



Mit Ganzttag mehr Zukunft.  
Das neue Ganzttagsgymnasium NRW

# DAS KONTROLLGRUPPENDESIGN IN GANZ IN 2

Diese Dokumentation beschreibt die Implementierung des empirisch gestützten Kontrollgruppendesigns für *Ganz In 2*.

Für Wirkungsanalysen sind geeignete Kontrolleinheiten nötig. Um geeignete Kontrolleinheiten zu identifizieren, wurden jeder Fachdidaktik (FD) und jedem themenspezifischen Zusatzangebot (TVA) eine Auswahl an Schulen genannt, die als Kontrolleinheiten (oder Pool für Kontrolleinheiten) dienen können. Um eine Überbelastung der Schulen zu vermeiden, wurden keine anderen Schulen als Kontrolleinheiten angefragt. Die Zuteilung erfolgte über die Ähnlichkeit von Hintergrundvariablen der Kontrollschulen zu den Interventionsschulen pro Schule und im Mittel der jeweiligen Interventionsschulen.

Die Setzungen sind dabei wie folgt:

- Jede FD und jedes TVA (in der Folge Interventionen) erhielt mindestens so viele Schulen als Kontrollgruppe wie es Schulen in der Interventionsgruppe gibt, aber maximal sechs Schulen, da es ansonsten zu Überbelastungen der Schulen und Interventionen kommen kann
- Jede Schule sollte maximal zwei Mal Kontrollgruppe sein, so dass jede Schule mit maximal vier Studien belastet wird (Ausnahmen sind die Schulen, die freiwillig an drei Interventionen teilnehmen)
- Die Kontrollgruppen sollen auf der Basis von ausgewählten Schulmerkmalen und Ergebnissen aus der Projektphase 1 vergleichbar sein

Es wurde ein *propensity score matching* (PSM) angewendet (z.B. Gangl, 2010). Matchingverfahren umfassen eine Klasse nichtparametrischer Verfahren zur statistischen Abschätzung kausaler Effekte. Matching-Verfahren gruppieren Einheiten mit ähnlichen Drittvariablen, als wenn es sich um Vergleichsgruppen in einem randomisierten Experiment handeln würde. Ein idealer Match wäre ein Partner mit exakt gleichen Drittvariablenausprägungen zwischen Kontrollschule und Interventionsschule. Die Drittvariablenkontrolle erfolgt durch die Konstruktion statistischer Vergleichsgruppen. Die empirische Analyse erfolgt dabei in drei Schritten: (a) Zunächst ist ein Zuweisungsmodell zu entwickeln, das die Verteilung des Kausalfaktors (Zugehörigkeit zu der Interventionsgruppe) in Abhängigkeit von theoretisch relevanten Kovariaten (Drittvariablen-set) beschreibt. Über die aus dem Zuweisungsmodell vorhergesagten *propensity scores* wird anschließend (b) die Balancierung der Drittvariablen-set in den Vergleichsgruppen angestrebt. Ist diese erreicht, erfolgt (c) die eigentliche Schätzung der interessierenden kausalen Effekte durch den nichtparametrischen Vergleich der Ereignisverteilungen in den Experimental- und Kontrollstichproben. Schritt (c) entfällt hier. Die Validität der Effektschätzungen hängt entscheidend davon ab, ob der kausal wirksame Faktor und die Ergebnisgröße unter Kontrolle der verfügbaren Variablen aus theoretischer Sicht als unabhängig betrachtet werden können. Für die Interventionsgruppen werden für uns zeitlich vorausgehende Drittvariablen-set definiert, die mit dem Effekt konfundieren. Hier wurden die Schulgröße, die mittlere Leistungsfähigkeit der Schülerinnen und Schüler in den Fächern Deutsch und Mathematik in der Projektphase 1, sowie



die soziale Zusammensetzung der Schülerschaft verwendet.

Es wird geschätzt, inwieweit die Zugehörigkeit zur Kontrollgruppe durch die Drittvariablen sets vorhergesagt wird. Eine lineare Kombination der diversen Variablen ist dabei hinreichend (Rosenbaum & Rubin, 1983). Dazu wurde auf das R Paket *MatchIt* (Ho, Imai, King & Stuart, 2011) zurückgegriffen.

Es wurden je nach Grundvoraussetzung „optimal matching“ Ansätze verwendet, die im Hinblick auf den Kontrollgruppenmittelwert und den exakten Partner ausgleichen. Es werden immer exakte Partner erstellt und zusätzlich wird auf das Gesamtmittel hin optimiert. Überschüssige Partner wurden in Hinblick auf die Belastungsgerechtigkeit ausgeschlossen.

Die Mittelwertsabweichungen sind die wichtigste Kennstatistik der Passung der Matchingpartner. Tests ergaben, dass für alle Interventionen und für alle Drittvariablen keine signifikante Abweichung in den Gruppenverteilungen zwischen den Partnern vorliegt. Die Kontrollgruppen werden als angemessen betrachtet. Die unvorteilhafteste Zuteilung liegt im Bereich des Vertiefungsangebotes Sprachförderung vor, da vor allem Schulen aus dem Randbereich der Leistungsverteilung dieses wählten und somit nicht genug mögliche Partner verblieben.

## Grenzen und Probleme

Ein Problem ist die korrekte Spezifikation des Zuweisungsmodells, da dieses die Balance und die Tragkraft der Kontrollgruppen bestimmt. Im Kontext der Schulforschung muss klar sein, dass die Zahl der Kovariaten quasi unbegrenzt ist und hier nur wenige, verhältnismäßig starke Einflussgrößen verwendet wurden. Zudem musste vermutet werden, dass diese Einflussgrößen für alle Interventionen in *Ganz In* gleichermaßen gelten. Das basale Problem, dass kausale Schlüsse auf der Basis nicht-experimenteller Daten erfolgen sollen, lässt sich nur theoriegeleitet rechtfertigen. Es kann auf dieser Basis nur eine Annäherung erfolgen. Dieses Matching ist kein Verfahren, mit dem automatisch gültige Schätzungen kausaler Effekte erfolgen können. Für ein vollständiges Matching muss eine Zuordnung auf Individualebene unter Einbezug der abhängigen Variablen pro Intervention erfolgen. Das vorliegende Matching kann als a-priori Zuteilung verstanden werden, die eine geringere Belastung aller Beteiligten erlaubt.

## Literatur

Gangl, M. (2010). Nichtparametrische Schätzung kausaler Effekte mittels Matchingverfahren. In: C. Wolf & H. Best. *Handbuch der sozialwissenschaftlichen Datenanalyse*. Wiesbaden: VS Verlag.

Ho, D. H., Imai, K., King, G. & Stuart, E. A. (2011). MatchIt: Nonparametric Preprocessing for Parametric Causal Inference. [MatchIt](#)

Rosenbaum, P. R. & Rubin, D. B. (1983). The central role of propensity scores in observational studies for causal effects. *Biometrika*. 70, 41-55.

**Kontakt:** Michael Schurig; [michael.schurig@tu-dortmund.de](mailto:michael.schurig@tu-dortmund.de)