JOURNAL OF E-LEARNING AND KNOWLEDGE SOCIETY
Vol. 22, No. 1 (2026)

Digital literacy programmes and social interventions for building resilience against hate speech among school students in Kerala

Genimon Vadakkemulanjanal Joseph[a,1], Sonia Philomena V A[b], Athira P[a]

[a]*Vimal Jyothi Institute of Management and Research, Chemperi. Kannur, Kerala (India)*

[b]*Department of English, Nirmalagiri College (Autonomous), Kerala (India)*

(submitted: 10/3/2026; accepted: 23/5/2026; published: 6/6/2026)

Abstract

Online hate speech against school students is increasing globally. This paper examines how digital literacy programmes and school interventions can help students aged 10 to 18 recognise, resist and respond to hate speech in online spaces. The study reviews research published between the years, 2020 and 2026. It expatiates the causes and effects of online aggression among adolescents. Qualitative data gathered through structured thematic discussions with 16 educationists drawn from government, aided and unaided schools across the State of Kerala. Inductive thematic analysis of these discussions informs the practical approaches presented, drawing on Social Emotional Learning (SEL) and critical media literacy frameworks. The paper highlights several initiatives used in Kerala schools. These include the KITE fake news detection curriculum, the State Council of Educational Research and Training (SCERT) value education framework, the Student Police Cadet programme and the Suraksha Mitram support system.

Two conceptual models guide the discussion. The first model explains how films, games and digital media can increase hostility and aggressive behaviour among students. The second model presents a four-component resilience framework that helps students develop responsible digital behaviour and stronger critical thinking. This framework is explicitly exploratory and normative in character: it synthesises existing evidence and programme documentation into a prescriptive architecture designed for iterative evaluation and refinement. The paper recommends a multi-level prevention approach. It focuses on student skill development, positive peer influence, teacher training, family participation and clear school policies. The approach suits government, aided and unaided schools in Kerala. The study is guided by Bronfenbrenner's ecological systems theory and Bandura's social learning theory. It concludes with practical recommendations to strengthen digital citizenship and reduce online hate among school students.

KEYWORDS: Digital Literacy, Hate Speech, Cyberbullying Prevention, Kerata KITE, Social Emotional Learning.

DOI
https://doi.org/10.20368/1971-8829/1136321

CITE AS
Joseph, G.V., Philomena V A, S. & P, A. (2026). Digital literacy programmes and social interventions for building resilience against hate speech among school students in Kerala. *Journal of e-Learning and Knowledge Society, 22*(1).
https://doi.org/10.20368/1971-8829/1136321

1. Introduction

The student community is greatly exposed to the digital world of information and connectivity. The seamless digital access creates new opportunities as well as risks to the students. The same platforms that provide educational content can also spread hate speech. Hate speech includes messages that attack or insult people based on caste, religion, gender, language or other social identities. Many young students face such content during an important stage of their psychological development (Castellano et al., 2023; Cedena-de-Lucas et al., 2026).

Indian schools bring together students from different social backgrounds and religions. Schools also include children from migrant families and students from different caste groups. In such a diverse environment, online hate speech can increase social tension among students. Studies in Indian educational settings show that verbal bullying and online harassment are common forms of peer victimisation. Religious and caste-based comments often act as the main triggers (Chakravarthi et

[1] corresponding author - email: jinuachan@vjim.ac.in

al., 2025). The impact of hate speech can be serious. Continuous exposure to hostile messages can increase anxiety and stress among students. It can also reduce their interest in studies and classroom participation. Some students may avoid school or withdraw from learning activities. These effects can influence academic performance and the overall social climate in schools (Philip, 2023).

Kerala has several institutional systems that can help address this problem. Educational programmes already work with students and teachers across the state. These include Kerala Infrastructure and Technology for Education (KITE) curriculum initiatives, the State Council of Educational Research and Training (SCERT) value education framework and the District Institutes of Education and Training (DIETs). Other initiatives include the Student Police Cadet (SPC) programme, the Suraksha Mitram support system and the Little KITEs student IT clubs. The monitoring structure of the Samagra Shiksha Abhiyan also supports school level programmes. Together these initiatives create a strong support network. However, there is still a need for a clear framework that connects these programmes to the issue of online hate speech.

What makes Kerala a particularly distinctive analytical case is the convergence of three structural factors that are rarely found together in any single Indian state. They are: near-universal school enrolment, exceptionally high baseline digital literacy (94 per cent as per Census 2011, approaching 100 per cent among current school students), and a dense institutional infrastructure for educational governance. These conditions mean that Kerala is not simply a convenient geographical focus; it is a context where systemic, scalable interventions are genuinely feasible within existing educational institutions. The applicability of the proposed framework is not limited to Kerala – similar conditions in other states or countries would support comparable implementation. But its specificity and the evidence underpinning it derive from the particular programmes, policies, and social dynamics of this state. Where the paper makes recommendations that are broadly generalisable.

It is also necessary to define hate speech carefully in the school context. Not every offensive comment can be treated as hate speech. Very broad definitions may restrict normal discussion or disagreement. This paper follows the definition used by UNESCO in 2019. Hate speech refers to communication that attacks a person or group based on protected characteristics such as religion, caste, gender, ethnicity, or disability. Such communication may insult, threaten or promote hatred. Hate speech may appear in schools in different forms. These include undermining value of life, caste-based insults shared through messaging apps, memes that stereotype religious communities, gender harassment in group chats and jokes about the accents or culture of migrant students. At the same time normal criticism, disagreement, or satire should not be treated as hate speech. Teaching students to understand this difference is an important part of digital literacy education.

It is equally important to distinguish hate speech from conceptually adjacent categories that appear in the same literature and policy discourse, including cyberbullying, online aggression, and misinformation. Cyberbullying refers to repeated, targeted harmful behaviour directed by an individual or group against another individual, and does not necessarily involve identity-based attacks on protected characteristics. Online aggression is a broader category encompassing hostile communication that may or may not be repeated, targeted, or identity-based. Misinformation concerns false or misleading content distributed irrespective of any intent to harm a particular social group. Hate speech, as this paper consistently uses the term in accordance with the UNESCO (2019) definition. It is distinctively characterised by its targeting of persons on the basis of protected identity characteristics as caste, religion, gender, ethnicity, disability and language in the Indian school context. These phenomena frequently co-occur in school digital environments and while some interventions address more than one category. The hate speech particularly damaging in diverse educational settings and prevents precise measurement of intervention outcomes.

Kerala had 94 percentage digital literacy as per census-2011 and it is nearing 100 percentage among school students. Digital learning tools have become common in classrooms after the COVID period. At the same time divisive content has increased on social media platforms. Some of this content circulates in messaging groups of school students. This situation increases the need for stronger digital awareness and responsible online behaviour.

This paper reviews research on how digital media exposure can influence hate speech and aggression among young people. The qualitative analysis discusses practical school-based methods that can build resilience among students and support dignified digital communication.

media promotes observational learning and social modelling (Bandura, 1977): when characters in films or games resolve conflicts through aggression and face no meaningful consequences, students absorb aggressive behaviour as a normative response to frustration or social threat. Repeated exposure reduces empathy by making suffering and revenge appear routine (Krahé, 2025; Wachs et al., 2025).

As illustrated in Figure 1, four reinforcing pathways are identified in the literature. First, exposure to violent media promotes observational learning and social modelling (Bandura, 1977): when characters in films or games resolve conflicts through aggression and face no meaningful consequences, students absorb aggressive behaviour as a normative response to frustration or social threat. Repeated exposure to these reduces empathy by making suffering and revenge appear routine (Krahé, 2025; Wachs et al., 2025).

2. Literature Review

2.1 Digital Media Impact on Hate and Aggression in Students

A growing body of research documents the mechanisms through which digital media exposure increases hostile attitudes and aggressive behaviour in adolescents. Figure 1 summarises these pathways, which have been identified across experimental, longitudinal and review-based studies conducted in India and internationally.

As illustrated in Figure 1, four reinforcing pathways are identified in the literature. First, exposure to violent

Second, interactive gaming introduces a qualitatively different dynamic. Unlike passive film viewing, games require players to practise aggressive responses repeatedly. Games are structured to reward aggressive choices through points, progress and social status within gaming communities. This reinforcement loop is immediate and direct. It strengthens associations between aggression and reward in ways that can generalise beyond the game environment (Chen et al., 2024).

Third, social media echo chambers expose students to concentrated streams of hate narratives, filtered through algorithmic recommendation systems that prioritise emotionally provocative content because it generates higher engagement. Online disinhibition – the reduction in social constraint associated with perceived anonymity – lowers the threshold for posting hateful content that students would not express face-to-face. Repeated exposure within a closed information environment normalises the attitudes being circulated (Sunstein, 2017; Törnberg & Törnberg, 2024; Vasconcellos-Silva & Castiel, 2025).

The combined effect of these four pathways produces the behavioural outcomes shown at the centre of Figure 1: increased hostile thoughts, desensitisation to violence, and normalisation of aggressive behaviour (Pittman, 2023).

The moderating factors identified in the research literature – parental involvement, critical discussion and education, and prosocial peer influence – correspond precisely to the intervention targets addressed in the second half of this paper.

Research affirms these findings and indicates the increasing online bullying among school students. A review by Suresh and Vijaya (2024) found that verbal and relational bullying are the most common forms in schools. Cyberbullying has increased sharply in studies after 2015. A major clinical study at the National Institute of Mental Health and Neurosciences (Ranjith, Vranda, & Kishore, 2023) examined 650 students. The study found cyberbullying victimisation in 14.5 percent of students and perpetration in 5.8 percent. About 13.8 percent were both victims and perpetrators. A global

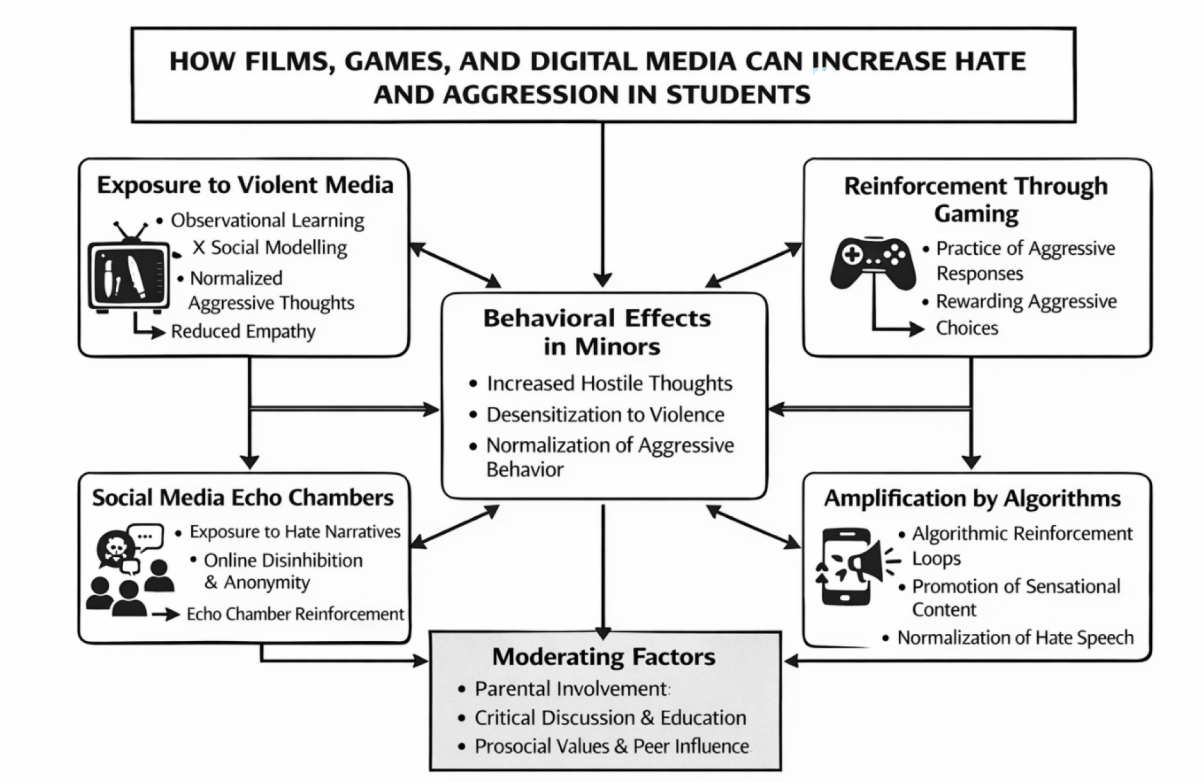

Figure 1 - How Films, Games, and Digital Media Can Increase Hate and Aggression in Students. [Note: This figure was developed by the author based on the studies using standard diagrammatic design tools and is not AI generated].

review by Zhu and colleagues (2021) shows that cyberbullying is linked with low self-esteem, social isolation and higher risk of suicidal thoughts.

2.2 Theoretical Frameworks

This analysis is based on two theoretical frames. First, Bronfenbrenner's (1979) ecological systems model advocates the integration of concentric layers of social context: the microsystem (classroom and peer group), the mesosystem (school–family interactions), the exosystem (institutions and policies) and the macrosystem (cultural values and norms). This model explains why a module delivered only to students without involving teachers, families, or school policy tends to produce limited and short-lived effects. Sustainable change requires addressing multiple levels of this ecology simultaneously (Bilz et al., 2024).

Second, Bandura's (1977) social learning theory provides a complementary account of how behaviours and norms are acquired and modified. Students learn through observation of peers and adults, through direct experience and through the social feedback they receive for their behaviour. An intervention that changes only the knowledge of individual students without incorporating their encircling social environment is unlikely to produce durable change in conduct.

2.3 Evidence for School-Based Interventions

A quasi-experimental trial of the Bullying Intervention Module (BIM) in Varanasi schools by Singh and Singh (2025) found a 25 percent improvement in students' awareness of bullying causes and consequences. Ranjith et al. (2025) validated components for Karnataka covering safe platform use, family communication strategies and school reporting systems directly adaptable for Kerala. Whole-school approaches consistently outperform single-component programmes. A review of Indian educational psychology literature found that programmes combining teacher training, peer norms change and family involvement reduced aggressive behaviour by approximately 20 percent relative to single-component alternatives (Joseph et al., 2026; Ranjith et al., 2023). UNESCO's (2019) digital literacy policy framework identifies critical media literacy as the capacity to evaluate rather than merely access digital content. This media literacy is protective against hate propaganda and reinforcing harmony.

Research on peer relationships also highlights the role of bystanders as a leverage point for intervention. In the majority of cyberbullying incidents, bystanders were present in some form – as witnesses to a public post, members of a group chat, or recipients of a forwarded message. Whether those bystanders reinforced, ignored, or challenged the hateful behaviour was a stronger predictor of incident escalation than any characteristic of the perpetrator (Joseph & Thomas, 2020; Rudnicki et al., 2023; Wachs et al., 2024). This finding has a direct and actionable implication for schools: programmes that activate the prosocial majority – the students who neither perpetrate nor are victimised but who witness hate speech regularly – have the potential to change the normative environment for the whole community.

UNICEF India's (2020) analysis of child protection in digital environments in India emphasises that the most effective national-level responses combine curriculum-based education with platform-level policy engagement and community-level awareness. Kerala's state government has engaged in policy dialogue with social media companies regarding communal content, and the Kerala Police's Cyberdome unit has conducted awareness programmes in schools and communities. Coordination between these institutional efforts and the school-based prevention framework proposed in this paper would enhance the coherence and reach of the overall response.

2.4 Social Emotional Learning as Affective Foundation

The implementation of Social Emotional Learning (SEL) provides the affective foundation for resilience of the students. The five core competencies – self-awareness, self-management, social awareness, relationship skills, and responsible decision-making – map directly onto the capacities required to navigate hate speech. The SCERT's existing value education slots (10–15 minutes) offer a practical timetable entry point for SEL activities without displacing academic content in government and aided schools of Kerala. In structured discussions, students articulate responses to realistic scenarios involving online exclusion or identity-based insults to build reflective habits. Role-playing exercises enable students to inhabit the perspectives of targets, bystanders and perpetrators in simulated online exchanges.

Assessment of SEL outcomes in school settings is most effectively accomplished through teacher observation rubrics, student self-report instruments and peer-assessment formats rather than written tests. Teachers use reflection on weekly prompts such as 'What did I notice this week about how online words affect people around me?' – instruments that serve simultaneously as learning tools and low-burden evidence-collection mechanisms. These continuous evaluations provide teachers and school administrators with data on the trajectory of students' developing awareness without requiring standardised testing infrastructure. The DIET network can support teachers in designing and interpreting these instruments as part of regular in-service training.

3. Analysis of Digital Literacy Programme Components in Kerala Schools

3.1 Qualitative Study: Methodology

Population, Sampling, and Data Collection

The study participants were school teachers and administrators serving in government, government-aided, and unaided private schools in Kerala. Participants were identified through a stratified purposive sampling procedure. The three school management categories (government, aided, unaided) formed the strata and within each stratum schools affiliated with the Kerala State syllabus, CBSE, and ICSE boards were represented to ensure curriculum diversity. Within this stratified frame, voluntary participation was invited through school principals who were briefed on the study's purpose. The resulting sample comprised 16 educators: 6 from government schools, 6 from government-aided schools, and 4 from unaided private schools, spanning primary, upper primary, and secondary levels.

Data were gathered through semi-structured group discussions conducted in Malayalam and English according to participant preference. Each discussion lasted between 60 and 90 minutes. A discussion protocol organised around three thematic domains was followed consistently: (i) the nature and manifestations of hate speech observed in the school's digital environment; (ii) the behavioural, social-normative, and psychological consequences of hate speech observed among students; and (iii) the interventions currently employed by the school to prevent or respond to hate speech, including participants' assessments of their effectiveness. Discussions were recorded with participant consent and subsequently analysed.

Data Analysis

The data of the discussions were subjected to inductive thematic analysis following the six-phase procedure described by Braun and Clarke (2006): familiarisation with the data, systematic generation of initial codes, organisation of codes into candidate themes, review of themes against the full dataset, definition and naming of final themes, and production of the analytic narrative. Coding was conducted manually. The emergent themes were cross-referenced against the experimental and epidemiological literature reviewed in Section 2. The apparent discrepancies between primary data and existing evidence were specifically interrogated rather than resolved by default in favour of either source.

Interpretive Criteria and Limitations

Trustworthiness was pursued through three procedures: prolonged engagement with the dataset, member-checking of key thematic summaries with two participants drawn from different school types, and reflexive acknowledgement of the researcher's

Table 1 - Consequences of Hate Speech in Educational Settings.
Source: Formulated from primary qualitative data (educator discussions) and literature synthesis.

Context	Behavioural Consequences	Social and Normative Consequences	Psychological and Emotional Consequences
Initial or limited exposure to hate speech	Ability to recognise derogatory language and hate-based content online; motivated avoidance of hostile interactions; appropriate use of reporting tools when prompted	Peer norms broadly align with school anti-discrimination expectations; bystanders respond with concern and support; class climate perceived as mostly inclusive	Emotional arousal (distress, discomfort) when exposed to slurs or hostile posts; empathy for peers who are targeted; sense of belonging and trust in school environment intact
Repeated or sustained exposure to hate speech	Reduced sensitivity to derogatory language; hate speech normalised in everyday digital exchange; increased risk of perpetrating bullying or passing on hostile content; avoidance of school and social settings; discrimination against peers from targeted groups	Descriptive norms shift: hateful behaviour perceived as common and therefore acceptable; bystander passivity increases; social hierarchies reinforced along caste, religious, or gender lines; erosion of inter-group trust in mixed-faith or multi-caste classrooms	Desensitisation to violence and hostility; diminished empathic response; elevated anxiety, depression, and emotional difficulties in victimised students (Zhu et al., 2021); reduced sense of school safety and belonging; increased risk of academic disengagement
Institutional level (school environment)	Increased disciplinary incidents linked to online conflicts spilling into school settings; decline in participation in co-curricular and community activities; teacher time diverted from instruction to conflict management	School's anti-hate norms weakened if policies are not enforced consistently; community trust in school as a safe environment erodes among minority-group families; peer group fragmentation along identity lines	Collective emotional climate shifts toward anxiety and vigilance; reduced teacher confidence in managing sensitive digital citizenship discussions; wider school community morale affected when high-profile incidents become public

institutional position within the Kerala educational system. Data saturation (the point at which additional discussion yielded no new themes) was reached after twelve discussions; the remaining four provided confirmatory depth rather than novel themes. The primary limitation of this qualitative component is its relatively small sample and its reliance on voluntary participation, which may have introduced a pro-intervention bias among participants. The data are best understood as providing practitioner-grounded corroboration of the literature-based evidence, not as a stand-alone empirical foundation for the proposed framework.

Table 1 below presents the consolidated output of the thematic analysis, organised by context of exposure to hate speech. The three-column structure (behavioural, social-normative, and psychological consequences) reflects the three thematic domains of the discussion protocol. Entries are indexed to supporting literature where convergence was identified.

3.2 Digital Literacy and Fake News Curriculum

Critical media literacy (CML) teaches students to interrogate the construction and circulation of digital content rather than treating it as neutral information (Bilz et al., 2024). This is especially important where communally divisive fabricated content circulates during election periods and religious festivals through social media. The SIFT method (Stop, Investigate the source, Find better coverage, Trace claims to their origin) provides a structured framework that can be scaffolded across year groups. For Classes 6 and 7, the investigation step can be simplified to checking whether a forwarded image appears on established fact-checking platforms such as Alt News or Boom. For Classes 9 and 10, students can engage with more demanding analysis of how hate speech often operates through partial truths.

KITE incorporated fake news detection modules into ICT textbooks for Classes 5 and 7 in 2024. The modules cover the mechanics of online misinformation, verification protocols adapted for the platforms students commonly use (WhatsApp, YouTube, Instagram) and the consequences of sharing unverified content with respect to recent communal disputes in Indian states. Teacher feedback indicated improved student engagement and measurable changes in verification habits.

The 2024 curriculum revision of the Kerala Syllabus introduced discussion of the social and legal consequences of spreading misinformation, including the provisions of the IT Act 2000 and the IT Rules 2021, bringing a rights and responsibilities dimension into the classroom that connects digital literacy to civic education more broadly.

3.3 Peer Mentoring and Bystander Training – Little KITEs

Peer-led mentoring is among the most cost-effective prevention components (Zhu et al., 2021). Kerala's Little KITEs clubs, operating across hundreds of governments and aided schools, provide existing infrastructure for peer digital citizenship leadership. Senior Little KITEs members trained in hate speech recognition and response can serve as first-line mentors for younger students encountering victimisation.

Bystander training is a frequently underemphasised dimension of prevention. Bystander behaviour is one of the strongest predictors of whether cyberbullying escalates or subsides (Kowalski et al., 2014; Rudnicki et al., 2023). The Little KITEs club activities train specific, low-risk bystander strategies – counter-posting, private supportive messages to targets, use of platform reporting mechanisms – to bridge the gap between awareness and action. This is integrated into the Little KITEs training calendar without displacing existing technical content.

3.4 Student Police Cadet Programme

The Student Police Cadet programme (SPC) is jointly managed by Kerala's Police Department and the General Education Department of Kerala since 2010. It engages approximately 86,000 students of Classes 8 to 12 at a time. Its two-year curriculum covers law, human rights, civic empathy, community service and rejection of social evils. The programme has documented reductions in disciplinary incidents in participating schools and improvements in self-reported empathy across social backgrounds. Students with SPC exposure integrate an explicit digital citizenship strand covering online rights, responsibilities and responses to hate speech (Chacko, 2020; Student Police Cadet, n.d.).

3.5 Suraksha Mitram and Safe Expression Structures

The Suraksha Mitram initiative was introduced by the Kerala General Education Department in 2025 to protect students from abuse. It ensures the availability of multiple confidential 'help boxes' in campus, weekly reviews of the feeds from the boxes, teachers' training, regular reporting, zero-hour sessions, and diary writing. It provides confidential multiple pathways for students to obtain protection from online victimisation (Kerala General Education Department, 2025). Research consistently shows that victimised adolescents rarely disclose to adults, citing fear of device confiscation, concern about parental overreaction or scepticism about effective intervention (Afrouz & Vassos, 2024; Jovanovic & Markovic, 2023). Normalising structured student expression within the school day lowers the social cost of disclosure.

Table 2 below presents a consolidated overview of the core activities performed in Kerala schools that constitute the empirical basis for the resilience framework proposed in Section 4.

These combined efforts strengthen empathy, responsible digital behaviour and safer school environments for diverse student communities.

4. A Proposed Framework for Building Resilience

Based on the review of the research and the analysis of the qualitative discussions, a model for building resilience among students is proposed. Figure 2 presents the integrated resilience framework proposed in this paper. It identifies four interconnected components – Education and Awareness, Critical Thinking Skills, Positive Online Behaviour, and Safe and Inclusive Spaces. The model is driven by the dual engines of Digital Literacy Skills and Empathy and Civic Values, and directed towards the outcome of resilient and responsible students.

This model is normative and exploratory in character: it synthesises evidence from the literature reviewed in Section 2 and from the qualitative data presented in Section 3.1 into a comprehensible prescriptive architecture. However, it has not been empirically validated through a controlled experimental or quasi-experimental evaluation. The four components and their inter-relationships are theoretically grounded and supported by practitioner-reported data. But require independent testing across different school contexts before strong causal claims can be made about the framework's effectiveness. The model is best understood as an existence claim: that the institutional infrastructure of Kerala's school system makes this kind of integrated framework feasible – not yet as a demonstrated efficacy claim.

Future validation strategies may include: pre-post survey designs measuring hate speech attitudes, bystander efficacy, and digital verification behaviour in schools that implement the framework. The structured teacher observation protocols using the SEL rubrics described in Section 2.4 and longitudinal tracking of anonymised incident data using the KITE platform's data management capabilities. Comparison between schools with high versus partial implementation would enable dose-response analysis of the framework's components.

The four components of the framework address the full range of moderating factors identified in Figure 1 – parental involvement, critical education, and prosocial peer influence – within a structured school-based architecture.

Together they operationalise the multi-level ecological approach described in Section 2.2, targeting the individual, peer, institutional and family dimensions simultaneously.

4.1 Education and Awareness

Students who cannot identify hate speech will neither seek help when victimised nor exercise restraint when tempted to perpetuate it. KITE's fake news detection curriculum and SCERT's value education framework jointly address this knowledge dimension. Classroom activities should include exposure to examples drawn from platforms students actually use, followed by structured analysis of why the content is harmful and what responses are available.

Table 2 - Summary of Core Intervention Activities in Kerala Schools.
Source: Formulated from primary data and programme documentation.

Initiative	Type	Core Activities	Relevance to Resilience	Source
Fake News Modules (KITE ICT, 2024)	Curriculum	Verification skills; source analysis; Classes 5 & 7	Reduces susceptibility to hate propaganda	KITE, GoK (2024)
First Bell Digital Teaching Platform	Digital education	Asynchronous content delivery; Grades 1–12	Safe information access regardless of bandwidth	KITE, GoK (2020–present)
Little KITEs Student IT Clubs	Student leadership	Peer mentoring; ICT projects; digital citizenship	Peer norm-setting; bystander capacity	KITE, GoK (ongoing)
SCERT Value Education (2023 revision)	Curriculum	Empathy activities; diversity narratives; social reform history	Inter-community respect; hate recognition skills	SCERT Kerala (2023)
Student Police Cadet (SPC) Programme	Civic character	Discipline; law; civic empathy; community service	Reduces violence; builds tolerance in Classes 8-12	Kerala Police / GED (ongoing)
Suraksha Mitram / Help Boxes / Zero-Hour	Student support	Confidential disclosure; safe expression; peer support	First-response support for victimised students	GED Kerala (2023–present)
Samagra Shiksha Abhiyan (Kerala)	National scheme	Inclusion support; DIET training; monitoring dashboard	Equitable implementation infrastructure	MoE GoI / GED Kerala (ongoing)

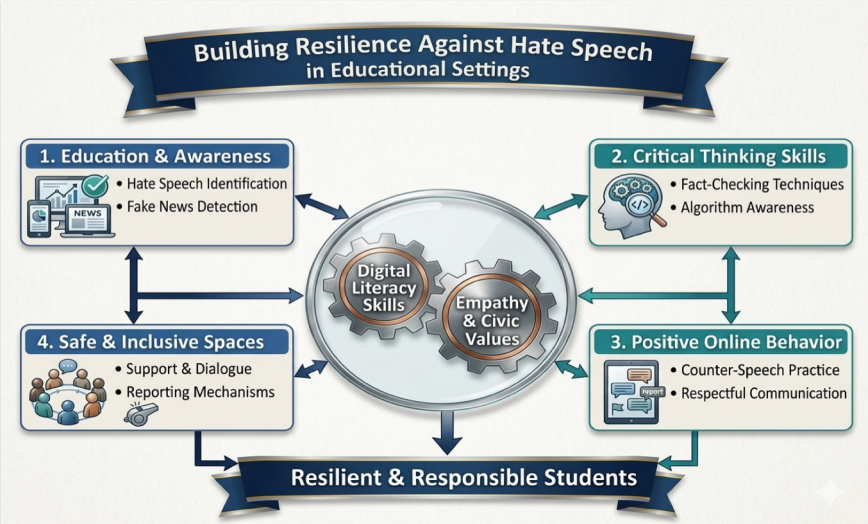

Figure 2 - Proposed Preventive Framework: Building Resilience Against Hate Speech in Educational Settings [Note: This figure was developed by the author using standard diagrammatic design tools and based on the earlier researches].

This component requires brief but specific professional development training for teachers. DIET in-service training structured as four to six sessions – using Malayalam-medium facilitator guides and case study cards mapped to SCERT's timetable – provides adequate preparation for primary and lower secondary levels. Training must include reflection on facilitators' own social positions, particularly regarding caste, to avoid inadvertently reproducing the dynamics the programme aims to address.

4.2 Critical Thinking Skills

The second component develops the analytical tools required to act on awareness. SIFT-based fact-checking pedagogy, adapted across year groups from Class 6 upward, is the core methodology. For senior secondary students, the analysis extends to understanding how algorithmic amplification works and what platform-level governance tools – reporting mechanisms, privacy settings and appeal procedures – are available to users.

The goal is to equip students with a calibrated critical evaluation habit – the automatic pause and questioning before engaging with content that triggers a strong emotional response, rather than a generalised scepticism that would undermine legitimate digital engagement. Singh and Singh's (2025) finding that structured reflective prompting was the most effective element of the BIM intervention supports the priority given to metacognitive approaches here.

4.3 Positive Online Behaviour

The third component moves from protective to generative: students do not merely avoid harmful behaviour but actively practise positive alternatives. Counter-speech training enables students to produce content that challenges hateful narratives with accurate, respectful information. Counter-speech is increasingly recognised as a more sustainable response than reporting or withdrawal alone (Prasannan et al., 2025), as it changes the information environment rather than merely removing individual posts (UNESCO, 2019). The positive online behaviour can be reinforced by cultivating a respectful communication culture within the student community.

Positive online behaviour can take the form of student-produced counter-narrative content in the native language. Students can draft unity pledges collaboratively in class and share them through school KITE platform channels. The Little KITEs peer leadership structure provides the social architecture through which positive digital norms are transmitted from experienced students to newer ones.

4.4 Safe and Inclusive Spaces

The fourth component addresses what happens when the first three are insufficient. Safe and inclusive spaces are defined by two features: accessible support and functional reporting mechanisms. The Suraksha Mitram initiatives and 'zero-hour' discussion provide the 'support and dialogue' dimension. The reporting dimension requires both a clear school policy specifying what constitutes reportable digital hate speech and a practical, low-stigma mechanism for making a report without fear of social consequences.

Families are essential to this component. PTA awareness sessions on how to respond to disclosure with supportive rather than punitive reactions increase the likelihood that victimised children will approach their parents. This family engagement dimension directly operationalises the mesosystem level of Bronfenbrenner's ecological model.

4.5 Implementation Across School Types

Kerala's school system comprises government schools, government-aided schools and unaided private schools, which differ in resource availability and governance. The framework is designed to be scalable across this spectrum. In well-resourced schools with reliable connectivity and trained teachers, full digital implementation using KITE platforms is feasible. The Samagra Shiksha monitoring dashboard provides a common accountability framework across school types, enabling district-level tracking of implementation quality and outcome indicators without imposing significant additional administrative burden.

4.6 Monitoring and Continuous Improvement

A prevention framework is only as strong as its monitoring system. Anonymous student surveys administered at the beginning, middle and end of each academic year provide data on exposure to hate speech, perceived school climate and bystander behaviour intentions. Quarterly review of incident reports, coupled with pattern analysis by grade, class and incident type, enables school leadership to direct targeted interventions where need is greatest. Schools participating in KITE's network can use the platform's data management capabilities to maintain anonymised records that inform continuous programme improvement without compromising student privacy.

External monitoring through district child protection units and UNICEF India's child safeguarding programme adds a layer of quality assurance, providing accountability to the prevention framework and credibility to the process beyond individual school boundaries.

The limitations of the proposed framework must be acknowledged explicitly. First, the model's feasibility depends on a level of institutional capacity – KITE infrastructure, DIET training capacity, SPC presence – that may not be uniformly available across all school types, particularly unaided private schools in rural areas. Second, the framework is primarily grounded in literature from Western, South Asian, and Kerala-specific contexts. Its components may require cultural adaptation before application in other Indian states with different social compositions, languages, or administrative structures. Third, the qualitative data underpinning the framework are practitioner-reported and subject to potential social desirability bias. These limitations do not undermine the framework's utility as an organising architecture for evidence-informed practice, but they do constrain the strength of claims that can be made about its generalisability and impact prior to independent evaluation.

5. Discussion

The programmes introduced in the schools of the State of Kerala are helpful in building resilience against hate speech and provide technological support systems to develop students' capacity to engage responsibly in digital environments. These include student skill development, positive peer influence, teacher preparation, family involvement and clear school policies. When these components work together they can reduce victimisation and improve student well-being. Kerala has strong institutional systems that can support such programmes. The real challenge is maintaining quality and consistency of implementation when they are scaled across many schools with different resource levels.

Evidence from pilot programmes and qualitative discussions shows that schools with systematic teacher orientation produced better outcomes than those without it. This suggests that training through the District Institutes of Education and Training (DIET) is a foundational, not optional, element of implementation.

Programmes designed only in English can exclude students who are already vulnerable to online abuse. Materials in Malayalam should go beyond simple translation; they must reflect the cultural experiences and reference points of students. Digital platforms change faster than school curricula. Because of this, programmes should focus on transferable analytical skills – the ability to recognise the structural patterns of hate speech – rather than platform-specific rules that become obsolete as platforms evolve. Classroom discussions on caste-related issues require particular care: teacher training must include structured reflection on personal attitudes and social positions to avoid inadvertently reproducing hierarchies the programme aims to address.

A concern that emerged from the qualitative discussions was the inequality in digital access: some students face online risks without adequate data connectivity support, meaning that the risk environment and the protective resources are unevenly distributed within the student population.

The current evidence base has limitations. Many Indian studies use small samples and short observation periods. The data from Kerala pilot programmes are useful but still limited in scope. More rigorous studies are needed, including larger school samples, validated research instruments, and follow-up over at least one academic year. Research should also examine hate speech separately from general cyberbullying, because identity-based attacks often involve different social, psychological, and normative dynamics.

6. Conclusions

Research shows that hate speech harms psychological well-being, reduces academic engagement and weakens

inter-group trust among students. However, Kerala's schools already have the systems and institutional support needed to address this problem if these efforts are organised within a clear and practical framework.

This paper proposes such a framework based on two conceptual models, existing research and discussion with educators. The approach has four main components: Education and Awareness, Critical Thinking Skills, Positive Online Behaviour and Safe and Inclusive Spaces. Together these elements help students understand harmful content, question misleading messages, respond responsibly online and support peers who face abuse. The aim is to build resilience and responsible digital citizenship among young learners.

The framework does not require new institutions or large financial investment. Instead, it calls for better coordination of programmes that already exist in Kerala's education system.

The distinctiveness of Kerala as a site for this framework lies in the institutional density and policy infrastructure already in place to address them. No other Indian state currently operates a comparable simultaneous combination of KITE's technical infrastructure for curriculum delivery, SCERT's curriculum development capacity, the SPC programme's civic character formation mandate, and the Suraksha Mitram support architecture within the same school system. This institutional convergence creates a genuine opportunity not only to implement but also to rigorously evaluate and iteratively refine a resilience model. This model could subsequently use in other states and international contexts where comparable. The framework presented here is offered as an initial architecture for that process of evidence-based development.

Several practical steps can move this effort forward. Schools can introduce a joint KITE–SCERT digital citizenship module for Classes 6 to 10 that clearly addresses hate speech. Teacher training through DIET programmes can support classroom discussions on digital behaviour. The Suraksha Mitram system can include clear reporting channels for online abuse. Monitoring indicators can also be added to Samagra Shiksha dashboards. Finally, schools should support rigorous research that evaluates these interventions across different types of institutions in Kerala.

Acknowledgements

All the authors are associated with this research. First author took a lead role in conceptualization, methodology, data curation, analysis and draft of the research. Second author supported editing and language correction. The third author associated with the refining of draft and final writing. All authors have read and approved the publication.

And there is no conflict of interest among them at any stage of this research. This research is not supported any funding from any agency. The technical support of the VJCRS, Vimal Jyothi Institute of Research is acknowledged.

All figures in this paper were developed by the author using standard diagrammatic design tools, and based on the relevant researches.

Informed Consent Statement

All subjects included in the study provided informed consent.

References

Afrouz, R., & Vassos, S. (2024). Adolescents' experiences of cyber-dating abuse and the pattern of abuse through technology, a scoping review. Trauma, Violence, & Abuse, 25(4), 2814-2828.

Akintayo, S. O. (2025). Impact of Social Media on Substance Abuse and Addiction among Nigerian Youth: A Call to Action for Awareness and Prevention. Available at SSRN 5599311.

Allport, G. W. (1954). The nature of prejudice. Addison-Wesley.

Bandura, A. (1977). Social learning theory. Prentice Hall.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative research in psychology*, *3*(2), 77-101.

Bilz, L., Fischer, S. M., Kansok-Dusche, J., Wachs, S., & Wettstein, A. (2024). Teachers' intervention strategies for handling hate-speech incidents in schools. Social Psychology of Education, 27(5), 2701-2724. https://doi.org/10.1007/s11218-024-09929-9

Bronfenbrenner, U. (1979). The ecology of human development: Experiments by nature and design. Harvard University Press.

Castellanos, M., Wettstein, A., Wachs, S., Kansok-Dusche, J., Ballaschk, C., Krause, N., & Bilz, L. (2023, April). Hate speech in adolescents: A binational study on prevalence and demographic differences. In Frontiers in Education (Vol. 8, p. 1076249). Frontiers Media SA. https://doi.org/10.3389/feduc.2023.1076249

Cedena-de-Lucas, B., Amate-García, M., Arco-Tirado, J. L., & Fernández-Martín, F. D. (2026). Service-learning as a strategy to prevent online hate speech perpetration in secondary education. International Journal of Educational Research, 137, 102964. https://doi.org/10.1016/j.ijer.2026.102964

Chakravarthi, B. R., Rajiakodi, S., Ponnusamy, R., Sivagnanam, B., Thakare, S. Y., & Thangasamy, S. (2025). Detecting caste and migration hate speech

in low-resource Tamil language: BR Chakravarthi et al. Language Resources and Evaluation, 59(3), 3051-3086. https://doi.org/10.1007/s10579-025-09848-x

Chen, S., Wei, M., Wang, X., Liao, J., Li, J., & Liu, Y. (2024). Competitive video game exposure increases aggression through impulsivity in Chinese adolescents: evidence from a multi-method study. Journal of youth and adolescence, 53(8), 1861-1874.

General Education Department, Government of Kerala. (2023). Curriculum framework for value education in Kerala schools. SCERT Kerala.

Hawke, J., & Puig Larrauri, H. (2026). Towards a third side on social media. Peacebuilding, 14(1), 14-29.

Joseph, G. V., & Thomas, K. A. (2020). Volatility of digital technology enabled learning through social media: educators' apprehensions. TEST Eng Manag, 82, 5832-9.

Joseph, G. V., Jose, D., Rajan, J., Shankaran, S., Vijay, D., & Navya, V. (2026). Management graduates' attitudes to green technology integration and AI tools for sustainable business practices. Journal of Learning for Development. https://doi.org/10.56059/jl4d.v13i1.2023

Jovanovic, S., & Markovic, L. (2023). Juvenile Offenders and Victims of Digital Violence. J. Crimin. & Crim. L., 61, 27.

Kerala General Education Department. (2025). Suraksha Mitram initiative for child protection in schools. Government of Kerala. https://www.education.kerala.gov.in/

Kerala Infrastructure and Technology for Education. (2024). ICT textbook revisions: Fake news detection modules for Classes 5 and 7. Government of Kerala.

Kerala Infrastructure and Technology for Education. (2025). First Bell: Digital class content for Kerala schools. Government of Kerala. https://firstbell.kite.kerala.gov.in

Kerala Infrastructure and Technology for Education. (n.d.). ICT-enabled education initiatives in Kerala. Government of Kerala. https://kite.kerala.gov.in

KITES (2026). Welcome to KITES. https://kite.kerala.gov.in/

Kowalski, R. M., Giumetti, G. W., Schroeder, A. N., & Lattanner, M. R. (2014). Bullying in the digital age: A critical review and meta-analysis of cyberbullying research among youth. Psychological Bulletin, 140(4), 1073–1137. https://doi.org/10.1037/a0035618

Krahé, B. (2025). The social psychology of aggression. Routledge.

Livingstone, S., & Helsper, E. J. (2010). Balancing opportunities and risks in teenagers' use of the internet: The role of online skills and internet self-efficacy. New Media & Society, 12(2), 309–329. https://doi.org/10.1177/1461444809342697

Ministry of Education, Government of India. (2020). National Education Policy 2020. Government of India.

Philip, R. (2023). Curbing Hate Speech through Education: Modalities for Equipping Future Teachers. Journal of Pedagogy and Education Science, 2(03), 209-220. https://doi.org/10.56741/jpes.v2i03.350

Pittman, S. K. (2023). Beliefs about aggression as mediators of relations between community violence exposure and aggressive behavior among adolescents: Review and recommendations. Clinical Child and Family Psychology Review, 26(1), 242-258.

Prasannan, P., Kumaresan, P. K., Rajiakodi, S., Subalalitha, C. N., & Chakravarthi, B. R. (2025). Counter-speech generation for homophobic and transphobic social media content in Malayalam. Social Network Analysis and Mining, 15(1), 87. https://doi.org/10.1007/s13278-025-01507-x

Ranjith, P. J., Vranda, M. N., & Kishore, M. T. (2023). Predictors, prevalence, and patterns of cyberbullying among school-going children and adolescents. Indian journal of psychiatry, 65(7), 720-728. https://doi.org/10.4103/indianjpsychiatry.indianjpsychiatry_313_23

Ranjith, P. J., Vranda, M. N., & Kishore, M. T. (2025). Development and validation of school-based intervention on cyberbullying for adolescents. Indian journal of psychiatry, 67(2), 252-255. https://doi.org/10.4103/indianjpsychiatry.indianjpsychiatry_777_23

Rudnicki, K., Vandebosch, H., Voué, P., & Poels, K. (2023). Systematic review of determinants and consequences of bystander interventions in online hate and cyberbullying among adults. Behaviour & Information Technology, 42(5), 527-544.

Singh, S., & Singh, S. (2025). Perceptions of Indian students towards bullying: Intervention through bullying intervention module (BIM). SAGE Open, 15(1). https://doi.org/10.1177/21582440241305199

State Council of Educational Research and Training Kerala. (n.d.). Academic guidelines and curriculum development programmes. https://scert.kerala.gov.in

Student Police Cadet. (n.d.). SPC: Overview. https://studentpolicecadet.info/overview/spc

Sunstein, C. R. (2017). #Republic: Divided democracy in the age of social media. Princeton University Press.

Suresh, S., & Vijaya, R. (2024). Mapping the literature on school bullying in India: A scoping review. Aggression and Violent Behavior, 77, 101978. https://doi.org/10.1016/j.avb.2024.101978

Törnberg, A., & Törnberg, P. (2024). From echo chambers to digital campfires: The making of an online community of hate in Stormfront. In Social processes of online hate (pp. 93-119). Routledge.

UNESCO. (2019). Digital literacy in education: Policy brief. United Nations Educational, Scientific and Cultural Organization. https://unesdoc.unesco.org/ark:/48223/pf0000369841

UNICEF India. (2020). Protecting children in the digital age: Challenges and responses in India. UNICEF. https://www.unicef.org/india

Vasconcellos-Silva, P. R., & Castiel, L. D. (2025). Chambers that echo hate, bubbles that distill fear: the constitution of Self and intolerance as roots of disinformation. Ciência & Saúde Coletiva, 30, e10492023.

Wachs, S., Schittenhelm, C., Kops, M., Gámez-Guadix, M., & Wright, M. F. (2025). Happiness Through HateLess? Examining the Direct and Indirect Effects of an Anti-Hate Speech Program on Victimized and Non-Victimized Youth. Journal of adolescence, 97(6), 1645-1655. https://doi.org/10.1002/jad.12525

Wachs, S., Wettstein, A., Bilz, L., Espelage, D. L., Wright, M. F., & Gámez-Guadix, M. (2024). Individual and contextual correlates of latent bystander profiles toward racist hate speech: A multilevel person-centered approach. Journal of youth and adolescence, 53(6), 1271-1286.

Zhu, C., Huang, S., Evans, R., & Zhang, W. (2021). Cyberbullying among adolescents and children: A comprehensive review of the global situation, risk factors, and preventive measures. Frontiers in Public Health, 9, Article 634909. https://doi.org/10.3389/fpubh.2021.634909