أثر التخمين على معامل ثبات الاختبار ودالة المعلومات بالاعتماد على نظرية استجابة الفقرة

حابس سعد موسى الزبون[1]

تاريخ وصول البحث: 2024/08/21 م **تاريخ قبول البحث: 2024/09/16م** **تاريخ نشر البحث: 2025/12/23م**

الملخص

هدفت هذه الدراسة إلى معرفة اثر التخمين على معامل ثبات الاختبار ودالة المعلومات بالاعتماد على نظرية استجابة الفقرة وفق نموذج راش، وللإجابة عن أسئلة الدراسة قام الباحث ببناء اختبار تحصيلي في مساق القياس النفسي والتقويم التربوي والذي يدرس في جامعة نزوى على الفصل الدراسي الأول للعام الجامعي (2023-2024) تكون الاختبار بصورته النهائية من (35) فقرة، تكونت عينة الدراسة من (838) طالبا وطالبة من طلبة التأهيل التربوي في جامعة نزوى تم اختيارهم بالطريقة العشوائية البسيطة من مجتمع الدراسة ، وبالاعتماد على البرنامج الإحصائي (SPSS) وبرنامج (Bilog-mg3) لتحليل استجابات عينة الدراسة. أظهرت نتائج الدراسة أن معامل الثبات للاختبار في حال استخدام معادلة التصحيح لأثر التخمين أعلى منه في حالة استخدام الطريقة التقليدية، وأظهرت وجود فروق ذات دلالة إحصائية بين الأوساط الحسابية لقيم دالة معلومات فقرات الاختبار لصالح معادلة التصحيح لأثر التخمين، وأن دالة المعلومات للاختبار ككل كانت أفضل عند استخدامع معادلة التصحيح لأثر التخمين، كما اظهرت النتائج عدم وجود فروق ذات دلالة إحصائية بين المتوسطات الحسابية بالقيم الخاصة لاحتمال الاجابة الصحيحة على الفقرة وفقا لطريقة التصحيح.

الكلمات المفتاحية: *التخمين، اختبار الاختيار من متعدد، نظرية استجابة الفقرة، دالة المعلومات، احتمال الاجابة الصحيحة للفقرة.*

The Effect of Guessing on the Test Reliability Coefficient and the Information Function Based on Item Response Theory

Abstract

This study aimed to determine the effect of guessing on the test reliability coefficient and the information function based on item response theory according to the Rasch model. To answer the study questions, an achievement test was constructed in the course of psychological measurement and educational evaluation which is taught at the University of Nizwa in the first semester of the academic year (2023-2024). The test in its final form consists of (35) items. The study sample consisted of (838) male and female educational rehabilitation students at the University of Nizwa who were selected by a simple random method from the study population, relying on the statistical program (SPSS) and the program (Bilog-mg3) to analyze sample responses. The results of the study showed that the reliability coefficient of the test when using the correction equation for the effect of guessing is higher than when using the traditional method. Result also showed that there were statistically significant differences between the arithmetic means of the values of the information function of the test items in

(1) أستاذ دكتور، قسم المناهج والتدريس، كلية العلوم التربوية، جامعة الحسين بن طلال، معان، الأردن.

** الباحث المستجيب:* habis.s.alzboon@ahu.edu.jo

Favor of the correction equation for the effect of guessing, and that the information function for the test as a whole was better when using the correction equation for the guessing effect.

Moreover, it was shown the results showed that there were no statistically significant differences between the arithmetic averages of the respective values of the probability of a correct answer to the item according to the correction method.

Keywords: *guessing, multiple choice test, item response theory, information function, probability of the correct answer to the item.*

مقدمة:

تعتبر الاختبارات هي أكثر أدوات القياس استخداما في العملية التدريسية، فقد اهتم التربويين بها اهتماما كبيرا من حيث بنائها وتطويرها ، وذلك لما تقدمه من بيانات موضوعية يمكن الاعتماد عليها في صنع القرارات السليمة التي تهدف الى تحسين عمليتي التعلم والتعليم (الكيلاني، وعدس، 1994).

إذ تُعد الاختبارات أحدى أهم أدوات القياس التي يتم الاعتماد عليها في قياس وتقويم قدرات الطلاب؛ وذلك للتعرف على مستواهم التحصيلي، والتعرف إلى درجة تحقق الأهداف السلوكية أو النواتج التعليمية، وما يقدمه المعلم من نشاطات تعليمية مختلفة تساعد على رفع الكفايات التحصيلية لدى الطلبة (النجار، 2010).

وتتنوع الاختبارات تبعاً للأهداف التي وضعت من أجلها، فما يصلح لمادة دراسية قد لا يصلح لأخرى، وما يصلح لهدف قد لا يصلح لآخر، وما يصلح لمستوى من الطلاب قد لا يصلح لمستوى آخر، فعلى المدرس أن يختار النوع المناسب، ومن أهم هذه الأنواع اختبارات التحصيل (Achievement Tests) التي تهدف لقياس تحصيل الطلبة (الكيلاني وعدس، 1986).

وقد انتشر استخدام الاختبارات الموضوعية بين المعلمين حيث يلاحظ أن المعلمين يستخدمون أنواعا متعددة منها، فمنهم من يستخدم نوع واحد من هذه الاختبارات ومنهم من يستخدم أكثر من نوع، ولكن اختبارات الاختيار من متعدد هي الأكثر استخداماً، وذلك لأن كثيرا من المعلمين والمقيمين يرون أنها أفضل أنواع الاختبارات الموضوعية وأكثرها مرونة في قياس مستويات متعددة من الأهداف السلوكية، أي أنها لا تقتصر على قياس تذكر المعلومات بل تتعدى ذلك الى قياس مستويات الفهم والاستيعاب والتطبيق والتحليل، وغيرها من النشاطات العقلية المركبة، كما يرون أيضا أنها تتمتع بدلالات صدق وثبات كبيرتين، تفوق دلالات صدق وثبات الأنواع الأخرى من الاختبارات الموضوعية (الكيلاني، 1994).

ويلجأ المفحوصون في أسئلة فقرات الاختيار من متعدد للغش أوالتخمين العشوائي، عندما لا تتوافر لديه المعلومات التي تمكنه من الوصول الى الإجابة الصحيحة من أكبر المشكلات التي تواجه هذا النوع من الاختبارات، حيث يقوم الطالب بالاختيار العشوائي لاحد البدائل حتى لو طلب منه ان يترك الفقرة بدون إجابة إذا لم يكن لدية معرفة بالإجابة الصحيحة. ويشير كنجر وهولشر وريف (Kubinger,Holocher&Reif,2010) الى ان التخمين في الاختبارات الموضوعية وبالأخص الاختيار من متعدد، يؤدي الى غياب العدالة عن نتائج الاختبار، على سبيل المثال اذا تقدم مفحوصان من مستوى قدرة متوسطة لأداء اختبار موضوعي من نوع الاختيار من متعدد فان هنالك احتمالاً بان يجتاز المفحوص الذي يلجا للتخمين الاختبار او يحصل على درجة اعلى مقارنة بالمفحوص الذي أجاب عن فقرات الاختبار بدون تخمين، وايضاً اذا كان هنالك مفحوصان متفاوتان في القدرة، فقد ينجح الطالب ذو القدرة المنخفضة نتيجة لجوئه للتخمين، ويفشل الطالب ذو القدرة المرتفعة الذي أجاب عن فقرات الاختبار بناءً على معرفته الحقيقية، مما يؤدي الى انتهاك اهم الخصائص الأساسية للاختبار وهي العدالة فتصبح نتائج الاختبار متحيزه.

فالتخمين العشوائي يعد من المهددات لكل من صدق وثبات اختبار الاختيار من متعدد، حيث يرى مسك (Messick,1995) ان لجوء الطلبة الى التخمين خلال اجاباتهم على فقرات اختبار الاختيار من متعدد يضيف تبايناً لا يرتبط بالتباين في القدرة الحقيقة، أي ان الاختبار لا يقيس ما وضع لقياسه، ويترتب على ذلك تخفيض صدق الاختبار بشكل ملحوظ، والى جانب ذلك فان التخمين يعد نوعاً من الخطأ العشوائي في القياس، فهو عامل مؤثر في دقة تقدير القدرة للمفحوصين، كما ويؤثر على جودة الاختبار حيث أن قبول الفقرات أو رفضها يعتمد على قيم معاملات الصعوبة والتمييز اللذان يتأثران بعامل التخمين، ويؤثر التخمين في ثبات الاختبار من ناحيتين، الأولى: أن وقت الاختبار غير الكافي قد يؤدي إلى لجوء الطالب إلى التخمين العشوائي الذي قد يؤدي الى عدم اتساق اجابته عن هذه الفقرات المتبقية مع الفقرات التي استغرق فيها وقت كافياً للإجابة في بداية الاختبار؛ اما الثانية، فتتمثل في انه عند تطبيق الاختبار مرة أخرى فمن المحتمل الا يحصل المفحوص الذي لجأ الى التخمين العشوائي على الدرجة نفسها؛ لأنه قد يغير اجابته عن فقرات الاختبار التي قام بتخمينها على أساس غير منطقي(McMillan, 2013). ونظراً لأهمية معالجة اثر التخمين في اختبارات الاختيار من متعدد، برزت جهود حثيثة لإيجاد طرق للمساعدة في الحد من أثر عامل التخمين على دقة وصدق وثبات نتائج الاختبارات الموضوعية، وتعرف طريقة معالجة أثر التخمين بانها طريقة لتحديد كمية المعلومات التي نحصل عليها من إجراءات القياس

دون الاعتماد على التخمين، لتكون مؤشر على مستوى الأداء على الاختبار عند المفحوصين. وفيما يلي عرض لطرق ضبط أثر التخمين المستخدمة في هذه الدراسة.

أولاً: معادلة التصحيح من أثر التخمين (العقاب) لتصحيح فقرات الاختبار من نوع الاختيار من متعدد والتي تنص على أن درجة المفحوص على الاختبار بالاعتماد على هذه الطريقة يعبر عنها بالاعتماد على المعادلة الاتية:

$$\text{الدرجة المصححة من أثر التخمين (العقاب)} = \text{عدد الإجابات الصحيحة} - \frac{\text{عدد الاجابات الخاطئة}}{\text{عدد البدائل} - 1}$$

ثانياً: معادلة التصحيح من أثر التخمين (المكافأة) لتصحيح فقرات الاختبار من نوع الاختيار من متعدد والتي تنص على أن درجة المفحوص على الاختبار بالاعتماد على هذه الطريقة يعبر عنها بالاعتماد على المعادلة الاتية:

$$\text{الدرجة المصححة من أثر التخمين (المكافأة)} = \text{عدد الإجابات الصحيحة} + \frac{\text{عدد الفقرات المحذوفة}}{\text{عدد البدائل}}$$

ثالثاً: معادلة التصحيح من أثر التخمين (المكافأة والعقاب) لتصحيح فقرات الاختبار من نوع الاختيار من متعدد والتي تنص على أن درجة المفحوص على الاختبار بالاعتماد على هذه الطريقة يعبر عنها بالاعتماد على المعادلة الاتية:

$$\text{الدرجة المصححة من أثر التخمين (العقاب والمكافأة)} = \text{عدد الإجابات الصحيحة} + \frac{\text{عدد الفقرات المحذوفة}}{\text{عدد البدائل}} - \frac{\text{عدد الاجابات الخاطئة}}{\text{البدائل عدد} - 1}$$

مشكلة الدراسة وأسئلتها:

تستخدم أسئلة الاختيار من متعدد في التقييم الصفي بشكل واسع نظرا لسهولة تصحيحها وموضوعيتها، بالاضافة إلى أنها تغطي عددا أكبر من الحقائق التحصيلية مقارنة بباقي أنواع الأسئلة، مما يجعلها النوع المفضل في حال كان عدد الطلبة كبير رغم الانتقادات الموجهة لهذا النوع من الأسئلة، والتي تتركز في وجود نسبة تخمين عالية بالاضافة الى أنها تركز على المستويات المعرفية الدنيا مما قد يؤثر على الدرجة الحقيقية للمفحوص مما يؤثر بشكل كبير على الخصائص السيكومترية للاختبار ودالة معلوماته وبالتالي التأثير على القرارات التي يتم اتخاذها بناء على نتائج هذه الاختبارات.

لذلك اقترح بعض علماء القياس بعض الطرق للتغلب على مشكلة التخمين في أسئلة الاختيار من متعدد والتي منها استخدام معادلة التصحيح لأثر التخمين، لذا جاءت هذه الدراسة لمعرفة اثر التخمين على معامل ثبات الاختبار ودالة

المعلومات بالاعتماد على نظرية استجابة الفقرة ، ويمكن أن تتحدد مشكلة الدراسة الحالية بالإجابة عن الاسئلة الآتية:

1. **هل يختلف معامل ثبات الاختبار باختلاف طريقة التصحيح (معادلة التصحيح لاثر التخمين، التقليدية) ؟**
2. **هل تختلف دالة معلومات الفقرات باختلاف طريقة التصحيح (معادلة التصحيح لاثر التخمين، التقليدية) ؟**
3. **هل تختلف دالة معلومات الاختبار باختلاف طريقة التصحيح (معادلة التصحيح لاثر التخمين، التقليدية) ؟**
4. **ما أثر التخمين على احتمال الاجابة الصحيحة عن الفقرة ؟**

أهداف الدراسة:

تهدف هذه الدراسة الى معرفة هل يختلف معامل ثبات الاختبار باختلاف طريقة التصحيح (معادلة التصحيح لاثر التخمين، التقليدية) وفق نموذج راش في نظرية استجابة الفقرة ، كما تهدف الى معرفة هل تختلف دالة معلومات الاختبار باختلاف طريقة التصحيح (معادلة التصحيح لاثر التخمين، التقليدية) وفق نموذج راش في نظرية استجابة الفقرة.

أهمية الدراسة:

تأتي أهمية هذه الدراسة **من** خلال استخدامها طرق احصائية مختلفة تعتمد على النظرية الحديثة في القياس لتقدير ثبات الاختبار ودالة المعلومات وتأتي أهميتها ايضا في الاسهام بتطوير البحث العلمي في هذا السياق، وكذلك لتعطي ملاحظة الى واضعي الاختبارت بأن بضرورة معالجة التخمين عند تطبيق الاختبارات التحصيلية.

الاطار النظري:

نظرية استجابة الفقرة (IRT)

جاءت النظرية الحديثة في القياس أو ما يعرف بنظرية استجابة الفقرة لمعالجة المشاكل التي كانت تعاني منها النظرية الكلاسيكية فيما يتعلق بانتقاء الفقرات وذلك بتقديم طريقة لانتقاء الفقرات وقدرة المفحوص كي يكون لمطور المقياس اختيار الفقرات الأكثر فاعلية في مدى يحدد بعلامة قطع على مقياس القدرة تساعد في فصل مستويات الإتقان والإتقان على المقياس (Hambleton & Rogers,1991).

نماذج نظرية استجابة الفقرة:

تمثل النظرية الحديثة في القياس الاتجاه الحديث في القياس النفسي والتربوي، ويطلق عليها نظرية استجابة الفقرة (Item Response Theory) ونظرية السمات الكامنة، ويعتبر لورد (Lord) أول من اسس نظرية استجابة الفقرة للمهتمين بالقياس النفسي والتربوي (علام، 2002) .

وأعتمدت نظرية استجابة الفقرة على نمذجه للعلاقة بين متغير غير ملاحظ يستخدم لقياس القدرة التي يقيسها الاختبار، واحتمالية الاستجابة الصحيحة على الفقرة ،وذلك بالاعتماد على دوال لوغاريتمية تربط قدرة المفحوص، ومعالم الفقرة باحتمالية الاستجابة الصحيحة عليها، وقد تم اشتقاق العديد من النماذج يالاعتماد على هذه النمذجة جميعها تفترض وجود قدرة واحدة تقيس الأداء على الاختبار، وأن القدرة يمكن تمثيلها على متصل مستمر، ولكنها تتغير في خصائصها التي تصفها الفقرات، وأن درجات الصعوبة والقدرة نظرياعلى يمكن تمثيلها على متصل يتراوح بين (∞−) إلى (∞+) ، ولكنه عمليا يتراوح بين (3− و3+)؛ وذلك لأنه من النادر توافر قيما تزيدعن (3+) أوتقل عن (3−) (Hambleton & Swaminathan,1985).

وقد طورت نماذج رياضية مختلفة؛ وذلك للتعبير عن العلاقة بين القدرة واحتمال الإجابة الصحيحة على الفقرة وتختلف هذه النماذج في صيغتها الرياضية وعدد معالمها، كما وتختلف من حيث عدد الافتراضات اللازم توافرها في البيانات المستخدمة.

وتقسم نماذج نظرية استجابة الفقرة إلى قسمين :-

- **النماذج ثنائية التدريج (Dichotomous IRT Models)** منها: نموذج راش (النموذج أحادي المعلمة) ، النموذج ثنائي المعلمة، والنموذج ثلاثي المعلمة.

- **النماذج متعددة التدريج (Polytomous IRT Models)**، وتميز هذه النماذج بين المفحوصين الذين لديهم معرفة جزئية ، والذين ليس لديهم معرفة جزئية، ومن هذه النماذج :نموذج الاستجابة المتدرجة وهو امتداد للنموذج ثنائي المعلمة ،ونموذج التقدير الجزئي، وسلم التقدير (RSM) وهما امتداد للنموذج أحادي المعلمة، ونموذج الاستجابة الاسمية (NRM) الذي يناسب الاستجابات التي لا يمكن تصنيفها في فئات رتبية (Embreston & Rise, 2000)

تتنوع نماذج نظرية الاستجابة للفقرة ثنائية التدريج بالاعتماد على المعالم الثلاثة للفقرة، وتنبثق عنها مجموعة من النماذج الرياضية عرفت باسم نماذج السمات الكامنة، يعتمد كل نموذج منها على معادلة رياضية تحدد علاقة أداء الفرد على فقرة بقدرته التي تكمن وراء هذا الأداء وتفسره. وتتضمن نظرية الاستجابة للفقرة النماذج اللوجستية(Hambleton & Rogers,1991)

- **النموذج اللوجستي الأحادي المَعْلَم(One Parameter Logistic Model)** :

يعد النموذج اللوجستي أحادي المعلمة (One Parameter Logistic Model) والذي يعرف كذلك بنموذج راش (Rasch Model) من أحد أوسع النمإذج المستخدمة في نظرية استجابة الفقرة ، و الذي يفترض أن جميع الفقرات لا تختلف عن بعضها إلا بمعلمة صعوبة الفقرة ويفترض تساوي معلمة التمييز لجميع الفقرات، في حين تقترب معلمة التخمين للفقرات من الصفر، و يتخذ النموذج الصيغة الرياضية الآتية لتعبر عن احتمالية الإجابة الصحيحة على الفقرة:

$$P_i(\theta) = \frac{e^{D(\theta - b_i)}}{1 + e^{D(\theta - b_i)}}, \quad i=1,2,3,\ldots,n \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(1)$$

- **النموذج اللوجستي الثنائي المَعْلَم (Two Parameter Logistic Model)**، وفي هذا النموذج تختلف فقرات الاختبار في صعوبتها وتمييزها.

$$P_i(\theta) = \frac{e^{Da_i(\theta - b_i)}}{1 + e^{Da_i(\theta - b_i)}}, \quad i=1,2,3,\ldots,n \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(2)$$

- **النموذج اللوجستي الثلاثي المَعْلَم(Three Parameter Logistic Model) بيرنبوم (Birnbaum Model)** أم في هذا النموذج تختلف فقرات الاختبار في صعوبتها وتمييزها والتخمين، ويميز هذا النموذج بأنه يحدد مَعْلَمة الموقع (Location Parameter)، ومَعْلَمة ميل المنحنى المميز (Slope Parameter)، ومَعْلَمة التخمين (Guessing) لكل فقرة، وبالتالي يؤدي إلى نتائج أكثر واقعية في العملية الاختبارية وبناء المقاييس(علام، 2002).

حيث أن:

$$P_i(\theta) = c_i + (1 - c_i)\frac{e^{Da_i(\theta - b_i)}}{1 + e^{Da_i(\theta - b_i)}}, \quad i=1,2,3,\ldots,n \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots(3)$$

a_i: معلمة تمييز الفقرة.

θ = قدرة مفحوص ما، D =1.702 وهو عامل تدريج، b_i= معلم صعوبة الفقرة i .

$p_i(\theta)$، احتمال الاجابة الصحيحة للمفحوص الذي قدرته (θ) عن الفقرة i

c_i: معلم تخمين الفقرة والذي يمثل الخط التقاربي الأدنى (Lower Asymptote). و يعبر عن احتمالية أن يجيب المفحوص ذو القدرة المتدنية إجابة صحيحة على الفقرة معتمدا على التخمين العشوائي، و الهدف من إضافة هذا المَعْلَم في النموذج هو مراعاة عدم مطابقة منحنيات خصائص الفقرة عند الطرف الأدنى لمتصل القدرة حيث يكون التخمين أحد المتغيرات التي تؤثر في الأداء في الاختبار.

دالة المعلومات:

يعد مفهوم دالة المعلومات من المفاهيم المهمة والتي يعتمد عليها في نظرية استجابة الفقرة التي تساعد في تحقيق دقة القياس، والتي يمكن من خلالها تحديد الخطأ المعياري في التقدير.حيث أن الخطأ المعياري في تقدير القدرة يساوي معكوس دالة المعلومات والذي يعطى بالاعتماد على المعادلة الاتية:

$$SSE= 1/ I(\theta)$$

حيث $I(\theta)$: دالة معلومات الاختبار

كما بين هامبلتون وسوامينثان (Hambleton & Swaminathan, 1985) الى ان دالة معلومات الفقرة عبارة عن اقتران يبين مدى مساهمة الفقرة في تحديد القدرة ، وفي العموم فان الفقرات ذات التمييز العالي تساهم بقوة أكبر في تأكيد دقة القياس حول قيمة صعوبتها (bi) على متصل القدرة.

وتستخدم دالة معلومات الفقرة لاختيار الفقرات عند بناء الاختبارات، بالاعتماد على النظرية الحديثة في القياس، على افتراض ان معلومات الفقرة تتغير عبر مستويات السمة المختلفة؛ وعليه فإنه من الممكن اختيار فقرات تقدم دقة قياس مرتفعة عند نقطة معينة على متصل السمة، كما أن الفقرات التي تتضمن معالم تمييز كبيرة، تقدم معلومات أكبر عن قدرة المفحوصين؛ وبالتالي الحصول على دقة أكبر، وتعطى دالة المعلومات للاختبار من خلال المعادلة الآتية:

$$I(\theta) = \sum_{i=1}^{n} \frac{P_i'(\theta)^2}{P_i(\theta)Q_i(\theta)}$$

علما بأن:

(θ) : قدرة المفحوص، I(θ) : دالة المعلومات للاختبار، $P_i(\theta)$: احتمال الاجابة الصحيحة للمفحوص الذي قدرته (θ) عن الفقرة i وتعرف بـ دالة استجابة الفقرة ، $Q_i(\theta)=1-P_i(\theta)$ ، $P_i'(\theta)$: المشتقة الأولى لدالة استجابة الفقرة.

وتتأثر كمية المعلومات بمعالم الفقرة، حيث تكون أعلى كمية معلومات في النموذج الأحادي والثنائي المعلمة عندما تتساوي قدرة المفحوصين مع الصعوبة؛ وذلك لان شكل دالة معلومات الفقرة بشكل عام قريب من الشكل الجرسي (Bell Shape) ، وعند الاعتماد على النموذج اللوجستي الثلاثي يمكن ملاحظة تأثير معالم الفقرة على دالة معلومات الاختبار من خلال الاطلاع على المعادلة التالية (Hambleton & Swaminathan,1993).

$$I_i(\theta)=\frac{2.89a_i^{\,2}(1-c_i)}{(c_i+e^{(1.7a_i(\theta-b_i))})(1+e^{(1.7a_i(\theta-b_i)})^2}$$

وبالاطلاع على المعادلة الرياضية السابقة نلاحظ أن كمية المعلومات تكون أكبر بشكل عام عندما تكون قيمة معلمة التمييز أكبر ما يمكن، وأن كمية المعلومات تزداد كلما كانت معلمة التخمين قريبة من الصفر، وأن كمية المعلومات عندما تكون (θ) قريبة من (b_i) أكبر من كمية المعلومات عندما تكون (θ) بعيدة من (b_i) .

المفهوم الإحصائي لمعامل الثبات

على الرغم من أن العلامات الحقيقية ودرجات الأخطاء العشوائية تعد من المفاهيم الافتراضية، إلا أنها تفيد بدرجة كبيرة في التوصل إلى المفهوم الإحصائي للثبات، فإذا كانت درجات الخطأ غير مرتبطة بالعلامات الحقيقية فإن تباين العلامات الملاحظة = تباين العلامات الحقيقية +تباين الخطأ؛ أي

$$\sigma_{\mathrm{T}}^2+\sigma_{\mathrm{E}}^2=\sigma_{\mathrm{X}}^2\ldots\ldots\ldots(1)$$

بقسمة طرفي المعادلة(1) على σ^2_x نحصل على

$$\frac{\sigma_{\mathrm{T}}^2}{\sigma_{\mathrm{X}}^2}+\frac{\sigma_{\mathrm{E}}^2}{\sigma_{\mathrm{X}}^2}=1$$

أي أن

$$\frac{\sigma_T^2}{\sigma_X^2} = 1 - \frac{\sigma_E^2}{\sigma_X^2} \ldots\ldots\ldots\ldots\ldots (2)$$

والقيمة $\frac{\sigma_T^2}{\sigma_X^2}$ تدل على نسبة التباين الحقيقي إلى التباين الملاحظ وهذا ما يعبر عنه بعامل ثبات درجات الاختبار

Coefficient of Reliability

أما $\frac{\sigma_E^2}{\sigma_X^2}$ فتدل على نسبة تباين الخطأ إلى التباين الملاحظ

وبالتالي فإن معامل الثبات من المنظور الإحصائي يمكن اعتباره نسبة تباين العلامات الحقيقية إلى تباين العلامات الملاحظة كالتالي :

$$\text{معامل الثبات} = \frac{\text{تباين العلامات الحقيقية}}{\text{تباين العلامات الملاحظة}}$$

وهذا يعني أن معامل الثبات هو ذلك الجزء من تباين العلامات الملاحظة الذي يعزى إلى تباين العلامات الحقيقية. وتتراوح قيم هذا المعامل بين صفر، 1، ولذلك يعبر عن قيمته بكسر عشري مثل معامل الارتباط. فكلما اقتربت قيمته من الواحد الصحيح دل ذلك على قدر أكبر من ثبات العلامات، وبالتالي تزداد الثقة في دقة هذه العلامات واتساقها والاعتماد عليها(علام،2002).

تعرفنا فيما سبق على مفاهيم العلامة الحقيقية وخطأ القياس، ولقد اصبح من الواضح انه عندما يعطي مدرسٌ أو باحثٌ أو طبيبٌ اختباراً فإنهم يعرفون العلامات الملاحظة فقط، على الرغم من انهم معنيون بالعلامة الحقيقية، وهنا يبرز السؤال حول مدى التقارب بين العلامة الحقيقية والملاحظة للمفحوص ؟! وأحد المؤشرات لهذه العلاقة هو الارتباط بين هذه المتغيرات. و يعرف معامل الارتباط الذي يبين العلاقة بين العلامة الحقيقية والعلامة الملاحظة في الاختبار بـ مؤشر الثبات، وعليه فإن العلامة الملاحظة للمفحوص يمكن أن تكون

$$X = T + E$$

وبالعلامة المعيارية

$$x = t + e \cdots\cdots\cdots\cdots\cdots\cdots\cdots (3)$$

ولإيجاد معامل الارتباط نستخدم الصيغة التالية :

$$\rho_{XT} = \frac{\sum xt}{N\sigma_X\sigma_T} \cdots\cdots\cdots\cdots\cdots\cdots (4)$$

وبتعويض قيمة x فان الصيغة(4) تصبح

$$\rho_{XT} = \frac{\sum (e+t)t}{N\sigma_X\sigma_T}$$

$$\rho_{XT} = \frac{\sum t^2}{N\sigma_X\sigma_T} + \frac{\sum te}{N\sigma_X\sigma_T} \cdots\cdots\cdots\cdots\cdots\cdots (5)$$

ولكن الارتباط بين العلامة الحقيقية وخطأ القياس يفترض انه يساوي صفر؛ أي $\frac{\sum te}{N\sigma_X\sigma_T} = 0$ وكذلك

$$\sigma_T^{\,2} = \frac{\sum t^2}{N}$$

وبالتعويض في المعادلة 5 نجد أن :

$$\rho_{XT} = \frac{(\sigma_T)^2}{\sigma_X\sigma_T} = \frac{\sigma_T}{\sigma_X} \cdots\cdots\cdots\cdots\cdots\cdots\cdots (6)$$

ولكن سبق توضيح أن : معامل الثبات $\rho = \frac{\sigma_T^2}{\sigma_X^2}$ فبالتعويض في المعادلة 6 نجد أن : $\rho_{XT}^2 = \rho$ ؛ أي

أن $\rho_{XT} = \sqrt{\rho}$

ويعني هذا أنه يمكن تقدير معامل الارتباط بين العلامات الحقيقية والعلامات الملاحظة لمجموعة من الأفراد في اختبار ما بإيجاد الجذر التربيعي لمعامل ثبات علامات هذا الاختبار. ويسمى معامل الارتباط عندئذ مؤشر الثبات Reliability Index (Crocker,1986؛علام،2002).

الدراسات السابقة

هناك العديد من الدراسات السابقة ذات صلة بالموضوع ومنها دراسة الزبون(2004)، حيث هدفت الى المقارنة بين ثلاث طرق لضبط التخمين على الخصائص السيكومترية لاختبارات الاختيار من متعدد، وتكونت عينة الدراسة من(406) من طلبة الصف الثاني ثانوي علمي، طبق عليهم اختبار تحصيلي في مادة الرياضيات تكون من ثلاث صور متكافئة ، وصححت كل صورة من صور الاختبار بإحدى الطرق الثلاث (التقليدية – عقاب التخمين – الاختيار الجزئي)، وتوصلت نتائج الدراسة الى وجود فروق دالة احصائيا في معاملات الثبات بين استخدام معادلة التصحيح لأثر التخمين وطريقة الاختيار الجزئي ولصالح معادلة التصحيح لأثر التخمين عند مستوى دلالة (0.05) كما أظهرت نتائج الدراسة أن هناك فروق ذات دلالة إحصائية معاملات الصدق بين طريقة استخدام معادلة التصحيح لأثر التخمين وطريقة الاختيار الجزئي لصالح طريقة الاختيار الجزئي عند مستوى دلالة 0.05.

وفي دراسة ل جيو (Chiu,2010) اعتمدت على اسلوب التصحيح لاثر التخمين بالاعتماد على النموذج الثلاثي من خلال اجراءات (Mantel–Haenszel) والانحدار اللوجستي وذلك باستخدام تقريب (Taylor) وطرق الاحتمال الشرطي، استخدمت الدراسة بيانات مولده، واظهرت النتائج كفاءة نظرية استجابة الفقرة في عدم فقدان الكثير من مقدار المعلومات، بسبب استخدام تقريب تايلر كما خفضت نسبة التحيز بمقدار تراوح بين 6% الى 10%.

وهدفت دراسة راشد(2011) التعرف الى أثر التعليمات الخاصة بالتخمين في استجابة الطلبة لفقرات اختبار اختيار من متعدد على درجاتهم، وعلى الخصائص السيكومترية للاختبار (الصدق والثبات)، وعلى الخصائص السيكومترية للفقرات (الصعوبة والتمييز)، ولتحقيق أهداف دراسته تم بناء اختبار تحصيلي من نوع الاختيار من متعدد بأربع بدائل لكل فقرة في مادة الحضارة العربية الإسلامية، وتكون الاختبار من (40) فقرة، طُبق على (666) طالباً وطالبةً من الملتزمين وغير الملتزمين بتعليمات التخمين من كلية التربية الاسلامية في جامعة الكويت، وأظهرت نتائج الدراسة وجود فروق دالة إحصائياً بين درجات التحصيل للمجموعتين، وذلك لصالح مجموعة الملتزمين بتعليمات التخمين.

وسعت دراسة صمصم (2012) الى مقارنة ثلاث طرق لتصحيح اختبارات الاختيار من متعدد (الطريقة التقليدية، معادلة التصحيح لأثر التخمين، الاختيار الجزئي) لضبط أثر التخمين على مؤشرات ثبات الفقرات وصدق الاختبار، حيث

تكونت عينة الدراسة من (90) طالبة من طالبات الصف الخامس الأساسي، طبق عليهم اختبار تحصيلي، واظهرت نتائج الدراسة ارتفاع ثبات الاختبار عندما يكون التصحيح بالاختيار الجزئي، يليه التصحيح لأثر التخمين واخيرا الطريقة التقليدية.

وفي دراسة ل بيك (Paek,2015) لدراسة اثر التخمين على ثبات الاختبار وفقا للنموذج الثلاثي المعلمة في نظرية استجابة الفقرة، وقد اعتمدت الدراسة ثلاث اختباراتلبيانات محاكاة (20،40،60) فقرة موزعين على (9000) فرد واتوصلت نتائج الدراسة أن التخمين يزداد بزيادة درجة صعوبة السؤال ويقل معه معامل تمييز السؤال، وعندما تكون الصعوبة مساوية لقدرة الافرادفان التخمين يكون منخفضاواحيانا لا يؤثر كما اظهرت نتائج الدراسة عندما يرتفع التخمين ويقترب من (0.50) فان الثبات ينخفض بينما عندما ينخفض التخمين فان معامل الثبات يزدادبغض النظر عن انخفاض القدرة(−2) او ارتفاعها (2+)

وسعت دراسة للخرشة (2016) إلى الكشف عن أثر طرق معالجة أثر التخمين في تقدير قدرات الأفراد ومعالم الفرات في اختبارات الاختيار من متعدد. حيث قام الباحث ببناء اختبار تحصيلي من نوع الاختيار من متعدد تكون من 30 فقرة في مساق مبادئ القياس والتقويم التي تدرس لطلبة كلية التربية بجامعة مؤتة ، و تضمن الاختبار ثلاث صور حسب طرق المعالجة تم تطبيقه على عينة تكونت من (200) طالب وطالبة، وتم الاعتماد على برنامج (3) Mg–Bilog وبرنامج (SPSS) لتحليل البيانات بالاعتماد على النموذج اللوجستي ثنائي المعلمة، وتوصلت نتائج الدراسة الى عدد من النتائج منها انه لا يوجد فروق دالة احصائيا بين متوسطات معالم التمييز للفقرات تعزى لطريقة المعالجة. كما بينت نتائج الدراسة الى عدم وجود فروق دالة إحصائيا بين متوسطات معلمة القدرة للأفراد تعزى لطريقة المعالجة وتوصلت ايضا إلى وجود فروق ذات دلالة إحصائية بين متوسطات معالم الصعوبة للفقرات تعزى لطريقة المعالجة لصالح طريقة المعالجة الأولى(العقاب).

وهدفت دراسة السعودي والنصراوين(2018) إلى مقارنة ثلاث طرق لضبط أثر التخمين في اختبارات الاختيار من متعدد وهي طريقة التصحيح لأثر التخمين التقليدية، طريقة تصحيح المموهات، طرقة استخدام البديل " لا اعرف الاجابة"، من خلال أثرها على خصائص الاختبار، ولتحقيق اهداف الدراسة تم بناء اختبار تحصيلي في الرياضيات تم تطبيقة على عينة تكونت من (159) طالبا وطالبة، واشارت نتائج الدراسة الى وجود فروق ذات دلالة احصائية عند مستوى الدلالة ($\alpha \geq$ 0.05) بين طريقة تصحيح المموهات وطريقة التصحيح لأثر التهمين لصالح طريقة تصحيح المموهات وفق محكات، هي :

الأداء على الاختبار، الصدق ، الثبات، نظرا لإعتمادها على المعرفة الجزئية لدى الطلبة، في حين لم تظهر النتائج فروق ذات دلالة احصائية بين طريقتي التصحيح لأثر التخمين وطريقة البديل لا أعرف الإجابة.

وكشفت دراسة الزهراني (2021) اثر عدد البدائل في اختبار الاختيار من متعدد على تقدير قدرة الفرد، ومعالم الفقرات وفقا للنظرية التقليدية والنموذج ثلاثي المعلمة، ولتحقيق اهداف الدراسة تم بناء اختبار تحصيلي في مادة الرياضيات تكون من (39) فقرة في مقرر الرياضيات للصف الثالث الثانوي تكون من ثلاثة نماذج متكافئة تماما عدا عدد البدائل (خمسة بدائل اربعة بدائل ، ثلاثة بدائل) تم تطبيقه على (3000) طالب. واظهرت نتائج الدراسة عدم وجود فروق ذات دلالة احصائية يعزى لعدد البدائل بين نماذج الاختبار الثلاثة في تقدير قدرات الافراد في ضوء النوذج الثلاثي الملعلمة، بينما يوجد اثر لعدد البدائل في تقدير معالم الصعوبة والتمييز والتخمين بين النماذج الثلاثة لصالح النموذج ذي الخمسة بدائل، اذ انه كلما زاد عدد البدائل زادت قيمة معامل الصعوبة ومعامل اتمييز وانخفض التخمين.

يتضح من خلال استعراض الدراسات السابقة أن كل طريقة من طرق ضبط أثر التخمين في اختبارات الاختيار من متعدد تمتعت بخصائص مختلفة، كما أنه لا يوجد اتفاق في نتائج الدراسات حول أي الطرق أكثر فاعلية كما ان معظم الدراسات كانت وفق التقليدية للقياس وان الدراسات التي كانت وفق النظرية الحديثة لم تدرس اثر التخمين على دالة معلومات افقرات والاختبار.

الطريقة والإجراءات:

مجتمع الدراسة:

تكون مجتمع الدراسة من طلبة التأهيل التربوي في جامعة نزوى في سلطنة عمان والبالغ عددهم (1214) طالبا وطالبة .

عينة الدراسة

تكونت عينة الدراسة من (838) طالبا وطالبة تم اختيارهم بالطريقة العشوائية البسيطة من مجتمع الدراسة.

أداة الدراسة

لجمع البيانات الخاصة بالدراسة بهدف الاجابة عن اسئلة الدراسة تم بناء اختبار تحصيلي في مساق القياس النفسي والتقويم التربوي المطروح للطلبة على الفصل الدراسي الأول للعام الجامعي (2023–2024) وفقا للخطوات الآتية وبالاعتماد

على الخطوات التي أوردها غرونلند ولين (Gronlund & Linn,1990):

أولاً: تحديد الغرض من الاختبار وهو قياس تحصيل الطلاب في مساق القياس النفسي والتقويم التربوي.

ثانيا: كتابة أهداف المحتوى ووبناء جدول المواصفات، تم فيه ربط مستويات الأهداف بمحتوى المادة الدراسية موضوع الاختبار.

ثالثا: كتابة (44) فقرة من نوع الاختيار من متعدد لكل فقرة منها اربعة بدائل أحدهما الإجابة الصحيحة، وقد روعيت فيها الأسس الفنية في كتابة هذا النوع من الفقرات، وملائمتها للهدف الذي تقيسه من حيث المحتوى والمستوى المعرفي.

رابعا: عرض الاختبار بصورته الأولية على مجموعة من المحكمين من أعضاء هيئة التدريس في كلية التربية في جامعة الحسين بن طلال وكلية العلوم والاداب قسم التربية والدراسات الانسانية في جامعة نزوى تخصص القياس والتقويم، وعلم النفس التربوي والمناهج العامة ، وطلب منهم إبداء أرائهم حول درجة ارتباط الفقرات بموضوع الدراسة، ومدى وضوح الصياغة اللغوية. لإبداء ملاحظاتهم على فقرات الاختبار من حيث وضوح الفقرات ودقتها والصياغة اللغوية، وشمولية فقرات الاختبار وإضافة ما يرونه مناسبا من تعديلات ومقترحات، وتم إعتماد معيار اتفاق 80% من آراء المحكمين. وبعد استرجاع الاختبار من المحكمين ومراجعة ملاحظات المحكمين وبناء على الملاحظات التي تفضل بها الزملاء المحكمون تم التعديل على بعض الفقرات وحذف (3) فقرات ليتكون الاختبار من (41) فقرة

خامسا: تطبيق الاختبار على عينة استطلاعية تكونت من (36) طالبا وطالبة والتأكد من الخصائص السيكومترية لفقرات الاختبار.وبالاعتماد على قيم معاملات الصعوبة والتمييز الناتجة فقد تم حذف (6) فقرات من فقرات الاختبار ذات معاملات التمييز المنخفض (أقل من 0.19) و الفقرات ذات معاملات الصعوبة المتدنية (أقل من 0.2) وذات معاملات الصعوبة المرتفعة (أعلى من 0.8) وبذلك أصبح الاختبار بصورته النهائية مكون من (35) فقرة . والجدول رقم(1) يبين قيم معاملات الصعوبة والتمييز لفقرات الاختبار.

جدول رقم (1) : معاملات الصعوبة والتمييز لفقرات الاختبار

رقم الفقرة	معامل الصعوبة	معامل التمييز	رقم الفقرة	معامل الصعوبة	معامل التمييز
1	0.75	0.35	**22**	0.31	0.32
2	0.61	0.32	**23**	0.36	0.41
3	0.47	0.38	**24**	0.35	0.66
4	0.39	0.48	**25**	0.44	0.51
5	0.28	0.28	**26**	0.42	0.65
6	0.50	0.33	**27**	0.53	0.33
7	0.31	0.27	**28**	0.23	0.39
8	0.36	0.34	**29**	0.61	0.22
9	0.31	0.06	**30**	0.21	0.15-
10	0.33	0.22	**31**	0.24	0.53
11	0.19	0.02	**32**	0.78	0.29
12	0.36	0.34	**33**	0.42	0.28
13	0.28	0.50	**34**	0.86	0.19-
14	0.22	0.05	**35**	0.58	0.24
15	0.39	0.23	**36**	0.34	0.24-
16	0.36	0.42	**37**	0.25	0.27
17	0.34	0.20	**38**	0.56	0.32
18	0.39	0.35	**39**	0.47	0.27
19	0.47	0.43	**40**	0.44	0.42
20	0.33	0.17	**41**	0.41	0.64
21	0.56	0.34			

سادسا: التأكد من الخصائص السيكومترية للاختبار بعد تطبيقه على أفراد عينة الدراسة الاستطلاعية وذلك بالاعتماد على كرونباخ الفا وطريقة التجزئة النصفية، وللتخلص من أثر التجزئة النصفية تم الاعتماد على معادلة سبيرمان بروان، وقد كانت قيمة معامل الثبات بالاعتماد على الطريقتين وعلى التوالي النصفية (0.81 ، 0.86) وهذا يدل على أن الاختبار يتمتع بدرجة عالية من الثبات مما يطمئن الباحث لتطبيقه على عينة الدراسة.

سابعا: إعداد ورقتي تعليمات الاختبار بحيث تضمنت الورقة الأولى على تعليمات التصحيح لأثر التخمين والورقة الثانية على تعليمات الطريقة التقليدية بالاضافة الى ورقة الإجابة النموذجية لكل نموذج من نموذجي الاختبار.

ثامنا: تطبيق الاختبار بصورته النهائية على مجموعتين من الطلبة بحيث تراعي المجموعة الأولى الاجابة وفق طريقة التصحيح من أثر التخمين والبالغ عددها (419) والمجموعة الثانية وفق الطريقة التقليدية والبالغ عددها (419) .

تاسعا: جمع استجابات عينة الدراسة وإدخالها بالاعتماد على البرنامج الاحصائي (SPSS).

عاشرا: التحقق من افتراضات نظرية الاستجابة للفقرة.

1- افتراض أحادية البعد Unidimensionality:

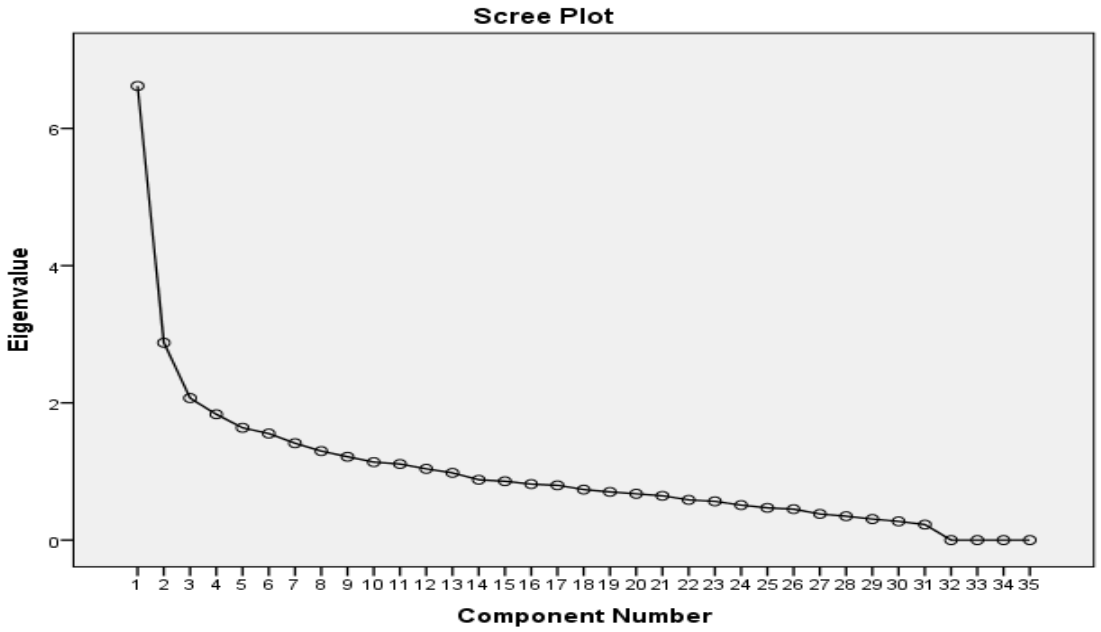
تم التحقق من افتراض أحادية البعد بالاعتماد على التحليل العاملي الاستكشافي لاستجابات المفحوصين (Factor Analysis) بطريقة المكونات الرئيسية (Principal Component) والاعتماد على طريقة (Varimax) لتدوير المحاور. ويظهر في الجدول (2) قيمة الجذر الكامن (Eigen Value) ونسبة التباين المفسر (Explained Variance) وكذلك التباين المفسر التراكمي لكل عامل من العوامل:

جدول (2): قيم الجذور الكامنة ونسبة التباين المفسر والتراكمي للاختبار.

العامل	الجذر الكامن	نسبة التباين المفسر	نسبة التباين المفسر التراكمي	الجذر الكامن الأول / الجذر الكامن الثاني
1	6.62	18.91	18.91	2.31
2	2.87	8.20	27.12	
3	2.070	5.91	33.04	
4	1.83	5.23	38.27	
5	1.63	4.67	42.95	
6	1.55	4.43	47.39	
7	1.41	4.03	51.42	

يتضح من الجدول رقم (2) أن العامل الأول يفسر أكبر نسبة تباين مقارنة ببقية العوامل. كما يتبين أن ناتج قسمة الجذر الكامن للعامل الأول على الثاني أكبر من (2). مما يدل على وجود سمة سائدة للاختبار، وبالتالي يمكن القول بتحقق افتراضية أحادية البعد للاختبار (Hattie, 1985).

كما تم الاعتماد على الرسم البياني باستخدام (Scree Plot) للعوامل مع الجذور الكامنة والشكل التالي يوضح ذلك.



الرسم البياني لقيم الجذور الكامنة للعوامل.

يتضح من التمثيل السابق أن العامل الأول يفسر أكبر نسبة من التباين مقارنة مع العوامل الأخرى ، وهذا يعتبر أيضًا مؤشرًا على أحادية البعد.

2- افتراض الاستقلال الموضعي Local Independence:

ينص هذا الافتراض أن إجابة المفحوص عن الفقرة لا تتأثر بإجابته عن فقرة أخرى والعامل الوحيد المؤثر على احتمال الإجابة عن الفقرة هو مقدار القدرة (θ) وخصائص الفقرة.وللتحقق من هذا الافتراض تم اعتماد أن تَحقُّق افتراض أحادية البُعد لفقرات الاختبار يُعَدّ دليلًا على تَحقُّق افتراض الاستقلال الموضعي ، إذ أن هذا الافتراض مكافئ لافتراض أحادية البعد، (Hambleton & Swaminathan, 1985).

3- التحرر من السرعة في الأداء:

يمكن التأكد من أن الاختبار ليس اختبار سرعة من خلال فحص نسبة المفحوصين الذين أكملوا الاختبار، وكذلك فحص الفقرات التي لم يجب عنها الطلاب. ووفقًا لما يراه هامبلتون وآخرون (Hambelton et al., 1991) فإنه إذا كان (75%) من الطلاب قد أكملوا الإجابة عن الاختبار، وإذا كان (80%) من فقرات الاختبار قد تمت الإجابة عنها من قبل الطلاب، فإن السرعة لن تعد في هذه الحالة عاملًا مهمًا في الأداء على الاختبار. وحيث إن نسبة الأفراد الذين أكملوا هذا

الاختبار هي (100%)، وأن نسبة الفقرات التي تمت الإجابة عنها أعلى من (90%) فإن هذا يعني أن الاختبار يقيس القوة وليس السرعة.

احدى عشر : التحقق من مطابقة الأفراد والفقرات للنموذج اللوجستي أحادي المعلمة:

استخدم برنامج (Bilog-MG3)؛ للتأكد من مطابقة الأفراد (Person-Fit)، ومطابقة الفقرات (Item-Fit) للنموذج اللوجستي أحادي المعلمة باستخدام اختبار مربع كاي (χ^2) عند مستوى دلالة ($\alpha = 0.01$)، وقد أظهرت نتائج التحليل عدم مطابقة استجابات ثمانية أفراد للنموذج أحادي المعلمة، حيث كانت القيم الاحتمالية لهم أقل من (0.01)؛ لذا تم حذف استجابتهم .

كما تم فحص مطابقة الفقرات للنموذج المستخدم حيث أُعيد التحليل باستخدام برنامج (Bilog-MG3) بعد حذف استجابة الافراد غير المطابقين للنموذج بناءً على مؤشر مربع كاي (χ^2) عند مستوى دلالة ($\alpha = 0.01$)؛ وأظهرت نتائج التحليل مطابقة جميع فقرات الاختبار للنموذج احادي المعلمة، حيث كانت القيمة الاحتمالية أكبر من (0.01) لجميع الفقرات.

نتائج الدراسة ومناقشتها:

<u>نتائج الدراسة المتعلقة بالسؤال الأول: هل يختلف معامل ثبات الاختبار باختلاف طريقة التصحيح (معادلة التصحيح لاثر التخمين، التقليدية) ؟</u>

للاجابة على هذا السؤال تم الإعتماد على النموذج أحادي المعلمة (نموذج) راش وتم ايجاد الثبات المتعلق بالأفراد والثبات المتعلق بفقرات الاختبار بفمهوم الثبات في إطار نموذج راش يشير إلى مدى الدقة في تقدير موقع كل من الفقرات والأفراد على متصل السمة، الذي نهدف إلى قياسه.

ويمكن تحديد مدى دقة الفقرات في تعريف المتصل وذلك بحساب النسبة بين الانحراف المعياري للقيم التقديرية المتحررة من صعوبة الفقرات ومتوسط الخطأ المعياري لهذه القيم، ويطلق على هذه النسبة معامل الفصل بين الفقرات (G_I) والذي يجب أن تكون قيمته أكبر ما يمكن، فإذا لم تزد قيمته عن (2) فانه يصعب قياس المتغير بواسطة هذه الفقرات، وقد بينت نتائج الدراسة أنه في حال استخدام الطريقة التقليدية في التصحيح أن معامل الفصل بين الفقرات في الاختبار بصورته النهائية (5.20) وتعد هذه القيمة جيده وتزيد عن (2) لذلك فإن فقرات الاختبار تعتبر كافية لتعريف متصل السمة الذي

تقيسه، وكذلك تبين أن قيمة معامل الفصل بين الافراد (G_P) تساوي(2.33)، وترمز هذه القيمة إلى النسبة بين الانحراف المعياري للقيم التقديرية المتحررة لقدرات الافراد، ومتوسط الخطأ المعياري لهذه القيم، وبما ان هذه القيمة أكبر من (2) فإن عينة الأفراد كافية للفصل بين الفقرات.

وتم حساب معامل الثبات المتعلق بالافراد من خلال العلاقة (G_P^2 +1)/ G_P^2 فكانت قيمته (0.85)، وتعتبر قيمة مقبولة وتدل على أن عينة الأفراد كافية للفصل بين الفقرات والتمييز بين مستويات الصعوبة للفقرات، اي ان الافراد يتوزعون بشكل مناسب على متصل السمة الممثل بفقرات الاختبار التي بنيت وفق نموذج اش، وتم حساب معامل الثبات للفقرات بالاعتماد على المعادلة (G_I^2 +1)/ G_I^2 فكانت قيمته (0.96) عند المستويات المختلفة للقدرة وهي قيمة مرتفعة وتدل على أن عينة الفقرات كافية للفصل بين الافراد بالتالي التمييز بين مستويات التحصيل لهؤلاء الأفراد.

وبينت نتائج الدراسة أنه في حال استخدام طريقة التصحيح لأثر التخمين في التصحيح أن معامل الفصل بين الفقرات في الاختبار بصورته النهائية (8.16) وتعد هذه القيمة جيده وتزيد عن (2) لذلك فإن فقرات الاختبار تعتبر كافية لتعريف متصل السمة الذي تقيسه، وكذلك تبين أن قيمة معامل الفصل بين الافراد (G_P) تساوي(3.04)، وترمز هذه القيمة إلى النسبة بين الانحراف المعياري للقيم التقديرية المتحررة لقدرات الافراد، ومتوسط الخطأ المعياري لهذه القيم، وبما ان هذه القيمة أكبر من (2) فإن عينة الأفراد كافية للفصل بين الفقرات.

وتم حساب معامل الثبات المتعلق بالافراد من خلال العلاقة (G_P^2 +1)/ G_P^2 فكانت قيمته (0.90) ، وتعتبر قيمة مقبولة وتدل على أن عينة الأفراد كافية للفصل بين الفقرات والتمييز بين مستويات الصعوبة للفقرات، اي ان الافراد يتوزعون بشكل مناسب على متصل السمة الممثل بفقرات الاختبار التي بنيت وفق نموذج اش، وتم حساب معامل الثبات للفقرات بالاعتماد على المعادلة (G_I^2 +1)/ G_I^2 فكانت قيمته (0.98) عند المستويات المختلفة للقدرة وهي قيمة مرتفعة وتدل على أن عينة الفقرات كافية للفصل بين الافراد بالتالي التمييز بين مستويات التحصيل لهؤلاء الأفراد.

ويلاحظ من النتائج أن معامل الثبات في حال استخدام معادلة التصحيح لأثر التخمين أعلى منه في حالة استخدام الطريقة التقليدية. وربما يعود السبب في ذلك إلى أنه في حالة استخدام معادلة التصحيح لأثر التخمين سيكون إعتماد المفحوصين على التخمين قليل مما يقلل من أخطاء القياس وبالتالي الدقة في تقدير الدرجة الحقيقية للمفحوص مما يزيد من قيم معامل الثبات للاختبار.

نتائج الدراسة المتعلقة بالسؤال الثاني: هل تختلف دالة معلومات الفقرات باختلاف طريقة التصحيح (معادلة التصحيح لاثر التخمين، التقليدية)؟

للاجابة على هذا السؤال تم الإعتماد على النموذج أحادي المعلمة (نموذج) راش وبرمجية (Bilog-mg3) لأيجاد دوال المعلومات الخاصة بالفقرات ثم **استخدام** اختبار ت للعينات المستقلة (Independent Samples Test) لمعرفة دلالة الفروق في دالة معلومات الفقرا باختلاف طريق التصحيح، والجدول(3) يبين هذه النتائج.

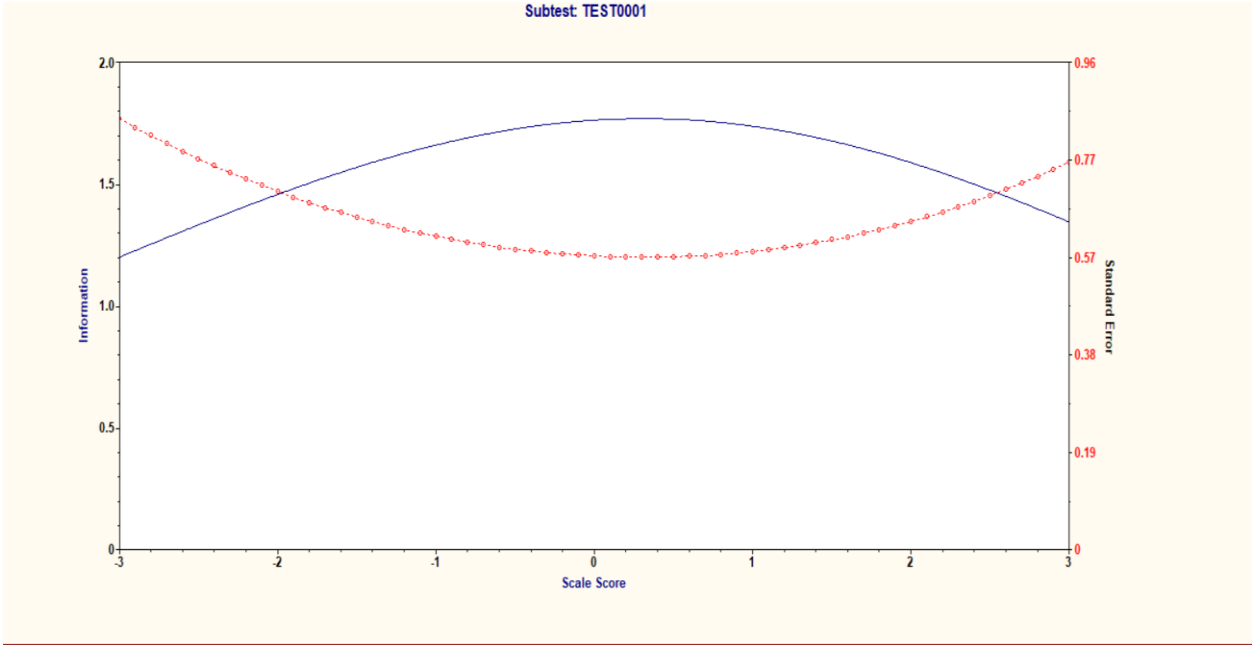
جدول (3): المتوسطات الحسابية والانحرافات المعيارية لدوال المعلومات لفقرات الاختبار وفقا لطريقة التصحيح

طريقة التصحيح	المتوسط الحسابي	الانحراف المعياري	قيمة "ت"	درجات الحرية	الدلالة الإحصائية
التقليدية	0.051	0.007	36.05	68	0.00
معادلة التصحيح لأثر التخمين	0.10	0.009			

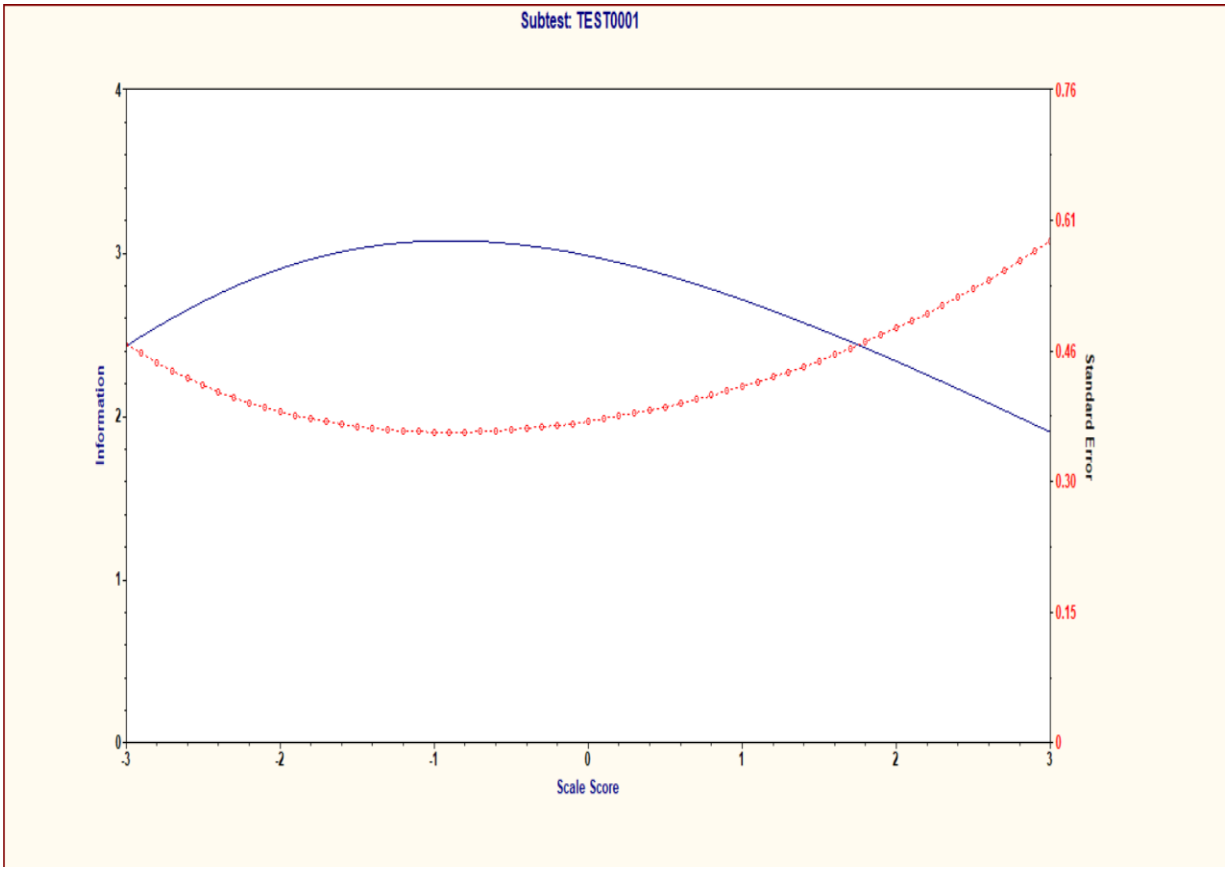
يتضح من الجدول (3) وجود فروق ذات دلالة إحصائية ($\alpha \leq 0.05$) بين المتوسطات الحسابية لدوال المعلومات الخاصة بالفقرات ولصالح طريقة معادلة التصحيح لأثر التخمين مما يؤكد أن متوسطات دالة المعلومات للفقرات وفق طريقة معادلة التصحيح لأثر التخمين أفضل من متوسط دوال المعلومات للطريقة التقليدية.

نتائج الدراسة المتعلقة بالسؤال الثالث :هل تختلف دالة معلومات الاختبار باختلاف طريقة التصحيح (معادلة التصحيح لاثر التخمين، التقليدية)؟

للاجابة على هذا السؤال تم الاعتماد على برمجية (Bilog-mg3) لايجاد المنحنى الخاص بدالة المعلومات وفقا للطريقتين والمنحنيين التاليان يوضحان ذلك.



شكل 1: منحنى دالة المعلومات الخاص الطريقة التقليدية



شكل 2: منحنى دالة المعلومات الخاص بطريقة التصحيح لأثر التخمين

يتضح من الشكلين السابقين أن كمية المعلومات التي يقدمها الاختبار بالاعتماد على معادلة التصحيح لأثر التخمين كانت تقريبا (3.4) وهي أكبر من كمية المعلومات التي يقدمها الاختبار بالاعتماد على الطريقة التقليدية والتي كانت تقريبا (1.8). مما يؤكد أن دالة معلومات الاختبار أفضل ف حال استخدام معادلة التصحيح لأثر التخمين وربما يعود السبب في ذلك إلى أنه في حالة استخدام معادلة التصحيح لأثر التخمين سيكون إعتماد المفحوصين على التخمين قليل مما يقلل من الخطأ المعياري في التقدير حيث أن العلاقة بين دالة معلومات الاختبار والخطأ المعياري للتقدير علاقة عكسية أي أنه كل ما قل الخطأ المعياري في التقدير زادت قيمة دالة المعلومات.

نتائج الدراسة المتعلقة بالسؤال الرابع: ما أثر التخمين على احتمال الاجابة الصحيحة عن الفقرة ؟

للاجابة على هذا السؤال تم الإعتماد على النموذج أحادي المعلمة (نموذج) راش وبرمجية (Bilog-mg3) لأيجاد احتمال الاجابة الصحيحة على الفقرات (p_i (θ)) عند أعلى مستوى لقدرات المفحوصين (θ =3) وفقا للمعادلة:

$$P_i(\theta) = \frac{e^{D(\theta - b_i)}}{1 + e^{D(\theta - b_i)}}, \quad i=1,2,3,...,n \quad .....(1)$$

وبالاعتماد على القيم الناتجة تم استخدام اختبار ت للعينات المستقلة (Independent Samples Test) لمعرفة دلالة الفروق في القيم الخاصة باحتمال الاجابة الصحيحة على الفقرة باختلاف طريقة التصحيح، والجدول(4) يبين هذه النتائج.

جدول (4): المتوسطات الحسابية والانحرافات المعيارية لاحتمال الاجابة الصحيحة على الفقرة وفقا لطريقة التصحيح

طريقة التصحيح	المتوسط الحسابي	الانحراف المعياري	قيمة "ت"	درجات الحرية	الدلالة الإحصائية
التقليدية	0.76	0.15	0.89	68	0.38
معادلة التصحيح لأثر التخمين	0.81	0.26			

يتضح من الجدول (4) عدم وجود فروق ذات دلالة إحصائية ($\alpha \geq 0.05$) بين المتوسطات الحسابية القيم الخاصة لاحتمال الاجابة الصحيحة على الفقرة وفقا لطريقة التصحيح، وربما يعود السبب في ذلك الى أن القدرة كانت متساويه لجميع المفحوصين عند استخدام كل من الطريقتين.

التوصيات:

بناء على النتائج التي توصلت اليها الدراسة يوصي الباحث بما يلي:

1. اعتماد طريقة التصحيح لأثر التخمين عند تصحيح الاختبارات التحصيلية من نوع الاختيار من متعدد.
2. دراسة دالة معلومات الاختبار عند تحليل نتائج الاختبارات.
3. اجراء المزيد من الدراسات بالاعتماد على النموذج ثنائي الثلاثي المعلمة.

المصادر والمراجع

المراجع العربية:

- 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). المجلد 30(12)، 22-99.*

- الزبون، حابس. (2004). *أثر استخدام ثلاث طرق للتصحيح لضبط أثر التخمين في الخصائص السيكومترية الاختبارات الاختيار من متعدد في ضوء نظرية الاستجابة للفقرة*، أطروحة ماجستير غير منشورة، جامعة مؤته، الكرك، الأردن.

- الزهراني، حسين (2021) . اثر عدد البدائل لاختبار الاختيار من متعدد على تقدير قدرة الفرد ومعالم الفقرة في كل من النظرية التقليدية والنموذج ثلاثي المعلمة في نظرية الاستجابة للمفردة، رسالة دكتوراه غير منشورة، قسم علم النفس، كلية التربية، جامعة الملك سعود.

- السعودي والنصراوين (2018). مقارنة ثلاث طرق ضبط أثر التخمين في اختبارات الاختيار من متعدد وأثرها على خصائص الاختبار، *مجلة دراسات*، العلوم التربوية، 45(4)، ملحق 6، ص ص 597-615.

- صمصم، ميساء (2012) اثر طرق التصحيح لضبط التخمين في اختبارات الاختيار من متعدد على كل من مؤشر ثبات الفقرة ومؤشر صدقها ، رسالة ماجستير غير منشورةظن جامعة مؤته، الاردن.

- علام، صلاح الدين. (2002). *القياس و التقويم التربوي والنفسي أساسياته وتطبيقاته وتوجيهاته*. دار الفكر العربي: القاهرة .

- الكيلاني، عبدالله .(1994). *التقويم التربوي واختبارات التحصيل*. الأنروا / اليونسكو. دائرة التربية والتعليم، معهد التربية، عمان.

- الكيلاني، عبدالله وعدس ، عبدالرحمن. (1994). *القياس والتقويم*. جامعة القدس المفتوحة: عمان، الأردن.

- الكيلاني، عبدالله وعدس، عبد الرحمن (1986). *القياس والتقويم في التربية وعلم النفس*. مركز الكتاب الأردني، الأردن، عمان.

- النجار، نبيل (2010). *القياس والتقويم منظور تطبيقي مع تطبيقات برمجية SPSS*. الأردن، عمان: دار الحامد للنشر والتوزيع.

المراجع الأجنبية:

- Al-Kilani, A & Adas, A. (1986). *Measurement and evaluation in education and psychology*. Jordanian Book Center, Jordan, Amman.

- Al-Kilani, A & Adas, A. (1994). *Measurement and evaluation*. Al-Quds Open University: Amman, Jordan.

- Al-Kilani, A (1994). *Educational evaluation and achievement tes*ts. UNRWA/UNESCO. Department of Education, Institute of Education, Amman

- Alkrusheh, T.(2016). The Effect of processing Guessing Method on the Estimating Item person statistics for Multiple choice Test According to Item Response Theory (IRT), *An-Najah University Journal for Research - B (Humanities)*,30(12).2348-2366.

- Allam, Saladin. (2002). *Educational and psychological measurement and evaluation, its basics, applications and directions*. Dar Al-Fikr Al-Arabi: Cairo.

- Al-Najjar, N (2010). *Measurement and evaluation applied perspective with SPSS software applications*. Jordan, Amman: Dar Al-Hamid for Publishing and Distribution.

- Al-Saudi,S & Al-Nasraween,M. (2018). Comparison of three methods for controlling the effect of guessing in multiple choice tests and their impact on test characteristics, *Dirasat Journal*, Educational Sciences, 45(4), Supplement 6, 597-615.

- Al-Zahrani, H. (2021). The effect of the number of alternatives for a multiple-choice test on estimating an individual's ability and item parameters in both the classical theory and the three-parameter model in item response theory, *unpublished PHD dissertation*, Department of Psychology, College of Education, King Saud University.

- Al-zboon, H.(2004). The effect of using Three Methods For Scoring To Control The Effect of

Guessing on The Psychometric Properties of Multiple choice Tests Upon The item Response Theory. *unpublished master's Thesis*, Mutah University, Jordan.

- Chiu, T. (2010). Correction for guessing in the framework of the 3PL item response theory (Order No. 3418414). Available from ProQuest Dissertations & Theses Global. (751246969)

- Crocker, L., & Algina, J. 1986. *Introduction to classical and modern test theory*. New York: Holt, Rinehart and Winston.

- Embretson, S.,& Reiase, S. (2000).“Item Response Theory forPsychologists”. New jersey: Lawrence Erlbaum Associates. Inc.

- Gronlund, N.E. and Linn, R.L. (1990) *Measurement and Evaluation in Teaching*. McMillan Company, New York.

- Hambleton, K.R., & Swaminathan .(1991). Fundamentals of Item Response Theory: Sage Publications.

- Hambleton, R., & Swaminathan, H. (1985). Item Response Theory: Principles and Applications. Kluwer· Nijh off Publishing.

- Hambleton,R.k., & Swaminathan,H, &Rogers. H.j.(1993). Fundamentals of Item Response Theory :International Educational and Professional. Publisher Newbury park

- Hattie, J (1985). Methodology Review: Assessing Unidimensionality of Tests and Items. Psychological Measurement, 9(2), 139-164.

- McMillan, James H. (2013). *Research on Classroom Assessment*. Sage Publication, USA.

- Merrit, Stephen. (2006). *Mastering Multiple Choice: The Definitive Guide to Better Grades on Multiple Choice Exams*. Brain Ranch, Canada.

- Messick ،S. (1995). Validation of Psychological Assessment: Validation of Inferences from person's response and performances as scientific inquiry into scoring meanings. *American Psychologist*, 50, 741-749.

- Paek I. (2015). An Investigation of the Impact of Guessing on Coefficient α and Reliability. Applied psychological measurement, 39(4), 264–277.

- Reif, Manuel. Holocher-Ertl, Stefana. Klaus D. Kubinger, and Frebort, Martina. (2010). On Minimizing Guessing Effects on Multiple-Choice Items: Superiority of two solutions

and three Distractors item format to a one Solution and Five Distractors Item Format. *International Journal of Selection and Assessment*. 18, (1): 110-115.

- Samsam, M (2012) The effect of correction methods to control guessing in multiple choice tests on both the item reliability index and the validity index*, unpublished master's thesis*, Mutah University, Jordan.