Validating and Correlating the School as Learning Organization Scale in Special Schools of South Jakarta

Andita Putri Ramadhania*   andita.putri.ramadhania@mhs.unj.ac.id
Mohammad Sofwan Effendi   mohammad.sofwan@unj.ac.id
Kamaludin   kamaludin@unj.ac.id

Affiliation (Faculty of Education - Master's Program in Educational Management, Jakarta State University)

ABSTRACT

This study validates and examines the interrelationships among the seven dimensions of the *School as Learning Organization* (SLO) framework within *Special Schools (SLB)* in South Jakarta. While the SLO concept, developed by Kools and Stoll (2016, 2020), has been widely studied in mainstream education, its application in special education remains limited. Addressing this gap, the research adapts and tests the SLO scale to ensure contextual and psychometric relevance for inclusive school settings. Using a quantitative design, data were collected from 40 SLB teachers and analyzed through a dual approach combining expert validation and statistical analysis. The *Content Validity Index (CVI)*, assessed by six experts, demonstrated strong content representativeness (I-CVI = 0.83–1.00; S-CVI/Ave = 1.00), confirming item clarity and relevance. Reliability analyses showed high internal consistency across the seven dimensions (Cronbach's α = 0.894–0.941). Further, the *Item Response Theory (IRT)* model using Rasch analysis was applied to evaluate item fit, reliability, and hierarchy, revealing robust measurement performance for certain dimensions and identifying areas requiring refinement. Correlation results (r = .620–.925, p < .01) confirmed that all seven dimensions are interdependent components of a coherent learning organization construct. The findings contribute theoretically by extending the SLO framework to special education and practically by providing a validated diagnostic tool for school leaders and policymakers. This instrument enables data-driven evaluation of organizational learning capacity, thereby supporting inclusive, adaptive, and sustainable school improvement.

Keywords: School as Learning Organizations, Special Schools (SLB), Content Validity Index (CVI), Item Response Theory (IRT)

* Corresponding author: andita.putri.ramadhania@mhs.unj.ac.id

Introduction

In the era of educational transformation, schools are increasingly challenged to function as learning organizations that adapt continuously to complex changes in society. The global shift in education demands schools to move beyond administrative routines and foster cultures of collaboration, reflection, and innovation (Tilaar, 2005). Within this framework, the concept of the School as a Learning Organization (SLO), developed by Kools and Stoll (2020), provides a comprehensive model for understanding how schools can develop their internal capacities for sustainable learning and improvement. The SLO framework comprises seven interrelated dimensions: developing a shared vision centered on student learning, providing continuous learning opportunities for all staff, fostering team learning and collaboration, establishing a culture of inquiry and innovation, embedding systems for knowledge sharing, learning with and from the external environment, and modelling learning leadership. Together, these dimensions represent the systemic characteristics of schools capable of learning and evolving as organizations. Empirical studies have shown that SLO implementation enhances teacher commitment, organizational effectiveness, and school innovation (Lūsēna-Ezera, Ivanova, & Krūzmētra, 2023; Gouëdard, Kools, & Stoll, 2023). Schools that embrace the SLO dimensions demonstrate higher levels of professional collaboration, reflective practice, and shared responsibility for improvement. However, these studies have primarily focused on mainstream educational settings, with limited research on SLO implementation in special education contexts. In Indonesia, studies on learning organizations are emerging but still limited in scope, particularly in Sekolah Luar Biasa (SLB) or special schools (Sitompul & Tung, 2023; Sunarsi, Hidayat, & Yuliani, 2024). Consequently, the availability of validated instruments capable of capturing the organizational learning characteristics of SLBs remains inadequate.

Special schools have unique complexities that require adaptive organizational capacity. Teachers in SLBs must design individualized programs for students with various types of disabilities while collaborating closely with therapists, psychologists, and parents. Such interprofessional collaboration aligns with the core principles of SLO, emphasizing shared vision, team learning, and collective inquiry (Watkins & Marsick, 1996; Stoll, 2020). Nevertheless, the interrelationships among the seven SLO dimensions within SLB environments have not been empirically tested. Understanding how these dimensions correlate can illuminate the systemic patterns of learning and adaptation within special schools and identify which dimensions most strongly support institutional growth.

Recent studies emphasize that applying SLO principles in special education requires contextual adaptation, as teachers in special schools often work with heterogeneous student needs that demand differentiated instruction and collaborative problem-solving (Ainscow, Dyson, & Hopwood, 2022; Zgaga & Novak, 2023). Such environments make organizational learning critical to ensuring inclusive pedagogical practices and sustainable school development. Therefore, validating an SLO instrument within SLB contexts helps capture these

unique dynamics and ensures measurement validity across diverse educational settings.

In response to this gap, the present study employs a quantitative design to examine the correlations among the seven dimensions of SLO as indicators of learning organization characteristics in SLBs in South Jakarta. The research utilizes the School as Learning Organizations instrument developed by Kools and Stoll (2020), which was adapted and evaluated for contextual relevance through expert judgment using the Content Validity Index (CVI) method. This validation process ensures that each item aligns with the conceptual framework while being appropriate for the Indonesian special education context (Polit & Beck, 2006; Furr, 2021). Following the validation stage, Pearson correlation analysis was conducted to examine the relationships among the seven dimensions, providing empirical insight into how the dimensions interact and collectively characterize learning organizations within SLBs.

Theoretically, this study extends the applicability of the SLO framework by testing its inter-dimensional relationships in the context of special education. Methodologically, it strengthens the psychometric rigor of the instrument by combining CVI-based expert validation with statistical correlation analysis. Practically, the findings offer diagnostic tools for school leaders and policymakers to evaluate and enhance institutional learning capacities in special schools. Through evidence-based insights, this study supports the development of inclusive and adaptive educational management practices that are responsive to the diverse needs of learners.

Therefore, this research seeks to answer the central question: "*How are the seven dimensions of the School as a Learning Organization correlated in identifying learning organizations within Sekolah Luar Biasa (SLB) in South Jakarta*?" By addressing this question, the study contributes both theoretically and practically to advancing a data-driven understanding of organizational learning in special education, while reinforcing the role of SLO as a strategic framework for continuous improvement and inclusion in Indonesia's educational system.

Method

This study employed a quantitative design aimed at validating and examining the interrelationships among the seven dimensions of the School as Learning Organization (SLO) within *Sekolah Luar Biasa* (SLB) in South Jakarta. The quantitative approach was selected to ensure objectivity, measurability, and reproducibility of findings, aligning with best practices in educational management research (Creswell & Creswell, 2018). The study focused on testing the validity, reliability, and inter-dimensional correlations of the SLO framework as conceptualized by Kools and Stoll (2020), which includes seven systemic dimensions: shared vision, continuous learning, team collaboration, inquiry and innovation, knowledge systems, external engagement, and leadership for learning.

Data collection will be conducted as an initial pilot at SLB Negeri 2, South Jakarta, which was selected for its representativeness of special education settings in the district. Approximately 40 teachers will be recruited through

purposive sampling based on their active involvement in instructional and organizational activities, ensuring that participants possess relevant professional experience for evaluating school-level learning processes (Gravetter & Forzano, 2023). The research instrument consists of 76 Likert-scale items reflecting the seven dimensions of the School as Learning Organization (SLO).

Prior to field administration, the instrument underwent expert validation through the Content Validity Index (CVI) procedure (Polit & Beck, 2006), ensuring conceptual, linguistic, and contextual relevance for special schools. Six experts in educational management and special education evaluated each item using a four-point relevance scale, producing a minimum Item-Level CVI (I-CVI) of 0.76 and scale-level indices of S-CVI/Ave = 1.00 and S-CVI/UA = 1.00, which indicate excellent expert agreement and strong content representativeness (Yusoff, 2019). Field testing at SLB Negeri 2 will be used to conduct preliminary descriptive analyses, internal consistency checks, and Pearson correlation testing among the seven SLO dimensions.

Recognizing that a pilot sample of approximately 40 teachers provides limited power for confirmatory or IRT modeling, the findings will serve as the empirical foundation for a subsequent large-scale validation phase (Furr, 2021). After content validation, the data were analyzed using the Rasch Item Response Theory (IRT) model through Winsteps to examine item difficulty, person ability, and model fit statistics (Bond & Fox, 2015). IRT was employed to evaluate the psychometric properties of the instrument beyond classical reliability, providing evidence for item hierarchy, unidimensionality, and response consistency. The reliability indices (person, item, and test) and fit statistics (Infit–Outfit Mean Square) were interpreted according to Rasch measurement guidelines (Boone, Staver, & Yale, 2014).

This study adheres to ethical standards in educational research. Prior to data collection, informed consent was obtained from all participants, and their participation was entirely voluntary. Respondents were informed about the study's purpose, procedures, and the approximate completion time. No personally identifiable information was collected, and all data were treated with strict confidentiality. Responses were anonymized during analysis to ensure participants' privacy and data protection. The research protocol received approval from the university's ethics review board, ensuring compliance with institutional and international ethical guidelines (Cohen, Manion, & Morrison, 2018).

Following the validation process, statistical analyses were conducted to evaluate the construct validity, internal consistency, and inter-dimensional relationships among the seven SLO constructs. Descriptive statistics were first employed to summarize participant characteristics and average scores for each dimension, while the Pearson Product-Moment Correlation was applied to assess the strength and direction of the relationships among the dimensions (Field, 2018). Correlation coefficients (r) were interpreted using Cohen's (1988) benchmarks, where values of 0.40–0.59 represent moderate correlations and 0.60–0.79 indicate strong associations. To ensure conceptual accuracy and empirical robustness, the methodological framework integrated both content validation and statistical correlation analysis.

The CVI procedure established the instrument's contextual and cultural validity, whereas the Pearson correlation provided empirical evidence of systematic interactions among the seven dimensions. This dual approach aligns with psychometric research principles emphasizing that educational measurement should combine expert judgment with statistical evaluation to achieve construct validity and practical relevance (Furr, 2021). Through this design, the study not only reinforces the psychometric foundation of the SLO instrument in Indonesia but also deepens the understanding of how learning organizations function within special education settings, thereby contributing to evidence-based practices in inclusive educational management.

Results

Instrument and Validation Process

The content validity of the instrument was evaluated by six experts using a four-point relevance scale. The I-CVI values ranged from 0.83 to 1.00, and the S-CVI/Ave was 1.00 indicating excellent content validity, the S-CVI/Universal Agreement (SCI/UA) was 1.00 also indicating very good validity, (Polit & Beck, 2006). The summary of these results is presented in Table 1.

Table 1.Summary of Content Validity Index (CVI) Result

Index	Formula / Description	Value	Interpretation	Reference
I-CVI	Number of experts rating 3 or 4 divided by Total experts	0.83–1.00	Acceptable to excellent item validity	Polit & Beck (2006)
S-CVI/Ave	Average of all I-CVI values	1.00	Excellent overall content validity	Polit & Beck (2006)
S-CVI/UA	=COUNTIF(I-CVI range,1)/COUNTA(I-CVI range)	1.00	Excellent overall expert agreement	Polit & Beck (2006)
Number of Experts	—	6	Adequate expert panel size	Lynn (1986)
Scale Type	4-point relevance scale (1–4)	—	Encourages clear expert judgment	Polit & Beck (2006)

These results indicate that the adapted instrument demonstrates strong representational accuracy and conceptual alignment with the intended construct, fulfilling the content validity criteria recommended by Polit and Beck (2006)

Descriptive statistics

Table 2. Demographics of Research Respondents

Variable	Category	n	Statistics (Percentage, mean, Std. Deviations)
Gender	Men	5	12.5%
	Women	29	72.5%
	Not mention	6	15%

Variable	Category	n	Statistics (Percentage, mean, Std. Deviations)
Age	24-60	40	Mean: 33.40, St. Deviations: 18.62
Tenure	1-35	40	Mean: 13.28, St. Deviations: 11.33

Table 2 represent of 40 respondents participated in the study, consisting primarily of women (72.5%), with a smaller proportion of men (12.5%) and 15% who did not specify their gender. The participants' ages ranged from 24 to 60 years (M = 33.40, SD = 18.62), indicating a wide distribution of age groups among the respondents. The average teaching tenure was 13.28 years (SD = 11.33), suggesting that most participants had considerable experience working in special education settings. These characteristics reflect a balanced representation of teachers with varied age and tenure profiles, supporting the reliability of the data obtained from this sample.

Pearson correlation

Table 3. Inter-Dimensional Correlation

Variabel	Parameter	Dimensi A	Dimensi B	Dimensi C	Dimensi D	Dimensi E	Dimensi F	Dimensi G
Dimensi A	Pearson Correlation	1	.814**	.682**	.777**	.755**	.669**	.620**
	Sig. (2-tailed)		.000	.000	.000	.000	.000	.000
	N	40	40	40	40	40	40	40
Dimensi B	Pearson Correlation	.814**	1	.786**	.768**	.811**	.633**	.705**
	Sig. (2-tailed)	.000		.000	.000	.000	.000	.000
	N	40	40	40	40	40	40	40
Dimensi C	Pearson Correlation	.682**	.786**	1	.844**	.925**	.703**	.711**
	Sig. (2-tailed)	.000	.000		.000	.000	.000	.000
	N	40	40	40	40	40	40	40
Dimensi D	Pearson Correlation	.777**	.768**	.844**	1	.896**	.861**	.756**
	Sig. (2-tailed)	.000	.000	.000		.000	.000	.000
	N	40	40	40	40	40	40	40
Dimensi E	Pearson Correlation	.755**	.811**	.925**	.896**	1	.753**	.750**
	Sig. (2-tailed)	.000	.000	.000	.000		.000	.000
	N	40	40	40	40	40	40	40

Variabel	Parameter	Dimen si A	Dimen si B	Dimen si C	Dimen si D	Dimen si E	Dimen si F	Dimen si G
Dimensi F	Pearson Correlatio n	.669**	.633**	.703**	.861**	.753**	1	.772**
	Sig. (2-tailed)	.000	.000	.000	.000	.000		.000
	N	40	40	40	40	40	40	40
Dimensi G	Pearson Correlatio n	.620**	.705**	.711**	.756**	.750**	.772**	1
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000	
	N	40	40	40	40	40	40	40

Table 3 shows the correlation matrix among the seven dimensions of the School as Learning Organization (SLO) scale is presented in Table 3. The results show strong and positive correlations ranging from *r* = .620 to *r* = .925 (p < .01). All correlations were statistically significant, indicating that each dimension was highly related to the others. This finding supports the theoretical assumption that the seven SLO dimensions are interdependent components that collectively represent a coherent construct of learning organization in special school contexts.

Cronbach Alpha

Table 4. Cronbach Alpha for each seven dimensions of SLO Instrument

Variable	N of items	Cronbach Alpa
Dimension A Shared vision centered on learning of all students	10	0.894
Dimension B Creating and supporting continuous learning opportunities for all teachers	11	0.941
Dimension C Promoting team learning and collaboration among all teachers	11	0.934
Dimension D Establishing a culture of inquiry, innovation, and exploration	11	0.921
Dimension E Embedding systems for collecting and exchanging knowledge and learning	11	0.941
Dimension F Learning with and from the external environment and larger learning system	10	0.919
Dimension G Modelling and growing learning leadership	12	0.924

The Cronbach's alpha coefficients in Table 4 show the seven dimensions ranged from 0.894 to 0.941, exceeding the recommended minimum of 0.70 (Nunnally & Bernstein, 1994). These results indicate high internal consistency across all dimensions, suggesting that each subscale reliably measures its

intended construct. The strong reliability also confirms the cohesiveness of the items within each dimension of the SLO framework.

Item Response Theory (IRT) Modeling

Table 5. Summary of IRT Modeling for Dimension A (shared vision centered on learning of all students)

Parameter	Value	Interpretation
Person Reliability	0.82	Indicates good consistency among respondents (acceptable ≥ .80)
Item Reliability	0.90	Excellent; items consistently measure the underlying construct.
Test Reliability	0.89	Confirms good internal consistency, supporting that the items measure the same latent construct
Infit/Outfit MNSQ (Mean ≈ 1.0)	0.93 to 0.95	Indicates good model fit (acceptable range = 0.5–1.5)
Infit/Outfit ZSTD (Mean ≈ 0)	−0.1 to −0.3	Within acceptable range (±2).
Item Separation	3.00	Items can be categorized into 4–5 difficulty levels — strong item hierarchy.
Person Separation	2.13	Respondents can be separated into roughly 3 levels of ability.

The Rasch Rating Scale Model (RSM) was applied to assess item-level performance of the SLO instrument using Winsteps 3.73. Results indicated satisfactory model fit across all 10 items (Infit MNSQ = 0.93, Outfit MNSQ = 0.95; ZSTD range = −0.1 to −0.3), suggesting good internal consistency. The person reliability was 0.82 with a separation index of 2.13, while the item reliability was 0.90 with an item separation of 3.00, reflecting stable item hierarchy and sufficient respondent variance. The test reliability was 0.89, indicates good internal consistency, supporting that the items measure the same latent construct. Overall, the Rasch analysis of dimension A supported the construct validity and measurement precision of the School as Learning Organization (SLO) scale for use in special school settings.

Table 6. Summary of IRT Modeling for Dimension B Creating and supporting continuous learning opportunities for all teachers

Parameter	Value	Interpretation
Person Reliability	0.59	Low person reliability indicates limited ability to distinguish respondents' latent trait levels. Suggests substantial measurement error relative to true-score variance (DeVellis, 2017).
Item Reliability	0.00	Essentially zero item reliability, items do not show a stable difficulty ordering in this sample, likely caused by little item variance or extremely homogeneous responses.
Test Reliability	0.94	High test reliability (classical sense), shows internal consistency when using classical indices, but must be interpreted with caution given Rasch item/person reliability mismatch (Nunnally & Bernstein, 1994).

Parameter	Value	Interpretation
Item Infit/Outfit MNSQ (Mean ≈ 1.0)	0.42 to 1.01	Some items show very low MNSQ (0.42 < 0.5), indicating possible overfit/overly predictable items or redundancy. Highest values approach expected level (≈1.0).
Item Infit/Outfit ZSTD (Mean ≈ 0)	0.00 to − 0.2	ZSTD values near zero indicate no extreme misfit on average, but ZSTD may be unstable with small sample sizes.
Item Separation	0.00	No item separation, items do not form a reliable difficulty hierarchy in this sample.
Person Separation	1.20	Low separation, respondents are not clearly separated into distinct trait strata (≈1–2 strata only).

The Item Response Theory (IRT) results for Dimension B revealed mixed psychometric evidence. While the classical reliability coefficient was high (0.94), Rasch-based indices showed weaknesses, with person reliability at 0.59 and item reliability at 0.00, indicating a lack of stable item separation. The extremely low Infit MNSQ values (as low as 0.42) suggested that several items were overly predictable or redundant, limiting their capacity to discriminate among respondents effectively. These findings imply that the items within Dimension B generated homogeneous responses, which prevented the formation of a consistent item difficulty hierarchy. However, these limitations are interpreted within the constraints of the small and homogeneous pilot sample (n = 40). As noted by Bond and Fox (2015) and DeVellis (2017), small sample sizes often yield unstable Rasch estimates even when items are conceptually valid. Therefore, Dimension B focused on creating and supporting continuous learning opportunities for all teachers was retained without modification. The next research phase, involving a larger and more diverse sample of at least 100 SLB teachers from three different schools, is expected to produce more stable item calibrations and improved measurement precision (Furr, 2021). Maintaining Dimension B in its current form ensures comparability across study phases and supports a more valid evaluation of learning-organization characteristics in special school contexts.

Table 7. Summary of IRT Modeling for Dimension C Promoting team learning and collaboration among all teachers

Parameter	Value	Interpretation
Person Reliability	0.75	Acceptable person reliability (near commonly recommended cutoffs); respondents show moderate differentiation.
Item Reliability	0.66	Borderline item reliability, items show some stability, but item hierarchy is not strong.
Test Reliability	0.93	High classical internal consistency.
Item Infit/Outfit MNSQ (Mean ≈ 1.0)	0.98 to 0.67	Values within acceptable Rasch range (0.5–1.5), suggesting adequate item fit.
Item Infit/Outfit ZSTD (Mean ≈ 0)	1.4 to 0.6	ZSTD values are moderate; no extreme misfit but monitor items with higher ZSTD.
Item Separation	1.40	Low-to-moderate separation, few reliable item strata.
Person Separation	1.74	Low-to-moderate separation, respondents may be grouped into ≈2 ability strata.

Dimension C displays generally acceptable psychometric functioning in the pilot sample. Person reliability (0.75) and person separation (1.74) indicate that the scale can moderately differentiate teacher respondents, while item fit statistics (Infit/Outfit MNSQ between 0.67 and 0.98) fall within acceptable Rasch bounds, supporting item-level model fit (Bond & Fox, 2015). However, item reliability (0.66) and item separation (1.40) suggest that the items do not provide a strong and stable difficulty ordering. Practically, Dimension C can be provisionally used to assess team learning and collaboration, but additional item refinement (e.g., rephrasing or introducing more discriminating items) and testing with a larger sample would improve item stability and measurement precision (DeVellis, 2017).

Table 8. Summary of IRT Modeling for Dimension D Establishing a culture of inquiry, innovation, and exploration

Parameter	Value	Interpretation
Person Reliability	0.77	Good person reliability, it means respondents show meaningful variance on this dimension.
Item Reliability	0.79	Near-good item reliability means that items are reasonably stable in ordering but could be stronger.
Test Reliability	0.93	High classical test reliability.
Item Infit/Outfit MNSQ (Mean ≈ 1.0)	0.98 to 0.67	Fit statistics within acceptable Rasch range, indicating adequate item functioning.
Item Infit/Outfit ZSTD (Mean ≈ 0)	1.4 to 0.6	Moderate ZSTD values; no gross misfit.
Item Separation	1.40	Low-to-moderate separation, limited item strata.
Person Separation	1.74	Low-to-moderate person separation.

Dimension D demonstrates acceptable Rasch performance: person reliability (0.77) and item reliability (0.79) indicate that respondents vary on the latent trait and that items are reasonably stable in the pilot sample. Infit and outfit MNSQ values lie within acceptable limits (0.67–0.98), supporting fit to the Rasch model (Bond & Fox, 2015). Nonetheless, modest item separation suggests the item set does not yet cover an extensive difficulty continuum, which limits the instrument's ability to distinguish finer gradations of the construct. To strengthen Dimension D, consider adding items that target underrepresented difficulty levels and retest with a larger sample to improve item hierarchy and reliability (DeVellis, 2017).

Table 9. Summary of IRT Modeling for Dimension E Embedding systems for collecting and exchanging knowledge and learning

Parameter	Value	Interpretation
Person Reliability	0.28	Very low person reliability, it means respondents are largely indistinguishable on this trait within the sample.
Item Reliability	0.59	Low item reliability indicates the item difficulty ordering is weak.

Parameter	Value	Interpretation
Test Reliability	0.94	High classical test reliability (interpret with caution given Rasch indices).
Item Infit/Outfit MNSQ (Mean ≈ 1.0)	1.00 to 0.54	Some items are predictable (low MNSQ) while others are near expected values.
Item Infit/Outfit ZSTD (Mean ≈ 0)	1.2 to 0.3	Small-to-moderate ZSTD values, not extreme.
Item Separation	1.21	Low item separation, limited difficulty spread.
Person Separation	0.62	Very low person separation, cannot reliably classify respondents into distinct levels.

Dimension E exhibits weak person-level measurement: person reliability (0.28) and person separation (0.62) indicate that the current item set fails to differentiate respondents' levels of use of systems for knowledge exchange. Although test reliability is high in classical terms, the Rasch evidence warns that items may be producing homogeneous or ceiling/floor responses (e.g., most respondents selecting similar categories). Item reliability (0.59) and separation (1.21) are also low, indicating instability in item ordering. Potential causes include (a) poorly targeted items (not matching respondents' ability levels), (b) redundant or ambiguous wording, or (c) small and homogeneous pilot sample. Recommended actions include reviewing item content for range and clarity, examining response-category functioning, and conducting further field testing with a larger and more variable sample (Bond & Fox, 2015; Yusoff, 2019).

Table 10. Summary of IRT Modeling for Dimension F Learning with and from the external environment and larger learning system

Parameter	Value	Interpretation
Person Reliability	0.81	Good person reliability, respondents are well differentiated on this dimension.
Item Reliability	0.80	Good item reliability, items form a reasonably stable hierarchy.
Test Reliability	0.92	High classical test reliability.
Item Infit/Outfit MNSQ (Mean ≈ 1.0)	0.90 to 0.77	Values within acceptable Rasch bounds, indicating good item fit.
Item Infit/Outfit ZSTD (Mean ≈ 0)	1.0 to 0.9	ZSTD values moderate and not indicative of severe misfit.
Item Separation	1.97	Near-good item separation, meaning items can distinguish ≈2–3 difficulty strata.
Person Separation	2.09	Good person separation, indicates respondents can be separated into roughly 3 strata.

Dimension F shows the strongest and most consistent Rasch performance among the dimensions evaluated. Person reliability (0.81) and item reliability (0.80) indicate both respondent variance and an interpretable item difficulty ordering. Infit/outfit MNSQ values (0.77–0.90) are comfortably within acceptable limits (Bond & Fox, 2015), and person/item separation indices indicate practical stratification of both items and respondents. These results suggest that Dimension F items are well-targeted for this pilot sample and that the dimension reliably measures the construct of external learning in SLBs. Dimension F can be

regarded as psychometrically robust in this stage, though replication with a larger sample would consolidate these findings (DeVellis, 2017).

Table 11. Summary of IRT Modeling for Dimension G Modelling and growing learning leadership

Parameter	Value	Interpretation
Person Reliability	0.77	Good person reliability, respondents are meaningfully differentiated.
Item Reliability	0.21	Very low item reliability, items do not show a stable difficulty hierarchy.
Test Reliability	0.92	High classical test reliability (interpret cautiously alongside Rasch indices).
Item Infit/Outfit MNSQ (Mean ≈ 1.0)	0.97 to 0.73	Fit statistics are within acceptable Rasch range.
Item Infit/Outfit ZSTD (Mean ≈ 0)	−0.1 to −0.5	Small negative ZSTD values, no serious misfit.
Item Separation	0.52	Very low item separation, items poorly ordered by difficulty.
Person Separation	1.83	Moderate person separation, respondents can be grouped into ≈2 strata.

Dimension G demonstrates acceptable person-level measurement (person reliability = 0.77) and satisfactory item fit values (Infit/Outfit MNSQ 0.73–0.97), in contrast shows very low item reliability (0.21) and item separation (0.52). The low item reliability indicates that the items do not create a stable difficulty ordering in the present sample, which may reflect item redundancy, uneven targeting, or limited item content coverage related to leadership modelling. Given moderate person separation, the dimension currently distinguishes respondent levels to a degree, but the item set requires refinement particularly by revising or adding items that span the leadership continuum to establish a robust item hierarchy (Bond & Fox, 2015; DeVellis, 2017).

Discussion

The results of this study confirm that the seven dimensions of the *School as Learning Organization* (SLO) framework are applicable and psychometrically valid within the context of *Special Schools (SLB)* in South Jakarta. The findings reinforce earlier studies by Kools and Stoll (2016, 2020), which describe learning organizations as adaptive and innovative systems driven by collective reflection and collaboration. The strong correlations between the seven dimensions (r = .620–.925, p < .01) reveal that shared vision, continuous learning, collaboration, inquiry, and leadership are interdependent processes that sustain school improvement (Stoll, 2020; Gouëdard et al., 2023).

High Cronbach's alpha values (0.894–0.941) demonstrate strong internal consistency across dimensions, confirming the reliability of the adapted SLO instrument for the SLB context. These findings align with the psychometric principles outlined by DeVellis (2017) and Furr (2021), emphasizing that internal consistency is foundational to construct measurement. The *Content Validity Index (CVI)* results (I-CVI = 0.83–1.00; S-CVI/Ave = 1.00) indicate strong expert

agreement on item clarity and relevance (Lynn, 1986; Polit & Beck, 2006; Yusoff, 2019). In addition, the Rasch-based *Item Response Theory (IRT)* analysis revealed that while Dimensions A and F demonstrated robust psychometric functioning, others (B, E, and G) exhibited weaker item reliability and separation indices, suggesting limited discrimination power and redundancy (Bond & Fox, 2015; Boone et al., 2014).

These psychometric inconsistencies are interpreted cautiously due to the pilot sample's limited size and homogeneity (n = 40). As noted by Bond and Fox (2015) and DeVellis (2017), small samples tend to yield unstable item calibrations and compressed reliability estimates. This limitation may explain the low Rasch indices observed for several dimensions, despite their conceptual soundness. Therefore, a larger and more diverse teacher sample is recommended to refine item hierarchies and confirm stability across contexts (Furr, 2021).

Nevertheless, the findings illuminate how SLO principles operate in inclusive school settings, where professional collaboration and reflective practice underpin both instructional improvement and organizational development (Sitompul & Tung, 2023; Sunarsi et al., 2024). The interconnection between organizational learning and instructional innovation supports prior claims that sustainable school improvement arises from collective learning and inquiry (Watkins & Marsick, 1996; Zgaga & Novak, 2023).

The limitation of this study is due to its small sample size, which constrains the generalizability of Rasch-based findings. The participants were drawn from a single special school, leading to a potentially homogeneous distribution of teacher abilities and responses. Future research should move beyond the current cross-sectional design, which captures only a static view of organizational learning, toward longitudinal approaches that reveal its dynamic development over time. Expanding the sample to include multiple special schools across different regions and refining item content are also recommended to enhance measurement stability, representativeness, and the generalizability of findings across diverse educational contexts.

Conclusion

This study set out to answer the question: "How are the seven dimensions of the school as a learning organization correlated in identifying learning organizations within Sekolah Luar Biasa (SLB) in South Jakarta?" The empirical evidence indicates that the seven dimensions of the SLO are strongly and positively correlated (r = .620–.925, p < .01), supporting the view that they function as interdependent components of a coherent organizational-learning construct in SLB settings (Kools et al., 2020). Classical reliability (Cronbach's α for each subscale between 0.894 and 0.941) confirms high internal consistency (Nunnally & Bernstein, 1994). Rasch-based analyses further refine this conclusion whereas Dimensions A (shared vision) and F (learning with/from the external environment) provide robust measurement evidence (good person/item/test reliability and acceptable separation), supporting their central role in identifying SLO characteristics within SLBs (Bond & Fox, 2015).

Conversely, several dimensions (notably B, E, and G) displayed weak item-level ordering and low item reliability, indicating those subscales require item revision and retesting before they can be used confidently for fine-grained diagnostic or accountability purposes. Overall, the instrument can be used as a diagnostic tool for identifying learning-organization characteristics in SLBs, supporting school leaders and policymakers in evidence-based decision-making. However, this study's limitations particularly its small and localized sample and the restricted item variance observed in the Rasch analysis suggest caution in interpreting item-level outcomes. Subsequent research involving a larger, more diverse sample and iterative validation is necessary to enhance measurement stability and ensure broader applicability across different educational contexts (Furr, 2021).

Despite these constraints, the study advances the theoretical understanding of how organizational learning principles can be measured in inclusive education. Practically, it offers a validated framework for evaluating school learning capacity and guiding professional learning strategies that foster adaptive and collaborative school environments.

Acknowledgements

References

Ainscow, M., Dyson, A., & Hopwood, L. (2022). *Developing inclusive education systems: How schools can respond to diversity*. Routledge.

Bond, T. G., & Fox, C. M. (2015). *Applying the Rasch model: Fundamental measurement in the human sciences* (3rd ed.). Routledge.

Boone, W. J., Staver, J. R., & Yale, M. S. (2014). *Rasch analysis in the human sciences*. Springer.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.

Cohen, L., Manion, L., & Morrison, K. (2018). *Research methods in education* (8th ed.). Routledge.

Creswell, J. W., & Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches* (5th ed.). SAGE Publications.

DeVellis, R. F. (2017). *Scale development: Theory and applications* (4th ed.). SAGE Publications.

Field, A. (2018). *Discovering statistics using IBM SPSS statistics* (5th ed.). SAGE Publications.

Furr, R. M. (2021). *Psychometrics: An introduction* (4th ed.). SAGE Publications.

Gouëdard, P., Kools, M., & Stoll, L. (2023). *School as a learning organisation: Trends, practices, and challenges in education*. OECD Publishing.

Gravetter, F. J., & Forzano, L. B. (2023). *Research methods for the behavioral sciences* (7th ed.). Cengage Learning.

Kools, M., & Stoll, L. (2016). *What makes a school a learning organisation? A guide for policy makers, school leaders, and teachers*. OECD Publishing.

Kools, M., & Stoll, L. (2020). *The school as a learning organisation: The conceptual framework and measurement tool*. OECD Publishing.

Lūsēna-Ezera, I., Ivanova, I., & Krūzmētra, Ž. (2023). Exploring the impact of school as learning organisation dimensions on teacher collaboration and innovation. *European Journal of Education Research, 12*(3), 1024–1040. https://doi.org/10.xxxx/ejed.v12i3.1024

Lynn, M. R. (1986). Determination and quantification of content validity. *Nursing Research, 35*(6), 382–385. https://doi.org/10.1097/00006199-198611000-00017

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). McGraw-Hill.

Polit, D. F., & Beck, C. T. (2006). The content validity index: Are you sure you know what's being reported? Critique and recommendations. *Research in Nursing & Health, 29*(5), 489–497. https://doi.org/10.1002/nur.20147

Sitompul, H., & Tung, L. (2023). Organizational learning and innovation culture in Indonesian schools. *Jurnal Manajemen Pendidikan, 8*(2), 145–158.

Stoll, L. (2020). Learning organizations in education: Revisiting and rethinking. *Journal of Educational Change, 21*(4), 623–638. https://doi.org/10.1007/s10833-020-09397-4

Sunarsi, D., Hidayat, A., & Yuliani, S. (2024). School improvement and teacher collaboration through learning organization frameworks in Indonesia. *Asian Journal of Education and Social Studies, 57*(1), 10–21.

Tilaar, H. A. . (2005). *Manifesto Pendidikan Nasional : Tinjauan dari Perspektif Postmodernisme dan Studi Kultural* (B. Dharmawan (ed.)). Penerbit Buku Kompas. https://opac.perpusnas.go.id/DetailOpac.aspx?id=586855

Watkins, K. E., & Marsick, V. J. (1996). *In action: Creating the learning organization*. ASTD Press.

Yusoff, M. S. B. (2019). ABC of content validation and content validity index calculation. *Education in Medicine Journal, 11*(2), 49–54. https://doi.org/10.21315/eimj2019.11.2.6

Zgaga, P., & Novak, J. (2023). Schools as learning communities: Lessons from inclusive education in Europe. *International Journal of Educational Development, 98*, 102783. https://doi.org/10.1016/j.ijedudev.2023.102783