

DOI: 10.37943/AVFP6158

Kenzhekul Tuganbekova

Candidate of Pedagogical Sciences, Associate Professor, Faculty of Pedagogy
klaratuganbekova@mail.ru, orcid.org/0000-0002-9434-3180
Buketov Karaganda National Research University, Kazakhstan

Zaryana Olexyuk

PhD, Associate Professor, Faculty of Pedagogy
Oleksyuk_Z@mail.ru, orcid.org/0009-0004-5810-1477
Buketov Karaganda National Research University, Kazakhstan

Zhanar Akhmetzhanova

Candidate of Philological Sciences, Associate Professor, Faculty of Pedagogy
zhanar-ahmetzhan@mail.ru, orcid.org/0000-0001-5212-162X
Buketov Karaganda National Research University, Kazakhstan

**TABLET-BASED AAC IN A KAZAKHSTANI SPECIAL SCHOOL:
IMPLEMENTATION, PRACTITIONER WORKLOAD AND INDIVIDUAL RESPONSE**

Abstract: Non-verbal children in Kazakhstani special schools communicate mainly through paper picture cards and gestures. This non-randomised, single-centre pilot study followed twelve children aged 6–10 with autism spectrum disorder or severe developmental language impairment at Special School No. 4 in Karaganda during the 2024–2025 academic year. Six children used Proloquo2Go or LetMe Talk on tablets, and six continued with cards and gestures. Outcomes were spontaneous communicative initiation, video-coded at three time points; a symbol comprehension test with taught and untaught items; and the Vineland-3 Communication Domain, rated separately by teachers and parents.

The tablet group improved more on all measures, although no difference was statistically significant. At endpoint, the tablet group averaged 7.4 initiations per 15-minute observation, compared with 5.1 in the card group. On untaught symbols, scores were 42% and 26% respectively, and the card group's result did not differ from chance. Vineland-3 change scores in the tablet group were 5.2 (teacher) and 7.8 (parent), compared with 2.9 and 3.7 in the card group. Because neither application supports Kazakh, the devices were set to Russian and practitioners spoke the child's home language alongside them. This did not affect the comparison between methods, but it reflects the real conditions of use in this setting. Individual outcomes within the tablet group varied more than the group means differed from each other. Setting up the devices took about 60 practitioner hours, mostly in the first two months. With six children per group, the study cannot show which children benefit most. It was designed to estimate effect sizes for a larger controlled trial, not to establish that one method is superior.

Keywords: non-verbal children, augmentative and alternative communication, AAC (augmentative and alternative communication), Proloquo2Go, LetMe Talk, implementation, Kazakh-Russian bilingual classroom, symbol acquisition, special education, Kazakhstan.

Introduction

Non-verbal children - those who do not produce functional speech due to autism spectrum disorder, cerebral palsy, or severe developmental language impairment - rely entirely on the adults around them to provide workable communication tools. In Kazakhstani special schools, those tools have historically meant paper picture cards and manual gesture systems. Both have been used for decades, both have supporting evidence, and both have the same practical ceiling: a child cannot initiate a request, reject an offer, or comment on an event unless a trained partner is present, the right card is physically available, and the gesture is part of a shared code that the partner recognizes. When any of these conditions is missing, communication stops.

The authors observed this pattern directly. Working with non-verbal children aged 6 to 10 at Special School No. 4 for children with speech disorders in Karaganda during the 2024-2025 academic year, we repeatedly saw the same sequence: children who had used paper PECS cards for one or two years could request familiar items during structured sessions but rarely initiated communication on their own, and rarely transferred what they had learned to a different room, a different adult, or a different time of day. Picture cards work within limits. The question this raised was whether a different tool could extend those limits.

Proloquo2Go and LetMe Talk are symbol-based AAC applications that run on tablets and smartphones. Each symbol press produces synthesized speech output immediately, without a partner needing to interpret a gesture or locate the correct card. Vocabulary is customizable, organized by category, and portable in principle, since the device can travel with the child - though whether it does depends on whether a school can let devices leave the building. These are not minor technical conveniences. Sterrett et al. [1] showed that the immediacy of auditory feedback from speech-generating devices is directly associated with increases in spontaneous vocal output in minimally verbal children with ASD - meaning the device creates conditions under which speech becomes more likely, and does more than substitute for it. Tamanaha and Olivatti [2] and Koudys et al. [3] confirmed that symbol-based intervention programs produce gains in communicative initiation when feedback is immediate and consistent, which is structurally built into both applications used in this study.

The comparison between digital and traditional AAC tools is not straightforward, because implementation quality shapes outcomes as much as the technology itself. May et al. [4] found that across special education settings using PECS, outcome variability was explained less by the method and more by how consistently it was applied. Denne et al. [5] reviewed 12 years of implementation data and identified weak practitioner training as the most consistent predictor of poor AAC outcomes - a finding that shaped how we structured the professional preparation component of this study. Alfuraih et al. [6] demonstrated that even in settings with limited resources, structured symbol-based requesting programs produce generalizable communicative gains when implementation is monitored and corrected in real time.

The evidence on speech-generating devices specifically is substantial. Yuan et al. [7] analyzed 18 studies involving children with ASD and severe communication impairment and found that 87% of participants showed gains in vocalizations following SGD-based intervention - a figure that holds across age groups, diagnostic profiles, and device types. Alzahrani [8] showed that apps adapted to the child's linguistic environment produce stronger engagement than generic tools, which is relevant in Kazakhstan where children operate in Kazakh, Russian, or both. Alzyoudi and Alghazali [9] found, in a controlled iPad-based study with minimally verbal students, that the most pronounced gains appeared in unprompted communicative acts - exactly the outcome that paper card systems struggle to produce.

Two findings from the training and design literature converge. Practitioner knowledge, skill and confidence improve most when instruction is paired with supervised device use [10], and programmes combining visual symbols with synthesized speech consistently outperform visual-only or speech-only approaches [11]. Ruiz-Martinez et al. [12] mapped current AAC technology and identified app-based systems as the most practically deployable option available to classroom teachers, ahead of more experimental modalities such as eye-tracking or brain-computer interfaces. Generalization is stronger where digital AAC is embedded in daily classroom activity instead of confined to therapy sessions [13]. Symbol acquisition speed and spontaneous initiation rate are the measures most sensitive to change in such programmes [14]. We took both points into the design: sessions ran inside the ordinary timetable, and these two measures became the primary outcomes.

Holyfield et al. [15] developed and preliminarily tested a prototype that uses AI to automate augmented input by recognising a communication partner's speech and surfacing relevant symbols in real time - a direction distinct from the applications used in this study, which do not include

such AI-driven automation. Suhr et al. [16] found that structured versus unstructured classroom contexts significantly affect how often nonspeaking children use their AAC devices throughout the day, which led us to standardize the physical setup across all intervention sessions. Joseph and Babu [17], in an opinion piece on generative AI and mental healthcare equity for non-verbal autistic individuals, argue for a cautious, ethically grounded approach to AI adoption in this population, distinct from evaluating any specific application. Beirat et al. [18] systematically reviewed AI-based diagnostic, therapeutic and educational technologies for autism spectrum disorder and highlighted communication and educational applications as one of the areas with clear potential, alongside diagnostic and robotic interventions. Proloquo2Go and LetMe Talk, the applications used in this study, are rule-based symbol-selection tools and do not incorporate the AI-driven functions discussed by Joseph and Babu [17] and Beirat et al. [18]; those works are cited here to situate app-based AAC within the broader assistive-technology landscape, not to characterise the technology evaluated in this study.

Kazakhstan registered 237,000 children with special educational needs at the start of the 2025-2026 academic year. The Karaganda region operates 10 specialised schools covering six categories of developmental and communicative disorders, including one school specifically for children with speech impairments.

Two features of this setting mean that findings from the studies reviewed above cannot be transferred to it directly. The first is linguistic. Proloquo2Go offers no Kazakh voice, and LetMe Talk depends on the speech engine installed on the device, where Kazakh support is partial at best. Children in Karaganda special schools live in Kazakh, Russian, or a mixture of both, and Alzahrani [8] found that applications matched to the child's linguistic environment produce markedly stronger engagement than generic ones. Soto and Clarke [19] followed fifteen school-based AAC providers over ten months and identified the main obstacles to bilingual implementation as systemic constraints and a monolingual default in service delivery, rather than limitations of the devices themselves. The providers kept bilingual AAC running through workarounds: collaboration with bilingual colleagues, interpreted sessions, and modelling of the home language alongside a device that could not produce it. The arrangement used in this study, a Russian-language device voiced in Kazakh by the practitioner, is the same workaround under a different pair of languages. Every deployment in this region therefore begins with a localisation problem that none of the published trials had to solve, and the outcomes those trials report were obtained under conditions that no Kazakhstani practitioner can reproduce without local adaptation. The second is preparation. Class teachers in these schools receive no pre-service instruction in augmentative and alternative communication, while Denne et al. [5] identify weak practitioner training as the most consistent predictor of poor AAC outcomes across twelve years of implementation data.

Against this background, no empirical study has examined how digital AAC applications behave in Kazakhstani special education. Practitioners have no local answer to three questions that precede any decision to adopt them: whether the applications can be configured for a bilingual Kazakh-Russian classroom at all, what that configuration costs a teacher in working hours, and what magnitude of change is realistic within a single academic year.

The published effect sizes answer the third question only in part. Therrien et al. [20] synthesised twenty SGD intervention studies with autistic preschoolers. Effects on requesting were small to high ($NAP = 0.65-1.0$) where device-mediated and spoken requests were counted together, and mixed or non-significant where spoken output alone was the outcome. The authors also rated the ecological validity of most included studies as limited, since few interventions ran inside ordinary classroom routines. Both results bear on what a single school year in a Karaganda special school can be expected to produce.

The present work is therefore designed as a non-randomised, single-centre pilot evaluation study with a small sample: its purpose is estimative, not confirmatory. Six children per condition at one school, allocated without randomisation, cannot establish that one method is superior to the

other. What such a design can show is whether change moves in the expected direction, roughly how large it is, and how much it varies from child to child - the parameters a larger multi-site trial would need before it could be planned. Jiang and Wang [21] analysed a corpus of naturalistic language samples from eighteen minimally speaking autistic children at three time points after aided AAC and caregiver training were introduced. Every child moved from minimal or prompt-dependent use towards spontaneous use, but the levels reached differed widely, and the gains concentrated in requesting, joint attention and commenting, while regulatory and emotional functions lagged. A group mean describes a distribution of that shape poorly, so individual trajectories are reported here alongside condition averages.

Materials and research methods

The study was designed as a non-randomised, single-centre pilot evaluation study with a small sample; its limitations are collected in a separate section below. It ran at Special School No. 4 for children with speech disorders in Karaganda through the 2024-2025 academic year. Participants were 12 non-verbal children aged 6 to 10. Inclusion required a diagnosis of autism spectrum disorder or severe developmental language impairment issued by the regional psychological-medical-pedagogical consultation, absence of functional speech as a primary communicative means at enrolment, and vision and hand function sufficient for either touch-screen or card use. Children with uncorrected visual impairment or motor limitations preventing both were not enrolled.

Allocation was not randomized. We ordered the children by baseline initiation rate, paired them off, and put one member of each pair on tablets. Within each matched pair, assignment to condition was made based on classroom schedule - which children's timetable slots aligned with device and practitioner availability - not by a randomization procedure; this is a further source of potential selection bias beyond the baseline matching, and we note it as a limitation. This procedure balances baseline communication level. It does not control for home language, parental involvement, or prior therapy, any of which could differ between the groups.

Table 1 gives the twelve participants at baseline.

Table 1. Participant characteristics at baseline

Child	Age	Diagnosis	Sex	Home language	Prior PECS (months)	Baseline initiation	Attendance (%)	Application
I1	6	ASD	M	Kazakh	8	2.1	91	LetMe Talk
I2	7	DLI	F	Russian	14	2.4	88	LetMe Talk
I3	7	ASD	M	Kazakh	11	2.9	94	LetMe Talk
I4	9	ASD	M	Russian	19	3.3	86	Proloquo2Go
I5	9	DLI	F	Kazakh	16	4.0	92	Proloquo2Go
I6	10	ASD	M	Russian	22	4.5	90	Proloquo2Go
C1	6	ASD	F	Kazakh	9	2.0	93	-
C2	7	ASD	M	Russian	12	2.3	87	-
C3	8	DLI	M	Kazakh	15	2.8	95	-
C4	8	ASD	F	Russian	17	3.2	89	-
C5	9	DLI	M	Kazakh	20	3.6	91	-
C6	10	ASD	M	Russian	23	4.1	88	-

The groups are close on age, diagnosis and prior PECS exposure, and home language is split three and three in each. Baseline initiation rates were the basis for pairing, so their similarity is by construction, not by chance.

Both groups received three sessions per week, 40 minutes each, throughout the academic year. All sessions were conducted by the same three practitioners to control for instructor variability. Fidelity to the session protocol was not formally monitored - no checklist or independent review of session recordings against the protocol was conducted; this is a limitation, since consistency of implementation across practitioners and across the two conditions cannot be confirmed from these data.

Six children worked on tablets. Three had iPads with Proloquo2Go and three had Android tablets with LetMe Talk, an even split determined by the devices the school possessed rather than by any characteristic of the child. Devices were distributed through the school's existing arrangements, with the result that the three youngest children in the tablet group (I1-I3, aged 6-7) used LetMe Talk and the three oldest (I4-I6, aged 9-10) used Proloquo2Go (Table 1). Application is therefore completely confounded with age. No comparison between the two applications is possible from these data, no claim is made about one being better than the other, and any age-related pattern in the tablet group cannot be separated from the application used. This confounding is treated as a limitation throughout the manuscript.

Neither application speaks Kazakh. Every symbol press produced Russian output, for children from Kazakh-speaking homes as much as for the others. The practitioners compensated for this gap verbally: after the device produced the Russian word, the practitioner immediately said the Kazakh equivalent. The symbol-word pairing a child heard therefore arrived in Russian from the device and in Kazakh only from the practitioner present. The effect of this arrangement on symbol learning cannot be determined from these data; it is considered further in the Discussion and in the Limitations section.

Vocabulary was built for each child separately, from a pre-assessment of what the child needed to communicate, what the day looked like, and what the child liked doing. Roughly 40 to 60 core symbols came first - requests, responses, social exchanges - before anything topic-specific. Grid layout stayed the same across both applications, so a child moving from one device to another found the symbols in the same places.

The six control children used ordinary PECS binders with printed cards and structured gesture instruction, on the same three-sessions-a-week schedule.

Method 1: Structured Observation with Communication Sampling

All sessions in both groups were video-recorded across the year. Communication sampling was done at three points: baseline in September 2024, mid-point in January 2025, endpoint in May 2025. At each point a 15-minute segment was coded from each of the three sessions immediately preceding the scheduled assessment date (baseline, mid-point, endpoint), selected by protocol rather than by content. The three segment scores per child at each time point were averaged before analysis, so each child contributed a single value per time point.

Coding was done by two people who did not run the sessions and were not told what the study expected to find. They worked from a written definition: a spontaneous initiation is a symbol selection, gesture, or vocalisation aimed at a communication partner with no prompt from the practitioner immediately before it. Both trained on footage from outside the study, then coded independently. Blinding was only partial and could not be otherwise, since a tablet is visible in the frame; we note this as a source of residual bias (see Limitations) and make no claim to have controlled it.

Agreement was measured with Cohen's kappa. Twenty per cent of segments at each time point were coded twice, independently, by both raters: $\kappa = 0.84$ at baseline, $\kappa = 0.75$ at mid-point, $\kappa = 0.80$ at endpoint. The mid-point value fell below the 0.80 threshold set before coding began. Most of the disagreement there concerned whether an act followed a prompt: by January several children in the tablet group were reaching for the device during pauses in practitioner talk, and the

two raters read those pauses differently. Agreement returned to threshold at endpoint. Mid-point figures should be read with this in mind. Where the two disagreed, they discussed the segment and settled on a code; segments where they could not agree were dropped. Of 73 segments coded at mid-point, 6 were dropped for unresolved disagreement, split evenly between the two groups (3 intervention, 3 control).

Spontaneous initiation rate was chosen as the primary outcome because it distinguishes between a child who communicates when prompted and a child who communicates on their own initiative - and that distinction is precisely what this study set out to measure.

Method 2: Symbol Acquisition Test

A symbol acquisition test was administered individually to each participant at the same three time points. The test held 30 symbols from that child's own vocabulary set: 20 that had been worked with in sessions and 10 that had not yet been introduced. The practitioner named the referent aloud and the child selected from an array of four. Correct selections were counted and expressed as a percentage.

The 20 familiar items measured retention of taught vocabulary; the 10 novel items measured whether symbol recognition extended to things never taught. Ten items is a coarse instrument. One answer moves the score by ten percentage points, and with four options a child answering at random will score near 25 per cent. Novel-symbol scores therefore indicate the direction and approximate magnitude of change; they do not support finer interpretation, and a score close to 25 per cent cannot be distinguished from guessing.

Referents were named in Russian for every child, in both groups and at all three time points. This keeps the procedure formally identical across participants, but it does not make the testing conditions equivalent: for children from Russian-speaking homes the test word was one they hear daily, for children from Kazakh-speaking homes it was not, and for the latter the test measured symbol recognition through a second language. Because testing language was uniformly Russian, any language-driven variability is not separable from treatment effect within this design, and future work should test in the child's home language or counterbalance language across conditions. Home language was evenly split between conditions, three Kazakh-speaking and three Russian-speaking children in each group, so this does not favour one group over the other, though it adds variability within groups that six children per condition cannot absorb. The language of testing is listed among the study's limitations.

Testing was done by the practitioners who ran the sessions, as no one else at the school knew the children's individual vocabulary sets, and they therefore knew which condition each child was in. Choosing from four fixed options gives a tester little room to steer the result, but the arrangement is not blind, and it is listed among the study's limitations.

Children in the tablet group were tested on screen, children in the card group on printed cards of the same size and layout, so that each child was tested in the format they worked in. This means symbol-test scores are confounded with format familiarity, not a pure measure of symbol knowledge; a design testing all children in both formats, with order counterbalanced, would be needed to separate the two.

Method 3: Vineland Adaptive Behavior Scales - Communication Domain

The Communication Domain of the Vineland Adaptive Behavior Scales, Third Edition, was completed for each child by the class teacher and by one parent, at baseline and at endpoint. Administration and scoring were carried out by a school psychologist licensed to administer standardized assessments, using the Vineland-3 scoring software. The scale covers receptive, expressive and written communication as they appear at home, at school and in the community, and produces age-referenced standard scores. Video observation and symbol testing tell us what a child does inside a session; this instrument was included to determine whether any change extended to the home and community settings.

Teacher and parent ratings are reported separately rather than as a composite. The two informants see the child in different settings, agreement between them on Vineland domains is

typically moderate at best, and a combined score would have hidden the comparison we most wanted: whether change reported at school matched change reported at home.

Two limits apply. Vineland standard scores carry a standard error of measurement of roughly four to five points on this domain, so differences smaller than that fall within the measurement error of the instrument. And neither teachers nor parents were blind to condition. Vineland results are therefore treated as supporting evidence for the direction of change, not as independent confirmation of it.

Data from all three methods were compiled into a single dataset per child. With six per group, tests were computed on exact distributions; normal approximations are unreliable at this size. Between-group comparisons used the exact Mann-Whitney U test, within-group change the exact Wilcoxon signed-rank test, whose smallest attainable two-tailed p at $n = 6$ is 0.031. Effect size is given as the rank-biserial correlation, which corresponds to the rank-based statistic and not to a difference of means.

No correction for multiple comparisons was applied. Since the design estimates parameters and does not test a superiority hypothesis, p -values are reported descriptively. Interpretation rests on the Hodges–Lehmann estimate with its exact confidence interval, and on the spread of individual scores. Individual trajectories are given alongside group means (Tables 2 and 4): with six children a mean can be driven by a single participant. Analyses were run in SPSS version 26.

The programme ran inside the school's regular timetable. Both conditions used methods already in routine use at Special School No. 4, and no child received less instruction than the standard curriculum provides; what the study changed was which of two accepted approaches a given child worked with. Authorisation, informed consent, child assent, data storage and ethical review are described in the Ethics Statement at the end of the article.

Results and discussion

Before presenting numerical results, it is worth describing what the intervention looked like in practice. Figure 1 shows the interface of Proloquo2Go and LetMe Talk as configured for participants in this study. Both applications display symbol grids on a touch screen; each symbol press triggers immediate synthesized speech output. Vocabulary was organized into three layers - core words on the home screen, activity-specific vocabulary one tap deeper, and a personal vocabulary section built around each child's daily routine. LetMe Talk used an open-source grid that teachers could edit directly on the device; Proloquo2Go required configuration through a companion app but offered finer control over symbol size, colour coding, and grid density. The three children on LetMe Talk were the three youngest, and the three on Proloquo2Go the three oldest, because that is how the school's devices were distributed (see Materials and research methods). Application and age are therefore fully confounded, no comparison between the two applications is possible from these data, and this confounding is carried through the interpretation of individual trajectories below.

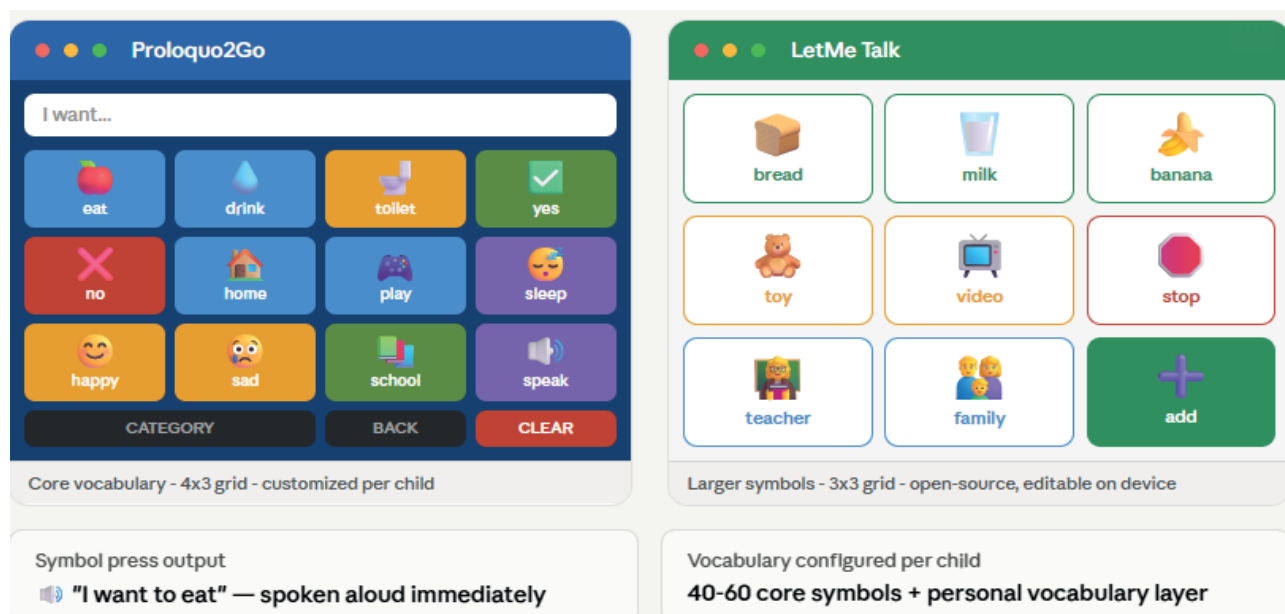


Figure 1. Interface of Proloquo2Go and LetMe Talk configured for study participants, core vocabulary grid layout (screenshot, 2024)

Table 2 gives initiations per 15-minute window at the three time points. Group means are shown with standard deviations and with the six individual scores behind each mean, since at this sample size an average can be driven by one child.

Table 2. Spontaneous communicative initiations per 15-minute observation window: mean, SD, range and individual values

Group	Baseline	Mid-point	Endpoint	Change
Intervention (n=6)	3.2 (SD 0.9), range 2.1-4.5	5.8 (SD 1.7), range 3.4-7.8	7.4 (SD 2.1), range 4.1-9.7	+4.2
Individual values	2.1, 2.4, 2.9, 3.3, 4.0, 4.5	3.4, 4.6, 5.5, 6.2, 7.3, 7.8	4.1, 6.1, 7.3, 8.0, 9.2, 9.7	
Control (n=6)	3.0 (SD 0.8), range 2.0-4.1	4.1 (SD 1.1), range 2.4-5.3	5.1 (SD 1.4), range 3.1-7.0	+2.1
Individual values	2.0, 2.3, 2.8, 3.2, 3.6, 4.1	2.4, 3.3, 4.0, 4.4, 5.2, 5.3	3.1, 4.0, 4.8, 5.4, 6.3, 7.0	

The two groups were equivalent at baseline. By May the tablet group averaged 7.4 initiations per window and the card group 5.1, but the spread in the tablet group had widened from 0.9 to 2.1, with individual endpoint scores running from 4.1 to 9.7.

The exact Mann-Whitney test puts the endpoint difference at $U = 6$, two-tailed $p = 0.065$. It does not reach the conventional threshold, and at six per group a result close to the boundary was to be expected: the smallest attainable two-tailed p here is 0.002 (2/924), and only near-complete separation of the two groups would clear 0.05 comfortably. The Hodges–Lehmann estimate of the endpoint difference is 2.55 initiations per window, with an exact 95.8% confidence interval of -0.2 to 4.9 . The interval includes zero, which is the same statement the p -value makes, and its width is the substantive result: these data are consistent with anything from no difference to a difference larger than the card group's entire year of change. The rank-biserial correlation is 0.67. Within-group change shows the same pattern, with the same caveat. Both groups improved over the year and the tablet group improved more, but at six pairs the Wilcoxon test cannot resolve

differences of this size; its smallest attainable two-tailed p is 0.031, which puts the whole question inside the resolution limit of the instrument, not inside the data.

The more informative result lies in the dispersion of individual scores. At baseline the two groups were similar, as were the children within each group. By May the tablet group had diverged: endpoint rates ran from 4.1 to 9.7 initiations per window. The card group stayed compact, from 3.1 to 7.0. Whatever produced this pattern in the tablet group - the device, the vocabulary redesign, or unmeasured child-level factors - it produced it unevenly. This is the finding that a study of this size can support, and it identifies the question a larger trial should address: for which children tablet-based AAC is effective.

Figure 2 shows the percentage of sessions containing at least one unprompted initiation across both groups and three time points.

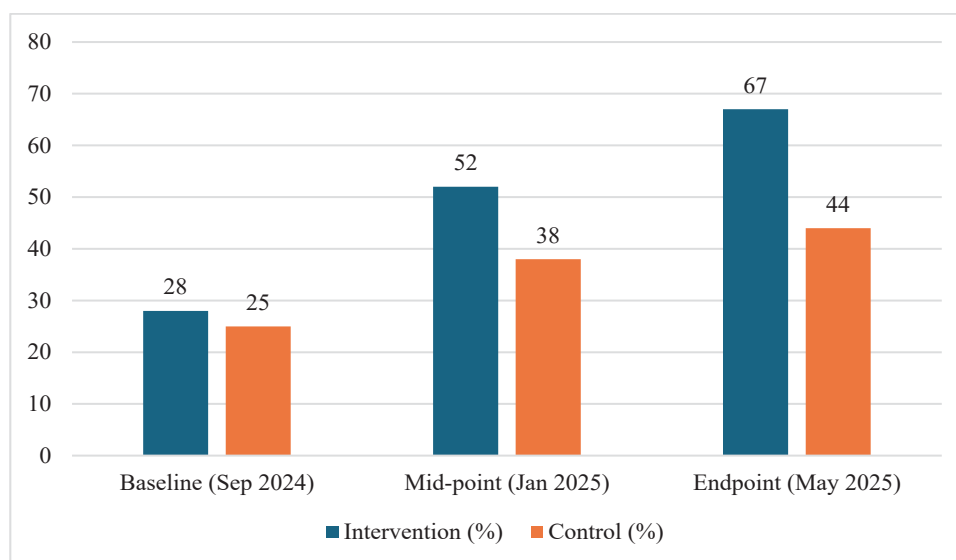


Figure 2. Percentage of sessions containing at least one unprompted communicative initiation by group and time point (%)

The gap between groups widened steadily across the year. By May 2025, two thirds of intervention sessions contained at least one unprompted initiation, compared to fewer than half in the control group. The control group also improved, which confirms that structured instruction with paper cards produces measurable gains - but at a slower rate and with a lower ceiling within the same time frame.

Figure 3 presents correct symbol selection rates for familiar and novel symbols across both groups and three time points.

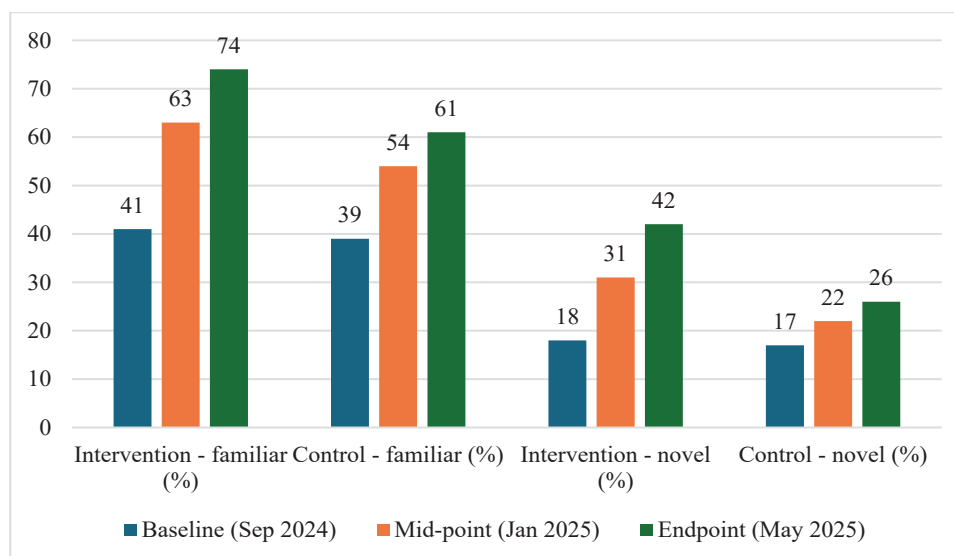


Figure 3. Correct symbol selection rate for familiar and novel symbols by group and time point (%)

On familiar symbols both groups improved, reflecting repeated exposure to trained symbol-referent pairs. The intervention group reached 74% correct by May 2025 against 61% in the control group - a gap of 13 percentage points. On novel symbols the gap was wider: 42% versus 26% at endpoint. Novel symbol performance is the more informative result because it shows whether children acquired a generalizable symbol-reading strategy or simply memorized specific trained pairs. The control group's 26 per cent on novel items is close to the chance level of 25% with four response options, though we did not run a formal test against this baseline; this should be read as suggestive of limited transfer rather than confirmed absence of it. The tablet group's 42 per cent is above chance, though with ten items that margin amounts to under two symbols per child. Whether immediate auditory feedback is what produced it cannot be established from a design that varies feedback and modality together.

The third measure addressed whether any of these changes extended beyond the session. Table 3 gives the change in Vineland-3 Communication Domain scores across the year, with teachers and parents reported apart.

Table 3. Vineland-3 Communication Domain, change in standard scores from baseline to endpoint, by informant

	Intervention (n=6)	Control (n=6)
Teacher rating	+5.2	+2.9
Parent rating	+7.8	+3.7

Teachers put the tablet group ahead by 2.3 points, parents by 4.1. Both gaps are close to the standard error of the instrument, which on this domain runs to four or five points, so neither figure carries a precise reading on its own.

The pattern across informants is more informative than the absolute differences. In the tablet group parents rated the change 2.6 points higher than teachers did; in the card group the two informants stayed within a point of each other. Teacher and parent ratings were not treated as interchangeable, and a gap that opened in one group and not the other is unlikely to be a scoring artefact. The teacher-parent gap in the tablet group is notable but cannot be attributed to transfer of communication to the home setting, given the absence of blinding and the lack of a formal parent-teacher agreement analysis. Informal comments from parents are noted here as anecdotal context, not as corroborating evidence, since no structured interview protocol was used to collect

them. The device itself cannot account for the pattern: tablets remained at school, so whatever parents observed at home was the child communicating without the device that had taught the symbols. This result was stronger than anticipated but also less robust, since parent ratings were not blind to condition.

Four implementation issues arose. Three were resolved; the fourth was accommodated.

Three of the six children initially refused to interact with the tablet. Two had sensory sensitivities, and the glass surface appeared to be the source of aversion; the third refused anything new in the first weeks of term, whether a tablet or not. The device was placed on the table and left there, with nothing asked of the child. The next step was a single press, with no communicative demand attached to it. Vocabulary work started only after that. Refusal ended at session 4 for the first child, session 6 for the second, and session 9 for the third, and from week four all six were working with devices. This delay cost between 4 and 9 sessions of instructional time for each of the three children and should be anticipated by anyone planning a term around this approach.

The vocabulary sets built over the summer did not match what the children wanted to say in September. Several searched the grid repeatedly for symbols that were not on it; two pushed the device away. A weekly review was added in the first month, and grid content was changed according to what children were visibly trying to communicate and what parents reported from home. Across the first six weeks 47 symbols were added and 22 removed, roughly 8 added and 4 removed per child. The problem had ceased by week six. A pre-assessment made in August does not fully anticipate the needs that emerge in a September classroom, and time for revision has to be budgeted from the outset, because it will not be found later.

Setting up Proloquo2Go for three children and keeping vocabulary current took about three extra hours a week through September and October. This was reduced by exporting configurations between devices and building a folder of grid templates for common categories: requests, feelings, classroom routines, food. From January it required under an hour a week. This figure is a retrospective estimate based on practitioner recall and weekly review notes rather than contemporaneous time logs, and does not separate research-related configuration time from routine lesson preparation that would have occurred regardless of the study. For a school considering adoption, the figure to plan against is the first eight weeks, when the load is heaviest and the programme has produced the least evidence of benefit. The steady state is comparatively undemanding.

The fourth issue was not resolved but accommodated. Tablets stayed at school. The school could not insure devices taken home, and no family was asked to take responsibility for one. This constrains what the home findings can mean: parents reported communicative gains, but those gains occurred without the device present, through gesture, vocalisation and pointing that the child had developed at school. Whether the effect would be larger if the device accompanied the child home is a question this design could not address.

Across the year the tablet group moved further than the card group on all three measures, and none of those differences reaches statistical significance. At six children per condition, power to detect effects of this size was limited. The card group did not stand still. Structured work with printed symbols produced measurable gains in initiation and in retention of taught vocabulary. These results should not be read as evidence that paper cards are ineffective. The two conditions differed in one respect only, on untaught symbols, where the tablet group went above chance and the card group did not - consistent with what is already known about card systems training particular requests rather than a code a child can extend.

A more notable finding is the dispersion of outcomes within the tablet group. By May the six children differed from each other more than the two group averages differed. Table 4 gives the individual trajectory of each child alongside the characteristics that might plausibly explain it. Individual gains in the tablet group ranged from +2.0 to +5.2 initiations per window. The largest was more than two and a half times the smallest, and the smallest fell inside the range observed in the card group, where gains ran from +1.1 to +2.9. The two largest gains belonged to the two oldest

children in the tablet group, both of whom used Proloquo2Go, and the smallest to the youngest child, who used LetMe Talk; because age and application are completely confounded in this sample, this pattern cannot be attributed to either factor. Home language, diagnosis and attendance show no consistent relation to the size of gain in either group, though with six children per condition the absence of a visible pattern is not evidence that none exists. A trial designed to test the average effect of tablets would collapse both of these patterns into a single figure. Which children respond, and why, is the question these data raise and cannot answer.

Table 4. Individual participant trajectories: characteristics, spontaneous initiations per 15-minute window at baseline (Base), mid-point (Mid) and endpoint (End), and correct symbol selection at endpoint (Diag. = diagnosis; Lang. = home language; App. = application; Att. = attendance; Fam. = familiar symbols; Nov. = novel symbols)

Child	Age	Diag.	Lang.	App.	Att. (%)	Base	Mid	End	Gain	Fam. (%)	Nov. (%)
I1	6	ASD	Kazakh	LetMe Talk	91	2.1	3.4	4.1	+2.0	60	30
I2	7	DLI	Russian	LetMe Talk	88	2.4	4.6	6.1	+3.7	70	30
I3	7	ASD	Kazakh	LetMe Talk	94	2.9	5.5	7.3	+4.4	70	40
I4	9	ASD	Russian	Proloquo2Go	86	3.3	6.2	8.0	+4.7	75	40
I5	9	DLI	Kazakh	Proloquo2Go	92	4.0	7.3	9.2	+5.2	80	50
I6	10	ASD	Russian	Proloquo2Go	90	4.5	7.8	9.7	+5.2	90	60
C1	6	ASD	Kazakh	-	93	2.0	2.4	3.1	+1.1	50	20
C2	7	ASD	Russian	-	87	2.3	3.3	4.0	+1.7	55	20
C3	8	DLI	Kazakh	-	95	2.8	4.0	4.8	+2.0	60	20
C4	8	ASD	Russian	-	89	3.2	4.4	5.4	+2.2	60	30
C5	9	DLI	Kazakh	-	91	3.6	5.2	6.3	+2.7	65	30
C6	10	ASD	Russian	-	88	4.1	5.3	7.0	+2.9	75	40

Limitations

This study is a non-randomised, single-centre pilot evaluation study with a small sample. It carries an elevated risk of selection bias and a limited capacity to establish causal relationships; its results are estimates of the direction, magnitude and variability of change, not tests of a superiority hypothesis. The limitations noted at the relevant points in the manuscript are collected here.

(1) Sample size. Six children per condition gives limited power to detect effects of the size observed. The smallest attainable two-tailed p is 0.002 for the exact Mann-Whitney test and 0.031 for the exact Wilcoxon test, confidence intervals are wide, and a group mean can be driven by a single child. Novel-symbol scores based on ten items move in steps of ten percentage points and cannot be distinguished from chance when close to 25 per cent.

(2) Absence of randomisation. Children were matched in pairs on baseline initiation rate, but assignment within each pair was made according to the compatibility of the child's timetable with device and practitioner availability. Matching balances baseline communication level only; home language, parental involvement and prior therapy were not controlled, and schedule-based assignment is an additional source of selection bias.

(3) Single-centre design. All participants attended one special school in one city and were taught by the same three practitioners. The findings describe this setting and cannot be generalised to other schools, regions or service models without replication.

(4) Absence of formal fidelity monitoring. No checklist or independent review of session recordings against the protocol was conducted, so consistency of implementation across practitioners and across conditions cannot be confirmed. The estimate of configuration time is retrospective, based on practitioner recall and weekly review notes rather than contemporaneous logs, and does not separate research-related configuration from routine lesson preparation.

(5) Incomplete blinding. Video coders were not told what the study expected to find, but a tablet is visible in the frame, so blinding was partial. The symbol acquisition test was administered by the practitioners who ran the sessions and therefore knew each child's condition. Teacher and parent Vineland-3 ratings were not blind to condition. Inter-rater agreement at mid-point ($\kappa = 0.75$) fell below the pre-set threshold of 0.80, and six segments were dropped for unresolved disagreement.

(6) Language of testing. Neither application produces Kazakh speech, so device output was Russian for every child, with the practitioner voicing the Kazakh equivalent for children from Kazakh-speaking homes. Referents in the symbol acquisition test were also named in Russian for all children. This standardised the procedure but did not make it equivalent: Kazakh-speaking children were tested through a second language, and any language-driven variability cannot be separated from treatment effect. Home language was balanced across conditions, so this does not favour one group, but it increases variability within groups.

(7) Confounding of age and application. The three youngest children in the tablet group used LetMe Talk and the three oldest used Proloquo2Go, so application is completely confounded with age. No comparison between the applications is possible, and the tendency for older children in the tablet group to show larger gains cannot be attributed to age or to application.

(8) Device access and outcome measurement. Tablets remained at school, so home-based change reported by parents occurred without the device present. Symbol-test scores are confounded with format familiarity, since each child was tested in the format they worked in. Finally, the programme was designed, delivered and largely measured by the same team, and no correction for multiple comparisons was applied.

Conclusion

The question with which this study began was whether tablet-based AAC outperforms picture cards. A non-randomised pilot study of twelve children cannot answer it. The tablet group showed larger change on all three measures, none of the differences reaches statistical significance, and at six per condition power to detect effects of this size was limited.

The finding of greater practical value lies instead in the dispersion of individual outcomes. By May the six tablet children were further from one another than the two group averages were from each other. Two changed substantially. The smallest individual gain in the tablet group was no larger than the mean gain in the card group. The group mean does not describe any individual child, and a larger trial comparing group means would produce a more precise estimate of the same uninformative quantity. For a practitioner in Karaganda, the relevant question is not whether tablets are effective on average but for which children in a given classroom they are effective. These data raise that question through the spread of individual results (Table 4) but cannot answer it: identifying which child-level factors (age, diagnosis, home language, attendance, prior intervention, or application assignment; age and application are fully confounded in this sample) drive that variation requires replicated single-case designs or a larger stratified trial.

Some conclusions can be stated without qualification. The applications run in an ordinary Kazakhstani special school with no technical support behind them. Refusal, vocabulary mismatch and configuration load all appeared and all were dealt with inside six weeks. The cost was about sixty practitioner hours across the year, nearly all of it in September and October. And neither application speaks Kazakh, so every session ran with the device talking Russian and the practitioner supplying the child's own language out loud - an arrangement that was not designed but imposed by the available technology, and one that practitioners in this region will recognise.

These conclusions are bounded by the limitations set out above: one school, one city, twelve children, non-randomised allocation, partial blinding, uniformly Russian-language testing, complete confounding of age with application, and measurement largely by the team that delivered the programme. Coder agreement fell below the pre-set threshold in January. The tablets never left the building, so what parents observed at home was the child, not the device. Larger multi-site

work with randomised allocation, powered to identify which children respond rather than whether any effect exists on average, is the next step.

Ethics Statement

The programme ran inside the school's regular timetable. Both conditions used methods already in routine use at Special School No. 4, and no child received less instruction than the standard curriculum provides; what the study changed was which of two accepted approaches a given child worked with. The work was authorised by the school administration. Parents were informed about the programme and about the video recording of sessions before it began and consented to both. Children were asked for assent in an accessible format at each session, and a child who signalled refusal was not recorded that day. Recordings were stored on a school computer and were seen only by the practitioners and the two coders.

The study underwent retrospective ethical review by the Ethics Committee of KSU “Special Boarding School No. 4” of the Department of Education, Karaganda region (Protocol No. 1, 18 August 2026; Order No. 32, 14 August 2026). The Committee confirmed that the procedures as implemented met the ethical requirements for research involving children, while noting that prospective ethical review had not been obtained before data collection and that this constitutes a procedural violation; the authors disclose it here accordingly.

References

- [1] Sterrett, K., Holbrook, A., Landa, R., Kaiser, A., & Kasari, C. (2023). The effect of responsiveness to speech-generating device input on spoken language in children with autism spectrum disorder who are minimally verbal. *Augmentative and Alternative Communication*, 39(1), 23–32. <https://doi.org/10.1080/07434618.2022.2120070>
- [2] Tamanaha, A. C., & Olivatti, D. O. F. (2023). Picture Exchange Communication System (PECS) implementation program for children with autism spectrum disorder. *CoDAS*, 35(4), e20210305. <https://doi.org/10.1590/2317-1782/20232021305en>
- [3] Koudys, J., Perry, A., Magnacca, C., & McFee, K. (2023). Predictors of Picture Exchange Communication System (PECS) outcomes. *Autism and Developmental Language Impairments*, 8. <https://doi.org/10.1177/23969415231221516>
- [4] May, R., Salman, H., O'Neill, S. J., Denne, L., Grindle, C., & Cross, R. (2025). Exploring the use of the Picture Exchange Communication System (PECS) in special education settings. *Journal of Autism and Developmental Disorders*, 55(2), 652–666. <https://doi.org/10.1007/s10803-023-06194-1>
- [5] Denne, L., Grindle, C., Apanasionok, M., & Hastings, R. P. (2024). Facilitators and barriers to the implementation of the Picture Exchange Communication System (PECS): A systematic review. *Frontiers in Education*, 9, 1206410. <https://doi.org/10.3389/feduc.2024.1206410>
- [6] Alfuraih, A., Almalki, A., & AlNemary, F. (2024). Effectiveness of picture exchange communication system in developing requesting skills for children with multiple disabilities. *Frontiers in Psychology*, 15, 1434478. <https://doi.org/10.3389/fpsyg.2024.1434478>
- [7] Yuan, W., Dunn, M., Kennedy, J., Hsiao, Y. J., & Poppen, M. (2024). Impact of speech generating device interventions on vocalizations of individuals with autism and severe communication impairment: A systematic review. *Education and Training in Autism and Developmental Disabilities*, 59(3), 233–256. <https://doi.org/10.1177/215416472405900302>
- [8] Alzahrani, A. N. (2024). Educational mobile apps enhancing communication skills in children with autism in Arab countries. *International Journal of Information and Communication Technology Education*, 20(1), 1–13. <https://doi.org/10.4018/IJICTE.359439>
- [9] Alzyoudi, M., & Alghazali, F. (2025). Enhancing communication skills in students with autism. *Educational Process: International Journal*, 15, e2025174. <https://doi.org/10.22521/edupij.2025.15.174>

- [10] Flores, C., & Dada, S. (2025). The effect of AAC training programs on professionals' knowledge, skills and self-efficacy in AAC: A scoping review. *Augmentative and Alternative Communication*, 41(1), 65–77. <https://doi.org/10.1080/07434618.2024.2381462>
- [11] Gomez-Asencio, D., Gago-Valiente, F. J., Arana-Alvarez, C., & Moreno-Sanchez, E. (2025). Pedagogical interventions in augmentative and alternative communication (AAC): Therapeutic benefits in healthcare and quality of life. *Frontiers in Communication*, 10, 1584169. <https://doi.org/10.3389/fcomm.2025.1584169>
- [12] Ruiz-Martinez, A. (2025). Future technologies in alternative and augmented communication: A scoping review of innovations. *Frontiers in Communication*, 10, 1607531. <https://doi.org/10.3389/fcomm.2025.1607531>
- [13] Zorzi, S., Lubkina, V., Jēkabsons, I., & Berteotti, L. (2025). Digital technology interventions for communication skills in persons with neurodevelopmental disorders: A scoping review. *Research in Developmental Disabilities*, 164, 105080. <https://doi.org/10.1016/j.ridd.2025.105080>
- [14] Jándula Justicia, A., Nistal Anta, D., & Peinado Rodríguez, M. (2026). Digital applications in the communicative development of people with ASD: A systematic review. *Education Sciences*, 16(2), 210. <https://doi.org/10.3390/educsci16020210>
- [15] Holyfield, C., MacNeil, S., Caldwell, N., O'Neill Zimmerman, T., & Lorah, E. (2024). Leveraging communication partner speech to automate augmented input for children on the autism spectrum who are minimally verbal: Prototype development and preliminary efficacy investigation. *American Journal of Speech-Language Pathology*, 33(3), 1174–1192. https://doi.org/10.1044/2023_AJSLP-23-00224
- [16] Suhr, M., Bean, A., Rolniak, J., Paden Cargill, L., & Lyle, S. (2024). The influence of classroom context on AAC device use for nonspeaking school-aged autistic children. *International Journal of Speech-Language Pathology*, 26(3), 434–444. <https://doi.org/10.1080/17549507.2023.2220992>
- [17] Joseph, A. P., & Babu, A. (2025). Redefining communication in mental healthcare: Generative AI for neurodivergent equity and non-verbal autistic inclusion. *Frontiers in Psychiatry*, 16, 1611101. <https://doi.org/10.3389/fpsy.2025.1611101>
- [18] Beirat, M. A., Algolaylat, A., Al Njadat, H., AlAbdallat, B., & Al-Makhzoomy, A. K. (2025). Utilization of artificial intelligence and assistive technology in autism: Diagnosis, treatment, and education applications - A systematic literature review. *Educational Process: International Journal*, 17, e2025350. <https://doi.org/10.22521/edupij.2025.17.350>
- [19] Soto, G., & Clarke, M. T. (2025). Experiences of school-based professionals delivering AAC to bilingual students in the US: Challenges and strategies. *Augmentative and Alternative Communication*, 1–13. Advance online publication. <https://doi.org/10.1080/07434618.2025.2531012>
- [20] Therrien, M. C. S., Whalon, K., Nunes, D., Marti, P., & Sellers, A. (2025). A systematic review of AAC interventions using speech generating devices for autistic preschoolers. *Augmentative and Alternative Communication*, 41(2), 114–128. <https://doi.org/10.1080/07434618.2025.2479768>
- [21] Jiang, Y., & Wang, T. (2026). Patterns of AAC use and communicative functions in minimally verbal autistic children following introduction of AAC tools and caregiver training: A corpus-based analysis. *Research in Developmental Disabilities*, 168, 105188. <https://doi.org/10.1016/j.ridd.2025.105188>