Received: 8 November 2025, Accepted: 7 October 2026,
Published on-line as Recently Accepted Paper: October 2026

DOI: https://doi.org/10.26529/cepsj.2371

Adaptation and Validation of the Literature Teaching Questionnaire Among Albanian Language and Literature and English as a Foreign Language Teachers in North Macedonia

Jehona Rushidi-Rexhepi[1], Kushtrim Ahmeti[2], Agon Memeti[3], Ibrahim Neziri[2] and Bujar Adili*[4]

Teachers' beliefs about the pedagogical value of literary texts shape whether that value is realised in the classroom, yet there was previously no validated instrument to measure these beliefs for Albanian Language and Literature or English as a Foreign Language teachers in North Macedonia. The present study adapted the Literature Teaching Questionnaire developed by Duncan and Paran (2017) and tested its four-factor structure among 224 teachers using confirmatory factor analysis. Starting from the pilot's 29-item structure, items with the weakest standardised loadings were removed one at a time in a uniform, pre-specified procedure, yielding an 18-item, four-factor model (Text Relevance, Linguistic Fit, Teacher Practice, Student Outcomes) with good fit and satisfactory reliability across subscales and overall. Measurement invariance testing across Albanian Language and Literature and English as a Foreign Language teachers supported configural and metric invariance (equivalent structure and loadings) but not scalar invariance (differing item intercepts). These findings support the Literature Teaching Questionnaire as a psychometrically sound instrument for this context.

Keywords: factor analysis, instrument validation, language development, literature teaching, teacher perceptions

1 Faculty of Pedagogy, University of Tetova, North Macedonia.
2 Faculty of Philosophy, University of Tetova, North Macedonia.
3 Faculty of Natural Sciences and Mathematics, University of Tetova, North Macedonia.
4 *Corresponding Author. Faculty of Pedagogy, University of Tetova, North Macedonia; bujar.adili@unite.edu.mk.

Prilagoditev in validacija vprašalnika o poučevanju književnosti med učitelji albanskega jezika in književnosti ter angleščine kot tujega jezika v Severni Makedoniji

Jehona Rushidi-Rexhepi, Kushtrim Ahmeti, Agon Memeti, Ibrahim Neziri in Bujar Adili

Prepričanja učiteljev o pedagoški vrednosti literarnih besedil vplivajo na to, ali se ta vrednost v celoti uresničuje tudi v učilnici, vendar do zdaj ni bilo validiranega instrumenta za merjenje teh prepričanj pri učiteljih albanskega jezika in književnosti ter angleščine kot tujega jezika v Severni Makedoniji. V tej študiji smo prilagodili vprašalnik o poučevanju književnosti, ki sta ga razvila Duncan in Paran leta 2017, ter s potrditveno faktorsko analizo preverili njegovo štirifaktorsko strukturo na vzorcu 224 učiteljev. Na podlagi strukture pilotne raziskave z 29 postavkami so bile postavke z najšibkejšimi standardiziranimi faktorskimi utežmi druga za drugo izločene skladno z enotnim, vnaprej določenim postopkom, tako da je nastal štirifaktorski model (relevantnost besedila, jezikovna ustreznost, praksa učitelja, dosežki učencev) z 18 postavkami, ki se je dobro prilegal podatkom in imel zadovoljivo zanesljivost na podlestvicah in v celoti. Preizkušanje merilne invariantnosti med učitelji albanskega jezika in književnosti ter angleščine kot tujega jezika je potrdilo konfiguracijsko in metrično invariantnost (primerljiva struktura in faktorske uteži), ne pa tudi skalarne invariantnosti (različne konstante postavk). Ti izsledki potrjujejo, da je vprašalnik o poučevanju književnosti v tem kontekstu psihometrično zanesljiv instrument.

Ključne besede: faktorska analiza, validacija instrumenta, jezikovni razvoj, poučevanje književnosti, zaznave učiteljev

Introduction

Language teachers' beliefs about how literary texts contribute to language development are not a peripheral detail of literacy instruction; they are a primary mechanism through which curricular potential becomes classroom practice. Teacher cognition research has long established that teachers' beliefs about a particular resource shapes whether and how they use that resource, often more than formal training (Borg, 2003; Phipps & Borg, 2009). The present study is concerned specifically with that belief system, investigating how Albanian Language and Literature (ALL) and English as a Foreign Language (EFL) teachers in North Macedonia perceive the linguistic, motivational and instructional value of literary texts, and whether that perception can be measured with a psychometrically sound instrument.

Literary texts are widely credited with three kinds of benefit: advancing vocabulary depth, syntactic awareness and discourse competence beyond the level achieved by grammar-focused instruction (Gottardo et al., 2018; Hedgcock & Ferris, 2018); sustaining higher-order skills, such as inference and critical evaluation (Bobkina & Stefanova, 2016; Mart, 2018); and fostering affective and intercultural growth through engagement with varied perspectives (Byram, 2021). These claims, however, describe the potential contribution of literature; whether that potential is realised depends on teachers, and specifically on how teachers understand and enact the role of literature in their own classrooms (Cunningham & Farmer, 2016; Li, 2020). This is the point at which measurement becomes necessary: without a reliable way of capturing teachers' perceptions, claims about the pedagogical value of literature remain disconnected from the practitioners responsible for delivering literary texts.

Existing efforts to measure this construct are limited. Duncan and Paran (2017) have developed the most directly relevant instrument, assessing UK secondary English teachers' beliefs about the contribution of literature to language skills and intercultural understanding across 17 thematic batteries, with reliability ranging from poor to excellent depending on the battery. Their work has established that teacher beliefs about literature are multidimensional rather than a single global attitude. This is consistent with broader practice in teacher cognition research, where constructs such as teacher self-efficacy and beliefs about curriculum are typically modelled and validated as multi-factor structures using confirmatory factor analysis (Boateng et al., 2018; DeVellis & Thorpe, 2021; Ploj Virtič et al., 2023). However, no equivalent instrument has been validated for Albanian-medium or EFL teacher populations, despite North Macedonia's linguistically and institutionally distinct schooling context. Duncan and Paran's

own instrument was built and normed on a UK sample, under different curricular structures, teacher training pathways and language-status conditions (Albanian as a minority-language medium of instruction alongside Macedonian, versus English taught as a foreign language rather than the first language). Adapting the instrument to this context, and re-establishing its factor structure and reliability empirically rather than assuming transferability, is therefore a necessary step before it can be used in regional research or policy work. This step is undertaken by the present study.

The rationale for examining ALL and EFL teachers as distinct groups instead of treating "language teachers" as a single population has empirical grounding rather than being an incidental sampling choice. Prior research indicates that teachers' disciplinary training shapes what they look for in a text: EFL teachers more often foreground functional language use, while first-language teachers foreground literary and interpretive analysis, even when working with comparable materials (Paran, 2008). If this distinction holds, ALL and EFL teachers may not only score differently on a literature-perception measure but may relate to its underlying factors differently. This is a measurement-invariance question, not simply a mean-difference question. The present study is positioned to test this question given that both groups of teachers are represented in comparable numbers in the sample.

Two structural obstacles have kept this gap open. First, teachers' engagement with literary texts is constrained in practice by time pressures, curricular emphasis on measurable competencies over interpretive work, and inconsistent training in literature-based pedagogy (Duncan & Paran, 2018; Levine et al., 2021; Mart, 2018). Teachers routinely adapt or select materials to bridge this gap (Rathert & Cabaroğlu, 2022). The Teacher Practice subscale used in the present study is designed to capture this practice-level behaviour as its own dimension, rather than folding it into general attitude items. Second, the absence of a validated regional instrument means that these constraints, and any differences in how ALL and EFL teachers experience them, have not previously been measurable at all in this context.

These gaps form a single chain: the benefits of literature for language development are well documented, but they are contingent on teacher belief and practice; teacher belief and practice differ plausibly by disciplinary group; and no validated instrument exists to measure any of this for Albanian-medium or EFL teachers in North Macedonia. Addressing this gap requires an instrument with demonstrated structural validity, not an assumed or borrowed instrument.

Purpose of the Study

The purpose of the present study is to adapt the Literature Teaching Questionnaire (LTQ) for ALL and EFL teachers in North Macedonia and to establish its psychometric properties in this context. Specifically, the study:

1. tests the four-factor structure identified in the pilot study using confirmatory factor analysis on the full sample;
2. refines this structure through a transparent, uniform item-trimming procedure where model fit is inadequate;
3. evaluates the reliability of the refined structure and its subscales;
4. examines measurement invariance of the refined structure across ALL and EFL teacher groups.

Method

Participants

Since no official data exist on the number of ALL and EFL teachers in schools where instruction is conducted in Albanian, estimates were derived using enrolment figures and instructional models. Based on an average class size of 18 students and applying a homeroom model for grades 1–5 and a specialist model for grades 6–12, the teacher population was projected to include approximately 973 ALL teachers and 603 EFL teachers, or about 1,576 teachers in total (State Statistical Office of the Republic of North Macedonia, 2024). Using this estimate as the population frame ($N \approx 1{,}576$), a sample size of 224 participants corresponds to a 95% confidence level with a sampling error of approximately ±6.1%, assuming maximum variability ($p = q = 0.5$) (Lohr, 2021; Scheaffer et al., 2011).

The final sample comprised 224 teachers: 108 Albanian L1 and 116 EFL. Of these, 189 were female and 35 male. Regarding teaching experience, 95 teachers had more than 20 years, while 129 had less than 20 years. In terms of qualifications, 56 held MA or PhD degrees, while 168 held bachelor's degrees. The participants represented both rural ($n = 98$) and urban ($n = 126$) school settings. By school level, 170 teachers worked in primary schools and 54 in secondary schools. With respect to professional development, 90 participants had received formal training in literature teaching, while 134 had not.

Instruments

The LTQ used in the present study was adapted from the instrument developed by Duncan and Paran (2017) to assess secondary English teachers' beliefs about the role of literature in language and intercultural development. The original instrument organised items into four conceptual domains: language suitability and challenge, the content and literary value of texts, teachers' practice and confidence in using literary material, and teachers' beliefs about the outcomes literature produces for students. These domains correspond, in adapted form, to the present study's four subscales (Linguistic Fit, Text Relevance, Teacher Practice and Student Outcomes), retaining Duncan and Paran's underlying conceptual structure while reflecting the linguistic and curricular adjustments described below. From an initial pool of 35 items, 32 were selected following consultation with two linguistic experts and three additional experts in education and assessment. At this stage, three items were identified as falling outside the instrument's four core conceptual domains and therefore removed: "I consider the visual presentation and layout of a literary text when selecting it for classroom use", "I prefer literary texts that are available in both printed and digital formats" and "I consider the availability of supplementary teaching materials when selecting a literary text".

In order to ensure clarity and reliability prior to the main-study administration, a pilot test was conducted with 30 teachers of ALL and EFL from primary and secondary schools (Rushidi-Rexhepi et al., 2025). The pilot aimed to evaluate item comprehensibility, response time and internal consistency. Feedback and reliability diagnostics informed the revision of several items and the removal of three items that performed weakly: "I find it difficult to find new literary texts to teach", "I prefer to use texts I have already used successfully in the past" and "Students should be challenged by the language used in the literary text". These items showed the weakest corrected item-total correlations in the pilot, resulting in the 29-item version of the LTQ. All of the pilot participants were excluded from the main study to avoid duplication and potential bias (DeVellis & Thorpe, 2021).

The LTQ was translated from English into Albanian following forward–backward translation procedures, expert review and pilot testing to ensure linguistic and cultural equivalence. Previous validations of translated scales demonstrate that careful translation, expert review and factor-analytic testing provide a sound basis for cross-cultural adaptation (Benić, 2021). Expert validation further strengthened content validity: one expert from the Albanian Language and Literature Department and one statistics expert reviewed clarity and expression,

while an education supervisor, four experts in educational administration and two lecturers in educational programmes assessed item content. The content validity ratio (CVR) of each item was calculated using Lawshe's technique, with all items exceeding the .99 minimum CVR threshold, indicating strong agreement among experts regarding item relevance (Ayre & Scally, 2014).

All items were measured on a 6-point Likert scale ranging from 1 (strongly disagree) to 6 (strongly agree); the forced-choice format, without a neutral midpoint, was adopted to reduce central tendency and social desirability bias (Lacko et al., 2022).

The items covered domains related to teachers' perceptions of literature in language education. Pilot testing confirmed a four-factor solution: Student Outcomes, Text Relevance, Linguistic Fit and Teacher Practice, with acceptable internal consistency (Cronbach's α = .728–.772; Taber, 2018).

Following the pilot, the 29-item instrument was administered to the main study sample. As detailed in the Results section, the four-factor structure was tested using confirmatory factor analysis. Eleven items were subsequently removed through an iterative trimming procedure to improve model fit, consistent with the criteria specified below. The resulting 18-item version retained items from all four original subscales.

Procedure

Data were collected between August and September 2025 using a standardised online procedure to maintain consistency across respondents. The administration period was intentionally chosen to avoid examination schedules and end-of-year administrative duties that could influence teachers' availability or responses. In order to minimise potential external influences, all of the participants received identical instructions and questionnaire versions, and no institutional training initiatives on literature teaching were reported during the data collection period. These measures enhanced the comparability and internal validity of the responses (Cohen et al., 2018).

Responses were gathered in a single wave via an online questionnaire distributed through professional groups and associations, using a snowball sampling approach appropriate for the relatively small population of ALL and EFL teachers in North Macedonia (fewer than 2,000). The inclusion criteria required participants to be actively teaching in Albanian-language schools, while the pilot study participants were excluded to avoid duplication. Data collection followed standard ethical procedures, with participants informed of the study's purpose, voluntary participation and the confidentiality of responses prior to completion.

Data Analysis

The data were analysed using SPSS v.25 and AMOS v.23. The four-factor structure identified in the pilot study was tested using confirmatory factor analysis (CFA) on the full sample of 224 teachers, with maximum likelihood estimation.

Model fit was evaluated using multiple indices, as reliance on a single index is discouraged (Kline, 2016). The following thresholds were used as benchmarks for acceptable fit: $\chi^2/df < 3$, RMSEA and SRMR < .08, Comparative Fit Index (CFI) > .90, and Tucker–Lewis Index (TLI) > .90 (Byrne, 2016; Hu & Bentler, 1999).

Where the initial model fit was inadequate, items were removed one at a time: at each step, the item with the weakest standardised loading was identified and removed, and the model was re-estimated, continuing until fit thresholds were met. This procedure was applied uniformly, independent of which specific items were affected. One residual covariance (between Q8SO and Q9SO) was subsequently added based on modification indices and the items' similar wording and shared thematic content, consistent with recommended practice for justifying correlated residuals on substantive rather than purely statistical grounds (Byrne, 2016).

Measurement invariance across teacher group (ALL vs. EFL) was tested using a sequence of nested multi-group confirmatory factor models in AMOS: configural (no equality constraints), metric (factor loadings constrained equal), scalar (item intercepts additionally constrained) and strict (residual variances additionally constrained) (Putnick & Bornstein, 2016). Each more constrained model was compared to the preceding less constrained model using the chi-square difference test and change in CFI, with $\Delta CFI \geq .01$ taken as an indicator of non-invariance, following recommended practice (Cheung & Rensvold, 2002).

The internal consistency reliability of the LTQ was assessed using Cronbach's alpha coefficients for each subscale and for the overall scale, with values above .70 interpreted as satisfactory (Taber, 2018).

Results

Confirmatory Factor Analysis

The four-factor structure identified in the pilot study was tested via CFA on the full sample of 224 teachers. The initial 29-item model showed inadequate fit (CFI = .712, TLI = .684, RMSEA = .092, SRMR = .089).

Following the procedure described in the Data Analysis section, items were removed one at a time based on standardised loading strength, with the model re-estimated at each step (Table 1). After eleven items were removed, the resulting 18-item model reached acceptable fit (CFI = .902, TLI = .884, RMSEA = .069, SRMR = .061).

Table 1
Item Removal Sequence

Step	Item removed	Factor	Loading	Content
1	Q16TP	Teacher Practice	.153	lack of training/guidance
2	Q5TR	Text Relevance	.419	supports critical thinking
3	Q4TR	Text Relevance	.398	age-appropriate themes
4	Q12LF	Linguistic Fit	.429	reflects current usage
5	Q1SO	Student Outcomes	.463	improves vocabulary
6	Q24LF	Linguistic Fit	.495	shorter texts preferred
7	Q6LF	Linguistic Fit	.503	vocabulary within ability
8	Q10TR	Text Relevance	.519	cultural comparison
9	Q19LF	Linguistic Fit	.520	grammar within ability
10	Q20TR	Text Relevance	.509	relevance to students' lives
11	Q17TR	Text Relevance	.497	suitable for discussion

One residual covariance, between Q8SO and Q9SO, was subsequently added based on modification indices and the substantive similarity of the two items' wording and content (see Data Analysis). This improved model fit further: $\chi^2 = 247.22$, $df = 128$, $p < .001$, $\chi^2/df = 1.93$, CFI = .914, TLI = .898, RMSEA = .065 (90% CI [.052, .077]), SRMR = .061, GFI = .893, AGFI = .857. All fit indices met or approached conventional thresholds (Hu & Bentler, 1999; Byrne, 2016). The final 18-item, four-factor structure comprised Text Relevance, Linguistic Fit, Teacher Practice and Student Outcomes. Table 2 presents the standardised loadings for the final 18-item model.

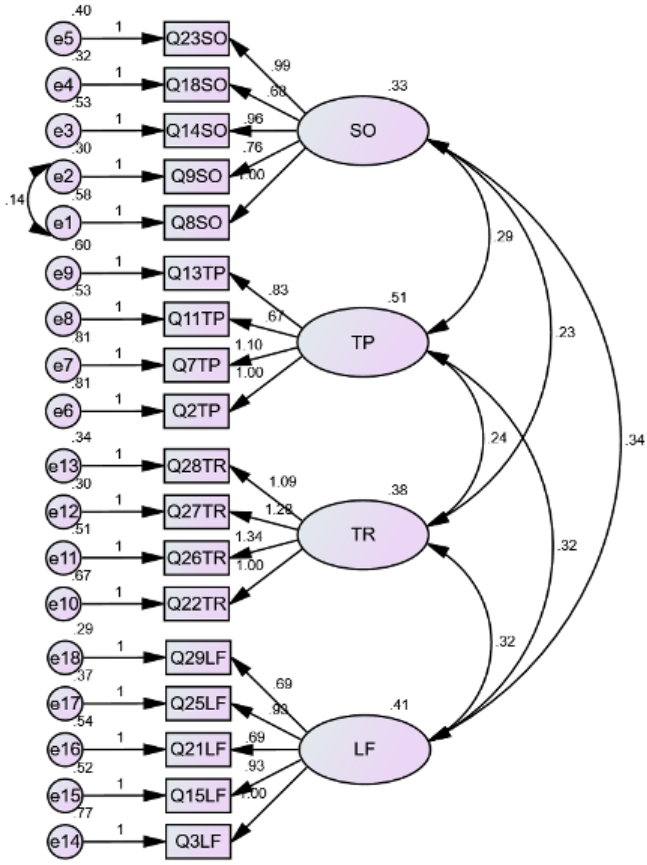
Table 2

Standardised Loadings for the Final 18-Item Model

Factor	Item	Description	Loading
Linguistic Fit	Q3LF	new grammar	.591
	Q15LF	teaching a language point	.638
	Q21LF	overall linguistic fit	.517
	Q25LF	linguistic challenge	.699
	Q29LF	specific language skills	.636
Text Relevance	Q22TR	global/social issues	.600
	Q26TR	literary quality	.756
	Q27TR	aesthetic/emotional value	.820
	Q28TR	national/world literature	.752
Teacher Practice	Q2TP	ease of integration	.623
	Q7TP	sufficient time	.660
	Q11TP	enjoyment	.550
	Q13TP	confidence with complex texts	.610
Student Outcomes	Q8SO	creative language use	.604
	Q9SO	multiple language skills	.624
	Q14SO	intercultural awareness	.602
	Q18SO	new vocabulary	.570
	Q23SO	motivation	.669

Note. All loadings significant at $p < .001$.

Figure 1
CFA Results for the LTQ



Reliability of the LTQ

Internal consistency was satisfactory across all four subscales and the overall scale. Compared to the pilot study (Rushidi-Rexhepi et al., 2025), the current results demonstrate improved reliability for the Text Relevance and Linguistic Fit subscales and for the overall scale, whereas the Teacher Practice and Student Outcomes subscales showed slightly lower coefficients.

Table 3

Reliability Coefficients (Cronbach's α) for the 18-Item LTQ

Subscale / Scale	Pilot α	Current study α
Text Relevance	.736	.817
Linguistic Fit	.730	.746
Teacher Practice	.728	.701
Student Outcomes	.772	.767
LTQ	.878	.891

Measurement Invariance

Measurement invariance across teacher group (ALL vs. EFL) was examined on the final 18-item model. The configural model showed acceptable fit (χ^2 = 423.71, df = 256, CFI = .881, TLI = .857, RMSEA = .054), supporting equivalence of the overall factor structure across groups. Constraining factor loadings (metric invariance) did not significantly worsen fit ($\Delta\chi^2$ = 13.28, Δdf = 14, p = .504; ΔCFI < .001), supporting metric invariance. Constraining item intercepts (scalar invariance) produced a significant decline in fit by both criteria ($\Delta\chi^2$ = 51.18, Δdf = 18, p < .001; ΔCFI = .024), indicating that scalar invariance was not supported. Constraining residual variances (strict invariance) likewise showed a significant decline ($\Delta\chi^2$ = 65.61, Δdf = 29, p < .001; ΔCFI = .026). These results indicate that the LTQ measures the same four constructs in the same way for ALL and EFL teachers, although direct comparison of raw subscale means between groups should be made with caution.

Table 4

Fit Indices for Measurement Invariance Models

Model	χ^2 (*df*)	CFI	TLI	RMSEA [90% CI]	$\Delta\chi^2$ (Δdf)	ΔCFI
Configural	423.714 (256)	.881	.857	.054	—	—
Metric	436.997 (270)	.881	.865	.053	13.283 (14), p = .504	.000
Scalar	488.173 (288)	.857	.848	.056	51.177 (18), p < .001	.024
Strict	553.786 (317)	.831	.837	.058	65.613 (29), p < .001	.026

Note. Measurement invariance tested across teacher group (ALL vs. EFL). CFI = Comparative Fit Index; TLI = Tucker–Lewis Index; RMSEA = Root Mean Square Error of Approximation; CI = confidence interval. $\Delta\chi^2$ and ΔCFI are compared with the previous, less constrained model.

Discussion and Conclusion

The present study examined the factor structure, reliability and cross-group measurement equivalence of the LTQ as adapted for ALL and EFL teachers in North Macedonia. Confirmatory factor analysis of the four-factor structure identified in the pilot study, following a transparent item-trimming procedure, produced an 18-item model with good fit and satisfactory reliability. The instrument also showed equivalent factor structure and item loadings, but not equivalent item intercepts, across the two teacher groups. All of these findings indicate that the LTQ is a psychometrically usable instrument for measuring literature-related teacher perceptions in this context, while also pointing to notable overlap between some of its four dimensions, which warrants attention in future validation work.

Interpretation of Findings

These results speak directly to the gap identified at the outset: the contribution of literature to language development is well documented, but realising this contribution depends on teachers' own beliefs and practices, and there was previously no validated instrument to measure these beliefs for Albanian-medium or EFL teachers. The four confirmed factors map onto recognisable components of this belief system. Text Relevance and Linguistic Fit reflect how teachers judge a text's suitability; Teacher Practice reflects confidence and instructional strategy; and Student Outcomes reflects perceived benefit to learners. This is consistent with the multidimensional view of teacher beliefs established by Duncan and Paran (2017) as well as with the broader teacher-cognition literature (Borg, 2003; Cunningham & Farmer, 2016).

Eleven items were removed through an iterative trimming procedure guided by standardised loading strength (see Method), and one residual covariance was retained between two Student Outcomes items sharing closely related wording. Even after this refinement, the Linguistic Fit factor correlated strongly with both Text Relevance (r = .815) and Student Outcomes (r = .915), suggesting that these dimensions, while statistically distinguishable, may not be fully independent in teachers' actual perceptions. One plausible explanation is cultural and curricular rather than purely psychometric: judgments about a text's linguistic accessibility, its thematic relevance and its value for students may be less separable for teachers working across two languages of instruction in North Macedonia's multilingual system than for the monolingual UK context in which the original instrument was built. This is an interpretive hypothesis,

not a confirmed finding, and it warrants future qualitative work rather than resolution from this dataset alone.

The measurement invariance results add a further, convergent piece of evidence at the same seam. ALL and EFL teachers showed equivalent factor structure and loadings (configural and metric invariance held; $\Delta CFI < .001$), indicating that the LTQ measures the same four constructs in the same way regardless of teaching subject. Item intercepts, however, were not equivalent across the two groups (scalar invariance did not hold; $\Delta CFI = .024$), meaning that at least some items carry a different baseline response tendency for ALL versus EFL teachers even at equal levels of the underlying trait. Practically, this means that the LTQ can be used with confidence to examine the relationships between Text Relevance, Linguistic Fit, Teacher Practice and Student Outcomes for either group, but direct comparison of raw subscale means between ALL and EFL teachers should be made cautiously, or supplemented with item-level scrutiny, until partial invariance is established in future work with a larger sample.

Beyond these structural questions, the substantive pattern of responses aligns with prior research on literature in language teaching. Higher endorsement of items concerning text accessibility and relevance is consistent with findings that accessible, culturally relevant texts sustain learner engagement (Calafato & Hunstadbråten, 2024), while the comparatively more moderate Teacher Practice scores echo persistent reports of limited training and curricular support for literature-based instruction (Levine et al., 2021; Paran, 2008). The preference implied by high Text Relevance and Linguistic Fit endorsement is also broadly consistent with evidence that printed, carefully selected literary texts support deeper comprehension than less curated digital material (Kovač & Kepic Mohar, 2022), and with findings that regionally produced EFL textbooks favour culturally proximate authentic texts over imported ones (Milković et al., 2025). Both of these findings align with the same emphasis on cultural and linguistic fit that the present study's factor structure was built to capture.

Limitations

Several limitations qualify these findings. The wider population of ALL and EFL teachers in North Macedonia, from which the sample was drawn, is itself small, estimated at approximately 1,576. Although the sample size of 224 respondents meets standard adequacy thresholds for confirmatory factor analysis, the item-trimming procedure used to reach the final 18-item model was conducted on this same sample; while this is standard practice in confirmatory

respecification (Byrne, 2016), the resulting model has not yet been tested on independent data, and some fit improvement may reflect sample-specific characteristics rather than a fully generalisable structure. The configural invariance model also showed only modest fit (CFI = .881), consistent with the reduced sample size within each teacher group (n = 108 and n = 116) relative to model complexity. Teachers from private schools were not represented, and reliance on self-report leaves the instrument subject to the usual social-desirability considerations associated with attitude measures. Replication with an independent sample remains the more definitive test of whether the 18-item structure, and particularly the elevated correlations between Linguistic Fit, Text Relevance and Student Outcomes, reflect a stable and generalisable structure.

Implications for Practice and Research

For research, the LTQ offers a validated basis for examining literature-related teacher beliefs among ALL and EFL teachers in North Macedonia, and for comparative or longitudinal work once the overlap between Linguistic Fit, Text Relevance and Student Outcomes is better understood. For practice, the moderate Teacher Practice scores reinforce calls for teachers to be supported as "critical and informed textbook users with the perspective to develop their own materials" (Rathert & Cabaroğlu, 2022, p. 185), rather than being left to adapt literary texts without guidance. This could be paired with institutional efforts towards equitable, non-discriminatory classroom practice around the texts literature raises (Sa'd & Eames, 2021).

Conclusion

The present study provides an adapted and empirically validated measure of ALL and EFL teachers' perceptions of the role of literary texts in language development in North Macedonia, addressing a gap left open by the absence of any context-specific instrument for this population. The LTQ's 18-item, four-factor structure, reliability and partial measurement equivalence across teacher groups support its use in future research and professional development contexts, while the elevated correlations between Linguistic Fit, Text Relevance and Student Outcomes, which surfaced consistently across the main analysis and the invariance testing, mark the clearest direction for the instrument's continued refinement.

Ethical Statement

The research study was approved by the Office for Scientific Research and Innovation at the University of Tetova (Decision No. 02-1975/1, 03 December 2025).

Data Availability Statement

The datasets generated during the current study are not publicly available due to the confidential nature of the data collected from participants, but they are available from the corresponding author on reasonable request.

Disclosure Statement

The authors have no conflict of interest to declare.

When preparing this article, the authors used Claude Sonnet 5 (Anthropic) on 15 and 16 August 2026 to revise manuscript text, cross-verify statistical analyses against results independently produced in SPSS and AMOS, and check reference-list accuracy. Representative prompts include: "Check whether these CFA fit indices meet standard thresholds" and "Verify that every in-text citation has a matching reference list entry". The authors reviewed and edited all of the output and accept full responsibility for the content and integrity of the publication.

Acknowledgements

The authors would like to thank Dr Sam Duncan and Dr Amos Paran (UCL Institute of Education, UK) for permission to adapt their 2017 questionnaire.

References

Ayre, C., & Scally, A. J. (2014). Critical values for Lawshe's content validity ratio: Revisiting the original methods of calculation. *Measurement and Evaluation in Counseling and Development*, *47*(1), 79–86.

Benić, M. Ž. (2021). Translation and validation of the Kaufman domains of creativity scale on a Croatian sample of early childhood and preschool education students. *Center for Educational Policy Studies Journal*, *11*(3), 163–179.

Boateng, G. O., Neilands, T. B., Frongillo, E. A., Melgar-Quiñonez, 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, Article 149. https://doi.org/10.3389/fpubh.2018.00149

Bobkina, J., & Stefanova, S. (2016). Literature and critical literacy pedagogy in the EFL classroom: Towards a model of teaching critical thinking skills. *Studies in Second Language Learning and Teaching, 6*(4), 677–696.

Borg, S. (2003). Teacher cognition in language teaching: A review of research on what language teachers think, know, believe, and do. *Language Teaching, 36*(2), 81–109. https://doi.org/10.1017/S0261444803001903

Byram, M. (2021). *Teaching and assessing intercultural communicative competence: Revisited*. Multilingual Matters. https://doi.org/10.21832/9781800410251

Byrne, B. M. (2016). *Structural equation modeling with AMOS: Basic concepts, applications, and programming* (3rd ed.). Routledge.

Calafato, R., & Hunstadbråten, S. (2024). Literature in language education: Exploring EFL learners' literary competence profiles. *English Teaching & Learning*, *49*(3), 537–559. https://doi.org/10.1007/s42321-024-00193-w

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, *9*(2), 233–255. https://doi.org/10.1207/S15328007SEM0902_5

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

Cunningham, B., & Farmer, A. (2016). An investigation of the relationships between teacher beliefs and student achievement: Teacher beliefs. *The International Journal of Educational Organization and Leadership*, *23*(2), 1–14. https://doi.org/10.18848/2329-1656/CGP/v23i02/1-14

DeVellis, R. F., & Thorpe, C. T. (2021). *Scale development: Theory and applications*. Sage Publications.

Duncan, S., & Paran, A. (2017). *The effectiveness of literature on acquisition of language skills and intercultural understanding in the high school context*. UCL Institute of Education, University College London.

Duncan, S., & Paran, A. (2018). Negotiating the challenges of reading literature: Teachers reporting on their practice. In J. Bland (Ed.), *Using literature in English language education: Challenging reading for 8–18 year olds* (pp. 243–260). Bloomsbury Academic.

Gottardo, A., Mirza, A., Koh, P. W., Ferreira, A., & Javier, C. (2018). Unpacking listening comprehension: The role of vocabulary, morphological awareness, and syntactic knowledge in reading comprehension. *Reading and Writing*, *31*(8), 1741–1764. https://doi.org/10.1007/s11145-017-9736-2

Hedgcock, J. S., & Ferris, D. R. (2018). *Teaching readers of English: Students, texts, and contexts*. Routledge.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Kline, R. B. (2016). *Principles and practice of structural equation modeling* (4th ed.). Guilford Press.

Kovač, M., & Kepic Mohar, A. (2022). The changing role of textbooks in primary education in the digital era: What can we learn from reading research? *Center for Educational Policy Studies Journal*, *12*(2), 11–27. https://doi.org/10.26529/cepsj.1290

Lacko, D., Čeněk, J., Točík, J., Avsec, A., Đorđević, V., Genc, A., Haka, F., Šakotić-Kurbalija, J., Mohorić, T., Neziri, I., & Subotić, S. (2022). The necessity of testing measurement invariance in

cross-cultural research: Potential bias in cross-cultural comparisons with individualism–collectivism self-report scales. *Cross-Cultural Research, 56*(2–3), 228–267. https://doi.org/10.1177/10693971211068971

Levine, S., Trepper, K., Chung, R. H., & Coelho, R. (2021). How feeling supports students' interpretive discussions about literature. *Journal of Literacy Research, 53*(4), 491–515. https://doi.org/10.1177/1086296X211052249

Li, L. (2020). *Language teacher cognition*. Palgrave Macmillan UK.

Lohr, S. L. (2021). *Sampling: Design and analysis* (3rd ed.). Chapman & Hall/CRC.

Mart, Ç. T. (2018). Literature in language teaching: A recipe to maximize learning. *L1-Educational Studies in Language and Literature, 18*(1), 1–25. https://doi.org/10.17239/L1ESLL-2018.18.01.09

Milković, I., Cindrić, I., & Cvitanović, D. (2025). In search of authentic and adapted literary texts in textbooks for young learners of English as a foreign language. *Center for Educational Policy Studies Journal,* Article 1906. https://doi.org/10.26529/cepsj.1906

Paran, A. (2008). The role of literature in instructed foreign language learning and teaching: An evidence-based survey. *Language Teaching, 41*(4), 465–496. https://doi.org/10.1017/S026144480800520X

Phipps, S., & Borg, S. (2009). Exploring tensions between teachers' grammar teaching beliefs and practices. *System*, *37*(3), 380–390. https://doi.org/10.1016/j.system.2009.03.002

Ploj Virtič, M., Du Plessis, A., & Sorgo, A. (2023). Development and validation of the 'Mentoring for Effective Teaching Practicum Instrument'. *Center for Educational Policy Studies Journal*, *13*(3), 233–260. https://doi.org/10.26529/cepsj.1315

Putnick, D. L., & Bornstein, M. H. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental Review*, *41*, 71–90. https://doi.org/10.1016/j.dr.2016.06.004

Rathert, S., & Cabaroglu, N. (2022). Theorising textbook adaptation in English language teaching. *Center for Educational Policy Studies Journal*, *12*(2), 169–188. https://doi.org/10.26529/cepsj.1287

Rushidi-Rexhepi J., Hamzai J., & Adili B. (2025). Literary texts and language development: Piloting a questionnaire among Albanian language and EFL teachers in North Macedonia. *British Journal of Education, 13*(10), 1–14. https://doi.org/10.37745/bje.2013/vol13n10114

Sa'd, S. H. T., & Eames, O. Q. (2021). A classroom survey of language teachers' discriminatory practices against students: Causes, consequences and keys. *Center for Educational Policy Studies Journal*, *11*(1), 91–114. https://doi.org/10.26529/cepsj.541

Scheaffer, R. L., Mendenhall, W., Ott, L., & Gerow, K. G. (2011). *Elementary survey sampling* (7th ed.). Cengage Learning.

State Statistical Office of the Republic of North Macedonia. (2024). *Education and science*. https://www.stat.gov.mk/oblastopsto.aspx?id=5

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

Appendix: Literature Teaching Questionnaire (LTQ)

This questionnaire is designed to assess teachers' perceptions of the role and impact of literary texts in students' language development. Participants are asked to rate each statement on a 6-point Likert scale:

1 = *Strongly Disagree/Aspak nuk pajtohem*; 2 = *Disagree/Nuk pajtohem*; 3 = *Somewhat Disagree/Disi nuk pajtohem*; 4 = *Somewhat Agree/Disi pajtohem*; 5 = *Agree/Pajtohem*; 6 = *Strongly Agree/Plotësisht pajtohem*

No.	Item Statement (English + Albanian)
1	I find it easy to integrate literary texts into my syllabus. E kam të lehtë t'i integroj tekstet letrare në planin tim mësimor.
2	Students should be able to learn some new grammar from the literary text. Teksti letrar duhet t'u mundësojë nxënësve të mësojnë diçka të re nga gramatika.
3	The time available is sufficient to explore literary texts in depth. Kam kohë të mjaftueshme në dispozicion për t'i trajtuar tekstet letrare në thellësi.
4	Teaching literature supports students' creative language use. Letërsia suporton përdorimin kreativ të gjuhës nga ana e nxënësve.
5	Literary texts help me address multiple language skills (reading, writing, speaking). Tekstet letrare më ndihmojnë të shqyrtoj aftësi të ndryshme gjuhësore te nxënësit (lexim, shkrim, të folur).
6	I enjoy teaching literature in my subject area. E shijoj ligjërimin e letërsisë në lëndën time.
7	I feel confident in teaching complex literary texts. Ndihem i/e sigurt kur ligjëroj tekste të ndërlikuara letrare.
8	Teaching literature helps students develop intercultural awareness. Letërsia ndihmon nxënësit të zhvillojnë ndërgjegjësimin ndërkulturor.
9	The literary text should lend itself to teaching a particular language point. Teksti duhet të mundësojë ligjërimin e një strukture gjuhësore të caktuar (psh. një rregull gramatikor, një fjalor specifik, një formë sintaksore apo një stil shprehjeje).
10	Students should be able to learn new vocabulary from the literary text. Nxënësit duhet të jenë në gjendje të mësojnë fjalor të ri nga teksti letrar.
11	Overall, the literary text should be well within the linguistic ability of my students. Në përgjithësi, teksti letrar duhet të jetë lehtësisht i kuptueshëm për nxënësit.
12	The literary text should explore global or social issues. Teksti letrar duhet të trajtojë çështje globale apo sociale.
13	Literary texts increase student motivation to learn the language. Tekstet letrare rrisin motivimin e nxënësve për ta mësuar gjuhën.
14	The literary text should challenge my students from a language point of view. Teksti letrar duhet t'i sfidojë nxënësit e mi nga aspekti gjuhësor.
15	The literary text should be of high literary quality. Teksti letrar duhet të jetë i një cilësie të lartë letrare.
16	The literary text should offer aesthetic or emotional value. Teksti letrar duhet të ofrojë vlera estetike ose emocionale.
17	The literary text should represent the diversity of local/national and world literature. Teksti letrar duhet të përfaqësojë larminë e letërsisë kombëtare dhe asaj botërore.
18	The literary text should support the development of specific language skills (e.g., reading, writing, grammar, speaking). Teksti letrar duhet ta ndihmojë zhvillimin e shkathtësive specifike gjuhësore (p.sh. lexim, shkrim, gramatikë, të folur).

Biographical note

Jehona Rushidi-Rexhepi, PhD, is an associate professor in the field of literature at the Faculty of Pedagogy, University of Tetova, North Macedonia. Her research interests include Albanian literature, particularly children's literature, intertextuality, and the relationships between literary texts, as well as the role of literature in language development, teaching, and educational practice.

Kushtrim Ahmeti, PhD, is a full professor in the field of philosophy at the Faculty of Philosophy, University of Tetova, North Macedonia. His research interests include political and moral philosophy, philosophical anthropology, and postmodern thought, particularly the philosophy of Michel Foucault, as well as moral development and the relationship between philosophy and education.

Agon Memeti, PhD, is a full professor in the field of computer science at the University of Tetova, North Macedonia. His research interests include software engineering, service-oriented architectures, web and mobile application development, cloud computing, and information systems. His recent work also focuses on educational technologies, particularly learning management systems, embedded computing, and the use of artificial intelligence in learning environments.

Ibrahim Neziri, PhD, is a full professor in the field of psychology at the Faculty of Philosophy, University of Tetova, North Macedonia. His research interests include organizational culture, work psychology and motivation, cross-cultural psychology, individualism and collectivism, the psychological effects of social media use among adolescents, and the relationship between workplace autonomy and employee commitment.

Bujar Adili, PhD, is an assistant professor in the field of professional pedagogy at the Faculty of Pedagogy, University of Tetova, North Macedonia. His academic work focuses on teacher education and professional development. His research interests include intercultural education and communication, teachers' intercultural competence and self-efficacy, inclusive education, and educational assessment and evaluation, particularly in school contexts.