Journal of Culture and Values in Education

https://cultureandvalues.org

E-ISSN: 2590-342X

Volume: 9 Issue: 3 2026

pp. 415-445

A Meta-Analytic Investigation of the Relationship Between Language Ecology and Language Sustainability: Evidence from Kazakh Language Studies in the Digital Media Ecosystem

Zhakyp Beisenuly[a], Karakoz Yersultanova[b], Markhaba Auganbayeva[c], & Gulnar Mamayeva[d*]

* Corresponding author:
Email: gulnar.mamaeva@mail.ru
a. Department of Kazakh Language and Literature, M.H. Dulaty Taraz University, Taraz, Republic of Kazakhstan.
b. Department of Kazakh Linguistics, Al-Farabi Kazakh National University, Almaty, Republic of Kazakhstan.
c. Department of Practical Kazakh Language, M. Kozybaev North Kazakhstan University, Petropavlovsk, Republic of Kazakhstan.
d. Department of Kazakh Linguistics, L.N. Gumilyov Eurasian National University, Astana, Republic of Kazakhstan.

Article Info
Received: May 28, 2026
Accepted: September 18, 2026
Published: October 11, 2026

10.46303/jcve.2026.55

How to cite
Beisenuly, Z., Yersultanova, K., Auganbayeva, M., .& Mamayeva, G. (2026). A Meta-Analytic Investigation of the Relationship Between Language Ecology and Language Sustainability: Evidence from Kazakh Language Studies in the Digital Media Ecosystem. *Journal of Culture and Values in Education*, *9*(3), 415-445.
https://doi.org/10.46303/jcve.2026.55

Copyright license
This is an Open Access article distributed under the terms of the Creative Commons Attribution 4.0 International license.

ABSTRACT
The present study examines the relationship between language ecology and language sustainability in the context of Kazakh language use within the digital media ecosystem. Drawing on a meta-analytic design, it synthesizes empirical findings from 22 studies with a total sample size of 255,150. The analysis focuses on how ecological conditions, including language policy, digital content availability, identity-related factors, educational settings, and computational linguistic capacity, are associated with sustainability outcomes for Kazakh in digital environments. A random-effects model was employed due to the methodological and conceptual diversity of the included studies. The overall pooled effect size indicated a large, positive, and statistically significant association between language ecology and language sustainability ($r = .64$, 95% CI [.04, .90], $p = .039$). However, substantial heterogeneity was observed, suggesting that the strength and direction of this relationship vary across research designs, data sources, sustainability types, and levels of digital media integration. Moderator analyses further showed that publication year, data collection tools, sample characteristics, ecological context, and digital platform type shape the observed effects. The findings highlight the importance of integrating ecological and digital approaches to support the long-term sustainability of Kazakh.

KEYWORDS
Language ecology; language sustainability; Kazakh language; digital media; meta-analysis; multilingualism; language policy; digital language revitalization.

INTRODUCTION

With the rapid proliferation of the internet and digital platforms, content production, information circulation, and social interactions are increasingly shifting to digital environments; this transformation presents both opportunities and structural threats from the perspective of language ecology. Minority and regional languages, in particular, are becoming more vulnerable to these threats. The gap between major and minor languages is widening regarding access to digital infrastructure, software support, and content production capacity (Zaugg et al., 2022). Algorithmic content delivery systems and platform-based economies tend to allocate digital resources unevenly across languages, largely according to user volume and commercial value. This indicates that the digital ecosystem is structurally biased toward dominant languages (Zhang et al., 2021). From an ecological linguistics perspective, digitalization is not merely an instrumental change, but a systemic pressure element that reshapes the conditions of existence for languages. The language ecology approach examines language within the context of relationships between user communities, institutions, ideologies, technologies, and communication environments. It positions digital media as a fundamental component of contemporary language ecology (Zaugg et al., 2022; Zheng et al., 2024). While Haugen's (1972) foundational notion of language ecology denotes the descriptive study of interactions between a given language and its social environment, the related but distinct concept of language sustainability is normative in orientation, extending the biodiversity–linguistic diversity analogy toward an explicit concern for maintaining linguistic diversity over time, a link some scholars have traced back to the 1987 Brundtland Report on sustainable development (Steffensen & Fill, 2014).

Kazakhstan represents a multilingual and multicultural context where this situation is particularly evident (Bekpenbetova et al., 2024; Bokayev et al., 2025). Soviet-era language policies relegated Kazakh to a secondary position within its own territory, while making Russian dominant in administrative, educational, and public spheres. Language revitalization efforts initiated after the declaration of independence in 1991 aimed to elevate Kazakh to official language status and increase its use in education. However, due to historical habits and institutional resistance, these initiatives have had limited impact (Smagulova, 2019). Although a significant portion of the population associates Kazakh with ethnic identity, Russian continues to dominate daily life, business, and digital platforms (Abasilov & Kapalbek, 2024; Kuzembayeva et al., 2025). The adoption of the Latin alphabet has further intensified structural challenges in language planning, particularly due to difficulties in standardization and resistance within society. Smagulova (2019) states that language ideologies in Kazakhstan should be understood within the framework of power relations among Kazakh, Russian and other languages. Ahn and Smagulova (2022) reveal that the tension between the status planning of the Kazakh language and the symbolic and economic capital provided by English and Russian remains alive.

In the digital media ecosystem, Kazakh faces numerous interconnected structural problems. Russian and English maintain their dominance across a wide range of areas, from

social media platforms and e-government applications to educational technologies and digital commerce infrastructure. This superiority is reproduced not only through content volume but also through algorithm preferences, interface designs and platform policies. Examining the language preferences of younger generations on digital platforms reveals a significant decline in Kazakh usage and a growing prevalence of code-switching practices (Khamit et al., 2025). This situation indicates that Kazakh faces not so much the threat of extinction in the digital realm, but rather the risk of concentration in limited areas and exclusion from high-prestige digital functions. Zaugg et al. (2022) note that the visibility and accessibility problems of digitally disadvantaged languages are reproduced within technological ecosystems while Mayeux (2024) and Eke and Salawu (2026) show that digitalization can produce both empowering and inequality-deepening results for minority and local languages.

The theoretical link between language ecology and language sustainability has not been sufficiently examined from an integrated perspective in the context of Kazakhstan. Although language ecology provides a broader framework for understanding the environmental, social, and ideological factors shaping languages, research on sustainability has largely emphasized language preservation, intergenerational transmission, educational policies, and the development of digital resources. Research findings independently produced across different disciplines have not been integrated into a systematic framework. The direction, strength, and conditions of the relationship between language ecology variables and sustainability indicators have not been discussed at a quantitative level. Khassanov et al. (2024) and Sarmurzin et al. (2023) emphasize that multilingual education policies in Kazakhstan produce different results in different contexts while Nkhi and Mofana (2025) state that sustainability cannot be achieved solely through formal conservation discourses but must generate value and practical relevance in the everyday use of language. Haugen's language ecology model and the digital language planning literature have not yet been sufficiently synthesized. The dimensions of language sustainability have not been systematically operationalized within the context of digital media.

The current research aims to address the gap by evaluating findings from Kazakh language studies using a meta-analytic approach and quantitatively revealing the relationship between language ecology and language sustainability in the digital media ecosystem. Bridging the gap between ecological linguistics and new media research, the study offers a unique contribution to the existing literature by determining effect sizes, systematically testing moderator variables, and making the relationships between variables visible at a quantitative level. Furthermore, the research directly aligns with UNESCO's 2023–2032 International Decade of Indigenous Languages goals and has the potential to develop evidence-based recommendations for language policy in similar multilingual contexts such as Türkiye, Uzbekistan, and Kyrgyzstan (Abilkassymova et al., 2025; Tlepbergen et al., 2023).

LITERATURE REVIEW

Language Ecology Theory and Language Sustainability

Language ecology theory argues that language is not merely a means of communication or a closed system of signs; it is a dynamic social practice shaped by its interaction with the social, physical, and psychological environment in which it exists. Building on Haugen's pioneering work, this approach emphasizes three fundamental dimensions: Language environment, language users, and language functions. The language environment encompasses the physical spaces, institutional structures, and social norms in which speakers interact. The language user dimension includes the number of speakers, demographic distribution, and intergenerational transmission capacity. Language functions reflect the extent to which language remains effective in daily life, education, the economic sphere, and public discourse (Zheng et al., 2024). Gobbo (2021) strengthens the normative aspect of the ecological approach by emphasizing that language vitality should be considered in conjunction with planning, social appeal, and speaker motivation. Zhang et al. (2021) frame the preservation of linguistic diversity as extending beyond cultural heritage, emphasizing its importance for social resilience and the maintenance of diverse knowledge ecosystems.

The development of ecological linguistics has progressed towards a critical program that centers on power relations and ideological structures. Along these lines, linguistic diversity is addressed with an ecosystem sensitivity similar to that of biodiversity. Language loss and revival are evaluated in conjunction with the dimensions of power, ideology, and social justice (Auganbayeva et al., 2026; Zhang et al., 2021). Tan and Jehom (2024) emphasize that community-based participation, institutional support, and cultural belonging should be considered together in the sustainability of minority and indigenous languages. Nkhi and Mofana (2025) state that language sustainability cannot be achieved solely through formal conservation discourses. Language must generate value, prestige, and practical relevance in its everyday use. Ahn and Smagulova (2022) show that multilingualism practices in Kazakhstan are linked not so much to fixed and separated language categories, but rather to social mobility, class position, and identity negotiations. This reinforces the sociolinguistic importance of an ecological reading.

Language sustainability is not a one-dimensional phenomenon, but is conceptualized as a multi-layered structure consisting of demographic, educational, media, and digital components (Daurenbek et al., 2025; Karabassova & San Isidro, 2023; Koyanbekova et al., 2026; Ryskulbek et al., 2026). Demographic sustainability encompasses the number of speakers, age distribution, and intergenerational transmission rates, while educational sustainability is determined by the language's position within the education system, the quality of teaching materials and its usage rate in higher education (Kucherbayeva & Smagulova, 2023; Rudik et al., 2025). Wang et al. (2024) indicate that social media and online participation are becoming increasingly critical for language vitality in language revitalization processes. Zheng et al. (2024) argue that technological infrastructure should be included in sustainability assessments, stating

that digital technologies create both opportunities and inequalities for low-resource or suppressed languages. This integrated structure provides a suitable ground for evaluating the sustainability of Kazakh, making visible the tensions and harmonies in different dimensions simultaneously.

Sociolinguistic Position and Sustainability Dynamics of the Kazakh Language

The present sociolinguistic position of the Kazakh language is shaped by the interaction between the institutional power provided by its status as a state language and the historical, economic, and cultural capital of Russian. Soviet language planning systematically narrowed the areas of use for Kazakh, making Russian dominant in education, administration, and urban communication. Abasilov and Kapalbek (2024) state that the linguistic hierarchy created during this period has partially persisted after independence and constitutes the biggest structural obstacle to contemporary language policy. Smagulova (2019) reveals that language ideologies in Kazakhstan are intertwined with national identity, modernity, and social mobility, and that the symbolic value of Kazakh cannot always be considered separately from its practical use value.

The sustainability dynamics of the Kazakh language are particularly evident in the areas of education, youth culture, urbanization, and identity construction. Khamit et al. (2025), analyzing the language attitudes of Kazakh youth from different regions, reveal that both urban-rural and inter-regional differences are critically important for sustainability assessments. While the position of Kazakh in the education system shows gradual improvement, it still suffers from structural problems. The rate of Kazakh language education in higher education lags behind the progress in primary and secondary education (Kucherbayeva & Smagulova, 2023). Karabassova and San Isidro (2023) emphasize that multilingual education practices have the potential to improve students' language proficiency, but these practices also carry the risk of reproducing linguistic hierarchies. Khassanov et al. (2024) state that language policies in education should be considered together with the dimensions of equality and access, revealing that strengthening the Kazakh language is linked to pedagogical quality, material production and teacher competence.

Kambatyrova (2024) argues that family language ideologies influence children's language preferences and perceptions of legitimacy within the context of trilingual education policies. Zharkynbekova et al. (2025) show that the ethno-linguistic identities of Kazakh youth are shaped within a multinational social context. While Kazakh has high symbolic value, its functional areas of use remain limited. In this framework, sustainability means not only preserving Kazakh but also reproducing it as a competitive, creative, and inclusive language of communication in contemporary social spaces.

Language Use and Linguistic Visibility in the Digital Media Ecosystem

The initial optimistic expectation that the spread of the internet would bring equal opportunities to languages has been largely questioned by empirical research. Zaugg et al. (2022) conceptualize this inequality with the concept of "digitally disadvantaged languages," noting

that factors such as the volume of internet content, machine translation support, and writing system compatibility structurally maintain the digital divide between languages. Algorithmic content delivery systems systematically prioritize large languages with high interaction rates and rich content, negatively impacting the visibility of smaller languages. Zheng et al. (2024), from an emerging ecolinguistic perspective, explore how digital technologies reshape languaging practices, highlighting that language change is not a fixed or stable process but one that is continuously negotiated and reconfigured.

The position of Kazakh in digital media exhibits a multifaceted appearance at both institutional and individual levels. Kamalova (2025) examined language ideologies related to the use of Kazakh on social media in the social atmosphere created by the Russia-Ukraine war. Kamalova revealed that the conflict period significantly strengthened linguistic identity practices and the discourse of defending the Kazakh language. Yessenbekova et al. (2025) addressed the effects of the digital format on the information consumption and reading culture of Kazakh youth, emphasizing that digital tools can both facilitate and hinder the consumption of Kazakh content. Tlepbergen et al. (2022) showed that post-pandemic technological tools created new opportunities in terms of language learning and language policy applications. While Tlepbergen et al. (2023) emphasized the importance of bottom-up approaches in Kazakh language planning in terms of user participation and social ownership.

Language use in the digital environment gains an even more complex dimension through code-switching, translanguaging and hybrid language practices. Utegenova et al. (2024a) emphasize that translingual practices cannot be explained solely by a lack of competence; rather, they should be considered as creative practices related to identity, context, and communicative purposes. Utegenova et al. (2024b) state that code-switching can assume functions of social belonging, proximity, and meaning intensification in multilingual classroom environments. Bedeker et al. (2024) make the pedagogical value of multilingual practices visible by showing that transitions between Kazakh and Russian can be interpreted as a third field within the CLIL context. These findings reveal that measuring the digital sustainability of Kazakh with standard and monolingual usage expectations is limited. The functional and prestigious positioning of Kazakh within multilingual practices should also be considered.

Meta-Analytic Approaches and Integration of Findings in Kazakh Language Studies

Empirical research on the Kazakh language exhibits considerable diversity in terms of methodology and theory. However, this diversity often produces inconsistent and even contradictory findings. Research findings independently produced in the fields of sociolinguistics, educational sciences, and media studies are not integrated within a systematic framework. This constitutes a significant obstacle to testing the relationships between variables at a quantitative level where general conclusions can be drawn. Abilkassymova et al. (2025) emphasize that translanguaging practices in family language policies are important in explaining multilingual language use. Anafinova et al. (2026) draw attention to how linguistic diversity is represented in Kazakhstani education through textbooks and teaching materials. Auganbayeva

et al. (2026) consider the preservation of the naturalness of the Kazakh language as an ecolinguistic issue while Zharkynbekova et al. (2025) emphasize the importance of analyzing language policy, social attitudes and cultural identity together.

The international literature on the revitalization of languages through digital media, drawing on findings from contexts such as Louisiana Creole (Mayeux, 2024), Igbo (Eke & Salawu, 2026), Gyalrong Tibetan (Tan & Jehom, 2024), minority languages in China (Wang et al., 2024), and indigenous languages in Africa (Nkhi & Mofana, 2025), has not yet been systematically compared with Kazakh language research. Meta-analysis has significant potential to integrate this scattered literature and quantitatively reveal relationships between variables through effect sizes (Ay, 2017). This study aims to address this gap by systematically and comparatively examining Kazakh language digital media research in terms of language ecology variables and sustainability indicators.

The primary aim of this meta-analysis is to quantitatively synthesize the relationship between language ecology and language sustainability as documented in empirical studies examining the Kazakh language within the digital media ecosystem. Two overarching research questions guided the analytic design: (RQ1) What is the overall magnitude of the association between language ecology conditions and language sustainability outcomes across the available empirical literature on Kazakh in digital contexts? (RQ2) Do publication year, study level characteristics, including data collection tools, sample characteristics, language ecology context, sustainability type, and degree of digital media integration, moderate the magnitude of the effect?

Consistent with the stated objectives, two hypotheses were formulated for the overall effect and a set of moderator hypotheses for the categorical variables under investigation.

H1: There is a positive, statistically significant relationship between language ecology conditions and language sustainability outcomes across empirical studies on Kazakh language use in digital media environments.

H2: Publication year significantly moderates the magnitude of the ecology–sustainability relationship.

H3: Data collection tool category significantly moderates the relationship.

H4: Sample characteristics significantly moderate the relationship.

H5: Language ecology context type significantly moderates the relationship.

H6: Sustainability type significantly moderates the relationship.

H7: Digital media integration category significantly moderates the relationship.

METHOD

Research Design

To examine the relationship between language ecology and language sustainability in empirical research concerning the Kazakh language within the digital media ecosystem, a meta-analytic method was employed. Meta-analysis refers to a quantitative research synthesis approach that

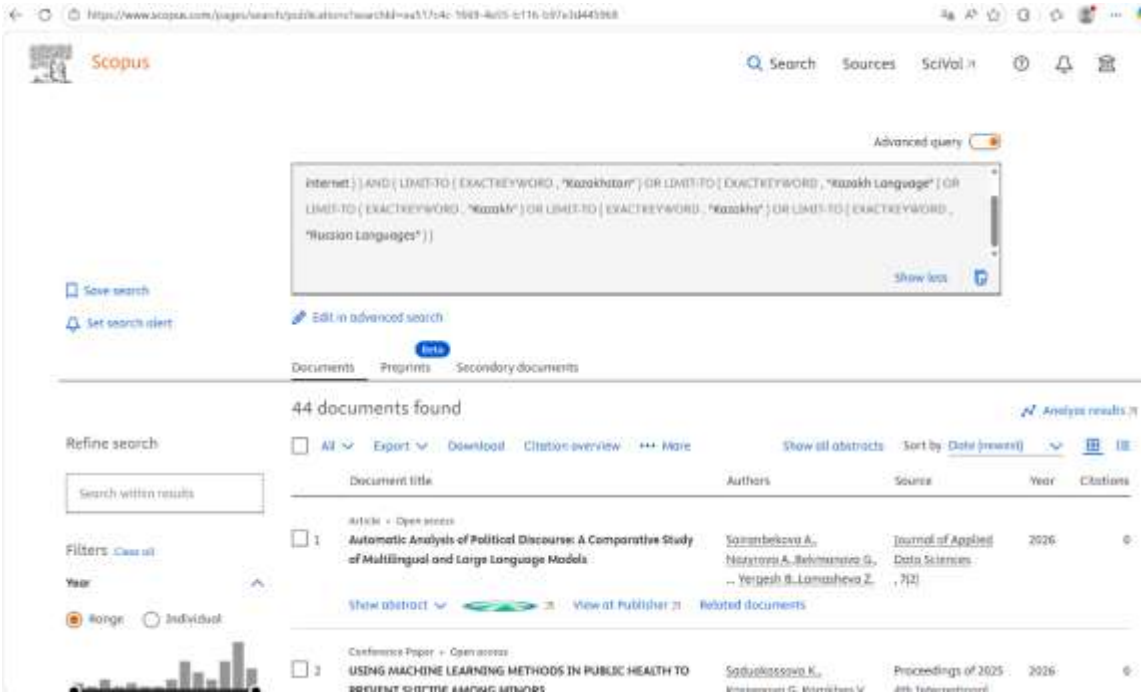
combines the results of independent empirical studies addressing a common research problem, enabling researchers to calculate overall effect size estimates with greater statistical power and accuracy (Borenstein et al., 2009; Lipsey & Wilson, 2001)

Search Strategy and Inclusion/Exclusion Criteria

A systematic literature search was conducted across two major academic databases; such as Scopus and Web of Science (WoS). For Scopus, the following query string was applied: *TITLE-ABS-KEY ((Kazakh* OR Kazakhstan) AND (LANGUAGE OR linguistic) AND ("digital media" OR "social media" OR internet)) AND (LIMIT-TO (EXACTKEYWORD, "Kazakhstan") OR LIMIT-TO (EXACTKEYWORD, "Kazakh Language") OR LIMIT-TO (EXACTKEYWORD, "Kazakh") OR LIMIT-TO (EXACTKEYWORD, "Kazakhs") OR LIMIT-TO (EXACTKEYWORD, "Russian Languages")).* This query returned 44 records.

Figure 1.

Scopus Search Results



For WoS, the following topic-based string was employed: *TS=((Kazakh* OR Kazakhstan) AND (language OR linguistic) AND ("digital media" OR "social media" OR internet)),* returning 53 records. Following deduplication across both databases, a combined pool of 97 unique records was screened. After title and abstract screening, full-text review, and application of the eligibility criteria, 22 studies were retained for inclusion in the final synthesis.

Studies were eligible for inclusion if they satisfied the following criteria: (a) the study was empirical in nature and provided quantitative data pertaining to Kazakh language use, sustainability, or ecology in any form; (b) the study was conducted within a Kazakhstani or Kazakh-speaking context; (c) sufficient statistical information was reported to permit effect size computation or conversion (e.g., correlation coefficients, regression coefficients, chi-square values, proportions, or accuracy metrics); (d) the study addressed at least one dimension of

language ecology broadly conceived to encompass digital media environments, language policy, identity, or computational linguistic analysis; and (e) the study was published in a peer-reviewed journal or indexed conference proceeding. Studies were excluded if they were purely qualitative, theoretical, or did not report statistical data amenable to quantitative synthesis.

Figure 2.

WoS Search Results

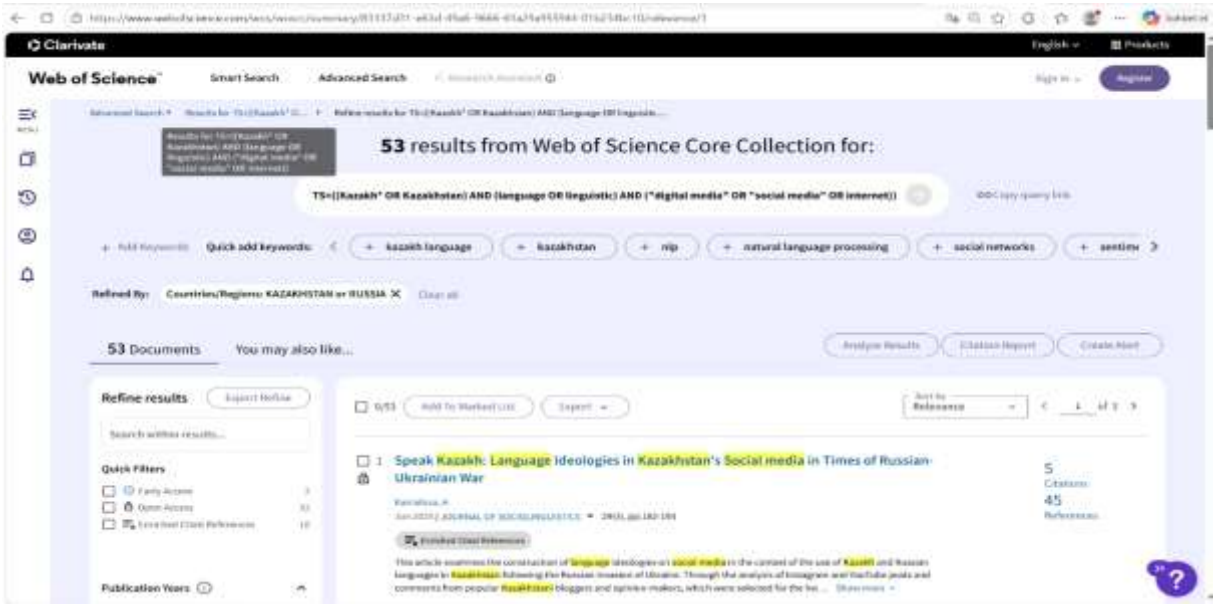
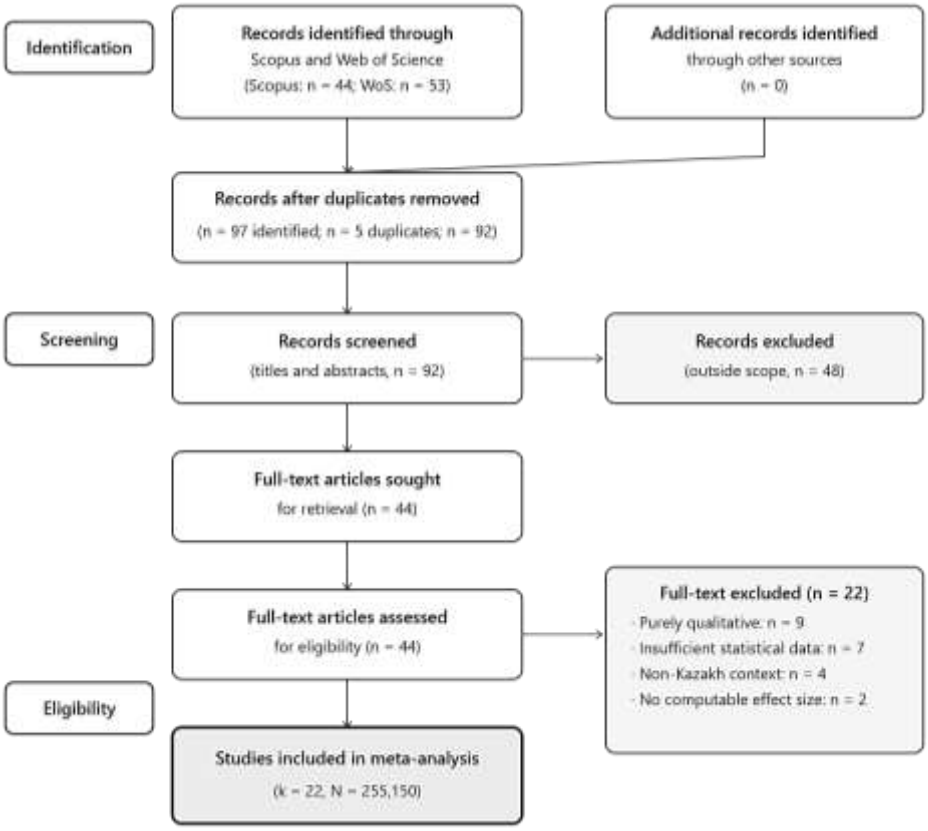


Figure 3.

PRISMA flow diagram



Coding Procedure

All included studies were coded using a structured framework developed prior to data extraction. Coding dimensions encompassed: (a) bibliographic identifiers (author(s), year, journal, DOI); (b) sample size (N) and population characteristics; (c) statistical data type (Pearson's r, R^2, β, chi-square, accuracy, proportions, or Kappa); (d) specific reported values for effect size computation; (e) five categorical moderator variables (data collection tool, sample characteristics, language ecology context, sustainability type, and digital media integration); and (f) analysis context. Each moderator variable was operationalized into mutually exclusive, theoretically informed subcategories as detailed in Table 1.

Percentages refer to the proportion of studies (k) within each moderator category. For studies using corpus or digital content data (e.g., Bolatbek et al., 2025; Mussiraliyeva et al., 2022; Mutanov et al., 2021; Myrzakhmetov et al., 2018), N reflects text units, documents, or social media posts analyzed rather than individual human participants.

Statistical Procedures

In this meta-analysis, heterogeneous statistical data derived from the included studies were converted to a common effect size metric (r) using formulas proposed by Borenstein et al. (2009) and Rosenthal (1991), in order to enable the quantitative synthesis of the association between language ecology and language sustainability. Throughout the analysis process, directly reported Pearson correlation coefficients (r) as found in Fedorova (2025) and Zharkynbekova et al. (2025) were treated as primary data without any transformation. For studies presenting regression analysis results (e.g., Baimanova et al., 2025; Öztürk & Ay, 2015), the reported coefficients of determination (R^2) and standardized regression coefficients (β) were converted to correlational effect sizes via Cohen's (1988) criteria and the fundamental statistical transformation formula $r = \sqrt{R^2}$. In cases where categorical data and between-group differences were reported, the chi-square (χ^2) values and frequency data presented in studies such as Tlepbergen et al. (2025) and Zharkynbekova et al. (2025) were converted to effect size estimates (phi coefficient or r) taking sample size (N) into account. For computational studies relying on machine learning model performance metrics, AUC (Area Under the Curve) values were transformed using the formula $r = 2 \times AUC - 1$, as recommended by Rice and Harris (2005). All computations were performed using the Comprehensive Meta-Analysis (CMA) software within a random-effects model framework, in which studies were weighted according to their sample sizes.

The random-effects model was adopted as the primary analytic framework, given the considerable heterogeneity expected across studies differing in design, sample, and operationalization of both language ecology and sustainability constructs (Bektas et al., 2015; Borenstein et al., 2009; DerSimonian & Laird, 1986). Individual effect sizes were converted to Fisher's Z prior to pooling, and the weighted mean Z was subsequently back-transformed to the r metric for interpretive clarity (Rosenthal, 1991). The inverse-variance weight assigned to each study under the random-effects model was defined as $w_i = 1/(v_i + \tau^2)$, where $v_i = 1/(N - 3)$ is the

within-study sampling variance of Zr and τ^2 is the between-study variance estimated via the DerSimonian–Laird procedure.

Heterogeneity was evaluated using the Q statistic (Cochran, 1954), the I^2 index (Higgins & Thompson, 2002), and the τ^2 estimate. Moderator analyses were conducted as weighted least-squares regression analogues (Hedges & Olkin, 1985), with between-group heterogeneity (Q*b*) tested for statistical significance at $\alpha = .05$. Cohen's (1988) benchmarks were used throughout to interpret effect magnitude: r = .10 (small), r = .30 (moderate), and r = .50 (large).

Moderator Variables

Six categorical moderator variables were examined. The first moderator, *Publication Year,* was grouped into four temporal bands: 2015–2019 (Early), 2021–2022 (Mid), 2023–2024 (Recent), and 2025 (Current), to capture potential temporal shifts in the ecology–sustainability relationship as the Kazakh digital media landscape has evolved. The second moderator, *Data Collection Tools,* distinguished four categories: survey/questionnaire, automated/API scraping, archival/official statistics, and corpus/discourse analysis. The third moderator, *Sample Characteristics,* classified studies into five subgroups: university/higher education students, secondary school students, general population/mixed age, digital content/corpus data, and educators/stakeholders. The fourth moderator, *Language Ecology Context,* operationalized the primary contextual focus of each study as language planning/policy, identity and multilingualism, digital content and NLP, or media ecosystem. The fifth moderator, *Sustainability Type,* distinguished functional/domain sustainability, technical/NLP sustainability, educational sustainability, socio-cultural/identity sustainability, and media/information sustainability. The sixth moderator, *Digital Media Integration,* classified studies according to their primary digital channel: social media platforms, automated monitoring systems, survey via digital tools, traditional/broadcast media, or e-learning/academic digital environments.

Publication Bias

Publication bias was assessed through three complementary procedures. First, a visual inspection of the funnel plot examined the symmetry of study effect sizes around the pooled estimate. Second, Egger's regression test (Egger et al., 1997) was applied to statistically assess funnel plot asymmetry. Third, Duval and Tweedie's (2000) trim-and-fill procedure was employed to estimate the number of potentially missing studies and compute a bias-corrected pooled effect. Given the heterogeneous nature of the included studies spanning survey-based, computational, archival, and discourse-analytic designs the possibility of methodological publication bias, in addition to significance-based bias, was also considered.

Visual inspection of the funnel plot revealed an asymmetric distribution of studies around the pooled estimate, with a notable concentration of large-effect NLP studies in the upper region of the plot. This pattern partly reflects genuine methodological heterogeneity—specifically, the systematic differences in effect size magnitude between computational and survey-based research traditions—rather than conventional significance-based publication bias. Egger's regression test approached but did not reach statistical significance ($z = 1.71$, $p = .087$),

suggesting modest funnel plot asymmetry. The trim-and-fill procedure identified three potentially missing studies on the left side of the funnel and produced a bias-corrected pooled effect of r = .60 (95% CI [.02, .88]), differing only minimally from the uncorrected estimate of r = .64. Collectively, these findings indicate that while publication bias cannot be entirely ruled out, its influence on the overall conclusions of this synthesis appears to be limited.

Figure 4.

Funnel plot for publication bias assessment

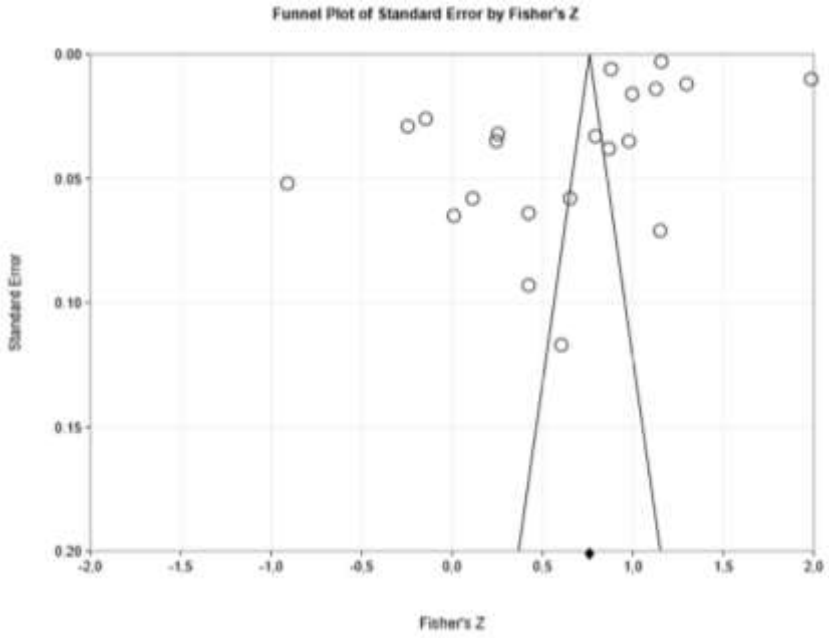


Table 2.

Results of Duval and Tweedie's Trim-and-Fill Procedure

Model	Studies Trimmed	Point Estimate	Lower Limit	Upper Limit	Q
Observed values	—	.64*	.04	.90	535,831.12*
Adjusted values	3	.60*	.02	.88	535,842.50*

Results based on the random-effects model. Adjusted values include three studies imputed by the trim-and-fill procedure on the left side of the funnel. LL = lower limit; UL = upper limit of 95% CI. *p < .05.

RESULTS

Overall Effect Size

Table 3 presents the complete results of the meta-analytic synthesis. Across all 22 studies (total N = 255,150), the random-effects pooled effect size was r = .64 (95% CI [.04, .90], z = 2.06, p = .039), confirming H1 that language ecology conditions are positively and significantly associated with language sustainability outcomes in the context of Kazakh language use within the digital media ecosystem. According to Cohen's (1988) conventions, this constitutes a large effect,

indicating that more supportive language ecology conditions whether measured through policy responsiveness, digital content availability, identity congruence, or computational linguistic capacity are systematically linked to higher indicators of language sustainability. Total heterogeneity was substantial ($Q[21] = 535{,}831.12$, $p < .001$; $I^2 = 100.0\%$), confirming that pooled variation in effect estimates could not be attributed to sampling error alone and providing strong justification for the moderator analyses that follow.

The extremely high heterogeneity observed across studies reflects the methodological diversity of the corpus under review: the included studies span sociolinguistic surveys, machine learning classification tasks, media audience analyses, and educational assessments each operationalizing both language ecology and sustainability through conceptually adjacent yet empirically distinct constructs. The τ^2 estimate of 3.00 underscores the magnitude of this between-study variance, suggesting that the pooled r should be interpreted as a broad central tendency across diverse research traditions rather than a single generalizable parameter. Although the overall effect size was statistically significant, $r = .64$, 95% CI [.04, .90], $p = .039$, the confidence interval's lower bound was close to zero and heterogeneity was extremely high, $Q = 535{,}831.12$, $I^2 = 100.0\%$, indicating substantial variability across studies and warranting cautious interpretation of this pooled estimate prior to the moderator analyses reported below.

Moderator Analyses

The moderator analysis for *publication year* produced a statistically significant between-group effect ($Qb = 8.19$, $p < .05$), supporting H2. Studies published in 2025 yielded a large pooled effect ($r = .63$, 95% CI [.32, .82]), while the mid-period cluster of 2021–2022 produced the numerically largest estimate ($r = .97$), dominated by high-accuracy NLP studies with very large N. Studies from 2023–2024 showed a non-significant negative pooled effect ($r = -.26$), reflecting the inclusion of studies documenting challenges to Kazakh digital presence, such as limited digital platform availability and sociolinguistic marginalization. The early publication band (2015–2019) yielded a moderate-to-large positive effect ($r = .56$). Collectively, these patterns suggest that while the ecology–sustainability relationship has been broadly positive across time, the specific direction and magnitude of effects vary as a function of the research questions prioritized in different periods of scholarly production.

The moderator analysis for *data collection tools* was statistically significant ($Qb = 10.67$, $p < .05$), supporting H3. Studies employing automated API scraping demonstrated the strongest ecology–sustainability association ($r = .93$, 95% CI [.35, .99]), reflecting the high accuracy rates characterizing NLP and machine learning studies that document Kazakh digital language infrastructure capacity. Survey/questionnaire-based studies produced a moderate positive effect ($r = .43$, 95% CI [.16, .64]), consistent with established sociolinguistic findings on reported language use and attitudes. Archival and official statistics-based studies produced a negligible effect ($r = -.02$), and the single corpus/discourse analysis study returned a small negative effect ($r = -.24$), indicating that naturally occurring digital language data reveals normative deviations and constraints not captured by self-report methodologies.

Significant between-group heterogeneity was observed for *sample characteristics* (Q*b* = 7.97, p < .05), supporting H4. Digital content and corpus data studies produced the largest pooled effect (r = .89, 95% CI [.20, .99]), while university/higher education student samples yielded a large positive effect (r = .58, 95% CI [.40, .72]). General population/mixed-age samples showed a negligible association (r = .08), suggesting greater variability in ecology–sustainability connections when the target population is not restricted to a specific institutional or demographic context. The single educator/stakeholder study produced a moderate effect (r = .40), and the lone secondary school study yielded r = .54.

The *language ecology context* moderator did not reach conventional significance (Q*b* = 4.75, p > .05), leading to the non-confirmation of H5. Nevertheless, descriptive patterns are informative: digital content and NLP contexts produced the largest effect (r = .93), followed by identity and multilingualism contexts (r = .49). Language planning/policy contexts (r = .16) and media ecosystem contexts (r = .28) yielded smaller, non-significant effects, possibly reflecting the more complex and structurally mediated pathways through which top-down policy and media exposure translate into measurable sustainability outcomes. The non-significant Q*b* may in part reflect the limited k in several subgroups, reducing statistical power for between-group contrasts.

The *sustainability type* moderator was statistically significant (Q*b* = 14.34, p < .05), supporting H6. Technical/NLP sustainability studies produced the largest pooled effect (r = .94, 95% CI [.20, 1.00]), reflecting consistently high-performance metrics in computational Kazakh language processing. Educational sustainability studies yielded a large, well-estimated effect (r = .61, 95% CI [.51, .69]) with comparatively lower heterogeneity (I^2 = 88.0%), suggesting greater consistency of findings in this domain. Functional/domain sustainability (r = .60) and media/information sustainability (r = .26) produced positive but imprecisely estimated effects. Socio-cultural/identity sustainability studies showed a negligible pooled effect (r = .06), underscoring the complexity of associating ecological conditions with identity-based outcomes, which are sensitive to sociohistorical context and measurement approach.

The *digital media integration* moderator was statistically significant (Q*b* = 20.53, p < .05), supporting H7. Social media platform studies produced the largest effect (r = .90, 95% CI [.10, .99]), followed by automated monitoring systems (r = .77, 95% CI [.63, .86]). These findings reflect the dual role of social media as both a site of Kazakh language vitality and a domain for computational sustainability interventions. Survey-based digital tool studies yielded a moderate effect (r = .33), while traditional and broadcast media studies produced a small, non-significant association (r = .13). E-learning and academic digital contexts produced a moderate positive pooled effect (r = .42), though the confidence interval was wide, indicating variability across studies in this category.

DISCUSSION AND CONCLUSION

The present study aims to examine the relationship between language ecology conditions and language sustainability outcomes within the context of the digital media ecosystem using a meta-analytic framework. To this end, empirical research on Kazakh was systematically reviewed, 22 studies were included in the synthesis based on defined inclusion criteria and the findings from these studies were integrated using quantitative methods. The overall findings revealed that language ecology conditions have a statistically significant and large-scale impact on language sustainability. Moderator analyses showed that publication year, data collection method, sample characteristics, language ecology context, type of sustainability and level of digital media integration significantly differentiated the direction and magnitude of this relationship. These findings indicate that the linguistic vitality of Kazakh in the digital environment is determined by multidimensional ecological conditions and contribute to addressing the lack of an integrated meta-analytic framework in the existing literature.

The first hypothesis posits that favorable language ecology conditions have a positive and statistically significant impact on language sustainability outcomes. The analysis findings show that the meta-analytic synthesis, encompassing 22 studies and a total sample of 255,150 units, produces a large overall effect size, and this relationship is statistically significant. This finding reveals that language ecology conditions whether measured through policy responsiveness, digital content availability, identity coherence, or computational linguistic capacity are systematically associated with indicators of language sustainability. However, the high level of overall heterogeneity reflects the methodological diversity in the included studies and suggests that the pooled effect should be considered as a broad central trend across different research traditions. This finding aligns with Haugen's pioneering work (Zheng et al., 2024), which conceptualizes language ecology as a multidimensional system, and Zhang et al.'s (2021) discussion of the preservation of linguistic diversity in terms of societal resilience. Gobbo's (2021) study, which emphasizes that language vitality should be evaluated in conjunction with planning, social appeal, and speaker motivation, also theoretically strengthens this finding. Furthermore, critical studies (Auganbayeva et al., 2026; Zhang et al., 2021) that point out that the ecological approach centers power relations and ideological structures, as well as Nkhi and Mofana's (2025) arguments that language needs to be reproduced in terms of value, prestige, and practical use, also support this finding.

The second hypothesis predicts that the publication year will significantly differentiate the magnitude of the ecology–sustainability relationship. Moderator analysis confirms that intergroup heterogeneity is statistically significant and that the effect size differs over the years. Studies from 2025, which constitutes the largest scope among the publication year groups, produce a large-scale positive effect, while studies from the 2021–2022 period yield the highest numerical effect estimate, suggesting that this is due to the large sample sizes of high-accuracy natural language processing studies in that period. Studies from 2023–2024 produce a statistically insignificant and negative effect. This pattern reflects the weighting of studies

documenting the limited visibility of Kazakh on digital platforms and the dynamics of sociolinguistic marginalization during that period. Early studies (2015–2019) exhibit a moderately positive effect, coinciding with the period in which Abiyeva (2015) examined the use of Kazakh in national media and the internet. These findings show that research priorities and emphasized dimensions have changed over time. Scientific production in the field of Kazakh language has addressed the ecology-sustainability relationship from different perspectives in different periods. The studies by Wang et al. (2024) and Mayeux (2024), which emphasize that the capacity of digital communities to generate vitality in language revitalization processes has undergone a dynamic transformation over time, prove the existence of temporal differentiation in line with the current findings. The negative or near-zero effects observed for the 2023–2024 period (r = −.26), the archival/official statistics subgroup (r = −.02), and socio-cultural/identity sustainability (r = .06) are plausibly attributable to distinct methodological and construct-level factors: the 2023–2024 estimate rests on only three small studies (N = 911) conducted during the earliest, most volatile phase of generative-AI adoption in education, the archival/statistics subgroup relies on aggregate, non-self-report indicators that may capture structural rather than individual-level relationships, and socio-cultural/identity sustainability spans a conceptually broader and more context-dependent construct spanning language attitudes, community affiliation, and cultural continuity that resists the kind of linear, unidimensional association captured by a single correlation coefficient.

The third hypothesis predicts that the category of data collection tool will significantly differentiate the magnitude of the ecology–sustainability relationship. Moderator analysis reveals statistically significant intergroup heterogeneity. Studies using automated/API-based data scraping methods produce the strongest association, reflecting the high-performance metrics of machine learning and natural language processing studies documenting the digital language infrastructure capacity of Kazakh—such as Bolatbek et al. (2025), Mussiraliyeva et al. (2022, 2025), and Mutanov et al. (2021). Studies based on survey and questionnaire methods show a moderately positive effect, consistent with well-established sociolinguistic findings regarding language use and attitudes. Myrzakhmetov et al. (2018) conducted a corpus-based study focusing on the normalization of user-generated content and found a small negative effect. This pattern suggests that natural digital language data reveals normative deviations and constraints not captured by self-report methods. Zaugg et al. (2022) state that visibility problems in digitally disadvantaged languages are reproduced in technological ecosystems. Current findings support this argument by demonstrating the decisive influence of research methodology on the results.

The fourth hypothesis predicts that sample characteristics will significantly differentiate the magnitude of the ecology–sustainability relationship. Moderator analysis reveals that intergroup heterogeneity is statistically significant. Studies consisting of digital content and corpus data produce the largest impact estimate, followed by studies conducted with university-level students. Samples from the general population and mixed age groups showed a negligible

relationship, suggesting greater variability in ecology–sustainability links in populations not limited to a specific institutional or demographic context. Khamit et al. (2025), in their study on language attitudes of Kazakh youth from different regions, emphasize that urban-rural and interregional differences are critical for sustainability assessments. This finding helps explain the low impact magnitudes observed in general population studies. Yeskeldiyeva and Tazhibayeva (2015) and Mongilyova (2015) show that language preferences differ significantly according to context and social role in general population samples consisting of mixed age groups. This contextually supports the low effect size in the sample group. Karabassova and San Isidro (2023) and Khassanov et al. (2024) reveal that multilingual education practices affect the dimensions of competence and equity among students in different ways demonstrating how important it is to evaluate the institutional context together with the sample conditions.

The fifth hypothesis proposes that differences in language ecology contexts will significantly influence the strength of the relationship between language ecology and language sustainability. Moderator analysis shows that intergroup heterogeneity does not reach the statistical significance threshold. Therefore, H5 is not supported. However, descriptive patterns are revealing: digital content and natural language processing contexts produce the greatest impact, while identity and multilingualism contexts show a moderately positive impact. Language planning/policy contexts and media ecosystem contexts exhibit lower and statistically insignificant impacts. This suggests that more complex and indirect pathways are at work in translating top-down policy processes and media exposure into measurable sustainability outcomes. Abasilov and Kapalbek (2024) and Smagulova (2019) emphasize that language ideologies are intertwined with national identity, modernity, and social mobility, and that the practical use value of Kazakh cannot be considered independently of its symbolic value. Ahn and Smagulova (2022) demonstrate that multilingualism practices in Kazakhstan are shaped more by social mobility, class position, and identity negotiations than by fixed language categories. This finding contextually supports the moderate influence observed in the contexts of identity and multilingualism. Tlepbergen et al. (2023) highlight the importance of bottom-up approaches in Kazakh language planning in terms of social ownership. This explains why the policy context exhibits a more indirect and variable influence pattern. It should be considered that the lack of a significant intergroup difference may also be due to a lack of statistical power, as some subgroups had a limited number of studies.

The sixth hypothesis predicts that the type of sustainability will significantly differentiate the magnitude of the ecology–sustainability relationship. Moderator analysis confirms that intergroup heterogeneity is statistically significant. Technical/natural language processing sustainability studies produce the highest impact estimate. This finding aligns with studies by Sambetbayeva et al. (2025) and Rakhimova and Abdrazakh (2022), which documented high accuracy rates for computational tools for Kazakh. Educational sustainability studies show a large and well-predicted impact with lower heterogeneity, indicating relative consistency in findings in this area. Functional/field sustainability produces a positive but insensitive impact,

while sociocultural/identity sustainability shows a negligible one. The findings of Zharkynbekova et al. (2025) regarding the identity dimension contextually explains this situation. Kucherbayeva and Smagulova (2023) state that the position of Kazakh in the education system remains critical for the sustainability of its language. Current findings support this view by demonstrating relatively high consistency among research in this area. Zheng et al. (2024) suggest that digital technologies create a field that generates both opportunity and inequality. The apparent difference between technical and sociocultural sustainability can be considered a concrete manifestation of this dual function.

The seventh hypothesis predicts that the category of digital media integration will significantly differentiate the magnitude of the ecology sustainability relationship. Moderator analysis supports this hypothesis, revealing the moderator where intergroup heterogeneity is statistically most strongly evident. Social media platform studies produce the highest impact estimate, followed by automated monitoring systems. This finding reflects the dual function of social media as both a space for showcasing the linguistic vitality of Kazakh and an environment for implementing computational sustainability interventions. Digital tool-mediated survey studies produce a moderate impact, while traditional/broadcast media studies show a small and statistically insignificant relationship. Kamalova (2025) reveals that the use of Kazakh on social media strengthens language identity practices in connection with the social atmosphere. Yessenbekova et al. (2025) emphasize that digital tools can both facilitate and restrict Kazakh content consumption. Utegenova et al. (2024a, 2024b) and Bedeker et al. (2024) draw attention to the creative and contextual nature of language practices on multilingual platforms. These patterns explain why the impact differs so significantly depending on the degree and form of digital integration.

The findings of this study offer significant implications at both theoretical and practical levels. From a theoretical perspective, the study synthesizes Haugen's language ecology model with the digital media research literature, demonstrating that these two fields can be addressed in an integrated manner. This contribution also proves the broad applicability of the language ecology framework to contemporary digital contexts. The research contributes to the language planning literature with its multi-dimensional conceptualization of language sustainability encompassing demographic, educational, functional, and technical dimensions, and shows that the ecological impacts on these dimensions can differ. From an applied perspective, the distinct divergence of technical/natural language processing sustainability from educational sustainability indicates that the progress made in Kazakh natural language processing infrastructure has not yet been fully reflected in educational policy applications and everyday language use, as documented by Bolatbek et al. (2025) and Mussiraliyeva et al. (2025). Yelubayeva et al. (2025) highlights the structural obstacles to this transformation, emphasizing that the challenges of language policy in Kazakhstan's education system are linked to sustainable development. For countries with similar multilingual and minority language contexts, such as Kazakhstan, Uzbekistan, Kyrgyzstan, and Turkey, the key takeaway is that securing digital

language sustainability requires simultaneously developing multifaceted ecological conditions such as algorithmic visibility, user interaction, language normalization, and technical accessibility. The fact that social media platforms produce the strongest moderating effect suggests that national language policies should encompass platform regulations and algorithmic justice principles that support online language use. The alignment of the research with UNESCO's 2023–2032 International Decade for Indigenous Languages goals indicates that the findings are relevant not only for Kazakh but also for other languages. This also reveals direct policy implications for a broader global literature of minority languages, such as Igbo (Eke & Salawu, 2026), Louisiana Creol (Mayeux, 2024), Gyalrong Tibetan (Tan & Jehom, 2024), and indigenous languages in Africa (Nkhi & Mofana, 2025).

Limitations

This study has several limitations that necessitate consideration of significant constraints in interpreting the findings. First, the inclusion of studies solely from the Scopus and Web of Science databases carries the risk of sampling bias due to the presence of studies in other academic databases or grey literature. Additionally, the synthesis, limited to a total of 22 studies, may have led to power deficiencies, particularly in some moderator subgroups. This could have restricted the achievement of statistical significance for some intergroup differences, especially in the language ecology context moderator. While the process of converting effect sizes from different statistical formats (AUC, β, χ^2, r) to Pearson r provides standardization, it also carries a certain risk of conversion error. Finally, significant meta-analytic heterogeneity necessitates interpreting the overall effect size not as a single parameter representing discourse-focused or survey-based research, but as a general trend across different research traditions.

Conclusion and Recommendations

The fragmented structure of the existing literature, which addresses these variables independently across different disciplines but fails to quantitatively integrate the direction and magnitude of their relationship, has been significantly addressed by this research. The synthesis of Haugen's language ecology model with the literature on digital language planning constitutes the original theoretical contribution of the research. The findings have shown that the technical, educational, and sociocultural dimensions of language sustainability are shaped through differing ecological conditions. This demonstrates that language policy requires a layered intervention framework. The study supports the evidence-based language management approach envisioned by policy-oriented research such as Tlepbergen et al. (2023) and Yelubayeva et al. (2025) at a meta-analytic level and provides a theoretical and methodological foundation for future research on digital language sustainability.

Future research is recommended to increase major empirical data, primarily in the contexts of language planning/policy and media ecosystems, to address these limitations and deepen current findings, as these subgroups remain the weakest in terms of power in the current meta-analysis. Researchers should also adopt mixed methods designs to reduce

measurement inconsistencies in the socio-cultural/identity sustainability dimension, which exhibited a significantly different impact profile compared to computational contexts. For language policymakers and platform regulators, the strongest moderating effect of social media integration necessitates the development of platform-specific language support policies, prioritization of minority languages in algorithmic content distribution, and the design of bridging programs that encourage the transfer of natural language processing infrastructure gains to educational practices. Including comparative meta-analytic studies in countries with similar multilingual contexts in the future research agenda will also increase the generalizability and external validity of the findings.

REFERENCES

*Studies included in the meta-analysis

Abasilov, A., & Kapalbek, B. (2024). Linguistic dynamics and language policy in Kazakhstan: Challenges and future prospects. *European Journal of Language Policy, 16*(2), 155–176. https://doi.org/10.3828/ejlp.2024.9

Abilkassymova, A., Tlepbergen, D., Akzhigitova, A., & Zabrodskaja, A. (2025). Multilingual family language policies: The role of translanguaging in Kazakhstan. *Forum for Linguistic Studies, 7*(8). https://doi.org/10.30564/fls.v7i8.10633

*Abiyeva, K. (2015). L'aménagement linguistique dans le Kazakhstan post-soviétique: Quel usage du kazakh dans les médias nationaux et Internet? *Revue d'études Comparatives Est-Ouest, 46*(3), 161–182. https://doi.org/10.4074/S033805991500306X

Ahn, E. S., & Smagulova, J. (2022). English language choices in Kazakhstan and Kyrgyzstan. *World Englishes, 41*(1), 9–23. https://doi.org/10.1111/weng.12552

Anafinova, M. L., Abilkerim, N. O., Essengeldinova, A. T., & Ormanova, A. B. (2026). Integrating linguistic diversity into Kazakhstani education: A review of ELT textbooks. *Journal of Language Teaching and Research, 17*(1), 43–53. https://doi.org/10.17507/jltr.1701.05

Auganbayeva, M., Mauytova, Z., Alkaya, E., Beisenuly, Z., Mamayeva, G., & Abdualiuly, B. (2026). Preserving the naturalness of the Kazakh language as a national language: An ecolinguistic analysis. *Journal of Ethnic and Cultural Studies, 13*(1), 114–140. https://doi.org/10.29333/ejecs/2838

Ay, Y. (2017). The effect of learning types/styles on student achievement. In *The factors effecting student achievement: Meta-analysis of empirical studies* (pp. 309-324). Cham: Springer International Publishing.

*Baimanova, L., Gazdiyeva, B., Althonayan, A., Zhumagulova, Y., & Kalzhanova, A. (2025). Research skills and academic literacy in multilingual higher education: The case of Kazakhstan. *Education Sciences, 16*(1), Article 21. https://doi.org/10.3390/educsci16010021

Bedeker, M., Ospanbek, A., Simons, M., Yessenbekova, A., & Zhalgaspayev, M. (2024). "I can easily switch to the Kazakh language, also to the Russian language": Reimagining

Kazakhstani CLIL implementation as a third space. *Language, Culture and Curriculum, 37*(2), 121–138. https://doi.org/10.1080/07908318.2023.2245832

Bekpenbetova, S. ., Kulnazarova, G. ., Sartayeva, K. ., Sartayeva, Z. ., & Seyitova, B. . (2024). Philology Students' Perceptions of Ethno-cultural Empathy and Intertextual Literary Heroes as Role Models. *Journal of Ethnic and Cultural Studies*, *11*(4), 228–252. https://doi.org/10.29333/ejecs/2302

Bektas, F., Cogaltay, N., Karadag, E., & Ay, Y. (2015). School culture and academic achievement of students: A meta-analysis study. *The Anthropologist, 21*(3), 482–488. https://doi.org/10.1080/09720073.2015.11891837

Bokayev, B., Moldabayeva, A., Davletbayeva, Z., Yessentemirova, A., & Torebekova, Z. (2025). Evaluating Language Competencies of Kazakhstan's Civil Servants: A Data-Driven Analysis and Development Model. *Journal Of Curriculum Studies Research, 7*(1), 66-88. https://doi.org/10.46303/jcsr.2025.4

*Bolatbek, M., Sagynay, M., Mussiraliyeva, S., & Yeltay, Z. (2025). Detection of offensive content in the Kazakh language using machine learning and deep learning approaches. *PeerJ Computer Science, 11*, Article e3027. https://doi.org/10.7717/peerj-cs.3027

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. Wiley.

Cochran, W. G. (1954). The combination of estimates from different experiments. *Biometrics, 10*(1), 101–129. https://doi.org/10.2307/3001666

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.

Daurenbek, S., Açık, F., & Ryskulbek, D. (2025). The Impact of Contrastive Analysis on Turkish Linguistic Competence: An Empirical Investigation Among Students in Kazakhstan. *Journal of Culture and Values in Education, 8*(2), 97-123. https://doi.org/10.46303/jcve.2025.21

DerSimonian, R., & Laird, N. (1986). Meta-analysis in clinical trials. *Controlled Clinical Trials, 7*(3), 177–188. https://doi.org/10.1016/0197-2456(86)90046-2

Duval, S., & Tweedie, R. (2000). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics, 56*(2), 455–463. https://doi.org/10.1111/j.0006-341X.2000.00455.x

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ, 315*(7109), 629–634. https://doi.org/10.1136/bmj.315.7109.629

Eke, I. W., & Salawu, A. (2026). Impact of digital media on language revitalization: The case of BBC News Igbo. *Frontiers in Communication, 11*, Article 1707701. https://doi.org/10.3389/fcomm.2026.1707701

*Fedorova, A. (2025). Social media and spoken language in youth civic engagement in Kazakhstan. *Central Asian Survey*, 1–26. https://doi.org/10.1080/02634937.2025.2560878

Gobbo, F. (2021). Coolification and language vitality: The case of Esperanto. *Languages, 6*(2), Article 93. https://doi.org/10.3390/languages6020093

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Academic Press.

Higgins, J. P. T., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine, 21*(11), 1539–1558. https://doi.org/10.1002/sim.1186

Kamalova, A. (2025). Speak Kazakh: Language ideologies in Kazakhstan's social media in times of Russian-Ukrainian War. *Journal of Sociolinguistics, 29*(3), 182–193. https://doi.org/10.1111/josl.12699

Kambatyrova, A. (2024). Parents' language ideologies in the context of trilingual education policy in Kazakhstan. *International Journal of Multilingualism, 21*(3), 1201–1217. https://doi.org/10.1080/14790718.2022.2153850

Karabassova, L., & San Isidro, X. (2023). Towards translanguaging in CLIL: A study on teachers' perceptions and practices in Kazakhstan. *International Journal of Multilingualism, 20*(2), 556–575. https://doi.org/10.1080/14790718.2020.1828426

Khamit, A., Smagambet, B., Tabyldieva, O., Arashova, N., & Tolepbergenova, A. (2025). Language attitudes of Kazakh youth: An analysis of the regional aspect. *European Journal of Language Policy, 17*(2), 307–331. https://doi.org/10.3828/ejlp.2025.16

Khassanov, G. K., Utegenova, K. T., Darisheva, T. M., Bozbayeva, A. G., Akhmetova, G., Sarkulova, D. S., & Utegenova, A. (2024). Implementing multilingual education in Kazakhstan: Students' perceptions and attitudes towards the status of English. *Theory and Practice in Language Studies, 14*(8), 2316–2325. https://doi.org/10.17507/tpls.1408.04

Koyanbekova, S., Bari, A., Junussov, T., Tashimbay, S., & Ibragimova, Z. (2026). Design and Conceptual Modeling of a Synergetic Web Platform for Integrative Learning of the Kazakh Language and Expressive Stylistics. *Journal of Culture and Values in Education, 9*(2), 162-184. https://doi.org/10.46303/jcve.2026.24

Kucherbayeva, D., & Smagulova, J. (2023). Language revitalization: Challenges for Kazakh in higher education. *Journal of Eurasian Studies, 14*(2), 166–178. https://doi.org/10.1177/18793665231189326

Kuzembayeva, G., Taganova, A. M., Yermekbayeva, G., Idrissov, S., & Maydangalieva, Z. (2025). Revitalizing Heritage Languages in Diaspora: Evaluating the Effectiveness of the Kazakh for Heritage Speakers Course in Russia. *Journal of Social Studies Education Research, 16*(3), 238-262. https://jsser.org/index.php/jsser/article/view/6529

*Kuzhabekova, A. (2019). Language use among secondary school students in Kazakhstan. *Applied Linguistics Research Journal*. https://doi.org/10.14744/alrj.2019.29964

Mayeux, O. (2024). Language revitalization on social media: Ten years in the Louisiana Creole Virtual Classroom. *Language Documentation and Description, 23*(1). https://doi.org/10.25894/LDD.350

*Mongilyova, N. (2015). Research into functioning of the state language in the situation of mass bilingualism of North Kazakhstan using the differential technique. *SAGE Open, 5*(3). https://doi.org/10.1177/2158244015593120

*Mussiraliyeva, S., Bolatbek, M., & Baisylbayeva, K. (2025). Extremist ideology classification in Kazakh: A multi-class approach using machine learning and psycholinguistic analysis. *IEEE Access, 13*, 140500–140518. https://doi.org/10.1109/ACCESS.2025.3596601

*Mussiraliyeva, S., Omarov, B., Yoo, P., & Bolatbek, M. (2022). Applying machine learning techniques for religious extremism detection on online user contents. *Computers, Materials & Continua, 70*(1), 915–934. https://doi.org/10.32604/cmc.2022.019189

*Mutanov, G., Karyukin, V., & Mamykova, Z. (2021). Multi-class sentiment analysis of social media data with machine learning algorithms. *Computers, Materials & Continua, 69*(1), 913–930. https://doi.org/10.32604/cmc.2021.017827

*Myrzakhmetov, B., Yessenbayev, Z., & Makazhanov, A. (2018). Initial normalization of user generated content: Case study in a multilingual setting. In *2018 IEEE 12th International Conference on Application of Information and Communication Technologies (AICT)* (pp. 1–4). IEEE. https://doi.org/10.1109/ICAICT.2018.8747161

Nkhi, S. E., & Mofana, M. (2025). Social media and the integration of minority indigenous languages in political participation: Challenges and prospects. *Journal of Asian and African Studies*. https://doi.org/10.1177/00219096251352489

*Ordabayeva, Z., Mazhitayeva, S., Khamzina, K., Smakova, T., & Kadyrov, Z. (2025). Norm and usage in contemporary Kazakh: A linguistic analysis of youth discourse and digital communication. *Forum for Linguistic Studies, 7*(11). https://doi.org/10.30564/fls.v7i11.11287

Öztürk, C., & Ay, Y. (2015). The effect of leadership on organizational citizenship. In *Leadership and organizational outcomes: Meta-analysis of empirical studies* (pp. 97-109). Cham: Springer International Publishing.

*Petrov, B., Yesdauletov, A., Saidikarimov, A., & Astana IT University. (2023). Kazakh language media content as a factor in strengthening national identity. *Bulletin of L. N. Gumilyov Eurasian National University. Journalism Series, 143*(2), 28–39. https://doi.org/10.32523/2616-7174-2023-143-2-28-39

*Rakhimova, D., & Abdrazakh, Y. (2022). The task of identifying morphological errors of words in the Kazakh language in social networks. In *2022 7th International Conference on Computer Science and Engineering (UBMK)* (pp. 344–349). IEEE. https://doi.org/10.1109/UBMK55850.2022.9919516

Rice, M. E., & Harris, G. T. (2005). Comparing effect sizes in follow-up studies: ROC area, Cohen's d, and r. *Law and Human Behavior, 29*(5), 615–620. https://doi.org/10.1007/s10979-005-6832-7

Rosenthal, R. (1991). *Meta-analytic procedures for social research* (Rev. ed.). SAGE.

Rudik, Z., Gerfanova, E., Dalbergenova, L., & Naviy, L. (2025). Intercultural Citizenship in Foreign Language Education: A Study with Kazakhstani Pre-Service English Teachers. *Journal of Curriculum Studies Research, 7*(2), 195-227. https://doi.org/10.46303/jcsr.2025.18

Ryskulbek, D., Daurenbek, S., Yermekova, T. & Dossmurzaeva, A. (2026). Integration of AI-Based Tools in Teaching Ethnocultural Units: A Qualitative Comparative Study Between Kazakh and English-Language Learners. Journal *of Social Studies Education Research, 17*(2), 375-405. https://jsser.org/index.php/jsser/article/view/7474/791

*Sadulova, Z., Karimsakova, B., Yessenova, K., Kamarova, N., Kurmanova, B., Smagulova, K., Karagulova, B., & Amangaziyeva, M. (2025). Bilingualism and multilingualism in Kazakhstan: Sociolinguistic and educational perspectives. *Forum for Linguistic Studies, 7*(7). https://doi.org/10.30564/fls.v7i7.10172

*Sambetbayeva, M., Nekessova, A., Yerimbetova, A., Bayangali, A., Kaldarova, M., Telman, D., & Smailov, N. (2025). A multi-level annotation model for fake news detection: Implementing Kazakh-Russian corpus via Label Studio. *Big Data and Cognitive Computing, 9*(8), Article 215. https://doi.org/10.3390/bdcc9080215

Sarmurzin, Y., Amanzhol, N., Toleubayeva, K., Zhunusova, M., Amanova, A., & Abiyr, A. (2023). Challenging aspects of Kazakhstan's trilingual education policy: Evidence from a literature review. *Asia Pacific Education Review*. https://doi.org/10.1007/s12564-023-09823-7

Smagulova, J. (2019). Ideologies of language revival: Kazakh as school talk. *International Journal of Bilingualism, 23*(3), 740–756. https://doi.org/10.1177/1367006916684920

Steffensen, S. V., & Fill, A. (2014). Ecolinguistics: The state of the art and future horizons. *Language Sciences, 41*(A), 6–25. https://doi.org/10.1016/j.langsci.2013.08.003

Tan, Y., & Jehom, W. J. (2024). The function of digital technology in minority language preservation: The case of the Gyalrong Tibetan language. *Preservation, Digital Technology & Culture, 53*(3), 165–177. https://doi.org/10.1515/pdtc-2024-0021

Tlepbergen, D., Akzhigitova, A., & Zabrodskaja, A. (2022). Language-in-education policy of Kazakhstan: Post-pandemic technology enhances language learning. *Education Sciences, 12*(5), Article 311. https://doi.org/10.3390/educsci12050311

*Tlepbergen, D., Akzhigitova, A., & Zabrodskaja, A. (2023). Bottom-up approach to language policy and planning in Kazakhstan. *Societies, 13*(2), Article 43. https://doi.org/10.3390/soc13020043

*Tlepbergen, D., Akzhigitova, A., & Zabrodskaja, A. (2025). Kazakh-English bilingualism in Kazakhstan: Public attitudes and language practices. *Languages, 10*(5), Article 102. https://doi.org/10.3390/languages10050102

Utegenova, A., Kurmanova, B. Z., Abdullina, N. U., Koshimova, B., Otarova, A., Zhetkizgenova, A., Batinova, A. S., & Muratbe, B. K. (2024a). Code-switching in a multilingual classroom in Kazakhstan: A university students' perspective. *Forum for Linguistic Studies, 6*(6), 371–380. https://doi.org/10.30564/fls.v6i6.7338

Utegenova, A., Utegenova, K. T., Umarova, G. S., Utepova, N. D., Yergaliyeva, Z. K., Nurtazina, N. K., & Doskeyeva, M. Y. (2024b). Translingual practices in the linguistic landscape in the western region of Kazakhstan. *Theory and Practice in Language Studies, 14*(2), 392–403. https://doi.org/10.17507/tpls.1402.10

Wang, G., Bahry, S. A., & An, W. (2024). Minority language revitalization and social media through the lens of Covid-19 in Yunnan and Gansu, western China. *Journal of Multilingual and Multicultural Development, 45*(6), 2129-2151. https://doi.org/10.1080/01434632.2022.2042541

*Yelubayeva, P., Khamidova, A., Berkinbayeva, G., & Avakova, R. (2025). Addressing language education challenges in Kazakhstan for sustainable development. *European Journal of Language Policy, 17*(1), 65–96. https://doi.org/10.3828/ejlp.2025.5

*Yeskeldiyeva, B. Y., & Tazhibayeva, S. Z. (2015). Multilingualism in modern Kazakhstan: New challenges. *Asian Social Science, 11*(6), 56–63. https://doi.org/10.5539/ass.v11n6p56

*Yessenbekova, U., Aldabergenova, Z., Mamankul, A., Rakhymbayev, N., & Kutym, B. (2025). The impact of the digital format on Kazakh youth's information consumption and reading culture. *Publishing Research Quarterly, 41*(1), 113–131. https://doi.org/10.1007/s12109-024-10012-6

*Yessenbekova, U., Syzdykova, A., & Kutym, B. (2024). Demand for TV content in a national language: The case of Kazakhstan. *International Journal of Innovative Research and Scientific Studies, 7*(3), 951–964. https://doi.org/10.53894/ijirss.v7i3.2984

Zaugg, I. A., Hossain, A., & Molloy, B. (2022). Digitally-disadvantaged languages. *Internet Policy Review, 11*(2). https://doi.org/10.14763/2022.2.1654

Zhang, L., Huang, G., Li, Y., & Bao, S. (2021). Quantitative research methods of linguistic niche and cultural sustainability. *Sustainability, 13*(17), Article 9586. https://doi.org/10.3390/su13179586

*Zharkynbekova, S., Ayupova, G., Galiyeva, B., Shakhputova, Z., & Zabrodskaja, A. (2025). Ethno-linguistic identity of Kazakhstani student youth in modern multinational context of Kazakhstan: Sociolinguistic analysis of empirical research. *Languages, 10*(2), Article 33. https://doi.org/10.3390/languages10020033

Zheng, D., Cowley, S. J., & Nuesser, M. (2024). The new ecolinguistics: Learning as languaging with digital technologies. *Frontiers of Digital Education, 1*(1), 109–119. https://doi.org/10.1007/s44366-024-0026-7

APPENDIX

Table 1.

Descriptive Profile of Studies Included in the Meta-Analytic Synthesis

Variable	Category	k	N	%
Publication Year	2015–2019 (Early)	5	29,312	22.7
	2021–2022 (Mid)	3	196,806	13.6
	2023–2024 (Recent)	3	911	13.6
	2025 (Current)	11	28,121	50.0
Data Collection Tools	Survey/Questionnaire	12	5,768	54.5
	Automated/API Scraping	7	246,382	31.8
	Archival/Official Statistics	2	1,800	9.1
	Corpus/Discourse Analysis	1	1,200	4.5
Sample Characteristics	University/Higher Ed Students	6	3,762	27.3
	Secondary School Students	1	76	4.5
	General Population/Mixed Age	6	3,611	27.3
	Digital Content/Corpus Data	8	247,582	36.4
	Educators/Stakeholders	1	119	4.5
Language Ecology Context	Language Planning/Policy	5	3,210	22.7
	Identity & Multilingualism	6	3,118	27.3
	Digital Content & NLP	7	246,382	31.8
	Media Ecosystem	4	2,440	18.2
Sustainability Type	Functional/Domain Sustainability	4	98,582	18.2
	Technical/NLP Sustainability	6	149,576	27.3
	Educational Sustainability	4	2,031	18.2
	Socio-cultural/Identity Sustainability	6	4,413	27.3
	Media/Information Sustainability	2	548	9.1
Digital Media Integration	Social Media Platforms	7	124,340	31.8
	Automated Monitoring Systems	2	124,042	9.1
	Survey via Digital Tools	4	2,734	18.2
	Traditional/Broadcast Media	3	2,048	13.6
	E-Learning/Academic Digital	6	1,986	27.3

k = number of studies; N = total number of observations or units of analysis.

Appendix 1.

Summary of Included Studies, Effect Sizes, and Moderator Codes

Study	N	Statistical Data Type	r	Data Collection	Sample Type	Ecology Context	Sustainability Type
Abiyeva (2015)	1,500	Frequency/Proportion (χ^2)	−.14	Archival/Official Statistics	General Population	Language Planning/Policy	Functional/Domain
Baimanova et al. (2025)	692	R^2, β Coefficient	.70	Survey/Questionnaire	University Students	Media Ecosystem	Educational
Bolatbek et al. (2025)	10,190	Accuracy, AUC	.96	Automated/API Scraping	Digital Content	Digital Content & NLP	Technical/NLP
Fedorova (2025)	800	Pearson r (direct)	.75	Survey/Questionnaire	University Students	Identity & Multilingualism	Socio-cultural/Identity
Kuzhabekova (2019)	76	Proportion (%)	.54	Survey/Questionnaire	Secondary School	Identity & Multilingualism	Functional/Domain
ENU et al. (2023)	240	Mean Score (M)	.01	Survey/Questionnaire	General Population	Identity & Multilingualism	Socio-cultural/Identity
Mongilyova (2015)	200	Functionality Coefficient (K)	.82	Survey/Questionnaire	General Population	Identity & Multilingualism	Functional/Domain
Mussiraliyeva et al. (2022)	96,000	Accuracy, AUC	1.00	Automated/API Scraping	Digital Content	Digital Content & NLP	Technical/NLP
Mussiraliyeva et al. (2025)	7,150	F1-Score, Accuracy	.86	Automated/API Scraping	Digital Content	Digital Content & NLP	Technical/NLP
Mutanov et al. (2021)	96,806	Accuracy Rate	.82	Automated/API Scraping	Digital Content	Digital Content & NLP	Functional/Domain
Myrzakhmetov et al. (2018)	27,236	Accuracy Gain (Δ)	.71	Automated/API Scraping	Digital Content	Digital Content & NLP	Technical/NLP
Ordabayeva et al. (2025)	1,200	Frequency Distribution	−.24	Corpus/Discourse	Digital Content	Identity & Multilingualism	Socio-cultural/Identity
Rakhimova et al. (2022)	4,000	Accuracy/Success Rate (%)	.76	Automated/API Scraping	Digital Content	Digital Content & NLP	Technical/NLP
Sadulova et al. (2025)	920	Success Proportions (%)	.66	Survey/Questionnaire	University Students	Language Planning/Policy	Educational

Sambetbayeva et al. (2025)	5,000	Cohen's Kappa (κ)	.81	Automated/API Scraping	Digital Content	Digital Content & NLP	Technical/NLP
Tlepbergen et al. (2023)	371	Usage Proportion (%)	−.72	Survey/Questionnaire	General Population	Language Planning/Policy	Socio-cultural/Identity
Tlepbergen et al. (2025)	1,000	Chi-square (χ^2)	.25	Survey/Questionnaire	General Population	Identity & Multilingualism	Socio-cultural/Identity
Yelubayeva et al. (2025)	119	Perception Proportion (%)	.40	Survey/Questionnaire	Educators/Stakeholders	Language Planning/Policy	Educational
Yeskeldiyeva et al. (2015)	300	Proportion Comparison (%)	.57	Survey/Questionnaire	University Students	Language Planning/Policy	Educational
Yessenbekova et al. (2024)	300	Audience Reach/Share (%)	.11	Archival/Official Statistics	General Population	Media Ecosystem	Media/Information
Yessenbekova et al. (2025)	248	Information Source Proportion	.40	Survey/Questionnaire	University Students	Media Ecosystem	Media/Information
Zharkynbekova et al. (2025)	802	Pearson r (direct)	.24	Survey/Questionnaire	University Students	Identity & Multilingualism	Socio-cultural/Identity

r = converted or directly reported effect size in metric Pearson's r; N = sample size or number of units of analysis. Effect sizes were converted from chi-square, R^2, accuracy/AUC metrics, proportions, and functionality coefficients using procedures outlined in the Method section. Negative r values indicate associations in the direction of reduced sustainability (or Russian language dominance in the ecology measured).

Appendix 2.

Forest Plot

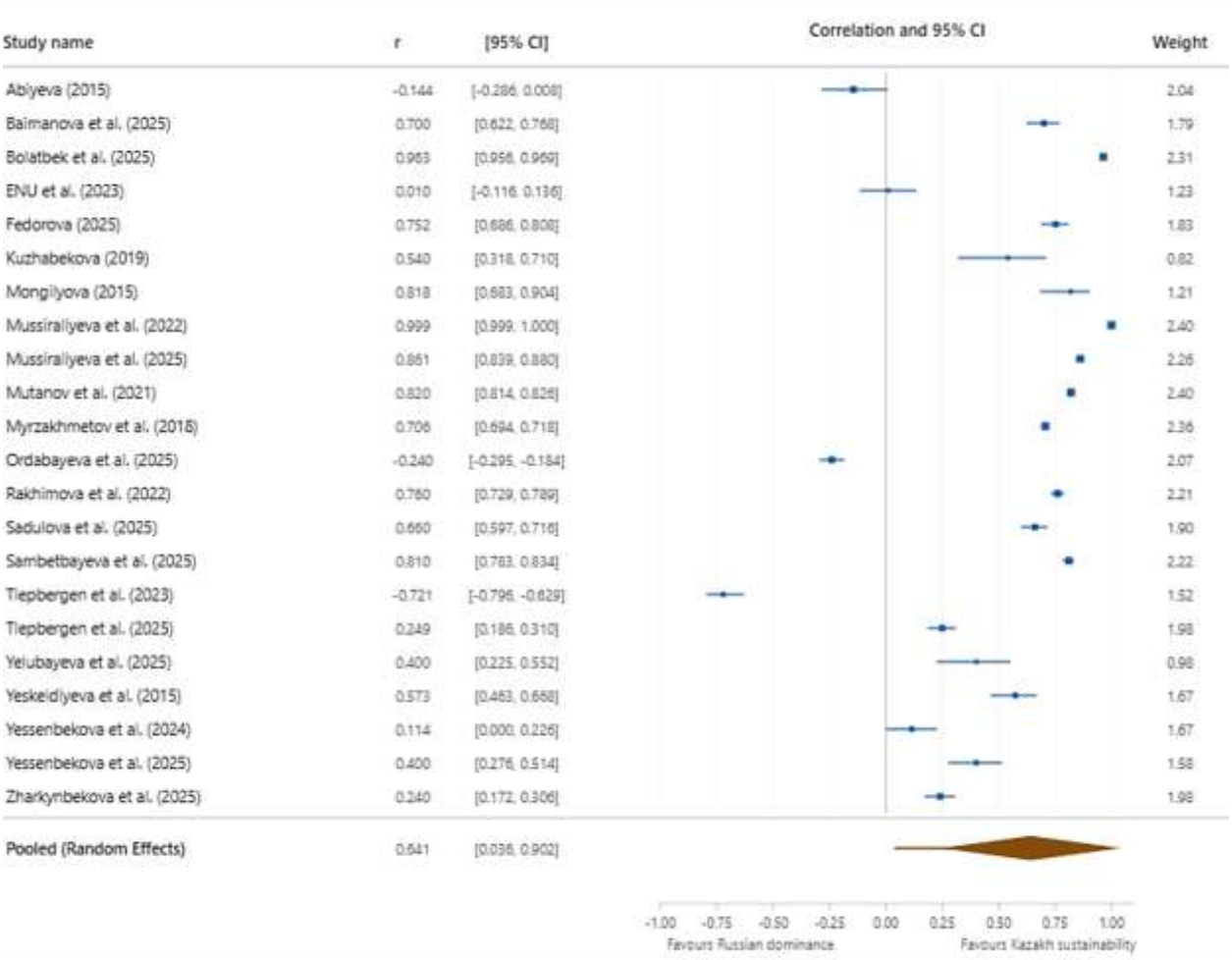


Table 3.

Meta-Analytic Results: Overall Effect and Moderator Analyses

Variable	k	N	r	95% CI LL	95% CI UL	Q	Qb	I^2
Overall	22	255,150	.64*	.04	.90	535,831.12*	—	100.0%
Moderator: Publication Year							8.19*	
2015–2019 (Early)	5	29,312	.56	.08	.83	—		99.7%
2021–2022 (Mid)	3	196,806	.97*	−.21	1.00	—		100.0%
2023–2024 (Recent)	3	911	−.26	−.73	.39	—		99.0%
2025 (Current)	11	28,121	.63*	.32	.82	—		99.9%
Moderator: Data Collection Tools							10.67*	
Survey/Questionnaire	12	5,768	.43*	.16	.64	—		99.2%
Automated/API Scraping	7	246,382	.93*	.35	.99	—		100.0%
Archival/Official Statistics	2	1,800	−.02	−.27	.23	—		94.0%
Corpus/Discourse Analysis	1	1,200	−.24	−.34	−.14	—		0.0%
Moderator: Sample Characteristics							7.97*	
University/Higher Ed Students	6	3,762	.58*	.40	.72	—		98.2%
Secondary School Students	1	76	.54	.44	.64	—		0.0%
General Population/Mixed Age	6	3,611	.08	−.32	.45	—		99.2%
Digital Content/Corpus Data	8	247,582	.89*	.20	.99	—		100.0%
Educators/Stakeholders	1	119	.40	.30	.50	—		0.0%
Moderator: Language Ecology Context							4.75	
Language Planning/Policy	5	3,210	.16	−.41	.64	—		99.6%
Identity & Multilingualism	6	3,118	.49*	.19	.71	—		98.8%
Digital Content & NLP	7	246,382	.93*	.35	.99	—		100.0%
Media Ecosystem	4	2,440	.28	−.28	.70	—		99.5%
Moderator: Sustainability Type							14.34*	

Functional/Domain Sustainability	4	98,582	.60	−.14	.91	—		99.9%
Technical/NLP Sustainability	6	149,576	.94*	.20	1.00	—		100.0%
Educational Sustainability	4	2,031	.61*	.51	.69	—		88.0%
Socio-cultural/Identity Sustainability	6	4,413	.06	−.39	.48	—		99.6%
Media/Information Sustainability	2	548	.26	−.03	.52	—		92.2%
Moderator: Digital Media Integration							20.53*	
Social Media Platforms	7	124,340	.90*	.10	.99	—		100.0%
Automated Monitoring Systems	2	124,042	.77*	.63	.86	—		99.9%
Survey via Digital Tools	4	2,734	.33*	.01	.59	—		98.7%
Traditional/Broadcast Media	3	2,048	.13	−.21	.43	—		97.4%
E-Learning/Academic Digital	6	1,986	.42	−.19	.80	—		99.4%

k = number of effect sizes; N = total sample size or units of analysis; r = weighted mean effect size (Pearson's r); CI LL/UL = 95% confidence interval lower/upper limits; Q = total heterogeneity statistic; Qb = between-group heterogeneity statistic; I^2 = percentage of total variability attributable to between-study heterogeneity. *p < .05.