

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 t_1 und t_2) 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 t_1 und t_2 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 Grieshaber (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 t_1 -Datensatz entsprach bei den Token $r_u = 0.99$ (99 %) und bei den Lemma-Types $r_{\bar{u}} = 0.99$ (99 %). Für den t_2 -Datensatz betrug die Übereinstimmung bei den Token $r_u = 0.99$

(99 %) und bei den Lemma-Types $rü = 0.99$ (99 %). Damit ließ sich die Reliabilität als sehr hoch einstufen.

Ergebnisdarstellung – Lexikalische Entwicklung

Die Ergebnisse zur lexikalischen Entwicklung werden im Folgenden für die Token und Lemma-Types beschrieben (Abb. 13).

Token

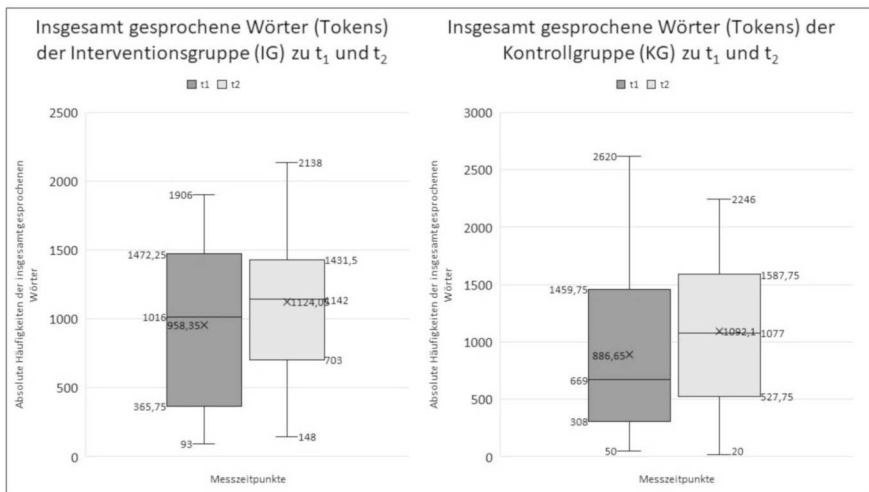


Abb. 13: Insgesamt gesprochene Wörter (Token) im Vergleich zwischen Interventionsgruppe und Kontrollgruppe (Lingk et al., 2023)

Die Interventionsgruppe nutzte zum *ersten Messzeitpunkt* (t₁) durchschnittlich mehr Token als die Kontrollgruppe (IG: $M = 958.35$, KG: $M = 886.65$). Dies bestätigte sich auch durch die Betrachtung des Interquartilsabstands. 50 % der Kinder aus der Interventionsgruppe sprachen zwischen $Q1 = 365.75$ Token und $Q3 = 1472.25$ Token. In der Kontrollgruppe sprachen 50 % der Kinder zwischen $Q1 = 308$ und $Q3 = 1459.75$ Token. Es wird deutlich, dass die Kinder aus beiden Untersuchungsgruppen bereits bei t₁ einen gewissen Wortschatz im Deutschen erworben hatten. Auf der lexikalischen Ebene waren die Kinder demnach in beiden Gruppen keine absoluten Sprachanfänger:innen. Die Interventionsgruppe sprach, gemessen an den drei Quartilen, insgesamt mehr Wörter (Token). Anhand der Maximalwerte ließ sich erkennen, dass sowohl in der Interventions- als auch in der Kontrollgruppe jeweils Kinder waren, die sehr viel sprachen (IG: $x_{\max} = 1906$; KG: $x_{\max} = 2138$), aber auch recht wenig (IG: $x_{\min} = 93$; KG: $x_{\min} = 50$).

Zum *zweiten Messzeitpunkt* (t₂) sprach die Interventionsgruppe durchschnittlich $M = 1124.05$ Wörter insgesamt. Die Kontrollgruppe nutzte durchschnittlich $M = 1092.1$ Wörter insgesamt. 50 % der Kinder aus der Interventionsgruppe sprachen zwischen $Q1 = 703$

und $Q_3 = 1\,431.5$ Token. In der Kontrollgruppe sprachen 50 % der Kinder zwischen 527.75 und 1 587.75 Token.

Vergleichsanalyse Token

Zwischen t_1 und t_2 war die Anzahl der genutzten Wörter (Token) in beiden Gruppen größer geworden. Der durchschnittliche Zuwachs der Token zwischen beiden Messzeitpunkten betrug bei der Interventionsgruppe $M = 166$ Token und bei der Kontrollgruppe $M = 205$ Token. Der Interquartilsabstand lag bei der Interventionsgruppe zu t_2 geringer als bei t_1 (t_1 : IQA = 1 106.5; t_2 : IQA = 728.5). Dabei blieb das dritte Quartil fast auf ähnlichem Niveau (t_1 : $Q_3 = 1\,472.25$; t_2 : $Q_3 = 1\,431.5$). Das erste Quartil stieg jedoch um 337.25 Token auf $Q_3 = 703$. Demnach sprachen zum zweiten Messzeitpunkt 75 % der Kinder mehr als 703 Token.

Der Interquartilsabstand der Kontrollgruppe verringerte sich gegenüber der Interventionsgruppe nur marginal von t_1 zu t_2 (t_1 : IQA = 1 151.75; t_2 : IQA = 1 060). Das dritte Quartil stieg bei der Kontrollgruppe von $Q_3 = 1\,459.75$ auf $Q_3 = 1\,587.75$ Token. Der Maximalwert nahm jedoch zwischen den beiden Messzeitpunkten ab (t_1 : max = 2 620; t_2 : max = 2 246). Das erste Quartil stieg um 219.75 auf $Q_1 = 527.75$ Token.

Bei der Betrachtung des arithmetischen Mittels war der Zuwachs bei der Kontrollgruppe etwas größer als bei der Interventionsgruppe – die Zuwächse lagen aber nah beieinander und können daher als ähnlich betrachtet werden. Eine Verringerung des Interquartilsabstands bei der Interventionsgruppe im Vergleich zur Kontrollgruppe ließ sich eindeutig herausarbeiten.

Die Veränderung der genutzten Token zwischen den beiden Messzeitpunkten lässt jedoch noch keine Rückschlüsse über die Variabilität des Wortschatzes zu. Dafür war die Betrachtung der durchschnittlichen Zuwächse an unterschiedlich genutzten Wörtern (Lemma-Types) beider Gruppen notwendig (Abb. 14).

Lemma-Types

Die Anzahl der unterschiedlichen Wörter im Gebrauchswortschatz war zu t_1 bei der Interventionsgruppe ($M = 164$ Lemma-Types) etwas geringer als bei der Kontrollgruppe ($M = 172$ Lemma-Types). Die Kontrollgruppe nutzte nur $M = 8$ Types mehr als die Interventionsgruppe und ließ sich daher als ein sehr ähnliches Ausgangsniveau deuten. 50 % der Kinder der Interventionsgruppe nutzten zum ersten Messzeitpunkt zwischen $Q_1 = 89$ und $Q_3 = 215.75$ unterschiedliche Wörter (Lemma-Types) (t_1 : IQA = 126.75). Die Hälfte der Kinder aus der Kontrollgruppe nutzte zwischen $Q_1 = 76$ und $Q_3 = 263.25$ Lemma-Types (t_2 : IQA = 187.25). Das sprachliche Niveau der Kontrollgruppe konnte aufgrund der Mini- und Maximalwerte im Vergleich zur Interventionsgruppe als heterogener eingestuft werden.

Bei der Interventionsgruppe stieg zum zweiten Messzeitpunkt die Anzahl der Lemma-Types auf $M = 210$ und bei der Kontrollgruppe auf $M = 185$ Lemma-Types an. Der Zuwachs (größere Anzahl) lag bei der Interventionsgruppe demnach bei $M = 46$ Lemma-Types und bei der Kontrollgruppe bei $M = 13$ Lemma-Types. Im Vergleich der Mediane lag

der Zuwachs für die Interventionsgruppe bei $x_{\text{Med}} = 40$ Types und für die Kontrollgruppe bei $x_{\text{Med}} = 13$ Lemma-Types.

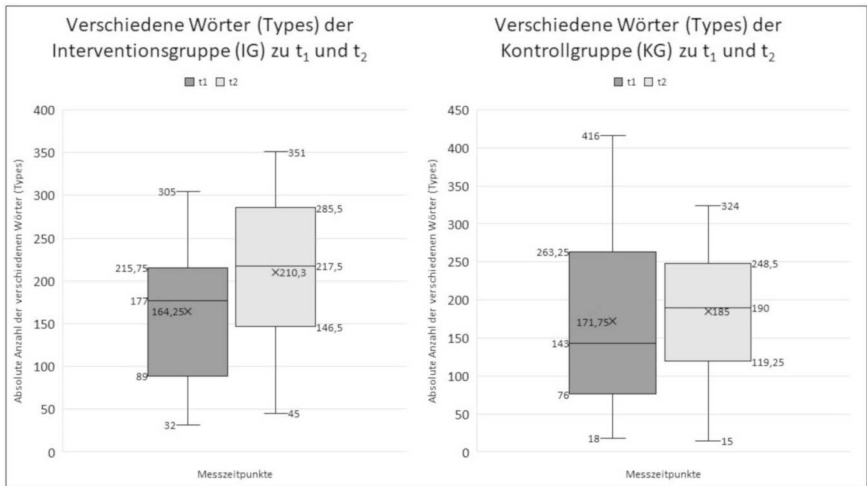


Abb. 14: Verschieden genutzte Wörter (Lemma-Types) im Vergleich zwischen Interventionsgruppe und Kontrollgruppe (Lingk et al., 2023)

Vergleichsanalyse Lemma-Types

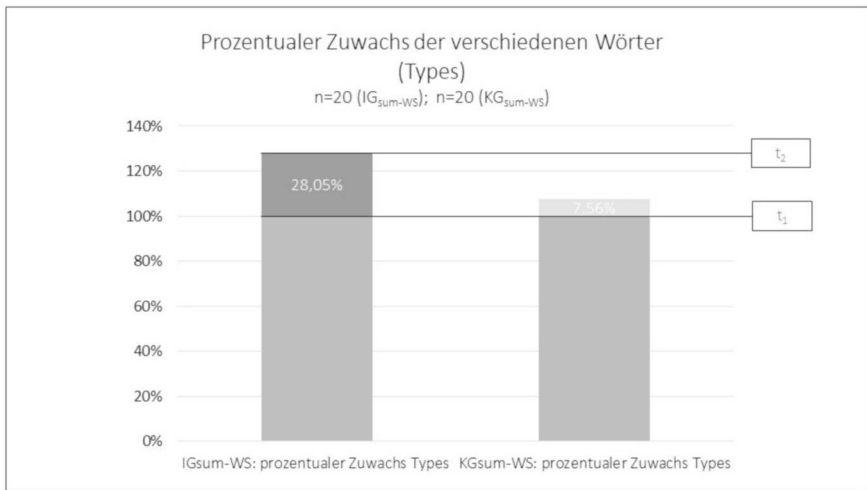


Abb. 15: Prozentualer Zuwachs Lemma-Types im Vergleich zwischen Interventionsgruppe und Kontrollgruppe (Lingk et al., 2023)

Der Interquartilsabstand blieb bei der Interventionsgruppe zu beiden Messzeitpunkten ähnlich groß (t_1 : IQA = 126.75; t_2 : IQA = 139). Sprachen bei t_1 75 % der Kinder mehr als $Q_1 = 89$ unterschiedliche Wörter (Lemma-Types), waren es bei t_2 bereits mehr als $Q_1 = 146.5$ Lemma-Types. Diese Veränderung bestätigte sich auch bei der Betrachtung der Mini- und Maximalwerte, denn auch diese stiegen in der Interventionsgruppe an.

In der Kontrollgruppe verringerte sich der Interquartilsabstand zu beiden Messzeitpunkten (t_1 : IQA = 187.25; t_2 : IQA = 129.25). Auch in der Kontrollgruppe veränderte sich das erste Quartil von t_1 zu t_2 positiv (t_1 : $Q_1 = 76$; t_2 : $Q_1 = 119.25$). Der Maximalwert sowie das dritte Quartil wurden jedoch kleiner. Der Maximalwert nahm um 98 Lemma-Types ab, der Minimalwert blieb auf ähnlichem Niveau.

In der Interventionsgruppe konnten zum zweiten Messzeitpunkt durchgängig mehr verschieden genutzte Wörter nachgewiesen werden. Die Interventionsgruppe hatte im arithmetischen Mittel mehr als ein Viertel ihres ursprünglichen Wortschatzes dazugewonnen, die Kontrollgruppe ein Zwölftel. Der Zuwachs der Interventionsgruppe war damit im Vergleich zur Kontrollgruppe fast viermal größer.

Diskussion und Ausblick Die Ergebnisse zur Wortschatzentwicklung zeigen, dass Kinder, die mit dem KvDaZ-Konzept gefördert wurden, innerhalb von vier Monaten deutlich mehr unterschiedliche Wörter (Lemma-Types) im Alltag sprechen als Kinder aus der Kontrollgruppe (+21 %). Unklar bleibt, welche Wörter genau gesprochen wurden (Funktions- oder Inhaltswörter). Neben der Wortschatzerweiterung (Lexik) kann dies auch als ein Hinweis für die Wirksamkeit des KvDaZ-Konzeptes im Bereich des flexiblen Sprachgebrauchs (pragmatische Kompetenz) gedeutet werden. Apeltauer (2020) verdeutlicht, dass Kinder beim Wortschatzerwerb auf viele Wiederholungen angewiesen sind, damit entsprechende Vernetzungen im mentalen Lexikon aufgebaut werden. »Ein seltener Gebrauch [von Wörtern] begünstigt auch sprachliche Vereinfachungen und das Entstehen defekter Formen (d.h. *Sprachkorrosion*)« (Apeltauer, 2020, S. 308). Es wird angenommen, dass Kindern aus der Kontrollgruppe vor allem Wörter aus dem Inhaltswortschatz (Nomen, Verben, Adjektive) vermittelt wurden. Die damit verbundenen anspruchsvolleren Verarbeitungsmechanismen (Seifart et al., 2018) sowie die seltene Aktivierung derselben (Boenisch, 2014b) können eine Erklärung für die geringere Anzahl unterschiedlich genutzter Wörter nach vier Monaten DaZ-Förderung sein.

Das didaktische Grundgerüst sowie die didaktischen Begleitmaterialien aus dem KvDaZ-Konzept sind so konzipiert, dass alle Kinder voraussetzungslos von den sprachförderlichen Angeboten profitieren können. Mit dem Fokus auf Kernvokabular in kommunikativ bedeutungsvollen Situationen können sich die Schüler:innen aktiv beteiligen. Weiterhin wird vermutet, dass auf dieser Grundlage bildungssprachliche Fähigkeiten leichter erworben werden können, weil den Kindern bereits ein sprachliches Fundament zur Ausdifferenzierung ihres Wortschatzes zur Verfügung steht (Kap. 7.2.1).

Aus der KvDaZ-Studie lassen sich erste Tendenzen zum Einfluss einer kernvokabularbasierten Sprachförderung auf den anfänglichen DaZ-Erwerb ableiten. Für eine abschließende Ursache-Wirkungs-Relation sind weitere Forschungen notwendig.

