



OPEN ACCESS

Development and Validation of an Instrument Measuring Elementary School Teachers' Readiness to Implement Multicultural Education in Social Studies (IPS) Learning

Khusnul Khotimah¹, Yuniawatika², Surayanah³

¹²³ Universitas Negeri Malang, Indonesia

Keywords:

teacher readiness;
multicultural education;
social studies (IPS);
instrument validation;
Cronbach's alpha

Correspondence to

Khusnul Khotimah, State
University of Malang,
Indonesia
email:
khusnul.khotimah.fip@um.ac.id

Received – August 10, 2026

Accepted – August 11, 2026

Revised – September 10,
2026

Published Online First -
September 14, 2026



Check for updates

© Author(s) (or their
employer(s)) 2025. Re-use
permitted under CC BY-NC.
No commercial re-use. See
rights and permissions.
Published
by JGA.

Abstract

The current study investigated the initial evidence for validity and reliability of the tool used for the measurement of elementary school teachers' readiness to apply multicultural education in IPS learning in Indonesia. The tool was based on the adaptation of DivePCK framework and previously validated multicultural attitude scales and consisted of 65 five-point Likert scale questions divided into four main constructs: diversity pedagogical content knowledge (15 questions), attitudes towards diversity (15), organizational support (15), and teachers' readiness (20). The data collection was conducted through the online survey administered to 79 elementary school teachers representing three different Indonesian areas (Malang and Blitar – East Java, Ende-Flores - East Nusa Tenggara). After excluding 13 careless straight line answers (16.5%), 66 valid responses were analyzed. The content validity of the tool was established by expert judgment. Item validity was evaluated by means of corrected item-total correlation (criterion $r > 0.242$, $df = 64$), reliability by means of Cronbach's alpha. Results showed that 63 of 65 items were valid; two items were flagged as problematic: a reverse-worded attitude item (SK9, $r = 0.04$) and a culturally responsive teaching item (KG8, $r = 0.21$). Cronbach's alpha per construct ranged from 0.891 to 0.934, indicating good to excellent internal consistency; removing SK9 raised the attitude subscale's alpha from 0.891 to 0.925. As a supplementary analysis, inter-construct correlations were positive and moderate ($r = 0.44-0.73$), consistent with the theorized structure but interpreted cautiously. The instrument demonstrates good preliminary reliability and item validity and is suitable for measuring teacher readiness after the two flagged items are revised or removed. However, the modest, non-probability, and geographically heterogeneous sample means the findings are preliminary and require exploratory and confirmatory factor analysis on a larger, representative sample.

To cite: Khotimah, K., Yuniawatika, & Surayanah. (2026). Development and validation of an instrument measuring elementary school teachers' readiness to implement multicultural education in social studies (IPS) learning. *As Sibyan: Jurnal Kajian Kritis Pendidikan Islam dan Manajemen Pendidikan Dasar*, 9(2). doi: https://doi.org/10.52484/as_sibyan.v9i2.1186

Introduction

Indonesia is one of the most culturally diverse countries in the world, with more than 1,300 ethnic groups and hundreds of local languages spread across its archipelago. This diversity makes the elementary school classroom a microcosm of the nation's diversity and demands deliberate pedagogical handling. Multicultural education has emerged as both an idea and a movement to ensure equal learning opportunities for students from diverse cultural, ethnic, linguistic, and religious backgrounds (Jayadi et al., 2022; Ladson-Billings, 1995; Suri & Chandra, 2021). Instilling the values of tolerance and appreciation of difference from the early-childhood and primary levels is crucial in shaping the foundation of students' attitudes (Qadafi et al., 2024). At the elementary level, Social Studies (IPS) occupies a strategic position as a vehicle for multicultural education because its content relates directly to social, cultural, and civic dynamics (Suwela Antara et al., 2024).

Indeed, such national diversity results in specific expectations regarding the work of elementary school teachers. In the same classroom, teachers have to deal with children who belong to various ethnicities, religions, and languages and hence have to be able to manage the possible tensions between groups of students, not to marginalize minority or newcomer students, to choose IPS samples and materials reflecting various cultures, and to modify language used in the classroom to be understandable for children speaking other home languages. This is complicated by the scarcity of culturally relevant educational resources and the lack of adequate preparation in terms of multicultural teaching practices among teachers, making them include multiculturalism into teaching practices occasionally (Jayadi et al., 2022; Suri & Chandra, 2021; Suwela Antara et al., 2024). It is this gap between the diversity teachers encounter and the preparation they receive that makes teacher readiness a central concern for multicultural education in Indonesian elementary schools.

The effectiveness of multicultural education depends heavily on teacher readiness as the spearhead for translating the curriculum into meaningful classroom practice. In this study, teacher readiness to implement multicultural education is conceptually defined as a teacher's overall state of preparedness to plan and enact culturally responsive IPS instruction. Readiness is treated as a higher-order, multidimensional construct rather than a single attribute: it integrates a cognitive component (diversity-related pedagogical content knowledge), an affective component (attitudes toward diversity), and a conative or behavioral component (perceived capability and intention to design inclusive lessons, deliver responsive instruction, assess fairly, and manage a diverse classroom), and it is enabled or constrained by the school's organizational support. In other words, readiness is understood here as the joint product of knowing, valuing, and being able to act, not knowledge or attitude alone (Aalto et al., 2024; Comstock et al., 2023; Dursun et al., 2021). Consistent with this definition, Dursun et al. (2021) formulated the Diversity Pedagogical Content Knowledge (DivePCK) framework, which extends the concept of pedagogical content knowledge into the domain of diversity, covering multicultural content

knowledge, pedagogical knowledge for diversity, and culturally responsive curricular knowledge.

Examining and strengthening teacher readiness requires a valid and reliable measurement tool. However, most of these instruments were constructed within Western settings, such as Teacher Multicultural Attitude Survey and DivePCK instrument, which means that their application to Indonesia requires their adaptation and psychometric validation (Chahar Mahali & Sevigny, 2022; Mou et al., 2025). Direct transfer of such instruments poses several problems related to four interdependent factors. First, the difference in the curricular setting lies in the fact that Indonesian elementary IPS curriculum is based on Merdeka Curriculum, Pancasila Student Profile, and IPS Learning Outcomes in Indonesia, while Western social studies curriculum instruments rely on different content standards and learning outcomes. Second, the difference in cultural referents consists of multiculturalism in Indonesia being based on the idea of *Bhinneka Tunggal Ika* and having a focus on interethnic, interreligious, and interlinguistic diversity of one nation, as opposed to multicultural scales in North America and Europe, which focus on race-related and immigration-related diversity. Fourth, and consequently, the very meaning of multicultural competence is culturally situated, so items validated elsewhere may carry different connotations or lose relevance when translated. For this reason, such tools need not only to be translated but also conceptually respecified and re-established through new psychometric evidence before being applied to Indonesian elementary IPS teachers since validity depends on quality measurement because a poor tool produces invalid findings (Boateng et al., 2018; Bolarinwa, 2015; Fatmawati et al., 2023; Sürücü & Maslakçı, 2020; Taber, 2018).

To identify the research gap in a systematic manner rather than making assertions, we performed a structured search in Scopus, Web of Science, ERIC, and the national Indonesian databases Garuda and the SINTA journal database portal with regard to articles published between 2015 and 2025. For this purpose, we used the keywords for the population, construct, and measurement goal, as well as Boolean strings such as ('multicultural education' OR 'culturally responsive teaching' OR 'diversity') AND ('teacher readiness' OR 'teacher competence' OR 'pedagogical content knowledge') AND ('instrument' OR 'scale' OR 'questionnaire' OR 'validation') AND ('elementary' OR 'primary' OR 'social studies' OR 'IPS') AND ('Indonesia*'). Studies were included when they reported the development or validation of a teacher-level instrument relevant to multicultural or culturally responsive education, and excluded when they measured student outcomes only, reported no psychometric evidence, or were set in a non-comparable schooling context. The search returned validated instruments developed largely in Western settings and a small number of Indonesian instruments focused on single dimensions, such as multicultural attitudes or general multicultural competence (Fatmawati et al., 2023; Subhan et al., 2026). It did not return any instrument that integrally measures elementary school IPS teachers' readiness for multicultural education by combining knowledge (DivePCK), attitudes, organizational support, and behavioral readiness, particularly one developed and validated within the Indonesian curriculum

context. This absence of an integrated, context-specific, and psychometrically examined instrument is the gap the present study addresses.

The DivePCK framework was selected as the cognitive backbone of the instrument for three reasons rather than adopted by default. Conceptually, DivePCK extends the well-established notion of pedagogical content knowledge specifically into the domain of diversity, so it captures not general multicultural awareness alone but the teachable, subject-embedded knowledge that IPS instruction requires, namely multicultural content knowledge, pedagogical knowledge for diversity, and culturally responsive curricular knowledge (Dursun et al., 2021). This subject-and-practice orientation fits IPS more closely than attitude-only frameworks such as the Teacher Multicultural Attitude Survey, which capture disposition but not knowledge, and more closely than broad culturally responsive teaching self-efficacy scales, which capture confidence but not content. Methodologically, DivePCK has an explicit three-subdimension structure that can be operationalized into items and mapped onto the Indonesian IPS Learning Outcomes. Finally, DivePCK is recent, has been validated across teacher-education programs, and has been applied cross-nationally (Dursun et al., 2021; Mou et al., 2025), which makes it a defensible cognitive anchor while still requiring the contextual re-validation this study provides. DivePCK therefore supplies the knowledge dimension, which is complemented by the affective (attitudes), contextual (organizational support), and behavioral (readiness) dimensions so that readiness is represented as a whole.

Against this background, this study aims to develop and examine the preliminary validity (content validity and item validity) and reliability of an instrument measuring elementary school teachers' readiness to implement multicultural education in social studies (IPS) learning. Specifically, the study answers three questions: (1) to what extent do the instrument's items meet item-validity criteria based on corrected item-total correlations; (2) to what extent does the instrument possess adequate internal consistency (reliability) for each construct and its subdimensions; and (3) as a supplementary question, what is the pattern of inter-construct correlations, interpreted cautiously as preliminary evidence of the instrument's internal structure and not as confirmatory evidence of its factor structure.

Methods

Research design. This study employed a quantitative, instrument development and validation design, implemented as a preliminary (pilot) validation study through a cross-sectional online survey. Consistent with staged approaches to scale development, the study is positioned at the item-analysis and internal-consistency stage that precedes full factor-analytic validation (Boateng et al., 2018). Given the sample size and the deliberate absence of factor analysis at this stage, the design is best characterized as a pilot instrument-validation study whose purpose is to generate preliminary content-validity, item-validity, and reliability evidence, not to confirm a factor structure. Accordingly, the analysis focused on evidence of content validity, item validity, and internal-consistency

reliability, with inter-construct correlations examined only as supplementary, preliminary evidence of construct relationships.

Instrument. The instrument consisted of 65 statement items in a five-point Likert format (1 = strongly disagree to 5 = strongly agree), measuring four constructs. The Diversity Pedagogical Content Knowledge construct (15 items) was adapted from the DivePCK framework (Dursun et al., 2021) with three subdimensions: multicultural content knowledge, pedagogical knowledge, and curricular knowledge. The attitudes-toward-diversity construct (15 items) was developed from the literature on teachers' multicultural attitudes (Abacioglu et al., 2020) with three subdimensions, cultural appreciation, perspective openness, and commitment to social justice, and it included one negatively worded item (SK9) as a control item. The organizational support construct (15 items) was built from a culturally responsive leadership perspective with three subdimensions: school policy, professional training, and organizational climate. The teacher readiness construct (20 items) measured four dimensions: inclusive lesson planning, responsive instruction, diversity-based assessment, and multicultural classroom management. To avoid conceptual confusion, this fourth construct, labeled Teacher Readiness (KG), operationalizes the behavioral or enactment facet of readiness, that is, teachers' perceived capability and intention to plan, deliver, assess, and manage culturally responsive IPS learning, whereas the superordinate concept of readiness is represented by the four constructs jointly. The items were localized to the Indonesian context by referring to Bhinneka Tunggal Ika, the Global Diversity (Berkebinekaan Global) dimension of the Pancasila Student Profile, and the IPS Learning Outcomes. Content validity was established through expert judgment by three experts (Yusoff, 2019; Zamanzadeh et al., 2015): an expert in multicultural education, an expert in measurement, and an expert in IPS education. The full specification of constructs, subdimensions, item sources, and adaptation status is presented in the instrument blueprint (Table 1).

Table 1. Instrument blueprint: constructs, subdimensions, item sources, and adaptation status

Construct	Subdimension	Item codes	n	Source / basis	Adaptation status
Construct 1. Diversity Pedagogical Content Knowledge (PCK), 15 items					
Diversity PCK	Multicultural content knowledge	PCK1–PCK5	5	DivePCK framework (Dursun et al., 2021)	Adapted and localized
Diversity PCK	Pedagogical knowledge for diversity	PCK6–PCK10	5	DivePCK framework (Dursun et al., 2021)	Adapted and localized
Diversity PCK	Culturally responsive curricular knowledge	PCK11–PCK15	5	DivePCK framework (Dursun et al., 2021)	Adapted and localized
Construct 2. Attitudes toward Diversity (SK), 15 items					
Attitudes	Cultural appreciation	SK1–SK5	5	Teachers' multicultural attitudes (Abacioglu et al., 2020)	Adapted
Attitudes	Perspective openness (incl. reverse item SK9)	SK6–SK10	5	Abacioglu et al. (2020); reverse control item newly written	Adapted + newly developed control item
Attitudes	Commitment to social justice	SK11–SK15	5	Abacioglu et al. (2020); Ladson-Billings (1995)	Adapted

Construct	Subdimension	Item codes	n	Source / basis	Adaptation status
Construct 3. Organizational Support (DO), 15 items					
Organizational support	School policy	DO1-DO5	5	Culturally responsive leadership perspective	Newly developed and localized
Organizational support	Professional training	DO6-DO10	5	Culturally responsive leadership; PD literature (Comstock et al., 2023)	Newly developed and localized
Organizational support	Organizational climate	DO11-DO15	5	Culturally responsive leadership perspective	Newly developed and localized
Construct 4. Teacher Readiness / perceived behavioral readiness (KG), 20 items					
Teacher readiness	Inclusive lesson planning	KG1-KG5	5	IPS Learning Outcomes; inclusive planning literature	Newly developed and localized
Teacher readiness	Responsive instruction (incl. KG8)	KG6-KG10	5	Culturally responsive teaching (Ladson-Billings, 1995)	Newly developed and localized
Teacher readiness	Diversity-based assessment	KG11-KG15	5	IPS Learning Outcomes; responsive assessment literature	Newly developed and localized
Teacher readiness	Multicultural classroom management	KG16-KG20	5	Culturally responsive classroom management literature	Newly developed and localized
<i>Note. Response format: five-point Likert scale (1 = strongly disagree to 5 = strongly agree). DivePCK = Diversity Pedagogical Content Knowledge; PD = professional development.</i>					

Participants and procedure. The questionnaire was distributed online via Google Forms to elementary school teachers on a voluntary basis. A total of 79 responses were collected from three regions: Malang Regency/City, Blitar Regency/City (East Java), and the Ende-Flores area (East Nusa Tenggara). The three regions have been selected specifically to add cultural diversity and geographic diversity in the pilot study as opposed to conducting a nationally representative sample survey. The Malang-Blitar is a Javanese majority, Muslim majority, urban and semi-urban area within Java, the most populated island in Indonesia, while the Ende-Flores region of East Nusa Tenggara is eastern, non-Javanese, Christian majority area with linguistic diversity. Sampling teachers across these contrasting cultural and religious compositions allowed a preliminary check on whether the items function meaningfully in more than one cultural context, which is central to a multicultural instrument. Sampling was non-probability (convenience/voluntary sampling), and the resulting heterogeneity is acknowledged as a limitation.

Data screening. Before analysis, the data were screened to detect careless responses following the recommendations of Meade & Craig, (2012). Thirteen respondents (16.5%) gave identical answers across all 65 items, including the negatively worded item, which is logically inconsistent, and were therefore excluded from the analysis. Accordingly, the number of analyzed responses was 66.

Data analysis techniques. Item validity was tested using corrected item-total correlations; an item was deemed valid when its correlation coefficient exceeded the *r*-table value, namely 0.242 (*df* = 64; α = 0.05; two-tailed). Internal-consistency reliability was tested using Cronbach's alpha, with a minimum acceptable threshold of 0.70 (Cortina, 1993; Souza et al., 2017; Taber, 2018; Tavakol & Dennick, 2011). Descriptive statistics and

Pearson correlations among construct scores were computed as supplementary, preliminary evidence of construct structure. Item SK9 was reverse-scored before analysis. All computations were performed using the Python programming language (pandas and SciPy libraries). Exploratory and confirmatory factor analysis were not conducted because the ratio of respondents to items did not meet the minimum requirement, and this is stated as a limitation.

Results

Respondent characteristics. The profile of the 66 analyzed respondents is presented in Table 2. Most respondents held a bachelor's degree (95.5%) and were certified (84.8%), with a mean age of 39.5 years. It should be noted that 68.2% of respondents had never attended multicultural education training and 34.8% did not specifically teach IPS content; both points are noted as sample-composition issues discussed in the limitations section.

Table 2. Respondent characteristics (N = 66)

Characteristic	Category	n (%)
Gender	Female	37 (56.1%)
	Male	29 (43.9%)
Age	Mean 39.5 years (SD 8.3; range 25–55)	
Highest education	Bachelor's (S1)	63 (95.5%)
	Diploma (D2/D3)	2 (3.0%)
	Master's (S2)	1 (1.5%)
Employment status	GTJ (permanent foundation teacher)	21 (31.8%)
	PPPK (government contract)	20 (30.3%)
	PNS (civil servant)	17 (25.8%)
	GTT/honorary (non-permanent)	8 (12.1%)
Teacher certification	Certified	56 (84.8%)
	Not yet	10 (15.2%)
Teaches IPS content	Yes, classroom teacher	40 (60.6%)
	Yes, subject teacher	3 (4.5%)
	No	23 (34.8%)
Multicultural education training	Never	45 (68.2%)
	1–2 times	14 (21.2%)
	3–5 times	5 (7.6%)
	>5 times	2 (3.0%)
Region	Malang (regency/city)	29 (43.9%)
	Ende-Flores, NTT	22 (33.3%)
	Blitar (regency/city)	15 (22.7%)

Item validity. The corrected item-total correlations are presented in Table 3. Of the 65 items, 63 met the validity criterion ($r > 0.242$), whereas two items were declared invalid: SK9 ($r = 0.038$) and KG8 ($r = 0.214$). These two items also had the highest standard deviations ($SD = 1.24$ and 1.19), indicating inconsistent respondent interpretation.

Table 3. Item validity based on corrected item-total correlations (N = 66; criterion $r > 0.242$)

Item code	M	SD	r item-total	Decision
Construct 1. Diversity Pedagogical Content Knowledge (PCK1–PCK15)				
PCK1	4.00	0.68	0.478	Valid
PCK2	3.76	0.82	0.562	Valid
PCK3	3.80	0.83	0.441	Valid
PCK4	3.65	0.77	0.648	Valid
PCK5	3.94	0.72	0.707	Valid
PCK6	3.86	0.65	0.674	Valid

Item code	M	SD	r item-total	Decision
PCK7	3.91	0.74	0.734	Valid
PCK8	3.91	0.65	0.723	Valid
PCK9	3.94	0.76	0.725	Valid
PCK10	3.94	0.70	0.745	Valid
PCK11	3.70	0.76	0.672	Valid
PCK12	3.64	0.78	0.713	Valid
PCK13	3.89	0.77	0.706	Valid
PCK14	3.85	0.73	0.687	Valid
PCK15	3.82	0.72	0.718	Valid
Construct 2. Attitudes toward Diversity (SK1-SK15)				
SK1	4.27	0.54	0.604	Valid
SK2	4.32	0.53	0.692	Valid
SK3	4.33	0.56	0.595	Valid
SK4	4.30	0.63	0.663	Valid
SK5	4.23	0.72	0.470	Valid
SK6	3.98	0.64	0.749	Valid
SK7	4.00	0.68	0.655	Valid
SK8	4.00	0.72	0.627	Valid
SK9 (R)	2.86	1.24	0.038	Dropped
SK10	3.79	0.67	0.471	Valid
SK11	4.30	0.70	0.775	Valid
SK12	4.03	0.68	0.543	Valid
SK13	4.32	0.66	0.804	Valid
SK14	4.26	0.64	0.753	Valid
SK15	4.09	0.70	0.686	Valid
Construct 3. Organizational Support (DO1-DO15)				
DO1	3.89	0.79	0.528	Valid
DO2	4.09	0.63	0.616	Valid
DO3	4.23	0.63	0.650	Valid
DO4	4.18	0.68	0.686	Valid
DO5	4.09	0.76	0.661	Valid
DO6	3.74	0.81	0.702	Valid
DO7	3.98	0.69	0.801	Valid
DO8	3.77	0.82	0.634	Valid
DO9	3.79	0.83	0.761	Valid
DO10	3.80	0.88	0.769	Valid
DO11	4.03	0.80	0.520	Valid
DO12	4.29	0.72	0.664	Valid
DO13	4.09	0.72	0.657	Valid
DO14	4.17	0.67	0.755	Valid
DO15	4.12	0.69	0.757	Valid
Construct 4. Teacher Readiness (KG1-KG20)				
KG1	4.02	0.64	0.773	Valid
KG2	4.03	0.68	0.763	Valid
KG3	4.05	0.62	0.736	Valid
KG4	3.94	0.65	0.664	Valid
KG5	4.08	0.71	0.731	Valid
KG6	4.06	0.63	0.659	Valid
KG7	4.09	0.65	0.609	Valid
KG8	3.23	1.19	0.214	Dropped
KG9	4.15	0.61	0.621	Valid
KG10	4.26	0.56	0.723	Valid
KG11	4.05	0.67	0.828	Valid
KG12	4.18	0.65	0.789	Valid
KG13	3.56	1.12	0.444	Valid
KG14	4.00	0.63	0.743	Valid
KG15	3.95	0.67	0.816	Valid
KG16	4.15	0.64	0.668	Valid
KG17	4.17	0.65	0.631	Valid
KG18	4.11	0.64	0.593	Valid
KG19	4.08	0.66	0.567	Valid
KG20	3.95	0.87	0.558	Valid

Reliability. Cronbach's alpha coefficients for the four constructs ranged from 0.891 to 0.934, categorized as good to excellent (Table 4). At the subdimension level, most were also satisfactory, except the perspective-openness subdimension ($\alpha = 0.551$) and the responsive-instruction subdimension ($\alpha = 0.617$), which were lower. These two subdimensions respectively contained the problematic items SK9 and KG8. As confirmation, removing SK9 raised the alpha of the attitude construct from 0.891 to 0.925.

Table 4. Cronbach's alpha reliability coefficients by construct and subdimension (N = 66)

Construct / Subdimension	No. of items	Cronbach's α	Interpretation
Diversity PCK (overall)	15	0.929	Excellent
Content knowledge (PCK1-5)	5	0.874	Good
Pedagogical knowledge (PCK6-10)	5	0.943	Excellent
Curricular knowledge (PCK11-15)	5	0.948	Excellent
Attitudes toward Diversity (overall)	15	0.891	Good
Cultural appreciation (SK1-5)	5	0.875	Good
Perspective openness (SK6-10)	5	0.551	Low *
Commitment to social justice (SK11-15)	5	0.903	Excellent
Organizational Support (overall)	15	0.934	Excellent
School policy (DO1-5)	5	0.879	Good
Professional training (DO6-10)	5	0.926	Excellent
Organizational climate (DO11-15)	5	0.917	Excellent
Teacher Readiness (overall)	20	0.932	Excellent
Inclusive planning (KG1-5)	5	0.939	Excellent
Responsive instruction (KG6-10)	5	0.617	Marginal *
Diversity-based assessment (KG11-15)	5	0.836	Good
Classroom management (KG16-20)	5	0.924	Excellent
<i>Note. * Subdimensions containing the flagged items SK9 (perspective openness) and KG8 (responsive instruction) showed lower internal consistency.</i>			

Descriptive statistics and inter-construct correlations. As a supplementary analysis, the mean scores of the four constructs ranged from 3.84 to 4.07 out of a maximum of 5 (Table 5), indicating a tendency toward high responses (clustering in the agree category). The inter-construct correlations were all positive and significant, ranging from 0.44 to 0.73, with the highest correlation between organizational support and teacher readiness ($r = 0.73$). The moderate magnitude of the correlations, not exceeding 0.85, provides preliminary evidence that the four constructs are interrelated yet remain distinguishable. Because these correlations are based on a small pilot sample and no factor model, they are reported here as supplementary and interpreted cautiously.

Table 5. Descriptive statistics and inter-construct correlation matrix (N = 66)

Construct	M	SD	1	2	3	4
1. Diversity PCK	3.84	0.53	1			
2. Attitudes toward Diversity	4.07	0.44	0.44	1		
3. Organizational Support	4.02	0.54	0.58	0.65	1	
4. Teacher Readiness	4.00	0.48	0.65	0.56	0.73	1
<i>Note. N = 66. All correlations are positive; the highest is between Organizational Support and Teacher Readiness ($r = 0.73$).</i>						

Discussion

Reliability and item quality

Overall, the instrument demonstrated good to excellent internal-consistency reliability ($\alpha = 0.891\text{--}0.934$) and high item validity, with 63 of 65 items meeting the criteria. These findings provide preliminary evidence that the instrument is suitable for measuring elementary school teachers' readiness for multicultural education in social studies (IPS) learning, consistent with the DivePCK framework (Dursun et al., 2021) and the literature on teachers' multicultural attitudes (Abacioglu et al., 2020). Reliability coefficients in this range are consistent with those commonly reported for multidimensional multicultural and culturally responsive teaching instruments, including the DivePCK measure (Dursun et al., 2021), teachers' multicultural attitude scales (Abacioglu et al., 2020), and Indonesian multicultural competence instruments (Fatmawati et al., 2023); in other words, the reliability pattern observed here does not deviate from prior instrument-development work in this field. Nevertheless, the very high alpha coefficients should be interpreted with caution. Alpha is influenced by the number of items, so a value above 0.90 for constructs with 15 to 20 items does not necessarily indicate ideal quality but may instead signal inter-item redundancy (Cortina, 1993; Taber, 2018; Tavakol & Dennick, 2011). Consistent with the parsimony principle in scale development, the instrument could therefore be shortened without meaningful loss of construct coverage (Boateng et al., 2018).

Failure of the negative item (SK9) and the ambiguity of KG8

Two items require special attention, and their verbatim wording is presented here so that readers can judge the source of their weak performance. The single negatively phrased item, SK9, stated, "It is better for teachers to avoid discussion on cultural differences so that the class remains orderly" and was reverse-coded prior to analysis. However, even with the reverse coding, it showed an item-total correlation coefficient approaching zero ($r = 0.038$). Such results are not inconsistent with empirical research which found that the inclusion of both positively and negatively phrased items into one scale decreases discrimination, reliability, and unidimensionality, especially owing to verbal comprehension problems in understanding negatively phrased statements (Suárez-Alvarez et al., 2018). Hence, in this case, the results confirm, rather than contradict, those findings. Indeed, the fact that the elimination of the item increased the reliability of the attitude subscale from 0.891 to 0.925 supports the poor performance of the item. Item KG8 reads "Saya menggunakan bahasa yang sensitif budaya saat berinteraksi dengan siswa" ("I use culturally sensitive language when interacting with students"). Despite its positive wording, it was positioned at the boundary ($r = 0.214$) and had the largest standard deviation ($SD = 1.19$), suggesting inconsistent interpretation on the part of respondents: the expression "culturally sensitive language" (bahasa yang sensitif budaya) is too general and could be understood differently by various teachers, as not using offensive words and using their native language by students, respectively. As such, we suggest modifying SK9 into a definite positive scale and rewriting KG8 in order to express a tangible behavior.

Notably, SK9 also serves as an indicator of poor-quality responses, since careless respondents respond to this negatively worded scale inconsistently (Meade & Craig, 2012); consequently, a controlled negative scale can be used as an independent attention test.

Construct structure and nomological validity

The pattern of inter-construct correlations, all positive and moderate ($r = 0.44\text{--}0.73$), indicates some degree of nomological validity: teacher readiness correlates positively with knowledge (DivePCK), attitudes, and organizational support, as theorized (Aalto et al., 2024; Mou et al., 2025). The highest inter-construct correlation was found between organizational support and teacher readiness ($r = 0.73$), consistent with the culturally responsive leadership approach and empirical findings about the impact of school climate, professional development, and community engagement on responsive practice (Comstock et al., 2023; Qadafi et al., 2024). On the other hand, the low correlation between PCK and attitudes ($r = 0.44$) suggests a certain empirical separation between the cognitive and affective roots of teacher readiness, which was expected based on theory (Abacioglu et al., 2020). Nonetheless, 0.73 is sufficiently high correlation coefficient for discriminant validity to be formally tested between organizational support and teacher readiness, e.g., using heterotrait-monotrait ratio method (Ab Hamid et al., 2017).

How these findings relate to prior research

Taken together, the results both converge with and diverge from previous studies in instructive ways. The convergences are clear: the psychometric failure of the reverse-worded item replicates concerns raised about reversed items in Likert scales (Suárez-Alvarez et al., 2018), the strong organizational-support-to-readiness link echoes culturally responsive leadership research (Comstock et al., 2023; Qadafi et al., 2024), and the overall reliability profile matches that of comparable instruments (Abacioglu et al., 2020; Dursun et al., 2021; Fatmawati et al., 2023). One notable divergence concerns training and self-rated readiness. Prior work associates multicultural professional development and initial teacher education with higher multicultural self-efficacy (Aalto et al., 2024; Mou et al., 2025), yet in this sample 68.2% of respondents had never attended multicultural training while still rating their readiness highly. Rather than contradicting that literature, this tension most plausibly reflects self-report inflation, namely social desirability and overconfidence in the absence of the very training that would calibrate self-assessment, which is why the present findings should be triangulated with observational or performance evidence before strong claims about readiness are made. The proportion of careless responses (16.5%) is likewise broadly in line with rates documented for unsupervised online surveys (Alarcon & Lee, 2022; Meade & Craig, 2012), reinforcing rather than challenging the case for routine attention checks.

High response tendency and potential bias

The scores of the four constructs clustered at the positive end of the scale (84.5% of responses fell in the agree and strongly agree categories; negative skew of -0.39 to -0.65 ; means of $3.84\text{--}4.07$). This tendency may reflect a genuine positive disposition, but it also

opens the possibility of social desirability bias, which is common in self-report measurement. As mentioned above, there is an internal inconsistency within the data set, which lies in the fact that there are a lot of inexperienced teachers, who, on the other hand, assess themselves as being well-prepared to teach. The consequence is that the future use of the instrument needs to triangulate self-assessment with observations and performance measures.

Careless responding in the online survey

The significant amount of carelessly answered questions (16.5%) represents an important methodology lesson for the future online surveys conducted in Indonesia. Otherwise, such answers will increase the reliability measures, making them higher than the real one (Alarcon & Lee, 2022; Meade & Craig, 2012). When collecting data in unsupervised online surveys, it is recommended to include attention-check questions, time constraints, and screening questions into the questionnaire to protect the integrity of collected data.

Theoretical implications

As a result, the present study provides the first adaptation of the DivePCK framework in relation to learning IPS in elementary schools in Indonesia, as well as expanding the idea of pedagogical content knowledge to the area of diversity. By integrating cognitive, affective, and contextual dimensions within one instrument and adapting it to the concepts of Bhinneka Tunggal Ika, the Pancasila Student Profile, and the IPS Learning Outcomes, the present instrument brings a new empirical component into the existing national literature review on this topic which consists mostly of normative studies and reviews (Fatmawati et al., 2023; Jayadi et al., 2022; Ladson-Billings, 1995; Suri & Chandra, 2021).

Practical implications

After revision, the instrument can be used as a teacher self-reflection tool and a diagnostic instrument for mapping professional-development needs; the high proportion of teachers who have never been trained (68.2%) indicates a real training gap. Since organizational support is the most robust predictor of teacher preparedness, it makes sense that principals and policy makers focus on fostering the enabling environment in the schools, including policies, training, and climate (Chahar Mahali & Sevigny, 2022; Comstock et al., 2023; Qadafi et al., 2024).

Limitations

Several limitations must be stated openly. First, the sample was small ($N = 66$), non-probability, and geographically heterogeneous (three regions), so the findings cannot be generalized and the respondent-to-item ratio was insufficient for exploratory or confirmatory factor analysis; consequently, evidence of structural, convergent, and discriminant validity could not be established in this study, which is why it is framed as a pilot. Second, the data came from self-reports and are therefore susceptible to social desirability bias and ceiling effects that limit variance. Third, the sample composition was less than ideal because 34.8% of respondents did not specifically teach IPS and 68.2% had

never attended multicultural training. Fourth, the cross-sectional design did not allow testing of test-retest reliability or measurement invariance. Fifth, the very high alpha indicates possible item redundancy. Accordingly, all findings are preliminary.

Directions for future research

Future research should collect data on a larger and more representative sample through probability sampling and then subject the instrument to full factor-analytic validation. A minimum sample size of 300 respondents per analysis is recommended based on psychometric sampling criteria, and not by choice: for instruments of this length with the expected communality values, samples that are close to 300 are considered to be good and sufficient for reliable exploratory and confirmatory analysis, while smaller samples result in unstable loadings and increased variability in the fit of the model (Boateng et al., 2018; Kyriazos, 2018). Because the same respondents should not be used both to explore and to confirm a structure, we recommend gathering a larger pooled sample (for example, $N \approx 600$) and randomly splitting it into two independent subsamples: one ($n \approx 300$) for exploratory factor analysis to examine the emergent structure, and a second, independent subsample ($n \approx 300$) for confirmatory factor analysis to test that structure and estimate model fit (Kyriazos, 2018). Confirmatory analysis should be complemented by tests of convergent validity (average variance extracted and composite reliability) and discriminant validity (heterotrait-monotrait; Ab Hamid et al., 2017). Testing measurement invariance across regions, certification status, and training experience, together with test-retest reliability, item shortening into a brief form, and validation against behavioral criteria such as classroom observation, is also recommended. Ultimately, a robust instrument can be used to test structural models of the relationships among knowledge, attitudes, organizational support, and teacher readiness.

Conclusion

The instrument for elementary school teachers' readiness to implement multicultural education in social studies (IPS) learning demonstrated good to excellent internal-consistency reliability ($\alpha = 0.891\text{--}0.934$) and high item validity (63 of 65 items valid). Two items, SK9 and KG8, are recommended for revision or removal. As a preliminary (pilot) study, these findings provide promising initial evidence but do not yet establish the validity of the factor structure. The recommended next step is to collect data on a larger and more representative sample through probability sampling, followed by exploratory and confirmatory factor analysis on independent subsamples, together with tests of convergent validity (average variance extracted, composite reliability) and discriminant validity (HTMT), as well as test-retest reliability. After revision, this instrument is suitable for use as a tool for reflecting on teacher competence and as a basis for further research on multicultural education in elementary schools.

Declarations

Author contribution statement

The first author designed the study, developed the instrument, led the data analysis, and drafted the manuscript. The second and third authors contributed to instrument development, coordination of the expert judgment, and editing and refinement of the manuscript. All authors approved the final version of the manuscript.

Funding statement

This research was funded by an internal research grant from Universitas Negeri Malang in 2026 (Research Decentralization Scheme, Faculty of Education/FIP).

Data availability statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Declaration of interests statement

The authors declare that they have no financial or personal conflicts of interest that could have influenced the results of this study.

References

- Aalto, S., Kankaanpää, R., Peltonen, K., Derluyn, I., Szelei, N., Verelst, A., De Haene, L., De Smet, S., Spaas, C., Jervelund, S. S., Skovdal, M., Andersen, A. J., Hilden, P. K., Opaas, M., Durbeej, N., Osman, F., Sarkadi, A., Soye, E., & Vänskä, M. (2024). The effect of teacher multicultural attitudes on self-efficacy and wellbeing at work. *Social Psychology of Education*, 27(5), 2527–2557. <https://doi.org/10.1007/s11218-024-09886-3>
- Ab Hamid, M. R., Sami, W., & Sidek, M. H. M. (2017). Discriminant validity assessment: Use of Fornell & Larcker criterion versus HTMT criterion. *Journal of Physics: Conference Series*, 890, 012163. <https://doi.org/10.1088/1742-6596/890/1/012163>
- Abacioglu, C. S., Volman, M., & Fischer, A. H. (2020). Teachers' multicultural attitudes and perspective taking abilities as factors in culturally responsive teaching. *British Journal of Educational Psychology*, 90(3), 736–752. <https://doi.org/10.1111/bjep.12328>
- Alarcon, G. M., & Lee, M. A. (2022). The relationship of insufficient effort responding and response styles: An online experiment. *Frontiers in Psychology*, 12, 784375. <https://doi.org/10.3389/fpsyg.2021.784375>
- Boateng, G. O., Neilands, T. B., Frongillo, E. A., Melgar-Quinonez, H. R., & Young, S. L. (2018). Best practices for developing and validating scales for health, social, and behavioral research: A primer. *Frontiers in Public Health*, 6, 149. <https://doi.org/10.3389/fpubh.2018.00149>
- Bolarinwa, O. A. (2015). Principles and methods of validity and reliability testing of questionnaires used in social and health science researches. *Nigerian Postgraduate Medical Journal*, 22(4), 195–201. <https://doi.org/10.4103/1117-1936.173959>
- Chahar Mahali, S., & Sevigny, P. R. (2022). Multicultural classrooms: Culturally responsive teaching self-efficacy among a sample of Canadian preservice teachers. *Education and Urban Society*, 54(8), 946–968. <https://doi.org/10.1177/00131245211062526>
- Comstock, M., Litke, E., Hill, K. L., & Desimone, L. M. (2023). A culturally responsive disposition: How professional learning and teachers' beliefs about and self-efficacy

- for culturally responsive teaching relate to instruction. *AERA Open*, 9. <https://doi.org/10.1177/23328584221140092>
- Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology*, 78(1), 98–104. <https://doi.org/10.1037/0021-9010.78.1.98>
- Dursun, H., Claes, E., & Agirdag, O. (2021). Diversity pedagogical content knowledge: A new conceptual framework and assessment across different teacher education programmes. *Multicultural Education Review*, 13(4), 303–322. <https://doi.org/10.1080/2005615X.2021.2006118>
- Fatmawati, L., Dewi, K. P., & Wuryandani, W. (2023). Multicultural competence of elementary teacher education students. *International Journal of Elementary Education*, 7(4), 721–730. <https://doi.org/10.23887/ijee.v7i4.62880>
- Jayadi, K., Abduh, A., & Basri, M. (2022). A meta-analysis of multicultural education paradigm in Indonesia. *Heliyon*, 8(1), e08828. <https://doi.org/10.1016/j.heliyon.2022.e08828>
- Ladson-Billings, G. (1995). Toward a theory of culturally relevant pedagogy. *American Educational Research Journal*, 32(3), 465–491. <https://doi.org/10.3102/00028312032003465>
- Meade, A. W., & Craig, S. B. (2012). Identifying careless responses in survey data. *Psychological Methods*, 17(3), 437–455. <https://doi.org/10.1037/a0028085>
- Mou, Q., Dursun, H., & Agirdag, O. (2025). A cross-national examination of teachers' multicultural self-efficacy: Can multicultural education in initial teacher education and professional development make a difference? *Education and Urban Society*. <https://doi.org/10.1177/00131245241278673>
- Qadafi, M., Alfian Sidik, M., & Mukhtar, Z. (2024). Community tolerance, habitus, and early childhood education in schools: A phenomenological approach. *Al-Athfal: Jurnal Pendidikan Anak*, 10(2), 165–178. <https://doi.org/10.14421/al-athfal.2024.102-06>
- Souza, A. C. de, Alexandre, N. M. C., & Guirardello, E. de B. (2017). Psychometric properties in instruments evaluation of reliability and validity. *Epidemiologia e Serviços de Saúde*, 26(3), 649–659. <https://doi.org/10.5123/S1679-49742017000300022>
- Suárez-Alvarez, J., Pedrosa, I., Lozano, L. M., García-Cueto, E., Cuesta, M., & Muñoz, J. (2018). Using reversed items in Likert scales: A questionable practice. *Psicothema*, 30(2), 149–158. <https://doi.org/10.7334/psicothema2018.33>
- Subhan, S., Puad, L. M. A. Z., & Listyowati, A. (2026). Assessing scientific thinking in early childhood: Cross-sectional evidence for a six-dimensional hierarchical structure in Indonesian preschoolers. *Al-Athfal: Jurnal Pendidikan Anak*, 12(1), 1–18. <https://doi.org/10.14421/al-athfal.2026.121-01>
- Suri, D., & Chandra, D. (2021). Teacher's strategy for implementing multiculturalism education based on local cultural values and character building for early childhood education. *Journal of Ethnic and Cultural Studies*, 8(4), 271–285. <https://doi.org/10.29333/ejecs/937>

- Sürücü, L., & Maslakçı, A. (2020). Validity and reliability in quantitative research. *Business & Management Studies: An International Journal*, 8(3), 2694–2726. <https://doi.org/10.15295/bmij.v8i3.1540>
- Suwela Antara, I. G. W., Yoga Swadnyana, I. B., Yudiana, K., Piki Pia Arini, N. P., Lasmawan, I. W., & Kertih, I. W. (2024). Integrating character education and social studies in elementary school: Prospective professional teachers' experiences. *Indonesian Values and Character Education Journal*, 7(1), 101–109. <https://doi.org/10.23887/ivcej.v7i1.81005>
- Taber, K. S. (2018). The use of Cronbach's alpha when developing and reporting research instruments in science education. *Research in Science Education*, 48(6), 1273–1296. <https://doi.org/10.1007/s11165-016-9602-2>
- Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education*, 2, 53–55. <https://doi.org/10.5116/ijme.4dfb.8dfd>
- 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>
- Zamanzadeh, V., Ghahramanian, A., Rassouli, M., Abbaszadeh, A., Alavi-Majd, H., & Nikanfar, A.-R. (2015). Design and implementation content validity study: Development of an instrument for measuring patient-centered communication. *Journal of Caring Sciences*, 4(2), 165–178. <https://doi.org/10.15171/jcs.2015.017>