أثر طريقة معالجة التخمين والبيانات المفقودة على الخصائص السيكومترية للفقرات وفقًا لنظرية استجابة الفقرة

محمد احمد المشرقي[i]
أ.د أحمد سليمان عودة[ii]

تاريخ الاستلام
2025/10/26
تاريخ القبول
2025/11/26

الملخص

هدفت الدراسة إلى الكشف عن أثر طريقة معالجة التخمين والبيانات المفقودة على الخصائص السيكومترية للفقرات وفقًا لنظرية استجابة الفقرة. تكونت عينة الدراسة من 612 طالب وطالبة من مجتمع الدراسة تم اختيارهم بشكل عشوائي بوحدة المدرسة، وهي بذلك تشكل ما نسبته 5% من مجتمع الدراسة، حيث تم تقسيم عينة الدراسة إلى 3 مجموعات فرعية عشوائية، وذلك بهدف ضمان ضبط ترتيب تطبيق نماذج الاختبار من خلال تطبيق 3 ترتيبات مختلفة على المجموعات الثلاث بعد التطبيق الافتراضي (أسئلة التكميل). ولتحقيق أهداف الدراسة تم اتباع المنهج شبه التجريبي، وبناء اختبار في الفيزياء للصف العاشر مكون من 30 فقرة من نوع اختيار من متعدد بأربعة نماذج، تختلف في تعليمات الاختبار ونوع الفقرات.

أظهرت النتائج أن نسب القيم المفقودة تختلف باختلاف طريقة معالجة التخمين، باستثناء طريقتي المعالجة من خلال المكافأة والعقاب عند مقارنتها بطريقة الاستجابة الجزئية. كما بينت النتائج وجود فروق في الأوساط الحسابية لقيم معلمة الصعوبة باختلاف طريقة معالجة أثر التخمين، إذ كانت تلك الفروق لصالح كل من طريقتي الثواب والعقاب بمقابل طريقة الاستجابة الجزئية، كما بينت النتائج عدم وجود فروق ذات دلالة إحصائية في الأوساط الحسابية لقيم معلمة التمييز باختلاف طريقة معالجة أثر التخمين وفقًا لنظرية الاستجابة للفقرة. وفي ضوء هذه النتائج توصلت الدراسة لعدد من التوصيات كان أهمها: أن يحرص القائمون على بناء الاختبارات الموضوعية، خصوصًا في المواد العلمية كالفيزياء، على اختيار طريقة معالجة التخمين التي تتناسب مع طبيعة الفقرات.

الكلمات المفتاحية: التخمين، القيم المفقودة، الخصائص السيكومترية، نظرية الاستجابة للفقرة.

[i] طالب دكتوراه/جامعة اليرموك/ القياس والتقويم التربوي
[ii] جامعة اليرموك/ القياس والتقويم التربوي

The Effect of Method of Handling Guessing and Missing Data on Item Psychometric Properties According to Item Response Theory

Abstract

This study aimed at investigating the effect of the method of handling guessing and missing data on the psychometric properties of test items within the framework of Item Response Theory (IRT). The sample of the study consisted of (612) male and female students randomly selected from the study population, using the school as the sampling unit, which represented (5%) of the total population. The sample of the study was divided into three random subgroups to control the administration order of the test forms by applying three different sequences to the three groups following the initial administration (completion questions). To achieve the objectives of the study, a quasi-experimental design was adopted, and a physics test for tenth-grade students was developed. The test consisted of (30) multiple-choice items presented in four different test forms that varied in test instructions and item types. The findings of the study showed that the proportion of missing data varied according to the method used to handle guessing, except between the reward and penalty methods when compared to the partial credit method. Furthermore, the findings showed significant differences in the mean values of the item difficulty parameter depending on the guessing treatment method, as these differences were in favor of both the reward and penalty methods versus the partial response method. However, no statistically significant differences were found in the mean values of the item discrimination parameter across the different guessing treatment methods according to IRT. In light of these findings, the study recommended that objective test developers-particularly in scientific subjects such as physics-should carefully select the method for handling guessing that best aligns with the nature of the test items.

Keywords: Guessing, Missing Data, Psychometric Properties, Item Response Theory.

مقدمة:

تُعد الاختبارات من نوع الاختيار من متعدد من أكثر أدوات التقييم استخدامًا في قياس وتقويم نواتج التعلم، نظرًا لما توفره من خصائص تميزها عن باقي أنواع أدوات القياس والتقويم حسب نوع الأسئلة، مثل سهولة التنفيذ، وموضوعية التصحيح، والقدرة على تغطية نطاق واسع من المحتوى التعليمي، ومع ذلك، فإن فعاليتها تعتمد بشكل كبير على جودة تصميمها ودقة تحليل نتائجها وتأثرها بمصادر أخطاء متعددة يمكن أن تقلل من مزاياها كالتخمين ووجود قيم مفقودة فيها.

يشير التخمين إلى استجابة الطلبة للفقرات بشكل عشوائي أو شبه عشوائي دون أن يكون لديهم معرفة تامة بالإجابة الصحيحة، هذا السلوك يُعد من الظواهر الشائعة في اختبارات الاختيار من متعدد، خاصة عندما يواجه الطلاب صعوبة في الإجابة؛ وهو ما يدفعهم إلى التخمين (السعودي والنصرأوين، 2018). وقد أوضحت دراسة ريوس وآخرون (Rios et al., 2022) أن التخمين يؤدي إلى تشويه تقديرات القدرة الحقيقية للطلاب، مما يقلل من دقة الاختبار. كما أشار ست هيلير وآخرون (St. Hilaire et al., 2024) إلى أن التخمين قد يؤدي إلى زيادة غير مبررة في تقديرات القدرة لدى بعض الأفراد، مما يؤثر على دقة القياسات المستخلصة من البيانات، وذلك على الرغم من أن بعض الاختبارات تُصمم لتقليل تأثير التخمين من خلال تصميم الخيارات، أو إضافة عقوبات للإجابات الخاطئة، أو تصحيح الدرجة الكلية بإحدى طرق معالجة التخمين.

من جهة أخرى، تُعد البيانات المفقودة مشكلة شائعة في الاختبارات، حيث يمكن أن تحدث بسبب عدم استجابة الطلاب لفقرات معينة لأسباب متعددة، مثل عدم الفهم أو نقص الوقت أو صعوبة الأسئلة، أو أية عوامل أخرى، إذ أشارت دراسة جيميتشي وآخرون (Gemici et al., 2012) إلى أن البيانات المفقودة تؤثر بشكل كبير على النتائج المستخلصة من التحليل، سواء باستخدام النظرية الكلاسيكية أو نظرية الاستجابة للفقرة؛ فإنّ تعدد طرق معالجتها يمكن أن يحدث فرقًا في دقة نتائج التحليل بأي من النظريتين. كما أشارت دراسة لودتكي وآخرون (Lüdtke et al., 2020) إلى أن استخدام نماذج تقدير معينة، مثل: التقدير باستخدام النماذج البايزية أو تقنيات الإكمال المتعدد، يمكن أن يقلل من تأثير هذه المشكلة مقارنة بطرق أخرى تعرف بالطرق التقليدية.

التخمين في اختبارات الاختيار من متعدد: يشير ست هيلير وآخرون (St. Hilaire et al., 2024) إلى أن تقييم الطلبة ينبغي أن يأخذ بالاعتبار فرصة التخمين، حيث يستخدم الطلبة نمطين لتخمين الاختيار الصحيح في أسئلة الاختيار من متعدد: ففي النمط الأول، لا يكون لدى الطالب أي رأي أو ترجيح في أي من الخيارات، وبالتالي، يختار أحد الخيارات عشوائيًا على أمل أن يكون الإجابة الصحيحة. في هذا النمط، المسمى بالتخمين الأعمى أو العشوائي التام، تكون لجميع الخيارات فرصة متساوية للاختيار، وفي النمط الثاني: لا يكون تخمين الطالب عشوائيًا تمامًا، ولكنه يزيل أو يختار الخيارات بناءً على معرفته الجزئية. في هذا النمط، المسمى بالتخمين المستنير أو التخمين الذكي القائم على المعرفة الجزئية يمكن الطالب من الإجابة الصحيحة إلى حد ما؛ بعبارة أخرى، باستخدام التفكير المنطقي أو الاعتماد على معرفته النسبية بالموضوع، يحاول الطالب تخمين الإجابة الصحيحة عن طريق حذف خيار أو أكثر واختيار الخيار الصحيح بدرجة معينة من اليقين من باقي الخيارات.

طرق ضبط أثر التخمين في اختبارات الاختيار من متعدد:

وفقًا لنظريتي القياس التقليدية والحديثة هناك العديد من الطرق التي يمكن استخدامها لضبط أو تصحيح أثر التخمين في اختبارات الاختيار من متعدد منها:

1. طريقة معادلة التصحيح لأثر التخمين القائمة على العقاب: تعتبر طريقة معادلة التصحيح واحدة من أكثر الطرق شيوعًا لضبط أثر التخمين في اختبارات الاختيار من متعدد، تهدف هذه الطريقة إلى تقليل التأثير الناتج عن الإجابة العشوائية على درجة الطالب، من خلال تعديل الدرجة النهائية بناءً على عدد الإجابات الصحيحة والخاطئة واحتمالية التخمين (Lesage et al., 2013).

2. طريقة الثواب تلك الطريقة التي تشجع الطلبة على عدم التخمين من خلال زيادة علاماتهم الظاهرية بما يتناسب مع درجة تحفظه على الإجابة، أو بمدى ابتعاده عن الإجابة عند عدم ترجيحه لأي بديل، أي من خلال إضافة جزء من العلامة التي يتوقع أن يحصل عليها فيما لو قام بتخمين الإجابة (السعودي والنصرويين، 2019).

3. طريقة الحذف الجزئي (Subset Elimination): تعتمد هذه الطريقة على إعطاء درجات جزئية للطلاب الذين يتمكنون من استبعاد بعض الخيارات الخاطئة، حتى لو لم يصلوا إلى الإجابة الصحيحة النهائية، وبذلك فهي تمتاز بأنها تقلل من أثر التخمين العشوائي ؛ لأنها تشجع الطلاب على التفكير في استبعاد الإجابات الخاطئة، كما تتيح قياسًا جزئيًّا لمرفة مستوى الطالب، حتى لو لم يكن متأكدًا تمامًا من الإجابة الصحيحة (Wu et al., 2019).

6. طريقة الاختيار الجزئي (Subset Selection): تعتمد هذه الطريقة على منح الطالب درجة جزئية بناءً على عدد الإجابات التي يحددها كإجابات محتملة بدلاً من اختيار إجابة واحدة فقط، إذ يُسمح للطالب بتحديد أكثر من إجابة محتملة إذا لم يكن متأكدًا تمامًا من الإجابة الصحيحة، ويتم توزيع الدرجة إذا اختار الطالب أكثر من إجابة، وهي تمتاز بأنها تعمل على توزيع الدرجة على جميع الإجابات المختارة، وتقلل من احتمالية النجاح عن طريق التخمين العشوائي، حيث إن اختيار أكثر من إجابة يقلل من الدرجة المكتسبة، كما تقيس مدى تأكد الطالب من الإجابة، فكلما كان أكثر ثقة، حصل على درجة أعلى (عودة، 2022).

معلمة التخمين (C_i) ودورها في ضبط أثر التخمين: معلمة التخمين (C_i) هي القيمة التي تمثل الحد الأدنى لاحتمال الإجابة الصحيحة بسبب التخمين العشوائي، فإذا كان السؤال يحتوي على 4 خيارات، فمن المتوقع أن يكون حوالي 0.25 (أي فرصة عشوائية بنسبة 25% للإجابة الصحيحة)، في حين أن المعادلة تأخذ في الحسبان أن التخمين لا يعتمد على قدرة المختبَر (θ)، وإنما يحدث كاحتمال مستقل (Lord, 2012).

على ذلك فإن معلمة التخمين في النموذج الثلاثي المعلمة تعزز جودة القياس بتقدير أكثر دقة لقدرات المختبرين، كما تساعد في تحسين تصميم الفقرات وتقييم الاختبارات، وتقلل من تأثير التخمين العشوائي على النتائج، مما يجعلها أكثر موثوقية (Brown & Maydeu-Olivares, 2010).

تأثير معلمة التخمين على منحنى الأداء: يتم تمثيل الأداء على الفقرة باستخدام منحنى خصائص الفقرة (Item Characteristic Curve -ICC)، والذي يظهر العلاقة بين مستوى قدرة الفرد (θ) واحتمال الإجابة الصحيحة، في حين أن معلمة التخمين (C_i) تؤثر على الجزء السفلي من المنحنى، فإذا كانت منخفضة جداً (أي قدرة ضعيفة جداً)، فإن احتمال الإجابة

الصحيحة لا يقترب من الصفر بل يقترب من (C_i)، على سبيل المثال: إذا كانت ($C_{i=0.25}$)، فإن احتمال الإجابة الصحيحة عند أدنى مستوى قدرة سيكون 25% (Erguven, 2013).

القيم المفقودة: تمثل القيم المفقودة تحديًا كبيرًا في الاختبارات والمقاييس التربوية؛ لأنها قد تؤثر على صدق وثبات القياسات وكذلك التأثير السلبي على جودة التحليل والنماذج التنبؤية، بالإضافة إلى تأثيرها على نتائج الطلبة وتفسير البيانات ويمكن أن تحدث هذه القيم لعدة أسباب منها:

عدم استجابة الطالب لبعض الأسئلة ؛ فقد يتجنب بعض الطلاب الإجابة عن أسئلة معينة بسبب عدم الفهم أو عدم الثقة في الإجابة، أو وجود أخطاء إدارية أو تقنية في تسجيل الدرجات، أو مشكلات تقنية في الاختبارات الإلكترونية، كذلك الغش أو عدم الالتزام بالقواعد فقد يتم استبعاد إجابات بعض الطلبة بسبب الاشتباه في الغش، أو غياب الطلبة عن بعض الاختبارات أو عن أجزاء منها ، فقد يتغيب الطالب عن جزء من الاختبار بسبب ظروف شخصية، صحية أو غيرها (Little & Rubin, 2019)، كما أن هناك بعض الأسباب التي تعود إلى تصميم الاختبار ففي بعض الاختبارات التكيفية (Adaptive Testing)، لا يتم تقديم جميع الأسئلة لكل الطلاب، مما قد يؤدي إلى بيانات غير مكتملة عن قصد، من الأسباب الأخرى لفقد البيانات عدم توافق بعض الأسئلة مع الخلفية الثقافية أو الاجتماعية للطلبة، فبعض الأسئلة قد تكون غيرمفهومة أو غير ذات صلة ببعض الفئات، مما يؤدي إلى تركها دون إجابة (Al-zboon et al., 2021).

كما أشار لودلو وأولري (Ludlow & Oleary, 1999) إلى أن القيم المفقودة تظهر. بسبب عدم الوصول إلى بعض الفقرات، أو عدم الإجابة عنها بسبب عدم. الاهتمام بها من قبل المستجيب، أو عدم قدرته على الإجابة عنها. كما أشار ميكنايت وآخرون (Mcknight et al., 2007) إلى ثلاث فئات لأسباب فقد القيم؛ وهي الأسباب التي تعود. إلى المستجيبين أنفسهم، أو إلى تصميم الاختبار، أو إلى التفاعل ما بين. المستجيبين وتصميم الدراسة.

وهكذا تتسبب مشكلة فقدان القيم في إثارة العديد من. المشاكل لعل أهمها وفقًا لما ذكره روث (Roth, 1994) احتمال تقليل البيانات المفقودة لقوة نتائج. الاختبار الإحصائي (Test Power)؛ نتيجة للخلل الحاصل في حجم العينة، والمشكلة الثانية تكمن في. تحيز تقدير معالم المقياس، وخاصة عند تقدير الصدق والثبات، كما يضيف الصرايرة (2018، ص485): "إن مشكلة فقدان استجابات. الأفراد على عدد من فقرات الاختبار له تأثير متوقع على دقة تقدير معالم الفقرات ومعلمة القدرة".

طرق معالجة القيم المفقودة:

أولًا: الطرق التي تقوم على الحذف (Methods Depends on Deletion)؛ وهي أقدم طرق التعامل. مع القيم المفقودة، وتتم قبل البدء بعملية التحليل. إما بحذف الحالة الكاملة، أي حذف أي مستجيب له قيمة مفقودة. لأي فقرة من فقرات الاختبار (Gemici et al., 2012)، وإما بحذف الحالة المتوفرة، أي التخلص من جميع. المستجيبين الذين لديهم قيم مفقودة على كلا المتغيرين، إلا أن هذه الطريقة قد تؤثر. على تمثيل العينة للمجتمع خاصة إذا كان عدد. المستجيبين الذين تم حذفهم كبيرًا (Graham, 2009).

ثانيًا: الطرق القائمة على. احتساب قيمة تعويضية (Methods Depends on Imputation)؛ في هذه الطريقة يتم. الاستعاضة عن القيم المفقودة في البيانات. بقيم معقولة

من خلال البيانات المتوفرة، تجنبًا لهدر أي معلومات. من أفراد عينة الدراسة والحفاظ. على دقة عالية من النتائج (Schafer & Graham, 2002).

الاختبارات: تعرف الاختبارات بأنها مجموعة من المثيرات التي تقدم إلى الأفراد لاستثارة استجابات بهدف الحصول على تقديرات كمية أو إعطاء درجات لهم، قائمة على عينة ممثلة لسلوك الفرد، كما أنها تعتبر بمثابة مؤشر للقدر الذي يمتلكه الفرد من الخاصية التي يقيسها الفرد، وهي ما تسمى بالسمات الكامنة لدى الفرد، إذ إن هذه الدرجات التي يحصل عليها الفرد تعد مؤشرًا للمفهوم الذي نرغب بقياسه من خلال تلك الاختبارات، وهي بذلك تعتبر دالة لموضوعية الاختبارات وصدقها وثباتها (السعودي ونصراويين، 2018).

كما تعرف الاختبارات بأنها: مجموعة من المثيرات التي أعدت لقياس العمليات العقلية والسمات والخصائص النفسية بطريقة كمية أو كيفية، فالمثير هنا قد يكون أسئلة شفوية كما في المقابلات المختلفة، أو كتابية كما في أغلب الاختبارات التي تطبق على الطلبة بمختلف مستوياتهم التعليمية، كما أنها قد تكون سلسلة من الأعداد أو الأشكال الهندسية أو النغمات الموسيقية أو الصور والرسومات، إذ إن كافة هذه الأشكال تُعد بمثابة مثيرات تؤثر على الفرد وتستثير استجابته (Brennan, 2023).

ويعرف نور الدين (2021) الاختبارات بأنها: مجموعة من الأسئلة أو الفقرات لكل واحد منها إجابة صحيحة واحدة، مثل: اختبارات التحصيل الموضوعية، أو اختبارات القدرات العقلية واختبارات الذكاء المختلفة وغير ذلك من الاختبارات التي تقيس الحقائق. في حين تعرف الاختبارات بأنها عبارة عن إجراء منظم لقياس عينة من السلوكيات من خلال عينة من المثيرات، أي أنه أداة للحصول على معلومات حول سلوك الفرد، ونقول عينة لأنه لا يمكن الحصول على جميع سلوكيات الفرد بالتالي لا يمكن إخضاعها لاختبار واحد فقط.

الاختبارات التحصيلية: هي تلك الاختبارات التي تقيس درجة التعلم والنجاح والإنجاز في أحد المواضيع الدراسية أو التعليمية المختلفة، فهي مُصمَّمة وموضوعة خصيصًا لقياس التعلم المدرسي أو النظامي والرسمي بالدرجة الأولى، إذ تُعرَّف الاختبارات التحصيلية بأنها: عبارة عن استثارة مقصودة تتطلب استجابة محددة، إذ تكون هذه الاستثارة من خلال مثير واحد أو عدة مثيرات على شكل أسئلة أو تساؤلات مجردة أو محسومة، لذلك فإن جودة الاختبارات التحصيلية تتوقف على مدى صلاحيتها في استدعاء الاستجابات الصحيحة (Burton, 2004).

وتعرف الاختبارات التحصيلية بأنها: عبارة عن اختبار صُمِّمَ بقصد قياس مدى إلمام الطلبة في معرفة محددة أو براعته في مهارة معينة، أي أنها مصممة بقصد قياس التحصيل الدراسي للطلبة. ويطلق عليها في بعض الأحيان اختبارات الورقة والقلم وتعد من أهم وسائل تقويم تحصيل الطلبة، فهي تعبر عن مستوى النجاح الذي يحققه الطالب في أحد المجالات المدرسية المقصودة، وتعني مدى اكتساب الطالب للمهارات والمعارف والقدرات المختلفة (Bush, 2015)، أما عودة (2022) فيعرفها بأنها: طريقة منظمة لتحديد مستوى تحصيل الطالب لمعلومات ومهارات في مادة دراسية معينة سبق وأن تعلمها بصورة رسمية وذلك من خلال إجاباته على عينة من الأسئلة التي تمثل محتوى تلك المادة الدراسية.

عرف جلفاردي يزدي وآخرون (Golvardi Yazdi et al., 2021) اختبار الاختيار من متعدد (Multiple choice) أو الاستجابة الموضوعية (Objective Response)، بأنه: شكل من أشكال التقييم الموضوعي حيث يُطلب من المستجيبين اختيار الإجابات الصحيحة فقط

من الخيارات المعروضة كقائمة أو شكل فقرات مرقمة. يتم استخدام اختبارات الاختيار من متعدد في أغلب الأحيان في الاختبارات التعليمية والتربوية المختلفة، وهي تستخدم على نطاق واسع لقياس نتائج المعرفة، والفهم، والتطبيق (Kubiszyn & Borich, 2024). إلا أن أحد العيوب الرئيسية لاختبارات الاختيار من متعدد هو إمكانية العثور على الإجابة الصحيحة من خلال التخمين، إذ يمكن للطالب الذي لا يستطيع الإجابة على سؤال معين ببساطة اختيار إجابة عشوائية، ولا يزال لديه فرصة للحصول على درجة ذلك السؤال، إذا تم اختيار إجابة سؤال مكون من أربعة خيارات من خلال التخمين العشوائي، فهناك عادةً فرصة بنسبة (25%) للحصول على الإجابة الصحيحة (Brennan, 2023). إذ يعتبر التخمين من الممارسات الشائعة التي يقوم بها الطلبة في الدقائق الأخيرة من الاختبار بإعطاء جميع الأسئلة المتبقية إجابات عشوائية على أمل أن يجيبوا على بعضها بشكل صحيح على الأقل (Vuorre & Metcalfe, 2022).

الدراسات السابقة:

بعد البحث والتقصي تبين لدى الباحثين وجود العديد من الدراسات السابقة ذات الصلة بطرق معالجة أثر التخمين، وفيما يلي عرض للدراسات العربية والأجنبية ذات الصلة بموضوع الدراسة:

هدفت دراسة عودة (2022) إلى معالجة أثر التخمين في اختبارات الاختيار من متعدد، عبر نموذج تدريج ثنائي الاتجاه، يربط بين المعرفة الجزئية والمخاطرة. ولتحقيق أهداف الدراسة تم اتباع المنهج التجريبي من خلال تطبيق 50 فقرة اختبارية في مساق للبحث التربوي بجامعة اليرموك على عينة مكونة من 64 طالبًا من طلبة الماجستير فيها، تم تحديد تعليمات الإجابة للطلبة فيها من خلال طريقتين هما: الطريقة التقليدية، وطريقة الاختيار الجزئي، وقد أظهرت النتائج أن إدراج المخاطرة كمصدر تباين يمكن أن يؤثر سلبًا على صدق النتائج؛ إذ إنه يضيف ثباتا زائفا على حساب الصدق، كما توصلت الدراسة إلى تقديم معادلة لتصحيح العلامات وفقًا لمتغير المخاطرة.

كما هدفت دراسة ريوس وآخرين (Rios et al., 2022) إلى معرفة أثر التخمين على خصائص الاختبار والفقرات وقدرات المفحوصين، ولتحقيق ذلك تم اتباع منهج تحليل المحتوى من خلال أسلوب التحليل البعدي (Meta-Analysis) لـ 25 دراسة و39 عينة مستقلة تتألف من 443264 مفحوصًا فريدًا، في الفترة الزمنية التي سبقت الرابع من أيلول (2020)، وقد بينت نتائج الدراسة أن اتباع أسلوب التخمين من قبل المفحوصين يعمل على تشويه خصائص الاختبار، وخصائص الفقرات وقدرات المفحوصين بانحراف معياري مقداره (0.13).

هدفت دراسة نور الدين (2021) إلى التعرف على أثر إضافة بديل (لا أعرف) في أسئلة الاختيار من متعدد على تقدير قدرة الفرد وخفض معلمة التخمين وتحسين كفاءة الاختبار في ضوء نظرية الاستجابة للمفردة، ولتحقيق ذلك اتبع الباحث المنهج شبه التجريبي من خلال تطبيق اختبار تحصيلي مكون من (20) فقرة اختيار من متعدد في مقرر (علم النفس الشخصية) لطلاب الدراسات الإسلامية في جامعة الجوف على مجموعتي الدراسة (التجريبية والضابطة) المكونتين من (97) طالبا من الذكور، إذ كان عدد بدائل الاختبار لأفراد المجموعة الضابطة أربعة بدائل، وقد تمت إضافة بديل خامس (لا أعرف) لأفراد المجموعة التجريبية يمنح لمن يختاره (¼) درجة، تضاف إلى إجابته الصحيحة للتَحقُّق من خفض أثر التخمين،

طبق على كل منهم اختبار تحصيلي يتألف من (20) سؤالاً من نمط الاختيار من متعدد، ذي أربعة بدائل، وقد تمت إضافة بديل خامس (لا أعرف) لأفراد المجموعة التجريبية يمنح لمن يختاره (¼) درجة، تضاف إلى إجابته الصحيحة للتحقق من خفض أثر التخمين. وقد أظهرت النتائج أن إضافة بديل (لا أعرف) قد خفض من معلمة التخمين لمفردات الاختبار، كما بينت النتائج أيضا أن إضافة بديل (لا أعرف) للمجموعة التجريبية واستخدامه في الاختبار قد حسّن من معلمة التمييز لفقرات الاختبار، كما كشفت النتائج أيضا أن المجموعة التجريبية كانت أقل بصورة دالة من المجموعة الضابطة في متوسط الخطأ المعياري لتقدير القدرة.

أما دراسة القرني وأبو هاشم (2021) فقد هدفت إلى التعرف على أثر طريقة معالجة القيم المفقودة على دقة معادلة الاختبارات في نظرية الاستجابة للفقرة، وتكونت عينة الدراسة من 2372 من تلاميذ الصفوف العليا بالمرحلة الابتدائية، وكذلك من 3276 من طلاب المرحلة المتوسطة بمدينة جدة. ولتحقيق هدف الدراسة تم استخدام اختبار أوتيس لينون للقدرات بصورتيه: (الابتدائي الثنائي والمتوسط) وتتكون كل صورة من 80 فقرة اختبارية تم تطبيقها على عينة استطلاعية بحجم 208 طلبا، وتم اختيار 10 فقرات من كل صورة بواقع 20 فقرة تمثل الجذع المشترك. ومن ثم طبقت الأدوات على العينة وتم الحصول على بيانات كانت نسبة القيم المفقودة 5.5% للمرحلة الابتدائية و(5.6%) للمرحلة المتوسطة. وتمت معالجة البيانات للقيم المفقودة من خلال ثلاث طرق للمعالجة وهي: خوارزمية تعظيم التوقعات (EM)، وحساب قيم تعويضية متعددة (MI)، وحساب المعالجة بطريقة دالة الاستجابة للفقرة (IRFI). وتم استخدام برنامج (X-calibre) في التقدير، وكذلك للتحقق من مطابقة البيانات للنموذج. وللكشف عن دقة المعادلة العمودية في نظرية الاستجابة للفقرة باختلاف طريقة المعالجة للقيم المفقودة وطريقة التقدير تم استخدام محك الخطأ المعياري. وقد أظهرت النتائج أن أفضل طريقة معالجة للقيم المفقود كانت طريقة المعالجة بدالة الاستجابة للفقرة (IRFI) وذلك عندما نستخدم الطريقة الافتراضية للتقدير في البرنامج وهي طريقة (EAP).

وأما دراسة الزبون وآخرين (Al-zboon et al., 2021) فقد هدفت إلى معرفة تأثير نسبة القيم المفقودة على تقدير الخطأ المعياري لمعالم الفقرات ودالة معلومات الاختبار وفقًا للنموذج الثلاثي المعلمة، ولتحقيق أهداف الدراسة تم توليد 1000 إجابة لمستجيبين افتراضيين على اختبار مكون من 50 فقرة ثنائية التدريج، وفقًا للنموذج اللوجستي ثلاثي المعلمة (PL3) باستخدام برنامج (WINGEN3). وباستخدام برنامج (EXCEL) تم الحصول على البيانات التي تحتوي على قيم مفقودة بنسب 0٪، 5٪، 10٪، 15٪. وقد أظهرت نتائج الدراسة وجود فروق ذات دلالة إحصائية بين متوسطات الأخطاء المعيارية لتقدير معلمتي الصعوبة والتمييز للاختبار وفقا للنموذج ثلاثي المعلمة باختلاف نسب القيم المفقودة البالغة 0٪، 5٪، 10٪، 15٪ ولصالح نسب القيم المفقودة الأقل، مما يعني أنه كلما قلت نسبة القيم المفقودة، من المتوقع أن تنخفض قيمة الخطأ المعياري في تقدير الصعوبة ومعامل التمييز للفقرة، وأن أفضل نسبة بيانات مفقودة كانت 5٪، كما أشارت النتائج إلى عدم جود فروق ذات دلالة إحصائية بين متوسطات الأخطاء المعيارية لتقدير معلمة للاختبار وفقا للنموذج ثلاثي المعلمة باختلاف نسب القيم المفقودة البالغة 0٪، 5٪، 10٪، 15٪، كما أشارات النتائج إلى أن دالة المعلومات للاختبار تزداد مع انخفاض نسبة البيانات المفقودة.

وفي دراسة للصرايرة (2018) هدفت لتقصي أثر نسبة القيم المفقودة، وطريقة معالجتها في دقة تقدير القيمة القصوى لدالة المعلومات للفقرة، وقيمة مؤشر ثبات الفقرة تحت افتراضات

النموذج اللوجستي ثلاثي المعلمة، ولفحص فرضيات الدراسة تم توليد استجابات تبعًا للنموذج ثلاثي المعلمة، وبحجم عينة بلغ 1000 مستجيب على اختبار مكون من 50 فقرة ثنائية الاستجابة. وأظهرت نتائج تحليل القياسات المتكررة وجود فروق دالة إحصائيًا في دقة تقدير القيمة القصوى لدالة المعلومات للفقرة تعزى لطريقة المعالجة وبأفضلية طريقة المعالجة المتعدد، ولنسبة القيم المفقودة 5% كما أظهرت وجود فروق دالة إحصائيًا في دقة تقدير القيمة القصوى لدالة المعلومات للفقرة تعزى للتفاعل بين نسبة القيم المفقودة وطريقة معالجتها وكانت لصالح نسبة الفقد 5% بأفضلية لطريقة المعالجة المتعدد. وأيضًا كشفت نتائج الدراسة عن وجود فروق دالة إحصائيًا في قيمة مؤشر ثبات الفقرة تعزى لطريقة المعالجة وبأفضلية طريقة المعالجة المتعدد، ولنسبة القيم المفقودة 5%، كما أظهرت النتائج عدم وجود فروق دالة إحصائيًا في قيمة مؤشر ثبات الفقرة تعزى للتفاعل بين نسبة القيم المفقودة وطريقة معالجتها.

وهدفت دراسة السعودي والنصراوين (2018) إلى مقارنة ثلاث طرق لضبط أثر التخمين في اختبارات الاختيار من متعدد، هي: طريقة التصحيح لأثر التخمين التقليدية، طريقة تصحيح المموهات، طريقة استخدام البديل (لا أعرف الإجابة)، من خلال أثرها على خصائص الاختبار، باستخدام اختبار رياضيات تحصيلي على عينة مكونة من (159) طالباً وطالبة من طلبة الثانوية العامة في العاصمة عمان، وذلك باتباع المنهج الوصفي التحليلي، وقد أشارت النتائج إلى وجود فروق دالة بين طريقة تصحيح المموهات وطريقة التصحيح لأثر التخمين لصالح طريقة تصحيح المموهات على وفق محكّات، هي: الأداء على الاختبار، الصدق، والثبات؛ نظرا 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.

كما هدفت دراسة النصراوين والبطش (2018) إلى الوقوف على فاعلية أربعة نماذج لمعالجة التخمين في الاختبارات الموضوعية /الاختيار من متعدد ضمن إطار النموذج اللوجستي ثلاثي المعلمة وأثرها على دقة تقدير معلمة القدرة باختلاف متغير طول الاختبار (25 فقرة و 50 فقرة) وحجم عينة المفحوصين (500 مفحوص و 1500 مفحوص)، والنماذج المستخدمة من نظرية الاستجابة للفقرة في هذه الدراسة هي: النموذج اللوجستي الثلاثي المعلمة، والنموذج الثنائي المعدل بإضافة قيمة ثابتة (0.20)، ونموذج عتبة التخمين، والنموذج القائم على الصعوبة، ولتحقيق غرض الدراسة تم استخدام المنهج الوصفي التجريبي من خلال توليد البيانات التي تمثل استجابات المفحوصين باستخدام أسلوب المحاكاة بحيث تكون مطابقة لافتراضات ومتطلبات النموذج الثلاثي المعلمة باستخدام برنامج (wingen3)، وقد أظهرت النتائج أن الفروق بين المتوسطات الحسابية للجذر التربيعي لمتوسط مربع الخطأ في تقدير معلمة القدرة كانت دالة إحصائيًا في حالة استخدام النموذج اللوجستي الثلاثي المعلمة مع كل نموذج من النماذج الأخرى، كما أظهرت نتائج الدراسة أن قيم الجذر التربيعي لمتوسط مربعات الخطأ، ومؤشر التحيز جميعها تقل بزيادة حجم العينة من (500) مفحوص إلى (1500) مفحوص، وبزيادة طول الاختبار من (20) فقرة إلى (50) فقرة.

وهدفت دراسة الخرشة (2016) إلى التحقق من أثر طرق معالجة أثر التخمين في تقدير إحصائيات الأفراد والفقرات في اختبارات الاختيار من متعدد. ولتحقيق الغرض من الدراسة تم استخدام المنهج الوصفي شبه التجريبي، من خلال بناء اختبار تحصيلي من نوع الاختيار من متعدد في مادة مبادئ القياس والتقويم التي تدرس لطلاب كلية التربية بجامعة مؤتة مكون من 30 فقرة، حيث تكونت عينة الدراسة من جميع الطلاب الذين درسوا مادة مبادئ في القياس

والتقويم بكلية التربية في جامعة مؤتة، وقد بلغ عددهم (276) طالبا وطالبة موزعين على خمس شعب. تم تطوير اختبار تحصيلي يشتمل على 30 فقرة من نوع الاختيار من متعدد ذات أربع بدائل لقياس المصطلحات والمفاهيم الأساسية في مقرر "مبادئ القياس والتقويم" الذي يدرس لطلاب كلية التربية بجامعة مؤتة. وقد كشفت النتائج بأنه يوجد فروق ذات دلالة إحصائية بين متوسطات معالم الصعوبة للفقرات تعزى لطرق معالجة التخمين. وقد جاءت تلك الفروق لصالح طريقة المعالجة التصحيح لأثر التخمين (العقاب)، وكذلك أظهرت النتائج أيضا بأنه لا يوجد فروق ذات دلالة إحصائية بين متوسطات معالم التمييز للفقرات ومعلمة القدرة للأفراد تعزى لطرق معالجة التخمين.

وهدفت دراسة شينيل وآخرون (Şenel et al., 2015) إلى معرفة أثر طرق معالجة التصحيح لأثر التخمين على خصائص الاختبار ومعالم الفقرات، ولتحقيق ذلك تم اتباع المنهج الوصفي الارتباطي، من خلال تطبيق أداة الدراسة والتي تمثلت باختبار للغة التركية مكون من (15) فقرة على عينة مقدارها (123) طالب وطالبة من طلبة المرحلة الثانوية في المدارس التركية، وقد بينت النتائج وجود فروق ذات دلالة إحصائية في صعوبة الفقرات وتميزها تعزى لطريقة التصحيح من أثر التخمين لصالح طريقة التصحيح (العقاب)، كما بينت النتائج أن قيم معامل الثبات قد ارتفعت عند استخدام طريقة التصحيح من أثر التخمين (العقاب).

وهدفت دراسة عبد الوهاب (2013) إلى المقارنة بين النظرية الكلاسيكية والنموذج ثلاثي المعلمة في معالجة أثر التخمين عند تقدير إحصائيات الفقرة (معلمتي الصعوبة والتمييز) وإحصائيات الأفراد –المتمثلة في القدرة (θ)، ولتحقيق ذلك تم اتباع المنهج الوصفي الارتباطي من خلال تطبيق اختبار تحصيلي من نوع الاختيار من متعدد في مادة القياس النفسي على عينة مكونة من (1010) طالب وطالبة من طلاب الفرقة الرابعة والدبلوم العام بكلية التربية بالمينا، وقد بينت النتائج عدم وجود اختلاف في تقدير إحصائيات الأفراد والفقرات في ضوء النظرية الكلاسيكية قبل وبعد تصحيح أثر التخمين، ولكن وجد اختلاف واضح في هذه الإحصائيات في ضوء النموذج الثلاثي المعلمة عن النظرية الكلاسيكية قبل تصحيح أثر التخمين، مما يشير إلى اختلاف كل منهما في معالجة أثر التخمين عند تقدير هذه الإحصائيات.

التعقيب على الدراسات السابقة:

تتفق كافة الدراسات السابقة على أن التخمين يؤدي إلى العديد من الآثار السلبية على الأفراد والاختبارات، فهو يؤثر على ثبات الاختبارات ومعالم فقراتها، وهذا ما أكدته دراسة ريوس وآخرون (Rio et al., 2022). إلا أن الدراسات السابقة قد تباينت في نتائجها حول أفضل طرق معالجة أثر التخمين، فقد أظهرت دراسة نور الدين (2021) أن إضافة بديل (لا أعرف) يخفض من معلمة التخمين لمفردات الاختبار، ويحسن من معلمة التمييز لفقرات الاختبار. على النقيض من ذلك أظهرت دراسة السعودي والنصراوين (2018) وجود فروق دالة بين طريقة تصحيح المموهات وطريقة التصحيح لأثر التخمين لصالح طريقة تصحيح المموهات على وفق عدة محكات محكّات.

ما يميز الدراسة الحالية عن باقي الدراسات:

تتميز الدراسة الحالية عن باقي الدراسات السابقة في أنها ستعمل على معالجة أثر التخمين والقيم المفقودة في اختبارات الاختيار من متعدد، من خلال النظرية الحديثة التي تستند إلى

الأساليب والإحصائيات الأكثر تقدمًا مثل نماذج استجابة الفقرة، هذا التمازج بين النظريتين قد يفتح أبوابًا جديدة لفهم الفرق بين معالجة التخمين والبيانات المفقودة من منظورين مختلفين.
كما تتميز هذه الدراسة في أنها ستطبق في مجال الفيزياء وهي إحدى المواد الدراسية التي تتطلب اختبارات دقيقة يمكن أن تتأثر بالأساليب المستخدمة في تحليل البيانات، لذا من المهم معرفة كيفية تأثير طرق معالجة البيانات سواء أكانت مفقودة أم تم تخمينها على دقة الاختبارات في هذا المجال. كما أن هذه الدراسة تجمع بين جوانب نظرية وإحصائية مهمة، وتقدم تحليلاً معمقًا قد لا يكون متاحًا بشكل شامل في الدراسات الأخرى المنشورة في نفس المجال.

مشكلة الدراسة وأسئلتها:
تواجه اختبارات الاختيار من متعدد العديد من التحديات التي تؤثر على دقة وصدق نتائجها، ومن أبرز هذه التحديات: التخمين، فالتخمين يمكن أن يؤدي إلى تضخيم تقديرات قدرات الطلبة دون أن يعكس مستواهم الحقيقي، ويسهم كذلك في زيادة نسبة البيانات المفقودة عند وجود تعليمات مختلفة للاختبار كاستخدام أسلوب التصحيح المعتمد على المكافأة، التي بدورها قد تشوه النتائج إذا لم تُعالج بطريقة صحيحة. ومع تطور القياس النفسي، ظهرت النظريات الحديثة التي طرحت بعض الحلول لمعالجة بعض المشكلات التي واجهة النظرية الكلاسيكية. ومع ذلك، لا تزال هناك فجوة معرفية حول تأثير معالجة التخمين والبيانات المفقودة على خصائص اختبارات الاختيار من متعدد لمادة الفيزياء.
وفي سياق معالجة أثر التخمين، تباينت نتائج الدراسات حول أفضلية النظرية الكلاسيكية للقياس ونظرية الاستجابة للفقرة، فقد أشارت بعض الدراسات كدراسة عبد الوهاب (2013) إلى أن النظرية الكلاسيكية تقدم إحصائيات عملية مهمة، كما أن نموذج الاستجابة للفقرة يقترح نماذج من شأنها معالجة أثر التخمين. كما أظهرت دراسات أخرى كدراسة النصراوين والبطش (2018) أن مقدارًا بسيطًا من التخمين يمكن أن يجعل تقديرات كل من معالم الفقرة ومعلمة القدرة للمفحوصين متحيزة إلى درجة كبيرة، في حين أظهرت دراسة فراري (Frary, 1989) أن استخدام معادلات التصحيح من أثر التخمين في النظرية الكلاسيكية من شأنه أن يزيد من درجة ثبات وصدق الاختبار، مما يشير إلى تحديات في كلا النظريتين.

من هنا، تتحدد مشكلة الدراسة في السؤال الرئيس التالي:
ما أثر طريقة معالجة التخمين والبيانات المفقودة على الخصائص السيكومترية للفقرات وفقًا لنظرية الاستجابة للفقرة؟
ينبثق عنه الأسئلة الفرعية الآتية:
- هل يوجد أثر لطريقة معالجة أثر التخمين على القيم المفقودة في استجابات الطلبة لاختبار اختيار من متعدد في مادة الفيزياء لطلبة الصف العاشر الأساسي؟
- هل يوجد أثر لطريقة معالجة أثر التخمين والقيم المفقودة على خصائص فقرات اختبار اختيار من متعدد في مادة الفيزياء لطلبة الصف العاشر الأساسي وفقًا لنظرية الاستجابة للفقرة؟

أهداف الدراسة:

1-التعرف إلى مدى استجابة الطلبة لتعليمات اختبارات الاختيار من متعدد المتعلقة بمعالجة أثر التخمين.

2-التعرف على أثر طريقة معالجة التخمين والقيم المفقودة على خصائص فقرات الاختبار في اختبار اختيار من متعدد لمادة الفيزياء تبعًا لنظرية الاستجابة للفقرة والنظرية الكلاسيكية.

أهمية الدراسة:

الأهمية النظرية: نظرًا لما سبق، تظهر الحاجة إلى دراسة أثر طريقة معالجة التخمين والبيانات المفقودة في اختبارات الاختيار من متعدد باستخدام النظرية الحديثة والكلاسيكية، فهذه الدراسة تهدف إلى تقديم فهم متعمق لتأثير كل من التخمين والبيانات المفقودة على خصائص فقرات الاختبار، مما يسهم في تحسين تصميم الاختبارات ورفع موثوقيتها وضمان عدالتها واستقرارها، إضافة إلى توفير إطار عملي لدعم القرارات التعليمية في هذا المجال. كما تبرز أهمية هذه الدراسة بشكل خاص في مجال تعليم مادة الفيزياء، حيث تعتمد اختبارات هذا المجال على قياس قدرات معرفية مختلفة تتطلب أدوات دقيقة وفعالة. كما تبرز الأهمية النظرية والتطبيقية للدراسة فيما يأتي:

الأهمية التطبيقية: قد توفر الدراسة توصيات عملية لمصممي الاختبارات حول كيفية تقليل تأثير التخمين ومعالجة البيانات المفقودة. كما قد تساعد نتائج الدراسة الجهات التعليمية في تحسين جودة الاختبارات المعيارية والمحققة للعدالة. وقد تقدم إرشادات للمعلمين والمعنيين بتطوير الاختبارات لتطبيق أفضل الممارسات في قياس قدرات الطلبة.

التعريفات الاصطلاحية والإجرائية:

طريقة معالجة أثر التخمين: تشير طريقة معالجة أثر التخمين إلى الأساليب الإحصائية والتربوية المستخدمة لتقليل أو تصحيح تأثير الإجابات العشوائية أو التخمينية في الاختبارات، خاصة في الاختبارات الموضوعية مثل: الاختيار من متعدد، إذ تهدف هذه الطرق إلى تحسين دقة تقدير مستوى معرفة المفحوص الحقيقي، بحيث لا تؤدي التخمينات العشوائية إلى تضخيم الدرجات أو تحريف النتائج (عبد الوهاب، 2013).

نظرية الاستجابة للفقرة (IRT): هي إحدى النظريات في مجال القياس والتقويم التي تفترض أن هناك سمات كامنة معينة يشترك فيها مجموعة من الأفراد، ولكنهم يختلفون في مقدارها، وبالرغم من أن هذه السمات غير ملاحظة، إلا أنه يمكن الاستدلال عليها من السلوك الملاحظ للفرد المتمثل في استجاباته على فقرات الاختبار، حيث تعتمد هذه النظرية على نماذج رياضية تستخدم في عمليات التقدير لكل من المعالم التي تصف الفقرة والقدرات الكامنة التي تصف الأفراد (علام، 2005).

القيم المفقودة (Missing value): قيم ناتجة عن عدم استجابة المفحوص عن بعض فقرات الاختبار، وترك هذه الفقرات دون إجابة، بصرف النظر عن سبب ذلك (Graham, 2009).

النموذج اللوجستي ثنائي المعلمة (Tow-parameter Logistic model): هو أحد نماذج نظرية الاستجابة للفقرة ثنائية التدريج، حيث يمكن لهذا النموذج تقدير معلمتي الصعوبة والتمييز (Swaminathan & Gifford, 1985).

معلمة الصعوبة (Threshold): مستوى القدرة الذي يقابل احتمال (0.5) للإجابة على الفقرة إجابة صحيحة عندما يكون معامل التخمين مسأويا للصفر (Nason, 195).
معلمة التمييز (Slope): ميل منحنى خصائص الفقرة، الذي يقابل النقطة التي تكون فيها علامة القدرة تسأوي صعوبة الفقرة (Salunkhe et al., 2017).
معلمة التخمين (Asymptote): هي خط المقاربة الأدنى (Lower Asymptote) من منحنى خصائص الفقرة، ويمثل احتمالية إجابة المفحوصين ذوي القدرة المتدنية على الفقرة إجابة صحيحة (Barton et al., 1981).

محددات الدراسة:
1. تقتصر الدراسة الحالية على استخدام النموذج الثنائي المعلمة في تقدير خصائص الفقرة عند معالجة أثر التخمين باستخدام الطريقة التقليدية (العقاب)، طريقة الثواب.
2. استخدام النموذج الجزئي (PCM) في تقدير خصائص الفقرة وخصائص الطلبة عند معالجة أثر التخمين باستخدام طريقة الاختيار الجزئي.
3. تقتصر الدراسة على عينة من طلبة الصف العاشر الأساسي في مديرية تربية قصبة إربد.
4. تقتصر الدراسة الحالية على استخدام طريقة الوسط الحسابي لمعالجة القيم المفقودة.

مجتمع الدراسة: تكون مجتمع الدراسة من جميع طلبة الصف العاشر الأساسي في المدارس الحكومية الأردنية التابعة لمديرية التربية والتعليم لقصبة إربد للعام الدراسي (2025/2024)، البالغ عددهم (13078) طالبة وطالبة.

عينة الدراسة: تكونت عينة الدراسة من 654 طالب وطالبة من مجتمع الدراسة تم اختيارهم بشكل عشوائي بوحدة المدرسة، وهي بذلك تشكل ما نسبته 5% من مجتمع الدراسة، حيث تم تقسيم عينة الدراسة إلى 3 مجموعات فرعية عشوائية؛ وذلك بهدف ضمان ضبط ترتيب تطبيق نماذج الاختبار من خلال تطبيق 3 ترتيبات مختلفة على المجموعات الثلاث بعد التطبيق الافتراضي (أسئلة التكميل).

المنهجية وتصميم الدراسة:
تم اتباع المنهج شبه التجريبي، حيث تم البحث في اختلاف خصائص فقرات الاختبار وخصائص الطلبة باختلاف طريقة معالجة أثر التخمين وتعويض القيم المفقودة من خلال ظروف تجريبية تتمثل في تعليمات الاختبار، وقد تم إجراء ذلك من خلال تصميم قياسات متكررة لكل من خصائص فقرات الاختبار باعتبار نموذج معالجة أثر التخمين هو متغير التكرار (العقاب، الثواب، الاختيار الجزئي) في ضوء نظرية الاستجابة للفقرة.

أداة الدراسة:
تم تحليل محتوى مبحث الفيزياء للصف العاشر الأساسي في المنهاج الأردني 2025/2024 ثم تم بناء اختبار في الفيزياء للصف العاشر الأساسي مكون من 30 فقرة من نوع اختيار من متعدد بأربعة نماذج، تختلف في تعليمات الاختبار ونوع الفقرات حيث تكون النموذج الأول من فقرات من نوع تكميل ولا يتضمن أي تعليمات، أما النموذج الثاني فيتكون من فقرات من نوع

الاختيار من متعدد، والتعليمات تحث وتشجع الطلبة على عدم الإجابة عن السؤال عند عدم معرفه الإجابة الصحيحة بالضبط وتجنب التخمين، وتتضمن تعليمات الاختبار توضيح لأثر التخمين السلبي على الدرجة الكلية التي ستنخفض بسبب التخمين وسمي هذا النموذج (نموذج العقاب). أما النموذج الثالث فإنه يتكون من فقرات من نوع الاختيار من متعدد، يحث ويشجع الطلبة على عدم الإجابة عن السؤال عند عدم معرفة الإجابة الصحيحة بالضبط وتجنب التخمين أي: (تركها فارغة)، وعدم الإجابة عنها للحصول على جزء من العلامة كمكافأة على عدم التخمين، وسمي هذا النموذج (نموذج المكافأة). أما النموذج الرابع فتضمن تعليمات مسبقة تحث وتشجع الطلبة على الإجابة عن السؤال من خلال اختيار مجموعة من البدائل التي يتيقن من احتوائها على البديل الصحيح بهدف الحصول على جزء من درجة السؤال، إذ تتضمن تعليمات الاختبار توضيحًا لأثر التخمين من خلال بديل واحد يقلل من فرصة الحصول على جزء من درجة السؤال.

تم التحقق من بناء الاختبار من خلال كتابة مجموعة من الفقرات عددها 45 فقرة، تشكّل ما مقداره مرّة ونصف من عدد الفقرات المضمّنة في الصورة النهائية للاختبار، إذ تم عرض مجموعة الفقرات الأولية على مجموعة من الخبراء في الفيزياء من معلمين ومشرفين تربويين ومعلمي لغة عربية، حيث تم التحقق من مدى صلاحية هذه الفقرات لقياس تحصيل طلبة الصف العاشر الأساسي في الفيزياء من حيث ارتباطها بالمحتوى، وقدرتها على تمييز الطلبة، وصلاحية البدائل، وفعالية المموهات، وسلامة الصياغة اللغوية والإملائية، كما تم الاسترشاد بآراء مختصي القياس والتقويم في تعليمات الاختبار المتعلقة بالتخمين، وبعد الأخذ بآراء المحكمين خرج الاختبار بالصورة المحكمة بطول 33 فقرة بعد حذف 12 فقرة.

تم تطبيق الاختبار بالصورة المحكمة على عينة حجمها 50 طالب من مجتمع الدراسة من خارج العينة الرئيسية للتحقق من ثبات الاختبار حيث بلغ معامل كرونباخ ألفا (0.84) وهي قيمة مقبولة لأغراض الدراسة، كما تم فحص صلاحية فقرات الاختبار من خلال حساب معاملات التمييز كما هو في جدول 1.

جدول 1: معاملات التمييز لفقرات الاختبار.

الفقرة	معامل التمييز	الفقرة	معامل التمييز	الفقرة	معامل التمييز
i1	0.369	i12	0.252	i23	0.447
i2	0.277	i13	0.388	i24	0.38
i3	0.199	i14	0.477	i25	0.366
i4	0.292	i15	0.347	i26	0.301
i5	0.408	i16	0.43	i27	0.358
i6	0.341	i17	0.347	i28	0.35
i7	0.174	i18	0.187	i29	0.269
i8	0.238	i19	0.341	i30	0.327
i9	0.44	i20	0.331	i31	0.205
i10	0.563	i21	0.355	i32	0.298
i11	0.371	i22	0.404	i33	0.37

من خلال معاملات التمييز يتضح أن معظم فقرات الاختبار تتمتع بثبات مقبول حيث تجاوزت قيمة التمييز الحد 0.2 وفقًا لميسيك (Messick, 1984)، مما يشير لصلاحيتها باستثناء الفقرات (3، 7، 18) التي كانت قيم معاملات تمييز كل منها أقل من 0.2 مما استدعى حذفها، ليصبح المقياس بالصورة النهائية بطول 30 فقرة لتطبيقه على العينة الرئيسية.

المعالجات الإحصائية:

للإجابة عن سؤالي الدراسة والتحقق من الخصائص السيكومترية للمقياسيين تم استخدام الإحصاءات التالية:

- تم حساب نسب القيم المفقودة في كل من نموذج العقاب ونموذج المكافأة ونموذج الإجابة الجزئية ومقارنتها بنسبة القيم المفقودة في النموذج الافتراضي من خلال اختبار اختلاف النسب باستخدام برمجية SPSS. ولفحص اختلاف نسب القيم المفقودة تم استخدام اختبار مربع كاي
- تم تقدير معالم الفقرة للاستجابات على كل من نماذج الاختبار باختلاف تعليماتها (الافتراضي (أسئلة التكميل)، العقاب، المكافأة) وفق النموذج الثنائي المعلمة باستخدام برمجية BILOG-MG3، بينما تم تحليل الاستجابات على نموذج الاستجابة الجزئي (الاختيار الجزئي) وفق النموذج المتعدد الجزئي المعمم GPCM باستخدام برمجية R. ثم تم بحث أثر طريقة معالجة القيم المفقودة على كل من معلمة الصعوبة والتمييز من خلال إجراء اختبار تحليل القياسات المتكررة على كل منها بالاعتماد على النتائج الواردة في البند السابق وباستخدام برمجية SPSS.
- تم فحص صلاحية فقرات الاختبار من خلال حساب معاملات التمييز، ومعامل ثبات كرونباخ ألفا.

عرض النتائج ومناقشتها:

السؤال الأول: الذي ينص على **"هل يوجد أثر لطريقة معالجة أثر التخمين على القيم المفقودة في استجابات الطلبة لاختبار اختيار من متعدد في مادة الفيزياء لطلبة الصف العاشر الأساسي؟"**

للإجابة عن هذا السؤال تم استخراج نسب القيم المفقودة في استجابات الطلبة على اختبار الفيزياء باختلاف طريقة معالجة أثر التخمين المتمثلة في تعليمات الاختبار بالإضافة الى طريقة إكمال الأسئلة كما يأتي:

جدول 2: نسبة القيم المفقودة باختلاف نموذج معالجة أثر التخمين

النموذج	عدد القيم المفقودة	عدد الاستجابات الكلي	نسبة القيم المفقودة
الاكمال	750	18360	0.041
الثواب	1320	18360	0.072
العقاب	1350	18360	0.074
الاستجابة الجزئية	270	18360	0.015

يتضح من خلال جدول 2 أن نسب القيم المفقودة تختلف باختلاف نموذج المعالجة، إذ يلاحظ أن أعلى نسبة للقيم المفقودة بلغت (0.074) عند استخدام طريقة العقاب، في حين بلغت أقل نسبة

للقيم المفقودة (0.015) عند استخدام طريقة الاستجابة الجزئية. ولفحص اختلاف نسب القيم المفقودة تم استخدام اختبار مربع كاي الذي بلغت قيمته (908.835) بدرجة حرية (3) وقيمة احتمالية أقل من (0.001) التي تشير لدلالة إحصائية مرتفعة مما يعني أن نسب القيم المفقودة تختلف باختلاف طريقة معالجة التخمين. ولمعرفة لصالح من كان الاختلاف تم إجراء مقارنات ثنائية باستخدام اختبار مربع كاي كما هو مبين في جدول 3.

جدول 3: المقارنات الثنائية لاختلاف نسب الفقد باختلاف طريقة معالجة التخمين.

النموذج	الإكمال	الثواب	العقاب	الاستجابة الجزئية
الإكمال	—	166.333	181.827	232.336
الثواب	<0.001	—	0.364	724.780
العقاب	<0.001	0.547	—	753.231
الاستجابة الجزئية	<0.001	<0.001	<0.001	—

*: القيم فوق القطر تمثل قيم اختبارمربع كاي. والقيم أسفل القطر تمثل القيم الاحتمالية لاختبار مربع كاي.

يلاحظ من جدول 3 أن جميع النسب مختلفة عن بعضها البعض باختلاف طريقة معالجة التخمين، باستثناء نسبتي القيم المفقودة عند معالجة التخمين بطريقتي الثواب والعقاب فقد كانتا متساويتين حيث بلغت القيمة الاحتمالية لاختبار مربع كاي (0.547) وهي أكبر من مستوى الدلالة (0.05) مما يشير لعدم اختلاف النسبتين. كما يلاحظ أن نسبة القيم المفقودة عند معالجة أثر التخمين باستخدام العقاب هي الأعلى يليها الثواب وهما تختلفان عن باقي النسب. وهذا يعني أن التخمين في أقل حالاته، أي أن طريقتي العقاب والثواب ساهمتا في خفض التخمين بشكل أفضل من طريقة الاستجابة الجزئية.

كما يلاحظ أن نسبة القيم المفقودة عند استخدام النموذج المرجعي "الإكمال" ارتفعت عند استخدام طريقتي العقاب والثواب، بينما انخفضت عند استخدام الاستجابة الجزئية. أي أن طريقة الاستجابة الجزئية شجعت الطلبة على التخمين أكثر فالطريقة توفر مرونة أكثر لمحاولة الوصول إلى الإجابة الصحيحة.

يمكن عزو هذه النتيجة إلى تأثير طريقتي المكافأة والعقاب على سلوكيات الاختبار، فقد أظهرت الأبحاث في مجال القياس التربوي أن طريقتي المكافأة والعقاب تؤثران بشكل كبير على سلوك التخمين. ووفقًا لبن-سيمون وآخرون (Ben-Simon et al., 1997)، فإن اتباع طريقة العقاب للإجابات الخاطئة تزيد من حذر الطلبة وتقلل من ميلهم للتخمين، حيث تتراجع الفائدة المتوقعة لديهم من الإجابات العشوائية.

وبالمثل، وجد روجرز وهارلي (Rogers & Harley, 1999) ان الطلبة الذين يخضعون للتقييم القائم على العقاب يميلون إلى حذف أسئلة أكثر تكرارًا لتجنب فقدان الدرجات، مما يؤدي إلى ارتفاع معدلات إغفال الإجابات، وهو ما يتوافق مع النتائج التي تم الوصول لها.

كما أن طريقة أو نهج المكافأة يعمل بطريقة مشابهة، وإن كانت أقل حدة، من خلال تعزيز الاستجابة الصحيحة دون إلغاء حافز التخمين تمامًا. إذ تتبع المساواة في نسب القيم المفقودة بين طريقتي المكافأة والعقاب (كما يتضح من اختبار مربع كاي غير الدال إحصائيًا) من توازن نفسي: فبينما تردع طريقة العقاب عن المخاطرة، تشجع طريقة المكافأة على بذل الجهد، وكلتاهما تعدّلان السلوك للحد من التخمين العشوائي (Van der Linden & Glas,2010)

في المقابل، تسمح طريقة الاستجابة الجزئية (Masters, 1982) بالتعرف الجزئي على الإجابات الصحيحة جزئيًا، إذ تُخفف هذه المرونة من قلق الاختبار وتزيد من المشاركة، مما يدفع الطلبة إلى محاولة حل المزيد من الأسئلة حتى في حال عدم تأكدهم (De Ayala, 2013) ونتيجةً لذلك، يُشجع سلوك التخمين بشكل غير مباشر، إذ يشعر المتقدمون للاختبار بمخاطر أقل عند محاولة الإجابة لاعتقادهم بأنهم قد يحصلون على بعض الدرجات بدلاً من لا شيء.

ووجد كو وهيراتا (Kuo & Hirata, 2014) أن طريقة التقييم الجزئي للدرجات في نماذج الاستجابة للفقرة (IRT) متعددة التدريج يُعزز "التخمين المُجهد"، حيث يستخدم الطلبة المعرفة الجزئية التي يمتلكونها لتقريب الإجابة الصحيحة، وبالتالي ينخفض معدل فقدان البيانات وفقًا لطريقة الاستجابة الجزئية.

كما يتوافق النمط المرصود في استجابات الطلبة مع نتائج نظرية الدافعية وقيمة التوقع لإكلس وويجفيلد (Eccles, & Wigfield, 2002) والتي تشير إلى أن سلوك أداء الاختبار يعتمد على التكلفة والقيمة المُدركة للإجابة، فعند تطبيق طريقة العقاب، تزداد "تكلفة" الإجابة الخاطئة، مما يدفع الطلبة إلى حذف الأسئلة التي لا يثقون تماما من قدرتهم على إجابتها بشكل صحيح. في المقابل، تُخفّض طريقة الاستجابة الجزئية التكلفة المُدركة، مما يزيد من معدلات الاستجابة عند توفر المعرفة الجزئية لدى الطلبة. بالإضافة إلى ذلك، لاحظ وايز وديمارس (Wise & DeMars, 2005) أن الشروط منخفضة المخاطر أو قواعد التقييم المرنة (مثل طريقة الاستجابة الجزئية) غالبًا ما تُؤدي إلى تخمين أو انسحاب أسرع، حيث يُدرك الطلبة عواقب أقل للإجابات الخاطئة.

تتفق هذه النتيجة مع ما توصلت إليه عدد من الدراسات السابقة؛ فقد أشارت نتائج الخرشة (2016) ودراسة شينيل وآخرون (Şenel et al., 2015) إلى أنّ طريقة العقاب ترفع من دقة تقدير خصائص الفقرات وتحد من التخمين، كما بيّنت دراسة نور الدين (2021) أنّ إتاحة خيارات تقلل من التخمين –مثل بديل (لا أعرف)– تخفّض من معلمة التخمين وتحسّن خصائص الاختبار. وفي الإطار ذاته، بيّنت دراسة ريوس وآخرون (Rios et al., 2022) أن التخمين يشوه خصائص الفقرات وقدرات المفحوصين، مما يعزّز أهمية اتباع نماذج تقلل منه. ومن جانب آخر، اتفقت نتائج عودة (2022) مع ما ظهر في هذه الدراسة حول التأثير المختلف لآليات التحكم بالتخمين، إذ أشار إلى أن استراتيجيات تقليل التخمين قد تؤثر في صدق النتائج إذا لم تُطبق بالشكل الملائم. كما أكدت نتائج الزبون وآخرون (2021) والصرايرة (2018) على أن ارتفاع نسب القيم المفقودة قد ينعكس سلبًا على دقة تقدير الخصائص الإحصائية للاختبار، مما يدعم ضرورة اختيار طرق معالجة توازن بين خفض التخمين وتقليل الفقد في البيانات. وعليه، فإن تفسير النتائج يشير إلى أن العقاب والثواب كانا الأكثر فاعلية في تقليل التخمين، لكنه جاء على حساب زيادة القيم المفقودة؛ في حين وفّرت الاستجابة الجزئية مرونة أكبر للطلبة وشجعتهم على الإجابة، وإن كان ذلك قد يسمح بظهور معدلات أعلى من التخمين، بما يتطلب مراعاة هذا التوازن عند تصميم الاختبارات ومعالجة بياناتها لتحقيق أفضل مستويات الدقة والموضوعية في التقدير.

السؤال الثاني: الذي ينص على **"هل يوجد أثر لطريقة معالجة أثر التخمين والقيم المفقودة على خصائص فقرات اختبار اختيار من متعدد في مادة الفيزياء لطلبة الصف العاشر الأساسي وفقًا لنظرية الاستجابة للفقرة؟"**

للإجابة عن هذا السؤال تم إجراء تحليل لاستجابات الطلبة على الاختبار، ولكل من النماذج باختلاف طريقة معالجة أثر التخمين، حيث تم تقدير معالم فقرات الاختبار باستخدام النموذج ثنائي المعلمة (2PLM) لكل من طرق معالجة أثر التخمين (الاكمال، الثواب، العقاب) في حين تم تحليل الاستجابات الجزئية باستخدام نموذج الاستجابة المتدرجة (GRM) كما هو في جدول 4.

جدول 4: معالم فقرات الاختبار وفقا لنظرية الاستجابة للفقرة.

الفقرة	a1	b1	a2	b2	a3	b3	a4	d1	d2	d3	d
1	0.861	-0.365	0.663	-0.230	0.650	-0.378	1.714	-1.421	-0.025	0.332	-0.371
2	0.930	1.159	0.866	1.361	0.927	1.079	1.647	-0.444	0.940	1.118	0.538
3	1.771	0.853	1.585	0.948	1.868	0.817	0.768	-0.671	0.255	1.454	0.346
4	1.694	-0.750	2.091	-0.702	1.617	-0.912	1.184	-0.924	-0.450	0.210	-0.388
5	1.364	-1.505	1.586	-1.364	1.958	-1.356	1.734	-1.276	-0.687	-0.401	-0.788
6	1.828	-1.113	1.865	-1.133	1.746	-1.180	1.454	-2.400	-1.282	2.047	-0.545
7	0.922	2.221	1.074	1.977	0.795	2.641	1.347	0.172	0.605	2.273	1.017
8	1.617	-0.371	1.755	-0.381	1.893	-0.248	1.752	-1.207	-0.689	0.748	-0.382
9	0.729	-0.172	0.915	-0.317	0.825	-0.305	0.668	-1.113	0.112	0.398	-0.201
10	1.897	-0.439	1.854	-0.292	1.742	-0.362	0.653	-1.465	-0.353	0.632	-0.395
11	0.749	0.771	0.682	1.027	0.687	0.893	1.798	-0.643	-0.079	1.415	0.231
12	1.740	0.917	1.569	1.031	1.565	1.060	1.065	0.097	0.179	0.624	0.300
13	1.116	0.065	0.920	0.080	1.028	-0.254	0.799	-0.514	-0.370	0.082	-0.268
14	1.367	-1.504	1.410	-1.622	1.345	-1.452	0.719	-2.617	-1.391	-0.848	-1.619
15	1.817	1.523	1.788	1.638	2.179	1.372	1.012	0.343	0.876	1.134	0.784
16	1.152	-1.592	1.069	-1.637	1.094	-1.365	1.608	-2.769	-1.262	-0.296	-1.442
17	1.868	1.423	2.008	1.365	2.109	1.296	1.965	0.197	0.459	0.667	0.441
18	1.763	-1.353	1.796	-1.356	1.793	-1.345	2.241	-1.022	-0.533	0.072	-0.494
19	0.749	0.186	0.924	0.249	1.025	0.324	1.054	-0.799	0.552	0.675	0.143
20	1.857	0.362	1.610	0.220	1.723	0.365	0.643	-0.430	-0.153	0.997	0.138
21	1.243	0.890	1.335	0.928	1.478	0.873	1.258	-0.752	0.701	0.878	0.276
22	0.766	0.057	0.820	-0.177	0.658	0.173	1.271	-0.629	-0.388	0.757	-0.087
23	0.977	0.538	0.927	0.583	1.109	0.670	1.749	-0.464	0.135	0.768	0.146
24	2.429	1.698	1.783	1.961	2.192	1.768	1.551	0.423	0.972	1.083	0.826
25	1.593	0.573	1.609	0.507	1.690	0.568	1.354	-0.806	0.325	0.958	0.159
26	0.843	-0.715	0.660	-0.663	0.793	-0.720	1.630	-1.183	-0.562	0.289	-0.486
27	0.764	1.336	0.681	1.110	0.608	1.365	1.380	-0.092	0.446	0.713	0.356
28	1.752	-1.019	1.726	-0.895	1.619	-0.978	1.769	-0.828	-0.676	-0.088	-0.531
29	1.672	1.032	2.134	0.859	1.524	0.888	1.153	-1.028	0.309	1.044	0.108
30	1.971	1.532	1.685	1.537	1.728	1.580	1.009	-0.099	0.802	1.143	0.615

a: معلمة التمييز b: معلمة الصعوبة d: العتبة (الصعوبة المتدرجة للنماذج المتعددة). d: متوسط عتبات الفقرة المتدرجة

يلاحظ من خلال جدول 4 وجود فروق في كل من معلمتي صعوبة الفقرة والتمييز باختلاف طريقة معالجة أثر التخمين، ولفحص دلالة تلك الفروق تم إجراء تحليل تباين للقياسات المتكررة باعتبار مرات التطبيق هي تكرار لنفس فقرات الاختبار، ولكل من معلمتي الفقرة كالاتي:

معلمة الصعوبة: تم استخراج الأوساط الحسابية لقيم معلمة الصعوبة باختلاف طريقة معالجة أثر التخمين كما هو في جدول 5.

جدول 5: الأوساط الحسابية لمعلمة الصعوبة باختلاف طريقة معالجة أثر التخمين.

النموذج	الوسط الحسابي	الانحراف المعياري	العدد
الإكمال	.2079	1.08096	30
الثواب	.2204	1.08375	30
العقاب	.2292	1.09230	30
الاستجابة الجزئية	-.0524	.60927	30

يلاحظ من خلال جدول 5 وجود فروق في الأوساط الحسابية لقيم معلمة الصعوبة باختلاف طريقة معالجة أثر التخمين، ولفحص دلالة الفروق تم إجراء تحليل القياسات المتكررة؛ لأنه تم تكرار تقدير معلمة الصعوبة باختلاف تعليمات الاختبار التي تمثل طريقة معالجة أثر التخمين، والجدول 6 يبين نتائج تحليل القياسات المتكررة.

جدول 6: تحليل التباين في القياسات المتكررة لمعلمة الصعوبة باختلاف طريقة معالجة أثر التخمين.

مصدر التباين	مجموع المربعات	درجة الحرية	متوسط المربعات	قيمة F	القيمة الاحتمالية	مربع ايتا
طريقة معالجة أثر التخمين	1.667	3	.556	6.756	.000	.189
خطأ الطريقة	7.155	87	.082			
الفقرات	2.746	1	2.746	.750	.394	.025
خطأ الفقرات	106.157	29	3.661			

من خلال تحليل القياسات المتكررة يلاحظ أن تقدير معلمة الصعوبة يختلف باختلاف طريقة معالجة أثر التخمين حيث بلغت القيمة الاحتمالية 0.000 وهي أقل من مستوى الدلالة ($\alpha \leq 0.05$)، كما يلاحظ أن تلك الفروق جوهرية ويمكن الوثوق بها حيث بلغ مربع ايتا 0.189 وهي قيمة أكبر من 0.083 حسب إشارة هاسز (Hasse's 1982). ولمعرفة لصالح من كانت الفروق وأي الاختبارات كانت أصعب تم إجراء مقارنات بعدية كما هو مبين في جدول 7.

جدول 7: المقارنات الثنائية لاختلاف معلمة الصعوبة باختلاف طريقة معالجة التخمين.

النموذج	الإكمال	الثواب	العقاب	الاستجابة الجزئية
الإكمال	—	-0.012	-0.021	0.260
الثواب	0.625	—	0.009-	*0.273
العقاب	0.413	0.815	—	*0.282
الاستجابة الجزئية	0.014	0.010	0.011	—

* القيم فوق القطر تمثل الفروق في الأوساط الحسابية. والقيم أسفل القطر تمثل القيم الاحتمالية للفروق.

من خلال المقارنات البعدية الواردة في جدول 7 يتبين أن الفروق في معلمة الصعوبة كانت لصالح كل من طريقتي الثواب والعقاب بمقابل طريقة الاستجابة الجزئية، أي أنه عند معالجة أثر التخمين بطريقة الاستجابة الجزئية يظهر الاختبار بشكل أسهل مما لو تم معالجة أثر التخمين من خلال كل من طريقتي العقاب والثواب.

معلمة التمييز: تم استخراج الأوساط الحسابية لقيم معلمة التمييز باختلاف طريقة معالجة أثر التخمين كما هو في جدول 8.

جدول 8: الأوساط الحسابية لمعلمة التمييز باختلاف طريقة معالجة أثر التخمين.

النموذج	الوسط الحسابي	الانحراف المعياري	العدد
الإكمال	1.3934	.47965	30
الثواب	1.3797	.47812	30
العقاب	1.3989	.50306	30
الاستجابة الجزئية	1.3316	.43136	30

يلاحظ من خلال جدول 8 وجود فروق في الأوساط الحسابية لقيم معلمة التمييز باختلاف طريقة معالجة أثر التخمين، ولفحص دلالة الفروق تم إجراء تحليل القياسات المتكررة لأنه تم تكرار تقدير معلمة التمييز باختلاف تعليمات الاختبار التي تمثل طريقة معالجة أثر التخمين، وجدول 9 يبين نتائج تحليل القياسات المتكررة.

جدول 9: تحليل التباين في القياسات المتكررة لمعلمة التمييز باختلاف طريقة معالجة أثر التخمين.

مصدر التباين	مجموع المربعات	درجة الحرية	متوسط المربعات	قيمة F	القيمة الاحتمالية	مربع ايتا
طريقة معالجة أثر التخمين	.084	3	.028	.230	.876	.008
خطأ الطريقة	10.646	87	.122			
الفقرات	227.172	1	227.172	428.076	.000	.937
خطأ الفقرات	15.390	29	.531			

من خلال تحليل القياسات المتكررة يلاحظ أن تقدير معلمة التمييز لا يختلف باختلاف طريقة معالجة أثر التخمين حيث بلغت القيمة الاحتمالية 0.876 وهي أكبر من مستوى الدلالة

(α≤0.05)، كما يلاحظ أن تلك الفروق ليست جوهرية ولا يمكن الوثوق بها حيث بلغ مربع ايتا 0.008 وهي قيمة أقل من 0.083 حسب إشارة هاسز (Hasse's 1982).

يمكن تفسير هذه النتيجة التي تتضمن وجود فروق دالة إحصائيًا في الأوساط الحسابية لقيم معلمة الصعوبة باختلاف طريقة معالجة أثر التخمين، في ضوء أن التغيير في قواعد التقييم أي طريقة معالجة أثر التخمين تُغير دوافع المُمتحنين وأنماط استجاباتهم، فعند اختلاف طرق معالجة التخمين كاستخدام طريقة العقاب أو الثواب أو طريقة الاستجابة الجزئية، يُغير الطلبة في نمط استجابتهم من حيث كيفية حل الأسئلة. إذ يزيد التقييم السلبي أو العقوبات الصريحة من التكلفة المتوقعة للتخمين العشوائي، وبالتالي يزيد من احتمالية إغفال الطالب لسؤال ما بدلاً من التخمين؛ كما يُمكن لطريقة المكافأة أن تُغير استراتيجية المحاولة (إما بتشجيع الإجابة بدقة أكبر أو بتغيير المكافأة المُتوقعة)، بينما تُخفض أنظمة الدرجات الجزئية التكلفة المُتوقعة للمحاولة غير المؤكدة، وبالتالي تُشجع على المزيد من المحاولات. تُغير هذه التحولات السلوكية النسب المُلاحظة للإجابات الصحيحة والقيم المفقودة، مما يُغير بشكل مباشر صعوبة الأسئلة المُقدرة في ظل اختبار الاستجابة للأحداث (تبدو الأسئلة أسهل عندما يُحاول الطلاب الأقل خبرة ويحصلون على درجات جزئية/أي درجات).

كما أن طريقة الاستجابة الجزئية تمنح سهولةً أكبر في حل الأسئلة، لما تقدمه من درجات غير صفرية للحلول غير الصحيحة أو الخطوات الوسيطة؛ مما يُقلل من عقوبة المخاطرة، وبالتالي يكون الممتحنون ذوو المعرفة الجزئية أكثر ميلاً للمحاولة والحصول على بعض النقاط، وبالتالي، سيتحول متوسط موقع الأسئلة (صعوبة) نحو قيم أسهل مقارنةً بطريقة العقاب في معالجة أثر التخمين ثنائي الأجزاء، نظرًا لزيادة احتمالية الحصول على درجات جزئية في مستويات القدرة المنخفضة. إذ إن هذه النتيجة تتوافق هذا مع صياغة طرق التقييم الجزئي الكلاسيكية والمناقشات العملية حول كيفية زيادة التقييم متعدد الأبعاد للدرجات الصحيحة/الجزئية المُلاحظة، وبالتالي تقليل الصعوبة المُقدرة (Frances & Solon, 2014).

كما تزيد كل من طريقتي العقاب والثواب من حالات الحذف، وبالتالي ترفع تقديرات الصعوبة، فعند تطبيق العقوبة/التصحيح السلبي، يحذف الممتحَنون وخاصةً الممتحَنون الذين يتجنبون المخاطرة أو ذوو القدرات المنخفضة المزيد من الفقرات الاختبارية أو الأسئلة، مما يقلل من معدلات الصواب الملحوظة لذوي القدرات المنخفضة ويرفع تقديرات معلمة صعوبة الاختبار (حيث تبدو الفقرات أكثر صعوبة). وبالمثل، يمكن لطريقة المكافأة التي تُغيّر المكاسب المتوقعة أن تُغيّر أنماط المحاولة/الحذف، إذ تُظهر النمذجة والأبحاث التجريبية الحديثة أن التصحيح السلبي (طريقة العقاب في معالجة أثر التخمين) يُقلل من التخمين ويزيد من حالات الحذف، مما يُحسّن غالبًا قياس القدرة الحقيقية ولكنه يُغيّر معالم الفقرات الملحوظة (Akyol et al., 2016). وهذا ما يتوافق مع فكرة أن قواعد التقييم (طرق معالجة أثر التخمين) تُحدث تغييرات جوهرية وذات مغزى عملي في معلمة الصعوبة المُلاحظة عند إعادة تقييم الفقرات نفسها وفقًا لقواعد تخمين مختلفة.

وقد جاءت هذه النتيجة متفقة مع ما جاء من نتائج في دراسة ريوس وآخرون (Rios et al., 2022) التي بينت أن اتباع أسلوب التخمين من قبل المفحوصين يعمل على تشويه خصائص الاختبار، وخصائص الفقرات وقدرات المفحوصين بانحراف معياري مقداره (0.13). وكذلك نتائج دراسة الخرشة (2016) التي بينت وجود فروق ذات دلالة إحصائية بين متوسطات معالم

الصعوبة للفقرات تعزى لطرق معالجة التخمين، وكذلك نتائج دراسة شينيل وآخرون (Şenel et al., 2015) التي بينت وجود فروق ذات دلالة إحصائية في صعوبة الفقرات وتميزها تعزى لطريقة التصحيح من أثر التخمين لصالح طريقة التصحيح (العقاب)، ونتائج دراسة عبد الوهاب (2013) التي بينت وجد اختلاف واضح في إحصائيات الفقرة في ضوء النموذج الثلاثي المعلمة عن النظرية الكلاسيكية قبل تصحيح أثر التخمين، مما يشير إلى اختلاف كل منهما في معالجة أثر التخمين عند تقدير هذه الإحصائيات.

أما عدم وجود فروق ذات دلالة إحصائية في معلمة التمييز فيمكن رده إلى أن التمييز مرتبط بمحتوى المادة وأهدافها أكثر من ارتباطه بالتقييم السطحي؛ إذ تعكس معلمة التمييز مدى قدرة المادة على التمييز بين الطلاب ذوي القدرات العالية والطلاب ذوي القدرات المنخفضة (ميل ICC). في حين أن طرق معالجة أثر التخمين تُغير مستوى الأداء (احتمالات التقييم)، إلا أنها لا تُغير دائمًا مدى ارتباط نجاح المادة بالقدرة إلا إذا أدت تغييرات التقييم إلى سلوك تفاضلي منهجي عبر مستويات القدرات (على سبيل المثال، الطلاب ذوو القدرات المنخفضة فقط هم من يُغيرون استراتيجيتهم). إذ إنه في العديد من الدراسات، تُعتبر تقديرات التمييز قوية نسبيًا عبر طرق معالجة أثر التخمين المعقولة، لأن ترتيب الممتحنين والاختلافات النسبية في أداء المادة تبقى متشابهة، لذا ينتج هذا الاستقرار نتائج تحليل غير دالة إحصائيًا لمعلمة التمييز (Espinosa & Gardeazabal, 2010).

كما تشير الدراسات السابقة إلى أن وجود أثر لطرق معالجة أثر التخمين في معلمة التمييز يحدث عندما تؤثر طرق معالجة أثر التخمين بشكل تفاضلي على الطلبة ذوي القدرات المنخفضة مقارنةً بالطلاب ذوي القدرات العالية (على سبيل المثال، إذا زاد الطلبة ذوو القدرات المنخفضة فقط من قدرتهم على التخمين في ظل طريقة الاستجابة الجزئية، بينما لم يتأثر الطلبة ذوو القدرات العالية)، فقد ينخفض التمييز (إذا كان التخمين يرفع معدلات الصواب للطلبة ذوي القدرات المنخفضة) أو يزداد (إذا قمعت العقوبات بشكل انتقائي تخمينات الطلبة ذوي القدرات المنخفضة) (Adedoyin et al., 2008). إلا أن طرق معالجة أثر التخمين التي تم الاعتماد عليها في هذه الدراسة أنتجت تحولات شبه موحدة عبر مستويات القدرات (تحويل نقاط عتبات الصعوبة) بدلاً من التأثير على علاقة الميل بالقدرة واحتمال النجاح. تُظهر الدراسات التجريبية حول ثبات معاملات IRT واستقرارها عبر إعادة التقييم أن التمييز يكون أحيانًا، وليس دائمًا، حساسًا لمثل هذه المعالجات (Adedoyin et al., 2008).

التوصيات:

في ضوء ما تم التوصل إليه من نتائج توصي الدراسة الحالية بـ:

- أن يحرص القائمون على بناء الاختبارات الموضوعية، خصوصًا في المواد العلمية كالفيزياء، على اختيار طريقة معالجة التخمين التي تتناسب مع طبيعة الفقرات ومستوى الطلبة، مع توعية المعلمين بأثر كل طريقة على الخصائص الإحصائية للفقرات.
- أن يعتمد الباحثون والمقوِّمون التربويون على أكثر من طريقة لمعالجة البيانات المفقودة قبل اتخاذ قرارات تتعلق بجودة الفقرات أو أداء الطلبة، كأن يشمل ذلك تطبيق طرائق التقدير البايزية أو خوارزميات الإكمال المتعدد (Multiple Imputation)، لما لها من أثر في تحسين دقة التقدير وتقليل تحيز النتائج.
- تنظيم دورات تدريبية وورش عمل للمعلمين والباحثين في مجال القياس والتقويم حول كيفية تطبيق وتحليل نماذج نظرية الاستجابة للفقرة وتفسير نتائجها ومقارنتها بالنتائج المستخلصة من التحليل الكلاسيكي.

المصادر والمراجع:

الخرشة، 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)، 2348-2366.

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

الصرايرة، راجي (2018). أثر نسبة القيم المفقودة وطريقة معالجتها في دقة تقدير القيمة القصوى لدالة المعلومات للفقرة وقيمة مؤشر ثبات الفقرة. دراسات، العلوم التربوية، 45(4)، 485-501.

الصرايرة، راجي. (2018). أثر نسبة القيم المفقودة وطريقة معالجتها في دقة تقدير القيمة القصوى لدالة المعلومات للفقرة وقيمة مؤشر ثبات الفقرة. دراسات-العلوم التربوية، 45(4)، 485-501.

عبد الوهاب، محمد. (2013). المقارنة بين النظرية الكلاسيكية والنموذج ثلاثي البارامتر في معالجة أثر التخمين عند تقدير إحصائيات الأفراد والمفردات في اختبار "اختيار من متعدد". المجلة التربوية، 28(109)، 169-209.

علام، صلاح الدين. (2005). نماذج الاستجابة للمفردة الاختبارية أحادية البعد ومتعددة الأبعاد وتطبيقاتها في القياس النفسي والتربوي. ط1، القاهرة: دار الفكر العربي للنشر والتوزيع.

عودة، أحمد. (2022). التصحيح لأثر التخمين في اختبارات الاختيار من متعدد بين المعرفة الجزئية والمخاطرة: نموذج مقترح لتدريج ثنائي الاتجاه. مجلة اتحاد الجامعات العربية للبحوث في التعليم العالي، 42(1)، 82-95.

عودة، أحمد. (2022). القياس والتقويم في العملية التدريسية. ط5، الأردن: مركز أمية.

القرني، أحمد وأبو هاشم، السيد (2021). أثر طريقة معالجة القيم المفقودة على دقة معادلة الاختبارات في نظرية الاستجابة للفقرة. دراسات-العلوم التربوية، 29(2)، 455-491.

النصراوين، معين والبطش، محمد. (2018). مقارنة أربعة نماذج لمعالجة التخمين في الأسئلة الموضوعية/الاختيار من متعدد في إطار النموذج اللوجستي ثلاثي المعلمة وأثرها على دقة تقدير معلمة القدرة. دراسات، العلوم التربوية، 45(4)، 332-353.

نور الدين، أمين. (2021). أثر إضافة بديل "لا أعرف" في أسئلة الاختيار من متعدد على تقدير قدرة الفرد وخفض بارامترات التخمين وتحسين كفاءة الاختبار في ضوء نظرية الاستجابة للمفردة. المجلة المصرية للدراسات النفسية، 32(114)، 126-164.

Resources:

Adedoyin, O., Nenty, H., & Chilisa, B. (2008). Investigating the invariance of item difficulty parameter estimates based on CTT and IRT. Educational Research and Reviews, 3(3), 83-93.

Akyol, Ş., Key, J., & Krishna, K. (2016). Hit or miss? Test taking behavior in multiple choice exams (No. w22401). National Bureau of Economic Research, 22(40), 2-53.

Al-zboon, H. s., Alnasraween, M. E. S., & Alkursheh, T. O. (2021). The effect of the percentage of missing data on estimating the standard error of the items' parameters and the test information function according to the three-parameter logistic model in the item response theory. Ilkogretim Online, 20(1), 887-898.

Barton, M. A., & Lord, F. M. (1981). An upper asymptote for the three-parameter logistic item-response model. ETS Research Report Series, 1981(1), 1-8.

Ben-Simon, A., Budescu, D. & Nevo, B. (1997). A comparative study of measures of partial knowledge in multiple-choice tests. Applied Psychological Measurement, 21(1), 65-88.

Brennan, R. L. (Ed.). (2023). Educational Measurement. Rowman & Littlefield.

Brown, A., & Maydeu-Olivares, A. (2010). How IRT can Solve Problems of Impassive Data (Doctoral dissertation, Doctoral dissertation). University of Barcelona, Spain. Retrieved from http://hdl. Handle. net/10803/80006).

Budescu, D., & Bar-Hillel, M. (1993). To guess or not to guess: A Decision-Theoretic view of formula scoring. Journal of Educational Measurement, 30(4), 277-291.

Burton, R. (2004). Multiple choice and true/false tests: reliability measures and some implications of negative marking. Assessment & Evaluation in Higher Education, 29(5), 585-595.

Bush, M. (2015). Reducing the need for Guesswork in Multiple-Choice tests. Assessment & Evaluation in Higher Education, 40(2), 218-231.

De Ayala, R. (2013). The theory and practice of item response theory. Guilford Publications.

Eccles, J., & Wigfield, A. (2002). Motivational beliefs, values, and goals. Annual Review of Psychology, 53(1), 109–132.

Erguven, M. (2013). Two Approaches to Psychometric Process: Classical Test Theory and Item Response Theory. Journal of Education, 2(2), 23-30.

Espinosa, M. P., & Gardeazabal, J. (2010). Optimal Correction for Guessing in Multiple-choice Tests. Journal of Mathematical psychology, 54(5), 415-425.

Frances, M., & Solon, T. (2014). Item response theory for measurement validity. Shanghai archives of Psychiatry, 26(3), 171-177.

Frary, R. B. (1989). Partial-Credit Scoring Methods for Multiple-Choice Tests. Applied Measurement in Education, 2(1), 79-96.

Gemici, S., Bednarz, A., & Lim, P. (2012). A Primer for Handling Missing Values in the. Analysis of Education and Training Data. International Journal of Training Research, 10(3), 233-250.
Golvardi Yazdi, M. S., Haghighat Shoar, S. M., Sobhani, G., Vafi Sani, F., Khoshkholgh, R., Mousavi Bazaz, N., & Mansourzadeh, A. (2021). Factors Affecting Students' Guesswork in Multiple Choice Questions and Corrective Strategies. Medical Education Bulletin, 2(4), 297-305.
Graham, J. (2009). Missing Data Analysis: Making. It Work in the Real world. Annual review of psychology, 60(1), 549-576.
Kubiszyn, T., & Borich, G. D. (2024). Educational Testing and Measurement. John Wiley & Sons.
Kuo, B., & Hirata, S. (2014). Partial credit scoring and examinee behavior in mixed-format tests. Educational and Psychological Measurement, 74(6), 1054–1073.
Lesage, E., Valcke, M., & Sabbe, E. (2013). Scoring Methods for Multiple-Choice Assessment in Higher Education–Is it Still, a Matter of Number Right Scoring or Negative Marking. Studies in Educational Evaluation, 39(3), 188-193.
Little, R. J., & Rubin, D. B. (2019). Statistical Analysis with. Missing Data (Vol. 793). John Wiley & Sons.
Lord, F. M. (2012). Applications of Item response Theory to. Practical Testing Problems. Routledge.
Ludlow, L, H., & Oleary, M. (1999). Scoring Omitted and Not-Reached Item: Practical Data Analysis Implication. Educational and psychological Measurement, 59(4), 615-630.
Lüdtke, O., Robitzsch, A., & West, S. G. (2020). Regression Models Involving Nonlinear effects with Missing Data: A Sequential Modeling Approach Using Bayesian Estimation. Psychological Methods, 25(2), 157-181.
Masters, G. N. (1982). A Rasch Model for Partial Credit Scoring. Psychometrika, 47(2), 149-174.
Mcknight, P., Mcknight, K., Sidani, S., & Figueredo. A. (2007). Missing Data: A gentle Introduction. New York: Guilford press.
Messick, S. (1984). The psychology of educational measurement. ETS Research Report Series, 1984(1), 1-55.
Nason, G. P. (1995). Choice of the threshold parameter in wavelet function estimation. In Wavelets and statistics (pp. 261-280). New York, NY: Springer New York.
Rios, J., Deng, J., & Ihlenfeldt, S. (2022). To what Degree does Rapid Guessing Distort Aggregated Test Scores? A Meta-Analytic Investigation. Educational Assessment, 27(4), 356-373.
Rogers, W., & Harley, D. (1999). An empirical comparison of three- and four-option multiple-choice items. Educational and Psychological Measurement, 59(3), 382–395.
Roth, P. (1994). Missing data: A conceptual. Review for Applied Psychologists. Personnel Psychology, 47, 537-560.

Salunkhe, D. P., Bartakke, R. N., Chvan, G., Kothavale, P. R., & Digvijay, P. (2017). An overview on methods for slope stability analysis. International Journal of Engineering Research & Technology (IJERT), 6(03), 2278-0181.
Schafer, J., & Graham, J. W. (2002). Missing Data: Ours. View of the State of the Art. Psychological Methods, 7(2),147–177.
Şenel, S., Pehlivan, E., & Alatlı, B. (2015). Effect of Correction-for-Guessing Formula on Psychometric Characteristics of Test. Procedia-Social and Behavioral Sciences, 191(2015), 925-929.
St. Hilaire, K., Chan, J., & Ahn, D. (2024). Guessing as a Learning Intervention: A Meta-Analytic Review of the Prequestion Effect. Psychonomic Bulletin & Review, 31(2), 411-441.
Swaminathan, H., & Gifford, J. (1985). Bayesian Estimation in the Two-Parameter Logistic Model. Psychometrika, 50, 349-364.
Thacker, B., Chapagain, G., Pattillo, D., & West, K. (2013). The Effect of Problem Format on Students Responses. Journal of Texas Tech University, 1(40), 1-42.
Van der Linden, W., & Glas, C. (Eds.). (2010). Elements of adaptive testing (Vol. 10, pp. 978-0). New York: Springer.
Vuorre, M., & Metcalfe, J. (2022). Measures of Relative Metacognitive Accuracy are confounded with Task Performance in Tasks that Permit Guessing. Metacognition and Learning, 17(2), 269-291.
Wise, S., & DeMars, C. (2005). Low examinee effort in low-stakes assessment: Problems and potential solutions. Educational Assessment, 10(1), 1–17.
Wu, Q., De Laet, T., & Janssen, R. (2019). Modeling Partial Knowledge on Multiple-Choice Items Using Elimination Testing. Journal of Educational Measurement, 56(2), 391-414.