

Feststellung der Wirksamkeit von Präventionsmaßnahmen am Beispiel des Mentorenprogramms „Balu und Du“:

Effektstärken, Netto-Effektstärken und die Funktion von Kontrollgruppen

Maren Schlüter und Hildegard Müller-Kohlenberg¹



Warum Kontrollgruppen? Welche Kontrollgruppen?

Um die Wirkung einer präventiven oder interventiven Maßnahme festzustellen, werden üblicherweise Kontrollgruppen herangezogen. Ein Kontrollgruppendesign kann Aufschluss darüber geben, ob festgestellte Unterschiede zwischen dem Beginn der Maßnahme und einem späteren Zeitpunkt tatsächlich auf die Behandlung (Intervention, *treatment*) und nicht auf andere Einflüsse zurückzuführen sind. Denn auch z. B. Reifung, Lernen oder anderweitige Erfahrungen während der entsprechenden Zeitspanne können zu Verhaltensänderungen führen. So wichtig Kontrollgruppen in der Evaluationskultur sind, so schwierig kann es in der Praxis werden, die mit dem Verfahren verbundenen strengen Ansprüche zu erfüllen: Die Kontrollgruppe muss bekanntlich (I) entweder als randomisierte Zufallsstichprobe gebildet werden oder so zusammengestellt werden, dass die

relevanten Kriterien für das *matching* – also die Parallelisierung der beiden Gruppen – berücksichtigt werden. Es muss (II) sichergestellt sein, dass die Beurteilung sowohl der Interventions- wie auch der Kontrollgruppe unvoreingenommen erfolgt; beispielsweise sollte keine Verzerrung durch Erwartungshaltungen in die Bewertung der Ergebnisse eingehen. Ein weiteres Problem ist die mangelnde Bereitschaft zur Teilnahme an einer Kontrollgruppe (III), wenn die Aufnahme in die Interventionsgruppe, also die Behandlung, verweigert wurde oder auf sich warten lässt.

Die genannten Schwierigkeiten führen bisweilen dazu, dass statt eines Kontrollgruppendesigns eine einfache Voher-Nachher-Messung der Interventionsgruppe durchgeführt wird. Dabei wird ein Mittelwertvergleich der Ergebnisse vorgenommen („Naturalistisches Eingruppen-Prä-Post-Design“) und die so ermittelten Unterschiede als Wirkung der Maßnahme interpretiert, ohne eventuelle weitere Einflussgrößen berücksichtigen zu können.

Effektstärken – Netto-Effektstärken

Aussagekräftiger als Mittelwertunterschiede sind Effektstärken, da diese auch die Standardabweichung der Differenzwerte berücksichtigen. Solange allerdings keine Kontrollgruppe existiert, geben auch die Effektstärken keine Auskunft darüber, ob und ggf. wie stark Wirkungen erzielt werden, die *nicht* auf die Maßnahme zurückzuführen sind. Etwaige positive Veränderungen der Kontrollgruppe sind demnach von den ermittelten Ef-

fektstärken abzuziehen oder – sofern diese sich in negativer Richtung entwickelt hat – den Effektstärken hinzuzurechnen. Damit erhält man Netto-Effektstärken, die sicherere Hinweise auf die Treatment-Wirkung geben.

Effektstärken werden häufig in graphischer Form als Balken oder Säulen ausgedrückt, da diese meist eine intuitiv verständliche Antwort auf die Frage nach der Wirkung anbieten.

Die Angemessenheit eines Kontrollgruppendesigns

Maier-Riehle und Zwingmann (2000)² gehen davon aus, dass Effektstärken in einem „Eingruppen-Prä-Post-Design“ im Allgemeinen größer ausfallen als in einem Kontrollgruppendesign. Diese Erwartung ist darin begründet, dass auch Teilnehmer(innen) der Kontrollgruppe positive Entwicklungen durchlaufen, die der Intervention realistischweise nicht zugerechnet werden dürfen. Entsprechendes gilt, wenn das gemessene Kriterium (Verhalten, Einstellungen, Kompetenzen ...) in der Zeitspanne keinen negativen Verlauf nimmt bzw. eher stabil ist. Andernfalls könnte das Kontrollgruppendesign dazu führen, dass die Netto-Effektstärken größer ausfallen als die Ergebnisse bei einem „Eingruppen-Prä-Post-Design“.

Da sich die Erwartung einer Verminderung der Netto-Effektstärken in der im Folgenden dargestellten Untersuchung nicht bestätigte, könnten Deskription und Analyse des Evaluations-

¹ Prof. em. Dr. H. Müller-Kohlenberg Universität Osnabrück, Fb 3; und M. Schlüter wissenschaftliche Mitarbeiterin

² Maier-Riehle, B., & Zwingmann, C. (2000). Effektstärkevarianten beim Eingruppen-Prä-Post-Design: Eine kritische Betrachtung. *Rehabilitation*, 39, 189–199.

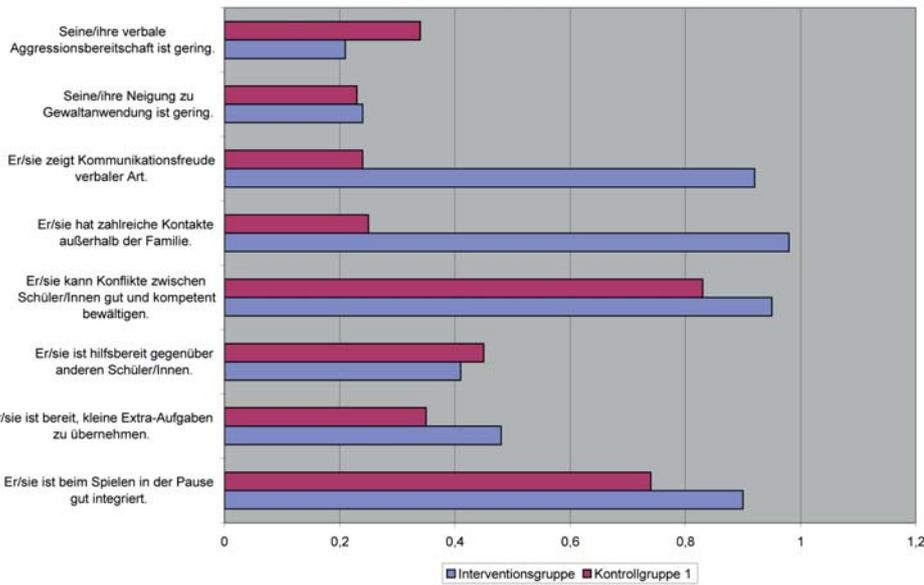


Abbildung 1: Vergleich der Effektstärken von Interventionsgruppe und Kontrollgruppe 1

verfahrens Einblick in Kontextvariablen geben, die Effektstärken zu beeinflussen in der Lage sind.

Ein Vergleich zweier Kontrollgruppen: Auswirkungen auf die Netto-Effektstärken

Bezogen auf die Stärkung des Sozialverhaltens von Grundschulkindern wurde das Mentorenprogramm „Balu und Du“ evaluiert. Lehrer(innen) beurteilten Kinder aus der Interventionsgruppe mittels der OSKAR-Skala³ in einer retrospektiven Vorher-Nachher-Evaluationsanordnung. Die Ergebnisse wurden mit denen einer parallelisierten Kontrollgruppe verglichen.

Mit den aufgezeigten Effektstärken der Interventionsgruppe (N = 51), die teilweise über den Durchschnittswerten einer klinischen Psychotherapie liegen, hätte man zufrieden sein können. Sie wurden bereits 2008 veröffentlicht⁴ und lieferten Hinweise auf die Wirksamkeit des Mentorenprogramms „Balu und Du“.

Allerdings fiel auf, dass Werte der Kontrollgruppe (N = 51) teilweise den erwartbaren Ergebnissen zuwider liefen. Aggressives und sozial unerwünschtes Verhalten ist im Kindes- und Jugendalter tendenziell stabil – und es nimmt mit dem Älterwerden eher zu als ab, sofern keine pädagogische Intervention erfolgt.⁵

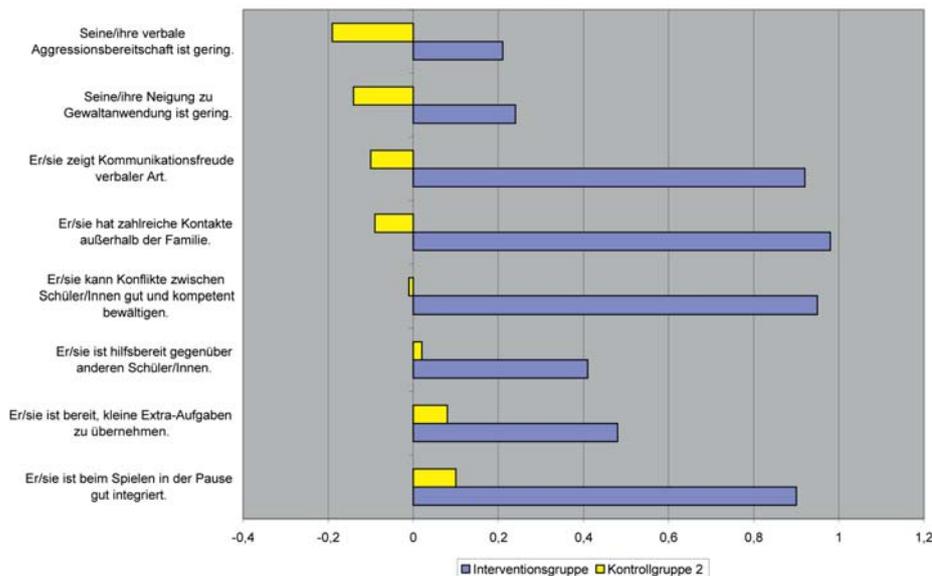


Abbildung 2: Vergleich der Effektstärken von Interventionsgruppe und Kontrollgruppe 2

Die teilweise überraschend hohen und erwartungswidrigen Effektstärken in der Kontrollgruppe gaben daher Anlass zur Überprüfung des Zustandekommens dieser Effekte. Als Hypothese wurde formuliert, dass die Beurteiler(innen), d. h. die Lehrer(innen) der Kontrollgruppenkinder, sich in einem subjektiv kompetitiven Verhältnis zu den durchweg jüngeren Mentor(innen) des Programms „Balu und Du“ befanden. Es war offensichtlich, dass die Wirkung des Mentorenprogramms mit den pädagogischen Einflussmöglichkeiten eines „normalen Schulalltags“ verglichen werden sollte. Entsprechende Bemerkungen („Sie glauben wohl, Sie können das besser als wir!“) veranlassten dazu, eine zweite Kontrollgruppenuntersuchung (N = 36) durchzuführen. Die Replikation erfolgte diesmal in Regionen, in denen das Projekt „Balu und Du“ nicht implementiert ist. Es wurde eine weitgehend neutrale Instruktion verwendet, mit der – wie bei der ersten Kontrollgruppe – um eine retrospektive Einschätzung der sog. Sorgenkinder gebeten wurde.

Ein Hinweis auf den Vergleich mit „Balu und Du“ wurde nicht gegeben. Es war also davon auszugehen, dass die beurteilenden Lehrer(innen) diesmal keine Konkurrenz zu ihrer eigenen pädagogischen Tätigkeit sahen, die die Bewertung hätte beeinflussen können.

Die Ergebnisse unterscheiden sich gravierend von denen der ersten Untersuchung, in der eine mentale „Wettbewerbssituation“ vermutet wurde.

³ Müller-Kohlenberg, H. (2008). Entwicklungsorientierte Prävention von Devianz im Jugendalter: Das Mentorenprojekt „Balu und Du“. In: Bundesministerium des Innern (Hrsg.), Theorie und Praxis gesellschaftlichen Zusammenhalts – Aktuelle Aspekte der Präventionsdiskussion um Gewalt und Extremismus. Berlin: Referat G 15.

⁴ ebenda

⁵ Caspi, A., Elder, G.H., & Bem, D. J. (1987). Moving against the world: life-course patterns of explosive children. *Developmental Psychology*, 23, 308–313. Farrington, D. P. (1991). Childhood aggression and adult violence: early precursors and later life outcomes. In: D. J. Pepler & K. H. Rubin (Eds.), *The development and treatment of childhood aggression*. Hillsdale, NJ: Lawrence Erlbaum Associates. Haapasalo, J. & Tremblay, R. E. (1994). Physically aggressive boys from ages 6 to 12: family background, parenting behavior, and prediction of delinquency. *Journal of Consulting and Clinical Psychology*, 62, 1044–1052. Loeber, R., & Hay, D. (1997). Key issues in the development of aggression and violence from childhood to early adulthood. *Annual Review of Psychology*, 48, 371–410. Loeber, R. & Stouthamer-Loeber, M. (1998). Development of Juvenile Aggression and Violence: Some Common Misconceptions and Controversies. *American Psychologist*, 53, 242–259. Olweus, D. (1979). Stability of aggressive reaction patterns in males: a review. *Psychological Bulletin*, 86, 852–857. Pulkkinen, L. (1992). The path to adulthood for aggressively inclined girls. In: K. Bjorkqvist & P. Niemela (Eds.), *Of Mice and Women: Aspects of Female Aggression*. San Diego, CA: Academic Press. Sherry, J. L. (2001). The effect of violent video games on aggression. A meta-analysis. *Human Communication Research*, 27, 409–431.

Basierend auf der zweiten Befragung von nicht mental involvierten Beurteiler(inne)n fallen die Netto-Effektstärken größer aus. Bedingt durch die teilweise negative Entwicklung der Kontrollgruppenkinder sind sie sogar noch stärker ausgeprägt als im „Eingruppen-Prä-Post-Design“.

Zur Frage der Generalisierbarkeit des Befunds

Die von Maier-Riehle und Zwingmann aufgestellte These, dass Effektstärken, die auf Grundlage eines Kontrollgruppendesigns ermittelt wurden, eher kleinere Werte aufweisen als solche aus „Eingruppen-Prä-Post-Designs“ kann mit diesem Befund si-

cherlich nicht allgemein widerlegt werden. Eine Konsequenz daraus wäre allerdings, dass – wo immer es möglich ist – mit dem Design des Doppelblindversuchs gearbeitet werden sollte. Die dargestellten Ergebnisse unterstreichen die Bedeutung des eingangs unter (II) genannten Anspruchs. Bei Berücksichtigung dieser Forderung wären Erwartungshaltungen, die die Einschätzung im negativen oder positiven Sinne verzerren könnten, weitgehend ausgeschlossen. In (sozial-)pädagogischen Settings ist diese Voraussetzung jedoch nicht immer gegeben, weil eine längerfristige genaue Kenntnis der Kinder meist auch mit dem Wissen um Interventionsprogramme einhergeht. Allerdings wäre darüber

nachzudenken, ob die teilweise enttäuschenden Ergebnisse der Sherman-Studie⁶, die seit 1997 eine fast epochal negative Auswirkung auf das Vertrauen in Präventionsprogramme hatten, zumindest teilweise durch Einflüsse der beschriebenen Art erklärt werden könnten.

Wir danken den Studentinnen Nicolé Adämmer, Katrin Eilers, Agnieszka Olow und Tina Vogel für ihr Engagement bei der Durchführung der Untersuchung.

⁶ Sherman, L. W., Gottfredson, D., MacKenzie, D., Eck, J., Reuter, P. & Bushway, S. (1997). Preventing Crime: What Works, What Doesn't, What's Promising. A report to the United States Congress. Prepared for the National Institute of Justice.

BUCHTIPP

Kirsten Heisig:

Das Ende der Geduld

Konsequent gegen jugendliche Gewalttäter
Herder, Freiburg 2010, 205 Seiten, 14,95 EUR

Wenige Wochen nach ihrem Tod – Kirsten Heisig nahm sich das Leben – ist das Buch der beachtlich engagierten Berliner Jugendrichterin erschienen. Es handelt sich um eine schonungslose Analyse der Jugendkriminalitätslage im Berliner Bezirk Neukölln, aber – anders als von



und zügige Zusammenarbeit der beteiligten Institutionen“ sind ihr Credo, das

einigen erwartet, vielleicht sogar erhofft – nicht um ein Plädoyer für gnadenlose Vergeltung, höhere Strafrahmen und schärfere Gesetze. „Die konsequente Anwendung des geltenden Rechts“ und die „strikte

bereits als „Neuköllner Modell“ bekannt wurde. Frühe Hilfe und Intervention, auch mehr Personaleinsatz in der sozialen Arbeit und vor allem der Wille, Verhalten und Verhältnisse beeinflussen zu wollen, werden eingefordert, ohne selbst die Patentrezepte für alle lebenswirklichen Probleme zu kennen und aufzeigen zu können. Erfordernisse für mehr Kooperation sind bereits seit Jahren auf der Agenda des fachlichen Diskurses, Kirsten Heisig hält diesem Anspruch die ernüchternde Wirklichkeit ihres beruflichen und sozialen Umfeldes



Baustoffe und Containerdienst Erd-, Ausschachtung + Abrißarbeiten

Hapelrath 1 • 40764 Langenfeld

☎ 0 21 73/1 77 00 und 0 21 75/96 24 • Telefax 0 21 75/7 36 27