JP3I (Jurnal Pengukuran Psikologi dan Pendidikan Indonesia)
DOI: http://dx.doi.org/10.15408/jp3i.v15i2.49207
http://journal.uinjkt.ac.id/index.php/jp3i

Development and Psychometric Validation of the Belonging-Supportive Deep Learning Environments (BSDLE) Instrument Using CFA and Multidimensional IRT

Oktavia[1], Wardani Rahayu[1], Iva Sarifah[1]
Educational Research and Evaluation, State University of Jakarta[1]
oktavia@uinjkt.ac.id

Abstract

This study aimed to develop and validate a multidimensional instrument for assessing Belonging-Supportive Deep Learning Environments (BSDLE) in higher education. A quantitative, cross-sectional instrument development design was employed. Data were collected from 825 university students from five Indonesian universities implementing Flipped Learning. Participants completed a 47-item Likert-scale questionnaire measuring six dimensions: student cohesiveness, involvement, investigation, task orientation, cooperation, and faculty–staff relations. Data analysis was conducted in several stages. Preliminary data screening indicated no missing data and adequate variability. Confirmatory Factor Analysis (CFA) using a multidimensional correlated model demonstrated acceptable model fit ($\chi^2 = 2966.51$, df = 999, $p < .001$; RMSEA = 0.049; CFI = 0.97; NFI = 0.95; RFI = 0.95), with most standardized factor loadings exceeding 0.50. Further psychometric evaluation was conducted using Multidimensional Item Response Theory (MIRT) with the Graded Response Model (GRM). The results indicated that most items demonstrated moderate to high discrimination parameters (approximately 0.85–2.10) and ordered threshold parameters, suggesting appropriate functioning of the response categories. Marginal reliability estimates (0.78–0.89) were generally higher than Cronbach's alpha estimates (0.63–0.80), indicating improved measurement precision under the IRT framework. The Test Information Function (TIF) analysis revealed that the Investigation (IN) and Involvement (IV) dimensions provided higher measurement precision across the ability range (θ = -1 to +2), whereas the student cohesiveness (SC), cooperation (CO), and faculty–staff relations (FR) dimensions showed relatively lower information values and require further refinement. The findings provide preliminary evidence supporting the construct validity and measurement reliability of the BSDLE instrument as a multidimensional measure of supportive learning environments in higher education. Further refinement, cross-validation, and additional psychometric evaluation are recommended before broader diagnostic applications.

Keywords: Belonging-Supportive Deep Learning Environments (BSDLE), confirmatory factor analysis, graded response model, instrument development, multidimensional item response theory.

Abstrak

Penelitian ini bertujuan untuk mengembangkan dan memvalidasi instrumen multidimensional untuk mengukur Belonging-Supportive Deep Learning Environments (BSDLE) pada pendidikan tinggi. Penelitian menggunakan desain pengembangan instrumen kuantitatif dengan pendekatan potong lintang (cross-sectional). Data dikumpulkan dari 825 mahasiswa yang berasal dari lima perguruan tinggi di Indonesia yang menerapkan pembelajaran Flipped Learning. Responden mengisi kuesioner berbentuk skala Likert yang terdiri atas 47 item yang mengukur enam dimensi, yaitu student cohesiveness, involvement, investigation, task orientation, cooperation, dan faculty–staff relations. Analisis data dilakukan melalui beberapa tahap. Hasil pemeriksaan awal data menunjukkan tidak terdapat data yang hilang dan memiliki variasi respons yang memadai. Analisis Faktor Eksploratori (Exploratory Analisis Faktor Konfirmatori (Confirmatory Factor Analysis / CFA) menggunakan model multidimensional berkorelasi menunjukkan kesesuaian model yang dapat diterima ($\chi^2 = 2966,51$, df = 999, $p < .001$; RMSEA = 0,049; CFI = 0,97; NFI = 0,95; RFI = 0,95), dengan sebagian besar nilai standardized factor loading berada di atas 0,50. Evaluasi psikometrik lanjutan dilakukan menggunakan Multidimensional Item Response Theory (MIRT) dengan model Graded Response Model (GRM). Hasil analisis menunjukkan bahwa sebagian besar item memiliki parameter diskriminasi sedang hingga tinggi (sekitar 0,85–2,10) serta nilai threshold yang berurutan, yang menunjukkan bahwa kategori respons berfungsi dengan baik. Nilai reliabilitas marginal (0.78–0.89) secara umum lebih tinggi dibandingkan nilai Cronbach's alpha (0,63–0,80), yang menunjukkan peningkatan ketepatan pengukuran dalam kerangka IRT. Analisis Test Information Function (TIF) menunjukkan bahwa dimensi Investigation (IN) dan Involvement (IV) memberikan tingkat ketepatan pengukuran yang lebih tinggi pada rentang kemampuan θ = -1 hingga +2, sedangkan dimensi student cohesiveness (SC), cooperation (CO), dan faculty–staff relations (FR) menunjukkan tingkat informasi yang relatif lebih rendah sehingga masih memerlukan penyempurnaan lebih lanjut. Temuan penelitian ini memberikan bukti awal (preliminary evidence) mengenai validitas konstruk dan reliabilitas pengukuran instrumen BSDLE sebagai instrumen multidimensional untuk mengukur lingkungan pembelajaran suportif di pendidikan tinggi. Penyempurnaan item, validasi silang, dan evaluasi psikometrik lanjutan masih diperlukan sebelum instrumen digunakan secara lebih luas untuk tujuan diagnostik.

Kata kunci: *Belonging-Supportive Deep Learning Environments (BSDLE), , confirmatory factor analysis, graded response model, pengembangan instrumen, multidimensional item response theory.*

This is an open access article under CC-BY-SA license (https://creativecommons.org/licenses/by-sa/4.0/)

Introduction

The quality of the learning environment in higher education has become a central concern in educational research because it influences students' academic engagement, persistence, motivation, and psychological well-being (Freeman, Anderman, & Jensen, 2017; Gopalan & Brady, 2020). Beyond instructional delivery, contemporary perspectives emphasise that effective learning environments should not only facilitate knowledge construction but also create conditions in which students experience social connection, academic support, and psychological safety. One instructional approach that reflects this paradigm is Flipped Learning, which reorganises traditional teaching by transferring initial content exposure to pre-class activities and maximising classroom time for interactive learning, collaboration, problem solving, and higher-order thinking (Bergmann & Sams, 2012; Zainuddin & Perera, 2019). However, empirical evidence suggests that the success of Flipped Learning depends not only on instructional structure but also on the quality of the learning environment that supports student participation and meaningful engagement (Lo & Hew, 2020; Bond, 2020).

Although previous studies have examined learning environments extensively, existing conceptualisations often separate environmental characteristics, social relationships, and students' psychological experiences. Traditional learning environment frameworks primarily focus on structural and pedagogical dimensions, such as classroom organisation, task orientation, cooperation, and teacher support (Fraser, 2012). Meanwhile, student belongingness research emphasises students' emotional attachment, acceptance, and identification with a learning community (Strayhorn, 2019; Slaten et al., 2018). Classroom engagement frameworks mainly capture behavioural, cognitive, and emotional participation during learning activities, whereas deep learning environment models focus on instructional approaches that encourage critical thinking, knowledge integration, and conceptual understanding.

Despite their contributions, these constructs are partially overlapping but theoretically distinct. A learning environment may provide opportunities for interaction without necessarily producing a sense of belonging. Similarly, students may demonstrate engagement without experiencing psychological connectedness or relational support. Deep learning environments emphasise cognitive processes but may not adequately explain the interpersonal and emotional conditions required for sustained participation in learning. Therefore, existing frameworks do not fully capture the integrated relationship between students' perception of learning support, social belonging, and deep learning opportunities.

To address this conceptual limitation, this study introduces Belonging-Supportive Deep Learning Environment (BSDLE) as a multidimensional construct that represents the extent to which a learning environment simultaneously supports deep cognitive engagement and students' psychological belonging. BSDLE is not conceptualised as a simple combination of existing constructs; rather, it represents an integrated educational environment in which relational support enables meaningful and sustained deep learning. The theoretical boundary of BSDLE lies in its emphasis on the interaction between learning processes (deep learning orientation), social relationships (belonging and connectedness), and institutional support mechanisms within a single measurement framework.

Specifically, BSDLE differs from existing constructs in several important ways. First, unlike general learning environment measures that primarily assess classroom structures and instructional characteristics, BSDLE incorporates students' perceived relational security and connectedness as essential components of learning quality. Second, unlike student belongingness scales that primarily measure emotional attachment to peers, institutions, or academic communities, BSDLE examines belongingness as an active mechanism that facilitates deeper learning processes. Third, unlike classroom engagement measures that focus on students' participation behaviours, BSDLE evaluates the environmental conditions that enable engagement. Fourth, unlike deep learning environment models that emphasise cognitive challenge and instructional design, BSDLE incorporates psychosocial dimensions that sustain students' willingness to engage in complex learning activities.

The relationship between Flipped Learning and BSDLE requires careful conceptual clarification. Although Flipped Learning provides an appropriate context for examining BSDLE because it emphasises collaboration, peer interaction, student autonomy, and active knowledge construction (Lo & Hew, 2020; Bond, 2020), BSDLE is not limited exclusively to Flipped Learning environments. Instead, BSDLE is designed as a broader higher education learning environment construct that can be applied across various

http://journal.uinjkt.ac.id/index.php/jp3i
This is an open access article under CC-BY-SA license (https://creativecommons.org/licenses/by-sa/4.0/)

student-centred instructional approaches, including collaborative learning, problem-based learning, and blended learning. Flipped Learning represents one context in which BSDLE characteristics are expected to become particularly visible because the model requires a supportive environment for successful implementation. Therefore, the present study does not position BSDLE as a Flipped Learning-specific instrument but as a general diagnostic framework applicable across diverse higher education contexts.
From a measurement perspective, BSDLE consists of six theoretically related dimensions: student cohesiveness, involvement, investigation, task orientation, cooperation, and faculty–student relations. These dimensions are grounded in established learning environment theories (Fraser et al., 1996; Chionh & Fraser, 2009) and extended by integrating belongingness perspectives (Slaten et al., 2018). Student cohesiveness and faculty–student relations reflect relational and belonging-oriented processes, whereas involvement, investigation, task orientation, and cooperation reflect cognitive and behavioural conditions that support deep learning. Thus, these dimensions are conceptually related but remain empirically distinguishable, supporting multidimensional measurement approaches.

Previous instruments developed internationally have contributed valuable insights into assessing learning environments. For example, the Classroom Environment Scale, What Is Happening in This Class (WIHIC), and Constructivist Learning Environment Survey (CLES) have examined classroom structures, student interaction, and constructivist practices (Fraser, 2012). Similarly, belongingness instruments such as the Psychological Sense of School Membership (PSSM) have focused on students' feelings of acceptance and inclusion (Goodenow, 1993), while engagement measures have assessed behavioural, emotional, and cognitive participation (Fredricks et al., 2004). However, these instruments generally examine learning environment, belongingness, engagement, or deep learning as separate domains rather than as an integrated multidimensional construct. Furthermore, most validation studies have originated in Western educational contexts, while evidence from non-Western higher education settings remains limited. Therefore, systematic development and validation of BSDLE are needed to extend current measurement approaches.

Recent advances in educational measurement emphasise that complex multidimensional constructs require analytical approaches beyond classical reliability estimation. Confirmatory Factor Analysis (CFA) enables researchers to evaluate the theoretical structure of latent dimensions, whereas Multidimensional Item Response Theory (MIRT), particularly the Graded Response Model (GRM), provides detailed information regarding item discrimination, category threshold functioning, and measurement precision (Embretson & Reise, 2000; Reckase, 2009; Cai, 2021). Combining CFA and MIRT provides stronger evidence for construct validity because it evaluates both the construct's structural representation and item-level functioning.

Therefore, this study aims to develop and validate the BSDLE instrument using an integrated CFA and MIRT approach. Specifically, this study examines whether BSDLE is represented by a theoretically justified six-factor structure and whether the items demonstrate adequate psychometric properties. Through this psychometric investigation, this study contributes to educational measurement literature by providing a theoretically differentiated and empirically validated instrument for assessing supportive deep learning environments. The BSDLE instrument is expected to help higher education institutions diagnose environmental conditions that promote student belonging, engagement, and meaningful learning across diverse instructional contexts.

Methodology

This study aimed to develop and validate a measurement model of the Belonging-Supportive Deep Learning Environment (BSDLE) as an indicator of the effectiveness of learning environments in higher education. The study employed an instrument-development design with a quantitative cross-sectional design to ensure systematic construct development and validation. The research was conducted from March to September 2025.

Participants

This study included 825 students from five higher education institutions in Indonesia, comprising three state universities and two private universities. The sample was obtained from diverse

This is an open access article under CC-BY-SA license (https://creativecommons.org/licenses/by-sa/4.0/)

institutional contexts to ensure sufficient variability in students' learning environment experiences. The participating institutions were State University A (n = 145, 17.6%), State University B (n = 131, 15.9%), State University C (n = 162, 19.6%), Private University A (n = 205, 24.8%), and Private University B (n = 182, 22.1%).

Regarding educational level, the participants represented various levels of higher education. The sample consisted of 108 students from Diploma programs (D3), 404 undergraduate students (bachelor's degree/S1), 185 postgraduate students (master's degree/S2), and 128 doctoral students (Doctoral degree/S3). This distribution enabled the validation of the BSDLE instrument across different academic levels and learning experiences.

The sample size of 825 participants was considered adequate for conducting Confirmatory Factor Analysis (CFA) and Multidimensional Item Response Theory (MIRT), as both approaches require sufficiently large samples to achieve stable parameter estimation and reliable evaluation of multidimensional measurement models (Hair et al., 2019; Finch & French, 2022).

Instrument Development

The BSDLE instrument was developed using a construct-based approach that integrated learning environment theory, belongingness theory, and deep learning perspectives. The development procedure followed a modified version of the instrument development framework proposed by Moore and Benbasat (1991), consisting of five sequential stages: (1) literature review, (2) item generation, (3) construct validation, (4) pilot testing, and (5) instrument administration.

The initial item pool was generated based on theoretical dimensions of BSDLE, including student cohesiveness, involvement, investigation, task orientation, cooperation, and faculty–staff relations. These dimensions were derived from established learning environment frameworks and expanded to include relational belongingness perspectives.

To establish content validity, an expert review procedure was conducted involving six experts with different areas of expertise, consisting of two psychometric experts, two educational psychology experts, one higher education expert, and one language expert. The experts evaluated each item's conceptual relevance, clarity, linguistic appropriateness, and representativeness in relation to the BSDLE construct.

Following the expert review, a quantitative content validity assessment was conducted by 10 educational practitioners and academics who served as expert panellists. The panellists evaluated each item using a five-point rating scale based on three criteria: (1) item relevance to the BSDLE indicators, (2) item clarity in measuring the intended dimension, and (3) item importance for representing the construct.

Content validity was quantified using Aiken's V coefficient. Items were reviewed and revised based on expert recommendations and quantitative validity evidence. Items showing insufficient relevance, unclear wording, or inadequate representation of the construct were revised before further empirical testing. The content validity analysis demonstrated satisfactory evidence. For item relevance, Aiken's V values ranged from 0.500 to 0.950 (M = 0.840). For item clarity, values ranged from 0.525 to 0.950 (M = 0.837), while item importance values ranged from 0.575 to 0.950 (M = 0.840). These findings indicated that the developed items demonstrated acceptable content validity before proceeding to construct validation.

The detailed test blueprint of the BSDLE instrument is included in the Appendix as supplementary material. The blueprint provides a comprehensive overview of the construct dimensions, number of items assigned to each dimension, and examples of representative items. This

http://journal.uinjkt.ac.id/index.php/jp3i
This is an open access article under CC-BY-SA license (https://creativecommons.org/licenses/by-sa/4.0/)

supplementary information is provided to ensure methodological transparency and to demonstrate the systematic process through which theoretical constructs were translated into measurable indicators.

Data Collection

Data were collected using an online questionnaire distributed to students across the participating universities. Participation was voluntary, and respondents completed the questionnaire anonymously. Prior to analysis, data screening was conducted to ensure completeness and data quality. Only complete responses were included in the analysis.

Data Analysis

1. Preliminary Analysis

Initial analysis included descriptive statistics and item analysis to examine response distribution and item consistency. This step aimed to ensure that the data met the assumptions required for further analysis.

2. Confirmatory Factor Analysis (CFA)

Confirmatory Factor Analysis (CFA) was conducted using LISREL 8.80 to examine the hypothesised six-factor structure of the Belonging-Supportive Deep Learning Environment (BSDLE) instrument. Because all BSDLE items were measured on a five-point Likert scale, the ordinal nature of the observed variables was considered before CFA estimation.

The inter-item association matrix was estimated using polychoric correlations through PRELIS in LISREL 8.80. This approach was selected because polychoric correlations better reflect ordinal response data by estimating relationships among underlying continuous latent variables represented by observed categorical responses. Therefore, the analysis considered the ordinal nature of Likert-scale measurements rather than treating the responses as purely continuous variables.

The CFA model consisted of six correlated latent factors representing student cohesiveness, task orientation, investigation, involvement, cooperation, and faculty–staff relations. The adequacy of the measurement model was evaluated using multiple goodness-of-fit indices, including chi-square (χ^2), Root Mean Square Error of Approximation (RMSEA), Comparative Fit Index (CFI), Normed Fit Index (NFI), and Relative Fit Index (RFI).

The CFA results indicated that the hypothesised six-factor model demonstrated acceptable model fit. The obtained fit indices were $\chi^2 = 2966.51$ with df = 999, RMSEA = 0.049, CFI = 0.97, NFI = 0.95, and RFI = 0.95. These results indicate that the proposed multidimensional structure adequately represented the empirical data.

Item retention decisions were based on standardized factor loadings (≥ 0.50), statistical significance of parameter estimates (t-value ≥ 1.96), and theoretical consistency with the BSDLE framework. Model modifications were only considered when supported by modification indices and theoretical justification.

3. Multidimensional Item Response Theory (MIRT)

Following CFA, Multidimensional Item Response Theory (MIRT) analysis was conducted to evaluate item-level characteristics. The analysis employed the Graded Response Model (GRM), which is appropriate for ordinal Likert-scale data.

A multidimensional model consisting of six correlated latent dimensions was specified. The analysis estimated item discrimination (a) and threshold (b) parameters. The discrimination

This is an open access article under CC-BY-SA license (https://creativecommons.org/licenses/by-sa/4.0/)

parameter reflects an item's ability to differentiate respondents, while the threshold parameter represents category transition points on the Likert scale. In addition, the study evaluated item fit, Test Information Function (TIF), and marginal reliability to assess measurement precision across different levels of respondent ability (Reckase, 2023; Cai, 2021).

4. Integration of CFA and MIRT

This study adopted a two-stage validation approach by integrating Confirmatory Factor Analysis (CFA) and Multidimensional Item Response Theory (MIRT). CFA was first conducted to examine the construct-level structure of the BSDLE instrument, while MIRT was subsequently applied to investigate item-level functioning and measurement precision.

Following the confirmation of the six-factor structure through CFA, MIRT analysis was performed using the Graded Response Model (GRM), which is suitable for analyzing polytomous ordinal responses obtained from Likert-type scales. The MIRT calibration was conducted using IRTPRO Version 6.0 (64-bit).

The model was parameterised using the graded response formulation with discrimination (a) and threshold (b) parameters. The discrimination parameter indicates each item's ability to differentiate respondents across levels of the latent trait, whereas threshold parameters represent the locations at which respondents transition between adjacent response categories.

The multidimensional model was specified according to the six BSDLE dimensions. Item-level evaluation was conducted using discrimination parameters, threshold ordering, marginal reliability, and item-fit statistics. Item fit was examined using summed score-based diagnostic statistics and chi-square indices provided by the MIRT estimation procedure.

The estimation procedure and convergence criteria were based on the default estimation settings implemented in IRTPRO Version 6.0. The estimation process was considered converged when the iterative parameter estimation procedure met the software's predefined convergence criterion.

Results and Discussion

Confirmatory Factor Analysis

Before conducting the Confirmatory Factor Analysis (CFA), preliminary data screening evaluated the dataset's suitability for structural modelling. The analysis was conducted using LISREL 8.80 based on responses from 825 participants and 47 observed indicators representing the six dimensions of the Belonging-Supportive Deep Learning Environment (BSDLE) construct.

The dataset was examined for completeness, and no missing data were identified. The distributional characteristics of the observed variables were evaluated through skewness and kurtosis values. The results indicated that the observed variables had acceptable distributional properties, supporting CFA estimation.

Multicollinearity among indicators was also assessed by examining inter-item correlations. The correlation coefficients remained within acceptable ranges, indicating that the indicators did not demonstrate problematic multicollinearity.

The CFA model was specified as a six-factor correlated model, consisting of:

1. Student Cohesiveness (SC)
2. Task Orientation (TO)
3. Investigation (IN)

http://journal.uinjkt.ac.id/index.php/jp3i
This is an open access article under CC-BY-SA license (https://creativecommons.org/licenses/by-sa/4.0/)

4. Involvement (IV)
5. Cooperation (CO)
6. Faculty–Staff Relations (FR)

The six latent variables were allowed to correlate because the BSDLE dimensions represent related but conceptually distinct components of a supportive deep learning environment.

Table 1. The CFA results demonstrated acceptable model fit

Fit Index	Value
χ^2	2966.51
Df	999
p-value	< .001
RMSEA	0.049
CFI	0.97
NFI	0.95
RFI	0.95

Sources : Personal data.

These findings indicate that the hypothesised six-factor structure adequately represents the empirical data.

To address construct validity concerns, item-level CFA parameters are presented in Table 1. The table reports standardized factor loadings, standard errors, t-values, R^2 values, and item decisions. Most items demonstrated statistically significant factor loadings ($t > 1.96$), indicating that the indicators significantly contributed to their respective latent dimensions.

Table 2. Item-Level CFA Results of the BSDLE Instrument

Item	Dimension	Loading	SE	t-value	R^2	Decision
SC1	Student Cohesiveness	.59	.04	16.76	.35	Retained
SC2	Student Cohesiveness	.64	.03	18.59	.41	Retained
SC3	Student Cohesiveness	.55	.04	15.28	.30	Retained
SC4	Student Cohesiveness	.51	.04	13.90	.26	Retained
SC5	Student Cohesiveness	.54	.04	15.00	.29	Retained
TO6	Task Orientation	.58	.03	16.97	.34	Retained
TO7	Task Orientation	.59	.03	17.26	.35	Retained
TO8	Task Orientation	.62	.03	18.30	.38	Retained
TO9	Task Orientation	.55	.03	16.06	.30	Retained
TO10	Task Orientation	.58	.03	17.04	.34	Retained
TO11	Task Orientation	.54	.03	15.62	.29	Retained
TO12	Task Orientation	.58	.03	16.80	.34	Retained
TO13	Task Orientation	.49	.03	13.95	.24	Retained

This is an open access article under CC-BY-SA license (https://creativecommons.org/licenses/by-sa/4.0/)

Item	Dimension	Loading	SE	t-value	R^2	Decision
IN14	Investigation	.48	.03	13.88	.23	Retained
IN15	Investigation	.61	.03	18.23	.37	Retained
IN16	Investigation	.54	.03	15.83	.29	Retained
IN17	Investigation	.49	.03	14.15	.24	Retained
IN18	Investigation	.56	.03	16.31	.31	Retained
IN19	Investigation	.56	.03	16.64	.31	Retained
IN20	Investigation	.63	.03	18.91	.40	Retained
IN21	Investigation	.40	.04	11.33	.16	Retained*
IN22	Investigation	.53	.03	15.42	.28	Retained
IN23	Investigation	.57	.03	16.91	.32	Retained
IN24	Investigation	.58	.03	17.28	.34	Retained
IN25	Investigation	.58	.03	17.15	.34	Retained
IN26	Investigation	.52	.03	15.04	.27	Retained
IV27	Involvement	.61	.03	18.20	.37	Retained
IV28	Involvement	.59	.03	17.16	.35	Retained
IV29	Involvement	.49	.03	14.03	.24	Retained
IV30	Involvement	.60	.03	18.04	.36	Retained
IV31	Involvement	.54	.03	15.74	.29	Retained
IV32	Involvement	.59	.03	17.28	.35	Retained
IV33	Involvement	.57	.03	16.79	.32	Retained
IV34	Involvement	.50	.03	14.48	.25	Retained
CO35	Cooperation	.49	.04	13.21	.24	Retained
CO36	Cooperation	.57	.04	15.93	.32	Retained
CO37	Cooperation	.58	.04	16.48	.34	Retained
CO38	Cooperation	.55	.04	15.38	.30	Retained
CO39	Cooperation	.63	.03	17.99	.40	Retained
FR40	Faculty–Staff Relations	.50	.03	14.24	.25	Retained
FR41	Faculty–Staff Relations	.58	.03	16.99	.34	Retained
FR42	Faculty–Staff Relations	.62	.03	18.49	.38	Retained
FR43	Faculty–Staff Relations	.61	.03	18.10	.37	Retained
FR44	Faculty–Staff Relations	.55	.03	15.84	.30	Retained
FR45	Faculty–Staff Relations	.57	.03	16.67	.32	Retained
FR46	Faculty–Staff Relations	.57	.03	16.69	.32	Retained

http://journal.uinjkt.ac.id/index.php/jp3i
This is an open access article under CC-BY-SA license (https://creativecommons.org/licenses/by-sa/4.0/)

Item	Dimension	Loading	SE	t-value	R^2	Decision
FR47	Faculty–Staff Relations	.25	.04	6.73	.06	Retained*

Sources : Personal data.

Although several items demonstrated relatively lower standardized factor loadings, they were retained based on theoretical considerations and dimensional representation. Specifically, Item FR47 was retained because it represents an important aspect of faculty–staff relations, which is theoretically central to the BSDLE framework. Removing the item would reduce the conceptual coverage of the dimension. Therefore, retention decisions were based not only on statistical criteria but also on the construct's theoretical completeness.

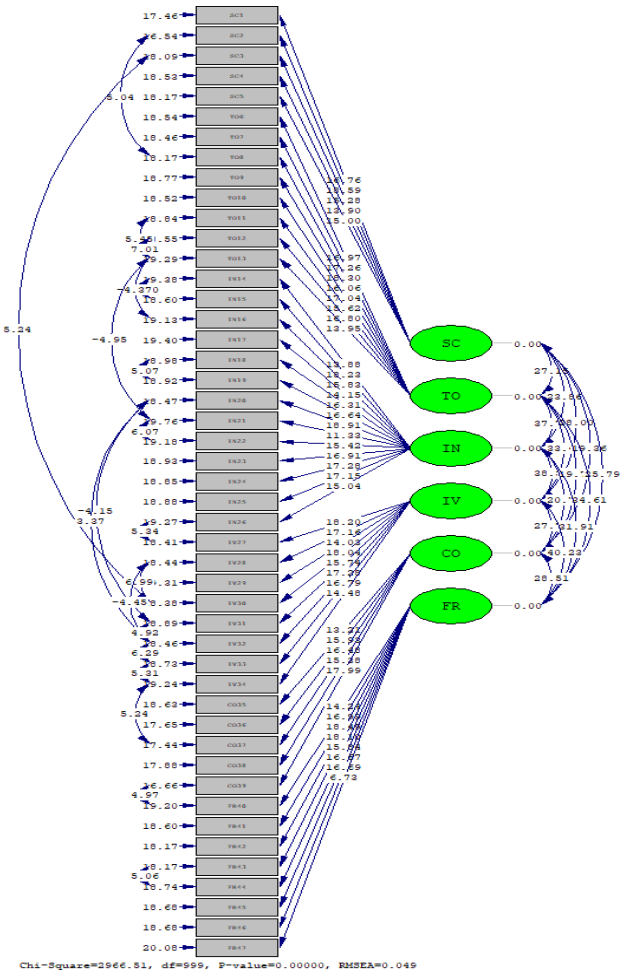
Table 3. Correlated Residual Modifications in the CFA Model

Item Pair	Modification Index	Interpretation
IN14 – TO6	35.64	Shared conceptual similarity between inquiry and task-oriented learning processes
SC2 – IN14	5.21	Related aspects of peer interaction and inquiry activities
IN24 – SC1	19.46	Similar social-learning experience
IV30 – IN16	11.85	Related active learning behavior

Sources : Personal data.

To further improve model fit, model modification procedures were examined using modification indices. We identified several correlated residuals between items and incorporated them into the model. These modifications were applied selectively and were theoretically justified, particularly for items within the same dimension that shared similar wording or conceptual content. Including correlated error terms improved model fit without undermining the model's theoretical coherence. The final CFA model is presented in Figure 1, which illustrates the standardized factor loadings and the correlations among the latent variables in the six-dimensional BSDLE construct.

This is an open access article under CC-BY-SA license (https://creativecommons.org/licenses/by-sa/4.0/)



Sources : Personal data.

Figure 1. Standardized Confirmatory Factor Analysis (CFA) Model of the Six-Dimensional BSDLE Construct

These findings support the conceptualisation of BSDLE as a multidimensional construct comprising six key dimensions. The results are consistent with recent studies emphasising the role of supportive and interactive learning environments in fostering student engagement and belonging (Bond, 2020; Pedler et al., 2022). Therefore, the BSDLE instrument demonstrates adequate construct validity for assessing learning environments that promote deep learning and belongingness.

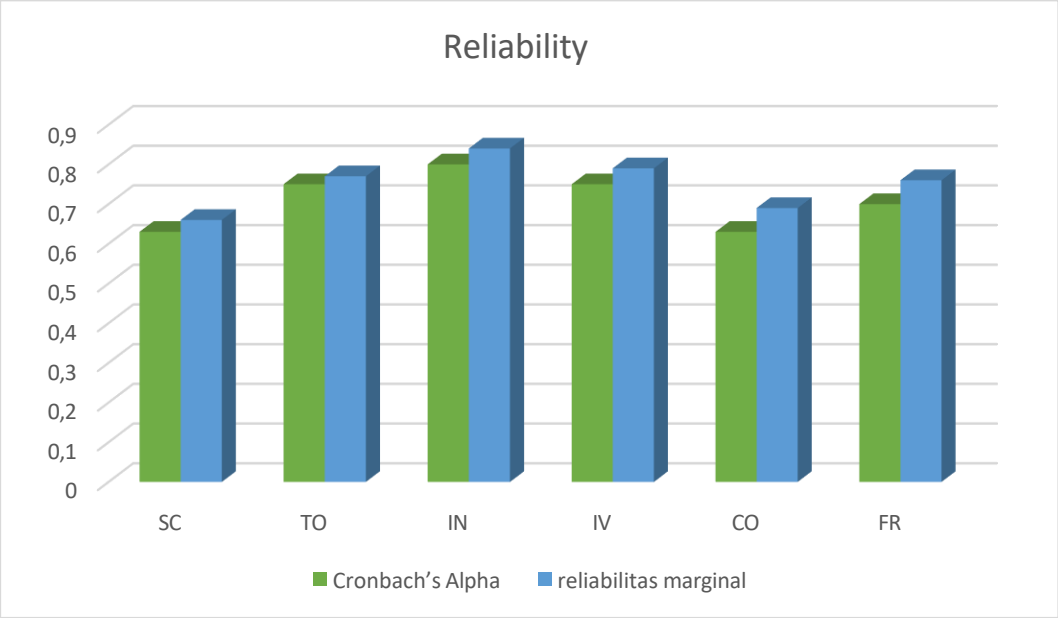
Multidimentional Item Respon Theory

An initial analysis was conducted on 47 items based on responses from 825 participants using a 5-point Likert scale. The dataset was complete, with no missing values, indicating adequate data quality for further analysis. Descriptive statistics showed that mean item scores ranged from 3.9 to 4.1, with standard deviations of approximately 0.8, indicating sufficient variability in responses. Exploratory analysis using Principal Component Analysis (PCA) identified multiple components; however, the authors retained a six-factor solution based on the predefined theoretical structure. Initial reliability analysis using Cronbach's Alpha yielded a coefficient of 0.91, indicating high internal consistency.

Item–rest correlation analysis showed that most items ranged from 0.22 to 0.39, while one item (FR8) had a lower value of 0.14. Factor loadings from the exploratory stage exceeded 0.40 for all items.

http://journal.uinjkt.ac.id/index.php/jp3i
This is an open access article under CC-BY-SA license (https://creativecommons.org/licenses/by-sa/4.0/)

Subscale scores, used as proxy theta values, showed approximately normal distributions, with standard deviations ranging from 0.45 to 0.59. MIRT analysis was conducted using a six-dimensional model with the Graded Response Model (GRM). The analysis estimated discrimination (a) and threshold (b) parameters for each item. The discrimination parameters varied across items, with most values ranging from approximately 0.85 to 2.10. Several items, particularly within the IN dimension, showed higher discrimination values (above 1.0), while one item (Item 47) showed a lower discrimination value below 0.50. The threshold parameters were found to be ordered across response categories. Marginal reliability analysis indicated that the IN and IV dimensions had reliability coefficients above 0.70, while other dimensions showed relatively lower values.

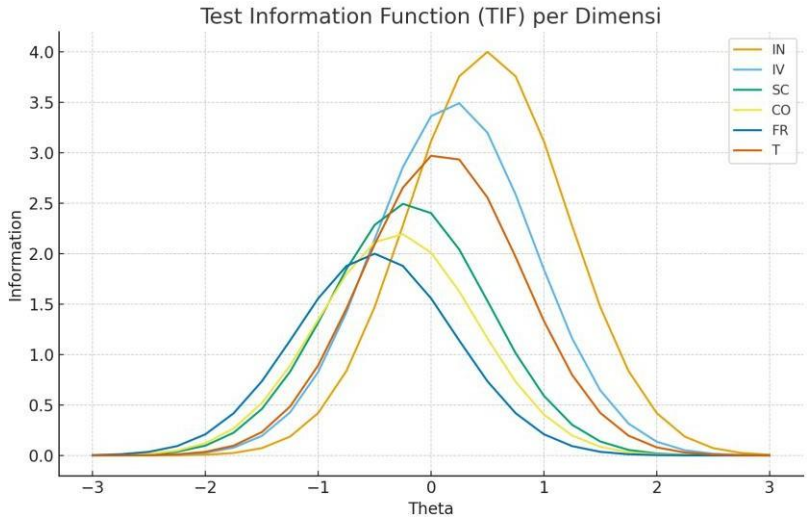


Sources : Personal data.

Figure 2. Differences Between Cronbach's Alpha and Marginal Reliability

The subscale reliability analysis compared Cronbach's Alpha and marginal reliability across the six dimensions (SC, T, IN, IV, CO, FR). The results showed that marginal reliability values were consistently higher than Cronbach's Alpha. For example, the IN dimension had a marginal reliability of approximately 0.84 compared with an Alpha of 0.80, while the IV dimension had a marginal reliability of approximately 0.79 compared with an Alpha of 0.75. The Test Information Function (TIF) analysis indicated different information patterns across dimensions. The IN dimension showed peak information from $\theta = 0$ to $\theta = +1$, whereas the FR dimension provided higher information below $\theta = 0$.

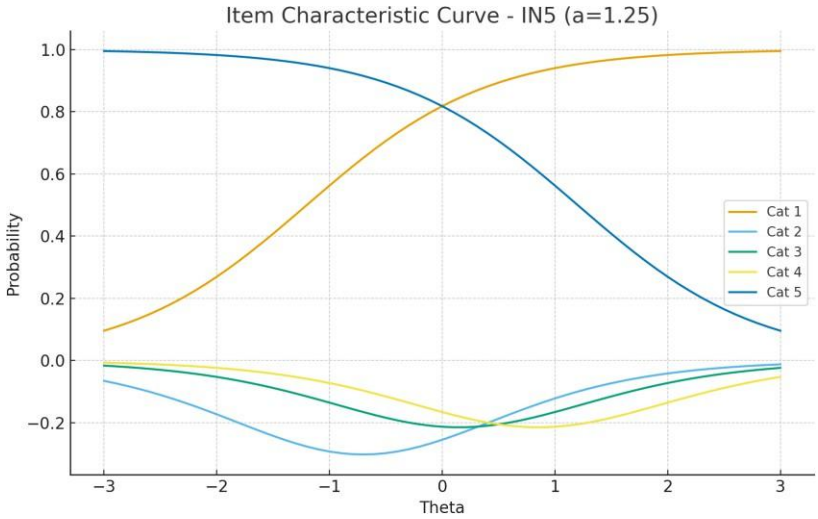
This is an open access article under CC-BY-SA license (https://creativecommons.org/licenses/by-sa/4.0/)



Sources : Personal data.

Figure 3. Test Information Function (TIF) Plot

The Test Information Function (TIF) analysis showed that each dimension provided peak information at different ranges of ability (θ). The IN and IV dimensions showed the highest information around θ = 0 to θ = +1, while the SC, CO, and FR dimensions showed relatively lower levels of information. Item fit analysis identified several items with misfit, particularly within the FR and CO dimensions. The MIRT analysis indicated variation in measurement performance across dimensions. The IN and IV dimensions showed relatively stronger measurement performance than the SC, CO, and FR dimensions. Moderate correlations were observed between several dimensions, including FR and CO. In addition to item parameter estimation, the MIRT model produced latent ability (θ) scores for each respondent across the six dimensions.



Sources : Personal data

Figure 4. Item Characteristic Curve (ICC)

This is an open access article under CC-BY-SA license (https://creativecommons.org/licenses/by-sa/4.0/)

The Item Characteristic Curve (ICC) for item IN5 showed a discrimination parameter of a = 1.25. The category response curves displayed an orderly pattern across the latent ability (θ) continuum, with transitions from lower to higher response categories. The MIRT analysis using the Graded Response Model (GRM) indicated that most items had discrimination parameters ranging from approximately 0.85 to 2.10. One item (Item 47) showed a lower discrimination value below 0.50. Threshold parameters (b) were ordered across response categories. The Test Information Function (TIF) analysis showed that the instrument provided the highest information within the ability range of θ between -1 and +2. The IN and IV dimensions showed higher information than the SC, CO, and FR dimensions. Item fit analysis identified several misfitting items, particularly within the FR and CO dimensions. Moderate correlations were observed among several dimensions, including between the FR and CO dimensions. Marginal reliability values ranged from approximately 0.78 to 0.89 across dimensions.

Discussion

T The findings of this study provide empirical support for the multidimensional conceptualisation of the Belonging-Supportive Deep Learning Environment (BSDLE). The Confirmatory Factor Analysis (CFA) results indicated that the hypothesised six-factor model demonstrated acceptable model fit, suggesting that BSDLE represents a complex educational environment in which cognitive learning processes and psychosocial support mechanisms operate simultaneously. This finding is consistent with contemporary perspectives in higher education that emphasise the interaction between academic engagement, social relationships, and students' psychological experiences in shaping meaningful learning environments (Bond, 2020; Pedler et al., 2022).

The MIRT analysis further extended the CFA findings by providing evidence regarding item-level functioning and measurement precision. Most items showed acceptable discrimination parameters, indicating that the instrument could differentiate students across different levels of perceived belonging-supportive deep learning environments. The stronger discrimination and information functions in the Investigation (IN) and Involvement (IV) dimensions suggest these dimensions may represent more observable, behaviorally expressed aspects of students' learning experiences.

The stronger performance of the IN dimension may be explained by the nature of inquiry-based learning processes. Investigation-related behaviours, such as exploring ideas, seeking explanations, and engaging in problem-solving activities, are closely associated with deep learning processes that are directly experienced by students during academic activities. Similarly, the IV dimension reflects students' active participation and engagement, which are often more visible and easier for students to evaluate based on their own learning experiences. Therefore, items within these dimensions may demonstrate stronger variability among respondents, resulting in higher discrimination and information values.

In contrast, the relatively lower information values observed in the Student Cohesiveness (SC), Cooperation (CO), and Faculty–Staff Relations (FR) dimensions require further consideration. These dimensions may involve more contextual and relational aspects of the learning environment that are not always directly observable or consistently experienced by students. For example, student cohesiveness and cooperation may depend heavily on classroom dynamics, group composition, instructional strategies, and institutional culture. Consequently, students from different learning contexts may interpret similar items differently, reducing response variability and limiting item discrimination.

The lower information values may also indicate that some items within these dimensions are relatively general in wording. Items that describe broad social support or interpersonal relationships may be endorsed similarly by many respondents, limiting the ability to distinguish between students with different levels of the latent trait. In psychometric terms, such items may function as less informative

This is an open access article under CC-BY-SA license (https://creativecommons.org/licenses/by-sa/4.0/)

indicators because they do not adequately cover a wide range of the latent continuum. Future refinement should therefore consider developing more context-specific items that capture variations in students' experiences of belonging, peer cooperation, and institutional support.

The identification of misfitting items, particularly within the Faculty–Staff Relations (FR) and Cooperation (CO) dimensions, should not be interpreted solely as statistical problems. Item misfit may reflect deeper conceptual issues, including ambiguity in item interpretation, overlap between dimensions, or differences in how students perceive relational experiences within their educational contexts. For example, faculty–staff relations involve multiple forms of interaction, including academic guidance, emotional support, administrative assistance, and institutional responsiveness. These aspects may not always represent a single homogeneous experience from the students' perspective. Similarly, cooperation may overlap conceptually with involvement and cohesiveness because collaborative learning involves both individual participation and social interaction.

Therefore, misfitting items provide valuable information for future instrument refinement. Rather than removing problematic items based solely on statistical criteria, future studies should examine whether item wording adequately represents the intended construct and whether additional contextual indicators are needed to capture the complexity of relational learning environments.

The variation in information functions across dimensions also highlights the importance of considering measurement precision across the latent trait continuum. The IN and IV dimensions provided higher information, particularly within the θ range of approximately -1 to +2, indicating stronger precision for students with moderate to high levels of the measured construct. Conversely, SC, CO, and FR showed lower information, suggesting that additional items or improved item targeting may be needed to achieve more balanced measurement precision across dimensions (Reckase, 2009; Finch & French, 2022).

The moderate to high correlations among BSDLE dimensions indicate that the six dimensions are empirically related while remaining conceptually distinct. These correlations are theoretically expected because learning environments are integrated systems in which academic activities, social relationships, and institutional support influence each other. However, these relatively high correlations also raise an important psychometric question about whether a broader general factor may underlie the six dimensions. Although the current six-factor model demonstrated acceptable fit, future studies should consider comparing alternative models, such as a higher-order factor model or bifactor model, to examine whether BSDLE consists of distinct dimensions or whether a general supportive learning environment factor explains substantial shared variance among dimensions.

The marginal reliability estimates obtained from MIRT analysis provide additional evidence regarding measurement precision beyond traditional internal consistency estimates. However, differences in reliability and information across dimensions suggest that not all subscales currently demonstrate equivalent measurement strength. Therefore, although the BSDLE instrument shows potential for research and preliminary diagnostic purposes, interpret claims about diagnostic assessment cautiously, particularly for dimensions with lower information and reliability. Further refinement and validation are necessary before the instrument can be applied for high-stakes individual-level diagnosis.

Several limitations should be acknowledged. First, the study employed purposive sampling, with participants recruited from selected higher education institutions. Although the sample size was adequate for CFA and MIRT analyses, the sampling approach may limit the generalizability of findings to broader higher education populations. Second, EFA, CFA, and MIRT analyses were conducted using the same dataset. Although the large sample size and multiple validation approaches provided initial evidence supporting the stability of the BSDLE structure, future research should use independent-sample validation or split-sample procedures to provide stronger cross-validation evidence. Third, this study did not examine measurement invariance across relevant groups, such as gender, academic level, institution

http://journal.uinjkt.ac.id/index.php/jp3i
This is an open access article under CC-BY-SA license (https://creativecommons.org/licenses/by-sa/4.0/)

type, or disciplinary background. Future studies should investigate whether BSDLE items function equivalently across different student populations. Fourth, Differential Item Functioning (DIF) analysis was not conducted. DIF analysis is necessary to determine whether certain items favour particular groups due to differences in interpretation rather than true differences in the latent construct. Fifth, the current study did not examine external validation using related constructs, such as student engagement, academic belongingness, psychological well-being, and learning outcomes. Future research should incorporate criterion-related and nomological validation to further establish the usefulness of BSDLE.

Finally, although the current findings suggest that BSDLE has potential for advanced measurement applications, including computerised adaptive testing (CAT), such implementation requires additional steps. Future research should first establish a sufficiently calibrated item bank, conduct simulation studies, and evaluate adaptive testing performance before considering CAT applications feasible.

Conclusion

This study aimed to develop and validate a multidimensional instrument to assess the Belonging-Supportive Deep Learning Environment (BSDLE). The findings indicate that the proposed six-factor structure is supported by the data. The Confirmatory Factor Analysis (CFA) results showed that the model achieved an acceptable fit, indicating that the six dimensions adequately represent the construct. The Multidimensional Item Response Theory (MIRT) analysis using the Graded Response Model (GRM) showed that most items functioned well. The discrimination parameters were generally in the moderate to high range, indicating that the items can differentiate respondents with different ability levels. The threshold parameters were ordered, suggesting respondents used the response categories consistently.

The Test Information Function (TIF) results showed that the instrument provides higher measurement precision within the ability range of θ between -1 and +2. The IN and IV dimensions showed stronger measurement performance, while the SC, CO, and FR dimensions showed lower levels of information. Some items, particularly in the SC, CO, and FR dimensions, showed lower discrimination or misfit and therefore require further evaluation. However, most items performed adequately, and the overall structure of the instrument is supported.

In summary, the BSDLE instrument demonstrates acceptable validity and reliability as a multidimensional measure. The instrument can be used to assess students' perceptions of learning environments, although further refinement is recommended to improve measurement consistency across all dimensions. Future research is needed to refine the items and examine the instrument's performance in different contexts.

Acknowledgements

This research was funded by the Research Grant Program of the Directorate of DPPM Kemdiktisaintek Indonesia for the Fiscal Year 2025 under Contract Number 083/C3/DT.05.00/PL/2025. The authors thank Universitas Negeri Jakarta, Universitas Islam Negeri Syarif Hidayatullah Jakarta, Universitas Teknologi Nusantara, Universitas Pamulang, and Universitas Terbuka for supporting access to student participants. Special appreciation is also extended to the experts who contributed to the instrument validation process.

This is an open access article under CC-BY-SA license (https://creativecommons.org/licenses/by-sa/4.0/)

References

Bergmann, J., & Sams, A. (2012). *Flip your classroom: Reach every student in every class every day*. International Society for Technology in Education.

Bond, M. (2020). Facilitating student engagement through educational technology: Towards a conceptual framework. *Journal of Interactive Media in Education, 2020*(1), Article 11. https://doi.org/10.5334/jime.528

Boone, W. J., Staver, J. R., & Yale, M. S. (2014). *Rasch analysis in the human sciences*. Springer. https://doi.org/10.1007/978-94-007-6857-4

Cai, L. (2021). Multidimensional item response theory. In P. Irwing, T. Booth, & D. J. Hughes (Eds.), *The Wiley handbook of psychometric testing: A multidisciplinary reference on survey, scale and test development* (pp. 1–35). Wiley. https://doi.org/10.1002/9781119716723.ch6

Chionh, Y. H., & Fraser, B. J. (2009). Classroom environment, achievement, attitudes and self-esteem in geography and mathematics in Singapore. *International Research in Geographical and Environmental Education, 18*(1), 29–44. https://doi.org/10.1080/10382040802591530

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Lawrence Erlbaum Associates.

Finch, W. H., & French, B. F. (2022). *Educational and psychological measurement* (2nd ed.). Routledge.

Fraser, B. J. (2012). Classroom learning environments: Retrospect, context and prospect. In B. J. Fraser, K. Tobin, & C. J. McRobbie (Eds.), *Second international handbook of science education* (pp. 1191–1239). Springer. https://doi.org/10.1007/978-1-4020-9041-7_79

Fraser, B. J., Fisher, D. L., & McRobbie, C. J. (1996). Development, validation, and use of personal and class forms of a new classroom environment instrument. *Paper presented at the Annual Meeting of the American Educational Research Association*, New York, NY.

Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research, 74*(1), 59–109. https://doi.org/10.3102/00346543074001059

Freeman, T. M., Anderman, L. H., & Jensen, J. M. (2007). Sense of belonging in college freshmen at the classroom and campus levels. *The Journal of Experimental Education, 75*(3), 203–220. https://doi.org/10.3200/JEXE.75.3.203-220

Goodenow, C. (1993). The psychological sense of school membership among adolescents: Scale development and educational correlates. *Psychology in the Schools, 30*(1), 79–90. https://doi.org/10.1002/1520-6807(199301)30:1%3C79::AID-PITS2310300113%3E3.0.CO;2-X

Gopalan, M., & Brady, S. T. (2020). College students' sense of belonging: A national perspective. *Educational Researcher, 49*(2), 134–137. https://doi.org/10.3102/0013189X19897622

Green, S. B., Yang, Y., & McNeish, D. (2021). Coefficient alpha and omega reliability coefficients: A theoretical and empirical comparison. *Psychological Methods, 26*(6), 714–732. https://doi.org/10.1037/met0000343

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). *Multivariate data analysis* (8th ed.). Cengage.

Hambleton, R. K., & Swaminathan, H. (1985). *Item response theory: Principles and applications*. Kluwer Academic Publishers.

Lo, C. K., & Hew, K. F. (2020). A comparison of flipped learning with traditional learning in higher education: Students' perspectives. *Educational Research Review, 30*, Article 100314. https://doi.org/10.1016/j.edurev.2020.100314

Moore, G. C., & Benbasat, I. (1991). Development of an instrument to measure the perceptions of adopting an information technology innovation. *Information Systems Research, 2*(3), 192–222. https://doi.org/10.1287/isre.2.3.192

Pedler, M., Willis, R., & Nieuwoudt, J. (2022). A sense of belonging at university: Student retention, motivation and enjoyment. *Journal of Further and Higher Education, 46*(3), 397–408. https://doi.org/10.1080/0309877X.2021.1955844

Freeman, T. M., Anderman, L. H., & Jensen, J. M. (2007). Sense of belonging in college freshmen at the classroom and campus levels. *The Journal of Experimental Education, 75*(3), 203–220. https://doi.org/10.3200/JEXE.75.3.203-220

Goodenow, C. (1993). The psychological sense of school membership among adolescents: Scale development and educational correlates. *Psychology in the Schools, 30*(1), 79–90. https://doi.org/10.1002/1520-6807(199301)30:1%3C79::AID-PITS2310300113%3E3.0.CO;2-X

This is an open access article under CC-BY-SA license (https://creativecommons.org/licenses/by-sa/4.0/)

Gopalan, M., & Brady, S. T. (2020). College students' sense of belonging: A national perspective. *Educational Researcher, 49*(2), 134–137. https://doi.org/10.3102/0013189X19897622

Green, S. B., Yang, Y., & McNeish, D. (2021). Coefficient alpha and omega reliability coefficients: A theoretical and empirical comparison. *Psychological Methods, 26*(6), 714–732. https://doi.org/10.1037/met0000343

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). *Multivariate data analysis* (8th ed.). Cengage.

Hambleton, R. K., & Swaminathan, H. (1985). *Item response theory: Principles and applications*. Kluwer Academic Publishers.

Lo, C. K., & Hew, K. F. (2020). A comparison of flipped learning with traditional learning in higher education: Students' perspectives. *Educational Research Review, 30*, Article 100314. https://doi.org/10.1016/j.edurev.2020.100314

Moore, G. C., & Benbasat, I. (1991). Development of an instrument to measure the perceptions of adopting an information technology innovation. *Information Systems Research, 2*(3), 192–222. https://doi.org/10.1287/isre.2.3.192

Pedler, M., Willis, R., & Nieuwoudt, J. (2022). A sense of belonging at university: Student retention, motivation and enjoyment. *Journal of Further and Higher Education, 46*(3), 397–408. https://doi.org/10.1080/0309877X.2021.1955844

This is an open access article under CC-BY-SA license (https://creativecommons.org/licenses/by-sa/4.0/)

Appendix: Test Blueprint

No	Aspect	Dimension	Question No	Sample Item
1	Student cohesiveness	Students feel accepted and valued by their peers.	1-5	*Saya nyaman saat berinteraksi dengan teman-teman di lingkungan universitas.*
2	Task Orientation	Students have a clear understanding of the objectives of each learning task.	6-9	*Saya memahami instruksi tugas yang diberikan dosen jelas dan terarah.*
		Students manage their time and complete assignments effectively.	10-13	*Saya membuat jadwal pribadi untuk mengerjakan tugas kuliah secara teratur.*
3	Investigation	Students ask questions and independently seek answers during the learning process.	14-17	*Saya bertanya tentang pembelajaran yang belum dimpahami.*
		Students actively seek additional information related to the topics they are studying.	18-22	*Saya membaca buku atau jurnal yang relevan di luar materi yang diberikan dosen.*
		Students conduct independent inquiries and research to deepen their understanding.	23-26	*Saya membuat rangkuman dari materi perkuliahan.*
4	Involvement	Students actively participate in classroom discussions.	27-30	*Saya senang mengemukakan pendapat dalam diskusi kelas.*
		Students participate in personal development activities organized by the university	31-34	*Saya mengikuti seminar atau pelatihan yang diselenggarakan oleh Universitas.*
5	Cooperation	Students share information and resources with fellow group members.	35-39	*Saya memberikan materi kepada anggota kelompok yang membutuhkan.*

http://journal.uinjkt.ac.id/index.php/jp3i

This is an open access article under CC-BY-SA license (https://creativecommons.org/licenses/by-sa/4.0/)

6	Faculty and staff relations	Students maintain positive relationships with lecturers during the learning process.	40-43	*Saya menyampaikan pendapat saya kepada dosen tanpa rasa takut.*
		Students feel heard and supported by lecturers and university staff.	44-47	*Saya mendapatkan tindak lanjut dari staf kampus setelah menyampaikan keluhan*

This is an open access article under CC-BY-SA license (https://creativecommons.org/licenses/by-sa/4.0/)