

مقارنة طريقة الترتيب الكامل والطريقة التقليدية والطريقة الاعتيادية لضبط أثر التخمين على الخصائص السيكمومترية لاختبار الاختيار من متعدد

د. مروان الصمادي
جامعة عمان العربية
تاريخ القبول: 2021/06/29

عمار خالد أحمد عقل
جامعة عمان العربية
تاريخ الاستلام: 2021/06/05

الملخص

هدفت هذه الدراسة إلى مقارنة طريقة الترتيب الكامل والطريقة التقليدية والطريقة الاعتيادية من أجل ضبط أثر التخمين على الخصائص السيكمومترية لاختبار الاختيار من متعدد، تكون مجتمع الدراسة من جميع طلبة الصف العاشر في المدارس الحكومية التابعة لمديرية التربية والتعليم عمان الرابعة (لواء ماركا) في الفصل الدراسي الأول من العام الدراسي 2021/2020، استخدمت الدراسة

المنهج الوصفي، وتم اختيار عينة من (180) طالب وطالبة بالطريقة القصيدة من مدرستي المعتمد الأساسية للبنين، ومدرسة باب الواد الثانوية للبنات، منهم (80) إناث، و(100) ذكور، ولتحقيق هدف الدراسة، والإجابة عن أسئلتها، تم بناء اختبار تحصيلي مكون من (27) فقرة اختيار من متعدد بأربعة بدائل، في وحدة (النسب المثلثية وحل المثلثات)، وتم تطبيق الاختبار على أفراد عينة الدراسة، توصلت الدراسة إلى النتائج الآتية: وجود فروق ذات دلالة إحصائية عند مستوى الدلالة ($\alpha = 0.05$) في متوسطات أداء الطلبة على الاختبار بين طرق التصحيح الثلاث لصالح طريقة الترتيب الكامل، تلاها الطريقة الاعتيادية، ثم طريقة التصحيح التقليدية (العقاب). وجود فروق ذات دلالة إحصائية عند مستوى الدلالة ($\alpha = 0.05$) في معاملات الصدق المحكي، لصالح طريقة الترتيب الكامل مقابل الطريقة الاعتيادية وطريقة التصحيح التقليدية (العقاب)، في حين لم تظهر فروق ذات دلالة إحصائية بين طريقتي التصحيح التقليدية (العقاب) والطريقة الاعتيادية، أما بالنسبة لمعاملات الثبات كرونباخ الفا لدرجات الاختبار، وجدت فروق بين طرق التصحيح الثلاث لصالح طريقة الترتيب الكامل تلاها الطريقة الاعتيادية ثم طريقة التصحيح التقليدية (العقاب).

الكلمات المفتاحية: طريقة الترتيب الكامل، والطريقة التقليدية، ضبط أثر التخمين، التخمين، الخصائص السيكمومترية، اختبار اختيار من متعدد.

Comparison between the Complete Ordering Method and the Traditional Method to Control the Effect of Guessing on the Psychometric Properties of the Multiple-Choice Test

Ammar Khaled Ahmed Aqel
Amman Arab University

Dr. Marwan Saleh AL Smadi
Amman Arab University

Abstract

This study aimed to compare between the Complete Ordering Method and the Traditional Method to Control the Effect of Guessing on the Psychometric Properties of the Multiple-Choice Test. The study population consisted of all 10th-grade students in the government schools of the Fourth Amman Directorate of Education (Lewaa-Marka) in the first semester of the academic year 2020/2021, The study used the descriptive and analytical approach. the sample is selected by the available method from the 10th-grade students, from (2) schools, al-Mu'tasim primary for male and Bab al-Wad secondary school for female, where their number reached (180) that are (100) males and (80) female students. To achieve the goal of the study, and to answer its questions, the researcher constructed an achievement test consisting of (27) multiple-choice items with four alternatives from the unit of (trigonometric ratios and solving triangles). The test is applied to the study sample members.

The results related the following: There are statistically significant differences at the level of significance ($\alpha \leq 0.05$) in the averages of students' performance on the test between the three methods of correction in favor of the Complete Ordering Method, followed by the standard method, then the traditional method of correction (punishment). Also are statistically significant differences ($\alpha \leq 0.05$) in Criterion – Referenced, in favor of the Complete Ordering Method versus the standard method and the traditional correction method which called (punishment), while no statistically significant differences appeared between the two traditional correction methods (punishment) and the standard method but, in case of Cronbach's alpha validity, there are statistically significant between the three methods of correction in favor of the Complete Ordering Method, followed by the standard method, then the traditional method of correction (punishment).

Key Words: Complete Ordering Method; the Traditional Method; the effect of guessing; Psychometric Properties; Multiple-Choice Test

المقدمة:

تُعتبر عملية التقويم التربوي عنصراً مهماً ومكوّناً رئيساً من مكونات الموقف التعليمي، حيثُ تتوفر الوسائل والأدوات المستخدمة فيها، كالاختبارات والملاحظة والمقابلة والاستبانة ودراسة الحالة وملف الإنجاز وغيرها، وفي مجال التحصيل تبقى الاختبارات التحصيلية الأوسع انتشاراً والأكثر استخداماً، فهي من أدوات القياس التربوي المهمة وهي متعددة الأغراض والأهداف، كما تعدُّ مؤشراً هاماً يُستند إليها في عملية التقويم، ولها دور أساسي في عملية التعليم والتعلم، وتستخدم الاختبارات التحصيلية في اتخاذ القرارات التربوية المختلفة، فعندما نمتحن الطلبة نسعى لقياس مستوى التحصيل لديهم، والوقوف على مدى التقدم الذي تمَّ إحرازه، كما تستخدم في مقارنة الطلبة بعضهم ببعض أو في تشخيص جوانب القوة والضعف لديهم، وتوزيعهم على شعب متجانسة من حيثُ قدرتهم العقلية والتحصيلية، وكذلك في عملية القبول والانتقاء والتوظيف والتنبؤ... وغير ذلك.

ومن الأمور المهمة التي يجب أن تؤخذ بعين الاعتبار، أنَّ الاختبارات التحصيلية هي في مجملها تطوير للعملية التعليمية، وهذا لا يتمُّ إلّا من خلال بناء اختبار يقيس مستوى التحصيل من ناحية، وأن يكون فيه ما يعمل على تقويم عملية التدريس من ناحية أخرى، فكلما زادت أهمية القرار الذي يتوقع اتخاذه، زادت الحاجة لأن تكون المعلومات التي يتمُّ الحصول عليها من الاختبار أكثر دقة، وصلة بالغرض الذي أعدَّ الاختبار من أجله (عبد الهادي، 2001).

لذلك؛ تعدُّ الاختبارات التحصيلية من أهمِّ الأدوات التي يلجأ من خلالها العاملون في المؤسسات التربوية لتقييم ما حققه الطلبة من نتائج التعليم أو لتحديد مسار الطلبة الأكاديمي، بما يتناسب مع قدراتهم ومهاراتهم، وتلجأ المؤسسات التربوية ضمن نظامها التربوي إلى عكس مدى النجاح الذي حققته من خلال نتائج الاختبارات التحصيلية، كما تستخدم الاختبارات التحصيلية كوسيلة لتشخيص الصعوبات وتصميم البرامج العلاجية للطلبة، ولعلَّ تحقُّق كلِّ ما سبق في وظائف الاختبارات التربوية واستخداماتها مرهوناً بدقة هذه الأدوات، وصدق وثبات وموضوعية النتائج المتمخضة عنها. ويرى عودة (2010) أنها تساعد في تحسين واتخاذ القرارات الخاصة بعملية التدريس، وتزيد من مستوى الاحتفاظ بالمعلومات، وانتقال أثر التعليم، وتحسّن من دافعية الطلبة، وتزيد من معرفة الطلبة بحقيقة أنفسهم، كما توفر التغذية الراجعة لفعالية العملية التدريسية. من هنا لا بدَّ أن يؤخذ بالاعتبار مجموعة من العوامل والاعتبارات عند إعدادها، كأن تكون ممثلة للمادة الدراسية المتضمنة في التدريس، وأن تكون متسقة مع الأهداف التدريسية، وأن تصاغ بطريقة تتناسب مع الغايات التي تستخدم النتائج من أجلها، وأن يتمتع الاختبار بدرجة مقبولة من الصدق الثبات، وأن يحتوي على نوعية الأسئلة التي الأكثر ملائمة من غيرها لقياس الناتج التعليمي المرغوب، فالنواتج التعليمية الخاصة بوحدة دراسية معيّنة هي التي تحدد أنواع السلوك الذي يمكن قبوله كدليل على تحقيق الأهداف التدريسية، وما الاختبار التحصيلي إلّا وسيلة لاستدعاء السلوك المطلوب. من هنا فإنَّ قياس التحصيل بشكل فعال يتأثر باختيار نوعية الأسئلة التي من شأنها أن تستثير الإجابة المطلوبة وتستبعد الإجابات الأخرى التي ليس لها صلة بالإجابات الصحيحة.

وللاختبارات التحصيلية أشكال متعددة ومتنوعة، منها الاختبارات الكتابية والاختبارات غير الكتابية، وتتضمن الاختبارات الكتابية: الاختبارات ذات الإجابة القصيرة (المقيدة) والاختبارات ذات الإجابة الحرة الطويلة (المفتوحة) واختبارات المزوجة، واختبارات الصواب والخطأ، واختبارات الاختيار من متعدد، واختبارات الصواب-الخطأ المتعدد (عبد الهادي، 2001؛ Thorndike، 1982).

وطالما هناك عدّة أنواع من الاختبارات التي يمكن أن تستخدم لأغراض اتخاذ القرارات، فمن الأهمية بمكان تحديد أي نوع من هذه الاختبارات هو الأفضل في تقديم معلومات مفيدة، وما هي الأسس التي يستند إليها في الاختيار،

لذلك فإنَّ تحديد نوع الأسئلة الذي يعتبر ملائماً للاختبار يعتمد إلى حد كبير على طبيعة المحتوى المراد قياسه، فهناك بعض المواد الدراسية يناسبها نوع معين من الأسئلة ولا يناسبها نوع آخر، كما يعتمد نوع الأسئلة أيضاً على الهدف المراد قياسه، فإذا كان الهدف يتطلب من الطالب، أن يعبر بلغته الخاصة عن مفهوم معين، فهنا يكون من الملائم استخدام الأسئلة المقالية، أما إذا كان الهدف يتطلب من الطالب أن يقارن بين عدد من الإجابات المعطاة وأن يختار من بينها الإجابة الصحيحة، فهنا يكون من الأفضل والملائم استخدام أسئلة الاختيار من متعدد، كما أنَّ مهارة المعلم في كتابة نوع من الأسئلة، لها دور في تحديد نوعية الأسئلة المناسبة (علام، 2010).

كما أنَّ اختيار نوعية الأسئلة، يعتمد كذلك على عوامل متعددة كعدد الطلبة الذين سيطبق عليهم الاختبار، والزمن المتوفر لإعداد الاختبار، والزمن المتوفر للتصحيح، فعندما يتساوى تأثير هذه العوامل، فإنَّ اختيار نوع الأسئلة الملائم يعتمد على الخصائص السيكمترية للاختبار، التي تتعلق بكل نوع من الأسئلة، فعلى المعلم أن يختار نوع الأسئلة التي يتوفر لها معاملات صدق وثبات مرتفعة والتي هي من الصفات الأساسية المرغوب توفرها في الاختبار، حيثُ أنَّ توفر هذه الخصائص يزيد من إمكانية تحقيق الاختبارات لأهداف القياس التربوي، كما أنَّ عمر المفحوص وسهولة التطبيق، والتصحيح، وعامل الكلفة، وتوفر الصور المتكافئة للاختبار ووجود المعايير هي من العوامل التي تدفع باتجاه اختيار نوع معين من الأسئلة (عودة، 2010)، ويرى فرزي وسويني (Frisbie & Sweeney، 1982) أنَّ فقرات الاختيار من متعدد هي الأكثر استخداماً لقياس تحصيل الطالب في العديد من الأغراض التربوية، فهي تتفوق على فقرات الصواب والخطأ في قياس تحصيل الطلبة، وتتسم بنتائجها بدرجة عالية من الصدق والثبات.

وتتكون فقرات الاختيار من متعدد من مشكلة معروضة في جملة تُسمى أصل الفقرة (المتن) وتوضع بعدها عدّة إجابات تُسمى البدائل، ويطلب من الطالب اختيار الإجابة الصحيحة من بين هذه البدائل. وتتمتع فقرات الاختيار من متعدد بمزايا عدّة فهي تحقق قدر كبير من الأهداف التربوية بدءاً من العمليات البسيطة كالذاكرة وانتهاءً بالعمليات المعقدة كالفهم والتقويم، كما يمكن تقدير إجابات الطلبة بموضوعية كاملة، وتقلل من أثر التورية والتلفيق إلى أدنى حدٍّ ممكن، إلا أنها تتأثر بعامل التخمين العشوائي، وتبقى أقلُّ تأثراً مقارنة بالاختبارات الموضوعية الأخرى (عودة، 2010).

ويُعدُّ التخمين العشوائي الذي يلجأ إليه الطالب عندما لا تتوافر لديه المعلومات التي تمكنه من الوصول إلى الإجابة الصحيحة من أكبر المشكلات التي تواجه هذا النوع من الاختبارات، حيثُ يقوم الطالب بالاختيار العشوائي لأحد البدائل حتى لو طُلب منه أن يترك الفقرة دون إجابة إذا لم تكن لديه معرفة بالإجابة الصحيحة، فبعض الطلبة يميلون إلى المجازفة أو المخاطرة بدرجات مختلفة حسب درجة التزامهم بالتعليمات، وحكمتهم في الإجابة، فأحياناً قد توجد منبهات أو مفاتيح في الفقرة أو البدائل تساعد الطالب على التوصل للإجابة الصحيحة، وبهذا لا يمكن للمصحح أن يقرر إذا كانت الإجابة الصحيحة للطالب عن فقرة ما تعكس قدرته الحقيقية على الإجابة للفقرة، خاصة عندما يكون الهدف من الاختبار تشخيص صعوبات التعلم أو الوصول إلى درجة الإتقان لمهارة معينة (Annie & Chan، 2012).

ويؤدي التخمين في اختبارات الاختيار من متعدد إلى غياب العدالة عن النتائج أحياناً، فإذا تقدم مفحوصان من مستوى قدرة متوسطة لأداء اختبار موضوعي من نوع الاختيار من متعدد فهناك احتمال بأن يجتاز المفحوص الذي يلجأ إلى التخمين الاختبار، في حين يفشل الآخر الذي أجاب عن فقرات الاختبار بناءً على معرفته الحقيقية، مما يؤدي إلى انتهاك أهمِّ الخصائص الأساسية للاختبار وهي العدالة فتصبح نتائج الاختبار متحيزة (Kubinger & Holocher، 2010).

مشكلة الدراسة: نظرًا لأهمية معالجة وضبط أثر التخمين في اختبارات الاختيار من متعدد، فقد برزت جهود كثيرة لإيجاد طرق تنطلق من أطر ومفاهيم نظرية متباينة تعتمد أساليب إحصائية مختلفة، للمساعدة في الحد من أثر التخمين على صدق وثبات نتائج الاختبارات الموضوعية. وفي إطار النظرية الكلاسيكية في القياس، تأخذ طرق ضبط أثر التخمين ثلاثة مواقف عند إجابة المفحوص على فقرات الاختيار من متعدد هي: أن المفحوص يعرف الإجابة الصحيحة، أو أن المفحوص لا يعرف الإجابة الصحيحة ويترك الفقرة دون إجابة، أو أن المفحوص لا يعرف الإجابة الصحيحة على الفقرة لكنه يجيب عليها بشكل عشوائي، وتعتبر أخطاء القياس من أبرز المشاكل التي يعاني منها القياس النفسي والتربوي، حيث يشكل التخمين أحد المصادر الأساسية لهذه الأخطاء، وبالتالي قد يحصل الطالب بالتخمين على درجة أعلى من الدرجة التي تمثل قدرته الحقيقية، مما يؤثر سلبًا على صدق الاختبار وثباته، وهما من أهم الخصائص السيكمترية للاختبار الجيد (اليامي، 2016).

لذلك حاول الكثير من الباحثين والتربويين معالجة مثل هذه العقبات، حيث أجريت العديد من الدراسات والأبحاث على خصائص الاختبار ومراحل تطبيقه، وتصحيحه، وتحليل نتائجه، وللحد من أثر التخمين في خصائص الاختبار السيكمترية، أستخدمت الكثير من الطرق لضبط أثر التخمين: كالطريقة التقليدية (العقاب)، وطريقة المكافأة وطريقة التصحيح لأثر التخمين (المكافأة والعقاب)، وطريقة التصحيح الجزئي، وطريقة تصحيح الموهبات، وطريقة لا أعرف الإجابة. من هنا جاءت هذه الدراسة للإجابة على بعض التساؤلات والتعرف على أثر طريقة الترتيب الكامل لضبط أثر التخمين مع طريقة التصحيح التقليدية (العقاب) والطريقة الاعتيادية على الخصائص السيكمترية لاختبارات الاختيار من متعدد. علمًا بأن طريقة الترتيب الكامل لم يسبق في ضوء علم الباحث مقارنة نتائجها بغيرها وأثرها على خصائص الاختبار،

عناصر مشكلة الدراسة (أسئلة الدراسة): أجابت الدراسة عن الأسئلة التالية:

السؤال الأول: هل توجد فروق ذات دلالة إحصائية عند مستوى $(\alpha = 0.05)$ بين المتوسطات الحسابية لأداء الطلبة في الاختبار تُعزى لطريقة ضبط أثر التخمين؟

السؤال الثاني: هل توجد فروق ذات دلالة إحصائية عند مستوى $(\alpha = 0.05)$ في الصدق المحكي لدرجات الاختبار تُعزى لطريقة ضبط أثر التخمين؟

السؤال الثالث: هل توجد فروق ذات دلالة إحصائية عند مستوى $(\alpha = 0.05)$ في ثبات درجات الاختبار تُعزى لطريقة ضبط أثر التخمين؟

أهمية الدراسة: تتمثل أهمية الدراسة في الأهمية النظرية والأهمية العملية:

الأهمية النظرية: ستوفر الدراسة أطر نظرية تخص متغيرات الدراسة كأثر التخمين، والاختبارات التحصيلية، والخصائص السيكمترية، لتفيد الباحثين والمهتمين في إجراء دراساتهم المستقبلية، كما أن نتائج الدراسة ستساهم في إيجاد الحلول الناجحة في التقليل من أثر التخمين. ونظرًا لانتشار اختبارات الاختيار من متعدد انتشارًا واسعًا والتأثير الملحوظ لأثر التخمين على نتائجها، زادت الحاجة لإجراء الدراسة الحالية لتلقي ضوءًا أكبر واهتمامًا أشد لمقارنة أثر طريقة الترتيب الكامل لضبط أثر التخمين مع طريقة التصحيح التقليدية (العقاب) والطريقة الاعتيادية على الخصائص السيكمترية لاختبارات الاختيار من متعدد.

الأهمية العملية: يمكن أن تسهم الدراسة الحالية في التعرف على أثر الطرق المستهدفة في الخصائص السيكمترية للاختبارات، وأنها توفر اختبار تحصيلي في الرياضيات للصف العاشر تتوفر فيه معايير الصدق والثبات

يمكن استخدامه من قِبَل باحثين وتطبيقه من قِبَل قسم الامتحانات في وزارة التربية والتعليم، كما أن النتائج التي توصلت إليها الدراسة يمكن أن تسهم في اهتمام الباحثين والمهتمين في تطوير اختبارات في مواد دراسية أخرى ستتوفر فيها خصائص سيكومترية جيدة، كذلك توضح معالم جديدة لمصممي الاختبارات في المجال التربوي، بأسلوب جديد في قياس تحصيل الطلبة، وتصحيح الاختبارات الموضوعية، واتخاذ القرارات التربوية المختلفة.

التعريفات الاصطلاحية الإجرائية:

التخمين: (Guessing) اختيار الطالب للبديل الصحيح من خلال الصدفة، ويُسمى تخمينًا عشوائيًا. أو من خلال المعرفة الجزئية ويُسمى تخمينًا ذكيًا (عودة، 2010).

ويعرف إجرائيًا: الاستجابة الخاطئة لفقرة بشكل عشوائي في حال لم يكن لدى الطالب معرفة بالإجابة الصحيحة. طرق ضبط أثر التخمين (Control the Effect of Guessing): طرق إحصائية يتم من خلالها تصحيح (تعديل) درجة الطالب حسب معادلات رياضية، للحد من أثر التخمين بناءً على عدد الإجابات الصحيحة وعدد الفقرات التي تم تخمينها.

طريقة الترتيب الكامل (Complete Ordering Method): وهي حالة خاصة من الطريقة الاحتمالية، حيث يُطلب من المفحوص وضع رُتب للبدائل لكل فقرة وفقًا للاحتمالات، وفي هذه الطريقة يمكن التمييز بين ثلاثة مستويات من المعرفة لدى المفحوصين، فالمفحوص الذي يمتلك معرفة تامة يعطي الرتبة الأولى للبديل الصحيح، والمفحوص الذي لديه معرفة جزئية يعطي رتبة متوسطة للبديل الصحيح والمفحوص الذي لا يمتلك المعرفة يعطي الرتبة الأخيرة للبديل الصحيح (أبو فودة و يونس، 2012).

وتعرف إجرائيًا: بالدرجة التي يحصل عليها الطالب باستخدام المعادلة أدناه:

$$F = \sum (A - B) \dots (1)$$

حيث: F: الدرجة النهائية المصححة.

B: الرتبة المعطاة للبديل الصحيح. A: عدد البدائل.

طريقة التصحيح التقليدية لأثر التخمين (العقاب) (Scoring Formula-SF): عرفها مِهْرَنْز وليهمان (Mehrens & Lehmann، 1991) بأنها "طريقة تقوم على حذف الدرجات التي يقدر أن المفحوص حصل عليها من خلال التخمين، إذ تفترض هذه الطريقة أن كل استجابة خاطئة هي نتيجة تخمين قام به المفحوص، حيث لا تأخذ هذه الطريقة بعين الاعتبار المعلومات الجزئية التي ربما يكون قد استند إليها المفحوص في إجابته عن فقرات الاختبار". وتعرف إجرائيًا بالمعادلة:

$$F = R - \frac{W}{A-1} \dots (2)$$

بحيث أن:

F: الدرجة النهائية المصححة. R: عدد الإجابات الصحيحة.

W: عدد الإجابات الخاطئة. A: عدد البدائل.

الطريقة الاعتيادية في التصحيح بدون ضبط التخمين (عدد الإجابات الصحيحة): حيث يُعطى الطالب درجة واحدة لكل إجابة صحيحة بغض النظر عن الإجابات الخاطئة أو المتروكة دون إجابة.

وتُعرف إجرائيًا: وتحسب بعدد الإجابات الصحيحة فقط دون اعتبار للإجابات الخاطئة كما في المعادلة أدناه:

$$F = \sum R$$

F: الدرجة النهائية المصححة. R: عدد الإجابات الصحيحة.

محددات الدراسة وحدودها:

الحدود الزمنية: تم تطبيق الاختبار أثناء الفصل الدراسي الأول من العام الدراسي (2020-2021م).

الحدود البشرية: اقتصر عينة الدراسة على طلبة الصف العاشر.

الحدود المكانية: المدارس الحكومية في محافظة العاصمة، لواء ماركا في مدرستي المعتصم الأساسية للبنين

وباب الواد الثانوية للبنات.

محددات الدراسة: إنَّ تعميم نتائج هذه الدراسة يبقى مرهونا بالمحددات الآتية:

- اقتصر المحتوى الدراسي للاختبار التحصيلي على مبحث الرياضيات للصف العاشر الأساسي الوحدة الخامسة (النسب المثلثية وحلّ المثلثات)، بحسب منهاج وزارة التربية والتعليم الأردنية.
 - اقتصر طرق الدراسة على طرق ضبط أثر التخمين الآتية: طريقة الترتيب الكامل، طريقة التصحيح التقليدية (العقاب)، طريقة التصحيح الاعتيادية
 - تعتمد إمكانية تعميم نتائج هذه الدراسة على مدى صدق استجابة الطلبة على الاختبار التحصيلي، والخصائص السيكومترية (الصدق والثبات) للاختبار التحصيلي
- الدراسات السابقة ذات الصلة:

فيما يلي عرض لأهم الدراسات السابقة التي تناولت طرق ضبط أثر التخمين في اختبارات الاختيار من متعدد،

مرتبة وفق الترتيب الزمني من الأقدم إلى الأحدث كالآتي:

قام النبهان (Alnabhan، 2002) بدراسة هدفت إلى تقصي أثر ثلاث طرق للتصحيح ضد أثر التخمين وهي:

(عدد الإجابات الصحيحة، ومعاقبة التخمين، ومكافأة التخمين الجزئي) على الخصائص السيكومترية (الثبات والصدق) للاختبار حيث تكونت عينة الدراسة من (120) طالباً وطالبة في مرحلة ما قبل التخرج في جامعة أبو ظبي في دولة الإمارات العربية المتحدة، وتألّف الاختبار من (40) فقرة اختيار من متعدد بواقع أربعة بدائل لكلّ فقرة في مساق منهجية البحث النفسي كما استخدم الباحث اختباراً قصيراً مكوناً من (10) فقرات، للمحتوى التعليمي نفسه الذي يغطيه الاختبار المكون من (40) فقرة وبدائل جميعها خاطئة لكلّ فقرة؛ من أجل تصنيف المفحوصين إلى: ذوي مخاطرة مرتفعة وذوي مخاطرة متدنية، فالطالب الذي أجاب عن عدد أكبر من الفقرات اعتُبر ذا مخاطرة مرتفعة والذي أجاب عن عدد أقلّ من الفقرات اعتُبر ذا مخاطرة منخفضة، وقد أظهرت نتائج الدراسة أنّ الطرق الثلاث المختلفة أعطت معاملات صدق وثبات مختلفة عن طريقة المعرفة الجزئية التي تقوم على افتراض أنّ المفحوص يمتلك معرفة تمكنه من حذف بعض البدائل عند إجابته عن الفقرة، حيثُ كان معامل الثبات للاختبار المصحح ضد أثر التخمين باستخدام طريقة المعرفة الجزئية (0.81) بينما كان (0.65) عند استخدام طريقة التصحيح باستخدام عدد الإجابات الصحيحة، كما بلغ (0.73) عند استخدام طريقة معاقبة التخمين، وأكدت الدراسة أنّ التخمين يُقلل من صدق وثبات الاختبار من خلال إعطاء تباين غير حقيقي في نتائج الاختبار.

كما هدفت دراسة الزبون (2004) إلى مقارنة أثر ثلاث طرق لتصحيح وضبط أثر التخمين لاختبارات الاختيار

من متعدد على الخصائص السيكومترية للاختبار والفقرة في ضوء نظرية الاستجابة للفقرة، حيثُ استخدم الطريقة التقليدية

ومعادلة التصحيح لأثر التخمين وطريقة الاختيار الجزئي، وقد تمّ إعداد ثلاث صور متكافئة لاختبار تحصيلي لطلبة مساق (الإحصاء التربوي) في جامعة مؤتة، وجرى تطبيقها على عينة مكونة من (406) طالباً من الصف الثاني ثانوي، جرى تصحيح الإجابات على كلّ صورة من صور الاختبار بإحدى طرق التصحيح الثلاث، ثمّ حللت البيانات باستخدام برمجية (One step) لتقدير معلمة القدرة (θ) والخطأ المعياري في القياس لقدرة كلّ فرد، كما تمّ استخراج معاملات الثبات لتباعد الأفراد، ومعاملات صدق المحك التلازمي، ومعاملات صعوبة الفقرة، ومعاملات تمييز الفقرة وأظهرت نتائج الدراسة وجود فروق ذات دلالة إحصائية عند مستوى الدلالة ($\alpha = 0.05$) بين معاملات الثبات لصالح استخدام كلّ من طريقة التصحيح لأثر التخمين وطريقة الاختبار الجزئي، كما أشارت النتائج إلى وجود فروق ذات دلالة إحصائية بين معاملات صدق الاختبار لصالح الطريقة التقليدية وطريقة الاختيار الجزئي، إضافة إلى ذلك أظهرت الدراسة أنّ هناك فروقاً ذات دلالة إحصائية بين معاملات صعوبة الفقرة، لصالح معادلة التصحيح لأثر التخمين، وأنّ هناك فروقاً ذات دلالة إحصائية بين معاملات تمييز الفقرة عند استخدام الطريقة التقليدية وطريقة (الاختبار الجزئي).

وقارن دودين (Dodeen, 2005) بين طريقة معادلة تصحيح الموهبات التي تأخذ بعين الاعتبار مستوى جاذبية المموه (البديل) في حساب درجة السؤال النهائية، والطريقة التقليدية لتصحيح أثر التخمين، كما قام بمقارنة فاعلية المعادلتين، لا سيّما تأثير كلّ منهما على أداء الطلبة على الاختبار، وقيمة الثبات، من خلال تطبيق المعادلتين على نتائج اختبارين معياريين في الرياضيات واللغة الإنجليزية، على عينة مكونة من (148) طالباً وطالبة من طلبة جامعة الإمارات العربية المتحدة، وقد دلّت النتائج على أفضلية استخدام معادلة تصحيح الموهبات في مكافأة الطلبة على حسن اختيارهم للبديل الأقرب إلى الصواب أو الأكثر صحة عند عدم التأكد تماماً من الجواب الصحيح، حيث أظهرت طريقة تصحيح الموهبات متوسط درجات أكبر من الطريقة التقليدية، كما أدت إلى خفض الانحراف المعياري لدرجات الاختبار، وكانت نسب التخمين أعلى للطلبة ذوي التحصيل المتوسط، بالإضافة إلى ذلك، فقد أدّى تطبيق هذه المعادلة إلى زيادة ثبات نتائج الاختبارين عندما حُسبت على العينة الكلية، وعينات فرعية مشتقة من العينة الكلية لكلّ اختبار، في حين كانت درجات ثبات الاختبار تتناقص مع تناقص حجم العينة للعينات الفرعية.

كما أجرى الصمادي (2012) دراسة هدفت لمعرفة أثر طريقة تصحيح اختبار الصواب والخطأ المتعدد في معاملات الصدق والثبات للاختبار، وأثرها في معاملات الصعوبة وتمييز فقرات الاختبار، ومعرفة أثرها في نتائج الطلبة وتحصيلهم، قام الباحث ببناء اختبار يتكوّن من (25) فقرة صواب وخطأ متعدّد لكلّ فقرة أربعة بدائل، تكوّنت عينة الدراسة من (122) طالب في جامعة إربد الأهلية، استخدم الباحث أربع طرق مختلفة للتصحيح تراعي المعرفة الجزئية واستخدم برنامج (SPSS) لحساب معاملي صدق الاختبار، واستخدم اختبار كرونباخ ألفا لحساب معامل ثبات الاختبار، وأظهرت نتائج الدراسة عدم وجود فروق ذات دلالة إحصائية في معاملات صعوبة الفقرة في تحصيل الطلبة، وكانت طريقة التصحيح الرابعة (إعطاء العلامة 0.25-0.50-0.75-1.00) لكلّ بديل يتمّ الإجابة عنه بشكل صحيح الأكثر مراعاة للمعرفة الجزئية لدى الطلبة ثمّ تلتها طريقة التصحيح الثالثة ثمّ الثانية ثمّ الأولى.

وأجرى عودة (2014) دراسة لتصحيح معامل صعوبة الفقرة لأثر التخمين في اختبار الاختيار من متعدد، صورة معدلة لمعادلة جيلفورد، مع الافتراض بأنّ التخمين يتمّ عشوائياً، إلّا أنّ التقدير لمعامل الصعوبة بهذه المعادلة يقوم على أساس رياضي بحت لا يتحقق فيه الافتراض الذي قام عليه، وأنّ المؤشرات العملية والمنطقية ترجح أهمية تقديم صورة معدلة لهذه المعادلة لتكون أكثر عدالة ومصادقية في تقدير معاملات الصعوبة لفقرات الاختبارات المعرفية بشكل عام، والتحصيلية بشكل خاص، تركز فيه هذه الصورة المعدلة على التوفيق والتكامل بين التخمين القائم على المعرفة الجزئية،

والفروق الفردية لدرجة المخاطرة في ضوء التعليمات ذات الصلة بالتخمين وما يترتب عليها من إجراءات في التصحيح، وتحليل الدرجات تحليلًا إجماليًا وتفصيليًا على مستوى الفقرات، وقد قدمت الدراسة هذه الصورة المعدلة مرفقة بجداول لاستخراج معاملات الصعوبة القائمة على التخمين العشوائي مقابل معاملات الصعوبة المقدرة بالمعادلتين الأصلية والمعدلة، وترجح نتائج المقارنة لهذه التقديرات توافقًا أعلى بين نتائج التقديرات لمعاملات الصعوبة المشتقة من الصورة المقترحة للمعادلة والأساس النظري الذي تركز عليه، وهي معادلة توفيقية تجمع بين نظرية القرار ونظرية القياس.

وأجرت رافسلوت ورفاقها (Ravesloot, et al، 2015) دراسة هدفت إلى الكشف عن خيار لا أعرف الإجابة في الاختبارات. حيثُ قورنت الطريقة التقليدية والعقاب مع طريقة لا أعرف الإجابة مع متغيرات الجنس وسنوات الخبرة. حيثُ تمَّ إعداد اختبار لأطباء الأشعة المتدربين في هولندا، مكوّن من (200) فقرة من ثلاثة اختيارات أحدها لا أعرف الإجابة موزعة على نموذجين حسب طريقة التصحيح، وتم تقسيم المشاركين إلى (337) إلى مجموعتين بشكل متساوٍ بالنسبة للجنس وسنوات الخبرة. وأشارت النتائج إلى أن طريقة العقاب تدرج علامات غير حقيقية، وأن الإناث استخدمن البديل لا أعرف الإجابة أكثر من الذكور، وكانت متوسطات طريقة العقاب أعلى من طريقة لا أعرف الإجابة وأقرب لمتوسطات الطريقة التقليدية، ولم تحرز طريقة لا أعرف الإجابة فروقًا ذات دلالة.

وأجرى أوساي (Ossai، 2015) دراسة هدفت إلى البحث عن الفرق بين ميل الجنسين للتخمين في الرياضيات، حيثُ تقدم (110) طالبًا وطالبة من جامعة دلتا في نيجيريا (56 ذكور، 45 إناث)، لاختبار مقنن في الرياضيات، مكوّن من (50) فقرة، وتم استخدام استبانة لقياس ميل الطلبة، وأظهرت الدراسة عدم وجود فروق في المتوسطات تعزى لاختلاف الجنس.

كما قام الخرشه (2016) بدراسة هدفت إلى التحقق من أثر طرق معالجة التخمين في تقدير إحصائيات الأفراد والفقرات في اختبارات الاختيار من متعدّد وفق النظرية الحديثة في القياس، تمت مقارنة معادلة التصحيح لأثر التخمين (العقاب) ومعادلة التصحيح لأثر التخمين (المكافأة) ومعادلة التصحيح من أثر التخمين (المكافأة والعقاب)، ولتحقيق الغرض من الدراسة تمَّ بناء اختبار تحصيلي من نوع الاختيار من متعدّد في مادة مبادئ القياس والتقويم التي تدرس لطلبة كلية التربية بجامعة مؤتة مكوّن من (30) فقرة بصورته النهائية، وقد اشتمل على ثلاثة نماذج حسب طرق المعالجة، تمَّ تحليل استجابات المفحوصين البالغ عددهم (200) طالبًا وطالبة لجميع نماذج الاختبار الثلاث، وفق النموذج اللوجستي الثنائي المعدل، كشفت نتائج تحليل التباين الأحادي عن وجود فروق ذات دلالة إحصائية بين متوسطات معالم الصعوبة للفقرات تعزى لطريقة المعالجة، وقد جاءت تلك الفروق لصالح طريقة المعالجة الأولى، كما أظهرت النتائج عدم وجود فروق ذات دلالة إحصائية بين متوسطات معالم التمييز للفقرات تُعزى لمتغير الدراسة، وبيّنت نتائج التحليل كذلك عدم وجود فروق ذات دلالة إحصائية بين متوسطات معلمة القدرة للأفراد تعزى لمتغير الدراسة.

وأجرى اليامي (2016) دراسة لمقارنة طريقتي (التصحيح من أثر التخمين وطريقة الاختيار الجزئي) وأثرهما على خاصيتي الصدق والثبات لاختبار الاختيار من متعدّد في مبحث الكيمياء للصف الثاني الثانوي في مدينة نجران في السعودية، حيثُ تمَّ اختيار أربع مدارس عشوائيًا تضمّ (32) طالبًا وتم تقسيم طلاب كلّ صف إلى مجموعتين عشوائيًا، المجموعة (أ) تمَّ تصحيح درجات أفرادها بمعادلة التصحيح من أثر التخمين، والمجموعة (ب) تمَّ تصحيح درجات أفرادها بطريقة الاختيار الجزئي، تمَّ التأكد من تكافؤ المجموعتين في التحصيل. وتمَّ التحقق من صدق المحتوى للاختبار من خلال عرضه على المحكمين، وحساب صدق المحك التلازمي والصدق التمييزي للاختبار. ثمَّ تمَّ حساب ثبات الاختبار من خلال معادلة كرونباخ ألفا، إذ بلغ معامل الاتساق الداخلي (0.89) وعن طريق التجزئة النصفية (سبيرمان-براون)

بلغت قيمته (0.86) وبعد تطبيق الاختبار تبين عدم وجود فروق بين طرق تصحيح المجموعتين في معاملات الثبات وباختلاف طرق حساب الثبات.

وأجرى صهيبي (2018) دراسة هدفت إلى قياس أثر التخمين في اختبارات الاختيار من متعدد، تكون مجتمع الدراسة من (2904) طالبًا وطالبة في الصف التاسع، من مدارس الأنوروا (شمال الأردن)، وتم اختيار العينة بالطريقة العشوائية، الذي بلغ عددهم (159)، تم تطبيق اختبار اختيار من متعدد تم بناءه في مادة الرياضيات، حيث طبق الاختبار بنموذجين: الأول يحتوي على أربعة بدائل، والثاني تم إضافة البديل "لا أعرف الإجابة" بدلاً من إحدى الإجابات، استخدم الباحث المنهج الوصفي التحليلي، حيث قارن الباحث بين ثلاث طرق للتصحيح من أثر التخمين وهي الطريقة التقليدية (العقاب) وطريقة تصحيح الموهبات وطريقة البديل "لا أعرف الإجابة"، ثم تم التحقق من صدق المحتوى عن طريق عرضه على مختصين في هذا المجال، كما تم التحقق من الصدق المرتبط بمحك من خلال إيجاد معامل ارتباط بيرسون بين درجات الاختبار ومتوسط درجات اختبار الشهر الثاني، وتم التحقق من ثبات الاختبار عن طريق تطبيق الاختبار على عينة استطلاعية، وبعد مرور (10) أيام تم إعادة تطبيق الاختبار وحساب كل من المتوسط الحسابي، والانحراف المعياري، وقيمة معامل الارتباط بيرسون ومعامل ثبات الإعادة، وقد أظهرت النتائج وجود فروق بين المتوسطات الحسابية لأداء الطلبة باختلاف طريقة التصحيح لأثر التخمين، وأكدت الدراسة أن للتخمين أثر سلبي على الخصائص السيكومترية للاختبار (الصدق والثبات).

كما وأجرى أيضًا السعودي، والنصراوي (2018) دراسة هدفت إلى مقارنة ثلاث طرق لضبط أثر التخمين في اختبارات الاختبار من متعدد، هي طريقة التصحيح لأثر التخمين التقليدية، وطريقة تصحيح الموهبات، وطريقة استخدام البديل لا أعرف الإجابة من خلال أثرها على خصائص الاختبار، باستخدام اختبار رياضيات تحصيلي على عينة مكونة من (159) طالبًا وطالبة، وأشارت النتائج إلى وجود فروق دالة إحصائية عند مستوى الدلالة بين طريقة تصحيح الموهبات وطريقة التصحيح لأثر التخمين لصالح طريقة تصحيح الموهبات وفق محكات هي: الأداء على الاختبار، الصدق، والثبات؛ نظرًا لاعتمادها على المعرفة الجزئية لدى الطلبة. في حين لم تظهر النتائج فروق ذات دلالة إحصائية بين طريقتي التصحيح لأثر التخمين وطريقة البديل لا أعرف الإجابة.

من خلال استعراض الدراسات السابقة يلاحظ أنها تباينت فيما بينها من حيث الأهداف والعينات والأدوات المستخدمة، كما بينت الدراسات السابقة جدلية التعامل مع الطرق المختلفة لضبط أثر التخمين وعدم الاتفاق حول الأنسب منها والأكثر فاعلية للتعامل مع نتائج الاختبار، وكذلك اختلاف تأثيرها على الخصائص السيكومترية للاختبار وفقراته، إلا أن الطرق الحديثة المعتمدة على المعرفة الجزئية مثل طريقة معادلة تصحيح الموهبات وطريقة التصحيح لأثر التخمين قد أثبتت فاعليتها بشكل أكبر من الطرق التقليدية الشائعة، وهذا يثبت دور تعليمات الاختبار وإجراءاته إضافةً للمعالجات الإحصائية المتبعة في ضبط أثر التخمين، في الحد من أثر التخمين، كما أوصت بعض الدراسات صهيبي (2018) والبيامي (2010)، بإجراء المزيد من الدراسات حول أثر طرق التصحيح لأثر التخمين على الخصائص السيكومترية للاختبارات. لذلك حاولت هذه الدراسة الاستجابة لهذه التوصيات وتغطية هذا الجانب من خلال مقارنة الطرق الثلاث، والتي ينطلق كل منها من أساس نظري وإجراء إحصائي مختلف، كما أنه لا يوجد (في حدود علم الباحث) دراسة حاولت الجمع والمقارنة بين الطرق المستهدفة في هذه الدراسة، والتي تميزت بطريقة التصحيح باستخدام الترتيب الكامل.

الطريقة والإجراءات:

منهج الدراسة: استخدمت هذه الدراسة المنهج الوصفي التحليلي، وذلك لملاءمته لتحقيق أهداف الدراسة.

مجتمع الدراسة: تكوّن مجتمع الدراسة من جميع طلبة الصف العاشر في المدارس الحكومية التابعة لمديرية التربية والتعليم عمان الرابعة (لواء ماركا) في الفصل الدراسي الأول من العام الدراسي 2021/2020، والبالغ عددهم (10272).

عيّنة الدراسة: تمّ اختيار عيّنة الدراسة بالطريقة المتيسرة من طلبة الصف العاشر، فقد تمّ اختيار مدرستي، المعتصم الأساسية للبنين نظراً لعمل الباحث فيها، ومدرسة باب الواد الثانوية للبنات، نظراً لقربها من مكان عمل الباحث وضمان تعاون إدارة المدرسة والعاملين فيها، حيث بلغ عددهم (180) طالباً وطالبة، منهم (80) إناث، و(100) ذكور. أداة الدراسة: لتحقيق هدف الدراسة، وللإجابة عن أسئلتها، تم بناء اختبار تحصيلي في مادة الرياضيات (الوحدة: الخامسة) للصف العاشر الأساسي بحسب منهاج وزارة التربية والتعليم الأردنية، من نوع الاختيار من متعدد بأربعة بدائل، مع مراعاة خطوات بناء الاختبار التحصيلي التي أوردها غرونلند ولين (Gronlund & Linn, 2000).

بعد الانتهاء من إعداد الصورة الأولية للاختبار والتي تضمنت (30) فقرة، تمّ عرضها مع جدول المواصفات على مجموعة من المحكمين والمتخصصين في مجال القياس والتقويم وأساليب تدريس الرياضيات والمناهج وطرق التدريس والمشرّفين التربويين، وقد بلغ عددهم (8) محكمين، وطلب منهم إبداء آرائهم وملاحظاتهم، من حيث صياغة فقرات الاختبار لغوياً وعلمياً، ومناسبتها لمستوى طلبة الصف العاشر الأساسي، ومدى انتمائها لكلّ مهارة من مهارات الاختبار، ومدى تمثيل الفقرات للعمليات المحددة، ووضوح التعليمات، أو حذف، أو إضافة، أو تعديل ما يروونه مناسباً، والدرجة المناسبة لكلّ فقرة على تدرج خماسي.

وقد أبدى المحكمون بعض الآراء في الاختبار منها: تعديل بعض الفقرات، وحذف بعضها، وتعديل الصياغة اللغوية، وقد تراوحت الأوساط الحسابية لتقديرات المحكمين على فقرات الاستبيان لجدول المواصفات ما بين 4.02 إلى 4.66، وللحكم على فقرات الاختبار ما بين 3.09 إلى 4.51، أمّا بالنسبة للاختبار ككلّ وجدول المواصفات فكان 4.46 من أصل 5 درجات، وفي ضوء تلك الآراء تمّ الأخذ بملاحظات المحكمين، كما تمّ حذف ثلاث فقرات كانت أوساطها الحسابية أقلّ من 4.00 أي أقلّ من (80%). وتكوّن الاختبار بصورته النهائية من (27) فقرة من نوع الاختيار من متعدد.

كما تم تطبيق الاختبار على عيّنة استطلاعية قوامها (30) طالباً وطالبة من مدرستي "المعتصم وباب الواد" في محافظة العاصمة عمان، وتم تطبيق الاختبار بشكل وجاهي يومي 6 و 13 من شهر أيلول عام 2020، حيث كان الطلبة قد درسوا المحتوى في الفصل السابق، لغرض التحليل الأولي للفقرات والتعرف على معاملات الصعوبة والتمييز للفقرات، وكذلك لمعرفة الزمن اللازم للاختبار، وقد روعي عند تطبيق الاختبار الاستطلاعي إعطاء الطلبة تعليمات واضحة عن كيفية الإجابة عن الأسئلة، والهدف منه. وتم تحليل البيانات للعيّنة الاستطلاعية، حيث تمّ حساب معامل الصعوبة لكلّ فقرة، وذلك بإيجاد نسبة الإجابات الصحيحة على الفقرة من بين الإجابات الكلية على هذه الفقرة، كما تمّ حساب معاملات التمييز للفقرات، والجدول (1) يوضّح ذلك.

جدول (1): قيم معاملات الصعوبة والتمييز لفقرات الاختبار

رقم الفقرة	معامل الصعوبة	معامل التمييز	رقم الفقرة	معامل الصعوبة	معامل التمييز	رقم الفقرة	معامل الصعوبة	معامل التمييز
1	0.733	0.754	10	0.70	0.765	19	0.533	0.774
2	0.60	0.742	11	0.733	0.779	20	0.35	0.763
3	0.40	0.766	12	0.70	0.761	21	0.40	0.744
4	0.60	0.782	13	0.667	0.762	22	0.70	0.787

0.774	0.733	23	0.740	0.533	14	0.749	0.467	5
0.780	0.70	24	0.745	0.533	15	0.748	0.60	6
0.783	0.767	25	0.752	0.333	16	0.747	0.733	7
0.777	0.733	26	0.770	0.733	17	0.767	0.70	8
0.766	0.767	27	0.754	0.667	18	0.754	0.667	9

يبين الجدول (1) قيم معاملات الصعوبة ومعاملات التمييز لكل فقرة من فقرات الاختبار، بعد تطبيقها على العينة الاستطلاعية، حيث تراوحت قيم معاملات صعوبة الفقرات بين (0.333-0.767)، بينما تراوحت قيم معاملات تمييز الفقرات بين (0.742-0.787)، وتعد هذه القيم مقبولة باعتبار أن فقرات الاختبار تتمتع بدرجات مناسبة من الصعوبة والتمييز، إذ إن أفضل معامل صعوبة للفقرة الاختبارية هو 50% وما حولها. أما معاملات التمييز فقد كانت جميعها ذات تمييز عالي وهي أعلى من (0.7).

صدق الاختبار: للتحقق من دلالات صدق الاختبار باستخدام النظرية الكلاسيكية في القياس فقد استخدمت الطرق الآتية:

- صدق المحتوى للاختبار: وذلك من خلال الإجراءات والخطوات التي اتبعت في بناء الاختبار، حيث تم تحليل المحتوى، وإعداد جدول للمواصفات، وتحكيم الاختبار من قبل خبراء مختصين.
- الصدق المحكي: تم حساب معامل الارتباط بين أداء العينة الاستطلاعية على الاختبار وعلاماتهم في مادة الرياضيات للفصل السابق كمحك، حيث بلغت قيمته (0.614) وهي علاقة ارتباطية جيدة وذات دلالة إحصائية.
- الصدق التمييزي: تم حساب الصدق التمييزي للاختبار؛ وذلك لمعرفة إذا كان الاختبار يميز بين المجموعات المختلفة، حيث قام بتطبيق الاختبار على عينة مكونة من (30) طالباً من طلاب الصف التاسع، وبعد تصحيح علاماتهم بالطريقة الاعتيادية تم حساب المتوسطات الحسابية لدرجاتهم ومقارنتها بدرجات عينة الطلاب من الصف العاشر، وذلك من خلال استخدام اختبار (ت)، والجدول (2) يبين ذلك.

جدول (2) نتائج اختبار (ت) للعينات المستقلة لمقارنة المتوسطات الحسابية لدرجات طلاب الصف التاسع ولدرجات طلاب

الصف العاشر

الصف	العدد	المتوسط الحسابي	الانحراف المعياري	قيمة (ت)	الدلالة الإحصائية
التاسع	30	53.77	12.90	-1.698	0.0448
العاشر	30	57.10	12.97		

تظهر النتائج في الجدول (2) وجود فروق ذات دلالة إحصائية عند مستوى دلالة (0.05)، وهذا يدل على أن الاختبار يمتلك صدق تمييزي بين المجموعات المختلفة.

ثبات الاختبار: تم التحقق من ثبات الاختبار من خلال تطبيق الاختبار على العينة الاستطلاعية من خارج عينة الدراسة الأصلية وقوامها (30) طالباً من طلاب الصف العاشر، ثم أعيد تطبيق الاختبار على نفس العينة بعد فاصل زمني مدته أسبوع من تاريخ التطبيق الأول، وتم استخراج معاملات الثبات على النحو الآتي:

- حساب معامل ارتباط بيرسون (Pearson Correlation) بين التطبيق الأول والثاني، حيث بلغت قيمة معامل ارتباط بيرسون (0.88).

- حساب معامل الثبات كرونباخ الفا والتي بلغت قيمته (0.79).

وتعتبر هذه المعاملات جيدة مما يجعل الاختبار صالحاً للتطبيق، وبعد تأكد الباحث من صدق وثبات الاختبار قام بعمل نموذجين للأداة تحتوي على نفس الأسئلة وتختلف فقط في طريقة التصحيح والتعليمات المرافقة لكل طريقة تصحيح، النموذج (أ) يصحح بطريقة الترتيب الكامل وبالطريقة الاعتيادية، والنموذج (ب) يصحح بطريقة العقاب.

تطبيق الاختبار وجمع البيانات: تم تطبيق نفس الاختبار مرتين على نفس أفراد عينة الدراسة باختلاف تعليمات الاختبار التي تتعلق بطريقة التصحيح المتبعة، وقبل بدء الطلبة بالإجابة على الأسئلة تم توضيح طبيعة الاختبار والغرض منه، كما تم شرح تعليمات الاختبار وطريقة الإجابة عليه، وتوضيح الاستراتيجية التي يفضل اتباعها لكل نموذج، وقد تم إبلاغ الطلبة بأنه سيكون لهذا الاختبار جزء من العلامة النهائية للمادة، وذلك لتوفير المزيد من اهتمام الطلبة بالاختبار، مما يزيد من صدق نتائج الاختبار، كما تم تقدير الوقت اللازم للإجابة عن فقرات الاختبار (45 دقيقة).

طريقة تصحيح الاختبار: تم استخدام ثلاث طرق للتصحيح بحيث تعطى نتيجة مختلفة لكل طريقة، حيث تم تصحيح النموذج الأول بطريقة التصحيح التقليدية (العقاب) والنموذج الثاني بالطريقة الاعتيادية وطريقة الترتيب الكامل، حيث تم رصد درجة واحدة لكل فقرة من فقرات الاختبار، حيث تضمن الاختبار (27) فقرة يتبعها أربعة بدائل من نوع الاختيار من متعدد، وبذلك تصبح الدرجة العظمى للاختبار (27) والصغرى (صفر درجة) ولتجاوز الكسور وتوحيد الدرجات، تم إجراء تحويل خطي للعلامات لتصبح قيم صحيحة وذلك بضرب الدرجة في 3 لتصبح الدرجة العليا (81). الأساليب الإحصائية:

- للإجابة عن السؤال الأول تم استخدام المتوسطات الحسابية والانحرافات المعيارية والالتواء والتفطح لأداء الطلبة على فقرات الاختبار، كما قام بحساب معاملات الارتباط (بيرسون) للدرجة الكلية بين طرق التصحيح المستخدمة، ولإيجاد أثر الطريقة المستخدمة لضبط أثر التخمين في متوسط أداء الطلبة على الاختبار استخدم الباحث تحليل التباين للقياسات المتكررة، واختبار أقل فرق دال (LSD) للمقارنات الثنائية بين المتوسطات الحسابية لدرجات الطلبة على الاختبار باختلاف طريقة التصحيح المستخدمة.
 - للإجابة على السؤال الثاني تم إيجاد معاملات الارتباط بين نتائج الاختبار وبين الدرجات الكلية وفق طرق التصحيح المستخدمة، كما استخدم اختبار فيشر لدلالة الفروق بين معاملات الصدق المرتبط بالمحك لطرق التصحيح المختلفة، وتم إيجاد نتائج التحليل العاملي.
 - للإجابة على السؤال الثالث تم إيجاد معاملات الثبات بطريقة الاتساق الداخلي كرونباخ الفا لكل طريقة، وتم استخدام تحليل التباين للقياسات المتكررة للتحقق من أثر الطريقة المستخدمة لضبط أثر التخمين في قيم معاملات ثبات الاختبار، واختبار أقل فرق دال (LSD) للمقارنات الثنائية بين قيم الثبات باختلاف طريقة التصحيح.
- النتائج المتعلقة بالسؤال الأول: هل توجد فروق ذات دلالة إحصائية عند مستوى $\alpha = 0.05$ بين المتوسطات الحسابية لأداء الطلبة في الاختبار يعزى لطريقة ضبط أثر التخمين؟، فلإجابة على هذا السؤال تم إيجاد مجموعة من الإحصاءات الوصفية لأداء الطلبة على فقرات الاختبار تبعاً للطرق المستخدمة لضبط أثر التخمين، والجدول رقم (3) يوضح ذلك.

جدول (3): المتوسطات الحسابية والانحرافات المعيارية والالتواء والتفطح لأداء الطلبة على فقرات الاختبار تبعاً للطرق المستخدمة لضبط أثر التخمين.

الطريقة	المتوسط الحسابي	الانحراف المعياري	الالتواء	التفطح
الاعتيادية	53.41	12.13	0.760	-0.483
الترتيب الكامل	62.21	9.34	0.363	-0.714
العقاب	50.58	12.27	0.816	-0.430

يظهر جدول (3) وجود فروق ظاهرية بين المتوسطات الحسابية لأداء الطلبة على الاختبار باختلاف طريقة التصحيح، حيثُ أظهرت طريقة التخمين (الترتيب الكامل) أعلى متوسط حسابي والبالغ (62.21) وبانحراف معياري (9.34)، تلاها طريقة التخمين (الاعتيادية) بمتوسط حسابي (53.41) وبانحراف معياري (12.13)، وأخيراً طريقة التخمين (العقاب) بوسط حسابي (50.58) وبانحراف معياري (12.27).

ويلاحظ من قيم التقلطح والالتواء، إنّ التوزيعات كان لها قيم تقلطح سالبة أمّا قيم الالتواء جميعها موجبة، وهذا يدلّ على أنّ الأفراد ذوي التحصيل المرتفع خمنوا الإجابات بشكل أفضل من ذوي التحصيل المنخفض. كما تمّ حساب معاملات ارتباط بيرسون بين الدرجة الكلية باختلاف طريقة التصحيح المستخدمة بدلالة عدد الإجابات الصحيحة والفرق بين الدرجات المعدلة لكلّ طريقتين، والجدول رقم (4) يوضّح ذلك.

جدول (4) معاملات الارتباط (بيرسون) للفروق بين الدرجات الكلية حسب طرق التصحيح المستخدمة.

الطريقة	معامل الارتباط للفروق
الاعتيادية*الترتيب الكامل	-0.734**
الترتيب الكامل*العقاب	-0.695**
الاعتيادية*العقاب	-0.577**

تظهر النتائج في جدول (4) بأنّ جميع معاملات الارتباط سالبة وذات دلالة إحصائية عند مستوى دلالة (0.01)، وهذا يدلّ على أنّ الطلبة ذوي التحصيل المرتفع لجأوا إلى التخمين بشكل أقلّ من الطلبة ذوي التحصيل المنخفض، وأنّ الطلبة ذوي التحصيل المنخفض حصلوا على درجات من خلال التخمين أكثر من ذوي التحصيل المرتفع، وكان الارتباط الأعلى ناتج عن الفرق بين الطريقة الاعتيادية والترتيب الكامل.

كما تم استخدام تحليل التباين للقياسات المتكررة لاختبار أثر الطريقة المستخدمة لضبط أثر التخمين في متوسط أداء الطلبة على الاختبار من خلال اختبار دلالة الفروق بين المتوسطات الحسابية، والجدول رقم (5) يبيّن ذلك.

جدول (5): تحليل التباين للقياسات المتكررة لأثر الطريقة المستخدمة لضبط أثر التخمين في متوسط أداء الطلبة على الاختبار

اختبار الأثر	مصدر التباين	مجموع المربعات	درجات الحرية	متوسط مجموع المربعات	ف	الدلالة الإحصائية
داخل الأفراد	الطريقة	13252.12	2	6626.06	17.03	0.000
	الخطأ	68867.28	177	389.08		

يلاحظ من الجدول (5) نتائج تحليل التباين للقياسات المتكررة لأثر الطريقة المستخدمة لضبط أثر التخمين في متوسط أداء الطلبة على الاختبار، حيثُ أظهرت النتائج وجود فروق ذات دلالة إحصائية بين الطرق المستخدمة لضبط أثر التخمين لأداء الطلبة على الاختبار، حيثُ بلغت قيمة ف (17.03) وبدلالة إحصائية (0.00)، ولمعرفة لأيّ طريقة تُعزى الفروق تمّ استخدام اختبار LSD للمقارنات الثنائية البعدية، والجدول رقم (6) يوضّح ذلك.

جدول (6): اختبار (LSD) للمقارنات الثنائية بين المتوسطات الحسابية لدرجات الطلبة على اختبار الرياضيات باختلاف طريقة التصحيح المستخدمة

طريقة التصحيح	متوسط الفروق بين الطريقتين	الخطأ المعياري	مستوى الدلالة
الاعتيادية	8.81*	1.194	0.000*
العقاب	2.83*	1.194	0.018*
الترتيب الكامل	11.63*	1.194	0.000*

يلاحظ من الجدول (6) بأنه يوجد فروق ذات دلالة إحصائية بين طريقة التصحيح الاعتيادية وطريقة الترتيب الكامل ولصالح طريقة الترتيب الكامل بمستوى دلالة (0.000)، وبين طريقة التصحيح الاعتيادية وطريقة العقاب ولصالح طريقة التصحيح الاعتيادية بمستوى دلالة (0.018)، وبين طريقة العقاب وطريقة الترتيب الكامل ولصالح طريقة الترتيب الكامل بمستوى دلالة (0.000).

النتائج المتعلقة بالسؤال الثاني: هل توجد فروق ذات دلالة إحصائية عند مستوى ($\alpha=0.05$) في الصدق المحكي لدرجات الاختبار يعزى لطريقة ضبط أثر التخمين؟، فلاجابة على السؤال الثاني تم مقارنة طرق التخمين من خلال طريقتين للصدق، وهما: الصدق المحكي باستخدام معامل الارتباط بين درجات الطلبة على الاختبار تبعاً لكل طريقة والدرجة الكلية لاختبار المحك المُعد من قِبَل المعلمين، كما تم استخدام صدق البناء من خلال إجراء التحليل العاملي لدرجات الطلبة وفق كل طريقة من طرق التصحيح من أثر التخمين المستخدمة، والجدول (7) يبين معاملات الصدق المرتبط بالمحك تبعاً لطريقة التصحيح.

جدول (7): معاملات الصدق المرتبط بالمحك تبعاً لطريقة التصحيح

الطريقة	الترتيب الكامل	العقاب
الترتيب الكامل	1	
العقاب	0.980**	1
الاعتيادية	0.985**	0.962**

تبيّن نتائج الجدول (7) وجود ارتباط عالي بين درجات الطلبة على الاختبار تبعاً لطرق التصحيح المختلفة، ولاختبار دلالة الفروقات بمعاملات الارتباط تم استخدام اختبار معادلة " فيشر " لتحويل معاملات الارتباط إلى قيم زائية وفحص دلالتها باستخدام معادلة فيشر (Fisher) وتبيّن أنّ قيمة $z = -4.80$ وكانت ذات دلالة إحصائية عند مستوى دلالة ($\alpha=0.05$). مما يؤكّد وجود فروق بين معاملات الارتباط، و تمّ مقارنة معاملات الارتباط الثلاث مع بعضها البعض والجدول (8) يوضّح ذلك.

الجدول (8) معاملات الارتباط وقيم ز المناظرة لها ومستوى الدلالة

المقارنة	قيم "z"	مستوى الدلالة
الاعتيادية*العقاب	-3.28	0.001
الاعتيادية*الترتيب الكامل	3.30	0.00
العقاب*الترتيب الكامل	6.59	0.00

يتبيّن من الجدول (8) وجود فروق دالة إحصائية بين معاملات الارتباط، وعند مقارنة معامل ارتباط طريقة العقاب مع الطريقة الاعتيادية لصالح الطريقة الاعتيادية، ولصالح معامل ارتباط طريقة الترتيب الكامل مقارنة مع الطريقة الاعتيادية ولصالح معامل ارتباط طريقة الترتيب الكامل مقارنة مع طريقة العقاب، أي أنّ الفروق لصالح معامل ارتباط طريقة الترتيب الكامل.

أما بالنسبة لصدق البناء فقد تمّ إجراء التحليل العاملي لدرجات الطلبة وفق طرق التصحيح المختلفة المستخدمة، والجدول (9) يبين ذلك.

جدول (9): نتائج التحليل العاملي

العامل	الاعتيادية			الترتيب الكامل			العقاب	
	النسبة التباين المفسر %	نسبة التباين المفسر %	النسبة التباين المفسر %	النسبة التباين المفسر %	النسبة التباين المفسر %	النسبة التباين المفسر %	النسبة التباين المفسر %	النسبة التباين المفسر %
1	5.738	21.250	21.250	3.861	14.302	14.302	5.632	20.859
2	3.932	14.563	35.813	3.530	13.075	27.377	3.829	14.183
3	3.556	13.172	48.985	2.995	11.092	38.468	3.408	12.622
4	2.897	10.731	59.717	2.850	10.554	49.023	2.756	10.207
5	2.157	7.988	67.705	2.441	9.040	58.063	2.009	7.440
6	2.004	7.421	75.125	2.269	8.404	66.467	1.886	6.985
7	1.720	6.371	81.496	2.262	8.377	74.844	1.571	5.818
8	1.317	4.878	86.375	2.196	8.133	82.977	1.146	4.243
9	1.107	4.099	90.474	2.024	7.497	92.70	.969	3.590

يظهر الجدول (9) نتائج التحليل العاملي، حيثُ أظهرت النتائج بأنَّ الطريقة الاعتيادية لها تسعة عوامل فسرت ما نسبته (90.47%) من التباين الكلي، في حين أظهرت طريقة الترتيب الكامل تسعة عوامل فسرت ما نسبته (92.7%) من التباين الكلي، وأظهرت طريقة العقاب تسعة عوامل فسرت ما نسبته (85.94%) من التباين الكلي. وعليه فإنَّ طريقة الترتيب الكامل تكون قد فسرت النسبة الأعلى من التباين الكلي، تلتها الطريقة الاعتيادية، ثمَّ طريقة العقاب.

النتائج المتعلقة بالسؤال الثالث: هل توجد فروق ذات دلالة إحصائية عند مستوى ($\alpha=0.05$) في ثبات درجات الاختبار يعزى لطريقة ضبط أثر التخمين؟، للإجابة عن السؤال الثالث تم مقارنة الطرق الثلاث المستخدمة من خلال معامل كرونباخ ألفا للاتساق الداخلي، حيثُ تمَّ حساب قيم ألفا للعينة الكلية لكلِّ طريقة ولثلاث عينات فرعية أخرى، وهي: الأفراد التي تقلَّ درجاتهم عن 45، والأفراد التي تقلَّ درجاتهم عن 60، والأفراد التي تقلَّ درجاتهم عن 72، وتمَّ تحديد هذه المجموعات من خلال المئينات (90، 75، 50) على التوالي وبشكل تقريبي، والجدول رقم (10) يوضِّح قيم كرونباخ ألفا لكلِّ طريقة من الطرق الثلاث.

جدول (10): قيم معاملات كرونباخ ألفا لكلِّ طريقة

الطريقة	أقل من 45	أقل من 60	أقل من 72	العينة ككل
الترتيب الكامل	0.81	0.75	0.80	0.83
الاعتيادية	0.44	0.61	0.65	0.70
العقاب	0.40	0.50	0.43	0.61

يوضِّح الجدول (10) قيم معاملات كرونباخ ألفا لكلِّ طريقة، حيثُ أظهرت النتائج بأنَّ طريقة الترتيب الكامل أظهرت أعلى قيم معاملات ثبات (0.81، 0.75، 0.80)، تلاها الطريقة الاعتيادية (0.44، 0.61، 0.65)، فيما أظهرت طريقة العقاب أقلَّ معاملات ثبات (0.40، 0.50، 0.43)، ولاختبار دلالة الفروق بين معاملات الثبات تم استخدام اختبار تحليل التباين للقياسات المتكررة، والجدول رقم (11) يوضِّح ذلك.

جدول (11): تحليل التباين للقياسات المتكررة لأثر الطريقة المستخدمة لضبط أثر التخمين في قيم معاملات ثبات الاختبار

اختبار الأثر	مصدر التباين	مجموع المربعات	درجات الحرية	متوسط مجموع المربعات	ف	الدلالة الإحصائية
داخل الأفراد	الطريقة	0.163	2	0.082	8.982	0.016*
	الخطأ	0.055	6	0.009		

يوضح الجدول (11) نتائج تحليل التباين للقياسات المتكررة لأثر الطريقة المستخدمة لضبط أثر التخمين في قيم معاملات ثبات الاختبار، حيث أظهرت النتائج وجود فروق ذات دلالة إحصائية في قيم معاملات ثبات الاختبار تبعاً لطريقة التخمين المستخدمة، ولمعرفة مصدر هذه الفروق تم استخدام اختبار LSD للمقارنات الثنائية البعدية، والجدول رقم (12) يوضح ذلك.

جدول (12): اختبار (LSD) للمقارنات الثنائية بين المتوسطات الحسابية لقيم لمعاملات ثبات الاختبار باختلاف طريقة التصحيح المستخدمة

طريقة التصحيح	متوسط الفروق بين الطريقتين	الخطأ المعياري	مستوى الدلالة
الترتيب الكامل	0.297*	0.07784	0.009*
العقاب	0.023	0.07784	0.774
العقاب	0.273*	0.07784	0.013*

يلاحظ من الجدول (12) وجود فروق ذات دلالة إحصائية بين طريقة التصحيح الاعتيادية وطريقة الترتيب الكامل بمستوى دلالة (0.009) ولصالح طريقة الترتيب الكامل، وبين طريقة التصحيح العقاب وطريقة الترتيب الكامل بمستوى دلالة (0.013) ولصالح طريقة الترتيب الكامل في حين لم تكن الفروق ذات دلالة إحصائية بين طريقة العقاب والطريقة الاعتيادية.

مناقشة النتائج:

السؤال الأول: هل توجد فروق ذات دلالة إحصائية عند مستوى $\alpha = 0.05$) بين المتوسطات الحسابية لأداء الطلبة في الاختبار يعزى لطريقة ضبط أثر التخمين؟

أظهرت النتائج المتعلقة بالسؤال الأول وجود فروق ذات دلالة إحصائية عند مستوى الدلالة ($\alpha = 0.05$) بين طريقة التصحيح الاعتيادية وطريقة تصحيح الترتيب الكامل بمستوى دلالة (0.00) ولصالح طريقة تصحيح الترتيب الكامل، وبين طريقة التصحيح الاعتيادية وطريقة التصحيح بالعقاب بمستوى دلالة (0.018) ولصالح طريقة التصحيح الاعتيادية، وبين طريقة التصحيح العقاب وطريقة التصحيح الترتيب الكامل بمستوى دلالة (0.00) ولصالح طريقة التصحيح الترتيب الكامل. فقد كانت طريقة الترتيب الكامل هي الأعلى متوسط تليها الطريقة الاعتيادية ثم طريقة العقاب، ويمكن تفسير هذه النتيجة بأن الطلبة ذوي الأداء المرتفع خمنوا الإجابات بشكل أقل من ذوي الأداء المنخفض، أي أن الطلبة ذوي الأداء المنخفض حصلوا على درجات تعزى للتخمين أكثر من ذوي الأداء المرتفع، إذ يقل التخمين بزيادة القدرة، وهذا يزيد من فعالية تطبيق طرق التصحيح للحد من التخمين الذي يؤثر في الخصائص السيكمترية للاختبار. وهنا تأخذ طرق ضبط أثر التخمين ثلاثة مواقف عند إجابة المفحوص على فقرات الاختبار من متعّد هي: أن المفحوص يعرف الإجابة الصحيحة، أو أن المفحوص لا يعرف الإجابة الصحيحة ويترك الفقرة دون إجابة، أو أن المفحوص لا يعرف الإجابة الصحيحة على الفقرة لكنه يجيب عليها بشكل عشوائي.

وقد أظهرت هذه الدراسة أفضلية طريقة تصحيح الترتيب الكامل، على الطريقة الاعتيادية، وعلى طريقة التصحيح بالعقاب، وفي طريقة الترتيب الكامل يمكن التمييز بين ثلاثة مستويات من المعرفة لدى الطلبة، فالطالب الذي يمتلك معرفة تامة يعطي الرتبة الأولى للبديل الصحيح، والمفحوص الذي لديه معرفة جزئية يعطي رتبة متوسطة للبديل الصحيح والمفحوص الذي لا يمتلك المعرفة يعطي الرتبة الأخيرة للبديل الصحيح (أبو فودة ويونس، 2012). وتفسر هذه النتائج بأن طريقة تصحيح الترتيب الكامل ذات فعالية كبيرة في ضبط التخمين العشوائي، وفي تشجيع الطلبة على استخدام معرفتهم التامة عند الإجابة عن فقرات الاختبار، الذي ينعكس على العلامة النهائية بما يتوافق مع معرفتهم

التامة، وقد أشارت قيمة الانحراف المعياري إلى زيادة نسبة التباين الحقيقي في قدرات الطلبة، بالمقابل كانت باقي الطرق تقلل من التخمين العشوائي.

وانتقلت نتائج هذه الدراسة مع دراسة السعودي والنصراوي (2018) التي جاء في نتائجها وجود فروق دالة إحصائية عند مستوى الدلالة بين طريقة تصحيح المموهات وطريقة التصحيح لأثر التخمين لصالح طريقة تصحيح المموهات وفق محكات هي: الأداء على الاختبار، الصدق، والثبات؛ نظراً لاعتمادها على المعرفة الجزئية لدى الطلبة. كما انتقلت نتائج هذه الدراسة مع دراسة صهيب (2018) التي جاء في نتائجها وجود فروق بين المتوسطات الحسابية لأداء الطلبة باختلاف طريقة التصحيح لأثر التخمين.

كما انتقلت نتائج هذه الدراسة مع دراسة بني عطا والشرفين (2017) التي جاء في نتائجها وجود فرق ذي دلالة إحصائية عند مستوى الدلالة بين متوسطي درجات أداء الطلبة في الاختبار التحصيلي في الحاسوب وبين النموذجين الأول والرابع، وعدم وجود فرق ذات دلالة إحصائية بين النماذج الأخرى.

وانتقلت مع نتائج الصمادي (2012) التي أظهرت وجود فروق ذات دلالة إحصائية في تحصيل الطلبة حيث كانت الطريقة الأكثر مراعاة للمعرفة الجزئية هي الأكثر فاعلية وتعطي العلامة الأعلى.

السؤال الثاني والذي ينص على: هل توجد فروق ذات دلالة إحصائية عند مستوى هل توجد فروق ذات دلالة إحصائية عند مستوى $(\alpha = 0.05)$ في الصدق المحكي لدرجات الاختبار يعزى لطريقة ضبط أثر التخمين؟

أظهرت النتائج المتعلقة بالسؤال الثاني وجود فروق ذات دلالة إحصائية عند مستوى $(\alpha = 0.05)$ في الصدق المحكي للاختبار يعزى لطريقة ضبط أثر التخمين، بين طريقتي التصحيح لأثر التخمين الترتيب الكامل والعقاب ولصالح طريقة الترتيب الكامل، وهنا كانت الأفضلية لطريقة الترتيب الكامل حيث ارتفع الصدق، ويعزو الباحث هذه النتيجة إلى أفضلية طريقة الترتيب الكامل، والتي تصحح التخمين العشوائي وتعطي من شأن المعرفة التامة لدى الطلبة عبر طرق التصحيح أثر التخمين، ثم تليها طريقة العقاب، والتي تمنع التخمين كلياً ولا تعير للمعرفة الجزئية وزناً، وقد يعزى ذلك إلى أنه عند استخدام هذه الطريقة، تخفض دافعية الطلبة، ويزداد قلق الاختبار فيقل التخمين، مما يفقد بعض الطلبة إلى ترك بعض الفقرات دون إجابة، خوفاً من العقاب عند الإجابة الخاطئة، مما يعني عدم الكشف عن معرفتهم الجزئية، وقد يعني انخفاض الصدق في هذه الطريقة، مما يؤدي إلى قيام الطلبة بترك بعض الفقرات وذلك يؤثر على صدق المحتوى لانخفاض تمثيل الاختبار للمادة الدراسية.

وانتقلت نتائج هذه الدراسة مع دراسة بني عطا والشرفين (2017) التي جاء في نتائجها وجود فرق ذي دلالة إحصائية عند مستوى الدلالة بين معاملي الصدق المرتبط بمحك الخاصين بالنموذجين الأول والرابع، وعدم وجود فرق ذي دلالة إحصائية بين قيم معاملات الصدق الخاصة بالنماذج الأخرى.

وتشير دراسة النبهان (2002، Alnabhan) التي جاء في نتائجها أن الطرق الثلاث المختلفة أعطت معاملات صدق مختلفة عن طريقة المعرفة الجزئية التي تقوم على افتراض أن المفحوص يمتلك معرفة تمكّنه من حذف بعض البدائل عند إجابته عن الفقرة.

السؤال الثالث والذي ينص على: هل توجد فروق ذات دلالة إحصائية عند مستوى $(\alpha = 0.05)$ في ثبات درجات الاختبار يعزى لطريقة ضبط أثر التخمين؟

أظهرت النتائج المتعلقة بالسؤال الثالث وجود فروق ذات دلالة إحصائية عند مستوى $(\alpha = 0.05)$ في قيم معاملات ثبات الاختبار تبعاً لطريقة التخمين المستخدمة، حيث أظهرت وجود فروق ذات دلالة إحصائية بين طريقة

التصحيح الاعتيادية وطريقة التصحيح الترتيب الكامل ولصالح طريقة التصحيح الترتيب الكامل، وبين طريقة التصحيح بالعقاب وطريقة التصحيح الترتيب الكامل ولصالح طريقة التصحيح الترتيب الكامل. ويعزو الباحث الاختلاف في قيم الثبات للطرق المختلفة للتصحيح إلى اختلاف الاستراتيجيات التي يستخدمها الطلبة، والتي تنعكس حتمًا على مدى استخدامهم للمعرفة التامة، ومدى المخاطرة بالتخمين بناءً على وجود العقاب من عدمه.

عند التصحيح بطريقة الترتيب الكامل، وامتلاك الطالب معرفة جزئية تقربه للإجابة الصحيحة، يتمكن الطالب من الحصول على أكبر قدر ممكن الدرجة لكل فقرة، وبالتالي ترتفع الدرجة الكلية للطالب على الاختبار، وهذا يؤدي إلى قياس أدق للمعرفة التامة لدى الطالب مقارنة بالطرق الأخرى، والذي يؤدي إلى ارتفاع نسبة الثبات، بالمقابل تساعد الطرق الباقية الطالب على ترك فقرات لا يتأكد من الإجابة عليها، مما يؤدي إلى زيادة التخمين، وزيادة التباين في الأخطاء، وبالتالي نقصان قيمة الثبات.

وبمقارنة نتائج الثبات مع الدراسات السابقة الأخرى نجد أنّ طرق التصحيح حققت ثباتًا مرتفعًا في الدراسة الحالية ودراسة النبهان (Alnabhan، 2002) ودراسة الزبون (2004)، ودراسة دودين (Dodeen، 2005)، ودراسة السعودي، والنصراوي (2018).

أما دراسة صهيبي (2018) فقد أكدت الدراسة أن للتخمين أثر سلبي على الخصائص السيكمترية للاختبار (الصدق والثبات). في حين لم تظهر دراسة اليامي (2016) والصمادي (2012) أية فروق ذات دلالة بين معاملات الثبات تعزى لاختلاف طرق التصحيح.

توصيات الدراسة: في ضوء النتائج التي تم التوصل إليها، توصي الدراسة بما يلي:

- ينبغي التعميم على المعلمين بوضع تعليمات للاختبار، بحيث توجه الطلبة عدم الاعتماد على التخمين، وقراءة السؤال بتمعن والتفكير فيه لتحديد البديل الصحيح.
- ضرورة تفعيل تطبيق طرق التصحيح لأثر التخمين في اختبارات الاختيار من متعدد، وخصوصًا طريقة الترتيب الكامل، وطريقة التصحيح بالعقاب.
- ضرورة إجراء المزيد من البحوث لدراسة جدوى تطبيق طرق التصحيح على الخصائص السيكمترية للاختبارات. ومقارنة طريقة الترتيب الكامل بطرق أخرى.
- إجراء البحوث المتعلقة بطرق التصحيح المستخدمة في الدراسة، ومقارنتها بمغيرات أخرى مثل: الجنس، والعمر، والمستوى التعليمي.
- الاستفادة من نتائج هذه الدراسة في الدراسات المستقبلية بإجرائها على مراحل دراسية أخرى ومواد دراسية جديدة ليتمّ تعميمها.

المراجع العربية:

- أبو فودة، باسل، ويونس، نجاتي (2012). الاختبارات التحصيلية. ط1. عمان: دار المسيرة.
- بني عطا، زايد، والشرفين، علي (2017). أثر موقع الفقرات الصعبة في اختبار اختيار من متعدد على خصائصه السيكمترية وأداء الطلبة عليه، مجلة اتخاذ الجامعات العربية وعلم النفس، 15(3)، ص 93-129.
- الخرشه، طه (2016). أثر طرق معالجة أثر التخمين على تقدير إحصائيات الأفراد والفقرات في اختبارات الاختيار من متعدد وفق النظرية الحديثة في القياس. مجلة جامعة النجاح للأبحاث (العلوم الإنسانية). 30(12). ص 23-66.
- الزبون، حابس (2004). أثر استخدام ثلاث طرق للتصحيح لضبط أثر التخمين في الخصائص السيكمترية للاختبارات الاختيار من متعدد في ضوء نظرية الاستجابة للفقرة. رسالة ماجستير غير منشورة، جامعة مؤتة، الكرك، الأردن.

السعودي، شريف، والنصراوي، معين (2018). مقارنة طرق ضبط أثر التخمين على اختبارات الاختيار من متعدد وأثرها على خصائص الاختبار، مجلة العلوم التربوية في الجامعة الأردنية، 45(4)، ص 597-615.

الصمادي، مروان (2012). أثر طريقة تصحيح اختبار الصواب-الخطأ المتعدد على الخصائص السيكومترية للاختبار وفقراته. مجلة العلوم التربوية والنفسية البحرين، 13(3)، ص 281-306.

صهيب، عبدالله (2018). قياس أثر التخمين في اختبارات الاختيار من متعدد في مادة الرياضيات للصف التاسع الأساسي في ضوء النظرية الكلاسيكية في القياس. رسالة ماجستير غير منشورة، جامعة عمان العربية، عمان، الأردن.

عبد الهادي، نبيل (2001). تدني مستوى التحصيل والانجاز المدرسي أسبابه وعلاجه، ط2. عمان: دار وائل للنشر والتوزيع.

علام، صلاح الدين (2010). القياس والتقويم التربوي في العملية التدريسية. عمان: دار المسيرة للنشر والتوزيع.

عودة، أحمد (2010). القياس والتقويم في العملية التدريسية. ط1. اربد: دار الأمل للنشر والتوزيع.

عودة، أحمد (2014). تصحيح معاملات صعوبة الفقرات لأثر التخمين في أسئلة الاختيار من متعدد: صورة معدلة لمعادلة جيلفورد. مجلة العلوم التربوية والنفسية، 8(2)، ص 248-257.

اليامي، محمد (2016). بعض طرق تصحيح التخمين وأثرها على خاصيتي الصدق والثبات لاختبار الاختيار من متعدد. مجلة عالم التربية، العدد (53)، ص 1-31.

المراجع الأجنبية:

Alnabhan, M. (2002). an empirical investigation of the effect of three methods of handling guessing and risk taking on the psychometric properties of a test. *Social Behavior and Personality*. 30 (7), 645-652.

Annie, W., & Chan, H. (2012). evaluation of the three multiple-choice assessment methods in a human factors engineering course. *Journal of the Chinese Institute of Industrial Engineers*. Vol. 29(7), PP. 466-476.

Dodeen, Hamzeh M. (2005). Using Distractors in Correcting for Guessing in Multiple Choice Tests. *Dirasat Educational Sciences*. Vol. 32(1), PP. 192-198.

Frisbie, D., & Sweeney, D. (1982). The Relative Merits of Multiple True-False achievement tests. *Journal of Educational Measurement*. Vol. 19(1), PP. 29-35.

Gronlund, N. & Linn, R. (2000). *Measurement and Assessment in Teaching* 8th Edition. U.S.A: Prentice-Hall.

Kubinger, K., & Holocher – Eftl, S., & Reif, M. (2010). On Minimizing Guessing Effects on Multiple-Choice Items: Superiority of a Two Solutions and Three Distractors | Item Format to a one Solution and Five Distractors Item Format. *International Journal of Selection and Assessment*. Vol (1) 18, PP. 111-115.

Mehrens, W., & Lehmann, I. (1991). *Measurement and Evaluation in Education and Psychology*. 4th. USA: Thomson Learning.

Ossai, P. (2015). Gender Differences in Guessing Tendencies in Mathematics. *British Journal of Education*. 3(8). Pp. 78 – 85.

Ravesloot, C. & Van der Schaaf, M. & Muijtjens, A. & Haaring, C. & Kruitwagen, C. & Beek f. & Bakker, J. vanschijk, j. & Ten Cate, T. (2015). The Dong Knows Option in Progress Testing. *Advance in Health Sciences Education*. Springer. 20(5), pp 1325-1338.

Thorndike, R. (1982). *Applied Psychometrics*. London: Houghton Mifflin Company Boston.