

Beyond Access: Development and Exploratory Pilot Evaluation of a Social Sustainability Framework for Interactive Technologies in Vulnerable Educational Communities

Juan Camilo Espinosa

Universidad Autónoma de Occidente

Juan Pablo Rivera

Universidad Autónoma de Occidente

Juan Fernández

Universidad Autónoma de Occidente

Carlos Peláez

Universidad Autónoma de Occidente

Andrés Solano

Universidad Autónoma de Occidente

Leonardo Saavedra

lsaavedra@uao.edu.co

Universidad Autónoma de Occidente

Juan Núñez

Universidad Autónoma de Occidente

Sandra Cano

Pontificia Universidad católica de Valparaíso

Huizilopoztli Luna-García

Universidad Autónoma de Zacatecas

Research Article

Keywords: Social Sustainability, Interactive Systems, Digital Divide, Emerging Technologies, Education, Pilot Study, Content Validity

Posted Date: September 9th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10060357/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Additional Declarations: No competing interests reported.

Beyond Access: Development and Exploratory Pilot Evaluation of a Social Sustainability Framework for Interactive Technologies in Vulnerable Educational Communities

Juan Espinosa¹, Juan Rivera¹, Juan Fernández¹, Carlos Peláez¹,
Andrés Solano¹, Leonardo Saavedra^{1*}, Juan Núñez^{1, 2},
Sandra Cano³, Huizilopoztli Luna-García⁴

¹Facultad de ingeniería y ciencias básicas, Universidad Autónoma de Occidente, Street 25115-85, Cali, 760030, Valle del Cauca, Colombia.

²BISITE Research Group, Universidad de Salamanca, Street, Salamanca, 37008, Comunidad autónoma de Castilla y León, España.

³Escuela de ingeniería Informática, Pontificia Universidad católica de Valparaíso, Avda. Brasil N°2950, Valparaíso, 2374631, Valparaíso, Chile.

⁴Unidad académica de Ingeniería Eléctrica, Universidad Autónoma de Zacatecas, Jardín Juárez 147, Zacatecas, 98000, Zacatecas, México.

*Corresponding author(s). E-mail(s): lsaavedra@uao.edu.co;

Contributing authors: juan_cam.espinosa@uao.edu.co; juan_p.rivera@uao.edu.co; juan_est.fernandez@uao.edu.co; capa@uao.edu.co; afsolano@uao.edu.co; jmnunez@uao.edu.co; sandra.cano@pucv.cl; hlugar@uaz.edu.mx;

Abstract

Purpose: This study reports the development and first exploratory pilot evaluation of a social sustainability framework intended to guide the integration of technology-mediated learning in vulnerable educational contexts in Santiago de Cali, Colombia. Building on a prior systematic review, the framework operationalizes seven social sustainability factors into 31 actionable recommendations. The pilot examined whether educational stakeholders could consistently associate real institutional comments with these recommendations, as an initial step toward content validation. **Methods:** The framework was derived from

a systematic literature review, field observations, and participatory consultation sessions with teachers and administrators of two primary schools. In the pilot, twelve participants organized into five working groups associated 28 institutional comments with framework recommendations. Multi-label inter-rater reliability was estimated with Krippendorff’s α under nominal, Jaccard, and MASI distance functions, with nonparametric bootstrap confidence intervals, permutation tests, leave-one-out sensitivity analysis, and explicit missing-data handling. Agreement with a hybrid human–LLM reference classification was assessed through exact match, partial overlap, consensus, Jaccard similarity, and precision–recall metrics, complemented by a qualitative analysis of participants’ written justifications. **Results:** Agreement among evaluators exceeded chance (permutation $p \approx 0.002$) but remained below pre-specified benchmarks: Krippendorff’s α was **0.150** (nominal; 95% CI [**0.079, 0.204**]), **0.296** (Jaccard distance; [**0.225, 0.343**]), and **0.246** (MASI; [**0.179, 0.299**]); none of the three pre-registered validation criteria was met (0/3). Item- and factor-level analyses located the lowest convergence in Factors 2 (cultural relevance), 4 (digital literacy), and 7 (psychosocial impact), while qualitative analysis revealed robust cross-group thematic convergence around family participation, continuous training, and digital safety. **Conclusion:** The framework shows conceptual promise and practical resonance with school stakeholders, but the present evidence corresponds to an early-stage pilot rather than a validated instrument. The study contributes a transparent, replicable evaluation protocol and a diagnostic map of where definitions, boundaries, and rater calibration must be strengthened before larger-scale validation and implementation testing.

Keywords: Social Sustainability, Interactive Systems, Digital Divide, Emerging Technologies, Education, Pilot Study, Content Validity

1 Introduction

Sustainable development in education increasingly demands attention to social sustainability, which emphasizes equity, inclusion, cultural diversity, and community well-being in learning contexts (Tafese and Kopp 2025). Globally, initiatives such as UNESCO’s Education for Sustainable Development and Sustainable Development Goal 4 (quality education for all) underscore the need for inclusive, equitable educational opportunities, particularly for marginalized populations (UNESCO 2023). Yet, compared to environmental and economic dimensions, the social pillar of sustainability in education has been underexplored (Tafese and Kopp 2025). Vulnerable groups, such as those in low-income, conflict-affected or remote communities, continue to face disproportionate educational inequalities. These disparities are not only a local concern, but a global challenge threatening to undermine social cohesion and justice if left unaddressed. In contexts like Santiago de Cali, Colombia, socioeconomically disadvantaged neighborhoods exemplify the educational gaps that social sustainability efforts seek to address. National programs (e.g., “Computadores para Educar” in Colombia) have expanded digital access in underserved areas, a need that became more acute

during the COVID-19 pandemic, when remote learning exposed stark digital divides (Edtech 2020).

However, such well-intentioned initiatives alone have proven insufficient to close these gaps (Peláez et al. 2025). The current era of emerging technologies, particularly the proliferation of Generative Artificial Intelligence (GenAI) in education, risks exacerbating existing divides, as advanced digital tools often remain out of reach for marginalized schools (J Cupac 2024). Without targeted interventions that embed social sustainability principles, technological progress may inadvertently widen educational inequalities, with vulnerable communities being the most adversely affected (Giorgi et al. 2022).

Interactive multimedia systems (IMS) and other educational technologies have emerged as promising tools for mediating and enriching learning experiences in primary education (Solano et al. 2024; Pelaez et al. 2023). In well-resourced settings, digital and interactive learning platforms have been shown to increase student engagement, personalize learning pathways, and improve educational outcomes (Tang et al. 2023). Multimedia storytelling and educational games can capture the attention of young learners and adapt to individual learning paces, potentially benefiting those who struggle with conventional teaching methods (Alotaibi 2024; Huynh et al. 2020). Interactive systems also facilitate new forms of communication and collaboration among teachers, students, and families, enabling remote learning, continuous feedback, and community participation in the educational process.

However, implementing interactive learning technologies in vulnerable communities presents unique challenges. Schools in low-income or rural areas often grapple with inadequate infrastructure, such as unstable electricity supply, limited internet connectivity, and a shortage of devices (Johnson et al. 2016; Afzal et al. 2023), as well as low levels of digital literacy among teachers and students (Werfhorst et al. 2022). Many educators in vulnerable areas receive minimal professional development in ICT, and ongoing technical support is scarce (Hennessy et al. 2020). Off-the-shelf e-learning content frequently does not reflect the lived experiences, languages, and values of learners in marginalized communities, undermining engagement and learning effectiveness (Helm et al. 2024). Moreover, resource-intensive emerging technologies raise ethical and equity concerns: data privacy, algorithmic bias, and the higher costs of advanced hardware can disproportionately impact vulnerable groups that lack safeguards and funding (Baker and Hawn 2022).

Considering these difficulties, it is increasingly recognized that technological solutions must be aligned with social sustainability if they are to deliver lasting benefits. A recent systematic review concluded that interactive educational technologies should be developed to promote genuine social sustainability rather than perpetuate existing inequalities (Peláez et al. 2025). This perspective resonates with “design justice” and inclusive design, which hold that those who have been historically marginalized should have a voice in shaping the tools and systems that affect them. As Faik et al. (2024) note, a more equal and sustainable digital future depends on including digitally marginalized communities in the opportunities created by digital technologies.

This article addresses two questions. **RQ1:** How can the social sustainability factors identified in prior literature be operationalized into a framework of actionable recommendations for technology-mediated learning in vulnerable primary schools? **RQ2:** To what extent can school stakeholders consistently and reproducibly associate real institutional situations with the framework’s recommendations, as preliminary evidence of its content validity, clarity, and usability?

Accordingly, the contributions of this work are threefold: (i) a seven-factor social sustainability framework that extends a prior systematic review (Pelález et al. 2025) by translating its descriptive factors into 31 operational recommendations co-refined with school stakeholders (Section 3.3); (ii) a transparent and replicable multi-metric protocol for the early-stage content evaluation of such frameworks, combining multi-label inter-rater reliability, reference-based agreement, and qualitative analysis (Section 3.4); and (iii) the empirical results of a first exploratory pilot in two schools of Santiago de Cali, reported in full, including uncertainty estimates, missing data, and negative results (Section 4). We deliberately frame this study as an *exploratory pilot*: the pre-registered validation criteria were not met, and the evidence presented should be read as diagnostic input for the next iteration of the framework rather than as a completed validation.

2 Background

This section synthesizes the principal findings of a systematic review conducted between January and May 2025, following the PRISMA methodology (Page et al. 2021) across five databases. The search yielded 194 records, from which 32 studies were retained after applying inclusion/exclusion criteria (Pelález et al. 2025). The main findings indicate that social sustainability is characterized by equity, justice, cohesion, participation, and community well-being (Woodcraft 2015; Eizenberg and Jabareen 2017). In education, this translates into fair access and culturally responsive learning experiences for learners facing structural disadvantages. While environmental and technical dimensions are often foregrounded in ICT/software sustainability (Naumann et al. 2011), the social pillar remains comparatively under-specified for interactive learning systems (ILS), motivating balanced, multidimensional approaches in design and evaluation (Duboc et al. 2020).

The review demonstrates that digital divides persist along three interlocking axes: access (connectivity, devices), skills (digital literacy of teachers and students), and support (maintenance, troubleshooting, training) (Qadri et al. 2025; Alhassan and Adam 2024). Although access expansion programs intensified during the pandemic, access alone has not ensured meaningful appropriation or improved outcomes. Moreover, the diffusion of advanced, data-intensive technologies risks widening disparities where infrastructure, funding, and human capacity are limited (Al-Emran et al. 2024).

Across the empirical corpus, the review converges on recurring factors shaping socially sustainable outcomes in vulnerable settings: accessibility and inclusion, cultural and contextual relevance, user acceptance and capacity, ethics, privacy and safety, and transformative approaches that democratize knowledge (Zárate et al. 2023). It

also documents promising operationalization strategies—keeping human needs central, embedding local context from the outset, empowering students, teachers, and communities (Amo et al. 2021)—and recurrent tensions: connectivity and device shortages; economic and policy logics that prioritize scale; organizational constraints such as limited training time, staff turnover, and resistance to change; privacy–personalization trade-offs; and misalignments between global standards and local sociocultural realities.

2.1 Positioning of this study relative to prior work

The seven factors presented in Section 3.3 are intentionally grounded in this prior evidence base, and several of them (accessibility and inclusion, cultural relevance, user capacity, ethics and privacy) overlap with categories already identified in the review and in the digital inclusion literature (Alhassan and Adam 2024; Qadri et al. 2025). The contribution of the present study is therefore not the discovery of new high-level categories, but their *operationalization and contextual grounding*. Three extensions distinguish this work from the prior review. First, each factor is decomposed into concrete, low-cost recommendations (31 in total) formulated for the institutional realities of vulnerable primary schools, including explicit implementation strategies (Appendix A). Second, the factor set was confronted with, and refined against, situated stakeholder input: institutional comments gathered in two schools were mapped onto the framework, and the wording and scope of factors and recommendations were adjusted accordingly (Section 3.2). Third, the framework is accompanied by an explicit, multi-metric evaluation protocol and a first empirical pilot of its interpretive stability—a step that, to our knowledge, prior frameworks in this space have not reported. Following content-validity guidance, full validation additionally requires evidence of relevance, comprehensiveness, and comprehensibility from both professionals and end users (Terwee et al. 2018); the pilot reported here addresses the comprehensibility and interpretive consistency dimension and is explicitly positioned as the first cycle of that larger program.

3 Materials and Methods

3.1 Research design and participatory approach

The study followed an iterative, participatory design process combining expert judgment (Uusitalo et al. 2015) and direct consultation with the target communities (Chevalier 2019). The process comprised four stages: (1) compilation of background experiences from previous engagements with schools in vulnerable communities, together with field observations of classroom dynamics and interviews with teachers and community leaders; (2) a systematic literature review (Section 2) identifying candidate social sustainability factors; (3) participatory consultation sessions in which teachers, school directors, and other educational actors discussed the relevance of the proposed factors to their daily realities, generating a corpus of institutional comments;

and (4) internal synthesis sessions in which the research team linked stakeholder comments to factors, examined redundancies and gaps, and consolidated factor definitions and recommendations.

We initially conceived this process within the tradition of participatory action research (PAR) (Chevalier 2019), whose long-term aim is for community members to act as co-researchers in problem definition, analysis, decision-making, and action (Cornish et al. 2023; Cockerham 2024). In this first cycle, however, community participation was limited to structured consultation and feedback: teachers and administrators contributed situated knowledge, discussed and contested the factors, and evaluated the framework, but did not co-lead the analysis or design decisions. We therefore describe the present methodology more precisely as a *participatory consultation and stakeholder-based pilot assessment*, which constitutes the trust-building and capability-building groundwork for fuller PAR cycles. Subsequent iterations are planned to progressively transfer analytical and decision-making roles to the school communities, consistent with PAR guidance on avoiding co-option and tokenistic participation (Cornish et al. 2023).

3.2 Derivation of the framework

The seven factors were derived through explicit traceability from the literature to the framework. The factor categories originate in the thematic synthesis of the systematic review (Peláez et al. 2025), which identified accessibility and inclusion, cultural and contextual relevance, user acceptance and capacity, ethics, privacy and safety, and transformative/knowledge-democratization approaches as recurrent determinants of socially sustainable outcomes. These categories were re-structured into seven factors by (a) splitting compound categories where stakeholder feedback indicated distinct institutional responsibilities (e.g., separating *skills development* from *digital inclusion*, and *infrastructure* from *safety and trust*), and (b) adding an explicit *psychosocial impact* factor, which emerged strongly from field observations and teacher interviews but was only implicit in the review corpus. Each factor was then operationalized into recommendations by combining strategies documented in the review with practices suggested by stakeholders during the consultation sessions. The full factor–recommendation mapping is given in Appendix A.

3.3 The social sustainability framework

The framework acknowledges that incorporating interactive technologies in primary education institutions located in vulnerable communities cannot be evaluated solely in terms of technological adoption or instrumental improvement of academic performance. Social sustainability is understood as the capacity of an educational intervention to endure over time, generate collective value, and strengthen processes of inclusion, cohesion, and community well-being without reproducing existing structural exclusions or creating new forms of vulnerability. This perspective positions technology as a relational and social medium that must respond to the cultural, material, organizational, and emotional realities of the school community (Qadri et al. 2025).

Figure 1 summarizes the seven factors and their interrelations. The factors operate at three complementary levels: *structural conditions* (F1 digital inclusion and social equity; F6 sustainable and context-appropriate infrastructure) define who can participate and under what material conditions; *appropriation processes* (F2 cultural relevance and contextualization; F3 social cohesion and collaboration; F4 skills development and digital literacy) determine whether technology becomes meaningful and collectively owned; and *protective conditions* (F5 safety, privacy, and trust; F7 psychosocial impact) safeguard the community against the risks that technology introduces. The levels are mutually dependent: structural gaps undermine appropriation, and weak protective conditions erode the trust on which both depend.

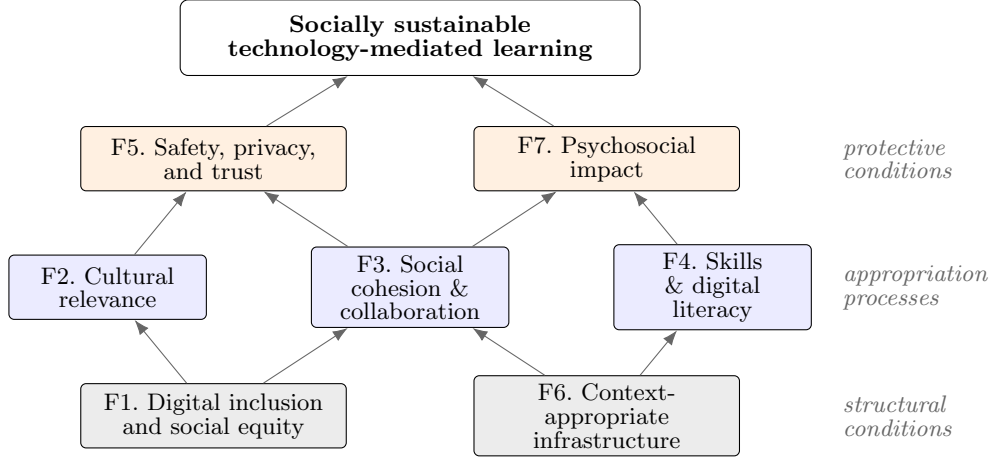


Fig. 1 Interrelation of the seven social sustainability factors. Structural conditions enable appropriation processes, which in turn depend on protective conditions to sustain trust and well-being.

In brief: **F1 (Digital inclusion and social equity)** demands that no group within the educational community be systematically excluded from technology-mediated learning, recognizing that the digital divide spans devices, connectivity, skills, confidence, language, disability, and intersecting vulnerabilities (Alhassan and Adam 2024; Eizenberg and Jabareen 2017). **F2 (Cultural relevance and contextualization)** requires that technologies respond to the languages, practices, memories, and aspirations of the community, treating interventions as co-creation processes rather than external products (Sanders and Stappers 2008; Giorgi et al. 2022). **F3 (Social cohesion and collaboration)** expects technological tools to foster mutual support, intergenerational collaboration, peer mentoring, and school–family–community ties (Sen et al. 2022; Kaliisa and Michelle 2019; Girela et al. 2021). **F4 (Skills development and digital literacy)** stresses that infrastructure without sustained capacity building does not lead to long-term appropriation, calling for continuous formative processes for teachers, students, and families (Livingstone and Blum-Ross 2020; Hartikainen et al. 2021). **F5 (Safety, privacy, and trust)**

addresses risks of surveillance, disproportionate data collection, harmful or biased content, and stigmatization, making transparency a condition of trust (Arpaci 2024). **F6 (Sustainable and context-appropriate infrastructure)** requires designing for real conditions—unstable electricity, intermittent connectivity, constrained budgets—through offline-capable, low-energy, repairable solutions (Padilla et al. 2023; Noman et al. 2024). **F7 (Psychosocial impact)** demands systematic attention to the emotional effects of technology adoption, which may enhance self-esteem and belonging or generate anxiety, isolation, and stigmatization (Afshari et al. 2022; Jung and Katz 2022).

The 31 recommendations operationalizing these factors are listed in Appendix A (Table A1). To support consistent interpretation—an issue that, as Section 4 shows, proved central in the pilot—Table 1 adds inclusion/exclusion boundary rules that delimit each factor from its conceptual neighbors.

Table 1 Boundary rules delimiting the seven factors (inclusion/exclusion guidance for coders).

Factor	Includes (primary focus)	Excludes (covered by)
F1 Inclusion & equity	Unequal access, participation, or benefit across groups; diagnosis of gaps; institutional leadership for equity	Training content and delivery (F4); physical/technical availability of equipment (F6)
F2 Cultural relevance	Fit between content/interfaces and local languages, knowledge, symbols; co-creation of local content	Generic stakeholder communication (F3); generic feedback on usability (F4)
F3 Cohesion & collaboration	Relationships and networks: peer mentoring, school–family–community ties, partnerships	Cultural adaptation of content (F2); family training in digital skills (F4)
F4 Skills & literacy	Building competencies of teachers, students, families; transfer and appropriation of solutions	Motivation/equity of access (F1); emotional barriers to use (F7)
F5 Safety, privacy, trust	Data protection, digital risks, transparency, trust-building mechanisms	Emotional well-being effects (F7); device security as hardware maintenance (F6)
F6 Infrastructure	Material conditions: energy, connectivity, spaces, maintenance, repair, reuse	Skills to use the infrastructure (F4); equitable allocation of access (F1)
F7 Psychosocial impact	Emotional and social effects of technology: anxiety, belonging, motivation, overload	Privacy-related harms (F5); social network building as pedagogy (F3)

3.4 Pilot content-evaluation study

The interpretive stability of the framework was examined through an exploratory pilot with primary education teachers and administrators. The objective was to assess whether stakeholders could consistently associate real institutional comments about educational technologies with the specific recommendations defined in the framework (Messick 1995; Rubio et al. 2003), a necessary, though not sufficient, condition for

content validity (Terwee et al. 2018). We emphasize that this exercise addresses interpretive consistency (comprehensibility and conceptual delimitation); complementary evidence of relevance and comprehensiveness, gathered through interviews, surveys, or Delphi procedures with professionals and end users (Terwee et al. 2018), is planned for subsequent cycles.

3.4.1 Participants, setting, and recruitment

Two private-subsidized primary schools serving vulnerable communities in Santiago de Cali participated: Colegio La Fontaine (LF) and Colegio Comfandi Calipso (CC). Both institutions had prior collaboration agreements with the research team. Participants were recruited through the school administrations using convenience sampling among teaching and administrative staff with experience using educational technologies in their institution. Twelve adults participated (five in LF, seven in CC), organized into five working groups, each producing one consensus coding (Table 2). Working in groups was a deliberate design decision: it mirrors how institutional decisions about technology are actually made in these schools and lowers literacy- and confidence-related barriers to participation; its implications for the statistical unit of analysis are described in Section 3.4.4. No students or family members participated in this pilot; their inclusion is part of the planned future work, and the present findings consequently speak only to the perspectives of teachers and administrators.

Table 2 Participants, working groups, and data completeness.

School	Group (rater)	Members	Items coded	Status in analysis
La Fontaine	G1LF	3	28/28	Included
La Fontaine	G2LF	2	28/28	Included
Comfandi Calipso	G1CC	2	28/28	Included
Comfandi Calipso	G2CC	2	16/28	Excluded (43% missing)
Comfandi Calipso	G3CC	3	28/28 ^a	Included

^{1a}G3CC returned four explicitly empty associations (all in Factor 7: CCF7C1, LFF7C3, LFF7C4, LFF7C5), recorded as “no association” as permitted by the protocol; a sensitivity analysis treating them as missing is reported in Section 4.2.

The LF session had no strict time limit; the CC session was constrained to 1 hour 30 minutes by institutional scheduling, which explains the incomplete coding of group G2CC (12 of 28 items missing). G2CC was excluded from all quantitative analyses; its available qualitative justifications were retained for the thematic analysis, where the asymmetry is explicitly flagged.

3.4.2 Materials: comments and reference classification

The evaluation corpus consisted of 28 institutional comments collected during the consultation stage in the two schools (e.g., “*De un grupo de 30 niños, aproximadamente la mitad tiene acceso a internet en casa*” — “Out of a group of 30 children, roughly half

have internet access at home”). Comments were distributed across the seven factors as follows: F1 (4), F2 (5), F3 (3), F4 (4), F5 (5), F6 (3), F7 (4).

A reference classification (comment $\rightarrow$ recommendation set) was constructed through a two-stage hybrid human–LLM process. *Stage 1 (human consensus)*: three members of the research team independently proposed associations for each comment and resolved disagreements by discussion until consensus. *Stage 2 (LLM augmentation)*: the GPT-5 large language model (OpenAI) was prompted—with the full factor definitions, the 31 recommendation descriptions, and each comment—to propose associations independently, in order to surface plausible associations overlooked by the team (Liao and Vaughan 2023). Every LLM-proposed association was accepted only if at least two team members judged it semantically justified; LLM proposals could add associations but never remove human-consensus ones. Nine of the 28 comments received multiple associations, reflecting the multidimensional nature of real institutional situations. The verbatim prompt, the raw LLM outputs, the acceptance log, and the final reference classification are provided as supplementary material in the project repository (see Data Availability). We acknowledge that using a partially LLM-augmented reference as the benchmark introduces a potential circularity and treat the reference as a *provisional* standard whose own validity must be established in future Delphi-style rounds; this limitation is discussed in Section 6.

3.4.3 Procedure

Each session began with informed consent and an explanation of the study’s purpose. For every factor, participant groups received the conceptual definition and a set of cards containing the related institutional comments. Groups then associated each comment with one or more of the factor’s recommendations, providing a brief written justification for each association. Participants were explicitly informed that multiple associations were possible, that some recommendations might remain unused, and that the absence of association was a valid response. Groups also recorded reflections on unused recommendations or comments they considered ambiguous.

3.4.4 Quantitative analysis

Unit of analysis and data structure.

The unit of analysis is the institutional comment (item; $n = 28$). Each rater is a working group (not an individual), and each observation is a *set* of recommendation labels (multi-label coding over 31 possible labels). Four raters entered the analysis (G1LF, G2LF, G1CC, G3CC).

Inter-rater reliability.

Because observations are label sets, Krippendorff’s α (Krippendorff 2018; Hayes and Krippendorff 2007) was computed in its generalized form $\alpha = 1 - D_o/D_e$ with three distance functions (Artstein and Poesio 2008): (i) *nominal* distance over entire sets (two sets agree only if identical), a conservative baseline; (ii) *Jaccard* set distance $\delta(A, B) = 1 - |A \cap B|/|A \cup B|$; and (iii) the *MASI* distance designed for set-valued annotations (Passonneau 2006), which weights the Jaccard ratio by a monotonicity

coefficient (1 for identical sets, 0.67 for subset relations, 0.33 for non-nested overlap, 0 for disjoint sets). A shared label-set encoding was used across raters; we note that an earlier internal version of the analysis encoded each rater’s responses independently, which artificially inflated the nominal coefficient (to $\alpha = 0.372$) and produced an outlier flag for G3CC; the corrected analysis reported here supersedes those estimates, and both analysis versions are preserved in the repository for transparency.

Uncertainty and inference.

95% confidence intervals for every α were obtained by nonparametric bootstrap over items (1,000 resamples, percentile method) (Efron and Tibshirani 1994). To test whether agreement exceeded chance, a permutation test (500 permutations) randomly reassigned each rater’s label sets across items, yielding an empirical null distribution for α .

Sensitivity and outliers.

Leave-one-out sensitivity analysis quantified each rater’s influence as $\Delta\alpha$ (change in α when the rater is excluded), with a pre-specified outlier criterion of $\Delta\alpha > 0.15$. Results are reported for the full rater set and for every leave-one-out subset, regardless of direction.

Missing data.

G2CC was excluded (12/28 items missing; Table 2). G3CC’s four empty responses were treated as a valid “no association” category in the primary analysis (consistent with the protocol) and as missing in a sensitivity analysis.

Agreement with the reference.

Alignment between each rater’s label sets and the reference was quantified by exact match (identical sets), the “at least one” criterion (non-empty intersection), mean Jaccard similarity, consensus rates (simple majority: $\geq 50\%$ of raters sharing at least one label with the reference per item; supermajority: $\geq 66\%$), and label-level precision, recall, and F1 per rater (treating each label as an independent binary decision).

Benchmark criteria.

Following the protocol defined a priori, the framework’s interpretive stability would be considered acceptable if at least two of three criteria were met: **A** (reliability): $\alpha \geq 0.667$ (Krippendorff 2018); **B** (consensus): $\geq 75\%$ of comments with at least one shared rater–reference recommendation, or $\geq 80\%$ with majority consensus, or $\geq 65\%$ with supermajority; **C** (overlap): mean Jaccard ≥ 0.55 or median exact agreement $\geq 50\%$ (Tan et al. 2018). In this pilot these thresholds function as benchmarks against which progress is measured, not as attained validation targets.

Software.

Analyses were implemented in Python 3 (pandas, NumPy, Matplotlib; the `krippendorff` package was used for cross-checking nominal coefficients). The complete analysis notebook, the standalone re-analysis script, and all raw anonymized coding data are available in the project repository.

3.4.5 Qualitative analysis

Participants' written justifications (one per association) were analyzed thematically. Two researchers independently performed open coding of the justifications, grouping them into themes linked to factors and recommendations; discrepancies were resolved by discussion with a third researcher. As a complementary step, an open-weight large language model (gpt-oss-120b) was used to produce a preliminary structuring of the corpus (recurrent problem–recommendation–justification patterns and inter-group contrasts); all model outputs were treated as candidate observations and were verified, corrected, or discarded by the human coders against the raw data before inclusion. The LLM therefore served as an augmentative pattern-surfacing aid, not as an autonomous analyst (Liao and Vaughan 2023). Reflexivity was supported by recording coding memos and by explicitly flagging the documentary asymmetry introduced by incomplete groups. Representative verbatim excerpts (translated from Spanish) are reported in Section 4.6.

4 Results

4.1 Field deployment and data completeness

In Colegio La Fontaine, the activity involved five participants organized into two working groups (Fig. 2). The session revealed a clear learning curve: the first factor required approximately 25 minutes to analyze, while subsequent factors averaged 15 minutes each. Participants described the activity as demanding due to the volume of information and the need to justify associations, but emphasized its value in fostering discussion of social sustainability challenges in their school.



Fig. 2 Evaluation activity at Colegio La Fontaine.

In Colegio Comfandi Calipso (Fig. 3), seven participants took part in three groups under a 1.5-hour constraint. One group (G2CC) was unable to complete the exercise (16/28 items coded) and was excluded from quantitative analysis; another (G3CC) completed all items but recorded four explicit “no association” responses, all in Factor 7 (Table 2). Participants reported that the activity was useful for reflecting on social sustainability challenges despite the time pressure.



Fig. 3 Evaluation activity at Colegio Comfandi Calipso.

4.2 Inter-rater reliability

Table 3 and Figure 4 summarize the corrected reliability estimates. Under the multi-label distances, Krippendorff’s α was 0.296 (Jaccard; 95% CI [0.225, 0.343]) and 0.246 (MASI; [0.179, 0.299]); under the conservative nominal (exact-set) treatment, $\alpha = 0.150$ [0.079, 0.204]. Permutation tests indicate that all three coefficients exceed chance-level agreement ($p \approx 0.002$; null mean ≈ 0), but all remain well below the $\alpha \geq 0.667$ benchmark. Treating G3CC’s four empty responses as missing rather than as a “no association” category changed the estimates only marginally ($\alpha_{\text{nominal}} = 0.162$; $\alpha_{\text{Jaccard}} = 0.326$; $\alpha_{\text{MASI}} = 0.269$).

Table 3 Inter-rater reliability (4 raters, $n = 28$ items). 95% CIs from 1,000 bootstrap resamples over items; p -values from permutation tests (500 permutations).

Distance function	Krippendorff’s α	95% CI	Permutation p
Nominal (exact set)	0.150	[0.079, 0.204]	0.002
Jaccard set distance	0.296	[0.225, 0.343]	0.002
MASI (Passonneau 2006)	0.246	[0.179, 0.299]	0.002
<i>Sensitivity: G3CC empty responses treated as missing</i>			
Nominal / Jaccard / MASI	0.162 / 0.326 / 0.269	—	—

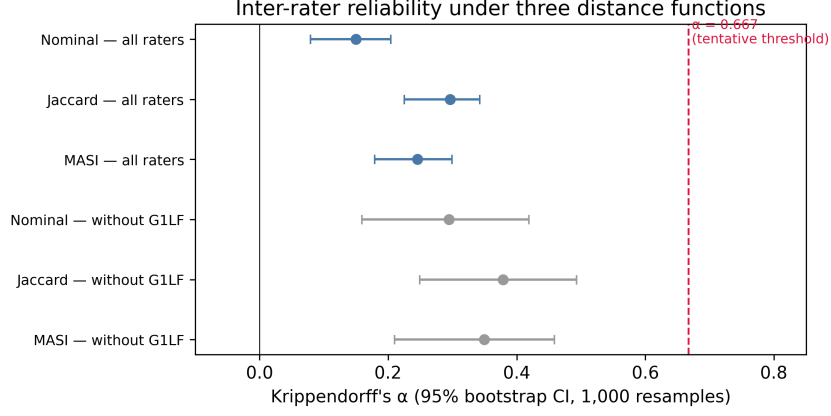


Fig. 4 Inter-rater reliability under three distance functions, with 95% bootstrap confidence intervals, for all raters (blue) and excluding the most influential rater, G1LF (gray). All estimates remain below the tentative $\alpha = 0.667$ threshold.

Leave-one-out sensitivity analysis (Fig. 5) showed that *no rater met the pre-specified outlier criterion* ($\Delta\alpha > 0.15$) under the corrected analysis. The largest influence corresponded to G1LF, whose exclusion increased α to 0.295 (nominal, $\Delta = +0.145$), 0.378 (Jaccard, $\Delta = +0.082$), and 0.350 (MASI, $\Delta = +0.104$)—still below the benchmark in every case. This finding supersedes the earlier, encoding-artifact-driven identification of G3CC as an outlier: under the corrected shared encoding, removing G3CC does not improve reliability (nominal $\Delta = -0.040$; MASI $\Delta = -0.002$). G1LF’s influence is consistent with its expansive coding style (high recall, low precision; Section 4.3) rather than with misunderstanding, and its data were therefore retained in all primary analyses, with the leave-one-out estimates reported alongside for completeness.

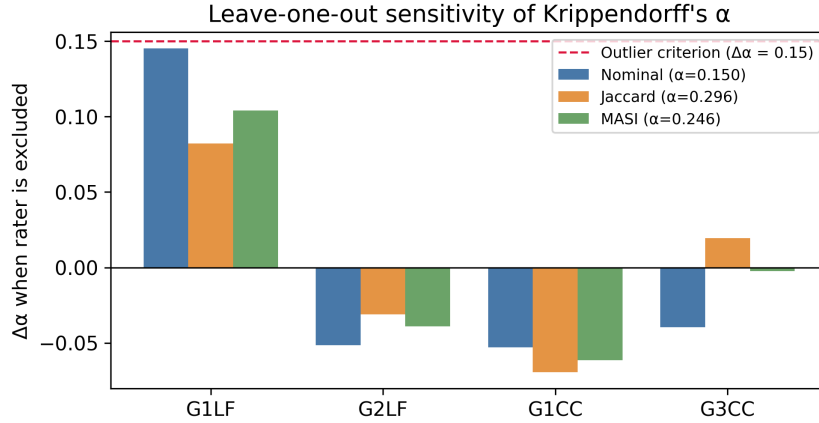


Fig. 5 Leave-one-out sensitivity of Krippendorff’s α under the three distance functions. No rater exceeds the pre-specified outlier criterion ($\Delta\alpha = 0.15$, dashed line).

At the institution level, the CC pair (G1CC–G3CC) showed higher internal consistency ($\alpha_{\text{MASI}} = 0.378$; $\alpha_{\text{nominal}} = 0.308$) than the LF pair (G1LF–G2LF; $\alpha_{\text{MASI}} = 0.184$; $\alpha_{\text{nominal}} = 0.029$), driven by the contrast between G1LF’s expansive and G2LF’s conservative styles. The pairwise similarity matrix (Fig. 6) confirms this: the highest mean Jaccard similarity was observed between G2LF and G1CC (0.45) and between G1CC and G3CC (0.43), and the lowest between G1LF and G3CC (0.15).

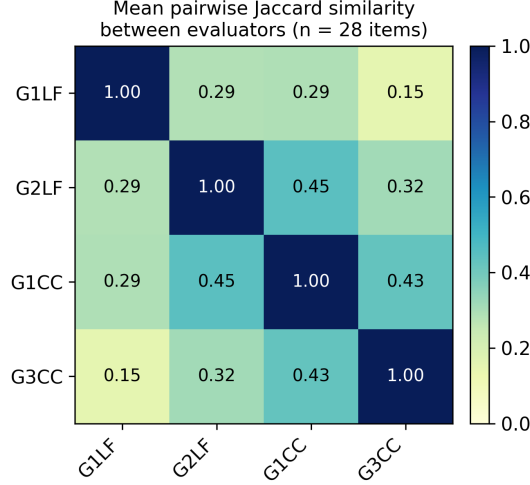


Fig. 6 Mean pairwise Jaccard similarity between evaluator groups across the 28 items.

4.3 Agreement with the reference classification

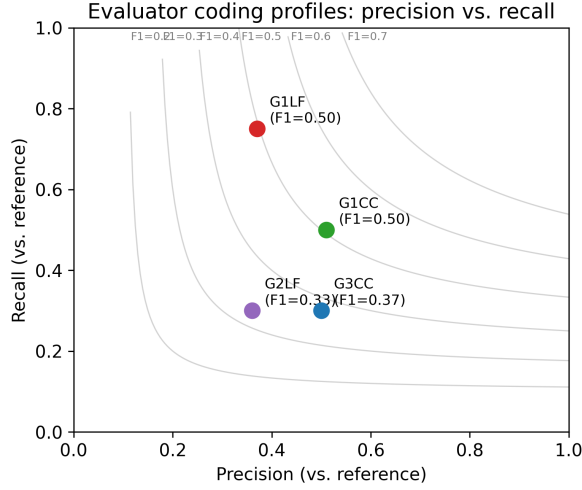
Exact agreement with the reference averaged 17.0% (median 19.6%), the “at least one overlap” criterion reached 58.0% (median 55.4%), and the mean Jaccard index was 0.343 (median 0.341). Simple-majority consensus was obtained for 67.9% of the items and supermajority consensus for 53.6%, with an average of 2.32 raters in agreement per item. Per-evaluator results (Table 4, Fig. 7) reveal distinct coding profiles. G1LF combined high recall (0.75) with low precision (0.37): an expansive strategy that maximizes partial matches (78.6% “at least one”) at the cost of exactness (0% exact match). G2LF showed the opposite, conservative pattern (precision 0.36, recall 0.30). G1CC achieved the most balanced profile (precision 0.51, recall 0.50, $F1 = 0.506$), and G3CC paired moderate precision (0.50) with low recall (0.30), a restrictive style. The average F1 across evaluators was 0.426.

4.4 Item- and factor-level analysis

Figure 8 maps inter-rater Jaccard similarity for every item and rater pair, grouped by factor. Items with the highest convergence exhibit clear conceptual boundaries and strong thematic specificity: CCF6C1 (infrastructure; mean pairwise Jaccard 0.75;

Table 4 Per-evaluator agreement with the reference classification ($n = 28$ items).

Rater	Exact (%)	≥ 1 overlap (%)	Jaccard	Precision	Recall	F1
G1LF	0.0	78.6	0.337	0.37	0.75	0.50
G2LF	17.9	42.9	0.274	0.36	0.30	0.33
G1CC	21.4	67.9	0.417	0.51	0.50	0.51
G3CC	28.6	42.9	0.345	0.50	0.30	0.38
Mean	17.0	58.0	0.343	0.44	0.46	0.43

**Fig. 7** Evaluator coding profiles: precision vs. recall against the reference classification, with iso-F1 curves. G1LF is expansive (high recall, low precision); G2LF and G3CC are conservative; G1CC is the most balanced.

F1 vs. reference 0.92), LFF1C2 (a quantified connectivity gap; 0.625; $F1 = 0.85$), and the three Factor 3 items (LFF3C1, LFF3C3, LFF3C5; all 0.50; $F1 \approx 0.75$). Conversely, the lowest convergence concentrated in Factor 2 (LFF2C2 and LFF2C6, 0.10; LFF2C7, 0.22), Factor 7 (LFF7C3, 0.11; LFF7C4, 0.18), and two Factor 1 comments of abstract, aspirational wording (CCF1C1, CCF1C5, 0.17; LFF1C3, F1 vs. reference 0.00).

Aggregating at the factor level (Table 5) localizes the interpretive difficulty: mean pairwise Jaccard was highest for F3 (0.50), F6 (0.39), and F5 (0.34), and lowest for F2 (0.20) and F7 (0.24). Exploratory per-factor MASI α values (to be read with caution given 3–5 items per factor) reproduce the pattern, with F2 (−0.001), F4 (−0.016), and F7 (0.049) indistinguishable from chance. Items permitting multiple plausible interpretations showed simultaneously low precision and low recall, indicating genuine boundary ambiguity rather than mere over- or under-selection. These results identify F2, F4, and F7—precisely the factors whose recommendations overlap most with neighboring factors (Table 1)—as the priority targets for rewording, boundary clarification, and coder calibration in the next iteration.

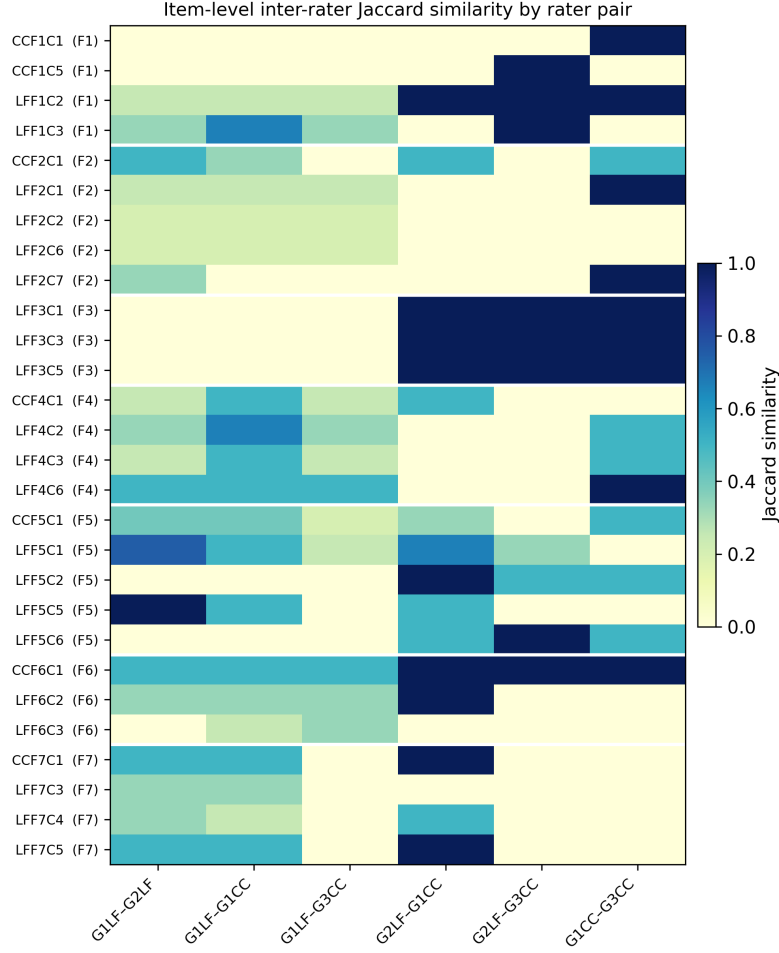


Fig. 8 Item-level inter-rater Jaccard similarity by rater pair, grouped by factor (white lines separate factors). Darker cells indicate higher agreement.

4.5 Benchmark attainment

Figure 9 contrasts every observed indicator with its pre-registered threshold. None of the three criteria was met (0/3): Criterion A ($\alpha = 0.296$ under the most favorable distance vs. 0.667), Criterion B (58% “at least one” vs. 75%; 67.9% majority vs. 80%; 53.6% supermajority vs. 65%), and Criterion C (mean Jaccard 0.343 vs. 0.55; median exact agreement 19.6% vs. 50%). Consequently, and in line with the framing adopted throughout this article, the pilot does not establish the content validity of the framework; it quantifies the distance to the validation targets and identifies where that distance is concentrated.

Table 5 Factor-level convergence (n = items per factor). Per-factor α is exploratory given the small number of items.

Factor	Items	Mean pairwise Jaccard	MA SI α (exploratory)
F1 Digital inclusion & equity	4	0.337	0.189
F2 Cultural relevance	5	0.204	−0.001
F3 Social cohesion	3	0.500	0.253
F4 Skills & digital literacy	4	0.306	−0.016
F5 Safety, privacy, trust	5	0.344	0.100
F6 Infrastructure	3	0.394	0.190
F7 Psychosocial impact	4	0.240	0.049

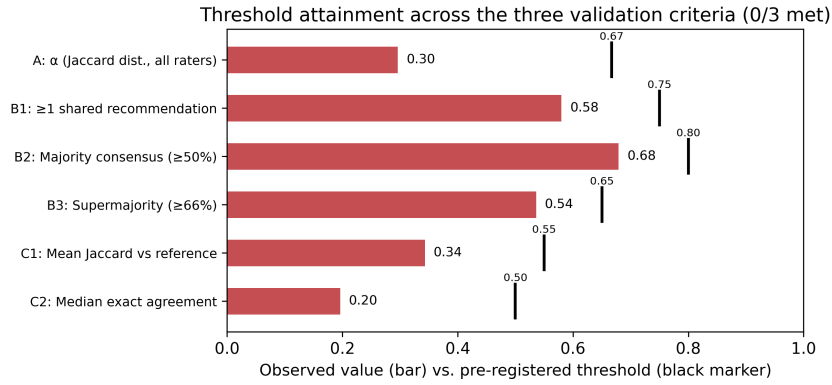


Fig. 9 Threshold-attainment dashboard. Bars show observed values; black markers show the pre-registered thresholds. Red bars indicate unmet criteria (0/3 met).

4.6 Qualitative findings

Across the corpus of justifications, three dominant thematic axes emerged consistently in both schools: family and community participation, the need for continuous training, and concerns surrounding digital safety. These axes were systematically expressed through the repeated selection of recommendations linked to digital divide diagnostics (F1R1), co-creation and cultural documentation (F2R4–F2R5), training spaces and technological transfer (F4R1–F4R4), and safe digital literacy (F5R2–F5R3). Representative justifications (translated from Spanish) illustrate the grounded character of these associations:

“Because it is important to integrate the family into educational spaces to foster communication in the school community.” (G1CC, comment LFF2C1 → F2R2)

“Because explaining to parents communicates their digital rights and data protection, bearing in mind that school leadership must recognize the importance of these rights and the risks.” (G2LF, LFF5C2 → F5R1, F5R2)

“Because it is important to train families so they can support students’ work at home.”
(G1CC, LFF4C6 → F4R2)

“Because spreading its purpose, benefit and impact makes the community understand the importance of co-creating.” (G2LF, LFF4C3 → F4R4)

Several recurrent argumentative structures emerged, including the sequences *diagnosis* → *intervention*, *training* → *appropriation*, and *co-creation* → *cultural relevance*, indicating internally coherent reasoning across groups. The analysis also identified implicit assumptions that shaped decisions without full justification: training was frequently assumed to lead automatically to appropriation; cultural co-creation was treated as sufficient for relevance; and observatory mechanisms were perceived as direct generators of opportunities. These implicit chains appeared most prominently in G1LF, consistent with its expansive quantitative profile, whereas G2LF and G3CC favored focused, operational selections and G1CC occupied an intermediate, community-oriented position. Variability in the completeness of justifications—especially in G1LF and in the partially documented G2CC—limits the traceability of some decisions and is itself a finding: future instruments should require a justification before an association can be recorded.

Overall, the qualitative corpus shows that participants recognize the structural challenges the framework names and reason about them in patterned, largely coherent ways, even where their label choices diverge. The divergences concentrate on *which* recommendation best operationalizes a shared concern—an instrument-level boundary problem—rather than on the underlying concepts.

5 Discussion

The integrated analysis offers a candid assessment of the current state of the framework and of the methodological process used to examine it. Quantitatively, agreement among evaluators exceeded chance but fell short of every pre-registered benchmark; under the corrected multi-label analysis, no indicator approached its threshold, and no rater qualified as an outlier whose removal would change that conclusion. The honest reading is that the framework, in its current wording and with untrained coders, does not yet support reproducible comment–recommendation mappings. At the same time, the structure of the disagreement is informative rather than random: convergence was systematically higher for factors with operational, materially anchored definitions (infrastructure, cohesion) and lower for factors whose boundaries are semantically porous (cultural relevance, literacy, psychosocial impact). This pattern points to instrument refinement—boundary rules, illustrative examples, coder calibration—rather than to a rejection of the underlying constructs.

Three methodological considerations qualify the interpretation. First, raters were untrained by design: the pilot measured the framework’s interpretability *as encountered*, after a brief introduction, by busy school staff. Reliability benchmarks such as $\alpha \geq 0.667$ were developed for trained coders applying a codebook (Krippendorff 2018; Hayes and Krippendorff 2007); their non-attainment here quantifies the calibration work ahead rather than invalidating the constructs. Second, the coexistence of divergent analytical styles (expansive vs. conservative; technical vs. community-oriented)

generated structured disagreement, visible in the precision–recall profiles; future sessions should include calibration rounds with worked examples and prioritization rules when multiple recommendations apply. Third, practical conditions—the CC session’s time constraint, simultaneous workload, and uneven familiarity with the materials—measurably affected completeness and depth, underscoring that evaluation logistics are part of the instrument.

The qualitative evidence complements this picture. Despite low coding consistency, robust thematic patterns cut across all groups: family involvement, continuing education, and digital safety were salient in every dataset, and participants’ justifications displayed coherent reasoning chains anchored in their institutional realities. Divergences reflected differences in pedagogical priorities and situated knowledge more than conceptual confusion. This convergence at the level of concerns, combined with divergence at the level of label assignment, is precisely the configuration one expects from a conceptually resonant but operationally under-specified instrument.

Finally, the comparison between institutions should be read cautiously. Under the corrected analysis the CC pair was internally more consistent than the LF pair, but with two rater-groups per school and different session conditions, institutional differences are confounded with group composition and time constraints, and we refrain from drawing school-level conclusions.

6 Limitations

This study has important limitations that bound its claims. (1) *Scale and sampling*: the pilot involved two schools and twelve participants recruited by convenience, organized into five group-raters of which four were analyzable; the findings describe these settings and cannot be generalized to vulnerable communities at large. (2) *Group-level coding*: consensus coding by groups reflects institutional decision practices but reduces the effective number of raters and conflates individual and group variance. (3) *Incomplete data*: one group was excluded for missingness and one session was time-constrained, introducing documentary asymmetries that we report but cannot fully correct. (4) *Missing stakeholder voices*: students and families did not participate in this cycle, so the evidence speaks only to teacher and administrator perspectives. (5) *Provisional reference standard*: the reference classification was produced by the research team with LLM augmentation; although every association was human-validated and all materials are released, the team’s involvement in both framework design and reference construction creates a potential circularity, and the reference itself requires independent validation. (6) *Scope of validity evidence*: the pilot addresses interpretive consistency only; relevance and comprehensiveness in the sense of COSMIN-type content validity (Terwee et al. 2018) remain to be assessed. (7) *Participation depth*: community members acted as consulted stakeholders rather than co-researchers, so the study does not yet meet the participation standards of mature PAR (Cornish et al. 2023; Cockerham 2024).

7 Conclusions and Future Work

This study developed a seven-factor social sustainability framework with 31 operational recommendations for technology-mediated learning in vulnerable primary schools, and subjected it to a first exploratory pilot of interpretive stability with school stakeholders in Santiago de Cali. The framework shows conceptual promise: stakeholders recognize its concerns, reason about them coherently, and value the evaluation exercise itself. However, the pre-registered validation criteria were not met (0/3), and the framework therefore requires further refinement, calibration, and larger-scale validation before it can be recommended as a validated instrument for institutional decision-making.

The pilot’s diagnostic value lies in localizing the work ahead. Future iterations will: (i) rewrite the lowest-convergence items and recommendations (concentrated in Factors 2, 4, and 7) using the boundary rules introduced in Table 1, with worked examples; (ii) introduce coder calibration sessions and require justifications at the moment of association; (iii) validate the reference classification independently, through Delphi-style expert rounds that remove the framework authors from the benchmark; (iv) expand the sample to additional schools and include students and families, gathering relevance and comprehensiveness evidence in line with content-validity guidance (Terwee et al. 2018); (v) deepen participation toward genuine co-research, transferring analysis and decision roles to community members (Cornish et al. 2023); and (vi) progress toward implementation testing, articulating diagnostics, training, and cultural co-creation into pilot interventions with capacity for scaling. The transparent protocol, open data, and uncertainty-aware reporting introduced here are intended to make each of those next steps auditable and replicable.

Statements and Declarations

Data availability. All raw anonymized coding data (`results.json`), the reference classification, the GPT-5 prompt and outputs with the acceptance log, the analysis notebook, the standalone re-analysis script reproducing every statistic and figure in this article, and the session materials (comment and recommendation cards, factor templates) are openly available in the Social Sustainability Framework repository: <https://github.com/JuanR5/Social-Sustainability-Framework>.

Ethics approval. The study protocol was reviewed and approved by the institutional ethics committee of [Institution] (approval no. [number]). The study involved adult professionals only and posed minimal risk.

Consent to participate. All participants provided written informed consent prior to the sessions, covering participation, audio/photographic recording, and the anonymized use of their contributions for research purposes.

Consent for publication. The individuals appearing in Figures 2 and 3 provided written informed consent for the publication of these images.

Declaration of generative AI use. Generative AI tools were used in three clearly delimited and human-supervised roles: (1) the GPT-5 model (OpenAI) acted as a secondary annotator in the construction of the reference classification (Section 3.4.2); every model-proposed association was reviewed and accepted or rejected by the research team. (2) An open-weight model (gpt-oss-120b) produced a preliminary structuring of the qualitative corpus (Section 3.4.5); all outputs were verified against the raw data by human coders. (3) AI-based tools assisted with language editing of the manuscript. No generative AI tool is an author; the authors take full responsibility for the content of this work.

Conflict of interest. The authors declare no competing interests.

Funding. [Funding information.]

Author contributions. [CRediT statement to be completed by the authors.]

References

- Afshari H, Agnihotri S, Searcy C, et al (2022) Social sustainability indicators: A comprehensive review with application in the energy sector. *Sustainable Production and Consumption* 31:263–286
- Afzal A, Khan S, Daud S, et al (2023) Addressing the digital divide: Access and use of technology in education. *Journal of Social Sciences Review* 3:883–895
- Al-Emran M, Abu-Hijleh B, Alsewari AA (2024) Exploring the effect of generative ai on social sustainability through integrating ai attributes, tpb, and t-eesst: A deep learning-based hybrid sem-ann approach. *IEEE Transactions on Engineering Management* 71:14512–14524. <https://doi.org/10.1109/TEM.2024.3454169>, URL <https://ieeexplore.ieee.org/document/10670553/>
- Alhassan MD, Adam IO (2024) Digital inclusion and sustainable development goals. *Journal of Science and Technology Policy Management* <https://doi.org/10.1108/JSTPM-09-2023-0152>, URL <https://www.emerald.com/insight/content/doi/10.1108/JSTPM-09-2023-0152/full/html>
- Alotaibi MS (2024) Game-based learning in early childhood education: a systematic review and meta-analysis. *Frontiers in Psychology* 15:1307881
- Amo D, García-Holgado A, Fonseca D, et al (2021) Creasteam. towards the improvement of diversity gaps through the compilation of projects, best practices and steam-lab spaces. In: *TEEM'21: Ninth International Conference on Technological Ecosystems for Enhancing Multiculturality*, pp 38–43, <https://doi.org/10.26754/CINAIC.2021.0007>
- Arpaci I (2024) A multianalytical sem-ann approach to investigate the social sustainability of ai chatbots based on cybersecurity and protection motivation theory. *IEEE*

- Transactions on Engineering Management 71:1714–1725. <https://doi.org/10.1109/TEM.2023.3339578>
- Artstein R, Poesio M (2008) Inter-coder agreement for computational linguistics. *Computational Linguistics* 34(4):555–596. <https://doi.org/10.1162/coli.07-034-R2>
- Baker RS, Hawn A (2022) Algorithmic bias in education. *International Journal of Artificial Intelligence in Education* 32:1052–1092
- Chevalier JM (2019) Participatory action research: Theory and methods for engaged inquiry. Routledge
- Cockerham D (2024) Participatory action research: building understanding, dialogue, and positive actions in a changing digital environment. *Educational Technology Research and Development* 72:2763–2791. <https://doi.org/10.1007/s11423-023-10294-1>
- Cornish F, Breton N, Moreno-Tabarez U, et al (2023) Participatory action research. *Nature Reviews Methods Primers* 3:34. <https://doi.org/10.1038/s43586-023-00214-1>
- Duboc L, Penzenstadler B, Porras J, et al (2020) Requirements engineering for sustainability: an awareness framework for designing software systems for a better tomorrow. *Requirements Engineering* 25:469–492
- Edtech G (2020) Colombia increases access to education technology. URL <https://global-edtech.com/colombia-increases-access-to-education-technology/>
- Efron B, Tibshirani RJ (1994) An Introduction to the Bootstrap. Chapman & Hall/CRC, New York
- Eizenberg E, Jabareen Y (2017) Social sustainability: A new conceptual framework. *Sustainability* 9:68
- Faik I, Sengupta A, Deng Y (2024) Inclusion by design: Requirements elicitation with digitally marginalized communities. *MIS Quarterly* 48:219–244
- Giorgi E, Cattaneo T, Guerrero KPS (2022) The principles of design for vulnerable communities: a research by design approach overrunning the disciplinary boundaries. *Buildings* 12:1789
- Girela FJG, Cervantes CT, García ER (2021) El sueño del voluntariado del grupo interactivo en educación física. *Revista de Educación* pp 31–46
- Hartikainen H, Orduña MC, Käsmä M, et al (2021) Make4change: Empowering unemployed youth through digital fabrication. In: *FabLearn Europe / MakeEd 2021 - An International Conference on Computing, Design and Making in Education*. ACM, pp 1–5, <https://doi.org/10.1145/3466725.3466763>, URL <https://dl.acm.org/doi/10.1145/3466725.3466763>

- Hayes AF, Krippendorff K (2007) Answering the call for a standard reliability measure for coding data. *Communication Methods and Measures* 1(1):77–89. <https://doi.org/10.1080/19312450709336664>
- Helm P, Bella G, Koch G, et al (2024) Diversity and language technology: how language modeling bias causes epistemic injustice. *Ethics and Information Technology* 26:8
- Hennessy S, D’Angelo S, McIntyre N, et al (2020) Technology use for teacher professional development in low- and middle-income countries: A systematic review. *Computers and Education Open* 3:100080
- Huynh E, Nyhout A, Ganea P, et al (2020) Designing narrative-focused role-playing games for visualization literacy in young children. *IEEE Transactions on Visualization and Computer Graphics* 27:924–934
- J Cupac TEH Schopmans (2024) Democratization in the age of artificial intelligence: introduction to the special issue. *Democratization* 31:899–921
- Johnson AM, Jacovina ME, Russell DG, et al (2016) Challenges and solutions when using technologies in the classroom, Routledge, pp 13–30
- Jung J, Katz RL (2022) Impacto del COVID-19 en la digitalización de América Latina. CEPAL
- Kaliisa R, Michelle P (2019) Mobile learning policy and practice in africa: Towards inclusive and equitable access to higher education. *Australasian Journal of Educational Technology* 35:1–14
- Krippendorff K (2018) *Content analysis: An introduction to its methodology*. Sage Publications
- Liao QV, Vaughan JW (2023) Ai transparency in the age of llms: A human-centered research roadmap
- Livingstone SM, Blum-Ross A (2020) *Parenting for a Digital Future: How Hopes and Fears about Technology Shape Children’s Lives*. Oxford University Press
- Messick S (1995) Validity of psychological assessment: Validation of inferences from persons’ responses and performances as scientific inquiry into score meaning. *American Psychologist* 50:741
- Naumann S, Dick M, Kern E, et al (2011) The greensoft model: A reference model for green and sustainable software and its engineering. *Sustainable Computing: Informatics and Systems* 1:294–304

- Noman H, Mahoto N, Bhatti S, et al (2024) Towards sustainable software systems: A software sustainability analysis framework. *Information and Software Technology* 169:107411. <https://doi.org/10.1016/j.infsof.2024.107411>, URL <https://linkinghub.elsevier.com/retrieve/pii/S0950584924000168>
- Padilla MO, Soto-Varela R, Berrocal LO (2023) From Digital Divide to Digital Inclusion in Colombia: A Systematic Literature Review of Strategies from Education, Springer, pp 123–147. https://doi.org/10.1007/978-981-99-7645-4_6, URL https://link.springer.com/10.1007/978-981-99-7645-4_6
- Page MJ, McKenzie JE, Bossuyt PM, et al (2021) The prisma 2020 statement: An updated guideline for reporting systematic reviews. *International Journal of Surgery* 88:105906. <https://doi.org/10.1016/j.ijssu.2021.105906>, URL <https://linkinghub.elsevier.com/retrieve/pii/S1743919121000406>
- Passonneau RJ (2006) Measuring agreement on set-valued items (MASI) for semantic and pragmatic annotation. In: *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC 2006)*. European Language Resources Association, pp 831–836
- Pelaez CA, Solano A, Rosa EADL, et al (2023) Learning analytics in the design of interactive multimedia experiences for elementary education: A systematic review. In: *Proceedings of the XXIII International Conference on Human Computer Interaction*
- Peláez CA, Espinosa JC, Fernández JE, et al (2025) Social sustainability in interactive learning systems: a systematic review of factors and tensions in vulnerable communities. In: *Proceedings of Interacción 25: XXV International Conference on Human-Computer Interaction*
- Qadri UA, Ghani MBA, Abbas U, et al (2025) Digital technologies and social sustainability in the digital transformation age: a systematic analysis and research agenda. *International Journal of Ethics and Systems* 41:142–169. <https://doi.org/10.1108/IJOES-08-2024-0239>, URL <http://www.emerald.com/ijoes/article/41/1/142-169/1240977>
- Rubio DM, Berg-Weger M, Tebb SS, et al (2003) Objectifying content validity: Conducting a content validity study in social work research. *Social Work Research* 27:94–104
- Sanders EBN, Stappers PJ (2008) Co-creation and the new landscapes of design. *Co-design* 4:5–18
- Sen K, Prybutok G, Prybutok V (2022) The use of digital technology for social wellbeing reduces social isolation in older adults: A systematic review. *SSM - Population Health* 17:101020. <https://doi.org/10.1016/j.ssmph.2021.101020>, URL <https://linkinghub.elsevier.com/retrieve/pii/S2352827321002950>

- Solano M, Moreira F, López J, et al (2024) Work route for the inclusion of learning analytics in the development of interactive multimedia experiences for elementary education. *Applied Sciences* 14:7645
- Tafese M, Kopp E (2025) Embedding social sustainability in education: A thematic review of practices and trends across educational pathways from a global perspective. *Sustainability* 17:4342
- Tan N, Steinbach M, Kumar V (2018) *Introduction to Data Mining* (2nd Edition). Pearson
- Tang X, Zainal SRBM, Li Q (2023) Multimedia use and its impact on the effectiveness of educators: a technology acceptance model perspective. *Humanities and Social Sciences Communications* 10:1–11
- Terwee CB, Prinsen CAC, Chiarotto A, et al (2018) COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study. *Quality of Life Research* 27(5):1159–1170. <https://doi.org/10.1007/s11136-018-1829-0>
- UNESCO (2023) Ensuring the right to education for vulnerable groups. URL <https://www.unesco.org/en/right-education/vulnerable-groups>
- Uusitalo L, Lehtikoinen A, Helle I, et al (2015) An overview of methods to evaluate uncertainty of deterministic models in decision support. *Environmental Modelling & Software* 63:24–31
- Werfhorst HGVD, Kessenich E, Geven S (2022) The digital divide in online education: Inequality in digital readiness of students and schools. *Computers and Education Open* 3:100100
- Woodcraft S (2015) Understanding and measuring social sustainability. *Journal of Urban Regeneration & Renewal* 8:133–144
- Zárate ZE, Trujillo CC, de-la Hoz JP (2023) Digitalization in vulnerable populations: A systematic review in latin america. *Social Indicators Research* 170:1183–1207. <https://doi.org/10.1007/s11205-023-03239-x>, URL <https://link.springer.com/10.1007/s11205-023-03239-x>

Appendix A Social sustainability factors and recommendations

Table A1: Social sustainability factors and associated recommendations for the development of learning experiences mediated by interactive technologies in vulnerable communities.

ID	Factor	Associated Recommendations
F1	Digital inclusion and social equity	F1R1. Conduct periodic analyses of the digital divide within the institution: Implement a multidimensional diagnostic process evaluating material access, digital skills, motivation, effective use, and tangible benefits obtained from technology. Utilize stratified sampling, differentiated instruments, and quantitative-qualitative indicators to identify gaps and direct resources equitably. F1R2. Apply Universal Design for Learning principles to educational content and platforms: Adopt a UDL framework based on cognitive neuroscience to optimize learning for diverse populations through multiple means of representation, action and expression, and engagement. Implement pilot experiences in fundamental subjects using existing resources and simplified assessment rubrics. F1R3. Strengthen pedagogical and technological capacities of teaching staff: Develop progressive, context-based training programs utilizing cascade methodologies and strategic partnerships. Ground professional development in authentic classroom challenges to enhance motivation and practical application of digital tools for designing context-appropriate pedagogical solutions. F1R4. Foster institutional leadership integrating ICT into the educational project: Establish leadership that positions ICT as a transversal axis in the institutional educational project, allocates resources strategically, and cultivates a continuous improvement culture. Implement recognition mechanisms for pedagogical innovation and intergenerational learning initiatives involving students as technological mediators. F1R5. Establish internal observatories for monitoring emerging technologies: Constitute interdisciplinary teams to systematically identify, analyze, and disseminate emerging technologies within the institution. Conduct periodic demonstrations evaluating strengths, limitations, and educational applications while maintaining community awareness through institutional communication channels and formative assessment mechanisms.

Continues on next page

ID	Factor	Associated Recommendations
F2	Cultural relevance and contextualization	F2R1. Organize collaborative design workshops involving students, families, and teachers: Implement collaborative design sessions integrating the educational community from initial phases of problem definition, ideation, prototyping, and validation. Establish partnerships with local universities to facilitate sessions through internships or capstone projects, reducing costs while strengthening institutional linkages and promoting community ownership. F2R2. Create ongoing feedback spaces maintaining open dialogue among educational stakeholders: Establish formal and informal feedback mechanisms through periodic meetings, anonymous suggestion channels, and brief satisfaction surveys. Implement inclusive participation formats with flexible schedules and ludic dynamics. Document contributions systematically and maintain continuous improvement cycles communicating implemented modifications transparently. F2R3. Ensure explicit cultural adaptation of content to local symbolic and linguistic contexts: Design and adapt technological solutions reflecting community languages, practices, and knowledge to enhance social appropriation. Identify cultural assets such as oral narratives, local art, ancestral knowledge, and cultural heritage. Integrate local cultural references into multimedia resources and train teachers in designing contextualized learning experiences recognizing linguistic, religious, and social diversity. F2R4. Co-create local digital content documenting and valuing community knowledge: Actively involve the educational community in creating digital content authentically reflecting cultural realities, territories, and local knowledge. Form intergenerational teams where elders narrate traditional stories, students document using accessible devices, and community leaders provide contextual insights. Guide documentation of local gastronomy, dances, and historical narratives using basic technological tools.

Continues on next page

ID	Factor	Associated Recommendations
F3	Social cohesion and collaboration	F2R5. Systematically integrate localized cultural knowledge into educational platforms and curricula: Collect and systematize community knowledge, practices, languages, and traditional learning forms for active integration into content design, interfaces, and technological activities. Include local visual or narrative elements in interfaces, adapt use cases to familiar contexts, co-design content with teachers and students based on their culture, and reinforce authenticity to foster belonging and confidence in technological solutions.
		F3R1. Design collaborative activities among students from different grades: Implement technology-mediated interactions fostering network construction, mutual learning, and shared belonging. Establish peer mentoring programs where upper-grade students guide younger peers in emerging technologies with rotating roles. Design integrated collaborative projects addressing shared objectives and real problems. Include collaboration indicators in assessment systems and publicly recognize collective achievements, as social interaction precedes individual internalization of learning.
		F3R2. Establish partnerships with local actors to share resources and co-design solutions: Develop strategic alliances based on mutual contribution of capacities, knowledge, and resources. Convene local leaders, private sector, academia, and community stakeholders to discuss common challenges and explore collaborative approaches for successful ICT implementation. Structure win-win agreements emphasizing institutional value as axis of community cohesion and future, facilitating acquisition of technological, human, and financial resources for emerging projects.
F3	Social cohesion and collaboration	F3R3. Integrate real community challenges into pedagogical work through low-cost technological solutions: Incorporate authentic community problems from family contexts and local action groups into collaborative working groups. Map and prioritize existing community problems with students, families, and leaders. Formulate clear time-bound challenges, co-design low-cost solutions with available resources, prototype and measure effectiveness, and disseminate results proposing improvement opportunities.

Continues on next page

ID	Factor	Associated Recommendations
F4	Skills development and digital literacy	F4R1. Develop ongoing teacher training based on locally identified needs: Design continuous training programs strengthening teachers' digital competencies beyond technical device use, focusing on pedagogical applications. Conduct diagnostic assessments to identify specific needs by knowledge area, structure progressive learning paths addressing technical, pedagogical, and social dimensions, and implement recognition systems documenting innovations and good practices. F4R2. Create joint training spaces for students and families to develop essential digital skills: Establish multigenerational workshops where students and families jointly develop basic digital competencies, strengthening intergenerational bonds and reducing technology-related anxiety. Diagnose family digital skills and household access, design accessible didactic materials addressing daily life needs, and establish communication channels linking school-based technological projects with home support. F4R3. Promote practical experiences in maker spaces or digital fabrication laboratories: Organize visits to nearby MakerLabs enabling hands-on, project-based collaborative learning with tools such as 3D printers, laser cutters, and robotics kits. Develop STEAM skills articulated with curriculum, facilitate community project participation, and provide access to emerging technologies for students addressing local challenges through structured pilot experiences. F4R4. Implement technology transfer strategies ensuring locally developed solutions remain accessible: Guarantee co-creation and appropriation of technological solutions from initial phases by actively involving students, teachers, and community members. Promote communication and dissemination spaces sharing purpose, benefits, potential impact, and associated risks through existing or low-cost institutional media, consolidating educational institutions as community leaders in information technology awareness.

Continues on next page

ID	Factor	Associated Recommendations
F5	Safety, privacy, and trust	F5R1. Draft clear, accessible data protection policies for students and families: Develop accessible policies in simple language adapted to local contexts using visual resources and situated examples. Implement differentiated communication materials for children, youth, and adults through multiple channels. Include digital rights, privacy, and security content in the curriculum, and ensure transparent communication of data usage and protection mechanisms to increase community trust. F5R2. Identify institutional vulnerabilities related to technology use and address them proactively: Recognize threats emerging from technology use including personal data misuse, privacy loss, access inequality, and ethical impacts. Organize workshops on digital privacy and identity when detecting uncontrolled social media use. Evaluate educational applications requesting excessive device permissions and train teachers to review privacy policies before classroom technology implementation. F5R3. Integrate ethical and safe digital literacy into the curriculum: Foster responsible technology use among students, teachers, and families through secure password creation, personal information protection, ethical information use avoiding plagiarism, and critical reflection on algorithmic biases in automated systems including generative AI. Promote awareness of digital risks and appropriate citation practices. F5R4. Select technological tools complying with privacy and transparency standards: Ensure classroom technologies meet minimum security and privacy standards, protecting student information and requiring authorization for personal data collection. Prioritize recognized educational platforms with clear privacy policies, avoid applications collecting unnecessary data, and obtain informed parental consent explaining data usage, protection measures, and storage duration.

Continues on next page

ID	Factor	Associated Recommendations
		F5R5. Foster trust through proactive engagement with families and creation of mixed monitoring committees: Generate trust by involving parents and caregivers through informative sessions and simple demonstrations explaining institutional technological solutions transparently. Organize quarterly workshops in accessible language, demonstrate digital tools before implementation, and establish mixed technology committees with students, teachers, and family representatives reviewing ICT use and proposing improvements monthly.
F6	Sustainable and context-appropriate infrastructure	F6R1. Prioritize solutions operating offline and synchronizing data when connectivity is available: Design architectures functioning offline by default with efficient synchronization when internet access exists. Document electrical supply and connectivity stability, identify environmental risks and available physical spaces. Prioritize low-energy consumption devices with integrated electrical protection, interchangeable parts, and repair possibilities. F6R2. Explicitly assess infrastructure conditions and design flexible, safe, low-cost spaces: Enable multipurpose spaces functioning as classrooms, laboratories, or community rooms for digital technology learning and experimentation. Utilize reusable furniture made from local materials, design spaces adapted to cultural peculiarities fostering ownership and collaborative learning, and ensure adequate conditions including electrical supply, ventilation, humidity control, and lighting. F6R3. Implement repair, reuse, and refurbishment practices through partnerships: Obtain donated devices from universities, companies, or foundations and recondition them with free software avoiding licensing costs. Organize community donation events collecting old computers, printers, and electronic parts. Create student and teacher groups with basic hardware repair and preventive maintenance knowledge, reducing costs and strengthening technological sustainability within the educational community.

Continues on next page

ID	Factor	Associated Recommendations
		F6R4. Establish institutional cooperation channels with community allies: Enable communication channels and collaborative initiatives with neighboring schools providing technology access and digital support. Partner with public libraries offering ICT access and digital accompaniment. Strengthen alliances with community partners and stakeholders to develop programs adapted to student needs and preferences. F6R5. Encourage family and community-based initiatives integrating technological creation with sustainability: Launch semester programs where teachers convene students and family representatives to identify local challenges, ideate technological solutions, and prototype using recycled materials and low-cost open-hardware kits under university mentorship. Exhibit projects for fundraising, document solutions through manuals and explanatory videos shared digitally, articulating innovation, family participation, and financial sustainability.
F7	Psychosocial impact of educational technologies	F7R1. Establish clear protocols for periodic evaluation of socio-emotional impact of technology use: Define psychosocial evaluation protocols identifying early signs of discomfort with technological solutions and activating preventive mechanisms. Establish baseline and continuous monitoring through simple indicators including general mood, task-related stress, and perceived teacher support. Define early warning signals, create risk-level intervention pathways, organize regular expression spaces, and train teachers in recognizing alert signals. F7R2. Ensure technological adoption has meaningful and publicly recognized purpose for the educational community: Ensure technologies provide real value to student and family needs rather than adoption by trend or imposition. Identify interests and needs through simple surveys or discussion spaces, define meaningful pedagogical objectives related to ICT use, design activities oriented toward final products satisfying identified needs, and select free or low-cost tools for prototype implementation.

Continues on next page

ID	Factor	Associated Recommendations
		F7R3. Create ICT-mediated spaces for collaboration and socialization fostering belonging and visibility: Promote collaborative ICT projects where students work in teams, build local solutions, and present results publicly to improve quality of life and social connectivity. Define academic calendar spaces enabling social integration, design science meetings, contests, and forums promoting collaboration and ICT empowerment, and document participation spaces identifying improvement opportunities. F7R4. Provide continuous emotional and pedagogical support reducing psychological barriers and technology-related anxiety: Implement training and accompaniment for teachers and students reducing technology use barriers and alleviating stress through face-to-face sessions and group programs in friendly environments. Recognize teachers' key role influencing student ICT appropriation.