Jurnal Teknik Informatika (JUTIF)
P-ISSN: 2723-3863
E-ISSN: 2723-3871
Vol. 7, No. 2, April 2025, Page. 1800-1817
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.2.5646

Benchmarking Brain-Training Apps Using DEGREE and Fuzzy Logic: Lumosity vs Elevate

Heri Sismoro*[1], Mei Parwanto Kurniawan[2]

[1]Manajemen Informatika, Universitas Amikom Yogyakarta, Indonesia
[2]Sistem Informasi, Universitas Amikom Yogyakarta, Indonesia

Email: [1]herisismoro@amikom.ac.id

Received : Jan 28, 2026; Revised : Feb 22, 2026; Accepted : Mar 26, 2026; Published : Apr 18, 2026

Abstract

This study provides an actionable benchmark of two popular brain-training apps—Lumosity and Elevate—by applying the 14-factor DEGREE framework as a structured UX evaluation tool and using fuzzy scoring to improve interpretability. We recruited 190 Computer Science undergraduates; each participant evaluated both apps, yielding 380 app evaluations using a counterbalanced two-sheet questionnaire. Fourteen factors covering usability, engagement, and perceived learning were rated on a five-point Likert scale. Reliability was strong for both apps (Cronbach's α = 0.822 for Lumosity; 0.847 for Elevate). Descriptive results showed mid-to-high perceptions overall, with mean scores of 3.51 (Lumosity) and 3.44 (Elevate). Fuzzy aggregation transformed subjective ratings into a 0–1 index, producing overall scores of 0.503 (Lumosity) and 0.490 (Elevate), indicating a small global advantage for Lumosity. At the factor level, Lumosity was slightly higher on most DEGREE dimensions, whereas Elevate showed relative advantages on Learnability and Confidence, suggesting potential benefits for early onboarding and self-efficacy. Overall, the proposed DEGREE–Fuzzy pipeline yields a transparent, reproducible benchmark that translates multi-factor perceptions into decision-ready recommendations for selecting apps aligned with instructional goals.

Keywords : *Brain-training apps, DEGREE, Fuzzy scoring, HCI, Usability evaluation*

This work is an open access article licensed under a Creative Commons Attribution 4.0 International License.

1. INTRODUCTION

Digital learning transformations have accelerated the integration of serious games and cognitive-training apps into classrooms and self-practice programs, as they offer adaptive, structured micro-exercises that are readily accessible on mobile devices. Recent reviews and controlled trials indicate that game-based interventions can improve attention, working memory, and executive functions when the user experience (UX) is well designed and aligned with specific instructional goals [1],[2]. In this context, educators require objective, concise, and auditable benchmarks to ensure that tool selection aligns with curricular aims—consistent with ongoing developments in ed-tech/XR benchmarking and contemporary usability evaluations of online learning applications [3],[4].

Among many choices, Lumosity and Elevate stand out as popular cognitive training apps; however, as the marketplace becomes increasingly diverse, educators need objective, concise, and auditable benchmarks to choose tools that fit their curricular needs. Although both apps pursue similar overarching goals, their design philosophies differ: some features privilege efficient onboarding and workflows, whereas others emphasise variety of mini-games and moment-to-moment motivation to sustain engagement [5],[6],[7]. These differences call for measurable, auditable justification through an evaluation framework with standard indicators and a scoring function that enables fair, transparent cross-app comparison—echoing studies on decision frameworks for educational software selection and usability benchmarking in recent mobile-learning research [3],[8].

P-ISSN: 2723-3863 https://jutif.if.unsoed.ac.id
E-ISSN: 2723-3871 DOI: https://doi.org/10.52436/1.jutif.2026.7.2.5646

To meet that need, this study adopts DEGREE (14 factors) as a credible evaluation tool that bridges usability dimensions—Aesthetics, Learnability, Operability, Accessibility, and User Error Protection—with player/learning experience—Focused Attention, Fun, Challenge, Social Interaction, Confidence, Relevance, Satisfaction, Perceived Learning, and Control. The coverage is therefore non-partial and easily communicated across stakeholders, in line with modern evaluations that combine usability/UX, player experience, and learning outcomes within a single audit-ready frame for educational games/apps [9],[10],[11],[12]. In this study, DEGREE is not the research object per se, but a systematic benchmarking instrument that translates user perceptions into factor-level decision indicators.

While Likert scales are standard in UX measurement, linear aggregation often compresses nuance—for example, the perceptual distance between "fair" and "good"—that matters for instructional decisions. We therefore enrich the analytics with Fuzzy Logic: Likert scores are mapped to Triangular Fuzzy Numbers (TFN) via Triangular Membership Functions (TMF), aggregated into a Fuzzy Weighted Average (FWA), and defuzzified (centroid) to yield stable 0–1 indices and linguistic labels (Low/Medium/High) using the Mean-of-Maxima (MoM) rule. This approach preserves meaningful ambiguity without sacrificing transparency or auditability, allowing results on the 14 DEGREE factors to be communicated clearly to diverse stakeholders [9],[13],[14]. In short, the DEGREE–Fuzzy integration maps Likert ratings to linguistic labels through membership functions and produces interpretable, noise-robust 0–1 indices for academic decision-makers via defuzzification (e.g., centroid/MoM) and weighted aggregation—consistent with modern fuzzy-Likert practice and fuzzy comprehensive evaluation in higher-education decision making [9],[15],[14].

Beyond content, successful adoption of learning apps depends on learnability, operational efficiency, and mobile accessibility. Recent HCI/mHealth work shows that clear microcopy, error prevention/protection, and informative feedback correlate with self-efficacy, engagement, and retention—especially during onboarding, when users form initial mental models and confidence to continue. Contemporary mHealth heuristics also include "error prevention" and "help users recognise, diagnose, and recover from errors," alongside principles of efficiency and clarity of instruction as keys to early-experience quality. These findings align with systematic reviews linking usability and engagement to eHealth/mHealth effectiveness and with development studies that emphasise real-time feedback as a driver of self-efficacy at first use [16],[17],[18]. Together, they underscore the need for indicators that can distinguish between operational/learning strengths and engagement/affective strengths, enabling selection decisions that align with teaching priorities.

A growing body of randomised trials and feasibility studies shows that cognitive app/game interventions are effective when UX design and challenge intensity are managed as a bundle rather than focusing solely on training content. Evidence spans diverse populations—from older adults to university students—using behavioural and cognitive outcome measures [1],[19],[20]. In higher education specifically, several studies report that optimising learnability and operability accelerates early adoption, while variety and adaptive challenge sustain engagement in repeated sessions [21],[17].

Building on that foundation, this study compares Lumosity and Elevate, positioning DEGREE as the primary measurement tool to reveal a factor-level performance map that can be translated directly into classroom policy. Rather than naming a single global "winner," we emphasise goal-based recommendations: which app better supports onboarding & control, and which better serves as an engagement & focus booster during reinforcement sessions [19]. This orientation ensures that outputs are actionable for instructors, programs, and end users.

By design, participants are Computer Science undergraduates who have completed core courses (logic, algorithms & programming, data structures, HCI) to ensure sufficient sensitivity to DEGREE dimensions. Quality control measures—anti-duplication, straight-lining detection, and reasonable

P-ISSN: 2723-3863
https://jutif.if.unsoed.ac.id
E-ISSN: 2723-3871
DOI: https://doi.org/10.52436/1.jutif.2026.7.2.5646

duration—were applied to safeguard data quality. At the same time, internal consistency was assessed with Cronbach's α in line with current psychometrics/HCI practice [21]. Data were organised per app, allowing descriptive and fuzzy analyses to be conducted symmetrically within each app and across factors.

The analytic pipeline proceeds in layers: (1) descriptive profiling (Mean & SD) for an intuitive summary, (2) fuzzification with TMF to capture nuance on the 1–5 scale, (3) defuzzification combining MoM for dominant labels and FWA as a rankable 0–1 index, and (4) a factor-level decision map (Δ and decision strength) to deliver classroom-ready recommendations without heavy interpretation [14],[22]. With this structure, results do not stop at "numbers" but are converted into practical rules that can be standardised in teaching SOPs.

The study's contribution is twofold. First, it provides an auditable benchmark that highlights a close overall match between the two apps, with Lumosity showing a small global advantage and Elevate showing relative strengths on Learnability and Confidence. Second, it demonstrates a transparent DEGREE–Fuzzy pipeline that converts multi-factor UX perceptions into interpretable 0–1 indices and factor-level recommendations for selecting brain-training apps aligned with instructional goals. This standard includes a factor-wise winners map to inform content sequencing, session duration, and the adaptive ordering of games [14]. Institutions can thereby systematically and reproducibly align app choices with curricular goals and class profiles.

Ultimately, the study compares two games using DEGREE as a credible measurement tool to deliver practical benefits: more accurate selection decisions, smoother learning experiences, and measurable impacts on students' engagement and perceived learning. Supported by recent literature across HCI, ed-tech, and fuzzy MCDM, this framework provides a transparent foundation for adopting learning apps aligned with the needs of contemporary higher education [19].

Accordingly, this study addresses two research questions: RQ1: Which DEGREE factors differentiate Lumosity and Elevate most clearly in a controlled, same-participant comparison? RQ2: Can fuzzy aggregation provide a more interpretable, decision-ready summary (0–1 index and factor-wise winners) than direct linear aggregation of Likert means?

2. METHOD

This research was designed as a descriptive–comparative quantitative study that combines the DEGREE model with a Fuzzy Weighted Average (FWA) approach to evaluate and compare the quality of two brain-training applications—Lumosity and Elevate—from the user experience perspective. The framework was chosen because it captures the inherent ambiguity of Likert-type ratings and converts them into more stable, interpretable metrics [9],[23].

2.1. Objects and Participants

The evaluation objects are two popular cognitive-training apps, Lumosity and Elevate, which were measured using the DEGREE instrument, comprising 14 factors. Participants were 190 undergraduate Computer Science students who had completed core courses in Logic, algorithms and programming, Data Structures, and human–computer interaction (HCI). Each participant evaluated both applications in a counter-balanced order, yielding 380 app evaluations. Representative screenshots of Lumosity and Elevate are provided in Figure 1.

2.2. Research Instrument

The evaluation instrument was adapted from the DEGREE model, encompassing 14 constructs: Aesthetics, Learnability, Operability, Accessibility, User Error Protection, Focused Attention, Fun, Challenge, Social Interaction, Confidence, Relevance, Satisfaction, Perceived Learning, and Control.

P-ISSN: 2723-3863
https://jutif.if.unsoed.ac.id
E-ISSN: 2723-3871
DOI: https://doi.org/10.52436/1.jutif.2026.7.2.5646

Each construction was measured with two items on a 5-point Likert scale (1 = strongly disagree to 5 = strongly agree). The questionnaire was delivered as two separate rating sheets (Lumosity and Elevate) in a counter-balanced order to minimise sequence bias. A concise summary of the questionnaire is provided in Figure 2.

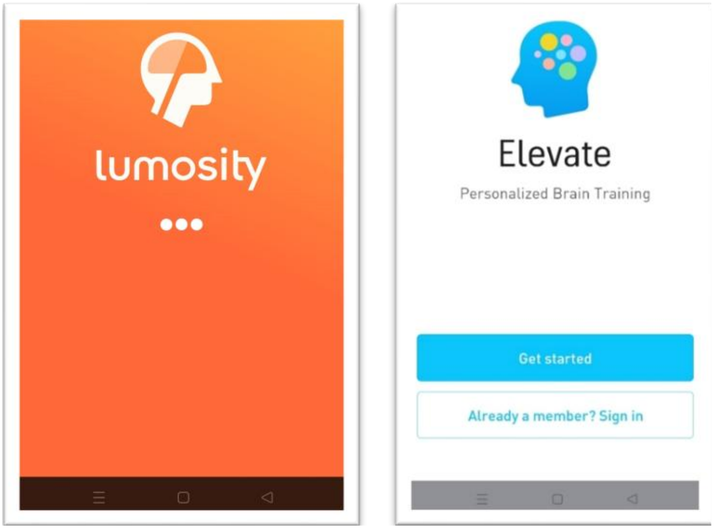

Figure 1. Screenshots of Lumosity and Elevate

C. Filling Instructions

Please mark the number 1–5 according to your level of agreement with each statement.
Scale: 1= Strongly Disagree, 2= Disagree, 3= Neutral, 4= Agree, 5= Strongly Agree.

1	2	3	4	5

D. Rating for LUMOSITY

Please mark 1–5 for each of the following statements regarding your experience using LUMOSITY.

Aesthetics

Statement	1	2	3	4	5
The game interface is attractive and consistent.	☐	☐	☐	☐	☐
The color scheme,	☐	☐	☐	☐	☐

Figure 2. DEGREE Questionnaire

Data collection was online via Google Forms in 2025, with two-level QC (anti-duplication token, straight-lining detection, reasonable completion time, and per-sheet completeness checks). In total, 190 respondents evaluated both apps, resulting in 380 app evaluations. Likert scores were then mapped to Triangular Fuzzy Numbers (TFN) on a 5-level grid within [0–1] and defuzzified (centroid) to obtain 0–1 fuzzy score for each factor–app combination.

Inclusion criteria were: (i) prior exposure to both apps for ≥3 sessions; (ii) age ≥ 18 years; (iii) informed consent and willingness to complete the online survey. Exclusion criteria included duplicate entries, incomplete responses, straight-lining, and unreasonable completion times, as per QC. This design ensures respondents are discipline-relevant (sensitive to learnability, operability, and control) and that comparisons across the two apps are fair within an informatics learning context. Inclusion criteria are summarised in Table 1.

Vol. 7, No. 2, April 2025, Page. 1800-1817
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.2.5646

Table 1. Inclusion and Exclusion Criteria

Category	Criterion	Operational Definition / Screening Question	Rationale / Source
Inclusion	Informatics undergraduates	Active status in the Informatics program	Target population for HCI evaluation in academic settings
Inclusion	Completed (Logic; Algorithms & Programming)	Verified via profile form	Ensures competence aligns with DEGREE dimensions
Inclusion	Tried Lumosity & Elevate ≥ 3 sessions	Screening items: "Lumosity ≥ 3 sessions"; "Elevate ≥ 3 sessions"	Guarantees minimum exposure to both apps
Inclusion	Age ≥ 18 years	Age (years) ≥ 18	Ethics compliance
Inclusion	Informed consent	Consent checkbox at survey start	Ethics compliance
Exclusion	Duplicate entry	Unique token/ID; timestamp/IP (optional)	Data quality (anti-duplicate)
Exclusion	Incomplete / straight-lining	All required items filled; reasonable duration	Data quality (attention & completeness)

2.3. DEGREE Model

The DEGREE model is an evaluation tool that balances usability and player/learning experience for educational applications. It comprises 14 factors: Aesthetics, Learnability, Operability, Accessibility, User Error Protection, Focused Attention, Fun, Challenge, Social Interaction, Confidence, Relevance, Satisfaction, Perceived Learning, and Control, as shown in Figure 3 [9].

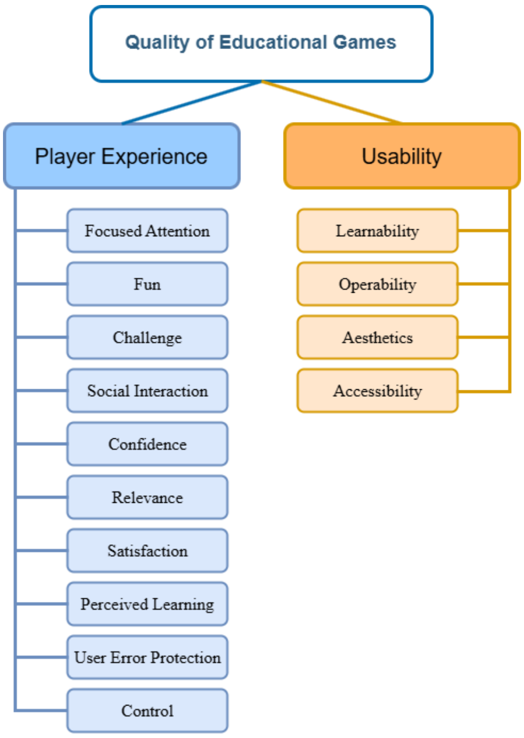

Figure 3. DEGREE Model

The first four factors represent usability (ease of learning, operation, access, and error protection), while the remainder capture engagement, motivation, relevance, perceived learning, and user control. In this study, DEGREE is positioned as a benchmarking instrument, not the subject of inquiry—thus all

Vol. 7, No. 2, April 2025, Page. 1800-1817
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.2.5646

Lumosity vs. Elevate comparisons derive from DEGREE scores (Likert 1–5 → TFN [0–1] → defuzzification), yielding interpretable, decision-ready metrics.

2.4. Quality CONTROL

Data quality was ensured through two-level screening, including duplicate prevention, completeness checks, straight-lining detection, and filtering for unreasonable completion times. The full QC rules are reported in Table 2.

Table 2. Quality Control Rules

Level	**Check**	**Rule/Threshold**	**Action**
Participant	De-duplication	Unique token (1 token = 1 respondent)	Remove duplicates
Participant	Total duration	Outlier <P5 or >P95 (IQR-based)	Review → remove if unreasonable
Participant	Global consistency	Anomalous pattern (all 1/5, zigzag)	Review → remove if invalid
Instrument (per sheet)	Straight-lining	≥80% identical consecutive answers	Remove the sheet
Instrument (per sheet)	Completeness	Item missingness = 0	Remove the incomplete sheet
Instrument (cross-sheet)	Counter-balanced order	Order must match design	Remove if the order is wrong

2.5. Reliability Test: Cronbach's Alpha

Reliability testing ensured that each dimension in the DEGREE model (14 factors) showed adequate internal consistency. We used Cronbach's alpha, a standard approach for evaluating the reliability of Likert-type instruments. In practice, α ≥ 0.70 is acceptable, while α ≥ 0.80 indicates good reliability. This study computed α separately per app (Lumosity and Elevate) at the composite (across-factor) level. The highest and most stable α values guided the final 2-item-per-factor configuration, which was carried into subsequent analyses. Internal consistency was computed using Equation (1).

$$\alpha = \frac{k}{k-1}\left(1 - \frac{\sum_{i=1}^{k} S_i^2}{S_t^2}\right) \quad (1)$$

where k is the number of items, (S_i^2) = item i, and (S_t^2) = the variance of total scores. Interpretation: α ≥ 0.90 (excellent), 0.70–0.89 (adequate–good), < 0.70 (needs revision). In our dataset, α {Lumosity} = 0.822 and α {Elevate} = 0.847, indicating good internal consistency. This method is standard in UX evaluations across health/education domains [24], [25].

2.6. Descriptive Profiling

At this stage, each DEGREE factor (14) × app is summarised with the mean and sample standard deviation (SD) on the 1–5 Likert scale, following contemporary reporting practices in reputable journals that explicitly present Mean ± SD for Likert responses [26],[27],[28]. This facilitates a quick view of response tendencies before fuzzy analysis and aligns with current methodological guidance on interpreting and reporting Likert data (Mean, SD, CI) [29]. We report two decimal places for both the mean and the SD. The mean for a factor–app pair with respondent scores $x_1, \ldots, x_n$(n respondents) is given by Equation (2).

Vol. 7, No. 2, April 2025, Page. 1800-1817
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.2.5646

$$\bar{x} = \frac{1}{n}\sum_{j=1}^{n} x_j \tag{2}$$

The mean estimates the central tendency for that factor on a 1–5 scale. The practice of reporting Mean ± SD for Likert responses is widely used in recent psychometrics and education research and is explicitly documented in Frontiers and JMIR articles [30],[31],[32]. The sample SD (denominator $n - 1$) we use follows open-access statistical references, see Equation (3); cf. C. Peter [33].

$$s = \sqrt{\frac{1}{n-1}\sum_{j=1}^{n}\left(x_j - \bar{x}\right)^2}. \tag{3}$$

The SD summarises dispersion; a smaller SD indicates more homogeneous perceptions on that factor. Reporting Mean–SD is recommended as a concise and informative summary for Likert-type data [34]. An overall mean (across all 14 factors) is used as a compact indicator of app-level experience, computed as the average across all factors in Equation (4).

$$\bar{x}_{\text{overall}} = \frac{1}{K}\sum_{k=1}^{K}\bar{x}_k\,,\ K = 14, \tag{4}$$

Here, $\bar{x}_k$is the meaning of factor k. This metric supports global app comparisons before drilling down to the factor level [34]. In factor-wise descriptive calculations, $x_j$denotes the Likert score (1–5) from respondent j, and $n\ is$ the count of valid respondents (in this study, $n = 190$ per app). Thus, $\bar{x}$ and s describe factor-level tendencies and spread, whereas $\bar{x}_{\text{overall}}$ summarises performance across all factors.

2.7. Fuzzification (TMF)

Fuzzification converts 1–5 Likert scores into a smoother fuzzy representation before aggregation and defuzzification. Each response $y \in \{1,2,3,4,5\}$is first normalised to the continuous domain [0,1] [35]; see Equation (5):

$$x = \frac{y-1}{4} \tag{5}$$

Membership degrees are then computed using a **Triangular Membership Function (TMF)** with parameters (a, b, c)denoting the left bound, modal peak, and right bound [36]. See **Equation (6)**.

$$\mu(x) = \begin{cases} 0, & x\ \leq a\ or\ x\ \geq c, \\ \frac{x-a}{b-a}, & a\ < x \leq b, \\ \frac{c-x}{c-b}, & b\ < x < c. \end{cases} \tag{6}$$

In this study, each Likert level is mapped to a **Triangular Fuzzy Number (TFN)** on **[0,1]** (an audit-friendly standard grid):

Likert 1: (0.0,0.0,0.2), Likert 2: (0.0,0.2,0.4),
Likert 3: (0.2,0.4,0.6), Likert 4: (0.4,0.6,0.8), Likert 5: (0.6,0.8,1.0).

For a normalised value x, the degree $\mu(x)$for a given level is obtained by substituting that level's TFN parameters (a, b, c)into (6) (linear rise from $a$to b, linear fall from b to c, and 0 outside $[a, c]$). Notation: y= raw Likert score; x= normalized score; $\mu(x)$= membership degree (0–1); (a, b, c)= TFN parameters. After projecting each response into TFN space, the factor-wise TFN $(\ell, m, u)$is averaged per factor × app, then defuzzified (centroid) to yield 0–1 scores for comparing Lumosity vs. Elevate.

2.8. Defuzzification (Centroid & Mean of Maxima)

Following aggregation into an average TFN $(\bar{\ell}, \bar{m}, \bar{u})$for each factor, the crisp 0–1 score is computed using the centroid method (primary index), Equation (7):

$$s = \frac{\bar{l}+\bar{m}+\bar{u}}{3} \in [0,1]. \quad (7)$$

The centroid represents the center of mass of the fuzzy output set and is widely used because it is stable and auditable [39]. For a compact linguistic label (Low/Medium/High), we apply Mean of Maxima (MoM) rule over three fuzzy sets $k \in \{L, M, H\}$ with triangular membership functions $\mu_k(x)$ defined by the TMF scheme [37],[38]. The maximum membership at x selects the label $= s$, Equation (8).

$$k^* = \underset{k\in\{L,M,H\}}{\arg\max}\ \mu_k(s). \quad (8)$$

If a tie occurs (two or more labels share the same maximum $\mu_k(s)$), MoM defines the set of maxima as in Equation (8′) and computes the average of peak points as in Equation (9) [38].

$$X_k^* = \{\ x \in [0,1] \mid \mu_k(x) = \mu_k^{\max}\ \}, \quad (9)$$

$$\mathrm{MoM}(k) = \frac{1}{|X_k^*|} \sum_{x \in X_k^*} x. \quad (10)$$

In practice, when a tie occurs at $x = s$, we adopt a conservative tie-break: if Medium is among the candidates, assign Medium; otherwise, choose the label nearest to s(e.g., High if $s \geq 0.66$, Low if $s < 0.50$). In short, the centroid provides a stable 0–1 index for comparative analysis, while MoM yields communicable linguistic labels consistent with the predefined TMF curves.

2.9. Decision Map

This stage identifies the winner per factor and the decision strength based on the difference of defuzzified (centroid) scores between Lumosity (L) and Elevate (E). Using differences of defuzzified indices as the basis for selection/ranking aligns with MCDA practice (e.g., fuzzy TOPSIS families), where decisions are derived from relative closeness to an ideal solution and the difference between closeness values clarifies preference [39]. For each factor f, define the margin in Equation (11).

$$\Delta_f = s_f^{(L)} - s_f^{(E)}, s_f^{(\cdot)} \in [0,1], \quad (11)$$

Where $s_f^{(L)}$and $s_f^{(E)}$is fuzzy score 0–1 (centroid) on factor f. The winner per factor is Determined by the sign Δ_f: lumosity if $\Delta_f > 0$, lumosity if $\Delta_f < 0$, and tie if $\Delta_f = 0$. For decision strength, a practical threshold based on the magnitude of the difference is used with the formula (12).

$$Strength_f = \begin{cases} Narrow, & \mid \Delta f \mid < 0.015, \\ Moderate, & 0.015 \leq \mid \Delta f \mid < 0.030, \\ Strong, & \mid \Delta f \mid \geq 0.015, \end{cases} \quad (12)$$

These Narrow/Moderate/Strong bands are pragmatic, defined on the 0–1 scale to attach practical meaning to $\mid \Delta_f \mid$ and facilitate communication. Setting a priori margins/ "acceptable advantage" as a

Vol. 7, No. 2, April 2025, Page. 1800-1817
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.2.5646

reporting rule is recommended in MCDA, provided thresholds and sensitivity checks are reported transparently. Optionally, the margin can be expressed as a percentage $m_f = 100 \cdot | \Delta_f |$for complementary reporting. The decision map is then presented as a compact table/figure listing $(\Delta_f, Winner_f, \text{Strength}_f$ for all 14 factors, making it clear where each app leads and how convincing the lead is. A global summary (e.g., win–loss–tie counts and/or Overall FWA) can be added to maintain linkage between factor-level results and program-level decisions [40].

2.10. Statistical Analysis

Because each participant rated both applications, all comparisons were treated as paired. In addition to descriptive profiling and fuzzy scoring, we tested whether Lumosity and Elevate differed statistically using a two-tailed Wilcoxon signed-rank test ($\alpha = 0.05$) for (i) the overall Likert mean (mean across 14 factors) and (ii) the overall fuzzy index (Overall FWA, 0–1). For factor-wise comparisons, the same paired test can be applied across the 14 factors; when multiple tests are reported, p-values should be adjusted using a family-wise procedure (e.g., Holm correction) to control false positives. Effect size was reported using rank-biserial correlation (r) to indicate the magnitude of the paired difference.

2.11. Research Workflow

The study workflow begins with data collection from 190 participants who evaluated two apps (a total of 380 app evaluations) using a two-sheet questionnaire (Lumosity–Elevate) in a counterbalanced order, along with the 14-factor DEGREE instrument (Likert scale, 1–5). Data were filtered through two-level QC (anti-duplication, straight-lining detection, reasonable duration, per-sheet completeness). We then tested reliability (Cronbach's α) at the composite level, followed by descriptive profiling (Mean & SD per factor × app). Likert scores were then fuzzified (5-point → TFN) and defuzzified (centroid) to yield 0–1 scores per factor. Finally, we constructed a decision map based on Δ between apps and decision strength (Narrow/Moderate/Strong) to determine factor-level winners and a global summary—the basis for HCI/learning recommendations. See Figure 4 for the research flow diagram.

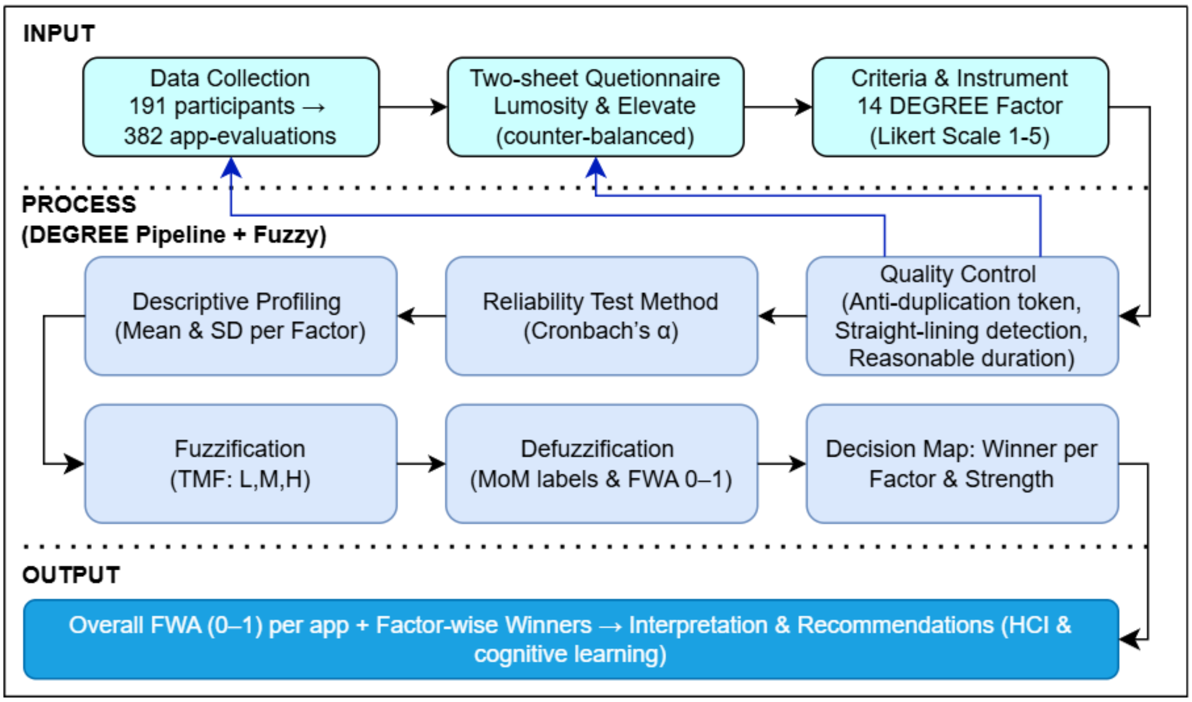

Figure 4. Research Flow Diagram

Vol. 7, No. 2, April 2025, Page. 1800-1817
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.2.5646

3. RESULT

3.1. Dataset and Reliability

A total of 190 undergraduate Informatics students participated; each evaluated both applications (Lumosity and Elevate), resulting in 380 app evaluations. The data were organised into two symmetric app sheets, each containing 190 respondent evaluations (one evaluation per respondent per app). The instrument's internal consistency (composite across the 14 DEGREE factors) was suitable for both apps, with Cronbach's α = 0.822 for Lumosity and 0.847 for Elevate, as summarised in Table 3.

Table 3. Cronbach's Alpha Summary.

Application	Respondents (n)	Scale	Cronbach's α	Interpretation
Lumosity	190	Composite across 14 DEGREE factors	0.822	Good
Elevate	190	Composite across 14 DEGREE factors	0.847	Good

Table 3 confirms that the DEGREE instrument achieves good internal consistency on both apps. The values α_{Lumosity} = 0.822 and α_{Elevate} = 0.847 indicate stable inter-item coherence for the same 14-factor construct set. Using ≥0.80 as "good," both values provide a sufficient psychometric basis for proceeding with comparative analysis—at both the descriptive level (Mean–SD) and the fuzzy level (0–1 scores). Given strong reliability, subsequent score variation is more credibly attributed to users' perceived differences than to measurement noise.

3.2. Descriptive Statistics (Likert 1–5)

Overall, the mean across all fourteen factors indicates that Lumosity (3.51) is slightly higher than Elevate (3.44) (Table 4). At the factor level, Lumosity is marginally higher on most dimensions, while Elevate shows advantages on Learnability and Confidence. This descriptive profile provides an intuitive baseline before the fuzzy scoring and decision-map analysis. At the factor level, Lumosity tended to score higher on most dimensions see Table 4 for details.

Table 4. Descriptive Statistics.

No	Factor	Lumosity (Mean±SD)	Elevate (Mean±SD)
1	Aesthetics	3.50 ± 0.99	3.45 ± 0.99
2	Learnability	3.50 ± 1.00	3.52 ± 0.96
3	Operability	3.46 ± 0.99	3.41 ± 1.01
4	Accessibility	3.52 ± 0.98	3.44 ± 1.02
5	User Error Protection	3.58 ± 1.05	3.34 ± 1.03
6	Focused Attention	3.52 ± 0.99	3.44 ± 0.99
7	Fun	3.54 ± 1.04	3.45 ± 0.99
8	Challenge	3.52 ± 0.99	3.43 ± 1.01
9	Social Interaction	3.48 ± 0.98	3.38 ± 0.99
10	Confidence	3.48 ± 0.99	3.52 ± 0.98
11	Relevance	3.53 ± 1.02	3.45 ± 0.99
12	Satisfaction	3.49 ± 1.01	3.43 ± 0.98
13	Perceived Learning	3.54 ± 1.00	3.46 ± 1.02
14	Control	3.49 ± 1.04	3.44 ± 0.97
	Overall (Mean of 14 factors)	3.51	3.44

Vol. 7, No. 2, April 2025, Page. 1800-1817
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.2.5646

Factor-wise winners (Likert means) were derived from Table 4. Lumosity showed higher mean scores across 12 of the 14 DEGREE factors, whereas Elevate led in two—Learnability and Confidence. This indicates that, within this cohort, Lumosity was perceived slightly better across most usability–experience dimensions, while Elevate showed a relative advantage in perceived ease of learning and self-efficacy.

The pattern mirrors Table 4: Lumosity is higher on 12 of the 14 factors, while Elevate is higher on Learnability and Confidence. The factor-wise comparison is visualized in Figure 5 (bar chart) and Figure 6 (radar plot) to provide a quick overview of the relative profiles.

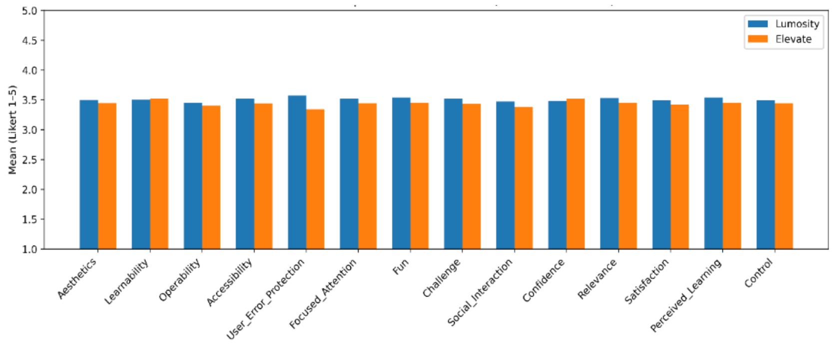

Figure 5. Factor-wise Comparison of Mean Scores (14 DEGREE Factors Lumosity vs. Elevate).

To complement the bar chart comparison, a radar plot is provided to visualise the 14-factor profile simultaneously. The radar view highlights that Lumosity exhibits a slightly broader advantage across most usability–experience factors, whereas Elevate's relative strengths are concentrated on Learnability and Confidence. The relative profiles of both apps are summarised in Figure 6.

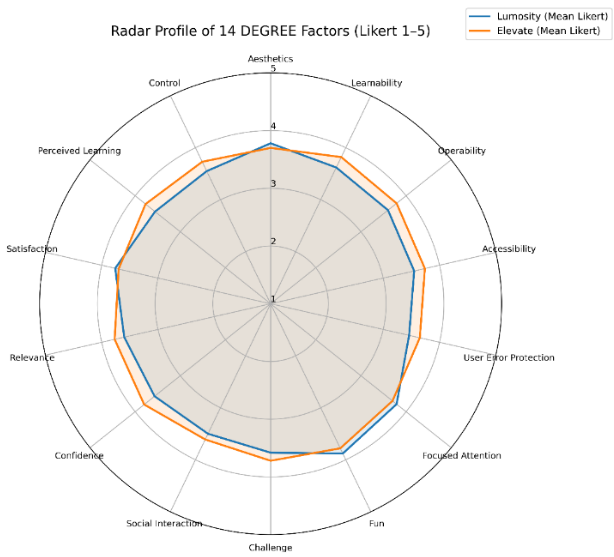

Figure 6. Radar Profile of 14 DEGREE Factors (Lumosity vs. Elevate).

3.3. Fuzzy Scoring (TFN from 5-Point Likert → Defuzzification 0–1)

Likert scores were mapped to 5-level TFN on [0–1], aggregated per factor–app, and defuzzified (centroid) into 0–1 scores. For readability, we also provide linguistic labels (Low/Medium/High) using bands: Low < 0.50; Medium 0.50 – <0.66; High ≥ 0.66.

Vol. 7, No. 2, April 2025, Page. 1800-1817
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.2.5646

For additional interpretability to non-technical readers, the 0–1 fuzzy index can be read as a practical quality band: values around 0.50 indicate a fair perception, values around 0.60 indicate good, and values ≥ 0.66 indicate very good perceived quality in the corresponding factor, consistent with the Low/Medium/High linguistic mapping used in this study. The fuzzy scoring results in Table 5 mirrored the descriptive pattern. Lumosity obtained higher fuzzy scores in most factors, while Elevate's advantages were concentrated in Learnability and Confidence. However, most gaps fell into the Narrow band, suggesting small factor-level differences.

Table 5. Fuzzy Scoring

Factor	Lumosity Score	Lumosity Label	Elevate Score	Elevate Label
Aesthetics	0.501228	Medium	0.491053	Low
Learnability	0.501404	Medium	0.505614	Medium
Operability	0.492281	Low	0.482982	Low
Accessibility	0.505263	Medium	0.49	Low
User_Error_Protection	0.517193	Medium	0.470351	Low
Focused_Attention	0.504912	Medium	0.490351	Low
Fun	0.509123	Medium	0.491579	Low
Challenge	0.505263	Medium	0.488947	Low
Social_Interaction	0.496491	Low	0.478421	Low
Confidence	0.497895	Low	0.505439	Medium
Relevance	0.507544	Medium	0.491579	Low
Satisfaction	0.500175	Medium	0.486316	Low
Perceived_Learning	0.508772	Medium	0.492982	Low
Control	0.500351	Medium	0.490351	Low

The TMF visualisation clarifies how mid-range Likert responses are translated into Low–Medium membership degrees under soft boundaries. In this dataset, most defuzzified scores fall around 0.47–0.52, so the mapping primarily supports stable, comparable indexing rather than producing High labels across factors.

As a companion, the TMF figure illustrates Triangular Membership Functions on the 1–5 scale (peaking at 2–3–4), clarifying how Likert values accumulate Low–Medium–High degrees gradually (with soft boundaries). The visualisation helps interpret 0–1 fuzzy scores in Table 5—for instance, In this dataset, scores are mostly concentrated around 0.47–0.52, so the mapping primarily supports stable indexing rather than frequent Medium-to-High transitions. See Figure 7 for TMF.

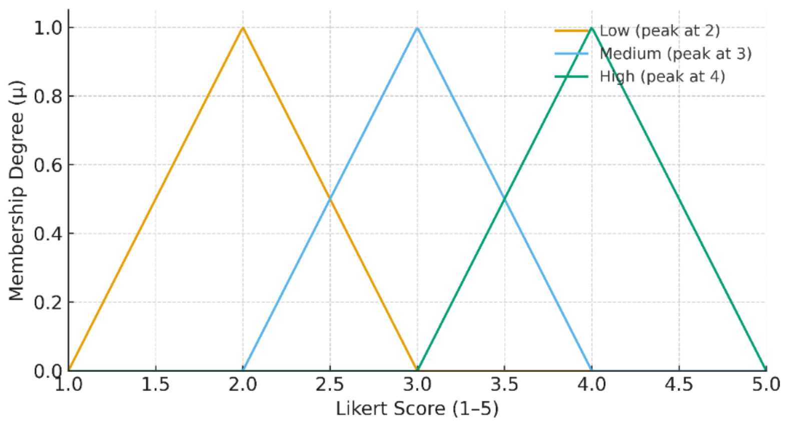

Figure 7. Triangular Membership Functions (TMF)

Vol. 7, No. 2, April 2025, Page. 1800-1817
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.2.5646

In the next step, factor-wise scores were mapped to a decision profile to support goal-based recommendations: Elevate can be prioritised when learnability and confidence building are the main needs, whereas Lumosity can be emphasised when broader experience consistency and engagement-related dimensions (e.g., focused attention and fun) are desired.

3.4. Global Fuzzy Score

A small difference is observed in the overall fuzzy scores, with Lumosity (0.503) slightly higher than Elevate (0.490), as shown in Table 6.

Table 6. Overall Fuzzy Score (FWA 0–1) for Lumosity and Elevate

Application	Overall Mean (Likert)	Overall FWA (0-1)
Lumosity	3.51	0.503
Elevate	3.44	0.490

A paired Wilcoxon signed-rank test indicated that this difference was not statistically significant ($Z = 1.132$, $p = 0.258$, $r = 0.058$), suggesting that the global gap should be interpreted as a small numerical variation rather than a consistent preference within the same-participant comparison. For clarity, the side-by-side bar comparison of Overall FWA is shown in Figure 8.

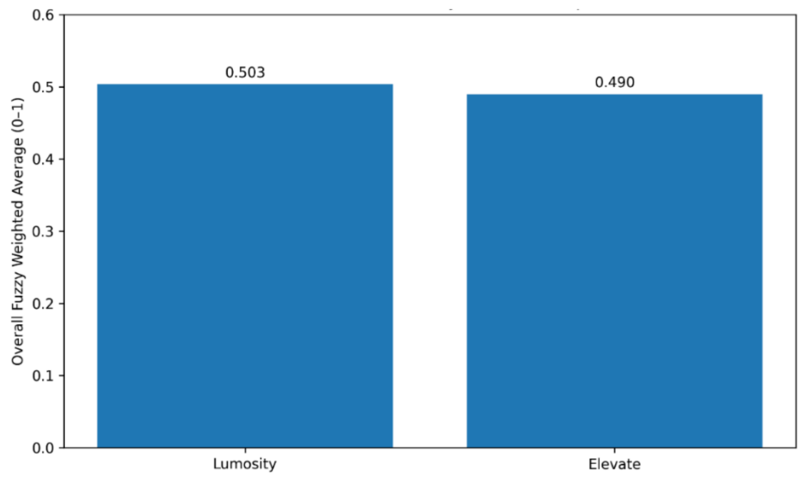

Figure 8. Overall FWA (0–1): Lumosity vs. Elevate.

3.5. Factor-Level Decision Map (Δ & Decision Strength)

Using $\Delta = FWA\_\{Lumo\} - FWA\_\{Elev\}$ and the decision-strength thresholds (Narrow < 0.015; Moderate 0.015–<0.030; Strong ≥ 0.030), Lumosity leads on 12 of 14 factors, whereas Elevate leads on 2 factors (Learnability and Confidence). Most factor-wise gaps fall within the Narrow band, indicating that although the direction of differences is observable, the magnitudes are generally small at the factor level.

3.6. Brief Interpretation

Elevate is recommended for early onboarding support, particularly where learnability and confidence building are prioritised. Lumosity is suitable as an engagement booster and shows stronger performance across most factors, including user error protection and perceived learning, which can help sustain focus and practice quality in repeated sessions. This enables instructors to design a goal-based strategy aligned with curricular needs and class profiles.

Vol. 7, No. 2, April 2025, Page. 1800-1817
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.2.5646

4. DISCUSSIONS

4.1. Overview of Findings and Meaning.

The results indicate a largely similar UX profile across the two apps, with Lumosity leading on 12/14 factors and Elevate leading on 2/14 factors (Learnability and Confidence). At the global level, the fuzzy index shows a slight advantage for Lumosity (Overall FWA: 0.503 vs 0.490), but the paired test indicates that this difference is not statistically significant, so it should be interpreted as a small numerical difference rather than a strong preference. The pattern aligns with the literature, which shows that the effectiveness of digital interventions depends on content and experience design that supports engagement, self-efficacy, and stepwise progress [41].

4.2. HCI Perspective: Onboarding, Control, and Engagement

Elevate's relative strengths on Learnability and Confidence suggest clearer onboarding cues and stronger perceived self-efficacy at early use. Meanwhile, Lumosity shows a slightly broader advantage across usability–experience factors (Table 4), including Operability and User Error Protection, which may support smoother repeated-session interaction, consistent with current HCI/mHealth usability practices [21]. Lumosity's higher scores on User Error Protection suggest stronger perceived support for error prevention/recovery, and clearer feedback during interaction. Meanwhile, the small gaps on Accessibility indicate that both apps are broadly usable on mobile devices, with only minor differences in perceived access-related experience, both of which directly contribute to stable mobile experiences [17]. Conversely, Lumosity maintains high levels of Focused Attention and Fun; the variety of mini-games and moment-to-moment rewards support short-burst engagement, which is necessary to sustain attention and provide immediate satisfaction [5].

4.3. Methodological Contribution: DEGREE as a Fuzzy "Tool" for Decisions

This study positions the DEGREE (14-factor) instrument as a credible measure spanning usability, experience, and learning, and then converts it into decision indicators using Fuzzy Logic. TMF modelling (Low–Medium–High) preserves perceptual nuance on the 1–5 scale; MoM provides communicable linguistic labels; and FWA (0–1) delivers interpretable, noise-robust indices—aligned with recent fuzzy/MCDM developments (intuitionistic, OWA, hesitant, generalised aggregation) [14]. Thus, differences that appear "small" in Likert averages become a decision map with strength (Narrow/Moderate/Strong) that can be used immediately in the classroom [42].

4.4. Implications: A "Dual-Track" Strategy

Implementation guidance can follow a goal-based strategy: use Elevate in introductory phases when rapid comprehension and confidence building are prioritised (Learnability and Confidence), and use Lumosity in reinforcement sessions to sustain engagement and interaction quality, supported by its slightly broader advantages across factors such as Operability, User Error Protection, Control, and perceived learning reinforcement (Table 4) [43]. At the program level, the per-factor decision map (Δ and decision strength) can serve as a dashboard to tune app choice based on factors such as course, session duration, game order, and task differentiation by class profile [21].

4.5. Experience Quality and Self-Efficacy

A practical link between onboarding clarity and learning confidence is reflected in Elevate's relative strengths on Learnability and Confidence. In contrast, Lumosity shows a slightly broader advantage across repeated-use experience factors, including Operability, User Error Protection, Control, and Perceived Learning (Table 4). Together, this profile suggests that Elevate may support earlier comprehension and self-efficacy, while Lumosity may better sustain interaction quality and perceived

Vol. 7, No. 2, April 2025, Page. 1800-1817
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.2.5646

learning reinforcement during ongoing practice. This mirrors mHealth/usability evaluations in which clear flows, error prevention, and informative feedback strengthen self-efficacy and sustained adoption [17]. At the same time, strong affective engagement (Fun/Satisfaction) in Lumosity acts as an attention booster, which, according to learning psychology and game-based training research, is associated with readiness for subsequent tasks [5].

4.6. Alignment with Current Evidence

Across domains, recent RCTs, protocols, and reviews support the idea that experience design (onboarding, adaptive challenges, feedback) is a key lever for cognitive intervention success in both young adults and older populations [1]. Interpretable, audit-ready metrics such as FWA 0–1 facilitate aligning app selection with curricular objectives and documenting quality for repeated evaluations [14].

4.7. Utility: From Scores to Action

The primary utility is converting multi-factor perceptions into practical rules: Elevate for onboarding and structured progress; Lumosity for engagement and short-term focus. With DEGREE as the tool and Fuzzy as the aggregation engine, outputs are transparent, replicable, and actionable—ready to be adopted as an evaluation standard and SOP for application selection in higher education [44],[45],[46].

4.8. Limitations and Future Work

This study reports a controlled same-participant benchmark within an Informatics undergraduate cohort, which strengthens internal comparability but may limit generalisability to non-technical populations. The outcomes are perception-based (UX ratings), and the fuzzy index is designed to improve interpretability rather than to replace objective cognitive outcomes. Future work should replicate the benchmark with broader demographics (e.g., non-CS students, older adults, or school-aged learners), triangulate survey-based UX with in-app telemetry (e.g., session persistence, task completion, error rates), and evaluate whether the factor-level advantages translate into measurable learning or cognitive gains under specific instructional protocols.

5. CONCLUSION

This study presents a clear, actionable choice between two popular cognitive training apps through a measured comparison using DEGREE (14 factors) and Fuzzy Logic as a credible evaluation engine. Based on 190 respondents (→ 380 app-evaluations), the instrument shows strong internal consistency ($\alpha_{Lumosity} = 0.822$; $\alpha_{Elevate} = 0.847$). Descriptive profiles place both apps in the mid-to-high range (typical factor means ≈ 3.4–4.0), with overall means 3.51 (Lumosity) and 3.44 (Elevate). Fuzzy aggregation yields interpretable indices: FWA 0.503 (Lumosity) and 0.490 (Elevate).

At the factor level, the comparison suggests that Lumosity is marginally higher on most dimensions (12/14), while Elevate shows advantages on Learnability and Confidence. However, because the overall paired difference is not statistically significant, the results are best interpreted as a close match between the two apps in this cohort, with only small factor-wise deviations.

Practically, the findings support a goal-based adoption strategy in higher education: use Elevate to reduce onboarding friction and support structured progression when courses emphasise task clarity and control and use Lumosity as an engagement booster when sessions require short-burst attention and enjoyment. While the benchmark is robust within the present cohort, future replications across diverse learner groups and the inclusion of behavioural and cognitive outcome measures will further strengthen the external validity of the proposed benchmarking pipeline.

Vol. 7, No. 2, April 2025, Page. 1800-1817
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.2.5646

CONFLICT OF INTEREST

The authors declare that they have no conflict of interest related to this research. The authors received no funding, gifts, premium accounts, or software support from Lumosity or Elevate, and have no financial ties or affiliations that could influence the interpretation of the results. Study design, data collection, analysis, and reporting were conducted independently and in accordance with ethical guidelines; all respondents provided informed consent, data were anonymised, and no personal/institutional interests conflicted with the scientific purpose of this study.

REFERENCES

[1] A. Pumpho, R. Boonsinsukh, and K. Srisim, "Mobile app-based cognitive decision-making and memory games enhance cognitive function in older adults with mild cognitive impairment," no. October, pp. 1–10, 2025, doi: 10.3389/fpsyg.2025.1633043.

[2] S. M. Cha, "Mobile Application Applied for Cognitive Rehabilitation: A Systematic Review," *Life*, vol. 14, no. 7, 2024, doi: 10.3390/life14070891.

[3] H. Kusuma, F. S. Rue, R. Y. Rumagit, and G. D. Pratama, "Usability evaluation of Ruangguru online learning mobile application using heuristic method," *Procedia Comput. Sci.*, vol. 245, no. C, pp. 176–184, 2024, doi: 10.1016/j.procs.2024.10.241.

[4] R. Lege, "A social presence benchmark framework for extended reality (XR) technologies," *Comput. Educ. X Real.*, vol. 4, no. December 2023, p. 100062, 2024, doi: 10.1016/j.cexr.2024.100062.

[5] M. Bang, C. W. Jang, H. S. Kim, J. H. Park, and H. E. Cho, "Mobile applications for cognitive training: Content analysis and quality review," *Internet Interv.*, vol. 33, no. July 2022, p. 100632, 2023, doi: 10.1016/j.invent.2023.100632.

[6] T. Mikolajczyk, F. Moldovan, I. Ciobanu, A. Chalupczak, and A. G. Marin, "Brain Research Using Computer Test," *Procedia Technol.*, vol. 22, no. October 2015, pp. 1113–1120, 2016, doi: 10.1016/j.protcy.2016.01.157.

[7] A. F. Silva *et al.*, "Cognitive Training with Older Adults Using Smartphone and Web-Based Applications: A Scoping Review," *J. Prev. Alzheimer's Dis.*, vol. 11, no. 3, pp. 693–700, 2024, doi: 10.14283/jpad.2024.17.

[8] A. Abid *et al.*, "An evaluation framework and comparative analysis of the widely used learning management systems," *PLoS One*, vol. 19, no. 12 December, pp. 1–30, 2024, doi: 10.1371/journal.pone.0311111.

[9] M. P. Kurniawan, M. Suyanto, and E. Utami, "DEGREE : Development and Validation of a User Experience Model for Digital Educational Games Using Cronbach ' s Alpha and Fuzzy Logic," vol. 6, no. 4, pp. 1831–1846, 2025, doi: https://doi.org/10.52436/1.jutif.2025.6.4.4942.

[10] A. Asadzadeh, H. Shahrokhi, B. Shalchi, Z. Khamnian, and P. Rezaei-Hachesu, "Serious educational games for children: A comprehensive framework," *Heliyon*, vol. 10, no. 6, p. e28108, 2024, doi: 10.1016/j.heliyon.2024.e28108.

[11] V. S. Sanjaya, N. V. Tanoto, N. Tandinata, and D. Ramdhan, "Player evaluation and learning outcomes of 'NusaVenturas': A serious game for enhancing player knowledge and interest in Indonesian traditional foods," *Procedia Comput. Sci.*, vol. 245, no. C, pp. 320–329, 2024, doi: 10.1016/j.procs.2024.10.257.

[12] G. Arastoopour Irgens, C. Bailey, T. Famaye, and A. Behboudi, "User experience testing and co-designing a digital game for broadening participation in computing with and for elementary school children," *Int. J. Child-Computer Interact.*, vol. 42, no. March, p. 100699, 2024, doi: 10.1016/j.ijcci.2024.100699.

[13] Konul Memmedova and Banu Ertuna, "Development of a Fuzzy Likert Scales to Measure Variables in Social Sciences," *Peer Rev.*, vol. 09, no. 5, pp. 1–11, 2023, [Online]. Available: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4495102

[14] J. Yang, "Fuzzy comprehensive evaluation system and decision support system for learning management of higher education online courses," *Sci. Rep.*, vol. 15, no. 1, pp. 1–19, 2025, doi: 10.1038/s41598-025-01782-w.

Vol. 7, No. 2, April 2025, Page. 1800-1817
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.2.5646

[15] T. Kyriazos and M. Poga, "Exploring Fuzzy Logic as an Alternative Approach in Psychological Scoring Abstract :," pp. 1–15, 2024, doi: 10.2174/0118743501337527241125044301.

[16] X. He *et al.*, "Mobile Health App for Adolescent Asthma Self-Management: Development and Usability Study of the Pulmonary Education and Knowledge Mobile Asthma Action Plan," *JMIR Form. Res.*, vol. 9, pp. 1–12, 2025, doi: 10.2196/64212.

[17] Z. Galavi, S. Norouzi, and R. Khajouei, "Heuristics used for evaluating the usability of mobile health applications: A systematic literature review," *Digit. Heal.*, vol. 10, 2024, doi: 10.1177/20552076241253539.

[18] Z. R. Li *et al.*, "Usability and Effectiveness of eHealth and mHealth Interventions That Support Self-Management and Health Care Transition in Adolescents and Young Adults With Chronic Disease: Systematic Review," *J. Med. Internet Res.*, vol. 26, no. 1, 2024, doi: 10.2196/56556.

[19] T. Tene, D. F. Vique López, P. E. Valverde Aguirre, N. I. Cabezas Oviedo, C. Vacacela Gomez, and S. Bellucci, "A systematic review of serious games as tools for STEM education," *Front. Educ.*, vol. 10, no. February, 2025, doi: 10.3389/feduc.2025.1432982.

[20] K. N. K. Fong *et al.*, "Efficacy of computerized cognitive training using mobile devices to promote cognitive functioning in community-dwelling older adults with or without risk of mild cognitive impairment: A multi-centre longitudinal study," *Digit. Heal.*, vol. 11, 2025, doi: 10.1177/20552076251362117.

[21] R. F. Adler, K. Baez, P. Morales, J. Sotelo, D. Victorson, and S. Magasi, "Evaluating the Usability of an mHealth App for Empowering Cancer Survivors With Disabilities: Heuristic Evaluation and Usability Testing," *JMIR Hum. Factors*, vol. 11, no. 1, pp. 1–14, 2024, doi: 10.2196/51522.

[22] S. Li, C. Wang, and Y. Wang, "Fuzzy evaluation model for physical education teaching methods in colleges and universities using artificial intelligence," *Sci. Rep.*, vol. 14, no. 1, pp. 1–17, 2024, doi: 10.1038/s41598-024-53177-y.

[23] A. Asemi and A. Asemi, "A judgment-based model for usability evaluating of interactive systems using fuzzy Multi Factors Evaluation (MFE)," *Appl. Soft Comput.*, vol. 117, p. 108411, 2022, doi: 10.1016/j.asoc.2022.108411.

[24] W. Husain *et al.*, "Reliability generalization meta-analysis of the internal consistency of the Big Five Inventory (BFI) by comparing BFI (44 items) and BFI-2 (60 items) versions controlling for age, sex, language factors," *BMC Psychol.*, vol. 13, no. 1, p. 20, 2025, doi: 10.1186/s40359-024-02271-x.

[25] P. A. Edelsbrunner, B. A. Simonsmeier, and M. Schneider, *The Cronbach's Alpha of Domain-Specific Knowledge Tests Before and After Learning: A Meta-Analysis of Published Studies*, vol. 37, no. 1. Springer US, 2025. doi: 10.1007/s10648-024-09982-y.

[26] M. Villar-Guevara, J. H. Livia-Segovia, E. E. García-Salirrosas, and I. Fernández-Mallma, "Student Evaluation of Teachers' Effectiveness (SETE) scale: translation, cross-cultural adaptation and psychometric properties in a Latin American sample," *Front. Educ.*, vol. 9, no. July, 2024, doi: 10.3389/feduc.2024.1401718.

[27] H. Takahashi *et al.*, "Educational Utility of Clinical Vignettes Generated in Japanese by ChatGPT-4: Mixed Methods Study," *JMIR Med. Educ.*, vol. 10, pp. 1–13, 2024, doi: 10.2196/59133.

[28] F. Saigí-Rubió, T. Romeu, E. H. Encuentra, M. Guitert, E. Andrés, and E. Reixach, "Design, Implementation, and Analysis of an Assessment and Accreditation Model to Evaluate a Digital Competence Framework for Health Professionals: Mixed Methods Study," *JMIR Med. Educ.*, vol. 10, 2024, doi: 10.2196/53462.

[29] J. R. Lindner and N. Lindner, "Interpreting Likert type, summated, unidimensional, and attitudinal scales: I neither agree nor disagree, Likert or not," *Adv. Agric. Dev.*, vol. 5, no. 2, pp. 152–163, 2024, doi: 10.37433/aad.v5i2.351.

[30] M. Carmona-Halty, K. Alarcón-Castillo, C. Semir-González, G. Sepúlveda-Páez, and W. B. Schaufeli, "Burnout Assessment Tool for Students (BAT-S): evidence of validity in a Chilean sample of undergraduate university students," *Front. Psychol.*, vol. 15, no. December, pp. 1–8, 2024, doi: 10.3389/fpsyg.2024.1434412.

[31] H. Sigmundsson, H. Hauge, J. S. R. Leversen, and M. Haga, "I CAN intervention to increase grit

Vol. 7, No. 2, April 2025, Page. 1800-1817
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.2.5646

and growth mindset: exploring the intervention for 15-year-olds Norwegian adolescents," *Front. Educ.*, vol. 9, no. November, 2024, doi: 10.3389/feduc.2024.1438280.

[32] M. Luo, Y. Gu, F. Zhou, and S. Chen, "Implementation of the Observational Medical Outcomes Partnership Model in Electronic Medical Record Systems: Evaluation Study Using Factor Analysis and Decision-Making Trial and Evaluation Laboratory-Best-Worst Methods," *JMIR Med. Informatics*, vol. 12, 2024, doi: 10.2196/58498.

[33] C. R. Peter, "statistics : Open for everyone," p. 353, 2025.

[34] A. Roquet, P. Martinelli, C. Lampraki, and D. S. Jopp, "Internet Use as a Moderator of the Relationship Between Personal Resources and Stress in Older Adults: Cross-Sectional Study," *JMIR Aging*, vol. 7, 2024, doi: 10.2196/52555.

[35] P. J. Muhammad Ali, "Investigating the Impact of Min-Max Data Normalization on the Regression Performance of K-Nearest Neighbor with Different Similarity Measurements," *ARO-The Sci. J. Koya Univ.*, vol. 10, no. 1, pp. 85–91, 2022, doi: 10.14500/aro.10955.

[36] M. Bergmann, "Fuzzy Membership Functions," *An Introd. to Many-Valued Fuzzy Log.*, no. xi, pp. 309–314, 2012, doi: 10.1017/cbo9780511801129.018.

[37] J. Yang, "Fuzzy comprehensive evaluation system and decision support system for learning management of higher education online courses," *Sci. Rep.*, vol. 15, no. 1, pp. 1–18, 2025, doi: 10.1038/s41598-025-01782-w.

[38] M. Mizumoto, "Defuzzification methods," *Handb. Fuzzy Comput.*, pp. 1–7, 2004, doi: 10.1887/0750304278/b438c24.

[39] G. Milutinović, *Computational and Visual Tools for Geospatial Multi-Criteria Decision-Making*. Faculty of Engineering and Sustainable Development Department of Computer and Geospatial Sciences SE-801 76 Gävle, Sweden, 2020. doi: 10.1061/41022(336)81.

[40] S. P. Baral, P. K. Parida, and S. K. Sahoo, "Fuzzy TOPSIS Approaches for Multi-criteria Decision-Making Problems in Triangular Fuzzy Numbers," *Springer Nat. Singapor*, no. February, pp. 1–28, 2023, doi: 10.1007/978-981-19-6631-6.

[41] A. P. Facchino, D. Marchetti, M. Colasanti, L. Fontanesi, and M. C. Verrocchio, "The use of serious games for psychological education and training: a systematic review," *Front. Educ.*, vol. 10, no. February, 2025, doi: 10.3389/feduc.2025.1511729.

[42] P. N. Kanimalar and R. Balakumar, "The Art of Fuzzy: Crafting Defuzzification with Centroid of Maxima and Minima," *Int. J. Fuzzy Syst.*, 2025, doi: 10.1007/s40815-025-01996-0.

[43] H. Schilling, S. Hirschberger, and S. Kauffeld, "A focus group study of students' expectations of digital onboarding tools in higher education," *Front. Educ.*, vol. 9, no. November, pp. 1–15, 2024, doi: 10.3389/feduc.2024.1287387.

[44] M. Kuppusamy, J. Singh, P. Raman, and N. Abdolrahimi, "Advanced Decision Analytics with Fuzzy Logic Integrating AI and Computational Thinking for Personnel Selection," *J. Fuzzy Ext. Appl.*, vol. 5, no. 4, pp. 679–699, 2024, doi: 10.22105/jfea.2024.477833.1621.

[45] H. Hidayat, A. Komariah, B. B. Wiyono, and Y. Huda, "Impact of the Use of Fuzzy Comprehensive Evaluation Applications towards Computational Thinking Skill Students in Engineering Education," *Int. J. Inf. Educ. Technol.*, vol. 15, no. 1, pp. 90–100, 2025, doi: 10.18178/ijiet.2025.15.1.2221.

[46] A. Alharbi *et al.*, "Selection of data analytic techniques by using fuzzy AHP TOPSIS from a healthcare perspective," *BMC Med. Inform. Decis. Mak.*, vol. 24, no. 1, 2024, doi: 10.1186/s12911-024-02651-8.