

Anke Lengning

Wünschenswerte Standards für die Evaluation Früher Hilfen

Der Beitrag „Wünschenswerte Standards für die Evaluation Früher Hilfen“ bezieht sich inhaltlich nahezu ausschließlich zum einen auf den Artikel „Goldstandards für einen wissenschaftlichen Nachweis der Wirksamkeit und Effektivität als Voraussetzung einer Dissemination einer Intervention im Bereich Früher Hilfen“ (Lengning, 2010) und zum anderen auf die Expertise „Evaluation von Interventions- und Präventionsmaßnahmen im Bereich Früher Hilfen: Internationaler Forschungsstand, Evaluationsstandards und Empfehlungen für die Umsetzung in Deutschland“ (Lengning & Zimmermann, 2009), so dass weitere Informationen, Literaturangaben etc. diesen beiden Publikationen zu entnehmen sind.

Es ist international und national anerkannt, dass Frühe Hilfen notwendig sind (Kindler, 2006; Carpenter, 2005), wobei das Ziel derartiger Präventions- und Interventionsmaßnahmen die Verbesserung der Entwicklungsbedingungen von Familien und Kindern sein sollte. Aus verschiedenen Gründen ist es wichtig, die Effektivität der Frühen Hilfen nachzuweisen. Hier sollen kurz nur drei Gründe genannt werden:

1. Wenn ineffektive Interventionen durchgeführt werden, kann es durchaus passieren, dass wichtige Zeitfenster in der kindlichen Entwicklung nicht genutzt werden (Rutter, 2002; Rutter, 2006).
2. Wenn bei Kindern Kompetenzzuwächse zu verzeichnen sind, muss dies nicht unbedingt für die Wirksamkeit einer Intervention sprechen, da Kinder auch in ihrer natürlichen Entwicklung Kompetenzzuwächse erfahren (Hager & Hasselhorn, 2000).
3. Präventions- und Interventionsmaßnahmen hängen mit einem hohen finanziellen Aufwand zusammen, so dass eine gesellschaftliche Rechtfertigung notwendig erscheint.

Einen möglichen, wünschenswerten Standard bei der Evaluation Früher Hilfen bildet das randomisierte Kontrollgruppendesign, das häufig als „Goldstandard“ zur Evaluation diskutiert wird (vgl. Lengning, 2010). Empfehlungen lassen sich hier aussprechen für:

1. das Studiendesign,
2. die Überprüfung des Gültigkeitsbereichs der Effekte sowie
3. die Überprüfung der Wirksamkeit und Effektivität.

Auf diese Punkte wird im Folgenden genauer Bezug genommen, wobei im Abschluss auf mögliche Empfehlungen zur Dissemination eingegangen wird. An dieser Stelle sei schon einmal darauf hingewiesen, dass diese Standards häufig kritisiert werden, da sie in der Praxis aus verschiedenen Gründen nicht immer umzusetzen sind (vgl. Ziegler, 2010). Dennoch sollten diese

Brücken bauen zwischen Wissen und Handeln – Strategien der Gesundheitsförderung

Mittwoch und Donnerstag, 6. und 7. März 2013 in der Technischen Universität Berlin

Standards m.E. bedacht werden, wenn eine Evaluation geplant wird, da das gewählte Design hierdurch optimiert werden kann (z.B. Lengning, 2010).

Zunächst zum Studiendesign: Ein Studiendesign sollte nach den angestrebten Standards eine Interventions- und eine Vergleichsgruppe in einem Pretest, in einem Posttest und in einem Follow-Up-Test berücksichtigen. Der Pretest sollte zu Beginn der Maßnahme, der Posttest nach Beendigung der Maßnahme und der Follow-Up-Test nach mindestens einem halben Jahr nach Beendigung der Maßnahme durchgeführt werden. Hierbei wird zwischen isolierten, vergleichenden und kombinierten Evaluationen unterschieden (vgl. Hager, 2002).

Zur Überprüfung des Gültigkeitsbereichs der Effekte: Hier sollten folgende Fragen beantwortet werden: Haben sich die gewünschten Eigenschaften wie beabsichtigt verändert? Zeigen sich Generalisierungen bzw. zeigt sich ein Transfer (Hager & Hasselhorn, 2000)? Zeigen sich die Effekte auch in der eigenständigen Lebensführung (Patry & Perrez, 2000)? Ein Beispiel hierzu: Wenn die Mutter-Kind-Interaktion durch eine Intervention verbessert werden soll, sollte sich diese Verbesserung nicht nur in der strukturierten Situation, sondern ebenfalls im Alltag zeigen. Bei der Überprüfung der Wirksamkeit und Effektivität sollte daher der Frage nachgegangen werden, ob sich die gewünschten Effekte auch unter Alltagsbedingungen zeigen (vgl. Flay, Biglan & Boruch, 2005; Hager, 2002). In diesem Zusammenhang sollten folgende Fragen beantwortet 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 weniger geschult waren, durchgeführt wurde und die Ausstattung nicht optimal war?“ (Lengning, 2010, S. 1058).

Im Folgenden wird auf die Standards zum Wirksamkeitsnachweis eingegangen. Es sollen hier sechs Punkte unterschieden werden:

1. „die Spezifizierung der Programmwirksamkeit,
2. die Programmdokumentation,
3. die Überprüfung der Durchführungsqualität der Maßnahme,
4. die Erfassung und Bewertung der erwarteten Veränderung,
5. die Schlussfolgerungen über die Kausalität der Ergebnisse und
6. die Generalisierbarkeit der Ergebnisse.“ (Lengning, 2010, S. 1058)

Diese Punkte werden im Folgenden ausgeführt.

Zunächst soll auf die Spezifizierung der Programmwirksamkeit eingegangen werden. Hier ist es wünschenswert, dass Aussagen darüber getroffen werden können, dass ein bestimmtes Programm (X) eine Effektvariable (Y) bei einer Population (Z) verändert.

Brücken bauen zwischen Wissen und Handeln – Strategien der Gesundheitsförderung

Mittwoch und Donnerstag, 6. und 7. März 2013 in der Technischen Universität Berlin

Die Programmdokumentation zielt darauf ab, dass Dritte das Programm replizieren können. Hierfür müssen die Durchführung und das Wirkungsmodell genau beschrieben werden.

Bei der Überprüfung der Durchführungsqualität muss genau untersucht werden, welche Maßnahme bei der Kontroll- und welche bei der Interventionsgruppe durchgeführt wird. Hierbei sollte es so sein, dass die Kontrollgruppe den Wirkfaktoren nicht ausgesetzt wird, damit eine Wirksamkeitsbeurteilung ermöglicht wird.

Der nächste Punkt betrifft die Erfassung und Bewertung der erwarteten Veränderung. Es muss darauf geachtet werden, dass die tatsächlichen Veränderungen der erwarteten Wirkung erfasst werden. Die eingesetzten Verfahren müssen hierfür den geforderten Gütekriterien (Reliabilität und Validität) entsprechen. Bei manchen Variablen sollten Proxy-Maße erfasst werden. Dies sind Maße, die mit bestimmtem Fehlverhalten in Zusammenhang stehen. Als Beispiel können hier Misshandlungen genannt werden, wo es fraglich sein kann, ob die gewonnenen Angaben immer glaubwürdig sind. Darüber hinaus kann es sehr wichtig sein, dass die Variablen auf verschiedene Weisen erhoben werden. Eingesetzt werden können Fragebögen, Interviews und Beobachtungen, die jeweils verschiedene Vor- und Nachteile mit sich bringen (vgl. hierzu Lengning & Zimmermann, 2009). Um die erwarteten Veränderungen weiter erfassen und bewerten zu können, ist es zusätzlich wichtig, dass die Stichproben der Interventions- und der Kontrollgruppe groß genug sind. Bei dem statistischen Vergleich soll in diesem Zusammenhang nicht nur auf die Signifikanz geachtet werden, sondern besonders auch auf die Effektstärke. Hiermit können Aussagen darüber getroffen werden, wie stark der Effekt einer Intervention und wie bedeutsam er hiermit für die Praxis ist. Zeigen sich trotz einer Randomisierung verschiedene Mittelwerte in den erhobenen Variablen zwischen der Kontroll- und der Interventionsgruppe zu Beginn der Intervention, müssen die Mittelwerte angepasst werden. Bei dem Evaluationsdesign muss - wie bereits erwähnt - eine Follow-Up-Erhebung aufgenommen werden. Diese ermöglicht es zu erfassen, wie langanhaltend die erzielten Effekte sind und ob sich Auswirkungen zeigen, die möglicherweise erst in späteren kindlichen Entwicklungsstadien deutlich werden. Darüber hinaus ist es dadurch möglich, sogenannte Schläfereffekte aufzudecken. Dies sind Effekte, die sich erst nach einem längeren Zeitraum entwickeln. Die erzielten Ergebnisse müssen veröffentlicht werden, wobei auch auf einer praktisch verständlichen Ebene publiziert werden sollte. Ein letzter Punkt bei der Erfassung und Bewertung der erwarteten Veränderung betrifft die Objektivität. Aus diesem Grund sollten von der Intervention unabhängige Personen die Evaluation durchführen. Insbesondere Auswertungen sollten von unwissenden Personen durchgeführt werden. D.h. die Auswerter sollten nicht wissen, ob die Probanden zur Kontroll- oder zur Interventionsstichprobe gehören.

Bei der Schlussfolgerung über die Kausalität der Ergebnisse ist die Randomisierung von besonderer Bedeutung, um letztlich nachzuweisen, dass die Intervention wirklich zu den ermittelten Effekten geführt hat.

Brücken bauen zwischen Wissen und Handeln – Strategien der Gesundheitsförderung

Mittwoch und Donnerstag, 6. und 7. März 2013 in der Technischen Universität Berlin

Um Angaben zur Generalisierbarkeit treffen zu können, müssen das Alter, die ethnische Zusammensetzung, das Geschlecht und der sozioökonomische Status der Stichprobe erfasst und dokumentiert werden.

Der nächste große Punkt bei Evaluationen betrifft die Effektivität. So ist es nicht nur wichtig, dass Standards zum Wirksamkeitsnachweis berücksichtigt werden, sondern es sollte ebenfalls gezeigt werden, dass eine erfolgreiche Durchführung des Interventionsprogramms auch unter Alltagsbedingungen möglich ist. Folgendes muss beachtet werden:

1. Die Intervention muss detailliert in einem Manual beschrieben werden, damit die Vorgehensweise vergleichbar überprüft werden kann. Informationen über verschiedene Punkte sind hier wichtig. Sie betreffen z.B. folgende Fragen: Wie ist das Netzwerk aufgebaut? Wie wird bei der Intervention vorgegangen?
2. Die Durchführung muss unter Alltagsbedingungen geprüft werden. Folgende Fragen sollten in diesen Zusammenhang beantwortet werden: Zeigen sich Effekte der Qualifizierung? D.h. zeigen sich Unterschiede, wenn z.B. Laienhelfer anstatt Fachpersonal das Programm durchführen? Welche Vorgaben des Programms können nicht eingehalten werden? Hier sollte z.B. festgehalten werden, ob die Dauer der Intervention nicht eingehalten werden konnte. Dies könnte dazu beitragen, zu klären, warum bestimmte Effekte nicht aufgetreten sind.
3. Es muss überprüft werden, wie hoch die Pro-Kopf-Kosten einer Intervention im Vergleich zu einer anderen Intervention sind, die eine vergleichbare Wirkung hat, um die Effektivität der Maßnahme zu beurteilen.

Abschließend sollen Empfehlungen zur Dissemination ausgesprochen werden. Die Voraussetzungen sind, dass zum einen die Wirksamkeit unter optimalen Bedingungen und zum anderen die Effektivität unter Alltagsbedingungen nachgewiesen ist. Aber dies alleine reicht nicht aus. Es muss z.B. zusätzlich gesichert werden, dass es Angebote gibt, um die Durchführung der Intervention zu erlernen, es bedarf deutlicher Beschreibungen, wann das Programm die gewünschten Effekte zeigt und die erwarteten Kosten müssen erkennbar sein.

Zum Fazit: Es ist ersichtlich, dass der Aufwand sehr groß ist, wenn diese Standards eingehalten werden sollen und dass diese Standards in der Praxis nicht immer zu erfüllen sind. Weiterhin wird häufig Kritik an diesen Standards geübt. So werden hier z.B. ethische Gründe und Kosten aufgeführt und es wird kritisiert, dass nicht alle Einflussvariablen kontrolliert werden können (vgl. z.B. Ziegler, 2010). Das Wissen über diese Standards ist dennoch sehr wichtig, damit die Planung von Evaluationsdesigns optimiert werden kann. So sollte bedacht werden, dass es nicht ethisch wäre, zum einen Programme zu finanzieren, die nicht effektiv sind und zum anderen, dass Entwicklungsmöglichkeiten von Kindern und Eltern verpasst werden könnten.

Für weitere Informationen sei nochmals auf den Artikel „Goldstandards für einen wissenschaftlichen Nachweis der Wirksamkeit und Effektivität als Voraussetzung einer Dissemination einer Intervention im Bereich Früher Hilfen“ (Lengning, 2010) und auf die Expertise „Evaluation von Interventions- und Präventionsmaßnahmen im Bereich Früher Hilfen: Internationaler Forschungs-

*Brücken bauen zwischen Wissen und Handeln –
Strategien der Gesundheitsförderung*

Mittwoch und Donnerstag, 6. und 7. März 2013 in der Technischen Universität Berlin

tand, Evaluationsstandrad und Empfehlung für die Umsetzung in Deutschland“ (Lengning & Zimmermann, 2009) hingewiesen. Diese beiden Publikationen bildeten die inhaltliche Grundlage für diesen Beitrag.

Anke Lengning

anke@lengning.de

www.lengning.de

Literatur / Quellen:

- Carpenter, B. (2005). Early Childhood Intervention: Possibilities and Prospects for Professionals, Families and Children' *British Journal of Special Education*, 32, 176-183.
- Flay, B. R., Biglan, A. & Boruch, R. F. (2005). Standards of Evidence: Criteria for Efficacy, Effectiveness and Dissemination. *Prevention Science*, 6, 151-175.
- Hager, W. & Hasselhorn, M. (2000). Psychologische Interventionsmaßnahmen: Was sollen sie bewirken können? In W. Hager, J. L. Patry & H. Brezing (Hrsg.), *Evaluation psychologischer Interventionsmaßnahmen* (S. 41-85). Bern: Huber.
- Hager, W. (2002). Wirksamkeits- und Wirksamkeitsunterschiedshypothesen, Evaluationsparadigmen, Vergleichsgruppen und Kontrolle. In W. Hager, J.L. Patry & H. Brezing (Hrsg.) *Evaluation psychologischer Interventionsmaßnahmen* (S. 180-201). Bern: Huber.
- Kindler, H. (2006). Frühe Prävention von Kindesmisshandlung und -vernachlässigung: Ein internationaler Forschungsüberblick. *Kindesmisshandlung und -vernachlässigung*, 9, 23-47.
- 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. Köln: NZFH.
- Lengning, A. (2010). Goldstandards für einen wissenschaftlichen Nachweis der Wirksamkeit und Effektivität als Voraussetzung einer Dissemination einer Intervention im Bereich Früher Hilfen: Randomisierte kontrollierte Studien. *Bundesgesundheitsblatt, Frühe Hilfen zum gesunden Aufwachsen von Kindern – Interdisziplinäre und intersektorale Zusammenarbeit Teil 1*, 1056-1060.
- Patry, J.-L. & Perrez, M. (2000). Theorie-Praxis-Probleme und die Evaluation von Interventionsprogrammen. In W. Hager, J.L. Patry & H. Brezing (Hrsg.), *Evaluation psychologischer Interventionsmaßnahmen* (S. 19-40). Bern: Huber.
- Rutter, M. (2002). Family influences on behaviour and development: Challenges for the future. In J. McHale & W. Grolnick, W. (Eds.), *Retrospect and prospect in the psychological study of families* (pp. 321-351). London: Erlbaum.
- Rutter, M. (2006). Is Sure Start an effective intervention? *Child and Adolescent Mental Health*, 11, 135-141.
- Ziegler, H. (2010). Ist der experimentelle Goldstandard wirklich Gold wert für eine Evidenzbasierung der Praxis Früher Hilfen? *Bundesgesundheitsblatt, Frühe Hilfen zum gesunden Aufwachsen von Kindern – Interdisziplinäre und intersektorale Zusammenarbeit Teil 1*, 1061-1066.