gab es nicht, jedoch zeigte sich ein negativer Zusammenhang zwischen Publikationsjahr und der Wirksamkeit sowie Effektstärkenunterschiede in Abhängigkeit von Merkmalen der Methodik, der präventiven Intervention und der Stichprobe. Insbesondere ergaben sich höhere Effektstärken für Vergleiche mit passiver Kontrollgruppe, individueller Randomisierung, nicht äquivalenten Gruppen und Per-protocol-Analysen. Kombinierte Programme insbesondere unter Einbeziehung der Eltern sowie Programme, die unter anderem auf internalisierende Probleme abzielten, intensiv supervidiert wurden, teilnehmerorientierte Maßnahmen zur Implementationsförderung unternahmen, eine sehr kurze oder sehr lange Dauer hatten, nicht durch Lehrer durchgeführt wurden und auch Einzelsitzungen enthielten, waren zudem wirksamer. Hinsichtlich der Stichprobe zeigte sich lediglich eine erhöhte Wirksamkeit für Studien, in denen die Gesamtstichprobe mehrheitlich Kinder und Jugendliche mit Migrationshintergrund umfasste.

Insgesamt macht diese Meta-Analyse deutlich, dass präventive Interventionen zur sozio-emotionalen Förderung bei Kindern und Jugendlichen mit Migrationshintergrund im Mittel eine kleine Wirksamkeit aufweisen, die jedoch von verschiedenen Faktoren abhängig ist. Diese Ergebnisse wurden in den Kontext bereits bestehender Arbeiten zur Wirksamkeit von Präventionsprogrammen gestellt. Es wurden Limitationen der Arbeit, aber auch Defizite der bisherigen Forschung benannt sowie Implikationen für die Praxis abgeleitet.

## 7. LITERATURVERZEICHNIS

(Mit \* gekennzeichnete Literatur wurde in die Meta-Analyse eingeschlossen)

- Abraha, I., Cherubini, A., Cozzolino, F., Florio, R. D., Luchetta, M. L., Rimland, J. M. et al. (2015). Deviation from intention to treat analysis in randomized trials and treatment effect estimates: meta-epidemiological study. *BMJ*, *350*, h2445.
- Adesope, O. O., Lavin, T., Thompson, T. & Ungerleider, C. (2011). Pedagogical strategies for teaching literacy to ESL immigrant students: a meta-analysis. *British Journal of Educational Psychology*, *81*, 629-653.
- Aksel, S., Gün, Z., Irmak, T. Y. & Cengelci, B. (2007). Migration and psychological status of adolescents in Turkey. *Adolescence*, *42*, 589-602.
- Alba, R. (1999). Immigration and the American realities of assimilation and multiculturalism. *Sociological Forum*, *14* (1), 3-16.
- \*Andrews, S. R. (1982). The skills of mothering: A study of Parent Child Development Centers. *Monographs of the Society for Research in Child Development*, *47* (6), 1-83. [Doppelpublikationen: Bridgeman, B., Blumenthal, J. B. & Andrews, S. R. (1981). *Parent Child Development Center: Final evaluation report*. Washington, DC: Department of Health and Human Services, Office of Human Development Services. Retrieved March 17, 2014, from ERIC database. (ED213764); Johnson, D. L. (2006). Parent-Child Development Center follow-up project: child behavior problem results. *Journal of Primary Prevention*, *27* (4), 391-407; Johnson, D. L. & Blumenthal, J. (2004). The Parent Child Development Centers and school achievement: A follow-up. *Journal of Primary Prevention*, *25* (2), 195-209; Johnson, D. L. & Breckenridge, J. N. (1982). The Houston Parent-Child Development Center and the primary prevention of behavior problems in young children. *American Journal of Community Psychology*, *10* (3), 305-316; Johnson, D. L. & Walker, T. (1987). Primary prevention of behavior problems in Mexican-American children. *American Journal of Community Psychology*, *15* (4), 375-385; Walker, T. & Johnson, D. L. (1988). A follow-up evaluation of the Houston Parent-Child Development Center: Intelligence test results. *Journal of Genetic Psychology: Research and Theory on Human Development*, *149* (3), 377-381.]

- Arnold, D. H. & Doctoroff, G. L. (2003). The early education of socioeconomically disadvantaged children. *Annual Review of Psychology*, 54, 517-545.
- Aronowitz, M. (1984). The social and emotional adjustment of immigrant children: A review of the literature. *International Migration Review*, 18 (2), 237-257.
- \*Austin, D. B. (2005). The effects of a strengths development intervention program upon the self-perceptions of students' academic abilities. *Dissertation Abstracts International*, 66 (05), 1631A. (UMI No. 3175080).
- \*Avcı-Werning, M. (2004). *Prävention ethnischer Konflikte in der Schule. Ein Unterrichtsprogramm zur Verbesserung interkultureller Beziehungen*, Münster: Waxmann.
- \*Bachay, J. (1998). Ethnic identity development and urban Haitian adolescents. *Journal of Multicultural Counseling and Development*, 26 (2), 96-109. [Doppelpublikation: Bachay, J. B. (1994). *A study of the effect of a preventive, psychoeducational, counselor intervention on the ethnic identity development of Haitian adolescents*. Retrieved December 21, 2015, from ERIC database. (ED404604).]
- \*Barnett, W. S., Jung, K., Yarosz, D. J., Thomas, J., Hornbeck, A., Stechuk, R. et al. (2008). Educational effects of the Tools of the Mind curriculum: A randomized trial. *Early Childhood Research Quarterly*, 23, 299-313.
- Baytop, C. M. (2006). Evaluating the effectiveness of programs to improve educational attainment of unwed African American teen mothers: A meta analysis. *Journal of Negro Education*, 75 (3), 458-477.
- Becker, B. & Biedinger, N. (2006). Ethnische Bildungsungleichheit zu Schulbeginn. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 58 (4), 660-684.
- Beelmann, A. (2006). Wirksamkeit von Präventionsmaßnahmen bei Kindern und Jugendlichen: Ergebnisse und Implikationen der integrativen Erfolgsforschung. *Zeitschrift für Klinische Psychologie Psychologie und Psychotherapie*, 35 (2), 151-162.
- Beelmann, A. & Bliesener, T. (1994). Aktuelle Probleme und Strategien der Metaanalyse. *Psychologische Rundschau*, 45, 211-233.
- Beelmann, A. & Karing, C. (2014). Implementationsfaktoren und -prozesse in der Präventionsforschung: Strategien, Probleme, Ergebnisse, Perspektiven. *Psychologische Rundschau*, 65 (3), 129-139.

- Beelmann, A., Pfof, M. & Schmitt, C. (2014). Prävention und Gesundheitsförderung bei Kindern und Jugendlichen: Eine Meta-Analyse der deutschsprachigen Wirksamkeitsforschung. *Zeitschrift für Gesundheitspsychologie*, 22 (1), 1-14.
- Beelmann, A. & Raabe, T. (2007). *Dissoziales Verhalten von Kindern und Jugendlichen. Erscheinungsformen, Entwicklung, Prävention und Behandlung*. Göttingen: Hogrefe.
- Begg, C. B. & Mazumdar, M. (1994). Operating characteristics of a rank correlation test for publication bias. *Biometrics*, 50 (4), 1088-1101.
- Belhadj Kouider, E., Koglin, U. & Petermann, F. (2014). Emotional and behavioral problems in migrant children and adolescents in Europe: A systematic review. *European Childhood & Adolescent Psychiatry*, 23, 373-391.
- Berry, J. W. (1997). Immigration, acculturation, and adaptation. *Applied Psychology: An International Review*, 46 (1), 5-34.
- Berry, J. W., Kim, U., Minde, T. & Mok, D. (1987). Comparative studies of acculturative stress. *International Migration Review*, 21, 491-511.
- Berry, J. W. & Sam, D. (1997). Acculturation and adaptation. In J. W. Berry, M. H. Segall & C. Kagitçibasi (Eds.), *Cross-cultural psychology: Volume 3 Social behavior and applications* (2<sup>nd</sup> ed., S. 291-326). Boston: Allyn and Bacon.
- Bhugra, D. (2004). Migration and mental health. *Acta Psychiatrica Scandinavica*, 109, 243-258.
- Birman, D. & Taylor-Ritzler, T. (2007). Acculturation and psychological distress among adolescent immigrants from the former Soviet Union: Exploring the mediating effect of family relationships. *Cultural Diversity and Ethnic Minority Psychology*, 13 (4), 337-346.
- \*Bjorknes, R., Kjobli, J., Manger, T. & Jakobsen, R. (2012). Parent training among ethnic minorities: Parenting practices as mediators of change in child conduct problems. *Family Relations*, 61 (1), 101-114. [Doppelpublikation: Bjorknes, R. & Manger, T. (2013). Can parent training alter parent practice and reduce conduct problems in ethnic minority children? A randomized controlled trial. *Prevention Science*, 14, 52-63.]
- Bledsoe, K. L. (2002). Effectiveness of drug prevention programs designed for adolescents of color: A meta-analysis. *Dissertation Abstracts International*, 63 (09), 4414B. (UMI No. 3065492).

- Bortz, J. (2005). *Statistik für Human- und Sozialwissenschaftler* (6., vollständig überarbeitete und aktualisierte Aufl.). Berlin: Springer.
- Bortz, J. & Schuster, C. (2010). *Statistik für Human- und Sozialwissenschaftler* (7., vollständig überarbeitete und aktualisierte Aufl.). Berlin: Springer.
- Bös, M. (2008). Ethnizität. In N. Baur, H. Korte, M. Löw & M. Schroer (Hrsg.), *Handbuch Soziologie* (S. 55-76). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Bosker, R. J. (1994). School effects on drop out: A multi-level logistic approach to assessing school-level correlates of drop out of ethnic minorities. *Tijdschrift voor Onderwijsresearch*, 19, 50-64.
- Bosnjak, M. & Viechtbauer, W. (2009). Die Methode der Meta-Analyse zur Evidenzbasierung von Gesundheitsrisiken: Beiträge der Sozial-, Verhaltens- und Wirtschaftswissenschaften. *Zentralblatt für Arbeitsmedizin, Arbeitsschutz und Ergonomie*, 59 (11), 322-333.
- Bourdieu, P. (1983). Ökonomisches Kapital, kulturelles Kapital, soziales Kapital. In R. Kreckel (Hrsg.), *Soziale Ungleichheiten* (S. 183-198). Göttingen: Schwartz.
- Bradley, R. H. & Corwyn, R. F. (2002). Socioeconomic status and child development. *Annual Review of Psychology*, 53, 371-399.
- Bradley, R. H., Corwyn, R. F., McAdoo, H. P. & Coll, C. G. (2001). The home environments of children in the United States part I: Variations by age, ethnicity, and poverty status. *Child Development*, 72, 1844-1867.
- Brock, R. L. & Lawrence, E. (2014). Marital processes, neuroticism, and stress as risk factors for internalizing symptoms. *Couple and Family Psychology: Research and Practice*, 3 (1), 30-47.
- \*Butler, M. M. (2008). Latinas taking the lead: The impact of a leadership intervention on the leadership behaviors and self-esteem of Latina girl scouts. *Dissertation Abstracts International*, 69 (04), 1228A. (UMI No. 3311985).
- Campbell, D. T. & Stanley, J. C. (1963). Experimental and quasi-experimental designs for research. In N. L. Gage (Ed.), *Handbook of research on teaching* (pp. 171-246). Chicago, IL: McNally & Company.
- Cantor-Graae, E. & Selten, J.-P. (2005). Schizophrenia and migration: A meta-analysis and review. *American Journal of Psychiatry*, 162, 12-24.
- Caplan, G. (1964). *Principles of preventive psychiatry*. New York: Basic Books.

- \*Cappella, E., Hamre, B. K., Kim, H. Y., Henry, D. B., Frazier, S. L., Atkins, M. S. et al. (2012). Teacher consultation and coaching within mental health practice: Classroom and child effects in urban elementary schools. *Journal of Consulting and Clinical Psychology, 80* (4), 597-610.
- Capps, R., Fix, M., Murray, J., Ost, J., Passel, J. S., Herwantoro, S. (2005). *The new demography of America's schools: Immigration and the No Child Left Behind Act*. Washington, D.C.: The Urban Institute.
- \*Cardemil, E. V., Reivich, K. J., Beevers, C. G., Seligman, M. E. & James, J. (2007). The prevention of depressive symptoms in low-income, minority children: Two-year follow-up. *Behaviour Research and Therapy, 45* (2), 313-327.
- Cardoza, V. J., Documet, P. I., Fryer, C. S., Gold, M. A. & Butler, J. (2012). Sexual health behavior interventions for US Latino adolescents: A systematic review of the literature. *Journal of pediatric and adolescent gynecology, 25* (2), 136-149.
- Carrasco, M., Volkmar, F. R. & Bloch, M. H. (2012). Pharmacologic treatment of repetitive behaviors in autism spectrum disorders: Evidence of publication bias. *Pediatrics, 129*, e1301-e1310.
- Carta, M. G., Bernal, M., Hardoy, M. G. & Haro-Abad, J. M. (2005). Migration and mental health in Europe (the state of the mental health in Europe working group: Appendix 1). *Clinical Practice and Epidemiology in Mental Health, 1*, 13. Retrieved November 12, 2014, from <http://www.cpementalhealth.com/content/1/1/13>
- Castro, F. G., Barrera, M., Pantin, H., Martinez, C., Felix-Ortiz, M., Rios, R. et al. (2006). Substance abuse prevention intervention research with Hispanic populations. *Drug Alcohol Dependence, 84* (Suppl. 1), S29-S42.
- Castro, F. G., Cota, M. K. & Vega, S. C. (1999). Health promotion in Latino populations: A sociocultural model for program planning, development, and evaluation. In R. M. Huff & M. V. Kline (Eds.), *Promoting health in multicultural populations: A handbook for practitioners* (pp. 137-168). Thousand Oaks, CA: Sage.
- Centers for Disease Control and Prevention (2006). Youth risk behavior surveillance—United States, 2005: Surveillance summaries. *Morbidity and Mortality Weekly Report 2006, 55* (SS-5), 1-108. Retrieved October 6, 2014, from <http://www.cdc.gov/mmwr/pdf/ss/ss5505.pdf>
- Charlton, B. G. (1996). The uses and abuses of meta-analysis. *Family Practice, 13* (4), 397-401.

- \*Chen, J.-L., Weiss, S. J., Heyman, M. B., Cooper, B. & Lustig, R. H. (2010). The active balance childhood program for improving coping and quality of life in Chinese American children. *Nursing Research*, 59 (4), 270-279.
- \*Cheng, S. H. (2004). Factors that influence the effectiveness of a parenting program for Asians and Pacific Islanders. *Dissertation Abstracts International*, 65 (05), 1963A. (UMI No. 3133127).
- Choi, Y., He, M. & Harachi, T. W. (2008). Intergenerational cultural dissonance, parent-child conflict and bonding, and youth problem behaviors among Vietnamese and Cambodian immigrant families. *Journal of Youth and Adolescence*, 37, 85-96.
- Chu, P. S., Saucier, D. A. & Hafner, E. (2010). Meta-analysis of the relationships between social support and well-being in children and adolescents. *Journal of Social and Clinical Psychology*, 29 (6), 624-645.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20 (1), 37-46.
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2<sup>nd</sup> ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Coie, J. D., Watt, N. F., West, S. G., Hawkins, J. D., Asarnow, J. R., Markman, H. J. et al. (1993). The Science of prevention. A conceptual framework and some directions for a national research program. *American Psychologist*, 48, 1013-1022.
- Cooper, H., Hedges, L. V. & Valentine, J. C. (2009). *Handbook of research synthesis and meta-analysis* (2<sup>nd</sup> ed.). New York: Russell Sage Foundation.
- Cooper, H. M. & Lemke, K. M. (1991). On the role of meta-analysis in personality and social psychology. *Personality and Social Psychology Bulletin*, 17 (3), 245-251.
- \*Costantino, G., Malgady, R. G. & Rogler, L. H. (1984). Cuentos folklóricos as a therapeutic modality with Puerto Rican children. *Hispanic Journal of Behavioral Sciences*, 6 (2), 169-178. [Doppelpublikation: Costantino, G., Malgady, R. G. & Rogler, L. H. (1986). Cuento therapy: A culturally sensitive modality for Puerto Rican children. *Journal of Consulting and Clinical Psychology*, 54 (5), 639-645.]
- \*Costantino, G., Malgady, R. G. & Rogler, L. H. (1994). Storytelling Through Pictures - culturally sensitive psychotherapy for Hispanic children and adolescents. *Journal of Clinical Child Psychology*, 23 (1), 13-20.
- \*Covey, D. D. (1973). *An analytical study of secondary freshmen bilingual education and its effects on academic achievement and attitudes of Mexican American students*



- (Doctoral dissertation). Retrieved October 23, 2013, from ProQuest Dissertations and Theses database. (UMI No. 73-5305).
- Cuijpers, P., van Straten, A., Bohlmeijer, E., Hollon, S. D. & Andersson, G. (2010). The effects of psychotherapy for adult depression are overestimated: A meta-analysis of study quality and effect size. *Psychological Medicine*, 40, 211-223.
- \*Davis, L. A. J. (1994). The effect of parent involvement training on the achievement of Hispanic students. *Dissertation Abstracts International*, 55 (08), 2262A. (UMI No. 9424368).
- De Vries, M. R. & Wolbers, M. H. J. (2004). Ethnic variation in labour market outcomes among school-leavers in the Netherlands: The role of educational qualifications and social background. *Journal of Youth Studies*, 7, 3-18.
- Dent, C. W., Sussman, S., Ellickson, P., Brown, P. & Richardson, J. (1996). Is current drug abuse prevention programming generalizable across ethnic groups. *American Behavioral Scientist*, 39, 911-918.
- Dickersin, K. (1990). The existence of publication bias and risk factors for its occurrence. *Journal of the American Medical Association*, 263 (10), 1385-1389.
- Durlak, J. A. & Wells, A. M. (1997). Primary prevention mental health programs for children and adolescents: A meta-analytic review. *American Journal of Community Psychology*, 25 (2), 115-152.
- Dwairy, M. & Dor, A. (2009). Parenting and psychological adjustment of adolescent immigrants in Israel. *Journal of Family Psychology*, 23 (3), 416-425.
- \*Eisen, M., Zellman, G. L., Massett, H. A. & Murray, D. M. (2002). Evaluating the Lions-Quest 'Skills for Adolescence' drug education program: First-year behavior outcomes. *Addictive behaviors*, 27 (4), 619-632. [Doppelpublikation: Eisen, M. (2002). Intermediate outcomes from a life skills education program with a media literacy component. In W. D. Crano & M. Burgoon (Eds.), *Mass media and drug prevention: Classic and contemporary theories and research* (S. 187-211). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.]
- \*Eisen, M., Zellman, G. & McAlister, A. (1990). Evaluating the impact of a theory-based sexuality and contraceptive education program. *Family Planning Perspectives*, 22, 261-271. [Doppelpublikation: Eisen, M. & Zellman, G. L. (1992). A health beliefs field experiment: Teen talk. In B. C. Miller, J. J. Card, R. L. Paikoff & J. L. Peterson (Eds.), *Preventing adolescent pregnancy: Model programs and evaluations* (Sage

- focus editions, Vol. 140, S. 220-264). Thousand Oaks, CA US: Sage Publications, Inc.]
- \*Elder, J. P., Wildey, M., Demoor, C., Sallis, J. F., Eckhardt, L., Edwards, C. et al. (1993). The long-term prevention of tobacco use among junior-high-school students - classroom and telephone interventions. *American Journal of Public Health*, 83 (9), 1239-1244. [Doppelpublikation: Elder, J. P., Woodruff, S. I., Sallis, J. F. & Moor, C. de (1994). Effects of health facilitator performance and attendance at training sessions on the acquisition of tobacco refusal skills among multi-ethnic, high-risk adolescents. *Health Education Research*, 9 (2), 225-233.]
- Erb, H.-P. & Bohner, G. (2006). Minoritäten. In H.-W. Bierhoff & D. Frey (Hrsg.), *Handbuch der Sozialpsychologie und Kommunikationspsychologie* (S. 494-503). Göttingen: Hogrefe.
- Europäische Kommission (2010). *Die Zusammenhänge zwischen Migration und Diskriminierung*. Luxemburg: Amt für Veröffentlichungen der Europäischen Union.
- \*Fang, L., Schinke, S. P. & Cole, K. C. A. (2010). Preventing substance use among early Asian-American adolescent girls: Initial evaluation of a web-based, mother-daughter program. *Journal of Adolescent Health*, 47 (5), 529-532. [Doppelpublikation: Fang, L. & Schinke, S. P. (2012). Two-Year Outcomes of a Randomized, Family-Based Substance Use Prevention Trial for Asian American Adolescent Girls. *Psychology of Addictive Behaviors*. Advance online publication. doi: 10.1037/a0030925]
- Farrington, D. P. (2003). Methodological quality standards for evaluation research. *ANNALS of the American Academy of Political and Social Science*, 587, 49-68.
- \*Feldman, B. A. (2012). An analysis of ethnicity in treatment response of parents in the national classroom literacy interventions and outcomes in even start study. *Dissertation Abstracts International*, 73 (01), 69A. (UMI No. 3477462).
- Fergusson, D. M., Swain-Campbell, N. R. & Horwood, L. J. (2002). Deviant peer affiliations, crime and substance use: A fixed effects regression analysis. *Journal of Abnormal Child Psychology*, 30 (4), 419-430.
- Finzi-Dottan, R., Bilu, R. & Golubchik, P. (2011). Aggression and conduct disorder in former Soviet Union immigrant adolescents: The role of parenting style and ego identity. *Children and Youth Services Review*, 33, 918-926.

- Foxcroft, D. R. & Tsertsvadze, A. (2011). Universal school-based prevention programs for alcohol misuse in young people. *Cochrane Database of Systematic Reviews*, 2011 (5). (CD009113).
- Franklin, C., Grant, D., Corcoran, J., Miller, P. O. & Bultman, L. (1997). Effectiveness of prevention programs for adolescent pregnancy: A meta-analysis. *Journal of Marriage and Family*, 59 (3), 551-567.
- Fuchs-Heinritz, W., Klimke, D., Lautmann, R., Rammstedt, O., Stäheli, U., Weischer, C. et al. (Hrsg.). (2011). *Lexikon zur Soziologie* (5., überarbeitete Aufl.). Wiesbaden: VS Verlag für Sozialwissenschaften.
- \*Galvez, D. M. (1989). *The effectiveness of the "thinking, changing, rearranging" program for fifth-grade students as measured by the behavioral academic self-esteem rating scale* (Master's thesis). Retrieved February 22, 2015, from ProQuest Dissertations and Theses database. (UMI No. 1335970).
- Gersten, R. & Baker, S. (2000). The professional knowledge base on instructional practices that support cognitive growth for English-language learners. In R. M. Gersten, E. P. Schiller & S. Vaughn (Eds.), *Contemporary special education research. Syntheses of the knowledge base on critical instructional issues. The LEA series on special education and disability* (S. 31-79). Mahwah, NJ US: Lawrence Erlbaum Associates Publishers.
- \*Ghosh Ippen, C. M. (1999). Testing the feasibility and effectiveness of a multicomponent program: Including parent-training, child competencies training, and tutoring. *Dissertation Abstracts International*, 60 (02), 828B. (UMI No. 9919133).
- Gil, A. G., Wagner, E. F. & Vega, W. A. (2000). Acculturation, familism, and alcohol use among Latino adolescent males: Longitudinal relations. *Journal of Community Psychology*, 28 (4), 443-458.
- Gilbody, S. M., Song, F., Eastwood, A. J. & Sutton, A. (2000). The causes, consequences and detection of publication bias in psychiatry. *Acta Psychiatrica Scandinavica*, 102, 241-249.
- Glass, G. V. (1976). Primary, secondary, and meta-analysis research. *Educational Researcher*, 5 (10), 3-8.
- Glass, G. V. (1977). Integrating findings: The meta-analysis of research. *Review of Research in Education*, 5, 351-379.

- Glass, G. V., McGaw, B. & Smith, M. L. (1981). *Meta-analysis in social research*. Newbury Park, CA: Sage Publications.
- \*Gonzales, N. A., Dumka, L. E., Millsap, R. E., Gottschall, A., McClain, D. B., Wong, J. J. et al. (2012). Randomized trial of a broad preventive intervention for Mexican American adolescents. *Journal of Consulting and Clinical Psychology, 80* (1), 1-16. [Doppelpublikation: Dillman Carpentier, F. R., Mauricio, A. M., Gonzales, N. A., Millsap, R. E., Meza, C. M., Dumka, L. E. et al. (2007). Engaging Mexican origin families in a school-based preventive intervention. *Journal of Primary Prevention, 28* (6), 521-546.]
- Gordon, R. S. (1983). An operational classification of disease prevention. *Public Health Reports, 98* (2), 107-109.
- Grabka, M. M. & Frick, J. R. (2010). Weiterhin hohes Armutsrisiko in Deutschland: Kinder und junge Erwachsene sind besonders betroffen. In Deutsches Institut für Wirtschaftsforschung (Hrsg.), *Wochenbericht, Jahrgang 77, Nr. 7*. Berlin: DIW.
- \*Graham, J. W., Johnson, A., Hansen, W. B., Flay, B. R. & Gee, M. (1990). Drug use prevention programs, gender and ethnicity: evaluation of three seventh grade project SMART cohorts. *Preventive Medicine, 19*, 305-313. [Doppelpublikation: Hansen, W. B., Johnson, C. C., Flay, B. R., Graham, J. W. & Sobel, J. L. (1988). Affective and social influences approaches to the prevention of multiple substance abuse among seventh grade students: Results from Project SMART. *Preventive Medicine, 17*, 135-154.]
- Grieco, E. M. (2010). *Race and Hispanic origin of the foreign-born population in the United States: 2007*. Washington DC: U.S. Census Bureau. Retrieved January 8, 2016, from <https://www.census.gov/prod/2010pubs/acs-11.pdf>
- Griffin, K. W., Epstein, J. A., Botvin, G. J. & Spoth, R. L. (2001). Social competence and substance use among rural youth: Mediating role of social benefit expectancies of use. *Journal of Youth and Adolescence, 30* (4), 485-498.
- Grossman, B., Wirt, R. & Davids, A. (1985). Self-esteem, ethnic identity, and behavioral adjustment among Anglo and Chicano adolescents in West Texas. *Journal of Adolescence, 8*, 57-68.
- Grünigen, R., Perren, S., Nägele, C. & Alsaker, F. D. (2010). Immigrant children's peer acceptance and victimization in kindergarten: The role of local language competence. *British Journal of Developmental Psychology, 28*, 679-697.

- \*Guilamo-Ramos, V., Bouris, A., Jaccard, J., Gonzalez, B., McCoy, W. & Aranda, D. (2011). A parent-based intervention to reduce sexual risk behavior in early adolescence: Building alliances between physicians, social workers, and parents. *Journal of Adolescent Health, 48* (2), 159-163.
- \*Guilamo-Ramos, V., Jaccard, J., Dittus, P., Gonzalez, B., Bouris, A. & Banspach, S. (2010). The Linking Lives Health Education Program: A randomized clinical trial of a parent-based tobacco use prevention program for African American and Latino youths. *American Journal of Public Health, 100* (9), 1641-1647.
- \*Gunn, B., Biglan, A., Smolkowski, K. & Ary, D. (2000). The efficacy of supplemental instruction in decoding skills for Hispanic and non-Hispanic students in early elementary school. *Journal of Special Education, 34* (2), 90-103. [Doppelpublikationen: Barrera, M., Biglan, A., Taylor, T. K., Gunn, B. K., Smolkowski, K., Black, C. et al. (2002). Early elementary school intervention to reduce conduct problems: A randomized trial with Hispanic and non-Hispanic children. *Prevention Science, 3* (2), 83-94; Gunn, B., Smolkowski, K., Biglan, A. & Black, C. (2002). Supplemental instruction in decoding skills for Hispanic and non-Hispanic students in early elementary school: A follow-up study. *Journal of Special Education, 36* (2), 69-79; Gunn, B., Smolkowski, K., Biglan, A., Black, C. & Blair, J. (2005). Fostering the development of reading skill through supplemental instruction: Results for hispanic and non-hispanic students. *Journal of Special Education, 39* (2), 66-85; Smolkowski, K., Biglan, A., Barrera, M., Taylor, T., Black, C. & Blair, J. (2005). Schools and homes in partnership (SHIP): Long-term effects of a preventive intervention focused on social behavior and reading skill in early elementary school. *Prevention Science, 6* (2), 113-125.]
- Gupta, S. K. (2011). Intention-to-treat concept: A review. *Perspectives in Clinical Research, 2* (3), 109-112.
- \*Harpine, E. C. & Reid, T. (2009). Enhancing academic achievement in a Hispanic immigrant community: The role of reading in academic failure and mental health. *School Mental Health, 1* (4), 159-170.
- \*Harrington, N. G., Giles, S. M., Hoyle, R. H., Feeney, G. J. & Yungbluth, S. C. (2001). Evaluation of the All Stars character education and problem behavior prevention program: Effects on mediator and outcome variables for middle school students. *Health Education & Behavior, 29* (5), 533-546. [Doppelpublikation: Hansen, W. B.

- (1996). Pilot test results comparing the All Stars program with seventh grade D.A.R.E.: Program integrity and mediating variable analysis. *Substance Use and Misuse*, 31 (10),1359-1377.]
- Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics*, 6 (2), 107-128.
- Hedges, L. V. (1982). Fitting categorical models to effect sizes from a series of experiments. *Journal of Educational Statistics*, 7 (2), 119-137.
- Hedges, L. V. & Olkin, I. (1985). *Statistical methods for meta-analysis*. San Diego, CA: Academic Press.
- Herbert-Myers, H., Guttentag, C. L., Swank, P. R., Smith, K. E. & Landry, S. H. (2006). The importance of language, social, and behavioral skills across early and later childhood as predictors of social competence with peers. *Applied Developmental Science*, 10 (4), 174-187.
- \*Hernandez, L. (2008). Identity as a moderator of guided-self change treatment outcomes with adolescents. *Dissertation Abstracts International*, 68 (08), 5610B. (UMI No. 3279226).
- Higgins, J. P. T. & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine*, 21, 1539-1558.
- \*Hovell, M., Blumberg, E., Sipan, C., Hofstetter, C. R., Burkham, S., Atkins, C. et al. (1998). Skills training for pregnancy and AIDS prevention in Anglo and Latino youth. *Journal of Adolescent Health*, 23 (3), 139-149.
- Hovey, J. D. (2000). Psychosocial predictors of acculturative stress in Mexican immigrants. *Journal of Psychology*, 134 (5), 490-502.
- Huedo-Medina, T. B., Sánchez-Meca, J., Marín-Martínez, F. & Botella, J. (2006). Assessing heterogeneity in meta-analysis: Q statistic or  $I^2$  index?. *Psychological Methods*, 11 (2), 193-206.
- Huey, S. J. & Polo, A. J. (2008). Evidence-Based Psychosocial Treatments for Ethnic Minority Youth. *Journal of Clinical Child & Adolescent Psychology*, 37 (1), 262-301.
- Hunter, J. E. & Schmidt, F. L. (1995). *Methods of meta-analysis: Correcting error and bias in research findings*. Newbury Park, California: Sage Publications.
- Huynh, Q.-L., Devos, T. & Smalarz, L. (2011). Perpetual foreigner in one's own land: Potential implications for identity and psychological adjustment. *Journal of Social and Clinical Psychology*, 30, 133-162.

- Iacono, W. G., Malone, S. M. & McGue, M. (2003). Substance use disorders, externalizing psychopathology, and P300 event-related potential amplitude. *International Journal of Psychophysiology*, *48*, 147-178.
- Jackson, K. F. (2009). Building cultural competence: A systematic evaluation of the effectiveness of culturally sensitive interventions with ethnic minority youth. *Children and Youth Services Review*, *31* (11), 1192-1198.
- Jani, J. S., Ortiz, L. & Aranda, M. P. (2009). Latino outcome studies in social work: A review of the literature. *Research on Social Work Practice*, *19* (2), 179-194.
- \*Jaycox, L. H., McCaffrey, D., Eiseman, B., Aronoff, J., Shelley, G. A., Collins, R. L. et al. (2006). Impact of a school-based dating violence prevention program among Latino teens: Randomized controlled effectiveness trial. *Journal of Adolescent Health*, *39* (5), 694-704.
- \*Johnson, C. A., Unger, J. B., Ritt-Olson, A., Palmer, P. H., Cen, S. Y., Gallaher, P. et al. (2005). Smoking prevention for ethnically diverse adolescents: 2-year outcomes of a multicultural, school-based smoking prevention curriculum in Southern California. *Preventive Medicine: An International Journal Devoted to Practice and Theory*, *40* (6), 842-852. [Doppelpublikation: Johnson, C. A., Cen, S., Gallaher, P., Palmer, P. H., Xiao, L., Ritt-Olson, A. et al. (2007). Why smoking prevention programs sometimes fail. Does effectiveness depend on sociocultural context and individual characteristics?. *Cancer Epidemiology Biomarkers & Prevention*, *16* (6), 1043-1049.]
- Johnston, L. D., O'Malley, P. M., Bachman, J. G. & Schulenberg, J. E. (2011). *Monitoring the future national results on adolescent drug use: Overview of key findings, 2010*. Ann Arbor: Institute for Social Research. The University of Michigan. Retrieved January 8, 2016, from <http://monitoringthefuture.org/pubs/monographs/mtf-overview2010.pdf>
- Jonas, K., Stroebe, W. & Hewstone, M. (Hrsg.). (2014). *Sozialpsychologie* (6., vollständig überarbeitete Aufl.). Berlin: Springer.
- Jung, W. S. & Stinnett, T. A. (2005). Comparing judgements of social, behavioural, emotional and school adjustment functioning for Korean, Korean American and Caucasian American children. *School Psychology International*, *26*, 317-329.
- \*Kelly, P. J., Lesser, J., Cheng, A.-L., Oscó-Sánchez, M., Martinez, E., Pineda, D. et al. (2010). A prospective randomized controlled trial of an interpersonal violence

- prevention program with a Mexican American community. *Family & Community Health: Journal of Health Promotion & Maintenance*, 33 (3), 207-215.
- Kemper, T. (2010). Migrationshintergrund – eine Frage der Definition!. *Die Deutsche Schule*, 102 (4), 315-326.
- \*Kirby, D., Baumler, E., Coyle, K., Basen-Engquist, K., Parcel, G., Harrist, R. et al. (2004). The "Safer Choices" intervention: its impact on the sexual behaviors of different subgroups of high school students. *Journal of Adolescent Health: Official Publication of the Society for Adolescent Medicine*, 35 (6), 442-52. [Doppelpublikationen: Coyle, K. Basen-Engquist, K., Kirby, D., Parcel, G., Banspach, S., Collins, J. et al. (2001). Safer Choices: Reducing teen pregnancy, HIV, and STDs. *Public Health Reports*, 116, 82-93; Lezin, N., Roller, L. A., Wilson, M. M., Taylor, J., Firpo-Triplett, R. & Coyle, K. K. (2011). *Safer Choices Adaptation Kit*. Santa Cruz, CA: ETR Associates.]
- Knight, G. P., Virdin, L. M., Ocampo, K. A. & Roosa, M. (1994). An examination of the cross-ethnic equivalence of measures of negative life events and mental health among Hispanic and Anglo-American children. *American Journal of Community Psychology*, 22, 767-783.
- Kong, G., Singh, N. & Krishnan-Sarin, S. (2012). A review of culturally targeted/tailored tobacco prevention and cessation interventions for minority adolescents. *Nicotine & Tobacco Research*, 14 (12), 1394-1406.
- Kraemer, H. C., Kazdin, A. E., Offord, D. R., Kessler, R. C., Jensen, P. S. & Kupfer, D. J. (1997). Coming to terms with the terms of risk. *Archives of General Psychiatry*, 54, 337-344.
- Kromrey, J. D. & Rendina-Gobioff, G. (2006). On knowing what we do not know: An empirical comparison of methods to detect publication bias in meta-analysis. *Educational and Psychological Measurement*, 66 (3), 357-373.
- Kuperminc, G. P., Blatt, S. J., Shahar, G., Henrich, C. & Leadbeater, B. J. (2004). Cultural equivalence and cultural variance in longitudinal associations of young adolescent self-definition and interpersonal relatedness to psychological and school adjustment. *Journal of Youth and Adolescence*, 33, 13-30.
- Landis, J. R. & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33 (1), 159-174.



- Laucht, M., Esser, G. & Schmidt, M. H. (2000). Externalisierende und internalisierende Störungen in der Kindheit: Untersuchungen zur Entwicklungspathologie. *Zeitschrift für Klinische Psychologie und Psychotherapie*, 29, 284-292.
- Lee, S. M. (1998). Asian Americans: Diverse and growing. *Population Bulletin*, 53 (2), 1-43.
- \*Leo, P. F. (1972). *The effects of two types of group counseling upon the academic achievement and self-concept of Mexican-American pupils in the elementary school* (Doctoral dissertation). Retrieved February 2, 2014, from ERIC database. (ED085683).
- \*Liber, J. M., Boo, G. M., Huizenga, H. & Prins, P. J. (2013). School-based intervention for childhood disruptive behavior in disadvantaged settings: A randomized controlled trial with and without active teacher support. *Journal of Consulting and Clinical Psychology*, 81 (6), 975-987.
- Lipsey, M. W. & Wilson, D. B. (1993). The efficacy of psychological educational, and behavioral treatment. *American Psychologist*, 48 (12), 1181-1209.
- Lipsey, M. W. & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage Publications.
- \*Litrownik, A. J., Elder, J. P., Campbell, N. R., Ayala, G. X., Slymen, D. J., Parra-Medina, D. et al. (2000). Evaluation of a tobacco and alcohol use prevention program for Hispanic migrant adolescents: Promoting the protective factor of parent-child communication. *Preventive Medicine: An International Journal Devoted to Practice and Theory*, 31 (2), 124-133. [Doppelpublikation: Elder, J. P., Litrownik, A. J., Slymen, D. J., Campbell, N. R., Parra-Medina, D., Choe, S. et al. (2002). Tobacco and alcohol use-prevention program for Hispanic migrant adolescents. *American Journal of Preventive Medicine*, 23 (4), 269-275.]
- Lösel, F. & Beelmann, A. (2003). Effects of child skills training in preventing antisocial behavior: A systematic review of randomized evaluation. *ANNALS of the American Academy of Political and Social Science*, 587, 84-109.
- \*Love, J. M., Chazan-Cohen, R., Raikes, H. & Brooks-Gunn, J. (2013). What makes a difference: Early Head Start evaluation findings in a developmental context. *Monographs of the Society for Research in Child Development*, 78 (1), vii-viii, 1-173. [Doppelpublikationen: Love, J. M. (2010). *Effects of Early Head Start prior to kindergarten entry: The importance of early experience*. Retrieved April 4, 2013,

- from ERIC database. (ED514662); Love, J. M., Kisker, E. E., Ross, C. M., Schochet, P. Z., Brooks-Gunn, J., Boller, K. et al. (2001). *Building their futures: How Early Head Start programs are enhancing the lives of infants and toddlers in low-income families. Volume I: Technical Report*. Retrieved March 3, 2014, from ERIC database. (ED454952); Love, J. M., Kisker, E. E., Ross, C. M., Schochet, P. Z., Brooks-Gunn, J., Boller, K. et al. (2001). *Building their futures: How Early Head Start programs are enhancing the lives of infants and toddlers in low-income families. Volume II: Technical Report Appendixes*. Princeton, NJ.: Mathematica Policy Research. Retrieved September 04, 2014, from <http://www.researchconnections.org/childcare/resources/5651#>; Love, J. M., Constantine, J., Paulsell, D., Boller, K., Ross, C., Raikes, H. et al. (2004). *The role of Early Head Start programs in addressing the child care needs of low-income families with Infants and toddlers: Influences on child care use and quality*. Princeton, NJ: Mathematica Policy Research. Retrieved September 05, 2014, from <http://www.acf.hhs.gov/programs/opre/resource/the-role-of-early-head-start-programs-in-addressing-the-child-care-needs-of>; Love, J. M., Kisker, E. E., Ross, C., Raikes, H., Constantine, J., Boller, K. et al. (2005). The effectiveness of early Head Start for 3-Year-old children and their parents: Lessons for policy and programs. *Developmental Psychology*, 41 (6), 885-901.]
- Lundahl, B., Risser, H. J. & Lovejoy, C. (2006). A meta-analysis of parent training: Moderators and follow-up effects. *Clinical Psychology Review*, 26, 86-104.
- Maichrowitz, S., Arnold, L., Schulz, S. & Beelmann, A. (2014). Die Wirksamkeit von psychosozialen und pädagogischen Programmen zur Prävention und Kompensation von Armutsfolgen bei Kindern und Jugendlichen: Eine multinationale Meta-Analyse [Abstract]. In O. Güntürkün (Hrsg.), *49. Kongress der Deutschen Gesellschaft für Psychologie: Supplement to Psychological Test and Assessment Modeling*, 21. Bis 25. September 2014, Ruhr-Universität Bochum. (S. 330-331). Lengerich: Pabst Science Publishers.
- \*Malgady, R. G. (1990). Hero/Heroin Modeling for Puerto Rican adolescents: A preventive mental health intervention. *Journal of Consulting and Clinical Psychology*, 58 (4), 469-474.
- Manongdo, J. A. & Ramírez García, J. I. (2007). Mother's parenting dimensions and adolescent externalizing and internalizing behaviors in a low-income, urban Mexican

- American sample. *Journal of Clinical Child and Adolescent Psychology*, 36 (4), 593-604.
- \*Markham, C. M., Tortolero, S. R., Peskin, M. F., Shegog, R., Thiel, M., Baumler, E. R. et al. (2012). Sexual risk avoidance and sexual risk reduction interventions for middle school youth: A randomized controlled trial. *Journal of Adolescent Health*, 50 (3), 279-288.
- \*Marsiglia, F. F., Yabiku, S. T., Kulis, S., Nieri, T. & Lewin, B. (2010). Influences of school Latino composition and linguistic acculturation on a prevention program for youths. *Social Work Research*, 34 (1), 6-19. [Doppelpublikationen: Gosin, M. N., Dustman, P. A., Drapeau, A. E. & Harthun, M. L. (2003). Participatory Action Research: creating an effective prevention curriculum for adolescents in the Southwestern US. *Health Education Research*, 18 (3), 363-379; Hecht, M. L., Graham, J. W. & Elek, E. (2006). The drug resistance strategies intervention: Program effects on substance use. *Health Communication*, 20 (3), 267-276; Hecht, M. L., Marsiglia, F. F., Elek, E., Wagstaff, D. A., Kulis, S., Dustman, P. et al. (2003). Culturally grounded substance use prevention: An evaluation of the keepin' it R.E.A.L. curriculum. *Prevention Science*, 4 (4), 233-248; Kulis, S., Marsiglia, F. F., Elek, E., Dustman, P., Wagstaff, D. A. & Hecht, M. L. (2005). Mexican/Mexican American adolescents and keepin' it REAL: An evidence-based substance use prevention program. *Children & Schools*, 27 (3), 133-145; Kulis, S., Yabiku, S., Marsiglia, F., Nieri, T. & Crossman, A. (2007). Differences by gender, ethnicity, and acculturation in the efficacy of the keepin' it REAL model prevention program. *Journal of Drug Education*, 37 (2), 123-44; Marsiglia, F. F., Kulis, S., Wagstaff, D. A., Elek, E. & Dran, D. (2005). Acculturation status and substance use prevention with Mexican and Mexican-American youth. *Journal of Social Work Practice in the Addictions*, 5 (1-2), 85-111; Yabiku, S., Kulis, S., Marsiglia, F. F., Lewin, B., Nieri, T. & Hussaini, S. (2007). Neighborhood effects on the efficacy of a program to prevent youth alcohol use. *Substance use & Misuse*, 42 (1), 65-87.]
- Martens, P. L. (2000). Immigrants as victims of crime. *International Review of Victimology*, 8, 199-216.
- Martinez, C. R. (2006). Effects of differential family acculturation on Latino adolescent substance use. *Family Relations*, 55, 306-317.

- \*Martinez, C. R. & Eddy, J. M. (2005). Effects of culturally adapted parent management training on Latino youth behavioral health outcomes. *Journal of Consulting and Clinical Psychology, 73* (5), 841-851.
- Martinez, C. R., McClure, H. H. & Eddy, J. M. (2009). Language brokering contexts and behavioral and emotional adjustment among Latino parents and adolescents. *Journal of Early Adolescence, 29*, 71-98.
- Matt, G. E. (2003). Will it work in Münster? Meta-analysis and the empirical generalization of causal relationships. In R. Schulze, H. Holling & D. Böhning (Eds.). *Meta-analysis: New developments and applications in medical and social sciences* (S. 113-140). Cambridge, MA: Hogrefe & Huber Publishers.
- Maynard, M. J. & Harding, S. (2010). Ethnic differences in psychological well-being in adolescence in the context of time spent in family activities. *Social Psychiatry and Psychiatric Epidemiology, 45*, 115-123.
- McKee, L., Colletti, C., Rakow, A., Jones, D. J. & Forehand, R. (2008). Parenting and child externalizing behaviors: Are the associations specific or diffuse?. *Aggression and Violent Behavior, 13* (3), 201-215.
- \*Meir, Y., Slone, M., Levis, M., Reina, L. & Livni, Y. B. D. (2012). Crisis intervention with children of illegal migrant workers threatened with deportation. *Professional Psychology: Research and Practice, 43* (4), 298-305.
- \*Mendelsohn, A. L., Valdez, P. T., Flynn, V., Foley, G. M., Berkule, S. B., Tomopoulos, S. et al. (2007). Use of videotaped interactions during pediatric well-child care: Impact at 33 months on parenting and on child development. *Journal of Developmental and Behavioral Pediatrics, 28* (3), 206-212. [Doppelpublikation: Mendelsohn, A. L., Dreyer, B. P., Flynn, V., Tomopoulos, S., Rovira, I., Tineo, W. et al. (2005). Use of videotaped interactions during pediatric well-child care to promote child development: A randomized, controlled trial. *Journal of Developmental and Behavioral Pediatrics, 26* (1), 34-41.]
- Merbach, M., Wittig, U. & Brähler, E. (2008). Angst und Depression polnischer und vietnamesischer MigrantInnen in Leipzig unter besonderer Berücksichtigung ihres Eingliederungsprozesses. *Psychotherapie Psychosomatik medizinische Psychologie, 58*, 146-154.

- Miranda, J., Bernal, G., Lau, A., Kohn, L., Hwang, W. C. & LaFromboise, T. (2005). State of the science on psychosocial interventions for ethnic minorities. *Annual Review of Clinical Psychology, 1*, 113-142.
- Moilanen, K. L., Shaw, D. S. & Maxwell, K. L. (2010). Developmental cascades: Externalizing, internalizing, and academic competence from middle childhood to early adolescence. *Developmental and Psychopathology, 22*, 635-653.
- Möllering, A., Schiefer, D., Knafo, A. & Boehnke, K. (2014). Acculturation and well-being among migrant and minority adolescents: A cross-national and cross-ethnic comparison. In R. K. Silbereisen, P. F. Titzmann & Y. Shavit (Eds.), *The challenges of diaspora migration: Interdisciplinary perspectives on Israel and Germany* (pp. 83-102). Farnham, UK: Ashgate.
- Morris, S. B. (2008). Estimating effect sizes from pretest-posttest-control group designs. *Organizational Research Methods, 11*, 364-386.
- Munoz, R. F., Mrazek & Haggerty (1996). Institute of Medicine report on prevention of mental disorders. *American Psychologist, 51*, 1116-1122.
- \*Murray, N. G., Kelder, S. H., Parcel, G. S., Frankowski, R. & Orpinas, P. (1999). Padres Trabajando por la Paz: A randomized trial of a parent education intervention to prevent violence among middle school children. *Health Education Research, 14* (3), 421-426. [Doppelpublikation: Murray, N. G. (1997). Design and evaluation of the parent education component of a comprehensive violence prevention project in middle schools. *Dissertation Abstracts International, 57* (12), 7481B. (UMI No. 9715366).]
- National Science Foundation, Division of Science Resources Statistics (2003). *Women, minorities, and persons with disabilities in science and engineering 2002*. Arlington, VA: National Science Foundation (NSF 03-312).
- Nauck, B., Clauß, S. & Richter, E. (2008). Zur Lebenssituation von Kindern mit Migrationshintergrund in Deutschland. In H. Bertram (Hrsg.), *Der UNICEF-Bericht zur Lage der Kinder in Deutschland* (S. 127-151). München: C.H.Beck.
- Newcombe, R. G. (1987). Towards a reduction in publication bias. *British Medical Journal, 295*, 656-659.
- \*Nickerson, K. F. (2004). Anger in adolescents: The effectiveness of a brief cognitive-behavioral anger management training program for reducing attitudinal and behavioral expressions of anger. *Dissertation Abstracts International, 65* (01), 425B. (UMI No. 3119161).

- Nykliček, I. & Kuijpers, K. F. (2008). Effects of mindfulness-based stress reduction intervention on psychological well-being and quality of life: Is increased mindfulness indeed the mechanism? *Annals of Behavioral Medicine*, 35 (3), 331-340.
- OECD (2007). *Education at a glance 2007: OECD indicators*. Paris: OECD Publishing.
- OECD (2011). *Education at a glance 2011: OECD-Indicators*. Paris: OECD Publishing.
- Offord, D. R. (2000). Selection of levels of prevention. *Addictive Behaviors*, 25 (6), 833-842.
- Offord, D. R., Kraemer, H. C., Kazdin, A. E., Jensen, P. S. & Harrington, R. (1998). Lowering the burden of suffering from child psychiatric disorder: Trade-offs among clinical, targeted, and universal interventions. *Journal of the American Academy of Child and Adolescent Psychiatry*, 37 (7), 686-694.
- Oltmer, J. (2013). Migration. In K.-H. Meier-Braun & R. Weber (Hrsg.), *Migration und Integration in Deutschland. Begriffe-Fakten-Kontroversen* (S. 31-34). Bonn: Bundeszentrale für politische Bildung.
- Orwin, R. G. (1983). A fail-safe  $N$  for effect size in meta-analysis. *Journal of educational statistics*, 8 (2), 157-159.
- \*Pantin, H., Coatsworth, J. D., Feaster, D. J., Newman, F. L., Briones, E., Prado, G. et al. (2003). Familias Unidas: The efficacy of an intervention to promote parental investment in Hispanic immigrant families. *Prevention Science*, 4 (3), 189-201.
- \*Pantin, H., Prado, G., Lopez, B., Huang, S., Tapia, M. I., Schwartz, S. J. et al. (2009). A randomized controlled trial of Familias Unidas for Hispanic adolescents with behavior problems. *Psychosomatic Medicine*, 71 (9), 987-995. [Doppelpublikation: Cordova, D., Huang, S., Pantin, H. & Prado, G. (2012). Do the effects of a family intervention on alcohol and drug use vary by nativity status?. *Psychology of Addictive Behaviors*, 26 (3), 655-660.]
- Patiño, C. & Kirchner, T. (2009). Estrés y calidad de vida en inmigrantes latinoamericanos en Barcelona [Stress und Lebensqualität lateinamerikanischer Einwanderer in Barcelona]. *Ansiedad y Estrés*, 15, 49-66.
- Perez-Johnson, I. & Maynard, R. (2007). The case for early, targeted interventions to prevent academic failure. *Peabody Journal of Education*, 82 (4), 587-616.
- Perrez, M. (1994). Optimierung und Prävention im erzieherischen Bereich. In K. A. Schneewind (Hrsg.), *Psychologie der Erziehung und Sozialisation* (Enzyklopädie der

- Psychologie, Serie Pädagogische Psychologie, Bd. 1, S. 585-617). Göttingen: Hogrefe.
- Perrez, M. (1998). Prävention und Gesundheitsförderung. In U. Baumann & M. Perrez (Hrsg.), *Lehrbuch Klinische Psychologie – Psychotherapie* (2. Aufl., S. 366-391). Bern: Huber.
- Perrez, M. & Baumann, U. (2011). Systematik der klinisch-psychologischen Intervention. In M. Perrez & U. Baumann (Hrsg.), *Lehrbuch klinische Psychologie – Psychotherapie* (4., aktualisierte Aufl.; S.341-351). Bern: Verlag Hans Huber.
- Pew Hispanic Center (2004). *Latino teens staying in high school: A challenge for all generations*. Washington, DC: Autor.
- Pew Research Center (2013). *Second-generation Americans: A portrait of the adult children of immigrants*. Washington, DC: Autor. Retrieved May 6, 2015, from <http://www.pewsocialtrends.org/2013/02/07/second-generation-americans/>
- Pigott, T. D. (2001). Missing predictors in models of effect size. *Evaluation & the Health Professions*, 24 (3), 277-307.
- Pinquart, M. & Silbereisen, R. K. (2014). Prävention und Gesundheitsförderung im Jugendalter. In K. Hurrelmann, T. Klotz & J. Haisch (Hrsg.), *Lehrbuch Prävention und Gesundheitsförderung* (4., vollständig überarbeitete Aufl., S. 70-78). Bern: Huber.
- \*Polansky, J. M., Buki, L. P., Horan, J. J., Ceperich, S. D. & Burows, D. D. (1999). The effectiveness of substance abuse prevention videotapes with Mexican American adolescents. *Hispanic Journal of Behavioral Sciences*, 21 (2), 186-198. [Doppelpublikation: Polansky, J. M. (1995). Common and specific effects of substance-abuse-prevention videotapes on Mexican American adolescents. *Dissertation Abstracts International*, 55 (09), 4104B. (UMI No. 9434396).]
- Porter, M. & Haslam, N. (2005). Predisplacement and postdisplacement factors associated with mental health of refugees and internally displaced persons: A meta-analysis. *JAMA*, 294, 602-612.
- \*Prado, G., Cordova, D., Huang, S., Estrada, Y., Rosen, A., Bacio, G. A. et al. (2012). The efficacy of Familias Unidas on drug and alcohol outcomes for Hispanic delinquent youth: Main effects and interaction effects by parental stress and social support. *Drug and Alcohol Dependence*, 125 (Suppl. 1), S18-S25. [Doppelpublikation: Prado, G., Pantin, H., Huang, S., Cordova, D., Tapia, M. I.,

- Velazquez, M.-R. et al. (2012). Effects of a family intervention in reducing HIV risk behaviors among high-risk Hispanic adolescents: a randomized controlled trial. *Archives of Pediatrics & Adolescent Medicine*, 166 (2), 127-133.]
- Prado, G., Huang, S., Schwartz, S. J., Maldonado-Molina, M. M., Bandiera, F. C., De la Rosa, M. et al. (2009). What accounts for differences in substance use among U.S.-born and immigrant Hispanic adolescents?: Results from a longitudinal prospective cohort study. *Journal of Adolescent Health*, 45, 118-125.
- \*Prado, G., Pantin, H., Briones, E., Schwartz, S. J., Feaster, D., Huang, S. et al. (2007). A randomized controlled trial of a parent-centered intervention in preventing substance use and HIV risk behaviors in Hispanic adolescents. *Journal of Consulting and Clinical Psychology*, 75 (6), 914-926. [Doppelpublikation: Prado, G. (2005). The efficacy of three interventions to prevent substance use and sex initiation in subgroups of Hispanic adolescents. *Dissertation Abstracts International*, 66 (03), 1705B. (UMI No. 3168675).]
- \*Prokhorov, A. V., Kelder, S. H., Shegog, R., Murray, N., Peters, R., Agurcia-Parker, C. et al. (2008). Impact of a smoking prevention interactive experience (ASPIRE), an interactive, multimedia smoking prevention and cessation curriculum for culturally diverse high-school students. *Nicotine & tobacco research*, 10 (9), 1477-1485.
- \*Puma, M., Bell, S., Cook, R., Heid, C., Shapiro, G., Broene, P. et al. (2010). *Head Start impact study: Final report*. Washington, DC.: Administration for Children & Families, US Department of Health and Human Services. Retrieved December 21, 2015, from <http://www.researchconnections.org/childcare/resources/17707?author=Puma%2C+Michael>
- Pyakuryal, A., Tausig, M., Subedi, S. & Subedi, J. (2011). Strangers in a familiar land: The psychological consequences of internal migration in a developing country. *Stress and Health*, 27, e199-e208.
- Quintana, S. M., Vogel, M. C. & Ybarra, V. C. (1991). Meta-analysis of Latino students' adjustment in higher education. *Hispanic Journal of Behavioral Sciences*, 13, 155-168.
- \*Raisner, A. (1967). *Science instruction in Spanish for pupils of Spanish speaking background, an experiment in bilingualism: Final report*. Retrieved December 20, 2013, from ERIC database. (ED017926).



- \*Ramirez, S. Z., Jain, S., Flores-Torres, L. L., Perez, R. & Carlson, R. (2009). The effects of Cuento therapy on reading achievement and psychological outcomes of Mexican-American students. *Professional School Counseling, 12* (3), 253-262.
- \*Raver, C. C., Jones, S. M., Li-Grining, C., Zhai, F., Metzger, M. W. & Solomon, B. (2009). Targeting children's behavior problems in preschool classrooms: A cluster-randomized controlled trial. *Journal of Consulting Clinical Psychology, 77* (2), 302-316.
- Reitz, E., Dekovic, M. & Meijer, A. M. (2005). The structure of externalizing and internalizing problem behavior during early adolescence. *Journal of Youth and Adolescence, 34* (6), 577-588.
- \*Rodriguez, O. (2012). *Efficacy of school-based mental health program on prosocial behavior and aggression among Mexican American children* (Doctoral dissertation). Retrieved June 8, 2013, from ProQuest Dissertations and Theses database. (UMI No. 3545192).
- \*Rodriguez, R., Marchand, E., Ng, J. & Stice, E. (2008). Effects of a cognitive dissonance-based eating disorder prevention program are similar for Asian American, Hispanic, and White participants. *International Journal of Eating Disorder, 41* (7), 618-625.
- Röhrle, B.(2008). Die Forschungslage zur Prävention psychischer Störungen und Förderung psychischer Gesundheit. *Prävention, 31* (1), 10-13.
- \*Rosenbaum, D. P., Flewelling, R. L., Bailey, S. L. & Ringwalt, C. L. (1994). Cops in the classroom: A longitudinal evaluation of drug abuse resistance education (DARE). *Journal of Research in Crime and Delinquency, 31* (1), 3-31. [Doppelpublikation: Ennett, S. T., Rosenbaum, D. P., Flewelling, R. L., Bieler, G. S., Ringwalt, C. L. & Bailey, S. L. (1994). Long-term evaluation of drug abuse resistance education. *Addictive Behaviors, 19* (2), 113-125.]
- Rosenthal, R. (1979). The 'File Drawer Problem' and tolerance for null results. *Psychological Bulletin, 86* (3), 638-641.
- Rosenthal, R. & Rubin, D. B. (1982). A simple, general purpose display of magnitude of experimental effect. *Journal of Educational Psychology, 74* (2), 168-169.
- \*Rousseau, C., Benoit, M., Gauthier, M.-F., Lacroix, L., Alain, N., Rojas, M. V. et al. (2007). Classroom drama therapy program for immigrant and refugee adolescents: A pilot study. *Clinical Child Psychology and Psychiatry, 12* (3), 451-465.

- \*Rousseau, C., Benoit, M., Lacroix, L. & Gauthier, M. F. (2009). Evaluation of a sandplay program for preschoolers in a multiethnic neighborhood. *Journal of Child Psychology and Psychiatry*, 50 (6), 743-750.
- Sáenz, L. M., Fuchs, L. S. & Fuchs, D. (2005). Peer-assisted learning strategies for English language learners with learning disabilities. *Exceptional Children*, 71 (3), 231-247.
- Sam, D. L., Vedder, P., Ward, C. & Horenczyk, G. (2006). Psychological and sociocultural adaptation of immigrant youth. In J. W. Berry, J. S. Phinney, D. L. Sam & P. Vedder (Eds.), *Immigrant youth in cultural transition: Acculturation, identity, and adaptation across national contexts* (pp. 117-141). Mahwah, New Jersey: Lawrence Erlbaum Associates.
- Schenk, L., Neuhauser, H., Ellert, U., Poethko-Müller, C., Kleiser, C. & Mensink, G. (2008). *Kinder- und Jugendgesundheitsurvey (KIGGS) 2003-2006: Kinder und Jugendliche mit Migrationshintergrund in Deutschland*. Berlin: Robert-Koch-Institut.
- \*Schulting, A. B. (2010). The Kindergarten Home Visit Project: A kindergarten transition intervention study. *Dissertation Abstracts International*, 71 (04), 2726B. (UMI No. 3397780).
- Schulz, S., Maichrowitz, S., Arnold, L. & Beelmann, A. (2014). Wirksamkeit von Präventionsprogrammen für Kinder und Jugendlichen mit Migrationshintergrund: Eine multinationale Meta-Analyse [Abstract]. In O. Güntürkün (Hrsg.), *49. Kongress der Deutschen Gesellschaft für Psychologie: Supplement to Psychological Test and Assessment Modeling, 21. Bis 25. September 2014, Ruhr-Universität Bochum*. (S. 330-331). Lengerich: Pabst Science Publishers.
- Schulz, S., Titzmann, P. F. & Michel, A. (2013). Jugendliche Übersetzer: Language Brokering in russischen Migrantenfamilien in Deutschland. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 45 (3), 161-171.
- Searle, W. & Ward, C. (1990). The prediction of psychological and sociocultural adjustment during cross-cultural transitions. *International Journal of Intercultural Relations*, 14, 449-464.
- Settelmeyer, A. & Erbe, J. (2010). *Migrationshintergrund. Zur Operationalisierung des Begriffs in der Berufsbildungsforschung. Wissenschaftliche Diskussionspapiere Heft 112*. Bonn: Bundesinstitut für Berufsbildung.

- Shanahan, T. & Beck, I. L. (2006). Effective Literacy Teaching for English-language learners. In D. August & T. Shanahan (Eds.), *Developing literacy in second-language learners: Report of the National Literacy Panel on Language-Minority Children and Youth* (S. 415-488). Mahwah, NJ US: Lawrence Erlbaum Associates Publishers.
- Sheeran, P., Abraham, C. & Orbell, S. (1999). Psychosocial correlates of heterosexual condom use: A meta-analysis. *Psychological Bulletin*, *125*, 90-132.
- Sheidow, A. J., Henry, D. B., Tolan, P. H. & Strachan, M. K. (2014). The role of stress exposure and family functioning in internalizing outcomes of urban families. *Journal of Child and Family Studies*, *23*, 1351-1365.
- \*Sherman, D. K., Hartson, K. A., Binning, K. R., Purdie-Vaughns, V., Garcia, J., Taborsky-Barba, S. et al. (2013). Deflecting the trajectory and changing the narrative: How self-affirmation affects academic performance and motivation under identity threat. *Journal of Personality and Social Psychology*, *104* (4), 591-618.
- Sherman, E. J. & Primack, B. A. (2009). What works to prevent adolescent smoking? A systematic review of the National Cancer Institute's Research-Tested Intervention Programs. *Journal of School Health*, *79* (9), 391-399.
- Slavin, R. E. & Cheung, A. (2005). A synthesis of research on language of reading instruction for English language learners. *Review of Educational Research*, *75* (2), 247-284.
- Sluzki, C. E. (1979). Migration and family conflict. *Family Process*, *18* (4), 379-390.
- Sluzki, C. E. (2010). Psychologische Phasen der Migration und ihre Auswirkungen. In T. Hegemann & R. Salman (Hrsg.), *Handbuch Transkulturelle Psychiatrie* (S. 108-123). Bonn: Psychiatrie-Verlag.
- Stanat, P., Becker, M., Baumert, J., Lüdtke, O. & Eckhardt, A. G. (2012). Improving second language skills of immigrant students: A field trial study evaluating the effects of a summer learning program. *Learning and Instruction*, *22* (3), 159-170.
- Steinbach, A. & Nauck, B. (2004). Intergenerationale Transmission von kulturellem Kapital in Migrantenfamilien. Zur Erklärung von ethnischen Unterschieden im deutschen Bildungssystem. *Zeitschrift für Erziehungswissenschaft*, *7* (1), 20-32.
- Steinhausen, H.-C., Bearth-Carrari, C. & Metzke, C. W. (2009). Psychosocial adaptation of adolescent migrants in a Swiss community survey. *Social Psychiatry and Psychiatric Epidemiology*, *44*, 308-316.

- Stevens, G. W. J. M., Vollebergh, W. A. M., Pels, T. V. M. & Crijnen, A. A. M. (2007). Parenting and internalizing and externalizing problems in Moroccan immigrant youth in the Netherlands. *Journal of Youth and Adolescence*, 36 (5), 685-695.
- Stice, E., Shaw, H., Bohon, C., Marti, C. N. & Rohde, P. (2009). A meta-analytic review of depression prevention programs for children and adolescents: Factors that predict magnitude of intervention effects. *Journal of Consulting and Clinical Psychology*, 77 (3), 486-503.
- Swinnen, S. G. H. A. & Selten J.-P. (2007). Mood disorders and migration. *British Journal of Psychiatry*, 190, 6-10.
- \*Szapocznik, J. (1989). Family Effectiveness Training: An intervention to prevent drug abuse and problem behaviors in Hispanic adolescents. *Hispanic Journal of Behavioral Sciences*, 11 (1), 4-27.
- Szapocznik, J. & Kurtines, W. M. (1993). Family psychology and cultural diversity: Opportunities for theory, research, and application. *American Psychologist*, 48, 400-407.
- Szapocznik, J. & Williams, R. A. (2000). Brief Strategic Family Therapy: Twenty-five years of interplay among theory, research and practice in adolescent behavior problems and drug abuse. *Clinical Child and Family Psychology Review*, 3 (2), 117-134.
- Teubert, D. & Pinquart, M. (2011). A meta-analytic review on the prevention of symptoms of anxiety in children and adolescents. *Journal of Anxiety Disorders*, 25, 1046-1059.
- Thornton, A. & Lee, P. (2000). Publication bias in meta-analysis: Its causes and consequences. *Journal of Clinical Epidemiology*, 53, 207-216.
- Timmermans, M., van Lier, P. A. C. & Koot, H. M. (2008). Which forms of child/adolescent externalizing behaviors account for late adolescent risky sexual behavior and substance use? *Journal of Child Psychology and Psychiatry*, 49 (4), 386-394.
- Titzmann, P. F. (2012). Growing up too soon? Parentification among immigrant and native youth in Germany. *Journal of Youth and Adolescence*, 41 (7), 880-893.
- Titzmann, P. F., Silbereisen, R. K., Mesch, G. & Schmitt-Rodermund, E. (2011). Migration-specific hassles among adolescent immigrants from the former Soviet Union in Germany and Israel. *Journal of Cross-Cultural Psychology*, 42 (5), 777-794.

- Tobler, N. S., Roona, M. R., Ochshorn, P., Marshall, D. G., Streke, A. V. & Stackpole, K. M. (2000). School-based adolescent drug prevention programs: 1998 meta-analysis. *Journal of Primary Prevention, 20* (4), 275-336
- \*Tortolero, S. R., Markham, C. M., Peskin, M. F., Shegog, R., Addy, R. C., Escobar-Chaves, S. L. et al. (2010). It's your game: Keep it real: Delaying sexual behavior with an effective middle school program. *Journal of Adolescent Health, 46* (2), 169-179.
- Treibel, A. (1999). *Migration in modernen Gesellschaften. Weinheim und München: Juventa Verlag.*
- Treibel, A. (2008). Migration. In N. Baur, H. Korte, M. Löw & M. Schroer (Hrsg.), *Handbuch Soziologie* (S. 295-318). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Tse, L. (1995). When students translate for parents: Effects of language brokering. *CABE Newsletter, 17*, 16-17.
- U.S. Census Bureau (2011). *Statistical abstract of the United States: 2012* (131st Edition). Washington, DC: United States Government Printing Office. Retrieved December 20, 2015, from <http://www2.census.gov/library/publications/2011/compendia/statab/131ed/2012-statab.pdf>
- Vadasy, P. F. & Sanders, E. A. (2010). Efficacy of supplemental phonics-based instruction for low-skilled kindergarteners in the context of language minority status and classroom phonics instruction. *Journal of Educational Psychology, 102* (4), 786-803.
- Vedder, P. (2005). Language, ethnic identity, and the adaptation of immigrant youth in the Netherlands. *Journal of Adolescent Research, 20* (3), 396-416.
- Vega, W. A., Zimmerman, R. S., Warheit, G. J., Apospori, E. & Gil, A. G. (1993). Risk factors for early adolescent drug use in four ethnic and racial groups. *American Journal of Public Health, 83*, 185-189.
- Viechtbauer, W. (2007). Accounting for heterogeneity via random-effects models and moderator analyses in meta-analysis. *Zeitschrift für Psychologie, 215* (2), 104-121.
- \*Villalba, J. A. (2003). Using group counseling to improve the self-concepts, school attitudes and academic success of limited-English-proficient (LEP) Hispanic students in English-for-speakers-of-other-languages/English-as-a-second-language (ESOL/ESL) programs. *Dissertation Abstracts International, 63* (09), 3118A. (UMI No. 3065994).

- \*Villarruel, A. M., Loveland-Cherry, C. J. & Ronis, D. L. (2010). Testing the efficacy of a computer-based parent-adolescent sexual communication intervention for Latino parents. *Family Relations*, 59 (5), 533-543.
- Waldron, H. B. & Turner, C. W. (2008). Evidence-based psychosocial treatments for adolescent substance abuse. *Journal of Clinical Child and Adolescent Psychology*, 37 (1), 238-261.
- Wang, H. (2008). *Family functioning and adolescent adaptation outcomes: A comparative study of race/ethnicity and immigrant status* (Doctoral dissertation). Retrieved November 11, 2011, from ProQuest Dissertations and Theses database. (UMI No. 3277762).
- Warzak, W. J., Grow, C. R., Poler, M. M. & Walburn, J. N. (1995). Enhancing refusal skills: Identifying contexts that place adolescents at risk for unwanted sexual activity. *Developmental and Behavioral Pediatrics*, 16 (2), 98-100.
- Weisburd, D., Lum, C. M. & Yang, S.-M. (2003). When can we conclude that treatments or programs “don’t work”? *Annals of the American Academy*, 587, 31-48.
- Wilson, D. B. (2010). *SPSS macros for performing meta-analytic analyses*. Retrieved December 7, 2015, from <http://mason.gmu.edu/~dwilsonb/ma.html>.
- 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 Counseling and Clinical Psychology*, 71 (1), 136-149.
- Wilson, S. J., Lipsey, M. W. & Soydan, H. (2003). Are mainstream programs for juvenile delinquency less effective with minority youth than majority youth? A meta-analysis of outcomes research. *Research on Social Work Practice*, 13, 3-26.
- Wilson, T. C. (1996). Cohort and prejudice: Whites’ attitudes toward Blacks, Hispanics, Jews, and Asians. *Public Opinion Quarterly*, 60, 253-274.
- Wirtz, M. & Nachtigall, C. (1998). *Deskriptive Statistik. Statistische Methoden für Psychologen, Teil 1*. Weinheim: Juventa Verlag.
- Yakhnich, L. (2008). Immigration as a multiple-stressor situation: Stress and coping among immigrants from the former Soviet Union in Israel. *International Journal of Stress Management*, 15 (3), 252-268.
- Yuen, R. K. (2004). The effectiveness of culturally tailored interventions: A meta-analytic review. *Dissertation Abstracts International*, 65 (08), 4311B. (UMI No. 3143928).

## ANHÄNGE

### Anhang A – Suchbefehle für die Online-Suche in den einzelnen Datenbanken

*EBSCOhost Suchbefehl (Datenbanken: PsycINFO, PSYINDEX, PsycARTICLES)*

((DE (Human Migration OR Immigration OR Social integration OR Refugees OR Migrant farm workers OR Foreign workers OR Expatriates OR Citizenship OR "Racial and Ethnic Differences" OR Cross Cultural Differences OR Cross Cultural Psychology OR Culture OR Culture Shock OR "Race (Anthropological)" OR "Racial and Ethnic relations" OR Multiculturalism OR Interethnic Family OR "Racial and Ethnic Groups" OR Minority Groups OR "Latinos/Latinas" OR Mexican Americans) OR KW (migrant\* OR migrat\* OR immigr\* OR "non-citi\*" OR "non-nativ\*" OR allocht\* OR foreigner\* OR (("born" OR birth\* OR origin\*) N8 (countr\* OR foreign\* OR "outside"))) OR (cultur\* N8 (foreign\* OR differen\* OR divers\* OR origin\* OR "other"))) OR multicultur\* OR ethnic\* OR ("minority group\*" NOT "sexual minority") OR refugee OR "English language learner#" OR bilingual\* OR "as a second language" OR Latino OR Latina OR "Latin Americans" OR "Mexican American#" OR Hispanic OR "Asian American#" OR "Chinese American#" OR "Vietnamese American#" OR "Filipino American#" OR "Indian American#" OR "Korean American#") OR TI (migrant\* OR migrat\* OR immigr\* OR "non-citi\*" OR "non-nativ\*" OR allocht\* OR foreigner\* OR (("born" OR birth\* OR origin\*) N8 (countr\* OR foreign\* OR "outside"))) OR (cultur\* N8 (foreign\* OR differen\* OR divers\* OR origin\* OR "other"))) OR multicultur\* OR ethnic\* OR ("minority group\*" NOT "sexual minority") OR refugee OR "English language learner#" OR bilingual\* OR "as a second language" OR Latino OR Latina OR "Latin Americans" OR "Mexican American#" OR Hispanic OR "Asian American#" OR "Chinese American#" OR "Vietnamese American#" OR "Filipino American#" OR "Indian American#" OR "Korean American#") OR AB (migrant\* OR migrat\* OR immigr\* OR "non-citi\*" OR "non-nativ\*" OR allocht\* OR foreigner\* OR (("born" OR birth\* OR origin\*) N8 (countr\* OR foreign\* OR "outside"))) OR (cultur\* N8 (foreign\* OR differen\* OR divers\* OR origin\* OR "other"))) OR multicultur\* OR ethnic\* OR ("minority group\*" NOT "sexual minority") OR refugee OR "English language learner#" OR bilingual\* OR "as a second language" OR Latino OR Latina OR "Latin Americans" OR "Mexican American#" OR Hispanic OR "Asian American#" OR "Chinese American#" OR "Vietnamese American#" OR "Filipino American#" OR "Indian American#" OR "Korean American#")

OR "Filipino American#" OR "Indian American#" OR "Korean American#") OR ((SU (Vietnames\* OR Chinese\* OR Asian\* OR Filipin\* OR "Turk" OR "Turks" OR "Turkish" OR Russian\* OR Korean\* OR "South Asian\*" OR (Indian\* NOT "American Indian#") OR Pakistani\* OR Chilean\* OR Ethiopian\* OR "Pacific Islander\*" OR Armenian\* OR Mexican\* OR Maghrebian\* OR Aussiedler OR "ethnic German\*" OR Antillean\* OR Surinames\* OR "Cape Verdean\*" OR Angolan\* OR Mozambican\* OR Timorese OR Kurd\* OR Hmong) OR TI (Vietnames\* OR Chinese\* OR Asian\* OR Filipin\* OR "Turk" OR "Turks" OR "Turkish" OR Russian\* OR Korean\* OR "South Asian\*" OR (Indian\* NOT "American Indian#") OR Pakistani\* OR Chilean\* OR Ethiopian\* OR "Pacific Islander\*" OR Armenian\* OR Mexican\* OR Maghrebian\* OR Aussiedler OR "ethnic German\*" OR Antillean\* OR Surinames\* OR "Cape Verdean\*" OR Angolan\* OR Mozambican\* OR Timorese OR Kurd\* OR Hmong)) AND (PL ("US" OR "England" OR "Scotland" OR "Ireland" OR "Australia" OR "Canada" OR "Israel" OR "New Zealand" OR "France" OR "Germany" OR "Netherlands" OR "Finland" OR "Norway" OR "Portugal" OR "Sweden")))) OR ((SU (Portugues\* OR "Central American\*" OR "Finns") OR TI (Portugues\* OR "Central American\*" OR "Finns")) AND (PL ("France" OR "Germany" OR "Sweden"))))

AND

(DE (\*Intervention\* OR Health education OR Adjunctive treatment OR Adventure therapy OR Behavior Modification OR \*Behavior Therapy OR Classroom behavior modification OR Contingency management OR Omission training OR Overcorrection OR \*Psychotherapy OR Aversion therapy OR Exposure therapy OR Implosive therapy OR Reciprocal inhibition therapy OR Response cost OR Systematic desensitization therapy OR Multisystemic therapy OR Autogenic training OR Brief relational therapy OR Play therapy OR Client centered therapy OR "Acceptance and commitment therapy" OR Emotion focused therapy OR Existential therapy OR Eye movement desensitization therapy OR Feminist therapy OR Gestalt therapy OR Encounter group therapy OR Marathon group therapy OR Guided imagery OR Hypnotherapy OR Insight therapy OR Logotherapy OR Narrative therapy OR Network therapy OR Persuasion therapy OR Primal therapy OR psychoanalysis OR psychodrama OR psychotherapeutic counseling OR Family therapy OR Reality therapy OR Relationship therapy OR Solution focused therapy OR Transactional analysis OR Prevention OR Accident prevention OR AIDS prevention OR Crime prevention OR Drug abuse prevention OR Fire prevention OR



Primary mental health prevention OR Relapse prevention OR Suicide prevention OR Program Evaluation OR Educational programs OR Educational Program Evaluation OR Mental health program evaluation OR Mental health programs OR Individual education Programs OR Literacy Programs OR Project follow through OR Project head start OR Upward Bound OR After school programs OR Government programs OR Job corps OR Home visiting programs OR Course Evaluation OR Treatment effectiveness evaluation OR Curriculum OR Career Education OR Compensatory Education OR Foreign Language Education OR Training OR Computer training OR Assertiveness training OR Memory training OR Motivation training OR Parent training OR Self instructional training OR Sensitivity training OR Social skills training OR Counseling OR Community counseling OR Cross cultural counseling OR Cross cultural treatment OR Educational counseling OR Group counseling OR Multicultural counseling OR Occupational guidance OR Pastoral counseling OR Peer counseling OR School counseling OR Bibliotherapy OR Computer assisted therapy OR Creative arts therapy OR Disease management OR Interdisciplinary treatment approach OR Language therapy OR Multimodal treatment approach OR Online therapy OR Psychotherapeutic techniques OR Relaxation therapy OR Progressive relaxation therapy OR Socioterapy OR Speech therapy OR Mentor OR Meditation OR Computer assisted instruction OR Individualized instruction OR Programmed instruction OR Teaching methods OR Directed discussion method OR Discovery teaching method OR educational field trips OR experiential learning OR group instruction OR Montessori method OR Nondirected discussion method OR Open classroom method OR Problem based learning OR Scaffolding OR Service learning OR Team teaching method OR tutoring OR peer tutoring) OR KW (prevent\* OR intervent\* OR program\* OR training OR counsel\* OR mentor\* OR tutor\* OR instruction OR treatment OR therapy OR psychotherapy) OR TI (prevent\* OR intervent\* OR program\* OR training OR counsel\* OR mentor\* OR tutor\* OR instruction OR treatment OR therapy OR psychotherapy) OR AB (((("prevention" OR "intervention" OR "program" OR "training" OR counsel\* OR mentor\* OR tutor\* OR "instruction" OR "treatment" OR "therapy" OR "psychotherapy") N8 (evaluat\* OR "effects" OR "efficacy" OR "effectiveness" OR "efficiency" OR influenc\* OR "improving" OR improve\* OR increas\* OR decreas\*))))

AND

(DE (Neonatal Development OR Neonatal Disorders OR Neonatal intensive Care OR Neonatal period OR "Infants (animal)" OR Child abuse OR Battered child syndrome OR Child abuse reporting OR Child attitudes OR Child behavior checklist OR Child Care OR Child Care workers OR Child custody OR Child Day care OR Child discipline OR Child guidance OR Child guidance clinics OR Child labor OR Child neglect OR Child psychiatry OR Child psychology OR Child psychopathology OR Child psychotherapy OR Child self care OR Child support OR Child visitation OR Child welfare OR Natural Childbirth OR Childbirth training OR Childhood development OR Early childhood development OR Childhood neurosis OR Childhood play behavior OR Childhood psychosis OR Symbiotic infantile psychosis OR Childhood schizophrenia OR Childrearing attitudes OR Childrearing practices OR Children of alcoholics OR Children apperception test OR Childrens manifest anxiety scale OR Childrens recreational games OR Adopted children OR Foster children OR Grandchildren OR Illegitimate children OR Wechsler intelligence scale for children OR Only children OR Stepchildren OR Kaufman Assessment Battery for children OR Kindergarten students OR Adolescent attitudes OR Adolescent development OR Adolescent fathers OR Adolescent mothers OR Adolescent pregnancy OR Adolescent psychology OR Adolescent psychopathology OR Adolescent psychotherapy OR Predelinquent youth OR Juvenile Delinquency OR Juvenile Gangs OR Juvenile Justice OR Preeschool Education OR Preeschool students OR Nursery school students OR High school students OR Reentry students OR Junior high school students OR Middle school students OR Special education students OR Transfer students OR Primary school students OR Elementary school students OR Intermediate school students OR Vocational school students OR Biological Family OR Interethnic family OR Interracial family OR Nuclear family OR Schizophrenogenic family OR Stepfamily OR Family Conflict OR Family crises OR Family therapy OR Strategic family therapy OR Structural family therapy OR Family relations OR Family intervention OR Family medicine OR Family systems theory OR Parent child communication OR Parent child relations OR Parent training OR Parent school relationship OR Parental absence OR Parental attitudes OR Parenting style OR Parental characteristics OR Parental expectations OR Parental investment OR Parental involvement OR Parental role OR Permissive Parenting OR Parenting OR Parenting skills OR Parenting style OR Parents OR Homosexual parents OR Authoritarian parenting OR Authoritative parenting OR Adoptive Parents OR Single parents OR Mothers OR Unwed mothers OR

Schizophrenogenic mothers OR Single mothers OR Mother absence OR Mother child communication OR Mother child relations OR Fathers OR Father child communication OR Father child relations OR Father absence OR Single fathers) OR KW (neonat\* OR baby OR infant OR (child NOT "adult children") OR childhood OR kid OR toddler OR boy OR girl OR adolescen\* OR teen\* OR youth\* OR "minors" OR preschool\* OR pupil# OR ((school OR grade) N8 student) OR grader OR family OR parent\* OR mother\* OR father\*) OR TI (neonat\* OR baby OR infant OR (child NOT "adult children") OR childhood OR kid OR toddler OR boy OR girl OR adolescen\* OR teen\* OR youth\* OR "minors" OR preschool\* OR pupil# OR ((school OR grade) N8 student) OR grader OR family OR parent\* OR mother\* OR father\*) OR AB ("neonates" OR "babies" OR "infants" OR ("children" NOT "adult children") OR "kids" OR "toddlers" OR "boys" OR "girls" OR "adolescents" OR "teens" OR "teenagers" OR "youths" OR "minors" OR "preschoolers" OR "pupils" OR ((school OR grade) N8 "students") OR "graders") OR AG ("childhood (birth-12 yrs)" OR "adolescence(13-17 yrs)" OR "neonatal" OR "infancy" OR "preschool age" OR "childhood" OR "school age" OR "adolescence"))

AND

(DE (Random Sampling OR Experiment controls OR Experimental Design OR Between groups design OR Clinical trials OR Experimental methods OR Quasi experimental methods) OR KW (experiment OR experimental\* OR "between group design" OR "contrast group\*" OR (("control" OR "comparison") N4 condition) OR (("control" OR "comparison") W0 (group\* OR "students" OR "pupils" OR class\* OR school\* OR "families" OR "children" OR "parents" OR "mothers" OR "fathers"))) OR "controls" OR (("wait-list" OR "non-intervention" OR "non-treatment") N4 (control OR condition OR group OR "participants")) OR random\* OR trial OR assign\* OR (("program" OR "intervention") W0 ("participation" OR "participants"))) OR TI (experiment OR experimental\* OR "between group design" OR "contrast group\*" OR (("control" OR "comparison") N4 condition) OR (("control" OR "comparison") W0 (group\* OR "students" OR "pupils" OR class\* OR school\* OR "families" OR "children" OR "parents" OR "mothers" OR "fathers"))) OR "controls" OR (("wait-list" OR "non-intervention" OR "non-treatment") N4 (control OR condition OR group OR "participants")) OR random\* OR trial OR assign\* OR (("program" OR "intervention") W0 ("participation" OR "participants"))) OR AB (experiment OR experimental\* OR "between group design" OR "contrast group\*" OR (("control" OR "comparison") N4 condition) OR (("control" OR

"comparison") W0 (group\* OR "students" OR "pupils" OR class\* OR school\* OR "families" OR "children" OR "parents" OR "mothers" OR "fathers")) OR "controls" OR (("wait-list" OR "non-intervention" OR "non-treatment") N4 (control OR condition OR group OR "participants")) OR random\* OR trial OR assign\* OR (("program" OR "intervention") W0 ("participation" OR "participants"))))

*ProQuest Suchbefehl (Datenbanken: ProQuest Dissertations & Theses Full Text, ERIC)*

SU,TI,AB(migrant\* OR migrat\* OR immigr\* OR "non-citi\*" OR "non-nativ\*" OR "foreign-born" OR "culturally diverse" OR multicultur\* OR ethnic\* OR "minority group\*" OR refugee OR "English language learner\*" OR bilingual\* OR "as a second language" OR Latino\* OR Hispanic\* OR "Asian American\*")

AND

SU,TI(prevent\* OR intervent\* OR program\* OR training OR counsel\* OR mentor\* OR tutor\* OR instruction OR treatment OR therapy OR psychotherapy)

AND

SU,TI(neonat\* OR babies OR infant\* OR child\* OR kids OR boys OR girl\* OR adolescen\* OR teen\* OR youth\* OR minors OR preschool\* OR "school students" OR "grade students" OR grader\* OR family OR families OR parent\* OR mother\* OR father\*)

AND

(SU,TI,AB(experiment OR experiments OR experimental\* OR "controlled trial" OR assign\* OR random\* OR "control condition\*" OR "control group\*") OR AB("contrast group\*" OR "control students" OR "control class\*" OR "control school\*" OR "control families" OR "control children" OR "control parents" OR "control mothers" OR "comparison condition\*" OR "comparison group\*" OR "comparison students" OR "comparison class\*" OR "comparison school\*" OR "comparison families" OR "comparison children" OR "comparison parents" OR "comparison mothers" OR "controls"))

*PubMed Suchbefehl*

((emigration and immigration [mh] OR Cultural Diversity [mh] OR Acculturation [mh] OR Cross-Cultural Comparison [mh] OR ethnic groups [mh:noexp] OR African Continental Ancestry Group [mh:noexp] OR Asian Continental Ancestry Group [mh] OR

Oceanic Ancestry Group [mh] OR Arabs [mh] OR Hispanic Americans [mh] OR Jews [mh]) OR ((migration\* [TIAB] NOT (cell movement [mh] OR "cell migration" [TIAB])) OR migrant\* [TIAB] OR immigr\* [TIAB] OR non-citi\* [TIAB] OR non-nativ\* [TIAB] OR allochth\* [TIAB] OR foreigner\* [TIAB] OR "country of origin" [TIAB] OR "born outside" [TIAB] OR "foreign-born" [TIAB] OR "culturally diverse" [TIAB] OR cultural divers\* [TIAB] OR cultural diff\* [TIAB] OR multicultur\* [TIAB] OR ethnic\* [TIAB] OR (minority group\* [TIAB] NOT sexual minority [TIAB]) OR refugee\* [TIAB] OR English language learner\* [TIAB] OR bilingual\* [TIAB] OR "as a second language" [TIAB] OR Latino\* [TIAB] OR Latina\* [TIAB] OR "Latin Americans" [TIAB] OR Mexican American\* [TIAB] OR Hispanic\* [TIAB] OR Asian American\* [TIAB] OR Chinese American\* [TIAB] OR Vietnamese American\* [TIAB] OR Filipino American\* [TIAB] OR Indian American\* [TIAB] OR Korean American\* [TIAB]))

AND

((("Therapeutics/Psychology" [mh:noexp] OR "therapeutics/Education" [mh:noexp] OR "behavior control/Psychology" [mh] OR Psychotherapy [mh] OR Early Intervention [mh] OR Preventive Health Services [mh:noexp] OR Health Education [mh:noexp] OR Patient Education as Topic [mh] OR Sex Education [mh] OR Health Promotion [mh] OR Primary Prevention [mh] OR Secondary Prevention [mh] OR Preventive Psychiatry [mh] OR Accident Prevention [mh] OR Program Evaluation [mh:noexp] OR Weight Reduction Programs [mh] OR Government Programs [mh] OR Computer-Assisted Instruction [mh] OR Voluntary Programs [mh] OR Postnatal Care [mh] OR Psychology, Applied [mh:noexp] OR Counseling [mh] OR Psychology, Educational [mh:noexp] OR child guidance [mh] OR education of intellectually disabled [mh] OR Neurolinguistic Programming [mh] OR remedial teaching [mh] OR vocational guidance [mh]) OR (prevent\* [TI] OR intervent\* [TI] OR program\* [TI] OR training\* [TI] OR counsel\* [TI] OR mentor\* [TI] OR tutor\* [TI] OR instruction\* [TI] OR psychotherapy [TI] OR psychotherapies [TI]) OR ((treatment\* [TI] OR therapy [TI] OR therapies [TI] OR "treatment effects" [TIAB] OR "therapy effects" [TIAB]) AND ((Psychology [sh] OR Education [sh] OR Psychology, Applied [mh] OR Psychology, Educational [mh]) NOT (general surgery [mh] OR surgical procedures [mh] OR "controlled substances" [mh]))) OR ("prevention effects" [TIAB] OR "intervention effects" [TIAB] OR "program effects" [TIAB] OR "training effects" [TIAB] OR "counseling effects" [TIAB] OR "tutoring

effects" [TIAB] OR "mentoring effects" [TIAB] OR "instruction effects" [TIAB] OR "psychotherapy effects" [TIAB]))

AND

((Infant [mh] OR Child [mh] OR Minors [mh] OR Adolescent [mh] OR Students [mh:noexp] OR Student Dropouts [mh] OR family [mh:noexp] OR parent-child relations [mh] OR parenting [mh] OR only child [mh]) OR (neonat\* [TI] OR neonates [TIAB] OR baby [TI] OR babies [TIAB] OR infant [TI] OR infants [TIAB] OR child [TI] OR childhood [TI] OR (children [TIAB] NOT "adult children" [TIAB]) OR kid [TI] OR kids [TIAB] OR toddler [TI] OR toddlers [TIAB] OR boy [TI] OR boys [TIAB] OR girl [TI] OR girls [TIAB] OR adolescen\* [TI] OR adolescents [TIAB] OR teen\* [TI] OR teens [TIAB] OR teenagers [TIAB] OR youth\* [TI] OR youths [TIAB] OR minors [TIAB] OR preschool\* [TI] OR preschoolers [TIAB] OR ((pupil [TI] OR pupils [TIAB]) NOT ("eye" [TIAB] OR "eyes" [TIAB])) OR school student\* [TI] OR "school students" [TIAB] OR grade student\* [TI] OR "grade students" [TIAB] OR grader\* [TIAB] OR graders [TIAB] OR family [TI] OR families [TI] OR parent\* [TI] OR mother\* [TI] OR father\* [TI]))

AND

((Psychology, Experimental [mh] OR Control Groups [mh] OR Random Allocation [mh] OR Matched-Pair Analysis [mh]) OR (experiment [TIAB] OR experiments [TIAB] OR experimental\* [TIAB] OR "between group design" [TIAB] OR contrast group\* [TIAB] OR control condition\* [TIAB] OR comparison condition\* [TIAB] OR control group\* [TIAB] OR comparison group\* [TIAB] OR "control students" [TIAB] OR "comparison students" [TIAB] OR "control pupils" [TIAB] OR "comparison pupils" [TIAB] OR control class\* [TIAB] OR comparison class\* [TIAB] OR control school\* [TIAB] OR comparison school\* [TIAB] OR "control families" [TIAB] OR "comparison families" [TIAB] OR "control children" [TIAB] OR "comparison children" [TIAB] OR "control parents" [TIAB] OR "comparison parents" [TIAB] OR "control mothers" [TIAB] OR "comparison mothers" [TIAB] OR "control fathers" [TIAB] OR "comparison fathers" [TIAB] OR "controls" [TIAB] OR wait-list [TIAB] OR non-intervention [TIAB] OR non-treatment [TIAB] OR random\* [TIAB] OR trial [TIAB] OR assign\* [TIAB] OR "program participation" [TIAB] OR "program participants" [TIAB] OR "intervention participation" [TIAB] OR "intervention participants" [TIAB]))

*The Cochrane Library Suchbefehl*

((migrant\* OR migrat\* OR immigr\* OR "non-citi\*" OR "non-nativ\*" OR allocht\* OR foreigner\* OR (("born" OR birth\* OR origin\*) NEAR/8 (count\* OR foreign\* OR "outside"))) OR (cultur\* AND (foreign\* OR differen\* OR divers\* OR origin\* OR "other")) OR multicultur\* OR ethnic\* OR ("minority group\*" NOT "sexual minority") OR refugee\* OR "English language learner\*" OR bilingual\* OR "as a second language" OR Latino OR Latina OR "Latin Americans" OR "Mexican American\*" OR Hispanic OR "Asian American\*" OR "Chinese American\*" OR "Vietnamese American\*" OR "Filipino American\*" OR "Indian American\*" OR "Korean American\*"):ti,ab,kw)

AND

((prevent\* OR intervent\* OR program\* OR training OR counsel\* OR mentor\* OR tutor\* OR instruction OR treatment OR therapy):ti,kw OR (("prevention" OR "intervention" OR "program" OR "training" OR counsel\* OR mentor\* OR tutor\* OR "instruction" OR "treatment" OR "therapy" OR "psychotherapy") NEAR/8 (evaluat\* OR "effects" OR "efficacy" OR "effectiveness" OR "efficiency" OR influenc\* OR "improving" OR improve\* OR increas\* OR decreas\*)):ab)

AND

((neonat\* OR baby OR infant OR (child NOT "adult children") OR childhood OR kid OR toddler OR boy OR girl OR adolescen\* OR teen\* OR youth\* OR "minors" OR preschool\* OR pupil OR ((school OR grade) NEAR/8 student) OR grader OR family OR parent\* OR mother\* OR father\*):ti,kw OR ("neonates" OR "babies" OR "infants" OR ("children" NOT "adult children") OR "kids" OR "toddlers" OR "boys" OR "girls" OR "adolescents" OR "teens" OR "teenagers" OR "youths" OR "minors" OR "preschoolers" OR "pupils" OR ((school OR grade) NEAR/8 students) OR "graders"):ab)

AND

((experiment OR experimental\* OR "between group design" OR "contrast group\*" OR (("control" OR "comparison") NEAR/4 condition) OR (("control" OR "comparison") NEXT/1 (group\* OR "students" OR "pupils" OR class\* OR school\* OR "families" OR "children" OR "parents" OR "mothers" OR "fathers"))) OR "controls" OR (("wait-list" OR "non-intervention" OR "non-treatment") NEAR/4 (control OR condition OR group OR "participants")) OR random\* OR trial OR assign\* OR (("program" OR "intervention") NEXT/1 ("participation" OR "participants"))):ti,ab,kw))

*Sowiport (Datenbanken: Sociological Abstracts, Social Services Abstracts)*

Folgende vier mit UND verknüpfte Suchfelder nach Schlagwörtern durchsucht

Suchfeld 1: migrant\* migrat\* immigr\* allochth\* foreigner\* "country of origin" "born outside" multicultur\* ethnic\* "minority group" "minority groups" "culturally diverse" "cultural diversity" "cultural differences" refugee refugees "English language learners" bilingual\* "as a second language" Latino\* Latina\* Hispanic\* "Latin Americans" "Asian American" "Asian Americans" "Chinese American" "Chinese Americans" "Vietnamese American" "Vietnamese Americans" "Filipino American" "Filipino Americans" "Indian American" "Indian Americans" "Korean American" "Korean Americans"

Suchfeld 2: prevent\* intervent\* program\* training counsel\* mentor\* tutor\* instruction instructions treatment treatments therapy therapies psychotherapy psychotherapies

Suchfeld 3: "neonates" "babies" "infants" "children" "kids" "youths" "toddlers" "boys" "girls" "adolescents" "teens" "teenagers" "minors" "preschoolers" "pupils" "school students" "grade students" "graders"

Suchfeld 4: experiment experiments experimental\* "between group design" "contrast group" "contrast groups" "control condition" "comparison condition" "control conditions" "comparison conditions" "control group" "control groups" "comparison group" "comparison groups" "control students" "comparison students" "control pupils" "comparison pupils" "control class" "comparison class" "control classes" "comparison classes" "control classroom" "comparison classroom" "control classrooms" "comparison classrooms" "control school" "comparison school" "control schools" " comparison schools" "control families" "comparison families" "control parents" "comparison parents" "control mothers" "comparison mothers" "control fathers" "comparison fathers" "wait-list" "non-intervention" "non-treatment" random\* trial trials assign\* "program participation" "intervention participation" "program participants" "intervention participants"

*Web of Science Suchbefehl*

((TS=(migrant\* OR migrat\* OR immigr\* OR "non-citi\*" OR "non-nativ\*" OR allocht\* OR foreigner\* OR (("born" OR birth\* OR origin\*) NEAR/8 (countr\* OR foreign\* OR "outside"))) OR (cultur\* NEAR/8 (foreign\* OR differen\* OR divers\* OR origin\* OR "other"))) OR multicultur\* OR ethnic\* OR ("minority group\*" NOT "sexual minority") OR refugee OR "English language learner\$" OR bilingual\* OR "as a second language"



OR Latino OR Latina OR "Latin Americans" OR "Mexican American\$" OR Hispanic OR "Asian American\$" OR "Chinese American\$" OR "Vietnamese American\$" OR "Filipino American\$" OR "Indian American\$" OR "Korean American\$"))

AND

(TI=(prevent\* OR intervent\* OR program\* OR training OR counsel\* OR mentor\* OR tutor\* OR instruction OR treatment OR therapy OR psychotherapy) OR AK=(prevent\* OR intervent\* OR program\* OR training OR counsel\* OR mentor\* OR tutor\* OR instruction OR treatment OR therapy OR psychotherapy) OR KP=(prevent\* OR intervent\* OR program\* OR training OR counsel\* OR mentor\* OR tutor\* OR instruction OR treatment OR therapy OR psychotherapy) OR TS=("prevention" OR "intervention" OR "program" OR "training" OR counsel\* OR mentor\* OR tutor\* OR "instruction" OR "treatment" OR "therapy" OR "psychotherapy") NEAR/8 (evaluat\* OR "effects" OR "efficacy" OR "effectiveness" OR "efficiency" OR influenc\* OR "improving" OR improve\* OR increas\* OR decreas\*))

AND

(TI=(neonat\* OR baby OR infant OR (child NOT "adult children") OR childhood OR kid OR toddler OR boy OR girl OR adolescen\* OR teen\* OR youth\* OR "minors" OR preschool\* OR pupil OR ((school OR grade) NEAR/8 student) OR grader OR family OR parent\* OR mother\* OR father\*) OR AK=(neonat\* OR baby OR infant OR (child NOT "adult children") OR childhood OR kid OR toddler OR boy OR girl OR adolescen\* OR teen\* OR youth\* OR "minors" OR preschool\* OR pupil OR ((school OR grade) NEAR/8 student) OR grader OR family OR parent\* OR mother\* OR father\*) OR KP=(neonat\* OR baby OR infant OR (child NOT "adult children") OR childhood OR kid OR toddler OR boy OR girl OR adolescen\* OR teen\* OR youth\* OR "minors" OR preschool\* OR pupil OR ((school OR grade) NEAR/8 student) OR grader OR family OR parent\* OR mother\* OR father\*) OR TS=("neonates" OR "babies" OR "infants" OR ("children" NOT "adult children") OR "kids" OR "toddlers" OR "boys" OR "girls" OR "adolescents" OR "teens" OR "teenagers" OR "youths" OR "minors" OR "preschoolers" OR "pupils" OR ((school OR grade) NEAR/8 "students") OR "graders"))

AND

(TS=(experiment OR experimental\* OR "between group design" OR "contrast group\*" OR (("control" OR "comparison") NEAR/4 condition) OR "control group\*" OR "comparison group\*" OR "control students" OR "comparison students" OR "control

pupils" OR "comparison pupils" OR "control class\*" OR "comparison class\*" OR "control school\*" OR "comparison school\*" OR "control families" OR "comparison families" OR "control children" OR "comparison children" OR "control parents" OR "comparison parents" OR "control mothers" OR "comparison mothers" OR "control fathers" OR "comparison fathers" OR "controls" OR (("wait-list" OR "non-intervention" OR "non-treatment") NEAR/4 (control OR condition OR group OR "participants")) OR random\* OR trial OR assign\* OR "program participation" OR "program participants" OR "intervention participation" OR "intervention participants"))))

AND

(SU=("Social Sciences" OR "Area Studies" OR "Biomedical Social Sciences" OR "Business & Economics" OR "Communication" OR "Criminology & Penology" OR "Cultural Studies" OR "Demography" OR "Education & Educational Research" OR "Ethnic Studies" OR "Family Studies" OR "Government & Law" OR "International Relations" OR "Linguistics" OR "Psychology" OR "Public Administration" OR "Social Issues" OR "Social Sciences Other Topics" OR "Social Work" OR "Sociology" OR "Urban Studies" OR "Women's Studies") OR WC=("Area Studies" OR "Asian Studies" OR "Behavioral Sciences" OR "Business" OR "Business, Finance" OR "Communication" OR "Computer Science, Information Systems" OR "Computer Science, Interdisciplinary Applications" OR "Cultural Studies" OR "Dance" OR "Demography" OR "Developmental Biology" OR "Economics" OR "Education & Educational Research" OR "Education, Scientific Disciplines" OR "Education, Special" OR "Engineering, Environmental" OR "Engineering, Multidisciplinary" OR "Environmental Sciences" OR "Environmental Studies" OR "Ergonomics" OR "Ethnic Studies" OR "Family Studies" OR "Film, Radio, Television" OR "Geography" OR "Geosciences, Multidisciplinary" OR "Health Care Sciences & Services" OR "Health Policy & Services" OR "Hospitality, Leisure, Sport & Tourism" OR "Humanities, Multidisciplinary" OR "Information Science & Library Science" OR "International Relations" OR "Language & Linguistics" OR "Linguistics" OR "Literature" OR "Management" OR "Multidisciplinary Sciences" OR "Music" OR "Nursing" OR "Operations Research & Management Science" OR "Pediatrics" OR "Political Science" OR "Psychiatry" OR "Psychology" OR "Psychology, Applied" OR "Psychology, Biological" OR "Psychology, Clinical" OR "Psychology, Developmental" OR "Psychology, Educational" OR "Psychology, Experimental" OR "Psychology, Mathematical" OR "Psychology, Multidisciplinary" OR "Psychology,

Psychoanalysis" OR "Psychology, Social" OR "Public Administration" OR "Public, Environmental & Occupational Health" OR "Rehabilitation" OR "Religion" OR "Social Issues" OR "Social Sciences, Biomedical" OR "Social Sciences, Interdisciplinary" OR "Social Sciences, Mathematical Methods" OR "Social Work" OR "Sociology" OR "Sport Sciences" OR "Statistics & Probability" OR "Substance Abuse" OR "Telecommunications" OR "Theater" OR "Transportation" OR "Urban Studies" OR "Women's Studies"))

## **Anhang B – Kodierschema**

Das Kodierschema wurde im Rahmen des Projektes „Wirksamkeit von Bildungs- und Interventionsprogrammen zur Prävention und Kompensation von Armuts- und Migrationsfolgen bei Kindern und Jugendlichen. Ein Forschungsüberblick und eine Meta-Analyse internationaler Evaluationsforschung“ (gefördert vom BMBF; Förderkennzeichen: 01JC1105; Projektleiter: Prof. Dr. Andreas Beelmann) erstellt und ist auf den folgenden Seiten im vollständigen englischsprachigen Original abgedruckt.



The effectiveness of psychosocial and educational programs for preventing or compensating the negative consequences in the development of children and adolescents caused by poverty or migration: A multinational meta-analysis.

## Coding Handbook

### General guidelines (please complete continuously)

- **Discussion:** Discuss all unclear codings within the team.
- **Optimal coding:** Always use the best estimator for the coding variables. If necessary also code on the basis of plausibility (e.g. if normal school classes are examined and there are no statements regarding gender, code 50:50).
- **Missings (not reported):** Only use “not reported” (-99) when it is unavoidable. Coding -99 is a strong statement and not an ambiguity rating. In case of insecurity, discuss variable in the weekly discussion group.
- **Missings (blank):** Please leave coding place blank, if coding does not apply.
- **Decimal places:** Concerning all demographic numeric variables: Assess two decimal places.

### Specific coding decisions

- **Booster Sessions:** Booster sessions are not included into the core-length of a program.
- **Several control groups:** If the author(s) use(s) more than one control group, discuss during team meeting which one to use (most suitable).
- **Several intervention groups:** If more than one intervention group has been assessed, code comparisons for each intervention, e.g. IG<sub>1</sub> vs. CG and IG<sub>2</sub> vs. CG. Exception: If results are only being reported for the whole sample AND the different programs can be subsumed to one group (e.g. early education programs).
- **Executor(s):** The executor(s) of the program is/are always the ones who first conducted it. This means that for example parents, who have learned a reading strategy from a coach and intensively taught the same to their children, are not executors (but rather the users of the program). If the parents are asked to apply the program at home (without any observation/control by the trainer), code this as homework (method). Don't take the homework time into account when coding the intensity, except when the trainer is present or the parents keep record.
- **Whole versus subsample:** If data is reported for both, the whole sample and the subgroups (e.g. regarding gender, age etc.), code only the data for the whole sample. This does not apply, if only certain subgroups are relevant for the question (e.g. separate listing of migrant groups).
- **Sample doubling:** Avoid coding the same sample more than once. If a publication reports results, for example, for the whole sample and subsamples, choose the sample for which more information are given or which better suits the inclusion criteria, respectively.
- **Teenage parents:** If the sample consists of teenage parents, define and code the teenage parents as parents and not as adolescents.
- **Poverty and migration sample:** If the whole sample of a publication suits the inclusion criteria for poverty/migration, use this sample, even if formed subgroups suit the criteria even better (e.g. for migration: ethnic background whole sample goes over first generation background subsamples; for poverty 75% poverty whole sample goes over 90% poverty subsamples).
- **Poverty versus/and migration:** If the sample meets migration AND poverty criteria, code the publication once and use a “B” at the end of the publication code. If a publication meets poverty and migration criteria, but relating to different samples (e.g. migration: whole sample, poverty: subsample), the publication needs to be coded twice (with an “A” or “M” at the end of the publication code, respectively).
- **IG-subsample gets an additional program component:** If some of the program components are only conducted with a subsample of the intervention group (e.g. a special component conducted only with high risk group), treat the special component as part of the program (i.e. when coding the program intensity) if the high

risk group is the majority of the sample ( $\geq 50\%$ ). If the high risk group is the minority, the special component is considered as an individualization component (i.e. not taken into account when coding the program intensity)

- **Post assessment time point:** If the outcome assessment took place at the end of a program (and not after), code a negative time point, e.g. if the program started in February and lasted until 1<sup>st</sup> of September and the outcome assessment period was in August, code -0,5 for the average post assessment time point.
- **Partial randomization:** If the majority of the posttest sample ( $> 50\%$ ) is non-randomly assigned to the groups, the study is to be excluded. This can be the case if for example a randomization procedure is only applied to one of three schools or if participants join the program after randomization and within the posttest sample these outnumber the participants originally randomized.
- **Program type: Strategies for coding difficulties on comparison level concerning teacher programs and parenting programs**
  1. Try to determine **one main** program type (**main** addressee of the program)
    - Teachers are rather executors than addressees
    - Parents are rather addressees than executors
    - **Coding consequences on comparison level:** Only code the program information of the main addressee (defined by program type). If for example a parent-as-teacher training is implemented and a small training component for the child itself is being conducted, consider these information when coding the addressee and content variable, select “parent training” as type of program and neglect the child training information on the variables following “program type”
  2. Coding **more than one** is an exception
    - Choose only several program types, if the program contains several equally intensive and important components, such as stand-alone child training and parent training
    - **Coding consequences on comparison level:** All the program information of the different program types is to be coded. If the different components would result in different codings (e.g. participation is high in child component and low in parent component), average the ratings (e.g. middle)

#### Effect size commentaries

- **Basic rule:** Positive effect sizes indicate that the intervention group did better than the control group.
- **Multiple measurements:** In case of multiple measurements over time, ignore intermediate measurements during the program. The earliest evaluation time point is directly after the program. Exception: If the program continues beyond the published study, the last available measurement should be interpreted as post-test.
- **Decimal places:** Assess three decimal places for effect sizes.
- **Uncertainty calculation:** In case of uncertainty of back calculation, please discuss in the team.
- **Detailed coding:** During effect size coding, always use the most accurate/detailed information (i.e. always use subscales instead of total test. Exception: If all ES are within the same outcome category)
- **SD versus SE:** Always use standard deviations to calculate effect sizes (Streuung, Standardabweichung; SD) and not standard errors (Standardschätzfehler; SE). Convert if necessary.

- **Cost-effectiveness:** Leave the effect size estimation for cost-effectiveness measures blank.
- **Follow-Up:** Always code follow-up effects, even if the sample falls under a sample size of 25 per group. The follow-up outcome needs to fit to the post outcome, otherwise code it as independent outcome with a late post assessment time point.
- **Structural equation model:** No effect size estimation is possible, because the whole model has been tested and not the program-control group comparison only.
- **Regressions:** The regression coefficient  $\beta$  of linear regressions can be interpreted as estimation for the effect size, if only the predictor(s) “group” (and “pre-test”) is(are) included. Concerning multiple regressions, only interpret the  $\beta$  coefficient if a hierarchical regression was calculated and if group is last included predictor.
- **Further formulas** (see Excel-Sheet “Effektstärkeberechnung”): Combining subgroups’ standard deviation, calculating F-value or t-value from p-value, calculating standard deviations from standard errors.
- **Relation of  $\eta^2$  and  $r^2$ :**  $\eta^2$  (variance analysis) =  $r^2$  (regression analysis)
- **Effect sizes of clusters:** If the effect is measured on cluster level, e.g. teacher behavior, classroom climate, code the effects even though the sample size will be most likely under  $n = 25$  (the minimum number of clusters per group still needs to be  $n = 5$ , see inclusion criteria). Make a note in the commentary section for the affected effect size.

#### Technical information or using the coding sheet

- **Variable type** /: This type is a dummy variable. Check the ones that apply. Multiple selections within variable are allowed.
- **Hierarchy:** The number written in brackets behind a variable serves as element to structure the next variable, e.g. if program orientation’s dummy variable child/adolescent promotion (1) is checked, then at least one of the program aim variables marked with a “1” (prior to the variable) needs to be checked, too. Also applies for the addressee and program content/focus.

#### File Maker information

- **Decimal numbers:** To insert decimal numbers, use a comma and not a dot.
- **ID change:** If a correction of the Publication ID on the comparison level is necessary, don’t forget to change it on study und publication ID, as well.

#### Abbreviations

ES = effect size, M = mean, SD = standard deviation, N = sample size



## DOCUMENT LEVEL

VARIABLE	CODING VALUES, DESCRIPTIONS AND NOTES	CODE
<b>Publication ID</b>	First four letters of the first authors' last name + first letter of the second and third authors last names + last two digits of the publication year + A/M/B for the study type (poverty/migration/both), e.g. poverty study from <b>McClanahan</b> , Wendy S., <b>Sipe</b> , Cynthia L., <b>Smith</b> , Thomas J. (2004) → McCISS04A Note: Omit letters when there are fewer than four authors, e.g. Ramey, Craig T. & Haskins, Ron (1981) → RameH81A	ID_Pub
<b>Date coded</b>	Present date (DD.MM.JJJJ)	date
<b>Coder</b>	Name of coder	coder
<b>Author(s)</b>	Write out the first six authors, then use "et al."; omit (initials of) surnames	authors
<b>Title</b>	Main title only	title
<b>Publication year</b>	JJJJ	pub_year
<b>Publication format</b>	1 = <b>Peer-reviewed journal article</b> (published; article published in a peer reviewed journal) 2 = <b>Journal article with unknown peer-review status</b> (published, article published in a journal whose peer review status is unknown) 3 = <b>Published scientific report</b> (published; scientific report that didn't undergo comprehensive independent peer review before publication; e.g. publication of a ministry) 4 = <b>Unpublished scientific report</b> (unpublished; any scientific work that is not publicly available (e.g. via internet or the ThULB) and only obtainable by contacting the author) 5 = <b>Doctoral thesis</b> (unpublished; Thesis for obtaining a doctor's degree) 6 = <b>Conference proceedings</b> (published; collection of academic papers published in the context of an academic conference) 7 = <b>Book</b> (published; scientific report published in form of a book) 8 = <b>Book chapter</b> (published; any scientific report published in a book) See appendix A for examples of peer-/not peer-reviewed journals	pub_form
<b>Name of journal</b>	Fill out only, if "Publication format" = 1 or 2, otherwise leave blank	journal
<b>Publication language</b>	1 = English 2 = German 3 = Other	lang

<b>Publication language: other</b>	Fill out only if “Publication language” = 3, otherwise leave blank	lang_o
<b>Country of performance</b>	1 = USA 2 = UK 3 = Canada 4 = Netherlands 5 = Germany 6 = Belgium 7 = Sweden 8 = Norway 9 = Other Country where the study took place; if no other information available, use first authors’ address	count_r
<b>Country of performance: other</b>	Fill out only if “Country of performance” = 9, otherwise leave blank	count_r_o
<b>Additional literature</b>	0 = <b>Only information from original study was used for coding the publication</b> (no additional literature was used) 1 = <b>Supplemental duplicate(s)</b> was/were used for coding (i.e. an eligible study referring to the same analysis and content as the original study, but contains additional information, e.g. further outcomes) 2 = <b>External publication(s)</b> was/were used for coding (i.e. a publication that did not meet the inclusion criteria, like a publication of the program curriculum or the sample description) 3 = <b>Both:</b> supplemental and external publications were used for coding. See appendix B for more information concerning literature types	add_lit
<b>Name supplemental duplicate(s)</b>	Fill out only if “Additional literature” = 1 or 3, otherwise leave blank Note complete citation.	sup_dub_tx
<b>Name external publication(s)</b>	Fill out only if “Additional literature” = 2 or 3, otherwise leave blank Note complete citation.	ext_pub_tx
<b>Study ID</b>	Number of study	ID_Stud

## COMPARISON LEVEL

### INFORMATION ABOUT THE CHARACTERISTICS OF THE PROGRAM

VARIABLE	CODING VALUES, DESCRIPTIONS AND NOTES	CODE
Publication ID		ID_Pub
Study ID		ID_Stud
Comparison ID	Number of comparison; if only subgroups are being analyzed, treat them as different comparisons	ID_Comp
<b>BASIC PROGRAM INFORMATION</b>		
Program name		pr_name
Program orientation <i>Who is supposed to benefit from the program according to the author?</i>	<input type="checkbox"/> /☒ <b>Child/Adolescent promotion</b> (1)	pr_or_ch
	<input type="checkbox"/> /☒ <b>Parent Training/Family promotion</b> (2)	pr_or_pf
	<input type="checkbox"/> /☒ <b>Environmental change</b> (a context outside the family home is changed) (3)	pr_or_en
Program aim (PA) <i>What is the aim of the program according to the author?</i>	Take into account all aims that the author <b>explicitly</b> mentions concerning the conducted program (with formulations like “the program intends to”, “aims at”, “wants to”, “focuses on”, we [the authors] examined" or "the purpose of the study..." and the like). If the author didn't explicitly mention a goal for the specific program that was conducted, BUT did formulate directed hypotheses, assess these as aims. If the author neither did explicitly mention a goal for the specific program that was conducted nor formulated directed hypotheses, BUT reports aims of a broader strategy (e.g. in a study about the effects of the parenting program “Familias Unidas” the author mentions that parenting programs in general are designed to...), check “not explicitly mentioned” and assess these aims. If problem behavior in general is reported, check the box for preventing internalizing and for externalizing problems, as well.	
	<input type="checkbox"/> /☒ <b>Not explicitly mentioned</b>	aim_nem
	1 <input type="checkbox"/> /☒ <b>Language promotion</b> (Language of the study country, the home country or bilingualism, e.g. reading or writing competencies, telling stories)	aim_c_lang
	1 <input type="checkbox"/> /☒ <b>Cognitive promotion</b> (e.g. (meta)cognitive competencies, promote inductive reasoning, decision making)	aim_c_cog
	1 <input type="checkbox"/> /☒ <b>School success improvement</b> (e.g. school performance, school dropout, higher education)	aim_c_scsu
1 <input type="checkbox"/> /☒ <b>(School) motivation enhancement</b> (e.g. school bonding, school attendance, college enrollment, mastery motivation, participation)	aim_c_scmo	

	1	<input type="checkbox"/> /☒ <b>Promotion of social/social-cognitive competence</b> (e.g. social skills, empathy, moral reasoning, social-problem solving, self-control, adaptive functioning)	aim_c_soc
	1	<input type="checkbox"/> /☒ <b>Emotion regulation enhancement</b>	aim_c_emo
	1	<input type="checkbox"/> /☒ <b>Promotion of self-esteem</b>	aim_c_se
	1	<input type="checkbox"/> /☒ <b>Ethnical identity</b> (sense of belonging)/inter-group relation support	aim_c_ethn
	1	<input type="checkbox"/> /☒ <b>Prevention of externalizing problems</b> (e.g. anti-social behavior, violence, aggressiveness)	aim_c_ext
	1	<input type="checkbox"/> /☒ <b>Prevention of internalizing problems</b> (e.g. depression, anxiety, stress)	aim_c_int
	1	<input type="checkbox"/> /☒ <b>Pregnancy/sexual prevention</b>	aim_c_sex
	1	<input type="checkbox"/> /☒ <b>Substance abuse prevention</b>	aim_c_suab
	1	<input type="checkbox"/> /☒ <b>Job training/professional success</b>	aim_c_job
	1	<input type="checkbox"/> /☒ <b>General developmental support</b> (generally rather applies to young aged children, school readiness, motor skills)	aim_c_gd
	1	<input type="checkbox"/> /☒ <b>Other child/adolescent related aim (1)</b>	aim_c_o
	2	<input type="checkbox"/> /☒ <b>Child care improvement</b> (meeting physical needs of the child, like feeding competence)	aim_p_chca
	2	<input type="checkbox"/> /☒ <b>Parenting skill enhancement</b> (e.g. constructive parenting behavior, parent-child communication)	aim_p_psk
	2	<input type="checkbox"/> /☒ <b>Reduction of violence in child-rearing</b>	aim_p_vio
	2	<input type="checkbox"/> /☒ <b>Parent-child relationship improvement</b> (e.g. warmth, bonding, parent involvement/investment, attachment)	aim_p_crel
	2	<input type="checkbox"/> /☒ <b>Family communication/family relationship modification</b> (e.g. communication style of the whole family)	aim_p_fcom
	2	<input type="checkbox"/> /☒ <b>Social support/social network enlargement of parents/family</b>	aim_p_sosu
	2	<input type="checkbox"/> /☒ <b>Academic/cognitive competence promotion</b> (e.g. parent as teacher)	aim_p_accu
	2	<input type="checkbox"/> /☒ <b>Parents' language competence promotion</b>	aim_p_lang
	2	<input type="checkbox"/> /☒ <b>Parental professional success</b>	aim_p_prof
	2	<input type="checkbox"/> /☒ <b>Promoting the process of acculturation</b>	aim_p_accu
	2	<input type="checkbox"/> /☒ <b>Other parent/family related aim (2)</b>	aim_p_o
	3	<input type="checkbox"/> /☒ <b>Teacher career-related skills promotion</b>	aim_e_tsk
	3	<input type="checkbox"/> /☒ <b>Peer relation modification/social network within classes/school</b> (e.g. classroom climate)	aim_e_peer
	3	<input type="checkbox"/> /☒ <b>Neighborhood climate/network/contact enhancement</b>	aim_e_neigh
	3	<input type="checkbox"/> /☒ <b>Curricular changes/school type change/special school lessons</b>	aim_e_curr
	3	<input type="checkbox"/> /☒ <b>Other environmental related aim (3)</b>	aim_e_o
<b>PA: other 1</b>		Fill out only if "Program aim" = Other child/adolescent related aim, otherwise leave blank	aim_c_otxt
<b>PA: other 2</b>		Fill out only if "Program aim" = Other parent/family related aim, otherwise leave blank	aim_p_otxt
<b>PA: other 3</b>		Fill out only if "Program aim" = Other environmental related aim, otherwise leave blank	aim_e_otxt

<b>Addressee</b> <i>Who was the program subject?</i>	<input type="checkbox"/> /☒ <b>Child/Adolescent related content</b>	addr_ch
	<input type="checkbox"/> /☒ <b>Parent Training/Family content</b>	addr_pf
	<input type="checkbox"/> <b>mother</b> (primarily mothers in the sample, i.e. more than 70% of the sample) <input type="checkbox"/> <b>father</b> (primarily fathers in the sample, i.e. more than 70% of the sample) <input type="checkbox"/> <b>both</b> (between 31% and 69% mothers/fathers in the sample or both parents are involved in the program; also other caregivers, like grandparents)	addr_p_spec
	<input type="checkbox"/> /☒ <b>Environmental change</b> (e.g. teacher promotion, institutional/system change)	addr_en
<b>Program content (PC)</b> <i>Assess the content of the program or its components, respectively.</i>	1 <input type="checkbox"/> /☒ <b>Language content</b> (In the program linguistic and meta-linguistic skills are trained, which means improving the ability of producing, recognizing, understanding, interpreting or translating written or spoken language (e.g. reading, writing, vocabulary or speech programs).)	co_c_lang
	1 <input type="checkbox"/> /☒ <b>Learn or cognitive content</b> (In the program participants work on their learning style and learning success itself. Included are also programs which focus on the meta/-cognitive development or the achievement motivation, e.g. math, science, school motivation, IQ, or replacement programs.)	co_c_lecog
	1 <input type="checkbox"/> /☒ <b>Anti-victimization content</b> (In the program anti-victimization strategies (psychological and physiological) are being trained and the consequences caused by victimization, like traumatization, are being broad up and worked on, e.g. child abuse program.)	co_c_vict
	1 <input type="checkbox"/> /☒ <b>Socio-emotional competence content</b> (In the program life skills and socio-emotional competences, such as strategies to keep up a healthy psycho hygiene and functional social relations are trained, e.g. conduct problems, social competence, emotional well-being, self-esteem.)	co_c_soce
	1 <input type="checkbox"/> /☒ <b>Substance abuse content</b> (In the program the reduction of the consumption of illegal substances is trained, e.g. alcohol or smoking program.)	co_c_suab
	1 <input type="checkbox"/> /☒ <b>Sexual behavior content</b> (In the program the modification of sexual behavior is being picked out as a central theme, e.g. births control or sexual education program.)	co_c_sex
	1 <input type="checkbox"/> /☒ <b>Career content</b> (In the program participants' careers are modified, e.g. work program, career coaching)	co_c_car
	1 <input type="checkbox"/> /☒ <b>General development content</b> (In the program the child's general (early) development is being trained, e.g. school readiness, early education or motor skills programs.)	co_c_gd
	1 <input type="checkbox"/> /☒ <b>Other child/adolescent related content</b> (1)	co_c_o
	2 <input type="checkbox"/> /☒ <b>Parenting skill content</b> (In the program constructive and consistent parenting or communication behavior <u>between parents and their children</u> are trained, respectively.)	co_p_psk

	2	<input type="checkbox"/> /☒ <b>Family communication/relation content</b> (In the program generally functional family communications and the improvement of family relations are being discussed, e.g. working on establishing constructive communication patterns between parents/partners.)	co_p_fcom
	2	<input type="checkbox"/> /☒ <b>Parents' academic and cognitive content</b> (In the program parents' academic and cognitive abilities, including job related competences, are trained.)	co_p_accog
	2	<input type="checkbox"/> /☒ <b>Parents' language competence content</b> (In the program parental language abilities are trained.)	co_p_lang
	2	<input type="checkbox"/> /☒ <b>Parental social support content</b> (In the program parents learn to contact sources that support them, e.g. other parents, teachers, doctors.)	co_p_socsu
	2	<input type="checkbox"/> /☒ <b>Other parent/family related content</b> (2)	co_p_o
	3	<input type="checkbox"/> /☒ <b>Teacher training content</b> (In the program pedagogical skills of the teacher are trained, e.g. teacher knowledge, didactic abilities.)	co_e_tt
	3	<input type="checkbox"/> /☒ <b>Peer communication content</b> (In the program the interaction/relation between peers/classmates and the modification of social networks within classes/schools, including classroom climate, are worked on.)	co_e_peer
	3	<input type="checkbox"/> /☒ <b>Neighborhood enhancement content</b> (In the program the neighborhood's climate and the contact within/between neighborhoods are picked out as central themes, e.g. new community centre with family activities being offered, socialization groups)	co_e_neigh
	3	<input type="checkbox"/> /☒ <b>School structure content</b> (In the program the curriculum or school type of a school is changed, e.g. introducing special school lessons, day care program, offering additional schooling.)	co_e_scst
	3	<input type="checkbox"/> /☒ <b>Other environmental related content</b> (3)	co_e_o
<b>PC: other 1</b>	Fill out only if "Program content" = Other child/adolescent related content, otherwise leave blank		co_c_otxt
<b>PC: other 2</b>	Fill out only if "Program content" = Other parent/family related content, otherwise leave blank		co_p_otxt
<b>PC: other 3</b>	Fill out only if "Program content" = Other environmental related content, otherwise leave blank		co_e_otxt

<b>Program type</b> <i>What type/kind of program has been used?</i>	<input type="checkbox"/> /☒ <b>Nursery-program</b> (A child around the age of one is the primary subject.)	prt_nurs
	<input type="checkbox"/> /☒ <b>Early education/prevention</b> (A child from about the age of one up to about the age of three is the primary subject.)	prt_eaed
	<input type="checkbox"/> /☒ <b>Child training program</b> (A child from about three years of age up to about the age of twelve is the primary subject.)	prt_chtp
	<input type="checkbox"/> /☒ <b>Adolescent training program</b> (An adolescent between thirteen and eighteen years of age is the primary subject.)	prt_adtp
	<input type="checkbox"/> /☒ <b>Preschool/School curriculum</b> (A curricular program for children from three years of age.)	prt_prsc
	<input type="checkbox"/> /☒ <b>Mentoring program</b> (e.g. Big brother/big sister program, tutoring)	prt_ment
	<input type="checkbox"/> /☒ <b>Parent training</b> (including parent-as-teacher)	prt_pt
	<input type="checkbox"/> /☒ <b>Family counseling</b>	prt_fcou
	<input type="checkbox"/> /☒ <b>Psychotherapeutic prevention</b>	prt_psynt
	<input type="checkbox"/> /☒ <b>Teacher training</b>	prt_tt
	<input type="checkbox"/> /☒ <b>Schooling concept</b> (e.g. special school type, establish a school)	prt_scco
	<input type="checkbox"/> /☒ <b>Transition program</b> (e.g. change of school, relocation)	prt_tran
	<input type="checkbox"/> /☒ <b>Job program</b> (e.g. job support or training)	prt_jobp
<input type="checkbox"/> /☒ <b>Other</b>	prt_o_c	
<b>Program type: other</b>	Fill out only if “Program type” = Other, otherwise leave blank	prt_o_tx
<b>TECHNICAL PROGRAM INFORMATION</b>		
<b>Theoretical foundation</b> <i>To which theoretical framework does the author primarily refer to (Why and how should the program work?)</i>	<p>1 = <b>General developmental theory</b>: Child’s <i>general</i> development (like motor development) is described by stages through which children typically progress (e.g. Erickson, Flavell, Griffith)</p> <p>2 = <b>Cognitive developmental theory</b>: Child’s <i>cognitive</i> development is described by stages through which children typically progress (e.g. Piaget)</p> <p>3 = <b>Social learning theory</b>: Underlines the importance of social and cultural factors in the child’s development (e.g. Bandura, Vygotsky, Forgatch)</p> <p>4 = <b>Cognitive-behavior theory</b>: Based on the premise that a cognitive change results in a behavioral change (e.g. Meichenbaum)</p> <p>5 = <b>Ecological-systemic theory</b>: the development of a child is influenced by factors within itself and by environmental influences (e.g. Bronfenbrenner, Brandstädter, Learner, Sameroff &amp; Chandler, Bertalanffy)</p> <p>6 = <b>Poverty/Migration specific theory</b></p> <p>7 = <b>Other</b> (also name theory if it can be assumed that a theory has been applied, however it is not well-known or labeled, e.g. theory on reading mastery)</p>	theo_fo

<b>Theoretical foundation poverty/migration: other</b>	Fill out only if “Theoretical foundation” = 6, otherwise leave blank Describe the theory and name the subgroup (poverty or migration)	theo_fo_pm
<b>Theoretical foundation: other</b>	Fill out only if “Theoretical foundation” = 7, otherwise leave blank	theo_fo_o
<b>Program method</b> <i>Which method(s) did the author use (explicitly reported)?</i>	<input type="checkbox"/> / <input checked="" type="checkbox"/> <b>Information transfer</b> (e.g. information leaflet, books, or instruction as intervention)	pm_info
	<input type="checkbox"/> / <input checked="" type="checkbox"/> <b>Psycho-education</b> (explaining to encourage understanding, rather than just giving information)	pm_psyed
	<input type="checkbox"/> / <input checked="" type="checkbox"/> <b>Incentives</b> (e.g. token system, vouchers, reward)	pm_incen
	<input type="checkbox"/> / <input checked="" type="checkbox"/> <b>Role play</b> (e.g. guided role plays)	pm_ropl
	<input type="checkbox"/> / <input checked="" type="checkbox"/> <b>Behavioral training with instruction</b> (modeling, rather behaviorally oriented)	pm_beht
	<input type="checkbox"/> / <input checked="" type="checkbox"/> <b>Video-feedback</b>	pm_vidfb
	<input type="checkbox"/> / <input checked="" type="checkbox"/> <b>Group discussion</b>	pm_grdis
	<input type="checkbox"/> / <input checked="" type="checkbox"/> <b>Exercises</b> (rather cognitively oriented e.g. problem task, repetition exercise, writing essays, also play-like exercised, e.g. number games, rhyming, songs)	pm_exerc
	<input type="checkbox"/> / <input checked="" type="checkbox"/> <b>Home work/keep a diary</b>	pm_howo
	<input type="checkbox"/> / <input checked="" type="checkbox"/> <b>Process diagnostics</b> (e.g. adapt the program to the participant, considering the child’s or parent’s abilities)	pm_pdiag
	<input type="checkbox"/> / <input checked="" type="checkbox"/> <b>Free or guided play</b> (e.g. in the context of play therapy or pedagogically guided, trading games)	pm_play
	<input type="checkbox"/> / <input checked="" type="checkbox"/> <b>Physical activity</b> (e.g. sport, dance, movement, yoga)	pm_phact
	<input type="checkbox"/> / <input checked="" type="checkbox"/> <b>Peer assistance</b> (e.g. peers as aid to executor, peer mediation)	pm_peerm
	<input type="checkbox"/> / <input checked="" type="checkbox"/> <b>Home visitation</b>	pm_hovis
	<input type="checkbox"/> / <input checked="" type="checkbox"/> <b>Multimedia</b> (e.g. videos, webpage)	pm_multm
	<input type="checkbox"/> / <input checked="" type="checkbox"/> <b>Enriching materials</b> (material placed at children’s or parents’ disposal, e.g. books or toys, handouts with age appropriate activities)	pm_enrm
	<input type="checkbox"/> / <input checked="" type="checkbox"/> <b>Performance</b> (e.g. through a film, theater, reading or learning together, building toys together, drawing/painting)	pm_perfo
	<input type="checkbox"/> / <input checked="" type="checkbox"/> <b>Internship/work sample</b> (e.g. job opportunity)	pm_insh
	<input type="checkbox"/> / <input checked="" type="checkbox"/> <b>Joint activities</b> (helps create a common context of experience between student and teacher or parents apart from everyday interactions, e.g. go swimming together, field trips)	pm_jact
	<input type="checkbox"/> / <input checked="" type="checkbox"/> <b>Networking</b> (e.g. mutual support of teenage mothers)	pm_netw
	<input type="checkbox"/> / <input checked="" type="checkbox"/> <b>Other</b>	pm_o_c



<b>Program method: other</b>	Fill out only if “Program method” = Other, otherwise leave blank	pm_o_tx
<b>Manualization</b> <i>Was a manual used for the program?</i>	<p>0 = <b>No</b></p> <p>1 = <b>Probably</b> (Whether a manual was used, has <u>not been explicitly mentioned</u>, but is very probable, e.g. a pre-version of a manual has been mentioned or a curriculum has been used)</p> <p>2 = <b>Yes</b> (The usage of a manual was <u>explicitly mentioned</u>.)</p> <p>Note: A curriculum rather consists of rough guidelines concerning the program, while a manual contains concrete information concerning the program, like contents, methods used or time planning.</p>	manual
<b>Structuredness</b> <i>To which degree is the program in terms of content and methodical-didactically determined?</i>	<p>1 = <b>Very low</b> (All open offering programs and programs with optional attendance, e.g. programs that use self-help materials (books/films))</p> <p>2 = <b>Low</b> (Largely free composition of content and program methods/didactics, at most largely the same number/duration (intensity) or regularity of the meetings/sessions/units)</p> <p>3 = <b>Medium</b> (Regularity of the contents and/or program methods/didactics, which go beyond the intensity/regularity in “low structuredness”, e.g. wide predetermination of topics without predetermining the exact procedure or sequence of the topics)</p> <p>4 = <b>High</b> (Detailed predetermination of contents and methods/didactics mostly realized through concrete manuals, e.g. exact procedure or sequence of the topics are predetermined)</p>	struct
<b>Individualization</b> <i>The degree to which the programs' content and methods/didactics are adaptable to individuals' needs or skills.</i>	<p>1 = <b>Very Low</b> (No recognizable individualization. Applies for example to programs which are exclusively implemented through films.)</p> <p>2 = <b>Low</b> (Everyone gets the same program, however it is likely that different skill levels/needs of the people were taken into account, e.g. if group sizes are low.)</p> <p>3 = <b>Middle</b> (Some adaptation of the contents/methods, like different strategies are proposed depending on the recipient or there is a specification based on group characteristics, e.g. gender, age.)</p> <p>4 = <b>High</b> (Individualizing as an explicit feature of the program, e.g. after systematic diagnostics certain contents are assigned, intervention components are administered based on skills/abilities/other personal traits.)</p>	individ
<b>Delivery format</b> <i>What is the format of delivery of the program?</i>	<p>1 = <b>Self-help</b></p> <p>2 = <b>Individual</b> (the program is being delivered one-by-one, e.g. individual counseling)</p> <p>3 = <b>Group</b> (the program is being delivered in a group setting, e.g. small group, entire classroom)</p> <p>4 = <b>Both</b> (individual and group delivery are being combined, e.g. a combination of individual mentoring and whole class teaching)</p>	del_form

<p><b>Level of participation</b> <i>The degree to which the participants take actively part in the program.</i></p>	<p>1 = <b>Low</b> (Program is primarily based on the frontal transfer of knowledge, e.g. only lecture is given or video/film shown)  2 = <b>Middle</b> (Some active participation is included, e.g. frontal transfer of knowledge combined with homework given, individually working on exercises)  3 = <b>High</b> (Participants take active part in the program by e.g. discussions, role play, behavioral training, or can influence the procedures of the intervention)</p>	l_partic
<p><b>Culturally tailored</b></p>	<p>0 = <b>No</b> (If no program characteristic is adapted to the participant's culture; also if only the measurement of the dependent variable is available in the demanded language)  1 = <b>Yes</b> (Author explicitly states, that the program is "Culturally tailored" or adapted to the ethnic/cultural/migrant group studied or the program addresses one of the following: Ethnic identity/identity conflict, cultural peculiarities or the acculturation gap between parents and child, official language of the host country as a foreign language or mother tongue of the immigrant group (if different from the host countries official language) or ethnic matching of the service delivery agent (mentor, tutor, trainer, counselor etc.) to the immigrant group); for example the executor speaks Spanish/English or program exercises are offered in English/Spanish depending on the participant's needs</p>	cult_tai
<p><b>Implementation control</b></p>	<p>Note: If not explicitly reported, check "no". nip = not in person; program has not been conducted in person, but by video tape or written information only.  The variables training, supervision, observance and additional services, strongly depend on how extensive the program itself was, i.e. for a simple and short program a one hour training/supervision/observance session might be totally sufficient (ergo: rate full).  For orientation: over 10% is considered as full.</p>	
	<p><input type="checkbox"/>nip/<input type="checkbox"/>no/<input type="checkbox"/>basic/<input type="checkbox"/>full      <b>Training of the executor</b> (<i>How intensive was the training of the executor?</i>)</p> <ul style="list-style-type: none"> <li>- <b>Basic:</b> The executor received some training, like basic instructions or an information sheet</li> <li>- <b>Full:</b> The executor received an extensive training for conducting the program, e.g. several training sessions with feedback, or already conducted the program in the past.</li> </ul> <p>Note: If the executor is the developer of the program and no further information are reported, than code "full".</p>	ic_train
	<p><input type="checkbox"/>nip/<input type="checkbox"/>no/<input type="checkbox"/>basic/<input type="checkbox"/>full      <b>Supervision during program</b> (<i>How intensive was the supervision during the program? Rate the quality of supervision rather than the quantity</i>)</p> <ul style="list-style-type: none"> <li>- <b>Basic:</b> Some supervision has been conducted, e.g. a small amount of all program sessions were observed and evaluated/discussed with the executor or there were meetings once in a while/monthly were the executor(s) could discuss conduct problems with a supervisor</li> </ul>	ic_supvi

	<ul style="list-style-type: none"> <li>- <b>Full:</b> The supervision during the program was extensive, e.g. meetings on a regular basis to discuss program methods, video recordings or discussions about program conduct with a supervisor</li> </ul>	
<input type="checkbox"/> nip/ <input type="checkbox"/> no/ <input type="checkbox"/> basic/ <input type="checkbox"/> full	<p><b>Observance while program conducted</b> (<i>To which degree has the executor been observed while conducting the program? Rate the quantity of observations</i>)</p> <ul style="list-style-type: none"> <li>- <b>Basic:</b> Once in a while the executor has been observed while conducting the program, e.g. at the beginning of the program random sessions have been observed</li> <li>- <b>Full:</b> Extensive observance of how the executor conducts the program, e.g. video recordings or personal observance of many sessions</li> </ul>	ic_obcon
<input type="checkbox"/> no/ <input type="checkbox"/> basic/ <input type="checkbox"/> full	<p><b>Documentation of program conduct</b> (<i>How well was the program conduct documented?</i>)</p> <ul style="list-style-type: none"> <li>- <b>Basic:</b> Some documentation is mentioned, e.g. some sessions have been recorded on video in order to check program fidelity, in some sessions records have been kept or some sessions were observed and documented (e.g. with a checklist)</li> <li>- <b>Full:</b> Extensive and detailed documentation of the program conduct, e.g. many session have been recorded and analyzed</li> </ul>	ic_dcond
<input type="checkbox"/> no/ <input type="checkbox"/> basic/ <input type="checkbox"/> full	<p><b>Documentation of study subjects' participation</b> (<i>How well was the study subject's participation documented?</i>)</p> <ul style="list-style-type: none"> <li>- <b>Basic:</b> The study subject's activity/on-task behavior has been rudimentarily documented, e.g. someone randomly kept record of the study subjects' compliance, eye contact or if they work quietly while doing exercises</li> <li>- <b>Full:</b> Extensive and detailed documentation of the study subjects' activity/on-task behavior, e.g. many sessions has been recorded and analyzed regarding the study subject's compliance, eye contact or if they work quietly while doing exercises</li> </ul>	ic_dpart
<input type="checkbox"/> no/ <input type="checkbox"/> basic/ <input type="checkbox"/> full	<p><b>Documentation of participants' attendance</b> (<i>How well was the participants' attendance documented?</i>)</p> <ul style="list-style-type: none"> <li>- <b>Basic:</b> The participants' attendance has been rudimentarily documented, e.g. for each session someone kept record how many participants attended but not who</li> <li>- <b>Full:</b> Extensive and detailed documentation of the participants' attendance, e.g. a detailed protocol or checklist was kept about the attendance of each participant</li> </ul>	ic_datd
<input type="checkbox"/> no/ <input type="checkbox"/> basic/ <input type="checkbox"/> full	<p><b>Offering additional services</b> (<i>To which degree were additional services offered in order to increase or insure program compliance?</i>)</p> <ul style="list-style-type: none"> <li>- <b>Basic:</b> Some services were offered, like a small amount of money as expense allowance</li> </ul>	ic_dadse

	<ul style="list-style-type: none"> <li>- <b>Full:</b> Many services were offered, like money, car services, child care</li> </ul>	
	<input type="checkbox"/> no/ <input type="checkbox"/> basic/ <input type="checkbox"/> full <b>Implementation problems</b> ( <i>To which degree were implementation problems mentioned?</i> ) <ul style="list-style-type: none"> <li>- <b>Basic:</b> Implementation problems can be assumed but are not explicitly mentioned. Assumption can be based on: participants didn't pay attention to the executor, sometimes participants didn't attend the program sessions, high drop-out rate.</li> <li>- <b>Full:</b> Implementation problems have been explicitly mentioned, e.g. parts of the program couldn't be conducted as planned, participants' attention was hard to get or many participants didn't attend the sessions</li> </ul>	ic_prob
<b>INTENSITY OF THE PROGRAM</b>		
<b>Length of program</b>	In days; Theoretically planned period of time over which the program will span (if reported in weeks x 7, if reported in months x 30, if reported in years x 365); school year in the USA: 300 days (36 weeks); estimation is allowed	leng_pr
<b>Number of sessions</b>	Number; Theoretically planned number of times the program will take place, e.g. 300 sessions; If conducted daily and reported in working weeks: x 5; ...and reported in months: x 20; ...and reported in years: x 200; If conducted weekly and reported in months: x 4; ...and reported in years: x 45; school year in the USA: 180 instructional days; estimation is allowed	N_ses
<b>Length of each session</b>	In minutes; Theoretically planned duration of each program session; in case of different session lengths, type in the average length; estimation is allowed	leng_ses
<b>Booster sessions used</b>	0 = <b>Not used</b> 1 = <b>Used</b>	boost_ses
<b>Frequency of sessions</b>	1 = <b>Several times a day</b> 2 = <b>Daily</b> (on working day basis, i.e. 5 days per week) 3 = <b>Several times a week</b> 4 = <b>Weekly</b> 5 = <b>Several times a month</b> 6 = <b>Monthly</b> 7 = <b>Irregular or no frequency</b> (e.g. only one session)	freq_ses
<b>GEOGRAPHICAL PROGRAM INFORMATION</b>		
<b>Area of program</b> <i>In which area did the program take place?</i>	1 = <b>Rural/Suburban</b> (e.g. village, small town, periphery of a city) 2 = <b>Urban</b> (e.g. city center) 3 = <b>Mixed</b> (e.g. participants from city X and its catchment area) Note: If not explicitly mentioned, use recruitment area or other similar hints.	area_pr

<b>Site of study</b>	1 = <b>Single site study</b> (The programs takes place within a small area, e.g. one school, two communities of the same city) 2 = <b>Multisite study</b> (The programs is being conducted in a wide geographical area, e.g. in two states of the USA, in two federal stated in Germany)	site_st
<b>Socially deprived area</b> <i>Is the program being conducted in an underprivileged district?</i>	0 = <b>No</b> 1 = <b>Yes</b> (for example in a ghetto, social hot spot, low-income areas or a socio-economically deprived area, black or Hispanic neighborhood, socially deprived school-district) Note: Also rate “no”, if no information concerning this variable is mentioned	socdepra
<b>Setting of program</b> <i>In which setting did the program take place?</i>	1 = <b>Educational facility</b> (e.g. preschool, kindergarten, nursery, school) 2 = <b>University/research facility or company</b> 3 = <b>Home</b> 4 = <b>Community centre</b> (e.g. church) 5 = <b>Hospital or special education unit</b> 6 = <b>Independently arranged setting</b> 7 = <b>Social Service Agency</b> (e.g. WIC centre, Upward Bound Centre, Head Start Centre) 8 = <b>Several settings</b> (e.g. home visitation and school setting) 9 = <b>Other</b>	set_pr
<b>Setting of program: other</b>	Fill out only if “Setting of program” = 9, otherwise leave blank	set_pr_o
<b>INFORMATION ABOUT THE EXECUTOR AND DEVELOPER OF THE PROGRAM</b>		
<b>Predominant professional orientation of executor</b> <i>Who executed the program in terms of profession?</i>	1 = <b>Not in person</b> (e.g. program delivery by video tape or written information only) 2 = <b>Educational staff</b> (people having educational experiences in working with children, e.g. kindergarten or school teacher) 3 = <b>Psychosocial professionals</b> (people with work experience in the psychosocial field, e.g. sociologist, psychologist, pedagogue, social worker) 4 = <b>Layperson or volunteers</b> (people with no professional experience in educating children, e.g. pensioner) 5 = <b>University staff</b> (e.g. university students as tutors) 6 = <b>Clinical professionals</b> (e.g. nurse, pediatrician) 7 = <b>Administrators/counselors</b> 8 = <b>Multiple personal</b> (several executors with different specializations)	prof_exe

<p><b>Interest executor</b> <i>Who executed the program in terms of conflict of interest?</i></p>	<p>1 = <b>Not in person</b> (e.g. program delivery by video tape or written information only)  2 = <b>Third party executes the program</b> (e.g. kindergarten or school teacher)  3 = <b>With the developer <u>loosely</u> associated person executes the program</b> (e.g. developer trains third party executor, like clinicians)  4 = <b>Developer of the program or <u>closely</u> associated people</b> (e.g. research assistants, Ph.D. students) <b>take part in executing the program</b>  5 = <b>Study author(s) has/have a close relationship to the executor</b> (e.g. his/her research assistant, Ph.D. student, executor receives a training certificate from the author)  6 = <b>Study author(s) is/are the executor(s)/one of the executors</b> (code also 6, if the executor is the study author AND the developer)</p>	<p>int_exe</p>
<p><b>Interest developer</b> <i>Is there an association between the program developer and the study author?</i></p>	<p>0 = <b>No conflict</b> (Program developer is not the study author/one of the study authors or advertized the program)  1 = <b>Light conflict</b> (The developer has a close relationship to the study author or with the author associated people (e.g. research students)  2 = <b>Strong conflict</b> (The study author took part in developing the program or advertizing it. If the author(s) doesn't/don't mention the developer's name, conclude the author(s) him/herself as the developer.)  Note: If there is no hint concerning the developer of the program, assume that the author is the developer.</p>	<p>int_dev</p>

INFORMATION ABOUT THE CHARACTERISTICS OF THE SAMPLE

→ FOR ASSESSING THE SAMPLE VARIABLES, USE THE TIME POINT BEFORE THE PROGRAM STARTED AS REFERENCE POINT

VARIABLE	CODING VALUES, DESCRIPTIONS AND NOTES	CODE
Type of group	1 = Poverty 2 = Migration 3 = Both	type_gr
<b>POVERTY AND ETHNICITY</b>		
70% of the sample needs to meet at least one or more of the 5 following poverty types. The information need to be explicitly quantified in the paper.		
Income poverty	<p>0 = Not Applied</p> <p>1 = Applied – parental</p> <p>2 = Applied – paternal</p> <p>3 = Applied – maternal</p> <p>4 = Applied (without hint about whose income the author assessed)</p> <p>The annual parental/paternal/maternal income falls below the poverty level, which can be found in appendix C - table A. If the family income is reported, divide it by 2,1 (equivalence size for an average family with two parents and two children: 1+0,5+0,3+0,3). If the threshold is not available for a certain year, use the next available estimator, e.g. if the study is published in Turkey in 1995, the best estimated threshold is 1.283US\$ (average in the time span 2000-2011 in Turkey). Use alternative sources if the table does not contain sufficient information (see appendix C - table B).</p> <p>If the mean or median income is reported, consider the standard deviation or maximum income, respectively, in order to estimate the income level of at least 70% of the sample:</p> <ul style="list-style-type: none"> <li>- Mean (or Median) and SD are reported: <math>income_{70} = M + SD</math></li> <li>- Mean (or Median) and maximum income is reported: <math>income_{70} = (M + income_{max}) / 2</math></li> </ul> <p>If only the mean or median are reported, no estimation for the income of 70% of the sample can be made.</p>	inc_pov
Occupation poverty	<p>0 = Not Applied</p> <p>1 = Applied – parental</p> <p>2 = Applied – paternal</p> <p>3 = Applied – maternal</p> <p>4 = Applied (without hint about whose occupation the author assessed)</p> <p>The pursued occupation is either a low-level profession, e.g. unskilled (i.e. porters, dustmen, laborer), semi/partly-skilled (i.e. postal workers, plant/machine operators) or non-professionals or no occupation/very few hours of occupation, e.g. 10h per</p>	occ_pov

	<p>week is accessible, e.g. occupation category 6 und 7 of the Warner Scale (6: Assembly Line Workers -auto and other; Attendants-gas stations, hospitals; Auto Body Repairmen-painting; Baggage Men; Bank Messengers; Body and Fender Workers; Car Deliverers; Car Inspectors, railroad; Carpenters' Assistants; Chauffeurs; Chemical Plant Workers; Clerks-duplicating, fleet-service, general, receiving room. mail and shipping, stock, order, supply, postal; Coffee Roasters; Construction, iron workers; Crate Followers; Delivery Men; Drivers, motor vehicles; Electrical Workers; Exterminator's Helpers; Factory Workers; Gardeners; Gear Grinders; Golf Course Inn Keepers; Green Keepers; Guards; Jailers; Landscapers; Lathers' metallic; Lift-drivers; Machine Operators; Mail Handlers; Mailmen (letter carriers); Mechanics-gas station, building; Metal Workers-cutters, sanders; Milkmen; Moulders; Night Policemen; Night Watchmen; Operators, highway maintenance equipment; Porters, hospital; Pressers--cleaning establishment; Salesmen, route; Semi-skilled Workers; Shipping Room Workers; Shop Workers; Tank Repairman; Taxi Drivers; Draw Bridge Operators; Truck Drivers; Waiters; Waitresses; Welders; Military Servicemen; 7: Caddies; Car Washers; Construction workers; Custodians; Dairy Laborers; Dishwashers; Helpers; Heavy Laborers; House-cleaning Servicemen; Institutional Aides; Janitors; Machine Cleaners; Maintenance Man; Migrant Workers; Miners; Movers; Newsboys; Odd-job Men; Orderlies-hospital; Porters, general; Scrubwomen; Window Washers)</p>	
<b>Education poverty</b>	<p>0 = <b>Not Applied</b>  1 = <b>Applied – parental</b>  2 = <b>Applied – paternal</b>  3 = <b>Applied – maternal</b>  4 = <b>Applied</b> (without hint about whose education the author assessed)  The number of years involved in the educational system falls below a certain level or a certain degree has not been achieved (see appendix D - table A. If the publication year of the study and the reference year of the threshold (see appendix D, table B) differ by more than 10 years, use alternative sources (see appendix D - table C).</p>	educ_pov
<b>Federal Assistance poverty</b>	<p>0 = <b>Not Applied</b>  1 = <b>Applied – parental</b>  2 = <b>Applied – paternal</b>  3 = <b>Applied – maternal</b>  4 = <b>Applied</b> (without hint about whose federal assistance the author assessed)  Eligible for any kind of governmental aid for poverty reasons, e.g. free or reduced lunch for children, aid to families with dependent children, Medicaid, MassHealth, Hartz IV, Arbeitslosengeld, Supplemental Nutrition Assistance Program (SNAP)/Food Stamps, Temporary Assistance for Needy Families (TANF), Alaska Temporary Assistance Program (ATAP)  Not automatically poverty sample: Medicare, Medi-Cal, Oregon Health Plan, Soonercare, TennCare, BadgerCare, State Children's Health Insurance Program, Program for Women, Infants, and Children (WIC), Hippy (Home Instruction for Parents and Preschool Youngsters), Head start (schools), Health Care Card (Australia)</p>	f_as_pov
<b>Composite poverty</b>	<p>0 = <b>Not Applied</b>  1 = <b>Applied – parental</b>  2 = <b>Applied – paternal</b>  3 = <b>Applied – maternal</b></p>	comp_pov



	4 = <b>Applied</b> (without hint about whose federal assistance the author assessed) Any kind of system for categorizing low social classes by several characteristics using sum scores, means or number of risk characteristics, e.g. High-Risk Index (based on Ramey, 2000: having at least 25/50 points equals poverty), Hollingshead Four-Factor Index of Socioeconomic Status, ad hoc measure, like combining education, occupation, housing and type of neighborhood and composing the categories “advantaged”, “average” and “disadvantaged” or meeting two out of three poverty risk characteristics (like unmarried, <12 years of education, unemployed), Schichtindex (Lampert & Kroll, 2006; 3-21, Cut-Off = 10% = 5)	
<b>Dominant ethnicity</b>	1 = <b>White</b> 2 = <b>African (American)/Black</b> 3 = <b>Hispanic/Latino</b> 4 = <b>Diverse</b> 5 = <b>Other</b> Note: Dominant means more than 50% of the sample. If no group reaches 50%, code 4 (divers).	d_eth
<b>Dominant ethnicity: other</b>	Fill out only, if “Dominant ethnicity” = 5, otherwise leave blank	d_eth_o
<b>% of dominant ethnicity</b>	Type in percentage of dominant ethnicity	p_d_eth
<b>Status of dominant ethnicity</b>	1 = <b>Minority</b> 2 = <b>Majority</b> (Central Intelligence Agency (US). The World Factbook. Ethnic Groups. Verfügbar unter <a href="https://www.cia.gov/library/publications/the-world-factbook/fields/2075.html#">https://www.cia.gov/library/publications/the-world-factbook/fields/2075.html#</a> )	st_d_eth
<b>MIGRATION</b> Note: Sample (or subsample for which separate results are reported) have to consist mainly (at least 50%) of children/adolescents with migration background (at least one parent is born in another country). If migration background is not specified and percentage of first generation immigrants is either below 50% or not specified, but more than 90% belong to an ethnic (e.g. Hispanic or Asian American) or linguistic (e.g. English Language Learner in US) minority, that usually consists mainly of people with migration background (NOT African American, since most of them have NO migration background; also indigenous ethnic minorities are NOT included) then the inclusion criterion for migration is still fulfilled.		
<b>Subsample of the study</b>	1 = <b>Whole sample</b> (if whole sample fulfills inclusion criteria) 2 = <b>Subsample</b> (if only a subsample for that results are reported fulfills inclusion criteria)	subsam

<b>Inclusion because of</b>	<p>1 = <b>1<sup>st</sup> Generation immigrants</b> (only if percentage is above 50%)</p> <p>2 = <b>Migration background</b> (only if percentage is above 50% and percentage of 1<sup>st</sup> generation immigrants is below 50% or not specified)</p> <p>3 = <b>Ethnic minority</b> (only if percentage is above 90%, migration background is not reported and percentage of 1<sup>st</sup> generation immigrants is either below 50% or not specified)</p> <p>4 = <b>Linguistic minority</b> (only if percentage is above 90%, neither ethnic minority nor migration background is reported and percentage of 1<sup>st</sup> generation immigrants is either below 50% or not specified)</p>	incl_bec
<b>Main linguistic minorities</b>	<p>Name; Use the most precise information (i.e. if only English Language Learners are mentioned as participants insert “English Language Learner”, but if the author also says that the home language of the participants is Spanish insert “Spanish speakers”); If more than one linguistic minority is included in the (sub-)sample insert the three linguistic minorities with the highest percentage in the sample, including the percentage of each minority in brackets and separating them by comma (e.g. “Spanish speakers (50%), Chinese speakers (40%), Portuguese speakers (10%)”); If not clear from the description of the author and “Inclusion because of” ≠ 4 then code “-99”.</p>	mig_lin
<b>Main ethnic minorities</b>	<p>Name; Use the most precise information (i.e. if only Asian Americans are mentioned as participants insert “Asian Americans”, but if the author also says that all the participants have Korean origin insert “Korean American”); If more than one ethnic minority is included in the (sub-)sample insert the three ethnic minorities with the highest percentage in the sample, including the percentage of each minority in brackets and separating them by comma (e.g. “Hispanic (50%), Chinese American (40%), Korean American (10%)”); If not clear from the description of the author and “Inclusion because of” ≠ 3 then code “-99”.</p>	mig_ethn
<b>Country of origin of main immigrant groups</b>	<p>Name; Use the most precise information (i.e. if only Latin America is mentioned as country of origin insert “Latin America”, if the author also says that all the participants are from Argentina insert “Argentina”); If more than one Country of origin is included in the (sub-)sample insert the three countries of origin with the highest percentage in the sample, including the percentage of each minority in brackets and separating them by comma (e.g. “Mexico (50%), Cuba (40%), Puerto Rico (10%)”); If not clear from the description of the author and “Inclusion because of” ≠ 1 and ≠ 2 then code “-99”.</p>	mig_coor
<b>Children/ adolescents with migration background in the sample</b>	<p>Percentage of children/adolescents in the sample with at least one parent born in another country; if data only is available for one parent use this percentage as best possible estimator.</p>	mig_bgp

<b>1. Generation immigrant children/adolescent in the sample</b>	Percentage of children/adolescents in the sample who were born in another country	mig_lgp
<b>Group is predominant in the recruitment area</b>	<p>1 = <b>No, ethnic majority predominant</b> (in the area of recruitment the ethnic majority of the country (e.g. White in US) is predominant)</p> <p>2 = <b>No, but minority/migrants predominant</b> (in the area of recruitment the minority/migrant group is not predominant but there are less majority members than minorities/ people with migration background of any origin in the area)</p> <p>3 = <b>Yes</b> (the specific minority/migrant group is predominant in the recruitment area)</p> <p>Note: If no information about the predominant ethnicity is reported by the author but a precise (i.e. city or county, NOT only state) recruitment area <u>within</u> the US is given and the study was conducted after 2000 check <a href="http://www.census.gov/2010census/popmap/ipmtext.php">http://www.census.gov/2010census/popmap/ipmtext.php</a> (e.g. if you are looking for Miami City select “Florida” as State, click “areas within”, choose “Places”, click “Search”, choose “Miami city” and click “Display”. If only a precise recruitment area <u>outside</u> the US is given or only a precise recruitment area within the US is given but the study was conducted before 2000 ask Sebastian.</p>	mig_pred
<b>SAMPLE SIZE, AGE, GENDER AND RISK FACTORS (THESE VARIABLES ALWAYS CONCERN THE CHILDREN)</b>		
<b>N IG</b>	Sample size at randomization; If sample size is only reported for the whole sample, use this divided by the number of comparison groups as best possible estimator for each group.	N_IG
<b>N CG</b>		N_CG
<b>Child mean age IG</b>	In months (If reported in weeks ÷ 4, if reported in years x 12); If mean age is only reported for the whole sample, use this as best possible estimator for each group. If only the median age is reported, use this as best possible estimator for each group. If the child is not born yet, count backwards, e.g. - 4,5 months for a mothers’ child in the second trimester of pregnancy For estimation based on school grade, use the following mean ages in months: Kindergarden → 66 (range: 60-72), 1 <sup>st</sup> class → 78 (range: 72-84), 2 <sup>nd</sup> class → 90 (range: 84-96), 3 <sup>rd</sup> class → 102 (range: 96-108), and so on	M_ag_IG
<b>Child mean age CG</b>		M_ag_CG
<b>Child min age IG</b>	In months (If reported in weeks ÷ 4, if reported in years x 12); If range age is only reported for the whole sample, use this as best possible estimator for each group. If only SD is reported, subtract 2 SDs from the mean age for the minimum range and add 2 SDs to the mean age for the maximum range as best possible estimator.	min_ag_IG
<b>Child max age IG</b>		max_ag_IG
<b>Child min age CG</b>		min_ag_CG
<b>Child max age CG</b>		max_ag_CG

<b>Child gender IG</b>	In percent of females; If gender is only reported for the whole sample, use this information as best possible estimator for each group.	ge_c_IG
<b>Child gender CG</b>		ge_c_CG
<b>Risk parameters of the sample: children</b> <i>Is at least 50% of the sample of high risk?</i>	<input type="checkbox"/> / <input checked="" type="checkbox"/> <b>Cognitive/academic developmental risk factors</b> (e.g. low IQ, academic problems, developmental disabilities, low birth weight)	risk_chco
	<input type="checkbox"/> / <input checked="" type="checkbox"/> <b>Problem behavior risk factors</b> (e.g. conduct problems, substance abuse, criminal background/behavior, teenage pregnancy, homelessness)	risk_chpb
<b>Risk parameters of the sample: parents</b> <i>Is at least 50% of the sample of high risk?</i>	<input type="checkbox"/> / <input checked="" type="checkbox"/> <b>Personal risk factors</b> (e.g. psychological disorders, mother's low IQ, criminal background, child abuse)	risk_pper
	<input type="checkbox"/> / <input checked="" type="checkbox"/> <b>Structural risk factors</b> (e.g. teenage mother/pregnancy, divorced/single/widowed parents, inadequate prenatal care)	risk_pstr
<b>Selectivity of the sample</b> <i>To which degree did the authors select the sample?</i>	1 = <b>Low</b> (Sample not selective, except for inclusion criteria poverty or migration; representative sample) 2 = <b>Middle</b> (Sample selective regarding additional demographic characteristics, e.g. only boys, only children of refugees, children of divorced parents; selective program) 3 = <b>High</b> (Study participant or parents already show some problem behaviors, e.g. children of depressed mothers, children having reading problems or children rated by teachers as problematic, teenage mothers; indicated program)	sel_sam
<b>RECRUITMENT RATE, ATTRITION RATE AND ATTENDANCE</b>		
<b>Recruitment rate</b>	In percent; Ratio of people who are willing to participate in the study/completed the randomization process in relation to the number of people who were asked to participate/who were addressed, for example if 30 people are being called to ask for participation and 15 people take actually complete the pretest, type in "50". If the authors don't report randomization information, use the number of participants completing the pretest. If there has been a public announcement for participation (e.g. radio announcement or advertisement on the blackboard in the school or in the newspaper) and no recruitment rate is being reported by the author, type in "-10"; If the potential participants were approached individually and no recruitment rate is being reported, type in "-30". If the potential participants were approached personally (i.e. face-to-face) and no recruitment rate is reported, type in "-40". In case of a combination of a personal/individual approach and an impersonal announcement, e.g. parent-teacher conference, type in "-20". If the sample has been methodologically preselected (e.g. interviews/screenings for preselection of a risk sample), type "-77".	recruitr

<b>Attrition rate post: IG</b>	In percentage; Ratio of participants (i.e. individuals who took part in the program) who <u>didn't</u> attend the posttest to participants who completed randomization;	attr_pIG
<b>Attrition rate post: CG</b>	If no randomization information is available, use the number of people who completed the pretest. If attrition rate is only reported for the whole sample, use this information as best possible estimator for each group. <i>Note: If different sample sizes are reported for different outcomes, use the highest as sample size post.</i>	attr_pCG
<b>Attrition rate FU1: IG</b>	In percentage; Ratio of participants who didn't attend the follow-up I to participants who completed randomization; If no randomization information is available, use the number of people who completed the pretest.	attr_f1IG
<b>Attrition rate FU1: CG</b>	If attrition rate is only reported for the whole sample, use this information as best possible estimator for each group. If no follow-up I has been assessed, leave blank.	attr_f1CG
<b>Attrition rate FU2: IG</b>	In percentage; Ratio of participants who didn't attend the follow-up II to participants who completed randomization; If no randomization information is available, use the number of people who completed the pretest.	attr_f2IG
<b>Attrition rate FU2: CG</b>	If attrition rate is only reported for the whole sample, use this information as best possible estimator for each group. If no follow-up II has been assessed, leave blank.	attr_f2CG
<b>Full attendance rate</b>	In percent; Percentage of people attending all program sessions (-10% tolerance) If program participation is obligatory in the intervention setting, e.g. intervention is a certain school curriculum, code "-88".	atdf_r
<b>Mean attendance rate</b>	In percent; Mean attendance (Ratio of mean attended number of sessions to maximal number of sessions, e.g. 10 sessions planned and on average participants attended 8,7 sessions → $8,7/10 = 0,87$ → attendance rate = 87%). If program participation is obligatory in the intervention setting, e.g. intervention is a certain school curriculum, code "-88".	atdm_r

VALIDITY EVALUATION

VARIABLE	CODING VALUES, DESCRIPTIONS AND NOTES	CODE
<p><b>Type CG</b>  <i>What kind of control group was applied?</i></p>	<p>1 = <b>No treatment</b> (The non-program group received <i>neither a treatment nor attention</i>, e.g. intensive reading program on top of the normal school curriculum versus normal school curriculum only)            2 = <b>Wait-list control group</b> (Control group will <i>receive the program after</i> the program group completed the posttests)            3 = <b>Treatment as usual</b> (Includes all control groups that are being given a <i>standard procedure</i>; e.g. counseling by a school psychologist on an as-needed basis)            4 = <b>Attention control group</b> (An <i>unspecific program that is unrelated</i> to the program’s mechanism of action is conducted, in order to give the <i>same amount of attention</i> to the control group as the program group receives; e.g. intensive reading program versus free reading time)            5 = <b>Minimal treatment control group</b> (Control group receives a <i>minimal form</i> of treatment, e.g. one session of parent training in the control group (versus 20 sessions in the program group))            6 = <b>Basic program control group</b> (Control group <i>already receives a program</i> component which is the same as in the program group, e.g. Head Start sample and the program group receives a parent training additionally)</p>	<p>type_CG</p>
<p><b>Contact between IG and CG</b>  <i>To which degree was an exchange between IG and CG possible concerning the content of the program strategies?</i></p>	<p>1 = <b>No</b> (There was no contact between the IG and CG, or very unlikely)            2 = <b>Few</b> (Few contact of program strategies took place or is probable, e.g. IG and CG live in the same neighborhood, attend the same school, executor mentioned program strategies to the CG)            3 = <b>Some</b> (Some contact of program strategies took place or is very probable, e.g. IG and CG are in the same class, executor explained/practiced program strategies to/with the CG)</p>	<p>IG_CG</p>
<p><b>Study design</b>  <i>What kind of randomization procedure was applied?</i></p>	<p>1 = <b>Cluster randomization</b> (Whole clusters (e.g. class, school, site, community) are assigned to IG and CG)            2 = <b>Individual randomization</b> (Randomization on individual level, e.g. each child is randomized to IG and CG)            3 = <b>Stratified randomization</b> (Participants are matched before individual randomization or randomization takes places within certain subgroup/strata, e.g. age, gender, grade, school, cohorts)</p>	<p>stud_des</p>

<p><b>Group equivalence</b>  <i>Were the groups equivalent on relevant variables prior to the program?</i></p>	<p>1 = <b>Not tested</b>  2 = <b>Tested but not equivalent</b> (<u>Significantly different</u> on relevant demographic variables (e.g. age, sex, poverty/migration, primary language, risk factors) and assessment relevant variables (e.g. pretest score)  3 = <b>Tested and equivalent</b> (<u>Not significantly different</u>) on relevant demographic variables (e.g. age, sex, poverty/migration, primary language, risk factors) and/or assessment relevant variables (e.g. pretest score)  Note: Use .05 as significance level to judge, if the groups are equivalent or not. If no tests are reported, take the author's interpretation as judgment.  If subgroups are being coded and the test for group equivalence is only reported for the whole sample, use this information as best possible estimator.</p>	<p>gr_equiv</p>
<p><b>Data analysis approach</b>  <i>How did the author handle missing values in the intervention and control group?</i></p>	<p>1 = <b>Per Protocol, strong</b> (The authors only include participants in the data analysis if they <u>received a certain amount of the program</u> e.g. at least 50%)  2 = <b>Per Protocol, soft</b> (The authors include participants in the data analysis who <u>completed the post-test</u>, no matter how much of the program has been received; <u>No imputation</u> procedure was applied in order to replace missing information due to drop-outs.)  3 = <b>Intention to treat, soft</b> (The authors include all participants in the data analysis <u>who completed the randomization procedure</u>, regardless of their attendance or attrition; Missing values due to drop-outs have been <u>imputed</u>.)  4 = <b>Intention to treat, strong</b> (The authors include all participants in the data analysis <u>who were approached</u> for study participation, regardless of agreement to randomization procedure (study participation), attendance or attrition.)  Note: If there is no attrition in the sample and no information about the analysis technique (per protocol or intention to treat) reported, <u>leave blank</u>.</p>	<p>dat_an_a</p>
<p><b>Selectivity of outcome reporting</b></p>	<p>1 = <b>Not selective</b> (Results for all assessed outcomes that were mentioned are reported for the (sub)sample.)  2 = <b>Selective</b> (For the (sub)sample only the results for some of the assessed outcomes are reported.)</p>	<p>sel_outr</p>

## EFFECT SIZE LEVEL

VARIABLE	CODING VALUES, DESCRIPTIONS AND NOTES	CODE
Publication ID		ID_Pub
Study ID		ID_Stud
Comparison ID		ID_Comp
<b>OUTCOME INFORMATION</b>		
Outcome	Specify outcome precisely; Use subscales if reported	outcome
Outcome category	1 = <b>Cognitive-developmental outcomes</b> (CDO) 2 = <b>Educational and professional outcomes</b> (EO) 3 = <b>Socioemotional outcomes</b> (PSO) 4 = <b>School environment outcomes</b> (SEO) 5 = <b>Home environmental and family outcomes</b> (HEO) 6 = <b>Cost-effectiveness outcomes</b> 7 = <b>Multiple Outcome categories</b>	out_cat
Outcome construct (OC)	Asses only the outcomes that can be assigned to one of the following categories, e.g. don't code physical parameters like body height or weight, heart rate or child's sicknesses. However, if a relevant outcome doesn't fit in one of the categories, discuss it in the weekly discussion group.	
	1 10 = <b>Language related parameter</b> (e.g. language skills, attitudes to language, use of language, reading skills, writing skills, PPVT-III, EOWPVT-R, WJ-R LW. GRTR, OLPT, frequency using library card, cognitive reading strategies) 1 11 = <b>Academic parameters</b> (e.g. math/science achievement/skills, WJ-RAP, social science studies) 1 12 = <b>Basic cognitive function</b> (e.g. intelligence, (general) knowledge, school readiness, WIPPSI, REEL, academic achievement, reasoning, BSID-MDI, coordination, attention, Bracken, distractibility) 1 13 = <b>General development</b> (e.g. developmental quotient, Bayley Scales of Infant Development (BSID), PSPCSA - competence subscale, Infant behavior scale) 1 14 = <b>Motor skills</b> 1 19 = <b>Other CDO</b> 2 20 = <b>Learning strategies</b> (e.g. learning strategies, learning problems, task orientation) 2 21 = <b>Learning motivation</b> (e.g. academic motivation, school attendance rate, learning engagement, tardiness) 2 22 = <b>Academic meta-strategies</b> (e.g. subjective academic abilities, academic self-concept, academic self-efficacy, educational	out_con



		expectations)	
2	23 =	<b>Educational career</b> (e.g. graduation rate, higher/further education, school biography, drop out, receiving financial aid for education)	
2	24 =	<b>Replacement</b> (e.g. special education placement, retention rate, eligibility of Early Intervention (EI), discipline referrals)	
2	25 =	<b>Grades</b> (reading/writing grades, math/science grades, social science grade, GPA)	
2	26 =	<b>Professional career</b> (Profession-related parameter, like knowledge about profession, employment type, employment rate, attitude toward work, earning; also income)	
2	29 =	<b>Other EO</b>	
3	30 =	<b>Internalizing behavior</b> (e.g. anxiety/fear, depression, withdrawal, emotion regulation, tension/relaxation, CBCL – internal, introversion, emotional adjustment)	
3	31 =	<b>Externalizing behavior</b> (e.g. aggressiveness, conduct problems, hyperactivity, detrimental leisure, delinquent/criminal behavior, getting arrested, incarceration, active lifestyle, CBCL – external, self-control, deviance, positive attitude behavior, impulsivity)	
3	32 =	<b>Self-related cognition</b> (self-esteem, self-concept, self-confidence, assertiveness, decision-making, goal-setting, planning skill, coping)	
3	33 =	<b>Social competence</b> (e.g. social interaction, empathy, socio-emotional success, assistance skills, child factor of the Parent Child Early Relational Assessment, PSPCSA - acceptance subscale, apathy, expressiveness, adapting skills, communication skills, acceptance of male on female aggression (v.v.))	
3	34 =	<b>Sexual parameter</b> (e.g. (risky) sexual activities, pregnancies, births, attitude concerning sexual activity)	
3	35 =	<b>Substance abuse parameter</b> (e.g. substance abuse, intention/attitudes towards substance consumption)	
3	36 =	<b>General behavior problems</b> (e.g. a combination of internal and external behavior, SSRS -behavioral subscale)	
3	39 =	<b>Other PSO</b>	
4	40 =	<b>Teacher-related parameter</b> (e.g. teacher content knowledge, teacher didactic abilities/instrumental practices, ECERS-R subscales “personal care”, “language reasoning”, “activities”, “program”)	
4	41 =	<b>Teacher relationship-related parameter</b> (e.g. student-teacher relationship, parent-teacher communication, ECERS-R subscales “interactions”, “parents and staff”)	
4	42 =	<b>General (school) climate</b> (e.g. peer discrimination, mobbing, bullying, classroom climate, stimulation, ECERS-R, SELA, CLASS)	
4	49 =	<b>Other SEO</b>	
5	50 =	<b>Parental attitudes &amp; knowledge</b> (self-efficacy, authoritarian educational beliefs, rejection of child-rearing role, appropriateness of developmental expectations, Belief in Use of corporal punishment, Mothers' Educational Attitudes	

	<p>Scale (EAS), Reading books is one of parent's 3 favorite things to do with their child, parenting warmth)</p> <p>5 51 = <b>Parental behavior</b> (e.g. sensitivity, neglect, victimization, abuse, parenting skills, parental involvement, parent factor of the Parent Child Early Relational Assessment, parent reads to child, HOME)</p> <p>5 52 = <b>Parent-child relationship</b> (e.g. quality of communication, attachment)</p> <p>5 53 = <b>Stimulation</b> (e.g. inventory at home like books, parent taking child to library)</p> <p>5 59 = <b>Other HEO</b></p> <p>6 60 = <b>Multiple Outcomes</b></p>	
<b>OC: other CDO</b>	Fill out only if "Outcome construct" = 19, otherwise leave blank	OC_o_CDO
<b>OC: other EO</b>	Fill out only if "Outcome construct" = 29, otherwise leave blank	OC_o_EO
<b>OC: other PSO</b>	Fill out only if "Outcome construct" = 39, otherwise leave blank	OC_o_PSO
<b>OC: other SEO</b>	Fill out only if "Outcome construct" = 49, otherwise leave blank	OC_o_SEO
<b>OC: other HEO</b>	Fill out only if "Outcome construct" = 59, otherwise leave blank	OC_o_HEO
<b>ASSESSMENT INFORMATION</b>		
<b>Effect size ID</b>		ID_EffS
<b>Method</b> <i>What kind of measurement has been applied?</i>	<p>1 = <b>Questionnaire</b> (Participants read a question (not rating- or knowledge-related) and answer it in a paper-and-pencil-format, e.g. "Did you work for pay during summer?", "Have you ever had sex?")</p> <p>2 = <b>Rating procedures</b> (Participant judges him/herself or another person does regarding a feature/trait by giving a rating, e.g. Child Behavior Checklist)</p> <p>3 = <b>Achievement test</b> (Participant answers knowledge questions or solves problems. The questions' answers and problems' solutions are either right or wrong. e.g. IQ-test, standardized math test)</p> <p>4 = <b>Observation</b> (Data is assessed through observation of relevant people, e.g. Strange situation, sensitivity of the mother playing with child)</p> <p>5 = <b>Interview</b> (A (trained) person talks to the participant to gather information about the participant or an associated person, e.g. SKID, Adult Attachment Interview, questionnaire asked in interview form)</p> <p>6 = <b>Archival data</b> (e.g. grades in school, number of births from hospital archives, number of incarcerations)</p> <p>Distinction between questionnaire and rating procedure: In contrast to questionnaires, rating procedures are test-theoretically well-founded and have a higher level of standardization, e.g. homogenous response format.</p>	method
<b>Assessor</b> <i>Who assessed the data (at post test)?</i>	<p>1 = <b>Child/Adolescent</b></p> <p>2 = <b>Parent or other relative</b></p> <p>3 = <b>(Kindergarten) teacher, Director of school</b></p> <p>4 = <b>Peers</b></p> <p>5 = <b>Author/developer of the program or associated personal</b> (e.g. (blinded) research assistant)</p>	assessor

	<p>6 = <b>Executor of the program</b> (If not teacher, author or developer)</p> <p>7 = <b>Specially trained or skilled personnel</b> (Person is either specially trained for this individual measurement, has professional qualification for this kind of measurement (e.g. psychologist) or professional qualifications are required to administer the test (e.g. HAWIK – Hamburg-Wechsler-Intelligence test for children))</p> <p>8 = <b>Other not specially trained or skilled person</b></p> <p>9 = <b>Divers/multiple assessors</b></p> <p>10 = <b>Not in person</b></p>	
<p><b>Assessment subject</b> <i>Who has been assessed?</i></p>	<p>1 = <b>Child/Adolescent</b></p> <p>2 = <b>Parent or other relative</b></p> <p>3 = <b>(Kindergarten) teacher</b></p> <p>4 = <b>Peers</b></p> <p>5 = <b>Multiple assessment subjects</b></p> <p>6 = <b>No person assessed</b> (e.g. inventory at home counted, archival data used)</p>	ass_subj
<b>Page</b>	Page where inserted effect size calculation information can be found	page
<b>EFFECT SIZE CALCULATION INFORMATION</b>		
<b>Assessment time point - post</b>	Time between completion of program and post-test in month after the program is completed; Include post program time point only (program needs to be completed).	astp_p
<b>IG Post Mean</b>		IG_postM
<b>IG Post SD</b>		IG_postSD
<b>IG Post N</b>	If N is only reported at randomization and exact attrition rate (AR) for post is known, calculate N under consideration of AR.	IG_postN
<b>CG Post Mean</b>		CG_postM
<b>CG Post SD</b>		CG_postSD
<b>CG Post N</b>	If N is only reported at randomization and exact attrition rate (AR) for post is known, calculate N under consideration of AR.	CG_postN
<b>IG Pre Mean</b>		IG_preM
<b>IG Pre SD</b>		IG_preSD
<b>IG Pre N</b>		IG_preN
<b>CG Pre Mean</b>		CG_preM
<b>CG Pre SD</b>		CG_preSD
<b>CG Pre N</b>		CG_preN

<b>Estimated effect size - post</b>	<p>Fill out only, if "Information for ES estimation - post" = 3, 4, 5, 6 or 2 in the case of change scores</p> <p>Note: Please insert the sample size for the estimated effect size in the fields for C1 effect size (for IG: "C1 IG Post N"; for CG: "C1 CG Post N").</p> <p>If estimation is based on results of an AN(C)OVA only use F-tests with 1 numerator degrees of freedom. If an interaction factor for group (IG vs. CG) x time (pre vs. post/FU) is reported, preferably use this F-value to calculate the effect size, otherwise use the F-value for the group factor.</p>	estES_po
<b>Polarity - post</b>	<p>-1 = <b>Negative</b> (The scale to measure the outcome follows the principle: Higher values stand for impairment)</p> <p>0 = <b>Not clear</b> (No direction indicated, e.g. when using opinion ratings)</p> <p>1 = <b>Positive</b> (The scale to measure the outcome follows the principle: Higher values stand for improvement)</p>	polar_po
<b>Information for ES estimation - post</b>	<p>1 = <b>Mean, SD and N for IG and CG prior and after the program are reported</b></p> <p>2 = <b>Mean, SD and N for IG and CG after the program are reported</b> (If no post SD (and post N), but pre SD (and pre N) is reported, use this as best possible estimator (Please still insert measures in the designated fields and rate missings with -99, more precisely, don't insert the preSD in postSD field even if it will be treated as such); If (in exceptional cases) comparable variables in terms of content are measured in pre-post-comparison, e.g. two versions of an IQ-test, then calculate ES from postES - preES)</p> <p>3 = <b>Effect size is reported</b> (consider the index and calculation, e.g. if Cohen's d is reported, convert into Hedges' g)</p> <p>4 = <b>Test statistic is reported</b></p> <ol style="list-style-type: none"> <li>Mean value test (e.g. <i>F</i>, <i>t</i>) <i>Note: For calculations based on F-value make sure to determine direction of effect!</i></li> <li>Measures of associations (e.g. <i>r</i>, <math>\chi^2</math>)</li> <li>Proportions (e.g. <i>OR</i>)</li> <li>Non-parametric statistics (e.g. <i>U</i>)</li> </ol> <p>5 = <b>Significance level is zero</b> (estimated effect size = 0; In case of a reported exact p-value, a more accurate back calculation can be made through inference on test statistics and degrees of freedom. Code 4a-d, depending on which applies)</p> <p>6 = <b>Estimation by the coder: d = 0.2(small)/0.5(medium)/0.8(high) effect</b> (from information in the text, e.g. description of effects (e.g. the effects of the program where small), graphs, statistical information insufficient for back calculation etc.; please note hints on corresponding estimates)</p> <p>Note: Final effect size is Hedges' g</p> <p>Note: The coding order corresponds to the dominance of one effect size calculation information over another, e.g. if an effect size can be estimated by proportions and by a mean value test, like t-test, choose the mean value test (4a before 4d).</p> <p>If the authors only mentioned that an effect is significant, without giving information about the test statistic, ask for solution in the weekly discussion groups.</p>	infES_po

<b>Significance - post</b> (explicitly mentioned)	0 = <b>Effect is not significant</b> ( $p \geq 0.05$ ) 1 = <b>Effect is significant</b> ( $p < 0.05$ ) If an AN(C)OVA without repeated measures was calculated and p-values for each factor reported, use preferably significance of group x time factor, otherwise use significance of group factor as best possible estimator. If an AN(C)OVA with repeated measures was calculated, discuss in weekly discussion group.	sig_po
<b>Assessment time point - follow-up I</b>	In months after program is completed	astp_fu1
<b>IG FUI Mean</b>		IG_fu1M
<b>IG FUI SD</b>		IG_fu1SD
<b>IG FUI N</b>	If N is only reported at randomization and exact attrition rate (AR) for FUI is known, calculate N under consideration of AR.	IG_fu1N
<b>CG FUI Mean</b>		CG_fu1M
<b>CG FUI SD</b>		CG_fu1SD
<b>CG FUI N</b>	If N is only reported at randomization and exact attrition rate (AR) for FUI is known, calculate N under consideration of AR.	CG_fu1N
<b>Estimated effect size - FUI</b>	Fill out only, if "Information for ES estimation - FUII" = 3, 4, 5 or 6 Note: Please insert the sample size for the estimated effect size in the fields for C1 effect size (for IG: "C1 IG FUI N"; for CG: "C1 CG FUI N"). If estimation is based on results of an AN(C)OVA only use F-tests with 1 numerator degrees of freedom. If an interaction factor for group (IG vs. CG) x time (pre vs. post/FU) is reported, preferably use this F-value to calculate the effect size, otherwise use the F-value for the group factor.	estES_fu1
<b>Polarity - FUI</b>	-1 = <b>Negative</b> (The scale to measure the outcome follows the principle: Higher values stand for impairment) 0 = <b>Not clear</b> (No direction indicated, e.g. when using opinion ratings) 1 = <b>Positive</b> (The scale to measure the outcome follows the principle: Higher values stand for improvement)	polar_fu1
<b>Information for ES estimation - FUI</b>	1 = <b>Mean, SD and N for IG and CG prior and after the program are reported</b> 2 = <b>Mean, SD and N for IG and CG after the program are reported</b> 3 = <b>Effect size is reported</b> 4 = <b>Test statistic is reported</b> a) Mean value test (e.g. $F$ , $t$ ) b) Measures of associations (e.g. $r$ , $\chi^2$ ) c) Proportions (e.g. $OR$ ) d) Non-parametric statistics (e.g. $U$ ) 5 = <b>Significance level is zero</b>	infES_fu1

	6 = <b>Estimation by the coder: d = 0.2(small)/0.5(medium)/0.8(high) effect</b>	
<b>Significance - FUI</b> <i>(explicitly mentioned)</i>	0 = <b>Effect is not significant</b> 1 = <b>Effect is significant</b> Note: See significance - post.	sig_fu1
<b>Assessment time point follow up II</b>	In months after program is completed	astp_fu2
<b>IG FUII Mean</b>		IG_fu2M
<b>IG FUII SD</b>		IG_fu2SD
<b>IG FUII N</b>	If N is only reported at randomization and exact attrition rate (AR) for FUII is known, calculate N under consideration of AR.	IG_fu2N
<b>CG FUII Mean</b>		CG_fu2M
<b>CG FUII SD</b>		CG_fu2SD
<b>CG FUII N</b>	If N is only reported at randomization and exact attrition rate (AR) for FUII is known, calculate N under consideration of AR.	CG_fu2N
<b>Estimated effect size - FUII</b>	Fill out only, if "Information for ES estimation - post" = 3, 4, 5 or 6 Note: Please insert the sample size for the estimated effect size in the fields for C1 effect size (for IG: "C1 IG FUII N"; for CG: "C1 CG FUII N"). If estimation is based on results of an AN(C)OVA only use F-tests with 1 numerator degrees of freedom. If an interaction factor for group (IG vs. CG) x time (pre vs. post/FU) is reported, preferably use this F-value to calculate the effect size, otherwise use the F-value for the group factor.	estES_fu2
<b>Polarity - FUII</b>	-1 = <b>Negative</b> (The scale to measure the outcome follows the principle: Higher values stand for impairment) 0 = <b>Not clear</b> (No direction indicated, e.g. when using opinion ratings) 1 = <b>Positive</b> (The scale to measure the outcome follows the principle: Higher values stand for improvement)	polar_fu2
<b>Information for ES estimation - FUII</b>	1 = <b>Mean, SD and N for IG and CG prior and after the program are reported</b> 2 = <b>Mean, SD and N for IG and CG after the program are reported</b> 3 = <b>Effect size is reported</b> 4 = <b>Test statistic is reported</b> a) Mean value test (e.g. $F$ , $t$ ) b) Measures of associations (e.g. $r$ , $\chi^2$ ) c) Proportions (e.g. $OR$ ) d) Non-parametric statistics (e.g. $U$ ) 5 = <b>Significance level is zero</b> 6 = <b>Estimation by the coder: d = 0.2(small)/0.5(medium)/0.8(high) effect</b>	infES_fu2

<b>Significance - FUII</b> <i>(explicitly mentioned)</i>	<b>0 = Effect is not significant</b> <b>1 = Effect is significant</b> Note: See significance - post.	sig_fu2
-------------------------------------------------------------	------------------------------------------------------------------------------------------------------------	---------

VALIDITY EVALUATION

VARIABLE	CODING VALUES, DESCRIPTIONS AND NOTES	CODE
<p><b>Standardization</b> <i>Did the author use a standardized measurement?</i></p>	<p>1 = <b>Standardized measurement</b> (The applied measure was standardized in another (cited) study to measure the target outcomes (e.g. the author applied the Strange Situation (Ainsworth, Blehar, Waters, &amp; Wall, 1978) to assess infant-mother attachment or cites Achenbach using the (corresponding scales of the) Child Behavior Checklist to measure internalizing and externalizing problems)</p> <p>2 = <b>Adapted standardized measurement</b> (A standardized measurement was adapted by the author for reasons of e.g. language (an English questionnaire translated for Spanish participants), cultural (different religious examples for Christian and Muslim children) or time reason (measurement is too long for the program in its original form); also code this category if a standardized interview (including e.g. a coding form and/or several trained coders) has been conducted</p> <p>3 = <b>Ad hoc</b> (The measure is self-constructed. This means either no other study is cited using this measure or cited scales are modified or applied to measure another construct, e.g. using the internalizing problems scale of the Child Behavior Checklist to measure externalizing problems; also medical record information, such as pregnancy test; also code this category if an unstandardized interview has been conducted)</p>	stand
<p><b>Level of effect</b> <i>On which level of impact is the outcome?</i></p>	<p>1 = <b>Learning level</b> (The impact of the program on knowledge, attitudes or skills, e.g. knowledge about risks of drug-taking, drug refusal skills, attitudes towards drugs, problem solving skills, attitudes towards math)</p> <p>2 = <b>Behavior Level</b> (The extent to which the acquired knowledge, attitude or skill is applied in real-life contexts and behavior outside a specific "testing situation" is changed; e.g. attending school, (frequency of) language usage, having unprotected sex within the last year, taking drugs within the last year, beating children within the last year, Becks Depression Inventory, strange situation)</p> <p>3 = <b>Results level</b> (The extent to which the program affects socially desired results that are partially but not exclusively influenced by participants knowledge, attitudes, skills or behavior; e.g. grades at school, get pregnant, get a job, get incarcerated, presence of books)</p>	loef
<p><b>Proximity</b> <i>How close is the measured outcome to the contents and aims of the program?</i></p>	<p>1 = <b>very distal</b> (If none of the three below mentioned questions can be answered with "yes")</p> <p>2 = <b>distal</b> (If only one of the three below mentioned questions can be answered with "yes")</p> <p>3 = <b>close</b> (If two of the three below mentioned questions can be answered with "yes")</p> <p>4 = <b>very close</b> (If all below mentioned questions can be answered with "yes")</p> <p><u>Use the following three aspects to rate:</u></p> <ul style="list-style-type: none"> <li>- Is the addressee of the program the tested person?</li> <li>- Yes, if for example a child training is conducted and the child's skill is also measured.</li> <li>- No, if for example a parent training is conducted and the child is tested.</li> </ul>	prox



- Is the outcome close to the aims of the program?
    - Yes, if author's intention is for example to improve reading and reading ability is measured.
    - No, if author's intention is for example the improvement of parenting skills and the aggressive behavior of the child is measured.
  - Is the outcome close to the contents of the program?
    - Yes, if for example a social competence training is being conducted and social competence is measured.
    - No, if for example a social competence training is being conducted and aggressive behavior is measured.
-

## APPENDIX A

## Peer reviewed journals

Acta Paediatrica	AUSTRALIAN JOURNAL OF PRIMARY HEALTH
Addiction	Australian Occupational Therapy Journal
Addictive Behaviors	BASIC AND APPLIED SOCIAL PSYCHOLOGY
Administration & Policy in Mental Health	Behavior therapy
AIDS and Behavior	Behaviour Research & Therapy
AIDS CARE-PSYCHOLOGICAL AND SOCIO-MEDICAL ASPECTS OF AIDS/HIV	Behavioural and Cognitive Psychotherapy
Ambulatory Pediatrics	Bilingual Research Journal
American Behavioral Scientist	BMC Public Health
American Educational Research Journal	BMJ (British Medical Journal)
American journal of community psychology	British Journal of Educational Psychology
American Journal of Health Behavior	BRITISH JOURNAL OF EDUCATIONAL TECHNOLOGY
American Journal of Health Promotion	Bulletin of the World Health Organization
American Journal of Mental Deficiency	Cancer Epidemiol Biomarkers Prev (Cancer epidemiology, biomarkers & program : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology)
American journal of obstetrics and gynecology	CESKOSLOVENSKA PSYCHOLOGIE
American Journal of Occupational Therapy	Chest
American Journal of Preventive Medicine	Child & Family Behavior Therapy
American Journal of Public Health	Child Abuse & Neglect
American Journal of Speech Language Pathology	CHILD AND ADOLESCENT MENTAL HEALTH
AMERICAN PSYCHOLOGIST	CHILD CARE HEALTH AND DEVELOPMENT
ANNALS OF EMERGENCY MEDICINE	Child Development
Annals of Family Medicine	Child Psychiatry and Human Development
Annual Progress in Child Psychiatry & Child Development	Child Study Journal
ANNUAL REVIEW OF CLINICAL PSYCHOLOGY	Child Welfare
APPLIED DEVELOPMENTAL SCIENCE	Child: Care, Health & Development
Applied Psycholinguistics	Children & Schools
Archives of Disease in Childhood	Children & Society
Archives of General Psychiatry	Children and Youth Services Review
Archives of Pediatrics & Adolescent Medicine	Children's Health Care
Asian Pacific Journal of Cancer Program: Apjcp	
Assessment for Effective Program	

Chinese Journal of Clinical Psychology  
 Chinese Mental Health Journal  
 CIN: Computers, Informatics, Nursing  
 CLINICAL OBSTETRICS AND GYNECOLOGY  
 CMAJ Canadian Medical Association Journal  
 COGNITION AND INSTRUCTION  
 Cognitive and Behavioral Practice  
 Communication Disorders Quarterly  
 COMMUNITY MENTAL HEALTH JOURNAL  
 Computers & Education  
 Contemporary Educational Psychology  
 CONTRACEPTION  
 Counselling Psychology Quarterly  
 Das Gesundheitswesen  
 Development & Psychopathology  
 Developmental Psychology  
 Drug & Alcohol Dependence  
 Early Child Development and Care  
 Early Childhood Education Journal  
 Early Childhood Research Quarterly  
 Early Education and Development  
 ECONOMIC JOURNAL  
 Economics of Education Review  
 Education  
 Education and Information Technologies  
 EDUCATIONAL ADMINISTRATION  
 QUARTERLY  
 Educational Evaluation and Policy Analysis  
 Educational Media International  
 EDUCATIONAL REVIEW  
 EDUCATIONAL STUDIES IN MATHEMATICS  
 Educational Technology Research and  
 Development  
 Elementary School Guidance & Counseling  
 Elementary School Journal  
 Enfance  
 Ethnicity & Disease  
 Exceptional Children  
 Exceptionality  
 Familiendynamik  
 Family & Community Health: The Journal of  
 Health Promotion & Maintenance  
 Family Process  
 Family Relations: An Interdisciplinary Journal of  
 Applied Family Studies  
 FOREIGN LANGUAGE ANNALS  
 Future of Children  
 Genetic Psychology Monographs  
 GRUPPENPSYCHOTHERAPIE UND  
 GRUPPENDYNAMIK  
 Health Care for Women International  
 Health Educ Res (Health education research)  
 HEALTH POLICY AND PLANNING  
 Health Promotion International  
 Health Promotion Practice  
 Hispanic Health Care International  
 Hispanic Journal of Behavioral Sciences  
 HOUSING POLICY DEBATE  
 Human Organization  
 Human Relations  
 Indian Journal of Clinical Psychology  
 Individual Psychologist  
 Infant and Child Development  
 Infant Behavior & Development  
 Infant Mental Health Journal  
 INFANTS AND YOUNG CHILDREN  
 INSTRUCTIONAL SCIENCE  
 Intelligence  
 INTERNATIONAL JOURNAL FOR EQUITY IN  
 HEALTH

International Journal of Behavioral Development  
 International Journal of Cognitive Education & Mediated Learning  
 International Journal of Evidence Based Healthcare  
 International Journal of Group Psychotherapy  
 International Journal of Language & Communication Disorders  
 International journal of nursing studies  
 International Journal of Play Therapy  
 INTERNATIONAL JOURNAL OF PSYCHOLOGY  
 INTERNATIONAL NURSING REVIEW  
 JAMA-JOURNAL OF THE AMERICAN MEDICAL ASSOCIATION  
 Journal of Abnormal Child Psychology  
 Journal of Adolescent Health  
 Journal of Advanced Academics  
 JOURNAL OF ALTERNATIVE AND COMPLEMENTARY MEDICINE  
 Journal of Applied Developmental Psychology  
 JOURNAL OF APPLIED SPORT PSYCHOLOGY  
 Journal of Behavioral Education  
 Journal of Child & Adolescent Substance Abuse  
 Journal of Child and Family Studies  
 Journal of Child Language  
 Journal of Child Psychology and Psychiatry  
 Journal of Classroom Interaction  
 Journal of Clinical Child & Adolescent Psychology  
 Journal of Clinical Child Psychology  
 Journal of Clinical Nursing  
 Journal of Clinical Psychology  
 Journal of College Student Personnel  
 Journal of Community Health Nursing  
 Journal of Community Psychology  
 Journal of Computer Assisted Learning  
 Journal of Consulting & Clinical Psychology  
 Journal of Counseling Psychology  
 JOURNAL OF CYSTIC FIBROSIS  
 JOURNAL OF DEVELOPMENT ECONOMICS  
 Journal of Developmental & Behavioral Pediatrics  
 Journal of Drug Education  
 Journal of Early Adolescence  
 Journal of Early Program  
 Journal of Education for Students Placed at Risk  
 Journal of Educational Computing Research  
 Journal of Educational Psychology  
 JOURNAL OF ENVIRONMENTAL HEALTH  
 Journal of Epidemiology & Community Health  
 Journal of Family Nursing  
 Journal of Family Psychology  
 Journal of Health Psychology  
 Journal of Korean Academy of Nursing  
 Journal of Learning Disabilities  
 Journal of Literacy Research  
 JOURNAL OF MARRIAGE AND FAMILY  
 Journal of Multicultural Counseling and Development  
 Journal of Music Therapy  
 Journal of Negro Education  
 Journal of Neural Transmission  
 Journal of Organizational Behavior  
 Journal of Pediatric Health Care  
 Journal of Pediatric Psychology  
 Journal of Perinatology  
 Journal of Personality & Social Psychology  
 Journal of Program & Program in the Community  
 Journal of Primary Program  
 JOURNAL OF PUBLIC ECONOMICS  
 Journal of Public Health  
 Journal of Public Health Medicine

Journal of Research in Childhood Education  
 Journal of Research in Music Education  
 JOURNAL OF RESEARCH IN PERSONALITY  
 Journal of Research in Reading  
 JOURNAL OF RESEARCH IN SCIENCE  
 TEACHING  
 Journal of Research on Educational Effectiveness  
 Journal of Scandinavian Studies in Criminology  
 and Crime Program  
 Journal of School Health  
 Journal of School Psychology  
 Journal of Science & Medicine in Sport  
 Journal of Social Issues  
 Journal of Social Service Research  
 Journal of Special Education  
 Journal of Speech & Hearing Research  
 JOURNAL OF SPEECH LANGUAGE AND  
 HEARING RESEARCH  
 Journal of Studies on Alcohol  
 Journal of the American Academy of Child &  
 Adolescent Psychiatry  
 Journal of the Indian Academy of Applied  
 Psychology  
 Journal of the National Medical Association  
 JOURNAL OF VOCATIONAL BEHAVIOR  
 Journal of Youth and Adolescence  
 Kind en Adolescent  
 Lancet  
 Language Arts  
 LANGUAGE LEARNING  
 Language, Speech, and Hearing Services in  
 Schools  
 Learning and Individual Differences  
 LEARNING AND INSTRUCTION  
 Learning Disabilities Research & Practice  
 Learning Disability Quarterly  
 Maternal & Child Health Journal  
 Mathematica Policy Research, Inc  
 Medical Care  
 Mental Retardation  
 Merrill-Palmer Quarterly  
 Middle Grades Research Journal  
 Monographs of the Society for Research in Child  
 Development  
 Music Perception  
 New Directions for Child & Adolescent  
 Development  
 NHSA Dialog  
 Nursing Research  
 Obesity  
 Omega (Westport) (Omega)  
 PARENTING-SCIENCE AND PRACTICE  
 Pediatrics  
 Perceptual and Motor Skills  
 Perspectives on Sexual & Reproductive Health  
 POPULATION RESEARCH AND POLICY  
 REVIEW  
 Program Science  
 Preventive Medicine: An International Journal  
 Devoted to Practice and Theory  
 Prevention Science  
 Psicologia: Reflexão e Crítica  
 Psychology in the Schools  
 Psychology of Addictive Behaviors  
 Psychology, Health & Medicine  
 Psychoneuroendocrinology  
 PSYCHOTHERAPY  
 Public Health Nursing  
 Public Health Reports  
 QUARTERLY JOURNAL OF ECONOMICS  
 Reading and Writing: An Interdisciplinary  
 Journal  
 Reading Psychology  
 Reading Research and Instruction

Reading Research Quarterly  
 Reading Teacher  
 Remedial and Special Education  
 Res Nurs Health (Research in nursing & health)  
 RESEARCH IN HIGHER EDUCATION  
 RESEARCH IN SCIENCE EDUCATION  
 Research in Social Stratification and Mobility  
 Research on Social Work Practice  
 Review of Educational Research  
 REVIEW OF HIGHER EDUCATION  
 REVISTA DA ESCOLA DE ENFERMAGEM DA USP  
 Revista de Psicología  
 Revista de Psicología Social Aplicada  
 Sao Paulo Medical Journal = Revista Paulista de Medicina  
 Scandinavian Journal of Public Health  
 School Mental Health  
 School Psychology Review  
 SCIENTIFIC STUDIES OF READING  
 Social Behavior and Personality  
 Social Development  
 Social Science & Medicine  
 Social Science Quarterly  
 SOCIAL SCIENCE RESEARCH  
 Social Service Review  
 Social work research  
 South African Journal of Psychology  
 Special Services in the Schools  
 Sportunterricht  
 Substance use & misuse  
 Support for Learning  
 Teachers College Record  
 TECHNOLOGY PEDAGOGY AND EDUCATION  
 The American Journal of Occupational Therapy  
 The Journal of Educational Research  
 The Journal of Genetic Psychology: Research and Theory on Human Development  
 The Journal of Mental Health Policy & Economics  
 Topics in Early Childhood Special Education  
 Transcultural Psychiatry  
 Trials  
 Universitas Psychologica  
 URBAN AFFAIRS REVIEW  
 Urban Education  
 Verhaltenstherapie und psychosoziale Praxis  
 Victims & Offenders  
 West J Nurs Res (Western journal of nursing research)  
 WORLD BANK ECONOMIC REVIEW  
 Young Children

**Probably non-peer-reviewed journals**

California Journal of Educational Research

Education and Training of the Mentally Retarded

Humanitas: Journal for Research in the Human Sciences

Improving Human Performance

Irish Journal of Education

Israeli Journal of Psychology & Counseling in Education

Journal of Child Psychology & Psychiatry & Allied Disciplines

Occupational Psychology

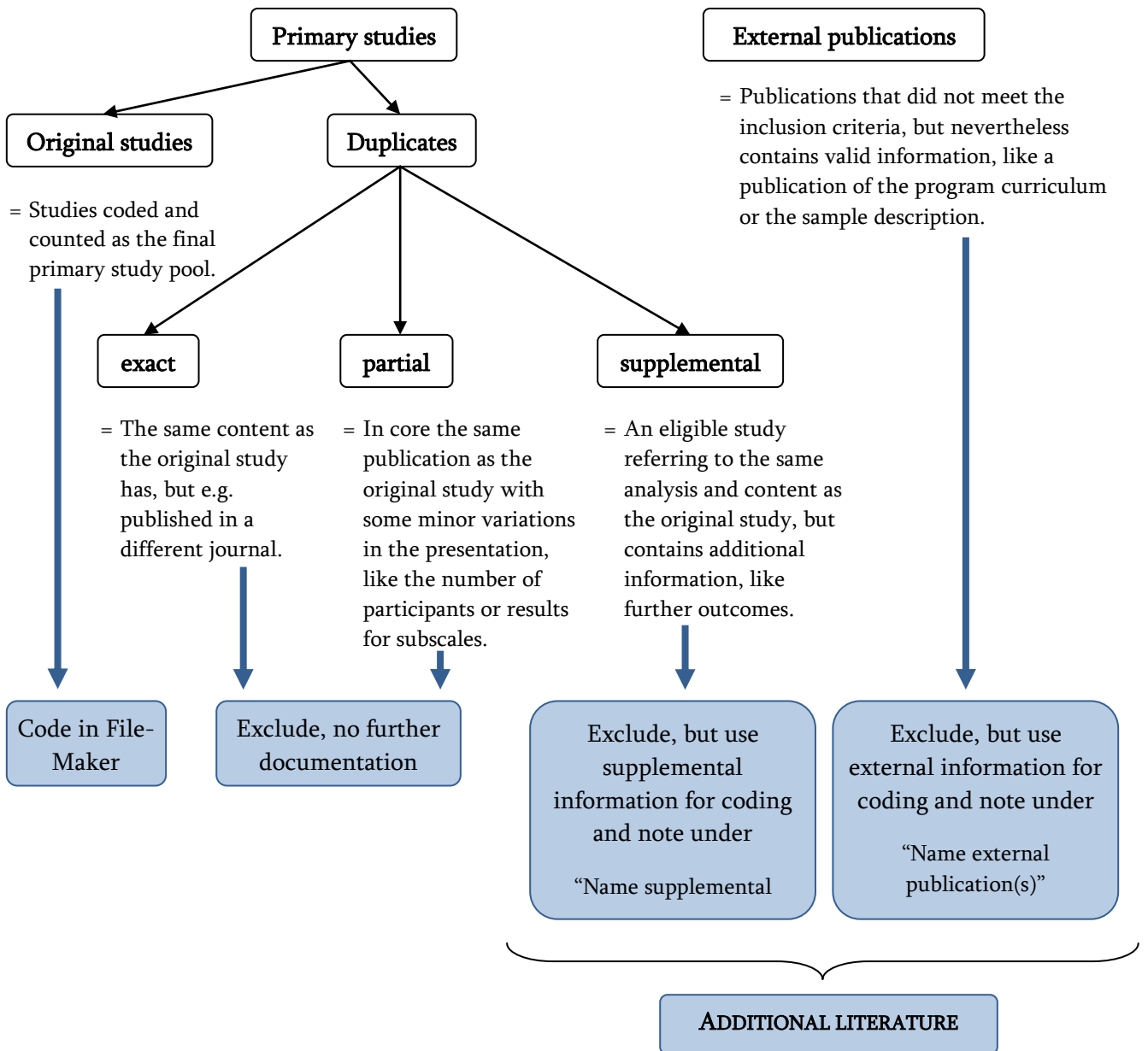
ODGOJNE ZNANOSTI-EDUCATIONAL SCIENCES

OPHTHALMIC EPIDEMIOLOGY

Papers in the Social Sciences

Psychiatry

APPENDIX B  
Literature types





## APPENDIX C

## Poverty thresholds for the operationalization by income

**Table A:** Income poverty threshold in US Dollar (annually and per person)

Land	bis 1979	1980-89	1990-99	2000-11
Australien	2 330	4 830	7 445	11 973
Belgien	2 526	5 467	8 966	13 313
Bulgarien	n.v.	n.v.	n.v.	1 436
Chile	n.v.	n.v.	3 090	4 158
Dänemark	3 282	7 030	11 141	17 199
Deutschland	2 438	5 419	8 791	12 994
Estland	n.v.	n.v.	1 383	2 859
Finnland	2 190	5 205	7 664	13 674
Frankreich	2 408	5 237	8 348	12 504
Griechenland	1 599	3 147	4 649	7 453
Indien	n.v.	n.v.	n.v.	1 019
Indonesien	n.v.	n.v.	n.v.	1 278
Irland	1 788	3 753	7 575	14 869
Island	3 969	9 732	15 063	20 137
Israel	n.v.	n.v.	7 281	8 926
Italien	2 170	5 004	8 175	11 518
Japan	1 827	4 526	7 731	10 268
Kanada	2 489	5 404	7 992	12 395
Korea	491	1 696	4 467	8 501
Kroatien	n.v.	n.v.	n.v.	3 655
Lettland	n.v.	n.v.	n.v.	3 144
Litauen	n.v.	n.v.	n.v.	2 862
Luxemburg	2 440	7 080	14 243	22 076
Malta	n.v.	n.v.	n.v.	7 861
Mexiko	1 043	2 098	2 973	4 770
Neuseeland	2 066	4 022	5 746	8 758
Niederlande	2 287	4 817	8 027	13 650
Norwegen	2 314	5 695	10 033	21 294
Österreich	2 559	5 788	9 388	14 112
Polen	n.v.	n.v.	1 356	2 615
Portugal	880	1 954	3 787	5 847
Rumänien	n.v.	n.v.	n.v.	1 640
Russland	n.v.	n.v.	1 932	4 823
Schweden	2 798	5 822	8 576	13 663
Schweiz	4 934	10 124	14 701	21 156
Slowakei	n.v.	n.v.	1 529	3 023
Slowenien	n.v.	n.v.	4 456	7 092

Spanien	1 361	2 808	4 964	8 370
Südafrika	n.v.	1 719	2 099	3 119
Tschechische Republik	n.v.	n.v.	2 679	4 200
Türkei	n.v.	n.v.	n.v.	1 283
Ungarn	n.v.	n.v.	1 601	2 813
Vereinigte Staaten von Amerika	2 895	6 336	10 355	15 753
Vereinigtes Königreich	2 190	4 638	7 896	13 294
Zypern	n.v.	n.v.	n.v.	12 330

*Anmerkungen.* n.v. = nicht verfügbar, dunkelbraun unterlegt = länderspezifische Schätzung (Schätzer A), hellbraun unterlegt = länderübergreifende Schätzung (Schätzer B), geld unterlegt = durchschnittliche Eurostat-Armutsschwelle.

**Table B:** Alternative sources for the income poverty threshold

Informationsquelle	Verfügbar unter
World Income Inequality Database (The United Nations University World Institute for Development Economics Research, UNU-WIDER)	<a href="http://www.wider.unu.edu/research/Database/en_GB/wiid/">http://www.wider.unu.edu/research/Database/en_GB/wiid/</a>
The World Top Income Database (Paris School of Economics)	<a href="http://mond.parisschoolofeconomics.eu/topincomes/">http://mond.parisschoolofeconomics.eu/topincomes/</a>
Country and Lending Groups by Income (The World Bank)	<a href="http://data.worldbank.org/about/country-classifications/country-and-lending-groups#Low_income">http://data.worldbank.org/about/country-classifications/country-and-lending-groups#Low_income</a>

*Note.* All websites accessed on 26.08.2013.

## APPENDIX D

**Table A:** Education poverty threshold in number of years spent in educational system or degree attained

number of years/ degree attained	Country
<b>0 years/no degree attained</b>	Afghanistan*, Algeria, Bahrain, Bangladesh, Benin, Bhutan, Bolivia, Burkina Faso, Cambodia, Cameroon*, Chad, Central African Republic*, Dominican Republic, Ecuador, Ethiopia, Gambia*, Guatemala, Guinea-Bissau*, Haiti*, Honduras, India*, Jordan, Kenya, Lebanon, Lesotho, Malawi, Maldives, Mali, Mauritius, Mozambique*, Namibia, Nepal*, Nicaragua*, Niger*, Oman, Pakistan, Papua New Guinea*, Portugal, Saudi Arabia, Senegal, Sierra Leone*, Sudan*, Suriname, Syrian Arab Republic, Togo, Tunisia, Turkey, United Arab Emirates, United Republic of Tanzania, Zambia, Zimbabwe
<b>6 years/primary finished</b>	Albania, Andorra, Anguilla, Argentina, Armenia, Barbados, Belgium, Bermuda, Bosnia and Herzegovina, Botswana*, Brazil, Chile, China, China, Hong Kong SAR, China, Macao SAR, Colombia, Costa Rica, Croatia, Cuba, Cyprus, Dominica, El Salvador, Fiji, France, Ghana*, Greece, Guyana, Indonesia, Iran (Islamic Republic of), Ireland, Israel, Italy, Kuwait, Liberia*, Malaysia, Malta, Mexico, Mongolia, Palestine, Panama, Paraguay, Peru, Philippines, Poland, Qatar, Republic of Korea, Romania, Samoa, Serbia, Seychelles, Singapore, South Africa, Spain, Sri Lanka, Swaziland*, Thailand, The former Yugoslav Republic of Macedonia, Tonga, Trinidad and Tobago, Uganda, Ukraine, Uruguay, Venezuela
<b>10 years/lower secondary finished</b>	Australia, Austria, Azerbaijan, Bahamas, Belarus, Bulgaria, Canada, Cayman Islands, Czech Republic, Denmark, Estonia, Finland, Georgia, Germany, Hungary, Iceland, Iraq*, Jamaica*, Japan, Kazakhstan, Kyrgyzstan, Latvia, Lithuania, Luxembourg, Montenegro, Netherlands**, New Zealand, Norway, Puerto Rico, Republic of Moldova, Russian Federation, Slovakia, Slovenia, Sweden**, Switzerland, Tajikistan, United Kingdom, United States of America

---

**No data available** Angola, Antigua and Barbuda, Aruba, Belize, British Virgin Islands, Brunei Darussalam, Burundi, Cape Verde, Comoros, Congo, Cook Islands, Côte d'Ivoire, Democratic People's Republic of Korea, Democratic Republic of the Congo, Djibouti, Egypt, Equatorial Guinea, Eritrea, Gabon, Gibraltar, Grenada, Guinea, Holy See, Kiribati, Lao People's Democratic Republic, Libya, Liechtenstein, Madagascar, Marshall Islands, Mauritania, Micronesia (Federated States of), Monaco, Montserrat, Morocco, Myanmar, Nauru, Antilles, Nigeria, Niue, Palau, Rwanda, Saint Kitts and Nevis, Saint Lucia, Saint Vincent and the Grenadines, San Marino, Sao Tome and Principe, Solomon Islands, Somalia, Timor-Leste, Tokelau, Turkmenistan, Turks and Caicos Islands, Tuvalu, Uzbekistan, Vanuatu, Viet Nam, Yemen

---

*Note.* Sources are the UNESCO Institute for statistics<sup>1</sup> (countries based on this data are not marked) or data by Barro, R. & Lee, J.W.<sup>2</sup> (2013, countries based on this data are marked with an asterisk \*); \*\* = based on UNSECO but change because auf plausibility based on Barro&Lee; All sources accessed on 26.08.2013.

---

<sup>1</sup> <http://stats.uis.unesco.org/unesco/ReportFolders/ReportFolders.aspx> and <http://stats.uis.unesco.org/unesco/TableViewer/tableView.aspx?ReportId=185>

<sup>2</sup> <http://www.barrolee.com/data/full1.htm>

**Table B:** Reference years for educational poverty threshold

Afghanistan	2000	Central African Republic	2000	Guyana	2002
Albania	2011	Chad	2004	Haiti	2000
Algeria	2006	Chile	2010	Honduras	2011
Andorra	2006	China	2010	Hungary	2011
Anguilla	2001	China, Hong Kong SAR	2010	Iceland	2005
Argentina	2003	China, Macao SAR	2006	India	2000
Armenia	2001	Colombia	2011	Indonesia	2009
Australia	2011	Costa Rica	2011	Iran (Islamic Republic of)	2010
Austria	2010	Croatia	2001	Iraq	2000
Azerbaijan	2009	Cuba	2002	Ireland	2010
Bahamas	2000	Cyprus	2010	Israel	2010
Bahrain	2001	Czech Republic	2011	Italy	2010
Bangladesh	2001	Denmark	2011	Jamaica	2000
Barbados	2000	Dominica	2001	Japan	2010
Belarus	2009	Dominican Republic	2011	Jordan	2010
Belgium	2010	Ecuador	2010	Kazakhstan	2007
Benin	2002	El Salvador	2010	Kenya	2010
Bermuda	2010	Estonia	2011	Kuwait	2008
Bhutan	2005	Ethiopia	2007	Kyrgyzstan	2009
Bolivia	2009	Fiji	2007	Latvia	2011
Bosnia and Herzegovina	2010	Finland	2009	Lebanon	2007
Botswana	2000	France	2010	Lesotho	2008
Brazil	2010	Gambia	2000	Liberia	2000
Bulgaria	2010	Georgia	2002	Lithuania	2011
Burkina Faso	2007	Germany	2010	Luxembourg	2010
Cambodia	2009	Ghana	2000	Malawi	1998
Cameroon	2000	Greece	2010	Malaysia	2010
Canada	2006	Guatemala	2006	Maldives	2006
Cayman Islands	2008	Guinea-Bissan	2000		

Mali	2006	Portugal	2010	Switzerland	2010
Malta	2011	Puerto Rico	2008	Syrian Arab Republic	2009
Mauritius	2000	Qatar	2011	Tajikistan	2000
Mexico	2010	Republic of Korea	2010	Thailand	2006
Mongolia	2000	Republic of Moldova	2010	The former Yugoslav	
Montenegro	2006	Romania	2010	Republic of	
Mozambique	2000	Russian Federation	2010	Macedonia	2002
Namibia	2001	Samoa	2001	Togo	2009
Nepal	2000	Saudi Arabia	2004	Tonga	2006
Netherlands	2000	Senegal	2006	Trinidad and Tobago	2009
New Zealand	2011	Serbia	2011	Tunisia	2010
Nicaragua	2000	Seychelles	2002	Turkey	2009
Niger	2000	Sierra Leone	2000	Uganda	2010
Norway	2010	Singapore	2011	Ukraine	2001
Oman	2008	Slovakia	2001	United Arab Emirates	2005
Pakistan	2009	Slovenia	2011	United Kingdom	2010
Palestine	2011	South Africa	2011	United Republic of	
Panama	2010	Spain	2010	Tanzania	2002
Papua New Guinea	2000	Sri Lanka	2009	United States of America	2010
Paraguay	2008	Sudan	2000	Uruguay	2010
Peru	2010	Suriname	2004	Venezuela	2009
Philippines	2008	Swaziland	2000	Zambia	2000
Poland	2011	Sweden	2010	Zimbabwe	2002

**Table C:** Alternative sources for the education poverty threshold

Informationsquelle	Verfügbar unter
Data set "Education attainment for population aged 25 and over" from Barro, R.J. & Lee, J.-W.	<a href="http://www.barrolee.com/data/full1.htm">http://www.barrolee.com/data/full1.htm</a>

**Commentary**

## **EHRENWÖRTLICHE ERKLÄRUNG**

Hiermit erkläre ich, Sebastian Schulz, ehrenwörtlich, dass mir die geltende Promotionsordnung der Fakultät für Sozial- und Verhaltenswissenschaften der Friedrich-Schiller-Universität Jena bekannt ist, dass ich die Dissertation selbst angefertigt habe, keine Textabschnitte eines Dritten oder eigener Prüfungsarbeiten ohne Kennzeichnung übernommen und alle von mir benutzten Hilfsmittel, persönlichen Mitteilungen und Quellen in meiner Arbeit angegeben habe.

Das im Anhang der Arbeit abgedruckte Kodierschema wurde in Zusammenarbeit mit Sabrina Maichrowitz und Louisa Arnold im Rahmen des vom BMBF geförderten und von Prof. Dr. Andreas Beelmann geleiteten Projektes „Wirksamkeit von Bildungs- und Interventionsprogrammen zur Prävention und Kompensation von Armuts- und Migrationsfolgen bei Kindern und Jugendlichen. Ein Forschungsüberblick und eine Meta-Analyse internationaler Evaluationsforschung“ (Förderkennzeichen: 01JC1105) erstellt. Zudem unterstützte mich Christiane Fischer unentgeltlich bei der formalen Korrektur (Rechtschreibung und Grammatik) der Arbeit. Darüber hinaus haben mich keine weiteren Personen bei der Auswahl und Auswertung des Materials sowie bei der Herstellung des Manuskriptes unterstützt.

Zudem erkläre ich ehrenwörtlich, dass ich die Hilfe eines Promotionsberaters nicht in Anspruch genommen habe und Dritte weder unmittelbar noch mittelbar geldwerte Leistungen von mir erhalten haben, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen.

Ich habe die Dissertation noch nicht als Prüfungsarbeit für eine staatliche oder andere wissenschaftliche Prüfung eingereicht und auch nicht eine gleiche, eine in wesentlichen Teilen ähnliche oder eine andere Abhandlung bei einer anderen Hochschule bzw. anderen Fakultät als Dissertation eingereicht.

Jena, 10.08.2016



## DANKSAGUNG

Am Ende dieser Arbeit möchte ich mich noch bei all jenen bedanken, die mich hierbei unterstützt und diese Dissertation erst ermöglicht haben.

Als erstes möchte ich meinem wissenschaftlichen Betreuer Prof. Dr. Andreas Beelmann danken, der als Leiter des Projektes, aus dem diese Dissertation hervorgeht, den Grundstein für diese Arbeit legte und mir bei Fragen stets mit seiner forschungs-integrativen Fachkompetenz zur Seite stand. Ebenfalls möchte ich Prof. Dr. Daniel Haun sowie Prof. Dr. Peter Noack für die unkomplizierte Übernahme der Gutachterfunktion danken.

Weiterhin danke ich allen an dem Projekt PrAMi beteiligten Personen. Insbesondere danke ich meinen Kolleginnen Sabrina Maichrowitz und Louisa Arnold sowie den wissenschaftlichen Hilfskräften Sandra Renas, Katrin Konold und Angelika Schulz für die Unterstützung bei der Kodierung der Studien sowie für die anregenden Diskussionen bezüglich inhaltlicher und kodierspezifischer Fragen, durch die sich mein Verständnis von Meta-Analysen weiterentwickelte.

Als letztes möchte ich schließlich allen danken, die mich während der Zeit privat begleitet haben. Meinen Eltern bin ich dankbar für die Finanzierung meines Studiums sowie die immerwährende, häufig sehr wichtige emotionale Unterstützung. Meiner Lebenspartnerin Christiane Fischer möchte ich dafür danken, dass sie da war, wenn sich kleinere und größere Sinnkrisen ergaben sowie für ihre konstruktive Begleitung dieser Arbeit, die zu ihrem Gelingen beigetragen hat.