

Knut Schwippert, Thorsten Klinger, Ingrid Gogolin

Evaluation im Modellprogramm FÖRMIG: Alles unter Kontrolle?

Sprachliche Förderung von Kindern und Jugendlichen mit Migrationshintergrund findet seit langem statt. Dies geschieht auf sehr unterschiedliche Weise und unter unterschiedlichen Bedingungen, unterschiedlich von Bundesland zu Bundesland, von Region zu Region und von Schule zu Schule. Oft geschieht es in gegenseitiger Unkenntnis anderer Aktivitäten mit gleicher Zielrichtung. Das führt dazu, dass kaum auf den Erfahrungen anderer aufgebaut wird. Hier mehr Transparenz zu schaffen und einer größeren Systematik in der Praxis sprachlicher Förderung und Bildung den Weg zu bereiten, war der Anlass für das Modellprogramm FÖRMIG.

FÖRMIG war als ein lebendiges Programm konzipiert, das einen Austausch von Erfahrungen initiieren und darauf aufbauend Raum für das kontinuierliche Ausprobieren neuer Wege lassen sollte. Wie im Einzelnen die Förderung jeweils organisiert und durchgeführt wurde, unterlag keiner zentralen Steuerung. Es wurde bewusst angestrebt, dass passgenaue regionale oder lokale Ansätze in lokalen Netzwerken, sogenannten Basiseinheiten, entwickelt werden, die die angemessene Reaktion auf jeweils spezifische Bedarfslagen darstellen. Im Rahmen der generellen Perspektive, dass die Förderung bildungssprachlicher Fähigkeiten im Zentrum der Maßnahmen stehen soll, wurden die Ziele und die Mittel zu ihrer Erreichung vor Ort in Diskussions- und Aushandlungsprozessen festgelegt. Die Rahmenvorgaben des Programms wurden im Programmverlauf modifiziert und auf die konkrete Praxis in den teilnehmenden Einrichtungen und Bundesländern zugeschnitten. Abgesehen von leitenden Prinzipien nahm FÖRMIG - wie andere Modellprogramme - erst in der Umsetzung und im programmübergreifenden Austausch die konkrete Gestalt an, die es heute besitzt.

Auf diese Bedingungen ist das Design der FÖRMIG-Evaluation zugeschnitten. Es wurde ein Weg gewählt, der den gewollt ‚unübersichtlichen‘ konkreten Handlungsbedingungen des Programms – als eines Abbilds von realer Bildungspraxis – Rechnung trägt, ohne auf

systematisch erhobene Informationen zu verzichten, die allein einen gerechten Vergleich der unterschiedlichen Förderaktivitäten erlauben.

Die FÖRMIG-Evaluation berücksichtigt zu diesem Zweck mehrere Ebenen des Vergleichs:

Alle FÖRMIG-Basiseinheiten sind angetreten, mit ihren jeweiligen Fördermaßnahmen die sprachlichen Kompetenzen der beteiligten Kinder und Jugendlichen zu verbessern. Eine erste Überprüfung der Zielerreichung besteht darin festzustellen, ob in dieser Hinsicht überhaupt eine positive Entwicklung eintritt. Mit einer Messung des Sprachstandes zu Beginn und zum Ende eines Förderjahres lassen sich Veränderungen identifizieren. Das kann auf der Ebene des einzelnen Kindes oder Jugendlichen geschehen, für eine Fördermaßnahme erwartet man aber systematische Änderungen, die sich über den Einzelnen hinaus auf möglichst die gesamte Fördergruppe beziehen. Dieser Vergleichsmodus ist der *ipsative Vergleich*, die Beobachtungseinheit wird mit sich selbst im Zeitverlauf verglichen.

Die Frage „Hat sich der Schüler oder die Fördergruppe in diesem Schuljahr überhaupt verbessert?“ ist für Lehrkräfte oft von großem Interesse, aber aus der Perspektive der Programmevaluation ist ein *sozialer Vergleich* der Ergebnisse der beteiligten Fördergruppen untereinander wesentlicher: FÖRMIG ist ein „innovatives multizentrisches Programm“ (Haubrich 2006), in dem die konkreten Formen der Förderung überhaupt erst lokal entstehen. In einem solchen Kontext ist ein programminterner Vergleich – also der Vergleich der Fördermaßnahmen – geradezu als Kernbestandteil der Evaluation zu betrachten. Er dient der Identifizierung von besonders erfolgversprechenden Förderstrategien; der Vergleich wird zu anderen Strategien gezogen, die im Programm eingesetzt wurden. Betrachtet werden also die an FÖRMIG beteiligten Einrichtungen und die Ergebnisse, die sie erzielt haben. Die Evaluation erbringt Informationen darüber, in welchen Fördergruppen die Teilnehmer in dem betrachteten Schuljahr mehr sprachliche Kompetenzen hinzugewonnen haben als in anderen, die in der gleichen Zeit das gleiche Ziel auf anderen Wegen verfolgen.

Die Veränderungen in den Kompetenzen können vielen Ursachen geschuldet sein, die nicht direkt auf die spezifische sprachliche Bildungsmaßnahme in FÖRMIG zurückzuführen sind. Zu berücksichtigen ist zum Beispiel, dass sich die Fördergruppen in Bezug auf die Eingangsselektivität (z.B. hinsichtlich der kognitiven Lernvoraussetzungen, des sozioökonomischen Status oder des familiären Bildungshintergrunds ihrer Schüler) unterscheiden. Möglicherweise erklären diese Unterschiede die Differenzen in den Lernfortschritten. Um den Einfluss solcher Faktoren auf die gemessenen sprachlichen Effekte zu neutralisieren, wurde in FÖRMIG mit Hilfe multivariater Verfahren ein „fairer

Vergleich“ durchgeführt. Dazu wurden für jede Fördergruppe Erwartungswerte für ihre Ergebnisse bei den Leistungsmessungen am Ende eines Förderjahres berechnet, die die sprachlichen Kompetenzen zu Beginn der Förderung sowie persönliche und familiäre Merkmale ihrer Teilnehmer berücksichtigen. Diese Erwartungswerte entsprechen den durchschnittlichen Ergebnissen einer aus der FÖRMIG-Gesamtgruppe statistisch errechneten Fördergruppe mit vergleichbaren Rahmenbedingungen. Die Residuen, also die Abweichungen von den Erwartungswerten, werden so zum Gegenstand des Vergleichs der Fördergruppen untereinander.

Die Identifizierung von Fördergruppen, in denen in Relation zu den individuellen Voraussetzungen ihrer Teilnehmer besonders deutliche Lernfortschritte erzielt wurden, ist ein Zwischenschritt für die Identifizierung erfolgversprechender Merkmale der Bildungsmaßnahmen. Die FÖRMIG-Evaluation verfolgt zu diesem Zweck zwei unterschiedliche Strategien: Zum einen werden zentral gesammelte Informationen auffällig abschneidender Basiseinheiten qualitativ gesichtet, um die besonderen Merkmale ihrer Förderstrategien herauszuarbeiten. Zum anderen wird über Mehrebenenanalysen quantitativ geprüft, inwieweit Lernfortschritte systematisch mit theoretisch bedeutsamen Merkmalen der Förderung und des Unterrichts in Verbindung gebracht werden können.

Neben FÖRMIG-internen Vergleichen wird in Teilbereichen auch ein externer, *kriterialer Vergleich* einbezogen. Es wurden Leseverständnistests am Übergang von der Grundschule in die Sekundarstufe und von der Sekundarstufe in den Beruf eingesetzt. Sie entstammen den (Querschnitts-)Untersuchungen IGLU und PISA. Dieses Vorgehen erlaubt punktuelle Vergleiche mit repräsentativen Leistungsdaten aus dem Regelschulsystem. FÖRMIG hat sich konsequent auf die Schülerinnen und Schüler konzentriert, die in Fördermaßnahmen aufgenommen waren, um zu sehen, welche unterschiedlichen Entwicklungen sie in den verschiedenen beteiligten Einrichtungen und Sprachbildungsnetzwerken gemacht haben. Der Vergleich mit den externen Daten gibt in Bezug auf das Leseverständnis einen Anhaltspunkt dafür, welche Leistungen sie im Vergleich mit Schülern aus repräsentativen Untersuchungen erzielten.

FÖRMIG und die Evaluation des Programms ist also von der Grundanlage her keine klassische Interventionsstudie. Geprüft wird nicht, ob Förderung als solche im Vergleich zu unterlassener Förderung überhaupt einen Beitrag zur sprachlichen Kompetenzentwicklung leistet. Gegenstand ist auch nicht eine begrenzte Auswahl bestimmter, ausschließlich theoretisch motivierter Fördermodelle, deren Wirkungen sich gegenüber den jeweils anderen Fördermodellen oder einer „Nichtförderung“ experimentell relativ einfach bestimmen ließen. Ausgangssituation von FÖRMIG und damit auch der Programmevaluation ist die faktische

Vielfalt von Förderung in der Praxis. Nicht allein die Vielfalt der unterschiedlichen Herangehensweisen ließ ein experimentelles Kontrollgruppendesign für ungeeignet erscheinen, auch waren die Bestandteile der in den Länderprojekten sich entwickelnden Fördermaßnahmen nicht im Vorhinein festgelegt. Eine trennscharfe Zuweisung zu einer Experimental- und einer Kontrollgruppe war weder vorgesehen noch angebracht - bezieht man in die Abwägungen das ethische Dilemma ein, das sich stellt, wenn einigen Förderbedürftigen eine regional und zielgruppenspezifisch angepasste Förderung vorenthalten wird, weil eine Non-Treatment-Gruppe für erforderlich gehalten wird.

Die FÖRMIG-Evaluation vergleicht die sprachlichen Fortschritte, die in verschiedenen Fördergruppen erzielt wurden, miteinander, und sie tut dies unter statistischer Kontrolle einer Reihe von personengebundenen Merkmalen, die sich in der Forschung als mögliche Störvariablen für einen Fördereffekt herausgestellt haben. Damit erbringt die Evaluation Ergebnisse, die zu einem systematischen und fairen Vergleich der vielfältigen Förderaktivitäten geeignet sind, die sich im Programm FÖRMIG als Abbild einer heterogenen Bildungslandschaft mit jeweils unterschiedlichen Randbedingungen entwickelt haben – nicht mehr, aber auch nicht weniger.

Literatur:

Haubrich, K. (2006): Die Konstruktion des Untersuchungsgegenstandes in der Evaluation innovativer multizentrischer Programme. In: Karl-Siegbert Rehberg (Hrsg.): Soziale Ungleichheit - Kulturelle Unterschiede, Verhandlungen des 32. Kongresses der Deutschen Gesellschaft für Soziologie in München 2004. Frankfurt: Campus, S. 3872-3881.

Klinger, T., Schwippert, K. & Leiblein, B. (Hrsg.) (2008): Evaluation im Modellprogramm FÖRMIG. Planung und Realisierung eines Evaluationskonzepts. FÖRMIG Edition Bd. 4. Münster, New York, München, Berlin: Waxmann.

Nickolaus, R. & Schnurpel, U. (2001): Innovations- und Transfereffekte von Modellversuchen in der beruflichen Bildung. Band 1. Bonn: Bundesministerium für Bildung und Forschung (BMBF).

Wottawa, H. & Thierau, H. (2003): Lehrbuch Evaluation, 3., korrigierte Auflage. Bern, Göttingen, Toronto, Seattle: Huber.

Hamburg, im Oktober 2009

Universität Hamburg, Institut für International und Interkulturell Vergleichende Erziehungswissenschaft