

Goldstandards für einen wissenschaftlichen Nachweis der Wirksamkeit und Effektivität einer Intervention im Bereich Früher Hilfen als Voraussetzung für ihre Verbreitung

Randomisierte kontrollierte Studien

Sowohl öffentliche Debatten über schwere Fälle der Kindeswohlgefährdung als auch wissenschaftliche Befunde, die empirisch bedeutsame Effekte der sozialen und ökonomischen Risikofaktoren auf die unmittelbare Gefährdung von Kindern sowie auf ihre langfristige Entwicklung im kognitiven und sozial-emotionalen Bereich nachweisen [1], führten dazu, dass die Notwendigkeit Früher Hilfen national und international anerkannt wurde [2, 3].

Ziel von Präventionsmaßnahmen sollte es sein, die Lebensbedingungen und die Entwicklungsbedingungen von Familien und deren Kindern effektiv und anhaltend zu verbessern. Somit ist es unabdingbar, die Wirksamkeit sowie die Effektivität von Maßnahmen im Bereich der Frühen Hilfen wissenschaftlich nachzuweisen. Dies verdeutlichen auch folgende Aspekte:

1. Die Optimierung der Lebenssituation und der Entwicklungschancen zahlreicher Kinder ist das Ziel von Präventions- und Interventionsmaßnahmen [4].
2. Werden ineffektive Interventionsmaßnahmen durchgeführt, können bedeutsame Zeitfenster in der Entwicklung der Kinder verfehlt werden [5, 6].

3. Der finanzielle Aufwand für Präventions- und Interventionsmaßnahmen muss gesellschaftlich gerechtfertigt sein, sodass Interventionsmaßnahmen gezielt eingesetzt werden müssen.
4. Es ist nicht nur wichtig, die Erfolge einer Präventionsmaßnahme zu kennen, sondern auch nachzuweisen, welche negativen Nebenwirkungen eine Intervention möglicherweise hat beziehungsweise welche Faktoren dazu führen, dass die Maßnahme abgebrochen wird [7].
5. Kinder erfahren im Verlauf ihrer natürlichen Entwicklung Kompetenzzuwächse im Bereich Kognition, Selbstkontrolle und im Sozialverhalten. Im Rahmen von Präventions- und Interventionsmaßnahmen muss daher nachgewiesen werden, dass die mit ihrer Hilfe erzielten Kompetenzzuwächse höher sind als die in einer interventionsfreien Kontrollgruppe [8].

Im internationalen Bereich sind zahlreiche Einzelstudien und auch Metaanalysen über Evaluationsergebnisse zu Präventions- und Interventionsmaßnahmen, die im Bereich der Frühen Hilfen anzusiedeln sind, erschienen. Evaluationsstudien gingen unter anderem folgenden

Fragen nach: Sinkt das Misshandlungsrisiko durch Präventions- und Interventionsmaßnahmen? Steigt die mütterliche Feinfühligkeit? Verbessert sich die Mutter-Kind-Interaktion? Für eine Übersicht beziehungsweise Zusammenfassung sei an dieser Stelle auf Lengning und Zimmermann [9] verwiesen. Die Evaluationsergebnisse weisen auf die Wirksamkeit Früher Hilfen hin, wobei die erzielten Effekte im niedrigen bis mittleren Bereich liegen. Leider finden sich aber bei den Untersuchungen auch Mängel. Einige Beispiele sind: 1. Es wurden keine randomisierten Kontrollgruppen aufgenommen. Dadurch ist es nicht möglich zu überprüfen, ob die ermittelten Effekte tatsächlich auf die Intervention zurückzuführen sind (zum Beispiel [10, 11, 12]). 2. Die Gütekriterien (Reliabilität und Validität) der eingesetzten Verfahren waren noch nicht nachgewiesen (zum Beispiel [13]), und 3. es wurde nicht überprüft, ob durch die Intervention Langzeiteffekte erzielt werden können, da keine Follow-up-Messungen durchgeführt wurden (zum Beispiel [14]).

In Deutschland sind vergleichende wissenschaftliche Analysen zur Evaluation im Bereich der Frühen Hilfen eher noch rar beziehungsweise stehen meist in den Anfängen. In der Literatur werden

Zusammenfassung · Abstract

Bundesgesundheitsbl 2010 · 53:1056–1060 DOI 10.1007/s00103-010-1129-5
© Springer-Verlag 2010

A. Lengning

Goldstandards für einen wissenschaftlichen Nachweis der Wirksamkeit und Effektivität einer Intervention im Bereich Früher Hilfen als Voraussetzung für ihre Verbreitung. Randomisierte kontrollierte Studien

Zusammenfassung

Die Notwendigkeit für Frühe Hilfen ist national und auch international anerkannt. Hiermit geht jedoch die legitime Forderung einher, die Wirksamkeit und Effektivität von Präventions- und Interventionsprogrammen nachzuweisen. International liegen hierzu zahlreiche Studien vor, die jedoch zum Teil einige Probleme aufweisen, sodass die Interpretierbarkeit der Ergebnisse häufig eingeschränkt ist. Eine wissenschaftliche Evaluation der Programme in Deutschland fehlt hingegen häufig, oder die Untersuchungen und/oder Auswertungen laufen derzeit. Der vorliegende Beitrag gibt am Beispiel des randomisierten Kontrollgruppendesigns (RCT) einen Überblick darüber, welche „Goldstandards“ für ein Evaluationsdesign, das heißt für den Nachweis der Wirksamkeit und Effektivität einer Prävention/Intervention im Bereich Früher Hilfen wünschenswert sind beziehungsweise wissenschaftlich gefordert

werden. Standards zum Wirksamkeitsnachweis nehmen Bezug auf 1. die Spezifizierung der Wirksamkeit des Programms, 2. die Programmdokumentation, 3. die Überprüfung der Qualität der Maßnahmendurchführung, 4. die Erfassung und Bewertung der erwarteten Veränderungen sowie 5. die Schlussfolgerungen über die Kausalität der Ergebnisse und die Generalisierbarkeit. Standards zum Effektivitätsnachweis beziehen sich auf 1. die Beschreibung der Intervention, 2. die Überprüfung unter Alltagsbedingungen und 3. die Kosten-Nutzen-Analyse. Aufbauend auf einer Darstellung dieser Standards werden Empfehlungen zur Verbreitung/Dissemination von Präventions-/Interventionsprogrammen ausgesprochen.

Schlüsselwörter

Frühe Hilfen · Wirksamkeit · Effektivität · Dissemination · Evaluation

Gold standards for scientific proof of efficacy and effectiveness as prerequisites for the dissemination of intervention programs in the field of early intervention. Randomized controlled trials

Abstract

The necessity of early intervention is recognized both nationally and internationally. This however goes hand in hand with the legitimate request for proving the efficacy and effectiveness of prevention and intervention programs. There are many international studies on this. Some of these exhibit a number of problems, which often restricts the validity of the studies' results. A scientific evaluation of programs in Germany, on the other hand, is often missing entirely or the studies and evaluations are currently in progress. The article at hand using the example of the randomized controlled trial (RCT) gives an overview of which gold standards of evaluation are desirable or should be scientifically demanded to prove efficacy and effectiveness of prevention/intervention in the field of early in-

tervention. Standards for efficacy refer to 1) specification of the program's efficacy, 2) program documentation, 3) evaluating the quality of how the program was carried out, 4) recording and evaluating expected changes, as well as 5) conclusions concerning the causality of results and generalizations. Standards for efficiency refer to 1) description of the intervention, 2) evaluation in real-life contexts and 3) cost-benefit analysis. Based on a presentation of these standards, recommendations are made for the dissemination of prevention and intervention programs.

Keywords

Early intervention · Efficacy · Effectiveness · Dissemination · Evaluation

unterschiedliche Standards diskutiert, die bei Evaluationen angesetzt werden sollen [15, 16, 17]. Als ein „Goldstandard“ wird häufig das randomisierte Kontrollgruppendesign (RCT) angeführt (für eine Definition siehe den Beitrag von H. Ziegler im vorliegenden Heft).

Der vorliegende Beitrag gibt anhand des RCT eine Übersicht darüber, welche Standards für die Evaluation von Präventions- und Interventionsprogrammen wünschenswert beziehungsweise wissenschaftlich notwendig sind in Bezug auf: das Studiendesign, die Überprüfung des Gültigkeitsbereichs der Effekte sowie die Überprüfung der Wirksamkeit und Effektivität. Darüber hinaus werden Empfehlungen zur Dissemination ausgesprochen. Kritisch anzumerken ist jedoch, dass diese Standards in der Praxis gerade im Bereich der Frühen Hilfen aus guten Gründen nicht immer einzuhalten sind und dass diese „Goldstandards“ nicht unumstritten sind (siehe unten). An dieser Stelle sei auf den Beitrag von H. Ziegler in diesem Heft verwiesen, der diese Aspekte kritisch diskutiert. Das Wissen über diese Standards ermöglicht es jedoch, bei der Planung einer Evaluation den Blick für die Optimierung ihres Designs zu schärfen.

Evaluation von Präventions- und Interventionsmaßnahmen

Das Studiendesign

Nationale [8] und internationale [18] Standards fordern ein Studiendesign, das eine Interventions- (beziehungsweise Experimental-) und eine Vergleichsgruppe (-programm) sowohl

- a) in einem Prätest (zu Beginn der Maßnahme),
- b) in einem Posttest (nach Beendigung der Maßnahme) und
- c) in einem Follow-up-Test (zumeist mindestens ein halbes Jahr nach der Beendigung der Maßnahme)

vorsieht. Es wird unterschieden zwischen isolierten, vergleichenden und kombinierten Evaluationen. Während bei den isolierten Evaluationen neben den Interventionsprogrammen Kontrollprogramme, deren Ziele anders als die der

Experimentalprogramme geartet sind, mit aufgenommen werden, beinhalten vergleichende Evaluationen Alternativprogramme, die die gleichen Ziele verfolgen wie das Experimentalprogramm. In kombinierten Evaluationen gehen sowohl Kontrollprogramme als auch Alternativprogramme ein [19].

Überprüfung des Gültigkeitsbereichs der Effekte

Im Rahmen von Evaluationen ist es unabdingbar zu überprüfen, ob sich die im Fokus stehenden Eigenschaften, Fähigkeiten und Begabungen der an einer Intervention beteiligten Personen durch die entsprechende Maßnahme wie beabsichtigt verändert haben. Idealerweise würde sich eine Generalisierung beziehungsweise ein Transfer zeigen [8]. Zum Beispiel könnte eine Maßnahme, die darauf abzielt, die Mutter-Kind-Interaktion in einer strukturierten Situation zu trainieren, sie nicht nur dort verbessern, sondern auch im Alltag. Für den Wirksamkeitsnachweis ist dies wichtig, da sich positive Effekte einer Intervention beziehungsweise Prävention nicht nur in der Interventionssituation und bei äußerer Unterstützung zeigen sollten, sondern auch in der folgenden eigenständigen Lebensführung [20].

Überprüfung der Wirksamkeit und Effektivität

Bei einer Evaluation von Präventions- und Interventionsmaßnahmen muss nicht nur die Wirksamkeit einer Maßnahme unter optimalen Voraussetzungen überprüft werden, sondern auch die Effektivität. So müssen die erzielten Effekte auch unter Alltagsbedingungen nachgewiesen werden [18, 19]. Um letztendlich die Wirkung von Präventions- und Interventionsmaßnahmen an verschiedenen Standpunkten, bei unterschiedlichen Klienten et cetera überprüfen zu können, ist es Voraussetzung, dass ihre kausale positive Wirkung auf die Situation der Familien beziehungsweise Kinder nachgewiesen ist.

Folgende Fragen sollten bei der Überprüfung einer Maßnahme bearbeitet werden:

1. War die Implementierung erfolgreich?
2. War sie wirksam,
 - a) wenn die Maßnahme von geschulten Fachkräften mit optimalem Material durchgeführt wurde?
 - b) wenn die Maßnahme von Personen, die eher wenig geschult waren, durchgeführt wurde und die Ausstattung nicht optimal war?

Wünschenswerte Standards zur Evaluation, die dem Nachweis der Wirksamkeit und Effektivität von Präventions- und Interventionsmaßnahmen dienen, bilden den Gegenstand der folgenden Ausführungen [8, 9, 18, 21, 22].

Wirksamkeit

Standards zum Wirksamkeitsnachweis beziehen sich auf 1. die Spezifizierung der Programmwirksamkeit, 2. die Programmdokumentation, 3. die Überprüfung der Durchführungsqualität der Maßnahme, 4. die Erfassung und Bewertung der erwarteten Veränderungen, 5. die Schlussfolgerungen über die Kausalität der Ergebnisse und 6. die Generalisierbarkeit.

Zu 1. Es müssen konkrete Aussagen zur Wirksamkeit des Programmes getroffen werden. Idealerweise: Das Programm X führt zu einer Veränderung der Effektvariable Y der Population Z.

Zu 2. Das Programm muss so exakt beschrieben werden, dass eine Replikation durch Dritte möglich ist. Dies umfasst die Beschreibung a) der Durchführung, das heißt zum Beispiel die Darstellung der Organisation, des Materials, der Dauer, der Intensität und des Trainingsaufwands, und b) des Wirkungsmodells, das heißt eine Ausführung der kausalen Faktoren.

Zu 3. Es ist notwendig, dass sowohl für die Interventions- als auch für die Kontrollgruppe erfasst wird, wie die entsprechenden Maßnahmen durchgeführt werden. Da erwartet wird, dass die Kontrollgruppe die vermuteten Wirkfaktoren nicht erhält, muss genau geprüft werden, welchen Interventionen sie tatsächlich ausgesetzt oder nicht ausgesetzt war. Nur so kann die Kontrollgruppe den Vergleichsmaß-

stab zur Wirksamkeitsbeurteilung darstellen.

Zu 4.

- a) *Erwartete Wirkung:* Bei Evaluationen ist es wichtig, dass die erwartete Wirkung erfasst wird. Folglich müssen tatsächliche Veränderungen, wie zum Beispiel Veränderungen in der Mutter-Kind-Interaktion oder im Ausmaß der Misshandlung, gemessen werden. Wichtig ist, dass hier Instrumente eingesetzt werden, deren Gütekriterien – also unter anderem ihre Reliabilität und Validität – ausreichend nachgewiesen sind. Unter Umständen kann es jedoch schwierig sein, valide Aussagen über die entsprechenden Variablen zu erhalten. So können Angaben zum Beispiel zur Misshandlung unter Umständen nicht glaubwürdig sein. Hier empfiehlt sich die Erhebung von Proxy-Maßen (dies sind Variablen, die im engen Zusammenhang mit dem Fehlverhalten stehen). Auch die verschiedenen Erhebungsarten (Fragebögen, Interviews und Beobachtungen) für die Variablen bergen unterschiedliche Vor- und Nachteile. Eine Übersicht hierzu findet sich bei Lengning und Zimmermann (2009) [9].
- b) *Stichprobe:* Um Unterschiede zwischen der Interventions- und der Kontrollgruppe zu zeigen, muss die Stichprobe groß genug sein, um ausreichend „Power“ zu haben. Hier ist es nicht nur wichtig, auf die Signifikanz der Effekte zu achten, sondern auch auf die Größe der Effektstärke. Diese zeigt die Stärke des Effektes einer Intervention an und gibt somit Auskunft über ihre Bedeutsamkeit für die Praxis.
- c) *Randomisierung:* Es ist möglich, dass sich in der Vormessung trotz einer Randomisierung Unterschiede in den Mittelwerten der gemessenen Variablen zeigen. Hier ist eine Anpassung der Mittelwerte erforderlich. Darüber hinaus muss eine Auflistung möglicher Unterschiede bei der Teilnehmerzahl im Verlauf der Intervention pro Messzeitpunkt erfolgen.
- d) *Follow-up:* Wie bereits beschrieben, muss ein Evaluationsdesign eine Follow-up-Erhebung einschließen. Erst

- dann können Aussagen darüber getroffen werden, wie lang anhaltend die Effekte der Maßnahme sind beziehungsweise ob sie Effekte erzielt, die sich erst nach einem gewissen Zeitpunkt zeigen (sogenannte Schläfereffekte). Darüber hinaus werden Auswirkungen deutlich, die sich unter Umständen erst in späteren Entwicklungsstadien der Kinder zeigen.
- e) **Veröffentlichungen:** Es ist wichtig, dass die Ergebnisse publiziert werden. Von besonderer Bedeutung sind hier Publikationen, die auch auf praktischer Ebene verständlich sind.
- f) **Objektivität:** Die Evaluation sollte von unabhängigen Personen, also von Personen, die die Intervention nicht umgesetzt haben, durchgeführt werden. Auch sollten die Auswertungen – zum Beispiel die Analyse von Effekten einer Intervention auf die Mutter-Kind-Interaktionen – durch Personen erfolgen, die nicht wissen, ob die Teilnehmer der Kontroll- oder der Interventionsstichprobe angehören.

Zu 5. Es ist wichtig nachzuweisen, dass die ermittelten Effekte wirklich auf die Intervention zurückzuführen sind. Daher ist die randomisierte Zuordnung von Personen zur Interventions- und Kontrollgruppe von besonderer Bedeutung. Diese sollte sicherstellen, dass sich die Kontroll- und Interventionsgruppe nicht bereits vor der Interventionsmaßnahme in den betrachteten Zielvariablen voneinander unterscheiden.

Zu 6. Angaben zur Generalisierbarkeit sind nur möglich, wenn folgende Informationen über die Stichprobe vorliegen:

- das Alter,
- das Geschlecht,
- die ethnische Zusammensetzung und
- der sozioökonomische Status.

Weiterhin sollten Informationen über die Art der Stichprobengewinnung vorliegen. Darüber hinaus wäre es wünschenswert, dass

1. eine Analyse der Effekte für die einzelnen Subgruppen vorliegt, damit die Wirksamkeit und Anwendbarkeit für verschiedene Risikostichproben nachgewiesen ist,

2. Replikationsstudien zur Intervention vorliegen, damit die Ergebnisse gesichert sind und eine Empfehlung auf nationaler Ebene möglich ist.

Effektivität

Es sind aber nicht nur die Standards zum Wirksamkeitsnachweis zu beachten. Zusätzlich muss gewährleistet sein, dass das Interventionsprogramm auch unter Alltagsbedingungen erfolgreich ist. Hier ist zu beachten, dass 1. eine genaue Beschreibung der Intervention vorliegt, 2. dass die Durchführung unter Alltagsbedingungen überprüft wird und 3. dass eine Kosten-Nutzen-Analyse erfolgt.

Zu 1. Manuale zur Beschreibung der Intervention sollten vorliegen, um eine vergleichbare Überprüfung der Vorgehensweise zu ermöglichen. Sie sollten zum Beispiel Informationen über:

- die Implementierung,
- den Aufbau des Netzwerkes und
- die Vorgehensweise der Intervention enthalten.

Zu 2. Da die Interventionen in vielen Fällen nicht nur vom Fachpersonal, sondern auch von Laienhelfern durchgeführt werden, ist zu überprüfen, ob die Qualifikation der Intervenierenden Effekte zeigt. Zudem ist es unter Alltagsbedingungen häufig nicht möglich, die Vorgaben zur Programmdurchführung genau einzuhalten. Folglich sollten Variablen wie unter anderem die Dauer und Häufigkeit der Intervention und die Akzeptanz sowie die Mitwirkung sorgfältig dokumentiert werden, um zum Beispiel das Ausbleiben gewünschter Effekte erklären zu können.

Zu 3. Die Beurteilung der Effektivität einer Maßnahme wird durch den Vergleich mit den Pro-Kopf-Kosten einer anderen Intervention mit einer vergleichbaren Wirkung ermöglicht.

Empfehlungen zur Dissemination

Die Voraussetzungen zur Empfehlung einer Maßnahme sind erfüllt, wenn sowohl ihre Wirksamkeit unter optimalen Bedingungen als auch ihre Effektivität unter Alltagsbedingungen nachgewiesen sind. Weitere Punkte müssen jedoch Be-

rücksichtigung finden, damit der Erfolg einer Intervention bei ihrer breiten Anwendung optimiert wird: a) Es müssen Trainingskurse zum Erlernen der reliablen Durchführung der Intervention angeboten werden. b) Es ist eine klare Darstellung des Wirkungsmodells der Maßnahme notwendig. c) Es muss deutlich formuliert werden, unter welchen Voraussetzungen, bei welchem Klienten und bei welchen Herangehensweisen positive Effekte beziehungsweise auch Widerstände zu erwarten sind. d) Es müssen Informationen über die zu erwartenden Kosten pro Kind beziehungsweise Familie vorliegen. e) Sowohl das Programm als auch mögliche Veränderungen müssen genau dokumentiert werden.

Fazit

Die vorangehenden Ausführungen verdeutlichen, dass es sehr aufwendig ist, die „Goldstandards“ für einen wissenschaftlichen Nachweis der Wirksamkeit und Effektivität von Präventions- und Interventionsmaßnahmen im Bereich Früher Hilfen zu erfüllen. Selbst wenn diese nachgewiesen sind, müssen weitere Aspekte berücksichtigt werden, um den Erfolg der Intervention auf breiter Ebene zu optimieren. Zu berücksichtigen ist jedoch, dass diese Standards häufig nur unter experimentellen Bedingungen einzuhalten sind. Weiterhin ist kritisch anzumerken, dass in der Frage, ob RCTs wirklich „Goldstandards“ sind und wenn ja, in welchen Fällen in der Fachwelt unterschiedliche Positionen vertreten werden. Kritiker der RCTs betonen dabei, dass diese zwar theoretisch bedeutsam, aber in der Praxis nicht umsetzbar sind, da sie unter anderem unethisch und zu kostspielig seien. Weiterhin wird kritisiert, dass auch in RCTs nicht alle Einflussvariablen zu kontrollieren sind und dass sie lediglich Kausalbeschreibungen und keine Kausalerklärungen liefern et cetera. Für eine ausführliche Pro-Contra-Diskussion sei auf den Beitrag von Cooh et al. im *American Journal of Evaluation* verwiesen [23]. Auch in der Praxis – besonders im Bereich der Frühen Hilfen – können zahlreiche Argumente – zum Beispiel der Datenschutz oder ethische

Probleme – gegen die Einhaltung dieser Standards sprechen (vergleiche hierzu auch den Beitrag von H. Ziegler im vorliegenden Heft). Dennoch dient das Wissen um diese Standards einer Optimierung der Planung von Evaluationsdesigns. Die Durchsicht internationaler Forschungsliteratur zeigt, dass Frühe Hilfen tatsächlich wirksam sind, wobei die ermittelten Effektstärken im niedrigen bis mittleren Bereich liegen. Außerdem wird deutlich, dass auch in anderen Ländern wissenschaftliche Evaluationen den Standard bilden und von Experten explizit gefordert und erläutert werden. In Deutschland fehlen wissenschaftliche vergleichende Analysen weitestgehend.

Im Beitrag von H. Ziegler werden die hier dargestellten experimentellen Standards des Wirksamkeits- und Effektivitätsnachweises kritisch beleuchtet. Es wird deutlich, dass neben dem Nachweis der Wirksamkeit und Effektivität weiter gehende Kausalerklärungen gerade im Bereich der Frühen Hilfen hohe praktische Bedeutung haben. Dennoch sollte beachtet werden, dass wissenschaftliche Standards bei der Evaluation von Präventions- oder Interventionsprogrammen unabdingbar sind, da es weder ethisch vertretbar wäre, die Gemeinschaft ineffektive Programme finanzieren zu lassen noch dass Kinder und Eltern dadurch altersgemäße Entwicklungsmöglichkeiten verpassen.

Korrespondenzadresse

Prof. Dr. A. Lengning

Fakultät Rehabilitationswissenschaften,
Rehabilitationspsychologie,
Technische Universität Dortmund
Emil-Figge-Str. 50, 44221 Dortmund
anke.lengning@tu-dortmund.de

Interessenkonflikt. Der korrespondierende Autor gibt an, dass kein Interessenkonflikt besteht.

Literatur

1. NICHD Early Child Care Research Network (2005) Child care and child development: results from the NICHD Study of Early Child Care and Youth Development. Guilford, New York

2. Kindler H (2006) Frühe Prävention von Kindesmisshandlung und -vernachlässigung: Ein internationaler Forschungsüberblick. *Kindesmisshandlung Vernachlässigung* 9:23–47

3. Carpenter B (2005) Early childhood intervention: possibilities and prospects for professionals, families and children. *Br J Spec Educ* 32:176–183

4. Goodson BD, Layzer JL, St Pierre RG, Bernstein LS (2000) Good intentions are not enough: a response to Gilliam, Ripple, Zigler, and Leiter. *Early Child Res Q* 15:61–66

5. Rutter M (2002) Family influences on behaviour and development: challenges for the future. In: McHale J, Grolnick W (Hrsg) *Retrospect and prospect in the psychological study of families*. Erlbaum, London, S 321–351

6. Rutter M (2006) Is sure start an effective intervention? *Child Adolescent Mental Health* 11:135–141

7. Brezing H (2000) Welche Bedürfnisse haben Anwender(innen), und wie werden sie in der Forschung abgedeckt? Die Bedeutung von Evaluationsstandards und von Effektivitätskriterien für die Praxis. In: Hager W, Patry JL, Brezing H (Hrsg) *Evaluation psychologischer Interventionsmaßnahmen*. Huber, Bern, S 8–18

8. Hager W, Hasselhorn M (2000) Psychologische Interventionsmaßnahmen: Was sollen sie bewirken können? In: Hager W, Patry JL, Brezing H (Hrsg) *Evaluation psychologischer Interventionsmaßnahmen*. Huber, Bern, S 41–85

9. Lengning A, Zimmermann P (2009) *Evaluation von Interventions- und Präventionsmaßnahmen im Bereich Früher Hilfen: Internationaler Forschungsstand, Evaluationsstandards und Empfehlungen für die Umsetzung in Deutschland. Expertise im Auftrag des Deutschen Jugendinstituts München. NZFH, Köln*

10. White C, Agnew J, Verduyn C (2002) The Little Hulton Project: a pilot clinical psychology service for pre-school children and their families. *Child Adolesc Ment Health* 7:10–15

11. Tuijl C van, Leseman P, Rispens J (2001) Efficacy of an intensive home-based educational intervention programme for 4- to 6-year-old ethnic minority children in the Netherlands. *Int J Behav Dev* 25:148–159

12. Davis H, Dusoit T, Papadopoulou K et al (2005) Child and family outcomes of the European Early Promotion Project. *Int J Ment Health Promot* 7:63–81

13. Beelmann A, Brambring M (1998) Implementation and effectiveness of a home-based early intervention program for blind infants and preschoolers. *Res Dev Disabil* 19:225–244

14. Blunt Bugental D, Crane Ellerson P, Lin EK et al (2002) A cognitive approach to child abuse prevention. *J Fam Psychol* 16:243–258

15. AERA, American Educational Research Association, National Council on Measurement in Education (1999) *Standards for educational and psychological testing*. AERA, Washington DC

16. Sanders JR, JCSEE, Joint Committee on Standards for Educational Evaluation (2006) *Handbuch der Evaluationsstandards. Die Standards des Joint Committee on Standards for Educational Evaluation, 3. erweiterte und aktualisierte Aufl.* VS, Wiesbaden

17. Stufflebeam DL, Shinkfield AJ (2007) *Evaluation theory, models, and applications*. Jossey-Bass, San Francisco

18. Flay BR, Biglan A, Boruch RF (2005) Standards of evidence: criteria for efficacy, effectiveness and dissemination. *Prev Sci* 6:151–175

19. Hager W (2002) *Wirksamkeits- und Wirksamkeitsunterschiedshypothesen, Evaluationsparadigmen, Vergleichsgruppen und Kontrolle*. In: Hager W, Patry JL, Brezing H (Hrsg) *Handbuch Evaluation psychologischer Interventionsmaßnahmen*. Huber, Bern

20. Patry JL, Perrez M (2000) *Theorie-Praxis-Probleme und die Evaluation von Interventionsprogrammen*. In: Hager W, Patry JL, Brezing H (Hrsg) *Evaluation psychologischer Interventionsmaßnahmen*. Huber, Bern, S 19–40

21. LeCroy & Milligan Associates Inc (2006a) *Healthy Families Arizona Longitudinal Evaluation – 1st annual report*. Project Report, Tucson

22. LeCroy & Milligan Associates Inc (2006b) *Healthy Families Arizona Longitudinal Evaluation – 2nd annual report*. Project Report, Tucson

23. Cook TD, Scriven M, Coryn CLS, Evergreen SDH (2010) *Contemporary thinking about causation in evaluation: a dialogue with Tom Cook and Michael Scriven*. *Am J Eval* 31:105–117