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.

Development and Psychometric Validation of the Early Childhood Education Policy Implementation Assessment Tool (ECE-PIAT): Evidence from Somaliland Public Preschools

Jibril Abdulkadir Ali

jibrilabdikadir@amoud.edu.so

Amoud University

Befekadu Zeleke

Addis Ababa University

Bahir Abdulahi Adem

Haramaya University

Dawit Negassa Golga

Haramaya University

Research Article

Keywords: Early Childhood Education, Policy Implementation, Leadership Commitment, Institutional Capacity, Psychometric Validation

Posted Date: March 26th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-8862387/v1

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

Additional Declarations: No competing interests reported.

Abstract

Effective education policies depend not only on sound design but also on strong implementation capacity. However, policy implementation processes are rarely measured systematically, particularly in fragile and resource-constrained education systems. This study developed and psychometrically validated the Early Childhood Education Policy Implementation Assessment Tool (ECE-PIAT), a diagnostic instrument designed to assess the latent drivers of policy implementation at the institutional and leadership levels. A cross-sectional validation study was conducted with 129 educators drawn from public preschools. The instrument comprised items measuring institutional capacity, leadership commitment, and policy implementation processes. Confirmatory factor analysis was applied using a robust maximum likelihood estimator. Internal consistency was assessed using Cronbach's alpha and McDonald's omega. Convergent validity was examined through composite reliability and average variance extracted. Results supported a coherent multi-dimensional measurement structure. Standardized factor loadings were strong across most indicators, and reliability coefficients exceeded accepted thresholds for all subscales. Evidence of convergent validity was observed for each construct. Model fit indices indicated an acceptable fit given the complexity of the measurement model and the contextual constraints of the study setting. The findings demonstrate that key drivers of education policy implementation can be measured reliably using a theoretically grounded instrument. The ECE-PIAT provides evaluators and policymakers with a diagnostic tool to identify implementation strengths and constraints, supporting more targeted and evidence-informed decision-making in fragile education systems.

1. Introduction

Educational policies rarely fail because they are poorly designed. More often, they fail because they are weakly implemented. This pattern is well documented across sectors, but it is particularly visible in education systems operating under institutional fragility, resource scarcity, and administrative volatility. In such contexts, the gap between formal policy commitments and everyday practice remains wide and persistent (Domorenok et al., 2021; M. J. Williams, 2020). Despite sustained reform efforts, implementation capacity continues to function as a hidden constraint rather than an explicit object of evaluation (Ali, Zeleke, Abdulahi, et al., 2025; Silva & Oliveira, 2021). This blind spot limits both accountability and learning (Cumar, Kidane, Dawit, et al., 2025; Nguyen, 2020).

Recent scholarship has increasingly emphasized that education policy outcomes are shaped less by policy content than by the institutional conditions under which policies are enacted (Brenton et al., 2022; M. J. Williams, 2020; Yan et al., 2022). Leadership commitment, organizational coordination, and resource mobilization consistently emerge as central drivers of implementation performance (Devonald et al., 2020; Jones et al., 2020; Metz et al., 2022). However, these drivers are rarely measured systematically within education systems(Balasubramanian et al., 2022; Cumar, Kidaneb, Golgac, et al., 2025). Instead, evaluations tend to focus on downstream outcomes, such as enrollment or learning achievement, without diagnosing the upstream mechanisms that enable or constrain implementation.

Leadership has been identified as a critical but under-specified factor in education reform processes. Studies across diverse contexts show that leaders' psychological attachment to policy goals shapes prioritization, persistence, and enforcement (Arar et al., 2022; Bipath et al., 2021; Chen & Wolf, 2021). Commitment operates through multiple motivational pathways, including moral obligation, perceived costs of noncompliance, and affective identification (Mayger & Provinzano, 2021; McHugh et al., 2020). Yet, in most education systems, leadership commitment remains inferred rather than measured (Madondo, 2020). This limits the ability of evaluators to distinguish between technical capacity failures and motivational breakdowns.

At the institutional level, implementation depends on more than individual leadership. Knowledge systems, relational networks, and material resources jointly shape organizational capacity (Abdi et al., 2025; Lemire et al., 2022; Mwesigwa et al., 2020; Weiner et al., 2020). When these elements are weak or misaligned, policies tend to fragment during execution (Powell et al., 2021; Ryan et al., 2024). Existing evaluation frameworks acknowledge these dynamics conceptually, but rarely operationalize them in measurable form (Cumar, Kidane, Golga, et al., 2025; Guo, 2022). As a result, policymakers often lack diagnostic tools to identify where implementation processes are breaking down.

The absence of validated instruments is especially problematic in fragile and low-resource education systems. In such contexts, informal practices frequently substitute for formal procedures, and community involvement often compensates for limited state capacity. Without context-sensitive measurement tools, implementation assessments risk importing assumptions that do not reflect local realities. This study responds to this gap by developing and validating a policy implementation assessment tool designed explicitly for use in a fragile education system. Rather than evaluating outcomes, the instrument focuses on the latent drivers of implementation: institutional capacity, leadership commitment, and implementation processes.

Accordingly, the study addresses the following objectives: (1) to develop a theoretically grounded instrument for assessing education policy implementation drivers, and (2) to evaluate its psychometric properties using confirmatory factor analysis and complementary reliability and validity tests. By shifting attention from outcomes to implementation capacity, the study contributes to educational evaluation by providing a diagnostic tool that can support more targeted and evidence-informed policy action.

2. Literature Review

2.1 Conceptualizing Policy Implementation in Education

Policy implementation is increasingly understood as a multi-dimensional process rather than a linear translation of decisions into action. Contemporary frameworks emphasize the interaction between organizational structures, actor motivation, and contextual constraints (Ali, Zeleke, Adem, et al., 2025; Peeters & Campos, 2022; Powell et al., 2021). Within education systems, implementation involves coordination across multiple levels, including ministries, school leadership, and frontline educators

(Goodwin & Jiménez, 2020; Helper et al., 2021). These interactions generate variability in how policies are interpreted and enacted (Allen et al., 2022).

Implementation scholarship has also highlighted the limits of outcome-based evaluation. Focusing solely on outputs obscures the processes through which policies succeed or fail (Ahmed et al., 2021; Ali, Abdi, Ali, Muse, et al., 2025; Khan & Ahmed, 2021; Ninsiima et al., 2019; Onoshakpokaiye, 2024). As a result, evaluators may misattribute poor outcomes to policy design rather than to implementation breakdowns. This has led to growing calls for diagnostic instruments that capture implementation mechanisms directly (Kaddoura et al., 2022).

2.2 Institutional Capacity as a Theoretical Construct

Institutional capacity provides a useful lens for understanding implementation performance. Capacity is commonly conceptualized as a combination of human resources, organizational routines, and material inputs (Chugh et al., 2023; Lycett et al., 2023). In education systems, capacity shapes the ability of institutions to absorb, adapt, and sustain policy interventions (Lima et al., 2022). However, capacity is not static. It evolves through learning, coordination, and resource mobilization (Mousa, 2021).

Recent theoretical work has extended capacity frameworks beyond formal state structures to include relational and networked dimensions (Di Prima et al., 2022; Hinton & Lambert, 2022; Rogal et al., 2020). These perspectives emphasize collaboration, trust, and stakeholder engagement as core components of capacity (Bond, 2023; Ruano-Borbalan, 2023). Yet, despite conceptual advances, empirical measurement of institutional capacity in education remains underdeveloped (Urlacher, 2021).

2.3 Leadership Commitment and Implementation Dynamics

Leadership commitment has emerged as a key explanatory factor in implementation research. Empirical studies show that committed leaders are more likely to allocate resources, enforce standards, and sustain reform efforts (Alvarez & Porac, 2020; Jackson et al., 2021; Tagtow et al., 2021). Commitment is not a unitary construct. Rather, it encompasses normative, continuance, and affective dimensions that influence behavior in distinct ways (Hall et al., 2020; Stauffer & Kuenzler, 2021).

In education settings, leadership commitment interacts with institutional norms and external pressures (Cumar, Kidane, Dawit, et al., 2025; Malin et al., 2020; Pilar et al., 2022). Where formal accountability is weak, moral obligation and community expectations often play a compensatory role (Ali, Zeleke, Abdulahi, et al., 2025; McLoughlin et al., 2021). Despite this complexity, most evaluation studies treat leadership as an exogenous variable or proxy it through demographic indicators Rupp, 2024. This limits explanatory power and constrains policy learning.

2.4 Measurement Approaches and Methodological Gaps

Measurement of policy implementation has lagged behind theoretical development. Many studies rely on ad hoc indicators or unvalidated survey items (Ali, Golga, Abdi, Ali, et al., 2025; Leresche et al., 2023).

While such approaches offer descriptive insights, they lack the psychometric rigor needed for comparative or longitudinal analysis (Alban Conto et al., 2021; Reuge et al., 2021). Confirmatory factor analysis and related techniques provide a means of addressing this limitation by testing the coherence and distinctiveness of latent constructs.

However, the transferability of instruments developed in high-income contexts remains questionable. Contextual differences in governance structures, resource availability, and social norms challenge assumptions embedded in many existing tools (Cumar, Kidane, Dawit, et al., 2025). This is particularly salient in fragile education systems, where informal practices and hybrid governance arrangements are common. The literature therefore points to a clear gap: the need for contextually grounded, psychometrically validated instruments that capture the drivers of policy implementation in education (Cumar, Kidane, Golga, et al., 2025).

2. Methodology

2.1. Research Design

This study employed a quantitative, cross-sectional research design to establish the psychometric properties of the *Early Childhood Education Policy Implementation Assessment Tool* (ECE-PIAT). A cross-sectional approach was selected as it allows for the simultaneous assessment of multiple latent constructs at a single point in time, providing a "snapshot" of the current structural relationships between institutional capacity and leadership commitment. The design is specifically categorized as a **methodological validation study**, distinct from pure hypothesis testing, as the primary aim was to confirm the factor structure and internal consistency of a novel instrument developed for post-conflict educational settings.

2.2. Participants and Setting

The study was situated within the public education sector of the Republic of Somaliland, a context characterized by emerging institutional structures and resource constraints.

- **Target Population**: The population of interest comprised educators currently employed in public preschools under the administration of the Ministry of Education.
- **Sample Characteristics**: The final analytic sample consisted of **129 educators**. While validation studies often aim for larger samples, this size represents a significant portion of the accessible workforce in this specific, niche sector.
- **Stratification by Role**: To ensure the tool captured the full spectrum of policy implementation, the sample included both **Headteachers (** $n = 32;\ 24.8\%$**)**, who are responsible for strategic oversight, and **Teachers (** $n = 97;\ 75.2\%$**)**, who are responsible for classroom execution. This stratification ensures the instrument is valid across different hierarchical levels.

- **Demographic Profile**: The sample was predominantly female (69.8%), accurately reflecting the gender distribution of the ECE workforce in Somaliland. The cohort was well-educated, with 54.3% holding Bachelor's degrees, ensuring that respondents possessed the literacy and cognitive capacity to engage meaningfully with the 90-item instrument.

2.3. Instrument Development and Operationalization

The **ECE-PIAT** comprises 90 items designed to measure the latent drivers of policy success. The instrument development was deductive, grounded in established theoretical frameworks to ensure content validity.

2.3.1. Theoretical Grounding

- **Institutional Capacity (27 Items)**: Operationalized using the **Resource-Based View (RBV)**. This theory posits that organizational performance is determined by the possession of valuable, rare, and non-substitutable resources. Accordingly, this scale measures *Knowledge* (intellectual capital), *Relational* (social capital), and *Mobilization* (physical/financial capital) resources.
- **Leadership Commitment (26 Items)**: Operationalized using **Meyer and Allen's Three-Component Model**. This framework was chosen for its ability to distinguish between different psychological motivations: *Normative* (acting out of duty), *Continuance* (acting out of necessity/cost), and *Affective* (acting out of emotional attachment).
- **Policy Implementation (37 Items)**: Developed to measure implementation fidelity, moving from strategic alignment to operational monitoring.

2.3.2. Scaling

All items were measured on a Likert scale. This format was chosen to capture the *degree* of agreement, treating the data as ordinal but approximating interval data for analysis, a standard practice in psychometrics.

2.4. Data Collection and Pre-processing

Data collection was administered via survey. To ensure the integrity of the dataset prior to statistical modeling, a rigorous data cleaning protocol was implemented:

- **Ambiguous Coding Handling**: In instances where respondents double-marked items (ambiguous coding), responses were averaged rather than discarded. This method preserves the sample size and assumes the respondent's true intent lay between the two selections.
- **Distributional Checks**: Item-level analysis was conducted to identify potential response sets or bias. The absence of floor or ceiling effects (extreme bunching of scores) indicated that the instrument successfully differentiated between respondents.

- **Normality Assessment**: Skewness and kurtosis values were calculated for all items. All values fell within the ± 2 threshold, suggesting that while the data was not perfectly normal (typical for Likert scales), deviations were not severe enough to invalidate parametric analysis, provided robust estimators were used.

2.5. Statistical Analysis

All analyses were conducted using **R Statistical Software (v4.5.1)**, utilizing the lavaan package for Latent Variable Modeling.

2.5.1. Confirmatory Factor Analysis (CFA)

CFA was selected over Exploratory Factor Analysis (EFA) because the instrument was built upon specific, pre-existing theoretical models (RBV and Meyer & Allen). The goal was to *confirm* that the data fit these theories, rather than to *explore* for patterns.

- **Estimator Selection**: The **Robust Maximum Likelihood (MLR)** estimator was deliberately chosen. Standard Maximum Likelihood (ML) assumes multivariate normality, an assumption often violated by ordinal Likert data. MLR provides robust standard errors and a Satorra-Bentler scaled test statistic, correcting for non-normality and ensuring more accurate parameter estimates.
- **Fit Indices**: Model fit was assessed using a multi-index strategy to avoid reliance on a single metric:
    - *CFI & TLI*: Compare the target model to a null model; values closer to 1.0 indicate better fit.
    - *RMSEA & SRMR*: Measure the discrepancy between observed and predicted covariance matrices; values closer to 0 indicate better fit.

2.5.2. Reliability Testing

To ensure the instrument consistently measured the target constructs, reliability was assessed using two distinct metrics:

- **Cronbach's Alpha (α)**: Reported as the standard benchmark for internal consistency.
- **McDonald's Omega (ω)**: Calculated to provide a more rigorous estimate. Unlike Alpha, Omega does not assume that all items contribute equally to the construct (tau-equivalence), making it a superior metric for multi-dimensional scales.

2.5.3. Convergent Validity

Convergent validity was evaluated to verify that items designed to measure the same construct were actually related.

- **Average Variance Extracted (AVE)**: This measured the amount of variance captured by a construct versus the amount due to measurement error. A threshold of ≥ 0.50 was applied, ensuring that the construct explained more variance than the error term.
- **Composite Reliability (CR)**: Used as a measure of reliability that draws on the factor loadings from the CFA, offering a more precise indication of internal consistency than Cronbach's alpha alone.

3. Results

3.1 Descriptive Statistics

Preliminary item-level diagnostics indicated no evidence of floor or ceiling effects across the study measures. Skewness and kurtosis values for all items fell within acceptable thresholds (± 2), supporting the assumption of approximate normality and justifying the use of maximum likelihood estimation in subsequent analyses.

Table 1
Demographic Characteristics of Participants (N = 129)

Characteristic	Category	*n*	%
Gender	Female	90	69.8
	Male	39	30.2
Current Role	Teacher	97	75.2
	Headteacher	32	24.8
Qualification	Diploma (2 years)	20	15.5
	Bachelor's Degree	70	54.3
	Short Training	38	29.5
	Other	1	0.8
Experience	Less than 1 year	9	7.0
	1–2 Years	36	27.9
	3–4 Years	43	33.3
	5 Years and above	41	31.8
Note. Percentages may not total 100% due to rounding.			

As shown in Table 1, the final sample comprised 129 participants, the majority of whom were female (69.8%), with males representing 30.2% of the respondents. Most participants were classroom teachers (75.2%), while nearly one-quarter held headteacher positions (24.8%). In terms of academic

qualifications, more than half of the sample possessed a bachelor's degree (54.3%), followed by participants with short-term professional training (29.5%) and diploma-level qualifications (15.5%). Only one respondent (0.8%) reported an alternative qualification. Participants exhibited substantial professional experience. Approximately one-third reported between three and four years of experience (33.3%), while a similar proportion had five years or more (31.8%). Those with one to two years of experience accounted for 27.9% of the sample, and a smaller segment reported less than one year of professional experience (7.0%). Collectively, these distributions indicate a professionally mature and educationally diverse sample, appropriate for examining leadership, institutional capacity, and policy implementation processes in the education sector.

3.2 Reliability Analysis

The measurement instrument demonstrated **excellent internal consistency reliability** across all constructs. As presented in Table 2, Cronbach's alpha (α) and McDonald's omega (ω) coefficients for every subscale exceeded the conservative threshold of .80, indicating a high degree of internal coherence among items.

Table 2
Internal Consistency Reliability Statistics for Study Constructs

Scale	Sub-construct	Items	Cronbach's α	McDonald's ω
Institutional Capacity	Knowledge Resource	10	.93	.93
	Relational Resource	9	.91	.91
	Mobilization Resource	8	.89	.90
Leadership Commitment	Normative	6	.90	.90
	Continuance	8	.91	.91
	Affective	12	.95	.95
Policy Implementation	Objective Alignment	8	.93	.93
	Implementation Plan	9	.93	.93
	Staff Training	6	.90	.90
	Stakeholder Involvement	5	.89	.89
	Monitoring & Evaluation	9	.93	.93

Note. Cronbach's alpha (α) and McDonald's omega (ω) were used to assess internal consistency reliability. Values of .70 or higher indicate acceptable reliability, values above .80 indicate good reliability, and values above .90 indicate excellent reliability. All coefficients were computed based on the final sample (N = 129).

Within the **Institutional Capacity** construct, reliability estimates ranged from .89 to .93, suggesting strong consistency across knowledge, relational, and mobilization resources. Similarly, all dimensions of **Leadership Commitment** exhibited robust reliability, with coefficients above .90. Notably, the **Affective Commitment** subscale produced the highest reliability estimates (α = .95; ω = .95), reflecting a particularly strong alignment among items capturing leaders' emotional attachment to early childhood education (ECE) policy initiatives.

Reliability was equally strong for the **Policy Implementation** construct. Subscales assessing objective alignment, implementation planning, staff training, stakeholder involvement, and monitoring and evaluation all demonstrated coefficients between .89 and .93. These results indicate that the instrument reliably captures both strategic and operational dimensions of policy implementation within the education context.

Overall, the consistency of α and ω estimates across all subscales provides strong evidence that the instrument yields stable and precise measurements. The use of McDonald's omega alongside Cronbach's alpha further strengthens confidence in the reliability of the scales, particularly given the multidimensional structure of the constructs assessed.

3.3 Confirmatory Factor Analysis (CFA)

Figure 1 presents the standardized confirmatory factor analysis (CFA) measurement model for the Institutional Capacity construct, comprising the Knowledge Resource, Relational Resource, and Mobilization Resource dimensions. The model illustrates the relationships between latent constructs and their observed indicators, as well as the correlations among the latent dimensions.

Figure 1 *Confirmatory Factor Analysis Measurement Model for Institutional Capacity*

he figure presents the standardized confirmatory factor analysis (CFA) measurement model for the Institutional Capacity construct, comprising three latent dimensions: Knowledge Resource, Relational Resource, and Mobilization Resource. Circles represent latent variables, rectangles indicate observed indicators, and single-headed arrows denote standardized factor loadings. Curved double-headed arrows represent correlations among latent constructs. All factor loadings are positive and exceed the recommended minimum threshold of .50, indicating satisfactory convergent validity. Error terms are shown for each observed indicator.

Figure 2 presents the standardized confirmatory factor analysis measurement model for Policy Implementation.

The model includes Objective Alignment, Implementation Planning, Staff Training, Stakeholder Involvement, and Monitoring and Evaluation. All observed indicators loaded strongly on their intended latent dimensions. Standardized factor loadings exceeded the .50 threshold. This supports convergent validity. Loadings within each dimension were consistent. The items captured closely related aspects of implementation practice. Correlations among the latent dimensions were positive and moderate to high.

This pattern reflects the interconnected nature of implementation processes in institutional settings. Stronger associations were observed among planning, training, and monitoring components. These elements appear to function as a coordinated implementation cycle. Overall, the results support the adequacy of the proposed measurement model and its use in subsequent structural analyses.

Figure 3 presents the standardized confirmatory factor analysis measurement model for Leadership Commitment.

The model comprises Affective, Continuance, and Normative commitment dimensions. All observed indicators loaded positively on their respective latent constructs. Standardized factor loadings exceeded the .50 threshold. This indicates adequate convergent validity. Loadings within each dimension were generally consistent. Items reflected coherent aspects of leaders' commitment to early childhood education policy. Correlations among the three latent dimensions were positive and moderate to high. Stronger associations were observed between affective and continuance commitment, suggesting emotional attachment and perceived costs operate jointly in leadership behavior. Despite these associations, the constructs remained theoretically distinct. Overall, the measurement model demonstrated satisfactory psychometric properties and supported its use in subsequent structural analyses.

Model Fit:

The confirmatory factor analysis models demonstrated an acceptable level of fit given the complexity of the measurement structure. As shown in Table 3, Comparative Fit Index (CFI) values ranged from .84 to .86 across the three models.

Table 2
Fit Indices for the Confirmatory Factor Analysis Models

Model	χ2\chi^2χ2	*df*	*p*	CFI	TLI	RMSEA	SRMR
Institutional Capacity	583.56	321	< .001	.85	.84	.09	.08
Leadership Commitment	542.78	296	< .001	.86	.85	.10	.07
Policy Implementation	1129.06	619	< .001	.84	.83	.09	.07

Note. χ2 = Chi-square statistic; *df* = degrees of freedom; CFI = Comparative Fit Index; TLI = Tucker-Lewis Index; RMSEA = Root Mean Square Error of Approximation; SRMR = Standardized Root Mean Square Residual. Estimator = Robust Maximum Likelihood (MLR).

The values in Table 2 fall slightly below the conventional .90 benchmark but remain defensible in pilot research with limited sample sizes. Tucker–Lewis Index (TLI) estimates followed a similar pattern. Root Mean Square Error of Approximation (RMSEA) values clustered around .09, indicating moderate approximation error rather than model misspecification. Standardized Root Mean Square Residual (SRMR) values were consistently below .08, suggesting adequate residual fit. The significant chi-square

statistics were expected given the large number of indicators relative to the sample size. Overall, the fit indices support the theoretical structure of the models while indicating opportunities to improve parsimony in future model refinements.

Factor Loadings

Factor loadings were robust. As illustrated in Error! Reference source not found., the vast majority of items loaded above 0.60 on their respective latent factors. Only two items in the Mobilization subscale (ICMR1 and ICMR2) fell slightly below 0.60 but remained statistically significant (p<.001). This suggests that the items are strong indicators of their underlying constructs. Figure 1 presents the standardized factor loadings for all observed indicators across the latent constructs included in the measurement model. Overall, the pattern of loadings demonstrates strong item–construct relationships and provides clear support for the adequacy of the measurement structure.

The vast majority of items loaded well above the conservative threshold of .60, indicating that the indicators capture substantial variance in their intended latent factors. This pattern was consistently observed across affective, normative, continuance, institutional capacity, and policy implementation dimensions.

Items associated with affective commitment, alignment, planning, training, monitoring, and relational resources exhibited particularly strong loadings, many exceeding .70. These results suggest high conceptual clarity and internal coherence within these subscales. The consistency of loadings within each factor further indicates that items measure closely related aspects of the same underlying construct rather than disparate traits.

Two items within the mobilization resource subscale (ICMR1 and ICMR2) displayed loadings slightly below .60. However, both remained statistically significant ($p < .001$). Their inclusion is therefore defensible, particularly in a pilot study context, as they still contribute meaningful information about the construct domain. Importantly, no item fell below the minimum .50 threshold. Taken together, the factor loading results provide strong evidence of convergent validity and support the retention of all items for subsequent structural modeling.

Table 3

Standardized Factor Loadings, Means, and Standard Deviations for Institutional Capacity

Item Code	Item Description (Abbreviated)	*M*	*SD*	Factor Loading (λ)
Knowledge Resources				
ICKR1	Qualified team of educators	3.40	1.51	.67***
ICKR2	Professional development opportunities	3.36	1.39	.63***
ICKR3	Comprehensive staff records	3.36	1.36	.78***
ICKR4	Quality assurance procedures	3.36	1.29	.84***
ICKR5	Regular evaluations conducted	3.28	1.35	.80***
ICKR6	Strategic leadership alignment	3.48	1.36	.79***
ICKR7	Departmental coordination	3.36	1.38	.79***
ICKR8	Performance measurement tools	3.44	1.39	.77***
ICKR9	Stakeholder feedback sought	3.30	1.39	.74***
ICKR10	Resource allocation based on needs	3.26	1.46	.77***
Relational Resources				
ICRR1	Effective communication channels	3.22	1.43	.78***
ICRR2	Stakeholder involvement in policy	3.44	1.32	.69***
ICRR3	Culture of teamwork	3.26	1.31	.77***
ICRR4	Regular meetings organized	3.03	1.38	.73***
ICRR5	Feedback used for improvement	3.14	1.38	.75***
ICRR6	Infrastructure well-maintained	3.33	1.37	.76***
ICRR7	Partnerships with external orgs	3.14	1.40	.72***
ICRR8	Collaboration with community	3.27	1.35	.68***
ICRR9	Partnerships with educational inst.	3.31	1.37	.73***
Mobilization Resources				
ICMR1	Adequate physical space	3.34	1.35	.59***
ICMR2	Financial resource mobilization	3.33	1.40	.58***

Note. N = 129. All factor loadings are standardized. ***p < .001.

Item Code	Item Description (Abbreviated)	*M*	*SD*	Factor Loading (λ)
ICMR3	Grants and funding opportunities	3.20	1.37	.75***
ICMR4	Partnerships with local business	3.09	1.38	.71***
ICMR5	Transparent financial management	3.23	1.42	.83***
ICMR6	Infrastructure development projects	3.30	1.34	.74***
ICMR7	Partnerships with NGOs	3.23	1.41	.75***
ICMR8	Stakeholder engagement strategies	3.41	1.33	.79***

Note. N = 129. All factor loadings are standardized. ***p < .001.

3.4 Convergent and Discriminant Validity

Convergent validity was unequivocally established. Table 3 reports the standardized factor loadings, means, and standard deviations for the Institutional Capacity construct, encompassing Knowledge Resources, Relational Resources, and Mobilization Resources. Overall, the results demonstrate strong psychometric performance and provide clear support for the proposed three-factor structure. All items loaded significantly on their intended latent dimensions ($p < .001$), with standardized loadings ranging from .58 to .84.

Table 3

Reliability and Convergent Validity of the ECE-PIAT Constructs

Construct	Subscale	No. of Items	α	ω	CR	AVE
Institutional Capacity	Knowledge Resources	10	.93	.93	.93	.57
	Relational Resources	9	.91	.91	.91	.54
	Mobilization Resources	8	.89	.90	.89	.52
Leadership Commitment	Normative	6	.90	.90	.90	.59
	Continuance	8	.91	.91	.91	.55
	Affective	12	.95	.95	.94	.59
Policy Implementation	Objective Alignment	8	.93	.93	.93	.63
	Implementation Plan	9	.93	.93	.93	.59
	Staff Training	6	.90	.90	.90	.61
	Stakeholder Involvement	5	.89	.89	.89	.63
	Monitoring & Evaluation	9	.93	.93	.93	.61

Note. α = Cronbach's alpha; ω = McDonald's omega; CR = Composite Reliability; AVE = Average Variance Extracted. Thresholds for acceptability: α, ω, CR>.70; AVE>.50.

Within the **Knowledge Resources** dimension, factor loadings were consistently high, with several items exceeding .75. Indicators related to quality assurance, regular evaluation, and leadership alignment exhibited particularly strong loadings, suggesting that formal systems and strategic coordination are central elements of institutional knowledge capacity. Mean scores for this dimension were moderate, indicating room for further institutional strengthening.

The **Relational Resources** dimension also demonstrated robust loadings, generally above .70. Items capturing communication, teamwork, and collaborative partnerships showed strong associations with the latent construct, underscoring the importance of social and organizational relationships in institutional functioning.

For **Mobilization Resources**, most items loaded strongly, although two indicators (ICMR1 and ICMR2) exhibited slightly lower loadings. These items remained statistically significant and conceptually relevant, particularly in a pilot context. Collectively, the findings provide strong evidence of convergent validity and support the retention of all items for subsequent structural analyses.

Correlational analysis (Fig. 1) revealed strong positive associations between constructs. Within the Leadership Commitment domain, Normative Commitment was highly correlated with Continuance

Commitment (r = 0.97), indicating that leaders' moral obligation to implement policy is closely linked to the perceived costs of noncompliance.

Affective Commitment was also strongly associated with both normative (*r* = .76) and continuance commitment (*r* = .80), suggesting that emotional attachment, obligation, and cost considerations form an integrated commitment profile in this context.

Strong correlations were likewise observed among **Policy Implementation** dimensions. Objective Alignment showed high associations with Implementation Planning (*r* = .92) and Staff Training (*r* = .81), highlighting the interdependence of strategic clarity and operational readiness. Particularly notable were the very high correlations among Staff Training, Stakeholder Involvement, and Monitoring and Evaluation (*r* values ≥ .96), reflecting a tightly coupled implementation cycle in which capacity building, participation, and oversight reinforce one another.

Correlations between **Institutional Capacity** components and policy implementation dimensions were moderate to strong, indicating that knowledge, relational, and mobilization resources are systematically linked to implementation effectiveness. Overall, the pattern supports theoretical coherence across constructs, while also underscoring the need to assess discriminant validity using complementary criteria.

4. Discussion

This study set out to develop and validate a diagnostic instrument capable of capturing the latent drivers of education policy implementation in a fragile system. The results provide strong evidence that policy implementation is best understood as a structured, multi-dimensional process rather than a diffuse organizational outcome. The validated factor structure demonstrates that institutional capacity, leadership commitment, and implementation processes can be measured reliably and meaningfully within a single framework (Dawson et al., 2018; Domorenok et al., 2021; Pont, 2020; M. J. Williams, 2020; N. J. Williams et al., 2022). This finding aligns with evaluation scholarship that calls for a shift from outcome-centric assessments toward process-sensitive diagnostics (Ali, Zeleke, Abdulahi, et al., 2025; Peeters & Campos, 2022). From a measurement perspective, the results confirm that implementation drivers can be operationalized as coherent latent constructs (Rupp, 2024).

The strong factor loadings observed across most subscales indicate high conceptual clarity. Items related to coordination, monitoring, and strategic alignment loaded particularly well, suggesting that these elements represent core features of implementation capacity rather than peripheral supports (Brenton et al., 2022; Eckhardt et al., 2022; M. J. Williams, 2020; Yan et al., 2022). This pattern is consistent with prior empirical work showing that implementation failures often originate in weak internal alignment rather than in policy ambiguity (Goodwin & Jiménez, 2020; Helper et al., 2021). The strength of these loadings reinforces the argument that implementation quality is embedded in organizational routines and decision structures Allen, 2022.

Leadership commitment emerged as a central and tightly integrated construct. The high correlations among affective, normative, and continuance commitment dimensions suggest that, in fragile systems, these motivational forms may operate as a unified psychological profile rather than as fully separable traits (Arar et al., 2022; Chen & Wolf, 2021; Matjokana & Bipath, 2024). Similar patterns have been observed in contexts where formal accountability mechanisms are limited and leaders rely on personal obligation and reputational considerations to sustain policy action. Rather than indicating poor discriminant validity, these correlations likely reflect contextual integration of leadership motivations (Norman, 2023).

From an evaluation standpoint, this finding has important implications. Treating leadership commitment as a single diagnostic domain may be more analytically useful in fragile settings than attempting to force artificial separation of motivational components (Abdi et al., 2025; Ali, Muse, Abdi, Ali, et al., 2025; Lemire et al., 2022; Mwesigwa et al., 2020; Weiner et al., 2020). This does not undermine the theoretical distinctiveness of commitment dimensions, but it does suggest that their empirical expression is shaped by institutional context (Di Prima et al., 2022; Rogal et al., 2020). Future modeling may therefore benefit from testing higher-order or bifactor representations (Urlacher, 2021).

Institutional capacity was also validated as a multi-dimensional construct encompassing knowledge, relational, and mobilization resources. The results show that these dimensions are closely linked but empirically distinguishable, supporting contemporary capacity frameworks that extend beyond material inputs (Ali, Zeleke, Adem, et al., 2025; Khan & Ahmed, 2021; Ninsiima et al., 2019). Relational resources, in particular, demonstrated strong associations with implementation processes, underscoring the importance of collaboration and stakeholder engagement (Ali, Golga, Abdi, Ali, et al., 2025; Pilar et al., 2022). This finding reinforces arguments that capacity in education systems is as much social as it is technical.

The policy implementation construct itself displayed a coherent internal structure, with strong linkages among planning, training, stakeholder involvement, and monitoring. These dimensions appear to function as an integrated implementation cycle rather than as isolated activities (Cumar, Kidane, Dawit, et al., 2025; Peeters & Campos, 2022). Empirical studies in comparable contexts suggest that when one component weakens, the entire cycle becomes unstable (Ali, Zeleke, Adem, et al., 2025; Goodwin & Jiménez, 2020; Helper et al., 2021). The high correlations observed among these dimensions therefore reflect functional interdependence rather than measurement redundancy (Allen et al., 2022).

Model fit indices were acceptable given the complexity of the measurement model and the context of application. While CFI values fell slightly below conventional thresholds, this pattern is well documented in CFA models with large item pools and theoretically constrained structures (Cumar, Kidane, Golga, et al., 2025; Pont, 2020). Importantly, strong factor loadings and satisfactory AVE values indicate that the constructs capture substantive variance Rupp, 2024. From an evaluation perspective, interpretability and diagnostic utility are more critical than marginal improvements in global fit (Ali, Golga, Abdi, Ali, et al., 2025; Leresche et al., 2023).

The findings also highlight the limits of transferring measurement tools across contexts without adaptation. Many existing implementation instruments were developed in systems with stable bureaucracies and abundant resources. Applying them uncritically in fragile systems risks misdiagnosis and policy misdirection (Alban Conto et al., 2021; Ali, Zeleke, Abdulahi, et al., 2025). The validated structure of the present instrument demonstrates that context-sensitive measurement is both feasible and necessary.

From a methodological standpoint, the study contributes to the growing use of CFA in educational evaluation. By grounding item development in theory and testing the structure explicitly, the study advances best practices in instrument validation (Cumar, Kidaneb, Golgac, et al., 2025; Lemire et al., 2022; Mwesigwa et al., 2020; Weiner et al., 2020). The use of multiple reliability indicators further strengthens confidence in the instrument's measurement properties. These methodological choices enhance the tool's suitability for benchmarking and comparative use (Rupp, 2024; Urlacher, 2021).

Substantively, the results suggest that improving policy implementation requires coordinated investment across leadership, institutional capacity, and operational processes. Focusing on one domain in isolation is unlikely to yield sustained gains (Brenton et al., 2022; M. J. Williams, 2020; Yan et al., 2022). The instrument provides a means of identifying which components are most constraining in a given setting, thereby supporting more targeted intervention. Overall, the discussion underscores the value of diagnostic measurement in educational evaluation. By making implementation drivers visible and measurable, the validated tool shifts evaluation practice toward a more nuanced understanding of how policies function in practice. This contribution is particularly relevant for fragile education systems, where conventional evaluation metrics often fail to capture the realities of policy enactment.

Conclusion

This study addressed a critical gap in educational evaluation by developing and validating a diagnostic instrument for assessing policy implementation drivers in a fragile education system. Rather than focusing on policy outcomes, the study concentrated on the underlying institutional and leadership factors that shape how policies are enacted in practice. This shift in focus responds to longstanding calls for evaluation approaches that move beyond outcome measurement toward process-sensitive diagnostics. The results provide strong evidence that institutional capacity, leadership commitment, and implementation processes can be operationalized as coherent latent constructs. The validated measurement structure demonstrates that these dimensions are not abstract concepts but empirically observable features of organizational functioning. High internal consistency across subscales further indicates that respondents interpreted the items consistently, despite contextual and resource constraints.

Importantly, the study shows that complex implementation dynamics can be captured using a structured instrument without oversimplifying the realities of fragile systems. The observed interrelationships among leadership commitment dimensions and implementation components reflect contextual

integration rather than conceptual ambiguity. This finding reinforces the need to interpret measurement results in light of institutional conditions rather than relying solely on generic benchmarks. By providing a validated assessment tool, the study contributes to the methodological advancement of educational evaluation. It offers a practical means of diagnosing implementation readiness and identifying areas where policy support is most urgently needed. While the instrument was developed in a specific context, its underlying structure is grounded in widely accepted theoretical frameworks, enhancing its relevance for adaptation in comparable settings. Overall, the study underscores the value of measurement as a foundation for learning, accountability, and improvement. Making implementation drivers visible and measurable is a necessary step toward more effective education policy execution, particularly in systems facing structural and institutional fragility.

Implications for Policy and Practice

The validated instrument has several important implications for education policy and practice. First, it enables policymakers to move from assumption-based judgments to evidence-based diagnostics. By identifying specific weaknesses in institutional capacity or leadership commitment, decision-makers can design interventions that address root causes rather than symptoms. Second, the instrument supports more strategic allocation of resources. Rather than applying uniform solutions across institutions, policymakers can prioritize investments where implementation constraints are most severe. This is particularly valuable in low-resource environments, where inefficient allocation carries high opportunity costs. Third, the tool can be used to support leadership development initiatives. By assessing different dimensions of commitment, education authorities can tailor professional development programs to strengthen motivational and organizational alignment with policy goals. This diagnostic approach complements traditional training models that focus primarily on technical skills.

Fourth, the instrument provides a foundation for monitoring implementation progress over time. Repeated administration can reveal whether capacity-building efforts are producing meaningful improvements. This enables learning-oriented evaluation rather than one-off assessments. At the institutional level, school leaders can use the findings to reflect on internal processes, coordination mechanisms, and stakeholder engagement strategies. The instrument encourages structured self-assessment and supports dialogue between schools and supervising authorities. Finally, the tool enhances accountability in a constructive manner. By focusing on implementation conditions rather than blame, it promotes a culture of improvement. This is especially important in fragile systems, where punitive accountability mechanisms may discourage honest reporting. Overall, the instrument offers a practical and scalable means of strengthening education policy implementation through informed action.

Study Limitations

Several limitations should be considered when interpreting the findings. First, the study employed a cross-sectional design, which restricts the ability to examine changes in implementation capacity over time. Longitudinal validation would be necessary to assess temporal stability and sensitivity to change. Second, the sample size was constrained by the size of the target population. While the sample was adequate for the intended validation purposes, larger samples would allow for more complex modeling and subgroup analyses. Third, the instrument was validated within a specific institutional and cultural context. Although the theoretical foundations are broadly applicable, direct transfer to other settings should be undertaken cautiously. Contextual adaptation and re-validation are recommended before wider use. Fourth, the study focused on internal consistency and convergent validity. Other forms of validity, such as predictive and criterion-related validity, were not assessed. These represent important areas for future research. Finally, the reliance on self-reported data introduces the possibility of social desirability bias. While this risk is common in organizational surveys, triangulation with external data sources could strengthen future assessments.

Recommendations for Future Research

Future research should extend the validation of the instrument in several directions. Longitudinal studies would allow researchers to examine how implementation capacity evolves over time and how it responds to targeted interventions. This would strengthen the instrument's utility for monitoring and evaluation. Further work is also needed to explore measurement invariance across institutional types and demographic groups. Establishing invariance would support meaningful comparisons across contexts and enhance the instrument's generalizability.

Scale refinement represents another important avenue. Item reduction techniques could be applied to develop shorter versions of the instrument for rapid assessments, while preserving psychometric quality. Future studies should also examine the relationship between implementation capacity scores and policy outcomes. Linking diagnostic measures to performance indicators would provide evidence of predictive validity and enhance the instrument's practical relevance. Finally, comparative applications across different education systems would deepen understanding of how institutional and leadership factors shape implementation in diverse settings. Such research would contribute to the development of a cumulative evidence base for implementation measurement in education.

Declarations

Ethics Approval

The study protocol was reviewed and approved by the Research Ethics Committee of the University of Hargeisa and received clearance from the Somaliland Ministry of Education and Science. All procedures performed in this study involving human participants were conducted in accordance with the ethical standards of the responsible institutional committees and with the principles of the Declaration of Helsinki.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Consent to Participate

Informed consent was obtained from all individual participants included in the study. Participation was voluntary, and respondents were informed about the purpose of the study, their right to withdraw at any time, and the confidentiality of their responses.

Consent to Publish

Consent to publish was obtained from all participants. No identifying personal information is included in the manuscript.

Conflict of Interest

The authors declare that they have no conflict of interest.

References

1. Abdi, M. K., Cumar, M. A., Muse, A. H., Ali, T. A., & Ali, J. A. (2025). Assessment of factors influencing student performance grades in Somaliland primary schools: Insights from an ordinal logistic regression analysis. *Equity in Education & Society*. https://doi.org/10.1177/27526461251393410
2. Ahmed, M., Mubarik, M., & Shahbaz, M. (2021). Factors affecting the outcome of corporate sustainability policy: a review paper. *Environmental Science and Pollution Research International*. https://doi.org/10.1007/s11356-020-12143-7
3. Alban Conto, C., Akseer, S., Dreesen, T., Kamei, A., Mizunoya, S., & Rigole, A. (2021). Potential effects of COVID-19 school closures on foundational skills a nd Country responses for mitigating learning loss. *International Journal of Educational Development*, *87*, 102434. https://doi.org/10.1016/j.ijedudev.2021.102434
4. Ali, J. A., Abdi, M. K., Ali, T. A., Muse, A. H., Omar, M. A., & Cumar, M. A. (2025). Multilingual competency and academic performance: a machine learning-based analysis of the 2022/2023 Somaliland national primary exam data. *Discover Data*, *3*(1), 21. https://doi.org/10.1007/s44248-025-00061-3
5. Ali, J. A., Golga, D. N., Abdi, M. K., Ali, T. A., Cumar, M. A., Ali, N., & Muse, A. H. (2025). Advancing early childhood education policy: a transformative case study in Hargeisa, Somaliland—insights from preprimary education development. *Cogent Education*, *12*(1). https://doi.org/10.1080/2331186X.2024.2442189
6. Ali, J. A., Muse, A. H., Abdi, M. K., Ali, T. A., Muse, Y. H., & Cumar, M. A. (2025). Machine learning-driven analysis of academic performance determinants: Geographic, socio-demographic, and

subject-specific influences in Somaliland's 2022–2023 national primary examinations. *International Journal of Educational Research Open*, *8*. https://doi.org/10.1016/j.ijedro.2024.100426

7. Ali, J. A., Zeleke, B., Abdulahi, B. A., & Golga, D. N. (2025). Drivers of success in Somaliland's preschools: a predictive machine learning model of leadership commitment and institutional capacity. *Frontiers in Education*, *10*. https://doi.org/10.3389/feduc.2025.1643452
8. Ali, J. A., Zeleke, B., Adem, B. A., & Golga, D. N. (2025). Normative commitment as the key to early childhood education policy implementation in Somaliland. *Discover Education*. https://doi.org/10.1007/s44217-025-01050-z
9. Allen, M. S., Iliescu, D., & Greiff, S. (2022). Single Item Measures in Psychological Science. *European Journal of Psychological Assessment*, *38*(1), 1–5. https://doi.org/10.1027/1015-5759/a000699
10. Alvarez, S., & Porac, J. (2020). *Imagination, Indeterminacy, and Managerial Choice at the Limit of Knowledge*. https://doi.org/10.5465/amr.2020.0366
11. Arar, K., Örücü, D., & Gümüş, S. (2022). Educational leadership and policy studies in refugee education: a syst ematic review of existing research. *Educational Review*, *76*(4), 1032–1056. https://doi.org/10.1080/00131911.2022.2066632
12. Balasubramanian, N., Ye, Y., & Xu, M. (2022). Substituting Human Decision-Making with Machine Learning: Implications for Organizational Learning. *Academy of Management Review*, *47*(3), 448–465. https://doi.org/10.5465/amr.2019.0470
13. Bipath, K., Tebekana, J., & Venketsamy, R. (2021). Leadership in Implementing Inclusive Education Policy in Early Childho od Education and Care Playrooms in South Africa. *Education Sciences*, *11*(12), 815. https://doi.org/10.3390/educsci11120815
14. Bond, M. (2023). <scp>BERJ</scp> 50th anniversary collection: Virtual issue 1995 to 200 4. *British Educational Research Journal*, *50*(6), 2570–2581. https://doi.org/10.1002/berj.3876
15. Brenton, S., Baekkeskov, E., & Hannah, A. (2022). Policy capacity: evolving theory and missing links. *Policy Studies*. https://doi.org/10.1080/01442872.2022.2043266
16. Chen, S., & Wolf, S. (2021). Measuring the Quality of Early Childhood Education in Low- and Middle-Income Countries. *Frontiers in Psychology*, *12*. https://doi.org/10.3389/fpsyg.2021.774740
17. Chugh, R., Turnbull, D., Cowling, M. A., Vanderburg, R., & Vanderburg, M. A. (2023). Implementing educational technology in Higher Education Institutions: A review of technologies, stakeholder perceptions, frameworks and metrics. *Education and Information Technologies*, *28*(12), 16403–16429. https://doi.org/10.1007/s10639-023-11846-x
18. Cumar, M. A., Kidane, B. Z., Dawit, N. G., & Dinsa, F. (2025). Influence of leadership styles on job satisfaction in the higher education institutions of Somaliland. *Cogent Education*, *12*(1). https://doi.org/10.1080/2331186X.2025.2480996
19. Cumar, M. A., Kidane, B. Z., Golga, D. N., & Dinsa, F. (2025). Leadership as cultural architect: a path-goal theory analysis of leadership styles and organizational culture in Somaliland higher education. *Discover Education*, *4*(1). https://doi.org/10.1007/s44217-025-00887-8

20. Cumar, M. A., Kidaneb, B. Z., Golgac, D. N., & Dinsa, F. (2025). Organizational culture and academic staff satisfaction in Somaliland higher education: evidence from structural equation modeling. *Frontiers in Education*, *10*. https://doi.org/10.3389/feduc.2025.1635607
21. Dawson, D., Hepworth, J., Bugaian, L., & Williams, S. (2018). The drivers of Higher Education leadership competence: a study of Mold ovan HEI's. *Studies in Higher Education*, *45*(6), 1217–1232. https://doi.org/10.1080/03075079.2018.1557135
22. Devonald, M., Jones, N., & Yadete, W. (2020). Addressing educational attainment inequities in rural Ethiopia: Leave no adolescent behind. *Development Policy Review*. https://doi.org/10.1111/dpr.12523
23. Di Prima, S., Wright, E. P., Sharma, I. K., Syurina, E., & Broerse, J. E. W. (2022). Implementation and scale-up of nutrition-sensitive agriculture in low- and middle-income countries: a systematic review of what works, what doesn't work and why. *Global Food Security*, *32*, 100595. https://doi.org/10.1016/j.gfs.2021.100595
24. Domorenok, E., Graziano, P., & Polverari, L. (2021). *Introduction: policy integration and institutional capacity: theoretical, conceptual and empirical challenges*. https://doi.org/10.1080/14494035.2021.1902058
25. Eckhardt, C. M., Madjarova, S. J., Williams, R. J., Ollivier, M., Karlsson, J., Pareek, A., & Nwachukwu, B. U. (2022). Unsupervised machine learning methods and emerging applications in hea lthcare. *Knee Surgery, Sports Traumatology, Arthroscopy*, *31*(2), 376–381. https://doi.org/10.1007/s00167-022-07233-7
26. Goodwin, A. P., & Jiménez, R. T. (2020). The Science of Reading: Supports, Critiques, and Questions. *Reading Research Quarterly*, *55*(S1). https://doi.org/10.1002/rrq.360
27. Guo, S. (2022). Exploring low-carbon pilot city policy implementation: evidence from China. *Climate Policy*, *23*(8), 1045–1057. https://doi.org/10.1080/14693062.2022.2140096
28. Hall, J. L., Zavattaro, S. M., Battaglio, R. P., & Hail, M. (2020). *Global Reflection, Conceptual Exploration, and Evidentiary Assimilation: COVID -19 Viewpoint Symposium Introduction*. https://doi.org/10.1111/puar.13261
29. Helper, S., Gray, J. V, Hughes, M. M., & Roman, A. V. (2021). Public policy and operations management. *Journal of Operations Management*, *67*(7), 780–802. https://doi.org/10.1002/joom.1160
30. Hinton, A. O., & Lambert, W. M. (2022). Moving diversity, equity, and inclusion from opinion to evidence. *Cell Reports Medicine*. https://doi.org/10.1016/j.xcrm.2022.100619
31. Jackson, C., Mohr-Schroeder, M. J., Bush, S. B., Maiorca, C., Roberts, T., Yost, C., & Fowler, A. (2021). Equity-Oriented Conceptual Framework for K-12 STEM literacy. *International Journal of STEM Education*, *8*(1). https://doi.org/10.1186/s40594-021-00294-z
32. Jones, N., Pincock, K., Baird, S., Yadete, W., & Hicks, J. H. (2020). Intersecting inequalities, gender and adolescent health in Ethiopia. *International Journal for Equity in Health*. https://doi.org/10.1186/s12939-020-01214-3

33. Kaddoura, S., Popescu, D. E., & Hemanth, J. D. (2022). A systematic review on machine learning models for online learning and examination systems. *PeerJ Computer Science*, *8*, e986. https://doi.org/10.7717/peerj-cs.986
34. Khan, M. J., & Ahmed, J. (2021). Child education in the time of pandemic: Learning loss and dropout. *Children and Youth Services Review*, *127*, 106065. https://doi.org/10.1016/j.childyouth.2021.106065
35. Lemire, C., Rousseau, M., & Dionne, C. (2022). A Comparison of Fidelity Implementation Frameworks Used in the Field of Early Intervention. *American Journal of Evaluation*. https://doi.org/10.1177/10982140211008978
36. Leresche, E., Hossain, M., De Rubeis, M. L., Hermans, V., Burtscher, D., Rossi, R., Lonsdale, C., & Singh, N. S. (2023). How is the implementation of empirical research results documented in conflict-affected settings? Findings from a scoping review of peer-reviewed literature. *Conflict and Health*, *17*(1). https://doi.org/10.1186/s13031-023-00534-9
37. Lima, M., Neto, M., Filho, T. S., & de A. Fagundes, R. A. (2022). Learning Under Concept Drift for Regression—A Systematic Literature Re view. *IEEE Access*, *10*, 45410–45429. https://doi.org/10.1109/access.2022.3169785
38. Lycett, K., Cleary, J., Calder, R., Frykberg, G., Hollonds, A., Dakin, P., Russell, S., Demaio, S., & Goldfeld, S. (2023). A framework for the Future Healthy Countdown 2030: tracking the health and wellbeing of children and young people to hold Australia to account. *Medical Journal of Australia*, *219*(S10). https://doi.org/10.5694/mja2.52145
39. Madondo, F. (2020). Perceptions on Curriculum Implementation: A Case for Rural Zimbabwean Early Childhood Development Teachers as Agents of Change. *Journal of Research in Childhood Education*, *35*(3), 399–416. https://doi.org/10.1080/02568543.2020.1731024
40. Malin, J., Brown, C. R., Ion, G., Ackeren, I. van, Bremm, N., Luzmore, R., Flood, J., & Rind, G. (2020). World-wide barriers and enablers to achieving evidence-informed practice in education: what can be learnt from Spain, England, the United States, and Germany? *Humanities and Social Sciences Communications*. https://doi.org/10.1057/s41599-020-00587-8
41. Matjokana, T. N., & Bipath, K. (2024). Unblocking the System to Strengthen Implementation of Early Childhood Development Policies and Practice in South Africa. *Research in Educational Policy and Management*, *6*(1), 32–51. https://doi.org/10.46303/repam.2024.3
42. Mayger, L., & Provinzano, K. (2021). Leadership for Family and Community Engagement: A Qualitative Policy Analysis of State Principal Evaluation Processes. *Educational Administration Quarterly*, *58*(1), 141–177. https://doi.org/10.1177/0013161x211052501
43. McHugh, S., Dorsey, C. N., Mettert, K., Purtle, J., Bruns, E., & Lewis, C. C. (2020). Measures of outer setting constructs for implementation research: A systematic review and analysis of psychometric quality. *Implementation Research and Practice*, *1*. https://doi.org/10.1177/2633489520940022
44. McLoughlin, G. M., Allen, P., Walsh-Bailey, C., & Brownson, R. C. (2021). A systematic review of school health policy measurement tools: implementation determinants and outcomes. *Implementation Science Communications*, *2*(1). https://doi.org/10.1186/s43058-021-00169-y

45. Metz, A., Jensen, T., Farley, A. B., Boaz, A., Bartley, L., & Villodas, M. L. (2022). Building trusting relationships to support implementation: A proposed theoretical model. *Frontiers in Health Services*. https://doi.org/10.3389/frhs.2022.894599
46. Mousa, M. (2021). The hidden curriculum in business schools: Does it initiate the implementation of responsible management education in fragile states? A conceptual analytical study. *Journal of Education for Business*, *97*(6), 393–400. https://doi.org/10.1080/08832323.2021.1973944
47. Mwesigwa, R., Tusiime, I., & Ssekiziyivu, B. (2020). Leadership styles, job satisfaction and organizational commitment amon g academic staff in public universities. *Journal of Management Development*, *39*(2), 253–268. https://doi.org/10.1108/jmd-02-2018-0055
48. Nguyen, C. (2020). Nationally standardized policy and locally interpreted implementation: how Vietnamese school leaders enact education reform. *International Journal of Leadership in Education*, *26*(3), 397–426. https://doi.org/10.1080/13603124.2020.1722249
49. Ninsiima, A. B., Coene, G., Michielsen, K., Najjuka, S., Kemigisha, E., Ruzaaza, G. N., Nyakato, V. N., & Leye, E. (2019). Institutional and contextual obstacles to sexuality education policy i mplementation in Uganda. *Sex Education*, *20*(1), 17–32. https://doi.org/10.1080/14681811.2019.1609437
50. Norman, A. (2023). Educational technology for reading instruction in developing countries: A systematic literature review. *Revista de Educación*. https://doi.org/10.1002/rev3.3423
51. Onoshakpokaiye, O. E. (2024). Students' psychological variables connection with secondary school students' academic performance in mathematics. *Arab Gulf Journal of Scientific Research*. https://doi.org/10.1108/agjsr-08-2023-0369
52. Peeters, R., & Campos, S. A. (2022). Street-level bureaucracy in weak state institutions: a systematic review of the literature. *International Review of Administrative Sciences*, *89*(4), 977–995. https://doi.org/10.1177/00208523221103196
53. Pilar, M. R., Jost, E., Walsh-Bailey, C., Powell, B. J., Mazzucca, S., Eyler, A., Purtle, J. I., Allen, P., & Brownson, R. (2022). Quantitative measures used in empirical evaluations of mental health policy implementation: A systematic review. *Implementation Research and Practice*. https://doi.org/10.1177/26334895221141116
54. Pont, B. (2020). A literature review of school leadership policy reforms. *European Journal of Education*, *55*(2), 154–168. https://doi.org/10.1111/ejed.12398
55. Powell, B., Mettert, K. D., Dorsey, C. N., Weiner, B., Stanick, C. F., Lengnick-Hall, R., Ehrhart, M. G., Aarons, G., Barwick, M., Damschroder, L., & Lewis, C. (2021). Measures of organizational culture, organizational climate, and implementation climate in behavioral health: A systematic review. *Implementation Research and Practice*. https://doi.org/10.1177/26334895211018862
56. Reuge, N., Jenkins, R., Brossard, M., Soobrayan, B., Mizunoya, S., Ackers, J., Jones, L., & Taulo, W. G. (2021). Education response to COVID 19 pandemic, a special issue proposed by U NICEF: Editorial review. *International Journal of Educational Development*, *87*, 102485. https://doi.org/10.1016/j.ijedudev.2021.102485

57. Rogal, S. S., Powell, B., Chinman, M., Agbalajobi, O., Kanwa, F., Gonzalez, R., Lamorte, C., Mellinger, J. L., Morgan, T., Neely, B., Park, A. M., Patel, A. A., Saini, S., Waljee, A., & Yakovchenko, V. (2020). Moving Toward Impact: An Introduction to Implementation Science for Gastroenterologists and Hepatologists. *Gastroenterology*. https://doi.org/10.1053/j.gastro.2020.07.063
58. Ruano-Borbalan, J.-C. (2023). Understanding and fostering the development of critical thinking education and competences. *European Journal of Education*. https://doi.org/10.1111/ejed.12572
59. Rupp, A. A. (2024). Commentary: Modernizing Educational Assessment Training for Changing Job Markets. *Educational Measurement: Issues and Practice*, *43*(3), 33–38. https://doi.org/10.1111/emip.12629
60. Ryan, A., Prieto-Rodriguez, E., Miller, A., & Gore, J. (2024). What can implementation science tell us about scaling interventions in school settings? A scoping review. *Educational Research and Reviews*. https://doi.org/10.1016/j.edurev.2024.100620
61. Silva, R. da, & Oliveira, J. (2021). Global education policy in African fragile and conflict-affected states: examining the Global Partnership for Education. *Globalisation, Societies and Education*, *20*(4), 508–522. https://doi.org/10.1080/14767724.2021.1947201
62. Stauffer, B., & Kuenzler, J. (2021). Introduction—Stories of the old world: The Narrative Policy Framework in the European context. *European Policy Analysis*, *7*(S2), 268–275. https://doi.org/10.1002/epa2.1128
63. Tagtow, A., Herman, D., & Cunningham-Sabo, L. (2021). Next-Generation Solutions to Address Adaptive Challenges in Dietetics Practice: The I+PSE Conceptual Framework for Action. *Journal of the Academy of Nutrition and Dietetics*. https://doi.org/10.1016/j.jand.2021.01.018
64. Urlacher, B. (2021). State Building and Peace Agreement Implementation. *Defence and Peace Economics*, *34*(1), 112–127. https://doi.org/10.1080/10242694.2021.1946894
65. Weiner, B., Mettert, K. D., Dorsey, C. N., Nolen, E. A., Stanick, C. F., Powell, B., & Lewis, C. (2020). Measuring readiness for implementation: A systematic review of measures' psychometric and pragmatic properties. *Implementation Research and Practice*. https://doi.org/10.1177/2633489520933896
66. Williams, M. J. (2020). Beyond state capacity: bureaucratic performance, policy implementation and reform. *Journal of Institutional Economics*. https://doi.org/10.1017/S1744137420000478
67. Williams, N. J., Hugh, M. L., Cooney, D. J., Worley, J. A., & Locke, J. (2022). Testing a Theory of Implementation Leadership and Climate Across Autism Evidence-Based Interventions of Varying Complexity. *Behavior Therapy*, *53*(5), 900–912. https://doi.org/10.1016/j.beth.2022.03.001
68. Yan, Y., Sano, H., & Sumiya, L. A. (2022). Policy Capacity Matters for Education System Reforms: A Comparative Study of Two Brazilian States. *Journal of Comparative Policy Analysis*. https://doi.org/10.1080/13876988.2022.2110472

Figures

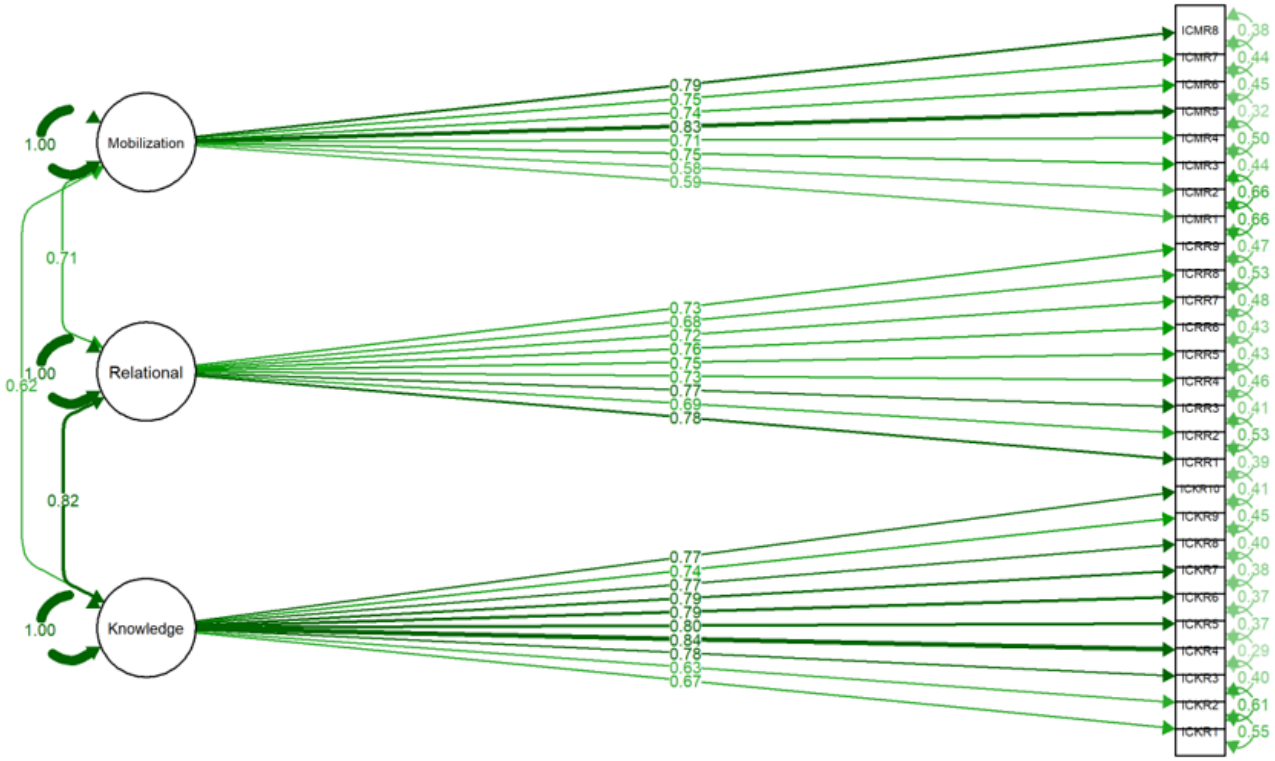


Figure 1

Confirmatory Factor Analysis Measurement Model for Institutional Capacity

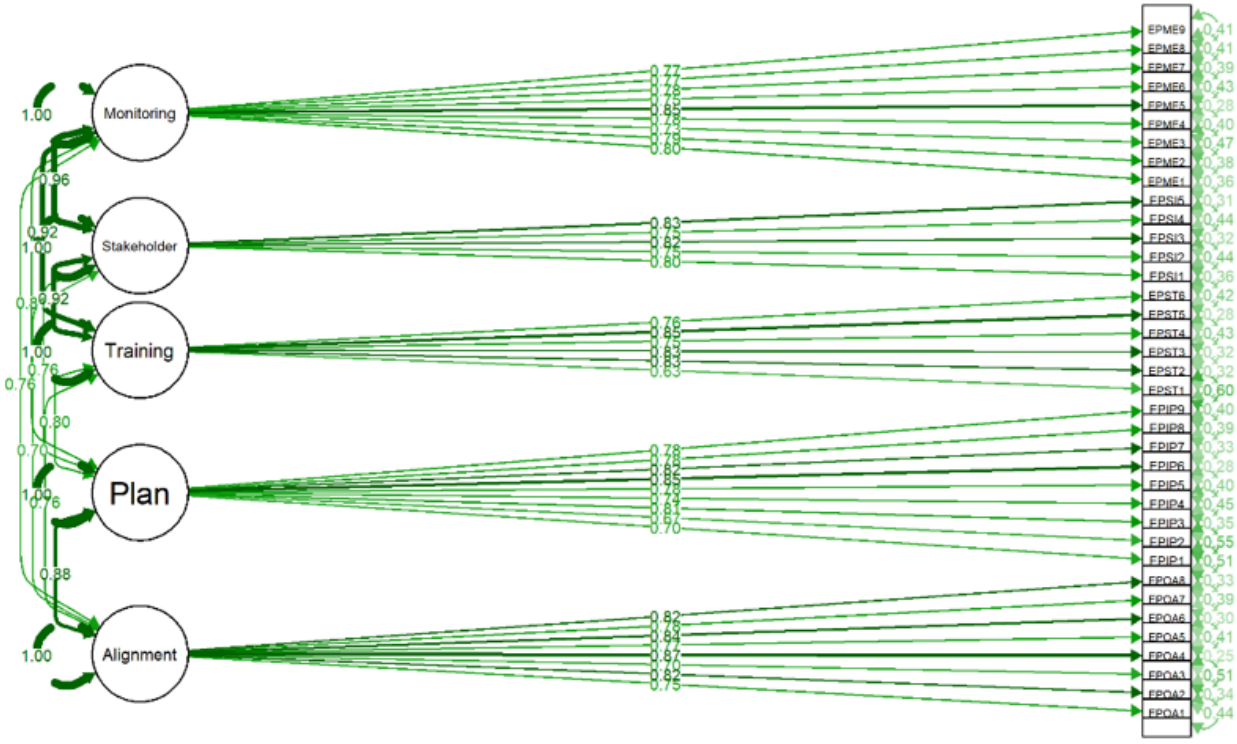


Figure 2

Confirmatory Factor Analysis Measurement Model for Policy Implementation

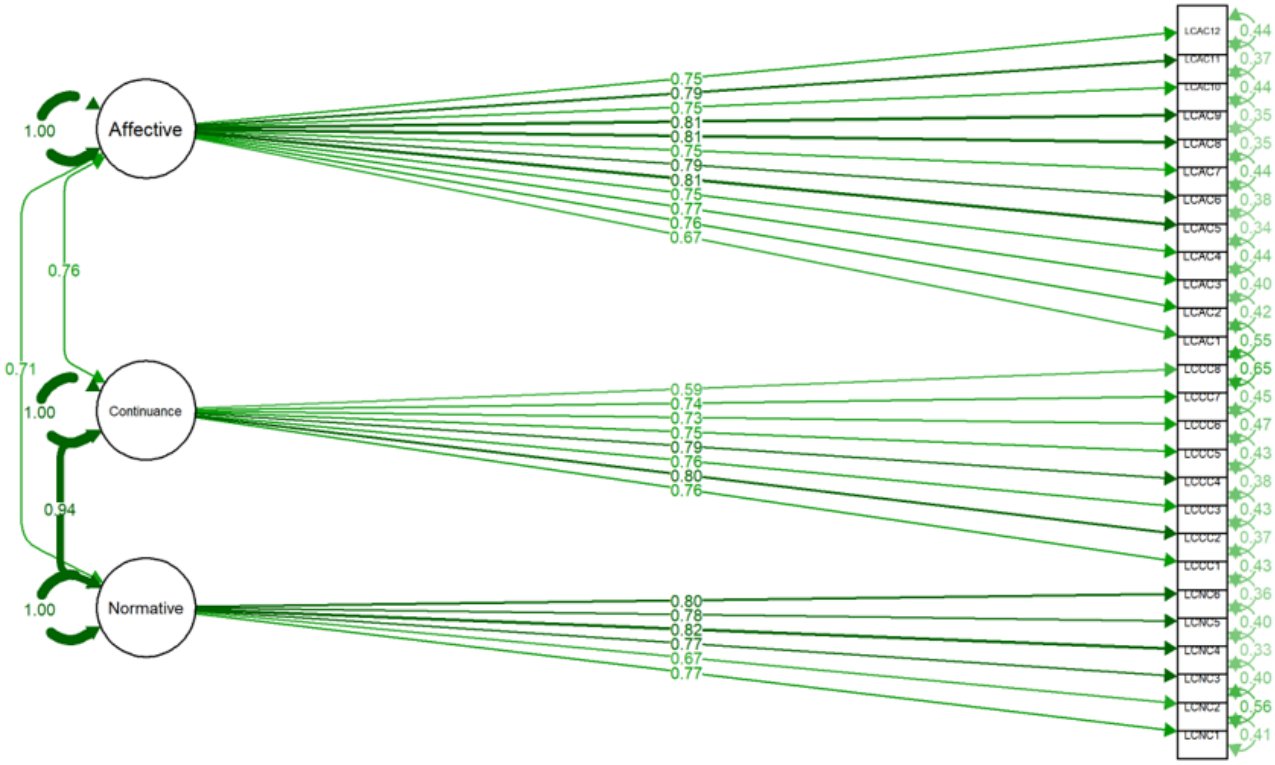


Figure 3

Confirmatory Factor Analysis Measurement Model for Leadership Commitment



Figure 4

Figure 1 *Standardized Factor Loadings for the Confirmatory Factor Analysis Measurement Model*

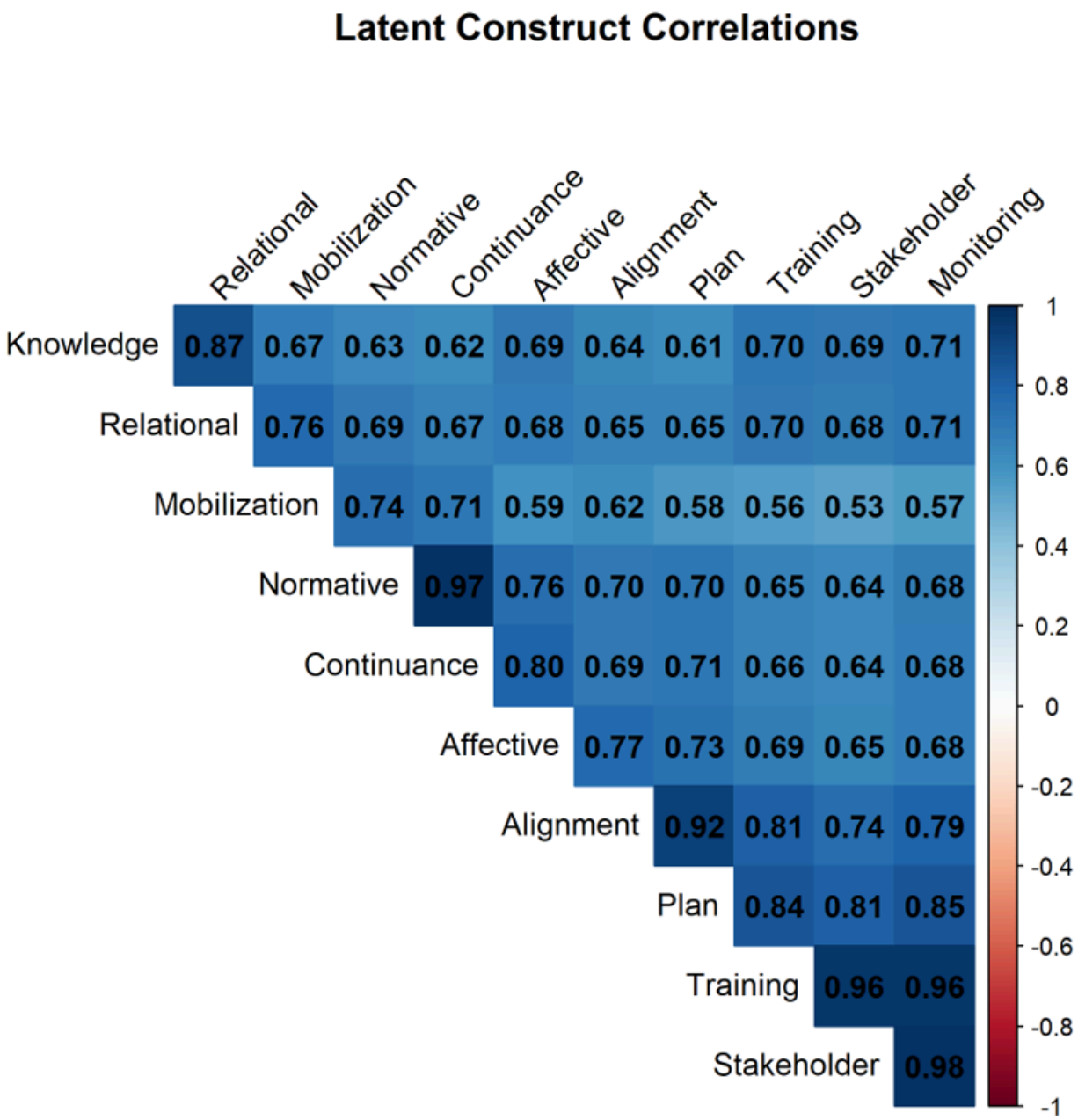


Figure 5

Figure 1 Latent Construct Correlations

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- AppendixI.docx