



ӘЛЕУМЕТТІК ЖӘНЕ ГУМАНИТАРЛЫҚ ҒЫЛЫМДАР
СОЦИАЛЬНО-ГУМАНИТАРНЫЕ НАУКИ
SOCIAL SCIENCES AND HUMANITIES

CENTRAL ASIAN SCIENTIFIC JOURNAL
ISSN 3135-3061
2026. No. 4 (32), Vol. 2. pp. 101–106
DOI: <https://doi.org/10.5281/zenodo.23776973>
Journal website: <https://cajournal.kz>

UDC 81

LANGUAGE OF SCHOOLING AND PERCEPTIONS OF ENGLISH
ACCENTED SPEECH IN KAZAKHSTAN

Saltanat Mussina, Torgyn Yeralinova

Ye. A. Buketov, Kazakhstan

Saltanat Mussina — Department of Foreign Philology, Karaganda National Research University named after academician, Ye. A. Buketov, Kazakhstan.

Torgyn Yeralinova — Department of Foreign Philology, Karaganda National Research University named after academician, Ye. A. Buketov, Kazakhstan.

© Saltanat Mussina, Torgyn Yeralinova, 2026

Abstract: This exploratory study examined whether language of secondary schooling was associated with university students' evaluations of English-accented speech in Kazakhstan. Twenty undergraduates, comprising 10 students with predominantly Kazakh-medium schooling and 10 with predominantly Russian-medium schooling, evaluated eight recordings by different speakers. Data were collected in September 2026 during a university phonetics course. Each recording lasted 30–60 seconds and was played once through classroom loudspeakers in a fixed order. Participants rated comprehensibility, perceived accent strength, and perceived speaker competence on separate nine-point scales. Complete responses yielded 160 listener–recording observations per outcome. Ratings were averaged within listeners before two-sided Welch tests, with Holm adjustment across three outcomes. Kazakh-medium listeners gave numerically higher ratings for comprehensibility (6.55 vs. 6.25), accent strength (5.78 vs. 5.58), and competence (6.85 vs. 6.38). All confidence intervals for mean differences included zero, and no Welch comparison was statistically significant after adjustment. Mann–Whitney sensitivity analyses likewise yielded no significant adjusted comparisons. The findings provide preliminary listener-level estimates but do not establish reliable group differences or a causal effect of schooling language. Interpretation is limited by the small convenience sample, fixed presentation order, unadjusted proficiency differences, and the absence of an objective intelligibility measure.

Keywords: accented English, comprehensibility, perceived accentedness, listener judgments, language of schooling, Kazakhstan

1. Introduction

Perception of additional-language English cannot be explained solely by the speaker's pronunciation. Research distinguishes accentedness (the perceived strength of a non-native accent), comprehensibility (the listener's perceived ease of understanding), and intelligibility (what the listener actually recognizes). Munro and Derwing (1995) found that heavily accented speech can nevertheless remain intelligible, and their later synthesis emphasizes that the three dimensions are related but not interchangeable (Munro & Derwing, 2022). Listener characteristics are therefore a legitimate object of investigation rather than merely background variation.

Expectations can also shape how listeners evaluate speech. Kang and Rubin (2009) examined reverse linguistic stereotyping and showed why information about a speaker's presumed group membership may influence speech evaluations. Research designs that separate auditory evidence from social expectations are consequently important when studying judgments of accented



speech. A rating of competence is especially sensitive to this issue: it represents a listener's perception and should not be treated as an objective measure of a speaker's language ability.

Kazakhstan provides a relevant multilingual educational context in which Kazakh, Russian, and English may occupy different positions in a person's language-learning history. Research on language learning in Kazakhstan has examined L1/L2 transfer in L3 English acquisition (Knol et al., 2026), while the present study addresses a related but distinct question: whether listeners with different school-language backgrounds evaluate the same English speech samples differently. The study does not assume in advance that either group will be more tolerant of an accent, and it is not intended to represent all students in Kazakhstan.

2. Research Questions

RQ1. Do students with Kazakh-medium and Russian-medium secondary-school backgrounds differ in their ratings of the comprehensibility of the same English speech recordings?

RQ2. Do the two groups differ in perceived accent strength and perceived speaker competence?

RQ3. Do the conclusions from mean-based comparisons remain consistent with rank-based sensitivity analyses?

3. Method

3.1. Design and participants

This exploratory study examined the association between language of secondary schooling and university students' evaluations of English-accented speech. Twenty undergraduate students enrolled in English Philology and Translation Studies at Karaganda Buketov University participated in the study. Recruitment and data collection took place in September 2026 during the course Theory of the Studied Language (Phonetics).

Participants received an oral invitation during class, with an explanation of the study's purpose and the voluntary nature of participation. Convenience sampling with quotas based on language of secondary schooling was used to recruit 10 students with predominantly Kazakh-medium backgrounds and 10 with predominantly Russian-medium backgrounds. Eligibility criteria were enrolment in one of the two-degree programs, consent to participate, and provision of information about the predominant language of secondary schooling.

Each schooling group included five English Philology students and five Translation Studies students. The mean age was 21.8 years in each group. Participants had studied English for a mean of 8.8 years ($SD = 0.42$) in the Kazakh-medium group and 8.6 years ($SD = 0.84$) in the Russian-medium group.

Schooling background was classified according to the predominant language of secondary education reported by each participant. These categories were not treated as indicators of ethnicity, first language, or monolingual status. Participants' home languages and language-use patterns varied within both groups. Self-rated English levels were Intermediate ($n = 4$), Upper-intermediate ($n = 5$), and Advanced ($n = 1$) in the Kazakh-medium group, and Intermediate ($n = 4$), Upper-intermediate ($n = 2$), and Advanced ($n = 4$) in the Russian-medium group.

3.2. Speech materials and measures

Eight English speech recordings, each lasting between 30 and 60 seconds, were selected from the author's own recording archive and used with the speakers' permission. The recordings featured eight different speakers and were selected based on audio quality, duration, and suitability for the phonetics listening task. They were assigned anonymous codes C01–C08. No comparisons between accent varieties were undertaken.

Participants evaluated each recording on three separate nine-point scales. Comprehensibility ranged from 1 (very difficult to understand) to 9 (very easy to understand). Perceived accent strength ranged from 1 (not noticeable) to 9 (very strong). Perceived speaker competence ranged from 1 (very low) to 9 (very high).

Comprehensibility represented the listener's subjective ease of understanding the speech. Accent strength referred to the perceived prominence of the accent, whereas perceived competence represented the listener's impression of the speaker's language ability. These judgments were not



treated as objective measures of proficiency. No usable transcription responses or scores were available; therefore, objective intelligibility was not assessed.

3.3. Procedure and data screening

Participants listened together in a teaching classroom during the course Theory of the Studied Language (Phonetics). The instructor played all eight recordings through classroom loudspeakers in the same fixed order for all participants. Each recording was played once. Instructions were provided in English. Participants recorded their evaluations on paper rating forms containing recording codes C01–C08 and the three nine-point scales. Responses were subsequently entered into an Excel workbook for analysis.

The dataset contained 160 participant–recording observations, corresponding to 20 listeners evaluating eight recordings. No duplicate participant–recording pairs were identified. All three primary rating variables were complete and within the permitted range of 1–9. Each participant contributed eight ratings per outcome, and all 20 participants were retained.

The optional transcription and exclusion-reason fields contained the label “Actual” rather than usable transcription responses, numeric transcription scores, or substantive exclusion reasons. These entries were treated as metadata and were not used as outcomes or as grounds for exclusion. The Clip_Order field contained values inconsistent with the fixed presentation procedure and was not used in the analyses.

3.4. Statistical analysis

The eight ratings for each outcome were averaged within each listener. The independent unit for between-group inference was therefore the listener ($n = 10$ per group), not the individual listener–recording observation. Group means and standard deviations were calculated from listener averages. Two-sided Welch independent-samples t tests compared groups, with Satterthwaite degrees of freedom and unadjusted 95% confidence intervals for mean differences. Differences were calculated as Kazakh-medium minus Russian-medium. Holm correction was applied across the three Welch tests, with a familywise significance threshold of .05.

Hedges’ g was calculated from the mean difference divided by the pooled listener-level standard deviation, using the small-sample correction $1 - 3/(4df - 1)$, where $df = 18$. Positive values indicate higher ratings in the Kazakh-medium group. Confidence intervals presented in Table 1 refer to raw mean differences, not to standardized effect sizes.

Two-sided asymptotic Mann–Whitney U tests, with tie correction and continuity correction, were used as a sensitivity analysis. Holm correction was applied separately across these three tests. Rank-based tests address relative ordering of observations rather than the same mean-difference estimand as Welch tests. All analyses were exploratory. No covariate adjustment was undertaken with this small sample, and no causal inference was intended. Inference is conditional on the eight selected recordings; the analysis does not estimate variation across a wider population of speakers.

Calculations were checked using Python with pandas 2.2.3, NumPy 2.3.5, and SciPy 1.17.0. Holm adjustment and the Hedges’ g correction were implemented directly.

3.5. Ethical considerations

Participation was voluntary, and all participants provided informed consent before taking part. The purpose of the study and the voluntary nature of participation were explained during recruitment. The speech recordings were drawn from the author’s own archive and used with the speakers’ permission. Participant codes were used instead of names in the analysis dataset. No claim of institutional ethics approval or formal exemption is made, as a documented institutional determination was not available for reporting.

4. Results

4.1. Data completeness

All 20 participants contributed complete ratings for all eight recordings and all three primary outcomes. Thus, 160 observations were available per outcome. No primary rating values were missing, and no participant–recording duplicates were detected. The mean-based comparisons below use 20 participant-level averages for each outcome, with 10 participants in each schooling group.



4.2. Descriptive and inferential comparisons

The Kazakh-medium group had a higher mean rating than the Russian-medium group on all three outcomes, although the size and uncertainty of the differences varied. The largest raw mean difference was observed for perceived competence (+0.48 scale points), followed by comprehensibility (+0.30) and perceived accent strength (+0.20). Table 1 summarizes the group-level participant averages and Welch tests.

Table 1. Participant-level rating averages and comparisons by language of schooling

Outcome	Kazakh-medium M (SD)	Russian-medium M (SD)	Mean difference (95% CI)	Welch test; p; Holm p; Hedges g
Comprehensibility	6.55 (0.48)	6.25 (0.40)	0.30 [-0.12, 0.72]	$t(17.54) = 1.52$; $p = .146$; $p_{\text{Holm}} = .293$; $g = 0.65$
Perceived accent strength	5.78 (0.76)	5.58 (0.43)	0.20 [-0.39, 0.79]	$t(14.21) = 0.72$; $p = .481$; $p_{\text{Holm}} = .481$; $g = 0.31$
Perceived competence	6.85 (0.46)	6.38 (0.66)	0.48 [-0.06, 1.01]	$t(16.12) = 1.88$; $p = .079$; $p_{\text{Holm}} = .237$; $g = 0.80$

Note. $n = 10$ per group. Higher scores indicate greater ease of understanding, stronger perceived accent, or higher perceived competence, respectively. Mean differences are Kazakh-medium minus Russian-medium. Confidence intervals are 95% Welch intervals. Holm-adjusted p values control familywise error across the three primary tests. Positive Hedges' g values indicate higher mean ratings in the Kazakh-medium group.

Welch's tests did not show statistically significant group differences at the .05 level after correction for the three primary outcomes. For comprehensibility, the estimated difference was 0.30 points, 95% CI [-0.12, 0.72], $t(17.54) = 1.52$, $p = .146$, Holm-adjusted $p = .293$. For perceived accent strength, the estimated difference was 0.20 points, 95% CI [-0.39, 0.79], $t(14.21) = 0.72$, $p = .481$, Holm-adjusted $p = .481$. For perceived competence, the estimated difference was 0.48 points, 95% CI [-0.06, 1.01], $t(16.12) = 1.88$, $p = .079$, Holm-adjusted $p = .237$. The confidence intervals include zero for every outcome, so the mean-based results are compatible with both negligible differences and potentially meaningful positive differences for the Kazakh-medium group.

4.3. Sensitivity analysis and variation across recordings

The Mann–Whitney sensitivity analysis yielded $U = 79.0$, $p = .029$, Holm-adjusted $p = .088$ for comprehensibility; $U = 62.5$, $p = .360$, Holm-adjusted $p = .360$ for perceived accent strength; and $U = 72.5$, $p = .095$, Holm-adjusted $p = .189$ for perceived competence. No comparison remained below .05 after adjustment. The smaller unadjusted rank-based p value for comprehensibility should be interpreted in light of the different estimand and the small sample, rather than as confirmation of a mean difference.

Descriptively, the Kazakh-medium group's mean comprehensibility rating was higher for seven recordings and equal for one. Mean perceived competence ratings were higher for the Kazakh-medium group for all eight recordings. In contrast, perceived accent-strength differences varied by recording, with the direction changing for some clips and no difference for one clip. These recording-level patterns are descriptive only; with eight stimuli and 20 listeners, they should not be interpreted as evidence of a general speaker- or accent-level effect.

5. Discussion

This exploratory analysis asked whether the language of secondary schooling was associated with how undergraduate listeners evaluated English-accented speech. Across the three ratings, the Kazakh-medium group gave descriptively higher scores than the Russian-medium group. However, the primary participant-level Welch comparisons were not statistically significant after correcting for the three outcomes, and each 95% confidence interval included zero. The data therefore do not



provide robust evidence that the two schooling groups differed in their mean evaluations in this sample.

The perceived-competence difference was largest in scale-point and standardized terms (Hedges' $g = 0.80$), but the confidence interval for the raw mean difference included zero and the adjusted p value was .237. In a sample of only 10 participants per group, effect sizes can be unstable and uncertainty is substantial. This pattern should not be translated into the claim that one group is more accurate or fair in evaluating speakers. Perceived competence is a social judgment, not an objective test of language proficiency.

The comprehensibility outcome merits particular caution. The mean-based test suggested a positive difference of 0.30 points in favor of the Kazakh-medium group, while the rank-based sensitivity test produced a smaller unadjusted p value. The latter result was no longer below .05 after Holm correction. This dependence on the analytical method is a reason to avoid a binary conclusion and to prioritize replication with a larger sample, prespecified analyses, and transparent documentation of the ratings.

The distinction between perceived comprehensibility and intelligibility remains important. Because no usable transcription responses were available, the present study cannot determine whether the groups differed in actual word recognition. Subjective ease-of-understanding ratings do not directly measure this ability. This distinction is consistent with the conceptual separation of accentedness, comprehensibility, and intelligibility in speech-perception research (Munro & Derwing, 1995, 2022). The present comparison also cannot establish causation: English proficiency, everyday exposure, accent familiarity, and other aspects of multilingual experience may be associated with both schooling history and listener judgments.

The study contributes preliminary evidence about listener evaluations in a local multilingual university setting. Its estimates concern the selected recordings and the sampled students, rather than all Kazakh- or Russian-medium learners. Future sample-size planning should allow for substantial uncertainty in these pilot effect estimates rather than relying directly on a single observed standardized difference.

6. Limitations and Future Research

Several limitations constrain interpretation. First, the sample comprised only 20 students from one university, recruited through convenience sampling with schooling-language quotas, which limits precision and generalizability. Second, schooling medium is a broad proxy for a complex multilingual biography. The groups differed descriptively in self-rated English proficiency, and proficiency and exposure were not controlled statistically. Third, inference is conditional on the eight selected recordings and cannot establish a general effect across speakers or accent varieties. Fourth, no usable transcription responses or scores were available, preventing objective intelligibility analysis. Fifth, all participants heard the recordings in the same fixed order, so recording-specific differences cannot be separated from possible order effects, such as fatigue or increasing familiarity with the task. Listening through classroom loudspeakers may also have introduced differences in audibility depending on seating position.

Future research should preregister the primary outcome and analysis plan, recruit a larger sample across institutions, and model the repeated-measures structure directly, ideally with mixed-effects models including listener and speaker/recording effects were supported by the design. Recording-level variation should be sampled systematically rather than relying on a small set of stimuli. Adding a transcription task would allow perceived comprehensibility to be compared with an objective intelligibility measure. Researchers should also document ethics approval or exemption, informed consent, stimulus permissions, and data-handling procedures.

7. Conclusion

In this 20-listener exploratory study, ratings by students with Kazakh-medium school backgrounds were descriptively higher than those by students with Russian-medium backgrounds for comprehensibility, perceived accent strength, and perceived competence. The group differences were not statistically robust after correction for multiple comparisons, and the comprehensibility result varied by analytical approach. The findings are therefore preliminary and do not establish a



reliable effect of schooling language. A larger, preregistered study combining subjective ratings with an objective transcription task is needed to determine whether schooling background is meaningfully associated with the perception of English-accented speech in Kazakhstan.

Data Availability Statement

The dataset contains coded participant characteristics and listener ratings. It is not publicly available because of participant privacy considerations. Requests for access may be directed to the corresponding author and will be considered subject to participant consent and applicable institutional requirements.

References:

1. Kang, O., & Rubin, D. L. (2009). Reverse linguistic stereotyping: Measuring the effect of listener expectations on speech evaluation. *Journal of Language and Social Psychology*, 28(4), 441–456. <https://doi.org/10.1177/0261927X09341950>
2. Knol, M., Assanova, D., Smagulova, G., Uteubayeva, E., & Akbaeva, G. (2026). Reducing L1 and L2 transfer in L3 English acquisition: An intervention study in Kazakhstani trilingual education. *Frontiers in Education*, 11, Article 1850753. <https://doi.org/10.3389/feduc.2026.1850753>
3. Munro, M. J., & Derwing, T. M. (1995). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning*, 45(1), 73–97. <https://doi.org/10.1111/j.1467-1770.1995.tb00963.x>
4. Munro, M. J., & Derwing, T. M. (2022). Foreign accent, comprehensibility and intelligibility, redux. In J. M. Levis, T. M. Derwing, & M. J. Munro (Eds.), *The evolution of pronunciation teaching and research: 25 years of intelligibility, comprehensibility, and accentedness* (pp. 7–32). John Benjamins. <https://doi.org/10.1075/bct.121.02mun>