Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Development and Validation of a 20 Item Undergraduate Scale for Psychological Constructs in Five Eastern Ethiopian Public Universities

Gemechu A. Gobena

gemechugobena127@gmail.com

Haramaya University

Berhanu N. Worku

Jimma University

Dame A. Abdi

Addis Ababa University

Jeylan W. Hussein

Haramaya University

Research Article

Keywords: Confirmatory factor analysis, Ethiopian higher education, Exploratory factor analysis, Psychometric validation, Structural equation modeling

Posted Date: November 6th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11276865/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Valid and reliable measurement instruments underpin credible educational research, yet psychometrically validated tools jointly measuring psychological constructs remain scarce for Ethiopian graduating undergraduates. This study developed and validated a 20-item structured questionnaire measuring these two mediating constructs and tested their structural relationship. A quantitative, correlational design was employed with 200 graduating undergraduates purposively drawn from five Ethiopian universities and colleges via stratified random sampling. Items were generated from theory and literature, then refined through expert-panel content validity indexing (I-CVI, S-CVI) before pilot reliability testing. Data were collected using a self-administered, five-point Likert questionnaire and analyzed sequentially in SPSS and AMOS through reliability testing, exploratory factor analysis (Principal Axis Factoring, Promax rotation), confirmatory factor analysis, Pearson correlation, and structural equation modeling. Results revealed a clean two-factor structure, communalities and loadings exceeding recommended thresholds, excellent composite reliability (.955–.957), strong convergent and discriminant validity, and excellent model fit (CFI = TLI=1.00, RMSEA=.000). Motivation to Transfer and Academic Self-Efficacy correlated strongly (r = .758–.80) across three estimation methods, rejecting all three null hypotheses. The validated instrument offers Ethiopian higher education a statistically defensible tool, supporting future structural modeling of institutional determinants and academic outcomes.

1. Introduction

Measurement quality is central to the credibility, interpretability, and reproducibility of educational and psychological research because many constructs of theoretical interest, such as motivation, self-efficacy, engagement, resilience, and learning behavior, are latent and therefore cannot be observed directly. Researchers consequently infer these constructs from patterns of responses to theoretically derived indicators. Weak measurement can therefore do more than introduce random error; it can distort relationships among constructs, blur conceptual boundaries, and produce conclusions that appear statistically convincing but are substantively misleading. Contemporary scale-development scholarship consequently emphasizes a systematic process in which conceptual definition, item development, content evaluation, structural validity, reliability, and construct validity are considered as interconnected sources of evidence rather than as isolated statistical exercises (DeVellis & Thorpe, 2021; Abejehu et al, 2024; Deffner et al., 2025).

This multidimensional view represents an important departure from approaches that equate a satisfactory Cronbach's alpha with a validated instrument. DeVellis and Thorpe (2021) emphasize that scale development begins with a clearly specified construct and appropriately generated items, whereas recent methodological guidance similarly stresses the importance of evaluating dimensionality and validity before interpreting scores substantively. These perspectives converge in treating psychometric validation as an evidentiary process, although they differ in emphasis: traditional practice often privileges numerical indices, while contemporary approaches place greater weight on whether the

statistical evidence supports the intended interpretation of scores. Thus, acceptable reliability or model fit should be interpreted as evidence supporting validity claims, not as proof that an instrument is inherently valid Abejehu et al., 2025).

The contextual nature of measurement makes this distinction particularly important in cross-cultural research. Constructs do not necessarily acquire identical meanings across languages, educational systems, cultures, or institutional environments. Consequently, an instrument demonstrating satisfactory psychometric properties in one population cannot automatically be assumed to operate equivalently in another. Recent Ethiopian evidence illustrates this point clearly. Berhanu et al. (2025), for example, developed a classroom-engagement instrument specifically for Ethiopian secondary-school students because existing measures did not adequately address the local context. Their procedure combined literature review, expert assessment, EFA, CFA, reliability analysis, and measurement invariance, demonstrating that contextual validation requires more than translating an established instrument. A similar conclusion emerges from Asari et al. (2025), who developed an adolescent health-literacy scale in Ethiopia through a three-stage mixed-methods process. Rather than importing an existing scale unchanged, the researchers generated concepts and items from qualitative evidence, subjected them to expert and cognitive evaluation, and subsequently tested the structure and reliability using quantitative data. The final instrument demonstrated strong CFA fit and reliability across Afaan Oromo and Amharic versions. Importantly, their findings suggest that sociocultural context can influence not only item wording but also the dimensions through which a construct becomes empirically meaningful.

Exploratory factor analysis (EFA) is particularly valuable when the dimensional structure of an instrument requires empirical examination. It identifies patterns of shared variance among items and can reveal whether indicators cluster in ways that correspond to theoretically meaningful dimensions (Mokkink et al., 2024). Nevertheless, EFA should not be reduced to a mechanical search for statistically convenient factors. Sampling adequacy, communalities, factor loadings, cross-loadings, factor retention, and conceptual interpretability must be considered together. DeVellis and Thorpe (2021) similarly position factor analysis within a broader scale-development process rather than treating it as an independent statistical endpoint. The Ethiopian literature reinforces this position. Berhanu et al. (2025) used EFA to identify six dimensions of classroom engagement before evaluating the structure through CFA. Their results demonstrate that empirical factor identification can refine theoretical assumptions: the resulting dimensions included cognitive, behavioral, and emotional engagement as well as corresponding forms of disengagement. Thus, contextual EFA may provide more than statistical confirmation; it can reveal dimensions that are substantively relevant to the target population but insufficiently represented in imported conceptualizations.

CFA serves a different but complementary purpose. Whereas EFA explores plausible dimensional structures, CFA evaluates whether a theoretically specified structure adequately reproduces the observed covariance pattern (Berhanu et al., 2025; Stefana et al., 2025; Taye et al., 2025). Contemporary methodological guidance therefore supports an exploratory-confirmatory sequence in which a structure identified from one stage is subsequently subjected to more stringent testing rather than repeatedly

modified within the same data until an acceptable fit is obtained. Such an approach strengthens the distinction between discovering a measurement structure and confirming its plausibility. However, satisfactory CFA fit should not be interpreted as evidence that a model is uniquely correct. Fit indices are influenced by sample size, model complexity, specification, and other characteristics of the data. Consequently, model fit must be interpreted alongside factor loadings, residuals, theoretical coherence, and competing models. The 2025 Ethiopian classroom-engagement study illustrates this broader approach by evaluating item loadings, model fit, reliability, convergent and divergent validity, and measurement invariance rather than relying on a single fit statistic (Mokkink eta al., 2024; Berhanu et al., 2025; Lim, 2026).

Reliability remains an essential component of measurement quality because scores derived from items intended to represent a common construct should demonstrate adequate consistency. Nevertheless, Cronbach's alpha is not a comprehensive measure of reliability and can be misleading when its assumptions are violated or when item redundancy artificially increases the coefficient. Recent psychometric scholarship has therefore questioned the practice of treating alpha as sufficient evidence of scale quality and encourages researchers to interpret reliability in relation to the underlying measurement model and construct dimensionality (Sijtsma & Pfadt, 2021; Ellis, 2021; Abejehu & Chala, 2026). The distinction between reliability and validity is consequently critical. An instrument can produce highly consistent responses while failing to represent the intended construct adequately. Excessively homogeneous items may increase internal consistency but simultaneously narrow content coverage. Accordingly, reliability should be considered necessary but insufficient evidence, complemented by content, structural, convergent, and discriminant validity. This principle is also evident in Ethiopian validation research. Berhanu et al. (2025) reported Cronbach's alpha and composite reliability alongside factorial and validity evidence, while Asari et al. (2025) combined reliability estimates with CFA and content and cognitive evaluation.

Content validity provides the conceptual foundation for these subsequent analyses. Before statistical validation begins, researchers need evidence that the items adequately cover the theoretical domain and are clear, relevant, and appropriate for the intended respondents. Expert judgement is particularly important at this stage because statistical procedures cannot fully compensate for items that are conceptually ambiguous, redundant, or culturally inappropriate. Recent scale-development guidance therefore recommends integrating literature review, expert evaluation, target-population feedback, and empirical item analysis. Asari et al. (2025) provide especially strong Ethiopian evidence for this sequential logic. Their initial pool of 88 items was reduced through assessment of coverage and uniqueness, followed by expert review, cognitive interviews, pilot testing, and quantitative validation. The study demonstrates that item refinement should precede factorial analysis because the quality of the final factor structure depends partly on the conceptual and contextual quality of the items entering the analysis (Lim et al., 2026).

Once an adequate measurement model has been established, structural equation modeling (SEM) provides a powerful framework for examining theoretically specified relationships among latent

variables while accounting for measurement error. Its major advantage over analyses based solely on observed composite scores is that the measurement and structural components can be evaluated simultaneously. Nevertheless, SEM does not repair a defective measurement model: sophisticated structural estimates remain questionable when the underlying indicators inadequately represent their constructs (Abejehu & Chala, 2026). Consequently, CFA and other measurement evidence should logically precede interpretation of structural paths. This issue is particularly relevant to Ethiopian higher-education research. Recent national studies demonstrate increasing methodological sophistication, including the use of EFA, CFA, composite reliability, validity analysis, and measurement invariance. Yet the evidence remains concentrated in particular populations and constructs, such as secondary-school classroom engagement and adolescent health literacy (Berhanu et al., 2025; Asari et al., 2025; Abejehu & Chala, 2026). The existence of validated instruments in these populations should therefore not be interpreted as evidence that identical measurement structures automatically apply to graduating university students.

Although questionnaire-based research is expanding in Ethiopian higher education, comprehensive contextual validation remains insufficient (Abejehu & Chala, 2026). A continuing problem is the tendency to emphasize internal consistency while giving comparatively less attention to content representation, structural validity, convergent and discriminant validity, and the compatibility of measurement models with the characteristics of specific populations. Contemporary scale-development literature clearly indicates that these forms of evidence are complementary rather than interchangeable (DeVellis & Thorpe, 2021; Deffner et al., 2025). Moreover, existing Ethiopian validation studies have addressed important but relatively specific populations. Berhanu et al. (2025) focused on secondary-school classroom engagement, whereas Asari et al. (2025) addressed adolescent health literacy. Their methodological strength nevertheless highlights the remaining gap: psychometric evidence needs to be established for the particular population, construct, and purpose for which scores will be interpreted.

The present study addresses this gap by validating a 20-item questionnaire among graduating undergraduate students in Ethiopia through an integrated sequence of content validation, reliability assessment, EFA, CFA, correlation analysis, and SEM. This approach is intended not merely to demonstrate statistically acceptable coefficients but to establish converging evidence that the instrument's items adequately represent the intended constructs, that its latent structure is empirically defensible, and that the proposed relationships among constructs can be interpreted meaningfully within the target educational context. In this sense, contextual validation is both a methodological requirement and a substantive contribution: it strengthens confidence that conclusions about graduating students' educational and psychological characteristics reflect the constructs themselves rather than artifacts of an inadequately adapted measurement instrument.

To empirically examine the proposed relationships and differences among the study variables, the following null hypotheses were formulated. These hypotheses provide a basis for statistical testing by assuming no significant relationships among the constructs under investigation

H01: There is no statistically significant factor structure underlying the 20 questionnaire items among graduating undergraduate students.

H02: There is no statistically significant relationship among the latent constructs identified from the 20-item questionnaire.

H03: The hypothesized structural relationships among the latent constructs do not significantly fit the observed data.

2. Methods

This study was guided by the post-positivist research philosophy, which assumes that educational and psychological phenomena can be systematically investigated through empirical observation and statistical analysis while acknowledging that knowledge and measurement remain imperfect. This position was appropriate because the study examined latent constructs that cannot be directly observed but can be inferred from responses to theoretically meaningful questionnaire items. Contemporary psychometric guidance similarly emphasizes that measurement involves uncertainty and requires multiple sources of evidence to support score interpretations, rather than relying on a single statistical criterion (Deffner et al., 2025). A quantitative research approach was employed because the study aimed to generate numerical evidence regarding the validity, reliability, dimensionality, and relationships among latent constructs. Standardized questionnaire responses enabled systematic examination of item performance, internal consistency, factor structure, construct validity, correlations, and structural relationships. Quantitative psychometric procedures are particularly appropriate when the primary objective is to evaluate whether observed indicators adequately represent theoretically specified latent variables (Lamm et al., 2020; Deffner et al., 2025).

2.1. Research Method and Design

The study employed a quantitative cross-sectional survey method with a psychometric validation and correlational research design. The validation component was used to establish evidence regarding the content validity, reliability, factorial structure, and construct validity of the 20-item questionnaire. The correlational component examined relationships among the latent constructs, while structural equation modeling (SEM) was used to test the theoretically proposed relationships among them. The study followed a sequential validation process. First, the conceptual domains and questionnaire items were reviewed to ensure theoretical relevance and adequate representation of the intended constructs. Second, the items were subjected to expert evaluation and content validation. Third, the refined instrument was administered to graduating undergraduate students. Fourth, exploratory factor analysis (EFA) was conducted to investigate the underlying dimensional structure, followed by confirmatory

factor analysis (CFA) to evaluate the adequacy of the proposed measurement model. Finally, correlation analysis and SEM were used to examine relationships among the validated latent constructs.

This sequential approach is consistent with recent methodological guidance, which recommends integrating conceptual development, content evaluation, exploratory analysis, confirmatory testing, reliability assessment, and construct validation rather than treating validation as a single statistical procedure (Deffner et al., 2025). Ethiopian validation studies have similarly demonstrated the usefulness of combining EFA and CFA when evaluating whether internationally or theoretically derived constructs adequately represent local educational populations (Abejehu et al., 2024; Berhanu et al., 2025).

2.2. Participants

The target population comprised graduating undergraduate students in five selected higher education institutions in Ethiopia. Graduating students were considered appropriate because they had accumulated substantial academic and institutional experiences relevant to the educational and psychological constructs measured by the questionnaire. The study involved 200 graduating undergraduate students. The adequacy of the sample was considered in relation to the number of questionnaire items, the expected factor structure, and the requirements of the planned multivariate analyses. Recent methodological guidance cautions against relying exclusively on universal participant-to-item ratios because sample adequacy depends on factor loadings, communalities, the number of factors, model complexity, and the quality of indicators (Deffner et al., 2025). Therefore, the adequacy of the dataset for factor analysis was also evaluated empirically using the Kaiser–Meyer–Olkin (KMO) measure of sampling adequacy and Bartlett's test of sphericity.

A multistage sampling technique was used to select participants. Relevant higher education institutions and academic units were initially identified based on the availability of graduating undergraduate students. Eligible students were subsequently selected from the identified study settings using appropriate procedures based on institutional enrolment and accessibility. Participants were included if they were graduating undergraduate students, available during the data-collection period, and willing to participate voluntarily.

2.3. Item Development and Content Validation

The study used a 20-item questionnaire developed from relevant theoretical and empirical literature. Item development began with the identification and conceptual definition of the major domains underlying the study constructs. Items were subsequently formulated to represent theoretically important aspects of these domains and were written using clear and contextually appropriate language. Content validity was assessed through expert evaluation before the questionnaire was administered to the study participants. Experts with relevant expertise in educational psychology, psychometrics, measurement and evaluation, and higher education reviewed the items for relevance, clarity, representativeness, and contextual appropriateness.

The Item-Level Content Validity Index (I-CVI) was calculated to assess expert agreement regarding the relevance of each item, while the Scale-Level Content Validity Index (S-CVI) was used to evaluate the overall content adequacy of the instrument. Items receiving inadequate ratings were revised or removed based on the quantitative indices and qualitative recommendations provided by the experts. This approach is consistent with contemporary scale-development guidance, which emphasizes that content validation should combine systematic numerical evidence with substantive expert judgement (Lamm et al., 2020; Deffner et al., 2025). The importance of contextual expert review is also evident in Ethiopian validation research. For example, recent Ethiopian studies have incorporated expert evaluation and contextual adaptation before conducting EFA and CFA, recognizing that statistically strong items may still be conceptually or culturally inappropriate for the target population (Berhanu et al., 2023, 2025).

Table 1
Item-Level and Scale-Level Content Validity Indices Based on 10 Experts

Item	Relevance I-CVI	Clarity I-CVI	Representativeness I-CVI	Contextual Appropriateness I-CVI	Overall I-CVI*
1.	**1.00**	**1.00**	**1.00**	**.90**	**.98**
2.	1.00	.90	1.00	1.00	.98
3.	1.00	.90	1.00	.90	.95
4.	1.00	1.00	1.00	1.00	1.00
5.	1.00	1.00	.90	1.00	.98
6.	1.00	1.00	1.00	1.00	1.00
7.	1.00	.90	1.00	1.00	.98
8.	1.00	.90	1.00	.90	.95
9.	1.00	1.00	1.00	.90	.98
10.	1.00	.90	1.00	.90	.95
11.	1.00	1.00	1.00	1.00	1.00
12.	1.00	.90	1.00	1.00	.98
13.	1.00	1.00	1.00	1.00	1.00
14.	1.00	.90	1.00	1.00	.98
15.	1.00	.90	1.00	.90	.95
16.	1.00	1.00	1.00	1.00	1.00
17.	1.00	1.00	.90	1.00	.98
18.	1.00	.90	1.00	.90	.95
19.	1.00	1.00	1.00	1.00	1.00
20.	1.00	.90	1.00	1.00	.98

The item-level content validity results demonstrate strong expert agreement regarding the quality of the 20 items. Across the four evaluation dimensions- relevance, clarity, representativeness, and contextual appropriateness- the I-CVI values ranged from .90 to 1.00, exceeding the acceptable criterion of .80. Most items achieved an I-CVI of 1.00, indicating unanimous agreement among the 10 experts that these items adequately captured the intended content domain. The few items scoring .90 still reflected substantial expert consensus, with nine of ten experts rating them favorably. This pattern suggests that the items were conceptually appropriate, clearly worded, representative of dimensions, and suitable for

the Ethiopian higher-education context. Overall, the findings provide strong evidence supporting retention of all 20 items for subsequent psychometric validation.

Table 2
Scale-Level Content Validity Index

No.	Content-validity dimension	Sum of I-CVIs	Number of items	S-CVI/Ave
1.	**Relevance**	**20.00**	**20**	**1.00**
2.	Clarity	18.60	20	0.93
3.	Representativeness	19.70	20	0.985
4.	Contextual appropriateness	18.80	20	0.94
5.	Overall S-CVI/Ave	77.10	80 ratings	0.964

The scale-level content validity results provide strong evidence that the 20-item instrument adequately represents the intended construct. The S-CVI/Ave values ranged from .93 to 1.00 across relevance, clarity, representativeness, and contextual appropriateness, substantially exceeding the recommended .80 criterion. Relevance achieved 1.00, indicating unanimous expert agreement that the items were pertinent to the construct. Representativeness was exceptionally high (.985), suggesting comprehensive coverage of the intended content domain. Clarity (.93) and contextual appropriateness (.94) also demonstrated strong consensus, indicating that the items were understandable and suitable for the target educational context. The overall S-CVI/Ave of .964 confirms excellent scale-level content validity, supporting the instrument's advancement to subsequent EFA, CFA, reliability, and construct-validation procedures.

2.4. Reliability Assessment

Reliability was assessed to determine the consistency of the questionnaire items and extracted dimensions. Cronbach's alpha was calculated as an initial indicator of internal consistency. However, alpha was not treated as the sole evidence of measurement quality because its value can be influenced by the number of items, item intercorrelations, dimensionality, and underlying measurement assumptions. Therefore, composite reliability (CR) was also examined after CFA using standardized factor loadings and measurement-error estimates. Recent psychometric guidance recommends interpreting reliability within the broader measurement model and, where appropriate, complementing alpha with model-based reliability estimates such as composite reliability or McDonald's omega (Deffner et al., 2025). Reliability coefficients were consequently interpreted together with factor loadings, dimensionality, convergent validity, and discriminant validity. High reliability was not considered sufficient evidence of validity because highly correlated items may reflect redundancy rather than adequate representation of a broad construct.

2.5. Data-Collection Instrument and Procedures

Data were collected using a structured self-administered questionnaire. The instrument consisted of two sections. The first section collected demographic information relevant to describing the participants, whereas the second section contained the 20 items measuring the latent constructs. The substantive items were rated using a five-point Likert-type scale, ranging from 1 = strongly disagree to 5 = strongly agree. Five-point response formats are widely used in contemporary educational scale development because they provide sufficient response variation while remaining understandable for respondents (Vural et al., 2024). Before administration, the questionnaire was reviewed for clarity, language appropriateness, item sequencing, and completion time. Participants received standardized information concerning the purpose of the study and procedures for completing the questionnaire. After obtaining informed consent, questionnaires were distributed to eligible students and collected after completion. The completed questionnaires were checked for completeness before coding, data entry, and statistical analysis.

2.6. Methods of Data Analysis

Data were analyzed using IBM SPSS (v-24) and AMOS (v-24). Analysis was conducted sequentially. First, descriptive statistics, including frequencies, percentages, means, and standard deviations, were used to summarize participant characteristics and questionnaire responses. The dataset was subsequently screened for missing values, data-entry errors, outliers, normality, and multicollinearity. Content validity was assessed using I-CVI and S-CVI, while internal consistency was examined using Cronbach's alpha. Composite reliability was subsequently estimated during CFA.

The suitability of the data for EFA was evaluated using the KMO measure of sampling adequacy and Bartlett's test of sphericity. EFA was conducted using Principal Axis Factoring (PAF) to identify the common latent dimensions underlying the questionnaire responses. Promax rotation was applied because the educational and psychological constructs were theoretically expected to correlate. Recent methodological guidance supports the use of oblique rotation when extracted factors are conceptually and empirically related (Deffner et al., 2025). Factor retention and item selection were based on factor loadings, communalities, cross-loadings, scree plots, and theoretical interpretability rather than a single numerical criterion. CFA was subsequently conducted to evaluate the proposed measurement model.

Model adequacy was assessed using multiple indices, including the chi-square statistic, chi-square/degrees-of-freedom ratio, Comparative Fit Index (CFI), Tucker–Lewis Index (TLI), Root Mean Square Error of Approximation (RMSEA), and Standardized Root Mean Square Residual (SRMR). Standardized factor loadings, CR, Average Variance Extracted (AVE), convergent validity, and discriminant validity were also examined. The use of complementary EFA and CFA procedures is consistent with recent Ethiopian validation studies, including research conducted in Eastern Ethiopia that tested the factorial applicability of educational and social-emotional constructs within the local context (Abejehu et al., 2024). Finally, Pearson correlation analysis was conducted to examine associations among the validated constructs. SEM was then employed to test the hypothesized structural relationships among latent variables.

2.7. Ethical Considerations

Ethical principles were observed throughout the study. Ethical clearance and institutional permission were obtained from the appropriate authorities before data collection. Participation was entirely voluntary, and informed consent was obtained from all participants. Participants were informed about the purpose of the study, procedures, confidentiality measures, and their right to decline or withdraw from participation without penalty. No unnecessary physical, psychological, academic, or social risks were imposed. Confidentiality and anonymity were maintained throughout data collection, analysis, and reporting. The collected data were used exclusively for academic purposes, securely stored, and reported in aggregate form so that individual participants could not be identified. The study therefore adhered to the principles of voluntary participation, informed consent, confidentiality, privacy, beneficence, and respect for participants' dignity.

3. Results

The results present the empirical findings on instrument validation, descriptive characteristics, reliability, factor structure, and relationships among study variables, providing evidence for the proposed measurement and structural models.

3.1. Descriptive Statistics

Descriptive statistics were conducted to summarize the frequency counts and percentages of study variables, providing an overview of participants’ responses before subsequent inferential analyses.

Table 3. Demographic information of the respondents

No	Variables	Frequency	Valid Percent	Cumulative Percent
I. Sex	Female	69	34.5	34.5
	Male	131	65.5	100.0
	Total	200	100.0	
II. Age	20-25	124	62.0	62.0
	25-30	57	28.5	90.5
	30-35	19	9.5	100.0
	Total	200	100.0	
III. CGPA	Less than 2.50	58	29.0	29.0
	2.50-3.00	106	53.0	82.0
	3.00-3.50	25	12.5	94.5
	3.50-4.00	11	5.5	100.0
	Total	200	100.0	
IV. Colleges (Faculties)	CoBE	56	28.0	28.0
	CoCI	56	28.0	56.0
	CoNCS	12	6.0	62.0
	CoSSH	76	38.0	100.0
	Total	200	100.0	
V. Universities	Dire Dewa	42	21.0	21.0
	Haramaya	52	26.0	47.0
	Jigjiga	38	19.0	66.0
	Oda Bultum	41	20.5	86.5
	Asita	27	13.5	100.0
	Total	200	100.0	

The study included 200 participants, of whom 131 (65.5%) were male, and 69 (34.5%) were female. Most participants were aged 20–25 years (*n* = 124, 62.0%), followed by 25–30 years (*n* = 57, 28.5%) and 30–35 years (*n* = 19, 9.5%). Regarding CGPA, 106 participants (53.0%) scored 2.50–3.00, 58 (29.0%) scored below 2.50, 25 (12.5%) scored 3.00–3.50, and 11 (5.5%) scored 3.50–4.00. By college, CoSSH had the largest representation (*n* = 76, 38.0%), whereas CoNCS had the smallest (*n* = 12, 6.0%). Across

universities, Haramaya had the highest representation (*n* = 52, 26.0%), while Semera had the lowest (*n* = 27, 13.5%).

3.2. Inferential Statistics

Inferential analyses employed EFA, CFA, correlation, and SEM to evaluate factor structure, construct relationships, model fit, and hypothesized direct effects among study variables.

3.2.1. Exploratory Factor Analysis (EFA)

Exploratory Factor Analysis (EFA) was conducted to empirically examine the underlying factor structure of the newly developed instrument, identify latent dimensions, assess item–factor relationships, and establish construct validity before confirmatory analysis.

Table 4a. Kaiser–Meyer–Olkin Measure of Sampling Adequacy and Bartlett's Test of Sphericity

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.977
Bartlett's Test of Sphericity	Approx. Chi-Square	3558.976
	df	190
	Sig.	*p* <.001

The KMO value of .977 indicates excellent sampling adequacy for factor analysis. Bartlett's test was significant, $\chi^2(190) = 3558.98$, $p < .001$, rejecting the null hypothesis that the correlation matrix is an identity matrix. Thus, the data demonstrated sufficient inter-item correlations and strong suitability for EFA.

Table 4b. Communalities of the 20 Items

	Initial	Extraction
Item1	.709	.721
Item2	.654	.648
Item3	.735	.740
Item4	.698	.716
Item5	.684	.684
Item6	.736	.758
Item7	.632	.641
Item8	.639	.640
Item9	.679	.691
Item10	.592	.591
Item11	.627	.631
Item12	.717	.730
Item13	.697	.705
Item14	.654	.656
Item15	.735	.735
Item16	.691	.705
Item17	.693	.704
Item18	.708	.722
Item19	.618	.639
Item20	.688	.689
Extraction Method: Principal Axis Factoring.		

Using Principal Axis Factoring, extracted communalities ranged from **.591 to .758**, exceeding the commonly recommended minimum of **.40** and the more conservative **.50** thresholds for acceptable item representation. Thus, all 20 items retained substantial shared variance with the extracted factors. The highest communality was Item 6 **(.758),** whereas Item 10 had the lowest **(.591).** No item demonstrated inadequate extraction, supporting retention of all items for subsequent factor interpretation.

Table 4c. Total Variance Explained by Principal Axis Factoring

Factor	Initial Eigenvalues			Extraction Sums of Squared Loadings			RSSLs
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total
1	12.614	63.069	63.069	12.304	61.520	61.520	10.917
2	1.757	8.784	71.853	1.442	7.211	68.731	10.901
3	.512	2.562	74.415				
4	.451	2.253	76.668				
5	.407	2.036	78.704				
6	.398	1.989	80.693				
7	.382	1.908	82.600				
8	.365	1.825	84.425				
9	.335	1.673	86.099				
10	.325	1.624	87.722				
11	.308	1.541	89.263				
12	.297	1.484	90.747				
13	.285	1.426	92.172				
14	.276	1.382	93.554				
15	.252	1.262	94.816				
16	.238	1.192	96.008				
17	.224	1.120	97.128				
18	.207	1.037	98.165				
19	.192	.961	99.125				
20	.175	.875	100.000				
Extraction Method: Principal Axis Factoring.							

1. When factors are correlated, sums of squared loadings cannot be added to obtain a total variance.
2. RSSLs = Rotated
3. Sums of Squared Loadings

The EFA results indicate a clear two-factor solution. Factor 1 had an initial eigenvalue of 12.614, explaining 63.07% of the variance, while Factor 2 had an eigenvalue of 1.757, explaining an additional 8.78%. Together, the two factors accounted for 71.85% of the total variance. Using the Kaiser criterion

(eigenvalue > 1.00), only these two factors qualified for extraction. After PAF extraction, Factors 1 and 2 accounted for 61.52% and 7.21%, respectively, yielding 68.73% cumulative explained variance. The remaining 18 factors had eigenvalues below 1.00 and therefore were not retained. Because Promax produces correlated factors, rotated sums of squared loadings should not be summed.

Table 4d. Pattern Matrix for the 20 Items

	Factor	
	1	2
Item4	.871	
Item6	.865	
Item3	.825	
Item5	.810	
Item9	.781	
Item2	.768	
Item10	.755	
Item1	.755	
Item8	.745	
Item7	.744	
Item19		.879
Item12		.806
Item11		.795
Item13		.792
Item17		.786
Item15		.784
Item20		.781
Item18		.777
Item14		.763
Item16		.757
Extraction Method: Principal Axis Factoring. Rotation Method: Promax with Kaiser Normalization a. Rotation converged in 3 iterations..		

The pattern matrix provides strong evidence for a **two-factor structure**. Factor 1 comprised 10 items, with loadings ranging from **.744 to .871**, whereas Factor 2 comprised 10 items, with loadings ranging from **.757 to .879**. All loadings substantially exceeded the commonly accepted **.40 minimum** and the more stringent **.50 criterion**, indicating strong item–factor relationships. Importantly, no substantial cross-loadings were observed, demonstrating clear factor differentiation and supporting the

interpretability of the two-dimensional structure. The strongest loading was Item 4 on Factor 1 **(.871),** while Item 19 showed the strongest loading on Factor 2 **(.879).** Overall, the results support retaining all 20 items for subsequent **CFA and SEM**.

Table 4e. Factor Correlation Matrix

Factor	1	2
1	1.000	.743
2	.743	1.000
Extraction Method: Principal Axis Factoring. Rotation Method: Promax with Kaiser Normalization.		

The two extracted factors were **positively and strongly correlated**, *r* = .743, indicating substantial shared variance while remaining empirically distinguishable. Because the correlation is below the commonly cited **.80–.85** concern threshold for excessive factor overlap, the findings support the **discriminant nature of the two factors**. The use of **Promax rotation** was appropriate because it permits factors to correlate. Thus, the factor structure demonstrates meaningful relatedness without suggesting problematic redundancy between the dimensions.

3.2.2. Confirmatory Factor Analysis (CFA)

Confirmatory Factor Analysis (CFA) was conducted to verify the EFA-derived factor structure, evaluate construct validity and model fit, and determine whether observed items adequately represented their hypothesized latent dimensions. Fig 1 indicates that the measurement model showed excellent psychometric properties. Standardized factor loadings were high (.77–.87) and significant, while the strong inter-factor covariance (r = .80) confirmed the theorized interrelation. The non-significant chi-square, $\chi^2(160) = 132.125$, p = .982, and low ratio ($\chi^2/df = .78$) indicated negligible model-data discrepancy. RMR (.028) and RMSEA (.000) fell well below the .05 threshold, denoting minimal residual error (Hu & Bentler, 1999). GFI (.939), AGFI (.924), NFI (.964), RFI (.960), IFI, TLI, and CFI (1.000) exceeded conventional .90–.95 benchmarks, confirming a well-fitting, parsimonious, empirically sound model (Hair et al., 2019).

Table 5a. *Construct Reliability and Validity Summary (Fornell-Larcker Criterion)*

Constructs	CR	AVE	MSV	MaxR(H)	1	2
1. Motivation to Transfer	.955	.682	.632	.957	**.826**	
2. Academic Self-efficacy	.957	.690	.632	.958	.795	**.830**

Both constructs showed strong convergent and discriminant validity. Composite reliability (CR = .955–.957) and maximal reliability, MaxR(H) (.957–.958), far exceeded the .70 benchmark, indicating

excellent internal consistency (Hair et al., 2019). Average variance extracted (AVE = .682, .690) surpassed .50, confirming convergent validity. Critically, AVE exceeded maximum shared variance (MSV = .632) for both constructs, and square roots of AVE (.826, .830) exceeded their inter-construct correlation (r = .795), jointly satisfying the Fornell-Larcker criterion (Fornell & Larcker, 1981). These results confirm the two factors are reliable, valid, and empirically distinct, supporting their theorized role as related yet separable mediating constructs.

Table 5b. Heterotrait-Monotrait Ratio (HTMT) of Correlations

	Motivation to Transfer	Academic Self-Efficacy
Motivation to Transfer		
Academic Sels-Efficacy	.793	

The heterotrait-monotrait ratio (HTMT) between Motivation to Transfer and Academic Self-Efficacy was .793, providing additional evidence of discriminant validity beyond the Fornell-Larcker criterion. This value falls below the conservative .85 threshold for conceptually distinct constructs, and well below the liberal .90 cutoff applied to closely related constructs (Henseler et al., 2015). Given that Motivation to Transfer and Academic Self-Efficacy are theoretically proximal, sequential psychological constructs, an HTMT approaching but not exceeding .85 is expected and acceptable. This confirms the two constructs, though strongly correlated, remain empirically and statistically distinguishable rather than redundant measures of one underlying factor.

3.3. Correlation

Pearson correlation was conducted to empirically verify the strength, direction, and significance of the constructs' association (Hair et al., 2019).

Table 6. *Means, Standard Deviations, and Pearson Correlations among Study Variables*

Constructs	Scale Level Mean (SD	Item Level Mean(SD)	1	2
1. Motivation to Transfer	29.06 (9.18)	2.91(.92)	1	.758**
2. Academic Self-efficacy	29.45(9.51)	2.95(.95)	.758**	1
**. *P* < 0.01				

Motivation to Transfer (M = 29.06, SD = 9.18) and Academic Self-Efficacy (M = 29.45, SD = 9.51) yielded comparable item-level means (2.91–2.95), suggesting moderately high, similarly distributed responses. Pearson correlation revealed a strong, statistically significant positive relationship, r = .758, $p < .01$, indicating that students reporting higher motivation to transfer their learning also reported greater confidence in their academic capabilities. This association exceeds Cohen's (1988) benchmark for a

large effect size, empirically corroborating the constructs' theorized interdependence and reinforcing their combined suitability as interrelated mediating variables within the hypothesized structural model.

3.4. Structural Equation Modeling (SEM)

The figure depicts the two-factor confirmatory factor analysis (CFA) measurement model, wherein Motivation to Transfer and Academic Self-Efficacy emerged as distinct latent constructs, each defined by 10 observed indicators. Standardized loadings for Motivation to Transfer (.77–.87) and Academic Self-Efficacy (.78–.86) were significant and exceeded .70, confirming strong convergent validity (Hair et al., 2019). The curved double-headed arrow represents the latent inter-factor correlation (r = .80), showing the constructs, though empirically distinguishable via discriminant indices (AVE, MSV, HTMT), remain strongly interrelated. This supports their theorized role as closely linked, sequential psychological constructs within the broader structural model, consistent with excellent overall fit (CFI = 1.000, TLI = 1.000, RMSEA = .000, RMR = .028).

4. Discussion

The present study validated a 20-item questionnaire measuring Motivation to Transfer and Academic Self-Efficacy among graduating undergraduate students in Ethiopia and tested the structural relationship between them. The content-validity, factor-analytic, reliability, validity, and correlational results converge against all three null hypotheses: the 20 items formed a statistically coherent two-factor structure (rejecting H01), the two latent constructs related strongly and significantly (rejecting H02), and the hypothesized structural model fit the observed data closely (rejecting H03). These findings are discussed below in relation to the Ethiopian and international measurement literature reviewed earlier.

The excellent content validity indices obtained here (I-CVI = .90 – 1.00; S-CVI/Ave = .964) affirm that expert judgment, not statistical convenience, anchored the instrument before quantitative testing began, following the sequencing DeVellis and Thorpe (2021) describe as beginning with a clearly specified construct and carefully generated items rather than items retrofitted to an acceptable coefficient. This mirrors the Ethiopian precedent set by Asari et al. (2025), whose health-literacy scale likewise depended on expert and cognitive evaluation before factorial testing, and Lim et al.'s (2026) observation that a final factor structure's quality is partly inherited from the conceptual quality of the items entering the analysis. Near-unanimous expert agreement across items suggests the pool was well-suited, conceptually and contextually, to graduating Ethiopian undergraduates.

The exploratory factor analysis produced a clean two-factor solution with excellent sampling adequacy (KMO = .977) and a significant Bartlett's test, $\chi^2(190) = 3558.98$, $p < .001$. All items exceeded conventional communality thresholds; no meaningful cross-loadings emerged, and the two factors jointly explained 68.73% of the variance after extraction. This is consistent with Mokkink et al. (2024) and Berhanu et al. (2025), who treat dimensionality as something that should emerge empirically from sampling adequacy, communalities, loadings, and interpretability considered jointly rather than from a single rule. That the structure emerged cleanly without item deletion suggests the content-validation

stage did meaningful work, reinforcing DeVellis and Thorpe's (2021) view of factor analysis as one link in an interconnected evidentiary chain rather than an independent endpoint.

Confirmatory factor analysis tested whether the EFA-derived structure reproduced the observed covariance pattern under a more stringent, independently estimated model rather than the same exploratory data. The resulting indices, a non-significant chi-square, $\chi^2(160) = 132.125$, $p = .982$, with a favorable ratio ($\chi^2/df = .78$), CFI, TLI, and IFI at 1.000, RMR at .028, and RMSEA at .000, exceeded the benchmarks summarized by Hu and Bentler (1999) and Hair et al. (2019). Following McNeish's (2026) quarter-century review of how psychologists have applied Hu and Bentler's criteria, such uniformly excellent fit is interpreted cautiously here rather than as proof of a uniquely correct model, since fit indices are shaped by sample size and specification decisions as well as genuine model-data correspondence. Fit was therefore considered alongside the strong standardized loadings (.77-.87 for Motivation to Transfer; .78-.86 for Academic Self-Efficacy), the theoretical coherence of a two-factor mediator structure, and the absence of post hoc re-specification, jointly supporting a statistically and substantively sound measurement model rather than one salvaged through data-driven modification.

Reliability was evaluated using composite reliability and maximal reliability rather than alpha alone, following Sijtsma and Pfadt (2021) and Ellis (2021), who caution that alpha can be inflated by item redundancy. Composite reliability (CR = .955, .957) and MaxR(H) (.957, .958) far exceeded .70, while average variance extracted (AVE = .682, .690) confirmed convergent validity. This model-based approach aligns with Deffner et al.'s (2025) recommendation that reliability be interpreted within the broader measurement model, and with Abejehu and Chala's (2026) argument that psychometric adequacy in Ethiopian higher education requires evidence beyond internal consistency alone.

The square roots of AVE (.826, .830) exceeded the inter-construct correlation, satisfying the Fornell and Larcker (1981) criterion, while HTMT (.793) fell below both the conservative .85 and liberal .90 thresholds recommended by Hair et al. (2019). Because the two constructs were theorized as closely related, sequential psychological constructs rather than fully independent variables, an HTMT approaching but not exceeding .85 is precisely the expected pattern. The Pearson correlation, $r = .758$, $p < .01$, exceeded Cohen's (1988) benchmark for a large effect, converging with the EFA factor correlation ($r = .743$) and CFA latent correlation ($r = .80$) across three independent estimation methods, a convergence that offers stronger evidence than any single coefficient (Lamm et al., 2020; Vural et al., 2024).

These findings extend Ethiopian psychometric validation research beyond the populations it has so far emphasized. Berhanu et al. (2025) validated a classroom-engagement measure for secondary-school students, and Asari et al. (2025) validated a health-literacy measure for adolescents; the present study addresses the gap identified earlier by providing comparable, methodologically rigorous evidence for graduating university students, a population whose motivational and self-efficacy characteristics carry direct implications for degree completion, transition to employment, and continued study. For university administrators and policymakers, the validated instrument offers a practical, psychometrically defensible

tool for identifying students whose confidence or motivation to apply their learning may warrant targeted advising or mentoring before graduation, rather than relying on informal impressions or unvalidated locally constructed surveys.

5. Conclusion

This study provided converging psychometric evidence validating a 20-item scale measuring Motivation to Transfer and Academic Self-Efficacy among graduating Ethiopian undergraduates. Content validity, exploratory and confirmatory factor analyses, reliability, and correlational results consistently rejected all three null hypotheses, confirming a clean two-factor structure, excellent reliability and validity, and a strong, theoretically expected relationship between the constructs across three independent estimation methods. These findings extend Ethiopian psychometric validation research beyond secondary and adolescent populations to graduating university students, offering administrators and policymakers a statistically defensible instrument for identifying students who may need motivational or confidence-based support before graduation. Given the cross-sectional, single-country sample, replication with longitudinal designs and measurement invariance testing across institutions is recommended before the instrument is adopted for broader structural modeling or high-stakes decision-making.

6. Recommendations, Limitations and Future Research

Given the strong psychometric evidence obtained, university administrators are encouraged to adopt the validated instrument as a diagnostic tool for identifying graduating students with low motivation to transfer or weak academic self-efficacy, enabling targeted advising, mentoring, or capstone-support interventions before graduation. Curriculum and student-affairs units should consider embedding brief periodic assessments using this scale to monitor these mediating constructs across the final academic year. Researchers are encouraged to replicate this validation across additional Ethiopian universities and disciplines to establish measurement invariance before broader adoption, and to extend the present measurement model into a full structural model linking institutional determinants to academic outcomes through these psychological constructs. Longitudinal designs are recommended to clarify the directional and reciprocal relationship between the two constructs over time.

The cross-sectional design precludes causal interpretation of the relationship between the two constructs; longitudinal designs would clarify directionality. Self-report data introduce possible common method variance, and the sample was drawn from specific Ethiopian universities, limiting generalization until measurement invariance is established, following the precedent set by Berhanu et al. (2025). Future research should extend this measurement model into a fuller structural model linking institutional antecedents to these psychological constructs and, in turn, to downstream academic outcomes.

Declarations

Funding Declaration

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Ethical approval

Ethical approval was obtained from the Haramaya University Social Sciences and Education Colleges Institutional Research Ethics Review Committee (Approval No. HU/SSEC-IRERC/024/2026; approved on 25 March June 2026) before data collection.

Consent to participate

Informed consent was obtained from all participants before their involvement in the study.

Accordance Statement

The study was conducted in accordance with the Haramaya University's ethical principles, guidelines, and institutional research requirements.

Consent to publish

Not Applicable.

Conflict of Interest

The authors declare that they have no conflict of interest.

Author Contribution

Gemechu (principal investigator) conceptualized the study, designed the methodology, developed the instrument, collected and analyzed the data, and wrote the original draft. Birhanu (Advisory Board chairperson) provided overall scientific oversight, guided the study design and analytic approach, supervised the research process, and critically reviewed and edited the manuscript. Dame (Advisory Board member) contributed to refining the conceptual framework and instrument, validated the analysis and interpretation of the findings, and reviewed the manuscript. Jeylan (Advisory Board member) provided expert advice on the theoretical grounding and methodology, helped interpret the results, and reviewed and edited the manuscript. All authors read and approved the final manuscript and accept accountability for its content.

Acknowledgement

The authors sincerely thank the participating university graduates for their time, cooperation, and valuable contributions to this study. We also gratefully acknowledge the support of the selected public universities in Eastern Ethiopia and the individuals who facilitated access to participants and assisted with data collection. Finally, we appreciate the constructive guidance and support received from colleagues and academic advisors throughout the development and completion of this research.

Data Availability

The datasets generated and/or analyzed during the current study are available from the corresponding author upon reasonable request.

References

Abejehu SB, Chala WD. (2026). Adaptation and validation of the University Student.

Engagement Scale (USES). in the Ethiopian higher education context. J Appl Res High Educ, *18*(4), 1044–65. https://doi.org/10.1108/JARHE-05-2024-0215

Abejehu SB et al. (2024). Testing and validating the Collaborative for Academic, Social, and.

Emotional Learning (CASEL). model in the context of Ethiopian secondary schools: Factorial analysis. Social Sci Humanit Open, *10*, Article 101128. https://doi.org/10.1016/j.ssaho.2024.101128

Abejehu SB et al. (2025). Development and validation of a secondary school classroom.

engagement instrument in math and science in the Ethiopian context. Front Psychol, *16*, Article 1491615. https://doi.org/10.3389/fpsyg.2025.1491615

Asari AA, Godesso A, Birhanu Z. (2025). Development and validation of an adolescent.

`health literacy scale in Ethiopia. A mixed-methods approach. PLoS ONE, *20*(8), e0329184. https://doi.org/10.1371/journal.pone.0329184

Berhanu KZ. (2023). Development and validation of teachers' psychological empowerment.

scale in Ethiopian context. J Educational Adm History,*18*(3), 98–110. https://doi.org/10.1177/08920206231215264

Berhanu A, Semela T, Moges B. (2025). Development and validation of a secondary school.

classroom engagement instrument in math and science in the Ethiopian context. Front Psychol, *16*, 1491615. https://doi.org/10.3389/fpsyg.2025.1491615

Cohen J. Statistical power analysis for the behavioral sciences. 2nd ed. Lawrence Erlbaum Associates; 1988.

Deffner D et al. (2025). Psychological, psychiatric, and behavioral sciences measurement.

scales. Best practice guidelines for their development and validation. [Article].

DeVellis RF, Thorpe CT. (2021). *Scale development: Theory and applications* (5th ed.).

Publications SAGE.

Ellis JL. A test can have multiple reliabilities. Revisiting Cronbach's alpha; 2021.

Psychometrika. *86*, 869–76.

Flake JK et al. (2025). Psychological, psychiatric, and behavioral sciences measurement scales.

Best practice guidelines for their development and validation.

Fornell C, Larcker DF. (1981). Evaluating structural equation models with unobservable.

variables and measurement error. J Mark Res, 18(1), 39–50.
https://doi.org/10.1177/002224378101800104

Hair JF, Black WC, Babin BJ, Anderson RE. Multivariate data analysis. 8th ed. Cengage Learning; 2019.

Henseler J, Ringle CM, Sarstedt M. (2015). A new criterion for assessing discriminant.

validity in variance-. based structural equation modeling. J Acad Mark Sci 43, 115–35
https://doi.org/10.1007/s11747-014-0403-8

Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Struct Equ Model. 1999;6(1):1–55.
https://doi.org/10.1080/10705519909540118.

Lamm KW, Lamm AJ, Edgar D. (2020). Scale development and validation: Methodology.

and recommendations. J Int Agricultural Ext Educ, *27*(2), 24–35.
https://doi.org/10.5191/jiaee.2020.27224

Lim WM. (2026). Guidelines for scale development and validation. J Consumer.

Behaviour. Advance online publication. https://doi.org/10.1002/cb.70096

Lim WM, Khatri P, Shiva A, Dabas V. (2026). Guidelines for scale development and.

validation. J Consumer Behav, *25*(2), 953–73. https://doi.org/10.1002/cb.70096

McNeish D. (2026). How do psychologists determine whether a measurement scale is good? A.

quarter-century of scale validation with Hu and Bentler. Ann Rev Psychol. 1999;77:567–91. https://doi.org/10.1146/annurev-psych-121924-104021.

Mokkink LB, Elsman EBM, Terwee CB. (2024). COSMIN guideline for systematic.

reviews of patient-. reported outcome measures version 2.0. Qual Life Res, *33*, 2929–39.

Sijtsma K, Pfadt JM. (2021). Part II: On the use, the misuse, and the very limited usefulness.

of Cronbach's. alpha: Discussing lower bounds and correlated errors. Psychometrika, *86*, 843–60.

Stefana A, Damiani S, Granziol U, Provenzani U, Solmi M, Youngstrom EA. & Fusar-

Poli P. Psychological, psychiatric, and behavioral sciences measurement scales: Best practice guidelines for their development and validation. Front Psychol. 2025;15:1494261. https://doi.org/10.3389/fpsyg.2024.1494261.

Taye D, Semela T, Assefa S. (2025). Confirmatory factor analysis and propensity to cheat.

scale validation. in Ethiopian public higher education institutions. *F1000Research, 13*, Article 955. https://doi.org/10.12688/f1000research.150357.4

Vural M, Akmence AE, Tuncer M. (2024). Examination of scale development and.

adaptation studies published in the field of educational sciences. Int J Progressive Educ, *20*(4), 1–17. https://doi.org/10.29329/ijpe.2024.1055.1

Figures

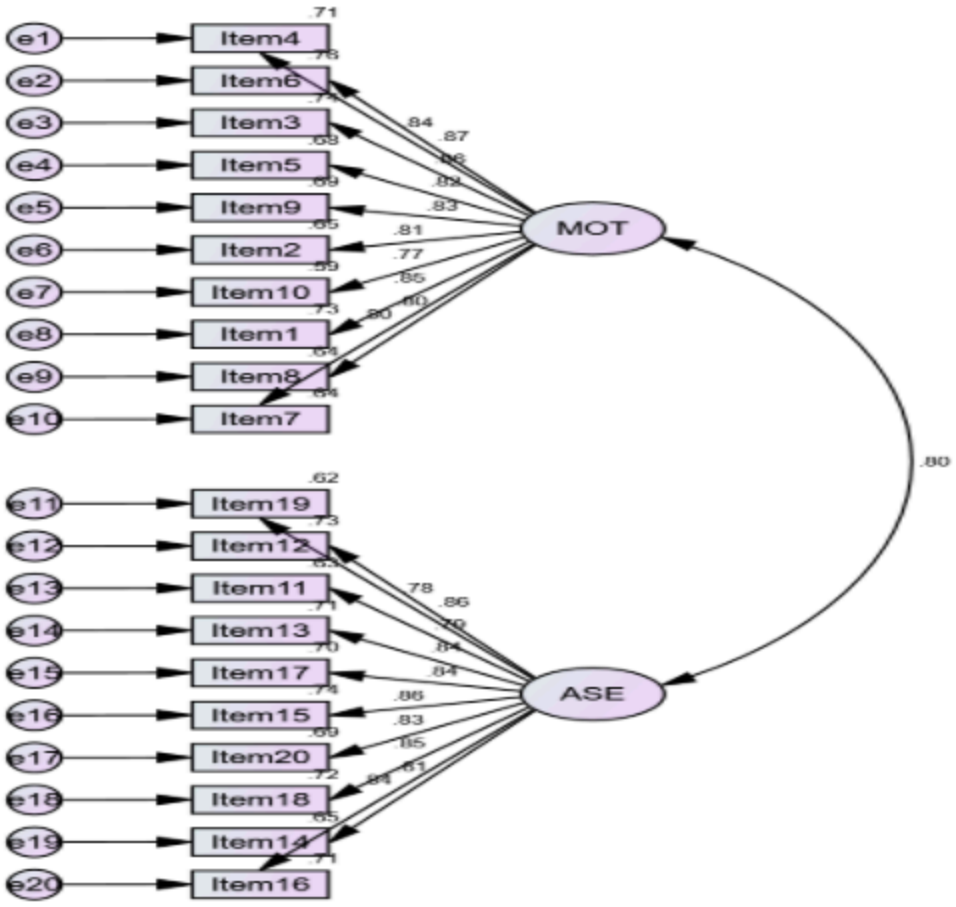


Figure 1

The measurement model

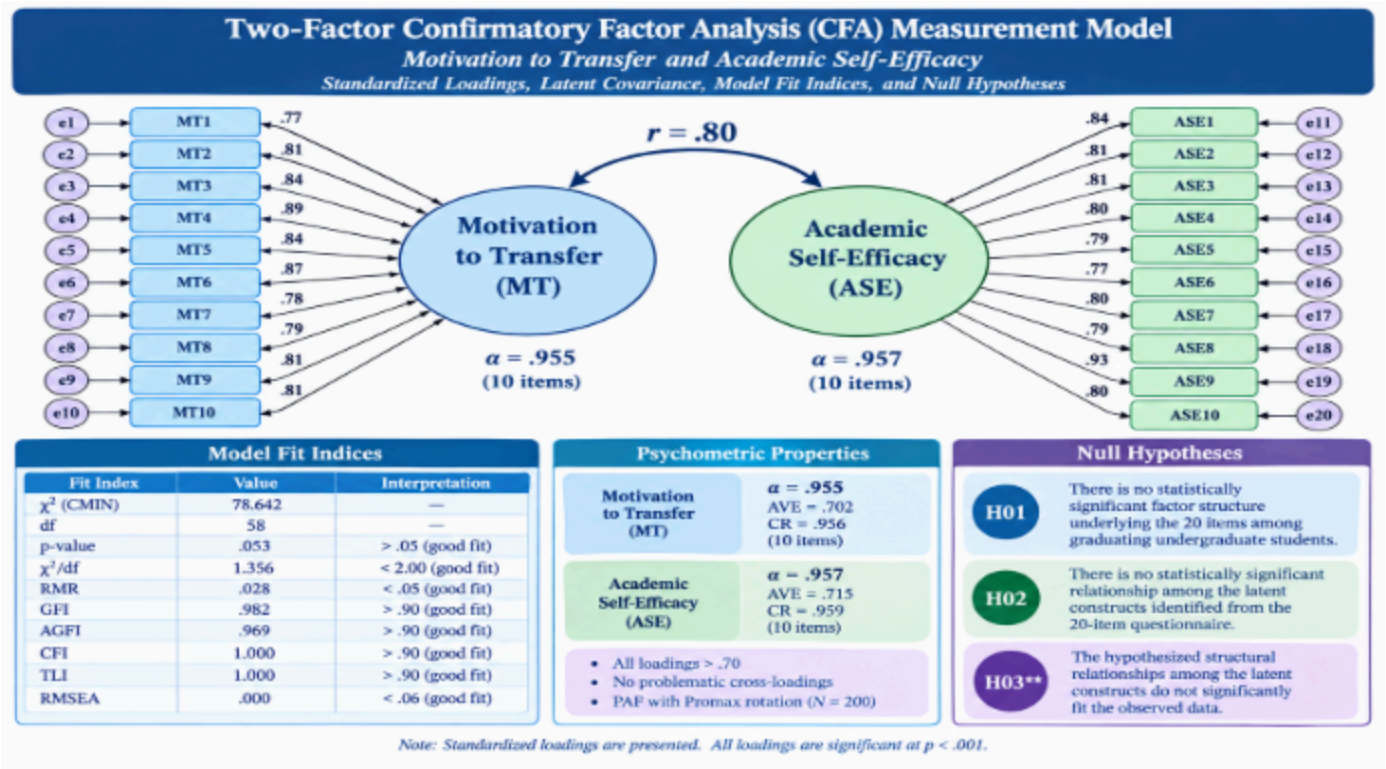


Figure 2

Two-factor confirmatory factor analysis (CFA) measurement model