JOURNAL OF CULTURAL ANALYSIS AND SOCIAL CHANGE

Journal of Cultural Analysis and Social Change,
2025, 10(4), 2902-2911
ISSN: 2589-1316

Development and Validation of Cultural Intelligence, Cultural Diversity Climate, and Intention to Implement Ethno-STEM Scales for Indonesian Novice Science Teachers: An Exploratory Factor Analysis Approach

Henie Poerwandar Asmaningrum[1*], Renuka V Sathasivam[2], Edy Hafizan Mohd Shahali[3]

[1] *Department of Mathematics and Science Education, Universiti Malaya, Kuala Lumpur, 50603, MALAYSIA, Email: s2138457@siswa.um.edu.my*
[2] *Department of Mathematics and Science Education, Universiti Malaya, Kuala Lumpur, 50603, MALAYSIA, Email: renukasivam@um.edu.my*
[3] *Department of Mathematics and Science Education, Universiti Malaya, Kuala Lumpur, 50603, MALAYSIA, Email: edyhafizan@um.edu.my*

*Corresponding Author: s2138457@siswa.um.edu.my

Citation: Asmaningrum, H. P., Sathasivam, R. V., & Shahali, E. H. M. (2025). Development and Validation of Cultural Intelligence, Cultural Diversity Climate, and Intention to Implement Ethno-STEM Scales for Indonesian Novice Science Teachers: An Exploratory Factor Analysis Approach, *Journal of Cultural Analysis and Social Change*, 10(4), 2902-2911. https://doi.org/10.64753/jcasc.v10i4.3360

Published: December 16, 2025

ABSTRACT

This pilot study developed and validated a three-scale measurement model investigating predictors of Indonesian novice science teachers' intention to implement Ethno-STEM approaches. The Cultural Intelligence Scale, Cultural Diversity Climate Scale, and newly developed Intention to Implement Ethno-STEM Scale underwent systematic adaptation through forward-backward translation, cultural contextualization, and two-phase expert validation. The adapted instruments were empirically tested with 105 novice science teachers using Exploratory Factor Analysis. Results demonstrated excellent sampling adequacy (KMO: 0.910-0.922) and significant correlations suitable for factor analysis. Seven items (14.9%) were eliminated due to low communality values, yielding validated scales: Cultural Intelligence (14 items, α=0.889), Cultural Diversity Climate (12 items, α=0.912), and Intention to Implement Ethno-STEM (14 items, α=0.835). All instruments maintained original theoretical structures with acceptable to excellent reliability. Critical findings revealed discrepancies between expert validation and empirical performance, particularly with reverse-coded items in Indonesian cultural contexts. The validated instruments provide Indonesian educational researchers with psychometrically sound tools for investigating culturally responsive STEM education, establishing methodological foundation for examining how cultural intelligence and diversity climate predict teachers' implementation intentions.

Keywords: Cultural Intelligence, Cultural Diversity Climate, Intention to Implement, Ethno-STEM, Instrument Validation.

INTRODUCTION

The 21st century has brought renewed attention to culturally responsive pedagogy, especially in culturally diverse nations like Indonesia with over 300 ethnic groups and 700 languages. Innovative frameworks like Ethno-STEM bridge traditional cultural practices with contemporary STEM concepts (Izzah et al., 2023; Rizki et al., 2022). While the Indonesian government has introduced supportive policies, translating these initiatives into classroom practice remains challenging (Sukmayadi & Yahya, 2020). Teachers stand at the heart of this transformation, with their intention to embrace culturally responsive approaches largely determining reform success. Understanding what shapes teachers' intentions requires examining both individual cultural competencies

and organizational climate factors. Despite its theoretical importance, few studies have systematically examined predictors of teachers' intention to implement Ethno-STEM approaches in diverse educational contexts.

This study proposes a conceptual framework positioning teachers' intention to implement Ethno-STEM approaches as the outcome variable (Y), predicted by two key antecedents: cultural intelligence (X_1) and cultural diversity climate (X_2). This predictor-outcome model is grounded in the Theory of Planned Behavior (Ajzen, 1991), which suggests behavioral intentions are determined by attitudes, subjective norms, and perceived behavioral control. Cultural intelligence represents perceived behavioral control—teachers' confidence in their capability to function effectively in culturally diverse settings. Cultural diversity climate represents the environmental condition that either facilitates or constrains implementation willingness. The rationale for examining these predictors is both theoretical and empirical: intention formation requires both capability beliefs ("I can do this") and environmental support ("The context enables me to do this").

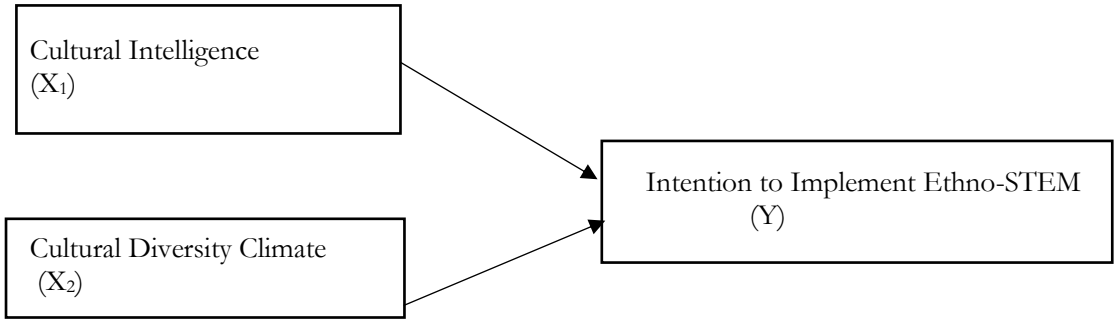

Figure 1. Conceptual Framework: Cultural Intelligence and Cultural Diversity Climate as Predictors of Intention to Implement Ethno-STEM

Cultural intelligence (CQ) is conceptualized as "an individual's capability to function and manage effectively in culturally diverse settings" (Earley & Ang, 2003), incorporating four dimensions: metacognitive CQ (conscious cognitive processes for acquiring cultural knowledge), cognitive CQ (knowledge of cultural norms), motivational CQ (directing attention toward learning about diversity), and behavioral CQ (adapting behaviors appropriately). The Cultural Intelligence Scale (Ang et al., 2007) demonstrated strong psychometric properties with Cronbach's alpha ranging from 0.70 to 0.87. Educational research shows teachers with higher CQ are more effective in multicultural classrooms (Dallman, 2020), positively affect student engagement (Ng et al., 2024), and demonstrate less pedagogical anxiety in Indonesian contexts (Alifuddin & Widodo, 2022). Theoretically, CQ represents perceived behavioral control that directly influences intention. Teachers with high CQ possess strategic thinking, cultural knowledge, intrinsic motivation, and behavioral repertoires that enhance confidence in implementing Ethno-STEM approaches. However, limited research has examined whether CQ specifically predicts teachers' intention to implement culturally integrated pedagogical approaches.

Cultural diversity climate is defined as "employees' shared perceptions that organization's policies, practices, and procedures communicate the extent to which fostering and maintaining diversity has been prioritized" (Mor Barak et al., 1998). The framework identifies four dimensions: organizational fairness, organizational inclusion, personal diversity value, and personal comfort with diversity. The Diversity Climate Perceptions Scale demonstrated overall reliability of α=0.83 and has been validated across organizational and educational contexts (McKay et al., 2009; Richards et al., 2007). Research reveals positive diversity climate associates with improved intergroup relations, enhanced learning outcomes (Hurtado et al., 2012), and teachers' implementation of inclusive practices (Sincer et al., 2021). The theoretical connection between diversity climate and intention draws from Social Cognitive Theory (Bandura, 1986), which emphasizes environmental influences that facilitate or constrain behaviors. When schools communicate that diversity is valued, teachers perceive greater organizational support, reducing anticipated barriers and strengthening implementation intentions. Despite theoretical rationale, limited research has examined whether school diversity climate specifically predicts teachers' intention to implement Ethno-STEM approaches in Indonesian contexts.

Behavioral intention occupies a central position in predicting behavior. According to the Theory of Planned Behavior, intentions capture motivational factors influencing behavior, representing how much effort individuals plan to exert (Ajzen, 1991). Intention serves as the most immediate antecedent of actual behavior, accounting for approximately 28% of variance in behavior (Sheeran et al., 1999). For implementing Ethno-STEM, teachers' intentions indicate their commitment to integrating indigenous knowledge with STEM content and willingness to invest effort despite obstacles. Implementation intention encompasses behavioral intention (willingness to

implement), planning intention (commitment to preparing materials), and persistence intention (determination to continue despite challenges). Focusing on intention rather than behavior is justified for novice teachers who may not yet have full curriculum autonomy, making intention a critical intervention target and forward-looking measure of readiness. Despite its importance, no validated instruments exist specifically measuring teachers' intention to implement Ethno-STEM approaches.

Despite growing recognition of culturally responsive STEM education's importance, several critical methodological gaps persist. First, there exists a significant absence of validated instruments measuring the complete predictor-outcome model for Ethno-STEM implementation in Indonesian contexts. While the Cultural Intelligence Scale (Ang et al., 2007) and Diversity Climate Perceptions Scale (Mor Barak et al., 1998) exist internationally, their psychometric properties have not been rigorously examined in Indonesian educational settings. More critically, no validated measure exists for intention to implement Ethno-STEM. Second, rigorous cross-cultural adaptation procedures are essential when applying instruments developed in Western contexts to Indonesia, as direct translation without proper validation may result in measurement errors (van de Vijver & Tanzer, 2004). Indonesian educational contexts differ substantially from Western settings where these instruments were originally developed. Third, minimal research exists on instrument validation specifically for novice science teachers in Indonesia, who face unique challenges including limited pedagogical content knowledge and inadequate professional development (Aslam et al., 2023).

This pilot study addresses these methodological gaps by developing and validating a complete measurement model. The specific objectives are threefold. First, to systematically adapt three instruments for Indonesian novice science teachers through forward-backward translation, cultural contextualization, Ethno-STEM integration, and expert validation. Second, to examine psychometric properties including content validity through Content Validity Index calculations, internal consistency reliability using Cronbach's alpha, and construct validity through Exploratory Factor Analysis to determine underlying factor structures. Third, to provide validated and reliable instruments for the main study investigating relationships among cultural intelligence, cultural diversity climate, and intention to implement Ethno-STEM approaches. By establishing measurement validity through this pilot study, subsequent research can confidently examine whether the hypothesized predictor-outcome relationships hold empirically, ultimately informing teacher preparation and professional development strategies for promoting culturally responsive pedagogy in Indonesia's diverse educational landscape.

MATERIAL AND METHODS

Design and Participants

This pilot study employed a quantitative cross-sectional design to validate instruments measuring cultural intelligence, cultural diversity climate, and intention to implement Ethno-STEM for Indonesian novice science teachers. The study involved 105 science and mathematics teachers from various schools across Indonesia. The sample consisted of 75 female (71.4%) and 30 male teachers (28.6%), with ages ranging from 22-30 years. Most participants were 22-24 years old (64.8%), with teaching experience of 1-3 years (42.9% had one year, 33.3% two years, 23.8% three years). The majority held degrees in mathematics and science education (74.3%) and taught mathematics (33.3%), chemistry (33.3%), biology (17.1%), or physics (16.2%). Geographical distribution included rural (55.2%), suburban (22.9%), urban (14.3%), and remote schools (7.6%). All participants volunteered after informed consent, with confidentiality assured.

Instruments and Adaptation

Three established instruments were adapted for Indonesian Ethno-STEM contexts. The Cultural Intelligence Scale (CQS) by Van Dyne et al. (2008) measures four dimensions: metacognitive CQ (4 items), cognitive CQ (6 items), motivational CQ (5 items), and behavioral CQ (5 items) through 20 items using 7-point Likert format. Original validation demonstrated strong psychometric properties (Cronbach's α: 0.70-0.87; CFA: NNFI=0.91, CFI=0.92, RMSEA=0.08). The Cultural Diversity Climate Scale (CDCS) by Mor-Barak and Cherin (1998) assesses organizational diversity climate through four dimensions: Organizational Fairness (6 items), Organizational Inclusion (4 items), Personal Diversity Value (3 items), and Personal Comfort (3 items) using 16 items with 6-point Likert format. Original validation showed strong reliability (α=0.83; subscales α=0.71-0.86). The Intention to Implement Ethno-STEM Scale (IIES) was newly developed grounded in Theory of Planned Behavior (Ajzen, 1991), comprising 15 items across four TPB dimensions: Behavioral Intention (4 items), Attitude toward Behavior (4 items), Subjective Norm (3 items), and Perceived Behavioral Control (4 items) using 5-point Likert format. Scale development involved literature review, item generation referencing Ethno-STEM behaviors, and expert consultation.

All instruments underwent systematic adaptation involving three phases. Phase 1 employed forward-backward translation procedures with bilingual researchers to ensure semantic equivalence. Phase 2 involved cultural contextualization, aligning terminology with Indonesian educational vocabulary, adapting examples to Indonesian school structures, and adjusting items to Indonesian values. Phase 3 integrated Ethno-STEM specificity: CQS items were recontextualized from international to Indonesian ethnic diversity, workplace to educational contexts, with four irrelevant items removed (final: 16 items); CDCS was reframed from organizational to educational settings, with six inappropriate items removed and four new items added measuring ethnic relations and knowledge integration (final: 16 items); IIES systematically replaced generic "STEM" with "Ethno-STEM" terminology across all items.

Content Validity

The adapted instruments underwent rigorous two-phase validation by five experts (two multicultural education specialists, two measurement experts, one science education professor). Face validation examined clarity, readability, and language appropriateness using dichotomous scale (1=acceptable, 0=not acceptable). The Index of Face Validity (IFV) was calculated with IFV ≥0.80 indicating acceptable validity (Yusoff, 2019). Phase 1 validation revealed areas requiring refinement (IFV=0.86). Following revisions, Phase 2 demonstrated enhanced clarity (IFV=1.00). Content validation used Item-level Content Validity Index (I-CVI) and Scale-level Content Validity Index (S-CVI) metrics. Each validator rated items as relevant (1) or not relevant (0). I-CVI was calculated using Lynn's (1986) formula with I-CVI ≥0.78 considered acceptable. S-CVI/Ave was calculated with S-CVI/Ave ≥0.90 indicating excellent validity (Polit & Beck, 2006). Phase 1 validation across 47 items showed S-CVI ranging from 0.94-0.96. Following expert feedback on ambiguous terminology and cultural specificity, Phase 2 achieved perfect I-CVI (1.00) for all items and S-CVI of 1.00 across all three variables.

Data Collection and Analysis

The questionnaire was administered online over two weeks. Each participant received a unique link after informed consent. The questionnaire took approximately 25-30 minutes to complete. All 105 participants successfully completed the survey (100% response rate). Data were screened for missing values, outliers, and careless responding before analysis using SPSS Version 26.0. Exploratory Factor Analysis (EFA) was conducted separately for each variable to examine underlying structure. Preliminary tests included Kaiser-Meyer-Olkin (KMO) Measure of Sampling Adequacy (acceptable ≥0.80) and Bartlett's Test of Sphericity (significant $p<0.05$). The extraction method was Principal Axis Factoring with Varimax rotation. The number of factors was determined using Kaiser's criterion (eigenvalues >1). Item retention decisions were based on factor loadings (minimum 0.50), cross-loading examination, and communality values (minimum 0.50). Items with communality values below 0.50 were flagged as problematic and removed (Hair et al., 2019). Internal consistency reliability was assessed using Cronbach's alpha, with values ≥0.70 considered acceptable and ≥0.80 indicating good reliability (Nunnally & Bernstein, 1994).

RESULT

Content Validity Results

Table 1 presents content validity indices across two validation phases. Phase 1 validation revealed areas requiring refinement across all three scales. The Cultural Intelligence scale achieved S-CVI of 0.95, with four items (CQ1-CQ4) receiving I-CVI of 0.80 and remaining items achieving perfect scores (1.00). Expert feedback focused on adding specificity to abstract cultural knowledge items and strengthening Ethno-STEM contextualization. The Cultural Diversity Climate scale achieved S-CVI of 0.95, with four items (CDC2, CDC5, CDC6, CDC13) receiving I-CVI of 0.80. Validators highlighted concerns about reverse-coded items causing confusion and the need for clearer connections to educational contexts. The Intention scale achieved highest Phase 1 S-CVI of 0.96, with only three items (IT2, IT10, IT14) receiving I-CVI of 0.80. Experts recommended focusing items specifically on implementation behaviors rather than learning intentions.

Following comprehensive revisions addressing all expert feedback, Phase 2 validation demonstrated substantial improvement. All three scales achieved perfect S-CVI of 1.00, with every item receiving unanimous expert approval (I-CVI=1.00). This progression from Phase 1 (S-CVI: 0.95-0.96) to Phase 2 (S-CVI: 1.00) indicates successful resolution of ambiguities, enhanced cultural appropriateness, and strengthened alignment with Ethno-STEM pedagogical principles. The iterative validation process proved essential for refining instruments adapted across substantially different cultural and educational contexts.

Table 1. Content Validity Index Results Across Two Phases

Variable	Initial Items	Phase 1 S-CVI	Phase 2 S-CVI	Status
Cultural Intelligence	16	0.95	1.00	Excellent
Cultural Diversity Climate	16	0.95	1.00	Excellent
Intention to Implement	15	0.96	1.00	Excellent

Psychometric Properties Using EFA

Table 2 presents comprehensive EFA results for all three variables. All variables met necessary assumptions for factor analysis. Kaiser-Meyer-Olkin (KMO) values exceeded the 0.80 threshold, ranging from 0.910 to 0.922, indicating excellent sampling adequacy. Bartlett's Test of Sphericity was significant ($p<0.001$) for all variables, confirming that correlation matrices were suitable for factor analysis. These results provided confidence that the data were appropriate for Exploratory Factor Analysis.

The Cultural Intelligence scale yielded KMO of 0.910. Communalities analysis revealed two items (CQ4, CQ7) with values below the 0.50 threshold, indicating inadequate contribution to factor structure. CQ4 assessed checking accuracy of cultural knowledge through comparison with elders or documentation, while CQ7 measured another metacognitive dimension. Despite achieving perfect I-CVI (1.00) in Phase 2 content validation, these items failed to perform adequately with actual respondents. These items were eliminated, leaving 14 items for the main study with an elimination rate of 12.5%. The refined scale achieved Cronbach's alpha of 0.889, indicating good internal consistency. The four-dimensional structure (metacognitive, cognitive, motivational, behavioral CQ) was maintained, suggesting successful adaptation of Earley and Ang's (2003) theoretical framework to Indonesian contexts.

The Cultural Diversity Climate scale demonstrated excellent sampling adequacy (KMO=0.922), highest among the three scales. Four items (CDC2, CDC7, CDC9, CDC16) failed to meet the communality threshold and were eliminated, leaving 12 items with an elimination rate of 25.0%. CDC2 was a reverse-coded item about making inappropriate stereotypical comments that proved problematic despite revision attempts. The higher elimination rate for CDC reflected the challenges of adapting organizational diversity climate measures to educational contexts, particularly when reframing items from workplace to school settings and from general diversity to Indonesian ethnic diversity. The refined scale demonstrated excellent reliability ($\alpha=0.912$), maintaining the four-dimensional structure (organizational fairness, organizational inclusion, personal diversity value, personal comfort with diversity).

The Intention to Implement Ethno-STEM scale showed excellent sampling adequacy (KMO=0.922). Only one item (IT6) failed to meet the communality criterion and was eliminated, resulting in 14 valid items with the lowest elimination rate of 6.7%. This minimal item loss suggests that the newly developed scale, grounded in Theory of Planned Behavior and specifically designed for Ethno-STEM contexts, functioned well psychometrically. The refined scale achieved acceptable reliability ($\alpha=0.835$), maintaining the four-dimensional TPB structure (behavioral intention, attitude toward behavior, subjective norm, perceived behavioral control). While the reliability coefficient was lower than the other two scales, it still met acceptable standards for newly developed instruments ($\alpha\geq0.70$).

Table 2. Summary of Exploratory Factor Analysis Results

Variable	Initial Items	KMO	Items Eliminated	Final Items	Cronbach's α
Cultural Intelligence	16	0.910	CQ4, CQ7	14	0.889
Cultural Diversity Climate	16	0.922	CDC2, CDC7, CDC9, CDC16	12	0.912
Intention to Implement	15	0.922	IT6	14	0.835

DISCUSSION

Overview of Validated Measurement Model

This pilot study successfully validated a three-scale measurement model through rigorous procedures combining expert validation and empirical testing. Figure 2 illustrates the validation progression from 47 initial items through two-phase expert validation (Phase 1: S-CVI=0.95-0.96; Phase 2: S-CVI=1.00) to empirical testing (n=105), ultimately yielding 40 validated items across Cultural Intelligence (14 items, α=0.889), Cultural Diversity Climate (12 items, α=0.912), and Intention to Implement Ethno-STEM (14 items, α=0.835). The 14.9% item elimination rate falls within expected ranges for cross-cultural adaptation (van de Vijver & Tanzer, 2004). All scales maintained original theoretical structures, establishing methodological foundation for investigating the predictor-outcome model.

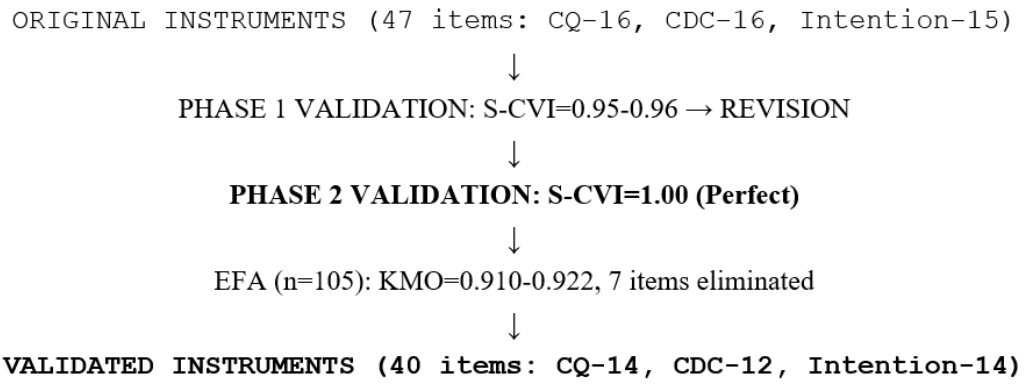

Figure 2. Validation Process Flow

Comparison with Original Validation Studies

The CQ scale's reliability (α=0.889) exceeds Ang et al.'s (2007) range, suggesting educational contextualization enhanced item cohesion. This finding aligns with research demonstrating that context-specific adaptation can improve rather than diminish psychometric properties when done thoughtfully (Lu et al., 2025; Volgemute et al., 2025). The CDC scale shows dramatic improvement (α=0.912 vs original 0.83), reflecting benefits of focused adaptation for specific populations and contexts. Reframing from organizational to school settings and from broad to ethnic diversity achieved greater conceptual clarity, consistent with findings that domain-specific measures often demonstrate stronger psychometric properties than generic measures (Nielsen et al., 2022). The Intention scale's acceptable reliability (α=0.835), while lower than Wu et al.'s (2022) unidimensional measure, meets acceptable standards for newly developed instruments (Nunnally & Bernstein, 1994) and reflects comprehensive four-dimension TPB coverage—a common tradeoff between breadth and internal consistency in multidimensional scale development (Ajzen, 2011).

Table 3. Comparison with Original Validation Studies

Scale	Original	Indonesian	Interpretation
CQ	20 items, α=0.70-0.87	14 items, α=0.889	Stronger reliability; educational contextualization enhanced cohesion
CDC	16 items, α=0.83	12 items, α=0.912	Exceeded original by 10%; focused adaptation enhanced item cohesion
Intention	15 items, α=0.931	14 items, α=0.835	Acceptable for new scale; comprehensive TPB coverage

Item Elimination and Cross-Cultural Adaptation Challenges

A critical finding emerges from the discrepancy between expert validation and empirical performance: all seven eliminated items achieved perfect I-CVI (1.00) in Phase 2, yet failed statistical criteria (communality <0.50) with actual respondents. This pattern exemplifies the well-documented limitation that expert judgment, while necessary for establishing content validity, cannot predict actual item functioning with target populations (DeVellis, 2017). CQ4, assessing verification of cultural knowledge through elders or documentation, proved conceptually appropriate to experts but not resonant with novice teachers' actual practices. This finding aligns with developmental expertise literature suggesting that novice professionals focus primarily on procedural skill acquisition rather than metacognitive monitoring of their knowledge (Ho & Lau, 2025; Persky & Robinson, 2017). The discrepancy underscores assertions by methodologists that pilot testing with adequate sample sizes constitutes a non-negotiable requirement rather than optional step in instrument development (Bujang et al., 2024).

The CDC scale's highest elimination rate (25%) reflects challenges adapting organizational measures to educational contexts and problems with reverse-coded items in Indonesian culture. CDC2 (making stereotypical

comments) remained problematic despite revision, consistent with research documenting that reverse-coded items often function poorly in non-Western contexts (Schmitt & Allik, 2005; Wong et al., 2003). Indonesian collectivistic norms emphasizing harmony maintenance (Hofstede, 2011) likely exacerbate social desirability bias for negatively-framed items. Additionally, cross-cultural psychology research demonstrates that Indonesian communication preferences favor direct, positive statements over negatively-framed self-assessments, creating cognitive complexity with reverse coding that may introduce measurement error rather than control acquiescence bias (Harzing, 2006). This finding supports recent recommendations that researchers consider eliminating reverse-coded items when adapting instruments for Asian contexts (Weijters & Baumgartner, 2012). The differential elimination rates—CQ (12.5%), CDC (25%), Intention (6.7%)—align with scale development literature suggesting newly developed indigenous instruments often demonstrate superior psychometric properties compared to heavily adapted measures requiring substantial reconceptualization (Cheung et al., 2011).

Validator feedback revealed four adaptation challenge categories documented in cross-cultural adaptation literature (Epstein et al., 2015): (1) terminological precision—"arts and crafts" inadequately captured indigenous artistic expressions, requiring "traditional arts, crafts, and cultural expressions"; (2) concreteness versus abstraction—abstract concepts needed specific examples ("traditional knowledge systems such as traditional medicine, farming techniques"), reflecting Indonesian educational culture's pragmatic orientation (Bjork, 2005); (3) outdated concepts—"learning styles" required updating to "student needs and preferences" to align with current educational research (Kirschner, 2017); (4) proactive versus passive language—"get accustomed to teaching" suggested passivity whereas "proactively integrate" emphasized agency consistent with transformative multicultural education philosophy (Nieto, 2017). These categories provide a taxonomy guiding systematic cross-cultural adaptation processes.

Implications for Main Study and Methodological Contributions

The validated instruments enable rigorous investigation of the predictor-outcome model, with strong predictor reliabilities (α=0.889, 0.912) providing confidence for structural equation modeling (Hair et al., 2019; Kline, 2016). The maintained four-dimensional structures enable nuanced analyses examining which specific dimensions drive relationships (e.g., motivational CQ predicting behavioral intention), following recommendations for dimension-level analyses in multidimensional construct research (Edwards, 2001). Future research should conduct Confirmatory Factor Analysis with larger samples ($n>200$) to verify factor structures (Brown, 2015), and conduct longitudinal studies establishing whether measured intention predicts actual implementation behavior as suggested by meta-analytic research on intention-behavior relationships (Webb & Sheeran, 2006).

Methodologically, this study contributes by: (1) providing empirical evidence for multi-method validation necessity through documented expert-empirical discrepancies, extending arguments by measurement theorists (DeVellis, 2017); (2) identifying reverse-coded items as culturally problematic in Indonesian contexts, supporting recent critiques of uncritical reverse-coding use in cross-cultural research (Weijters et al., 2013); (3) demonstrating indigenous instrument development may outperform heavy adaptation (Intention's 6.7% vs CDC's 25% elimination), consistent with cultural psychology recommendations (Cheung et al., 2011); (4) providing taxonomy of adaptation challenges grounded in systematic validator feedback analysis (Guillemin et al., 1993); and (5) documenting organizational-to-educational and workplace-to-ethnic diversity translation challenges, contributing to domain-specific adaptation literature (Herdman et al., 1998). These contributions advance cross-cultural instrument development methodology while providing Indonesian researchers with psychometrically sound tools for investigating culturally responsive STEM education, ultimately supporting evidence-based teacher preparation and professional development strategies (Darling-Hammond et al., 2017).

CONCLUSION

This pilot study successfully developed and validated a complete measurement model for investigating predictors of Indonesian novice science teachers' intention to implement Ethno-STEM approaches. Through rigorous cross-cultural adaptation procedures, three instruments were validated: the Cultural Intelligence Scale (14 items, α=0.889), the Cultural Diversity Climate Scale (12 items, α=0.912), and the newly developed Intention to Implement Ethno-STEM Scale (14 items, α=0.835). All instruments demonstrated acceptable to excellent psychometric properties, maintaining their original theoretical structures while proving contextually appropriate for Indonesian settings. The elimination of seven items (14.9%) despite achieving perfect content validity indices highlights a critical methodological insight: expert validation alone is insufficient for ensuring instrument quality. Empirical pilot testing revealed functional problems that expert panels could not anticipate, particularly with reverse-coded items in Indonesian cultural contexts. The validated instruments now provide Indonesian educational researchers with psychometrically sound tools for rigorous investigation of culturally responsive STEM education, ultimately informing teacher preparation and professional development strategies.

LIMITATIONS AND FUTURE DIRECTIONS

Several limitations warrant consideration. First, the sample size of 105 participants, while adequate for EFA, represents a modest sample that may limit generalizability. Second, this pilot study employed only Exploratory Factor Analysis and assessed only internal consistency reliability through Cronbach's alpha. Third, convergent and discriminant validity were not examined. Fourth, all data were collected through self-report questionnaires, potentially introducing common method bias. Finally, the cross-sectional design precludes examination of whether measured intention predicts subsequent implementation behavior.

Future research should address these limitations through three priorities. First, conducting Confirmatory Factor Analysis with larger independent samples (n>200) to verify factor structures and establish measurement invariance across teacher subgroups. Second, expanding validity evidence by examining test-retest reliability, convergent and discriminant validity with theoretically related constructs, and incorporating multiple data collection methods including classroom observations. Third, conducting longitudinal studies to establish predictive validity by tracking whether teachers with higher intention scores actually implement Ethno-STEM approaches more frequently in their classrooms. Despite these limitations, this study successfully established the methodological foundation for investigating factors influencing Ethno-STEM implementation in Indonesian contexts.

ACKNOWLEDGEMENT

The greatest appreciation is given to Indonesia Endowment Fund for Education (LPDP) Ministry of Finance, Indonesia, for the support provided for the success of this research and publication.

REFERENCES

Ajzen, I. (1991). The Theory of planned behavior. Organizational Behavior and Human Decision Processes, 50, 179–211.

Ajzen, I. (2011). The theory of planned behaviour: Reactions and reflections. Psychology & Health, 26(9), 1113–1127. https://doi.org/10.1080/08870446.2011.613995

Alifuddin, Moh., & Widodo, W. (2022). How Is Cultural Intelligence Related to Human Behavior? Journal of Intelligence, 10(1), 3. https://doi.org/10.3390/jintelligence10010003

Ang, S., Van Dyne, L., Koh, C., Ng, K. Y., Templer, K. J., Tay, C., & Chandrasekar, N. A. (2007). Cultural Intelligence: Its Measurement and Effects on Cultural Judgment and Decision Making, Cultural Adaptation and Task Performance. Management and Organization Review, 3(3), 335–371. https://doi.org/10.1111/j.1740-8784.2007.00082.x

Aslam, S., Alghamdi, A. A., Abid, N., & Kumar, T. (2023). Challenges in Implementing STEM Education: Insights from Novice STEM Teachers in Developing Countries. Sustainability (Switzerland), 15(19). https://doi.org/10.3390/su151914455

Bandura, A. (1986). Social foundations of thought and action: A social cognitive theory. Prentive-Hall Inc.

Bjork, C. (2005). Indonesian Education. Routledge. https://doi.org/10.4324/9780203959015

Brown, T. A. (2015). Confirmatory Factor Analysis for Applied Research (2nd ed.). The Guilford Press.

Bujang, M. A., Omar, E. D., Foo, D. H. P., & Hon, Y. K. (2024). Sample size determination for conducting a pilot study to assess reliability of a questionnaire. Restorative Dentistry & Endodontics, 49(1). https://doi.org/10.5395/rde.2024.49.e3

Cheung, F. M., van de Vijver, F. J. R., & Leong, F. T. L. (2011). Toward a new approach to the study of personality in culture. American Psychologist, 66(7), 593–603. https://doi.org/10.1037/a0022389

Dallman, L. L. (2020). The Role of Teacher Emotional Intelligence and Cultural Intelligence in Establishing Positive Social-Emotional Dynamics with Diverse Student Group. George Mason University.

Darling-Hammond, L., Hyler, M., & Gardner, M. (2017). Effective Teacher Professional Development. https://doi.org/10.54300/122.311

DeVellis, R. F. (2017). Scale Development: Theory and Applications (4th ed.). Sage.

Earley, P. C., & Ang, S. (2003). Cultural Intelligence. Stanford University Press. https://doi.org/10.1515/9780804766005

Edwards, J. R. (2001). Multidimensional Constructs in Organizational Behavior Research: An Integrative Analytical Framework. Organizational Research Methods, 4(2), 144–192. https://doi.org/10.1177/109442810142004

Epstein, J., Santo, R. M., & Guillemin, F. (2015). A review of guidelines for cross-cultural adaptation of questionnaires could not bring out a consensus. Journal of Clinical Epidemiology, 68(4), 435–441. https://doi.org/10.1016/j.jclinepi.2014.11.021

Guillemin, F., Bombardier, C., & Beaton, D. (1993). Cross-cultural adaptation of health-related quality of life measures: Literature review and proposed guidelines. Journal of Clinical Epidemiology, 46(12), 1417–1432. https://doi.org/10.1016/0895-4356(93)90142-N

Hair, J. F., L.D.S. Gabriel, M., da Silva, D., & Braga Junior, S. (2019). Development and validation of attitudes measurement scales: fundamental and practical aspects. RAUSP Management Journal, 54(4), 490–507. https://doi.org/10.1108/RAUSP-05-2019-0098

Harzing, A.-W. (2006). Response Styles in Cross-national Survey Research. International Journal of Cross Cultural Management, 6(2), 243–266. https://doi.org/10.1177/1470595806066332

Herdman, M., Fox-Rushby, J., & Badia, X. (1998). A model of equivalence in the cultural adaptation of HRQoL instruments: the universalist approach. Quality of Life Research, 7(4), 323–335. https://doi.org/10.1023/A:1024985930536

Ho, W. W. Y., & Lau, Y. H. Y. (2025). Role of reflective practice and metacognitive awareness in the relationship between experiential learning and positive mirror effects: A serial mediation model. Teaching and Teacher Education, 157, 104947. https://doi.org/10.1016/j.tate.2025.104947

Hofstede, G. (2011). Dimensionalizing Cultures: The Hofstede Model in Context. Online Readings in Psychology and Culture, 2(1). https://doi.org/10.9707/2307-0919.1014

Hurtado, S., Alvarez, C. L., Guillermo-Wann, C., Cuellar, M., & Arellano, L. (2012). A Model for Diverse Learning Environments (pp. 41–122). https://doi.org/10.1007/978-94-007-2950-6_2

Izzah, S. N., Sudarmin, Wiyanto, & Wardani, S. (2023). Analysis of Science Concept Mastery, Creative Thinking Skills, and Environmental Attitudes After Ethno-STEM Learning Implementation. International Journal of Instruction, 16(3), 777–796. https://doi.org/10.29333/iji.2023.16342a

Kirschner, P. A. (2017). Stop propagating the learning styles myth. Computers & Education, 106, 166–171. https://doi.org/10.1016/j.compedu.2016.12.006

Kline, R. B. (2016). Principles and Practice of Structural Equation Modeling (4th ed.). The Guilford Press.

Lu, H., Li, X., & Li, K. (2025). Adaptation and Validation of the Scale for Chinese Preschool Teachers' Self-Efficacy (SCPTSE): Based on Classical Test Theory and Item Response Theory. Behavioral Sciences, 15(6), 741. https://doi.org/10.3390/bs15060741

McKay, P. F., Avery, D. R., & Morris, M. A. (2009). A Tale of Two Climates: Diversity Climate From Subordinates' and Managers' Perspectives and Their Role in Store Unit Sales Performance. Personnel Psychology, 62, 767–791.

Mor Barak, M. E., Cherin, D. A., & Berkman, S. (1998). A Tool to Expand Organizational Understanding of Workforce Diversity: Exploring a Measure of Inclusion-Exclusion. Administration in Social Work, 22(1), 47–64. https://doi.org/https://doi.org/10.1300/J147v22n01_04

Ng, H. K. Y., Chen, S. X., & Lam, B. C. P. (2024). An examination of the impact of cultural intelligence on life satisfaction: Insights from Hong Kong, Mainland China, and the United States. International Journal of Intercultural Relations, 100. https://doi.org/10.1016/j.ijintrel.2024.101971

Nielsen, T., Martínez-García, I., & Alastor, E. (2022). Exploring First Semester Changes in Domain-Specific Critical Thinking. Frontiers in Education, 7. https://doi.org/10.3389/feduc.2022.884635

Nieto, S. (2017). Language, Culture, and Teaching. Routledge. https://doi.org/10.4324/9781315465692

Nunnally, J. C., & Bernstein, I. H. (1994). The Assessment of Reliability. Psychometric Theory, 3, 248–292.

Persky, A. M., & Robinson, J. D. (2017). Moving from Novice to Expertise and Its Implications for Instruction. American Journal of Pharmaceutical Education, 81(9), 6065. https://doi.org/10.5688/ajpe6065

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

Richards, H. V., Brown, A. F., & Forde, T. B. (2007). Addressing Diversity in Schools: Culturally Responsive Pedagogy.

Rizki, I. A., Suprapto, N., & Admoko, S. (2022). Exploration of physics concepts with traditional engklek (hopscotch) game: Is it potential in physics ethno-STEM learning? Jurnal Ilmiah Pendidikan Fisika Al-Biruni, 11(1), 19–33. https://doi.org/10.24042/jipfalbiruni.v11i1.10900

Schmitt, D. P., & Allik, J. (2005). Simultaneous Administration of the Rosenberg Self-Esteem Scale in 53 Nations: Exploring the Universal and Culture-Specific Features of Global Self-Esteem. Journal of Personality and Social Psychology, 89(4), 623–642. https://doi.org/10.1037/0022-3514.89.4.623

Sheeran, P., Orbell, S., & Trafimow, D. (1999). Does the Temporal Stability of Behavioral Intentions Moderate Intention-Behavior and Past Behavior-Future Behavior Relations? Personality and Social Psychology Bulletin, 25(6), 724–734. https://doi.org/10.1177/0146167299025006007

Sincer, I., Volman, M., van der Veen, I., & Severiens, S. (2021). The relationship between ethnic school composition, school diversity climate and students' competences in dealing with differences. Journal of Ethnic and Migration Studies, 47(9), 2039–2064. https://doi.org/10.1080/1369183X.2020.1846508

Sukmayadi, V., & Yahya, A. H. (2020). Indonesian Education Landscape and the 21 st Century Challenges. Journal of Social Studies Education Research, 11(4), 219–234. www.jsser.org

van de Vijver, F., & Tanzer, N. K. (2004). Bias and equivalence in cross-cultural assessment: an overview. European Review of Applied Psychology, 54(2), 119–135. https://doi.org/10.1016/j.erap.2003.12.004

Van Dyne, L., Ang, S., & Koh, C. (2008). Development and Validation of The CQS: The Cultural Intelligence Scale. In Handbook of Cultural Intelligence.

Volgemute, K., Ulme, G., Perepjolkina, V., Licis, R., Abele, A., Lavins, R., Klonova, A., Tihija, A., & Grants, E. (2025). Adaptation and psychometric properties of Psychological Skills Inventory for Sport (PSIS-R5) in Latvian athletes: Insights and implications for practice. PLOS One, 20(5), e0325225. https://doi.org/10.1371/journal.pone.0325225

Webb, T. L., & Sheeran, P. (2006). Does changing behavioral intentions engender behavior change? A meta-analysis of the experimental evidence. Psychological Bulletin, 132(2), 249–268. https://doi.org/10.1037/0033-2909.132.2.249

Weijters, B., & Baumgartner, H. (2012). Misresponse to Reversed and Negated Items in Surveys: A Review. Journal of Marketing Research, 49(5), 737–747. https://doi.org/10.1509/jmr.11.0368

Weijters, B., Baumgartner, H., & Schillewaert, N. (2013). Reversed item bias: An integrative model. Psychological Methods, 18(3), 320–334. https://doi.org/10.1037/a0032121

Wong, C. A., Eccles, J. S., & Sameroff, A. (2003). The Influence of Ethnic Discrimination and Ethnic Identification on African American adolescents' School and Socioemotional Adjustment. Journal of Personality, 71(6), 1197–1232. https://doi.org/10.1111/1467-6494.7106012

Wu, P., Yang, L., Hu, X., Li, B., Liu, Q., Wang, Y., & Huang, J. (2022). How K12 Teachers' Readiness Influences Their Intention to Implement STEM Education: Exploratory Study Based on Decomposed Theory of Planned Behavior. Applied Sciences (Switzerland), 12(23). https://doi.org/10.3390/app122311989

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.

