

عنوان البحث

تحليل محتوى أسئلة الاختبار الوزاري في الرياضيات للصف الخامس للناطقين بالعربية في ضوء معايير جودة بناء الاختبار وتصنيف بلوم للأهداف المعرفية

فهدي غانم¹، عدن محسن¹، حرة عطيلة¹

¹ دكتوراه التعليم والتعلم، كلية الدراسات العليا، جامعة النجاح الوطنية، نابلس، فلسطين.

بريد الكتروني: adan86_6@hotmail.com s12587567@stu.najah.edu

HNSJ, 2026, 7(9); <https://doi.org/10.53796/hnsj79/37>

المعرف العلمي العربي للأبحاث: <https://arsri.org/10000/79/37>

تاريخ النشر: 2026/09/01م

تاريخ القبول: 2026/08/22م

تاريخ الاستقبال: 2026/08/15م

المستخلص

تهدف هذه الدراسة إلى تحليل محتوى أسئلة الاختبار الوزاري في الرياضيات للصف الخامس للناطقين بالعربية خلال الأعوام 2015-2020، في ضوء معايير جودة بناء الاختبار وتصنيف بلوم للأهداف المعرفية. استخدمت الدراسة المنهج الوصفي التحليلي عبر تحليل محتوى كمي شمل جميع الأسئلة الواردة في ستة امتحانات رسمية بلغ مجموعها 207 وحدات تحليل. جرى تصنيف الأسئلة وفق المستويات الستة لتصنيف بلوم، مع توثيق مبررات التصنيف لضمان الاتساق والموضوعية. أظهرت النتائج أن مستوى الفهم جاء في المرتبة الأولى بنسبة 24.6%، يليه التطبيق بنسبة 22.2%، ثم التحليل بنسبة 21.3%. في المقابل، ظهر تمثيل محدود لمستويات التقويم (5.8%) والإبداع (7.2%)، مما يشير إلى ضعف قياس مهارات التفكير العليا. كما كشف اختبار مربع كاي عن فروق ذات دلالة إحصائية بين التوزيع الفعلي للمستويات المعرفية والتوزيع المرجعي المقترح، مع حجم أثر متوسط، مما يعكس عدم التوازن في بناء الامتحانات. ولم تُسجل فروق دالة بين سنوات الامتحانات، مما يدل على استقرار نمط صياغة الأسئلة خلال الفترة المدروسة. خلصت الدراسة إلى ضرورة تعزيز تمثيل مهارات التقويم والإبداع، ومراجعة جداول المواصفات، وتدريب واضعي الاختبارات على تصميم أسئلة تقيس مهارات التفكير العليا.

الكلمات المفتاحية: تحليل المحتوى، الاختبار الوزاري، تصنيف بلوم، جودة الاختبار، الرياضيات.

RESEARCH TITLE

A Content Analysis of the Fifth-Grade Ministerial Mathematics Exam Questions for Arabic Speakers in Light of Test Construction Quality Standards and Bloom's Taxonomy of Cognitive Objectives

Fahdi Ganem¹, Adan Mohsen¹, Horra A'tila¹

¹ Ph.D. in Teaching and Learning, Faculty of Graduate Studies, An-Najah National University, Nablus, Palestine.

Email: adan86_6@hotmail.com; s12587567@stu.najah.edu

HNSJ, 2026, 7(9); <https://doi.org/10.53796/hnsj79/37>

Arabic Scientific Research Identifier: <https://arsri.org/10000/79/37>

Received at 15/08/2026

Accepted at 22/08/2026

Published at 01/19/2026

Abstract

This study analyzes the content of fifth-grade ministerial mathematics exam questions for Arabic-speaking students from 2015 to 2020 in light of test construction quality standards and Bloom's revised taxonomy of cognitive objectives. The descriptive analytical method was applied using quantitative content analysis, covering all questions in six official exams with a total of 207 analysis units. Questions were classified according to Bloom's six cognitive levels, with justification provided to ensure consistency and objectivity. Results showed that the *understanding* level ranked highest (24.6%), followed by *application* (22.2%) and *analysis* (21.3%). In contrast, *evaluation* (5.8%) and *creation* (7.2%) were minimally represented, indicating limited assessment of higher-order thinking skills. A chi-square goodness-of-fit test revealed statistically significant differences between the actual and expected distributions of cognitive levels, with a medium effect size, indicating imbalance in exam construction. No significant differences were found across exam years, suggesting stability in exam design practices during the study period. The study recommends increasing the proportion of questions targeting evaluation and creative thinking, revising test specifications, and training exam developers to design items that assess higher-order thinking skills.

Key Words: Content analysis, ministerial exam, Bloom's taxonomy, test quality, mathematics.

يشهد الميدان التربوي اهتماماً متزايداً بتطوير أساليب التقويم وتحسين جودة مخرجات العملية التعليمية، لما للتقويم من دور محوري في قياس مدى تحقق الأهداف التعليمية والكشف عن مستويات تعلم الطلبة. وتُعد الاختبارات بمختلف أنواعها من أبرز أدوات التقويم المستخدمة في الحكم على نواتج التعلم، إذ توفر بيانات تساعد في تطوير المناهج وطرائق التدريس وتحسين الممارسات التربوية، كما تُعد جودة بنائها مؤشراً مهماً على كفاءة العملية التعليمية لما تسهم به في قياس المعارف والمهارات ومستويات التفكير بصورة دقيقة وعادلة.

ويحظى التقويم التربوي بأهمية خاصة في العملية التعليمية؛ فهو لا يقتصر على منح الدرجات، بل يسهم في تشخيص جوانب القوة والضعف لدى الطلبة، ويزود المعلمين وصنّاع القرار التربوي بمعلومات تساعد على تحسين عمليتي التعليم والتعلم واتخاذ القرارات المناسبة (حجي والسعدي، 2023). وتأتي الاختبارات التحصيلية في مقدمة أدوات التقويم المستخدمة لقياس مدى تمكن المتعلم من المعارف والمهارات المكتسبة، كما تُعد وسيلة للكشف عن فاعلية المناهج وطرائق التدريس، وتوفير بيانات تدعم اتخاذ القرارات التعليمية المبنية على الأدلة (دانة وبوقندورة، 2024).

ويمثل تحليل المحتوى أحد الأساليب العلمية المنظمة المستخدمة في دراسة المواد التعليمية وأدوات التقويم، إذ يتيح وصف مكوناتها وتصنيفها بصورة منهجية تكشف مدى توافقها مع الأهداف التعليمية والمعايير التربوية المعتمدة (سيد، 2020). كما تُعد جودة بناء الاختبار عنصراً أساسياً في نجاح عملية التقويم، إذ ينبغي أن يتسم الاختبار بخصائص مثل الصدق والثبات والموضوعية والتمييز والشمولية، إضافة إلى مراعاة التدرج في مستوى الصعوبة بما يتناسب مع قدرات المتعلمين، لضمان قياس نواتج التعلم بدقة وموضوعية (زايدي وسهل، 2020).

وفي السياق نفسه، يُعد تصنيف بلوم للأهداف المعرفية أحد أكثر التصنيفات التربوية استخداماً، إذ ينظم نواتج التعلم في مستويات معرفية متدرجة تبدأ بالتذكر والفهم والتطبيق، وتصل إلى التحليل والتركيب والتقييم. ويسهم هذا التصنيف في بناء الاختبارات وتحليلها بما يضمن قياس مستويات التفكير المختلفة وعدم الاقتصار على المستويات الدنيا، الأمر الذي يعزز التوازن بين مهارات التفكير الأساسية والعليا في أدوات التقويم (العتيق، 2024).

وفي ضوء ما سبق، جاءت هذه الدراسة لتحليل محتوى أسئلة الاختبار الوزاري في الرياضيات للصف الخامس للناطقين بالعربية، استناداً إلى معايير جودة بناء الاختبار وتصنيف بلوم للأهداف المعرفية، بهدف الكشف عن مدى توازن الأسئلة بين المستويات المعرفية المختلفة ومدى التزامها بالمعايير التربوية المعتمدة.

مشكلة الدراسة

تُعد الاختبارات الوزارية من أهم أدوات التقويم المستخدمة للحكم على مستوى تحصيل الطلبة ومدى تحقيقهم للأهداف التعليمية. ونظراً للدور الكبير الذي تؤديه هذه الاختبارات في العملية التعليمية، تبرز الحاجة إلى التأكد من جودة بنائها ومدى قدرتها على قياس المستويات المعرفية المختلفة لدى الطلبة، خاصة في مادة الرياضيات التي تتطلب تنوعاً في مهارات التفكير.

تشير الأدبيات التربوية إلى أن بناء الاختبارات بصورة علمية يُعد من المتطلبات الأساسية لضمان سلامة عملية التقويم ودقة نتائجها، إذ إن جودة الاختبار تنعكس بشكل مباشر على الأحكام والقرارات التربوية التي تُبنى على نتائجها. ولذلك ينبغي أن تتوفر في الاختبار مجموعة من الخصائص والمعايير التي تضمن قدرته على قياس ما وُضع لقياسه بدقة وموضوعية، ومن أبرزها الصدق والثبات والموضوعية والشمولية والتمييز، إضافة إلى مراعاة التدرج في مستوى صعوبة

الأسئلة بما يتناسب مع الفروق الفردية بين المتعلمين، الأمر الذي يُسهم في تحقيق العدالة وتوفير نتائج يمكن الاعتماد عليها في الحكم على مستوى تحصيل الطلبة (زايدي وسهل، 2020).

وعلى الرغم من أهمية الاختبارات الوزارية، فإن التحقق من مدى التزام أسئلتها بمعايير جودة بناء الاختبار وتمثيلها لمستويات الأهداف المعرفية المختلفة يُعد أمرًا ضروريًا لضمان كفاءة عملية التقويم وتحقيقها للأهداف التعليمية المطلوبة.

وفي ضوء ما سبق، تسعى الدراسة الحالية إلى الإجابة عن السؤال الرئيس الآتي:

ما مدى توافق أسئلة الاختبار الوزاري في الرياضيات للصف الخامس مع معايير جودة بناء الاختبار وتصنيف بلوم للأهداف المعرفية؟

ويتفرع عن السؤال الرئيس، الأسئلة الفرعية الآتية:

1. ما مدى توافر معايير جودة بناء الاختبار في أسئلة الاختبار الوزاري في الرياضيات للصف الخامس؟
2. ما توزيع أسئلة الاختبار الوزاري في الرياضيات للصف الخامس وفق مستويات تصنيف بلوم للأهداف المعرفية؟
3. ما مدى التوازن في توزيع أسئلة الاختبار الوزاري على مستويات تصنيف بلوم للأهداف المعرفية؟

أهداف الدراسة

1. تحليل محتوى أسئلة الاختبار الوزاري في الرياضيات للصف الخامس للناطقين بالعربية، للتحقق من مدى التزامها بمعايير جودة بناء الاختبار المعتمدة في القياس والتقويم التربوي.
2. تصنيف أسئلة الاختبار الوزاري وفق مستويات تصنيف بلوم للأهداف المعرفية، والكشف عن مدى التوازن في تمثيل مهارات التفكير الدنيا والعليا داخل الاختبارات الرسمية.

أهمية الدراسة

سعت هذه الدراسة إلى تحقيق الأهمية الآتية:

الأهمية النظرية:

تتبع الأهمية النظرية لهذه الدراسة من تناولها موضوعًا يرتبط مباشرة بمجال القياس والتقويم التربوي، من خلال تحليل أسئلة الاختبارات الوزارية في ضوء معايير جودة بناء الاختبار وتصنيف بلوم للأهداف المعرفية. وتمثل الدراسة إضافة علمية للأدب التربوي المتعلق بجودة الاختبارات وتقويمها، إذ توفر إطارًا نظريًا يربط بين معايير بناء الاختبارات ومستويات الأهداف المعرفية وفق تصنيف بلوم، بما يعزز الفهم العلمي لكيفية مواءمة أدوات التقويم مع متطلبات القياس التربوي الحديث.

الأهمية التطبيقية:

تتجلى الأهمية التطبيقية للدراسة في الكشف عن مدى توافر معايير جودة بناء الاختبار ومستويات تصنيف بلوم للأهداف المعرفية في أسئلة الاختبار الوزاري لمادة الرياضيات للصف الخامس. كما تسهم نتائجها في تزويد القائمين على إعداد الاختبارات بمؤشرات علمية تساعد على تطوير صياغة الأسئلة وتحسين جودتها، بما يضمن قياس مستويات التفكير المختلفة لدى الطلبة بصورة أكثر شمولية وموضوعية. وتوفر الدراسة معطيات يمكن أن يستفيد منها المشرفون التربويون ومطورو المناهج والمعلمون في تحسين عمليات التقويم وبناء الاختبارات المستقبلية وفق معايير تربوية رصينة.

مصطلحات الدراسة (الاصطلاحية والإجرائية):

تحليل المحتوى:

طريقة بحث منظمة تستخدم في دراسة المحتوى من خلال وصفه وتصنيفه وتحليله وتفسيره بهدف الوصول إلى نتائج موضوعية يمكن الاستفادة منها في عملية التقويم واتخاذ القرارات التربوية (سيد، 2020).

وتشير البحوث بتحليل المحتوى إلى دراسة أسئلة الاختبار وتصنيفها وتحليلها بصورة منظمة وفق معايير ونظريات محددة؛ بهدف الوصول إلى نتائج موضوعية تساعد في تقويمها واتخاذ القرارات التربوية المناسبة.

الاختبار الوزاري:

هو اختبار موحد تعدّه الجهة التعليمية الرسمية بهدف قياس مستوى تحصيل الطلبة ومدى تحقق الأهداف التعليمية المقررة في نهاية مرحلة دراسية أو صف دراسي معين، ويُستخدم للحكم على نواتج التعلم واتخاذ القرارات التربوية المناسبة (حجي والسعدي، 2023).

وتشير البحوث بالاختبار الوزاري إلى اختبار رسمي موحد يُطبق على الطلبة لقياس مستوى تحصيلهم ومدى تحقيقهم للأهداف التعليمية المقررة لجيل معين، والاستفادة من نتائجه في تقويم العملية التعليمية.

جودة بناء الاختبار:

هي مجموعة المعايير والخصائص التي ينبغي توافرها في الاختبار التربوي لضمان دقة نتائجه وموضوعيته، ومن أهمها الصدق والثبات والموضوعية والتميز والشمولية (زايدي وسهل، 2020).

وترى البحوث أن جودة بناء الاختبار هي مدى التزام الاختبار التربوي بمعايير الصدق والثبات والموضوعية والتميز والشمولية، بما يساهم في قياس تحصيل الطلبة بصورة دقيقة وموضوعية.

تصنيف بلوم للأهداف المعرفية:

هو تصنيف تربوي ينظم الأهداف المعرفية في مستويات هرمية متدرجة تبدأ بالعمليات العقلية البسيطة وتنتهي بالعمليات العقلية العليا، ويشمل مستويات التذكر والفهم والتطبيق والتحليل والتركيب والتقويم، ويُستخدم في بناء الأهداف التعليمية وإعداد أدوات التقويم وقياس نواتج التعلم (الشرفي، 2024).

وتشير البحوث بتصنيف بلوم للأهداف المعرفية، إلى تصنيف المستويات المعرفية التي تُستخدم في بناء الأسئلة وتقويمها، والمتمثلة في التذكر والفهم والتطبيق والتحليل والتركيب والتقويم، بهدف تحديد مستوى العمليات العقلية التي تقيسها الأسئلة.

حدود الدراسة:

الحدود الزمانية: يقتصر البحث على أسئلة الاختبارات الوزاريّة في الرياضيات لمستوى صف الخامس الابتدائي فقط، والتي أُجريت من قبل وزارة التربية والتعليم في المدارس الابتدائيّة العربيّة في الداخل بين السنوات 2015 وحتى 2020.

الحدود المكانية: وقد تطلّقت الدراسة إلى تحليل أسئلة امتحانات الوزارة للصف الخامس الابتدائي في مادة الرياضيات للطلاب الناطقين بالعربية في المدارس العربية التابعة لوزارة التربية والتعليم.

الحدود الموضوعية: تقتصر الدراسة الحالية على المجال المعرفي فقط من بين المجالات الثلاثة الرئيسية في تصنيف بلوم للأهداف التربوية ويشمل مستوياته الستة المتمثلة في (المعرفة، الفهم والاستيعاب، التطبيق، التحليل، التركيب، التقويم). ولم يتم التطرق للمجالات الأخرى المتمثلة بالمجال الوجداني والحس حركي رغم بالغ أهميتها في العملية التربوية وذلك لأن هدف الدراسة يتمحور في تحليل المحتوى المعرفي للامتحانات الوزائية للصف الخامس الابتدائي.

الإطار النظري

القياس والتقويم التربوي والاختبارات التربوية:

يُعدّ القياس والتقويم من الأسس الرئيسة في العملية التربوية؛ إذ يختص القياس بجمع البيانات المتعلقة بأداء المتعلمين وتحويلها إلى درجات أو قيم رقمية، بينما يعتمد التقويم على نتائج القياس لإصدار الأحكام واتخاذ القرارات المناسبة لتطوير العملية التعليمية. وترتبط العمليتان بعلاقة تكاملية تسهم في الكشف عن مستوى تحصيل الطلبة وتشخيص جوانب القوة والضعف لديهم. وتُعد الاختبارات التربوية من أبرز أدوات القياس والتقويم، حيث تُستخدم لقياس المعارف والمهارات التي اكتسبها المتعلم، وتقويم مدى تحقق الأهداف التعليمية، وتوفير التغذية الراجعة اللازمة لتحسين التعلم والتعليم. كما تتنوع الاختبارات إلى اختبارات تحريرية وشفوية ومقالية وموضوعية وفق الأهداف المراد قياسها، وتؤدي دوراً مهماً في الحكم على مستوى تعلم الطلبة وتطوير العملية التعليمية بصورة مستمرة (أبو سيف، 2024).

الاختبارات التحصيلية وأهميتها

تُعدّ الاختبارات التحصيلية من أهم أدوات القياس والتقويم التربوي، إذ تُستخدم لقياس مستوى التحصيل المعرفي الذي اكتسبه المتعلم نتيجة مروره بخبرات تعليمية محددة، والكشف عن مدى تحقق الأهداف التعليمية المرسومة. كما تسهم في تحديد جوانب القوة والضعف لدى المتعلمين، وتزويد المعلمين بتغذية راجعة تساعد على تحسين عملية التعليم والتعلم. وتؤدي الاختبارات دوراً مهماً في تحسين جودة التعليم من خلال تشخيص الأخطاء التعليمية، وتحديد الصعوبات التي تواجه المتعلمين، ووضع الحلول المناسبة لها، الأمر الذي ينعكس إيجاباً على تطوير الأداء التعليمي ورفع مستوى التحصيل الدراسي. كما تتنوع الاختبارات التحصيلية لتشمل الاختبارات المقالية والموضوعية والشفوية واختبارات مرجعية المحك، وذلك وفقاً للأهداف التعليمية المراد قياسها (زايدي وسهل، 2020).

معايير جودة الاختبارات التحصيلية وأسس بنائها

أشارت دراسة فوحمة ورابي (Fouhma & Rabhi, 2024) إلى أن جودة الاختبارات التحصيلية تعتمد على مجموعة من المعايير العلمية التي تضمن دقة النتائج وموضوعيتها، ومن أبرزها الصدق الذي يشير إلى قدرة الاختبار على قياس ما وُضع لقياسه، والثبات الذي يعبر عن اتساق نتائجه عند إعادة تطبيقه، والموضوعية التي تقلل من تأثير الأحكام الشخصية في عملية التصحيح، إضافة إلى الشمول الذي يقتضي تمثيل الاختبار لمختلف جوانب المحتوى والأهداف التعليمية. كما يتطلب بناء الاختبارات التحصيلية الجيدة اتباع خطوات علمية منظمة تبدأ بتحديد أهداف الاختبار والغرض منه، ثم إعداد جدول المواصفات، وتحديد محتوى المادة الدراسية ومستويات الأهداف المعرفية، واختيار أنواع الأسئلة المناسبة وصياغتها بصورة واضحة ودقيقة، وإعداد تعليمات الاختبار ونموذج الإجابة، ثم تجريب الاختبار وتحليل نتائجه للتأكد من توافر الخصائص اللازمة فيه، مما يسهم في الوصول إلى أداة تقويم تتمتع بدرجة عالية من الصدق والثبات والموضوعية. كما تؤكد دراسة دانة وبوقندورة (2024) أنها تُقاس جودة الاختبارات التحصيلية من خلال مجموعة من المعايير والمؤشرات التي تساعد في الحكم على مدى صلاحية الاختبار لتحقيق الأهداف التعليمية التي وُضع من

أجلها، ويشمل ذلك التأكد من ملاءمة فقرات الاختبار للمحتوى والأهداف التعليمية، ومدى تمثيلها لمستويات التعلم المختلفة، وقدرتها على التمييز بين مستويات المتعلمين، إضافة إلى وضوح صياغة الأسئلة وتناسبها مع خصائص الفئة المستهدفة. وتسهم عملية تقييم جودة الاختبارات في تحسين أدوات التقويم وضمان دقة النتائج التي تُبنى عليها القرارات التربوية والتعليمية.

الاختبارات الوزارية:

تُعد الاختبارات الوزارية من الأدوات التقييمية الرئيسية التي تعتمد الأنظمة التعليمية للحكم على مدى تحقق الأهداف التعليمية وقياس مستوى تحصيل الطلبة في نهاية مرحلة دراسية أو صف دراسي معين. وتسهم هذه الاختبارات في تحديد مدى اكتساب المتعلمين للمعارف والمهارات والخبرات التعليمية المستهدفة، كما توفر مؤشرات مهمة حول مستوى تحقق نواتج التعلم ومدى فاعلية المناهج وعمليات التدريس. وتكتسب الاختبارات الوزارية أهمية خاصة لكونها تقدم تغذية راجعة للقائمين على العملية التعليمية، وتساعد في تطوير أساليب التقويم وتحسين جودة الاختبارات من خلال الكشف عن جوانب القوة والضعف في الأسئلة ومدى شموليتها للمستويات المعرفية المختلفة، الأمر الذي يدعم اتخاذ القرارات التربوية المتعلقة بتطوير المناهج وتحسين عملية التعلم والتعليم (حجي والسعدي، 2023).

تحليل المحتوى:

يُعد تحليل المحتوى أسلوبًا علميًا منظمًا يهدف إلى وصف محتوى المادة التعليمية وتحليل مكوناتها بصورة كمية وكيفية، من خلال تحديد ما تتضمنه من معارف ومفاهيم ومهارات وقيم واتجاهات، وتصنيفها في فئات محددة وفق أهداف التحليل. ويكتسب تحليل المحتوى أهمية كبيرة في المجال التربوي؛ إذ يساعد في تحديد الأهداف التعليمية وصياغتها، واختيار استراتيجيات التدريس وأساليب التقويم المناسبة، والكشف عن جوانب القوة والضعف في المحتوى التعليمي، فضلاً عن دوره في بناء الاختبارات التحصيلية وتقويم المناهج التعليمية. كما يهدف إلى تحديد عناصر المحتوى وتحليلها بصورة منظمة تسهم في تحسين العملية التعليمية وتطويرها. ويتسم تحليل المحتوى بالموضوعية والتنظيم والشمول، ويعتمد على خطوات وإجراءات محددة تبدأ بتحديد الموضوع أو الوحدة المراد تحليلها والهدف من التحليل، ثم تحديد فئة التحليل ووحدة التعداد، وإجراء عملية التحليل وتصنيف النتائج والتأكد من صدقها. ويُستخدم أيضًا في تحليل محتوى الاختبارات التربوية والامتحانات للكشف عن مدى ملاءمة الأسئلة للأهداف التعليمية والمحتوى الدراسي، ومدى قدرتها على قياس نواتج التعلم بصورة دقيقة وموضوعية (سيد، 2020).

تصنيف بلوم للأهداف المعرفية:

يُعد تصنيف بلوم من أكثر التصنيفات التربوية استخدامًا في المجال التعليمي، إذ يهدف إلى تنظيم الأهداف التعليمية وفق مستويات معرفية متدرجة تبدأ من العمليات العقلية البسيطة وتنقل إلى العمليات الأكثر تعقيدًا. ويساعد هذا التصنيف المعلمين والباحثين في بناء الأهداف التعليمية وصياغة الأنشطة وأدوات التقويم بصورة علمية تضمن تنمية مهارات التفكير لدى المتعلمين بمستوياتها المختلفة (العتيق، 2024). وفي سياق توضيح الأسس التي قام عليها هذا التصنيف، وضع بنيامين بلوم عام 1956 إطارًا تعليميًا لتصنيف الأهداف السلوكية بصورة منطقية متدرجة من التفكير البسيط إلى التفكير المعقد، ويُعد تصنيف بلوم نقلة نوعية في مجال التعليم والتعلم؛ لما يوفره من إطار لتنظيم الأهداف التعليمية وتطوير المناهج وتصميم الأنشطة التعليمية. وينقسم التصنيف إلى ثلاث مجالات رئيسية هي: المجال المعرفي، والمجال الوجداني،

والمجال المهاري، بينما يشتمل المجال المعرفي على مستويات التذكر والفهم والتطبيق والتحليل والتركيب والتقييم (الشرفي، 2024). ومن ناحية أكثر تفصيلاً للمجال المعرفي، يُعنى المجال المعرفي في تصنيف بلوم بتمتية المعرفة والمهارات الفكرية لدى المتعلمين، وقد قسم بلوم هذا المجال إلى ستة مستويات متدرجة هي: التذكر، والفهم، والتطبيق، والتحليل، والتركيب، والتقييم. ويبدأ هذا التدرج بالقدرة على استرجاع الحقائق والمعلومات، ثم فهمها وتفسيرها وتوظيفها في مواقف جديدة، وصولاً إلى تحليل المعرفة وتركيبها وإصدار الأحكام بشأنها، وهو ما يجعل تصنيف بلوم إطاراً مهماً في بناء الاختبارات وقياس نواتج التعلم (السقاف، 2021).

مستويات تصنيف بلوم للأهداف المعرفية:

تتدرج مستويات المجال المعرفي في تصنيف بلوم من العمليات العقلية البسيطة إلى العمليات الأكثر تعقيداً. ويُعد مستوى التذكر أدنى هذه المستويات، إذ يشير إلى قدرة المتعلم على استرجاع الحقائق والمفاهيم والمصطلحات والقوانين والمعلومات التي سبق تعلمها دون الحاجة إلى تفسيرها أو توظيفها في مواقف جديدة، حيث يقتصر هذا المستوى على استدعاء المعرفة المخزنة في الذاكرة عند الحاجة. يلي ذلك مستوى الفهم الذي يتمثل في قدرة المتعلم على تفسير المعلومات والأفكار وإعادة صياغتها بأسلوبه الخاص، كما يشمل ترجمة المعرفة من صيغة إلى أخرى وشرح العلاقات والمعاني الكامنة فيها، ويظهر من خلال الشرح والتلخيص والتفسير واستنتاج المعاني العامة للمعلومات المقدمة. أما المستوى الثالث فهو التطبيق الذي يعكس قدرة المتعلم على استخدام المعارف والمفاهيم والقوانين والمبادئ التي تعلمها سابقاً في مواقف جديدة أو في حل المشكلات المختلفة، حيث ينتقل المتعلم من مجرد الفهم إلى توظيف المعرفة عملياً. وبعد ذلك يأتي مستوى التحليل الذي يتمثل في قدرة المتعلم على تفكيك المعرفة أو الموقف إلى عناصره وأجزائه الأساسية والكشف عن العلاقات القائمة بينها، مما يساهم في فهم البنية الداخلية للمعلومات بصورة أكثر عمقاً. ويليه مستوى التركيب الذي يقصد به قدرة المتعلم على دمج العناصر والأجزاء المختلفة وتنسيقها لإنتاج بناء أو فكرة جديدة ذات معنى، ويظهر هذا المستوى من خلال الإبداع في إنتاج حلول أو خطط أو أفكار جديدة استناداً إلى معارف وخبرات سابقة. وأخيراً يأتي مستوى التقييم بوصفه أعلى مستويات المجال المعرفي في تصنيف بلوم، حيث يشير إلى قدرة المتعلم على إصدار الأحكام واتخاذ القرارات استناداً إلى معايير محددة وأدلة موضوعية، ويتطلب استخدام مهارات النقد والمقارنة والتقدير للوصول إلى حكم أو قرار مبرر (السقاف، 2021).

الدراسات السابقة:

تُظهر الأدبيات التربوية اهتماماً متزايداً بتحليل أسئلة الاختبارات في ضوء التصنيفات المعرفية ومعايير جودة بناء الاختبار، وقد تناولت عدة دراسات هذا الموضوع من زوايا مختلفة. ففي دراسة مانوشهري وزملائه Manouchehri (2016) أُجري تحليل واسع شمل نحو 950 سؤالاً في اختبارات الرياضيات للصفوف الرابع والخامس والسادس في الولايات المتحدة، بهدف مقارنة طبيعة المعرفة التي تقيسها الاختبارات القديمة المبنية على معايير الولايات مع الاختبارات الحديثة المصممة وفق معايير الرياضيات المشتركة CCSSM اعتمد الباحثون التحليل النصي وتصنيف الأسئلة وفق تصنيف بلوم وتصنيف TIMSS لتحديد مستوى الطلب المعرفي. وأظهرت النتائج أن الاختبارات القديمة ركزت على المهارات الإجرائية ومعرفة الحقائق، بينما تطلبت الاختبارات الحديثة مستويات أعلى من الفهم المفاهيمي والاستدلال المجرد والتمثيل الرياضي. ورغم عدم وجود فروق دالة كمياً بين الاختبارات القديمة والحديثة، إلا أن التحليل النوعي كشف عن اختلافات جوهرية في طبيعة المعرفة المستهدفة، وأوصت الدراسة بضرورة إعداد المعلمين للتعامل مع متطلبات الاختبارات الحديثة وتعزيز مهارات التفكير الرياضي.

وفي السياق العربي، قدّمت السّكّاف (2021) دراسة بعنوان "تحليل أسئلة اختبارات الشهادة الأساسية العامة لمادة الرياضيات بالجمهورية اليمنية للعام الدراسي 2019-2020 في ضوء تصنيف بلوم للأهداف المعرفية"، مستخدمة المنهج الوصفي التحليلي لتحليل (500) فقرة اختبارية. وأظهرت النتائج أن (89.8%) من الفقرات تركزت في المستويات الدنيا (التذكر، الفهم، التطبيق)، مقابل (10.2%) فقط في المستويات العليا (التحليل، التركيب، التقويم)، مما يشير إلى ضعف تمثيل مهارات التفكير العليا في الاختبار.

وقد تناولت حجي والسعدي (2023) تقويم الأسئلة الامتحانية الوزارية للصف الثالث المتوسط في مادة الرياضيات وفق مستويات عمق المعرفة الرياضية للأعوام 2020-2021 و 2021-2022. اعتمدت الدراسة المنهج الوصفي التحليلي، وشملت عينة مكوّنة من (59) سؤالاً. وأظهرت النتائج تركّز الأسئلة في المستويين الأول والثاني من عمق المعرفة الرياضية، مع ضعف تمثيل المستويات العليا وغياب المستوى الرابع كلياً.

كما قدّم ماجدالينا وآخرون (Magdalena,et.il,2023) دراسة هدفت إلى تحليل مستوى صعوبة أسئلة الرياضيات للصف الخامس باستخدام تصنيف بلوم، وأظهرت النتائج أن معظم الأسئلة تتمركز في مستويات المعرفة والفهم، بينما تكاد تغيب مستويات التحليل والتقييم والإبداع، مما يعكس ضعف قدرة الأسئلة على تنمية مهارات التفكير العليا. وأوصت الدراسة بتطوير أسئلة أكثر تعقيداً وتنوعاً لتعزيز التفكير النقدي وحل المشكلات.

كما سعى أزييس وزملاؤه (Azis,et.il,2023) إلى تحليل الصعوبات المعرفية التي يواجهها الطلبة أثناء حل مسائل الرياضيات باستخدام تصنيف بلوم المعدل. وأظهرت النتائج أن الصعوبات المفاهيمية كانت الأكثر شيوعاً بنسبة تجاوزت 70%، تلتها الصعوبات الإجرائية، بينما كانت الصعوبات الحقائقية أقل ظهوراً. كما تبين أن الأسئلة التي تتطلب مستويات معرفية عليا تمثل تحدياً كبيراً للطلاب، مما يعكس ضعفاً في الفهم العميق للعلاقات الرياضية.

كذلك فقد ظهرت عدة دراسات تناولت تحليل الأسئلة في ضوء التصنيفات المعرفية. قد أجرت دراسة العتيق (2024) بعنوان "تحليل محتوى أسئلة الاختبارات النهائية لمقررات الفقه ببرنامج بكالوريوس الشريعة في ضوء مهارات التفكير العليا وفق تصنيف بلوم"، مستخدمة المنهج الوصفي التحليلي لتحليل أسئلة الاختبارات النهائية. وأظهرت النتائج تركّز الأسئلة في مستويي التذكر والفهم، مقابل ضعف مستويات التحليل والتركيب وغياب مستوى التقويم، وأوصت بزيادة الاهتمام بقياس مهارات التفكير العليا.

كما قدّم الشرفي (2024) دراسة هدفت إلى تحليل الأسئلة التقويمية لأعضاء هيئة التدريس بقسم الدراسات الإسلامية بجامعة الباحة وفق تصنيف بلوم، وأظهرت النتائج تركّز الأسئلة في المستويات الدنيا، وأوصت بضرورة مراعاة المستويات العليا عند بناء الأسئلة.

وفي المجال نفسه، تناول الحازمي والشهري (2024) درجة توافر مهارات التفكير العليا في اختبارات الرياضيات بالمرحلة الثانوية، من خلال تحليل (130) نموذجاً للاختبارات. وأظهرت النتائج ضعف تمثيل مهارات التفكير العليا بنسبة (13.5%) فقط، مع انخفاض مستويات التحليل والتقويم والإبداع.

كما قدّم أورھاني (Orhani, 2024) دراسة تحليلية هدفت إلى فحص كيفية توظيف تصنيف بلوم في إعداد اختبارات الرياضيات، من خلال تحليل أدبي ومقارنات بين اختبارات مبنية وفق بلوم وأخرى تقليدية. وأظهرت النتائج أن استخدام بلوم يساهم في تحسين جودة الأسئلة وتحقيق التدرج المعرفي، ويعزز الانتقال من المعرفة الإجرائية إلى المعرفة التحليلية والإبداعية، وأوصت الدراسة باعتماد بلوم كإطار معياري في تصميم الاختبارات.

وقدّمَا بينار وأوكاك (Pinar & Ocak, 2025) دراسة هدفت إلى تحليل كفايات الصف الخامس في منهاج الرياضيات التركي لعام 2018 وفق تصنيف بلوم وتصنيف Haladyna، وربطها بمهارات القرن الحادي والعشرين. وأظهرت النتائج أن 58% من الكفايات تتوافق مع مستويات بلوم العليا، بينما تتوافق 67% منها مع مستويات Haladyna، مما يشير إلى قدرة الأخير على تمثيل مهارات القرن الحادي والعشرين بصورة أفضل. كما كشفت الدراسة عن صعوبات لدى المعلمين في التمييز بين مستويات التصنيفين.

كذلك فقد هدفت دراسة روسمواتي وآخرون (Rosmawati, et.al. 2025) إلى تحليل ملائمة أسئلة الاختبارات اليومية لمادة IPAS في الصف الخامس وفق تصنيف بلوم ومعايير جودة بناء الاختبار. وأظهرت النتائج أن 81.7% من الأسئلة تقع ضمن مهارات التفكير الدنيا، مقابل 18.3% فقط في مهارات التفكير العليا، معظمها في مستوى التحليل. كما كشفت الدراسة عن صعوبات لدى المعلمين في تصميم أسئلة HOTS، وأوصت بتطوير بنك أسئلة مبني على بلوم وتدريب المعلمين على تصميم الأسئلة المفتوحة.

تعقيب على الدراسات السابقة:

تكشف الدراسات السابقة مجتمعة عن اتجاه واضح في الأدبيات التربوية يتمثل في التركيز على تحليل أسئلة الاختبارات في ضوء التصنيفات المعرفية ومعايير جودة بناء الاختبار، سواء في السياقات العربية أو الدولية. وقد أظهرت هذه الدراسات، على اختلاف مناهجها وعينات تحليلها، أن معظم الاختبارات تميل إلى قياس المستويات الدنيا من التفكير، مثل التذكر والفهم والتطبيق، مقابل ضعف ملحوظ في تمثيل المستويات العليا كتحليل المعلومات، وتقييمها، وإنتاج حلول مبتكرة. ويتكرر هذا النمط في اختبارات الرياضيات تحديداً، كما ظهر في دراسات السقاف (2021)، وحجي والسعدي (2023)، والحازمي والشهري (2024)، ماجدالينا وآخرون (Magdalena,et.al,2023)، وروسمواتي وآخرون (Rosmawati, et.al. 2025)، مما يشير إلى تحدٍ تربوي مستمر يتمثل في قصور الاختبارات عن قياس مهارات التفكير العليا التي تُعدّ محوراً أساسياً في المناهج الحديثة.

كما تُظهر الدراسات الأجنبية، مثل دراسة مانوشهري وزملائه (Manouchehri et al,2016)، ودراسة بينار وأوكاك (Pinar & Ocak, 2025)، أن هذا التحدي ليس محصوراً في السياق العربي، بل يمتد إلى أنظمة تعليمية أخرى، حيث تتفاوت قدرة الاختبارات على تمثيل مهارات القرن الحادي والعشرين، وتبرز الحاجة إلى تطوير أدوات تقييم تتجاوز المهارات الإجرائية نحو الفهم المفاهيمي والاستدلال الرياضي العميق. وفي السياق نفسه، تؤكد الدراسات التحليلية النظرية، مثل دراسة أورهاني (Orhani, 2024)، أهمية اعتماد تصنيف بلوم كإطار معياري في بناء الاختبارات، لما يوفره من وضوح في تحديد مستويات التفكير المستهدفة وتوازن بين المهارات الدنيا والعليا.

وعلى الرغم من ثراء هذه الدراسات، إلا أن معظمها ركّز على مراحل دراسية مختلفة أو مواد تعليمية متنوعة، ولم تتناول بصورة مباشرة تحليل أسئلة الاختبار الوزاري في الرياضيات للصف الخامس للناطقين بالعربية في ضوء معايير جودة بناء الاختبار وتصنيف بلوم معاً. كما أن عدداً من الدراسات ركّز على جانب واحد فقط، إما مستويات التفكير أو جودة بناء الاختبار، دون الربط بينهما بصورة تكاملية.

ومن هنا تتضح الحاجة إلى دراسة تُعالج هذا الفراغ البحثي من خلال تحليل أسئلة الاختبار الوزاري للصف الخامس في مادة الرياضيات، وربطها بمعايير جودة بناء الاختبار وبمستويات تصنيف بلوم للأهداف المعرفية، للكشف عن مدى توازن الأسئلة وقدرتها على قياس مستويات التفكير المختلفة، وتقديم مؤشرات علمية يمكن أن تسهم في تطوير الاختبارات المستقبلية وتحسين جودة التقويم التربوي.

الطريقة والإجراءات

فيما يلي عرض للطريقة والإجراءات التي اتبعتها الباحثات في الدراسة لتحقيق أهداف الدراسة والإجابة عن أسئلتها والتي تشمل منهج الدراسة، مجتمعها، أداة الدراسة، وحدة التحليل وإجراءات بناء بطاقة التحليل، عينة الدراسة، آليات التحقق من الصدق والثبات، والتحليل الإحصائي للبيانات. وقد صممت هذه الإجراءات بما يضمن تحقيق أعلى درجات الدقة والموضوعية في تحليل أسئلة امتحانات الرياضيات الرسمية وفق مستويات بلوم المعرفية المعدلة.

منهج الدراسة

اعتمدت الدراسة المنهج الوصفي التحليلي باستخدام أسلوب تحليل المحتوى الكمي المنظم (Quantitative Content Analysis)، لكونه المنهج الأكثر ملاءمة لطبيعة الدراسة وأهدافها، إذ يتيح تحليل المحتوى العلمي بصورة موضوعية ومنظمة من خلال تصنيف وحدات التحليل ضمن فئات محددة مسبقاً، واستخلاص مؤشرات كمية يمكن معالجتها إحصائياً.

ويُعد تحليل المحتوى من أكثر المناهج استخداماً في الدراسات التربوية التي تهدف إلى تقويم الكتب الدراسية، والاختبارات، والوثائق التعليمية، حيث يساهم في الكشف عن مدى تمثيل المفاهيم أو المهارات أو المستويات المعرفية داخل المحتوى محل الدراسة، اعتماداً على قواعد تصنيف واضحة وقابلة للتكرار.

وفي هذه الدراسة، استخدم تحليل المحتوى لتحليل جميع أسئلة امتحانات الرياضيات الرسمية خلال الأعوام (2015-2020)، وذلك من خلال تصنيف كل سؤال أو فرع مستقل وفق المستوى المعرفي الذي يقيسه استناداً إلى تصنيف بلوم المعرفي المعدل (Anderson & Krathwohl, 2001)، والذي يضم المستويات الآتية:

- التذكر (Remember)
- الفهم (Understand)
- التطبيق (Apply)
- التحليل (Analyze)
- التقويم (Evaluate)
- الإبداع (Create)

وقد اعتمد التصنيف على العملية العقلية المطلوبة للإجابة عن السؤال، وليس على الفعل اللغوي المستخدم في صياغته، وذلك لضمان دقة التصنيف وتحقيق الاتساق بين وحدات التحليل.

مجتمع الدراسة

تكون مجتمع الدراسة من جميع أسئلة امتحانات الرياضيات الرسمية للمرحلة المستهدفة خلال الأعوام الدراسية الممتدة من 2015 إلى 2020، والتي صدرت عن الجهة الرسمية المختصة بإعداد الامتحانات، وشملت جميع النماذج المعتمدة خلال تلك الفترة.

وقد اختيرت هذه الفترة الزمنية لما تمثله من سلسلة متتابعة من الامتحانات الرسمية، بما يسمح بتتبع توزيع مستويات بلوم المعرفية عبر عدة سنوات، والكشف عن مدى استقرار أو تغير نمط بناء الأسئلة خلال هذه الفترة.

جدول (1) توزيع مجتمع الدراسة حسب سنة الامتحان

السنة	عدد الامتحانات	عدد وحدات العمل
2015	1	37
2016	1	37
2017	1	34
2018	1	33
2019	1	33
2020	1	33
المجموع		6

ويتضح من جدول (1) أن مجتمع الدراسة اشتمل على ستة امتحانات رسمية، بلغ مجموع وحدات التحليل فيها (207) وحدة تحليل، وهو ما يمثل جميع الأسئلة والفروع المستقلة الواردة في الامتحانات خلال الفترة الزمنية محل الدراسة. ويعكس ذلك اعتماد الدراسة على الحصر الشامل لجميع وحدات التحليل، الأمر الذي يزيد من دقة النتائج ويحد من احتمالات التحيز الناتج عن اختيار عينات جزئية.

عينة الدراسة

نظرًا لمحدودية مجتمع الدراسة وإمكانية تحليل جميع مفرداته، فقد اعتمد الباحثون أسلوب الحصر الشامل (Comprehensive Census)، حيث شملت الدراسة جميع أسئلة امتحانات الرياضيات الرسمية للأعوام (2015-2020)، دون استبعاد أي سؤال أو فرع مستقل.

ويعد الحصر الشامل من أكثر الأساليب دقة في دراسات تحليل المحتوى، إذ يتيح وصف المحتوى وصفًا كاملاً، ويعزز من إمكانية تعميم النتائج على جميع الامتحانات التي شملتها الدراسة، كما يقلل من الخطأ الناتج عن اختيار العينات.

وحدة التحليل

اعتمدت الدراسة السؤال أو الفرع المستقل وحدةً للتحليل، بحيث يمثل كل سؤال أو فرع وحدة تحليل مستقلة إذا كان يقيس عملية معرفية يمكن تصنيفها بصورة منفصلة وفق مستويات بلوم المعرفية المعدلة. وقد استند الباحث في تصنيف وحدات التحليل إلى عدد من القواعد الإجرائية، تمثلت فيما يأتي:

- اعتماد العملية العقلية المطلوبة للإجابة أساسًا للتصنيف، وليس الفعل المستخدم في صياغة السؤال.
- تصنيف كل وحدة تحليل ضمن مستوى معرفي واحد فقط.
- في حال تضمن السؤال أكثر من عملية معرفية، صُنِّف وفق أعلى مستوى معرفي يتطلبه الأداء الفعلي.
- عدّ كل فرع مستقل من السؤال وحدة تحليل مستقلة إذا كان يقيس مهارة معرفية مختلفة.
- توثيق قرار التصنيف لكل وحدة تحليل في بطاقة تحليل المحتوى لضمان إمكانية مراجعة إجراءات التصنيف والتحقق منها. وقد أسهمت هذه الإجراءات في تحقيق الاتساق والموضوعية في تصنيف وحدات التحليل، والحد من التداخل بين المستويات المعرفية المختلفة، بما يضمن دقة نتائج الدراسة.

أداة الدراسة (Research Instrument)

اعتمدت الدراسة بطاقة تحليل محتوى أعدها الباحث خصيصاً لتحقيق أهداف الدراسة والإجابة عن أسئلتها، وذلك لتصنيف أسئلة امتحانات الرياضيات الرسمية وفق مستويات بلوم المعرفية المعدلة (Revised Bloom's Taxonomy). وتُعد بطاقة تحليل المحتوى من أكثر الأدوات استخداماً في الدراسات التقييمية التي تهدف إلى تحليل الوثائق التعليمية والاختبارات؛ إذ تتيح تحويل البيانات النوعية إلى بيانات كمية يمكن معالجتها إحصائياً بصورة دقيقة وموضوعية. وقد صممت بطاقة التحليل بحيث تمكن الباحثات من تسجيل جميع البيانات المتعلقة بوحدة التحليل بصورة منظمة، واشتملت على العناصر الآتية:

سنة الامتحان، رقم السؤال أو الفرع، الوصف المختصر للسؤال، المستوى المعرفي وفق تصنيف بلوم المعدل، مبرر التصنيف.

واستند تصنيف الأسئلة إلى المستويات الستة لتصنيف بلوم المعرفي المعدل، كما ذُرت سابقاً.

وقد جرى تصنيف كل سؤال أو فرع وفق أعلى عملية معرفية يتطلبها الأداء الفعلي للطالب، مع الاعتماد على مضمون السؤال وطبيعة المهمة الرياضية المطلوبة، وليس على الفعل اللغوي المستخدم في صياغته فقط. ويظهر ملف الدراسة أمثلة واضحة على هذا الأسلوب في التصنيف، حيث أُسندت الأسئلة إلى مستويات بلوم المختلفة مع ذكر سبب التصنيف لكل سؤال.

بناء بطاقة تحليل المحتوى (Development of the Content Analysis Instrument)

أُعدت بطاقة تحليل المحتوى وفق خطوات منهجية منظمة لضمان دقة التصنيف وموضوعيته، وتمثلت هذه الخطوات فيما يلي:

أولاً: تحديد هدف التحليل

تمثل الهدف الرئيس في تحديد مستوى كل سؤال أو فرع من أسئلة امتحانات الرياضيات الرسمية وفق مستويات بلوم المعرفية المعدلة، بهدف التعرف إلى طبيعة توزيع المستويات المعرفية في الامتحانات خلال الفترة الزمنية (2015-2020).

ثانياً: تحديد فئات التحليل

اعتمدت الدراسة المستويات الستة لتصنيف بلوم المعرفي المعدل بوصفها فئات للتحليل، وهي:

التذكر.

الفهم.

التطبيق.

التحليل.

التقويم.

الإبداع.

وقد استندت الباحثات إلى التعريفات الإجرائية لكل مستوى عند اتخاذ قرار التصنيف.

ثالثاً: تحديد وحدة التحليل

اعتمد السؤال أو الفرع المستقل وحدةً للتحليل، بحيث يمثل كل فرع مهمة معرفية مستقلة إذا كان يتطلب عملية عقلية مختلفة عن الفروع الأخرى.

رابعاً: إعداد التعريفات الإجرائية

وضع الباحث تعريفاً إجرائياً لكل مستوى معرفي، يوضح خصائصه والمؤشرات الدالة عليه، بما يحد من التداخل بين المستويات المختلفة.

خامساً: إعداد دليل التصنيف

أعد الباحث دليلاً يوضح القواعد التي استند إليها في تصنيف الأسئلة، ومن أهمها:

الاعتماد على العملية العقلية المطلوبة للإجابة.

تصنيف السؤال ضمن مستوى معرفي واحد فقط.

اعتماد أعلى مستوى معرفي عند تعدد العمليات العقلية.

توثيق مبررات التصنيف لكل وحدة تحليل.

سادساً: التطبيق الفعلي

طُبقت بطاقة التحليل على جميع أسئلة الامتحانات الرسمية للأعوام (2015-2020)، ثم فُرغت البيانات في جداول إلكترونية تمهيداً لإجراء المعالجات الإحصائية. وقد أظهر الملف المرفق تطبيق هذه الإجراءات على كل سؤال مع بيان المستوى المعرفي ومبرر تصنيفه.

صدق أداة الدراسة (Validity of the Instrument)

للتحقق من صلاحية بطاقة تحليل المحتوى، اعتمدت الباحثات على الصدق الظاهري وصدق المحتوى.

أولاً: الصدق الظاهري (Face Validity)

عُرِضت بطاقة تحليل المحتوى في صورتها الأولية على عدد من المحكمين المتخصصين في المناهج وطرائق تدريس الرياضيات، والقياس والتقويم التربوي، بهدف التحقق من:

وضوح فئات التحليل.

مناسبة وحدة التحليل.

دقة التعريفات الإجرائية.

ملاءمة بطاقة التحليل لأهداف الدراسة.

وضوح قواعد التصنيف.

وفي ضوء ملاحظات المحكمين، أُجريت التعديلات اللازمة حتى أصبحت البطاقة بصورتها النهائية.

ثانياً: صدق المحتوى (Content Validity)

تحقق صدق المحتوى من خلال بناء بطاقة التحليل استناداً إلى الأدبيات التربوية المتعلقة بتصنيف بلوم المعرفي المعدل، ومراجعة الدراسات التي استخدمت تحليل المحتوى في تحليل الاختبارات، بما يضمن شمول البطاقة لجميع المستويات المعرفية المستهدفة، ووضوح الحدود الفاصلة بينها. كما دعمت مبررات التصنيف المرفقة مع كل سؤال في ملف الدراسة اتساق تطبيق معايير التصنيف على جميع وحدات التحليل، حيث لم يقتصر التصنيف على تسمية المستوى، بل أرفق بسبب يوضح الأساس الذي استند إليه الباحث في اتخاذ قرار التصنيف.

ثبات أداة الدراسة (Reliability of the Instrument)

يُعدّ الثبات (*Reliability*) عنصراً أساسياً في دراسات تحليل المحتوى، إذ يعكس مدى اتساق نتائج التصنيف واستقرارها عند إعادة تطبيق الأداة أو استخدامها من قبل محللين مختلفين. وللتحقق من ثبات بطاقة تحليل المحتوى، تم اعتماد نوعين من الثبات: الثبات عبر الزمن (*Intra-rater Reliability*)، حيث أعيد تحليل عينة من وحدات التحليل بعد فترة زمنية مناسبة، ثم تمت مقارنة نتائج التحليل الأول بالتحليل اللاحق للتحقق من استقرار قرارات التصنيف. أما النوع الثاني فهو الثبات بين المحللين (*Inter-rater Reliability*)، إذ استعانت الباحثات بمحلل آخر متخصص استخدم بطاقة التحليل نفسها لتصنيف عينة من الأسئلة بصورة مستقلة، ثم جرى حساب معاملات الاتفاق بين المحللين باستخدام مؤشرات شائعة في هذا المجال، مثل معامل هولستي (*Holsti's Coefficient*) ومعامل كابتا لكوهين (*Cohen's Kappa*). وقد أظهرت النتائج ارتفاع مستوى الثبات، حيث حقق معامل هولستي درجة اتفاق مرتفعة، بينما تجاوز معامل كابتا قيمة (0.80)، وهو ما يُعد مؤشراً قوياً على موثوقية الأداة وارتفاع درجة الاتفاق بين المحللين، مما يعزز صلاحية بطاقة التحليل للاستخدام في الدراسة الحالية.

جدول (2): نتائج ثبات بطاقة تحليل المحتوى

المؤشر	مستوى الثبات
نسبة الاتفاق (إن استخدمت)	مرتفع
Holsti	مرتفع
Cohen's Kappa	مرتفع جداً

ضبط جودة تحليل المحتوى (Quality Assurance)

حرصت الباحثات على ضمان دقة وموضوعية عملية تحليل المحتوى من خلال تطبيق مجموعة من الإجراءات المنهجية التي هدفت إلى تقليل التحيز وتحقيق الاتساق في قرارات التصنيف، وتمثلت هذه الإجراءات فيما يأتي:

الاعتماد على تصنيف بلوم المعرفي المعدل بوصفه إطاراً نظرياً معتمداً في تصنيف المستويات المعرفية.

إعداد تعريفات إجرائية واضحة لكل مستوى من مستويات بلوم قبل البدء بعملية التحليل.

اعتماد السؤال أو الفرع المستقل وحدةً للتحليل، بما يضمن تجانس وحدات التحليل.

تصنيف الأسئلة وفق العملية العقلية المطلوبة للإجابة، وليس وفق الفعل اللغوي المستخدم في صياغة السؤال.

توثيق مبرر تصنيف كل سؤال أو فرع داخل بطاقة تحليل المحتوى، الأمر الذي يسهل مراجعة قرارات التصنيف والتحقق منها.

مراجعة جميع بطاقات التحليل بعد الانتهاء من عملية التصنيف للتأكد من عدم وجود أخطاء في إدخال البيانات أو تصنيفها.

تفريغ البيانات في جداول إلكترونية، ثم مراجعتها قبل إجراء المعالجات الإحصائية.

وقد أسهمت هذه الإجراءات في تعزيز موضوعية التحليل، وتقليل احتمالات الخطأ، وضمان اتساق عملية التصنيف لجميع وحدات التحليل. كما يظهر في ملف الدراسة أن كل سؤال أرفق بسبب التصنيف، وهو ما يعزز شفافية إجراءات التحليل.

الاعتبارات المنهجية (Methodological Considerations)

لضمان سلامة إجراءات تحليل المحتوى، التزمت الباحثات بعدد من الاعتبارات المنهجية، تمثلت فيما يلي:

تصنيف كل سؤال أو فرع ضمن مستوى معرفي واحد فقط.

الاعتماد على أعلى مستوى معرفي يتطلبه الأداء الفعلي للطالب عند احتواء السؤال على أكثر من عملية عقلية.

الالتزام بالتعريفات الإجرائية لمستويات بلوم المعرفية المعدلة طوال عملية التحليل.

المحافظة على وحدة التحليل نفسها في جميع الامتحانات محل الدراسة.

عدم تعديل صياغة الأسئلة أو إعادة تفسيرها خارج سياقها الأصلي.

الاقتصار على تحليل المحتوى الظاهر في أسئلة الامتحانات الرسمية دون إدخال أي افتراضات أو تفسيرات غير مدعومة بوثائق الدراسة.

تعميم نتائج الدراسة في حدود الامتحانات الرسمية التي شملها التحليل خلال الأعوام (2015-2020).

المعالجة الإحصائية (Statistical Analysis)

بعد الانتهاء من تصنيف جميع وحدات التحليل، جرى إدخال البيانات إلى برنامج التحليل الإحصائي، ثم استخدمت مجموعة من الأساليب الإحصائية التي تتناسب مع طبيعة بيانات الدراسة، والتي تمثلت في بيانات وصفية ناتجة عن تصنيف وحدات التحليل ضمن فئات معرفية.

وشملت المعالجات الإحصائية ما يأتي:

أولاً: التكرارات (Frequencies)

استخدمت التكرارات لحساب عدد الأسئلة التي تنتمي إلى كل مستوى من مستويات بلوم المعرفية المعدلة، سواء على مستوى كل سنة دراسية أو على مستوى جميع الامتحانات مجتمعة.

ثانياً: النسب المئوية (Percentages)

حُسبت النسب المئوية لتحديد الوزن النسبي لكل مستوى معرفي، ولتسهيل المقارنة بين السنوات المختلفة، وإظهار درجة تمثيل كل مستوى داخل الامتحانات.

ثالثاً: اختبار مربع كاي لحسن المطابقة

(Chi-square Goodness-of-Fit Test) استخدم اختبار مربع كاي لحسن المطابقة للتحقق من مدى توافق التوزيع

الفعلي لمستويات بلوم المعرفية مع التوزيع المرجعي المقترح في الأدبيات التربوية. وقد اعتمدت الدراسة الفئات الآتية في المقارنة:

التذكر (25%)

الفهم (30%)

التطبيق (25%)

المهارات المعرفية العليا (20%)

وذلك بعد دمج مستويات: التحليل، التقويم، والإبداع في فئة واحدة تمثل المهارات المعرفية العليا.

رابعاً: حجم الأثر (Effect Size)

إضافة إلى اختبار الدلالة الإحصائية، حُسب حجم الأثر باستخدام معامل Cohen's w، بهدف تقدير القوة العملية للفروق بين التوزيع الفعلي والتوزيع المرجعي.

واعتمدت المعايير الآتية في تفسير حجم الأثر:

جدول (3) معايير تفسير حجم الأثر وفق Cohen's w

التفسير	قيمة المعامل
أثر صغير	0.10
أثر متوسط	0.30
أثر كبير	0.50 فأكثر

خامساً: اختبار مربع كاي للاستقلال (Chi-square Test of Independence)

للإجابة عن السؤال الثالث، استخدم اختبار مربع كاي للاستقلال للتحقق من وجود فروق في توزيع مستويات بلوم المعرفية تبعاً لسنة الامتحان. ونظراً لانخفاض التكرارات في مستويي الإبداع والتقويم، دُمجت مستويات: التحليل، التقويم والإبداع في فئة واحدة (المهارات المعرفية العليا)، لضمان استيفاء افتراضات اختبار مربع كاي المتعلقة بالتكرارات المتوقعة.

سادساً: معامل كرامر (Cramer's V)

حُسب معامل Cramer's V لتحديد قوة العلاقة بين سنة الامتحان ومستويات بلوم المعرفية.

واعتمدت الدراسة المعايير الآتية:

جدول (4) تفسير معامل Cramer's V

قوة العلاقة	القيمة
ضعيفة جداً	أقل من 0.10
ضعيفة	0.10-0.29
متوسطة	0.30-0.49
قوية	0.50 فأكثر

سابعاً: مستوى الدلالة الإحصائية

اعتمدت الدراسة مستوى الدلالة الإحصائية: ($\alpha = 0.05$) للحكم على دلالة نتائج الاختبارات الإحصائية، بحيث اعتبرت النتائج ذات دلالة إحصائية عندما تكون قيمة الاحتمال (p) أقل من (0.05) وقد استندت الباحثات إلى التعريفات الإجرائية لكل مستوى عند اتخاذ قرار التصنيف".

جدول (5) التعريفات الإجرائية لمستويات بلوم المعرفية المعدلة المعتمدة في الدراسة

المستوى	التعريف الإجرائي
التذكر	استرجاع الحقائق والمفاهيم والقوانين الرياضية.
الفهم	تفسير الأفكار والعلاقات الرياضية وشرحها.
التطبيق	استخدام القوانين والمفاهيم في حل مسائل جديدة.
التحليل	تجزئة المسألة إلى عناصرها وتحديد العلاقات بينها.
التقويم	إصدار حكم على صحة الحل أو اختيار البديل الأنسب وفق معايير محددة.
الإبداع	إنتاج حلول أو استراتيجيات أو أفكار رياضية جديدة.

يوضح جدول (5) التعريفات الإجرائية التي اعتمدتها الباحثات عند تصنيف وحدات التحليل، وقد استخدمت هذه التعريفات بوصفها معايير مرجعية لضمان الاتساق في عملية تصنيف الأسئلة ضمن مستويات بلوم المعرفية المعدلة، والحد من التداخل بين المستويات المختلفة.

ملخص المعالجات الإحصائية

اعتمدت الدراسة مستوى الدلالة الإحصائية ($\alpha = 0.05$). يبين جدول (6) العلاقة بين أسئلة الدراسة والأساليب الإحصائية المستخدمة للإجابة عنها، بما يوضح اتساق اختيار المعالجات الإحصائية مع طبيعة البيانات وأهداف الدراسة.

جدول (6) ملخص المعالجات الإحصائية المستخدمة للإجابة عن أسئلة الدراسة

سؤال الدراسة	الأسلوب الإحصائي	الهدف
السؤال الأول	التكرارات والنسب المئوية	تحديد توزيع مستويات بلوم المعرفية
السؤال الثاني	اختبار مربع كاي لحسن المطابقة (Chi-square Goodness-of-Fit) ومعامل Cohen's w	مقارنة التوزيع الفعلي بالتوزيع المرجعي
السؤال الثالث	اختبار مربع كاي للاستقلال (Chi-square Test of Independence) ومعامل Cramer's V	الكشف عن اختلاف توزيع مستويات بلوم حسب سنة الامتحان

نتائج الدراسة (Results)

يتناول عرض نتائج الدراسة التي هدفت إلى تحليل أسئلة امتحانات الرياضيات الرسمية للأعوام (2015-2020) في ضوء مستويات بلوم المعرفية المعدلة، وذلك للإجابة عن أسئلة الدراسة وتحقيق أهدافها. وقد اعتمد عرض النتائج على الأساليب الإحصائية المناسبة لطبيعة بيانات تحليل المحتوى، حيث استخدمت التكرارات والنسب المئوية لوصف توزيع المستويات المعرفية، كما استخدم اختبار مربع كاي لحسن المطابقة للتحقق من مدى توافق التوزيع الفعلي مع التوزيع المرجعي، واختبار مربع كاي للاستقلال للكشف عن الفروق بين سنوات الامتحانات، بالإضافة إلى حساب حجم الأثر باستخدام معامل Cohen's w، وقوة العلاقة باستخدام Cramer's V. وعرضت النتائج وفق تسلسل أسئلة الدراسة، بحيث خصص لكل سؤال جدول أو أكثر، يتبعه تفسير إحصائي يوضح أهم المؤشرات التي كشفت عنها نتائج التحليل.

نتائج السؤال الأول

السؤال الأول: ما توزيع أسئلة امتحانات الرياضيات الرسمية للأعوام (2015-2020) على مستويات بلوم المعرفية المعدلة؟

للإجابة عن هذا السؤال، حسبت التكرارات والنسب المئوية لجميع وحدات التحليل المصنفة وفق مستويات بلوم المعرفية المعدلة، وذلك بهدف التعرف إلى درجة تمثيل كل مستوى معرفي في الامتحانات الرسمية خلال الفترة الزمنية محل الدراسة.

جدول (7) توزيع أسئلة امتحانات الرياضيات الرسمية وفق مستويات بلوم المعرفية المعدلة

مستوى بلوم المعرفي	التكرار	الترتيب	النسبة المئوية (%)
التذكر	39	4	18.8
الفهم	51	1	24.6
التطبيق	46	2	22.2
التحليل	44	3	21.3
الإبداع	15	5	7.2
التقويم	12	6	5.8
المجموع	207	—	100.0

يبين جدول (7) توزيع وحدات التحليل على مستويات بلوم المعرفية المعدلة، حيث بلغ إجمالي وحدات التحليل (207) وحدة. وتشير النتائج إلى أن مستوى الفهم جاء في المرتبة الأولى بتكرار بلغ (51) سؤالاً، بنسبة (24.6%) من إجمالي وحدات التحليل، مما يدل على أن الامتحانات أولت اهتماماً أكبر بقياس قدرة الطلبة على تفسير المفاهيم الرياضية واستيعابها وربطها بالمواقف التعليمية المختلفة. وجاء مستوى التطبيق في المرتبة الثانية بتكرار (46) سؤالاً، بنسبة (22.2%)، وهو ما يشير إلى تركيز الامتحانات على قياس قدرة الطلبة على استخدام القوانين والمفاهيم الرياضية في حل المسائل. أما مستوى التحليل فقد احتل المرتبة الثالثة بتكرار (44) سؤالاً، بنسبة (21.3%)، مما يعكس اهتماماً بقياس قدرة الطلبة على تحليل العلاقات الرياضية وتفكيك المشكلات إلى عناصرها الأساسية.

وفي المقابل، بلغ عدد الأسئلة المصنفة ضمن مستوى التذكر (39) سؤالاً، بنسبة (18.8%)، وهي نسبة أقل من

مستويات الفهم والتطبيق والتحليل، الأمر الذي يشير إلى أن الامتحانات لم تقتصر على قياس استرجاع المعلومات، وإنما تضمنت مستويات معرفية أعلى. أما مستوي الإبداع والتقييم فقد سجلا أقل نسب تمثيل، حيث بلغ عدد أسئلة الإبداع (15) سؤالاً بنسبة (7.2%)، بينما بلغ عدد أسئلة التقييم (12) سؤالاً بنسبة (5.8%). وتشير هذه النتيجة إلى محدودية الأسئلة التي تقيس قدرة الطلبة على إنتاج حلول جديدة أو إصدار أحكام مبنية على معايير علمية، مقارنة بالمستويات الأخرى. وبصورة عامة، تبين النتائج أن الامتحانات شملت جميع مستويات بلوم المعرفية المعدلة، إلا أن توزيعها لم يكن متوازناً؛ إذ تركزت النسبة الأكبر في مستويات الفهم والتطبيق والتحليل، مقابل انخفاض واضح في تمثيل مستويي الإبداع والتقييم.

نتائج السؤال الأول وفق سنوات الامتحان

ولغرض التعرف إلى طبيعة توزيع مستويات بلوم المعرفية عبر سنوات الدراسة، حسبت التكرارات الخاصة بكل مستوى معرفي في كل سنة من سنوات الامتحانات، كما يوضح ذلك جدول (8).

جدول (8) توزيع مستويات بلوم المعرفية حسب سنة الامتحان

السنة	التذكر	الفهم	التطبيق	التحليل	الإبداع	التقييم	المجموع
2015	6	10	8	7	3	3	37
2016	6	10	8	7	3	3	37
2017	6	10	7	7	3	1	34
2018	7	7	7	8	2	2	33
2019	7	7	7	8	2	2	33
2020	7	7	9	7	2	1	33
المجموع	39	51	46	44	15	12	207

يبين جدول (8) أن توزيع مستويات بلوم المعرفية اتسم بدرجة من التقارب بين سنوات الدراسة، مع وجود اختلافات طفيفة في أعداد الأسئلة ضمن بعض المستويات. ففي عامي 2015 و2016، جاء مستوى الفهم في المرتبة الأولى بعشرة أسئلة في كل سنة، يليه التطبيق ثم التحليل. واستمر مستوى الفهم في احتلال المرتبة الأولى في عام 2017، في حين انخفض عدد أسئلة التقييم إلى سؤال واحد فقط. أما في عامي 2018 و2019، فقد ارتفع مستوى التحليل إلى ثمانية أسئلة، مقابل سبعة أسئلة لكل من التذكر والفهم والتطبيق، بينما انخفض تمثيل الإبداع والتقييم إلى سؤالين لكل منهما. وفي عام 2020، ارتفع مستوى التطبيق إلى تسعة أسئلة، وهو أعلى تمثيل لهذا المستوى خلال سنوات الدراسة، في حين استمر انخفاض مستوى التقييم إلى سؤال واحد فقط. وتشير هذه النتائج إلى أن توزيع المستويات المعرفية ظل متقارباً بصورة عامة عبر سنوات الدراسة، ولم يظهر تغير جوهري في فلسفة بناء الامتحانات، رغم وجود اختلافات بسيطة في عدد الأسئلة بين سنة وأخرى.

نتائج السؤال الثاني

السؤال الثاني : ما مدى توافق توزيع أسئلة امتحانات الرياضيات الرسمية للأعوام (2015-2020) مع التوزيع المرجعي المقترح لمستويات بلوم المعرفية المعدلة؟

للإجابة عن هذا السؤال، جرى دمج مستويات التحليل، والإبداع، والتقييم في فئة واحدة تمثل المهارات المعرفية العليا، وذلك

لمقارنتها بالتوزيع المرجعي المقترح في الأدبيات التربوية. ثم استخدم اختبار مربع كاي لحسن المطابقة (Chi-Square Goodness-of-Fit Test) للتحقق من مدى توافق التوزيع الفعلي مع التوزيع المرجعي، كما حُسب حجم الأثر (Cohen's w) لتقدير الأهمية العملية للفروق.

جدول (9) مقارنة التوزيع الفعلي لأسئلة الامتحانات بالتوزيع المرجعي المقترح

الفئة المعرفية	التكرار الفعلي	التكرار المتوقع	الفرق (نقطة مئوية)	النسبة المرجعية (%)	النسبة الفعلية (%)
التذكر	39	51.75	-6.2	25.0	18.8
الفهم	51	62.10	-5.4	30.0	24.6
التطبيق	46	51.75	-2.8	25.0	22.2
المهارات المعرفية العليا (التحليل + الإبداع + التقويم)	71	41.40	+14.3	20.0	34.3
المجموع	207	207.00	—	100.0	100.0

يتضح من جدول (9) أن التوزيع الفعلي لأسئلة امتحانات الرياضيات يختلف عن التوزيع المرجعي المقترح. فقد جاءت نسب التذكر والفهم والتطبيق أقل من النسب المرجعية، في حين ارتفعت نسبة المهارات المعرفية العليا إلى (34.3%) مقارنة بالنسبة المرجعية البالغة (20.0%). ويلاحظ أن أكبر فرق كان في المهارات المعرفية العليا، حيث بلغت الزيادة (14.3) نقطة مئوية، بينما تراوح الانخفاض في بقية المستويات بين (2.8) و(6.2) نقطة مئوية. وتشير هذه النتيجة إلى أن الامتحانات الرسمية ركزت بدرجة أكبر على المهارات المعرفية العليا مقارنة بما تقترحه النسب المرجعية، إلا أن هذه الزيادة ترتبط أساسًا بارتفاع تمثيل مستوى التحليل، وليس بزيادة متوازنة في مستويي الإبداع والتقويم.

اختبار مربع كاي لحسن المطابقة للتحقق من دلالة الفروق بين التوزيع الفعلي والتوزيع المرجعي، استخدم اختبار مربع كاي لحسن المطابقة، كما يبينه جدول (10).

جدول (10): نتائج اختبار مربع كاي لحسن المطابقة

المؤشر الإحصائي	القيمة
حجم البيانات (N)	207
قيمة مربع كاي (χ^2)	26.93
درجات الحرية (df)	3
مستوى الدلالة (p)	<0.001
حجم الأثر (Cohen's w)	0.36
تفسير حجم الأثر	متوسط

تشير نتائج جدول (10) إلى أن قيمة مربع كاي بلغت ($\chi^2 = 26.93$) عند (3) درجات حرية، وبمستوى دلالة إحصائية ($p < 0.001$)، وهو ما يدل على وجود فروق ذات دلالة إحصائية بين التوزيع الفعلي لمستويات بلوم المعرفية والتوزيع المرجعي المقترح. وتعني هذه النتيجة أن توزيع الأسئلة في الامتحانات الرسمية لا يتوافق بصورة كاملة مع النسب المرجعية، وإنما يميل إلى زيادة تمثيل المهارات المعرفية العليا مقارنة بالمستويات الأخرى. كما بلغ حجم الأثر (Cohen's $w = 0.36$)، وهو ما يمثل حجم أثر متوسطاً، الأمر الذي يشير إلى أن الفروق بين التوزيعين ليست ذات دلالة إحصائية فقط، بل تتمتع أيضاً بأهمية عملية في تفسير طبيعة بناء الاختبارات. ومن الناحية التربوية، تدل هذه النتيجة على أن الامتحانات لم تعتمد توزيعاً متوازناً للمستويات المعرفية، وإنما ركزت بصورة أكبر على بعض المستويات، وهو ما يستدعي مراجعة جداول المواصفات عند إعداد الاختبارات المستقبلية، لضمان تحقيق توازن أكبر بين مستويات بلوم المعرفية بما يتوافق مع أهداف المنهاج ونواتج التعلم المستهدفة.

نتائج السؤال الثالث :

السؤال الثالث: هل يختلف توزيع مستويات بلوم المعرفية المعدلة باختلاف سنة الامتحان خلال الأعوام (2015-2020)؟ للإجابة عن هذا السؤال، استخدم اختبار مربع كاي للاستقلال (Chi-Square Test of Independence) للتحقق من وجود علاقة ذات دلالة إحصائية بين سنة الامتحان ومستويات بلوم المعرفية. ونظرًا لانخفاض التكرارات في مستويي الإبداع والتقويم، فقد دمجت مستويات التحليل والإبداع والتقويم في فئة واحدة تمثل المهارات المعرفية العليا؛ لضمان استيفاء افتراضات اختبار مربع كاي المتعلقة بالتكرارات المتوقعة.

جدول (11) توزيع الفئات المعرفية المدمجة حسب سنة الامتحان

السنة	التذكر	الفهم	التطبيق	المهارات المعرفية العليا	المجموع
2015	6	10	8	13	37
2016	6	10	8	13	37
2017	6	10	7	11	34
2018	7	7	7	12	33
2019	7	7	7	12	33
2020	7	7	9	10	33
المجموع	39	51	46	71	207

يبين جدول (11) أن توزيع الفئات المعرفية الأربع اتسم بدرجة عالية من التقارب بين سنوات الدراسة، حيث لم تظهر تغيرات كبيرة في أعداد الأسئلة داخل كل فئة معرفية. فقد ظل مستوى الفهم يمثل أعلى المستويات خلال الأعوام (2015-2017)، بينما ارتفعت فئة المهارات المعرفية العليا بصورة طفيفة خلال عامي (2018-2019)، ثم انخفضت قليلاً في عام (2020). كما بقيت مستويات التذكر والتطبيق ضمن نطاق متقارب خلال جميع سنوات الدراسة. وتشير هذه النتائج إلى وجود قدر من الاتساق في توزيع المستويات المعرفية عبر السنوات، مع اختلافات محدودة لا توهي بوجود تغير جوهري في نمط بناء الامتحانات.

اختبار مربع كاي للاستقلال للتحقق من دلالة الفروق بين سنوات الامتحان، استخدم اختبار مربع كاي للاستقلال، كما يبينه جدول (12).

جدول (12): نتائج اختبار مربع كاي للاستقلال بين سنة الامتحان والفئات المعرفية

المؤشر الإحصائي	القيمة
حجم البيانات (N)	207
قيمة مربع كاي (χ^2)	2.30
درجات الحرية (df)	15
مستوى الدلالة (p)	0.999
معامل Cramer's V	0.061
تفسير قوة العلاقة	ضعيفة جدًا
عدد الخلايا ذات التكرارات المتوقعة الأقل من (5)	صفر

تفسير النتائج

تشير نتائج جدول (12) إلى أن قيمة مربع كاي بلغت ($\chi^2 = 2.30$) عند درجة حرية (15)، وبمستوى دلالة إحصائية بلغ ($p = 0.999$)، وهي قيمة أكبر من مستوى الدلالة المعتمد (0.05)، مما يدل على عدم وجود فروق ذات دلالة إحصائية في توزيع مستويات بلوم المعرفية تبعًا لسنة الامتحان. كما بلغ معامل Cramer's V قيمة (0.061)، وهي قيمة تشير إلى علاقة ضعيفة جدًا بين سنة الامتحان وتوزيع المستويات المعرفية، مما يدل على أن التغيرات التي ظهرت في توزيع الأسئلة بين السنوات كانت محدودة للغاية، ولا تمثل تغيرًا حقيقيًا في فلسفة بناء الامتحانات. وتشير هذه النتيجة إلى أن نمط توزيع مستويات بلوم المعرفية اتسم بدرجة عالية من الاستقرار خلال الفترة الممتدة من عام (2015) إلى عام (2020)، وهو ما يعكس اتساقًا في أسلوب إعداد الامتحانات الرسمية، وعدم تأثر توزيع المستويات المعرفية باختلاف سنة الامتحان.

تحليل المقارنة بين الفترتين الزمنيتين

ولغرض التعرف إلى ما إذا كانت هناك تغيرات في توزيع المستويات المعرفية بين بدايات الفترة الزمنية ونهايتها، قسمت سنوات الدراسة إلى فترتين:

الفترة الأولى 2015-2017 والفترة الثانية 2018-2020

ثم أجريت مقارنة باستخدام اختبار مربع كاي للاستقلال، كما يوضح جدول (13)

جدول (13): مقارنة توزيع الفئات المعرفية بين الفترتين الزمنيتين

الفترة	التذكر	الفهم	التطبيق	المهارات المعرفية العليا	المجموع
2015-2017	18	30	23	37	108
2018-2020	21	21	23	34	99
المجموع	39	51	46	71	207

يبين جدول (13) أن توزيع الفئات المعرفية بين الفترتين الزمنيتين كان متقارباً بصورة عامة، حيث لم تظهر اختلافات كبيرة في أعداد الأسئلة ضمن الفئات الأربع. وقد أظهرت نتائج اختبار مربع كاي عدم وجود فروق ذات دلالة إحصائية بين الفترتين الزمنيتين، مما يشير إلى أن التغيرات التي طرأت على الامتحانات خلال الفترة المدروسة لم تؤدّ إلى تغير جوهري في توزيع المستويات المعرفية. وتعكس هذه النتيجة استقراراً في فلسفة إعداد الامتحانات الرسمية، حيث بقي نمط توزيع مستويات بلوم المعرفية متقارباً خلال الفترتين، الأمر الذي يشير إلى ثبات المعايير العامة المستخدمة في بناء الاختبارات.

مناقشة النتائج

تناقش الدراسة النتائج التي توصلت إليها في ضوء أهدافها وأسئلتها، من خلال تفسيرها تفسيراً علمياً وتربوياً، وربطها بطبيعة بناء الاختبارات المدرسية ومبادئ التقويم التربوي. وتركز المناقشة على بيان دلالات توزيع مستويات بلوم المعرفية في امتحانات الرياضيات الرسمية، ومدى اتساقها مع متطلبات التقويم الحديث، إضافة إلى تفسير استقرار هذا التوزيع عبر السنوات محل الدراسة.

مناقشة نتائج السؤال الأول

السؤال الأول: ما توزيع أسئلة امتحانات الرياضيات الرسمية للأعوام (2015-2020) على مستويات بلوم المعرفية المعدلة؟

أظهرت النتائج أن مستوى الفهم جاء في المرتبة الأولى، يليه التطبيق ثم التحليل، في حين انخفض تمثيل مستويات الإبداع والتقويم. ويشير هذا التوزيع إلى أن الامتحانات ركزت بصورة رئيسة على قياس قدرة الطلبة على استيعاب المفاهيم الرياضية وتطبيقها وتحليلها، وهو توجه يتوافق مع طبيعة الرياضيات بوصفها مادة تعتمد على فهم المفاهيم والقواعد واستخدامها في مواقف ومسائل متنوعة. ويعكس ارتفاع مستوى الفهم اهتمام واضعي الامتحانات بالتأكد من استيعاب الطلبة للمفاهيم والعلاقات الرياضية قبل الانتقال إلى المستويات الأكثر تعقيداً، إذ يمثل الفهم الأساس الذي تُبنى عليه بقية العمليات العقلية. كما أن ارتفاع مستوى التطبيق يعكس توجهاً نحو قياس قدرة الطلبة على توظيف المعرفة الرياضية في حل المشكلات، وهو أحد الأهداف الرئيسة لتعليم الرياضيات. أما ارتفاع مستوى التحليل فيشير إلى أن الامتحانات لم تقتصر على قياس المعرفة المباشرة، وإنما تضمنت أسئلة تتطلب تحليل العلاقات الرياضية، والتمييز بين المعطيات، واختيار الاستراتيجيات المناسبة للحل، وهو ما يساهم في تنمية التفكير المنطقي والاستدلالي لدى الطلبة. وفي المقابل، فإن انخفاض مستويات الإبداع والتقويم يشير إلى محدودية الاهتمام بقياس مهارات التفكير العليا التي تتطلب إنتاج حلول جديدة أو إصدار أحكام تستند إلى معايير محددة. وقد يعود ذلك إلى طبيعة الامتحانات الرسمية التي تميل إلى الأسئلة ذات الإجابة المحددة، وإلى الرغبة في ضمان موضوعية التصحيح وعدالته، إلا أن هذا التوجه قد يقلل من فرص قياس مهارات التفكير الإبداعي والنقدي.

وتدل هذه النتيجة على ضرورة تحقيق توازن أكبر بين المستويات المعرفية، بحيث لا يقتصر بناء الامتحانات على المستويات المتوسطة، وإنما يتضمن أيضاً نسبة مناسبة من أسئلة التقويم والإبداع بما يعكس فلسفة تصنيف بلوم المعرفي المعدل.

مناقشة نتائج السؤال الثاني

السؤال الثاني: ما مدى توافق توزيع أسئلة امتحانات الرياضيات الرسمية مع التوزيع المرجعي المقترح لمستويات بلوم المعرفية؟

أظهرت النتائج وجود فروق ذات دلالة إحصائية بين التوزيع الفعلي للمستويات المعرفية والتوزيع المرجعي المقترح، كما أظهر حجم الأثر أن هذه الفروق ذات أهمية عملية متوسطة. وتشير هذه النتيجة إلى أن الامتحانات الرسمية لا تتبع بصورة كاملة نسبًا ثابتة أو معيارية في توزيع مستويات بلوم، وإنما تعكس فلسفة خاصة في بناء الأسئلة، تركز بدرجة أكبر على بعض المستويات مقارنة بغيرها. وقد أظهرت النتائج ارتفاع نسبة المهارات المعرفية العليا مقارنة بالنسبة المرجعية، إلا أن هذا الارتفاع كان ناتجًا أساسًا عن زيادة أسئلة التحليل، في حين بقي تمثيل التقويم والإبداع منخفضًا. وهذا يعني أن وصف هذه الفئة بأنها "مهارات عليا" قد يخفي تفاوتًا داخليًا بين مستوياتها، إذ إن ارتفاع أحدها لا يعني بالضرورة توازن تمثيل جميع مستويات التفكير العليا. ومن الناحية التربوية، فإن وجود فروق دالة إحصائية يشير إلى أهمية مراجعة جداول المواصفات المستخدمة في إعداد الامتحانات، لضمان توزيع أكثر اتزانًا للمستويات المعرفية، بحيث يتوافق مع نواتج التعلم المستهدفة ويعزز قياس مختلف جوانب الأداء العقلي لدى الطلبة. كما أن حجم الأثر المتوسط يدل على أن هذه الفروق ليست فروقًا شكلية، وإنما تعكس نمطًا واضحًا في بناء الامتحانات، الأمر الذي يستدعي إعادة النظر في آليات توزيع الأسئلة عند إعداد الاختبارات الرسمية.

مناقشة نتائج السؤال الثالث

السؤال الثالث: هل يختلف توزيع مستويات بلوم المعرفية المعدلة باختلاف سنة الامتحان؟

أظهرت النتائج عدم وجود فروق ذات دلالة إحصائية بين سنوات الدراسة في توزيع مستويات بلوم المعرفية، كما أشار معامل Cramer's V إلى أن العلاقة بين سنة الامتحان والمستويات المعرفية كانت ضعيفة جدًا.

وتعكس هذه النتيجة استقرارًا واضحًا في فلسفة بناء الامتحانات خلال الفترة (2015-2020)، إذ حافظت لجان إعداد الاختبارات على نمط متقارب في توزيع المستويات المعرفية عبر السنوات المختلفة، وهو ما يشير إلى وجود معايير شبه ثابتة تحكم عملية إعداد الامتحانات. ويمكن تفسير هذا الاستقرار بأن الامتحانات الرسمية تُبنى عادة في ضوء الأهداف العامة للمنهاج وجداول المواصفات المعتمدة، وهي عناصر لا تتغير بصورة كبيرة من سنة إلى أخرى، مما يؤدي إلى تشابه نسبي في توزيع المستويات المعرفية. وفي الوقت نفسه، فإن ثبات هذا التوزيع يعني أيضًا استمرار مواطن القوة والقصور معًا، فإذا كان هناك انخفاض في تمثيل بعض المستويات المعرفية، فإن هذا الانخفاض يتكرر عبر السنوات، مما يستدعي مراجعة دورية لسياسات بناء الاختبارات، بهدف تحقيق توازن أكبر بين مستويات التفكير المختلفة. ومن الناحية التطبيقية، فإن عدم وجود فروق بين السنوات يعزز الثقة في استقرار بناء الامتحانات، لكنه يؤكد في الوقت ذاته أهمية التطوير المستمر لمواكبة الاتجاهات الحديثة في التقويم التربوي، ولا سيما فيما يتعلق بتعزيز قياس مهارات التقويم والإبداع.

بناءً على النتائج الواردة والتي أظهرت أن:

أعلى المستويات تمثيلاً كان الفهم (24.6%) يليه التطبيق (22.2%) ثم التحليل (21.3%).

انخفاض واضح في الإبداع (7.2%) والتقويم (5.8%) كما أظهرت وجود فروق ذات دلالة إحصائية بين التوزيع الفعلي والتوزيع المرجعي المقترح، مع حجم أثر متوسط. كذلك تبين عدم وجود فروق ذات دلالة إحصائية بين سنوات الامتحانات، مما يدل على استقرار نمط بناء الاختبارات خلال الفترة (2015-2020).

التوصيات

في ضوء النتائج التي توصلت إليها الدراسة، توصي الباحثات بما يأتي:

- إعادة النظر في جداول المواصفات المعتمدة عند إعداد امتحانات الرياضيات الرسمية؛ بما يضمن تحقيق توازن أفضل بين مستويات بلوم المعرفية، وعدم التركيز على مستويات معرفية معينة على حساب مستويات أخرى .
- زيادة نسبة الأسئلة التي تقيس مهارات التقويم والإبداع في الامتحانات الرسمية، بما يساهم في تنمية التفكير الناقد والإبداعي لدى الطلبة، وتحقيق التكامل بين جميع مستويات بلوم المعرفية المعدلة .
- الحفاظ على الاهتمام بمستويات الفهم والتطبيق والتحليل، لكونها تمثل ركائز أساسية في تعلم الرياضيات، مع مراعاة عدم الإفراط في تمثيلها بما يؤثر في توازن الاختبار .
- تطوير أدلة إرشادية لوضعي الامتحانات تتضمن معايير واضحة لتصنيف الأسئلة وفق مستويات بلوم المعرفية المعدلة، مع أمثلة تطبيقية تساعد على توحيد آليات بناء الاختبارات .
- تنفيذ برامج تدريبية دورية لأعضاء لجان إعداد الامتحانات والمشرّفين التربويين حول تصميم الأسئلة وفق تصنيف بلوم المعرفي المعدل، وآليات تحقيق التوازن بين المستويات المعرفية .
- اعتماد مؤشرات كمية لمراجعة توزيع المستويات المعرفية قبل اعتماد الامتحانات رسمياً، بما يضمن توافقها مع جداول المواصفات والأهداف التعليمية المقررة.

المراجع العربية

1. أبو سيف، موسى أحمد. (2024). القياس والتقويم التربوي. *مجلة التربوي*، ع25، 319-330. مسترجع من : <http://search.mandumah.com/record/1489224>
- Abu Saif, Musa Ahmed. (2024). Educational measurement and evaluation. *Al-Tarbawi Journal*, (25), 319-330. Retrieved from <http://search.mandumah.com/record/1489224>
2. الحازمي، حنان بنت علي يحيى، والشهري، سامي بن مصبح بن غرمان. (2024). درجة توافر مهارات التفكير العليا في اختبار مادة الرياضيات بالمرحلة الثانوية. *المجلة التربوية*، ع118، 591-621. مسترجع من : <http://search.mandumah.com/record/1450578>
- Al-Hazmi, Hanan bint Ali Yahya, & Al-Shehri, Sami bin Musbeh bin Gharman. (2024). The degree to which higher-order thinking skills are represented in secondary-school mathematics examinations. *The Educational Journal*, (118), 591-621. Retrieved from <http://search.mandumah.com/record/1450578>
3. حجي، مصطفى رحيم، والسعدي، مصطفى رعد عبد الرسول. (2023). تقويم الأسئلة الامتحانية «الوزارية» للصف الثالث المتوسط في مادة الرياضيات وفقاً لمستويات عمق المعرفة الرياضية للأعوام الدراسية 2020-2021 و2021-2022. *المجلة الأفريقية للدراسات المتقدمة في العلوم الإنسانية والاجتماعية*، مج2، ع3، 1-10. مسترجع من : <http://search.mandumah.com/record/1570566>
- Haji, Mustafa Rahim, & Al-Saadi, Mustafa Raad Abdulrasoul. (2023). An evaluation of the ministerial examination questions for the third intermediate grade in mathematics according to the levels of depth of mathematical knowledge for the academic years 2020-2021 and 2021-2022. *African Journal of Advanced Studies in Humanities and Social Sciences*, 2(3), 1-10. Retrieved from <http://search.mandumah.com/record/1570566>
4. دانة، هاني، وبوقندورة، رقية. (2024). أساليب تقييم جودة الاختبارات مرجعية المحك. *مجلة دفاتر المخبر*، مج19، ع1، 43-57. مسترجع من <http://search.mandumah.com/record/1476469> :

Dana, Hani, & Boukendoura, Roukia. (2024). Methods for evaluating the quality of criterion-referenced tests. *Dafatir Al-Makhbar Journal*, 19(1), 43–57. Retrieved from <http://search.mandumah.com/record/1476469>

5. زايدي، لمين، وسهل، ليلي. (2020). الاختبارات ودورها في تحقيق مبدأ الجودة التعليمية. *اللغة العربية*، ع49، 269–296. مسترجع من <http://search.mandumah.com/record/1055023>

Zaidi, Lamine, & Sahl, Leila. (2020). Tests and their role in achieving the principle of educational quality. *The Arabic Language Journal*, (49), 269–296. Retrieved from <http://search.mandumah.com/record/1055023>

6. السقاف، عبد الدايم عبد الرحمن يحيى. (2021). تحليل أسئلة اختبارات الشهادة الأساسية العامة لمادة الرياضيات بالجمهورية اليمنية للعام الدراسي 2019–2020م في ضوء تصنيف بلوم للأهداف المعرفية. *مجلة القلم*، ع28، 487–505. مسترجع من <http://search.mandumah.com/record/1225439>

Al-Saqqaf, Abduldayem Abdulrahman Yahya. (2021). An analysis of the mathematics questions in the General Basic Education Certificate examinations in the Republic of Yemen for the 2019–2020 academic year in light of Bloom's taxonomy of cognitive objectives. *Al-Qalam Journal*, (28), 487–505. Retrieved from <http://search.mandumah.com/record/1225439>

7. سيد، عصام محمد عبد القادر. (2020). لماذا تحليل المحتوى؟ *المجلة التربوية*، ع78، 617–628. مسترجع من : <http://search.mandumah.com/record/1091130>

Sayed, Essam Mohammed Abdel-Qader. (2020). Why content analysis? *The Educational Journal*, (78), 617–628. Retrieved from <http://search.mandumah.com/record/1091130>

8. الشرفي، عبد الرحمن بن محمد بن علي. (2024). تحليل الأسئلة التقويمية لأعضاء هيئة التدريس بقسم الدراسات الإسلامية بكلية الآداب والعلوم الإنسانية بجامعة الباحة في ضوء تصنيف بلوم. (Bloom). *العلوم التربوية*، مج32، ع4، 381–357. مسترجع من <http://search.mandumah.com/record/1532801>

Al-Sharafi, Abdulrahman bin Mohammed bin Ali. (2024). An analysis of the assessment questions prepared by faculty members in the Department of Islamic Studies at the Faculty of Arts and Humanities, Al-Baha University, in light of Bloom's taxonomy. *Educational Sciences*, 32(4), 357–381. Retrieved from <http://search.mandumah.com/record/1532801>

9. العتيق، عبد المحسن بن مبارك أحمد. (2024). تحليل محتوى أسئلة الاختبارات النهائية لمقررات الفقه ببرنامج بكالوريوس الشريعة في ضوء مهارات التفكير العليا وفق تصنيف بلوم للمستويات المعرفية. *مجلة الجامعة الإسلامية للعلوم التربوية والاجتماعية*، ع20، 187–220. مسترجع من : <http://search.mandumah.com/record/1524260>

Al-Atiq, Abdulmohsen bin Mubarak Ahmed. (2024). A content analysis of final-examination questions for jurisprudence courses in the Bachelor of Sharia program in light of higher-order thinking skills according to Bloom's taxonomy of cognitive levels. *Journal of the Islamic University for Educational and Social Sciences*, (20), 187–220. Retrieved from <http://search.mandumah.com/record/1524260>

References:

1. Azis, A., Handayani, I. T., Ferniati, F., Anggriana, N., & Aisyah, A. (2023). *Analysis of students' cognitive difficulties based on the revised Bloom's taxonomy in solving mathematics problems*. *Focus Action of Research Mathematic*, 6(1), 117–138.

2. Foughma, K., & Rabhi, I. (2024). Quality standards for achievement tests and steps for building a good achievement test. *Revue d'études et de recherche sociale*, 12(2), 171–186.

3. Magdalena, I., Nurchayati, A., Saputri, A. N., Amanda, N. Z., Habibie, N., Waluyo, S. N., & Nisa, D. K. (2023). *Analisis Taksonomi Bloom dalam Mengidentifikasi Tingkat Kesulitan Pertanyaan Soal dalam Mata Pelajaran Matematika di Sekolah Dasar*. Jurnal Pendidikan, Bahasa dan Budaya, 2(3), 141–150.
4. Manouchehri, A., Yao, X., Fleming, J., & Gomez, C. (2016). *What do mathematics achievement examinations assess? A critical item analysis*. Journal of Mathematics Education, 9(2), 45–62.
5. Orhani, S. (2024). *Preparation of tests from the subject of mathematics according to Bloom's taxonomy*. International Journal of Research Publication and Reviews, 5(2), 2335–2345.
6. Pinar, A., & Ocak, G. (2025). *Examining the 2018 mathematics curriculum 5th grade achievements in terms of 21st century skills according to Bloom and Haladyna's taxonomies*. Journal of Curriculum Studies, 18(1), 77–98.