Creative Education, 2026, 17(11), 2096-2123
https://www.scirp.org/journal/ce
ISSN Online: 2151-4771
ISSN Print: 2151-4755

Dynamic-Static Integrated Assessment of Learning Dynamics in Regular Undergraduate and Top-Up Degree Students

Mo Shi*, Lujie Wang, Xun Liang

School of Economics and Management, Ankang University, Ankang, China
Email: *shimo0204@outlook.jp

How to cite this paper: Shi, M., Wang, L. J., & Liang, X. (2026). Dynamic-Static Integrated Assessment of Learning Dynamics in Regular Undergraduate and Top-Up Degree Students. *Creative Education, 17,* 2096-2123.
https://doi.org/10.4236/ce.2026.1711117

Received: September 12, 2026
Accepted: November 7, 2026
Published: November 10, 2026

Copyright © 2026 by author(s) and Scientific Research Publishing Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
http://creativecommons.org/licenses/by/4.0/
Open Access

Abstract

Conventional higher education assessment systems rely predominantly on terminal examinations, which obscure distributed learning processes and individual heterogeneity. To address this limitation, this study proposes and empirically validates a dynamic-static integrated assessment framework combining longitudinal self-evaluation, chapter-level formative homework, and terminal final examinations across regular undergraduate (n = 38) and top-up degree (n = 83) cohorts with an identical course in the School of Economics and Management, Ankang University. Longitudinal self-evaluation revealed positive linear growth in attentional engagement and knowledge acquisition, yet demonstrated that the top-up cohort exhibited 103.62% - 130.21% larger score dispersion despite comparable mean levels. A matching distributional pattern emerged in formative homework, showing minimal mean accuracy differences (79.94 vs. 80.88 for single-choice) but substantially wider within-group variability. Terminal examination outcomes exhibited pronounced target and format dependence, with the largest disparity concentrated in Target 1 multiple-choice items (33.10% relative difference), whereas overall normalized cumulative knowledge-absorption indices converged closely (79.78 pts·p vs. 77.29 pts·p). Basically, the findings under dynamic-static assessment in this study underscore the diagnostic necessity of triangulating perceptions with summative achievement to capture hidden student heterogeneity and inform targeted pedagogical interventions.

Keywords

Higher Education, Dynamic-Static Assessment, Triangulating Perceptions, Self-Evaluation, Formative Homework, Terminal Examination

1. Introduction

In recent years, higher education has increasingly emphasized continuous, process-oriented learning over traditional, outcome-centric paradigms. However, many previous studies note that university teaching evaluation has historically relied heavily on summative, end-of-semester examinations as the primary metric of academic achievement (Boud & Falchikov, 2006; Bulut et al., 2025; Cartwright, 2025; French et al., 2024; Gratchev et al., 2024; Montano & Sobrejuanite, 2025). As Dumas et al. (Dumas et al., 2020) indicate, while such static assessments provide a standardized endpoint for evaluating terminal proficiency, they inherently compress temporally distributed learning processes into a single, retrospective observation. Consequently, instructors are afforded limited opportunities to identify localized learning fluctuations, monitor student engagement, or implement timely pedagogical interventions during the instructional cycle. To address this limitation, recent advancements in learning analytics and educational evaluation advocate for dynamic assessment architectures that capture the longitudinal trajectory of knowledge acquisition (Bonte & Brem, 2024; Diao et al., 2026; Khosla et al., 2026; Tong & Ren, 2025; Ueno et al., 2025).

As Wondim & Dessie (2024) emphasize, despite the theoretical consensus on the pedagogical value of formative assessment, empirical implementations frequently struggle to systematically integrate continuous process data with static terminal outcomes. Furthermore, contemporary university classrooms are characterized by growing demographic heterogeneity, notably with the integration of regular undergraduate students and top-up degree (transfer) students (Ayoubi et al., 2025; Cheung et al., 2022; Ivins et al., 2017; Marks & Tuytens, 2026). These cohorts often possess distinct educational backgrounds, knowledge reserves, and learning habits. However, existing systems that predominantly rely on mean cohort scores frequently obscure within-group dispersion and localized fluctuations that differentiate these populations. Based on the reality noted above, there remains a critical research gap regarding the development of a comprehensive assessment framework capable of capturing both the temporal evolution of learning and the multidimensional nature of student heterogeneity without conflating subjective perception with objective performance.

To bridge this methodological gap, this study proposes a novel dynamic-static integrated assessment framework that systematically triangulates learning processes and outcomes. As shown in **Figure 1**, including longitudinal self-evaluation, chapter-level formative homework, and a target-specific summative final examination, the architecture consists of three complementary layers. The self-evaluation module captures the temporal evolution of students' subjective attentional engagement, perceived knowledge acquisition, and recognition of curricular significance. Concurrently, the chapter-level homework provides repeated objective measurements of short-term performance. Moreover, to enhance the analytical resolution of terminal achievement, this study introduces the concept of a cumulative knowledge-absorption index. Adapted from the mathematical principles of

hysteresis energy dissipation in structural engineering (Shi & Choi, 2026; Taborda et al., 2016; Villalobos Luna et al., 2026; Wang et al., 2024; Yang et al., 2026), this cross-disciplinary index integrates score magnitude with population distribution, providing a more robust quantification of cohort-level learning accumulation than conventional central tendency metrics.

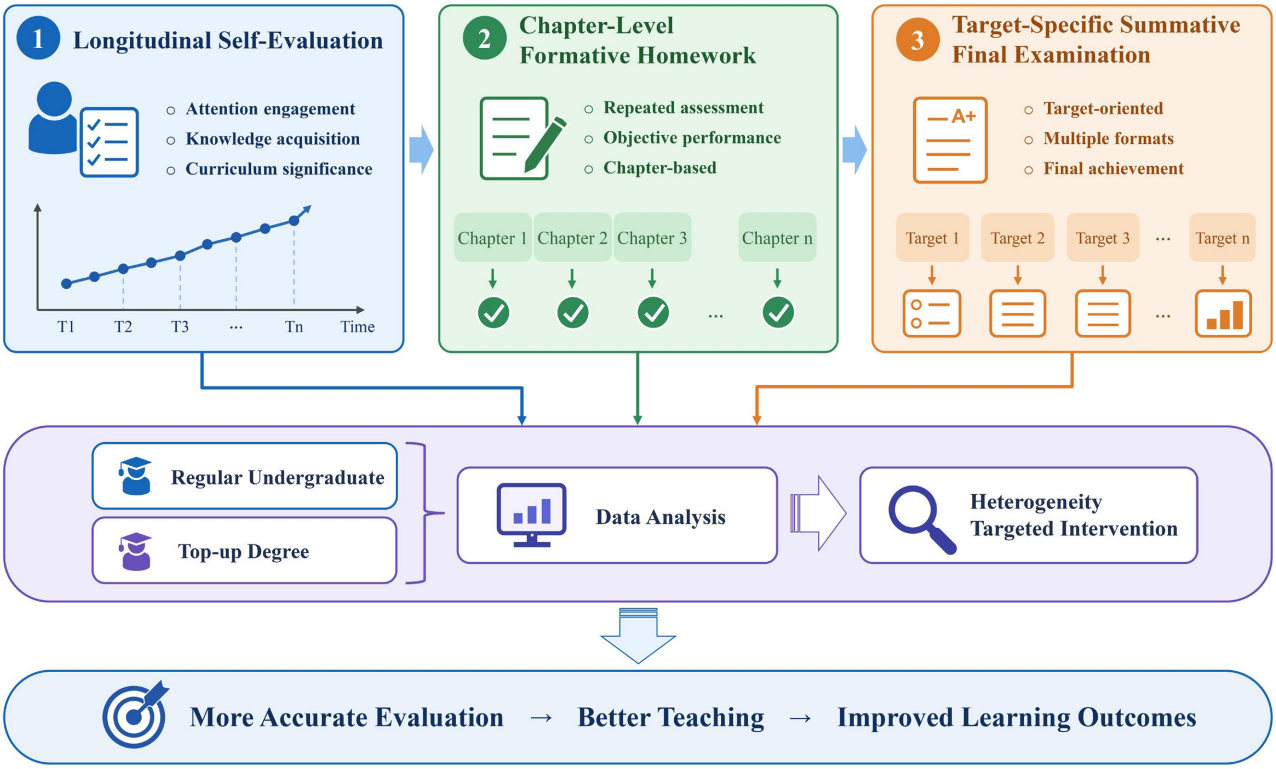


Figure 1. Structure of the dynamic-static integrated assessment.

The integrated framework in this study was empirically applied and evaluated during an Engineering Safety and Environmental Protection course at Ankang University, encompassing 121 students divided into a regular undergraduate cohort and a top-up degree cohort. Both cohorts were subjected to identical instructional conditions, materials, and instructor oversight across the semester. By employing the dual-cohort observational design, this study seeks to characterize the temporal and distributional learning patterns inherent to different educational pathways. Given the naturalistic setting of the study, the analytical approach is deliberately descriptive and comparative rather than strictly causal, aiming to demonstrate the diagnostic utility of the framework in a realistic instructional environment.

Based on the analytical results in this study, preliminary application of the framework reveals that relying solely on terminal average scores is insufficient for characterizing comprehensive educational performance. While both cohorts exhibited parallel upward trajectories in perceived engagement and demonstrated comparable mean formative performance, the dynamic process data revealed that

the top-up degree cohort possessed substantially greater within-group heterogeneity. Furthermore, terminal achievement proved to be highly dependent on assessment targets and specific response formats, underscoring the inadequacy of single-metric evaluations. By transforming dispersed learning-process observations and static examination outcomes into an analytically coherent structure, this dynamic-static framework provides a high-resolution evidence base for targeted instructional optimization and course redesign.

2. Materials and Methods

2.1. Study Participants and Instructional Context

The empirical investigation in this study was conducted during the second semester of the 2025-2026 academic year (spanning from March to July 2026) at Ankang University, China, utilizing the undergraduate core course Engineering Safety and Environmental Protection. To capture comprehensive learning dynamics within an actual educational environment, this study examined two naturally occurring student cohorts enrolled in the same curriculum, including 38 regular undergraduate students from the 2023 grade of Engineering Management and 83 students enrolled in the 2025 grade of Engineering Management within the top-up degree course. Both cohorts received instruction from the same faculty member, utilized identical course syllabi and lecture materials, and studied from the same prescribed core textbook, thereby maintaining a highly controlled and consistent instructional context for comparative evaluation.

Despite sharing a common instructional environment, the two cohorts represent distinct educational pathways and program structures. As illustrated in Table 1, the regular undergraduate cohort entered higher education through the national college entrance examination and followed a conventional four-year undergraduate curriculum, whereas the top-up degree cohort entered through a specialized articulation pathway following prior higher vocational education, completing their undergraduate training within a condensed two-year timeframe. In accordance with rigorous educational research standards, the institutional differences in this study are categorized strictly as variations in educational pathways and curriculum pacing rather than direct measures of individual intellectual capacity or baseline academic ability. Due to the absence of pre-experimental baseline equivalence testing, as prior GPAs or pre-admission academic performance records were not independently controlled, the observed differences in score dispersion and performance profiles between the regular undergraduate and top-up degree cohorts cannot be interpreted as direct causal consequences of the two-year educational pathway alone because pre-existing variations in foundational knowledge reserves and academic preparedness may also contribute. Given the naturalistic observational constraints above, the empirical findings provide descriptive and comparative evidence of dynamic learning dynamics within a common instructional setting, rather than definitive causal attribution.

Table 1. Comparative analysis of educational profiles and curriculum structures.

Comparative Dimension	Regular Undergraduate (UC)	Top-Up Degree (TC)
Target Admission Group	High school graduates (National College Entrance Exam) (Gaokao)	Vocational college students (Articulation Selection Exam) (Zhuanshengben)
Educational Duration	4 Academic Years	2 Academic Years
Instruction & Train	Four-year continuous education	Two-year continuous education
Certificate Type	Standard Undergraduate Diploma (Unannotated)	Undergraduate Diploma (Annotated)
Degree Awarded	Bachelor of Science (B. S.)	Bachelor of Science (B. S.)
Nature of Education	Full-time Regular Higher Education	Full-time Regular Higher Education
Primary Academic Degree	Bachelor's Degree	Bachelor's Degree
Foundational Coursework	Four-year continuous curriculum (Full foundational training)	Two-year condensed curriculum (Abbreviated foundational training)

The instructional content of the course was structured around a comprehensive 16-chapter textbook covering two major thematic domains. As shown in Table 2, fundamental management frameworks, regulatory requirements, risk control mechanisms, emergency response protocols, accident investigation procedures, environmental protection, and occupational health for Part 1; and specialized construction-site safety technologies, including deep foundation pit operations, primary structural construction, scaffolding, work-at-height, temporary electrical systems, and fire safety for Part 2. For the actual instructional cycle of the semester, two scheduled chapters (Chapter 15 & Chapter 16) were omitted during the semester due to unforeseen academic calendar interruptions caused by university-level institutional activities and the May Day national holiday, while the empirical analyses were systematically restricted to the first 14 chapters of the textbook. This boundary was applied uniformly to both cohorts, ensuring that all longitudinal process observations and summative assessments were derived from an identical set of instructional units.

Table 2. Thematic structure and chapter distribution of the course curriculum.

Parts	Chapter	Thematic Contents & Instructional Focus
Part 1	Chapter 1	Safety Production Management in Construction Engineering
	Chapter 2	Laws and Regulations on Construction Safety and Environmental Protection

Continued

	Chapter 3	Operational and Institutional Policies for Construction Safety Management
	Chapter 4	Management of High-Risk Operations & Construction Site Safety Inspections
	Chapter 5	Hazard Source Control, Identification, and Emergency Response Protocols
	Chapter 6	Investigation, Handling, and Reporting Procedures of Construction Accidents
	Chapter 7	Civilized Construction Practices and Environmental Protection Measures
	Chapter 8	Occupational Health, Safety Management, and Disease Prevention
Part 2	Chapter 9	Safety Technology for Deep Foundation Pit Engineering
	Chapter 10	Safety Technology for Primary Structural Construction
	Chapter 11	Safety Technology for Scaffolding Operations
	Chapter 12	Safety Technology for Work-at-Height Operations
	Chapter 13	Safety Technology for Temporary Electrical Systems on Construction Sites
	Chapter 14	Safety Technology for Construction Site Fire Prevention
	Chapter 15*	Safety Technology for Hoisting and Lifting Operations
	Chapter 16*	Safety Technology for Construction Machinery and Equipment

Note: *represents the two scheduled chapters not covered during the instructional period.

2.2. Assessment Framework and Statistical Analysis

To move beyond traditional terminal evaluations, a tripartite assessment architecture was established to capture student performance across both continuous process and terminal outcome dimensions, comprising repeated longitudinal self-evaluation, chapter-level formative homework assessments, and a target-specific summative final examination. Referring to Pinkosova et al. (2023), the self-evaluation component accompanied each instructional chapter. It measured three distinct psychological and cognitive dimensions, including attentional engagement (perceived ability to maintain cognitive focus), perceived knowledge acquisition (self-assessed comprehension and assimilation of course content), and curricular significance (perceived relevance and importance of the material). To mitigate potential self-reporting bias and measurement error, the longitudinal self-evaluation instruments underwent psychometric validation before deployment. Following Zhoc et al. (2019) and Chanaka Ushantha & Samantha Kumara (2016), internal consistency reliability, assessed via Cronbach's alpha, yielded high coefficients ($\alpha = 0.87$ for attentional engagement, $\alpha = 0.89$ for knowledge acquisition, and $\alpha = 0.85$ for curricular significance). Furthermore, Confirmatory Factor Analysis (CFA) (Ayaneh et al., 2021; Kostiv et al., 2025; Mohammad & Kamran, 2023;

Zhang et al., 2021) confirmed that the three sub-constructs exhibited satisfactory discriminant validity, verifying that they capture statistically distinct dimensions of student learning perceptions. Concurrently, chapter-level formative homework served as an objective performance indicator, consisting of three single-choice and two true-or-false questions per chapter (Brassil & Couch, 2019; Couch et al., 2015; Enders et al., 2021; Freeman et al., 2011; Johnson & Kiviniemi, 2009). The final examination served as the summative endpoint, structured around three predefined curriculum targets: Target 1 (foundational knowledge and basic management competence, evaluated via single-choice, true-or-false, and multiple-choice items); Target 2 (hazard identification and safety measure formulation, evaluated via short-answer questions); and Target 3 (theoretical foundations and core engineering principles, evaluated via essay questions).

To quantify the observed learning patterns, group-level descriptive statistics and longitudinal trend estimations were integrated (Abd Manap & Mohd Zain, 2026; Arja et al., 2018; Avila-Andino et al., 2026; Garzón et al., 2026; Henriquez-Machado et al., 2026; Jones & Oh, 2024; Na et al., 2026; Peluso et al., 2026; Saparbayeva et al., 2025; Shi & Tang, 2026). For any given cohort g, instructional unit j, and evaluation dimension k, the cohort mean score was defined as:

$$\bar{X}_{gjk} = \frac{1}{n_g} \cdot \sum_{i=1}^{n_g} X_{igjk} \tag{1}$$

where n_g denotes the total student population within cohort g. Inter-cohort disparities were evaluated using a relative difference index (D), calculated as:

$$D = \left(\frac{\bar{X}_{TC} - \bar{X}_{UC}}{\bar{X}_{UC}} \right) \times 100\% \tag{2}$$

where UC and TC represent the regular undergraduate and top-up degree cohorts, respectively. Response dispersion and within-group heterogeneity were characterized using the sample standard deviation (SD_g):

$$SD_g = \sqrt{\frac{\sum_{i=1}^{n_g} \left(X_{ig} - \bar{X}_g \right)^2}{n_g - 1}} \tag{3}$$

Furthermore, temporal trajectories across the semester were estimated using a linear regression model:

$$Y_t = \beta_0 + \beta_1 t + \varepsilon_t \tag{4}$$

where t signifies the instructional sequence order, β_1 denotes the estimated rate of change (slope), and R^2 represents the coefficient of determination, interpreted strictly as a goodness-of-fit metric indicating the proportion of variance explained by the linear temporal trend.

For formative homework evaluation, chapter-level accuracy was computed for each item format as:

$$A = \frac{N_{\text{correct}}}{N_{\text{attempted}}} \times 100\% \tag{5}$$

This structure allowed the analytical model to decouple central tendency from score dispersion, addressing situations where cohorts exhibited comparable mean accuracy alongside markedly different variability. For summative evaluation, the study introduces a cross-disciplinary quantitative metric termed the cumulative knowledge-absorption index (K_A) (Shi & Choi, 2026; Taborda et al., 2016; Villalobos Luna et al., 2026; Wang et al., 2024; Yang et al., 2026). Following the calculative details in Teruna et al. (2015), adapted mathematically from hysteresis energy dissipation principles in structural engineering, where mechanical force and displacement are analogously reflected to examination scores and ordered student population coordinates, this index integrates score magnitude with student distribution. In its continuous form, cumulative knowledge absorption is expressed as:

$$K_A = \int_0^{n_{\max}} S(n)\,\mathrm{d}n \tag{6}$$

where $S(n)$ denotes the score profile along the ordered population coordinate n, and $n_{\max}$ represents the cohort size. Because educational data are inherently discrete, K_A was computed numerically using the trapezoidal approximation rule:

$$K_A \approx \sum_{i=1}^{N-1} \frac{S_i + S_{i+1}}{2} \cdot \left(n_{i+1} - n_i\right) \tag{7}$$

where S_i and S_{i+1} represent scores at adjacent ordered student positions n_i and n_{i+1}. This formulation corrects potential transcription errors in interval width notation, ensuring absolute mathematical reproducibility. The resulting index is measured in point-person units (pts·p) and functions as a population-score accumulation measure rather than a physical energy quantity.

3. Discussion

3.1. Self-Evaluation

According to Ren et al. (2026), self-evaluation provides a vital process-oriented perspective that diverges fundamentally from terminal achievement testing by capturing students' real-time perceptions of engagement and assimilation while instruction is actively underway. Within the proposed framework in this study, the longitudinal dimension was operationalized across three distinct constructs: attentional engagement, which reflects students' self-assessed ability to maintain cognitive focus during lectures; perceived knowledge acquisition, which measures their subjective assimilation of instructional content; and curricular significance, which evaluates their recognition of the relative importance of individual learning units. This tripartite architecture allows for the continuous tracking of learning-related perceptions across the semester rather than reducing cognitive progression to a single, retrospective end-of-course score.

The longitudinal trajectories of attentional engagement in **Figure 2(a)** reveal a clear positive temporal growth trend across the 14 assessed chapters for both student cohorts. The linear regression equations for the attentional engagement

yielded slopes of $T_{UC} = 0.0058x + 0.8263$ ($R^2 = 0.6645$) for the regular undergraduate cohort and $T_{TC} = 0.0076x + 0.8154$ ($R^2 = 0.9117$) for the top-up degree cohort, indicating that the top-up degree students experienced an approximately 31.0% steeper rate of increase in self-rated attentional focus over time. Furthermore, the substantially higher coefficient of determination (R^2) for the top-up cohort shows that linear regression more closely approximated its longitudinal progression. As emphasized in the methodological parameters, this higher R^2 value reflects a better fit to the linear trend rather than intrinsic data stability or lower fluctuation.

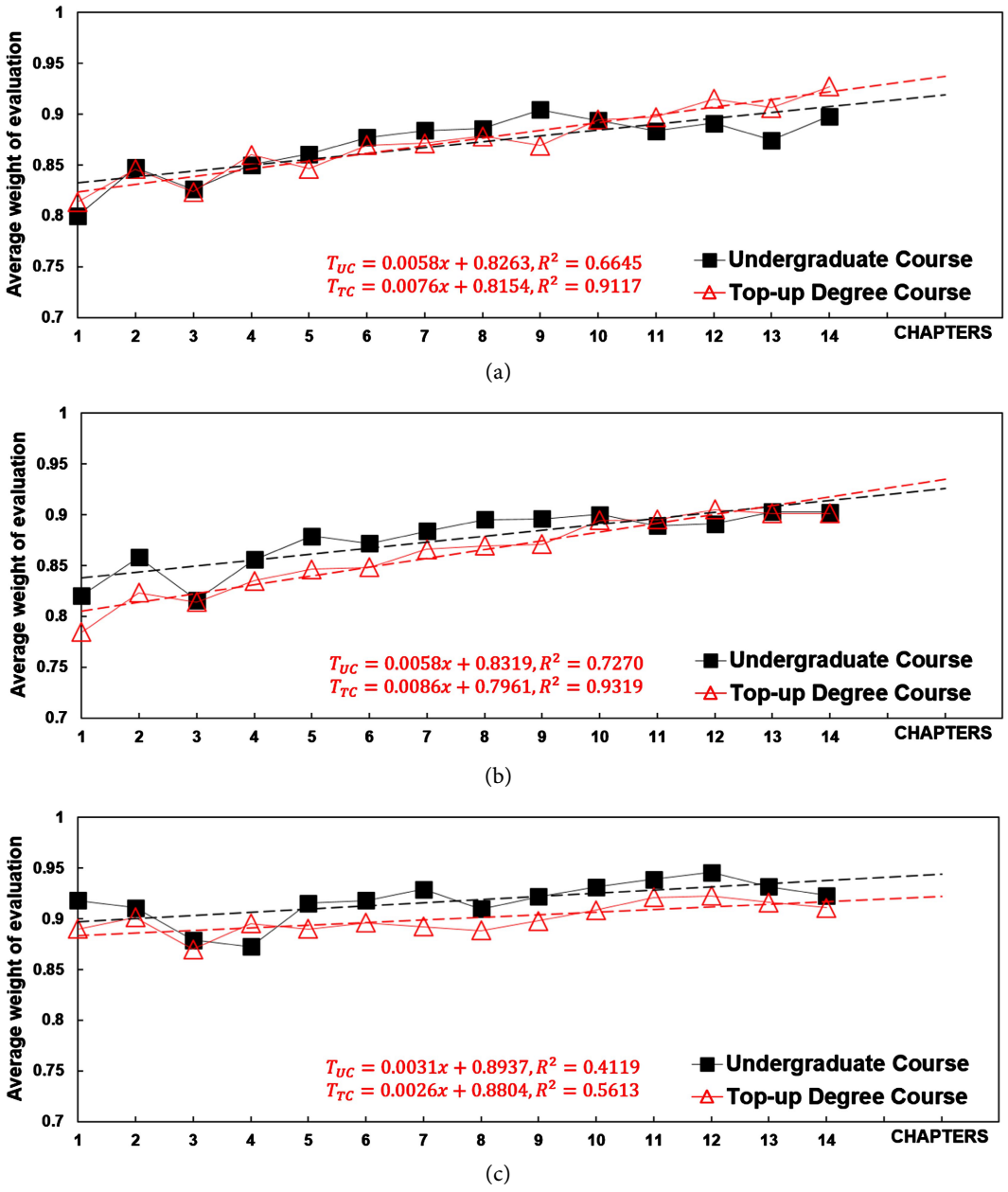


Figure 2. Dynamic evaluation across a full semester. (a) Attentional engagement, (b) Knowledge acquisition, (c) Curricular significance.

As depicted in Figure 2(b), a parallel longitudinal pattern emerged for perceived knowledge acquisition, accompanied by an even more pronounced inter-cohort contrast in growth velocity. The regular undergraduate cohort exhibited a fitted slope of 0.0058 ($R^2 = 0.7270$), whereas the top-up degree cohort recorded a steeper slope of 0.0086 ($R^2 = 0.9319$), representing an estimated 48.3% increase in slope magnitude. The close alignment in initial intercept values (0.8319 for regular undergraduates versus 0.7961 for top-up students) suggests that both cohorts commenced the semester with comparable baseline perceptions of knowledge assimilation, with the steeper trajectory of the top-up cohort leading to a gradual divergence in later instructional units.

On the other hand, Figure 2(c) illustrates a qualitatively distinct yet positively oriented temporal trend for curricular significance. The estimated linear regression equations yielded slopes of $T_{UC} = 0.0031x + 0.8937$ ($R^2 = 0.4119$) for regular undergraduates and $T_{TC} = 0.0026x + 0.8804$ ($R^2 = 0.5613$) for top-up degree students. The values of the linear regression analysis demonstrate more gradual temporal growth, with the top-up slope being roughly 16.1% lower than that of the undergraduate cohort, alongside lower coefficients of determination, which indicates that students' judgments regarding the broader importance of individual course chapters evolve at a slower pace and are less constrained by a simple linear function compared to immediate engagement and comprehension metrics.

Synthesizing the findings from Figure 2, the three self-evaluation dimensions underscore that learning perceptions do not operate as interchangeable manifestations of a single latent construct. Particularly among top-up degree students, while attentional engagement and perceived knowledge acquisition exhibited strong, steady linear upward trends, curricular significance remained relatively stable and high throughout the semester. Upon discussion of Figure 2, the multi-dimensional dissociation in this study validates the design choice to decouple process perceptions into specialized sub-constructs, ensuring that nuanced shifts in student psychology are not prematurely collapsed into an ambiguous aggregate score.

Transitioning from longitudinal trends to aggregate descriptive metrics, Table 3 demonstrates that the primary distinction between the two student cohorts lies in score dispersion rather than central tendency. The semester-level mean evaluations for attentional engagement were highly comparable at 86.96 for regular undergraduates and 87.26 for top-up degree cohorts, reflecting an absolute difference of merely 0.30 percentage points (a relative difference rate of 0.34%). Similarly, mean knowledge-acquisition scores stood at 87.57 versus 86.10 (1.72% relative difference), while curricular-significance scores reached 91.73 versus 89.99 (1.94% relative difference). Across all three dimensions in Table 3, inter-cohort discrepancies in central tendency remained negligible, ranging between 0.30 and 1.94 percentage points.

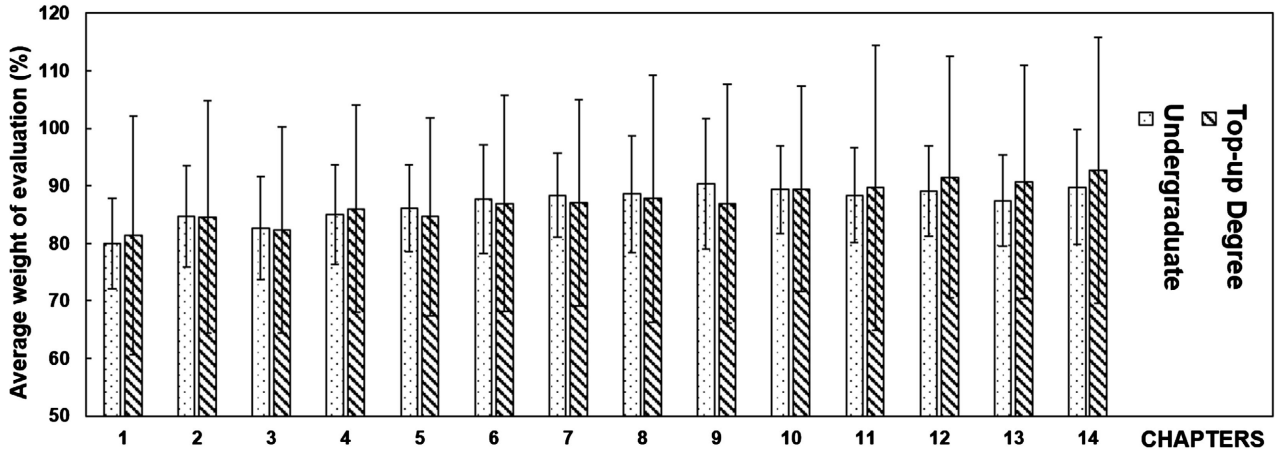
Despite these closely aligned mean scores, Table 3 reveals stark contrasts in score dispersion, as quantified by sample standard deviations. For attentional en-

gagement, the standard deviation surged from 8.69 in the regular undergraduate cohort to 19.99 in the top-up degree cohort, representing a 130.21% relative increase. Comparable expansions in dispersion were observed for perceived knowledge acquisition (from 9.42 to 19.19, a 103.62% increase) and curricular significance (from 9.73 to 20.66, a 112.35% increase). Expressed via coefficients of variation, relative dispersion levels for the top-up degree cohort (approximately 22.3% to 23.0%) were double those of the regular undergraduate cohort (approximately 10.0% to 10.8%), confirming that top-up students exhibited substantially greater internal response heterogeneity.

Table 3. Comparison analysis of average dynamic evaluation.

		Attentional Engagement	Knowledge Acquisition	Curricular Significance
Average Evaluation	Undergraduate	86.96	87.57	91.73
	Top-up Degree	87.26	86.10	89.99
	Rate	0.34%	1.72%	1.94%
Standard Deviation	Undergraduate	8.69	9.42	9.73
	Top-up Degree	19.99	19.19	20.66
	Rate	130.21%	103.62%	112.35%

The chapter-level distributional patterns illustrated in Figure 3 corroborate these aggregate dispersion statistics across the entire 14-chapter sequence. For every assessed instructional chapter, regular undergraduate students consistently displayed narrower error bars, whereas the top-up degree cohort exhibited markedly wider confidence intervals and standard deviation bounds across all three self-evaluation dimensions. The visual evidence in Figure 3 confirms that the heightened variance observed in Table 3 is a persistent, chapter-level phenomenon rather than an artifact of semester-level aggregation.



(a)

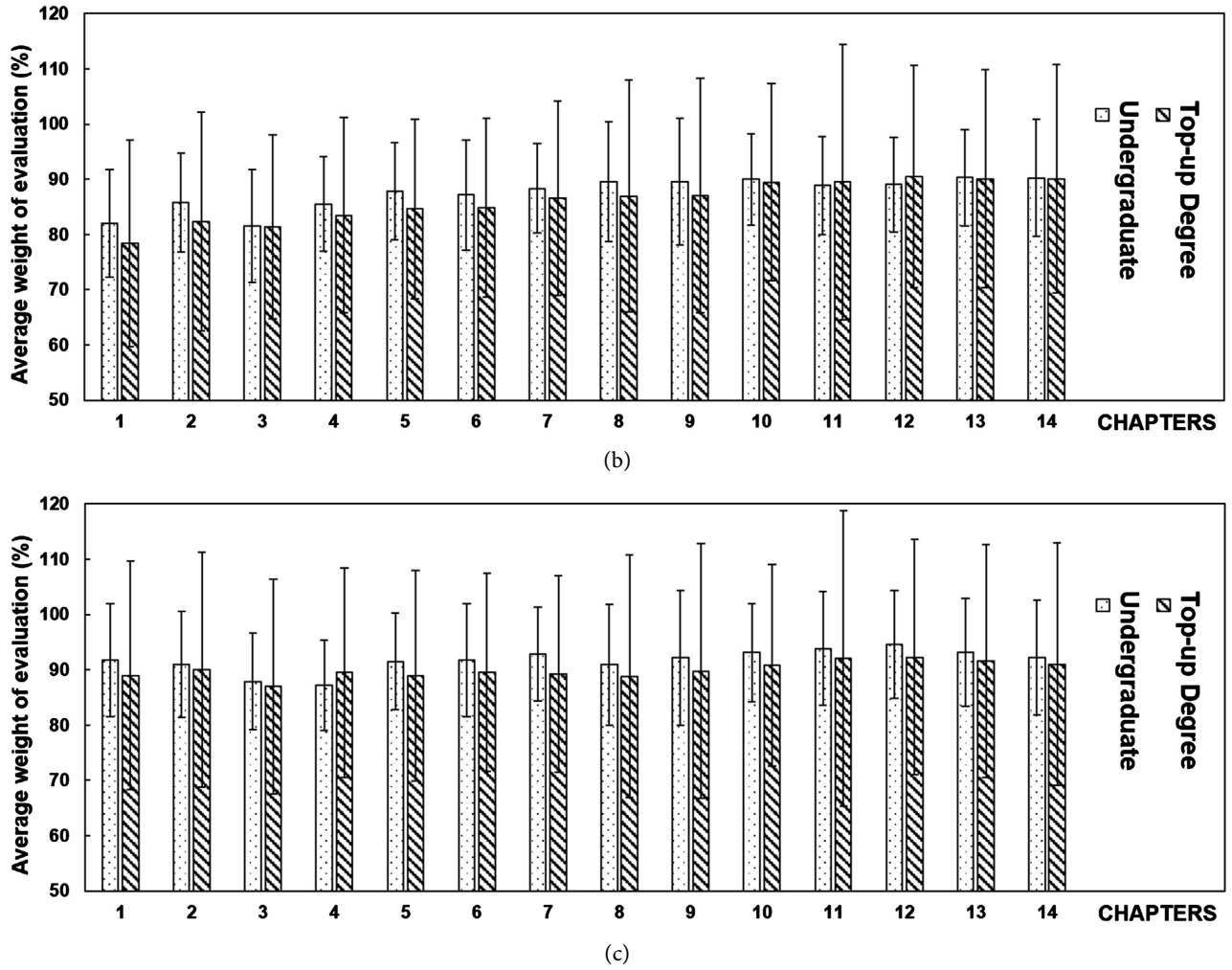


Figure 3. Quantitative analysis of the dynamic evaluation. (a) Attentional Engagement, (b) Knowledge Acquisition, (c) Curricular Significance.

More significantly, these pronounced differences in standard deviation must be interpreted with methodological caution from an analytical standpoint. As emphasized by Ost et al. (2017) and Soffer-Vital & Glassner (2026), a larger standard deviation cannot be construed as direct evidence of inferior learning capability, deficient knowledge reserves, or lower educational quality in the top-up degree cohort. Standard deviation merely measures the dispersion of observed scores around the cohort mean without identifying the underlying causal mechanism. Due to this, the data support the conservative and objective conclusion that the top-up degree cohort demonstrated greater within-group heterogeneity in subjective learning states, whereas the regular undergraduate cohort exhibited more homogeneous behavioral and psychological responses under identical instructional conditions.

For the summarization, the self-evaluation analyses establish the initial tier of evidence within the dynamic-static assessment framework, revealing that longitudinal process tracking can uncover critical distributional variations that remain

entirely obscured by terminal mean scores. The findings in this study demonstrate that while both cohorts shared parallel upward trajectories in perceived engagement and assimilation, the top-up degree students were characterized by substantially broader response heterogeneity. The process-oriented insights in this study provide a vital methodological foundation and interpretive baseline for subsequent chapters, setting the stage to investigate whether similar distributional discrepancies persist when subjective perceptions are triangulated against objective formative homework performance and summative final examination performance.

3.2. Homework Evaluation

To complement the subjective longitudinal information obtained from self-evaluation, a chapter-level formative homework assessment was implemented following the completion of each instructional chapter. Comprising three single-choice questions and two true-or-false questions per assessed chapter, this objective instrument provided repeated performance-based observations of students' short-term content mastery across the semester. Unlike self-evaluation, which captures internal perceptions of learning states, homework accuracy serves as an observable metric of academic performance, enabling valuable comparisons across instructional chapters and between student cohorts, which is essential for decoupling central tendency from score variability (Chapman et al., 2023; Dorko, 2020).

As illustrated in **Figure 4(a)**, the single-choice question accuracy trajectories exhibit substantial chapter-to-chapter fluctuations for both student cohorts, accompanied by weak overall positive linear trends. The linear regressions yielded $T_{UC} = 0.0122x + 0.7080$ ($R^2 = 0.1013$) for regular undergraduate cohort and $T_{TC} = 0.0065x + 0.7602$ ($R^2 = 0.0307$) for top-up degree cohort. Although the regular undergraduate cohort manifested an estimated slope approximately 87.7% steeper than that of the top-up cohort, and linear regression analyses were employed to estimate overall temporal directions ($Y_t = \beta_0 + \beta_1 t + \epsilon_t$), the low coefficients of determination ($R^2 < 0.25$) underscore that chapter-to-chapter performance is inherently non-linear. As shown in **Figure 4(a)**, the raw trajectories exhibit localized performance drops followed by rapid rebounds, reflecting shifting cognitive demands across distinct engineering safety topics rather than a monotonic linear accumulation. Due to this, linear trendlines and coefficients of determination (R^2) in this study are interpreted as macro-level goodness-of-fit indicators capturing general directional trajectories, rather than implying absolute dynamic stability or simple linear progression.

A related yet distinct dynamic pattern emerged in the true-or-false assessment component presented in **Figure 4(b)**. The linear regression equations yielded $T_{UC} = 0.0114x + 0.7609$ ($R^2 = 0.2308$) and $T_{TC} = 0.0127x + 0.7453$ ($R^2 = 0.1601$) for the regular undergraduate and top-up degree cohorts, respectively. In contrast to the single-choice results, the top-up degree cohort displayed a slightly steeper linear regression slope (an 11.4% increase relative to undergrad-

uates). Nevertheless, the modest R^2 values reaffirm that chapter sequence alone explains a limited portion of performance variability, which underscores that formative achievement fluctuates in response to chapter-specific content characteristics and question complexity rather than adhering to a strict linear progression.

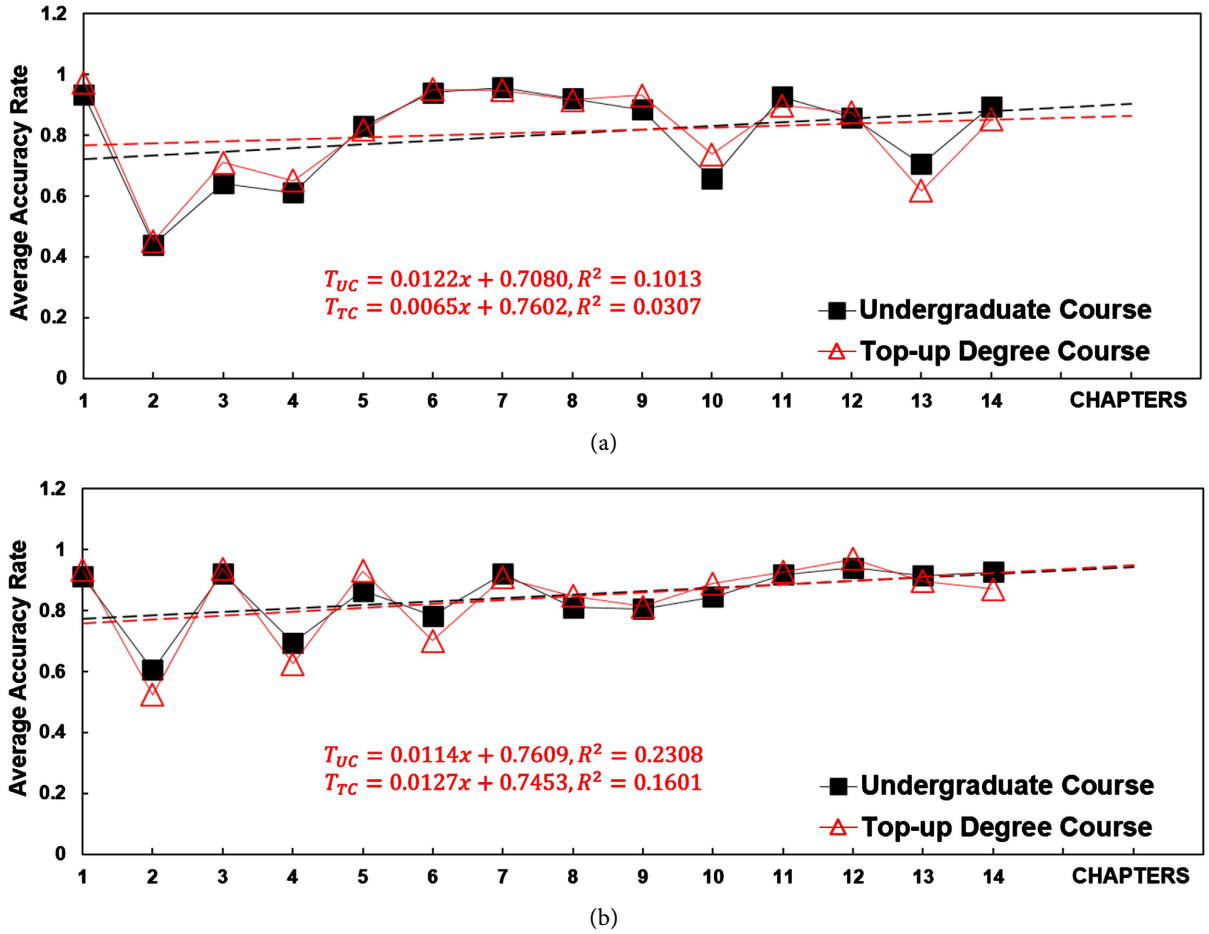


Figure 4. Dynamic accuracy rate evaluation. (a) Single-Choice Question, (b) True or False Question.

The aggregate quantitative results detailed in Table 4 demonstrate that the central tendencies of formative homework performance were remarkably similar between the two cohorts, contrasting sharply with the visual variability of chapter curves in Figure 4. For single-choice questions (Q1 - Q3), the regular undergraduate cohort achieved mean scores of 87.13, 70.63, and 82.06, compared with 87.15, 73.53, and 81.97 for the top-up degree cohort, yielding an overall average accuracy of 79.94 versus 80.88 (a negligible relative difference of 1.18%). For true-or-false questions (Q4 - Q5), the regular undergraduate cohort recorded mean scores of 90.02 and 79.28, while the top-up degree cohort obtained 88.77 and 79.28, resulting in overall averages of 84.65% and 84.02% (a 0.75% relative difference). At the individual question level, inter-cohort divergence remained minimal, peaking at 4.11% for single-choice Q2 and dropping to 0.0045% for true-or-false Q5.

Table 4. Comparison analysis of dynamic homework evaluation.

		Single-Choice Question				True or False Question		
		Q1	Q2	Q3	Average	Q4	Q5	Average
Average Evaluation (pts)	UC	87.13	70.63	82.06	79.94	90.02	79.28	84.65
	TC	87.15	73.53	81.97	80.88	88.77	79.28	84.02
	Rate	0.02%	4.11%	0.12%	1.18%	1.41%	0.0045%	0.75%
Standard Deviation	UC	15.91	13.84	14.93	14.89	20.92	16.62	18.77
	TC	35.91	32.06	33.53	33.83	49.04	41.57	45.31
	Rate	125.75%	131.69%	124.68%	127.23%	134.44%	150.15%	141.40%

Conversely, the primary distinction between cohorts emerged in score dispersion and within-group heterogeneity, as quantified in **Table 4**. For single-choice assessments, standard deviations for the regular undergraduate cohort stood at 15.91, 13.84, and 14.93 across Q1 - Q3, whereas those for the top-up degree cohort reached 35.91, 32.06, and 33.53, culminating in an average standard deviation increase of 127.23% (from 14.89 to 33.83). An even more pronounced dispersion contrast materialized in the true-or-false component, where average standard deviations expanded from 18.77 for the regular undergraduate cohort to 45.31 for the top-up degree cohort by a 141.40% relative increase. The statistical disparities in this study confirm that while both cohorts maintained comparable average accuracy, the top-up degree cohort exhibited substantially greater internal score dispersion.

A comparative examination of the two question formats provides further insight into students' formative performance profiles. Overall mean accuracy was consistently higher for true-or-false questions than for single-choice items across both cohorts (84.65% vs. 79.94% for the regular undergraduate cohort; 84.02% vs. 80.88% for the top-up cohort). The absolute difference between format means was 4.71 percentage points for regular undergraduates and 3.14 percentage points for top-up students. The format-level convergence through the discussion indicates that broad achievement levels were stable, while including the discriminating power of single-choice Q2, question-level variations demonstrate the diagnostic necessity of granular, chapter-resolved monitoring over aggregate scoring.

The graphical distributions depicted in **Figure 5** visually corroborate the statistical findings above across the 14 instructional chapters. Throughout the semester, error bars and confidence intervals for the top-up degree cohort were consistently wider than those for the regular undergraduate cohort, particularly within the true-or-false assessment panels. In accordance with previous studies (Ost et al., 2017; Soffer-Vital & Glassner, 2026), this heightened dispersion cannot be interpreted as an inherent indicator of inferior student capability or lower academic quality. Due to this, standard deviations measure score spread around the mean without identifying causative factors, leading to the objective conclusion that top-up students displayed greater behavioral and performance heterogeneity under identical instructional conditions.

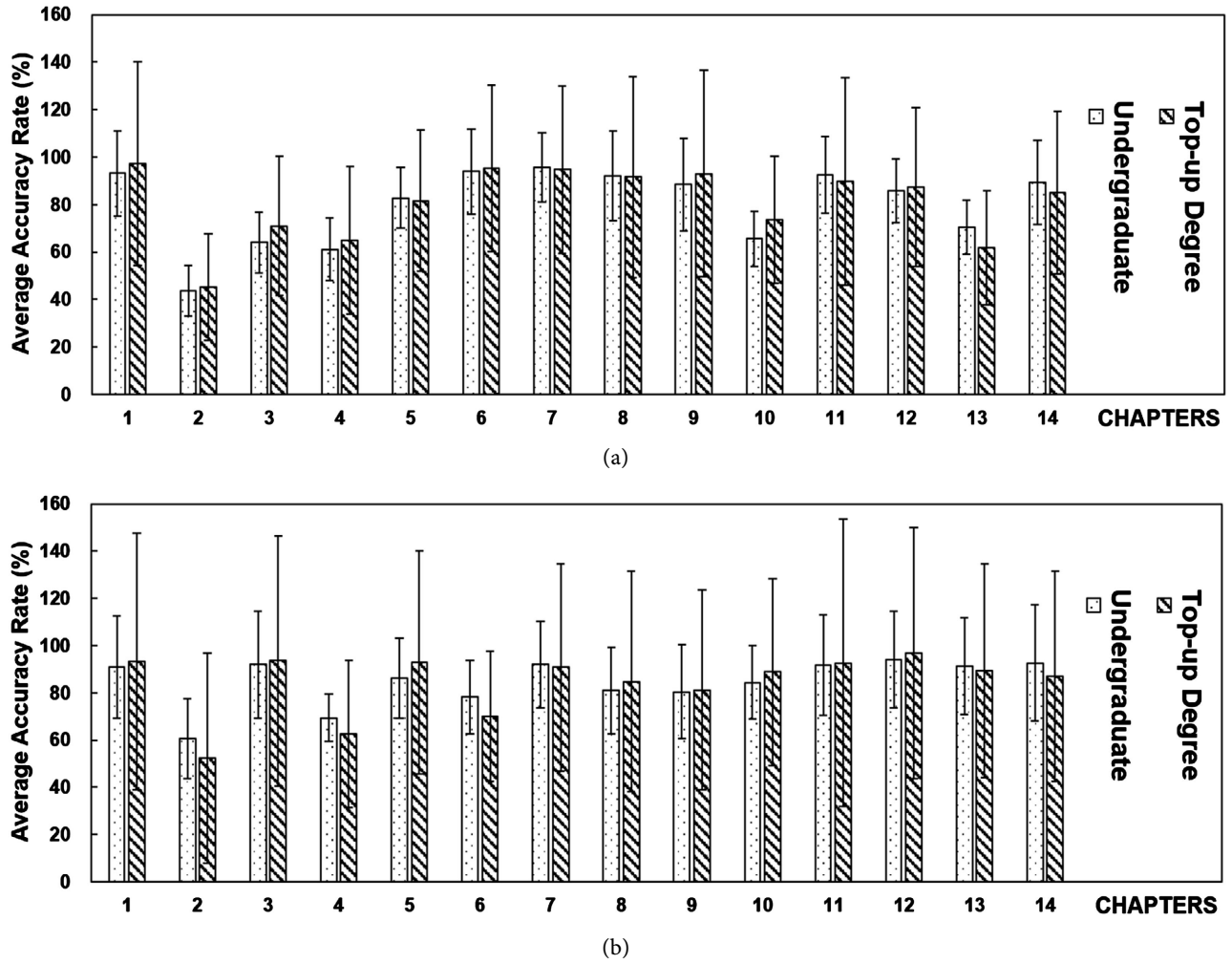


Figure 5. Quantitative analysis of the accuracy rate. (a) Single-choice question, (b) True or false question.

The alignment between the formative homework results and the preceding self-evaluation data yields compelling convergent descriptive evidence. Both assessment tiers indicate that the two cohorts shared highly comparable central tendencies paired with markedly larger score dispersion in the top-up degree cohort (with relative standard-deviation expansions exceeding 125% across corresponding metrics). Although the observational dataset precludes formal student-level paired correlation analyses or causal claims regarding whether engagement directly drove homework accuracy, the consistent distributional patterns across subjective and objective indicators reinforce the validity of a multi-layered evaluation framework in this study.

Across the discussion above, the chapter-level homework assessment establishes a crucial intermediate analytical bridge between longitudinal self-evaluation and terminal summative examinations. The findings demonstrate that relying solely on mean accuracy scores conceals profound differences in internal score distributions between student cohorts. By capturing localized performance fluctuations and response heterogeneity, formative homework provides actionable di-

agnostic insights for pedagogical adjustment, supplying the empirical foundation for examining terminal outcomes across curriculum targets.

3.3. Final Exam Evaluation

The final examination constituted the summative component of the proposed integrated assessment framework, designed to ascertain whether the learning patterns and trajectories observed through longitudinal self-evaluation and chapter-level formative homework were reflected in end-of-course academic achievement (Axelsen et al., 2025; Dong et al., 2026; Gao et al., 2025; Kausar, 2026; McTaggart & Hortsch, 2024; Otieno Juma et al., 2026). Referring to Broadbent et al. (2018) and Panadero et al. (2019), unlike the process-sensitive indicators examined in the discussions regarding self and homework evaluation, which captured real-time student engagement and short-term content mastery, the final examination provided a standardized terminal evaluation of demonstrated student performance following the completion of the instructional sequence. The terminal phase of this study served as a crucial reference for evaluating cumulative learning performance across diverse cognitive and response demands.

The final examination was structurally organized around three distinct educational targets to ensure comprehensive curricular coverage in **Table 5**. Target 1 was established to evaluate foundational knowledge acquisition and the development of basic competence in engineering safety management, assessed through single-choice, true-or-false, and multiple-choice questions. Target 2 focused on the systematic identification of workplace hazards and the formulation of practical safety management measures, assessed via short-answer questions. Target 3 addressed the theoretical foundations and core principles required for advanced engineering safety management, evaluated through essay questions. The target-oriented architecture for studying performance evaluation in this study enabled the summative evaluation to transcend a simplistic aggregate score, allowing achievement to be differentiated across varying levels of cognitive complexity and knowledge expression.

Table 5. Comparison analysis of the final exam score.

		Target 1			Target 2	Target 3
		Single-Choice	True or False	Multiple-Choice	Short-Answer	Essay
Average Score (pts)	Undergraduate	91.05	90.26	86.84	84.12	62.28
	Top-up Degree	87.07	82.56	65.24	85.61	65.04
	Rate	4.57%	9.33%	33.10%	1.77%	4.43%
Standard Deviation	Undergraduate	8.63	11.50	22.31	11.92	15.64
	Top-up Degree	14.36	12.75	24.14	11.28	16.42
	Rate	66.30%	10.80%	8.18%	5.63%	4.98%

The quantitative results detailed in Table 5 reveal a pronounced dependence of cohort disparities on both the assessment target and the specific response format. For Target 1, the regular undergraduate cohort achieved mean scores of 91.05, 90.26, and 86.84 on single-choice, true-or-false, and multiple-choice components, respectively, whereas the top-up degree cohort recorded corresponding scores of 87.07, 82.56, and 65.24. Applying the relative difference convention, between-group gaps were 4.57%, 9.33%, and a striking 33.10% for the multiple-choice section. Conversely, performance on Target 2 short-answer questions exhibited a negligible relative difference of 1.77% (84.12 vs. 85.61), while Target 3 essay performance showed a modest relative difference of 4.43% (62.28 vs. 65.04). The findings in the discussion above demonstrate that terminal achievement cannot be generalized as a uniform cohort-level hierarchy, while performance discrepancies are highly format-specific.

A closer examination of the Target 1 components underscores the critical importance of assessment granularity as single-choice and true-or-false items yielded moderate between-cohort contrasts, the multiple-choice format exposed an exceptionally large disparity (33.10%). Because these three formats were consolidated under the same foundational curriculum target, a single aggregated target score would have completely masked this internal divergence. The pronounced performance gap in Target 1 multiple-choice questions (33.10% relative difference) suggests that tasks requiring the simultaneous discrimination of multiple plausible distractors pose a comparative challenge for students in the top-up degree cohort. However, the absence of item-level psychometric indicators (such as Item Difficulty and Item Discrimination indices) is considered a limitation, while indicting that this study cannot rule out the alternative possibility that specific distractor designs or item ambiguities contributed to the observed score disparity, emphasizing the need for future item-level psychometric verification. Although the current observational dataset lacks item-level psychometric diagnostics to isolate the exact cognitive bottlenecks, the descriptive evidence confirms that Target 1 differentiation was concentrated in the multiple-choice domain.

In the constructed-response assessments, the results in Table 5 illustrate that the top-up degree cohort achieved marginally higher mean scores than the regular undergraduates in both Target 2 short-answer items (85.61 vs. 84.12) and Target 3 essay questions (65.04 vs. 62.28), indicating that the absolute superiority of regular undergraduates observed in objective testing does not extend universally to open-ended, written expression formats. Upon discussion regarding the three different targets in this study, the final examination data point toward a differential proficiency profile rather than a pervasive deficit, reinforcing the premise that single terminal scores are insufficient for capturing student learning distributions.

The standard deviation metrics presented in Table 5 also introduce a vital interpretive dimension regarding score dispersion. For single-choice questions, standard deviations rose from 8.63 for the regular undergraduate cohort to 14.36 for the top-up degree cohort (a 66.30% increase), while true-or-false and multiple-choice standard deviations exhibited moderate expansions of 10.80% and 8.18%.

On the other hand, standard deviations for short-answer scores were 11.92 versus 11.28 with 5.63%, and 15.64 versus 16.42 for essay scores with 4.98%. This structural shift reveals that the pronounced within-group heterogeneity previously identified in dynamic self-evaluation and homework is not uniformly replicated across all final exam components, attenuating significantly in multiple-choice items and vanishing in short-answer responses.

To advance conventional test statistics, the study introduces the cumulative knowledge-absorption index (K_A) as a cross-disciplinary quantitative metric. By reflecting mechanical force and displacement in examination scores and ordered student population coordinates, this study conceptualizes this index via the mathematical analogy of hysteresis energy dissipation in structural engineering (Shi & Choi, 2026; Taborda et al., 2016; Villalobos Luna et al., 2026; Wang et al., 2024; Yang et al., 2026) and integrates score magnitude with student distribution. Although traditional statistical metrics (e.g., standard deviation, skewness, or interquartile ranges) characterize distribution shapes independently, they fail to integrate individual score magnitudes with cumulative population placement. The cumulative knowledge-absorption index (K_A), formulated via numerical integration over ordered student population coordinates (n), bridges the analytical gap by treating the score profile as an area-based accumulation function, thereby offering a more comprehensive metric of cohort-level achievement than central tendency statistics alone. As established in the previous chapter, the continuous formulation of this area-under-the-score-distribution metric is defined by **Equation (6)**, integrating the score profile along the ordered population axis to capture both individual achievement levels and population-level spread.

Because educational examination records are discrete rather than continuous, the practical evaluation of cumulative knowledge absorption was performed numerically via the trapezoidal approximation rule articulated in **Equation (7)**. The discrete expression in this study computes the cumulative area beneath the ordered score distribution across adjacent student position intervals ($n_{i+1} - n_i$), correcting potential notation ambiguities and ensuring complete mathematical reproducibility. Measured in point-person units (pts·p), K_A functions as a population-score accumulation metric rather than a physical energy quantity, maintaining strict dimensional consistency and avoiding erroneous thermodynamic interpretations.

The aggregate and target-specific knowledge-absorption findings summarized in Table 6 demonstrate that cumulative absorption is not merely proportional to average test scores. Mean absorption values for the regular undergraduate cohort across Targets 1 - 3 were 35.91 pts·p, 25.18 pts·p, and 18.69 pts·p, totaling 79.78 pts·p, compared with 32.16 pts·p, 25.67 pts·p, and 19.45 pts·p for the top-up degree cohort, totaling 77.29 pts·p (an overall relative difference of 3.22%). Consistent with Table 5, Target 1 accounted for the largest between-cohort contrast (11.65%), whereas Targets 2 and 3 exhibited close alignment (1.97% and 4.07% relative differences). This internal consistency bridges conventional examination statistics with the proposed cumulative framework.

Table 6. Knowledge absorption analysis.

		Target 1	Target 2	Target 3	Total
Average Knowledge Absorption (pts·p)	Undergraduate	35.91	25.18	18.69	79.78
	Top-up Degree	32.16	25.67	19.45	77.29
	Rate	11.65%	1.97%	4.07%	3.22%
Standard Deviation	Undergraduate	2.45	2.56	3.19	6.19
	Top-up Degree	3.39	2.36	3.32	6.26
	Rate	38.28%	8.45%	4.15%	1.01%
Cumulative knowledge absorption (pts·p)	Undergraduate	1328.75	931.50	691.50	2951.75
	Top-up Degree	2605.25	2079.50	1575.50	6260.25

The dispersion statistics associated with cumulative knowledge absorption in Table 6 offer further analytical refinement. For Target 1, standard deviations were 2.45 for the regular undergraduate cohort and 3.39 for the top-up degree cohort (a 38.28% difference), whereas Target 2 showed an inversion (2.56 vs. 2.36), indicating slightly lower variability for the top-up cohort. Target 3 demonstrated comparable dispersion (3.19 vs. 3.32). Also, the analytical results in Table 6 illustrate that the standard deviations converged to 6.19 and 6.26, differing by only 1.01% for the total knowledge-absorption index. The remarkable convergence in Table 6 proves that once performance is aggregated across the complete examination target structure, overall cohort variability becomes nearly identical, refuting the notion of universal heterogeneity in top-up students.

Figure 6 and Figure 7 provide comprehensive visual representations of these target-specific and cumulative absorption trajectories across student population coordinates (n) for the regular undergraduate and top-up degree cohorts. Due to the different total numbers of participant students for the two cohorts, the raw cumulative totals reported in Table 6 (2951.75 pts·p for regular undergraduates and 6260.25 pts·p for top-up students) must be interpreted with caution. Because the two cohorts possessed unequal sample sizes ($n = 38$ vs. $n = 83$), raw cumulative sums are intrinsically scaled by cohort magnitude. Consequently, raw cumulative absorption appropriately describes total cohort-level accumulation but cannot serve as a direct comparative measure of individual effectiveness without normalization, rendering the normalized mean index (79.78 pts·p vs. 77.29 pts·p) the statistically rigorous basis for cohort evaluation.

The average accuracy rates against average cumulative knowledge-absorption indices across the three curriculum targets are visually juxtaposed in Figure 8, highlighting that while percentage-based accuracy rates can sometimes exaggerate or obscure performance gaps due to scaling artifacts, the knowledge-absorption metric enables to integrate both central tendency and distributional properties. This dual representation confirms that Target 1 represents the primary locus of inter-cohort differentiation, whereas Targets 2 and 3 reflect substantial parity in advanced competency application.

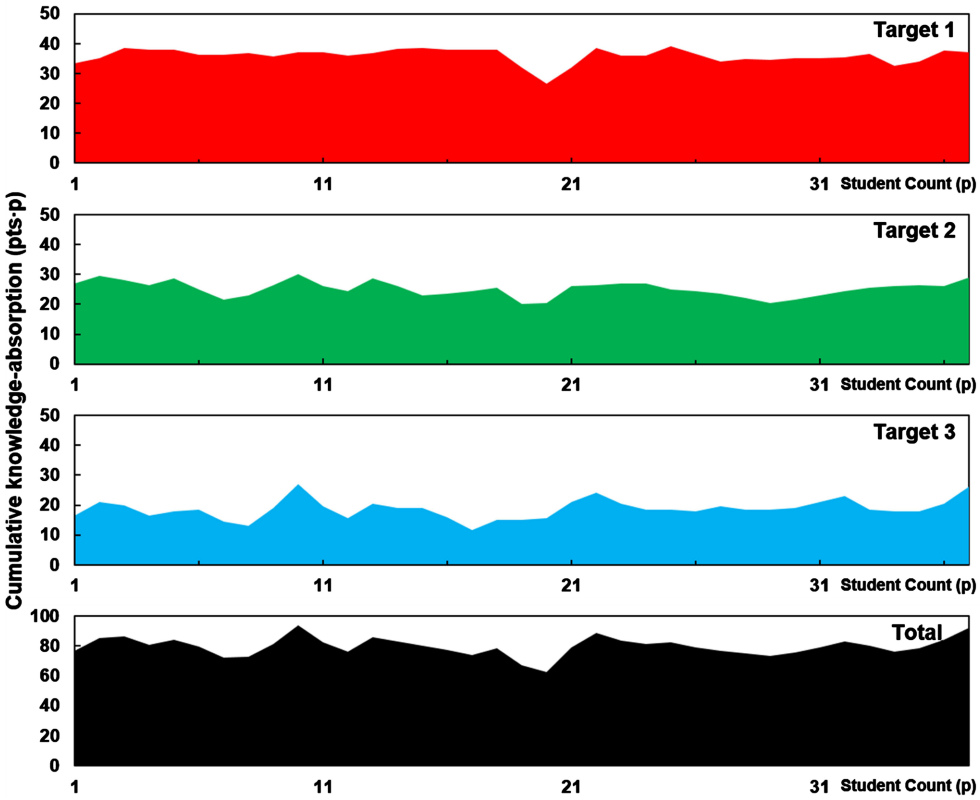


Figure 6. Cumulative knowledge absorption (Undergraduate Course).

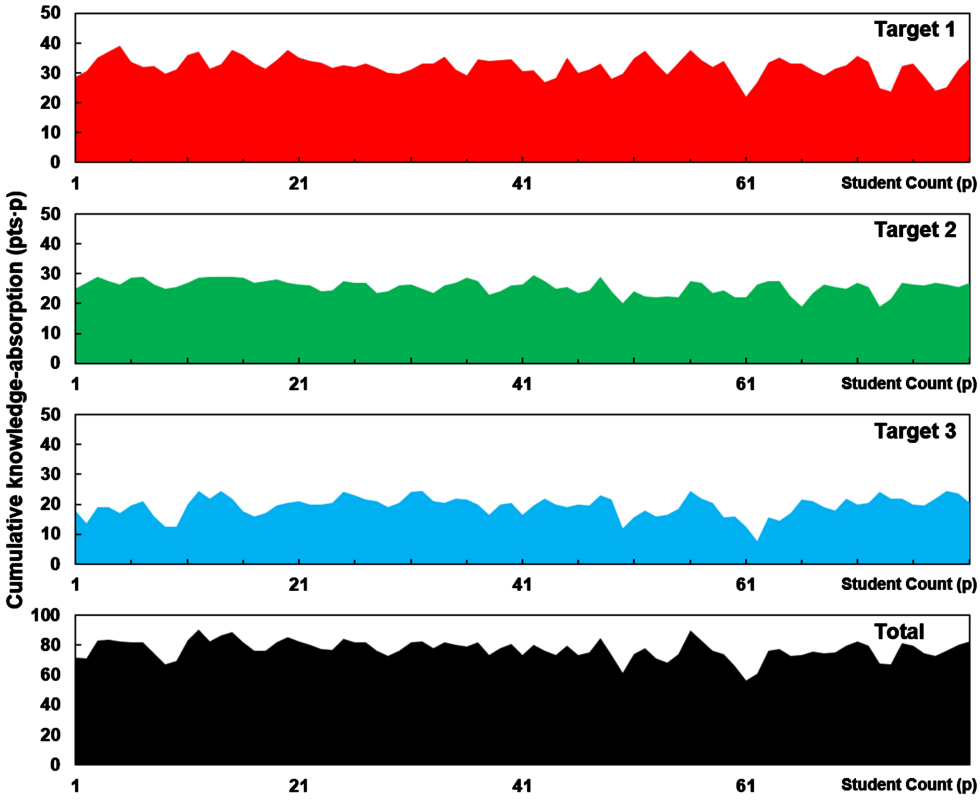


Figure 7. Cumulative knowledge absorption (Top-up Degree Course).

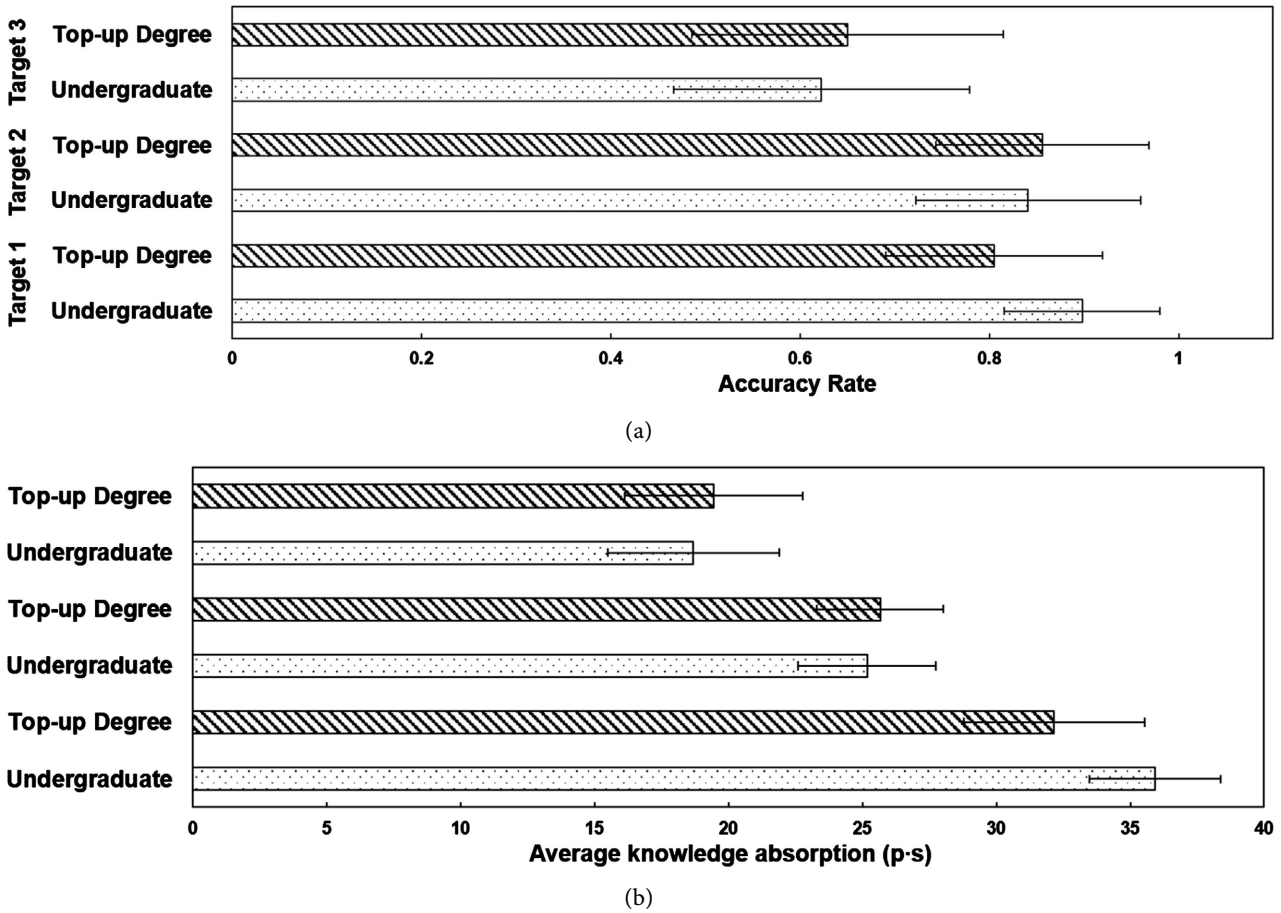


Figure 8. Accuracy rate and Cumulative knowledge absorption. (a) Average accuracy rate, (b) Average cumulative knowledge absorption.

Synthesizing the final examination findings with the preceding chapters, the multi-stage narrative indicates that while dynamic self-evaluation and formative homework established that the top-up degree cohort possessed greater response dispersion around comparable means, the summative examination demonstrates that performance divergence is highly sensitive to cognitive targets and item formats. The alignment between traditional exam metrics and the novel cumulative index establishes the dynamic-static framework as an advanced triangulation architecture, wherein process perceptions, formative performance, and terminal achievement yield complementary insights.

Furthermore, the incorporation of the cumulative knowledge-absorption index bridges structural engineering analytics with educational measurement through explicit mathematical calculations. By repurposing an area-based accumulation principle to quantify population score distributions, the index enhances interpretive depth without invoking untenable physical analogies. In conjunction with dynamic observations, the final examination results demonstrate that educational performance varies across temporal trajectories, assessment formats, and curriculum targets, offering an empirically grounded foundation for advanced instructional optimization.

4. Conclusion

This study established and empirically evaluated a dynamic-static integrated assessment framework within a university engineering safety and environmental protection course, systematically linking longitudinal self-evaluation and chapter-level formative homework with terminal summative exam performance across regular undergraduate ($n = 38$) and top-up degree ($n = 83$) cohorts. Longitudinal self-evaluation demonstrated distinct positive temporal trajectories in attentional engagement ($T_{UC} = 0.0058x + 0.8263$, $R^2 = 0.6645$ & $T_{TC} = 0.0076x + 0.8154$, $R^2 = 0.9117$) and perceived knowledge acquisition ($T_{UC} = 0.0058x + 0.8319$, $R^2 = 0.7270$ & $T_{TC} = 0.0086x + 0.7961$, $R^2 = 0.9319$), whereas perceived curricular significance exhibited more gradual growth ($T_{UC} = 0.0031x + 0.8937$ & $T_{TC} = 0.0026x + 0.8804$). More significantly, while semester-level mean self-evaluations remained remarkably close between cohorts (0.30 to 1.74 percentage point absolute differences), the top-up degree cohort exhibited 103.62% to 130.21% larger standard deviations across all three process dimensions, highlighting substantial within-group response heterogeneity beneath nearly identical central tendencies.

A convergent distributional pattern emerged in objective formative homework assessment, where mean accuracy rates differed minimally between cohorts (79.94% vs. 80.88% for single-choice; 84.65% vs. 84.02% for true-or-false), yet standard deviations were 127.23% and 141.40% higher in the top-up degree cohort. On the other hand, terminal summative achievement demonstrated marked target and format dependence, as regular undergraduates outperformed top-up students in foundational Target 1 objective formats, most notably in multiple-choice questions (86.84 vs. 65.24, a 33.10% relative discrepancy), whereas top-up students achieved slightly higher mean scores in Target 2 short-answer (85.61 vs. 84.12) and Target 3 essay questions (65.04 vs. 62.28). Cross-disciplinary evaluation via the population-score integral framework further revealed that normalized cumulative knowledge-absorption indices remained comparable overall (79.78 pts·p vs. 77.29 pts·p, a 3.22% gap), with total score variability converging to within 1.01% between cohorts.

The empirical findings confirm that relying exclusively on terminal mean examination scores obscures vital temporal dynamics, format-specific proficiencies, and within-cohort dispersion. By triangulating subjective engagement perceptions, short-term formative accuracy, and target-differentiated outcome indices, the dynamic-static assessment architecture provides advanced diagnostic resolution to inform real-time pedagogical adjustments and curriculum optimization. Given the non-experimental, observational nature of the pre-existing cohort design and the absence of baseline equivalence control (e.g., prior GPA or baseline competency alignment), the results in this study are better interpreted as descriptive, practice-based comparative evidence of learning heterogeneity within a common educational environment, rather than causal proof of educational pathway efficacy or pedagogical intervention effects. Within these constraints, the frame-

work is still expected to offer an actionable, multidimensional evaluation paradigm for modern higher education.

Funding

This work was supported by the Project of the Shaanxi Provincial Educational Science Planning Program, "Research on Dynamic Modeling of Undergraduate Students' Learning Behaviors and Academic Performance Prediction Based on Artificial Intelligence", under Grant AK-25-30, and the 2025 science and technology project of Ankang Municipal Science and Technology Bureau, "AI-based Plush Toy Image Generation and Design Using Deep Learning Neural Networks", under Grant AK2025-GY-24, which provided essential financial funding for the computational resources and survey administration utilized in this work.

Acknowledgements

This work benefited significantly from the comprehensive resources provided by the Ankang Science & Technology Innovation Center (Qinchuangyuan Platform), which facilitated the execution and provided technical assistance for this scientific investigation.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

Abd Manap, N., & Mohd Zain, A. S. (2026). Enhancing Student Centered and Technology Supported Learning in Higher Education through the BEAM Framework. *Discover Education, 5,* Article No. 503. https://doi.org/10.1007/s44217-026-01546-2

Arja, S. B., Acharya, Y., Alezaireg, S., Ilavarasan, V., Ala, S., & Arja, S. B. (2018). Implementation of Formative Assessment and Its Effectiveness in Undergraduate Medical Education: An Experience at a Caribbean Medical School. *MedEdPublish, 7,* Article 131. https://doi.org/10.15694/mep.2018.0000131.1

Avila-Andino, M. P., González-Vázquez, K. Y., Pérez-Manzo, O. A., Cárdenas-Cobo, J. D. P., Guevara-Viejó, J. F., & Rodas-Silva, J. L. (2026). Progressive Integration of Meaningful Summative Assessment and Formative Assessment: A Longitudinal Analysis of Student Performance in Higher Education. *Frontiers in Education, 11,* Article 1906986. https://doi.org/10.3389/feduc.2026.1906986

Axelsen, T., King, R., & Curtis, E. (2025). Visualising and Evaluating Learning/Achievement Consistency in Introductory Statistics. *Cogent Education, 12,* Article ID: 2492727. https://doi.org/10.1080/2331186x.2025.2492727

Ayaneh, M. G., Dessie, A. A., & Fetene, D. M. (2021). Psychometric Properties of Student Evaluation of Teachers' Performance Scale: Evidence from Debre Markos University Students' Evaluation Dataset. *Education Research International, 2021,* Article ID: 5543317. https://doi.org/10.1155/2021/5543317

Ayoubi, R. M., Solwa, B., Ayoubi, M. T., & Crawford, M. (2025). Bridging the Gap: Academic Performance and Experiences of Technical College Students in Top-Up University Degree Programme. *Cogent Education, 12,* Article ID: 2441059.

https://doi.org/10.1080/2331186x.2024.2441059

Bonte, M., & Brem, S. (2024). Unraveling Individual Differences in Learning Potential: A Dynamic Framework for the Case of Reading Development. *Developmental Cognitive Neuroscience, 66,* Article ID: 101362. https://doi.org/10.1016/j.dcn.2024.101362

Boud, D., & Falchikov, N. (2006). Aligning Assessment with Long-Term Learning. *Assessment & Evaluation in Higher Education, 31,* 399-413. https://doi.org/10.1080/02602930600679050

Brassil, C. E., & Couch, B. A. (2019). Multiple-True-False Questions Reveal More Thoroughly the Complexity of Student Thinking than Multiple-Choice Questions: A Bayesian Item Response Model Comparison. *International Journal of STEM Education, 6,* Article No. 16. https://doi.org/10.1186/s40594-019-0169-0

Broadbent, J., Panadero, E., & Boud, D. (2018). Implementing Summative Assessment with a Formative Flavour: A Case Study in a Large Class. *Assessment & Evaluation in Higher Education, 43,* 307-322. https://doi.org/10.1080/02602938.2017.1343455

Bulut, O., Gorgun, G., & Yildirim-Erbasli, S. N. (2025). The Impact of Frequency and Stakes of Formative Assessment on Student Achievement in Higher Education: A Learning Analytics Study. *Journal of Computer Assisted Learning, 41,* e13087. https://doi.org/10.1111/jcal.13087

Cartwright, N. (2025). There Is No Baby, Just Lots of Bathwater: An Argument against Summative Assessment. *The Law Teacher, 59,* 450-469. https://doi.org/10.1080/03069400.2025.2531720

Chanaka Ushantha, R. A., & Samantha Kumara, P. A. P. (2016). A Quest for Service Quality in Higher Education: Empirical Evidence from Sri Lanka. *Services Marketing Quarterly, 37,* 98-108. https://doi.org/10.1080/15332969.2016.1154731

Chapman, K. E., Davidson, M. E., Azuka, N., & Liberatore, M. W. (2023). Quantifying Deliberate Practice Using Auto-Graded Questions: Analyzing Multiple Metrics in a Chemical Engineering Course. *Computer Applications in Engineering Education, 31,* 916-929. https://doi.org/10.1002/cae.22614

Cheung, K., Ho, C. S. W., Tsang, H., & Lau, E. (2022). Comparing the Learning Approaches of Transfer Students and Direct Entrants in an Asian Higher Education Context. *Sustainability, 14,* Article 2523. https://doi.org/10.3390/su14052523

Couch, B. A., Wood, W. B., & Knight, J. K. (2015). The Molecular Biology Capstone Assessment: A Concept Assessment for Upper-Division Molecular Biology Students. *CBE—Life Sciences Education, 14,* ar10. https://doi.org/10.1187/cbe.14-04-0071

Diao, X., Hu, R., Zeng, Q., Song, Z., & Duan, H. (2026). Dynamic Domain Learning Ability Enhanced Knowledge Tracing with Stability. *Scientific Reports, 16,* Article No. 23576. https://doi.org/10.1038/s41598-026-53677-z

Dong, D., Chen, M., Zhang, S., Chen, Y., & Ji, F. (2026). Effects of Formative-Summative Assessment Weight Configurations on Evaluation Validity and Discriminatory Power in Landscape Architecture Planning and Design Courses. *Scientific Reports, 16,* Article No. 19299. https://doi.org/10.1038/s41598-026-50554-7

Dorko, A. (2020). Red X's and Green Checks: A Model of How Students Engage with Online Homework. *International Journal of Research in Undergraduate Mathematics Education, 6,* 446-474. https://doi.org/10.1007/s40753-020-00113-w

Dumas, D., McNeish, D., & Greene, J. A. (2020). Dynamic Measurement: A Theoretical-Psychometric Paradigm for Modern Educational Psychology. *Educational Psychologist, 55,* 88-105. https://doi.org/10.1080/00461520.2020.1744150

Enders, N., Gaschler, R., & Kubik, V. (2021). Online Quizzes with Closed Questions in

Formal Assessment: How Elaborate Feedback Can Promote Learning. *Psychology Learning & Teaching, 20,* 91-106. https://doi.org/10.1177/14757257209712O5

Freeman, S., Haak, D., & Wenderoth, M. P. (2011). Increased Course Structure Improves Performance in Introductory Biology. *CBE—Life Sciences Education, 10,* 175-186. https://doi.org/10.1187/cbe.10-08-0105

French, S., Dickerson, A., & Mulder, R. A. (2024). A Review of the Benefits and Drawbacks of High-Stakes Final Examinations in Higher Education. *Higher Education, 88,* 893-918. https://doi.org/10.1007/s10734-023-01148-z

Gao, Z., Yan, H., Liu, J., Cui, C., Wang, J., Zhang, X. et al. (2025). Untangling Complexity of Competency Evolution: Multi-Channel Learning Trajectories of Score, Engagement and Problem-Solving Efficiency. *British Journal of Educational Technology*. https://doi.org/10.1111/bjet.70017

Garzón, B., Berger, S., Driver, C. C., & Tomasik, M. J. (2026). Prediction of Performance in Standardised Assessments from Computer-Based Formative Assessment Data. *Technology, Knowledge and Learning, 31,* 967-995. https://doi.org/10.1007/s10758-025-09947-2

Gratchev, I., Howell, S., & Stegen, S. (2024). Academics' Perception of Final Examinations in Engineering Education. *Australasian Journal of Engineering Education, 29,* 20-29. https://doi.org/10.1080/22054952.2023.2284484

Henriquez-Machado, R., Muñoz-Villamizar, A., Florez-Jimenez, M. P., & Santos, J. (2026). Strategy Transitions in an Operations Management Course: A Longitudinal Case Study of Student Performance and Teaching Evaluations in Colombia. *Interactive Learning Environments*. https://doi.org/10.1080/10494820.2026.2680275

Ivins, T., Copenhaver, K., & Koclanes, A. (2017). Adult Transitional Theory and Transfer Shock in Higher Education: Practices from the Literature. *Reference Services Review, 45,* 244-257. https://doi.org/10.1108/rsr-08-2016-0048

Johnson, B. C., & Kiviniemi, M. T. (2009). The Effect of Online Chapter Quizzes on Exam Performance in an Undergraduate Social Psychology Course. *Teaching of Psychology, 36,* 33-37. https://doi.org/10.1080/00986280802528972

Jones, D., & Oh, S. (2024). Correlation between Student Performances on Case-Based Constructed-Response Formative Assessment and Summative Assessment. *Journal of Medical Education and Curricular Development, 11,* 1-7. https://doi.org/10.1177/23821205241239496

Kausar, N. (2026). Validity Evidence and Equity of a Multi-Component Assessment Framework in a Developmental Anatomy Elective among Undergraduate Medical Students across Parallel Sections. *Clinical Anatomy*. https://doi.org/10.1002/ca.70217

Khosla, A., Singh, H., Aggarwal, A., & Bedi, J. (2026). A Knowledge Tracing Based Intelligent Framework for Formative Assessment of Students Skills. *Complex & Intelligent Systems, 12,* Article No. 22. https://doi.org/10.1007/s40747-025-02149-4

Kostiv, O., Rodríguez-Hernández, A. F., & Hernández, J. D. (2025). Teachers' Emotional Commitment: The Emotional Bond That Sustains Teaching. *Journal of Intelligence, 13,* Article 158. https://doi.org/10.3390/jintelligence13120158

Marks, P., & Tuytens, P. (2026). Technology-Enhanced Differentiation in Public Administration Education: A Flipped Classroom Approach to Student Heterogeneity. *Teaching Public Administration*. https://doi.org/10.1177/01447394261472515

McTaggart, S., & Hortsch, M. (2024). Early Practice Makes Histology Masters: The Use of a Formative Assessment Quiz to Prepare Histology Learners for a High-Stakes Final Examination. *Anatomical Sciences Education, 17,* 1265-1274.

https://doi.org/10.1002/ase.2472

Mohammad, R. F., & Kamran, M. (2023). Examining the Efficacy of Online Learning in Nurturing Students' Learning: An Analysis of Students' Experiences. *Asian Association of Open Universities Journal, 18,* 218-232. https://doi.org/10.1108/aaouj-11-2022-0163

Montano, V. E., & Sobrejuanite, G. C. (2025). Evaluating Teaching Effectiveness and Student Performance across Diverse Courses: An Analysis of Final Exam Scores and Teaching Techniques. *Journal of Learning Theory and Methodology, 6,* 24-32. https://doi.org/10.17309/jltm.2025.6.1.03

Na, W., Jung, S., & Kim, M. (2026). Longitudinal Associations between Homework and Examination Performance in Engineering Mathematics. *Education Sciences, 16,* Article 1610. https://doi.org/10.3390/educsci16101610

Ost, B., Gangopadhyaya, A., & Schiman, J. C. (2017). Comparing Standard Deviation Effects across Contexts. *Education Economics, 25,* 251-265. https://doi.org/10.1080/09645292.2016.1203868

Otieno Juma, Z., Doz, D., Wastalas D'Assise Dogbalou, M., & Lewanski, D. (2026). Evaluating the Impact of Stack-Based Formative Assessment on Final Examination Performance in First Year Linear Algebra. *International Journal of Mathematical Education in Science and Technology*. https://doi.org/10.1080/0020739x.2026.2723184

Panadero, E., Broadbent, J., Boud, D., & Lodge, J. M. (2019). Using Formative Assessment to Influence Self- and Co-Regulated Learning: The Role of Evaluative Judgement. *European Journal of Psychology of Education, 34,* 535-557. https://doi.org/10.1007/s10212-018-0407-8

Peluso, A., Coretto, P., Niglio, M., Parrella, M. L., & Restaino, M. (2026). OPeDi: A Shiny Application to Monitoring Students' Career. *Social Indicators Research, 184,* Article No. 6. https://doi.org/10.1007/s11205-026-03891-z

Pinkosova, Z., McGeown, W. J., & Moshfeghi, Y. (2023). Moderating Effects of Self-Perceived Knowledge in a Relevance Assessment Task: An EEG Study. *Computers in Human Behavior Reports, 11,* Article ID: 100295. https://doi.org/10.1016/j.chbr.2023.100295

Ren, H., Zhang, W., & Cao, H. (2026). Formative Assessment as a Self-Regulation Engine in Blended Language Learning: A Case Study with Top-Tier Engineering Talents. *Acta Psychologica, 264,* Article ID: 106505. https://doi.org/10.1016/j.actpsy.2026.106505

Saparbayeva, E., Abdualiyeva, M., Torebek, Y., Kostangeldinova, A., Tursynbayev, A., Takibayeva, G. et al. (2025). Transforming Mathematics Education in Kazakhstan: Evaluating the Impact of Innovative Teaching Methods on Student Outcomes in Technical Universities. *Cogent Education, 12,* Article ID: 2461978. https://doi.org/10.1080/2331186x.2025.2461978

Shi, M., & Choi, Y. (2026). Comparative Quasi-Static Analysis of the Seismic Performance of Piloti-Type RC Structures with and without Seismic Isolation. *Structural Engineering and Mechanics, 98,* 93-116.

Shi, M., & Tang, B. (2026). Quantifying the Learning Process of Professional English with an Importance Weighted Absorption Metric. *Creative Education, 17,* 1046-1074. https://doi.org/10.4236/ce.2026.176064

Soffer-Vital, S., & Glassner, A. (2026). Ongoing and Holistic Self Assessment as a Developmental Practice: Learner Agency, Reflection and Identity Formation in Higher Education. *Assessment & Evaluation in Higher Education*. https://doi.org/10.1080/02602938.2026.2643334

Taborda, D. M. G., Potts, D. M., & Zdravković, L. (2016). On the Assessment of Energy

Dissipated through Hysteresis in Finite Element Analysis. *Computers and Geotechnics, 71,* 180-194. https://doi.org/10.1016/j.compgeo.2015.09.001

Teruna, D. R., Majid, T. A., & Budiono, B. (2015). Experimental Study of Hysteretic Steel Damper for Energy Dissipation Capacity. *Advances in Civil Engineering, 2015,* Article ID: 631726. https://doi.org/10.1155/2015/631726

Tong, C., & Ren, C. (2025). Deep Knowledge Tracing and Cognitive Load Estimation for Personalized Learning Path Generation Using Neural Network Architecture. *Scientific Reports, 15,* Article No. 24925. https://doi.org/10.1038/s41598-025-10497-x

Ueno, M., Miyazawa, Y., & Tsutsumi, E. (2025). Probability-Based Scaffolding System Using Sliding Hidden Markov IRT for Longitudinal Learning. In A. I. Cristea, E. Walker, Y. Lu, O. C. Santos, & S. Isotani (Eds.), *Artificial Intelligence in Education* (pp. 448-462). Springer. https://doi.org/10.1007/978-3-031-98417-4_32

Villalobos Luna, J. D. J., Lopez Cruz, P., Ledezma Ramírez, D. F., & Puente Córdova, J. G. (2026). Experimental Analysis of Energy Dissipation in Structural Adhesive Joints: Relationship between Hysteresis and Modal Damping. *Experimental Mechanics, 66,* 615-626. https://doi.org/10.1007/s11340-026-01278-2

Wang, J., Men, J., & Cui, Y. (2024). Hysteretic Model and Application of a Metallic Damper with Two-Phased Energy Dissipation. *Structures, 70,* Article ID: 107850. https://doi.org/10.1016/j.istruc.2024.107850

Wondim, M. G., & Dessie, B. A. (2024). Continuous Assessment: The Missing Link in Ethiopian Universities. *Cogent Education, 11,* Article ID: 2404778. https://doi.org/10.1080/2331186x.2024.2404778

Yang, Y., Zuo, Z., & Gong, M. (2026). Estimation of Structural Hysteretic Energy Dissipation Based on Normalized Energy Spectra Considering Ground Motion Duration. *Earthquake Engineering & Structural Dynamics, 55,* 1416-1433. https://doi.org/10.1002/eqe.70148

Zhang, G., Yue, X., Ye, Y., & Peng, M. Y. (2021). Understanding the Impact of the Psychological Cognitive Process on Student Learning Satisfaction: Combination of the Social Cognitive Career Theory and SOR Model. *Frontiers in Psychology, 12,* Article 712323. https://doi.org/10.3389/fpsyg.2021.712323

Zhoc, K. C. H., Webster, B. J., King, R. B., Li, J. C. H., & Chung, T. S. H. (2019). Higher Education Student Engagement Scale (HESES): Development and Psychometric Evidence. *Research in Higher Education, 60,* 219-244. https://doi.org/10.1007/s11162-018-9510-6