ACADEMIA International Journal for Social Sciences
Volume 4, Issue 5, 2025 ISSN-L (Online): 3006-6638

Development And Psychometric Validation of The University Students' Productivity Scale (USPS): Evidence from Public-Sector Universities of Punjab, Pakistan

Muhammad Shahzad
msmpu@yahoo.com
PhD Scholar, University of the Punjab, Lahore, Pakistan.

Dr. Muhammad Shahid Farooq
Professor, University of the Punjab, Lahore, Pakistan.

Corresponding Author: Muhammad Shahzad msmpu@yahoo.com

Received: 19-09-2025 **Revised:** 03-10-2025 **Accepted:** 16-10-2025 **Published:** 30-10-2025

ABSTRACT

Self-report instruments that capture how university students experience academic productivity in artificial intelligence (AI)-supported study contexts are scarce. This study reports the development and psychometric evaluation of the University Students' Productivity Scale (USPS), a 50-item instrument covering ten theoretically derived dimensions: time management, task completion, goal achievement, efficiency, quality of work, focus and concentration, collaboration and teamwork, stress management, resource utilization and problem-solving ability. Items were rated as percentage agreement (0–100%) by 1,190 undergraduates from 12 public-sector universities in Punjab, Pakistan. Item analysis, parallel analysis, Velicer's minimum average partial test, confirmatory factor analysis of four competing models (one-factor, ten correlated factors, second-order and bifactor) with split-sample cross-validation, reliability estimation, differential item functioning screening and convergent–discriminant analyses were conducted. All items contributed to a single, well-fitting general factor (CFI = .948, RMSEA = .013, SRMR = .028; ω = .86, α = .86), and the solution replicated in two random halves. The ten dimensions, however, were not empirically distinguishable: parallel analysis and the MAP test each supported one factor, dimension scores had low internal consistency (α = .32–.42), and their disattenuated intercorrelations averaged 1.04. Scores did not differ by gender, background or discipline, and no item showed meaningful differential functioning. The USPS correlated strongly with scales of AI practices (r = .87) and metacognition (r = .85), but discriminant validity relative to these scales was not established. The USPS is supported as a reliable single-score measure of perceived AI-supported academic productivity for group-level research; subscale scoring is not recommended.

Keywords: *scale development; academic productivity; artificial intelligence; higher education; confirmatory factor analysis; reliability; discriminant validity*

INTRODUCTION

Artificial Intelligence (AI) has become an integral part of educational technology, transcending its status as a specialized field to become a central component of everyday learning. Students at universities are now utilizing generative AI, intelligent tutoring systems, writing assistants, reference management and planning applications to write and edit assignments, organize their work time, search the literature, and solve problems (Kasneci et al., 2023; Zawacki-Richter et al., 2019; Castillo-Martínez et al., 2024). The key point that is being debated, both from a practical standpoint for universities and from a policy perspective, is whether such tools increase student productivity and, if they do, how.

There is a tentative positive trend in evidence, but it's not conclusive. In experiments performed in the workplace, having access to a generative AI assistant can save approximately 40% of time for completing professional writing tasks and improve rated output quality by approximately 18% (Noy & Zhang, 2023). Learning contexts are varied. High-school students in a large-scale experiment

showed a higher performance rate when they practiced with an unrestricted GPT-4 tutor compared to those who did not have access to a GPT-4 tutor, but lower performance rates on subsequent unaided tests (Bastani et al., 2025). The scores for the essays written with ChatGPT support did not improve knowledge gai3 or transfer, and the authors cautioned for "metacognitive laziness" in the study of randomly selected university students (Fan et al., 2025). There is also a systemic connection between overreliance on AI dialogue systems and reduced critical thinking and analytical capacity (Zhai et al., 2024). Productivity—in other words, something other than learning—should by no means be confused with learning, and should be equally carefully measured as any other educational outcome.

This care is currently lacking. Most of the empirical studies on student AI use are based on single items and ad hoc questionnaires or tools developed from existing instruments whose dimensional structure, reliability and invariance have not been tested in the population under study. Existing measures like the Time Management Behavior Questionnaire (Macan et al., 1990) and the Motivated Strategies for Learning Questionnaire (Pintrich et al., 1993) focus on general study behaviour and self-regulation and were not developed to address students' experience of their productivity when AI tools are integrated into the study process. If there is no validated measure then the comparisons across studies and groups becomes unstable and inferences about the "effect" of AI on productivity are made on the basis of scores that are open to interpretation (Clark & Watson, 2019; Messick, 1995).

But the context is important too. The current body of evidence regarding the use of AI in higher education comes from a small number of countries, and from technical fields (Zawacki-Richter et al., 2019). Large and diverse student population is enrolled by universities in public sector of Pakistan in Punjab, including the students from the rural areas in general, women and medical, engineering, agricultural and vocational universities. There are few known instruments that measure academic productivity using AI, which have been developed and tested in such an environment.

The present study thus constructs and tests the University Students' Productivity Scale (USPS) which consists of 50 items, each requiring self-report, based on 10 theoretically derived dimensions of academic productivity. Responses of 1,190 undergraduates are used to explore the item properties, dimensionality, reliability, cross-validated measurement structure, equivalence across gender, social background and discipline, and relationships with related constructs of the scale.

LITERATURE REVIEW

Academic productivity as a multidimensional construct

Academic productivity is the ratio of time and effort spent to the amount of good quality work produced. It is not just about quantity of output, but also learning behaviour, cognitive engagement and self-management (Zimmerman, 2002; Credé & Kuncel, 2008). It is the best-researched component of time management: Planning, prioritising, and scheduling study time has been shown to correlate with improved academic results and stress levels (Macan et al., 1990; Britton & Tesser, 1991); a meta-analysis further supports small but consistent relationships between time management and academic performance as well as well-being (Aeon et al., 2021). In addition, self-regulated learning theory introduces goal setting and monitoring and persistence as other components, and students with specific goals are more likely to succeed in completing tasks, as are students who monitor progress to complete tasks (Zimmerman, 2002; Bandura, 1997).

Also, attention and work load. According to cognitive load theory, working memory has limited capacity, and thus, when it is cluttered, interruptions and disorganized materials can reduce the working memory available to learn (Sweller, 1988).

In today's university context, where much work is done in groups and workload stress is a concern (Macan et al., 1990), collaboration and stress management round it off the picture. This literature

suggests a multi-faceted definition of productivity that includes time use, task completion, goal attainment, efficiency, work quality, concentration, collaboration, stress regulation, resource use and problem solving.

AI and academic productivity

AI tools have a potential to directly impact all of these factors. Planning: Scheduling and reminder applications can support this; Writing and language assistants can speed the drafting and revision process; Reference managers and AI-assisted search tools can reduce the effort of locating and organizing sources; conversational agents can provide on-demand assistance with problems (Kasneci et al., 2023; Chauke et al., 2024). This range of applications is reflected in the reports of AI in Higher Education (Castillo-Martínez et al., 2024; Zawacki-Richter et al., 2019), as well as in the studies of students' productivity and stress when using AI (Li & Jan, 2023). Learning analytics (Uzir et al., 2020) has been studied in the context of time management in technology-supported learning.

Yet the benefits may be accompanied by costs. Gains in task performance can coexist with weaker later unaided performance (Bastani et al., 2025), with reduced engagement in self-regulatory processes (Fan et al., 2025) and with over-reliance (Zhai et al., 2024). These findings do not make productivity measures less useful; they make it more important to measure productivity as a distinct construct, separable from learning and from the frequency of AI use, so that its relationships with these other variables can be examined rather than assumed.

Measuring perceived productivity: requirements and gaps

he process of good scale development starts with a clear construct definition, continues with item generation and then expert review, then moves to pilot testing and evaluation in a larger independent sample, with explicit attention to dimensionality, reliability and invariance and validity (DeVellis & Thorpe, 2022; Boateng et al., 2018; Clark & Watson, 2019). Three points are highly relevant here. First a multidimensional scale must demonstrate that its dimensions are empirically distinct not just conceptually distinct; otherwise the scores on the subscales do not contribute to the total score. Second, the reliability estimates should be as close as possible to the measurement model, ideally omega along with alpha (McDonald, 1999; Hayes & Coutts, 2020; Sijtsma, 2009). Third, validity evidence must contain discriminant evidence of the absence of a mere rewording of related measures (Fornell & Larcker, 1981; Henseler et al., 2015). Self-report scales that share a response format and have a uniformly favorable item direction are also susceptible to common-method variance (Podsakoff et al., 2003, 2012) that can cause an overestimate of relationships among scales.

Conceptual basis of the USPS and the present study

The USPS was designed to measure students' perceived academic productivity in AI-supported study contexts. Drawing on the literature above, ten dimensions were specified, each represented by five items (Table 1). Because every item refers to AI-supported work, the construct is best described as perceived AI-supported academic productivity rather than general academic productivity. The present study addresses five research questions. RQ1: What are the item-level properties of the USPS? RQ2: How many factors underlie the 50 items? RQ3: Which measurement model (one-factor, ten correlated factors, second-order or bifactor) best represents the data, and does it replicate across independent halves of the sample? RQ4: Is the scale reliable, and does it function equivalently across gender, social background and discipline? RQ5: What are its convergent, discriminant and nomological relationships with AI practices, metacognition and the frequency of AI-tool use?

Table 1. Dimensions of the University Students' Productivity Scale (USPS), conceptual definitions and example items.

Dimension	Conceptual definition	Example item
A. Time Management	Planning, prioritizing and allocating study time with AI support	I use AI to prioritize tasks like homework and project deadlines.
B. Task Completion	Finishing assignments, essays and projects on time with AI support	I can complete my group projects faster with AI collaboration tools.
C. Goal Achievement	Meeting academic goals, deadlines and learning objectives with AI support	I can achieve high grades in exams with AI study aids.
D. Efficiency	Working faster and with less effort on research, analysis and writing	AI practices help me manage my study schedule more efficiently.
E. Quality of Work	Improving the accuracy, coherence and quality of academic output	I rely on AI to check grammar and coherence in my written work.
F. Focus and Concentration	Sustaining attention and reducing distractions during study	I maintain concentration during research work by using AI-based task managers.
G. Collaboration and Teamwork	Coordinating group work through AI-enabled collaboration platforms	I contribute actively to team discussions using AI collaboration tools like Zoom.
H. Stress Management	Using AI-based tools to relax, schedule breaks and avoid stress	I track my stress levels using AI tools like Mood fit during exam preparations.
I. Resource Utilization	Finding, organizing and managing academic resources and references	I make use of AI-based language tools like Grammarly to improve my written work.
J. Problem-Solving Ability	Solving academic, data-analytic, programming and engineering problems	AI-based platforms streamline my problem-solving approach, saving time on assignments.

Note. Each dimension comprises five items rated from 0% to 100% agreement in steps of 10%. The full item list is given in Appendix A.

METHOD

Participants and procedure

The population comprised full-time undergraduate students in public-sector universities of Punjab, Pakistan. A multi-stage random sampling design was used. Public universities were grouped into six categories (gender-free general, women's, medical, engineering, agricultural and vocational universities); two universities were randomly selected from each category, giving 12 universities; and students were then selected from each university by simple random sampling. The sampling plan targeted 1,320 students (110 per university); the analytic sample comprised 1,190 students with complete data on all USPS items. Table 2 describes the sample.

Table 2. Sample characteristics (N = 1,190).

Characteristic	Category	n	%
Gender	Male	478	40.2
	Female	712	59.8
Social background	Rural	836	70.3
	Urban	354	29.7
Discipline	Social Sciences	201	16.9
	Engineering	203	17.1
	Arts & Humanities	196	16.5
	Other Sciences	590	49.6

Note. Percentages are of the full sample. Other Sciences refers to the science-related fields not covered by the other three categories.

Data were collected with a structured questionnaire administered by the researcher, with a digital (Google Forms) option for students who preferred it. The questionnaire began with demographic items and then presented four self-developed scales in sequence: the AI Practices Scale (AIPS), the University Students' Metacognition Scale (USMS), the USPS and the AI Tools Usage Scale (AITUS). Participation was voluntary; participants were informed of the study's purpose and of their right to withdraw without penalty, and no identifying information was collected. Permission to approach students was obtained from the relevant university authorities through a letter from the research supervisor.

Instrument development and measures

The USPS items were written by the research team on the basis of the literature on academic productivity, time management and AI in education, in consultation with a subject expert. Ten dimensions were specified with five items each (Table 1). Face and content validity were reviewed by national and international experts, who received the scale by e-mail, and necessary improvements were made before piloting. A pilot study with 300 university students then examined internal consistency and structure (Cronbach's $\alpha = .821$; Kaiser–Meyer–Olkin = .887; principal component analysis with varimax rotation yielding ten components; confirmatory model fit CFI = .828, RMSEA = .027). The revised scale was administered to the main sample reported here.

Respondents indicated their level of agreement with each of the 50 statements as a percentage from 0% (no agreement) to 100% (complete agreement) in steps of 10%, giving an 11-point response scale (Preston & Colman, 2000). All items are positively worded. The USPS total score is the mean of the 50 items (range 0–100), and each dimension score is the mean of its five items. For validity analyses, the main sample also provided scores on three related measures: the AIPS (50 items, ten dimensions of AI-related study practices; $\alpha = .89$ in this sample), the USMS (50 items, ten dimensions of metacognition; $\alpha = .87$), and the AITUS, in which students rated how often they use each of 128 named AI tools on a five-point scale from 1 (never) to 5 (almost daily). The AIPS and USMS scores are item means on the same 0–100 metric, and the AITUS score is the mean frequency across tools.

Data analysis

All analyses used the full sample (N = 1,190) and were run in Python 3.12 (pandas, SciPy, stats models, factor analyzer and semopy). Item-level descriptive statistics included means, standard deviations, skewness, excess kurtosis, floor and ceiling percentages (responses of 0% and 100%) and corrected item–total correlations. Sampling adequacy was examined with the Kaiser–Meyer–Olkin (KMO) index (Kaiser, 1974) and Bartlett's test of sphericity. The number of factors was assessed by Horn's (1965) parallel analysis, comparing observed eigenvalues of the polychoric correlation matrix

(estimated by bivariate-normal maximum likelihood) with the 95th percentile of eigenvalues from 40 random permutations of the data, by Velicer's (1976) minimum average partial (MAP) test, and by inspection of the scree plot.

Four measurement models were compared by confirmatory factor analysis (CFA): M1, a single general factor; M2, ten correlated first-order factors corresponding to the intended dimensions; M3, a second-order model in which the ten first-order factors load on a general factor; and M4, a bifactor model with one general and ten orthogonal specific factors (Reise, 2012; Rodriguez et al., 2016). Models were estimated by maximum likelihood on item scores treated as continuous, which is acceptable for scales with more than five response categories and moderate skewness and kurtosis (Rhemtulla et al., 2012). Fit was judged by CFI and TLI (≥ .95), RMSEA (≤ .06) and SRMR (≤ .08) (Hu & Bentler, 1999), and models were compared with chi-square difference tests and χ^2-based AIC ($\chi^2 + 2q$) and BIC ($\chi^2 - df \cdot \ln N$). To check replication, all four models were also fitted separately to two random halves of the sample (n = 595 each).

Reliability was estimated as Cronbach's α (Cronbach, 1951) and McDonald's ω from the one-factor model (McDonald, 1999; Revelle & Zinbarg, 2009), with bootstrap 95% confidence intervals (1,000 resamples for α; 300 for ω), odd–even split-half reliability with Spearman–Brown correction, and the standard error of measurement. Dimension-level reliability, dimension intercorrelations and their disattenuated values were computed to evaluate the separability of the ten dimensions. Group differences by gender and social background were tested with Welch's t-test (Cohen's d; Cohen, 1988) and by discipline with one-way ANOVA (η^2), with Mann–Whitney and Kruskal–Wallis tests as robustness checks. Similarity of the factor structure across groups was assessed with Tucker's congruence coefficient computed from one-factor loadings estimated within each group (Lorenzo-Seva & ten Berge, 2006). Differential item functioning (DIF) was screened by regressing each item on the rest score and group membership, with a Bonferroni-adjusted threshold (α = .05/50 = .001). Formal multi-group invariance testing (Cheung & Rensvold, 2002; Putnick & Bornstein, 2016) was not undertaken.

Convergent and discriminant validity were examined through Pearson correlations (with Fisher-z confidence intervals) between the USPS and the AIPS, USMS and AITUS, correlations corrected for attenuation, and a heterotrait–monotrait ratio computed from item-level correlations (Henseler et al., 2015). A multiple regression of USPS scores on the AIPS, USMS, AITUS and demographic variables provided nomological evidence. As a diagnostic for method variance, the proportion of variance explained by the first unrotated factor (Harman's single-factor test) was calculated, recognising its limitations (Podsakoff et al., 2003).

RESULTS

Data screening and item analysis

There were no missing responses and no duplicate response patterns, and no respondent gave identical answers across items (the smallest within-person standard deviation was 20.5 points). Item means ranged from 52.3 (item A4) to 68.7 (item B3), and item standard deviations from 32.1 to 36.0 (Appendix A). Item distributions were negatively skewed (skewness −.74 to −.06) and flat (excess kurtosis −1.35 to −.67), with floor effects of 5.5–16.0% and ceiling effects of 18.4–36.3%; respondents therefore used the whole response range. Corrected item–total correlations ranged from .21 to .41 (mean = .30), all above the .20 benchmark, and removing any single item would have lowered α slightly (α if item deleted = .854–.858, against .858 for the full scale). All 50 items were therefore retained. The total score (M = 58.59, SD = 12.22, range 25.4–85.2) was approximately symmetric (skewness = −.19, excess kurtosis = −.88; Figure 1a). The 5th, 25th, 50th, 75th and 95th percentiles were 38.6, 48.6, 59.6, 68.5 and 76.6. Dimension means were nearly identical, from 58.0 to 59.2 (Table 3), and the distribution of responses was almost the same across dimensions (Figure 1b).

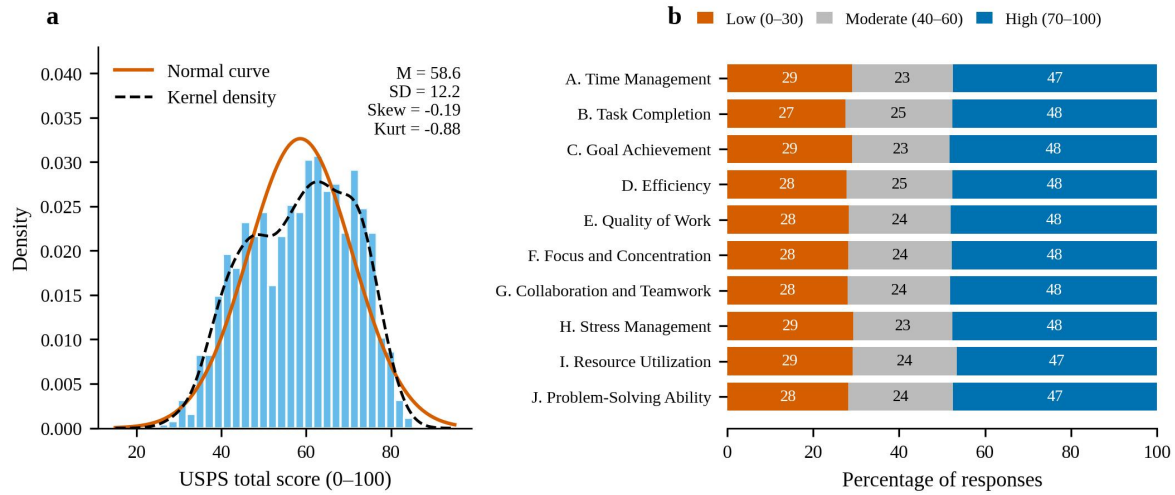

Figure 1: Distribution of USPS scores. (a) Histogram of total scores with normal and kernel-density curves (N = 1,190). (b) Percentage of responses in the low (0–30), moderate (40–60) and high (70–100) bands for each dimension.

Table 3. Descriptive statistics and reliability of the ten USPS dimension scores.

Dimension	M	SD	α	Mean inter-item r	Item–total r (range)	r with remaining 45 items
A. Time Management	58.1	19.0	.42	.13	.29–.41	.60
B. Task Completion	59.2	18.0	.35	.10	.23–.34	.51
C. Goal Achievement	58.9	18.4	.36	.10	.21–.39	.58
D. Efficiency	58.8	17.9	.33	.09	.24–.37	.57
E. Quality of Work	58.8	18.8	.42	.12	.29–.34	.56
F. Focus and Concentration	58.7	18.5	.39	.11	.29–.33	.57
G. Collaboration and Teamwork	58.9	17.9	.32	.09	.25–.37	.58
H. Stress Management	58.0	18.4	.36	.10	.27–.34	.59
I. Resource Utilization	58.0	18.3	.38	.11	.29–.32	.58
J. Problem-Solving Ability	58.5	18.3	.36	.10	.25–.33	.55

Note. Dimension scores are means of five items on a 0–100 metric. Item–total correlations are corrected for the item itself and computed against the 50-item total. The last column correlates the dimension score with the mean of the other 45 items.

Dimensionality

The correlation matrix was suitable for factor analysis (KMO = .91; Bartlett's $\chi^2(1,225) = 5,984.8$, $p < .001$), although inter-item correlations were modest (mean r = .11). The first eigenvalue was 6.41 from the Pearson matrix (12.8% of variance) and 6.99 from the polychoric matrix, compared with 1.41 and 1.44 for the second (Table 4; Figure 2a). Parallel analysis retained one factor: the first observed eigenvalue (6.99) far exceeded the 95th percentile of random eigenvalues (1.49), whereas the second (1.44) did not exceed its threshold (1.44). Velicer's MAP criterion was minimised at one factor (Figure 2b). The first ten Pearson components together accounted for only 34.8% of the variance, so no ten-component solution is supported by the full sample.

Table 4. Eigenvalues of the USPS item correlation matrices and parallel analysis.

Component	Eigenvalue (Pearson)	% variance (Pearson)	Eigenvalue (polychoric)	Parallel analysis, 95th percentile	Retained
1	6.41	12.8	6.99	1.49	Yes
2	1.41	2.8	1.44	1.44	No
3	1.35	2.7	1.38	1.41	No
4	1.27	2.5	1.30	1.38	No
5	1.23	2.5	1.25	1.35	No

Note. Parallel analysis used 40 random permutations of the item data and the polychoric correlation matrix. The Pearson-based first-factor percentage is also the Harman single-factor statistic.

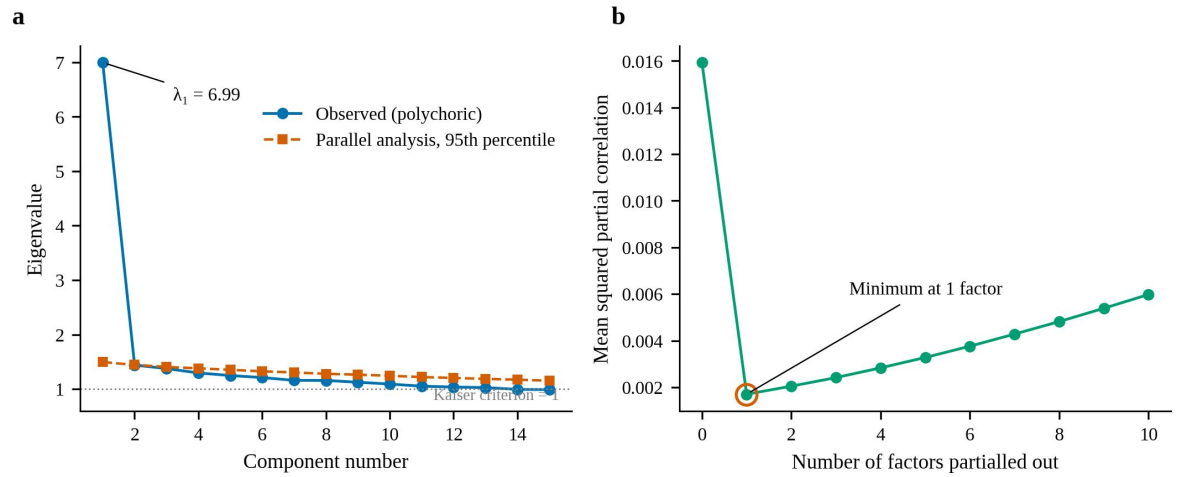

Figure 2: Dimensionality diagnostics for the 50 USPS items (N = 1,190). (a) Scree plot of polychoric eigenvalues with the 95th percentile of eigenvalues from permuted data. (b) Velicer's minimum average partial (MAP) test; the average squared partial correlation is smallest after one factor.

The correlation structure explains why. Items belonging to the same intended dimension were no more strongly correlated than items from different dimensions (mean polychoric r = .117 within dimensions and .121 between dimensions; Figure 3). Block means of within-dimension correlations ranged from .09 to .14, while block means of between-dimension correlations ranged from .10 to .14. In short, the data show one dominant source of shared variance and no dimension-specific clusters.

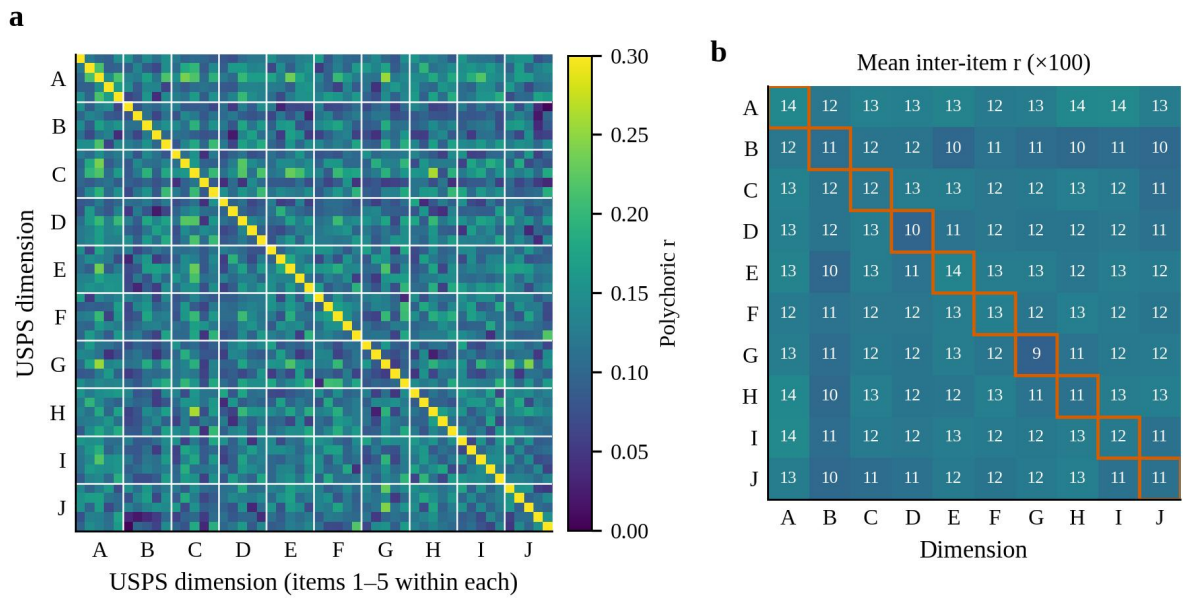

Figure 3: Inter-item correlations. (a) Polychoric correlation matrix of the 50 items ordered by intended dimension (A–J); white lines separate dimensions. (b) Mean inter-item correlation (×100) within dimensions (outlined diagonal) and between dimensions.

Competing measurement models and cross-validation

Table 5 reports the four competing models. The one-factor model fitted well ($\chi^2(1,175) = 1429.4$, CFI = .948, TLI = .945, RMSEA = .013, SRMR = .028). The more elaborate models improved fit only marginally. Relative to M2, the one-factor model's chi-square was higher by 61.6 on 45 degrees of freedom (p = .051); relative to the second-order model (M3) the difference was 3.0 on 10 degrees of freedom (p = .981); and relative to the bifactor model (M4) it was 63.7 on 50 degrees of freedom (p = .092). Information criteria favoured the one-factor model (lowest AIC and BIC; Table 5). In M2, the standardised correlations among the ten factors ranged from .84 to 1.22 (mean = 1.04); correlations above 1.0 are inadmissible and indicate that the factors cannot be separated. The bifactor solution was also improper: the standardised loadings on the specific factors were unstable and ranged from −.73 to .91, with several negative, so its general-versus-specific variance indices were not interpreted (Rodriguez et al., 2016).

Table 5. Fit of competing confirmatory factor models in the full sample and in two random halves.

Model	Sample	χ^2	df	CFI	TLI	RMSEA	SRMR	AIC	BIC
M1 One-factor	Full (N = 1,190)	1429.4	1175	.948	.945	.013	.028	1629	-6,892
M2 Ten correlated factors	Full (N = 1,190)	1367.8	1130	.951	.947	.013	.027	1658	-6,635
M3 Second-order	Full (N = 1,190)	1426.3	1165	.946	.943	.014	.028	1646	-6,824
M4 Bifactor	Full (N = 1,190)	1365.6	1125	.950	.946	.013	.027	1666	-6,601
M1	Half A (n = 595)	1322.9	1175	.940	.938	.015	.037	—	—

M2	Half A (n = 595)	1271.3	1130	.943	.938	.015	.036	—	—
M3	Half A (n = 595)	1321.0	1165	.937	.934	.015	.037	—	—
M4	Half A (n = 595)	1255.0	1125	.947	.943	.014	.036	—	—
M1	Half B (n = 595)	1334.9	1175	.935	.932	.015	.037	—	—
M2	Half B (n = 595)	1288.8	1130	.935	.930	.015	.037	—	—
M3	Half B (n = 595)	1332.0	1165	.932	.928	.016	.037	—	—
M4	Half B (n = 595)	1266.8	1125	.942	.937	.015	.036	—	—

Note. M1 = one general factor; M2 = ten correlated first-order factors; M3 = second-order model; M4 = bifactor model (improper solution). AIC = $\chi^2 + 2q$ and BIC = $\chi^2 - df \times \ln(N)$, where q is the number of free parameters; lower values indicate better fit. Information criteria are reported for the full sample only. Maximum likelihood estimation; items treated as continuous.

The structure replicated across independent halves. In Half A the one-factor model gave CFI = .940, RMSEA = .015 and SRMR = .037; in Half B it gave CFI = .935, RMSEA = .015 and SRMR = .037. CFI values were slightly below .95 in the halves, as expected with smaller samples, while RMSEA and SRMR were well within accepted limits, and the ordering of the models was the same as in the full sample (Figure 5a–c). All freely estimated loadings in M1 were statistically significant (smallest z = 6.0, p < .001). Standardised loadings ranged from .22 to .45 (mean = .33; Figure 4, Appendix A); no item fell below .20, and the items with the weakest loadings were C4 (.22), B1 (.25) and D1 (.26), and the strongest were A3 (.45), C3 (.43) and D3 (.41).

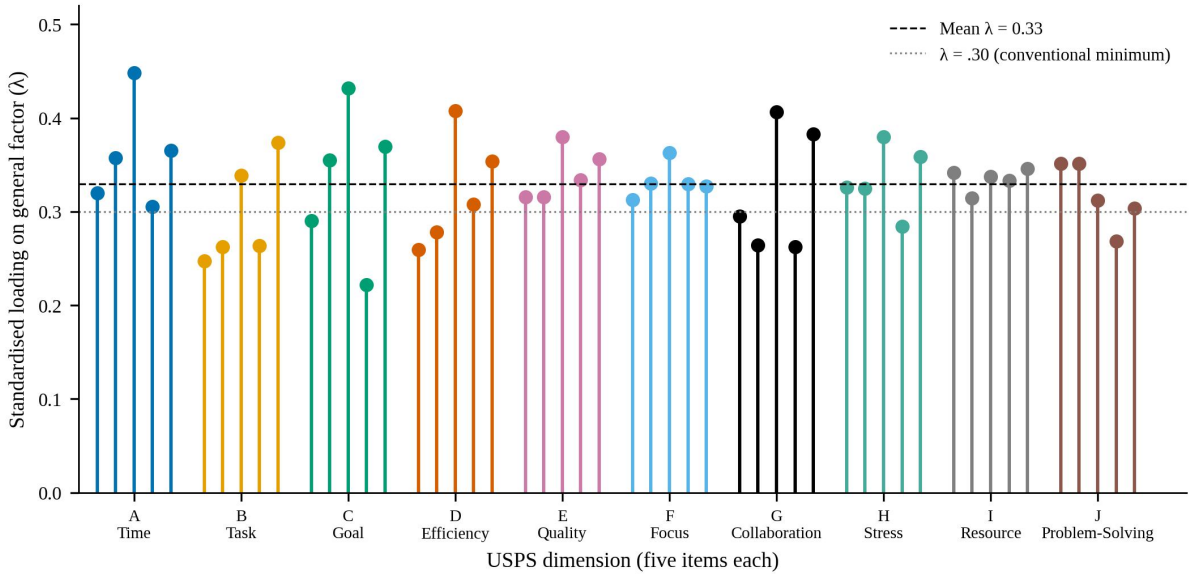

Figure 4: Standardized loadings of the 50 items on a single general factor (Model M1). Colours denote the ten intended dimensions (A–J).

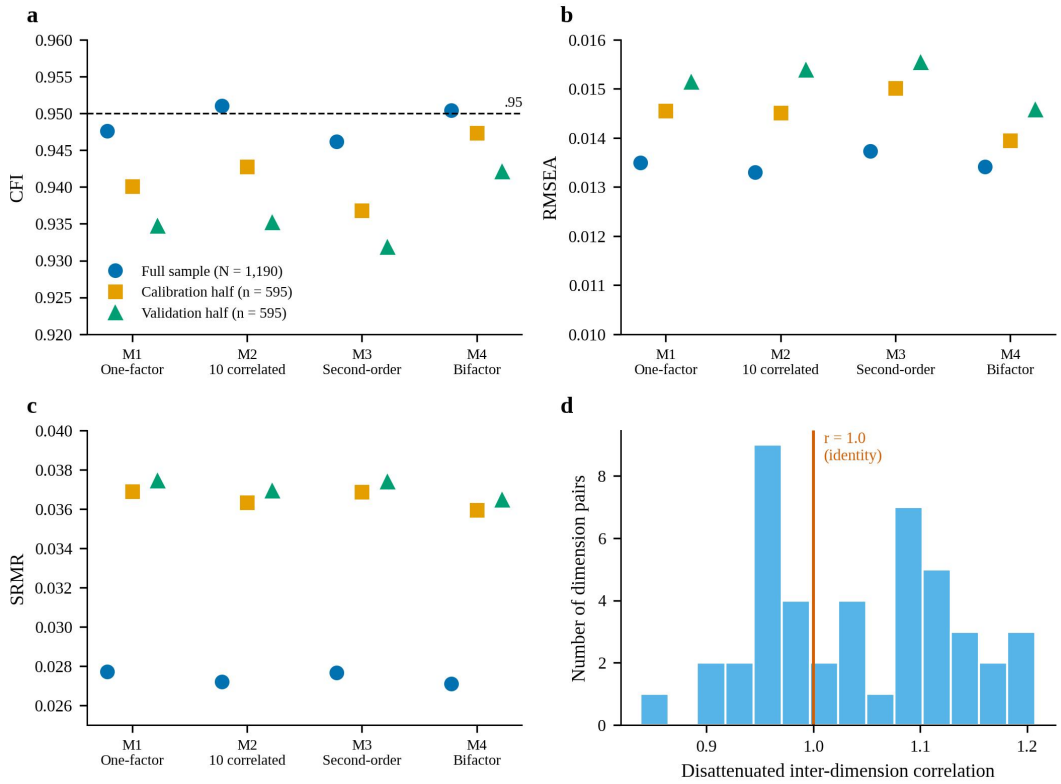

Figure 5: Comparison of competing measurement models and separability of the dimensions. (a–c) CFI, RMSEA and SRMR for models M1–M4 in the full sample and in two random halves (axes are truncated; the conventional RMSEA ≤ .06 and SRMR ≤ .08 limits lie outside the plotted range). (d) Distribution of the 45 disattenuated correlations among the ten dimension scores; values cluster around 1.0.

Reliability

Table 6. Reliability estimates for the USPS total score (50 items).

Index	Estimate	95% CI
Cronbach's α	.858	.849–.868
McDonald's ω (one-factor model)	.859	.848–.868
Odd–even split-half r	.765	—
Spearman–Brown corrected split-half	.867	—
Mean inter-item correlation	.109	—
Mean squared standardised loading (λ^2)	.111	—
Standard error of measurement (0–100 metric)	4.60	—

Note. Confidence intervals are percentile bootstrap intervals (1,000 resamples for α; 300 for ω).

The total score was reliable (α = .858, 95% CI .849–.868; ω = .859, 95% CI .848–.868; Spearman–Brown split-half = .867; Table 6). This is higher than the pilot value (α = .821; n = 300). The standard error of measurement was 4.60 points, so an individual's 95% confidence band is about ±9.0 points on the 0–100 metric. Reliability was achieved largely through scale length: the mean inter-item correlation was only .11 and the average item shared about .11 of its variance with the general factor.

Dimension scores, by contrast, were unreliable (α = .32–.42; Table 3). Raw correlations among the ten dimension scores ranged from .32 to .43; after correction for attenuation they averaged 1.04 (range .84–1.21; Figure 5d), meaning that the dimensions are statistically indistinguishable from one another once measurement error is taken into account. Each dimension score correlated .51–.60 with the mean of the other 45 items (Table 3).

Group differences, structural similarity and differential item functioning

Mean USPS scores were virtually identical for men and women (M = 58.17 vs 58.86; Welch's $t(1032.7) = -.96$, $p = .339$, $d = -.06$) and for rural and urban students (M = 58.36 vs 59.12; $t(674.6) = -.99$, $p = .325$, $d = -.06$; Table 7, Figure 6). Differences between disciplines were small and not statistically significant ($F(3, 1{,}186) = 2.32$, $p = .074$, $\eta^2 = .006$), with the highest mean among Social Sciences students (60.56) and the lowest among Other Sciences students (58.00); no pairwise Tukey comparison was significant (smallest adjusted $p = .050$). Non-parametric tests agreed (Mann–Whitney $p = .326$ for gender and .351 for background; Kruskal–Wallis $p = .099$), and there was no gender × background interaction ($F(1, 1{,}183) = 0.09$, $p = .769$).

Table 7. USPS total scores, reliability and factor similarity by gender, social background and discipline.

Group	n	M	SD	α	Tucker's φ (vs. total sample)
Male	478	58.17	12.12	.856	.996
Female	712	58.86	12.29	.860	.998
Rural	836	58.36	12.28	.859	.998
Urban	354	59.12	12.09	.856	.991
Social Sciences	201	60.56	11.43	.839	.981
Engineering	203	58.14	12.80	.869	.991
Arts & Humanities	196	58.79	11.96	.852	.988
Other Sciences	590	58.00	12.33	.861	.998
Total sample	1190	58.59	12.22	.858	—

Note. Tucker's φ compares each group's one-factor loading vector with that of the full sample; values ≥ .95 indicate that the factors can be considered equal (Lorenzo-Seva & ten Berge, 2006). Pairwise φ was .990 (male–female) and .982 (rural–urban); across the six discipline pairs it ranged from .960 to .984.

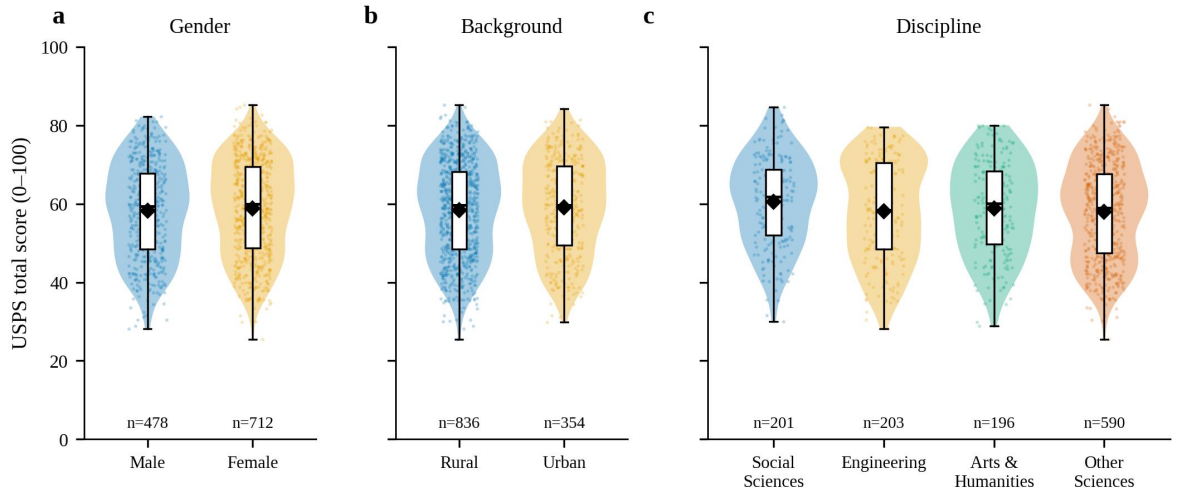

Figure 6: USPS total scores by gender, social background and discipline. Violins show the density of scores, boxes the interquartile range with median, diamonds the mean, and dots individual respondents.

Reliability was similar across groups (α = .839–.869), and the one-factor loading patterns were highly congruent with the full-sample pattern (Tucker's φ = .981–.998; Table 7). DIF screening flagged no item for gender, social background or discipline at the Bonferroni-adjusted threshold. The largest rest-score-adjusted differences were .14 SD between men and women and .16 SD between rural and urban students, and the largest discipline effect accounted for .006 of item variance beyond the rest score. Although formal multi-group invariance models were not fitted, these results indicate no evidence of non-equivalence.

Convergent, discriminant and nomological validity

The USPS correlated strongly with the AI Practices Scale (r = .87, 95% CI .85–.88) and the Metacognition Scale (r = .85, 95% CI .83–.86), supporting convergent validity (Table 8; Figure 7a–b). Correlations of the individual dimension scores with the AIPS (.54–.60) and USMS (.51–.59) were uniformly moderate to strong (Figure 7d). By contrast, the USPS was unrelated to the mean frequency of AI-tool use (r = .00, 95% CI −.06 to .06, p = .982; Figure 7c).

Table 8. Correlations among USPS, AIPS, USMS and AITUS scores, with disattenuated correlations and heterotrait–monotrait ratios.

Scale	1	2	3	4	α
1. USPS	—				.86
2. AIPS	.87	—			.89
3. USMS	.85	.87	—		.87
4. AITUS (tool-use frequency)	.00	.01	.02	—	—

Note. N = 1,190. Pearson correlations; those among USPS, AIPS and USMS are significant at p < .001, and correlations with AITUS are not significant (all p ≥ .43). Correlations corrected for attenuation: USPS–AIPS = .99; USPS–USMS = .98. Heterotrait–monotrait ratios: USPS–AIPS = .99; USPS–USMS = .98; AIPS–USMS = .99 (values above .90 indicate inadequate discriminant validity).

Discriminant evidence was lacking. After correction for attenuation, the USPS–AIPS and USPS–USMS correlations were .99 and .98, and the heterotrait–monotrait ratios (.99 and .98) far exceeded the usual .85–.90 ceiling (Henseler et al., 2015). The three scales were equally difficult to separate

from each other (AIPS–USMS ratio = .99). In a regression with all three related measures and demographic covariates (Table 9), the model explained 78.8% of USPS variance ($F(8, 1,181) = 547.4$, $p < .001$); the AIPS ($\beta = .54$) and USMS ($\beta = .37$) were independent positive predictors, the variance inflation factors were about 4.2 for both, and tool-use frequency, gender, background and discipline were not significant. In a bivariate model, AI practices alone explained 75.5% of USPS variance ($\beta = .87$). Harman's single-factor statistic for the USPS alone was 12.8%, but it does not rule out shared variance across scales completed in the same session in the same format.

Table 9. Multiple regression of USPS scores on related measures and demographic variables.

Predictor	β	SE	95% CI	p
AI Practices Scale (AIPS)	.54	.03	.49 to .60	< .001
Metacognition Scale (USMS)	.37	.03	.32 to .43	< .001
AI-tool use frequency (AITUS)	−.01	.01	−.04 to .02	.420
Gender (male vs. female)	−.02	.03	−.08 to .03	.367
Background (urban vs. rural)	.01	.03	−.05 to .07	.782
Discipline: Engineering	−.05	.05	−.14 to .04	.258
Discipline: Other Sciences	−.00	.04	−.08 to .07	.937
Discipline: Social Sciences	.01	.05	−.08 to .10	.816

Note. Continuous variables are standardised; the reference categories are female, rural and Arts & Humanities. $R^2 = .788$, adjusted $R^2 = .786$. Because the three scales are strongly collinear and were completed together, these coefficients describe associations among self-reports and should not be read as effects.

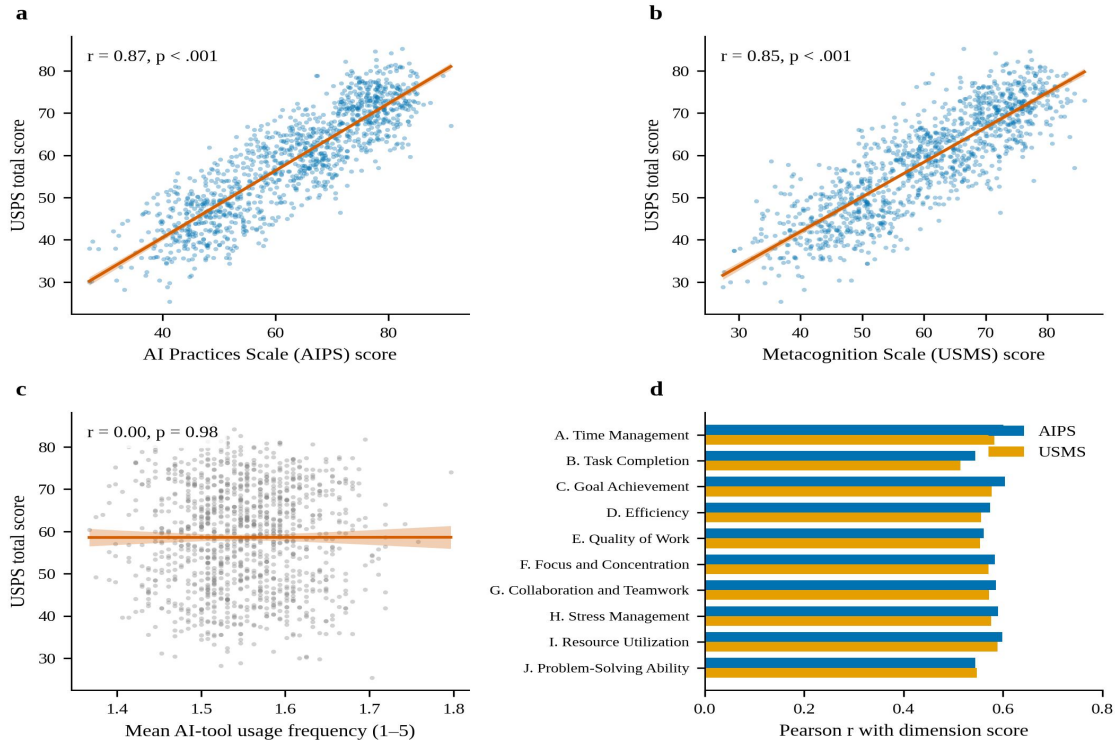

Figure 7: Convergent and nomological evidence. (a–c) USPS total score against AIPS, USMS and AI-tool use frequency, with least-squares lines and 95% confidence bands. (d) Correlations of the ten USPS dimension scores with the AIPS and the USMS.

DISCUSSION

This study aimed to create and examine the productiveness scale in study contexts supported by AI and to examine the scale in a large sample to see if its ten hypothesized dimensions are empirically separate. Three conclusions emerge. First, the 50 items are coherent, reliable and replicable; one general factor is a good fit in the full sample and in both halves, there is a good reliability (α = .86; ω = .86), the items are reasonably spread out across the various response categories, and there is no need to remove any item. Second, that coherence is not dimensional: results from parallel analysis, the MAP test, the correlation structure, the improper solutions of the multi-factor models, and the near unit disattenuated correlations among dimension scores suggest that the ten dimensions are not separable. Third, USPS is nearly identical to the scales of AI practices and metacognition that were also administered in the same session; thus its discriminant validity has not been established.

The former statement is practical. Total scores can be used in group-level research with reasonable confidence, were similar by gender, social background and discipline and no item had meaningful differential functioning. In addition, the lack of differences among groups is informative for a scale that has items referring to discipline-specific activities (e.g., coding, lab reports and engineering projects): students from all disciplines answered the items in a manner similar to the rest of the scale. However, the confidence interval for an individual score is 95% of ±9 points for the score, which is not too large to make comparisons between groups, but is too large to make decisions about an individual.

The second conclusion has implications for the scoring and reporting of the scale. The ten dimensions were selected from a coherent literature, and were read sensibly (Table 1), but they share the same stem structure (AI helps me …) and a uniformly favourable direction, and they have only modest unique information: the average item shares about 11% of its variance with the general factor, and the dimension-specific variance is negligible. Therefore, it is not appropriate to compare the Dimension scores separately in studies utilizing the USPS (α = .32 - .42) because each Dimension is not valid for comparison with one another. This is different from what was reported in the pilot stage—the 10-component solution. In the pilot, a principal component analysis produced ten components with very high loadings (above .98) and accounting for nearly all of the variance. Arithmetically, such values are incompatible with the modest inter-item correlations found here, with the first ten of them accounting for 35% of the variance. The pilot components are therefore considered an artefact of this analysis and of the item groupings, and the ten dimensions are thus not considered as empirical dimensions, but are recommended as a content-coverage framework only.

There are no simple fixes for short forms. The length-dependency of the reliability is a function of length, as individual items are weak indicators (mean standardised loading of about .33), a 20-item version would have an expected α of about .71 and a 10-item version about .55 (Spearman–Brown). Items should instead be made stronger, such as by eliminating brand nomenclature (Calm, Headspace, Moodfit, Mendeley, Slack and Trello) that rely on the name of the tools students know, by minimising double-barreled language, by changing the direction of the items, and by writing items that draw on the dimensions specific experience rather than a generic "AI helps me" answer.

The third conclusion must lead to a revision of the use of the scale in substantive research. The USPS correlated .87 with the AIPS and .85 with the USMS, and the corrected correlations and heterotrait–monotrait ratios were at or near 1.0. There are two possible explanations, which cannot be distinguished on the basis of the data. One is a substantive, namely that students who use AI extensively in their learning processes could actually perceive themselves as becoming more metacognitive and more productive. The latter is methodological as it is in all three scales the respondents, in all scales the same occasion, in all scales the same format (percentage agreement), and a common AI-benefit framing and therefore a general favorable response tendency (Podsakoff et al., 2003, 2012). The actual rate of tool use did not correlate with USPS scores (r = .00), and the rate itself

was strongly skewed towards "never" use for most of the 128 tools, so it does not have the same level of sensitivity. Whatever the reason is, the regression or structural coefficients relating the USPS scores to the practices of AI (e.g., $\beta = .87$ in a bivariate model) should be viewed as associations between self-perceptions, not as estimates of the effect of AI on productivity. This warning is important since the literature provides experimental evidence that there is a difference between perceived and actual gains (Bastani et al., 2025; Fan et al., 2025).

There are some caveats to be considered. This scale is self-report based and has not been used in a cross-sectional way, and test–retest reliability or sensitivity to change is unknown. There were no objective criteria available (grades, time on task or independently rated work quality), so there is no criterion validity data. Items were treated as continuous in the maximum likelihood CFA; despite the moderate skewness (Rhemtulla et al., 2012) the analysis of item functioning in an ordinal estimator's model and item response models could provide a more complete picture of item functioning, especially as ceiling effects (18–36% of responses at 100%) were obvious. No formal multi-group invariance tests were performed to arrive at group equivalence; however, congruence coefficients and DIF screening were used. Claims of invariance are therefore provisional and subject to further research. The respondents are from a public-sector university in one province of Pakistan and it is therefore essential that the findings should be generalized to private universities in other provinces, post graduate students and other countries. Lastly, since all items mention AI, comparisons cannot be made between students who use AI and those who do not, or for isolating the impact of using AI and students' productivity.

Future research should involve (a) a rewording of the item pool and stronger dimension specific content; (b) test–retest stability and measurement invariance with ordinal multi-group models; (c) addition of behavioral or institutional criteria and discriminant comparators collected at different times or from different sources; and (d) an investigation of whether a parsimonious general score, when supplemented by conceptually distinct scales for specific domains (e.g., AI-assisted collaboration or stress regulation), are better suited for research and practice.

CONCLUSION

In a sample of 1,190 undergraduates from public-sector universities in Punjab, Pakistan, the University Students' Productivity Scale behaves as a reliable, replicable and group-equivalent measure of a single general construct, perceived AI-supported academic productivity. Its ten conceptual dimensions are useful for content coverage but not as separately scorable subscales, and its strong overlap with related AI-practice and metacognition scales calls for caution when inferring effects among them. Researchers can use the total score for group-level studies, while revision of the item pool and independent criterion evidence are needed before the scale is used to draw conclusions about distinct dimensions or about the effects of AI on productivity.

DECLARATIONS

Ethics approval and consent to participate. Participation was voluntary and informed; permission to approach students was obtained from the relevant university authorities, and no identifying information was collected.

Data availability. The anonymised item-level dataset (N = 1,190) is available from the corresponding author on reasonable request.

REFERENCES

Aeon, B., Faber, A., & Panaccio, A. (2021). Does time management work? A meta-analysis. *PLOS ONE, 16*(1), e0245066.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. American Educational Research Association.

Bandura, A. (1997). *Self-efficacy: The exercise of control*. W. H. Freeman.

Bastani, H., Bastani, O., Sungu, A., Ge, H., Kabakcı, Ö., & Mariman, R. (2025). Generative AI without guardrails can harm learning: Evidence from high school mathematics. *Proceedings of the National Academy of Sciences, 122*(26), e2422633122.

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

Britton, B. K., & Tesser, A. (1991). Effects of time-management practices on college grades. *Journal of Educational Psychology, 83*(3), 405–410.

Castillo-Martínez, I. M., Flores-Bueno, D., Gómez-Puente, S. M., & Vite-León, V. O. (2024). AI in higher education: A systematic literature review. *Frontiers in Education, 9*, 1391485.

Chauke, T. A., Mkhize, T. R., Methi, L., & Dlamini, N. (2024). Postgraduate students' perceptions on the benefits associated with artificial intelligence tools on academic success: In case of ChatGPT AI tool. *Journal of Curriculum Studies Research, 6*(1), 44–59.

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

Clark, L. A., & Watson, D. (2019). Constructing validity: New developments in creating objective measuring instruments. *Psychological Assessment, 31*(12), 1412–1427.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum.

Credé, M., & Kuncel, N. R. (2008). Study habits, skills, and attitudes: The third pillar supporting collegiate academic performance. *Perspectives on Psychological Science, 3*(6), 425–453.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*(3), 297–334.

DeVellis, R. F., & Thorpe, C. T. (2022). *Scale development: Theory and applications* (5th ed.). SAGE.

Fan, Y., Tang, L., Le, H., Shen, K., Tan, S., Zhao, Y., Shen, Y., Li, X., & Gašević, D. (2025). Beware of metacognitive laziness: Effects of generative artificial intelligence on learning motivation, processes, and performance. *British Journal of Educational Technology, 56*(2), 489–530.

Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research, 18*(1), 39–50.

Hayes, A. F., & Coutts, J. J. (2020). Use omega rather than Cronbach's alpha for estimating reliability. But… *Communication Methods and Measures, 14*(1), 1–24.

Henseler, J., 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.

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika, 30*(2), 179–185.

Hu, L., & 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.

Inam-ul-Haque, M., & Farooq, M. S. (2025). Beyond collegiality: Development and validation of the University Teachers' Competitiveness Inventory (UTCI). *ACADEMIA International Journal for Social Sciences, 4*(6), 181-194. https://doi.org/10.63056/academia.4.6.2025.2588

Inam-ul-Haque, M., & Farooq, M. S. (2026). Development and Psychometric Validation of University Teachers' Trust Scale (UTTS): A Large-Sample Confirmatory Study. *Research Journal for Social Affairs, 4*(3), 1195-1202. https://doi.org/10.71317/RJSA.004.03.1216

Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika, 39*(1), 31–36.

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., … Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences, 103*, 102274.

Li, E. Y., & Jan, A. (2023). Impact of artificial intelligence (AI) in enhancing productivity and reducing stress among students. In *ICEB 2023 Proceedings*. AIS Electronic Library.

Lorenzo-Seva, U., & ten Berge, J. M. F. (2006). Tucker's congruence coefficient as a meaningful index of factor similarity. *Methodology, 2*(2), 57–64.

Macan, T. H., Shahani, C., Dipboye, R. L., & Phillips, A. P. (1990). College students' time management: Correlations with academic performance and stress. *Journal of Educational Psychology, 82*(4), 760–768.

McDonald, R. P. (1999). *Test theory: A unified treatment*. Lawrence Erlbaum.

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist, 50*(9), 741–749.

Noy, S., & Zhang, W. (2023). Experimental evidence on the productivity effects of generative artificial intelligence. *Science, 381*(6654), 187–192.

Pintrich, P. R., Smith, D. A. F., Garcia, T., & McKeachie, W. J. (1993). Reliability and predictive validity of the Motivated Strategies for Learning Questionnaire (MSLQ). *Educational and Psychological Measurement, 53*(3), 801–813.

Podsakoff, P. M., MacKenzie, S. B., & Podsakoff, N. P. (2012). Sources of method bias in social science research and recommendations on how to control it. *Annual Review of Psychology, 63*, 539–569.

Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology, 88*(5), 879–903.

Preston, C. C., & Colman, A. M. (2000). Optimal number of response categories in rating scales: Reliability, validity, discriminating power, and respondent preferences. *Acta Psychologica, 104*(1), 1–15.

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

Reise, S. P. (2012). The rediscovery of bifactor measurement models. *Multivariate Behavioral Research, 47*(5), 667–696.

Revelle, W., & Zinbarg, R. E. (2009). Coefficients alpha, beta, omega, and the glb: Comments on Sijtsma. *Psychometrika, 74*(1), 145–154.

Rhemtulla, M., Brosseau-Liard, P. É., & Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychological Methods, 17*(3), 354–373.

Rodriguez, A., Reise, S. P., & Haviland, M. G. (2016). Evaluating bifactor models: Calculating and interpreting statistical indices. *Psychological Methods, 21*(2), 137–150.

Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika, 74*(1), 107–120.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12*(2), 257–285.

Uzir, N. A., Gašević, D., Matcha, W., Jovanović, J., & Pardo, A. (2020). Analytics of time management strategies in a flipped classroom. *Journal of Computer Assisted Learning, 36*(1), 70–88.

Velicer, W. F. (1976). Determining the number of components from the matrix of partial correlations. *Psychometrika, 41*(3), 321–327.

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education – where are the educators? *International Journal of Educational Technology in Higher Education, 16*, 39.

Zhai, C., Wibowo, S., & Li, L. D. (2024). The effects of over-reliance on AI dialogue systems on students' cognitive abilities: A systematic review. *Smart Learning Environments, 11*, 28.

Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. *Theory Into Practice, 41*(2), 64–70.

APPENDIX A. USPS ITEMS AND ITEM-LEVEL STATISTICS

Table A1. Item wording, descriptive statistics, corrected item–total correlation (rit) and standardised loading (λ) on the general factor (N = 1,190).

ID	Item	M	SD	Skew	Kurt	r it	λ
A1	I effectively manage my study time using AI scheduling tools.	54.8	34.3	−.16	−1.25	.30	.32
A2	AI helps me allocate time for assignments and exams efficiently.	55.6	35.0	−.20	−1.30	.33	.36
A3	I use AI to prioritize tasks like homework and project deadlines.	68.3	33.1	−.71	−.76	.41	.45
A4	AI time management tools help me plan for daily study sessions.	52.3	35.3	−.06	−1.33	.29	.31
A5	AI-driven schedules enable me to complete tasks within allocated timeframes.	59.5	35.0	−.38	−1.19	.33	.37
B1	I complete assignments on time using AI tools for research.	55.8	34.8	−.21	−1.26	.23	.25
B2	AI helps me finish essays and projects more efficiently.	55.7	34.5	−.20	−1.24	.25	.26
B3	I can complete my group projects faster with AI collaboration tools.	68.7	32.7	−.70	−.80	.31	.34
B4	AI assists me in finishing my thesis research in less time.	53.3	34.5	−.14	−1.27	.24	.26
B5	I finish my homework faster with AI-based time management tools.	62.2	34.0	−.46	−1.07	.34	.37
C1	I achieve my academic goals using AI tools for research projects.	56.6	34.7	−.24	−1.28	.27	.29
C2	AI helps me meet assignment deadlines more effectively.	54.3	35.4	−.15	−1.34	.33	.36
C3	I can achieve high grades in exams with AI study aids.	67.3	33.2	−.65	−.87	.39	.43
C4	AI-based tools support me in meeting course project deadlines.	55.6	35.3	−.17	−1.34	.21	.22
C5	AI tools support my progress toward achieving my learning objectives.	60.4	34.8	−.38	−1.19	.34	.37
D1	I efficiently complete research papers with the help of AI tools.	54.8	34.6	−.15	−1.25	.24	.26
D2	AI-assisted methods help me work faster on assignments and projects.	56.5	34.6	−.20	−1.26	.26	.28
D3	AI practices help me manage my study schedule more efficiently.	68.4	32.8	−.74	−.67	.37	.41
D4	I utilize AI to complete data analysis faster than before.	54.1	35.1	−.16	−1.29	.29	.31
D5	AI improves my ability to finish essays on time.	60.2	34.1	−.36	−1.14	.33	.35
E1	I use AI tools to enhance the quality of my assignments and essays.	58.0	34.7	−.29	−1.23	.30	.32
E2	AI helps me produce high-quality research papers with thorough analysis.	54.7	34.8	−.15	−1.30	.29	.32
E3	I rely on AI to check grammar and	66.3	33.5	−.63	−.89	.34	.38

	coherence in my written work.						
E4	AI-powered tools help me improve the accuracy of my lab reports.	54.0	35.0	−.12	−1.30	.31	.33
E5	I use AI to refine presentations and make them more impactful.	61.0	34.0	−.38	−1.15	.33	.36
F1	I use AI tools to stay focused during long study sessions for exams.	54.6	35.0	−.14	−1.31	.29	.31
F2	AI reminders help me stay focused while working on assignments or projects.	55.8	34.3	−.23	−1.22	.31	.33
F3	I maintain concentration during research work by using AI-based task managers.	67.5	32.9	−.67	−.81	.33	.36
F4	AI tools help me reduce distractions during online lectures and study sessions.	54.7	34.7	−.15	−1.27	.31	.33
F5	I stay focused on my writing tasks with AI-driven productivity apps.	60.9	34.5	−.41	−1.15	.30	.33
G1	I use AI tools like Microsoft Teams to collaborate effectively on group projects.	58.9	33.4	−.30	−1.15	.28	.29
G2	AI-driven platforms, such as Slack, help me coordinate tasks in group assignments.	55.5	34.7	−.19	−1.27	.25	.26
G3	I contribute actively to team discussions using AI collaboration tools like Zoom.	65.6	34.1	−.59	−.97	.37	.41
G4	AI tools support real-time collaboration during virtual group meetings for assignments.	54.3	35.8	−.16	−1.35	.25	.26
G5	AI-powered tools like Trello enhance my collaboration with peers on joint tasks.	60.0	34.6	−.38	−1.16	.36	.38
H1	I use AI tools like Calm to manage stress during intense study sessions.	54.5	36.0	−.18	−1.34	.30	.33
H2	AI-based apps like Headspace help me relax during study breaks.	53.4	35.2	−.13	−1.31	.31	.32
H3	I track my stress levels using AI tools like Moodfit during exam preparations.	67.6	33.7	−.70	−.82	.34	.38
H4	AI-driven reminders in my calendar help me schedule breaks, reducing stress.	53.7	34.6	−.14	−1.28	.27	.28
H5	AI tools help me stay organized, preventing stress during busy project periods.	60.9	34.8	−.42	−1.15	.33	.36
I1	I use AI tools like Zotero to manage research references effectively.	54.8	35.2	−.15	−1.32	.32	.34
I2	I utilize AI-powered search engines like Google Scholar to find academic resources.	54.1	34.7	−.14	−1.27	.29	.31
I3	I make use of AI-based language tools like Grammarly to improve my written work.	67.2	32.1	−.62	−.84	.31	.34
I4	AI tools such as Mendeley help me organize academic resources efficiently.	53.3	34.9	−.11	−1.30	.31	.33

I5	I utilize AI-driven time management apps like Trello to plan my study sessions.	60.9	34.0	−.40	−1.13	.32	.35
J1	I use AI tools like ChatGPT to solve academic problems faster and more efficiently.	56.9	34.3	−.23	−1.24	.33	.35
J2	AI-powered systems help me analyze complex data, increasing my productivity in research tasks.	54.2	34.5	−.15	−1.26	.33	.35
J3	AI-based platforms streamline my problem-solving approach, saving time on assignments.	67.3	33.4	−.69	−.80	.29	.31
J4	AI algorithms assist me in resolving programming issues quickly, improving my coding productivity.	55.0	34.8	−.20	−1.26	.25	.27
J5	I use AI-driven solutions to simplify problem-solving in engineering projects, enhancing productivity.	59.2	35.1	−.34	−1.24	.28	.30

Note. Items are rated 0–100% agreement. Item IDs: letter = dimension (A = Time Management, B = Task Completion, C = Goal Achievement, D = Efficiency, E = Quality of Work, F = Focus and Concentration, G = Collaboration and Teamwork, H = Stress Management, I = Resource Utilization, J = Problem-Solving Ability); number = item within dimension. Skew = skewness; Kurt = excess kurtosis.