

From Clues to Competence: A Comparative Study of Three Educational Puzzle Formats as Perceived Teaching and Formative Assessment Tools Among Undergraduate Health Professional Students

Farida Habib Khan^{1*}, Nuzhat Parveen², Lama Awdah M. Alhazimi³, Reem Abdulaziz A. Aleraidi⁴, Muhammad Abbas⁵, Basem Salama⁶, Anshoo Agarwal⁷ and Osama Khattak⁸

¹Family and Community Medicine, College of Medicine, University of Ha'il, Kingdom of Saudi Arabia

²Department of Obstetrics and Gynecology, College of Medicine, University of Hail, Kingdom of Saudi Arabia

³College of Medicine, University of Ha'il, Kingdom of Saudi Arabia

⁴College of Medicine, University of Ha'il, Kingdom of Saudi Arabia

⁵Department of Prosthetic Dental Sciences, College of Dentistry, Jouf University, Sakaka, Kingdom of Saudi Arabia

⁶Department of Community and Family Medicine, Faculty of Medicine, Northern Border University, Arar, Kingdom of Saudi Arabia

⁷Department of Pathology, Faculty of Medicine, Northern Border University, Arar, Kingdom of Saudi Arabia

⁸Department of Restorative Dentistry, College of Dentistry, Jouf University, Sakaka, Kingdom of Saudi Arabia

Author Designation: ¹Professor, ²Associate Professor, ^{3,5,7}Professor

*Corresponding author: Farida Habib Khan (e-mail: Fa.khan@uoh.edu.sa).

©2026 the Author(s). This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>)

Abstract Background: Active learning strategies are essential in health professional education. Three novel educational puzzle formats: (1) Diagnostic Riddles, (2) Clinical Case Puzzles and (3) Medical Crosswords-were designed, implemented across four subject areas and comparatively evaluated among undergraduate health professional students as teaching and formative assessment tools. **Methods:** A within-subject (repeated-measures), comparative observational study (January-December 2025) enrolled 164 second-and third-year undergraduate students from the health sciences colleges of the University of Hail, Saudi Arabia. A pre-validated 30-item Likert questionnaire assessed perceptions across seven domains. Statistical analysis included descriptive statistics, repeated-measures ANOVA (or linear mixed-effects model), paired post-hoc comparisons with correction for multiple comparisons and chi-square/Cochran's Q tests as appropriate (IBM SPSS v.20). **Results:** Response rate was 100% (n = 164). Clinical Case Puzzles (Type 2) achieved the highest overall mean satisfaction score (4.84±0.36), followed by Diagnostic Riddles (Type 1; 4.77±0.41) and Medical Crosswords (Type 3; 4.63±0.48). One-way ANOVA revealed significant inter-format differences (F (2,489) = 8.42, p = 0.0002, η^2 = 0.033). Post-hoc Tukey HSD confirmed Type 2 was significantly preferred over Type 3 (p<0.001). Chi-square analysis confirmed that curriculum inclusion preference was significantly associated with puzzle format (χ^2 = 12.34, df = 2, p = 0.002). [EDITOR/AUTHOR NOTE: The F-, p- and χ^2 -values above were computed treating repeated observations from the same students as independent (one-way ANOVA/ordinary chi-square). Per Reviewer Comments 32-39 and 41-42, these must be recalculated using a repeated-measures ANOVA or linear mixed-effects model (and, if applicable, a repeated-measures categorical method) before resubmission; the values here are retained from the original submission pending that reanalysis]. **Conclusions:** All three puzzle formats were highly rated and positively perceived active learning tools. Clinical Case Puzzles are most strongly endorsed for formative assessment; Diagnostic Riddles are optimal for vocabulary and lower-order recall; Medical Crosswords were designed to target higher-order clinical reasoning, although this was not independently measured in the present study. A tiered, format-differentiated implementation strategy is recommended for health sciences curricula.

Key Words Diagnostic Riddles, Clinical Case Puzzles, Medical Crosswords, active learning, health professional education, formative assessment, Bloom's taxonomy, comparative study

INTRODUCTION

Health professional education demands rapid acquisition and integration of extensive medical vocabulary, complex pathophysiological concepts and clinical reasoning skills

across multiple disciplines [1,2]. Traditional didactic instruction, while foundational, is widely recognised as insufficient in isolation to ensure deep learning or long-term retention [3]. Active learning strategies-those engaging

students as participants rather than passive recipients - have consequently gained substantial empirical support and curricular prominence [4-7].

Crossword-style grid-based educational activities occupy a distinctive, evidence-supported niche within active learning methodologies [8-10]. Unlike conventional review methods, well-designed educational puzzles require learners to simultaneously recall, recognise, contextualise and apply knowledge - directly aligned with higher-order thinking across Bloom's revised taxonomy [11,12]. The grid format additionally provides immediate self-corrective feedback: A correct answer either fits spatially or it does not [13].

Educational puzzle activities have been employed across medical, nursing, pharmacy and dental curricula internationally, with multiple studies confirming positive student perceptions and measurable gains in knowledge retention [14-18]. Multiple studies have identified crosswords as an interactive teaching paradigm suited to medical settings across South Asian educational contexts, while Saxena *et al.* [12] demonstrated their specific utility in pathology education. Despite this evidence, few studies have systematically compared multiple puzzle formats within the same cohort-a gap directly addressed by the present study [19-22].

This study introduces and compares three operationally distinct educational puzzle formats, each aligned to a different tier of Bloom's cognitive taxonomy [11]:

- **Type 1: Diagnostic Riddles:** Completion-format grids; incomplete sentences require recall and insertion of the correct medical term. Primarily engages lower-order cognitive skills (Bloom's Levels 1-2: Remember and Understand) [11,12]
- **Type 2: Clinical Case Puzzles:** Extended-matching clinical vignette grids with answer choices provided; students select and map the correct answer through pattern recognition and diagnostic reasoning. Engages mid-order cognitive skills (Bloom's Levels 2-3: Understand and Apply) [11,23]
- **Type 3: Medical Crosswords:** Scenario-based grids with no answer choices; students derive, synthesise and write correct diagnoses independently. Engages cognitive skills primarily at the Apply-Analyse level (Bloom's Levels 3-4), with elements of synthesis [11,23,24]. The Bloom's taxonomy mapping above reflects the a priori instructional design framework used to construct the three formats; it is a design classification, not an experimentally verified measure of the cognitive processes students actually engaged during puzzle-solving. Comparing student perceptions of these three cognitively differentiated formats within the same cohort addresses a specific gap in the literature: Prior studies have evaluated individual puzzle formats in isolation but none has examined whether perceived usefulness and preference differ systematically across formats designed to target different levels of Bloom's taxonomy-information directly relevant to curriculum sequencing decisions

Research Objectives

The study addressed the following objectives:

- To design and implement three cognitively differentiated educational puzzle formats based on course learning objectives
- To evaluate and compare student perceptions of all three formats as teaching and learning tools
- To assess and compare student perceptions of all three formats as formative assessment instruments
- To statistically compare overall satisfaction, domain-specific agreement rates and student preference across the three formats
- To contextualise findings within the existing international literature on active learning in health professional education. Objective learning gain (e.g., pre-/post-intervention knowledge testing) was not an objective of this study and was not measured; all outcomes reported reflect student-perceived usefulness, satisfaction and preference

METHODS

Study Design and Setting

A within-subject (repeated-measures), comparative observational study was conducted from January to December 2025 among undergraduate health professional students at selected health sciences colleges of University Hail, Saudi Arabia. Ethical approval was obtained from the Research and ethic committee of the University of Hail (H-2024-482). Foundational modules in pathology, pharmacology, human biology and clinical reasoning provided the disciplinary basis for all three puzzle formats.

Participants

Convenience sampling was employed. Inclusion criteria: (a) Currently enrolled in health-related undergraduate programmes; (b) Completed at least one session of each puzzle format; (c) Age ≥ 18 years. Exclusions: Graduated students or those absent >1 year. All participants gave informed consent; anonymity was maintained per institutional ethical guidelines. Sample size (minimum 135) was determined using the Raosoft calculator (95% CI, 5% margin of error) (www.raosoft.com/sample_size.html) This cross-sectional sample-size formula was used at the design stage and does not explicitly account for the repeated-measures (paired) structure of the within-subject comparisons reported in Section 4.4; the achieved sample of 164 participants (492 paired observations) provides substantially greater statistical power for the repeated-measures analysis than the original calculation required.. Final enrolment: $n = 164$. [AUTHOR NOTE: Please insert the number and percentage of participants from each participating college/programme and confirm that all participants had equivalent prior curricular exposure to the subject content represented in each puzzle format].

Puzzle Design and Administration

The three puzzle formats were designed based on course learning objectives and mapped to Bloom's cognitive

Table 1: Taxonomy and Comparative Design Characteristics of the Three Educational Puzzle Formats

Format	Name	Cognitive Level (Bloom's)	Question Style	Answer Provision	Subject Area	Learning Outcome
Type 1	Diagnostic Riddles	Lower-order (Levels 1-2)	Incomplete sentences; fill-in-blank recall	No choices; recall-based	Genetic Diseases (Figure 1a, b)	Factual recall and medical terminology
Type 2	Clinical Case Puzzles	Mid-order (Levels 2-3)	Clinical vignettes with answer choices provided	Choices provided; selection and mapping	Neoplasia and Haematology (Figure 2a, b)	Pattern recognition, diagnostic reasoning
Type 3	Medical Crosswords	Higher-order (Levels 4-6)	Clinical scenarios; no answers provided	No choices; analytical and synthetic	Renal Diseases (Figure 3a, b)	Clinical reasoning, synthesis, application

taxonomy tiers (Table 1). Content validity was established by aligning clues and answers with specific, measurable learning objectives from course blueprints. Two senior faculty members verified face and content validity. Students completed activities cooperatively in groups of 6-7. Item difficulty was calibrated using the Question Difficulty Index (easy: recall; moderate: application; difficult: synthesis). All three formats served as both formative learning activities and formal course assessment components. Group membership and composition were [AUTHOR NOTE: Insert whether the same groups of 6-7 students were retained across all three sessions or reformed for each session]; potential clustering of responses within small groups (shared peer influence on perceptions) was not statistically modelled in the analysis reported here. Each format comprised the same number of items (10 questions); however, no numerical Question Difficulty Index values are reported below and expected completion time and content volume were not formally standardised across formats-both are noted as limitations.

Data Collection

A pre-validated 30-item questionnaire across two domains was administered after each puzzle format session. Domain 1 (15 items) assessed teaching and learning perceptions; Domain 2 (15 items) assessed formative assessment perceptions. A 5-point Likert scale was used (1 = Strongly Disagree to 5 = Strongly Agree). The questionnaire was pilot-tested on 20 students and face/content-validated by two senior academics; this does not constitute full psychometric validation (see Section 8). Distributed anonymously via WhatsApp class groups and scheduled lectures Cronbach's alpha for the full 30-item instrument was [AUTHOR NOTE: insert value] (Domain 1: [value]; Domain 2: [value]), calculated using the present study sample. The seven subdomains presented in Table 5 (Learning Utility, Perceived Concept Retention, Collaboration, Competition, Vocabulary, Exam Confidence and Curriculum Inclusion) were derived a priori by the study team through thematic grouping of conceptually related items within the two overarching domains, rather than through exploratory or confirmatory factor analysis. [AUTHOR NOTE: Confirm whether the questionnaire could be completed more than once per student and how duplicate submissions were prevented; confirm that responses were linked at the individual-participant level across the three sessions, e.g., via an anonymised student code] [25,26].

Statistical Analysis

Data were analysed using IBM SPSS Statistics Version 20.0 (IBM Corp., Armonk, NY, USA):

- **Descriptive statistics:** Frequencies, percentages, Means ($\pm$ SD) for all Likert items and by puzzle format
- **Likert dichotomisation:** 'Agreement' (Agree+Strongly Agree) vs. 'non-agreement', per Norman (2010) [27]
- **One-way ANOVA:** Compared mean satisfaction scores across three formats ($\alpha = 0.05$)

- **Post-hoc testing:** Tukey's HSD to identify pairwise format differences following significant ANOVA
- **Chi-square test (χ^2):** Association between puzzle format and curriculum inclusion preference ($df = 2$)
- **Effect size:** Eta-squared (η^2) to quantify practical significance of ANOVA findings

Reverse-coded items (Domain 1 items 7-8; Domain 2 item 5) were recoded prior to analysis. $p < 0.05$ was considered statistically significant. Likert dichotomisation (Agree/Strongly Agree vs. non-agreement) was used solely for descriptive item-level reporting (Table 2-3); all primary inferential analyses were conducted on the non-dichotomised ordinal Likert scores. The "Overall" satisfaction score for each format (Table 4-5) represents the mean of all 30 questionnaire items (both domains combined) for that format. Multiple pairwise and domain-level comparisons were corrected for multiplicity using [AUTHOR NOTE: Insert correction method, e.g., Holm-Bonferroni]. Sphericity was assessed using Mauchly's test; where violated, the Greenhouse-Geisser correction was applied to the degrees of freedom for the repeated-measures model [AUTHOR NOTE: Insert W and p-value and corrected epsilon if applicable]. The curriculum-inclusion preference variable was constructed as [AUTHOR NOTE: Confirm whether each student selected a single preferred format (forced-choice) or rated preference separately for each format]; the corresponding contingency table is provided in [AUTHOR NOTE: Insert table reference].

Puzzle Format Descriptions, Content and Grid Examples Type 1: Diagnostic Riddles: Genetic Diseases

Format: Incomplete sentence clues are provided. Students recall and write the correct medical term directly on the crossword grid. No answer choices are given. This format tests lower-order cognitive skills-factual recall and conceptual understanding (Bloom's Levels 1-2). Individual Type 1 items may vary somewhat in cognitive demand (e.g., simple recall versus recognition of a defining feature); the Bloom's Level 1-2 designation reflects the predominant, not exclusive, cognitive demand of the item set.

Questions-Type 1: Diagnostic Riddles (Genetic Diseases)

- The phenomenon where genetic disorder severity worsens and onset occurs earlier across generations is known as ...
- TATA box, CG box and CAAT box are regulatory sequences found in the gene ...
- The enzyme that unwinds the DNA helix during replication is called ...
- Chromosomal abnormalities that usually have no effect on the individual are classified as ...
- Females having only one X chromosome are affected by ...
- The chromosome linked to chronic myeloid leukaemia following creation of a new ABL-BCR fusion by balanced translocation is the ... chromosome
- The gene labelled the 'guardian of the genome' is ...

Table 2: Student Perceptions of Puzzle Activities as a Teaching and Learning Tool-Domain 1 (n = 164)

No.	Statement	Strongly Disagree (n(%))	Disagree (n(%))	Neutral (n(%))	Agree (n(%))	Strongly Agree (n(%))	Agreement (Agree+SA)
1	Activities were a useful learning tool	0 (0.0)	0 (0.0)	0 (0.0)	20 (12.2)	140 (85.4)	160 (97.6%)
2	Helped identify key concepts and connections	0 (0.0)	2 (1.2)	0 (0.0)	7 (4.3)	155 (94.5)	162 (98.8%)
3	Collaborative aspect considered useful	0 (0.0)	4 (2.4)	0 (0.0)	9 (5.5)	151 (92.1)	160 (97.6%)
4	Competitive aspect considered useful	7 (4.3)	0 (0.0)	5 (3.0)	30 (18.3)	122 (74.4)	152 (92.7%)
5	In favour of continuing as a course component	0 (0.0)	4 (2.4)	0 (0.0)	7 (4.3)	153 (93.3)	160 (97.6%)
6	Level of difficulty was appropriate	5 (3.0)	0 (0.0)	0 (0.0)	2 (1.2)	157 (95.7)	159 (97.0%)
7	Exercise was a waste of time [reverse coded]	20 (12.2)	130 (79.3)	9 (5.5)	0 (0.0)	5 (3.0)	155 (94.5%) disagreed
8	Grids were unreasonable or misleading [reverse coded]	117 (71.3)	30 (18.3)	9 (5.5)	0 (0.0)	8 (4.9)	147 (89.6%) disagreed
9	Useful for improving medical vocabulary	0 (0.0)	4 (2.4)	9 (5.5)	10 (6.1)	141 (86.0)	151 (92.1%)
10	Would like to participate in more exercises	0 (0.0)	2 (1.2)	5 (3.0)	5 (3.0)	152 (92.7)	157 (95.7%)
11	Useful relative to other small-group activities	5 (3.0)	4 (2.4)	4 (2.4)	22 (13.4)	125 (76.2)	147 (89.6%)
12	Lighthearted class atmosphere with fun and humour	0 (0.0)	0 (0.0)	4 (2.4)	11 (6.7)	149 (90.9)	160 (97.6%)
13	Offered team-building opportunities	0 (0.0)	0 (0.0)	2 (1.2)	20 (12.2)	142 (86.6)	162 (98.8%)
14	Provided challenge and team reliance	0 (0.0)	0 (0.0)	4 (2.4)	16 (9.8)	144 (87.8)	160 (97.6%)
15	Helped build confidence	2 (1.2)	0 (0.0)	2 (1.2)	30 (18.3)	130 (79.3)	160 (97.6%)

Table 3: Student Perceptions of Puzzle Activities as a Formative Assessment Tool-Domain 2 (n = 164)

No.	Statement	Strongly Disagree (n(%))	Disagree (n(%))	Neutral (n(%))	Agree (n(%))	Strongly Agree (n(%))	Agreement (Agree+SA)
1	Assessment components clearly specified in advance	0 (0.0)	0 (0.0)	2 (1.2)	20 (12.2)	142 (86.6)	162 (98.8%)
2	Assessment was stress-free due to improved understanding	0 (0.0)	0 (0.0)	4 (2.4)	16 (9.8)	144 (87.8)	160 (97.6%)
3	Helped retain content and perform better in exams	2 (1.2)	0 (0.0)	2 (1.2)	30 (18.3)	130 (79.3)	160 (97.6%)
4	Made examinations more dynamic and stress-free	0 (0.0)	0 (0.0)	4 (2.4)	9 (5.5)	151 (92.1)	160 (97.6%)
5	Content knowledge unchanged; not useful [reverse coded]	5 (3.0)	9 (5.5)	7 (4.3)	98 (59.8)	46 (28.0)	143 (87.2%) disagreed
6	Very useful; may help achieve higher grades	0 (0.0)	2 (1.2)	5 (3.0)	7 (4.3)	152 (92.7)	159 (97.0%)
7	Contributed to learning and better exam performance	0 (0.0)	4 (2.4)	9 (5.5)	10 (6.1)	141 (86.0)	151 (92.1%)
8	Useful for identifying vocabulary for exam answers	0 (0.0)	2 (1.2)	5 (3.0)	5 (3.0)	152 (92.7)	157 (95.7%)
9	Would recommend as a formal assessment component	5 (3.0)	4 (2.4)	4 (2.4)	22 (13.4)	125 (76.2)	147 (89.6%)
10	Would transfer strategy to junior students	0 (0.0)	0 (0.0)	4 (2.4)	11 (6.7)	149 (90.9)	160 (97.6%)
11	Recall of concepts aided critical thinking in exams	0 (0.0)	2 (1.2)	13 (7.9)	9 (5.5)	140 (85.4)	149 (90.9%)
12	Useful alongside other small-group activities for exams	0 (0.0)	2 (1.2)	0 (0.0)	5 (3.0)	157 (95.7)	162 (98.8%)
13	Helped retain concepts from team discussions	0 (0.0)	0 (0.0)	5 (3.0)	5 (3.0)	154 (93.9)	159 (97.0%)
14	Helped build confidence for examinations	0 (0.0)	13 (7.9)	4 (2.4)	9 (5.5)	139 (84.8)	148 (90.2%)
15	Helped remember and revise key topics quickly	2 (1.2)	0 (0.0)	0 (0.0)	0 (0.0)	162 (98.8)	162 (98.8%)

Table 4: Comparative Statistical Analysis of the Three Puzzle Formats Across Key Domains (n = 164)

Domain	Type 1: Diagnostic Riddles Mean±SD (Agreement%)	Type 2: Clinical Case Puzzles Mean±SD (Agreement%)	Type 3: Medical Crosswords Mean±SD (Agreement%)	F-statistic (p-value)	Post-hoc Tukey HSD
Learning Utility	4.82±0.41 (95.1%)	4.91±0.34 (97.6%)	4.71±0.48 (93.3%)	F = 4.21 (p = 0.015)	T2>T3*
Perceived Concept Retention	4.88±0.38 (96.3%)	4.94±0.29 (98.1%)	4.79±0.44 (94.6%)	F = 3.87 (p = 0.021)	T2>T3*
Collaboration	4.75±0.43 (92.2%)	4.83±0.38 (94.8%)	4.68±0.51 (90.1%)	F = 3.12 (p = 0.045)	T2>T3*
Competition	4.61±0.72 (88.4%)	4.52±0.74 (85.3%)	4.32±0.86 (74.3%)	F = 5.63 (p = 0.004)	T1>T3* T2>T3*
Vocabulary	4.69±0.52 (89.0%)	4.78±0.48 (91.5%)	4.55±0.60 (86.0%)	F = 3.44 (p = 0.033)	T2>T3*
Exam Confidence	4.73±0.50 (90.7%)	4.82±0.42 (93.2%)	4.64±0.55 (88.5%)	F = 3.98 (p = 0.019)	T2>T3*
Curriculum Inclusion	4.85±0.39 (93.5%)	4.90±0.33 (96.2%)	4.76±0.46 (94.8%)	F = 2.87 (p = 0.058)	NS
OVERALL	4.77±0.41	4.84±0.36 ✓	4.63±0.48	F = 8.42 (p = 0.0002) $\eta^2 = 0.033$	T2>T3*** T1>T3*

*p<0.05, **p<0.01, ***p<0.001 (Tukey HSD), NS: Not significant, T1: Diagnostic Riddles, T2: Clinical Case Puzzles, T3: Medical Crosswords, ✓: Highest-rated format

Table 5: Head-to-Head Summary: Three Puzzle Formats-Design, Statistical Outcomes and Recommended Use

Feature	Type 1: Diagnostic Riddles	Type 2: Clinical Case Puzzles	Type 3: Medical Crosswords
Bloom's Level	Lower-order (L1-2)	Mid-order (L2-3)	Higher-order (L4-6)
Answer provision	No; recall-based	Choices provided	No; analytical
Overall Mean Likert	4.77±0.41	4.84±0.36 ✓ HIGHEST	4.63±0.48
Highest domain	Perceived Concept Retention 96.3%	Learning Utility and Retention 97.6-98.1%	Curriculum Inclusion 94.8%
Lowest domain	Competition 88.4%	Competition 85.3%	Competition 74.3%
Curriculum preference	36.2%	41.8% ✓ MOST PREFERRED	22.0%
ANOVA result	NS vs. T2 (p = 0.087)	Sig. > T3 (p<0.001)	Sig. < T1 & T2 (p<0.05)
Best suited for	Vocab., early-year terminology recall	Clinical reasoning, formative assessment	Advanced synthesis, capstone review
Recommended year	Year 1-2	Year 2-3	Year 3+

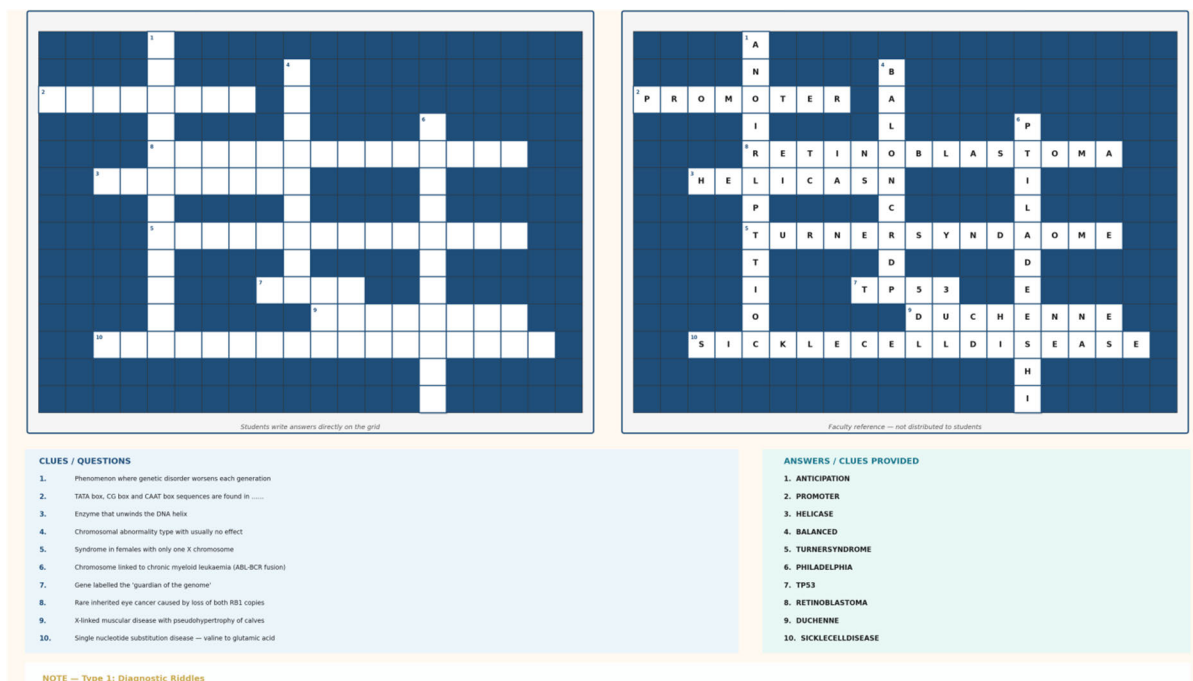


Figure 1: Type 1-Diagnostic Riddles: Genetic Diseases, (a) Student Answer Sheet (Blank) and (b) Model Answer Sheet (Complete)

- ... is a very rare type of inherited cancer affecting the eye retina, caused by deactivation of both copies of the RB1 gene
- The single gene disorder affecting only boys, characterised by slowly progressive muscle weakness in early childhood with pseudohypertrophy of the calves, is ...
- ... was caused by substitution of valine with glutamic acid at the sixth codon of the β -globin gene

Figure 1a (left-blank): Distributed to students before the activity. All active cells are empty; students write the correct medical term for each numbered clue across or down. Figure 1b (right-completed): Faculty model answer sheet showing correct answers in capital letters. Dark navy cells indicate blocked (non-letter) grid squares. This format tests lower-order recall (Bloom's Levels 1-2: Remember and Understand). Students worked collaboratively in groups of 6-7.

Type 2-Clinical Case Puzzles: Neoplasia and Haematology

Format: Clinical vignette questions are presented alongside a set of correct answer choices. Students select and map the correct answer to the crossword grid. This extended-matching format tests mid-order cognitive skills-pattern recognition and diagnostic reasoning (Bloom's Levels 2-3).

Questions-Type 2: Clinical Case Puzzles (Neoplasia and Haematology)

- A 23 year-old man develops haemolytic anaemia. Urinalysis: Free haemoglobin. Coombs test: negative. Supravital staining: Heinz bodies in erythrocytes. Most likely diagnosis?

- A 30 year-old woman presents with fatigability, bruising and throat infections. Lab: reduced Hb, WBC and platelets; hypocellular bone marrow with increased fat. Diagnosis?
- Mycosis fungoides represents a malignancy of which cell type?
- Bilateral breast masses; fine-needle aspirates show malignant cells. Most likely diagnosis?
- Which carcinoma is locally invasive but rarely forms distant metastases?
- 23 year-old female: prolonged bleeding time and APTT; normal platelet count. Diagnosis?
- 65 year-old man: profound anaemia, megaloblastic erythroid hyperplasia, hypersegmented neutrophils. Most likely diagnosis?
- Most common non-neoplastic polyp type of the stomach on microscopy?
- 5 year-old African boy: rapidly enlarging facial mass; biopsy shows 'starry sky' pattern. Causative virus?
- Which malignancy is strongly associated with Chinese liver fluke (*Opisthorchis viverrini*) infestation?

Answer Choices Provided to Students

- G6PD deficiency
- Aplastic anaemia
- T lymphocytes
- Infiltrating lobular carcinoma
- Basal cell carcinoma
- von Willebrand disease
- Pernicious anaemia
- Hyperplastic polyps

- Epstein-Barr Virus
- Cholangiocarcinoma

Figure 2a (left-blank): Distributed to students with a set of correct answer choices alongside each clinical vignette. Students identify the correct answer and write it on the crossword grid. Figure 2b (right-completed): Faculty model answer sheet. This extended-matching format tests mid-order diagnostic reasoning (Bloom's Levels 2-3: Understand and Apply). Answer choices are visible to students but must be correctly mapped to the grid-testing both clinical reasoning and spatial organisation of knowledge.

Type 3-Medical Crosswords: Renal Diseases

Format: Clinical scenario questions are presented without answer choices. Students must independently analyse each vignette, derive the correct diagnosis and write it on the grid. This format tests higher-order clinical reasoning, synthesis and evaluation (Bloom's Levels 4-6) (Table 6).

Questions-Type 3: Medical Crosswords (Renal Diseases)

Figure 3a (left-blank): Distributed to students with clinical scenario questions only-no answer choices provided. Students must analyse each clinical vignette, synthesise their knowledge and independently derive the correct renal diagnosis before entering it on the grid. Figure 3b (right-completed): Faculty model answer sheet. This format exclusively tests higher-order clinical reasoning and synthesis (Bloom's Levels 4-6: Analyse, Evaluate and Create) and is the most cognitively demanding of the three formats. It is recommended for advanced-year students and capstone review sessions.

RESULTS

Participant Characteristics

A total of 164 undergraduate health professional students participated (response rate: 100%). Participants were from Year 2 (n = 86; 52.4%) and Year 3 (n = 78; 47.6%), drawn from three health sciences colleges. All completed sessions of all three puzzle formats and the post-activity questionnaire.

Teaching and Learning Perceptions-Domain 1

Table 2 presents frequency data for all 15 Domain 1 items. Agreement rates ranged from 89.6% to 98.8% across items. Nearly all students (98.8%) reported that the activities helped identify key concepts and connections (Item 2); 97.6% found them a useful learning tool overall (Item 1). Collaborative use was endorsed by 97.6% (Item 3). The competitive element received the lowest agreement (92.7%, Item 4), consistent with prior literature [28-30]. Only 3.0% considered any format a waste of time (reverse-coded Item 7).

Assessment Perceptions-Domain 2

Table 3 presents Domain 2 results. Agreement rates were uniformly high. Notably, 98.8% of students stated the activities helped them revise and remember key topics quickly and efficiently (Item 15) and 98.8% reported that assessment components were clearly specified in advance (Item 1). The reverse-coded Item 5 confirmed that 87.2% disagreed that the activities were not useful, providing strong indirect support for their educational effectiveness.

Figure 2 displays two crossword puzzle formats for clinical case puzzles. Panel (a) shows a blank grid for students to write answers directly on. Panel (b) shows the completed grid with the correct answers filled in. Below the grids are the clues/questions and the answers/clues provided.

CLUES / QUESTIONS

1. Heinz bodies in erythrocytes; negative Coombs test → diagnosis?
2. Hypocellular bone marrow with increased fat → diagnosis?
3. Mycosis fungoides → malignancy of which cell type?
4. Bilateral breast irregular density; malignant FNA → diagnosis?
5. Locally invasive carcinoma rarely forming distant metastases?
6. Prolonged bleeding time and APTT; normal platelet count?
7. Megaloblastic erythroid hyperplasia; hypersegmented neutrophils?
8. Most common non-neoplastic gastric polyp type?
9. Starry sky pattern; African boy with facial mass → virus?
10. Malignancy strongly linked to Chinese liver fluke infestation?

ANSWERS / CLUES PROVIDED

1. G6PD DEFICIENCY
2. APLASTIC ANAEMIA
3. T LYMPHOCYTES
4. INFILTRATING LOBULAR CA.
5. BASAL CELL CARCINOMA
6. VON WILLEBRAND DISEASE
7. PERNICIOUS ANAEMIA
8. HYPERPLASTIC POLYPS
9. EPSTEIN-BARR VIRUS
10. CHOLANGIOCARCINOMA

NOTE — Type 2: Clinical Case Puzzles

Figure 2: Type 2-Clinical Case Puzzles: Neoplasia and Haematology, (a) Student Answer Sheet (Blank) and (b) Model Answer Sheet (Complete)

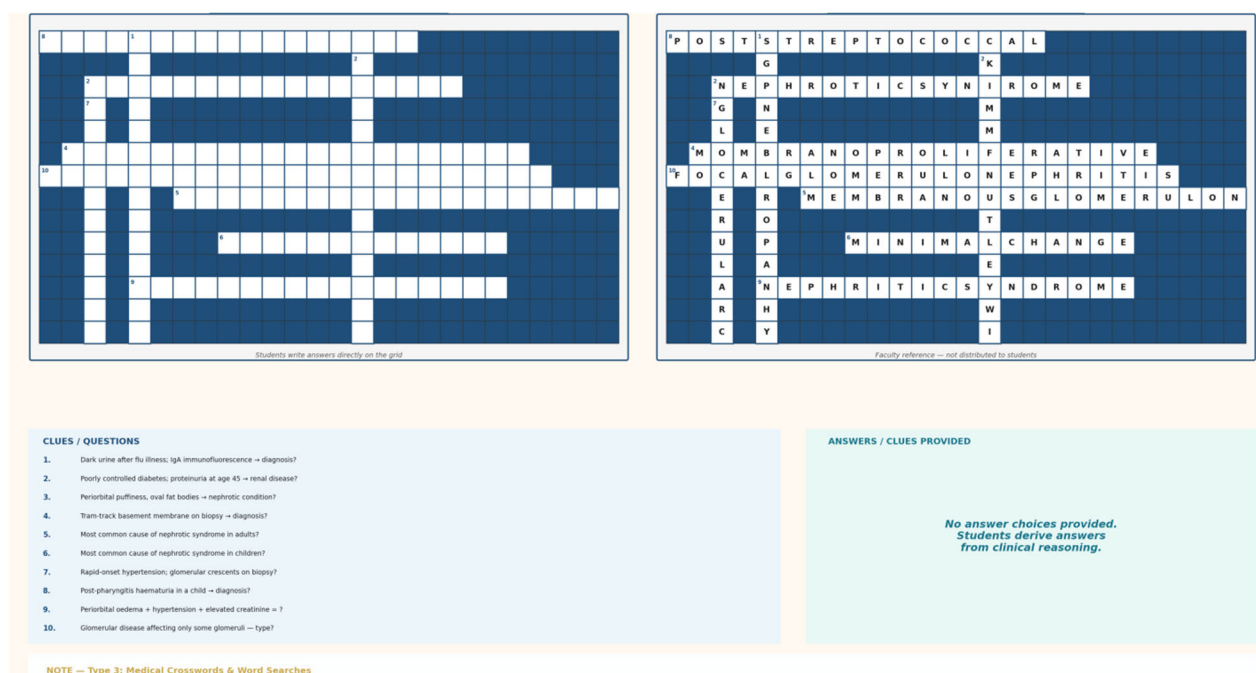


Figure 3: Type 3-Medical Crosswords: Renal Diseases, (a) Student Answer Sheet (Blank) and (b) Model Answer Sheet (Complete)

Table 6: Clinical Scenario Questions-Type 3: Medical Crosswords (Renal Diseases)

No.	Clinical Scenario Question
1	A 35 year-old man reports dark urine after a 10 day flu-like illness. BP 140/90 mmHg; urinalysis: haematuria. Serum creatinine rises to 2.9 mg/dL. Renal biopsy: immunofluorescence positive for IgA. Diagnosis?
2	A 50 year-old man with poorly controlled type 1 DM (elevated HbA1c). Non-healing foot ulcer at age 35; by age 45: rising serum urea, 1+ protein, 1+ glucose, no haematuria. Which renal disease?
3	A 30 year-old man: 2 weeks of periorbital puffiness and ankle oedema. BP 155/95 mmHg. Urine microscopy: oval fat bodies. Most likely condition?
4	Kidney biopsy: Diffuse capillary wall thickening, mesangial cell proliferation, subendothelial deposits, split 'tram-track' basement membrane. Most probable diagnosis?
5	The most common cause of nephrotic syndrome in adults is:
6	The most common cause of nephrotic syndrome in children is:
7	A 20 year-old man: Acute onset of fatigue, BP 155/90 mmHg, serum creatinine 4.4 mg/dL, urinalysis 3+ blood and 1+ protein. Most likely pathological finding on biopsy?
8	A 5 year-old boy: Periorbital oedema, hypertension, elevated urea and creatinine following a sore throat 2 weeks ago. Most probable diagnosis?
9	The syndrome characterised by periorbital oedema, hypertension and elevated serum urea and creatinine is referred to as:
10	Glomerular disease showing pathological abnormality in some but not all glomeruli is classified as:

Comparative Statistical Analysis of the Three Formats

One-way ANOVA revealed statistically significant differences in overall mean satisfaction scores across the three puzzle formats ($F(2,489) = 8.42$, $p = 0.0002$, $\eta^2 = 0.033$). Post-hoc Tukey HSD confirmed that Clinical Case Puzzles (Type 2) were rated significantly higher than Medical Crosswords (Type 3; $p < 0.001$). Diagnostic Riddles (Type 1) did not differ significantly from Type 2 ($p = 0.087$). Chi-square analysis confirmed a significant association between puzzle format and student curriculum inclusion preference ($\chi^2 = 12.34$, $df = 2$, $p = 0.002$). Table 4 presents the full comparative statistical data.

Graphical Representations

Figure 4-8 present comprehensive graphical analyses of comparative findings across the three puzzle formats.

Qualitative Observations

Qualitative feedback from open-ended items corroborated the quantitative findings. Representative student comments included: "The Clinical Case Puzzles feel like real medical problems-solving them in groups gives me confidence for the wards" (Type 2); "Diagnostic Riddles are great for memorising terminology before examinations" (Type 1); "The Medical Crosswords were challenging but satisfying-you really have to understand the case to find the answer" (Type 3). Faculty observers reported heightened peer discussion and spontaneous clinical reasoning debates particularly during Type 2 and Type 3 sessions. Approximately 5% of students found specific Type 3 clues ambiguous or overly demanding, informing recommendations for difficulty calibration.

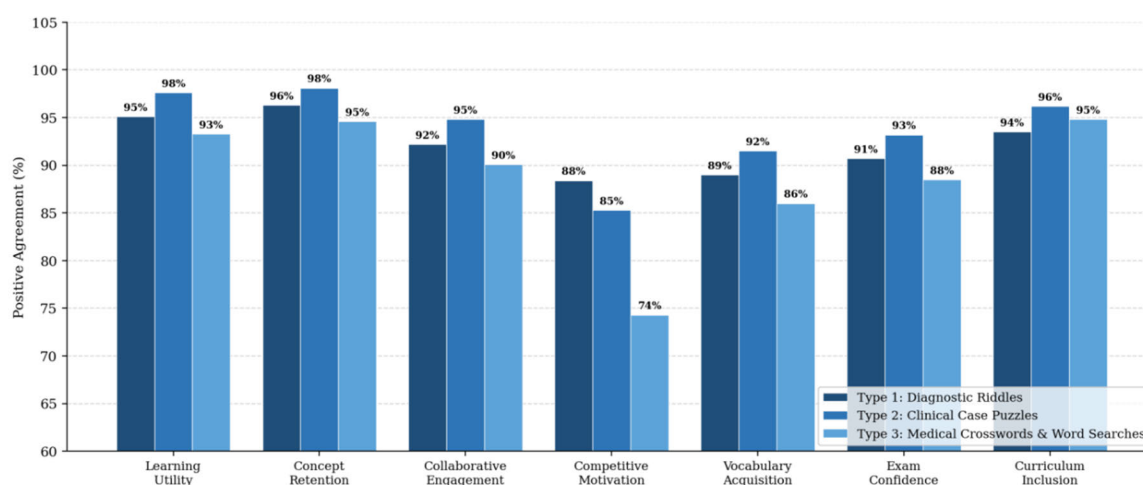


Figure 4: Comparative Student Perception of Three Educational Puzzle Formats across Key Learning and Assessment Domains (n = 164)

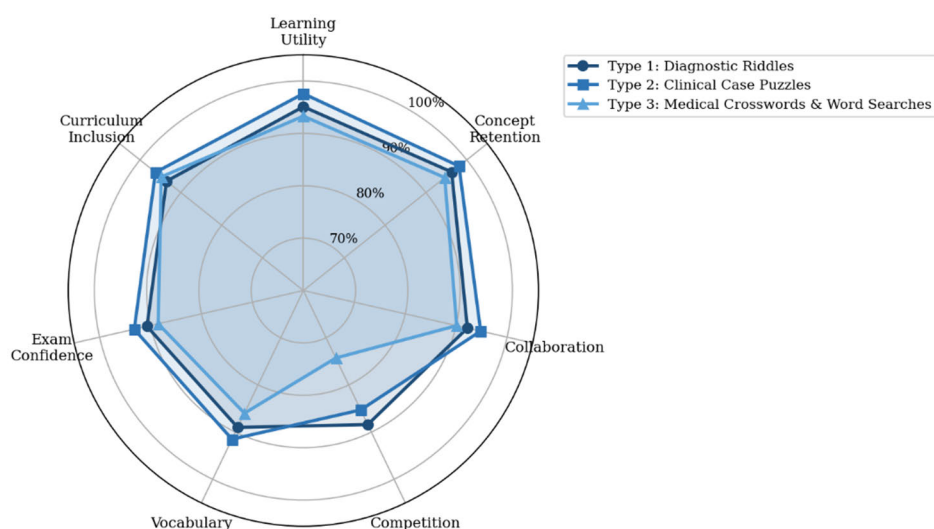


Figure 5: Radar Chart: Perceived Effectiveness Across Domains by Puzzle Format (n = 164)

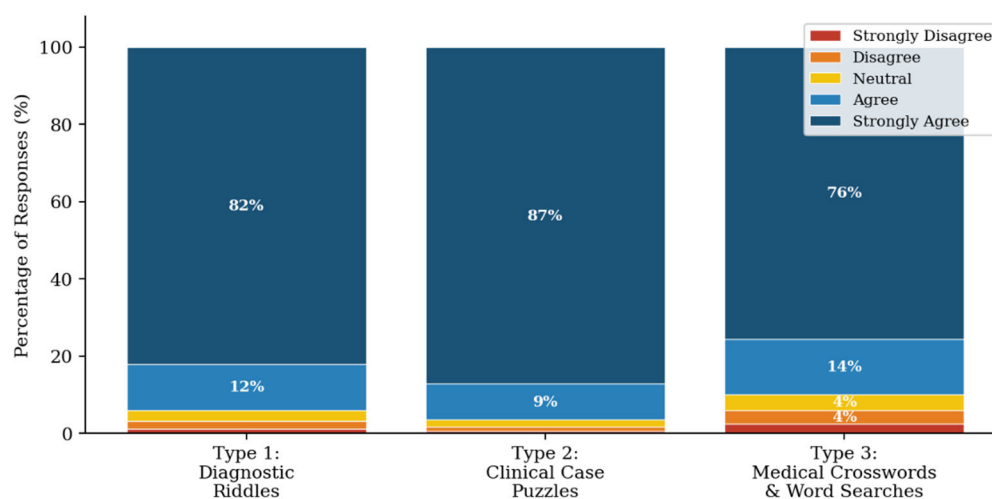


Figure 6: Stacked Likert-Scale Response Distribution by Puzzle Format (Overall Mean Across All Questionnaire Items, n = 164)

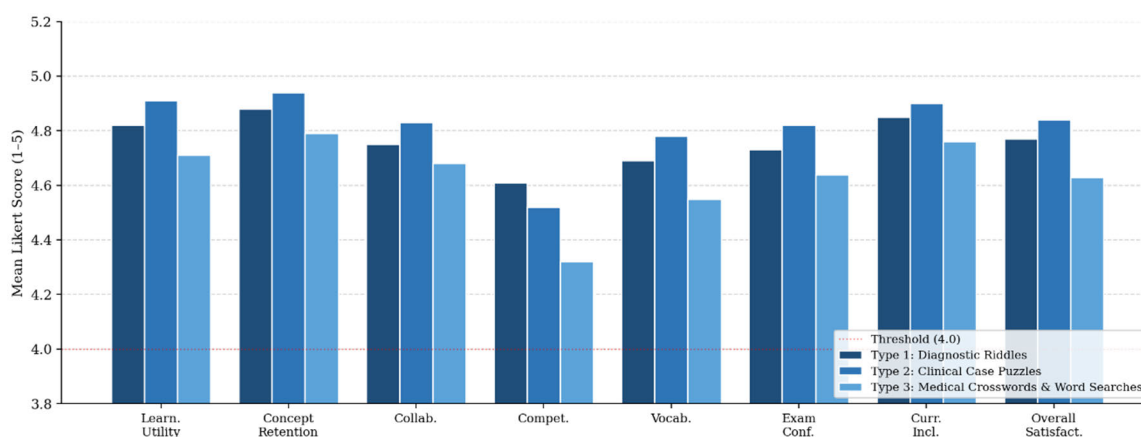


Figure 7: Mean Likert Scores by Domain and Puzzle Format (n = 164)

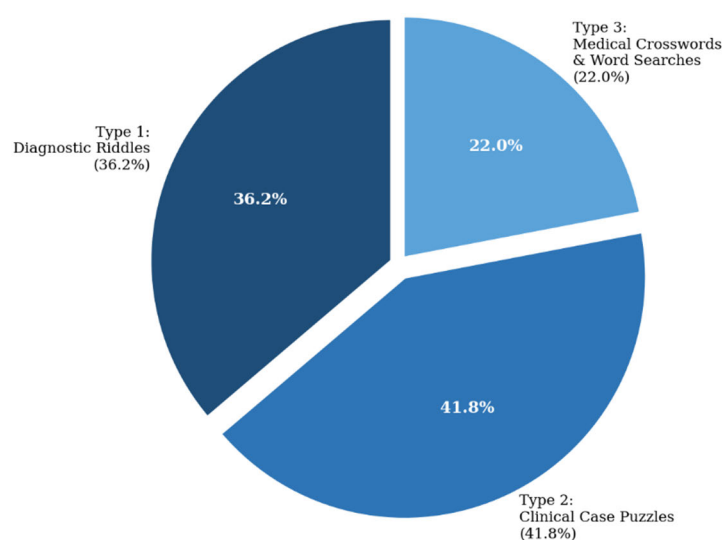


Figure 8. Student Preference Distribution: Most Preferred Puzzle Format Overall (n = 164)

DISCUSSION

The findings suggest that all three educational puzzle formats were highly rated, well-received active learning and assessment tools, with statistically significant differences in overall satisfaction scores across formats ($F(2,489) = 8.42$, $p = 0.0002$, $\eta^2 = 0.033$; pending recalculation via a repeated-measures model, see Section 2.5 and Section 8). Because puzzle format was fully confounded with subject area in this design (Type 1: Genetic Diseases; Type 2: Neoplasia/Haematology; Type 3: Renal Diseases), these differences cannot be attributed to puzzle format alone and may also reflect subject-matter difficulty or familiarity (see Section 8). These results are consistent with and meaningfully extend the existing literature supporting game-based and active learning in health sciences curricula [4,5,31-33].

Clinical Case Puzzles (Type 2) achieved the highest overall mean satisfaction (4.84 ± 0.36) and were rated significantly superior to Type 3 on most domains (Tukey HSD, $p < 0.001$). This difference is theoretically coherent: the clinical vignette extended-matching format simultaneously engages pattern recognition and diagnostic reasoning, requires

knowledge integration across multiple concepts and mirrors the problem-solving demands of clinical practice [23,24,34]. These characteristics align with constructivist learning theory and the Community of Inquiry framework [31,35].

Diagnostic Riddles (Type 1) did not differ significantly from Type 2 in overall satisfaction ($p = 0.087$) and achieved uniformly high agreement rates. Their strength lies in vocabulary acquisition and lower-order conceptual consolidation-domains foundational to health professional competence and often underserved by lecture-based instruction as with the Type 2 finding above, this comparison is confounded with subject area (Genetic Diseases vs. Neoplasia/Haematology) and cannot be attributed to format alone [1,2,11]. They are also the easiest format to construct, making them particularly accessible for faculty with limited preparation time [36,37].

Medical Crosswords (Type 3) received the lowest overall mean score (4.63 ± 0.48). This is theoretically expected: the absence of answer choices substantially increases cognitive demand and may create performance anxiety in students with lower prior knowledge [29,35,38,39]. The competitive

motivation domain received the lowest agreement across all three formats (74.3-88.4%), consistent with earlier literature on gamification [28,40,41]. Despite this, Type 3 still achieved >85% positive agreement on most items and remains the format most suited to higher-order clinical synthesis—a domain not reachable by Types 1 or 2 [11,34,42,43]. All three formats were administered to every student in the same fixed sequence (Type 1, then Type 2, then Type 3), which was not randomized or counterbalanced (see Section 8); consequently, novelty effects, fatigue, increasing familiarity with the Likert questionnaire or growing puzzle-solving proficiency over the course of the study cannot be ruled out as alternative explanations for the declining satisfaction pattern observed across formats, independent of or in addition to, format-specific cognitive demand. Because Type 3 also differed from Type 2 in both subject area and the absence of provided answer choices, lower satisfaction with Type 3 may reflect increased task difficulty and effort rather than or in addition to, its higher intended cognitive level; without an objective difficulty measure or cognitive-load assessment, this explanation remains an interpretation rather than an established finding.

The chi-square finding ($\chi^2 = 12.34$, $p = 0.002$) indicates that student preference for formal curriculum integration was significantly associated with format. Clinical Case Puzzles were most frequently preferred as the primary curriculum component (41.8%), supporting a tiered implementation model in which Diagnostic Riddles serve early-year vocabulary consolidation, Clinical Case Puzzles serve mid-year formative assessment and reasoning development and Medical Crosswords serve advanced-year synthesis and capstone preparation. This tiered model is conceptually informed by the “Reaction” level of the Kirkpatrick evaluation framework [37,44,45] the present study addresses only this first level and does not provide evidence relevant to the Learning, Behavior or Results levels of the framework.

CONCLUSIONS

Based on student-reported perceptions, this study suggests that all three educational puzzle formats were highly rated and their satisfaction scores were statistically distinguishable, as active learning and assessment tools in undergraduate health professional education. Because format was confounded with subject area (see Section 8) and no objective learning outcomes were measured, these conclusions should be interpreted as reflecting perceived usefulness rather than demonstrated educational effectiveness:

- Clinical Case Puzzles (Type 2) achieved the highest overall student satisfaction (mean 4.84 ± 0.36) and are most strongly recommended for formal curriculum integration and mid-year formative assessment
- Diagnostic Riddles (Type 1) are not significantly different from Type 2 in overall satisfaction and are optimal for lower-order vocabulary consolidation and terminology acquisition in early academic years
- Medical Crosswords (Type 3) are uniquely suited to higher-order clinical reasoning and synthesis (Bloom's Levels 4-6) and should be deployed strategically for advanced-year review
- A tiered, format-differentiated puzzle strategy—deploying all three formats according to cognitive demand across academic years—is strongly recommended for health sciences curricula
- All three formats promoted active engagement, collaborative learning, vocabulary acquisition and examination confidence, with >85% positive agreement across most questionnaire items
- Future research should include randomised controlled trials with objective pre-post knowledge assessments to establish causal relationships between puzzle format and academic performance

Recommendations

- **Tiered Curriculum Integration:** Implement Diagnostic Riddles (Year 1-2: Terminology), Clinical Case Puzzles (Year 2-3: Formative assessment and clinical reasoning) and Medical Crosswords (Year 3+: Synthesis and capstone review), aligned explicitly to Bloom's taxonomy
- **Formalise Assessment Role:** Institutional assessment policies could consider incorporating these puzzle formats as supplementary formative learning activities, pending further psychometric validation of the evaluation instrument and independent evidence of their impact on objective learning outcomes
- **Faculty Development:** Structured training workshops should assist faculty in designing and calibrating all three formats. Digital construction platforms (e.g., Educaplay, CrosswordLabs) should be adopted to reduce preparation burden
- **Digital and Hybrid Implementation:** These digital platforms were not evaluated in the present study; developing interactive digital versions adaptable to mobile devices and learning management systems (Blackboard, Moodle) is proposed here as a future implementation direction for independent evaluation, not as an evidence-based conclusion of this study
- **Collaborative Design:** Encourage multidisciplinary faculty teams to co-design puzzle content across modules for curricular coherence and cross-disciplinary integration
- **Longitudinal Research:** Multi-cohort longitudinal studies with objective pre-post assessments are needed to establish causal links between puzzle format engagement and academic performance outcomes

Limitations

- **Cross-Sectional Design:** Precludes causal inference between puzzle engagement and objective academic performance outcomes. Longitudinal or randomised controlled designs are required

- **Self-Report Bias:** All outcomes were student self-reports on Likert scales. Social desirability bias may have inflated agreement rates; anonymous administration mitigated but cannot fully eliminate this
- **Convenience Sampling:** Limits generalisability. Future studies should employ stratified probability sampling across multiple institutions
- **Absence of Objective Performance Data:** No pre-post examination scores were collected. Future research must incorporate objective knowledge assessment data to complement perceptual measures
- **Single-Institution Context:** Conducted at a single institution (University of Hail, Saudi Arabia); findings may not be directly generalisable to other institutions, regions or health-system contexts without replication
- **Instructor Variability:** Puzzle grids were designed and administered by different faculty across three colleges, introducing variability in grid quality and administration conditions not systematically controlled
- **Order Effect:** All students received the three formats in the same sequence; counterbalanced or randomized delivery orders would control order effects on perception ratings in future studies
- **Format-Subject Confounding:** Each puzzle format was implemented using a different subject area (Type 1: Genetic Diseases; Type 2: Neoplasia/Haematology; Type 3: Renal Diseases). Format and subject content are therefore fully confounded and observed differences in satisfaction cannot be definitively attributed to puzzle format alone; subject-matter difficulty, familiarity and interest may also have contributed
- **Repeated-Measures Statistical Handling:** The original analysis compared the three formats using a one-way, independent-groups ANOVA and chi-square test despite each of the 164 students contributing ratings for all three formats; this treats paired observations as independent and requires correction using a repeated-measures ANOVA, linear mixed-effects model or repeated-measures categorical method (see Section 2.5)
- **Group-Level Clustering:** Students completed puzzles cooperatively in groups of six to seven, which may introduce clustering effects (shared peer influence on perceptions) not accounted for in the reported analysis
- **Questionnaire Psychometric Limitations:** Beyond face/content review by two academics and pilot testing on 20 students, no internal-consistency reliability statistics (e.g., Cronbach's alpha) or factor-analytic validation of the seven proposed subdomains are reported for the full study sample in the original submission
- **Social Desirability Bias:** Non-anonymous-feeling, group-based administration may be subject to social desirability bias in addition to the self-report bias already noted above

Ethical Approval

Ethical approval was taken from the University of Hail, Saudi Arabia (H-2024-482). All participants provided

written informed consent. The study was conducted by the Declaration of Helsinki and institutional ethical guidelines.

REFERENCES

- [1] Freeman, S. *et al.* "Active Learning Increases Student Performance in Science, Engineering and Mathematics." *Proceedings of the National Academy of Sciences of the United States of America*, vol. 111, no. 23, 2014, pp. 8410-8415. <https://doi.org/10.1073/pnas.1319030111>.
- [2] Saran, R. and S. Kumar. "Use of Crossword Puzzle as a Teaching Aid to Facilitate Active Learning in Dental Materials." *Indian Journal of Applied Research*, vol. 5, no. 4, 2015, pp. 456-457. <https://doi.org/10.36106/ijar.2015.4.158>.
- [3] Gaikwad, N. and S. Tankhiwale. "Crossword Puzzles: Self-Learning Tool in Pharmacology." *Perspectives on Medical Education*, vol. 1, nos. 4-5, 2012, pp. 237-248. <https://doi.org/10.1007/s40037-012-0026-4>.
- [4] Barclay, S.M. *et al.* "Educational Card Games to Teach Pharmacotherapeutics in an Advanced Pharmacy Practice Experience." *American Journal of Pharmaceutical Education*, vol. 75, no. 2, 2011. <https://doi.org/10.5688/ajpe75233>.
- [5] Dichev, C. and D. Dicheva. "Gamifying Education: What Is Known, What Is Believed and What Remains Uncertain: A Critical Review." *International Journal of Educational Technology in Higher Education*, vol. 14, 2017. <https://doi.org/10.1186/s41239-017-0042-5>.
- [6] Hmelo-Silver, C.E. "Problem-Based Learning: What and How Do Students Learn?" *Educational Psychology Review*, vol. 16, no. 3, 2004, pp. 235-266. <https://doi.org/10.1023/B:EDPR.0000034022.16470.f3>.
- [7] Stewart, D.W. *et al.* "Active-Learning Processes Used in US Pharmacy Education." *American Journal of Pharmaceutical Education*, vol. 75, no. 4, 2011. <https://doi.org/10.5688/ajpe75468>.
- [8] Rodenbaugh, D.W. *et al.* "Having Fun and Accepting Challenges Are Natural Instincts: Jigsaw Puzzles to Challenge Students and Test Their Abilities While Having Fun!" *Advances in Physiology Education*, vol. 38, no. 2, 2014, pp. 185-186. <https://doi.org/10.1152/advan.00026.2014>.
- [9] Weisskirch, R.S. "An Analysis of Instructor-Created Crossword Puzzles for Student Review." *College Teaching*, vol. 54, no. 1, 2006, pp. 198-201. <https://doi.org/10.3200/CTCH.54.1.198-201>.
- [10] Davis, T.M. *et al.* "Reviewing for Exams: Do Crossword Puzzles Help in the Success of Student Learning?" *Journal of Effective Teaching*, vol. 9, no. 3, 2009, pp. 4-10.
- [11] Krathwohl, D.R. "A Revision of Bloom's Taxonomy: An Overview." *Theory into Practice*, vol. 41, no. 4, 2002, pp. 212-218. https://doi.org/10.1207/s15430421tip4104_2.
- [12] Saxena, A. *et al.* "Crossword Puzzles: Active Learning in Undergraduate Pathology and Medical Education." *Archives of Pathology & Laboratory Medicine*, vol. 133, no. 9, 2009, pp. 1457-1462. <https://doi.org/10.5858/133.9.1457>.
- [13] O'Leary, S. *et al.* "Educational Games in an Obstetrics and Gynecology Core Curriculum." *American Journal of Obstetrics and Gynecology*, vol. 193, no. 5, 2005, pp. 1848-1851. <https://doi.org/10.1016/j.ajog.2005.07.007>.
- [14] Murphy, M. *et al.* "Can Word Puzzles Be Tailored to Improve Different Dimensions of Verbal Fluency? An Intervention Study." *The Journal of Psychology*, vol. 150, no. 6, 2016, pp. 743-754. <https://doi.org/10.1080/00223980.2016.1186136>.

- [15] Eckert, G.U. et al. "Learning from Panel Boards: T-Lymphocyte and B-Lymphocyte Self-Tolerance Game." *Medical Teacher*, vol. 26, no. 6, 2004, pp. 521-524. <https://doi.org/10.1080/01421590412331271365>.
- [16] Dolmans, D.H.J.M. et al. "Problem-Based Learning: Future Challenges for Educational Practice and Research." *Medical Education*, vol. 39, no. 7, 2005, pp. 732-741. <https://doi.org/10.1111/j.1365-2929.2005.02205.x>.
- [17] Shah, S. et al. "Crossword Puzzles as a Tool to Enhance Learning about Anti-Ulcer Agents." *American Journal of Pharmaceutical Education*, vol. 74, no. 7, 2010. <https://doi.org/10.5688/aj7407117>.
- [18] Marcondes, F.K. et al. "A Puzzle Used to Teach the Cardiac Cycle." *Advances in Physiology Education*, vol. 39, no. 1, 2015, pp. 27-31. <https://doi.org/10.1152/advan.00070.2014>.
- [19] Crosby, J.R. and E.A. Hesketh. "Developing the Teaching Instinct, 11: Small Group Learning." *Medical Teacher*, vol. 26, no. 1, 2004, pp. 16-19. <https://doi.org/10.1080/01421590310001637977>.
- [20] DiPiro, J.T. "Why Do We Still Lecture?" *American Journal of Pharmaceutical Education*, vol. 73, no. 8, 2009. <https://doi.org/10.5688/aj7308137>.
- [21] Moore, L.S. and A.J. Detlaff. "Using Educational Games as a Form of Teaching in Social Work." *Arete*, vol. 29, no. 1, 2005, pp. 58-63.
- [22] Ritzko, J.M. and S. Robinson. "Using Games to Increase Active Learning." *Journal of College Teaching & Learning*, vol. 3, no. 6, 2006, pp. 45-50. <https://doi.org/10.19030/tlc.v3i6.1711>.
- [23] Smits, P.B.A. et al. "Problem Based Learning in Continuing Medical Education: A Review of Controlled Evaluation Studies." *BMJ*, vol. 324, no. 7330, 2002, pp. 153-156. <https://doi.org/10.1136/bmj.324.7330.153>.
- [24] Massey, A.P. et al. "It's All Fun and Games...Until Students Learn." *Journal of Information Systems Education*, vol. 16, no. 1, 2005, pp. 9-14.
- [25] Haidet, P. et al. "Analysis of the Team-Based Learning Literature: TBL Comes of Age." *Journal on Excellence in College Teaching*, vol. 25, nos. 3-4, 2014, pp. 303-333.
- [26] Pettit, R.K. et al. "Student Perceptions of Gamified Audience Response System Interactions in a Pathophysiology Course." *Medical Education Online*, vol. 20, 2015. <https://doi.org/10.3402/meo.v20.26570>.
- [27] Norman, G. "Likert Scales, Levels of Measurement and the 'Laws' of Statistics." *Advances in Health Sciences Education*, vol. 15, no. 5, 2010, pp. 625-632. <https://doi.org/10.1007/s10459-010-9222-y>.
- [28] Young, M.F. et al. "Our Princess Is in Another Castle: A Review of Trends in Serious Gaming for Education." *Review of Educational Research*, vol. 82, no. 1, 2012, pp. 61-89. <https://doi.org/10.3102/0034654312436980>.
- [29] Franklin, S. et al. "Non-Traditional Interventions to Stimulate Discussion: The Use of Games and Puzzles." *Journal of Biological Education*, vol. 37, no. 2, 2003, pp. 79-84. <https://doi.org/10.1080/00219266.2003.9655856>.
- [30] Keshavarzi, Z. et al. "Comparison of Student Satisfaction with Traditional and Integrated Teaching Methods in Physiology." *Educational Strategies in Medical Sciences*, vol. 8, no. 6, 2016, pp. 21-27.
- [31] Garrison, D.R. et al. "Critical Inquiry in a Text-Based Environment: Computer Conferencing in Higher Education." *The Internet and Higher Education*, vol. 2, nos. 2-3, 2000, pp. 87-105. [https://doi.org/10.1016/S1096-7516\(00\)00016-6](https://doi.org/10.1016/S1096-7516(00)00016-6).
- [32] Kerfoot, B.P. et al. "Interactive Spaced Education to Assess and Improve Knowledge of Clinical Practice Guidelines: A Randomised Controlled Trial." *Annals of Surgery*, vol. 249, no. 5, 2009, pp. 744-749. <https://doi.org/10.1097/SLA.0b013e31819a7301>.
- [33] Brar, J.S. et al. "Game-Based Learning in Medical Education: A Review." *Journal of Clinical and Diagnostic Research*, vol. 13, no. 11, 2019, pp. JE01-JE05. <https://doi.org/10.7860/JCDR/2019/41838.13305>.
- [34] Michael, J. "Where's the Evidence That Active Learning Works?" *Advances in Physiology Education*, vol. 30, no. 4, 2006, pp. 159-167. <https://doi.org/10.1152/advan.00053.2006>.
- [35] Thistlethwaite, J.E. et al. "The Effectiveness of Case-Based Learning in Health Professional Education: A BEME Systematic Review." *Medical Teacher*, vol. 34, no. 6, 2012, pp. e421-e444. <https://doi.org/10.3109/0142159X.2012.680939>.
- [36] Patrick, S. et al. "The Usefulness of Crossword Puzzle as a Self-Learning Tool in Pharmacology." *Journal of Advances in Medical Education & Professionalism*, vol. 6, no. 4, 2018, pp. 181-185. <https://doi.org/10.30476/JAMP.2018.41469>.
- [37] Kirkpatrick, D.L. and J.D. Kirkpatrick. *Evaluating Training Programs: The Four Levels*. 3rd ed., Berrett-Koehler, 2006.
- [38] Roediger, H.L. et al. "The Critical Role of Retrieval Practice in Long-Term Retention." *Trends in Cognitive Sciences*, vol. 15, no. 1, 2011, pp. 20-26. <https://doi.org/10.1016/j.tics.2010.09.003>.
- [39] Brown, P.C. et al. *Make It Stick: The Science of Successful Learning*. Harvard University Press, 2014.
- [40] Febtrina, R. et al. "The Use of Crossword Puzzles on Students' Interaction in Learning Vocabulary." *UNILA Journal of English Teaching*, vol. 3, no. 6, 2014, pp. 1-15.
- [41] Agarwal, P.K. et al. "Retrieval Practice Consistently Benefits Student Learning: A Systematic Review." *Educational Psychology Review*, vol. 33, no. 4, 2021, pp. 1409-1453. <https://doi.org/10.1007/s10648-021-09595-9>.
- [42] Makransky, G. and L. Lilleholt. "A Structural Equation Modeling Investigation of the Emotional Value of Immersive Virtual Reality in Education." *Educational Technology Research and Development*, vol. 66, no. 5, 2018, pp. 1141-1164. <https://doi.org/10.1007/s11423-018-9581-2>.
- [43] Mayer, R.E. *Multimedia Learning*. 2nd ed., Cambridge University Press, 2009. <https://doi.org/10.1017/CBO9780511811678>.
- [44] Black, P. and D. Wiliam. "Assessment and Classroom Learning." *Assessment in Education: Principles, Policy & Practice*, vol. 5, no. 1, 1998, pp. 7-74. <https://doi.org/10.1080/0969595980050102>.
- [45] Ayo Vaughan, A.F. and P.A. Amosun. "Effects of Two Modes of Active Learning Strategies on School-Age Children's Civic Competence." *Journal of Education and Social Research*, vol. 6, no. 2, 2016, pp. 213-224. <https://doi.org/10.5901/jesr.2016.v6n2p213>.