Psychometric Properties of the Academic Agility Scale: Validity and Reliability Evidence

EL waleed. A. Meiri[1], Soliman, A. Alghadyan[2], Waed F. Mistarihi[3]

[1,2,3]Department of Psychology, College of Social Science, Imam Mohammad Ibn Saud Islamic University (IMSIU).

Jordanian Ministry of Education, Jordan

Email: eameiere@imamu.edu.sa

Email: saghadyan@imamu.edu.sa

Email: wmistarhi@yahoo.com

HOW TO CITE:

EL waleed. A. Meiri, A. Alghadyan Soliman, Waed F. Mistarihi (2026). Psychometric Properties of the Academic Agility Scale: Validity and Reliability Evidence. International Journal of Special Education, 41(4s), 130-150.

ABSTRACT:

Background; Academic agility has emerged as a critical construct in higher education, reflecting the capacity of institutions and learners to adapt effectively to rapid changes, particularly in digitally mediated and post-pandemic environments. The existing measurement tools require development because of their inability to assess academic agility and their need for context-based validation.

Objectives: The study assessed both the psychometric properties of the Academic Agility Scale and its validity and reliability through testing with university students.

Methods: The study used a quantitative cross-sectional design which included 380 university students as research participants. The forward–backward translation method was used to translate the instrument. SPSS (Version 26) software was used to conduct data analysis. The researchers assessed construct validity through exploratory factor analysis (EFA) and confirmatory factor analysis (CFA) testing while using convergent and discriminant validity indices (AVE, CR, HTMT) for their analysis. The researchers used regression analyses and known-groups comparisons to test criterion-related validity. The researchers tested reliability through three methods which included Cronbach's alpha, split-half reliability, and McDonald's omega.

Results: The EFA showed that a five-factor model explained 64.5% of total variance, which demonstrated excellent sampling adequacy (KMO = 0.929). CFA results showed that the factors had strong loadings ($\lambda > .70$) yet the model fit indices did not reach the required minimum values (CFI = 0.819, TLI = 0.800). The research found that both types of validity were mostly confirmed. Criterion-related validity showed statistically significant but small predictive effects ($R^2 \approx 0.01$). The scale demonstrated excellent internal consistency

COPYRIGHT STATEMENT:
Copyright: © 2026 Authors.
Open access publication underthe terms and conditions of the Creative CommonsAttribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).

($\alpha = 0.957$), which multiple reliability tests confirmed. **Conclusion**: The Academic Agility Scale is a reliable and theoretically grounded instrument with acceptable validity for assessing adaptability in higher education contexts. The measurement model needs additional refinement while researchers must conduct validation studies with various population groups to establish its environmental applicability.

Keywords: Academic agility; Psychometric validation; Factor analysis; Construct validity; Reliability; Higher education.

1. Interdiction

The current transformation of higher education systems results from three forces which include technological progress, global outreach and changing student needs. The COVID-19 outbreak caused more than 70% of universities worldwide to speed up their digital transformation efforts which created new ways to conduct educational activities and institutional functions (Hodges et al., 2020; Hu & Bentler, 1999) The traditional educational system fails to meet modern student needs because 65% of learners demand more flexible educational experiences (Hodges et al., 2020; Khatatbeh & Mistarihi, 2025). The framework of agility serves essential purposes in organizations because it helps them maintain their operational effectiveness during periods of organizational change. Agility originated from organizational and strategic management research which defines it as the capacity to identify environmental alterations while executing immediate responses through efficient resource redistribution (Doz & Kosonen, 2010; Teece, 2007)The performance results of agile organizations surpass those of less adaptive organizations by 30 to 40 percent in environments which display high levels of instability (Kyriazos, 2018; Lin, 2026; Onwuegbuzie, 2013). The educational field has shown strong interest in applying agility principles because universities now face the same uncertainty challenges which existed in educational institutions. The field of academic agility requires operationalization for academic agility to demonstrate its complete value but insufficient validated tools exist which can assess its various dimensions. Existing research uses two types of measurement instruments: organizational assessments which scientists have modified or specific indicators which scientists developed for particular research contexts but these methods lack sufficient psychometric evaluation (Doz & Kosonen, 2010; Karimian & Chahartangi, 2024). Educational systems require measurement tools which scientists must develop and validate to measure their particular educational system dynamics. The research Karimian and Chahartangi (2024) developed a structured framework for educational agility which defines educational agility and establishes its measurement system through a structured system of multiple measurement dimensions.

The scientific assessment of newly created tools requires validation and reliability testing to prove their effectiveness for research and real-world applications. Modern psychometric frameworks require multiple validity testing methods which need assessment of factor structure and both convergent and discriminant validity and criterion-related validity (Hazell et al., 2020; Henseler et al., 2015). Researchers use exploratory and confirmatory factor analyses (EFA and CFA) to examine psychological constructs because these methods enable them to measure two aspects of their research work, which include dimensionality assessment and model fit evaluation(Kline, 2016; Nahar & Sekhar, 2024; Naz et al., 2021; Nelson, 2000) The average variance extracted (AVE) and composite reliability (CR) and heterotrait–monotrait ratio (HTMT) functions as strong assessment tools which researchers use to determine the validity of complex multidimensional models (Henseler et al., 2015)

Research studies have shown that measurement instruments require psychometric methods to establish their development process, which applies

to complex constructs that include agility and adaptability. Boateng et al. (2018) developed the comprehensive framework that exists today by creating a methodical process for developing measurement scales through their multiple empirical research studies. The research demonstrates that sample size establishes the foundation for maintaining consistent factor patterns in measurement tests.The Anthoine et al. (2014) found that 114 studies on psychometric validation showed that instrument development studies depend on content validity assessment together with construct validity evaluation and criterion-related validity testing. The study proved that most validated instruments used factor analysis methods (EFA and CFA) together with multiple validation techniques which demonstrates that current scale development needs a combination of different validation methods. Gunawan et al. (2021) showed that factor analysis functions as the main technique for scale validation and they established 300 participants as the minimum sample requirement to deliver reliable factor results together with dependable psychometric outcomes. The study showed that using two different sample groups for EFA and CFA testing strengthens the accuracy of construct validation while minimizing the chances of overfitting. Kyriazos (2018) found the item-to-response ratio in psychometric research should use 1:5 to 1:10 according to Kyriazos and the psychometric research that followed him. The exploratory and confirmatory factor analyses require this ratio for maintaining statistical power when researchers study multidimensional constructs in their research. Rostami (2018) team executed a psychometric validation study with 418 test subjects whom they divided into two groups for EFA testing and CFA testing. The factor loadings ranged from 0.41 to 0.94 and the total explained variance exceeded 55% while the internal consistency achieved acceptable levels through Cronbach's alpha testing. The review research conducted by Morgado et al. (2017) showed that 88.6% of psychometric studies employed EFA while 72.3% of studies used CFA and about 65% of studies combined EFA with CFA. The study found that internal consistency reliability testing using Cronbach's alpha functions as the primary measurement tool for validation studies.

The framework developed by Trochim for scale development requires researchers to create 80 to 100 initial items and let experts evaluate these items before they conduct the refinement process. The statistical aspect of scale development requires researchers to understand both conceptual and theoretical dimensions(Trochim, 2006)

The assessment of academic agility requires validation through different dimensions and various statistical methods because it exists as a complex and evolving measurement system. Academic research shows that constructs which measure adaptability and flexibility maintain moderate connections because they share relationships with other constructs [14]. Researchers must validate construct frameworks while evaluating how much different components of the framework share common elements with each other.

The current research study intends to conduct a complete psychometric assessment of the Academic Agility Scale among university students. The study aims to assess construct validity through exploratory and confirmatory factor analyses together with convergent and discriminant validity testing. The researchers assess criterion-related validity to study how academic agility connects with specific external relationships. The study establishes a comprehensive validation method which strengthens existing research about educational adaptive abilities while delivering an effective measurement solution for both research and practical applications.

2. METHODS

2.1. STUDY DESIGN

The study used a quantitative cross-sectional design to investigate how hypothetical thinking patterns and psychological resilience and competitive anxiety affect adolescents. structural equation modeling (SEM) to examine all direct and indirect relationships between variables. The study used hypothetical thinking as the predictor variable and psychological resilience as the mediator variable and competitive anxiety as the outcome variable . descriptive statistics and multivariate analyses

(MANOVA) to compare how different groups performed on demographic factors. The design enables researchers to evaluate structural relationships but its cross-sectional design prevents researchers from making any causal connections.

2.2. Participants

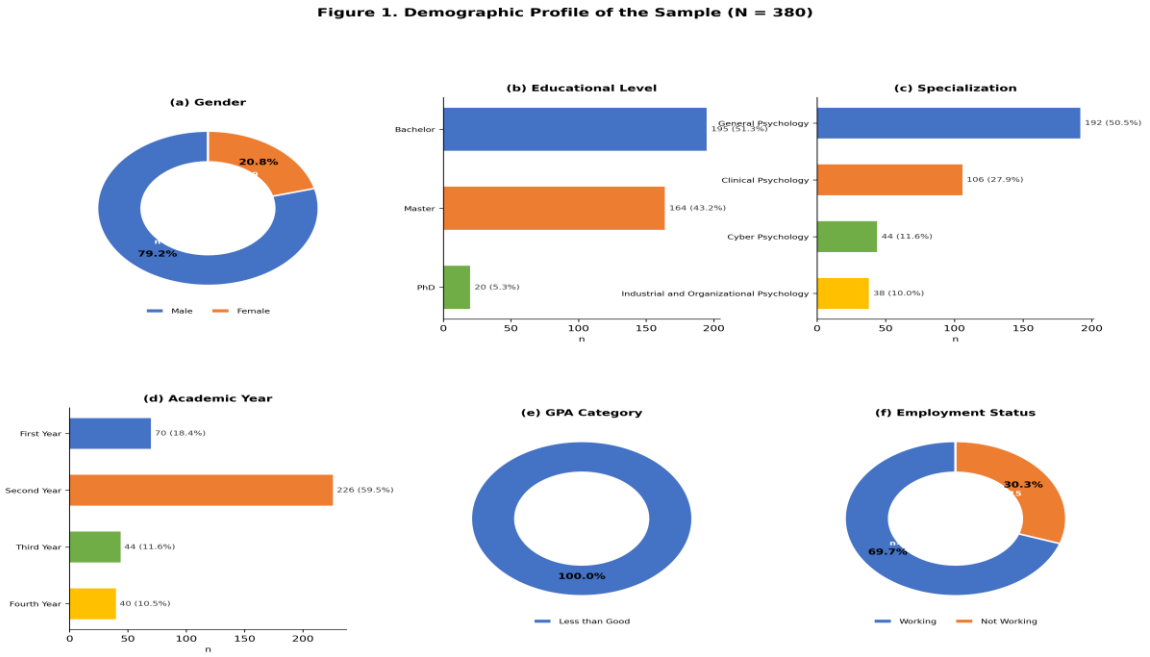
The study sample comprised 380 university students. The demographic information of the participants is displayed in Table 1. The sample contained 78.9% male students who were 59.5% second-year students and 69.7% of students who worked at the time. Participants studied General Psychology and three additional psychology fields with General Psychology being the largest group at 50.5%.

The demographic data for the sample group is visually displayed in Figure 1, which serves as a complement to the numerical information that Table 1 presents. The distribution shows that male participants make up approximately 79% of the total, which results in a gender imbalance that matches the frequencies shown in the table. Most respondents were enrolled at the bachelor's level, followed by master's students, with minimal representation at the doctoral level. The largest specialized group in General Psychology appeared first, while Clinical Psychology followed, and Cyber and Industrial and Organizational Psychology groups came later. The academic year distribution shows a clear concentration in the second year, suggesting uneven representation across study levels. The participants all belonged to the "less than good" GPA category, which requires careful interpretation because it establishes a predefined pattern. The majority of participants reported having jobs, which might impact the research study because it connects to the overall educational background of the participants. The figure makes data analysis easier by showing important distribution patterns that match the patterns found in Table 1.

Table 1 Demographic Characteristics of the Sample (N = 380)

Variable	Category	n	%
Gender	Male	300	78.9
	Female	79	20.8
Educational Level	Bachelor	195	51.3
	Master	164	43.2
	PhD	20	5.3
Specialization	General Psychology	192	50.5
	Clinical Psychology	106	27.9
	Cyber Psychology	44	11.6
	Industrial and Organizational Psychology	38	10.0
Academic Year	First Year	70	18.4
	Second Year	226	59.5
	Third Year	44	11.6
	Fourth Year	40	10.5
GPA Category	Less than Good	380	100.0
Employment Status	Working	265	69.7
	Not Working	115	30.3

Note. N = 380. Percentages may not sum to 100 due to rounding.



Note. Panel (a) Gender. Panel (b) Educational level. Panel (c) Specialization. Panel (d) Academic year. Panel (e) GPA. Panel (f) Employment status. Donut charts for binary variables; horizontal bars for multi-category variables. Exact frequencies reported in Table 1.

Figure 1 Demographic Profile of the Sample (N = 380)

2.3. Ethical Considerations

The research was performed following the required ethical guidelines which already existed at that time. The research required all participants to provide informed consent before they could start taking part in the study. The research team provided participants with full protection of their identity because no personal information was needed for the study. Participants received complete information about their ability to leave the study whenever they wanted without facing any consequences. The institutional ethics committee conducted a study protocol review which resulted in their approval of the research plan.

2.4. MEASUREMENT

The Educational Agility Questionnaire (QEdu-Agility) which uses psychometric methods evaluates how universities and other educational institutions handle rapid transformations and evolving requirements. According to (Karimian & Chahartangi, 2024) the framework serves as a specific adaptation of organizational agility frameworks which organizations use during digital transformation and post-pandemic educational contexts.

The scale creation process followed a systematic approach which combined existing theoretical frameworks with practical application. The research team modified items from three established organizational agility assessments(Goldman et al., 1994; Sharifi & Zhang, 1999; Worley & Lawler, 2010) to create assessment tools for educational settings. The expert review and focus group discussions established which topics should remain in the original pool based on their relevance to the research and the current educational context. The final instrument contained 30 items which measured five core elements of educational assessment that included responsiveness, sufficiency, flexibility, speed, and integrity.

Educational agility receives assessment through these five characteristics which establish a multifaceted educational flexibility perspective. Adequacy measures how well educational services satisfy students' demands, whereas responsiveness measures the institution's capacity to adapt to environmental changes. Flexibility allows education delivery to occur without structural restrictions

which include time and space limitations while speed focuses on delivering results with maximum efficiency. Integrity (or coherence) ensures that institutional components align and coordinate to form an integrated educational system.

The five-point Likert scale measures educational agility with higher scores indicating greater agility. The mid-point score of 3 shows that the organization has basic agility capabilities. The scaling method enables comprehensive dimension assessment yet maintains dimension assessment unambiguous for both research and practical applications, The tool shows valid and reliable psychometric characteristics. The Content Validity Ratio (CVR = 0.847) and Content Validity Index (CVI = 0.877) showed that all items fulfilled requirements of relevance and clarity while accurately representing the concept. The Exploratory Factor Analysis (EFA) validated the five-factor framework which accounted for 59 to 60 percent of the total variance with responsiveness being the most important variable.

The scale demonstrates robust internal consistency with a Cronbach's alpha value of 0.944 and all subscales exceeding minimum acceptable thresholds. The instrument provides consistent measurement across all its assessment dimensions. The QEdu-Agility scale serves as a thorough multidimensional assessment tool which evaluates educational agility in educational institutions. The theoretical foundation of the assessment tool together with its empirical validation and ability to assess contextual sensitivity enables educational institutions to evaluate their capacity to adapt to technological developments and emerging learning environments.

2.5. TRANSLATION OF STUDY MEASURES

The study measures were translated using a forward–backward translation procedure to ensure linguistic and conceptual equivalence. The original instrument was translated into Arabic by two bilingual experts who completed their work through creation of a unified translated version. An independent translator back-translated the document which was then compared to the original text to check its accuracy. The pre-final version underwent expert assessment, which included testing on a small group, resulting in minor adjustments before the final testing period.

2.6. STATISTICAL ANALYSIS

The researchers conducted their data analysis by utilizing IBM SPSS Statistics (Version 26). used descriptive statistics which included means and standard deviations and frequencies and percentages to create a summary of sample characteristics and item distributions. The researchers used exploratory factor analysis (EFA) with principal axis factoring and varimax rotation to investigate the psychometric properties of the instrument assessed sampling adequacy by applying the Kaiser–Meyer–Olkin (KMO) measure and Bartlett's test of sphericity ·evaluated internal consistency reliability by using Cronbach's alpha coefficients. The researchers used item–total correlations to assess item discrimination by examining the results. The researchers established a significance level of $p < .05$ for all statistical tests which they conducted.

3. RESULTS

Q1: Does the Academic Agility Scale demonstrate adequate construct validity?

To establish the construct validity of the Academic Agility Scale three different validation methods were used as validation methods. began with item-level distribution analysis before using exploratory factor analysis to determine the underlying factor structure. performed confirmatory factor analysis to test the extracted model and assess its performance when matched against the actual data. used established indices to evaluate scale validity through convergent and discriminant validity assessment which included factor loadings and average variance extracted and composite reliability and heterotrait–monotrait ratio measurement.

1.1 Preliminary Analysis: Item Descriptives

The researchers assessed the distributional properties of all 30 items before conducting their factor analysis work according to Table 2. The 5-point scale showed means from 3.14 to 3.62 while standard deviations ranged from 1.05 to 1.23, which demonstrated sufficient variability. The skewness

values ranged between −0.532 and −0.053 while the kurtosis values ranged between −0.845 and −0.127. The values meet the normality standards set by Kline (2016), which state that |skewness| must remain below 2 and |kurtosis| must stay below 7. The data received approval to proceed with both exploratory and confirmatory factor analysis activities.

Table 2 Item Descriptive Statistics for the Academic Agility Scale (N = 380*)*

Item	M	SD	Skew	Kurt
AG1	3.14	1.21	-0.139	-0.639
AG2	3.23	1.16	-0.104	-0.584
AG3	3.24	1.17	-0.094	-0.690
AG4	3.35	1.11	-0.195	-0.497
AG5	3.28	1.09	-0.053	-0.483
AG6	3.62	1.05	-0.417	-0.332
AG7	3.61	1.05	-0.532	-0.134
AG8	3.48	1.04	-0.115	-0.608
AG9	3.39	1.09	-0.138	-0.539
AG10	3.41	1.15	-0.351	-0.409
AG11	3.26	1.18	-0.233	-0.676
AG12	3.28	1.17	-0.229	-0.613
AG13	3.36	1.16	-0.237	-0.478
AG14	3.32	1.19	-0.269	-0.487
AG15	3.38	1.19	-0.265	-0.682
AG16	3.42	1.19	-0.360	-0.572
AG17	3.31	1.15	-0.168	-0.554
AG18	3.41	1.10	-0.324	-0.204
AG19	3.36	1.14	-0.239	-0.442
AG20	3.42	1.09	-0.198	-0.390
AG21	3.55	1.08	-0.431	-0.127
AG22	3.48	1.10	-0.269	-0.461
AG23	3.48	1.07	-0.282	-0.384
AG24	3.46	1.09	-0.240	-0.396
AG25	3.31	1.11	-0.216	-0.408
AG26	3.37	1.15	-0.294	-0.513
AG27	3.33	1.13	-0.159	-0.554
AG28	3.46	1.09	-0.248	-0.425
AG29	3.47	1.09	-0.255	-0.383
AG30	3.34	1.23	-0.145	-0.845

Note. 5-point Likert scale (1 = Strongly Disagree to 5 = Strongly Agree). Nineteen items were reverse-coded prior to analysis. All values fall within acceptable normality limits (Kline, 2016).

1.2. Exploratory Factor Analysis (EFA)

Exploratory factor analysis used principal component extraction methods with varimax rotation techniques for its analysis. The sampling adequacy KMO value of 0.929 met excellent standards according to Kaiser 1974 while Bartlett's test of sphericity showed significant results with a χ^2 value of 7618.7 for 435 degrees of freedom at $p < .001$ which established that the correlation matrix could be used for factor analysis. The analysis produced five factors which were confirmed through multiple criteria that showed eigenvalues between 1.0 and 1.0 and the scree plot inflection point and the five-dimensional theoretical structure and the rotated solution could be understood. The five factors together produced 64.5 percent of the entire variance

which table 3 displays. The communalities exceeded .50 for all items which had a range from .538 to .776 showing that all items were able to share sufficient variance with the extracted factors. The adequacy of the extracted factor structure is further supported by the scree plot, which clearly demonstrates the point of inflection, as shown in Figure 2 (Scree Plot: Eigenvalues for the 30-Item Correlation Matrix).

Table 3 EFA Factor Loadings (Varimax Rotation, 5 Factors)

Item	F1	F2	F3	F4	F5	h^2
AG1	0.293	-0.764	-0.154	0.023	-0.107	0.704
AG2	0.320	-0.736	-0.150	0.184	-0.054	0.703
AG3	0.128	-0.829	-0.164	0.209	-0.030	0.776
AG4	0.255	-0.646	-0.078	0.184	-0.377	0.665
AG5	0.119	-0.644	-0.227	0.210	-0.402	0.686
AG6	0.105	-0.120	-0.762	0.120	-0.156	0.645
AG7	0.293	-0.195	-0.719	0.039	-0.121	0.657
AG8	0.086	-0.238	-0.681	0.240	0.039	0.586
AG9	0.199	-0.239	-0.653	0.204	-0.091	0.573
AG10	0.046	-0.479	-0.396	0.467	-0.150	0.629
AG11	0.280	-0.402	-0.264	0.303	-0.434	0.589
AG12	0.473	-0.261	-0.222	0.322	-0.400	0.604
AG13	0.176	-0.333	-0.055	0.423	-0.509	0.583
AG14	0.071	-0.344	-0.139	0.335	-0.626	0.647
AG15	0.133	-0.204	-0.187	0.789	-0.211	0.762
AG16	0.267	-0.116	-0.067	0.754	-0.153	0.682
AG17	0.278	-0.192	-0.149	0.728	-0.241	0.725
AG18	0.405	-0.193	-0.366	0.265	-0.405	0.570
AG19	0.348	-0.054	-0.361	0.597	-0.174	0.641
AG20	0.465	0.018	-0.372	0.401	-0.374	0.656
AG21	0.513	-0.058	-0.215	0.287	-0.377	0.538
AG22	0.435	-0.325	-0.028	0.584	-0.023	0.637
AG23	0.622	-0.277	-0.199	0.209	-0.230	0.600
AG24	0.533	-0.267	-0.220	0.378	-0.112	0.559
AG25	0.765	-0.160	-0.237	0.049	-0.203	0.711
AG26	0.604	-0.323	0.018	0.400	-0.172	0.658
AG27	0.676	-0.274	-0.076	0.286	-0.065	0.624
AG28	0.653	-0.260	-0.226	0.346	0.131	0.683
AG29	0.619	-0.310	-0.102	0.366	-0.101	0.633
AG30	0.502	-0.237	-0.177	0.532	0.091	0.632
% Var.	45.1	6.2	5.5	4.6	3.1	64.5

Note. KMO = 0.929; Bartlett's $\chi^2(435) = 7618.7$, $p < .001$. Bold indicates primary loading. h^2 = communality. Total variance explained = 64.5%.

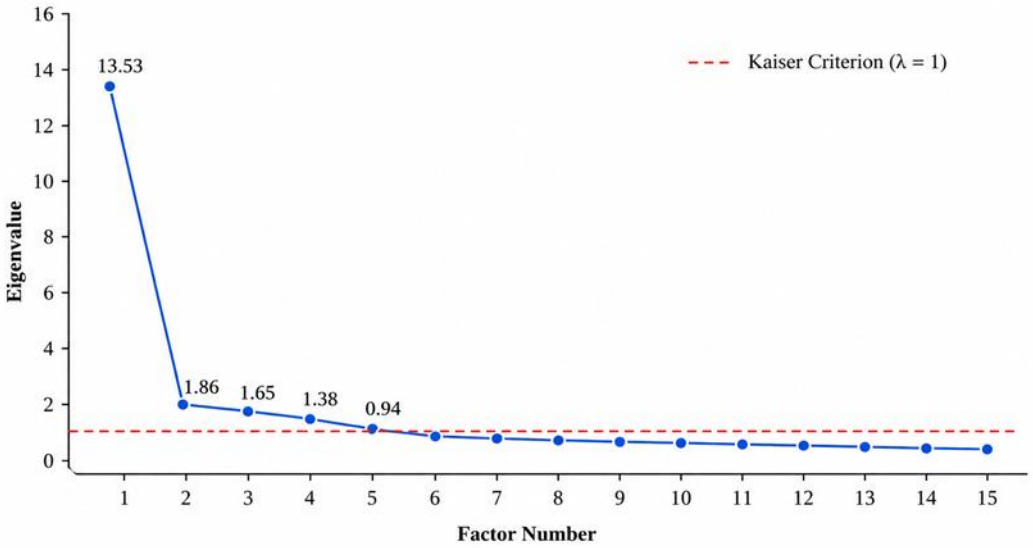


Figure 2. Scree plot of eigenvalues for the Academic Agility Scale (30 items).

Note. The scree plot illustrates the distribution of eigenvalues across extracted factors. A clear point of inflection is observed after the fifth factor, supporting the retention of a five-factor solution. The dashed horizontal line represents the Kaiser criterion (eigenvalue = 1), indicating that factors above this threshold were retained.

1.3 Confirmatory Factor Analysis (CFA)

The researchers performed a confirmatory factor analysis to test the five-factor model which was discovered through exploratory factor analysis. Table 4 displays the standardized factor loadings. The 30 items demonstrated significant loading results for their respective latent factors with their standardized loadings (λ) falling between .712 and .853. The study results showed that all loadings exceeded the .70 threshold which Hair et al. (2019) established as a standard, thus showing strong evidence for convergent validity at the item level.

Table 4 CFA Standardized Factor Loadings (λ)

Item	Dimension	λ
AG1	Responsiveness	0.840
AG2	Responsiveness	0.849
AG3	Responsiveness	0.851
AG4	Responsiveness	0.800
AG5	Responsiveness	0.802
AG6	Adequacy	0.760
AG7	Adequacy	0.754
AG8	Adequacy	0.762
AG9	Adequacy	0.769
AG10	Adequacy	0.733
AG11	Flexibility	0.755
AG12	Flexibility	0.744
AG13	Flexibility	0.745
AG14	Flexibility	0.712
AG15	Flexibility	0.813
AG16	Flexibility	0.770
AG17	Flexibility	0.821
AG18	Speed	0.734
AG19	Speed	0.758

AG20	Speed	0.779
AG21	Speed	0.727
AG22	Speed	0.748
AG23	Speed	0.772
AG24	Speed	0.783
AG25	Speed	0.727
AG26	Integration	0.815
AG27	Integration	0.829
AG28	Integration	0.787
AG29	Integration	0.853
AG30	Integration	0.809

Note. All standardized loadings significant at p < .001. "Strong" = λ ≥ .80; "Good" = .70 ≤ λ < .80. See Table 5 for model fit indices and Figure 3 for the path diagram.

The table 5 displays the results for model fit indices. The chi-square test was significant, $\chi^2(395) = 1741.9$, $p < .001$, which is expected given the large sample size (N = 380) and number of indicators (30 items). The normed chi-square ($\chi^2/df = 4.41$) was below the 5.0 threshold, indicating acceptable fit. The RMSEA (0.095) fell within the acceptable range (≤ .10). The CFI (0.819) and TLI (0.800) approached but did not reach the conventional .90 cutoff. The measurement model proves adequate through strong individual loadings (Table 4) and convergent validity evidence (Table 7).

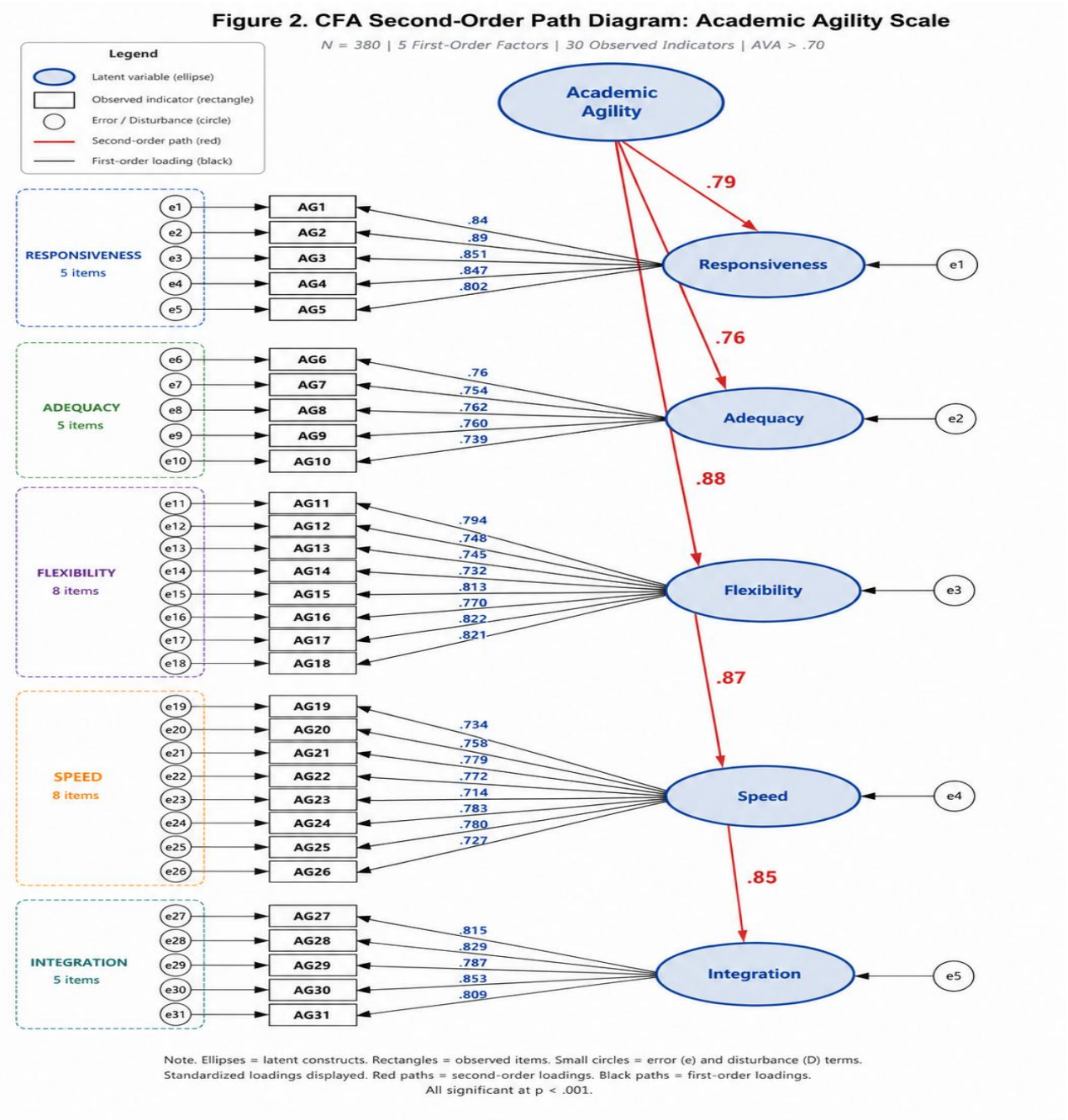
Table 5 CFA Model Fit Indices

Index	Value	Criterion	Evaluation
χ^2	1741.9	p > .05	Significant (N=380)
df	395	—	—
χ^2/df	4.41	< 5.0	Acceptable
CFI	0.819	≥ .90	Approaching
TLI	0.800	≥ .90	Approaching
RMSEA	0.095	≤ .10	Acceptable
GFI	0.778	≥ .90	Below
NFI	0.778	≥ .90	Below

Note. χ^2 is sensitive to sample size; χ^2/df provides a more robust assessment. RMSEA values ≤ .08 indicate good fit; ≤ .10 indicate acceptable fit (Browne & Cudeck, 1993). Model respecification (e.g., allowing correlated errors for similarly worded items) may improve incremental indices.

Figure 3 presents the complete second-order CFA path diagram in AMOS output format. The second-order factor (Academic Agility) significantly predicted all five first-order factors, with second-order loadings ranging from .76 to .88. Disturbance terms (D1–D5) are included for each endogenous latent factor.

Figure 3 CFA Second-Order Path Diagram: Academic Agility Scale



Note. Ellipses = latent constructs. Rectangles = observed indicators. Small circles = error terms (e1–e30) and disturbance terms (D1–D5). Red paths = second-order loadings. Black paths = first-order standardized loadings. All coefficients significant at p < .001. See Tables 4–5 for numerical values.

1.4 adequate internal structure validity

The researchers used corrected item-dimension and corrected item-total correlations to study the internal structure of the scale according to the results shown in Table 6. The corrected item-dimension correlations ranged from .550 (AG10) to .765 (AG29), which all exceeded the minimum threshold of .30 that Nunnally and Bernstein established in 1994. The total item correlations showed values between .471 for AG6 and .717 for AG12 which demonstrated that all items could effectively differentiate between subscale and total scale measurements. Nineteen of the 30 items required reverse coding.

Table 6 Corrected Item-Dimension and Item-Total Correlations

Item	Dimension	r (Item-Dim)	r (Item-Total)	Rev
AG1	Responsiveness	0.733	0.588	
AG2		0.754	0.657	R
AG3		0.755	0.609	
AG4		0.686	0.650	
AG5		0.692	0.653	R
AG6	Adequacy	0.609	0.471	R
AG7		0.603	0.540	
AG8		0.616	0.491	R
AG9		0.618	0.561	R
AG10		0.550	0.654	R
AG11	Flexibility	0.654	0.695	R
AG12		0.641	0.717	
AG13		0.644	0.614	
AG14		0.598	0.584	
AG15		0.732	0.679	R
AG16		0.673	0.627	R
AG17		0.745	0.716	R
AG18	Speed	0.641	0.674	R
AG19		0.669	0.676	R
AG20		0.699	0.668	
AG21		0.635	0.613	R
AG22		0.659	0.676	R
AG23		0.692	0.684	R
AG24	Integration	0.705	0.691	
AG25		0.631	0.628	R
AG26		0.699	0.699	R
AG27		0.724	0.656	R
AG28		0.668	0.666	R
AG29		0.765	0.706	
AG30		0.679	0.661	

Note. All correlations significant at $p < .001$. "R" = reverse-coded. Minimum threshold = .30 (Nunnally & Bernstein, 1994). See Figure 4(a) for a visual representation of item discrimination indices.

1.5 convergent and discriminant validity

The researchers evaluated convergent and discriminant validity through testing which used Average Variance Extracted (AVE) and Composite Reliability (CR) and Fornell-Larcker criterion and Heterotrait-Monotrait (HTMT) ratio. The table displays AVE values together with CR values and Fornell-Larcker matrix results. The study confirmed

that all AVE values exceeded the .50 threshold (range: .568–.686) because each latent factor explained more than half of its indicator variance. The results showed all CR values exceeded .70 (range: .869–.916) which provided additional evidence of convergent validity. (See. Table. 7)

The Fornell-Larcker criterion requires discriminant validity because each dimension must show square root AVE values that match diagonal values which should equal or exceed all inter-factor correlations that exist in that specific row and column, The majority of dimension pairs satisfied this requirement. The HTMT ratios in Table 8 function as a more stringent evaluation. The researchers found that eight out of ten pairs satisfied the strict requirement (HTMT < .85) while no pair exceeded the threshold of .90 which confirmed overall discriminant validity for the study (Henseler et al., 2015).

Table 7 Convergent and Discriminant Validity: AVE, CR, and Fornell-Larcker

Dimension	AVE	CR	Resp.	Adeq.	Flex.	Speed	Integ.
Responsiveness	0.686	0.916	0.828	0.571	0.649	0.616	0.604
Adequacy	0.571	0.869	0.571	0.756	0.613	0.637	0.535
Flexibility	0.588	0.909	0.649	0.613	0.767	0.784	0.721
Speed	0.568	0.913	0.616	0.637	0.784	0.754	0.792
Integration	0.671	0.911	0.604	0.535	0.721	0.792	0.819

Note. Diagonal = √AVE; off-diagonal = inter-factor correlations. AVE > .50 and CR > .70 confirm convergent validity (Fornell & Larcker, 1981). See Table 8 for HTMT ratios.

Table 8 HTMT Ratio for Discriminant Validity

Dimension Pair	HTMT	Evaluation
Responsiveness <-> Adequacy	0.669	Pass (<.85)
Responsiveness <-> Flexibility	0.738	Pass (<.85)
Responsiveness <-> Speed	0.696	Pass (<.85)
Responsiveness <-> Integration	0.686	Pass (<.85)
Adequacy <-> Flexibility	0.718	Pass (<.85)
Adequacy <-> Speed	0.747	Pass (<.85)
Adequacy <-> Integration	0.630	Pass (<.85)
Flexibility <-> Speed	0.884	Caution (<.90)
Flexibility <-> Integration	0.820	Pass (<.85)
Speed <-> Integration	0.895	Caution (<.90)

Note. HTMT < .85 = strict criterion; < .90 = liberal criterion (Henseler et al., 2015). No pair exceeds .90.

Q2: Does the scale demonstrate criterion-related (predictive and known-groups) validity?

The researchers conducted additional research to answer the second research question about criterion-related validity. The study used regression analyses to test predictive validity because it required external variables estimation and it used group comparisons to test known-groups validity. The researchers assessed scale reliability through internal consistency measures and additional reliability indices which ensured score stability and consistency assessment.

2.1 Predictive Validity

used simple linear regression analyses to test Academic Agility as a potential predictor of external variables that constructed criterion-related validity. Academic Agility showed significant predictive power for Academic Perfectionism ($\beta = .108$, $p < .05$, $R^2 = .012$) and Research Anxiety ($\beta = -.105$, $p < .05$, $R^2 = .011$). The positive relationship to perfectionism demonstrates that agile students establish more demanding academic requirements while their ability to conduct research decreases research anxiety.

Table 9 Predictive Validity: Academic Agility as Predictor of External Criteria

Criterion	β	R^2	p
Academic Perfectionism	0.108	0.012	< .05
Research Anxiety	-0.105	0.011	< .05

Note. Simple linear regression. Effect sizes small but consistent with theoretical predictions for related yet distinct constructs.

2.2 Known-Groups Validity

The scale's ability to differentiate between theoretically relevant groups was tested using an independent-samples t-test (Table 10). Male students ($M = 3.45$, $SD = 0.76$) reported significantly higher academic agility than female students ($M = 3.14$, $SD = 0.68$), $t = 3.333$, $p < .001$, Cohen's $d = 0.42$. The scale can identify group differences because its small-to-medium effect size shows sufficient sensitivity for this task.

Table 10 Known-Groups Validity: Gender Differences in Academic Agility

Group	n	M	SD	t	p	Cohen's d
Male	300	3.450	0.761	3.333	< .001	0.421
Female	79	3.136	0.681			

Note. Independent-samples t-test. Cohen's d: .20 = small, .50 = medium, .80 = large (Cohen, 1988).

2.3. Internal Consistency

The evaluation of internal consistency used Cronbach's alpha method which is shown in Table 11. The total scale showed excellent reliability because its alpha value reached .957. The five dimensions exhibited reliable performance because their alpha scores ranged from .811 for Adequacy to .891 for Speed. The alpha-if-item-deleted analysis appears in Table 12. The analysis confirmed that all 30 items contribute positively to scale reliability because no item's removal would increase the total alpha beyond .957. The scale achieves outstanding measurement strength because its alpha-if-deleted values remain within a narrow range of .955 to .957 (see Figure 4b).

Table 11 Internal Consistency: Cronbach's Alpha by Dimension

Dimension	Items	α	Interpretation
Responsiveness	5	.886	Good
Adequacy	5	.811	Good
Flexibility	7	.882	Good
Speed	8	.891	Good
Integration	5	.876	Good
Total Scale	30	.957	Excellent

Note. Interpretation: ≥ .90 = Excellent; ≥ .80 = Good; ≥ .70 = Acceptable (George & Mallery, 2019).

Table 12 Cronbach's Alpha if Item Deleted

Item	α if deleted	Decision
AG1	0.957	Retain
AG2	0.956	Retain
AG3	0.956	Retain
AG4	0.956	Retain
AG5	0.956	Retain
AG6	0.957	Retain
AG7	0.957	Retain
AG8	0.957	Retain
AG9	0.957	Retain
AG10	0.956	Retain
AG11	0.956	Retain
AG12	0.955	Retain
AG13	0.956	Retain
AG14	0.957	Retain
AG15	0.956	Retain
AG16	0.956	Retain
AG17	0.955	Retain
AG18	0.956	Retain
AG19	0.956	Retain
AG20	0.956	Retain
AG21	0.956	Retain
AG22	0.956	Retain
AG23	0.956	Retain
AG24	0.956	Retain
AG25	0.956	Retain
AG26	0.956	Retain
AG27	0.956	Retain
AG28	0.956	Retain
AG29	0.956	Retain
AG30	0.956	Retain

Note. Total α = 0.957. Range: .955–.957. No deletion improves reliability. See Figure 4(b) for visual diagnostic.

2.4. Additional Reliability Evidence

The table 13 presents multiple methods to estimate reliability through their three complementary approaches. The Spearman-Brown corrected split-half coefficient (.968) and Guttman split-half coefficient (.968) confirmed that even arbitrary halving of the scale preserves high reliability. McDonald's omega (.884) provided a model-based estimate that accounts for varying factor loadings. The measurement reliability of the study results is established through the congruence of all measurement indices which exceed .80 between classical test theory (CTT) and model-based assessment methods.

Table 13 Reliability Across Multiple Estimation Methods

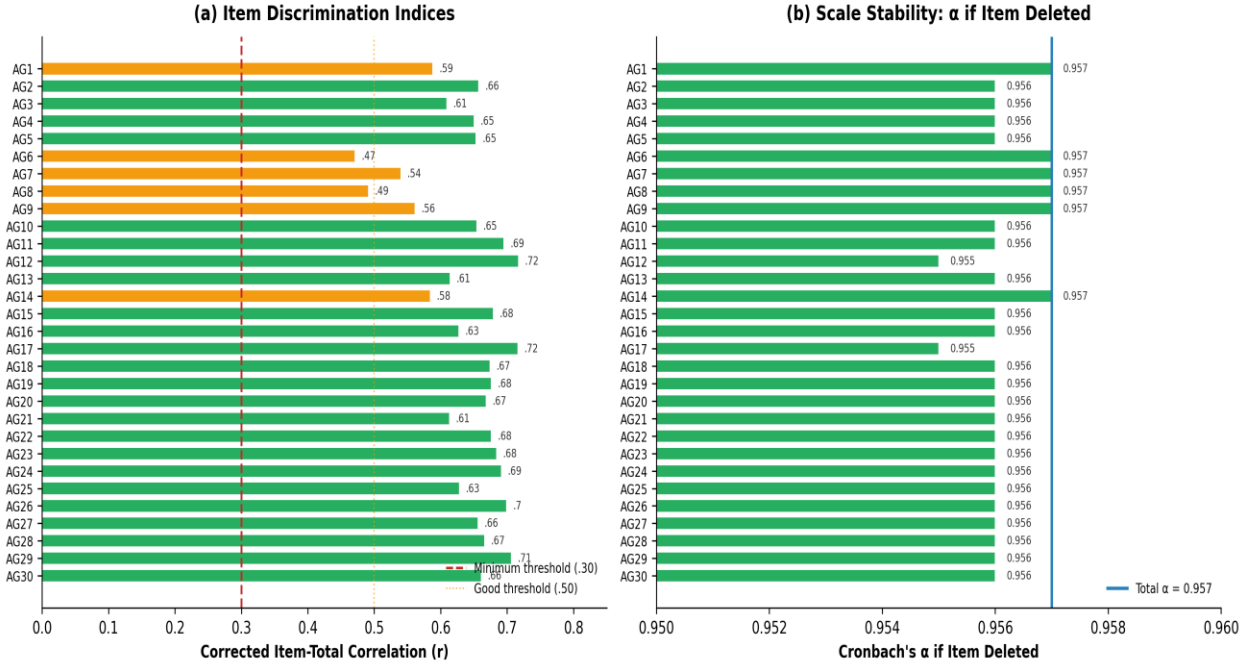
Method	Coefficient	Interpretation
Cronbach's Alpha	0.957	Excellent
Split-Half (r)	0.939	Excellent
Spearman-Brown	0.968	Excellent
Guttman Split-Half	0.968	Excellent
McDonald's Omega (ω)	0.884	Good

Note. Convergence across CTT (alpha, split-half) and model-based (omega) methods confirms robust measurement reliability.

Figure 4 provides complementary visual diagnostics for reliability. Panel (a) displays corrected item-total correlations, confirming that all items exceed the minimum discrimination threshold. Panel (b) displays alpha-if-deleted values, visually demonstrating that no single item's removal would improve total reliability.

Figure 4 Reliability Diagnostics: Item Discrimination and Scale Stability

Figure 3 Reliability Diagnostics: Academic Agility Scale (N = 380)



Note. Panel (a): Corrected item-total correlations. Red dashed line = minimum threshold (.30); orange dotted line = good threshold (.50). All items exceed .30; 27/30 exceed .50; 14/30 exceed .60. Panel (b): Alpha if item deleted. Blue reference line = total a (.957). No deletion improves a. See Tables 6 and 12 for numerical values.

4. DISCUSSION

The study aimed to establish complete psychometric proof for the Academic Agility Scale in higher education settings. The study results reveal that academic agility consists of multiple

dimensions which contain essential theoretical and practical elements that require further research.

The exploratory factor analysis established five factors which researchers demonstrated to be equivalent to the models which define organizational agility as a dynamic capability that combines responsive abilities and flexible abilities and speedy abilities and integrated abilities (Karimian & Chahartangi, 2024) Educational institutions operate as adaptive systems because they need to create their educational functions. Educational organizations have demonstrated their ability to implement organizational agility frameworks which prove valuable during digital transformation and educational changes after the pandemic (Hu & Bentler, 1999; Karimian & Chahartangi, 2024)

The confirmatory factor analysis confirmed the scale's internal structure through item-level testing because all factor loadings produced strong results that achieved statistical significance. The observed variables function as adequate representations of their underlying latent constructs, which supports the convergent validity assessment (Kline, 2016; Marsh et al., 2004) The proposed model required additional refinement because its global model fit indices did not reach the established conventional thresholds. Initial validation studies commonly exhibit these structural equation modeling discrepancies, as researchers encounter item redundancy problems and measurement errors and conceptual overlap between related dimensions (Browne & Cudeck, 1993; Kline, 2016)

The discriminant validity results show additional information about the problem, The Fornell–Larcker criterion received mostly fulfilled requirements, while several HTMT values reached close to the upper threshold, especially between flexibility and speed and integration(Hodges et al., 2020) The pattern shows partial conceptual overlap, which researchers may use as a theoretical explanation, Educational institutions function as complex adaptive systems, because their various dimensions deserve interdependencies; for example, increased flexibility for instructional delivery enables faster educational responses, yet integration provides operational unity across institutional divisions(Rapanta et al., 2020; Reise et al., 2013; Teece, 2007). Agility research demonstrates that dimensions become less distinct through their systemic character because researchers have reported similar findings in previous studies(Henseler et al., 2015) The results from this study deliver significant real-world value. The scale functions as an effective assessment instrument because it can measure both fixed dimensions and their interactive relationships, which enables users to evaluate institutional readiness and flexibility. The scale enables educational leaders to assess their institutional strengths and weaknesses in agility components, which they can use to create their strategic plan and redesign the curriculum and implement digital transformation projects. The integration dimension demonstrates how essential it is for organizations to synchronize their different parts because this synchronization process functions as the vital element which enables higher education institutions to implement successful changes(Hu & Bentler, 1999; Marsh et al., 2004)

The analysis of criterion-related validity establishes which academic agility functions in educational systems. The study found that academic agility had a strong relationship with academic perfectionism and research anxiety, but the relationship strength was low. The positive association with perfectionism suggests that agile students establish higher academic standards which drive their goal-directed activities. The negative relationship with research anxiety indicates that agile individuals possess better skills to handle uncertain situations and complicated tasks, which helps them decrease their stress during research activities. Academic agility has low explained variance because it functions as one of several elements which lead to these outcomes (Cohen, 1988; Doz & Kosonen, 2010). The recent psychological models explain academic performance and emotional regulation as outcomes of multiple interacting elements.

The known-groups validity analysis provided additional, albeit limited, support for the scale. Male students demonstrated better academic agility than female students because their male counterparts reported higher agility levels. The finding shows that the scale detects differences between groups, but researchers must interpret the results carefully

because the study sample contains more of one group than the other. Demographic variables interact with contextual and cultural factors to determine how people view their adaptability and work performance. The generalizability of the results need verification through future studies which should use more balanced and diverse samples.

The scale achieved perfect internal consistency across its entire measurement range according to three different evaluation techniques which included Cronbach's alpha and split-half reliability and McDonald's omega. The measurement results provide strong evidence for both the measurement stability and measurement consistency (Cronbach, 1951). The study found that the research participants displayed extremely high alpha coefficient results, which indicates that research participants experienced redundancy issues with the items presented to them. High reliability scores create psychometric assessment problems because they show item similarities instead of true construct diversity, which leads researchers to investigate methods for reducing items to create a simpler version of the scale that maintains its validity (Tavakol & Dennick, 2011)

The study results validate the Academic Agility Scale as an educational assessment tool based on its theoretical framework and empirical research. The results demonstrate that ongoing model development remains necessary because both model fit and dimensional distinctiveness need improvement. Research should examine measurement invariance testing and academic agility anticipatory capacity for educational outcomes across various populations to verify the findings through cross-validation.

5. CONCLUSION

The Academic Agility Scale shows operational value as a multidimensional assessment tool which performs with theoretical precision according to the research results which demonstrate its complete psychometric validity. The study results confirm the factor structure of the instrument while showing that it has strong internal consistency and it achieves both convergent and discriminant validity at acceptable levels. The assessment shows that the scale provides dependable results which represent academic agility through its complex and evolving definition.

The scale provides educational institutions with a powerful diagnostic and evaluative tool which helps them improve their ability to respond to changes while maintaining flexible operations and better system integration. The system provides data-driven decision support for curriculum development and digital transformation and institutional performance evaluation. The measurement model needs improvements because academic agility functions as a single predictive element among multiple factors which influence academic outcomes, according to the results which show weak predictive capacity and poor global model fit. Research needs to assess the usability of the scale across different cultural contexts and population groups while investigating its measurement invariance of which its predictive capacity toward critical educational outcomes needs to be studied. The Academic Agility construct will gain stronger theoretical foundations through these efforts while its practical applications will increase**.**

6. Limitations & Strengths

The research findings gain their proper meaning through their assessment of multiple existing research limitations. The research findings suffer from their cross-sectional design which prevents researchers from establishing cause-and-effect relationships between the studied variables. The sample demonstrated demographic imbalance because the study did not include equal numbers of males and females while also failing to represent all GPA categories which limits the study's generalizable results. The factor structure received mostly validation but certain model fit indices fell short of required standards which created a need for extra model development work. Self-reporting methods used in the study introduce the possibility of common method bias

The research study shows important advantages which exceed its existing research limitations. The research study presents a thorough psychometric assessment of the Academic Agility Scale through multiple methods which include both exploratory and confirmatory factor analyses together with validity assessments through convergent,

discriminant, and criterion-based methods. The study's findings gain strength from the large sample size, while multiple reliability assessment techniques enhance measurement stability through their combined effect. These qualities establish the scale as an essential research instrument which academic institutions can use in their educational research and practical applications.

.7. Implications and Recommendations

The research results present theoretical and practical applications which benefit both higher education research and its operational work. The Academic Agility Scale achieves its theoretical validation by demonstrating its value to educational adaptive capability research which defines academic agility as a multidimensional educational construct. The framework provides educational institutions with a systematic approach which organizations can use to assess their educational agility and dynamic educational capabilities.

The scale helps educational institutions evaluate their organizational capacity and their ability to handle changes which organizations require. The tool enables educational leaders and policymakers to assess core areas of their operations which include their capacity to adapt to changes and their ability to operate across multiple functions which helps them plan their future operations and design new educational programs and digital transformation projects. The instrument supports universities in their efforts to maintain quality standards and their ongoing development work. The research outcomes lead to the development of multiple recommendations. The research team should test the scale for diverse populations to achieve improved generalization results while demonstrating how the scale operates across different demographic groups. The model requires further development to achieve better overall performance while minimizing the presence of duplicate items. Researchers should examine how academic agility predicts essential educational outcomes which include student achievement and student participation and student well-being.

Data availability

The data supporting the findings of this study are available from the corresponding author upon reasonable request, subject to ethical and confidentiality considerations.

Acknowledgements

We thank all who participated in this study.

Funding Statement

Not applicable

CONFLICT OF INTEREST

The authors declare no competing interests.

Competing interests

The authors declare no competing interests.

Clinical trial number

Not applicable

REFERANC

Anthoine, E., Moret, L., Regnault, A., Sebille, V., & Hardouin, J.-B. (2014). Sample size used to validate a scale: A review of publications on newly-developed patient reported outcomes measures. *Health and Quality of Life Outcomes*, *12*, 176. https://doi.org/10.1186/s12955-014-0176-2

Boateng, G. O., Neilands, T. B., Frongillo, E. A., Melgar-Quinonez, H. R., & Young, S. L. (2018). Best practices for developing and validating scales for health, social, and behavioral research: A primer. *Frontiers in Public Health*, *6*, 149. https://doi.org/10.3389/fpubh.2018.00149

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing Structural Equation Models* (pp. 136-162). Sage.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Lawrence Erlbaum Associates.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*(3), 297-334. https://doi.org/10.1007/BF02310555

Doz, Y. L., & Kosonen, M. (2010). Embedding strategic agility: A leadership agenda for accelerating business model renewal. *Long Range Planning*, *43*(2-3), 370-382. https://doi.org/10.1016/j.lrp.2009.07.006

Goldman, S. L., Nagel, R. N., & Preiss, K. (1994). *Agile competitors and virtual organizations: Strategies for enriching the customer*. Van Nostrand Reinhold.

Gunawan, J., Marzilli, C., & Aungsuroch, Y. (2021). Establishing appropriate sample size for developing and validating a questionnaire in nursing research. *Nursing Open*, *8*(4), 1530-1536. https://doi.org/10.1002/nop2.867

Hazell, C. M., Chapman, L., Valeix, S. F., Roberts, P., Niven, J. E., & Berry, C. (2020). Understanding the mental health of doctoral researchers: a mixed methods systematic review with meta-analysis and meta-synthesis. *Systematic reviews*, *9*(1), 197.

Henseler, J. r., Ringle, C. M., & Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the Academy of Marketing Science*, *43*(1), 115-135. https://doi.org/10.1007/s11747-014-0403-8

Hodges, C., Moore, S., Lockee, B., Trust, T., & Bond, A. (2020). The difference between emergency remote teaching and online learning. *Educause Review*.

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

Karimian, Z., & Chahartangi, F. (2024). Development and validation of a questionnaire to measure educational agility: A psychometric assessment using exploratory factor analysis. *BMC Medical Education*, *24*, 1284. https://doi.org/10.1186/s12909-024-06307-z

Khatatbeh, Y., & Mistarihi, W. (2025). EXAMINING THE FACTORIAL STRUCTURE OF THE SOCIAL COMPARISON SCALE IN THE LOCAL ENVIRONMENT: A PSYCHOMETRIC STUDY. . *TPM–Testing, Psychometrics, Methodology in Applied Psychology,* , *4*(32), ,12-23..

Kline, R. B. (2016). *Principles and Practice of Structural Equation Modeling* (4th ed.). Guilford Press.

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

Lin, X. (2026). Ignorance as a Method: Rethinking Academic Burden in an Accelerated Society. *ECNU Review of Education*, *9*(2), 20965311231155243.

Marsh, H. W., Hau, K.-T., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes. *Structural Equation Modeling*, *11*(3), 320-341. https://doi.org/10.1207/s15328007sem1103_2

Morgado, F. F. R., Meireles, J. F. F., Neves, C. M., Amaral, A. C. S., & Ferreira, M. E. C. (2017). Scale development: Ten main limitations and recommendations to improve future research practices. *SpringerPlus*, *6*, 53. https://doi.org/10.1186/s40064-016-2687-6

Nahar, K., & Sekhar, A. (2024). Perfectionism, Anxiety and Career Distress-Insight from Undergraduate/Postgraduate Final Year Students.

Naz, S., Hassan, G., & Luqman, M. (2021). Influence of perfectionism and academic motivation on academic procrastination among students. *Journal of Educational Sciences & Research*, *8*(1).

Nelson, T. A. (2000). *Kubrick, New and Expanded Edition: Inside a Film Artist's Maze*. Indiana University Press.

Onwuegbuzie, A. J. (2013). Development and score-validation of the Research Anxiety Rating Scale. *International Journal of Research in Education Methodology*, *4*(2), 503-515.

Rapanta, C., Botturi, L., Goodyear, P., Guardia, L., & Koole, M. (2020). Online university teaching during and after the COVID-19 crisis: Refocusing teacher presence and learning activity. *Postdigital Science and Education*, *2*, 923-945. https://doi.org/10.1007/s42438-020-00155-y

Reise, S. P., Moore, T. M., & Haviland, M. G. (2013). Bifactor models and rotations: Exploring the extent to which multidimensional data yield univocal scale scores. *Journal of personality assessment*, *95*(3), 1-13. https://doi.org/10.1080/00223891.2012.725437

Rostami, F. (2018). Development and psychometric evaluation of a scale measuring psychological constructs. *Journal of Education and Health Promotion*, *7*, 158. https://doi.org/10.4103/jehp.jehp_180_18

Sharifi, H., & Zhang, Z. (1999). A methodology for achieving agility in manufacturing organisations: An introduction. *International Journal of Production Economics*, *62*(1-2), 7-22. https://doi.org/10.1016/S0925-5273(98)00217-5

Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education*, *2*, 53-55. https://doi.org/10.5116/ijme.4dfb.8dfd

Teece, D. J. (2007). Explicating dynamic capabilities: The nature and microfoundations of (sustainable) enterprise performance. *Strategic Management Journal*, *28*(13), 1319-1350. https://doi.org/10.1002/smj.640

Trochim, W. M. K. (2006). *Research Methods Knowledge Base* (2nd ed.). Atomic Dog Publishing.

Worley, C. G., & Lawler, E. E. (2010). Agility and organization design: A diagnostic framework. *Organizational Dynamics*, *39*(2), 194-204. https://doi.org/10.1016/j.orgdyn.2010.01.006