

Developing and Validating A Students' Underperformance Indicator Instrument For The Ethiopian Secondary School Leaving Certificate Examination: Evidence From Eastern Ethiopia

Abera Assefa^{1*}, Garkebo Basha², Wondifraw Dejene³, Ambissa Kenea⁴, Mathivanan Dhamodaran⁵

¹ Research Scholar, Department of Curriculum Studies, Haramaya University, Ethiopia.

² Associate Professor, Department of Curriculum Studies, Haramaya University, Ethiopia.

³ Associate Professor, Department of Curriculum Studies, Dire Dawa University, Ethiopia.

⁴ Professor, Department of Curriculum Studies, Addis Ababa University, Ethiopia.

Corresponding author: Mathivanan Dhamodaran-MU-SSA.

Abstract

This study aims to develop and validate a multi-item instrument designed to measure the students' underperformance indicators on the Ethiopian secondary school-leaving certificate examination. Accordingly, a cross-sectional retrospective design was adopted. Data were collected from 602 students enrolled in TVET and remedial programs across four major cities in eastern Ethiopia: Chiro, Harar, Dire Dawa, and Jigjiga, using stratified random sampling. These participants were among those who failed to meet the 50% threshold for the 2022/23 national examination and enrolled in either TVTE or remedial programs. In the first instrument development stage, literature reviews, expert opinions, and content evaluations were conducted, and a questionnaire consisting of 30 items, divided into three components, rated on a 5-point Likert scale was developed. In the subsequent validation stage, the total sample was divided into two halves ($n=301$), with one half designated for either EFA or CFA and analyzed using the JASP for Windows 10 application. The results of the EFA indicated a sampling adequacy ($KMO = 0.93$; Bartlett's test, $\chi^2 = 4572.82$, $p < .001$), leading to retaining all 30 items with a three-factor solution, all having item loadings of $\geq .60$. The CFA supported the proposed structure, depicting acceptable fit indices: ($\chi^2/df = 1.05$, $RMSEA = 0.013$, $SRMR = 0.041$, $CFI = 0.995$, and $TLI = 0.995$). The reliability and validity of the instrument were established with Cronbach's alpha ranging from 0.909 to 0.916, composite reliability greater than 0.70, and average variance extracted remaining ≥ 0.50 . Discriminant validity was also confirmed. Generally, the finalized instrument exhibited strong psychometric properties, making it a valid and reliable tool for assessing students' underperformance in Ethiopia's secondary school national examination context.

Keywords: Students Underperformance, ESSSCE, Psychometric Validation, Exploratory Factor Analysis, Confirmatory Factor Analysis.

1. Introduction

The Ethiopian Secondary School Leaving Certificate Examination (ESSSCE), also known as the matriculation exam, is given at the end of secondary school, grade 12. This national exam is administered across the country under the auspices of the Ministry of Education [MoE] by the Educational Assessment and Examination Services [EAES]. The exam performance determines the students' employment, admission to higher education and, in many cases, what they will study. This places significant weight on students, combined with intense competition driven by a substantial rise in student enrollment, as reported by the MoE (2022). For example, prior to the 2021/22 educational policy, one study showed that 80.5% of secondary school students admitted to some level of academic dishonesty (Chala, 2021).

As part of combating this long-standing and widespread cheating trend in ESSSCE and strengthening its credibility, the MoE relocated the exam administration from secondary schools to university campuses, accompanied by strengthened supervision. During a one-week-long campus stay of examiners and examinees, all expenses are covered by the MoE, including accommodation and transportation. Following this exam administration, Tefera et al. (2024) reported a significant decrease in the number of cheating cases.

However, this reform was accompanied by a sharp drop in passing rates, from 40–60% before the 2020/21 reform to just 3–8% afterward (Alford, 2024; Tefera & Zelalem, 2025), and this is also reflected at the school level. For instance, in the 2021/22 academic year, 1,161 out of 2,959 secondary schools reported having no students who passed. This figure increased to 1,328 of 3,106 schools in 2022/23 (Borkena, 2023), and by 2023/24, 1,363 of 3,792 schools recorded no passing students (Borkena, 2024; MoE, 2025).

This prompted extensive discussion among the general public, who characterized the situation as systemic crisis rather than individual exam failure (Monitor, 2023; Tamrat, 2023b). Moreover, Ethiopia's Parliament, during its 2023/24–2024/25 sessions, further identified the decline as a pressing national development concern. Responding to this issue also requires identifying the underlying reasons for students' underperformance in the ESSLCE, rather than simply tracking pass and failure rates after the national examination results are announced. And so, this study aims to develop and validate diagnostic instruments that capture underperformance from multiple perspectives, rather than merely measuring underperformance as an outcome. Considering this, the study integrates two complementary systems-based models: Scheerens' (1990) Context–Input–Process–Output (CIPO) framework and Astin's (1993) Input–Environment–Output (I-E-O) model. Both frameworks conceptualize educational outcomes as products coming through a system but at different levels of educational process detail. The two models complement each other in that CIPO provides a structural map outlining where conditions originate (context), while I-E-O focuses on how these conditions are transformed into learning experiences (process).

Within this perspective, when used together, the CIPO and I-E-O frameworks, directly applied to this study, inform the three domains: the (SU-II) student-related items reflect the input dimension, teacher-related items operationalize the process, and contextual items capture the broader systemic context. This framework ensures that the instrument is not only empirically validated but also conceptually coherent within established systems-based models of schooling. The contextual layer is particularly relevant, as the ESSLCE result varies across Ethiopia's regions, and a single, unified explanation of underperformance is unlikely to apply equally to all regions. Even within these major cities of Eastern Ethiopia, average ESSLCE scores have often exceeded the national mean in Harar and Dire Dawa following the new exam administration in 2021/22, while remaining comparatively low in Chiro and Jijiga (MoE, 2022).

Closely related to this topic, two recent studies were identified. One of these, by Taye et al. (2025), addressed students' tendency to cheat in Ethiopian universities and was grounded in the Theory of Planned Behavior. In fact, this instrument demonstrates rigorous psychometric development and validation, but its focus was only on academic dishonesty rather than the broader students' underperformance contributing factors. The other instrument comes from the Philippines, where Honra (2022) developed a scale for students' academic resilience. However, the study relied solely on exploratory factor analysis, without confirmatory analysis or consideration of related factors such as teacher and system-level content. Neither instrument has been developed or validated within secondary school national exam contexts, nor do they integrate student, teacher, and contextual factors into a single validated structure.

Thus, no validated tool currently exists that may capture students' underperformance in ESSLCE involving student-, teacher, and context-related indicators. Such an instrument would situate underperformance within a systems-based model of schooling and validate this definition through both exploratory and confirmatory analyses in a national exam context. Accordingly, the purpose of this study was to develop and validate the Students' Underperformance Indicator Instrument (SU-II) within the ESSLCE context in Eastern Ethiopia.

2. Materials and methods

2.1. Study Design

In this research, a cross-sectional retrospective design was used. In this context, Pusbelina and Sutrisno (2025) highlight that this type of survey involves distributing questionnaires in a specific location. A deliberate decision was made to employ this retrospective self-report design. Specifically, it allowed students who scored below 50% in the 2022/23 ESSLCE to recall and assess the reasons for the underperformance. The recall window was kept relatively short, accordingly engaging first-year students in remedial and TVET programs in the year immediately following the exam to limit recall decay.

2.2. Population and Setting

The target population consisted of first-year students who failed to meet the 50% threshold in the 2022/23 ESSLCE and subsequently enrolled in remedial or TVET programs across four major eastern Ethiopian cities: Chiro, Harar, Dire Dawa, and Jijiga. These urban centers serve as educational hubs due to the presence of large universities and polytechnic colleges, together accounting for an accessible population of $N = 11,403$. Related with this, it is important to note that the scope excludes students who exited the education system or joined the workforce after

exam failure. Thus, the findings reflect the retrospective perceptions of those who remained within formal education pathways.

2.3. Sampling

The research was conducted in February in the first semester of 2024/25. To determine the sample size, Kyriazos's (2018) absolute lower-limit fixed standard was used. Related with exploratory factor analysis (EFA) and confirmatory factor analysis (CFA), the absolute lower-limit fixed standard set was $N = 200$, with thresholds defined as poor (< 100), moderate (< 150), adequate (< 200), good (< 250), very good (> 300), and excellent (> 400). Considering this, a sample of 602 participants was recruited.

After this, with a split-sample cross-validation method, half ($n=301$) was used for EFA and the other half was used for CFA. In support of this, Kyriazos (2018) asserted that randomly dividing one large sample into two independent subsamples is a well-established practice, specifically to avoid the inflated fit that comes from running EFA and CFA on the same data. This also satisfies a common rule of thumb for EFA or CFA is the 10:1 item-to-participant ratio per subsample (Hair et al., 2018; Tabachnick and Fidell, 2019).

In keeping with this, among the program participants, 326 (54.2%) were enrolled in remedial programs, while 276 (45.8%) were enrolled in the TVTE program. As mentioned previously, the sample was collected from four cities, employing proportional stratified random sampling, with participants from Harar (160, 26.58%), Dire Dawa (198, 32.72%), Jijiga (111, 18.44%), and Chiro (134, 22.26%). The sample included secondary school students from both urban and rural schools; gender balance and program distinctions were preserved within subsamples following random allocation to EFA and CFA.

2.4. Questionnaire Development

The questionnaire should develop through a systematic, multi-step process, as outlined by Ngah et al. (2022). It has to encompass item generation, expert review, and cognitive interviewing. Considering this rigor process, this questionnaire was developed as followed:

2.4.1. Item Generation: The initial pool of (35) items was generated by reviewing a comprehensive body of literature, especially the most existing systematic review works and few full flagged articles with commentaries in various electronic media published after the new examination procedure was undertaken in 2021/22 (e.g., Borkena, 2023; Demissie, 2024; MoE, 2023; Molla and Tiruneh 2023; Monitor, 2024; Mosisa et al., 2024; Tamrat, 2023a,b; Tefera and Zelalem, 2025; Tefera et al., 2024; Tiruneh and Molla, 2024) and international studies (e.g., Reyno & Guzman, 2025; Umeh & Hamzah, 2025). The items were mapped accordingly, based on the combination of the domains of the CIPO/I-E-O, namely $CR (11) + TR (12) + SR (12) = 35$. After this, the first pool was looked at by the researchers.

2.4.2. Expert Review: As a preliminary screening prior to expert review, the items were reduced as follows: $CR (11 \rightarrow 10)$, $TR (12 \rightarrow 10)$, and $SR (12 \rightarrow 10)$. This reduction step was guided by the researchers' judgment, considering redundancy among items measuring the same indicator, clarity of wording, and an initial construct-fit check against the CIPO/I-E-O domain. Then, these 30 items were subjected to expert review to establish content validity. Accordingly, a panel of five experts evaluated the instrument content validity: two curriculum specialists, two educational psychologists, and one expert in educational measurement, all experienced researchers. Polit and Beck (2006) recommended the applications of CVI method commonly use panels of five to ten. Based on expert qualitative feedback and quantitative scoring, items were revised, and a content validity index (CVI) was calculated using a 4-point relevance scale (1 = not relevant, 2 = somewhat relevant, 3 = quite relevant, 4 = highly relevant) to quantify expert agreement. The minimum threshold of 0.80 for 5-member expert panels was expected, and for the initial 30-item draft, it has yielded an S-CVI/Ave of 0.86. This confirmed high expert consensus and strong content validity before pilot testing.

2.4.3. Cognitive Interviewing: it was conducted prior to the actual distribution of the survey to identify the items with ambiguous wording, potential misinterpretations, and difficulties in responding. Subsequently, using Willis (2004) probing method, the interviewers asked targeted questions such as: "What does this term mean to you?" - Comprehension /definition probe, "Can you tell me in your own words what this question is asking?, paraphrasing probe, "Was there any word or phrase that felt confusing or unclear?", - general/difficulty probe, and "Do you think this question concerns your personal effort, your teacher's role, or another factor ?" - Category-selection. In this regard, six first-year diploma program students from Harar Health Science College (four males, two females) and three male secondary school teachers participated in the cognitive interviewing sessions. This number of

participants was justified by saturation, as cognitive interviewing typically yields diminishing returns beyond 5–10 interviews, a threshold commonly cited in the methodological literature (Willis, 2004). These single-round interviews were conducted in Amharic and Oromifa, in both health Science College and secondary school settings, with each session lasting approximately 40 minutes. Based on participant feedback, ambiguous wording was subsequently refined; for example, "absence of commitment" was changed to "lack of commitment," "students' failure rate" to "students' underperformance rate," and "absence of study habits" to "poor study habits."

2.5. Item pool

After the necessary steps of the instrument development process described in Section 3, the research instrument was developed. As indicated in Table 1, the instrument consisted of 30 items grouped into three sub-dimensions: SR, TR and CR factors. These elements are assumed to contribute to students' underperformance in the ESSLCE, as supported in the local literatures (Demissie, 2024; Mosisa, et al., 2024; Tiruneh & Molla, 2024; Tamrat, 2023a; Tefera et al., 2024; Tefera & Zelalem, 2025). The students' underperformance indicator instrument (SU-II) for the ESSLCE was developed considering a 5-point Likert scale (1 = strongly disagree, 2 = disagree, 3 = neutral, 4 = agree, and 5 = strongly agree).

Table 1. The Students' Underperformance Indicator Draft Instrument for the ESSLCE

Item	Diminutions	Statement
Q1S	Student-Related (SR)	I experienced a significant lack of commitment necessary to study for the examination.
Q2S		My result was due to gaps in my prior learning or foundational knowledge necessary for the exam material.
Q3S		I struggled significantly with poor study habits and did not allocate sufficient preparation time for the examination.
Q4S		Work commitments (income-earning activities or house chores) frequently led to my absenteeism or inability to dedicate time to studying.
Q5S		I struggled with tasks requiring high-level skills, such as critical thinking and analytical reasoning.
Q6S		An overload of fear of failure significantly hindered my ability to perform during the examination..
Q7S		I am generally not interested in continuing my study or often consider dropping out, which affected my effort level.
Q8S		I struggled because I was unfamiliar with the ESSLCE test format
Q9S		My insufficient competency in the instructional language of the curriculum made it difficult to fully understand lessons and examination questions.
Q10S		My performance suffered because I engaged in indiscipline, truancy, or unruly behaviors, resulting in frequent absenteeism
Q1T	Teacher – Related(TR)	My teachers often came to class unprepared for the lesson
Q2T		My teachers did not seem to know the subject well enough to teach it
Q3T		My teachers did not prepare me well for the exam using practice tests or relevant assessment techniques,
Q4T		My teachers rushed through the lessons or did not cover all the important chapters of the curriculum
Q5T		The way my teachers taught (instructional style) was ineffective and did not help me understand the concepts,
Q6T		My teachers struggled to keep the class orderly (manage routines), which made it hard for me to learn and focus,
Q7T		My Teachers avoid monitoring /checking on my academic progress or helping me closely.
Q8T		I believe my teachers seemed unhappy or lacked motivation to teach well
Q9T		My teacher was frequently late or absent (absenteeism/tardiness), causing us to lose

		valuable learning time,
Q10T		My teachers rarely used technology to support their teaching,
Q1C		My class is too large to get individual help from the teacher.
Q2C		My classes do not prepare me well for the national examinations.
Q3C		My school lacks necessary facilities (classrooms, labs, or library) including textbooks or reference materials for the exam
Q4C	Contextual-Related (CR)	My parents or guardians rarely involve themselves in my education or school progress
Q5C		My school provided career counseling or psychological support related to the stress of preparing for the ESSLCE.
Q6C		The policy that prevents students from repeating a grade hurt my performance by moving me ahead unprepared.
Q7C		What I study is not relevant to real-world jobs or my future.
Q8C		Weak leadership and management at my school hurt my academic performance.
Q9C		The government focuses too much on enrolling new students rather than ensuring high quality of teaching
Q10C		The constant and quick changes in government educational policies make it hard to focus and succeed.

Note: Q3S & Q4S were negatively worded items and reversed during data entry

2. 6. Pilot Testing

2.6.1. Reliability Analysis (Cronbach's α , McDonald's ω)

The Cronbach's alpha coefficients for the three dimensions were calculated in Dire Dawa using 60 students, 30 from the preparatory and 30 from the TVTE sections, randomly selected. To avoid overlap with the main study, only students excluded from the primary sample participated in this pilot study. As indicated in Table 2, the Student-Related (SR) factor (10 items) revealed good internal consistency ($\alpha = .82$, $\omega = .83$), while the Teacher-Related (TR) factor (10 items) showed acceptable reliability ($\alpha = .79$, $\omega = .80$). Similarly, the Contextual-Related (CR) factor (10 items) exhibited good internal consistency ($\alpha = .85$, $\omega = .86$). These three dimensions were above 0.70, and confirmed that are internally consistent and suitable for subsequent explanatory factor analysis.

Table 2: Reliability Analysis

Factor	Items	Cronbach's α	McDonald's ω	Interpretation
SR	10	.82	.83	Good internal consistency
TR	10	.79	.80	Acceptable internal consistency
CR	10	.85	.86	Good internal consistency

3. Results

3.1. Psychometric Validation Procedures

In this psychometric analysis, EFA and CFA were mainly conducted consecutively with the distinct purpose to identify the underlying structure of the variables and a cluster of constraints that together measure the same latent factor, and then through CFA to validate the predefined components identified in the first phase of EFA, respectively. In doing so, JASP (Jeffrey's Amazing Statistics Program) version 0.96 Software was used.

3.1.1. Preliminary Tests of EFA Assumptions

Prior to proceeding to uncover the underlying factor structure and evaluate the psychometric quality of each item, it was found essential to first determine how the variables are organized and classified as factors. Accordingly, preliminary assumption tests for EFA was conducted on 30 items designed to measure three factors under each set 10 items, namely SR, TR, and CR factors. As reported in Table 3, the initial test results of the assumption confirmed, Bartlett's sphericity test was significant, $\chi^2 (435) = 4572.82$, $p < .001$. This shows that the observed

correlations are adequate for extracting latent factors. Furthermore, the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy was 0.93, which exceeds the recommended threshold of .60.

Table 3: Preliminary Tests to EFA Factorability

Tests	Value	df	P	Threshold for "Good Fit"
Bartlett's Test of Sphericity	$\chi^2 = 4572.82$	435.00	< .001	correlations adequate
Kaiser-Meyer-Olkin (KMO)	.93	—	—	≥ .60 acceptable

As an additional check, Table 4 reports that Cronbach's alpha for each 10-item factor ($n = 301$), demonstrates excellent internal consistency in all three dimensions. These high alpha coefficients are also supported by the EFA structure, in which each group of ten items loads strongly on a single factor.

Table 4: Reliability of the three factors ($n=301$)

Factor	Items	Cronbach's α	Interpretation
SR	Q1S–Q10S	.91	Excellent
TR	Q1T–Q10T	.92	Excellent
CR	Q1C–Q10C	.92	Excellent

3.1.2. Factor Structure and Variance

As depicted in Table 5 factor refinement was performed to identify the significant variables within each category that contribute to the model. Factor extraction was conducted with 30 items ($n=301$) using maximum likelihood (ML) with oblique rotation, considering the assumption that factors expected to correlate. A minimum factor loading threshold of 0.60 was applied, and all items demonstrated communalities above the acceptable threshold of .40. This confirms that each item shared sufficient variance with its respective factor. In addition, the Chi-Square test yielded a non-significant result, $\chi^2 (348) = 346.39$, $p > .05$. This implies that there is no significant difference between the covariance matrix of actual data and the covariance matrix of the proposal factor model. Accordingly, there was no empirical base to remove any of the 30 items at the EFA stage (see Table 5). Additionally, all communalities exceeded the recommended ($\geq .40$) threshold, which indicates each item contributed adequately to the factor solution. This supports the appearance of the three-factor structure was empirically justified.

Consistent with this, both Matsunaga (2010) and Tabachnick and Fidell (2019) argue that a factor loading above 0.50 thresholds was preferred for scale refinement, and 0.70 thresholds is strong indicator for the convergent validity. The threshold of 0.60 adopted here lies between these two ranges and accounts for approximately 36% of the variance in each item.

Table 5: Factor Loading in EFA

Factor Loadings				
Items	Factor 1	Factor 2	Factor 3	Uniqueness
Q7C	0.768			0.413
Q8C	0.763			0.428
Q9C	0.757			0.421
Q6C	0.736			0.442
Q2C	0.734			0.465
Q1C	0.725			0.467
Q5C	0.696			0.490
Q3C	0.695			0.510
Q4C	0.685			0.537

Q10C	0.661	0.563
Q4T	0.766	0.416
Q1T	0.759	0.424
Q7T	0.737	0.455
Q3T	0.734	0.457
Q2T	0.730	0.468
Q10T	0.707	0.453
Q5T	0.706	0.502
Q6T	0.691	0.525
Q8T	0.688	0.525
Q9T	0.688	0.532
Q4S	0.744	0.449
Q3S	0.739	0.456
Q2S	0.727	0.469
Q10S	0.725	0.472
Q7S	0.721	0.475
Q5S	0.703	0.506
Q6S	0.684	0.525
Q1S	0.682	0.522
Q8S	0.681	0.537
Q9S	0.657	0.556

Note. Applied rotation method is oblimin.

As shown in Table 6, the three factors extracted (SR, TR and CR) explained 51 % of the variance. The CR factor accounted for 16% of the variance, with all 10 'C' items (Q1C–Q10C) loading cleanly onto it. Factor loadings ranged from .661 (Q10C) to .768 (Q7C). The TR factor accounted for 18% of the variance, with all 10 'T' items (Q1T–Q10T) loading onto it. Factor loadings ranged from .688 (Q8T, Q9T) to .766 (Q4T). The SR factor accounted for 17% of the variance, with all 10 'S' items (Q1S–Q10S) loading onto it. Factor loadings ranged from .657 (Q9S) to .744 (Q4S). This indicates that the proportional variance explained by the three factors was relatively close one another. In addition to this, the three extracted factors, depicted in the scree plot in Figure 2, identified based on the eigenvalue criterion of being greater than 1 (with eigenvalues of CR 5.91, TR 5.25, and SR 4.38).

Table 6: Factor Structure and Variance

Factors	Factor Characteristics					
	Eigenvalue	Unrotated solution			Rotated solution	
		Proportion var.	Cumulative	Sum Sq. Loadings	Proportion var.	Cumulative
CR-Factor	5.91	0.20	0.20	5.26	0.16	0.16
TR-Factor	5.25	0.18	0.38	5.24	0.18	0.34
SR-Factor	4.38	0.15	0.52	5.04	0.17	0.51

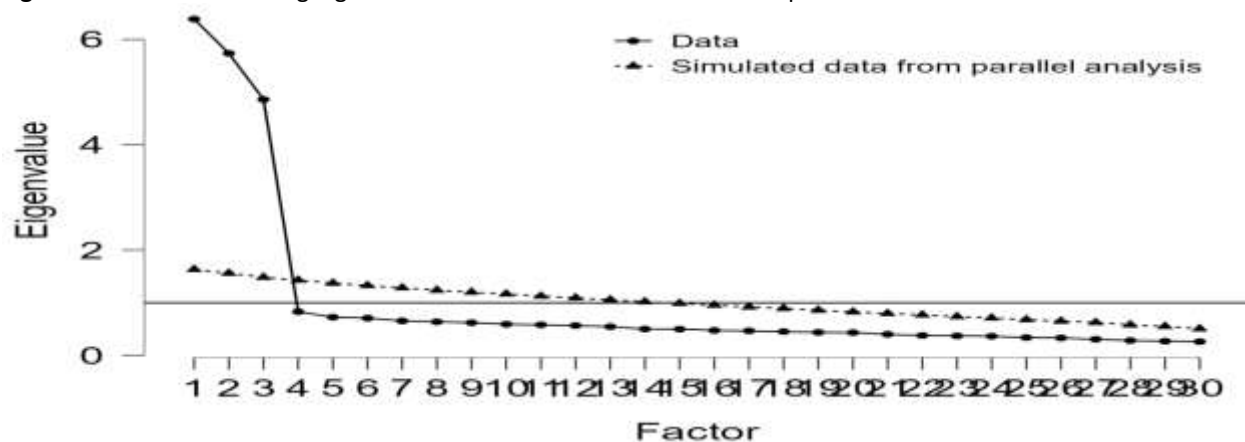
In addition, eigenvalues from the parallel analysis was performed to confirm the number of factors is still three, independent of the Kaiser >1 rule, as shown in Table 5. As shown in Table 7, the first three eigenvalues seen are significantly larger than the random-data eigenvalues. This lends strong support to a three factor solution, as the

fourth observed eigenvalue (0.83) is below the parallel analysis criterion (1.47). So, this method also supports retaining three factors. The eigenvalue differences between Table 6 and Table 7 arise due to PCA eigenvalues are systematically larger than those from ML extraction. PCA accounts for the total variance in each item, whereas ML/common-factor extraction captures only the shared variance.

Table 7: parallel Analysis Factor Structure

Factor	Observed eigenvalue	95th-percentile random eigenvalue	Decision
1	6.39	1.72	Retain
2	5.74	1.61	Retain
3	4.86	1.53	Retain
4	0.83	1.47	Reject
5	0.73	1.42	Reject
6	0.71	1.37	Reject
7	0.66	1.32	Reject

Figure 2: Scree Plot Showing Eigenvalues for Extracted Three Factor Components



Note: Eigenvalue > 1 indicates three factors

3.1.3. Item Retention Criteria

In this initial EFA phase, 30 Items were retained under three categories (SR, TR, CR; 10 items each). The reliability coefficient ($\alpha > .70$) confirmed internal consistency and all retained items demonstrated loadings ($\geq .60$) on their respective factors (see Table 2). This indicates that convergent validity and internal consistency across SR, TR, and CR dimensions were maintained. This yields a clear and interpretable factor solution that is the basis for the subsequent CFA to validate this predetermined structure.

3.2. Validation Procedures — CFA Phase

This section presents the CFA analysis results, which is conducted to validate the hypothesized measurement model derived from the first-phase EFA. The CFA was applied with a separate subsample ($n = 301$). This phase helps translate what was discovered in EFA into model confirmation in CFA.

3.2.1. CFA Assumptions

CFA analysis included robust estimation methods, and a maximum-likelihood model-fit index was used to assess each factor's viability. In line with this, Hou et al. (2014) highlighted that several fundamental assumptions are necessary before the confirmation of model fit in CFA. In the same vein, the authors further affirmed that these recommended requirements included the following criteria: a Chi-square value (χ^2) / df value less than 2.00 suggests the model does not deviate significantly from the observed data. The Root Mean Square Error of Approximation (RMSEA), with values $\leq .08$, reflects how closely the model approximates the population covariance matrix. Other incremental fit indices, such as the Comparative Fit Index (CFI) and Tucker–Lewis Index (TLI), are expected to exceed .90, with TLI additionally penalizing for model complexity. The Standardized Root Mean Square

Residual (SRMR), with values < .08, indicates the average discrepancy between observed and predicted correlations. In addition to these fit indices, Composite Reliability (CR ≥ .70) and Average Variance Extracted (AVE ≥ .50) were examined to ensure reliable measurement. A standardized loading was fixed to determine adequate variance explanation. These checks also help ensure the convergent and discriminant validity of the final model.

3.2.2. Model Fit Indices

The results of the CFA regarding the factors contributing to students' underperformance in ESSLCE were evaluated in accordance with the aforementioned (Section 3.2.1) several goodness-of-fit indices. As for this, the results of the Chi-square test presented in Table 8 show a Chi-square value of $\chi^2/df \approx 1.05$, and well under the 2.00 threshold.

Table 8: Chi-square Test

Model	χ^2	df	p
Baseline model	4759.950	435	
Factor model	423.482	402	.221
Note. The estimator is ML. The test statistic is standard. The standard error method is standard. Model $\chi^2=423.482$, $df=402 \rightarrow$ ratio = 1.05.			

Additionally, as reported in Table 9, the fit indices demonstrate a strong fit alignment between the model and the data. The RMSEA was 0.013 (90% CI = 0.000–0.025), both the CFI and, the TLI were 0.995, and the SRMR was 0.041.

Table 9: Additional fit Indices

RMSEA	RMSEA 90% confidence	SRMR	TLI	CFI
0.013	0.000, 0.025	0.041	0.995	0.995

3.2.3. Reliability and Validity of the Measurement Model

In this section, after completing the CFA refinement of the students' underperformance indicators scale in ESSLCE, it was essential to assess the reliability (consistency) and validity (accuracy) of the scale. To accomplish this, several metrics were utilized, including Cronbach's alpha and Composite Reliability (CR), as well as measures for Convergent Validity and Discriminant Validity.

3.2.4. Reliability Coefficient and Composite Reliability (CR)

As indicated in Table 10 the internal consistency was assessed using Cronbach's alpha and McDonald's omega. Both coefficients exceeded the recommended threshold of .70 across all three factors.

Table 10: Reliability Coefficient and Composite Reliability

Reliability		
Factors	Coefficient ω	Coefficient α
SR	0.909	0.909
TR	0.915	0.915
CR	0.916	0.916

3.2.5. Convergent Validity (CV)

The Average Variance Extracted (AVE) values were used to provide evidence of the convergent validity (CV). The result of the SR (0.500), TR (0.520), and CR (0.523) in the Table 11 shows that all the AVE values closely met or met the conventional .50 criterion.

Table 11: The Average Variance Extracted (AVE) values

Average variance extracted	
Factor	AVE
SR	0.500
TR	0.520

3.2.6. Discriminant Validity (MSV/ASV/HTMT)

Discriminant validity was determined by comparing the Average Variance Extracted (AVE) against the Maximum Shared Variance (MSV) and Average Shared Variance (ASV), commonly referred to as the Fornell–Larcker criterion (AVE, MSV, and ASV). The result in Table 12 depicted that the AVE for all three factors (SR = 0.500, TR = 0.520, CR = 0.523) were higher than the respective the MSV and ASV values. This result demonstrates that each construct retains more variance in its own indicators than it shares with other constructs. This supports the presence of discriminant validity for all three latent constructs (SR, TR, CR)..

Table 12: Discriminant Validity

Factor	AVE	MSV (Highest r^2)	ASV (Average r^2)	Result
SR	0.500	0.004	0.003	Valid
TR	0.520	0.011	0.006	Valid
CR	0.523	0.011	0.008	Valid

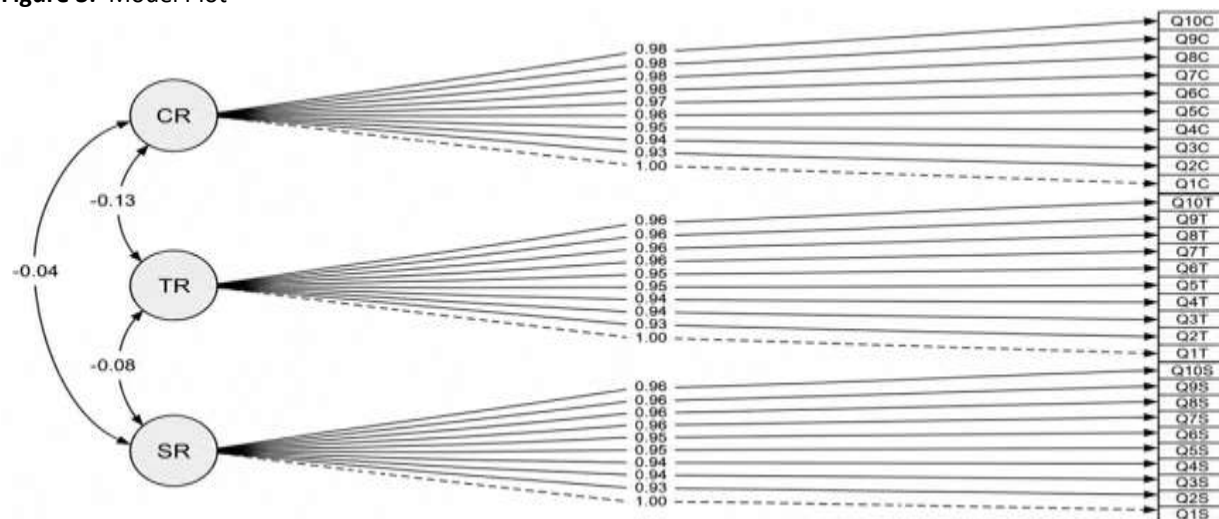
Additionally, the inter-construct correlation matrix among the latent constructs, particularly the Heterotrait-monotrait (HTMT) criterion, was used to assess the discriminant validity. The values were found below the recommended threshold of 0.85 and indicated sufficient discriminant validity among the constructs (see Table 13). Based on these results, it is statistically valid to conclude that the students' underperformance indicators scale demonstrates good internal consistency and acceptable accuracy.

Table 13: Heterotrait-monotrait ratio of predictors

Heterotrait-monotrait ratio		
SR	TR	CR
1.000		
0.060	1.000	
0.064	0.087	1.000

Figure 3 supports that the discriminant validity and the subscales are measuring distinct constructs.

Figure 3: Model Plot



4. Discussion

The study sought to develop and validate a multidimensional instrument designed to identify the Students' Underperformance Indicator Instrument (SU II) within the Secondary School Leaving Certificate Examination (ESSLCE) context in Easter part of Ethiopia. The structures of the three factors in each of the 10 items, student-related (SR), teacher-related (TR), and contextual (CR) were discovered and confirmed empirically in the EFA and CFA phases, respectively. The three-factor structure (SR, TR, and CR) confirmed through EFA and CFA provides strong evidence of reliability, and validity, indicating the robustness of SU-II.

Building on this, the alignment with the CIPO and IEO frameworks clarifies how each domain functions within the educational system. The CIPO framework describes the structural causes of underperformance, and the I-E-O model describes the process by which underperformance is translated into learning experiences and outcomes. Related to this study, Taye et al. (2025) validated a rigorous instrument limited to academic dishonesty, whereas Honra (2022) captured academic resilience through EFA alone without a confirmatory stage and gave attention to only teacher- and system-level factors. This student underperformance indicator instrument (SU-II) in ESSLCE combines the student, teacher, and a contextual aspect, which provides a comprehensive diagnostic tool as compared with previous instruments that had a narrow focus on dishonesty or resilience.

Beyond its psychometric properties, the instrument's stability across the four study sites, with scores higher in Harar and Dire Dawa than the national average but Chiro and Jijiga being comparatively low, provides a holistic diagnostic tool capable of situating underperformance in settings that have different profiles. The low inter-factor correlations further imply that interventions must be multidimensional, reinforcing the systemic framing that has characterized public and parliamentary discourse around the ESSLCE decline. In general, these findings support the study's central premise, which is that underperformance in ESSLCE is better understood as a multidimensional, pre-outcome phenomenon than as a sole outcome of student effort. Thus, the SU II provides both empirical rigor and conceptual coherence, enabling policymakers and educators to identify leverage points for intervention.

5. Conclusions and Implications

This study developed and validated the Students' Underperformance Indicator Instrument (SU II) within the Ethiopian Secondary School Leaving Certificate Examination (ESSLCE) context. Through rigorous procedures, in the first development phase, item generation, expert review, cognitive interviews, and pilot testing were carried out consecutively. In the second, statistical validation through EFA and CFA was carried out. Accordingly, the results of the construct validation demonstrated strong psychometric properties with high reliability, convergent validity, and discriminant validity across three distinct domains: (SR), (TR), and (CR). The findings confirm that this validated instrument, which comprises 30 items divided into three dimensions, SR, TR, and CR on a 5-point Likert scale, provides a promising psychometrically supported tool for assessing indicators associated with students' underperformance in ESSLCE.

Additionally, the SU II carries significant practical relevance that the validated instrument is essential for educators, policymakers, and future researchers for diagnostic use, policy monitoring, and intervention design. The instrument encourages evidence-based decision-making aimed at improving students' ESSLCE outcomes by situating underperformance within systems based frameworks (CIPO and I E O). It is clear that this helps to foster ongoing educational quality. Future studies should extend this validation by examining the instrument's predictive capacity in longitudinal designs, particularly focusing on whether SR, TR, and CR domains forecast later academic outcomes. This evidence would allow policymakers to use SU-II not only diagnostically but also proactively to identify the risk factors before exam failure occurs.

6. Limitations

In this study, the rigorous procedures were followed in the development and validation of the SU-II instrument. It is not possible to say that the study was not without limitations. First, the sample consisted of students who did not achieve the 50% cutoff in the 2022/23 exam year and those who were permitted to attend TVET and remedial programs. This may limit the applicability of the results to students who scored below 50% and are ineligible for either the TVET or remedial program cohort. Second, the results in the Eastern Ethiopia geographical and cultural context may restrict the generalizability of the findings to other settings. Third, it is appropriate to acknowledge

that the retrospective self-report approach is a source of potential recall and attribution bias, particularly in sensitive areas such as teacher misconduct or student indiscipline. Fourth, while the instrument demonstrated strong psychometric properties, its predictive validity and applicability across diverse educational settings remain to be tested in future studies.

7. Funding

The authors received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

8. Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

9. Data Availability Statement

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

10. AI Use Disclosure

Artificial intelligence tools were only used for language editing and manuscript refinement. All interpretations, analyses, and conclusions remain the responsibility of the authors.

References

1. Alford, Z. (2024, November 16). Current Issues Ethiopia's Education Problem: 96% of students fail final school exam. - Broken Chalk. <https://brokenchalk.org/current-issues-ethiopias-education-problem-96-of-students-fail-final-school-exam/>
2. Astin, A. W. (1993). What matters in college? Four critical years revisited. San Francisco, CA: Jossey-Bass.
3. Borkena. (2023, October 9). Ethiopian high school leaving exam result highlights ongoing education disaster. <https://borkena.com/>
4. Borkena. (2024, August 1). Ethiopia: Failed or made to fail? The shocking results of ESLCE. <https://borkena.com/ethiopia-failed-or-made-to-fail-the-shocking-results-of-eslce/>
5. Chala, W. D. (2021). Perceived seriousness of academic cheating behaviors among undergraduate students: an Ethiopian experience. *International Journal for Educational Integrity*, 17(1).
6. Demissie, B. (2024, August 22). New Ethiopia's exam system: Successes and challenges. Org.Et. <https://www.psi.org.et/index.php/blog/331-new-ethiopia-s-exam-system-successes-and-challenges>
7. Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2018). *Multivariate data analysis* (8th ed.). Cengage Learning.
8. Honra, J. R. (2022). Development and Validation of Filipino Learners' Academic Resilience Scale (FLARS). *American Journal of Education and Technology*, 1(2), 107–113. <https://doi.org/10.54536/ajet.v1i2.664>
9. Hou, J., Wang, H., & Zhou, T. (2014). Evaluating model fit in confirmatory factor analysis: Guidelines and practices. *Psychological Test and Assessment Modeling*, 56(2), 219–239.
10. Kyriazos, T. A. (2018). Applied Psychometrics: Sample Size and Sample Power Considerations in Factor Analysis (EFA, CFA) and SEM in General. *Psychology*, 9(8), 2207–2230. <https://doi.org/10.4236/psych.2018.98126>
11. Matsunaga, M. (2010). How to factor-analyze your data right: do's, don'ts, and how-to's. *International Journal of Psychological Research*, 3(1), 97–110. <https://doi.org/10.21500/20112084.854>
12. Ministry of Education. (2022). Education sector annual abstract (ESAA 2022–23). Federal Democratic Republic of Ethiopia
13. Ministry of Education. (2023). Ethiopia Education Transformation Program: Teacher licensing and relicensing proposal. Ethiopian Ministry of Education Retrieved from <https://www.dpgethiopia.org/wp-content/uploads/2024/03/Ethiopia-Education-Transformation-Program-proposal-.pdf>

14. Ministry of Education. (2025, July 16). Press statement on Grade 12 national examination results. Educational Assessment and Examinations Service. <https://eaes.gov.et>
15. Ministry of Education. (2025, July 16). Press statement on Grade 12 national examination results. Educational Assessment and Examinations Service. <https://eaes.gov.et>
16. Molla, T., & Tiruneh, D. (2023). Ethiopia's education system is in crisis – now's the time to fix it. In N. Joseph (Ed.), *The Conversation*. <https://doi.org/10.64628/aaj.avgqnsxue>
17. Monitor, E. (2023, October 9). Grade 12 Exam Result: Only 27,000 Students Hit 50% Mark. <https://ethiopianmonitor.com/2023/10/09/eslce-report-card-only-27000-grade-12-students-hit-50-mark/>
18. Monitor, E. (2024, September 9). Grade 12 exams: Over 36, 400 students score 50% and above. [Ethiopianmonitor.com. https://ethiopianmonitor.com/2024/09/09/grade-12-exams-over-36-400-students-score-50-and-above/](https://ethiopianmonitor.com/2024/09/09/grade-12-exams-over-36-400-students-score-50-and-above/)
19. Ngah, H., Mohd Hairon, S., Hamzah, N. A., Noordin, S., & Shafei, M. N. (2022). Development and Validation of Knowledge, Attitude, and Practice Questionnaire: Toward Safe Working in Confined Spaces. *International Journal of Environmental Research and Public Health*, 19(3), 1242. <https://doi.org/10.3390/ijerph19031242>
20. Polit, D. F., & Beck, C. T. (2006). The content validity index: Are you sure you know what's being reported? critique and recommendations. *Research in Nursing & Health*, 29(5), 489–497. <https://doi.org/10.1002/nur.20147>
21. Pusbelina, A. R., & Sutrisno, H. (2025). Development Instrument Independence Study in Chemistry Learning Using Exploratory Factor Analysis (EFA) and Factor Analysis Confirmatory (CFA). *Jurnal Penelitian Pendidikan IPA*, 11(5), 869–876. <https://doi.org/10.29303/jppipa.v11i5.9790>
22. Reyno, R., & Guzman, R. (2025). Factors affecting the performance of students in the National Achievement Test. *Journal of Interdisciplinary Perspectives*, 3(6). <https://doi.org/10.69569/jip.2025.255>
23. Scheerens, J. (1990). School effectiveness research and the development of process indicators of school functioning. *School Effectiveness and School Improvement : An International Journal of Research, Policy and Practice*, 1(1), 61–80. <https://doi.org/10.1080/0924345900010106>
24. Tabachnick, B. G., & Fidell, L. S. (2019). *Using multivariate statistics* (7th ed.). Boston, MA: Pearson.
25. Tamrat, W. (2023a, February 1). HE to feel the impact of shocking school-leaving exam results. *University World News*. <https://www.universityworldnews.com/post.php?story=20230201195745831>
26. Tamrat, W. (2023b, October 19). Mass failure in national exam calls for urgent intervention. *University World News*. <https://www.universityworldnews.com/post.php?story=20231018103529310>
27. Taye, D., Semela, T., & Assefa, S. (2025). Confirmatory Factor Analysis and Propensity to Cheat Scale Validation in the Ethiopian Public Higher Education Institutions. *F1000Research*, 13, 955. <https://doi.org/10.12688/f1000research.150357.3>
28. Tefera T. & Zelalem T. (2025). Ethiopia's secondary school leaving examinations system: A glimpse inside its governance, management, and potential consequences. *Bahir Dar Journal of Education*. 25 (1). <https://doi.org/10.4314/bdje.v25i1.7>
29. Tefera, T., Belay, H., and Shewatatek, G. (2024). Analysing Ethiopia's Two-Year General Secondary School Leaving Exams: Using Cheating and Assessment Lenses. *The Ethiopian Journal of Education*, 44(1), 223–234. <https://doi.org/10.63990/eje.v44i1.10033>
30. Tiruneh, D. T., & Molla, T. (2024, May 1). Spotlight on Ethiopia's secondary education challenges. *The Education and Development Forum (UKFIET)*. <https://www.ukfiet.org/2024/spotlight-on-ethiopias-secondary-education-challenges/>
31. Umeh, B. O., & Hamzah, A. (2025). Determinants of academic achievement in Nigerian secondary schools: Leadership, environment, and motivation. *International Journal of Academic Research in Business and Social Sciences*, 15(5). <https://doi.org/10.6007/ijarbss/v15-i5/25333>
32. Willis, G. B. (2004). *Cognitive Interviewing*. SAGE Publications.