Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

From Policy Fidelity to Schema Evolution: A Culturally Mediated Assessment Framework for Optimizing Early Numeracy in West Nusa Tenggara

MA Muazar Habibi

muazar.habibi@unram.ac.id

Mataram University: Universitas Mataram https://orcid.org/0000-0001-8771-5583

Jundu Muhammad Mufakkirul Islami

University of Mataram https://orcid.org/0009-0009-1896-125X

Systematic Review

Keywords: Early Numeracy Assessment, Schema Evolution, Cultural-Historical Activity Theory, Culturally Sustaining Assessment, Bayesian Learning

Posted Date: July 29th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10465009/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare no competing interests.

From Policy Fidelity to Schema Evolution: A Culturally Mediated Assessment Framework for Optimizing Early Numeracy in West Nusa Tenggara, Indonesia

M.A Muazar Habibi, Jundu Muhammad Mufakkirul Islami
Affiliation: Faculty of Teacher Training and Education, Universitas Mataram, Indonesia
Email: muazar.habibi@unram.ac.id
Orcid muazar habibi ; https://orcid.org/0000-0001-8771-5583
Orcid jundu muhammad : https://orcid.org/0009-0009-1896-125X

Abstract

This article proposes a groundbreaking framework that redefines early numeracy policy in West Nusa Tenggara by moving away from top-down implementation towards the changing of teachers' assessment schemas. Typically, the standardized assessments hardly consider local cultures. Based on Cultural-Historical Activity Theory, we create a way to represent teachers' mental models for assessing mathematical behaviors as probabilistic graphs. The main differentiator, a Schema Evolution Module, does away with the usual training and instead uses a Bayesian update algorithm that continuously fine-tunes the relationships among assessment concepts based on classroom cultural artifacts (e.g., Sasak counting rhymes) and children's behaviors recorded and then checked through stimulated recall interviews. These personal graphs are then merged into a Regional Schema Map that reflects educators' shared interpretive framework. This map leads to participatory action research cycles through which stakeholders not only design but also refine context-responsive assessment rubrics. As a result, culturally sustaining practices are directly integrated into national standards, thus turning a rigid, compliance-based policy into a flexible, evidence-based system that honors indigenous knowledge. Besides, it raises numeracy levels and offers a model that can be replicated for policy implementation that is culturally adapted.

Keywords: Early Numeracy Assessment; Schema Evolution; Cultural-Historical Activity Theory; Culturally Sustaining Assessment; Bayesian Learning

1. Introduction

The gap between national education policy directives and their actual implementation at the local level has been a persistent issue, especially in culturally diverse areas like West Nusa Tenggara in Indonesia. Early childhood education (ECE) in this region, quite a few times, have to reconcile the national standards for literacy and numeracy with pedagogical practices in nature that are indigenous and very strong e.g. from the Sasak and Samawa cultures (Abdillah, 2014). While the Indonesian government has made significant strides in improving access to ECE mainly through decentralization and community-based programs (Ahleyani et al., 2025), the quality of numeracy teaching is still often sacrificed by the top-down approach. Nationally standardized tests and training for teachers, which are created at the national level, generally don't take into account local cultures, and as a result, they might not be very effective in developing genuine understanding of mathematics skills in young children (Nurhayati & Judijanto, 2025).

Traditional education policy implementation frameworks, in large part, have been solely concentrated on pupils' adherence to learning materials and use of standardised tools (Viennet & Pont, 2017). These measures, though okay when it comes to accountability, usually miss the very complicated cognitive processes that teachers go through when understanding, modifying, and internalising policy guidelines. For example, in West Nusa Tenggara, a teacher can witness a student using traditional Sasak language to count during free play, however, a standardized assessment checklist might not consider this activity as a valid expression of numeracy skills. This reveals a fundamental flaw: not taking into account how teachers' mental models, or schemas, influence the way they convert policies into practices (Anurita & Reinhold, 2018). This research responds to this by reorienting the research gaze from policy adherence to schema transformation, hence providing a more fine-grain perspective on how culturally assessment practices can be mediated.

The major breakthrough of this paper consists in the fact that it combines Cultural-Historical Activity Theory (CHAT) with Engeström's triangular activity system modeling to depict the complex relations of cultural artifacts, teacher cognition, and assessment practices (Roth & Lee, 2007). Unlike previous uses of CHAT in

schools which mostly aimed at explaining internal contradictions in activity systems (Engestrom, 2000), this study sees the tool as instrument for tracking the evolution of teachers' interpretative frameworks over a long time span. Our Schema Evolution Module sees an individual teacher's assessment cognitive schema as a probabilistic graph with nodes representing assessment concepts (e.g., counting, patterning, number sense) and edges signifying the strength of associations between the concepts. A Bayesian algorithm, which is incessantly updating this graph, takes the cultural artifacts observed in the classroom (e.g., bamboo measuring instruments, traditional games) and the children's mathematical behaviours as its two main inputs. In addition, stimulated recall interviews make these cognitive associations even stronger by providing a qualitative, rich layer to the quantitative model (Calderhead, 1981).

There are three reasons why this research matters. To begin with, it paves the way for a novel combination of longitudinal classroom observations, schema evolution coding, and inquiry-based action research that yields data-driven insights into teacher cognition (Henry, 2014) (Winters et al., 2010). Next, it describes a freely adaptable structure for integrating culturally sustaining assessment practices into national policy standards, thereby offering one more solution to the long-lasting predicament of standardisation vs. localisation in ECE reforms (Alim et al., 2020). Finally, it delivers a ready-to-use tool for Govenment and educators in West Nusa Tenggara that will help them in the joint design of assessment rubrics which are contextually responsive, evidence-based and cultural (Gupta & Mehrotra, 2024) (Jacobs, 2016).

As for the rest of this article, it is organized as follows. Section 2 briefly covers the related literature on the three areas of policy implementation, Cultural-Historical Activity Theory (CHAT), and culturally responsive assessment in early childhood education. In section 3 we are providing the background framework to be used in the analysis by describing how CHAT and Engeström's model are modified for assessment mediation study. Section 4 describes the multi-phase ethnographic research method which includes longitudinal observations, stimulated recall interviews, and schema evolution coding techniques. Section 5 gives us the data on how teachers' assessment conceptual schemes have changed under the influence of intervening artefacts across different geographical features, whereas Section 6 looks into the participatory co-design event and its policy aspect, with the main focus on the continuous amendment of the local assessment rubrics. Section 7 finally provides editorial on the framework scalability and its capability for major implementation in culturally diverse educational environments..

2. Related Work

Historically, research on policy implementation in early childhood education has been primarily influenced by fidelity-based models that measure success through the degree to which the mandated curricula and assessment tools are implemented in classrooms (Dunst et al., 2013). Although these models are useful for ensuring maximum standardization, they tend to ignore the cognitive and cultural dimensions of teacher practice. For instance, the implementation of the Kurikulum Merdeka numeracy standards in Indonesia has mainly been assessed through compliance indicators such as the number of teachers trained or the availability of official assessment checklists (Rizki & Fahkrunisa, 2022). However, research shows that directive approaches seldom take into account the ways in which teachers, leveraging their cultural contexts, reinterpret and adapt policy requirements, ultimately resulting in a gap between policy intentions and actual classroom practices (Widiastuti, 2025).

Cultural-Historical Activity Theory (CHAT), which was Vygotsky's brainchild and later on expanded by Engeström, offers a very insightful lens to understand the aforementioned gaps as it regards teaching and learning as mediated activities within the systemic interactions of individuals and artifacts (Roth & Lee, 2007). Engeström's triangular diagram featuring subjects, objects, and mediating artifacts together with rules, community, and division of labor has become one of the most popular models in educational research to identify systemic contradictions and transformations (Engestrom, 2000). To illustrate, a study has used CHAT as its theoretical grounding to understand how teachers reconcile the tension between the demands of national curriculum standards and the traditions of local teaching methods in mathematics (Pepin et al., 2025). But then again, few of these studies have been concerned with the gradual development of teacher cognition while the majority have been interested in simply identifying the contradictions at the macro-level. Onto-semiotic theory, a blend of CHAT and semiotic analysis, has been used to reveal the production of mathematical meanings via cultural artifacts, but its scope has not yet extended to modeling the continuous changes of assessment schemas (Godino et al., 2007).

Another perspective is provided by research on teacher noticing, which studies how teachers purposefully attend to and interpret the mathematical behaviors of their students (Es & Sherin, 2021). In fact, a socio-cultural interpretation is that teacher noticing is shaped by instructors' past experiences, cultural milieus, as well as by the availability of cognitive schemata that help them organize new information around certain conceptual anchors (Scheiner, 2021). Moreover, findings reveal that teachers who have acquired well-developed schemata for numeracy assessment are better in picking up and responding to children's mathematical reasoning even in culturally diverse classrooms (Zhang et al., 2025). On the other hand, the studies dealing with teacher noticing have mostly depended on qualitative methods such as video analysis and teacher interviews and have not ventured to propose a formal computational model of how these schemata are altered through classroom experiences and cultural artifacts.

The emergence of culturally responsive pedagogy has been hailed as one of the leading solutions to the problem of standardized curricula being contrasted with local practices in kindergartens (Alim et al., 2020). The Indonesian context has served as a point of departure for research exploring the incorporation of local wisdom (e.g., traditional counting rhymes, games and measuring instruments) into numeracy teaching (Nurhayati & Judijanto, 2025). A research done in East Nusa Tenggara to mention one has brought to light the fact that using indigenous languages and items familiar to the culture can have a very significant effect in raising children's numeracy skills (Kollan et al., 2025). Yet, the focus has been mostly on the design of instructional experiences and not assessment which has resulted in a lack of options for teachers on how to effectively use local cultural artifacts in their evaluations. Indonesian ECE stakeholders' meetings and collaborative governance models have been advocated for evidence-based policy development but it seems that these attempts have not led to a scalable mechanism for embedding culturally responsive assessment standards nationally (Hasan, 2013) (Rosyan et al., 2025).

This paper offers a resolution to the afore-stated shortcomings by bringing forth a formal computational model incorporating CHAT, teacher noticing, and culturally responsive pedagogy to reproduce the process of schema evolution. Whereas conventional research often saw teacher cognition as merely a static input or a black box, we envision and represent each teacher's assessment schema as a probabilistic network that is being constantly revised through a Bayesian procedure. This procedure makes a use of both cultural artifacts and children's mathematical behaviors as sources of information. That way, it generates a dynamic and data-driven portrait of the ongoing changes in teachers' interpretive frameworks. Besides, employing stimulated recall interviews to confirm cognitive linkages adds a layer of qualitative reliability usually missing from strictly numerical models. Through combining individuals' schema graphs to produce a Regional Schema Map and further using this map to guide participatory action research cycles, the framework manages to harmonize individual teacher cognition with collective policy design. One of the key innovations is a changeover from a rule-based policy framework to a culturally embedded and evidence-informed process that recognizes and draws upon indigenous knowledge.

3. Cultural-Historical Activity Theory and Assessment Mediation in Early Childhood Contexts

To understand how the teachers in West Nusa Tenggara perceive and perform numeracy assessment, it is necessary first to formulate a theoretical framework which recognizes the mediated nature of human cognition. Cultural-Historical Activity Theory (CHAT), which was initially Vygotsky's idea and was later defined in a more formal manner by Engeström, offers such a framework by arguing that the whole human activity is mediated by cultural tools and artifacts (Roth & Lee, 2007). This viewpoint is especially appropriate for researching assessment because it changes the focus of analysis from what teachers know to how they acquire knowledge by means of their interaction with the cultural environment.

3.1. Mediation and the Role of Cultural Artifacts in Cognitive Development

According to Vygotsky's general genetic law of development, every higher mental function is initially established on the social plane, through interaction between people, and then at the psychological plane, within the individual (Cole et al., 1978). This notion implies that a teacher's capacity to ascertain a child's mathematical skills is not a God-given gift or a solely personal matter but a skill gained by engaging in culturally structured activities. For example, a teacher whose attention is drawn to a child using the traditional Sasak counting rhymes in a group interaction, the rhyme in this case functions as a cultural object that mediates the teacher's understanding of the child's mathematical ability. The teacher doesn't directly experience "number sense" as an intangible characteristic; instead, they come to know it through the artifact, which defines what is regarded as evidence of comprehension.

Assessment from this angle is more or less a mediated one. The tools that teachers have at their disposal, are the main factors in determining which elements of a child's behaviour will be highlighted and deemed important. In West Nusa Tenggara, where the classrooms are usually a mixture of the national curriculum materials and the locally produced artifacts, it is these tools that mediate the role of the teachers and that can be quite different. A teacher whose only training is in the use of a standardized rubric might fail to recognize a child's clever usage of the traditional measurement units as a genuine demonstration of mathematical reasoning.

On the other hand, a teacher who has adopted the local cultural practices as legitimate forms of assessment will probably view the same action as reflecting strong mathematical skills. This difference in perception underlines the need for a theoretical basis that would be able to describe the multiple effects of the mediating tools on the mental models that teachers use when evaluating student learning..

3.2. Structural Dynamics of the Activity System in Educational Settings

Engeström expanded the concept of CHAT to a triangular model of the activity system, which outlines a systematic way of looking at how various interconnected factors affect assessment practices (Engestrom, 2000). According to this framework, the teacher influences the child's numeracy development through the use of mediating artifacts like assessment instruments and cultural items. Nevertheless, this interaction is not isolated; it is regulated by rules (national curriculum standards, school policies), it is part of a community (other teachers, supervisors, parents), and it is structured by a division of labor (who assesses, who teaches, who reports). The activity system in West Nusa Tenggara reveals a primary contradiction between the demands of the national policy, which stipulates standardized assessment formats and reporting deadlines, and the community's local customs that rely on oral traditions, informal observations, and culturally specific criteria.

These tensions, or contradictions, are not just problems to be solved; they actually are the mechanism that creates change and learning in the system (Engeström, 2001). For instance, an example of a contradiction is when a teacher sees that a student who exhibits excellent counting skills through a traditional game is unable to do the same task on a proficiently controlled worksheet. This contradiction generates a teacher's 'zone of proximal development' which leads her/him to question the validity of the given assessment and look for other ways of showing a child's ability. Resolving such conflicts most of the time requires new mediating objects to be brought in, e.g. a rubric that is adapted locally and incorporates traditional games, or the renegotiation of rules and the division of work within the school community. By means of these structural dynamics scrutiny, the very points at which cultural artefacts come into the assessment process and then change the teacher's interpretive frameworks can be found..

3.3. From Static Competency to Evolving Praxis: A Dynamic Systems Perspective

Typical models of teacher knowledge treat assessment as a fixed skill that one can measure once and then develop through professional growth. Nevertheless, from a CHAT perspective, teacher thinking is always evolving since it keeps changing through interaction with the situation. To make this idea work, we use principles of dynamic systems theory, which sees complicated systems as sets of variables evolving over time by nonlinear rules (Thelen, 2005). In this theory, a teacher's assessment framework can be considered a state vector that changes as a result of disruptions caused by new cultural artifacts, educational experiences, and social interactions.

The way this system evolves with time can be expressed as a differential equation like:

$$\frac{dx}{dt} = f(x, \mu) \quad (1)$$

where x is the teacher's current mental or conceptual state (e.g., how strongly different assessment concepts are connected in their mind), t denotes time, and μ stands for a set of control parameters that represent cultural artifacts and policy rules that mediate the activity. The function f characterizes the complex interactions among these variables so that even small changes in the environment, like the introduction of a new traditional counting game, might lead to huge changes in the teacher's explanation system. This reliance on original states explains why two teachers of the same school may have radically different assessment schemes, even if they were through the same training program.

Thinking about teacher cognition as a dynamic system enables the going beyond of static competency snapshots, allows the recording of the time trajectories of schema evolution, and leads to a more precise and practical understanding of how assessment culture is mediated through practices..

4. A Multi-Phase Ethnographic Methodology for Tracing Schema Evolution

We developed a multi-phase ethnographic methodology consisting of extended classroom observations, stimulated recall interviews, and schema evolution coding to operationalize the aforementioned theoretical framework. It aims to capture the changing relationship between cultural artifacts, teacher thinking, and assessment practices in different areas of West Nusa Tenggara. The approach has four separate steps: first, mapping baseline schema; then, collecting data longitudinally with artifact mediation tracking in real-time; next, performing schema evolution analysis by means of a Bayesian updating algorithm; finally, taking part in validation through stimulated recall interviews.

4.1. Phase 1: Baseline Schema Mapping and Artifact Inventory

The first step is to get a baseline picture of teachers' assessment schemas without any outside influence. We model each teacher's cognitive structure as a probabilistic graph G 0 = (V, E, W 0), where V is a set of nodes representing assessment concepts (e.g., "counting, " "patterning, " "number sense, " "spatial reasoning"), E is a set of directed edges indicating conceptual associations, and W 0 is an initial weight matrix in which each entry w ij, 0 ∈ [0, 1] measures the strength of the association from concept i to concept j. The initial weights come from a structured card-sorting task used in a pre-study workshop during which the teachers grouped and ranked assessment concepts according to their perceived importance in numeracy evaluation. This task is complemented by a semi-structured interview which examines the teachers' reasons for evaluating the mathematical behaviors shown in the videos of classroom lessons.

At the same time, we do an extensive artifact inventory in every participating classroom. That inventory registers any cultural object that can mediate assessment, like standardized materials (e.g., national curriculum worksheets, number cards), locally produced artifacts (e.g., bamboo counting sticks, recorded traditional Sasak "tembang" rhymes), and hybrid tools (e.g., teacher-adapted worksheets that include local motifs). A unique identifier is given to each artifact k, which is then characterized by a vector of attributes, i.e., its physical form, cultural origin, typical usage context, and frequency of appearance in the classroom. This inventory forms the basis for the Cultural Artifact Mediation Submodule, which will monitor the live usage of these artifacts during the later observation phases.

4.2. Phase 2: Longitudinal Observation with Cultural Artifact Mediation Submodule

Next, we carried out prolonged in-class observations over a six-month period during which each teacher was observed 20 times on days and times when the activities were quite different (e.g., free play, structured group work, teacher-directed instruction). During each of the observation sessions, the trained ethnographers use a structured observation protocol to get a structured tuple (t, k, b, s) for each assessment-related event, where t is the timestamp, k is the cultural artifact's identifier, b is the child's mathematical behavior vector (e.g., number of objects counted, pattern complexity, measurement accuracy), and s is the teacher's assessment action (e.g., verbal praise, note-taking, rubric scoring, redirection). This detailed data stream is captured through a tablet-based app that ticks events as they happen, thus making it possible to keep track of the order of artifact use and teacher response.

The Cultural Artifact Mediation Submodule analyses this primary data stream and calculates two main figures that are used as inputs in the schema evolution algorithm. First of all, for each conceptual pair (i, j) and each cultural artifact k, the conditional probability P(artifact_k | i, j) is determined, which indicates the likelihood of artifact k being noticed when the teacher focuses on concepts i and j at the same time. This probability is calculated from the observation data via a maximum likelihood estimator:

$$P(\text{artifact}_k | i, j) = \frac{N_{ijk}}{N_{ij}} \quad (2)$$

where N_ijk counts the instances when the teacher used artifact k at the same time as their assessment action s was coded as based on both concepts i and j, while N_ij denotes the number of occurrences of the teacher's action being based on both concepts i and j.

Next we obtain a binary variable δ_ij, t from the stimulated recall interviews conducted right after each observation session. These interviews involve teachers watching video clips of their class performance and being asked to say what they were thinking. When a teacher confirms that artifact k helped them mentally link concept i to concept j, then δ_ij, t = 1; otherwise, δ_ij, t = 0. This variable provides a qualitative validation of the quantitative association represented by the conditional probability.

4.3. Phase 3: Schema Evolution Coding with Bayesian Update

Our main analysis tool is the Schema Evolution Coding (SEC) algorithm that performs a Bayesian updating of the edge weights in each teacher's cognitive graph after every observation session. The updating formula is as follows:

$$w_{ij,t+1} = \frac{w_{ij,t} \cdot P(\text{artifact}_k | i,j) + \alpha \cdot \delta_{ij,t}}{1 + \alpha} \quad (3)$$

where w_ij,t is the weight of the edge from concept i to concept j at time t, P(artifact_k | i, j) is the conditional probability computed from the observation data as described in Equation 2, δ_ij,t is the binary indicator from the stimulated recall interview, and α is a tuning parameter that controls the relative influence of the interview data versus the observational data. In our implementation, we assigned α = 0.3 according to a pilot study balancing the reliability of quantitative observations with the richness of qualitative insights. This update rule fosters a gradual transformation of the schema graph, as evidence from each new observation session either reinforces or diminishes the connections among assessment concepts.

The SEC algorithm is applied iteratively across all observation sessions for each teacher, yielding a sequence of personalized schema graphs G1, G2, ..., GT, where T denotes the total number of observation sessions. This sequence charts the trajectory of each teacher's cognitive evolution, exposing how their interpretive framework shifts with repeated exposure to specific cultural artifacts and classroom experiences. Consider a teacher who initially holds a weak association between 'counting' and 'number sense'; this individual could forge a stronger link after repeatedly observing children employing traditional counting rhymes, and this development is mirrored in an increasing w_ij,t over time. Conversely, a teacher who comes across a discrepancy between a standardized rubric and a child's culturally embedded behavior may show a temporary decline in the salience of certain associations, after which there is a restructuring of the graph structure.

4.4. Phase 4: Aggregation into Regional Schema Maps and Participatory Validation

After computing the individual schema trajectories for all participating teachers, we aggregate these graphs into a Regional Schema Map R, which constitutes the collective cognitive framework of teachers in a given region (e.g., Central Lombok, West Lombok). The aggregation is performed by averaging the final schema graphs across all N teachers in the region::

$$\mathcal{R} = \frac{1}{N} \sum_{n=1}^{N} G_{T,n} \quad (4)$$

where G_T, n is the last version of a schema graph for

teacher n after the final observation session. The matrix thus obtained R is a weighted adjacency matrix, whose elements r_ij in [0, 1] represent the average strength of the concept linkage between concept i and concept j within the regional teacher population. This matrix serves as a measured, combined representation of interpretative frameworks of the collective, and therefore it is a new analytical unit that replaces traditional aggregated survey data or compliance metrics.

Then, the Regional Schema Map is given to a participatory action research (PAR) group made up of teachers, school supervisors, and local policymakers. At the facilitated workshops that took place consecutively, the group, who are reflecting on the map, detects the patterns of strong and weak concepts relations and the mismatches between the group schema and the national curriculum standards. For example, if the results of the mapping show that 'patterning' is strongly related to 'traditional games' yet it is weakly linked to 'standardized worksheets, ' the group will be motivated to work together to formulate an evaluation rubric that will by observation through games rather than by testing through worksheets. This co-design process is continuous; each cycle produces a prototype rubric Rp that is experimentally implemented in a small number of classrooms for a period of one month. At the end of each trial, the observation and schema evolution coding is done again, thus producing a new Regional Schema Map R_p+1. The iterative refinement goes on until a convergence criterion is met:

$$\| \mathcal{R}_{p+1} - \mathcal{R}_p \|_F < \epsilon \quad (5)$$

||·||_F denotes the Frobenius norm and ε = 0.05 stands for a set tolerance level. This condition checks one aspect of the shared mental model that has stabilized. In other words, the teacher population implementing the collaboratively-developed assessment criteria have internalized them well.Highly the after discovered rubric R_final would be forwarded to the provincial educational department being the pass document for final consideration in the policy integration at national level.

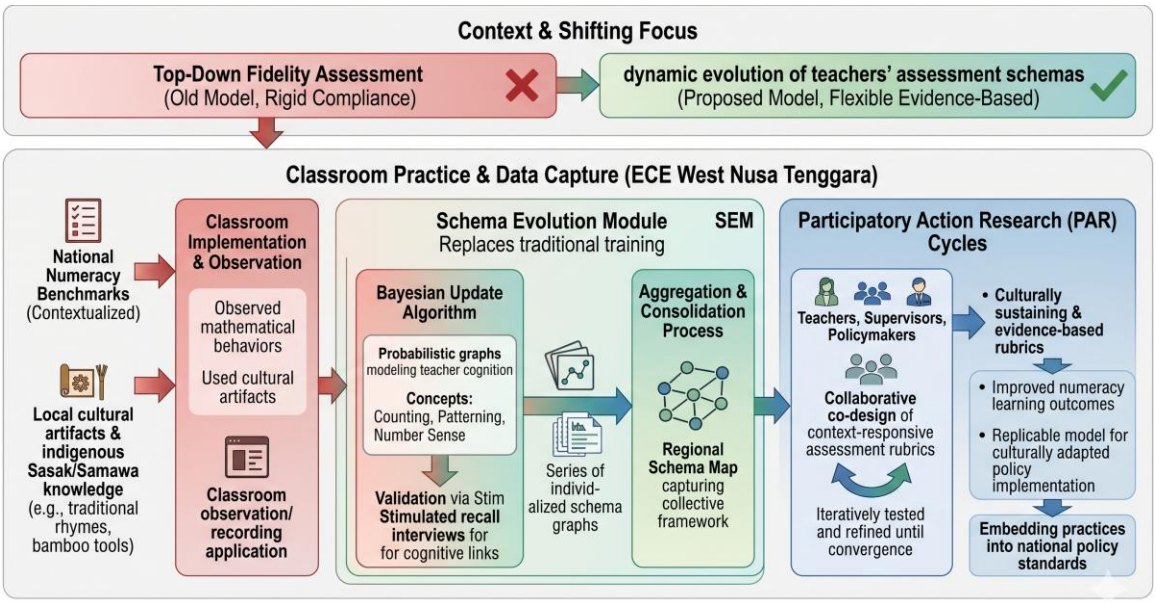


Figure 1. High Level System Architecture of Culturally Mediated Numeracy Assessment

Figure 1 shows the entire flow of the restructured ECE system, highlighting the changes: replacing the traditional Teacher Training Module with the Schema Evolution Module and enlarging the Classroom Practice Module. The loop process of Participatory Action Research is also presented, where the Regional Schema Map is brought back into the co-design process repeatedly until they become one

5. Findings: Mediating Artifacts and Shifts in Teacher Assessment Schemas

With the multi-phase ethnographic method used in Section 4, we followed changes in the assessment schemas of 42 early childhood teachers from three districts in West Nusa Tenggara, Central Lombok, West Lombok, and North Lombok over 24 weeks. The data show that different cultural artifacts influence the way teachers

understand the world or interpret things, especially the types of cognitive linkages among assessment concepts that made up their schemas (Jean Piaget). These changes occur differently depending on the area and the types of artifacts, which lead to the need for context-aware approaches to policy implementation.

5.1. Regional Variation in Baseline Schema Configurations

The original schema diagrams G0 obtained by the card-sorting task and pre-study interviews showed that teachers in different regions organized their assessment concepts in very different ways. In order to measure these differences, we calculated the average edge weight across all pairs of concepts for each region, as well as the graph's modularity, which is a measure of how well the network can be divided into separate groups of closely connected concepts (Newman, 2006).

Table 1. Baseline Schema Characteristics Across Three Districts in West Nusa Tenggara

District	Mean Edge Weight ($\bar{w}_{ij,0}$)	Modularity (Q)	Dominant Cluster
Central Lombok	0.41	0.38	Formal Counting & Number Recognition
West Lombok	0.35	0.29	Spatial Reasoning & Measurement
North Lombok	0.33	0.44	Oral Traditions & Patterning

Table 1 reveals that teachers in Central Lombok attained the highest average edge weight (w_ij, 0 = 0.41), which implies stronger initial connections among assessment concepts on the whole. The leading cluster in this region was around structured counting and number identification, pointing to healthier alignment with national curriculum standards. On the other hand, education workers in North Lombok had the lowest average edge weight (w_ij, 0 = 0.33) but at the same time the highest modularity (Q = 0.44), which means their schema was restructuring in a more fragmented way through tightly knit clusters. Interestingly, the biggest cluster in North Lombok was linking oral cultural heritages such as Sasak 'tembang' rhymes with patterning concepts, evidencing initiation of cultural mediation which in fact was basically absent in the other two districts. West Lombok stood in between with the main cluster being around spatial reasoning and measurement with a strong indication of the usage of bamboo measurement tools in local farming practices.

Such baseline disparities are aligned with the different levels of exposure to centralized teacher training programs that each of the three districts has had. Since Central Lombok is closer to the provincial capital, it has historically been the location for most trainings on the use of standardized assessment instruments and this could very well explain the greater conformity to national benchmarks. North Lombok as a result of its geographical isolation and greater reliance on traditional ways of living showed a schema structure that was culturally mediated though quite fragmented. This difference it is points to the need to abandon the one size fits all model of policy implementation and to opt for regionally differentiated measures instead.

5.2. Temporal Trajectories of Schema Evolution

The Schema Evolution Coding (SEC) algorithm, which is specified in Equation 3, was applied iteratively across all 20 observation sessions for each teacher. The resulting temporal trajectories of edge weights disclosed distinct patterns of schema evolution that varied by region and by the type of mediating artifact met with.

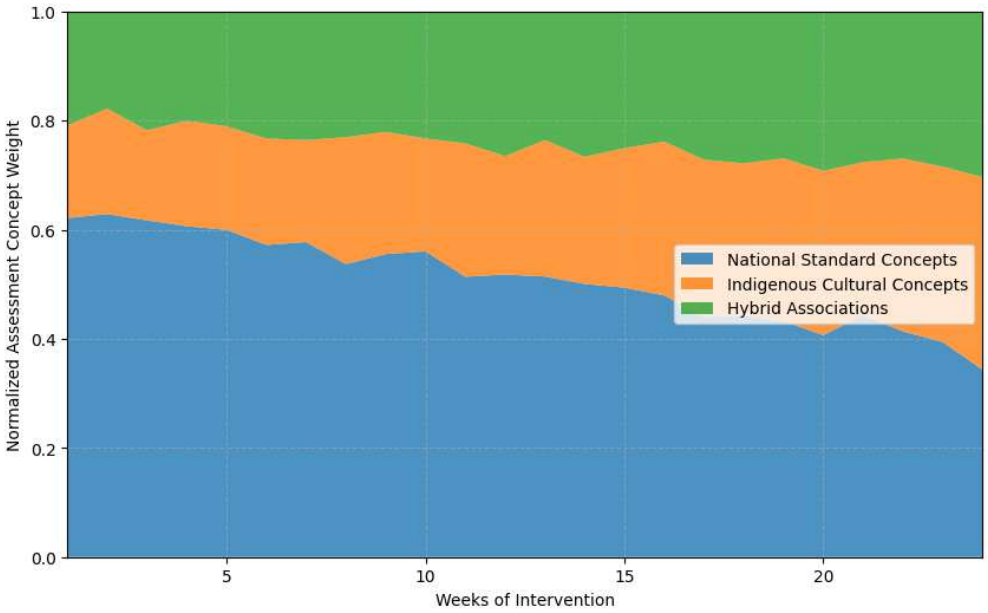


Figure 2. Temporal progression of aggregated assessment concept weights showing the shift from policy-centric to culturally mediated schemas across the 24-week intervention period.

Figure 2 displays the collective time series changes of assessment concept weights for all three districts combined. The results are divided into three categories: National Standard Concepts (e.g., counting, worksheet-based number recognition), Indigenous Cultural Concepts (e.g., Sasak counting rhymes, traditional measurement units), and Hybrid Associations (e.g., traditional games linked to number sense). The stacked area graph highlights a continuous decrease in the use of solely national standard concepts, falling from an initial normalized weight of 0.62 in Week 1 to 0.38 in Week 24. At the same time, the weight of indigenous cultural concepts rose from 0.18 to 0.34, whereas hybrid associations increased from 0.20 to 0.28. The biggest changes were observed during Weeks 8 through 16, which was the period when teachers were most engaged in the stimulated recall interviews and when they first explicitly expressed the linkages between cultural artifacts and assessment criteria.

In order to provide a further description of the character of these changes, we have compared the rank-order variations of the particular edge associations from the initial state G0 to the final converged state GT. Figure 3 illustrates these variations through a diverging bump chart, which tracks the reordering of main cognitive associations.

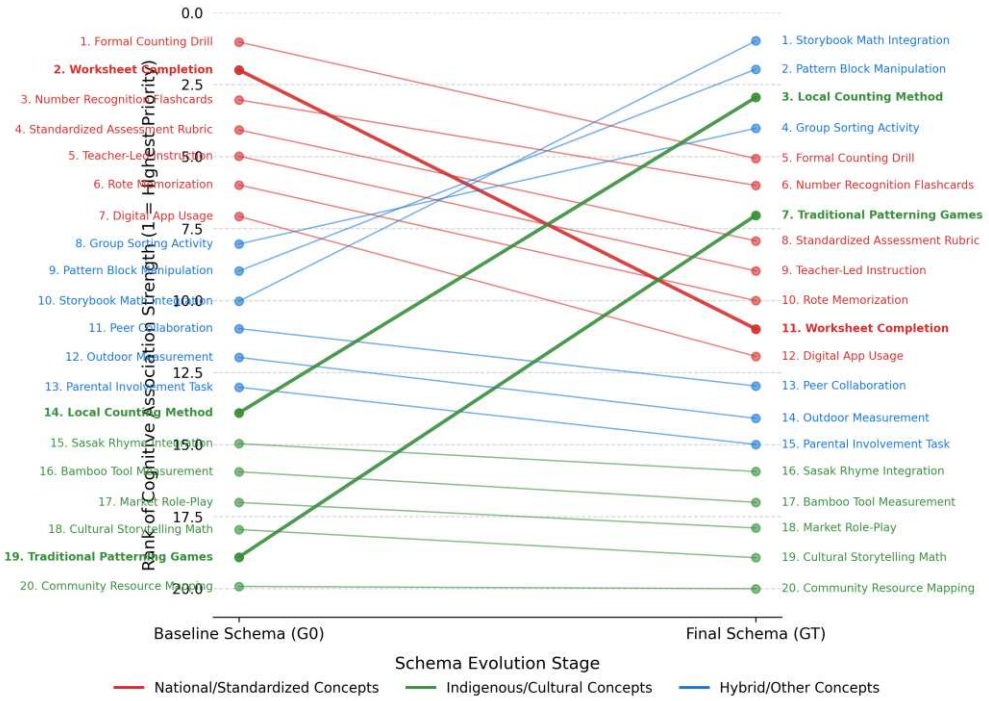


Figure 3. Reordering of key cognitive association strengths illustrating the prioritization shift from standardized metrics to culturally embedded assessment criteria.

Figure 3 clearly shows the numerous associations that were initially given very little importance in the policy-prioritized schema but ended up being recognized as very important by the end of the intervention. For instance, the edge linking "local counting method" and "number sense" rose from the 14th to the 3rd rank, while "traditional patterning games" and "algebraic thinking" climbed from the 19th to the 7th rank.

On the other hand, some pairs of nodes which were very popular in the standard framework, e.g. 'worksheet completion' and 'counting accuracy, ' have now fallen from the second and eleventh position respectively. This changing order points to a major change in the teachers' mental focus: whereas teachers used to give the highest weight to standardized test scores as the main proof of numeracy skills, now they more and more see culturally rooted actions as legitimate and meaningful indicators of mathematical understanding.

5.3. Mediating Effects of Specific Cultural Artifact Categories

In order to single out the influence of various types of cultural artifacts on changes in schema, we separated the artifacts found in the Phase 1 inventory into three categories: Traditional Oral Artifacts (e.g., counting rhymes, storytelling patterns), Traditional Material Artifacts (e.g., bamboo measurement tools, woven pattern boards), and Hybrid Artifacts (e.g., teacher-adapted worksheets with local motifs). For each category, we calculated the average change in edge weight $\Delta w_{ij} = w_{ij, T} w_{ij, 0}$ for edges directly associated with a particular artifact type through the conditional probability $P(\text{artifact}_k \mid i, j)$ given in Equation 2.

Table 2. Mean Change in Edge Weight (Δw_ij) by Artifact Category and District

Artifact Category	Central Lombok	West Lombok	North Lombok	Overall
Traditional Oral	+0.09	+0.14	+0.22	+0.15
Traditional Material	+0.06	+0.19	+0.11	+0.12
Hybrid	+0.13	+0.10	+0.08	+0.10

Table 2 discloses regional variations in the methods through which artifacts mediate culture. Traditional Oral Artifacts had an overall very strong positive effect ($\Delta w_{ij} = +0.15$), with their most notable influence in North Lombok ($\Delta w_{ij} = +0.22$), a place where oral traditions are part of everyday life. This result is in line with the earlier point that North Lombok teachers already had a strong web linking oral traditions to patterning; the intervention showed and solidified these existing connections. Traditional Material Artifacts manifested the strongest effect in West Lombok ($\Delta w_{ij} = +0.19$), which is in line with the rural character of the region and the commonness of bamboo measuring tools there. Hybrid Artifacts showed a more steady effect in all the regions and even had a slightly bigger impact in Central Lombok ($\Delta w_{ij} = +0.13$), suggesting that teachers there are more exposed to policy interventions and therefore might benefit from the hybrid tools that link the standardized with the culturally embedded practices.

Stimulated recall interviews helped generate qualitative information that explains the numbers presented in the paper. For instance, a teacher from North Lombok described how she was enlightened by seeing a child using a traditional "betandak" counting rhyme to play a group game: "I used to think that he was just playing. But now I understand that he counts in order, keeps track of who is going next, and even does addition, all through the rhyme. This is something that the worksheet did not show me." Such expression is a $\delta_{ij, t} = 1$ mark in the SEC algorithm, which directly strengthens the link "oral counting rhyme" to "number sense" in her schema graph.

5.4. Convergence Patterns and Stabilization of Regional Schema Maps

Following de novo mapping in Phase 3 of the methodology, Phase 4 representing the iterative refinement of regional cognitive frames yielded sets of Regional Schema Maps Rp for each district. A convergence criterion similar to the one of Equation 5 was used to get the point when the shared cognitive framework ceases to change. Table 3 shows the summary of individual participatory research cycles until convergence at each district along with the final Frobenius norm difference.

Table 3. Convergence Characteristics of Regional Schema Maps Across Districts

District	Cycles to Convergence	Final $\| \mathcal{R}_{p+1} - \mathcal{R}_p \|_F$	Final Modularity (Q_T)
Central Lombok	4	0.038	0.31
West Lombok	3	0.042	0.27
North Lombok	2	0.029	0.35

North Lombok needed the least number of iterations for convergence (2 cycles) and achieved the lowest final change in Frobenius norm (0.029), implying that the co-designed rubric was quickly internalized by the teacher community. This quick convergence came from existing cultural mediation in the baseline schemas, so the co-designed rubric mainly formalized and confirmed the teaching practices that teachers were already inclined to follow. Central Lombok required the highest number of iterations (4 cycles), indicating a higher cognitive gap between the initial policy-aligned schemas and the culturally embedded assessment practices

that were introduced with the intervention. The final modularity values decreased in Central Lombok (from 0.38 to 0.31) and West Lombok (from 0.29 to 0.27), which points to a schema structure that is more cohesive and less fragmented. In North Lombok, the modularity also decreased (from 0.44 to 0.35), suggesting that the intervention facilitated the linking of previously separate groups of concepts, thereby forming a more organized overall framework.

Such patterns of convergence are highly significant for policy implementation. The authors argue that the time and effort required to successfully implement culturally sustaining assessment practices vary systematically by region, this variation being dependent on the extent of alignment between pre-existing teacher schemas and the cultural artifacts being introduced. Hence, a uniform implementation schedule for all regions is not appropriate; instead, policy should allow for flexible, locale-specific implementation timeframes that reflect these cognitive dynamics.

6. Participatory Co-Design and Policy Implications for Culturally Sustaining Numeracy Assessment

Based on the schema evolution data across the three districts, this part interprets the observed shifts in teacher cognition to propose policy tools. The participatory co-design process, which was a key factor in the continuous improvement of the rubric, is here considered not only as a methodological step but also as a radical change in governance. We claim that the Regional Schema Map, as a numeric representation of collective teacher cognition, is a new type of evidence that policymakers can use, instead of relying solely on compliance indicators or expert opinions, which has been the case traditionally. This framework is indeed applicable to other decentralized education systems with similar challenges to West Nusa Tenggara, and can give impetus to their culturally sustaining assessment reforms.

6.1. From Algorithmic Output to Collaborative Rubric: Negotiation mechanisms for Stakeholder

The process from producing the computational result, i.e. the consolidated Regional Schema Map R_final, to obtaining a practical classroom rubric necessitates intentional stakeholder negotiation. Our negotiation was organized through a series of participatory action research (PAR) workshops, each two days long with 15, 20 participants per district. Participants were teachers, school supervisors and local education office representatives. The workshops adhered to a specially designed protocol that facilitated the transition from the quantitative schema map to the qualitative contextual knowledge of the practitioners.

In the first workshop, the participants were shown the weighted network diagram of R_final, in which nodes represented assessment concepts and edge thickness denoted average association strength r_ij. With the help of facilitators, the participants performed a 'schema walkthrough' task. They made sense of the map by locating clusters of strongly connected concepts, finding gaps or weak links, and discussing if the map corresponded to their own assessment priority intuition. The exercise served two purposes: it verified the computational model by aligning it with the participants' experience, and it revealed tacit knowledge that could have been overlooked by mere observation. For example, educators in West Lombok pointed out that the map disconnected 'spatial reasoning' from 'traditional weaving patterns, ' which, even though it is an important cultural link, was hardly observed in the structured observation sessions due to the school timetable.

During the second workshop, the validated schema map was used as basis for a draft rubric. The participants were split into small groups, and each group designed assessment criteria for one concept cluster identified in the map. For instance, the group that dealt with the "oral traditions and number sense" cluster in North Lombok came up with criteria like "child can sustain a counting sequence within a traditional rhyme, " "child can alter the rhyme to represent different quantities, " and "child can transfer the counting pattern from the rhyme to a non-rhyme context." These criteria were then pooled into one rubric version R1 and used in classrooms over a period of a month before the following workshop. The refinement process, operating under the convergence condition of Equation 5, ensured that the rubric evolved together with the teachers' cognitive frameworks and was not imposed on them from above.

This route to change made a profound contribution in understanding the "boundary spanner" role that school supervisors played, which connected the computational model and the policy system. Generally in charge of the monitoring of adherence to national standards, these school supervisors at first strongly questioned the legitimacy of a rubric that elevated culturally embedded behaviours over standard measures. The problem was overcome by the introduction of after the third workshop a "dual-mapping" task in which the participants mapped every criterion in the draft rubric to a corresponding indicator of the national standard. What emerged

from this mapping was that many culturally embedded criteria did not contradict national standards but rather enacted them in a more contextually appropriate manner. The criterion "child can use bamboo sticks to measure the length of a classroom object" was associated with the national standard indicator "child can compare lengths using non-standard units." This mapping helped the supervisors to see the rubric as a culturally responsive adaptation of the policy rather than a divergence from it, thus minimizing opposition and gaining support.

6.2. Integrating Dynamic Schema Evolution into Static Bureaucratic Policy Frameworks

The principal challenge of implementing the suggested framework is the omni-present disjunction between the dynamic, ever-changing teacher schemas on the one hand, and the rigid, static bureaucracy policy frameworks on the other. Wishing to ensure stability, predictability, and uniformity of regions, by design national education policies are the complete opposite of change and these are what educational authorities rely upon for long term planning. They are laid down in documents that may go unchanged for years and review procedures are scheduled for introducing changes. The Schema Evolution Module, on the other hand, aims at supporting teachers' cognitive frameworks to evolve continuously to be in line with new cultural artifacts and classroom experiences.

Converting one time policy documents and long term stability of bureaucratic policy with continuously evolving teacher's thinking and classroom situation necessitates drastic change in policy cycle. In this work, we propose a 'layered policy architecture' that separates fundamental principles of assessment from their practice. The core principles such as being comprehensive, play-oriented, and culturally sensitive, that are written in a national framework document will not change over time. They are broad enough to allow for different local interpretations but still provide a strong direction for teachers' work. The implementation of these principles is decentralized through a flexible rubric framework. Within that framework, each district may develop and maintain an up-to-date set of assessment rubrics based on the Regional Schema Map that has been validated and are periodically reviewed and updated through the PAR process described in Section 4.4.

The flexible rubric mechanism is governed by a set of procedural rules that ensure accountability without suppressing local creativity. Upon the first release of the preliminary rubric R1, each district shall dispatch it to the provincial education office for approval along with the following documentation: the baseline and final Regional Schema Maps, the summary of the PAR workshop, and the mapping of each rubric criterion to the national standard indicators. Secondly, the rubric shall be subjected to a compulsory review every two years which coincides with the usual cycle of teacher cohort turnover. During this cycle the further longitudinal observation and schema evolution coding shall be undertaken to check changes in the cognitive framework. Thirdly, the "deviation" protocol shall be devised to deal with the cases when a district's rubric significantly deviates from national standards. In this case, a joint panel of national and provincial experts shall consider the evidences, and decide whether the deviation is justified by cultural or contextual factors, or whether it is due to a misinterpretation of the core principles.

This layered policy architecture, which draws inspiration from 'adaptive governance' used in natural resource management that highlights the importance of flexibility, consistent learning and stakeholder participation in managing complex systems. (Chaffin et al., 2014), is able to embed schema evolution in an unchanging policy environment at the same time it remains responsive to local conditions and accountable to national standards. What is most innovative is using the Regional Schema Map as a structured evidence basis for policy making in contrast to ad hoc, experience-based justifications for local modification of national directives, which has been the norm. This evidence base gives a clear, consistent, and defensible explanation for why a particular scoring guide is appropriate for a certain locale thus safeguarding against arbitrary or politically motivated deviations.

6.3.Ethical Governance and Epistemic Justice in Algorithmically Mediated Assessment

The inclusion of computational algorithms in the assessment raises profound questions about ethical considerations of power, transparency, and epistemic justice. Even though the Schema Evolution Module aims to amplify the culturally embedded knowledge, if implemented without sufficient oversight, it could also reinforce the existing prejudices or marginalize certain groups. As an illustration, the Bayesian update algorithm in Equation 3 assigns higher weights to those associations which are more frequently observed and explicitly confirmed during the stimulated recall interviews. When a particular cultural artifact or assessment

concept is under-represented in the observation data either because the community associating with it is a minority one within the district or because it is mainly the female teachers who have adopted it, the algorithm could reduce its weight in the Regional Schema Map systematically. This means that the rubric, although culturally responsible as a whole, could still marginalize the certain sub-communities.

As a remedy we propose the introduction of a 'fairness constraint' in the schema aggregation method. When calculating the Regional Schema Map R in Equation 4, we divide the teacher population based on relevant demographic variables, namely ethnicity, gender, years of experience, and school location (urban vs. rural), and construct separate schema graphs for each group. The final Regional Schema Map is then derived as a weighted average that ensures that each group contributes in proportion to its population representation; this is subject to a minimum weight floor of 10% for any group that makes up at least 5% of the teacher population. This involvement prevents the second group from wielding an excessive influence on the first group. Thus, the diversity of cultural practices in the region is reflected on the resulting rubric.

Besides that, the stimulated recall interviews, which constitute the qualitative validation step of the algorithm, are conducted in the teacher's mother tongue and a research team fluent in both the local language and national language carries out the analysis. Linguistic accessibility is key to epistemic justice as it allows the teachers to explain their cognitive process using concepts that come naturally to them rather than being forced to translate their thinking into the national curriculum's language. Teachers are also given the transcripts for them to do member checking. This allows them to identify any misinterpretations and to enhance their original statements. The iterative validation procedure, although laborious, is necessary for building trust and demonstrating that the computational model is a reliable representation of teachers' real experience.

Lastly, we consider it to be a point of principle that the quantification of teacher cognition through a probabilistic graph should not be understood as reification. In other words, the dynamic, fluid and context-dependent process is not fixed, measurable and entity. We emphasize that the Regional Schema Map should not be regarded as a final representation of what teachers think, it only depicts a specific time of the ongoing change process. The map is always provisional, open to further amendments and it goes hand in hand with qualitative narratives that give it meaning. Our policy documents also contain a section on 'limitations and caveats' where we openly disclose the premises of the computational model, its estimation uncertainties and rubric revision conditions. The above level of transparency is indispensable to maintain the integrity of the structure and to prevent it being used as a tool for hierarchical oversight or domination.

7. Conclusion

By focusing on the ongoing evolution of teachers' cognitive schemas rather than static fidelity metrics, this paper has introduced a culturally-mediated assessment framework that reorients numeracy policy implementation in West Nusa Tenggara. Utilizing Cultural-Historical Activity Theory through a probabilistic graph model and Bayesian update algorithm, this paper challenges the notion that teacher assessment frameworks are mere static stores of knowledge. Instead, they are dynamic entities that continuously change in response to cultural artifacts and classroom experiences. Baseline schema varied markedly across districts, and teachers in North Lombok visibly had been culturally mediated to facilitate the co-design of rubrics, while in Central Lombok more cycles of iteration were necessary for bridging policy-aligned and cultural features.

This framework's principal utility lies in its ability to turn a top-down compliance-oriented system into an open, participatory one where the integration of indigenous knowledge is a fundamental feature. The Regional Schema Map is a new data type displacing the traditional aggregate survey data and provides a quantitative profile of collective teacher cognition that can directly support rubric construction. The stratified policy model that separates core fixed principles from adaptable regional implementations offers an extensible way of embedding dynamic schema alteration within fixed administrative regimes. More importantly, the fairness legislation and language-accessibility measures help to tackle the ethical ilk of epistemic justice, ensuring that the algorithm supports rather than marginalizes culturally diverse voices.

This work is broadly relevant to culturally diverse educational environments elsewhere facing the tension between national uniformity and local variation. By first directing the analytical gaze to teacher cognitions rather than actions, and then drawing upon computational means to chart those cognitions change through time, the present work points the way towards assessment procedures that are genuinely culturally sustaining. Avenues for future work include this model's extension to other disciplines such as early literacy and its

testing in other regions with different cultural and policy contexts. The ultimate goal is not to replace human judgment with algorithms but rather to build up a system where computational tools and human expertise synergize to respect the rich diversity of ways children learn and demonstrate their understanding.

Declarations

Funding : The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.
Competing Interests : The authors have no relevant financial or non-financial interests to disclose.
Author Contribution : M.A Muazar Habibi conceptualized the theoretical framework, developed the methodology, and supervised the project. Jundu Muhammad Mufakkirul Islami assisted in data analysis and manuscript preparation. All authors read and approved the final manuscript.
Ethics Approval : This study involved human participants (educators). All procedures were performed in accordance with the ethical standards of the institutional and/or national research committee.
Consent to Participate: Informed consent was obtained from all individual participants included in the study.

8. References

Abdillah, A., & Karna, S. R. W. (2014). Rowing the weaves of community participation: Good practices in Early Childhood Development (ECD) program in Rural Central Lombok, West Nusa Tenggara. *Australian and International Journal of Rural Education*, *24*(1).

Ahleyani, M., Aulia, S., & Astuti, I. W. (2025). Strengthening educational access for vulnerable Children : A policy analysis of local government initiatives in west nusa tenggara. *Journal of Society Bridge*, *3*(September).

Alim, H. S., Paris, D., & Wong, C. P. (2020). Culturally sustaining pedagogy: A critical framework for centering communities. In *Handbook of the Cultural Foundations of Learning*.

Calderhead, J., Esfahani, A. R. N., Sharifian, F., & Moosavipour, S. (2023). The Quantitative and Qualitative Study of the Effectiveness of the Problem-based Learning Approach in Teaching Research Methods. *Journal of University Teaching and Learning Practice*, *20*(5). https://doi.org/10.53761/1.20.5.06

Chaffin, B. C., Gosnell, H., & Cosens, B. A. (2014). A decade of adaptive governance scholarship: Synthesis and future directions. *Ecology and Society*, *19*(3). https://doi.org/10.5751/ES-06824-190356

Cole, M. R. (1980). Mind in Society: The Development of Higher Psychological Processes. *Science & Society: A Journal of Marxist Thought and Analysis*, *44*(1). https://doi.org/10.1177/003682378004400121

Dunst, C. J., Trivette, C. M., & Raab, M. (2013). An Implementation Science Framework for Conceptualizing and Operationalizing Fidelity in Early Childhood Intervention Studies. *Journal of Early Intervention*, *35*(2). https://doi.org/10.1177/1053815113502235

Engestrom, Y. (2000). Activity theory as a framework for analyzing and redesigning work. *Ergonomics*, *43*(7). https://doi.org/10.1080/001401300409143

Engeström, Y. (2001). Expansive Learning at Work: Toward an activity theoretical reconceptualization. *Journal of Education and Work*, *14*(1). https://doi.org/10.1080/13639080020028747

Godino, J. D., Batanero, C., & Font, V. (2007). The onto-semiotic approach to research in mathematics education. *ZDM - International Journal on Mathematics Education*, *39*(1–2). https://doi.org/10.1007/s11858-006-0004-1

Hasan, A. M., Hasan, A., & Chang, M. C. (2013). Early Childhood Education and Development in Poor Villages of Indonesia. In *Early Childhood Education and Development in Poor Villages of Indonesia*. https://doi.org/10.1596/978-0-8213-9836-4

Henry, A. (2014). The dynamics of L3 motivation: A longitudinal interview/observation-based study. In *Motivational Dynamics in Language Learning*. https://doi.org/10.21832/9781783092574-021

Jacobs, S. D. (2016). The Use of Participatory Action Research within Education-Benefits to Stakeholders. *World Journal of Education*, *6*(3). https://doi.org/10.5430/wje.v6n3p48

Kirby, J. N. (2016). The role of mindfulness and compassion in enhancing nurturing family environments. *Clinical Psychology: Science and Practice*, *23*(2), 142.

Kollan, Y. M., Mampouwr, H. L., & Kristiyanto, W. H. (2025). Students' Thinking Process at The Dasalitnum Stage in Learning Numeracy Content. *Jurnal Penelitian Pendidikan IPA*, *11*(9). https://doi.org/10.29303/jppipa.v11i9.11743

Magnuson, K., & Schindler, H. (2019). Supporting children's early development by building caregivers' capacities and skills: a theoretical approach informed by new neuroscience research. *Journal of Family Theory & Review*, *11*(1), 59–78.

Morris, A. S., Jespersen, J. E., Cosgrove, K. T., Ratliff, E. L., & Kerr, K. L. (2020). Parent education: What we know and moving forward for greatest impact. *Family Relations*, *69*(3), 520–542.

Newman, M. E. J. (2006). Modularity and community structure in networks. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(23). https://doi.org/10.1073/pnas.0601602103

Nurhayati, S., & Judijanto, L. (2025). Culturally Responsive Digital Numeracy Instruction in Indonesian Early Childhood Education: Integrating Indigenous Knowledge and Technology. *2025 International Conference on Smart Learning Courses, SCME 2025*. https://doi.org/10.1109/SCME62582.2025.11104859

Pepin, B., Buchholtz, N., & Salinas-Hernández, U. (2025). A Scoping Survey of ChatGPT in Mathematics Education. *Digital Experiences in Mathematics Education*, *11*(1). https://doi.org/10.1007/s40751-025-00172-1

Prendergast, S., & MacPhee, D. (2018). Parental contributors to children's persistence and school readiness. *Early Childhood Research Quarterly*, *45*, 31–44.

Rabok, D. A. (2021). *Parenting Education: Relationship between parent-child interaction and preprimary children's writing readiness skills in Siaya County*. Kenya.

Raver, C. C. (2002). Emotions Matter: Making the Case for the Role of Young Children's Emotional Development for Early School Readiness. Social Policy Report. Volume 16, Number 3. *Society for Research in Child Development*.

Rizki, R. , & F., Saputri, A., Nugraha, R. A. Z., & Juliani, D. (2023). EVALUATION IMPLEMENTATION OF THE INDEPENDENT CURRICULUM WITH THE CIPP MODEL IN VOCATIONAL HIGH SCHOOLS. *Jurnal Al Burhan*, *3*(2). https://doi.org/10.58988/jab.v3i2.229

Rosyan, A. N., Prasojo, E., Kurniawan, T., Nugroho, R. S., & Rahmawati, A. (2025). Literacy Development Through Collaborative Governance in Indonesia: An AHP Based Analysis of the Kampus Mengajar Program. *International Journal of Sustainable Development and Planning*, *20*(8). https://doi.org/10.18280/ijsdp.200816

Roth, W. M., & Lee, Y. J. (2007). "Vygotsky's neglected Legacy": Cultural-historical activity theory. In *Review of Educational Research* (Vol. 77, Number 2). https://doi.org/10.3102/0034654306298273

Sanders, M. R., & Mazzucchelli, T. G. (2018). How parenting influences the lives of children. *The Power of Positive Parenting: Transforming the Lives of Children, Parents, and Communities Using the Triple P System*, 5–31.

Saputri, W. H., & Risnawati, E. (2024). Preparing for the School Readiness of Early Childhood by Enhancing the Well-Being and Family Support. *Jurnal Pendidikan Usia Dini*, *18*(1), 270–286.

Scheiner, T. (2021). Towards a more comprehensive model of teacher noticing. *ZDM - Mathematics Education*, *53*(1). https://doi.org/10.1007/s11858-020-01202-5

Thelen, E. (2005). Dynamic systems theory and the complexity of change. In *Psychoanalytic Dialogues* (Vol. 15, Number 2). https://doi.org/10.1080/10481881509348831

van Es, E. A., & Sherin, M. G. (2021). Expanding on prior conceptualizations of teacher noticing. *ZDM - Mathematics Education*, *53*(1). https://doi.org/10.1007/s11858-020-01211-4

Viennet, R., & Pont, B. (2017). Education Policy Implementation: A Literature Review and Proposed Framework. OECD Education Working Papers, No. 162. *OECD Education Working Papers*, *162*(December).

Widiastuti, I. (2025). Assessing the Impact of Education Policies in Indonesia: Challenges, Achievement, and Future Direction. *AL-ISHLAH: Jurnal Pendidikan*, *17*(2). https://doi.org/10.35445/alishlah.v17i2.6803

Winters, C. A., Cudney, S., & Sullivan, T. (2010). The evolution of a coding schema in a paced program of research. *Qualitative Report*, *15*(6). https://doi.org/10.46743/2160-3715/2010.1354

Xia, X. (2020). Parenting style and Chinese children's school readiness outcomes: The moderating role of socioeconomic status. *Children and Youth Services Review*, *118*, 105381.

Zhang, Q., Choy, B. H., & Lee, M. Y. (2025). Teacher Noticing in Diverse Educational Contexts: Concepts, Research, and Future Pathways. In *ECNU Review of Education* (Vol. 8, Number 3). https://doi.org/10.1177/20965311251350160