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 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 ($SUM_{WS1\text{-}3}$) randomisiert ausgewählt, erneut transkribiert und als Wortlisten aufbereitet. Der Übereinstimmungskoeffizient der Wortschatzlisten für den t_1-Datensatz entsprach bei den Token rü = 0.99 (99 %) und bei den Lemma-Types *rü* = 0.99 (99 %). Für den t_2-Datensatz betrug die Übereinstimmung bei den Token *rü* = 0.99