



Article

Measuring Curriculum-Assessment Regulation Coherence: Learning Outcomes-Criteria Coupling, Thematic Emphasis, and Global Discourse Alignment in Kazakhstan and Finland

Aliya Assylbekova

Center for pedagogical measurements, Nazarbayev Intellectual Schools

Email: assylbekova_a@cpn.nis.edu.kz

Aiman Ali

National Scientific and Practical Institute for Child Well-Being

Email: aliyevaaiman7@gmail.com

Abstract

National curricula regulate educational expectations and assessment through their statement genres and the linkage between expectations and criteria. This article compares Kazakhstan and Finland through cross-lingual computational document analysis of formal curriculum and assessment-regulation texts; it does not examine classroom assessment or implementation. Atomic normative statements were embedded and compared using directional mean-best similarity, threshold coverage, one-to-one matching, and size-matched bootstrapping. Both systems show strong internal textual coherence through different architectures: Kazakhstan exhibits close outcomes-to-criteria coupling, whereas Finland links a broader objective layer to criteria. Cross-system convergence is higher for expectations/objectives than for criteria, indicating that evaluative phrasing is more system-specific than curricular intentions. Finland shows higher semantic alignment with literacy, civic and social learning, sustainability, and inclusion frames, whereas Kazakhstan shows higher semantic alignment with digital-competence language in the assessment-related



©2026 Aliya Assylbekova, Aiman Ali. This is an Open Access article distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), allowing third parties to copy and redistribute the material in any medium or format and to remix, transform, and build upon the material for any purpose, even commercially, provided the original work is properly cited and states its license.

layer. The findings show that global curriculum discourses travel more readily at the level of stated aims than assessment rules, where evaluative frameworks remain less portable.

Keywords: national curricula, learning outcomes, global competencies, STEM, 21st-century skills

Introduction

National curricula and associated assessment requirements go beyond specifying intended learning outcomes: they establish a governance architecture that defines legitimate knowledge, assigns evaluative authority, and regulates how evidence of learning is produced and interpreted. Over the last two decades, international organizations have increased convergence pressures by endorsing competency-oriented curricula, coherence across curriculum–teaching–assessment, and stronger international comparability. However, convergence in language does not necessarily translate convergence in regulatory design. As policy is contextualized into national instruments, systems may embrace similar thematic goals while encoding them through different genres (e.g., outcomes, objectives, criteria) and different relational logics between expectations and assessment rules (OECD, 2025; Steiner-Khamsi, 2004).

The article considers how Kazakhstan and Finland articulate educational expectations and regulate assessment in lower-secondary policy texts, with particular attention to convergence or divergence across layers of official documentation. The object of comparison is the intended curriculum and its assessment-regulation language, not classroom assessment practice, teacher judgement, examination administration, nationally administered testing, or student performance. The contrast is theoretically significant. Both systems participate in global curriculum discourses while representing different governance traditions: Kazakhstan relies more on a centralized, outcome-led regulatory model, whereas Finland is characterized by a professionalized, criteria-anchored curriculum and assessment culture. Rather than assuming one model as preferable, we treat the comparison as a way to examine how different systems maintain internal coherence and where attempts at cross-national transfer are likely to break down (Priestley et al., 2021; van den Akker, 2003).

The article is guided by five research questions that connect directly to comparative and international education discussions on curriculum governance, policy translation, and the portability of assessment models across national contexts. RQ1 focuses on how expectations are formulated (Kazakhstan’s learning outcomes versus Finland’s learning objectives) and how assessment regulation is articulated and connected to expectations. RQ2 explores convergence and divergence patterns across expectation/objective and criteria layers. RQ3 investigates what these patterns suggest about how global curriculum ideas are recontextualized into different national governance architectures. RQ4 compares thematic priorities, including STEM, languages/literacy, civic/social competencies and digital literacy. RQ5 evaluates alignment

with global educational discourses, particularly 21st-century competencies and sustainable development, and explores implications for cross-national policy learning.

For clarity, three terms are used consistently throughout the article. An atomic statement is the smallest stand-alone policy segment that expresses one educational expectation or assessment rule, such as a numbered outcome, objective, criterion, or performance descriptor. Expectation statements refer to Kazakhstan's stated learning outcomes and Finland's learning objectives, although these genres are not identical in regulatory function. Criteria refer to formal assessment-regulation language, including criteria, performance-level descriptors, and grading formulations. We operationalize coupling as the degree of semantic linkage between these statement sets. A stronger linkage means that statements in one layer have close semantic matches in the other layer; a weaker linkage means that a larger share of statements cannot be closely matched at the selected threshold. These terms describe textual alignment in policy documents only and should not be read as direct evidence that assessment is better, fairer, or more effectively enacted in practice.

Methodologically, the analysis compares Kazakhstan's expectation statements ($n = 25$) and criteria ($n = 56$) with Finland's objectives ($n = 210$) and criteria ($n = 62$). Because the corpora differ substantially in size and granularity, the results combine directional mean-best similarity, threshold coverage, one-to-one optimal matching, and size-matched bootstrapping. The methods section provides the technical details; the introduction uses these terms only to define the comparison in accessible terms.

This study maps shared curricular and evaluative space across national policy texts. It provides a reproducible statement-level procedure for estimating curriculum–assessment-regulation coupling. The procedure is documented through an anonymized peer-review repository containing derived corpora, code, similarity matrices, and outputs. Rather than producing a universal measure of curriculum quality, it identifies patterns of textual linkage that remain traceable to original policy segments. The article first reviews the conceptual and methodological literature, then describes the corpus, measurement strategy, and safeguards, and finally presents results and discusses their implications for policy translation and cross-national learning (Baden et al., 2022; Grimmer et al., 2022).

Literature review

Global curriculum agendas and convergence pressures

In contemporary international policy discourse, international organizations increasingly promote national curricula through the language of competency-oriented goals, measurable learning, and system comparability. The OECD's Education 2030/2040 work casts curricula as levers for developing

“transformative competencies” and learner agency, while curriculum design itself is framed as political and value-laden rather than purely technical (Karseth et al., 2024; OECD, 2018; 2020a). UNESCO likewise regards curriculum as a core policy instrument for addressing current social and economic challenges, arguing that curriculum development should be iterative and evidence-informed, with particular attention to coherence across curriculum design, implementation, and evaluation (UNESCO International Bureau of Education [IBE], 2017).

Scholarship on world society, policy borrowing, and curriculum governance explains how global agendas can create convergence pressure through shared vocabularies (e.g., competence, agency, inclusion, sustainability), common reference points (such as international large-scale assessments and digital-competence and sustainability frameworks), and increasingly comparable curriculum architectures (Meyer et al., 1997; OECD, 2020a; 2020b; UNESCO IBE, 2017). However, the same literature cautions that convergence in discourse can coexist with divergence in the ways curricula are operationalized, regulated, and enacted. Global policy “models” are seldom imported intact; instead, they are selectively translated into domestic policy language, and adapted to institutional histories, governance arrangements, and professional cultures (Meyer et al., 1997; Pettersson et al., 2017; Steiner-Khamsi, 2004). Consequently, Kazakhstan and Finland are considered here as a theoretically informative contrast in curriculum governance and policy translation, rather than as a normative ranking of educational systems.

National curriculum frameworks as governance instruments

A national curriculum framework operates as a governance instrument in at least three dimensions. First, it specifies legitimate knowledge and skills by defining what constitutes valued learning. Second, it distributes decision-making authority by allocating interpretive discretion across system levels (central agencies, municipalities, schools, teachers). Third, it structures accountability by determining what is assessable and the terms under which the judgements should be made (Ball, 1993; UNESCO IBE, 2017).

Curricula therefore should be understood as policy texts that simultaneously articulate educational intentions and shape the space of pedagogical action through their language, structure, and relationship to assessment regulation (Ball, 1993; Priestley et al., 2021). This perspective underscores the distinction between an intended curriculum (policy expectations), the implemented/enacted curriculum (realized through local interpretation and practice), and the attained curriculum (learning as evidenced through assessment), a distinction that is essential for interpreting cross-national comparisons without conflating text with practice (van den Akker, 2003).

This distinction also sets the normative baseline for the comparison. The article does not begin with a single universal definition of a ‘good’ curriculum. It treats curriculum quality as contested and system-dependent,

while using constructive alignment as a limited analytic principle: in a formally coherent curriculum, stated aims and assessment-relevant criteria should be legibly connected. Different political and professional traditions can distribute this connection differently. More centralized systems may codify expectations and criteria more explicitly in national documents, whereas more decentralized or professionally mediated systems may leave greater interpretive work to municipalities, schools, and teachers. The empirical question is therefore not which system is normatively superior, but how each system textualizes the relationship between aims and judgement (Biggs, 1996; van den Akker, 2003).

Comparative work treats coherence as the legibility of relations among aims, content, pedagogy, and assessment across governance levels. Formal textual coupling does not prove enacted coherence: tightly worded policy may still be reinterpreted locally, while broader texts may be enacted coherently through professional norms. Statement-level comparison therefore identifies the architecture of the intended curriculum rather than implementation quality (Priestley et al., 2021; UNESCO IBE, 2017; van den Akker, 2003).

Policy borrowing rarely transfers complete institutional designs. Competence, agency, inclusion, and assessment-for-learning often travel as legitimating vocabularies but are reassembled in national governance contexts. Assessment criteria, grading language, and performance descriptors are particularly embedded in accountability traditions and professional understandings of judgement. Statement-level comparison therefore distinguishes shared aspiration from shared regulation (Ball, 1993; Meyer et al., 1997; Pettersson et al., 2017; Steiner-Khamsi, 2004).

Taken together, these perspectives position curriculum expectations and assessment criteria as governance texts rather than neutral lists of statements. Their relationship makes visible how national systems organize educational judgement and provides the basis for the statement-level comparison used in this study to distinguish shared curricular aspirations from nationally specific regulatory forms.

Outcomes, objectives, and regulatory consequences

A key analytic concern is how curricular expectations are expressed and what these formulations mean for assessment regulation. In policy and qualifications frameworks, learning outcomes are typically presented as statements specifying what learners are expected to demonstrate—knowledge, skills, and competences that can, in principle, be assessed (Adam, 2004; European Commission, 2015). When curricula foreground outcome statements framed around observable performances, they may foster transparency and facilitate monitoring, yet they may also deepen accountability logics by prioritizing what can be measured and compared (Clark & Hsu, 2023).

By comparison, goals or objectives often operate as instructional steering devices: they guide teachers' planning, organize learning progressions, and inform task design without necessarily expressing end-state performance in a fully measurable form (Siegesmund & Smilan, 2023). This difference carries regulatory consequences. Outcome-dominant curricula may implicitly shift assessment design to downstream instruments such as tests, rubrics, reporting systems, since outcomes invite measurement. Objective-and-criteria formulations, however, can embed assessment expectations directly within the curriculum text by pairing objectives with performance descriptors or criteria, thereby specifying not only "what to learn" but also "how quality is judged" (Ventista & Brown, 2023).

Thematic salience in curricula and global frames

Thematic priorities are also central to cross-national comparisons. Research on policy-text shows that curricular "salience" is not simply whether a topic appears, but how strongly it is foregrounded, how it is framed, and how directly it is connected to assessable expectations. Framing theory conceptualizes salience as selective emphasis—foregrounding problem definitions, causal interpretations, moral evaluations, or solution frames (Entman, 1993). Applied to curricula, framing points to how domains such as languages/literacy, civic and social learning, digital competence, or sustainability are positioned: as core aims, cross-cutting competencies, or marginal add-ons.

Substantively, recent international agendas elevate four thematic clusters that are especially relevant for Kazakhstan–Finland comparison. First, digital competence has become a cross-system priority, increasingly supported by shared reference frameworks such as DigComp 2.2 (Vuorikari et al., 2022). Second, multiliteracy and language education are widely treated as foundational for equity and participation, with Nordic curriculum debates often emphasizing broad literacy practices rather than narrow language proficiency. Third, civic and citizenship education is framed internationally as a pathway to inclusive, democratic engagement (Borhan, 2025). Fourth, education for sustainable development (ESD) has been strengthened through the ESD for 2030 agenda, which calls for embedding sustainability competencies across curricula and assessment (UNESCO, 2020).

Computational curriculum analysis and measurement validity

A growing methodological literature argues that computational "text-as-data" approaches can support curriculum comparison by enabling scalable measurement of (i) curriculum–assessment coupling and (ii) thematic alignment, provided that measurement validity and auditability are treated as design constraints rather than afterthoughts (Grimmer et al., 2022; Baden et al., 2022). For curriculum documents, this entails (a) keeping an auditable link from quantitative scores back to specific text segments, (b) applying sampling or matching strategies to avoid corpus-size artefacts, and (c) triangulating automated similarity measures

with qualitative interpretation of policy language (Grimmer et al., 2022). Accordingly, the documents are treated as structured corpora that encode educational expectations, assessment regulation, thematic salience, and policy translation (OECD, 2020b; UNESCO IBE, 2017).

Research gap and research questions

The literature points to a clear gap in comparative curriculum research. Existing scholarship discusses global curriculum agendas, outcomes-based reform, curriculum coherence, and thematic framing, but far fewer studies compare, at statement level and across languages, how national systems formulate expectations, how they anchor assessment criteria to those expectations, and how global discourse is translated across different layers of policy text. This gap matters for comparative and international education because claims of convergence often rest on shared vocabulary rather than on directly comparable regulatory wording.

The study addresses this gap through the following research questions:

1. How are educational expectations formulated in Kazakhstan's and Finland's curricula (learning outcomes vs learning objectives), and what kinds of regulatory logics do these formulations imply?
2. How is formal assessment regulation specified in policy texts (e.g., criteria, performance levels/descriptors, grading language), and how tightly is it linked to the stated expectations?
3. What patterns of convergence and divergence across expectation/objective and criteria layers reveal the translation of global curriculum ideas into different national governance architectures?
4. How do thematic priorities (e.g., STEM, languages/literacy, civic education, digital literacy, sustainability, inclusion) differ in emphasis across Kazakhstan's and Finland's curricula?
5. To what extent do Kazakhstan's and Finland's curricula align with global educational discourses, including 21st-century competencies and sustainable development, and what do these alignments imply for cross-national policy learning and transfer?

Theoretical framework

This study uses four complementary interpretive lenses. Together they distinguish description from interpretation: the empirical analysis measures textual coupling and thematic/discursive alignment, while the framework explains how those patterns may be read in comparative terms.

First, constructive alignment provides the main lens for interpreting curriculum–assessment–regulation coupling. It treats coherence as the degree to which intended learning and assessment provisions are connected in the formal curriculum, making it useful for reading strong or weak linkage between expectations and criteria while avoiding claims about enacted assessment practice (Biggs, 1996; Hailikari et al., 2022; Hamdoun, 2023).

Second, comparative education and policy-translation perspectives help interpret convergence and divergence across Kazakhstan and Finland. These perspectives direct attention to how similar global agendas may be recontextualized through different governance arrangements, institutional histories, and nationally embedded regulatory traditions (Meyer et al., 1997; Pettersson et al., 2017; Steiner-Khamsi, 2004).

Third, framing and thematic salience guide interpretation of relative emphasis across domains such as STEM, literacy, civic/social learning, digital competence, and sustainability. Here, salience is understood not simply as topic presence but as the selective foregrounding of themes in expectation statements and assessment-relevant language (Entman, 1993; Porter, 2002).

Fourth, alignment with OECD and UNESCO discourses is interpreted as policy translation rather than literal adoption. National texts may reproduce, adapt, or omit global competency- and sustainability-oriented frames, and the relevant question is whether such frames appear only as aspirations or are also coupled to evaluative language (OECD, 2020b; UNESCO, 2020).

Taken together, these lenses support interpretation of the findings at three levels: curriculum–assessment coupling, thematic emphasis, and the selective translation of global discourse into nationally specific curriculum and assessment design.

Methods

We use cross-lingual computational text analysis of national curriculum and assessment-regulation documents of Kazakhstan and Finland, using multilingual sentence embeddings and similarity-based matching to estimate coupling and alignment (Creswell & Creswell, 2022; Grimmer et al., 2022). Curriculum and assessment texts are treated as normative regulatory artifacts. The analysis is therefore confined to the formal textual encoding of intended learning and assessment rules in official documents; it does not examine classroom enactment, teacher judgement, assessment tasks, student performance, or the administration of national examinations. The tightness of coupling between (i) expectation statements (outcomes/objectives) and (ii) assessment criteria is operationalized via cross-lingual semantic similarity, with each quantitative result traceable to the original text segments (Morgan, 2022).

Data sources and corpus construction

This study analyzed official lower-secondary curriculum/standards and assessment-regulation texts from Kazakhstan and Finland: the Kazakhstani criteria for assessing students' academic achievement at the level of basic secondary education (Орта, техникалық және кәсіптік, орта білімнен кейінгі білім

алушылардың білімін бағалау өлшемшарттарын бекіту туралы, 2016), the State standards for lower-secondary education (Мектепке дейінгі тәрбие мен оқытудың, бастауыш, негізгі орта, жалпы орта, техникалық және кәсіптік, орта білімнен кейінгі білім берудің мемлекеттік жалпыға міндетті стандарттарын бекіту туралы, 2022), and the National core curriculum for primary and lower secondary education (Opetushallitus/Utbildningsstyrelsen, 2016). The analysis was conducted on the original-language documents rather than on English translations. The Kazakhstani curriculum and assessment-regulation texts were processed in Russian, while the Finnish curriculum and assessment-regulation texts were processed in Finnish. No full English translation was used as an input for embedding, matching, or cosine-similarity estimation. Cross-lingual comparability was instead addressed computationally by using the multilingual Sentence-Transformers model paraphrase-multilingual-mpnet-base-v2, which represents texts from different languages in a shared semantic vector space. This design allowed Russian and Finnish statements to be compared directly at the embedding level. The computational stage did not require prior translation of Finnish segments into English; Finnish and Russian segments were treated as original-language textual data and compared through their multilingual embeddings.

English translations were used only after the computational analysis had been completed, for reporting and interpretive transparency. Specifically, translated excerpts were used to make illustrative matched pairs, top-match examples, and audit samples accessible to readers who do not read Russian or Finnish. These translations were not part of the analytical corpus and did not affect the calculation of similarity scores. They were used to support reader-facing interpretation of selected examples, not to create the similarity matrix. Post hoc English reporting translations were generated with machine-translation tools used only for reporting support, not by a human translator and not as analytical input.

Unit of analysis and unitization rules

The analysis was limited to sections specifying educational expectations (KZ learning outcomes/expectations; FI learning objectives) and assessment regulation (criteria, performance-level descriptors, grading language).

The unit of analysis is the atomic normative statement: the smallest stand-alone text segment that expresses an educational expectation or an assessment rule. This unit was used in both countries because the documents differ in format and length, but in each case the normative function is carried by short list items, numbered statements, table cells, or criterion descriptors rather than by entire sections or pages. Using this level makes the comparison substantively comparable: each item represents a discrete claim about what students should learn or how performance should be judged (see Table 1) (Morgan, 2022).

Texts were parsed and segmented using structural cues (bullets, numbering, rubric headings, performance-level labels). To prevent under-segmentation, long multi-part bullets were split only when they contained clearly separable obligations or descriptors that could stand on their own without loss of normative meaning. To prevent over-segmentation, phrases were not split when they depended grammatically or semantically on the same statement frame, rubric label, or performance descriptor. Length thresholds (minimum 25 characters for expectations/objectives; minimum 15 for criteria) and a 700-character cap were applied to exclude headings and overly long prose blocks.

Ambiguous cases were resolved conservatively. When numbering or layout cues alone were insufficient, the segment was checked against surrounding syntax and section function; items were kept intact when splitting would have created fragments without independent normative meaning. Each retained segment was assigned a stable audit identifier and linked to its source-document identifier, corpus file, and row position. This enables each row and column of the similarity matrices, as well as the reported matched pairs, to be traced back to the corresponding original-language statement included in the reproducibility package (Morgan, 2022).

Statement-level corpus sizes were as follows:

KZ expectations: n = 25

FI objectives: n = 210

KZ assessment criteria: n = 56

FI assessment criteria: n = 62

Table 1. Illustrative unitization examples from the analytic corpus

Country/layer	Excerpt from corpus	Analytic treatment
Kazakhstan expectation	“применяет математические знания для решения практических задач” [applies mathematical knowledge to solve practical tasks]	Expectation statement: intended capability.
Kazakhstan criterion	“обучающийся понимает ... тексты ... создает ... высказывания ... аргументирует ... мнение” [understands texts, creates statements, argues an opinion]	Criterion statement: evidence of achievement.

Country/layer	Excerpt from corpus	Analytic treatment
Finland objective	“Oppilaat oppivat arvioimaan, ylläpitämään ja kehittämään omaa fyysistä toimintakykyään.” [Students learn to assess, maintain, and develop physical capacity.]	Objective statement: intended learning.
Finland criterion/grading language	“Oppilas saa arvosanan kahdeksan (8), mikäli hän osoittaa ... kriteerien määrittämää osaamista.” [Grade 8 is awarded when the pupil demonstrates competence defined by the criteria.]	Grading/criterion language: performance judgement.

Computational coupling analysis: regulatory form and linkage

This study implements comparative document analysis as text-as-data, extracting *atomic statements* from official curriculum and assessment documents and quantifying how tightly expectations are linked to criteria across systems. Kazakhstan’s expectations and Kazakhstan’s assessment criteria are extracted from separate DOCX files; Finland’s objectives and criteria are extracted from a curriculum PDF using PyMuPDF block ordering, with an optional pdfplumber fallback. Statements are segmented with rule-based splitting for numbered and bulleted lists and table cells, then cleaned by removing headings, normalizing whitespace, and de-duplicating.

Granularity is controlled with length guards that match the script: expectations and objectives must be at least 25 characters, criteria at least 15 characters, and all items are capped at 700 characters to prevent long paragraphs from dominating similarity.

For cross-lingual comparison, each statement is embedded with the Sentence-Transformers multilingual model paraphrase-multilingual-mpnet-base-v2 (normalized embeddings), and pairwise cosine similarities form a similarity matrix S , where each cell s_{ij} ranges from -1 to 1 and higher values indicate greater semantic proximity between statement i and statement j . Directional mean-best similarity from corpus A to corpus B is calculated by selecting, for each statement in A, its highest similarity score in B and averaging these maxima. Thus, A to B and B to A can differ when one corpus is broader, more granular, or more repetitive than the other. Coverage at threshold τ reports the proportion of statements whose best match is at or above τ ; the primary threshold is $\tau = 0.50$, with sensitivity checks at 0.35 to 0.60. These numerical

indicators are interpreted as degrees of textual semantic alignment, not as statistical tests of student achievement or implementation quality.

To mitigate inflation from many-to-one matches, we compute a one-to-one matching score using Hungarian assignment when SciPy is available, otherwise a greedy fallback. To address corpus-size imbalance, we run a size-matched bootstrap (500 iterations; only when the matched size is at least 20), sub-sampling the larger corpus without replacement and reporting percentile intervals.

Robustness and validity safeguards

To reduce over-interpretation risks in computational text measures, we implement checks aligned with recommended validity priorities for computational text analysis (Baden et al., 2022). These include size-matched bootstrap confidence intervals, which subsample the larger corpus to the smaller corpus size and recompute metrics over 500 iterations with percentile-based 95% intervals, and an audit of top matches and a sample of low-similarity pairs to verify face validity and detect segmentation or ‘generic criterion’ artifacts (Grimmer et al., 2022; Morgan, 2022;). The audit was performed on the original segment IDs and post hoc English reporting translations; the translations served only to make examples inspectable and were not used to generate embeddings or scores. Sensitivity to alternative encoders remains an important avenue for further robustness testing (Ni et al., 2022).

Software and reproducibility

All preprocessing and analyses were implemented in Python with fixed random seeds and versioned outputs. The pipeline exports statement-level corpora, similarity matrices, metric summaries, bootstrap draws and confidence intervals, threshold-sensitivity tables, and top-match files. An anonymized reproducibility package is available for peer review¹. The package includes the statement-level corpora, analysis code, source-document metadata and official URLs, similarity matrices, metric summaries, bootstrap outputs, threshold-sensitivity tables, and reporting tables. Full official DOCX/PDF source documents are not redistributed; instead, the repository provides their official public locations and the derived statement-level corpora used in the analysis.

Ethics

All materials were publicly available policy and curriculum documents; the study involved no human participants, interventions, identifiable records, or personal data. Under the authors’ institutional ethics procedures, this documentary analysis did not require human-subject ethics approval. The study was

¹ Anonymized peer-review repository: https://osf.io/yrx8b/?view_only=9d114ef5ae0742b7a9e52417f4026121

conducted in accordance with the principles of reliability, honesty, respect, and accountability set out in the 2023 revised European Code of Conduct for Research Integrity (All European Academies - ALLEA, 2023). The data-handling design was also consistent with the Law on Personal Data and Their Protection (2013) in the Republic of Kazakhstan, as no personal data were collected, processed, or stored. The principal ethics risk was interpretive bias in cross-lingual comparison; it was mitigated through traceable IDs, match audits, sensitivity analyses, and transparent reporting of methodological limitations (Baden et al., 2022).

Results

The results show a consistent layer effect. Across the analyses, overlap is stronger within systems than across systems, and stronger at the expectations/objectives layer than at the criteria layer. The reported values are cosine-similarity summaries derived from text embeddings. A value closer to 1 indicates that statements are semantically closer in the model space; it does not mean that a curriculum is better or that assessment is more valid in practice. Directionality is substantively meaningful because it indicates whether one system's statements are broad enough to absorb another system's wording or narrow enough to exclude part of it.

Formally, mean best A to B is the average of the highest s_{ij} score obtained by each statement in corpus A when compared with all statements in corpus B. Coverage reports the share of statements whose best match exceeds the selected threshold τ . One-to-one matching forces unique pairings so that a single generic statement cannot serve as the best match for many statements. Size-matched bootstrapping repeatedly reduces the larger corpus to the size of the smaller corpus and recalculates the indicators, which helps assess whether the interpretation is driven by corpus imbalance.

Table 2. Summary of coupling indicators (primary results)

Comparison	nA	nB	mean best A→B	mean best B→A	harmonic symmetry	cov(A) ≥ 0.50	cov(B) ≥ 0.50	one-to- one mean
KZ								
expectations								
↔ FI	25	210	0.6565	0.5338	0.5888	0.9600	0.6381	0.6346
objectives								

Comparison	nA	nB	mean best A→B	mean best B→A	harmonic symmetry	cov(A) ≥ 0.50	cov(B) ≥ 0.50	one-to- one mean
KZ expectations ↔ KZ criteria	25	56	0.9479	0.7787	0.8550	1.0000	0.9643	0.9479
FI objectives ↔ FI criteria	210	62	0.8351	0.9401	0.8845	0.9952	0.9839	0.9352
KZ expectations ↔ FI criteria (cross-system extension)	25	62	0.5823	0.5091	0.5432	0.7600	0.5323	0.5367

Both systems are internally coherent, but they achieve coherence differently (Table 2). Kazakhstan's outcomes are tightly mirrored by its criteria, whereas Finland's criteria are anchored to a broader objective layer. This matters because it shows that strong curriculum–assessment coherence can coexist with different regulatory designs, and it sets up the directional asymmetry examined in the next table.

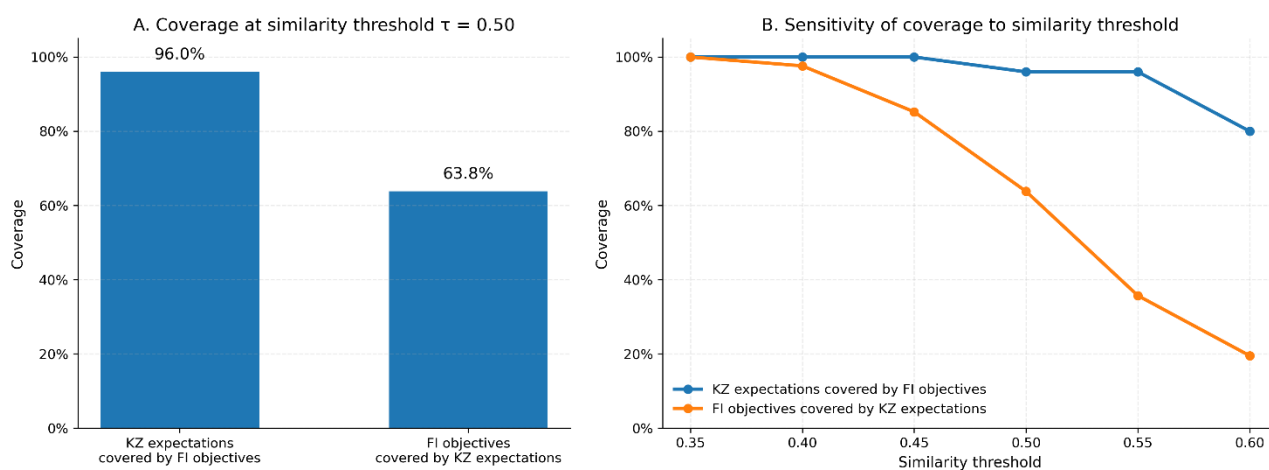
Table 3. Threshold sensitivity profile (coverage at stricter cutoffs)

Direction	cov ≥ 0.50	cov ≥ 0.55	cov ≥ 0.60
KZ expectations → FI objectives	0.9600	0.9600	0.8000
FI objectives → KZ expectations	0.6381	0.3570	0.1950

The asymmetry is not merely technical but structural (Table 3). Kazakhstan's compact expectation set fits relatively well within Finland's broader objective space, while many Finnish objectives do not compress cleanly into Kazakhstan's smaller set of expectations. In substantive terms, Finland's objective layer appears

broader and less compressible, whereas Kazakhstan's expectations operate as a more compact set of assessable formulations.

Figure 1. Asymmetric alignment between Kazakhstan's expectations and Finland's objectives



Note. Panel A shows coverage at the primary similarity threshold $\tau = 0.50$. Panel B shows how coverage changes across alternative similarity thresholds.

Table 4. Size-matched bootstrap estimates (conservative checks)

Comparison (size-matched setup)	Mean A→B	95% interval	Mean B→A	95% interval
KZ expectations → FI objectives (FI subsampled to 25)	0.5606	[0.5144, 0.6087]	—	—
FI objectives → KZ expectations (FI subsampled to 25)	—	—	0.5357	[0.5056, 0.5680]
KZ expectations → KZ criteria (KZ criteria subsampled to 25)	0.8561	[0.7876, 0.8955]	—	—
KZ criteria → KZ expectations (KZ criteria subsampled to 25)	—	—	0.7786	[0.7267, 0.8364]
FI objectives → FI criteria (FI subsampled to 25)	0.8343	[0.8096, 0.8590]	—	—

Comparison (size-matched setup)	Mean A→B	95% interval	Mean B→A	95% interval
objectives subsampled to 62)		0.8611]		
FI criteria → FI objectives (FI objectives subsampled to 62)	—	—	0.8713	[0.8419, 0.8919]
KZ expectations → FI criteria (FI criteria estimate under size control)	0.5337	[0.4930, 0.5704]	—	—

The conservative resampling checks do not alter the main interpretation (Table 4). The difference between Kazakhstan's tighter outcomes-to-criteria mirroring and Finland's broader objective-to-criteria anchoring remains visible even after size control.

Table 5. Theme scores by system (expectations/objectives layer)

Theme	Kazakhstan (KZ)	Finland (FI)
STEM	0.5548	0.5395
Language & literacy	0.3495	0.7168
Civic/social competencies	0.3143	0.5186
Digital literacy	0.3659	0.5025

Thematic overlap exists, but Finland gives greater textual prominence to literacy, civic/social learning, and digital themes (Table 5). The result is not a different curriculum core so much as a different pattern of emphasis across domains.

Table 6. Alignment with global discourses (reported scores)

Discourse	KZ expectations	FI objectives	KZ criteria	FI criteria
Critical thinking / 21st-century skills	0.4678	0.6239	—	—
Sustainability / SDG framing	0.3926	0.6846	0.3664	0.3655
Inclusion / equity	0.2742	0.5059	—	—
Digital competence	0.5320	0.4585	0.5705	0.4232

Global-discourse alignment is selective rather than uniform (Table 6). Finland is more strongly aligned with critical thinking, sustainability, and inclusion at the objective layer, while Kazakhstan stands out for more explicitly evaluative digital competence language.

Across the tables, one substantive contrast remains stable: Finland's objective layer is broader and less compressible, whereas Kazakhstan's outcomes are more tightly mirrored by criteria. Put differently, the two systems share substantial curricular space, but they encode that space through different relationships between broad aims and evaluative wording.

Discussion

This study identifies a clear layer effect in cross-national curriculum comparison. Both systems show strong within-system coupling, but cross-system similarity is stronger for expectations/objectives than for criteria. This pattern suggests that broadly stated curricular aims travel more readily across systems than evaluative formulations used to anchor judgement. For comparative and international education, the key implication is that claims of convergence depend on which layer of policy text is being compared (OECD, 2025).

A necessary qualification is that the word assessment in this article refers to formal assessment regulation in policy documents. It does not refer to observed classroom assessment, teacher feedback practices, school-based grading decisions, or national high-stakes testing as enacted. The findings therefore speak to the textual architecture that frames judgement, not to the empirical quality of assessment practice. This distinction is important because a tightly codified policy text may still be implemented unevenly, while a

less tightly codified text may be enacted coherently through strong professional norms (van den Akker, 2003; Priestley et al., 2021).

Outcomes versus objectives and directional asymmetry

Cross-system mapping is asymmetric. The results show that Kazakhstan's compact expectations are more easily absorbed by Finland's larger objective space than the reverse. Substantively, this suggests a containment pattern: most Kazakh expectations fall within the semantic range of Finnish objectives, whereas part of the Finnish objective layer extends beyond what Kazakhstan encodes as a stand-alone expectation.

The corpus examples illustrate this structural difference. Kazakhstan's expectation and criterion statements often use compact performance verbs such as 'applies,' 'analyzes,' or 'evaluates,' followed by assessable content. By contrast, Finnish objective statements frequently combine intended learning with pedagogical context, learning environments, or broader competence development. For example, the Kazakh expectation "applies mathematical knowledge to solve practical tasks" specifies learner performance, whereas the Finnish mathematics objective "to guide the pupil to strengthen reasoning and mental-calculation skills and encourage use of those skills in different situations" combines intended learning with a teacher-guidance directive. This illustration concerns statement form rather than one-to-one semantic equivalence.

This asymmetry is consistent with comparative scholarship showing that shared global vocabularies can be translated through different statement formats and degrees of granularity (Meyer et al., 1997; Steiner-Khamisi, 2004). In that sense, Finland's objectives may include a wider mix of assessable competences and broader steering aims, while Kazakhstan's expectations appear closer to compact attainment formulations. Because the one-to-one result remains close to the unconstrained estimate, the pattern is unlikely to be explained mainly by many-to-one inflation.

Two coupling architectures and where divergence concentrates

Both systems therefore appear internally coherent, but the coherence is organized differently. Kazakhstan's pattern is consistent with very tight mirroring between expectations and criteria, whereas Finland's pattern is consistent with criteria that remain strongly anchored to, but not fully exhaustive of, a broader objective layer.

This architecture can be read through constructive alignment, but the evidence should be interpreted cautiously. Kazakhstan's near-isomorphic pattern may indicate a more tightly specified relationship between expected and assessed learning, while Finland's anchored-breadth pattern may indicate greater

space for broader curricular aims alongside assessment anchors. The divergence concentrated in the Finland-to-Kazakhstan direction plausibly reflects differences in what each system treats as an “atomic” normative statement, rather than a simple difference in curricular ambition.

Thematic priorities as curricular salience

The thematic results suggest overlap in core academic domains alongside different patterns of emphasis. Finland more clearly foregrounds literacy and civic/social learning, while Kazakhstan retains a comparatively stronger digital signal than might be expected from the broader discourse profile. Interpreted through framing theory, this may indicate different ways of prioritizing cross-curricular themes in formal text.

Alignment with global discourses as translation

Alignment with global competency and sustainability discourses also appears selective. Finland is more strongly aligned with several OECD/UNESCO-associated frames at the objective layer, but this does not automatically carry through to equally distinctive criteria-level language. The findings therefore suggest that alignment with global discourse is not identical to alignment in evaluative design.

Kazakhstan’s stronger digital-competence alignment is the main exception. This may indicate that digital competence is translated into more explicit evaluative wording in the Kazakh materials, rather than remaining primarily an aspirational or integrative frame. More broadly, the pattern is consistent with policy translation rather than uniform adoption (Meyer et al., 1997; Steiner-Khamsi, 2004).

These findings also have implications for cross-national policy learning. If comparability is stronger at the level of curricular intentions than at the level of criteria, then policy borrowing is likely to be more straightforward for broad aims than for performance descriptors or grading language. This does not mean transfer is impossible, but it suggests that crosswalks between systems need to preserve nationally specific evaluative conventions rather than assume that criteria travel intact.

Implications, robustness, and limitations

From a comparative governance perspective, the findings also suggest that coherence should not be equated with uniformity. Two systems may be similarly coherent while distributing interpretive work differently between curriculum statements and criteria. In Finland, broader objectives appear to leave more room for professional mediation before evaluative judgement is fixed, whereas in Kazakhstan the closer mirroring between outcomes and criteria reduces that distance and makes the evaluative logic more explicit in the formal text. This distinction is important for policy interpretation because systems can look

convergent at the level of thematic vocabulary while still differ in where judgement is institutionally located across national settings. (Ball, 1993; Priestley et al., 2021).

A second implication concerns portability. Cross-national learning often proceeds by comparing headline curricular aims, but the present results show why this can be misleading if the criteria layer is ignored. Borrowing outcome language without the accompanying evaluative grammar may produce superficial resemblance rather than functional equivalence. For that reason, comparative analyses of curriculum reform should examine not only whether systems mention similar competencies, but also whether those competencies are translated into criteria, descriptors, and grading formulations in similar ways. This is precisely where nationally embedded assessment traditions remain most resistant to direct transfer, even under strong global convergence pressures (Pettersson et al., 2017; Steiner-Khamsi, 2004).

Methodologically, the study shows the value of directional comparison, one-to-one matching, and size-matched bootstrapping when corpora differ in granularity. At the same time, the interpretation remains bounded by the limits of text-based evidence: semantic similarity does not capture legal force, classroom enactment, teacher professional judgement, high-stakes assessment administration, nationally administered testing, assessment tasks, or all nuances introduced by language and model choice.

A limitation of this approach is that cross-lingual semantic similarity depends on the quality and language coverage of the multilingual embedding model. The study therefore does not claim to replace close bilingual curriculum interpretation. Rather, it uses cross-lingual modelling as a structured comparison of policy-text segments, with all extracted statements retained through traceable IDs linking each similarity result back to the original-language source document. A further limitation is that the study did not include an independent Finnish-language expert audit of the original Finnish segments or the selected matched examples; consequently, claims about fine-grained Finnish terminology, modality, and language-specific curricular nuance are treated cautiously. Future work could strengthen this design through independent bilingual review of a stratified sample of matched and low-similarity pairs. For that reason, the findings are best read as evidence about curriculum architecture in text, not as direct evidence about practice or implementation quality (Baden et al., 2022; Grimmer et al., 2022; van den Akker, 2003).

Conclusion

Comparative patterns in curriculum-assessment architecture

This study demonstrates that Kazakhstan and Finland achieve curriculum–assessment–regulation coherence through distinct textual architectures. Regarding the formulation and regulatory role of educational expectations (RQ1), Kazakhstan’s lower-secondary curriculum is characterised primarily by compact,

performance-oriented statements, whereas Finland's objectives more frequently combine intended learning with pedagogical guidance and broader competence development. These differences are substantive rather than merely terminological: they shape the extent to which curriculum statements function as attainment specifications, pedagogical orientations, or both.

The analysis of formal assessment regulation (RQ2) indicates that both systems establish close textual connections between expectations or objectives and the criteria, descriptors, and grading-related formulations through which achievement is judged. The architecture of this connection differs across the two cases. In Kazakhstan, criteria closely mirror a relatively compact set of expected outcomes. In Finland, criteria are firmly anchored in a broader objective layer while leaving part of that layer outside direct evaluative specification. The comparison therefore identifies contrasting configurations of curriculum–assessment coupling rather than a simple difference in the degree of coherence.

At the cross-system level, convergence is clearly layer-specific (RQ3). Similarity is stronger at the expectation/objective layer than at the criteria layer, and Kazakhstan's compact expectation set maps more readily onto Finland's broader objective space than the reverse. The two systems thus share substantial curricular terrain while organizing it through different statement genres, degrees of granularity, and relations between educational aims and evaluative provisions. This pattern is consistent with selective policy translation: common curricular ideas are retained but are reworked through nationally specific regulatory arrangements.

The analysis of thematic priorities (RQ4) reveals both convergence and differentiation. STEM receives broadly comparable emphasis in the two expectation/objective corpora, whereas Finland gives greater prominence to language and literacy, civic and social competencies, and digital literacy. Sustainability and inclusion also appear more strongly in the Finnish policy corpus. Taken together, these findings point to a more pronounced Finnish emphasis on literacy-, civic-, inclusion-, and sustainability-oriented concerns, while Kazakhstan's assessment-related language gives greater prominence to digital competence.

Considered in relation to external global-discourse frames (RQ5), the pattern remains selective rather than uniform. Finnish objectives show closer alignment with the critical-thinking, sustainability, and inclusion reference frames, whereas Kazakhstan's materials show stronger alignment with the digital-competence frame, particularly in criteria and other assessment-related wording. The different digital patterns reflect the distinction between thematic prominence within national curriculum texts and proximity to an external reference framework. For cross-national policy learning, the findings suggest that broad curricular aspirations may travel more readily than criteria, performance descriptors, and grading language, which remain embedded in nationally specific conventions of judgement.

Taken together, the analysis addresses all five research questions at the level of formal policy text. The next analytical stage concerns how these textual arrangements are interpreted and enacted through assessment tasks, grading decisions, classroom practice, and institutional routines; how they developed historically; and how their consequences vary across regional, linguistic, institutional, and school contexts.

Learning outcomes as situated regulatory constructs

The findings contribute to a relational understanding of learning outcomes. Learning outcomes acquire meaning not through wording alone, but through their relationship to criteria and descriptors, the actors authorized to interpret them, and the institutional, linguistic, and spatial conditions in which they become consequential. A performance-oriented statement embedded in a centrally regulated standard and a broader objective interpreted through municipalities, schools, and teachers may occupy overlapping semantic space while allocating responsibility, professional discretion, and educational opportunity in markedly different ways.

Global organizations, national authorities, and local professional actors therefore participate in the selection, recontextualization, and authorization of learning outcomes. Recontextualization may preserve the wording of a statement while altering its institutional function. Depending on the surrounding regulatory architecture, a learning outcome may operate as an attainment claim, a pedagogical steering device, a criterion for judgement, or a professional reference point. Convergence in vocabulary can consequently coexist with divergence in institutional meaning and practical consequence.

The study provides a transparent and auditable approach to cross-national curriculum measurement. Explicit unitization rules, multilingual embeddings, similarity thresholds, directional and one-to-one matching, size-matched bootstrapping, and traceable audit trails render the comparisons open to inspection, replication, and refinement. Together, these procedures strengthen the credibility of inferences about patterned relations among expectations, objectives, criteria, and grading-related language, while reducing the influence of corpus imbalance, generic phrasing, and repeated many-to-one matches.

The resulting indicators illuminate how formal expectations and evaluative provisions are organized within each system. Their interpretation gains depth when integrated with evidence on legal force, policy hierarchy, teacher judgement, assessment-task design, institutional capacity, and the opportunities available to learners across regional, linguistic, and school contexts. Research bringing these sources together can clarify how national assessment practices amplify, modify, or counter global evaluation agendas. The present findings provide a formal regulatory baseline for this line of inquiry.

The central implication is that semantic similarity should not be treated as evaluative equivalence. Systems

may employ comparable language and align with similar international frames while assigning different institutional roles to those formulations and locating evaluative authority in different places. Curriculum coherence is therefore best understood as a situated relationship among policy statements, assessment rules, institutional authority, and opportunities for enactment.

Policy trajectories and research agenda

The findings also support a more nuanced account of the relationship between national curriculum traditions and global educational agendas. The observed patterns are more consistent with selective recontextualization than with straightforward adoption. International vocabularies are incorporated into nationally specific statement genres, assessment conventions, and institutional arrangements, challenging any assumption that global agendas move intact into domestic policy.

Reciprocal trajectories merit equal analytical attention. National curriculum traditions, professional cultures, and policy innovations can shape global agendas through expert networks, comparative reviews, policy borrowing, and the circulation of influential models. Historical and documentary research can trace how particular concepts, formulations, and assessment approaches move across national and international policy spaces, thereby clarifying both the influence of global agendas on national curricula and the contribution of national frameworks to international discourse.

Future research should shift attention from the extent of alignment with global models to the processes through which learning outcomes are interpreted, enacted, and given practical significance across national, regional, and school contexts. Such work could combine statement-level analysis with historical examination of policy development, documentary analysis of actor networks, bilingual expert review, and evidence from assessment tasks, grading practices, and classroom enactment. It could also examine how international organizations, national authorities, municipalities, schools, and professional communities shape these processes, and how institutional capacity, regional conditions, language, and school context influence the opportunities that formally shared outcomes create for different learners.

This study contributes to debates on comparative governance, policy translation, and the construction of learning outcomes by showing that curriculum comparison becomes more informative when it distinguishes semantic similarity from the nationally situated evaluative arrangements through which learning outcomes acquire meaning. Curriculum coherence is neither a single universal property nor an automatic consequence of shared terminology; it is a situated relationship among policy statements, assessment rules, institutional authority, and opportunities for enactment. Comparative curriculum research is strengthened when it treats national systems as sites where global ideas are actively reworked into locally meaningful forms.

References

- Adam, S. (2004). *Using learning outcomes: A consideration of the nature, role, application and implications for European education of employing learning outcomes at the local, national and international levels*. United Kingdom Bologna Seminar.
- All European Academies - ALLEA. (2023). The European code of conduct for research integrity (Revised ed.). <https://allea.org/code-of-conduct/>
- Baden, C., Pipal, C., Schoonvelde, M., & van der Velden, M. A. C. G. (2022). Three gaps in computational text analysis methods for social sciences: A research agenda. *Communication Methods and Measures*, 16(1), 1-18. <https://doi.org/10.1080/19312458.2021.2015574>
- Ball, S. J. (1993). What is policy? Texts, trajectories and toolboxes. *Discourse: Studies in the Cultural Politics of Education*, 13(2), 10-17. <https://doi.org/10.1080/0159630930130203>
- Biggs, J. (1996). Enhancing teaching through constructive alignment. *Higher Education*, 32(3), 347-364. <https://doi.org/10.1007/BF00138871>
- Borhan, H. (2025). *Civic education as a pathway to inclusive societies: Exploring the role of education in fostering civic and social engagement (OECD Education Working Papers No. 326)*. OECD Publishing. <https://doi.org/10.1787/3f128be8-en>
- Clark, N., & Hsu, J. L. (2023). Insight from biology program learning outcomes: Implications for teaching, learning, and assessment. *CBE-Life Sciences Education*, 22(1). <https://doi.org/10.1187/cbe.22-09-0177>
- Creswell, J. W., & Creswell, J. D. (2022). *Research design: Qualitative, quantitative, and mixed methods approaches* (6th ed.). SAGE.
- Entman, R. M. (1993). Framing: Toward clarification of a fractured paradigm. *Journal of Communication*, 43(4), 51-58. <https://doi.org/10.1111/j.1460-2466.1993.tb01304.x>
- European Commission. (2015). *ECTS users' guide 2015*. Publications Office of the European Union. <https://doi.org/10.2766/87192>
- Grimmer, J., Roberts, M. E., & Stewart, B. M. (2022). *Text as data: A new framework for machine learning and the social sciences*. Princeton University Press.
- Hailikari, T., Virtanen, V., Vesalainen, M., & Postareff, L. (2022). Student perspectives on how different elements of constructive alignment support active learning. *Active Learning in Higher Education*, 23(3), 217-231. <https://doi.org/10.1177/1469787421989160>
- Hamdoun, W. M. A. (2023). Constructive alignment approach: Enhancing learning and teaching. *British Journal of Multidisciplinary and Advanced Studies*, 4(2), 162-170. <https://doi.org/10.37745/bjmas.2022.0173>
- Karseth, B., Bernotaite, S., & Sundby, A. H. (2024). The OECD's narrative on the future curriculum: Issuing values. *Nordic Journal of Studies in Educational Policy*, 10(2), 79-90. <https://doi.org/10.1080/20020317.2024.2361937>
- Law on personal data and their protection. (2013). *Law on personal data and their protection - Law No. 94-V*. Republic of Kazakhstan/Adilet Legal Information System. <https://adilet.zan.kz/eng/docs/Z1300000094>
- Мектепке дейінгі тәрбие мен оқытудың, бастауыш, негізгі орта, жалпы орта, техникалық және кәсіптік, орта білімнен кейінгі білім берудің мемлекеттік жалпыға міндетті стандарттарын бекіту туралы [On approval of the state compulsory standards for preschool education and training, primary, lower secondary and general secondary, technical and vocational, and post-secondary education], Қазақстан Республикасы Оқу-ағарту министрінің № 348 бұйрығы [Order of the Minister of Enlightenment of the Republic of Kazakhstan No. 348] (2022). <https://adilet.zan.kz/kaz/docs/V2200029031>

- Meyer, J. W., Boli, J., Thomas, G. M., & Ramirez, F. O. (1997). World society and the nation-state. *American Journal of Sociology*, 103(1), 144-181. <https://doi.org/10.1086/231174>
- Morgan, H. (2022). Conducting a qualitative document analysis. *The Qualitative Report*, 27(1), 64-77. <https://doi.org/10.46743/2160-3715/2022.5044>
- Ni, J., Hernández Ábrego, G., Constant, N., Ma, J., Hall, K. B., Cer, D., & Yang, Y. (2022). Sentence-T5: Scalable sentence encoders from pre-trained text-to-text models. In *Findings of ACL 2022* (pp. 1864-1874). <https://doi.org/10.18653/v1/2022.findings-acl.146>
- OECD. (2018). *The future of education and skills: Education 2030*. <https://doi.org/10.1787/54ac7020-en>
- OECD. (2020a). *Curriculum (re)design: A series of thematic reports from the OECD Education 2030 project* [Overview brochure]. <https://www.oecd.org/content/dam/oecd/en/about/projects/edu/education-2040/2-1-curriculum-design/brochure-thematic-reports-on-curriculum-redesign.pdf>
- OECD. (2020b). *Curriculum analysis of the OECD future of education and skills 2030: Technical report* (OECD Education Policy Perspectives No. 106). <https://doi.org/10.1787/fec47a82-en>
- OECD. (2025). *Education policy outlook 2025: Nurturing engaged and resilient lifelong learners in a world of digital transformation*. <https://doi.org/10.1787/c3f402ba-en>
- Opetushallitus/Utbildningsstyrelsen [Finnish National Board of Education]. (2016). *Perusopetuksen – Opetussuunnitelman Perusteet 2014* [National core curriculum for primary and lower secondary education 2014]. https://www.oph.fi/sites/default/files/documents/perusopetuksen_opetussuunnitelman_perusteet_2014.pdf
- Орта, техникалық және кәсіптік, орта білімнен кейінгі білім алушылардың білімін бағалау өлшемшарттарын бекіту туралы [On approval of the criteria for assessing the knowledge of students in secondary, technical and vocational, and post-secondary education], Қазақстан Республикасы Білім және ғылым министрінің № 52 бұйрығы [Order of the Minister of Education and Science of the Republic of Kazakhstan No. 52] (2016). <https://adilet.zan.kz/kaz/docs/V1600013137>
- Pettersson, D., Prøitz, T. S., & Forsberg, E. (2017). From role models to nations in need of advice: Norway and Sweden under the OECD's magnifying glass. *Journal of Education Policy*, 32(6), 721-744. <https://doi.org/10.1080/02680939.2017.1301557>
- Porter, A. C. (2002). Measuring the content of instruction: Uses in research and practice. *Educational Researcher*, 31(7), 3-14. <https://doi.org/10.3102/0013189X031007003>
- Priestley, M., Alvunger, D., Philippou, S., & Soini, T. (Eds.). (2021). *Curriculum making in Europe: Policy and practice within and across diverse contexts*. Emerald Publishing. <https://doi.org/10.1108/9781838677350>
- Siegesmund, R., & Smilan, C. (2023). Educational aims, goals, and objectives: Balancing instructional objectives and expressive outcomes. In R. Siegesmund & C. Smilan (Eds.), *Authentic secondary art assessment* (pp. 11-28). Routledge.
- Steiner-Khamsi, G. (Ed.). (2004). *The global politics of educational borrowing and lending*. Teachers College Press.
- UNESCO. (2020). *Education for sustainable development: A roadmap*. <https://unesdoc.unesco.org/ark:/48223/pf0000374802>
- UNESCO International Bureau of Education. (2017). *Prototype of a national curriculum framework: Training tools for curriculum development*. UNESCO. https://www.um.edu.mt/library/oar/bitstream/123456789/108649/1/Prototype_of_a_National_Curriculum_Framework.pdf
- van den Akker, J. (2003). Curriculum perspectives: An introduction. In J. van den Akker, W. Kuiper, & U. Hameyer (Eds.), *Curriculum landscapes and trends* (pp. 1-10). Springer.

- Ventista, O. M., & Brown, C. (2023). Teachers' professional learning and its impact on students' learning outcomes: Findings from a systematic review. *Social Sciences & Humanities Open*, 8(1).
<https://doi.org/10.1016/j.ssaho.2023.100565>
- Vuorikari, R., Kluzer, S., & Punie, Y. (2022). *DigComp 2.2: The digital competence framework for citizens: With new examples of knowledge, skills and attitudes*. Publications Office of the European Union.
<https://doi.org/10.2760/115376>