

Stichprobe

Die Gesamtstichprobe bestand aus einer Interventionsgruppe (IG: Einsatz KvDaZ-Konzept) und einer Kontrollgruppe (KG: bisherige DaZ-Förderung). Für die Sicherstellung einer strukturähnlichen Stichprobe wurden vier Einschlusskriterien formuliert:

1. Sprachanfänger:innen: Zunächst durch Einschätzung der Lehrpersonen. Anhand eines Screeningverfahrens wurde zusätzlich eine vereinfachte Profilanalyse nach Grießhaber (2013) durchgeführt. Sprachaufnahmen mit mindestens fünf Äußerungen auf Profilstufe 0 oder 1 wurden in die Stichprobe aufgenommen.
2. Kontaktduer zum Deutschen: Unter Einbezug von Elternfragebögen mit Angaben zu personbezogenen und sprachbiographischen Daten der Kinder (Kap. 11.1.1) wurden in der Regel Fälle in die Stichprobe aufgenommen, welche zum ersten Messzeitpunkt maximal zwölf Monate in Deutschland waren.
3. Datengrundlage: Der vorhandene Datensatz wurde auf Vollständigkeit geprüft. Es wurden lediglich die Fälle aufgenommen für die zu beiden Messzeitpunkten die freien Sprachproben vorlagen sowie alle notwendigen Erhebungsinstrumente weitestgehend vorhanden waren.
4. Aufnahmelänge: Durch den Vergleich von Excel-Listen wurde eine ähnliche Aufnahmelänge beider Stichproben angestrebt. Die Reliabilität der Angleichung wurde durch eine unabhängige Person geprüft. Der Übereinstimmungskoeffizient nach Hirschmann (2019) betrug $r_{ii} = 67\%$ (IG) und $r_{ii} = 100\%$ (KG) (S. 98).

Aufgrund dieser Einschlusskriterien wurden in einem ersten Analyseverfahren von ursprünglich 63 Kindern im Alter von 6.0 bis 10.11 Jahren insgesamt 48 Kinder aus zwölf Schulen für die summative Evaluation der Grammatikentwicklung berücksichtigt sowie 50 Kinder für die Wortschatzanalyse ($\text{Sum}_{\text{WS } 1-3}$). Aufgrund einer erforderlichen Angleichung der Aufnahmelängen zum Vergleich der Wortschatzanalysen wurde das vierte Einschlusskriterium auf die Stichprobe $N = 50$ angewendet. Für die Wortschatzanalyse konnten daher 40 Fälle berücksichtigt werden (Tab. 19).

Tab. 19: Stichprobe KvDaZ-Studie (IG = Interventionsgruppe, KG = Kontrollgruppe)

Stichprobe	N	IG	KG
Anzahl Schule	12	4	8
Anzahl Sprachaufnahmen	63	38	25
Summativ(Sum_{CR})	48	29	19
Summativ($\text{Sum}_{\text{WS } 1-3}$)	50	29	21
Summativ($\text{Sum}_{\text{WS } 1-4}$)	40	20	20

Die Stichprobenauswertung bezieht sich im Folgenden auf den Datensatz $N = 40$: Das Durchschnittsalter der Kinder lag in der Interventionsgruppe bei 7.00 Jahre, in der Kontrollgruppe bei 7.11 Jahren. Die Geschlechterverteilung war zwischen beiden Gruppen ähnlich verteilt. 53 % der Interventionsgruppe waren weiblich, entsprechend 47 % männlich. Auch bei der Kontrollgruppe war der Anteil der weiblichen Schülerinnen mit 57 % leicht höher als der Anteil der männlichen Schüler (43 %). In beiden Gruppen wurde mit der systematischen Förderung von Deutsch als Zweitsprache meist zum Schuljahr 2017/2018 begonnen. Die durchschnittliche Aufenthaltsdauer betrug zu diesem Zeitpunkt für die Interventionsgruppe 21 Monate und für die Kontrollgruppe 16 Monate. 70 % der Schüler:innen hatten seit maximal zwölf Monaten Kontakt zum Deutschen. Ein besonderes Merkmal beider Stichproben war die Sprachenvielfalt (IG_{WS}: $n = 9$; KG_{WS}: $n = 10$). Zu den drei häufigsten Familiensprachen zählte Kurdisch, Arabisch, Griechisch (IG_{WS}) sowie Arabisch, Italienisch, Türkisch (KG_{WS}).

Datenerhebung und -aufbereitung

Als Datengrundlage wurden freie Sprachproben zu zwei Messzeitpunkten (4 Monate zwischen t₁ und t₂) jeweils für die Interventions- und Kontrollgruppe erhoben (ausführliche Beschreibung der Datenerhebung und -aufbereitung in Kap. 11.2, 11.3). Die Beachtung mündlicher (spontansprachlicher) Daten wird in der schulbezogenen Forschung bisher als randständig wahrgenommen, da der Fokus vor allem auf schriftlichen Daten liegt (Schramm & Marx, 2017, S. 212).

Zur indirekten Veränderungserfassung des Wortschatzes wurde die Anzahl der insgesamt verwendeten Wörter (Token) sowie die Anzahl der verschiedenen Wörter (Lemma-Types) pro Schüler:in, jeweils für t₁ und t₂ über alle kommunikativen Kontexte hinweg ermittelt. Auf Grundlage dieser Vorgehensweise konnten Boxplots für jeweils die Interventions- und Kontrollgruppe pro Messzeitpunkt berechnet werden. Die Grammatikentwicklung wurde mithilfe der Profilanalyse nach Grießhaber (2013) und den daraus gewonnenen Profilstufen gemessen. Die Ergebnisse der Grammatikentwicklung werden in dieser Arbeit nicht näher beschrieben. Weitere forschungsbasierte Informationen und Ergebnisse zum KvDaZ-Projekt sind bei Boenisch et al. (2021) sowie Lingk et al. (2023) zu finden.

Reliabilitätsprüfung

Die Interrater-Reliabilität wurde über den Übereinstimmungskoeffizienten (Anzahl der übereinstimmenden Fälle geteilt durch Gesamtheit der analysierten Fälle) berechnet (Hirschmann, 2019, S. 98). Dafür wurden 25 % ($n = 10$) der Sprachaufnahmen aus dem KvDaZ-Datensatz (SUM_{WS1-3}) randomisiert ausgewählt, erneut transkribiert und als Wortlisten aufbereitet. Der Übereinstimmungskoeffizient der Wortschatzlisten für den t₁-Datensatz entsprach bei den Token $r_{ü} = 0.99$ (99 %) und bei den Lemma-Types $r_{ü} = 0.99$ (99 %). Für den t₂-Datensatz betrug die Übereinstimmung bei den Token $r_{ü} = 0.99$