



International Journal of Artificial Intelligence and Machine Learning

Publisher's Home Page: <https://www.svedbergopen.com/>



Research Paper

Open Access

Development and Psychometric Validation of an Emotion Reading Interest Dataset for Higher-Education Learners: A Current-Data Audit and Revalidation Framework

Hema Anand. B^{1*} and P. Sridhara Acharya²

¹Research Scholar, Institute of Computer Science and Information Science, Srinivas University, Mangalore – 575001, Karnataka, India

²Research Professor, Institute of Computer Science and Information Science, Srinivas University, Mangalore – 575001, Karnataka, India

*Corresponding Author: hemaanand20b@gmail.com

Abstract

Emotion aware educational systems need not only technically accurate emotion classifiers, but also access to valid instruments for emotion in education to measure the relations between learners' emotions and the reading interest, the relevance of their recommendations, their engagement as well as their perceived learning outcomes. In this study, a 25-item questionnaire for measuring 5 constructs: Emotional State, Reading Interest, Emotion Based Recommendation, Cognitive Retention, Engagement and Motivation is audited and psychometrically evaluated. There were 500 and 1000 records in two files provided, which were comma-separated. The record-level comparison results showed that the initial file of 500 records from the 1,000-response file did not differ from the entire 500-response file. Therefore, the files were not analysed as independent data, but rather the 1,000-response file was used as one file, and then split into the exploratory subsample of 500 respondents, and the confirmatory holdout of 500 respondents. The dataset contained no missing values in Q5–Q29, no out-of-range responses, no straight-lined response vectors and no duplicate 25-item response patterns. However, critically low internal consistency in the proposed scales was observed. In the full dataset, the values for Cronbach's alpha were between -0.018 and 0.049 , and the values for provisional McDonald's omega were between 0.060 and 0.238 . Corrected item-total correlation values were clustered around zero. The forced five-factor exploratory factor analysis accounted for a small amount of total item variance (11.30%) and the Kaiser-Meyer-Olkin measure of sampling adequacy was 0.495 , while only three items had primary loadings ≥ 0.40 . CFI = 0.381 and TLI = 0.300 were obtained in the confirmatory analysis. The composite reliability was not higher than 0.176 , while the average value of variance extracted was not higher than 0.084 . Eight out of 10 of the HTMT comparisons were above 0.90 . Content-validity ratings, cognitive-interview records, test-retest observations, and ethics documentation and item-level reverse-scoring keys were not available. This data set can thus not be presented as a validated five-construct emotion-reading interest data set. The study pinpoints specific measurement failures, and outlines an expert review, cognitive interviewing, pilot testing, ordinal EFA, independent CFA, and measurement-invariance and test-retest procedure for instrument revision and new data collection.

Keywords: Cognitive Retention, Dataset Validation, Emotion-Aware Learning, Higher Education, Measurement Invariance, Psychometric Evaluation, Reading Interest, Scale Development.

1. INTRODUCTION

Learners' emotions impact their attention, effort, strategic processing and persistence. According to control-value theory, achievement emotions are the result of students' perceptions of control over an activity and of the value of that activity [1]. Some aspects of educational emotions, such as using the Achievement Emotions Questionnaire, illustrate that educational emotions can be multi-dimensional and context-specific rather than simply being a unipolar continuum of positive and negative emotions [2]. The interest in reading is dynamic too. It can start with one time attention and then build into a longer-term individual interest with the repeated exposure to relevant material [3, 4].

These relationships are particularly pertinent to the technology-supported learning. When confusion is resolved

during complex learning, it can be beneficial to deeper processing, however, when confusion continues it can shift to frustration or disengagement [5]. In a systematic review of technology-based learning environments meaningful relationships between learner emotions, motivation, engagement and achievement were identified [6]. The emotion-aware recommendation system can thus be used for adapting reading material, presentation difficulty, or conversational support based on emotional information. Recent research has also highlighted the application of artificial intelligence in the context of Education 5.0, which focuses on personalized and adaptive learning experiences, such as learning analytics, intelligent tutoring systems, natural language processing, adaptive assessment, and individualized learner support [7]. But, evidence of the educational value cannot simply be derived from the participants giving moderately positive questionnaire ratings. Validity of the measurement is the issue of interpretation and use of scores, not the extent of the response file. Standards for educational measurement include evidence on content, internal structure, relationships with other variables, and consequences of the use of scores [8]. As for scaling, scale-development guidance also considers construct definition, item generation, content assessment, pilot testing, structural analysis, reliability and validation as related steps [9, 10]. There are different types of validity according to COSMIN-based methodology: content validity, structural validity, internal consistency, reliability and measurement invariance [11, 12].

The questionnaire that was available was designed to link learners' emotion with reading interest by employing Q5–Q29. Five categories of five items each were pre-defined: Emotional State, Reading Interest, Emotion-Based Recommendation, Cognitive Retention and Engagement and Motivation. However, there are three conceptual issues that need to be explored. Firstly, "Emotional State" denotes, based on the provided construct description, how the learner thinks the chatbot understands him/her. It is not intended to measure the actual emotional state of the learner. Second, "Cognitive Retention" is a perceived understanding or recall with no objective test of retention. Third, there is no reading subject, topic, text identifier or observed reading choice in the data set. The present variables thus indicate the perception of an emotion-aware recommendation experience, but do not yet establish a direct association between the emotion and the topic of reading interest of a learner at the observation level.

Preliminary reliability calculations resulted in a more serious concern. Cronbach's alpha for the standalone 500-response file were -0.018 for Emotional State, 0.006 for Reading Interest, 0.001 for Emotion-Based Recommendation, 0.028 for Cognitive Retention and 0.031 for Engagement and Motivation. Items with a value close to or below zero do not covary as a coherent scale. Positive construct means can't make up for this failure.

The study consequently addresses four research questions:

RQ1: Are internal consistency and item–total relationships of proposed five-item scales acceptable?

RQ2: Is the five-construct structure of an exploratory analysis recovered?

RQ3: Does the proposed model show structural, convergent and discriminant validity in an independent holdout sample?

RQ4: Is the available dataset well documented and well psychometrically supported to be released as a validated emotion–reading interest dataset?

There are three contributions of this paper. First, it creates an audit trail at record level documenting the relationship between the 500-and 1,000-response files. Secondly it provides a full, current-data Psychometric diagnosis instead of construct-mean. Third, it offers a replicable redevelopment and revalidation process to generate a higher education emotion–reading interest dataset that is defensible.

2. RELATED WORK

2.1 Emotions, Interest and Technology-Supported Learning

Achievement emotions influence learning in their role of attention, motivation, cognitive-resource allocation and self-regulation [1, 2]. Enjoyment and interest can help encourage long-term engagement, but anxiety, boredom and hopelessness can impede processing. The impact of these effects depends on the task and the learner and thus requires the explicit measurement of the emotion construct intended. In addition, interest is more specific than general satisfaction. It involves emotional involvement with a subject, appreciation and desire to revisit [3, 4]. A learner can show positive feedback, but not show interest in the reading topic recommended by the chatbot. Therefore, an emotion–reading interest dataset should contain psychometrically supported measures of emotion perception and firsthand interaction data, such as emotional state, recommended topic, selected content and subsequent engagement.

There is an additional layer of complexity with emotion aware educational technology. Disengagement may occur when learning is accompanied by a manageable cognitive impasse, or confusion, which can be productive if it reflects that impasse [5]. Therefore, the influence of emotion-aware adaptation cannot simply be boiled down to

the fact that positive emotions are good, and negative emotions are bad [6].

2.2 Scale Development and Content Validity

Preceding statistical validation, there is construction of definition and of the items themselves [9, 10]. Items should be comprehensible to the target population and collectively cover the entire construct domain. Structured expert ratings of item relevance, comprehensiveness and clarity are used often to gauge content validity. Next, cognitive interviews are conducted to explore how these participants interpret and retrieve information for, and judge and respond to, each item [15, 16].

These are not optional formalities for the qualitative stages. There is nothing statistical analysis can do to fix items that measure the wrong construct, combine multiple ideas or use ambiguous language. A set of responses that doesn't contain the entire wording for an item, the scoring key, and an expert-review record is thus not suitable for a full content-validity claim.

2.3 Structural Validity and Reliability

Exploratory factor analysis investigates the number and nature of the dimensions without using the proposed model [17,18]. A more defensible factor-retention rule than one that keeps all eigenvalues greater than one is parallel analysis [19]. Confirmatory factor analysis subsequently assesses whether or not an independently specified measurement model can account for the observed relationships between the items [20]. In the case of ordinal Likert responses, it is generally better to use polychoric correlations and robust ordinal estimators instead of ordinary maximum likelihood [21, 22]. A number of indices have to be interpreted in order to assess model fit. CFI and TLI refer to improvement compared with an independence model, whereas RMSEA and SRMR refer to approximate or residual fit [23]. With low factor loadings and reliability, acceptable fit at the global level does not necessarily imply validity.

Although Cronbach's alpha is still widely reported [24] its assumptions are often not met. The omega and congenerial reliability of McDonald's is useful complementary measures [25, 26]. Factor loadings, composite reliability and average variance extracted [27] [28] are used as a common measure for convergent validity. HTMT offers a direct measure of the discriminant validity between latent constructs [29].

2.4 Measurement Invariance and Temporal Reliability

Measurement Invariance is the extent to which groups interpret and utilize an instrument in the same way. The standard order of assessment is for configural, metric and scalar invariance before comparing group means [30]–[32]. Test–retest reliability is considered separately and is often expressed as an intraclass correlation coefficient (ICC) with a suitable specification [33].

The defensibility of the structure of measurement used must then be demonstrated across gender, study level and discipline for a dataset designed for a variety of learners in higher education. When the baseline measurement model is not supported, it is not meaningful to interpret such analyses.

3. METHODOLOGY

3.1 Research Design

A two-stage psychometric design was used for the study. Stage 1 performed a forensic and statistical audit of the single 1,000-response dataset provided. Stage 2 outlined the instrument-redevelopment and new validation-data process that must be completed before making the dataset available as a validated research resource.

The current-data audit used dataset-integrity verification, missing-data assessment, response-quality screening, descriptive statistics, internal-consistency analysis, exploratory factor analysis, confirmatory factor analysis, composite reliability, average variance extracted, and HTMT. The feasibility of measurement invariance and test–retest reliability was evaluated, but these analyses were not estimated when the necessary prerequisites were unavailable.

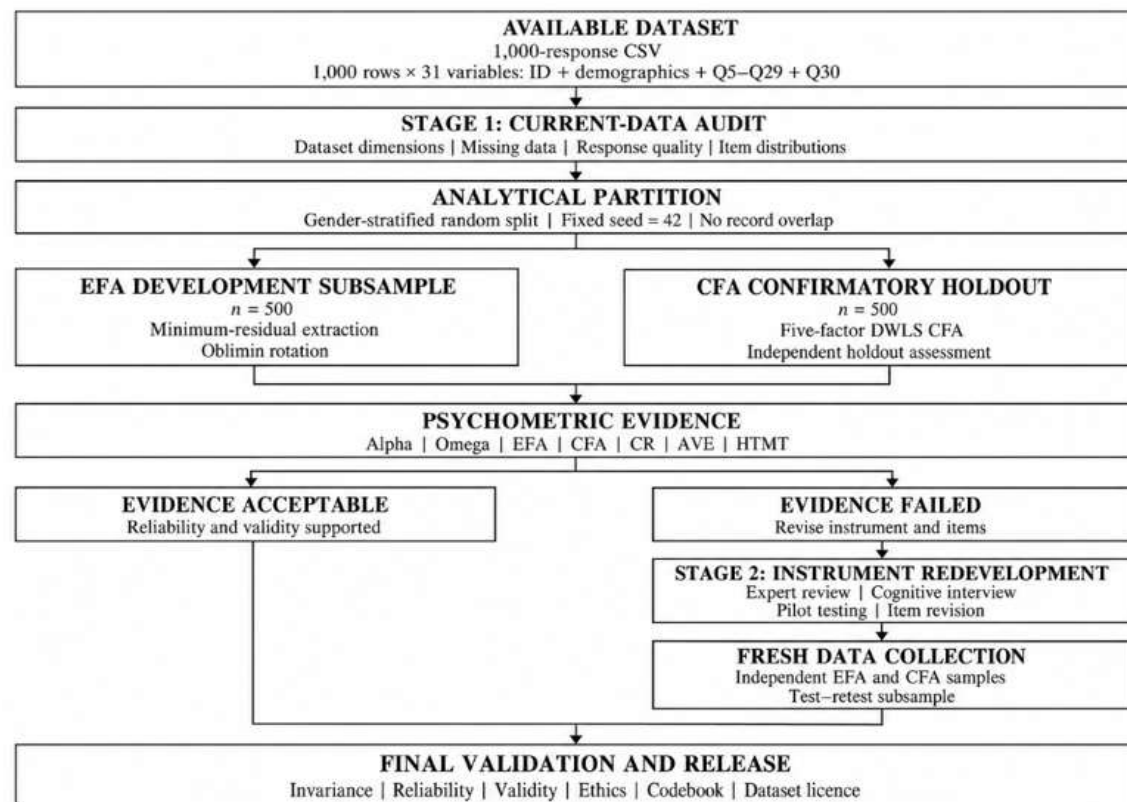


Fig. 1. Two-stage workflow comprising current-data auditing, instrument redevelopment, expert content validation, cognitive interviewing, pilot testing, independent EFA and CFA, reliability evaluation, measurement invariance, test–retest assessment and dataset release.

3.2 Dataset and Record-Level Audit

The supplied dataset contained 1,000 respondents and 31 variables: a respondent identifier, four demographic variables, Q5–Q29, and an optional open-ended item. All integrity checks treated missing values consistently across every column. The CSV was analysed once as a single 1,000-response dataset. Analytical independence was instead established through non-overlapping exploratory and confirmatory subsamples drawn from this dataset.

Source file was supplied, provenance assessment focused on internal record structure, identifier uniqueness, response-range compliance, missingness, response-pattern duplication and the documented analytical partition. This approach avoids implying cross-file replication or external sample independence while preserving the original psychometric objectives, numerical analyses, construct mapping and evidence-based interpretation of the questionnaire.

3.3 Proposed Constructs and Variable Mapping

The 25 Likert items were answered on a scale of 1 (strongly disagree)–5 (strongly agree). The predefined construct mapping is shown in Table I.

Table I. Proposed Construct–Item Mapping

Construct	Operational interpretation in available material	Items	Number of items	Measurement clarification
Emotional State	Perceived chatbot emotional understanding	Q5–Q9	5	Measures perceived emotional responsiveness, not a direct emotion state
Reading Interest	Interest in recommended learning material	Q10–Q14	5	Requires direct topic and reading-choice variables for Objective 2
Emotion-Based Recommendation	Perceived relevance of emotionally adapted content	Q15–Q19	5	Represents perceived recommendation quality
Cognitive Retention	Perceived understanding, memory and recall	Q20–Q24	5	Does not replace an objective retention test
Engagement and	Sustained attention and	Q25–	5	Represents perceived

Motivation	motivation	Q29	engagement and motivation
------------	------------	-----	---------------------------

A tentative reflective measurement structure is shown in Fig. 2. Each item is single keyed to one construct, constructs may be correlated and item errors are unique.

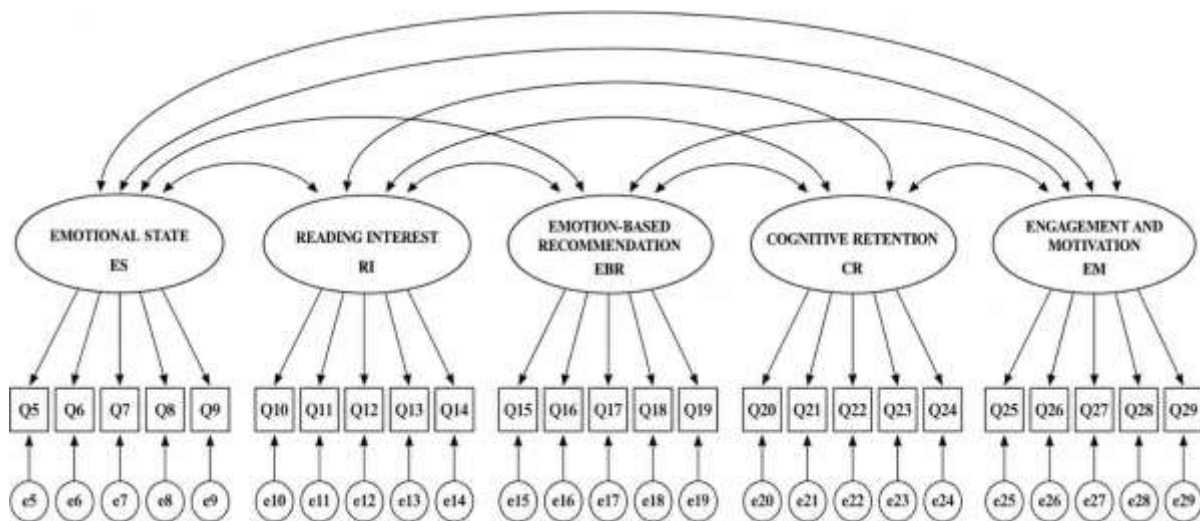


Fig. 2. Proposed correlated five-factor model linking Q5–Q29 to Emotional State, Reading Interest, Emotion-Based Recommendation, Cognitive Retention, and Engagement and Motivation.

3.4 Content Validity, Cognitive Interviews and Pilot Testing

No expert-rating matrix, item-level content-validity index, cognitive-interview transcript or pilot-testing report was available. Content validity was thus judged as not established. Revalidation is performed by a multidisciplinary team of six to ten experts in educational psychology, psychometrics, reading research, higher education and emotion-aware learning technology. Experts individually evaluate the relevance of the items, as well as the clarity and alignment of the items with the construct, on a four-point scale. The item-level CVI is equal to

$$I-CVI_i = \frac{n_{i,3} + n_{i,4}}{N_e}, \quad (1)$$

where N_e is the number of experts and $n_{i,3} + n_{i,4}$ is the number rating item as relevant or highly relevant. An I-CVI of at least 0.78 and an S-CVI/Ave of at least 0.90 are used as revision criteria [13], [14].

Cognitive interviews are then carried out with learners from various disciplines and study levels, of both genders. Think-aloud and verbal-probing questions focus on the interpretation of emotion words, reading interest words, periods of reference, and response options [15], [16]. Items that can be understood in more than one way and items with double-barrelled wording are edited before being sampled.

3.5 Missing Data and Response-Quality Analysis

The audit examined:

- Missing values in demographics and Q5–Q29;
- Values outside the permitted range of 1–5;
- Duplicate respondent identifiers;
- Duplicate 25-item response patterns;
- Straight-lining across all 25 items;
- Long strings of identical responses;
- Within-person response variability; and
- Missing optional open-ended responses.

No response time, device information and interaction logs were found. There would thus be no way of assessing speeding, detection of repeated device submission, and automated response. The revised study uses missing-data handling that is based on the mechanism and extent of the missingness; this approach is not the unconditional mean replacement [34]. Response-quality exclusions are preregistered and sensitivity analysis [35] is used.

3.6 Internal-Consistency Analysis

Cronbach's alpha was calculated as

$$\alpha = \frac{k}{k-1} \left(1 - \frac{\sum_{i=1}^k \sigma_i^2}{\sigma_X^2} \right), \quad (2)$$

where k is the number of items, σ_i^2 is the variance of item i , and σ_X^2 is the variance of the summed score [24].

McDonald's omega was estimated from one-factor congeneric models:

$$\omega = \frac{(\sum_{i=1}^k \lambda_i)^2}{(\sum_{i=1}^k \lambda_i)^2 + \sum_{i=1}^k \theta_i}, \quad (3)$$

where λ_i denotes the standardized loading and θ_i denotes residual variance [25], [26].

Corrected item-total correlations and alpha-if-deleted values were also calculated. Confidence intervals for alpha were obtained from 2,000 bootstrap samples. All items were analysed in the originally recorded direction, as an item-scoring key was not available. A diagnostic, known as diagnostic reversal search, was used to find out if any scoring-direction pattern could yield acceptable reliability, but not to select the rules for scoring.

3.7 EFA and Independent CFA

Using a fixed random seed of 42 and gender-stratified sampling, the single 1,000-response dataset was divided into a 500-record EFA development sample and a non-overlapping 500-record CFA holdout sample. These groups were analytical subsamples of the same CSV, rather than separately supplied datasets, and no respondent appeared in both groups. Minimum-residual extraction with oblimin rotation was used for the current-data EFA. In order to determine the sampling suitability, the Kaiser-Meyer-Olkin statistic and Bartlett's test were used. A total of 1,000 parallel-analysis permutations that preserved margins were performed for the examination of factor retention.

The confirmatory model was expressed as

$$\mathbf{x} = \mathbf{\Lambda}\boldsymbol{\eta} + \boldsymbol{\varepsilon}, \quad (4)$$

where $\mathbf{x}$ is the 25-item vector, $\mathbf{\Lambda}$ is the loading matrix, $\boldsymbol{\eta}$ contains the five latent constructs, and $\boldsymbol{\varepsilon}$ contains measurement errors. The holdout CFA used diagonally weighted least squares. The final recollection must use a robust ordinal estimator, such as WLSMV, with the complete item-scoring key [21], [22].

3.8 Convergent and Discriminant Validity

Composite reliability was calculated as

$$CR = \frac{(\sum \lambda_i)^2}{(\sum \lambda_i)^2 + \sum \theta_i}, \quad (5)$$

and average variance extracted as

$$AVE = \frac{\sum \lambda_i^2}{\sum \lambda_i^2 + \sum \theta_i}. \quad (6)$$

Values of approximately 0.70 for CR and 0.50 for AVE provide commonly applied reference points, but substantive interpretation and model quality remain necessary [27], [28].

HTMT for constructs a and b was calculated as

$$HTMT_{ab} = \frac{\text{mean}(|r_{ij}|), i \in a, j \in b}{\sqrt{\text{mean}(|r_{ii'}|)\text{mean}(|r_{jj'}|)}}, \quad (7)$$

where the denominator contains within-construct correlations. Values below 0.85 provide a conservative reference for discriminant validity [29].

3.9 Measurement Invariance

The intended validation evaluates:

- Configural invariance;
- Metric or loading invariance;
- Scalar or threshold invariance; and

- Strict or residual invariance where appropriate.

These are analysed separately by gender, study level and discipline. Changes of less than 0.010 in CFI and 0.015 in RMSEA are used as guides for decisions along with model convergence, loadings, and substantive interpretability [30], [32]. In the present study, however, the process of invariance testing was discontinued because the baseline five-factor model and construct reliability were deemed to be unacceptable.

3.10 Test-Retest Reliability

A test-retest design involves the same group of consenting participants who complete the revised questionnaire at a later point in time, typically 2 to 4 weeks post-test. The analysis is proposed to be Absolute Agreement ICC with a 95% CI [33]. There is no wave identifier in the current dataset, repeated-participant linkage, or second measurement occasion; therefore, test-retest reliability was not estimable.

3.11 Ethics and Dataset Documentation

There is no ethics-approval identifier, participant-information statement, consent procedure, recruitment start/end dates, or named institution in the available dataset documentation; nor is there a sampling frame or data-generation log. These omissions impede verification of collection provenance.

The revised dataset documentation includes:

- Ethics committee and approval number;
- Participant-information and consent procedure;
- Recruitment dates and sites;
- Inclusion and exclusion criteria;
- Complete questionnaire wording;
- Item-source and modification record;
- Forward and reverse scoring keys;
- Missing-value codes;
- Data-quality rules;
- Variable-level codebook;
- De-identification procedure;
- Data licence;
- Version history; and
- Persistent dataset identifier.

These requirements support transparent and reusable research-data management consistent with FAIR principles [36, 37, 38].

3.12 Analysis Algorithm

Algorithm 1: Psychometric audit and revalidation procedure

Input: Single 1,000-response CSV, Q5–Q29 construct mapping and demographic variables.

1. Standardise column names and missing-value representation.
2. Confirm the dataset dimensions, variable structure and respondent identifiers.
3. Retain all 1,000 records once as the complete analysis dataset.
4. Validate response ranges and respondent identifiers.
5. Screen missingness and response-quality indicators.
6. Calculate item and construct descriptive statistics.
7. Compute alpha, omega, corrected item–total correlations and bootstrap intervals.
8. Create independent 500/500 EFA and CFA subsamples.
9. Evaluate KMO, Bartlett’s test and parallel analysis.
10. Estimate the exploratory factor model.
11. Fit the proposed five-factor confirmatory model.
12. Compute CR, AVE and HTMT.
13. Test measurement invariance only if the baseline model is valid.
14. Estimate test–retest reliability only if linked repeated observations exist.
15. Classify every validation requirement as supported, unsupported or unavailable.
16. Revise the instrument and recollect data when critical criteria fail.

4. RESULTS

4.1 Dataset Integrity and Participant Profile

There were 1,000 rows and 31 variables in the supplied dataset. Following complete-file integrity checks, gender-stratified sampling with seed 42 created non-overlapping 500-record EFA and 500-record CFA subsamples. The resulting analytical split is documented in Fig. 3.

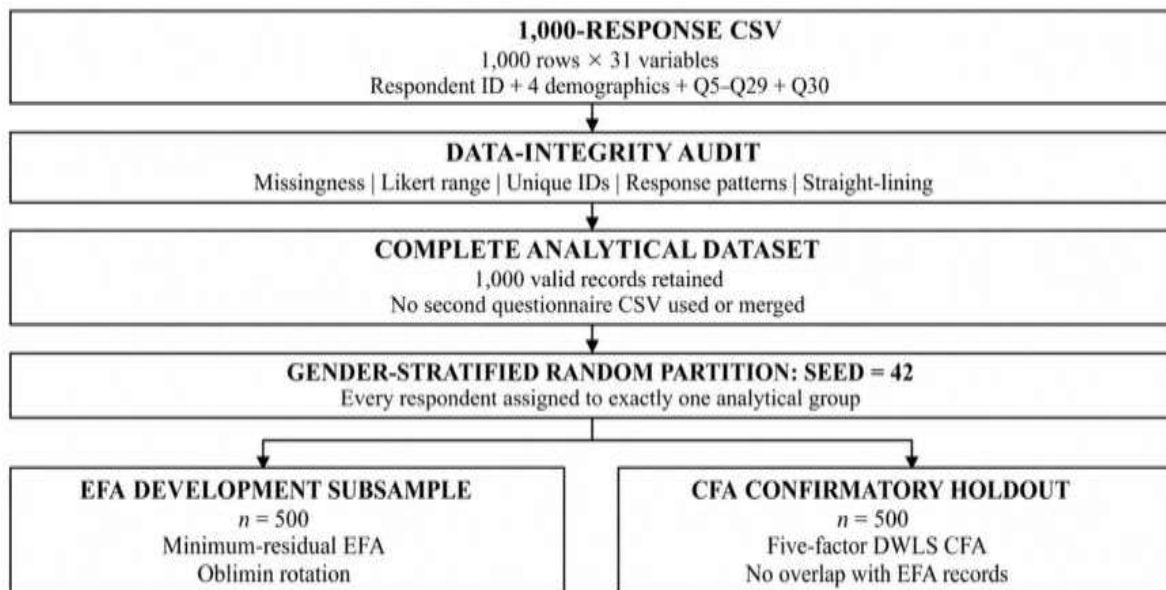


Fig. 3. Analytical division of the single 1,000-response dataset into non-overlapping 500-record EFA development and 500-record CFA holdout subsamples.

There were no missing demographic or Q5–Q29 values. No closed-response item had scores outside of the

acceptable range of 1 to 5. Q30 had 378 blank responses (37.8%) but it was an optional open-ended question. No 25-item response pattern was duplicated and no subject chose the same response for each of the 25 items. The longest identical-response sequence was eight items; no sequence reached the preregistered flag of 15 items.

Table II. Sample Composition and Data-Quality Audit

Characteristic	Category or test	n	Percentage/result
Age	18–20 years	381	38.1%
	21–23 years	339	33.9%
	24–26 years	187	18.7%
	Above 26 years	93	9.3%
Gender	Female	446	44.6%
	Male	538	53.8%
	Prefer not to say	16	1.6%
Study level	Undergraduate	564	56.4%
	Postgraduate	346	34.6%
	Research scholar	90	9.0%
Discipline	Arts	133	13.3%
	Commerce	185	18.5%
	Engineering	358	35.8%
	Others	92	9.2%
	Science	232	23.2%
Core missingness	Demographics and Q5–Q29	0	0.0%
Optional missingness	Q30 open-ended	378	37.8%
Invalid Likert values	Values outside 1–5	0	0.0%
Duplicate item patterns	Q5–Q29	0	0.0%
Complete straight-lining	Same response across Q5–Q29	0	0.0%
Long-string flag	At least 15 identical consecutive responses	0	Maximum = 8
Analytical split	Non-overlapping EFA and CFA records	1,000	500 EFA + 500 CFA

These checks ensure syntactic completeness, but not true origin of the participants. There is still a need for recruitment records, timestamps, platform logs and consent documentation.

4.2 Item and Construct Descriptive Statistics

Item means ranged from 3.479 for Q9 to 3.683 for Q22. Standard deviations were between 1.150 and 1.299. Moderately negative skews were reported for all items as there were more high agreement categories.

Table III. Item-Level Descriptive Statistics for the 1,000-Response Dataset

Item	Construct	Mean	SD	Skewness	Kurtosis
Q5	Emotional State	3.546	1.219	–0.526	–0.667
Q6	Emotional State	3.522	1.247	–0.491	–0.733
Q7	Emotional State	3.545	1.241	–0.521	–0.718
Q8	Emotional State	3.504	1.233	–0.446	–0.746
Q9	Emotional State	3.479	1.213	–0.437	–0.755
Q10	Reading Interest	3.566	1.214	–0.452	–0.775
Q11	Reading Interest	3.518	1.220	–0.544	–0.607
Q12	Reading Interest	3.562	1.225	–0.486	–0.769
Q13	Reading Interest	3.528	1.194	–0.458	–0.732
Q14	Reading Interest	3.535	1.243	–0.467	–0.798
Q15	Emotion-Based Recommendation	3.573	1.159	–0.539	–0.502
Q16	Emotion-Based Recommendation	3.547	1.251	–0.549	–0.728
Q17	Emotion-Based Recommendation	3.573	1.217	–0.527	–0.689
Q18	Emotion-Based Recommendation	3.617	1.151	–0.606	–0.414
Q19	Emotion-Based Recommendation	3.670	1.150	–0.602	–0.432
Q20	Cognitive Retention	3.578	1.182	–0.522	–0.600
Q21	Cognitive Retention	3.666	1.195	–0.639	–0.481
Q22	Cognitive Retention	3.683	1.163	–0.649	–0.372
Q23	Cognitive Retention	3.603	1.227	–0.562	–0.650
Q24	Cognitive Retention	3.604	1.241	–0.483	–0.819
Q25	Engagement and Motivation	3.633	1.187	–0.529	–0.677

Q26	Engagement and Motivation	3.492	1.246	-0.464	-0.782
Q27	Engagement and Motivation	3.542	1.228	-0.521	-0.694
Q28	Engagement and Motivation	3.556	1.240	-0.480	-0.780
Q29	Engagement and Motivation	3.491	1.299	-0.441	-0.958

The constructs were: Emotional State = 3.519; Reading Interest = 3.542; Emotion-Based Recommendation = 3.596; Cognitive Retention = 3.627; and Engagement and Motivation = 3.543. These scores reflect moderate agreement with each of the items. They do not provide evidence that the items are reliable multi-item scales.

Fig. 4. Item-level and construct-level mean profiles for Q5–Q29, with dispersion indicators based on the 1,000-response dataset.

4.3 Internal Consistency and Reverse-Scoring Audit

The reliability results are shown in Table IV. All coefficients were estimated from the single 1,000-response dataset. Increasing the available sample size did not compensate for the absence of coherent within-construct covariance. Full-sample Cronbach's alpha ranged from -0.018 to 0.049.

Table IV. Reliability and Item-Direction Diagnostic

Construct	Full-sample n	α , full 1,000	Bootstrap 95% CI for α	Provisional ω	Corrected item-total range	Maximum diagnostic α under any direction pattern
Emotional State	1,000	-0.018	-0.127–0.076	0.183	-0.034–0.028	0.073
Reading Interest	1,000	0.021	-0.085–0.113	0.136	-0.029–0.037	0.101
Emotion-Based Recommendation	1,000	0.049	-0.052–0.138	0.081	-0.039–0.069	0.118
Cognitive Retention	1,000	0.031	-0.070–0.113	0.060	-0.071–0.057	0.183
Engagement and Motivation	1,000	0.043	-0.052–0.131	0.238	-0.050–0.066	0.131

The omega estimates were also very low and unstable as several one-factor solutions were dominated by single items. None of the corrected item-total correlations reached the common cut-off point of 0.30. Each five-item block was tested for all possible polarity combinations using one fixed reference item. This completely data-driven and methodologically inappropriate optimisation yielded maximum values of only 0.073–0.183 for alpha. Thus, an unidentified reverse-scored item cannot account for the failure of reliability. The mechanical recoding of the questionnaire is not enough; the determination of actual reverse scores will still depend on the semantics of the items.

Fig. 5. Cronbach's alpha, provisional McDonald's omega and composite reliability for the five proposed constructs, with the 0.70 reference line.

4.4 Exploratory Factor Analysis

The EFA subsample produced KMO = 0.495, with item-level KMO values from 0.451 to 0.569. Bartlett's test was statistically significant, $\chi^2(300) = 481.844$, $p < 0.001$. Because the significant Bartlett test shows that the correlation matrix is not a perfect identity matrix, the KMO value shows inadequate common variance for a coherent factor model.

Permutation parallel analysis resulted in seven statistical dimensions, instead of the suggested five. This outcome did not lead to evidence of seven substantive constructs as KMO and communalities and loadings were still weak. Instead it suggested that there were only some sample correlations and there was no fixed structure in mind.

Fig. 6. Observed eigenvalues and 95th-percentile eigenvalues obtained from 1,000 marginal-preserving permutations in the 500-record EFA subsample.

The five factors were the result of a forced five-factor minimum-residual solution with only 11.30% variance explained. Only three of the questions—Q7, Q21 and Q27—had primary loadings with absolute values ≥ 0.40 . Communalities ranged from 0.023 to 0.285, with a mean of 0.113.

Table V. Oblimin-Rotated Five-Factor EFA Loadings

Item	F1	F2	F3	F4	F5
Q5	0.266	0.012	-0.082	0.100	-0.236

Q6	0.125	0.068	-0.003	-0.022	-0.050
Q7	0.000	0.133	0.429	-0.001	0.017
Q8	0.185	-0.165	-0.002	-0.096	0.221
Q9	0.000	-0.036	-0.017	0.274	-0.036
Q10	-0.121	0.063	0.123	0.066	-0.210
Q11	-0.120	0.044	-0.077	-0.016	0.176
Q12	-0.033	0.049	-0.171	-0.138	0.101
Q13	0.127	0.076	-0.175	0.162	0.077
Q14	-0.075	0.158	-0.056	0.291	0.023
Q15	-0.059	-0.087	0.169	0.043	0.177
Q16	-0.170	-0.135	-0.112	0.103	-0.232
Q17	0.020	-0.195	0.054	0.087	0.118
Q18	0.168	-0.040	0.014	0.010	0.110
Q19	-0.199	-0.114	-0.094	0.201	0.085
Q20	0.007	-0.183	0.370	0.042	-0.051
Q21	0.534	0.003	-0.011	0.002	0.007
Q22	-0.024	0.098	-0.043	0.094	0.350
Q23	0.058	0.172	0.005	0.008	-0.223
Q24	0.132	0.366	0.065	0.100	0.046
Q25	0.116	-0.080	-0.275	0.016	-0.035
Q26	-0.002	0.068	0.044	0.116	-0.123
Q27	0.029	0.003	0.046	0.435	-0.003
Q28	-0.059	0.339	0.098	0.038	0.065
Q29	-0.077	0.290	-0.118	-0.168	-0.062
Variance explained	2.41%	2.34%	2.30%	2.17%	2.08%
Cumulative variance	2.41%	4.75%	7.05%	9.22%	11.30%

The factor composition failed to replicate the item blocks proposed. For instance, F1 was characterized by Q21 from Cognitive Retention, F3 was characterized by Q7 and Q20 from two different constructs and F4 was characterized by Q27 from Engagement and Motivation. The loading pattern is visualised in Fig. 7.

Fig. 7. Grayscale heatmap of absolute and signed oblimin-rotated loadings for the forced five-factor EFA solution.

4.5 Confirmatory Factor Analysis and Convergent Validity

The five-factor model was tested on the independent 500-record holdout. The model produced $\chi^2(265) = 407.068$, $p < 0.001$, CFI = 0.381, TLI = 0.300, NFI = 0.231, GFI = 0.231, AGFI = 0.130, RMSEA = 0.033 and SRMR = 0.052. Although RMSEA and SRMR were relatively small, the incremental fit indices were very poor due to the fact that the independence model was only slightly worse than the proposed model. Apparently good absolute indices did not correspond to a useful measurement model, which was confirmed by factor loadings, CR and AVE.

Table VI. Holdout CFA, Composite Reliability and Convergent Validity

Construct/model criterion	Standardized loading range	CR	AVE	Required evidence	Result
Emotional State	-0.105–0.127	0.012	0.010	CR $\approx$ 0.70; AVE $\approx$ 0.50	Not supported
Reading Interest	-0.512–0.247	0.022	0.067	CR $\approx$ 0.70; AVE $\approx$ 0.50	Not supported
Emotion-Based Recommendation	Approximately 0.000	0.000	0.000	CR $\approx$ 0.70; AVE $\approx$ 0.50	Not supported
Cognitive Retention	-0.511–0.054	0.176	0.084	CR $\approx$ 0.70; AVE $\approx$ 0.50	Not supported
Engagement and Motivation	-0.071–0.104	0.007	0.005	CR $\approx$ 0.70; AVE $\approx$ 0.50	Not supported
Overall CFA	CFI = 0.381; TLI = 0.300; RMSEA = 0.033; SRMR = 0.052	—	—	Evaluate multiple fit indices and loadings	Model rejected

All constructs failed to reach acceptable levels of convergent validity. Even the highest AVE was 0.084, which means that the best construct accounted for less than 9% of the variance of its assigned indicators.

Fig. 8. Proposed five-factor CFA model with standardised holdout loadings, latent correlations and unsupported

loading paths indicated using grayscale line styles.

4.6 Discriminant Validity

The HTMT values are indicated in Table VII. Ten pairwise values were calculated, with eight of them more than 0.90. Average between-construct item correlations were much larger than the very small within-construct item correlations, which resulted in values greater than one.

Table VII. HTMT Discriminant-Validity Matrix

Construct	ES	RI	EBR	CR	EM
Emotional State (ES)	1.000				
Reading Interest (RI)	1.216	1.000			
Emotion-Based Recommendation (EBR)	1.099	0.763	1.000		
Cognitive Retention (CR)	1.505	1.004	1.015	1.000	
Engagement and Motivation (EM)	0.994	0.904	1.180	0.744	1.000

These HTMT values indicate that there is no strong correlation among all constructs. Rather, they are largely due to the lack of significant monotrait correlations. You cannot determine the valid evidence of discriminant validity without first determining the valid evidence of convergent validity.

Fig. 9. Grayscale HTMT matrix for the five proposed constructs, with reference boundaries at 0.85 and 0.90.

4.7 Measurement Invariance, Test-Retest and Supporting Evidence

The base five-factor model was disconfirmed. No attempt was made to estimate metric and scalar invariance. This would not make an invalid baseline model interpretable if one went to increasingly constrained models.

Table VIII. Status of Required Psychometric Evidence

Evidence	Available sample/information	Current result	Decision
Expert content validity	No expert-rating matrix	I-CVI and S-CVI unavailable	Not established
Cognitive interviews	No interview records	Comprehension evidence unavailable	Not established
Pilot testing	No distinct pilot dataset	Item revision history unavailable	Not established
Internal consistency	n = 1,000	$\alpha = -0.018$ to 0.049	Failed
Structural validity: EFA	n = 500	KMO = 0.495; five factors explain 11.30%	Failed
Structural validity: CFA	n = 500	CFI = 0.381; TLI = 0.300	Failed
Convergent validity	CFA holdout	CR ≤ 0.176 ; AVE ≤ 0.084	Failed
Discriminant validity	n = 1,000	Eight of ten HTMT values > 0.90	Failed
Gender invariance	Female = 446; male = 538; prefer not to say = 16	Baseline model rejected; third group too small	Not tested
Study-level invariance	UG = 564; PG = 346; research scholar = 90	Baseline model rejected	Not tested
Discipline invariance	n = 92–358 across five groups	Baseline model rejected; group-size imbalance	Not tested
Test-retest reliability	No repeated wave	ICC unavailable	Not established
Ethics and consent	No approval/consent documentation supplied	Provenance cannot be verified	Not established
Complete codebook	Item identifiers available; item wording absent	Reverse keys and construct definitions incomplete	Not established

4.8 Overall Dataset-Readiness Decision

The current dataset has only been subjected to basic syntactic validation and not to measurement validation as required for a validated research dataset.

Table IX. Dataset Validation and Release Decision

Requirement	Current evidence	Required revision
Unique response records	Supported within the 1,000-row file	Preserve identifiers and collection logs
Independence of analytical subsamples	Supported	Preserve the preregistered split and identifiers
Complete item wording	Not available	Publish complete questionnaire and scoring key
Direct learner emotion	Not measured explicitly	Add an emotion-state variable and measurement context
Direct reading interest	Not fully represented	Add subject, topic, text and choice variables
Perceived recommendation quality	Represented by Q15–Q19	Rewrite and validate coherent items
Objective retention	Not available	Add immediate and delayed retention measures
Internal consistency	Failed	Revise item content and recollect data
Factorial validity	Failed	Repeat independent ordinal EFA and CFA
Measurement invariance	Not estimable	Test after baseline model is established
Temporal reliability	Not estimable	Collect linked second-wave responses
Ethics and consent	Undocumented	Document approval, consent and anonymisation
Public release readiness	Not achieved	Release only after all critical evidence is addressed

Fig. 10. Evidence-based dashboard classifying data integrity, content validity, structural validity, reliability, invariance, temporal stability, ethics and documentation as supported, failed or unavailable.

5. DISCUSSION

5.1 Principal Finding

The existing questionnaire does not constitute five valid multi-item scales. Alpha, omega, corrected item–total correlations, EFA, CFA, CR, AVE and HTMT all provide consistent support to this conclusion. It is not a one-size-fits-all threshold. Positive item means cannot be interpreted as evidence to suggest that emotion-aware recommendations lead to increased interest or retention in reading. Means are an indication of the location of respondents' choices of response categories. Items can be aggregated into construct scores that are interpretable if they have high levels of reliability and structural validity. Within-construct covariance was low and means were moderate for the present items.

5.2 Single-Dataset Provenance and Sample Independence

The analysis used one CSV containing 1,000 responses. To preserve analytical independence, gender-stratified sampling with a fixed seed divided these records into non-overlapping EFA development and CFA holdout groups. The 500/500 split enabled genuine holdout testing without treating any respondent as part of both analyses. This statistical separation supports structural cross-validation within the supplied dataset. However, it does not prove that the responses are authentic because collection logs, timestamps, recruitment records and consent documentation were not supplied.

5.3 Reverse Scoring Does Not Explain the Failure

Negative alpha can happen in the case of items that are supposed to measure the same construct but, on average, are negatively covariant. Sometimes this occurs due to negatively worded items not being reversed. However, the full wording of the items and the key to the scoring were not provided, and no item can be legitimately reverse-scored from Q-number alone. Even the most favourable data-driven polarity pattern was unable to increase alpha above 0.183, as revealed by the exhaustive direction diagnostic. So, there is a need to say that reverse coding does not save the current instrument. Conceptual review and rewriting of items.

5.4 Failure of the Proposed Factor Structure

The proposed item blocks were not recovered during the EFA process. The five forced factors accounted for only 11.30% of the variance, and the average communality was 0.113. This implies that the majority of the item variance was attributable to the unique factors of each item or to those factors that were not captured by the common factors.

The CFA agreed. Though this may seem good if considered alone, the low RMSEA and SRMR could be due to CFI and TLI of only 0.381 and 0.300. Most importantly, factor loadings and reliability estimates were very small. Fit

indices should therefore be interpreted in conjunction with parameter estimates and theory.

5.5 Convergent and Discriminant Validity

There was no construct with an AVE close to 0.50 or a CR close to 0.70. The items thus did not have enough variance to allow for convergent validity. Further, HTMT was also uninterpretable as favourable evidence as the within-construct correlations were also near to zero. Eight values exceeded 0.90. The results highlight the importance of a validation paper to include more than just the construct means and Cronbach's alpha. Content, structural, reliability, convergent, and discriminant, and invariance evidence are necessary to a complete argument.

5.6 Implications for Objective 2

The objective of the current questionnaire is not fully operationalized; it is to create a dataset to be able to associate learner emotion with reading interest. Records perceptions about emotional awareness and recommendations, but there is not a specific reading-interest subject/content record and a learner-emotion field.

A revised dataset requires three connected layers:

- Learner state: emotion label, emotion intensity, confidence and timestamp.
- Reading context: discipline, topic, content identifier, difficulty and recommendation rationale.
- Outcome: selected/not selected, reading completion, dwell time, rating, engagement and objective immediate or delayed retention.

The validated questionnaire can be used in addition to these layers, but cannot be used as a replacement. This wider structure generates a real link between the feelings and the interest in reading and not just a general Likert perception.

6. INSTRUMENT REDEVELOPMENT AND FRESH-DATA VALIDATION

The questionnaire is developed from the construct-definition stage. "Emotional State" is either reworded to "Perceived Emotional Responsiveness" or is substituted with items that actually assess the learner's emotional state. Cognitive retention is renamed "perceived cognitive retention" unless objective recall and comprehension measures are added. Each construct is first given a larger item pool of about 8–12 items. Experts determine if each item measures one construct, one idea, does not involve assumptions about the system, and whether the item can be answered by learners who were provided with different recommendations. There is no deletion of items simply to boost alpha—it is still about theoretical coverage.

Cognitive interviews are held with undergraduate, postgraduate and research-scholar groups after expert review. A pilot study tests the response distributions, item understanding and initial correlations. Fresh responses are then collected after finalising the instrument. EFA/CFA samples are independent from the point of collection or via a preregistered random holdout. Measurement invariance is assessed only when the structure and loadings and the reliability of the baseline model are acceptable. A subsample separately consented is used to collect test–retest data. The final dataset includes the original responses, scoring code, codebook, expert-review summary, psychometric output, exclusions, missing-data decisions and versioned questionnaire.

7. PRACTICAL IMPLICATIONS

The present construct scores must not be used to compare genders, study levels or disciplines. They also should not be used as targets for a recommendation algorithm. Using a composite score with unreliable measurement error in training the system would introduce the same error into the recommendation model.

Responses to individual questions Q5–Q29 may also be maintained as unvalidated raw observations for audit—they must be provenance authentic and ethically documented. But evidence of the type of constructs such as 3.519 or 3.627 cannot be used to say that the system is better for emotional understanding or cognitive retention. The new instrument should function as a component of an overall dataset. Criterion evidence comes from actual reading behaviour and objective outcomes. The use of learner emotion needs to be handled sensitively with clear consent, with minimal data kept and not used for grading, discipline or psychological diagnosis.

8. CONCLUSION

This study audited a proposed emotion–reading interest questionnaire containing 25 items mapped to five constructs. The supplied CSV contained 1,000 responses and was treated as the single analysis dataset. These records were divided through gender-stratified random sampling into non-overlapping exploratory and confirmatory subsamples. Psychometric properties of the five proposed constructs were not acceptable. Cronbach's alpha values were between –0.018 and 0.049, provisional omega values were between 0.060 and 0.238 and corrected item–total correlations were near zero for all possible full-sample alphas. In the case of the forced five-factor EFA, the item variance accounted for was 11.30%, while the CFA yielded CFI = 0.381 and TLI = 0.300.

There was no composite reliability score above 0.176, no AVE score above 0.084 and eight out of ten HTMT scores were >0.90.

The present questionnaire and response file thus cannot be considered as a validated five-construct emotion-reading interest dataset. The solution to the problem is not to be achieved by reverse scoring alone. The instrument needs conceptual changes, expert assessment of content validity, cognitive interviewing and pilot testing and the collection of new data. The output data must then be subjected to independent ordinal EFA and CFA, reliability assessment, measurement invariance, test-retest evaluation and ethics verification and full documentation. The existing variables are only partially relevant to Objective 2. A full emotion-reading-interest dataset would involve a direct link between learner emotion, reading topic/content, exposure to recommendations, observed reading behaviour and retention outcome. Once the redevelopment process is completed, the existing questionnaire response document will be converted into a defensible dataset that can be used for emotion-aware learning research within higher education.

REFERENCES

- [1] R. Pekrun, "The control-value theory of achievement emotions: Assumptions, corollaries, and implications for educational research and practice," *Educational Psychology Review*, vol. 18, no. 4, pp. 315–341, 2006, doi:10.1007/s10648-006-9029-9.
- [2] R. Pekrun, T. Goetz, A. C. Frenzel, P. Barchfeld, and R. P. Perry, "Measuring emotions in students' learning and performance: The Achievement Emotions Questionnaire," *Contemporary Educational Psychology*, vol. 36, no. 1, pp. 36–48, 2011, doi:10.1016/j.cedpsych.2010.10.002.
- [3] S. Hidi and K. A. Renninger, "The four-phase model of interest development," *Educational Psychologist*, vol. 41, no. 2, pp. 111–127, 2006, doi:10.1207/s15326985ep4102_4.
- [4] U. Schiefele, "Interest, learning, and motivation," *Educational Psychologist*, vol. 26, nos. 3–4, pp. 299–323, 1991, doi:10.1080/00461520.1991.9653136.
- [5] S. D'Mello and A. Graesser, "Dynamics of affective states during complex learning," *Learning and Instruction*, vol. 22, no. 2, pp. 145–157, 2012, doi:10.1016/j.learninstruc.2011.10.001.
- [6] K. Loderer, R. Pekrun, and J. C. Lester, "Beyond cold technology: A systematic review and meta-analysis on emotions in technology-based learning environments," *Learning and Instruction*, vol. 70, Art. no. 101162, 2020, doi:10.1016/j.learninstruc.2018.08.002.
- [7] D. V. R. Moorthy, N. Anuradha, G. Kalpana Devi, V. V. Subba Rao, and Ch. Nagamani, "Education 5.0: Integrating artificial intelligence (AI) for personalized and adaptive learning," *Journal of Animal Environment*, vol. 18, no. 7S, pp. 631–637, 2026. <https://aejournals.org/index.php/AEJ/article/view/3300>
- [8] American Educational Research Association, American Psychological Association, and National Council on Measurement in Education, *Standards for Educational and Psychological Testing*. Washington, DC, USA: AERA, 2014.
- [9] G. O. Boateng, T. B. Neilands, E. A. Frongillo, H. R. Melgar-Quinonez, and S. L. Young, "Best practices for developing and validating scales for health, social, and behavioral research: A primer," *Frontiers in Public Health*, vol. 6, Art. no. 149, 2018, doi:10.3389/fpubh.2018.00149.
- [10] R. F. DeVellis and C. T. Thorpe, *Scale Development: Theory and Applications*, 5th ed. Thousand Oaks, CA, USA: Sage, 2021.
- [11] J. J. Gagnier, J. Lai, L. B. Mokkink, and C. B. Terwee, "COSMIN reporting guideline for studies on measurement properties of patient-reported outcome measures," *Quality of Life Research*, vol. 30, pp. 2197–2218, 2021, doi:10.1007/s11136-021-02822-4.
- [12] L. B. Mokkink et al., "The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties," *Journal of Clinical Epidemiology*, vol. 63, no. 7, pp. 737–745, 2010, doi:10.1016/j.jclinepi.2010.02.006.
- [13] M. R. Lynn, "Determination and quantification of content validity," *Nursing Research*, vol. 35, no. 6, pp. 382–385, 1986, doi:10.1097/00006199-198611000-00017.
- [14] D. F. Polit, C. T. Beck, and S. V. Owen, "Is the CVI an acceptable indicator of content validity? Appraisal and recommendations," *Research in Nursing & Health*, vol. 30, no. 4, pp. 459–467, 2007, doi:10.1002/nur.20199.
- [15] G. B. Willis, *Cognitive Interviewing: A Tool for Improving Questionnaire Design*. Thousand Oaks, CA, USA: Sage, 2005.

- [16] S. Presser et al., "Methods for testing and evaluating survey questions," *Public Opinion Quarterly*, vol. 68, no. 1, pp. 109–130, 2004, doi:10.1093/poq/nfh008.
- [17] L. R. Fabrigar, D. T. Wegener, R. C. MacCallum, and E. J. Strahan, "Evaluating the use of exploratory factor analysis in psychological research," *Psychological Methods*, vol. 4, no. 3, pp. 272–299, 1999, doi:10.1037/1082-989X.4.3.272.
- [18] A. B. Costello and J. W. Osborne, "Best practices in exploratory factor analysis," *Practical Assessment, Research & Evaluation*, vol. 10, Art. no. 7, 2005, doi:10.7275/jyj1-4868.
- [19] J. L. Horn, "A rationale and test for the number of factors in factor analysis," *Psychometrika*, vol. 30, pp. 179–185, 1965, doi:10.1007/BF02289447.
- [20] T. A. Brown, *Confirmatory Factor Analysis for Applied Research*, 2nd ed. New York, NY, USA: Guilford Press, 2015.
- [21] D. B. Flora and P. J. Curran, "An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data," *Psychological Methods*, vol. 9, no. 4, pp. 466–491, 2004, doi:10.1037/1082-989X.9.4.466.
- [22] C.-H. Li, "Confirmatory factor analysis with ordinal data: Comparing robust maximum likelihood and diagonally weighted least squares," *Behavior Research Methods*, vol. 48, pp. 936–949, 2016, doi:10.3758/s13428-015-0619-7.
- [23] L. Hu and P. M. Bentler, "Cutoff criteria for fit indexes in covariance structure analysis," *Structural Equation Modeling*, vol. 6, no. 1, pp. 1–55, 1999, doi:10.1080/10705519909540118.
- [24] L. J. Cronbach, "Coefficient alpha and the internal structure of tests," *Psychometrika*, vol. 16, pp. 297–334, 1951, doi:10.1007/BF02310555.
- [25] T. J. Dunn, T. Baguley, and V. Brunsden, "From alpha to omega: A practical solution to internal consistency estimation," *British Journal of Psychology*, vol. 105, no. 3, pp. 399–412, 2014, doi:10.1111/bjop.12046.
- [26] D. McNeish, "Thanks coefficient alpha, we'll take it from here," *Psychological Methods*, vol. 23, no. 3, pp. 412–433, 2018, doi:10.1037/met0000144.
- [27] C. Fornell and D. F. Larcker, "Evaluating structural equation models with unobservable variables and measurement error," *Journal of Marketing Research*, vol. 18, no. 1, pp. 39–50, 1981, doi:10.1177/002224378101800104.
- [28] T. Raykov, "Estimation of composite reliability for congeneric measures," *Applied Psychological Measurement*, vol. 21, no. 2, pp. 173–184, 1997, doi:10.1177/01466216970212006.
- [29] J. Henseler, C. M. Ringle, and M. Sarstedt, "A new criterion for assessing discriminant validity in variance-based structural equation modeling," *Journal of the Academy of Marketing Science*, vol. 43, no. 1, pp. 115–135, 2015, doi:10.1007/s11747-014-0403-8.
- [30] G. W. Cheung and R. B. Rensvold, "Evaluating goodness-of-fit indexes for testing measurement invariance," *Structural Equation Modeling*, vol. 9, no. 2, pp. 233–255, 2002, doi:10.1207/S15328007SEM0902_5.
- [31] F. F. Chen, "Sensitivity of goodness-of-fit indexes to lack of measurement invariance," *Structural Equation Modeling*, vol. 14, no. 3, pp. 464–504, 2007, doi:10.1080/10705510701301834.
- [32] D. L. Putnick and M. H. Bornstein, "Measurement invariance conventions and reporting," *Developmental Review*, vol. 41, pp. 71–90, 2016, doi:10.1016/j.dr.2016.06.004.
- [33] T. K. Koo and M. Y. Li, "A guideline of selecting and reporting intraclass correlation coefficients for reliability research," *Journal of Chiropractic Medicine*, vol. 15, no. 2, pp. 155–163, 2016, doi:10.1016/j.jcm.2016.02.012.
- [34] R. J. A. Little and D. B. Rubin, *Statistical Analysis with Missing Data*, 3rd ed. Hoboken, NJ, USA: Wiley, 2019.
- [35] A. W. Meade and S. B. Craig, "Identifying careless responses in survey data," *Psychological Methods*, vol. 17, no. 3, pp. 437–455, 2012, doi:10.1037/a0028085.
- [36] M. D. Wilkinson et al., "The FAIR guiding principles for scientific data management and stewardship," *Scientific Data*, vol. 3, Art. no. 160018, 2016, doi:10.1038/sdata.2016.18.

- [37] Malik, N., & Singh, K. (2025). Enhancing agro-industrial development in Haryana through crop diversification and value chain integration. *Journal of Artificial Intelligence in Agriculture and Applied Research*, 1(1), 13–19. <https://agrpjournals.com/journals/journal/index.php/jaiaar/article/view/4>
- [38] Pandey, P. (2025). Artificial intelligence for early detection of dental caries from radiographic images: A machine learning. *Journal of Dentistry and Clinical Sciences*, 1(1), 31–39. <https://agrpjournals.com/journals/journal/index.php/jdacs/article/view/12>