The Effect of Multiple Choice Test Modality (Paper, Electronic) on the Test Psychometric Properties and the Item Parameters of the Test Utilizing Item Response Theory

Mahdi M. Olimat[(1)*] **Zainab S. Al Shdefat**[(2)]

(1) Educational Advisor, .Ministry of Education, Jordan.
(2) Educational Advisor, .Ministry of Education, Jordan.

Received: 23/03/2024
Accepted: 20/04/2024
Published: 30/12/2024

* *Corresponding Author:*
drmahdiolimat@gmail.com

DOI:https://doi.org/10.59759/educational.v3i4.837

Abstract

The objective of this study was to investigate the effect of multiple choice test modality (paper, electronic) on the test psychometric properties and the item parameters of the test utilizing item response theory. To achieve this objective, an achievement test was constructed consisting of (25) items of multiple choice with four alternatives. The test was applied on the two samples of the study (a sample got a paper test, while the other got an electronic one). Each sample had (500) examinees (250 examinees for each test form) from the eight basic grade in the Directorate of Education Kasabat Al-Mafraq in the second semester of the academic year 2023/2024.). The assumptions of the Item Response Theory were verified: unidimensionality and local independence for the two test forms. The data was analyzed using BILOG-MG3, and the analysis showed that item 19 in the paper test and item 22 in the paper test did not fit the model used (three-parametric model), so that the test in its final form consisted of (23) items. The results showed that there were no statistically significant differences due to the test form (paper, electronic) . It was also shown that the difficulty value of the information function provided by the test presented electronically is higher than the test presented on paper. It was also found that there were no statistically significant differences in empirical reliability coefficients between the two methods of presenting the test, and that there were no statistically significant differences at the significance level ($\alpha = 0.05$) between the averages of the parameters of difficulty, discrimination, and guess, according to the three-parameter model, due to the difference in the method in which the test was presented. The researchers recommended presenting the teste electronically as it has some qualities such as saving time and effort, as well as the efficient psychometric properties for the test and the items.

Keywords: Item Response Theory, Test Psychometric Properties of the Test, Item Parameters.

A Special Issue on the Conference on Learning and Teaching in the Digital Age.

أثر طريقة تقديم اختبار الاختيار من متعدد (ورقي، إلكتروني) على الخصائص السيكومترية للاختبار وعلى معالم فقرات الاختبار وفقا نظرية الاستجابة للفقرة

مهدي محمد عليمات[١] زينب سالم الشديفات[٢]

(١) مشرف تربوي، وزارة التربية والتعليم، الأردن.

(٢) مشرف تربوي، وزارة التربية والتعليم، الأردن.

ملخص

هدفت هذه الدِّراسة إلى التحقق من أثر طريقة تقديم اختبار الاختيار من متعدد (ورقي، إلكتروني) على الخصائص السيكومترية للاختبار وعلى معالم فقرات الاختبار وفقاً نظرية الاستجابة للفقرة. ولتحقيق هذا الهدف تم بناء اختبار في اللغة الإنجليزية للصف الثامن الأساسي مكوَّن من (٢٥) فقرة من نوع الاختيار من متعدد بأربعة بدائل، وطُبِّق الاختبار على عينتي الدِّراسة (عينة تم تقديم الاختبار ورقياً، والأخرى إلكترونياً) وبلغ حجم كل عينة (500) مفحوصاً، بواقع (٢٥٠ مفحوصاً لكل صورة من صور الاختبار) من طلبة الصف الثامن الأساسي في المدارس الحكومية التَّابعة لمديرية تربية وتعليم لواء قصبة المفرق في الفصل الدِّراسي الثاني للعام الدِّراسي ٢٠٢٣/٢٠٢٤م. وتم التحقق افتراضات نماذج نظرية الاستجابة للفقرة: أُحادية البُعد (Unidimensionality)، والاستقلال الموضعي (Local Independence) لصورتي الاختبار. وحللت الفقرات باستعمال البرنامج المحوسب BILOG-MG3، حيث أظهرت نتائج التحليل عدم مطابقة الفقرة ١٩ في الاختبار الورق والفقرة رقم ٢٢ في الاختبار الورقي للنَّموذج المستخدم (ثلاثي المعلمة)، ليصبح الاختبار بِشكله النِّهائي مكوَّناً من (٢٣) فقرة. وأظهرت النتائج عدم وجود فروق ذات دلالة إحصائية بين متوسطات معالم الصعوبة والتمييز والتخمين باختلاف طريقة تقديم الاختبار (ورقي، إلكتروني)، كما تبين أن قيمة دالة المعلومات التي يقدمها الاختبار المقدم إلكترونياً أعلى من الاختبار المقدم ورقياً. وتبين أيضاً عدم وجود فروق ذات دلالة إحصائية في معاملات الثبات الامبريقي بين طريقتي تقديم الاختبار. كما تبين عدم وجود فروق ذات دلالة إحصائية عند مستوى الدلالة (α=٠.٠٥) بين متوسطات معالم الصعوبة والتمييز والتخمين، وفق النموذج ثلاثي المعلمة يعزى لاختلاف الطريقة التي تم فيها تقديم الاختبار. وأوصى الباحثان باستخدام طريقة تقديم الاختبار الإلكترونية لما لها من مميزات كتوفير الوقت والجهد ولما أظهرته من خصائص سيكومترية جيدة للاختبار والفقرات.

الكلمات المفتاحية: نظرية الاستجابة للفقرة، الخصائص السيكومترية للاختبار، معالم الفقرات

مقدمة

يشير العمري (٢٠٠٥) إلى أن التطورات النوعية والسريعة التي حدثت في مجالات التكنولوجيا أدت إلى ظهور العديد من المستجدات التي دفعت العاملين في المجال التربوي إلى مواكبة هذه التطورات، بما في ذلك التكنولوجيا والتغيرات المنبثقة عنها. ونتيجة لذلك، ظهرت نظريات واستراتيجيات وأجهزة ومعدات قابلة للاستخدام وذات فائدة كبيرة للعملية التعليمية التعلمية؛ لما تتمتع به من ميزات فريدة تسهم في تحقيق أهداف التعليم والتعلم بوقت وجهد أقل، بالإضافة إلى دورها في تحسين اتجاه المتعلمين نحو العملية التعليمية التعلمية.

وقد أدركت العديد من المؤسسات التعليمية أهمية الاستفادة من التطور الهائل في تكنولوجيا المعلومات والاتصالات، ودورها في تعزيز الكفاءة والفعالية في العملية التعليمية التعلمية. وبات التطور نهجًا ضروريًا والتغيير أمرًا حتميًا لمختلف الأنظمة والمؤسسات في المجتمع.

يُعد التقويم أحد أهم عناصر المناهج؛ ولذلك أولى القائمون على بناء المناهج اهتمامًا كبيرًا بتحسين عملية التقويم في تعليم وتعلم المواد الدراسية، وتشجيع المعلمين على تطبيق أساليب حديثة إلكترونية لتقويم تحصيل الطلبة من خلال اختبارات تتسم بمواصفات جيدة.

وتُعد الاختبارات الإلكترونية من الاتجاهات الحديثة في التقويم التربوي، والتي يمكن من خلالها تقييم مخرجات العملية التعليمية باستخدام الحاسب الآلي. وتمتاز هذه الاختبارات بالمرونة والكفاءة في جميع مراحل تطبيقها، إضافة إلى قدرتها على الاحتفاظ بالسجلات والبيانات لمدة طويلة، فضلاً عن سهولة مراجعتها، واكتشاف الأخطاء، وتصحيح بنودها عند الحاجة (Bennett, 1999).

بذل القائمون على بناء الاختبارات والمقاييس جهودًا كبيرة للحصول على فقرات اختبارية عالية الجودة تقيس تحصيل المفحوصين بطريقة موضوعية، مما يضمن الحصول على اختبارات ومقاييس تتسم بخصائص سيكومترية جيدة.

وقد أدى هذا السعي إلى ظهور نظريات تهدف إلى بناء اختبارات ومقاييس تحقق معايير عالية الجودة. وسعت هذه النظريات إلى إيجاد مؤشرات إحصائية دقيقة لقياس مدى دقة وملاءمة أدوات القياس للأهداف المراد قياسها، مما يسهم في اتخاذ قرارات صائبة بناءً على نتائج هذه المقاييس. ومن بين هذه النظريات: **النظرية التقليدية في القياس** (Classical Measurement Theory)، التي تعرضت لاحقًا لانتقادات عديدة لقصورها عن تفسير بعض القضايا في القياس، الأمر الذي

أدى إلى ظهور **نظرية الاستجابة للفقرة** (Item Response Theory) كبديل للنظرية الكلاسيكية.

تُعد نظرية الاستجابة للفقرة من أهم إنجازات علم القياس؛ إذ تتميز بالعديد من الخصائص، من أبرزها أن تقدير معالم الفقرات باستخدامها لا يتأثر بخصائص المفحوصين الذين طبق عليهم الاختبار **(Person-Free)**، كما أن تقدير قدرات المفحوصين لا يتأثر بخصائص الفقرات المقدمة لهم **(Item-Free)**.

تفترض نظرية الاستجابة للفقرة إمكانية التنبؤ بأداء الأفراد المفحوصين من خلال خاصية أو خصائص مميزة تُفسر هذا الأداء، وهي ما يمكن أن يُطلق عليه السمة أو السمات الكامنة **(Latent Traits)** ونظرًا لصعوبة ملاحظة هذه السمات الكامنة مباشرة، فإنه يتم تقدير قدرات الأفراد المفحوصين بناءً على السمة موضع القياس، من خلال الأداء الملاحظ لهؤلاء الأفراد على مجموعة من الفقرات الاختبارية (Hambleton & Swaminathan, 1985).

وتتضمن نظرية الاستجابة للفقرة مجموعة من النماذج الرياضية تُعرف بـ**نماذج الاستجابة للمفردة (Item Response Models)** وتُحقق هذه النماذج قياسًا موضوعيًا وتعالج العديد من مشكلات القياس النظرية والتطبيقية التي عجزت النظرية التقليدية عن حلها. ومن أبرز هذه النماذج المستخدمة في مجال القياس والتقويم النفسي والتربوي: النموذج اللوغاريتمي أحادي المعلمة، والنموذج اللوغاريتمي ثنائي المعالم، والنموذج اللوغاريتمي ثلاثي المعالم.

نماذج نظرية استجابة الفقرة ثنائية الإجابة

تُقسم نماذج نظرية استجابة الفقرة ثنائية الإجابة بناءً على عدد المعالم إلى ثلاثة نماذج، كما أوردها أمبرينسون وريز (Embretson & Reise, 2000):

١. **النموذج اللوجستي أحادي المعلمة (1PLM):**

يحتوي هذا النموذج على معلمة الصعوبة (bj) فقط، حيث يُفترض أن معلمة التخمين تساوي صفرًا، وأن جميع الفقرات لها نفس معلمة التمييز. ويُعبر عن هذا النموذج بالمعادلة الرياضية التالية:

$$P_i(\theta) = \frac{1}{1 + e^{-1.7(\theta - b_i)}} \quad \dots\dots\dots (1)$$

٢. **النَّموذج اللوجستي ثنائي المعلمة (2PLM):**

وتظهر في معادلته الرياضيّة معلمة الصعوبة (bj)، والتمييز (aj) فقط، ويفْترض أنَّ معلمة التَّخمين تساوي صفراً، وتكون معادلته على الصّورة:

$$P_i(\theta) = \frac{1}{1 + e^{-1.7a_I(\theta - b_i)}} \quad \ldots\ldots\ldots (2)$$

٣. **النَّموذج اللوجستي ثلاثي المعلمة (3PLM):**

وتظهر في معادلته الرياضيّة المعالم الثلاث الصعوبة (bj)، والتمييز (aj)، والتَّخمين (cj)، وتكون معادلته على الصّورة:

$$P_i(\theta) = c_i + \frac{1}{1 + e^{-1.7(\theta - b_i)}} \quad \ldots\ldots\ldots (1)$$

مشكلة الدراسة وأهميتها

يتوقف النجاح في تطبيق نظرية الاستجابة للفقرة على الظروف التجريبية والإجراءات المستخدمة في تقدير معالم الفقرة والقدرة. ومن هنا، فقد كان لا بد من دراسة أثر طريقة تقديم الاختبار (ورقي، إلكتروني) على الخصائص السيكومترية للاختبار وعلى معالم فقرات الاختبار وفقًا لنظرية الاستجابة للفقرة. وتحديدًا، فإن الدراسة سعت للإجابة عن الأسئلة الآتية:

١. ما درجة ملاءمة بيانات الدراسة مع نماذج نظرية الاستجابة للفقرة؟
٢. ما أثر طريقة تقديم اختبار الاختيار من متعدد (ورقي، إلكتروني) على معالم فقرات الاختبار وفق نظرية الاستجابة للفقرة؟
٣. ما أثر طريقة تقديم اختبار الاختيار من متعدد (ورقي، إلكتروني) على خصائص الاختبار السيكومترية وفق نظرية الاستجابة للفقرة؟

وجاءت هذه الدراسة استكمالًا للجهود المبذولة لزيادة دقة القياس والتقديرات لمعالم الفقرات الاختبارية التي بُنيت عليها نظرية الاستجابة للفقرة؛ حيث تمثّل قضية طريقة تقديم الاختبارات أحد الظروف التجريبية والإجراءات المستخدمة في تقدير معالم الفقرة والقدرة، لذا فهي تمثل أحد المتغيرات التي قد تؤثر في نجاح تطبيق نماذج نظرية الاستجابة للفقرة.

من هنا، تظهر أهمية الدراسة من الناحية النظرية والعملية بما يلي:

١. **من الناحية النظرية:** تبين فيما إذا كان هناك أثر لطريقة تقديم الاختبار (ورقي، إلكتروني) على معالم فقرات الاختبار وعلى خصائصه السيكومترية.
٢. **من الناحية العملية:** قد تسهم هذه الدراسة في توجيه القائمين على الاختبارات في وزارة التربية والتعليم لتحديد الطريقة الأمثل في تقديم الاختبار للمفحوصين.

فرضيات الدراسة

١. لا توجد فروق ذات دلالة إحصائية عند مستوى ($\alpha = 0.05$) بين متوسطات معالم فقرات الاختبار (الصعوبة، التمييز، والتخمين) تُعزى لطريقة تقديم الاختبار (ورقي، إلكتروني) وفق نظرية الاستجابة للفقرة.

٢. لا توجد فروق ذات دلالة إحصائية بين معاملي الثبات الإمبريقي تُعزى لطريقة تقديم الاختبار (ورقي، إلكتروني).

مصطلحات الدراسة

- **طريقة تقديم الاختبار**: هي الطريقة التي تم فيها تقديم اختبار اللغة الإنجليزية للصف الثامن الأساسي، وإجابة الطلبة على فقراته، وذلك على النحو التالي:
  - **ورقي**: وهي الطريقة التقليدية التي قام فيها الطلبة بالإجابة على فقرات الاختبار باستخدام القلم والورقة.
  - **إلكتروني**: باستخدامGoogle Drive Form ، حيث تم تقديم فقرات الاختبار لجميع الطلبة بالترتيب ذاته وتمت الإجابة عليها باستخدام الحاسوب.
- **نظرية الاستجابة للفقرة** (Item Response Theory): نظرية حديثة في القياس النفسي والتربوي يتم فيها تحديد العلاقة بين أداء المفحوص والسمة الكامنة موضع القياس وفق دالة رياضية محددة. وتعتمد هذه النظرية على عدد من النماذج تُسمى نماذج السمات الكامنة (Latent Trait Models) التي يتم من خلالها الربط بين الأداء على الفقرة وقدرة المفحوص (Hambleton & Swaminathan, 1985).
- **النموذج ثلاثي المعلمة**: هو أحد نماذج نظرية الاستجابة للفقرة ثنائية التدريج، حيث يمكن لهذا النموذج تقدير أربع معلمات، هي: قدرة الفرد، ومعامل صعوبة الفقرة، ومعامل التمييز، ومعامل تخمين الفقرة (بني عطا، ٢٠١٤).
- **الخصائص السيكومترية**: يُقصد بالخصائص السيكومترية وفق نظرية الاستجابة للفقرة: معالم الصعوبة، والتمييز، والتخمين للفقرات، ومعاملي الصدق والثبات للاختبار (علام، ٢٠٠٥).

الدراسات السابقة

أجريت بعض الدراسات التي تناولت أثر طريقة تقديم اختبار الاختيار من متعدد (ورقي، إلكتروني) على معالم فقرات الاختبار وعلى خصائص الاختبار السيكومترية وفق نظرية الاستجابة للفقرة.

فقد هدفت دراسة المومني (Al-Momany, 2012) إلى التحقق من وجدود أثر لطريقة تقديم اختبار في مادة اللغة الانجليزية للصف الخامس الأساسي على تقديرات القدرة وخصائص الاختبار وخصائص فقراته السيكومترية. وحيث تم تطبيق اختبار في اللغة الانجليزية للصف الخامس الأساسي مكون من (٣٠) فقرة من نوع الاختيار من متعدد، حيث قدم الاختبار في ثلاث طرق: (القلم والورقة)، (حاسوب نموذج (١)): قدم فيها الاختبار على الحاسوب باستخدام نموذج واحد، الطريقة الثالثة (حاسوب نموذج٢) قدم فيها الاختبار على الحاسوب باستخدام خمسة نماذج للاختبار تم فيها تغيير ترتيب الفقرات عشوائياً في كل نموذج. وتكونت عينة الدراسة من (٦٠٠) طالباً وطالبة يدرسون في (٢١) شعبة مختارة بالطريقة العشوائية العنقودية حيث تم توزيع هذه العينة إلى (٣) مجموعات تبعاً للطريقة التي قدم فيها الاختبار. وقد أظهرت نتائج الدراسة: عدم وجود فروق ذات دلالة إحصائية في دالة المعلومات التي يقدمها الاختبار عند مستوى الدلالة (α=٠.٠٥). وعدم وجود فروق ذات دلالة إحصائية في معاملات الثبات الامبريقي بين الطرق الثلاث لتقديم الاختبار عند مستوى الدلالة (α=٠.٠٥). كما أظهرت النتائج وجود علاقة بين مستوى القدرة للطلبة وبين طريقة تقديم الاختبار لصالح الطريقة الثالثة. وعدم وجود تغير في فترات الثقة حول معلمة الصعوبة، وفق نموذج راش، يعزى لاختلاف الطريقة التي تم فيها تقديم الاختبار.

وهدف دراسة محمود (Mahmod, 2018) إلى مقارنة بين الخصائص السيكومترية للاختبارين التحصيلي الإلكتروني والورقي، عند تقديمه بثلاث طرق مختلفة: الطريقة الإلكترونية أثناء المحاضرة، والطريقة الإلكترونية عن بعد، والطريقة الورقية. واستخدمت الباحثة اختباراً تحصيليا تكوينيا من عشرين فقرة من نوع الاختيار من متعدد ذات أربع بدائل في مقرر "مهارات التفكير"، واستخدم المنهج التجريبي، حيث تم اختيار ثلاث شعب متكافئة التحصيل من طالبات كلية التربية بجامعة الإمام عبد الرحمن بن فيصل. تم تقديم الاختبار للمجموعة الأولى إلكترونياً أثناء المحاضرة من خلال نظام البلاك بورد (Blackboard)، مع ضبط خيارات النظام، بحيث يتيح للممتحن محاولة واحدة، وتم توقيت الاختبار بتحديد ٢٥ دقيقة كزمن للإجابة، وتم تقديم فقرات الاختبار في صفحة واحدة؛ مما يتيح للممتحن التنقل بين الأسئلة ومراجعة الإجابات. وتم تقديم الاختبار للمجموعة الثانية الاختبار إلكترونياً عن بعد من خلال نظام البلاك بورد، وتم ضبط خيارات تقديم الاختبار بحيث يكمل الممتحن الاختبار في جلسة واحدة، مع حظر الرجوع؛ مما يمنع الممتحن من التنقل بين الأسئلة، وتم توقيت الاختبار بتحديد ٢٥

دقيقة كزمن للإجابة. وقدم الاختبار للمجموعة الثالثة في صورة ورقية. وأظهرت النتائج وجود فروق ذات دلالة إحصائية بين معاملات صعوبة المفردات عند تقديمه إلكترونيا عن بعد، وعند تقديمه إلكترونياً أثناء المحاضرة أو ورقيا، وكانت الفروق لصالح الصورة الأولى "الإلكترونية عن بعد"، بينما لم توجد فروق ذات دلالة إحصائية بين معاملات تمييز مفردات الاختبار التحصيلي بصوره الثلاثة "إلكترونياً عن بعد وإلكترونياً أثناء المحاضرة وورقيا"، وتم الحصول على أفضل معامل ثبات للاختبار وأفضل معامل للصدق التلازمي عند تقديمه إلكترونيا أثناء المحاضرة ، وكان أقل معامل للثبات وأقل معامل للصدق التلازمي عند تقديم الاختبار الكترونيا عن بعد. وأوصت الباحثة بتقديم الاختبار الكترونياً بتواجد المعلم، وعدم تقديمه إلكترونيا عن بعد، كما أوصت بتطوير إجراءات ضبط تقديم الاختبار عن بعد؛ ليتيح للمعلم مراقبة أداء الممتحنين أثناء تقديم الاختبار.

ومن الدراسات ذات الصلة دراسة الدلالعة ورفاقه (Al-Dalalah etal, 2019) والتي كان هدفها معرفة أثر الاختبارات الإلكترونية ونمط التفكير على قلق الاختبار والكفاءة الذاتية لدى طلبة الجامعة الاردنية. حيث تكونت عينة الدراسة من (١٢٣) طالباً وطالبة والمسجلين لمساق مهارات الحاسوب موزعين على أربعة شعب من طلبة الجامعة. وأظهرت نتائج الدراسة وجود فروق ذات دلالة إحصائية تعزى لطريقة المعالجة فأظهرت أن الطلبة الذين لديهم كفاءة ذاتية منخفضة أداؤهم على الاختبار الإلكتروني أفضل منه على الاختبار الورقي، والطلبة ذوي القلق المرتفع أداؤهم على الاختبار الإلكتروني أفضل من أدائهم على الاختبار الورقي، وأن الاختبارات الإلكترونية أفضل من الاختبارات الورقية فتخفيضها لنسبة القلق، ونسبة الكفاءة الذاتية تزداد في الاختبارات الإلكترونية ويزيد التحصيل، وأن الطلبة ذوي التفكير المستقل انخفض القلق الديهم وارتفعت كفاءتهم الذاتية، وجود فروق ذات دلالة إحصائية ولصالح الاختبارات الإلكترونية مقارنة بالاختبارات الورقية.

وفي دراسة عبدالوهاب (Abd-Alwahhab, 2022) والتي هدفت إلى التعرف على أثر طريقة عرض مفردات الاختبار الإلكتروني الطريقة التتابعية مع حظر الرجوع – الطريقة التتابعية بدون حظر الرجوع) وعدد مفرداته (٣٠–٤٠) مفردة في كل من الخصائص السيكومترية للاختبار، والأداء عليه، وذلك لدى عينة قوامها (٤٥٠) طالبة من طالبات كلية العلوم الإقتصادية والموارد الإسلامية بجامعة أم القرى، تم تقسيمها إلى (٦) مجموعات، وكان قوام كل مجموعة (٧٥) طالبة، بحيث تتعرض كل منها إلى اختبار تحصيلي إلكتروني مختلف في طريقة عرض مفرداته و / أو عددها، وبالمقارنة بين المجموعات تبين وجود فروق دالة إحصائيا تعزى إلى اختلاف طريقة عرض المفردات عند تثبيت

عددها – في كل من معاملات ارتباط المفردات المصححة بالدرجة الكلية ومعاملات صعوبة المفردات ومعاملات تمييزها وأداء الطالبات على الاختبار ، كما تبين عدم وجود فروق دالة إحصائيا تعزى إلى عدد مفردات الاختبار – عند تثبيت طريقة عرضها – في معاملات ارتباط المفردات المصححة بالدرجة الكلية وأداء الطالبات على الاختبار ، وقد ظهرت فروق دالة إحصائيا تعزى لعدد مفردات الاختبار عند عرض المفردات بالطريقة التتابعية بدون حظر الرجوع فقط في كل من معاملات صعوبة المفردات ومعاملات ،تمييزها، كما أسفرت النتائج عن اختلاف بين الصور الاختبارية في معاملات ارتباط درجات الطالبات عليها بمعدلاتهن التراكمية، وفي معاملات ثباتها.

التعليق على الدراسات السابقة

اتضح للباحثين من خلال مراجعة الدراسات السابقة ندرة الدراسات التي تناولت أثر طريقة تقديم الاختبار (ورقي، إلكتروني). وأن أغلب الدراسات السابقة درست أثر طريقة تقديم الاختبار (ورقي، إلكتروني) على خصائص الاختبار السيكومترية وفق مفاهيم النظرية التقليدية في القياس (Classical Measurement Theory)، وأن الدراسات التي تناولت أثر طريقة تقديم الاختبار (ورقي، إلكتروني) على خصائص الاختبار السيكومترية استخدمت النموذج أحادي المعلمة كدراسة المومني (Al-Momany, 2012)؛ ومن هنا جاءت هذه الدراسة استكمالاً لهذه الدراسات، حيث درست أثر ترتيب فقرات الاختبار أثر طريقة تقديم الاختبار الاستجابة للفقرة (Item Response Theory) النموذج ثلاثي المعلمة.

الطريقة والإجراءات

منهجية الدراسة

تبع هذا البحث المنهج الوصفي التحليلي في تحليل نتائج نماذج الاختبارات وفقراتها وفق النموذج ثلاثي المعلمة من نماذج نظرية الاستجابة للفقرة، باستخدام برمجيات (Spss)، (Jasp) و(MG-BILOG3).

عينة الدراسة

تكونت عينة الدِّراسة من طلاب وطالِبات الصف الثامن الأساسي في المدارس التابعة لمديرية التربية والتعليم للواء قصبة المفرق، للعام الدِّراسي ٢٠٢٣/٢٠٢٤م. وقد بلغ مجموع عينة الدراسة

(٥٠٠) مفحوصاً، تقدم منهم (٢٥٠) مفحوصاً للاختبار الورقي و (٢٥٠) مفحوصاً للاختبار الإلكتروني.

أداة الدراسة

تمَّ بناء اختبار تحصيلي في مبحث اللغة الإنجليزية للصَّف الثامن الأساسي في وحدة "I know now" من كتاب اللغة الإنجليزية، حيث تمَّت الاستعانة بدليل المعلِّم في ذلك المبحث، وفيما يأتي وصفاً للإجراءات التي اتبعت في إعداد وبناء الاختبار:

١) **تحديد الغرض من الاختبار:** والذي تمثَّل بقياس تحصيل طلاب وطالِبات الصف الثامن الأساسي في المدارس التابعة لمديرية التربية والتعليم للواء قصبة المفرق، للعام الدِّراسي ٢٠٢٣/٢٠٢٤م، في مبحث اللغة الإنجليزية في وحدة "I know now" من كتاب اللغة الإنجليزية المقرَّر.

٢) **تحديد الموضوعات الدَّاخلة في الاختبار:** وذلك من خلال الأهداف التفصيليَّة لمادة الاختبار، وتحليل محتواها، ثم تحديد المستوى للأهداف من خلال ربط مستويات النتاجات التعليميَّة الخاصَّة بمحتوى المادة الدِّراسية، ليتمَّ بعد ذلك إعداد جدول مواصفات. والجدول (١) يبين جدول المواصفات للاختبار.

جدول رقم (١)

جدول مواصفات الاختبار

مستوى النتاج / المحتوى	معرفة		فهم وتطبيق		مهارات عقلية عليا		المجموع	
	عدد الفقرات	النسبة المئوية	عدد الفقرات	النسبة المئوية	عدد الفقرات	النسبة المئوية	عدد الفقرات	النسبة المئوية
Give a good title for the article.	1	%4	0	% 0	0	% 0	**1**	%4
Guess the meanings.	2	%8	1	%4	١	% 4	**4**	% 16
Missing linking word.	1	% 4	0	% 0	0	% 0	**1**	%4
Using simple future for predication	1	%4	3	%12	3	%12	**7**	% 28
Using vocabulary items in context.	1	%4	2	%8	1	%4	**4**	% 16
Determining pronoun reference.	1	%4	1	%4	0	0%	**2**	%8
Expressing opinions and justify reasons .	1	%4	1	%4	1	%4	**3**	%12
Expressing their preference.	1	%4	1	%4	1	%4	**3**	% 12
المجموع	9	%36	9	% 36	7	%28	25	%100

٣) **صياغة فقرات الاختبار بِصورتِه الأوليَّة:** بالاعتِماد على جدول مواصفات الاختبار، تمَّت كتابة فقرات الاختبار بِصورتِهِ الأوليَّة حسب الأسس العلميَّة والفنيَّة المتَّبعة في كتابة فقرات الاختبار، مع مراعاة ألَّا تعتمد إجابة إحدى هذه الفقرات على إجابة الفقرات الأخرى، حيث تكوَّن الاختبار بِصورتِهِ الأوليَّة من (٢٥) فقرة من نوع الاختيار من متعدد بأربعة بدائل لكل فقرة، بديل واحدة منها فقط يحتمل الإجابة الصحيحة.

٤) **التأكد من صدق محتوى الاختبار:** تمَّ عرض الصُّورة الأَوليَّة للاختبار، وجدول المواصفات، والنتاجات التعليميَّة التفصيليَّة، على (٩) محكمين ممن هم من ذوي الخبرة والاختصاص، واحد منهم من ذوي تخصص القياس والتقويم، واثنين من ذوي تخصص أساليب تدريس اللغة الإنجليزية، وثلاثة من مشرفي اللغة الإنجليزية في وزارة التربية والتعليم، وثلاثة مدرِّسين ممن يدرِّسون المبحث، حيث طلب منهم الإجابة على استبانة للحكم على مدى ملائمة جدول المواصفات، وفقرات الاختبار.

وفي ضوء تقديرات المحكِّمين ومناقشة مقترحاتهم تمَّ إجراء التعديلات اللازمة على فقرات الاختبار، حيث تمَّت إعادة صياغة إحدى بدائل الفقرة رقم (١٥)، كما تمَّت إعادة صياغة متن الفقرتين رقم (١٩) و(٢٠)، ولم تحذف أي فقرة، ليبقى الاختبار بِصورتِه الأَوليّة بعد أَخِذ تعديلات المحكِّمين بعين الاعتبار مكوَّناً من (٢٥) فقرة.

٥) **التجريب الأولي للاختبار:**

تمَّ تطبيق الاختبار بِصورتِه الأوليّة والمكوَّن من (٢٥) فقرة، على العينة الاستطلاعيَّة، للتَأكُّد من وضوح الصياغة اللغوية للفقرات، وتحرِّي الفقرات التي يتمُّ التساؤل حولها للكشف عمَّا إذا كان هنالك غموض فيها، وضبط الزمن الذي يستغرقه الطلبة في إنهاء الاختبار.

وبناءً على عملية تطبيق الاختبار على العينة الاستطلاعيَّة تمَّ تحديد الزمن المناسب للإجابة على فقرات الاختبار: حيث وجِدَ أنَّ معدَّل (٩٥%) من الطلبة تقريباً أنهوا الإجابة على فقرات الاختبار بعد مرور (٤٥) دقيقة، وعليه تمَّ اعتِماد زمن الاختبار بفقراته الخمس والثلاثين ليكون (٤٥) دقيقة. وتبين وضوح فقرات وتعليمات الاختبار للمفحوصين.

٦) **إعداد صور الاختبار:** بعد الخروج بالصورة النهائية للاختبار، تم إعداد صورتي الاختبار (الورقي والإلكتروني).

٧) **إعداد تعليمات الاختبار والإجابة النموذجية:** تم إعداد ورقة تعليمات الاختبار.

٨) تطبيق الاختبار بنموذجيه على عينة الدراسة:

لتحقيق أهداف الدراسة طبق الاختبار على عينة الدراسة والتي كان عددها (٥٠٠) مفحوصاً، تقدم منهم (٢٥٠) مفحوصاً للاختبار الورقي و(٢٥٠) مفحوصاً للاختبار الإلكتروني. وصححت فقراته تمهيدا لاستكمال إجراءات التحليل والإجابة على أسئلة الدراسة.

إجراءات الدراسة

المعالجة الإحصائية

من أجل الإجابة عن أسئلة الدراسة، تم استخدام المعالجات الإحصائية التالية:

- التحقق من تحقق افتراضات نظرية الاستجابة للفقرة من خلال التحليل العاملي ومؤشر Q.
- التحقق من ملاءمة بيانات الدراسة مع نماذج نظرية الاستجابة للفقرة باستخدام برمجية BILOG-MG لصورتي الاختبار.
- استخدام اختبار t-test عن الفروق في بين متوسطات معاملات الصعوبة والتخمين باختلاف صور الاختبار (ورقي، إلكتروني) واختبار Mann-Whitney اللامعلمي للكشف عن الفروق في بين متوسطات معاملات التمييز باختلاف صور الاختبار(ورقي، إلكتروني).
- إيجاد دالة المعلومات لفقرات الصور الثلاث للاختبار والاختبار ككل باستخدام برمجية BILOG-MG.
- استخدام اختبار M للكشف عن الفروق في معاملات الثبات الامبريقي لصور الاختبار (ورقي، إلكتروني).

النتائج ومناقشتها

للإجابة عن السؤال الأول: ما درجة ملاءمة بيانات الدراسة مع نماذج نظرية الاستجابة للفقرة؟

التحقق من افتراضات النموذج

تم التحقق من افتراضات النموذج اللوجستي ثلاثي المعلمة؛ لما لها من أثر على دقة التقديرات وفق النموذج. وهي:

أحادية البعد: وتم التحقق منه بطريقتين:

أولاً: باستخدام التحليل العاملي لكل صورة من صور الاختبار باستخدام طريقة المكونات الرئيسة (Analysis Principle Component) ويبين الجدول (2) قيم الجذر الكامن(Eigen Values)

ونسب التباين المفسر (Explained Variance) للعوامل الخمسة الأولى في صور الاختبار.

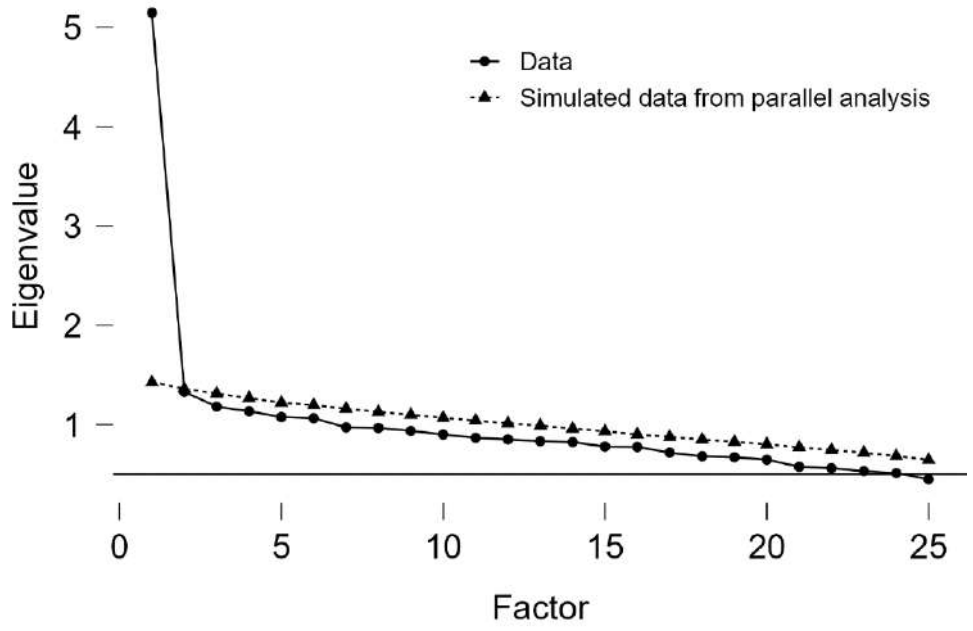
جدول رقم (2)

قيم الجذر الكامن ونسب التباين المفسر للعوامل الخمسة الأولى بصورها الثلاث

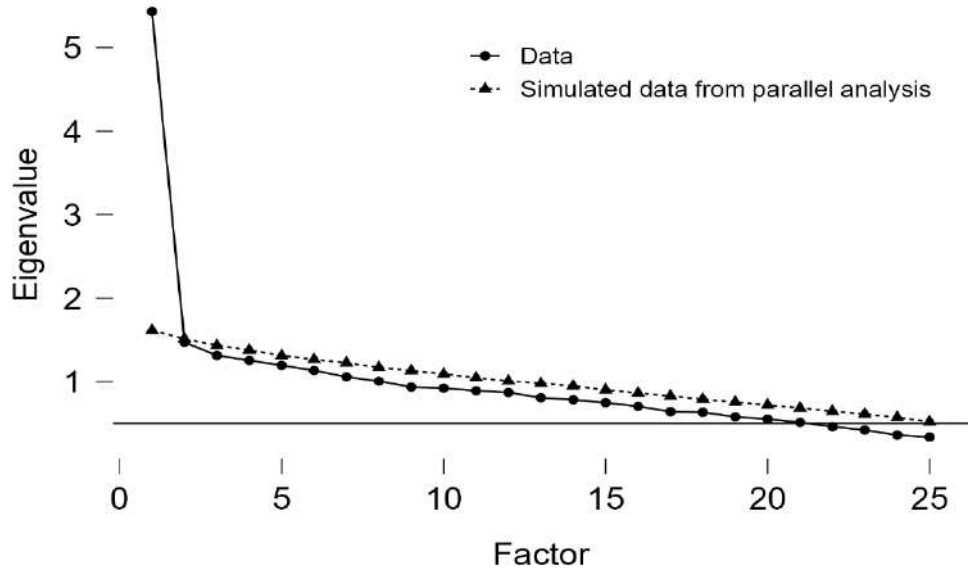
رقم العامل	نماذج الاختبار			
	إلكتروني		ورقي	
	الجذر الكامن	نسبة التباين المفسر %	الجذر الكامن	نسبة التباين المفسر %
1	5.431	20.2	5.148	18.6
2	1.473	4.5	1.331	3.1
3	1.316	3.6	1.181	2.5
4	1.256	3.3	1.135	2.3
5	1.197	3.3	1.077	2.3

يتضح من الجدول (2) تحقق شرط أحادية البعد في صورتي الاختبار؛ وذلك اعتماداً على محك الجذر الكامن كمؤشر على أحادية البعد، حيث يتبين من الجدول (٢) أنه وبالرغم من أن نسبة التباين المفسر للعامل الأول ليست مرتفعة – أقل من ٢٠%–للاختبار الورقي، إلا أن الجذر الكامن للعامل الأول يطغى بشكل جليّ على الجذور الكامنة لبقية العوامل، حيث أن نسبة الجذر الكامن للعامل الأول إلى الجذر الكامن للعامل الثاني تزيد عن (٢). حيث ذكر لورد (Lord) (المشار إليه في دراسة (Albanese and Forsyth, 1984) بأن الفقرات تكون أحادية البعد إذا كانت قيمة الجذر الكامن للعامل الأول كبيرة بالمقارنة مع قيمة الجذر الكامن الثاني، بحيث تكون قيمة الجذر الكامن الثاني ليست أكبر بكثير من أي عامل آخر، إذ أن نسبة الجذر الكامن للعامل الأول إلى الجذر الكامن الثاني كبيرة وتزيد عن (٢).

كما تطابقت نتائج اختبار فرز العوامل بالرسم البياني (Scree plot) الشكل (١)، والشكل (٢) مع نتائج التحليل العاملي، حيث يتبين من شكل (١)، والشكل (٢) إن العامل الأول يمثل عاملا سائدة وإن المنحنى قد تغير اتجاهه وانعطف انعطافا ملحوظا عند العامل الثاني، مما يؤكد تحقق افتراض أحادية البعد لأغراض التحليل وفق نظرية الاستجابة للفقرة.



شكل (١) التمثيل البياني للجذور الكامنة للاختبار الورقي



شكل (٢) التمثيل البياني للجذور الكامنة للاختبار الإلكتروني

ثانياً: مؤشِّر الجذر التربيعي للوسط الحسابي لمربعات البواقي: Root Mean Square of Residuals (RMSR)

فقد بلغت قيمته المحسوبة لدى مجموعة الطلبة لكلا صيغتي لاختبار صفراً وهي أقل من القيمة الحرجة له البالغة قيمتها (٠.٢٥٩٣) والتي تحسب من خلال المعادلة:

$$C.V. = 4.1/\sqrt{N}$$

حيث N: حجم العينة. وهذا مؤشِّر كاف على تحقق افتراض أُحادية البُعد (Fraser & McDonald, 1988).

الاستقلال الموضعي Local independence:

ويشير هذا الافتراض أن استجابات المفحوصين الذين لديهم نفس القدرة لفقرة ما مستقلة إحصائياً، أي أن استجابة المفحوص على فقرة ما يجب ألا تؤثر في استجابته على أي فقرة أخرى، بمعنى أن إجابة الفقرة لا تُعطي تلميحات أو معيقات لإجابة فقرة أخرى. وهذا الافتراض يعني أن لا شيء يؤثر في التحصيل في الاختبار سوى قدرة المفحوص وخصائص الفقرة. وعندما يتحقق هذا الافتراض، يصبح احتمال الحصول على علامة لمجموعة من الفقرات هو حاصل ضرب الاحتمال لكل من هذه الفقرات. ويشير هامبلتون وسوامنيثان (Hambleton & Swaminatan, 1985) إلى أن هذا الافتراض يكافئ افتراض أحادية البعد، وهذا يعني أنه إذا تحقق افتراض أحادية البعد في الاختبار فإن الاختبار يحقق افتراض الاستقلال لموضعي. وهذا ما حصل في الدراسة الحالية. إلا أن العديد من علماء القياس في تكافؤ افتراضي أحادية البعد والاستقلال الموضعي، فقد ذكر Meara & Sired (2000) المشار إليهم في حمادنة (٢٠١١)، وعبدالوهاب (٢٠١٦) إلى عدم تكافؤ الافتراضين.

لذلك تم التحقق من افتراض الاستقلال الموضعي باستخدام طريقة أخرى وهي مؤشِّر (Yen's Index) المعروف بـمِؤشر (Q_3) حيث تم استخدام برنامج (LDID)، (Local Dependence Indices for Dichotomous Items)، حيث تم حساب نسبة ارتباطات البواقي بين كل زوج من فقرات الاختبار التي حققت الاستقلال الموضعي والتي لم تحققه وفقاً لمؤشر Q3 للاستقلال الموضعي للنموذين، ويبين الجدول (٣) نتائج افتراض الاستقلال الموضعي لفقرات الاختبار لدى بيانات عينة الدِّراسة.

جدول (٣)

التكرارات والنِّسب المئويَّة لحالتي الاستقلال الموضعي لدى بيانات عينة الدِّراسة

حالة الاستقلال الموضعي	(إلكتروني)		(ورقي)	
	عدد الأزواج	النسبة المئوية	عدد الأزواج	النسبة المئوية
معتمد	5	%1.7	6	%2
مستقل	295	%98.3	294	%98
الكلي	300	%100.0	300	%100.0

يلاحظ من الجدول (٣) تحقق الاستقلال الموضعي لفقرات المقياس، حيث بلغت النسبة المئوية للحالات التي تحققت لها الاستقلالية ٩٨.٣% للاختبار الإلكتروني، و٩٨% للاختبار الورقي؛ وهي بذلك تجاوزت نسبة ٨٠% التي تعد النسبة الدنيا لتحقق فرض الاستقلال الموضعي الشريفين وأبو وردة (٢٠٢٠).

افتراض التحرر من السرعة في الأداء.

أشار الزبيدي، (٢٠١٨) إلى أنه يمكن التأكد من أن الاختبار ليس اختبار سرعة من خلال فحص نسبة الذين أنهوا الاختبار، وفحص الفقرات التي لم يجب عنها الأفراد. ويرى هامبلتون (Hambleton,1994)، إنه إذا كان (٧٥%) من الأفراد أكملوا الإجابة على الاختبار، و(٨٠٪) من فقرات الإجابة قد تمت الإجابة عليها، فإن السرعة لن تعتبر في هذه الحالة عاملا مهما في الأداء على الاختبار، وحيث إن نسبة الطلاب والطالبات الذين أكملوا هذا الاختبار هي (١٠٠%)، وأن نسبة الفقرات التي تمت الإجابة عنها أيضا تساوي (١٠٠%)، فإن هذا الاختبار يقيس القوة، ولا يقيس السرعة.

مطابقة الأفراد والفقرات للنَّموذج اللوجستي ثلاثي المعلمة وفق نظرية الاستجابة للفقرة:

وبعد ذلك تمت مطابقة الاستجابات عن فقرات كل صورة من صورتي الاختبار لنماذج نظرية الفقرة، حيث تم تحليل استجابات أفراد عينة الدراسة الذين أجابوا عن فقرات صورتي الاختبار باستخدام برنامج (BILOG-MG) المصمم لتحليل البيانات وفقاً لنماذج نظرية الفقرة. وتبين مطابقة جميع الطلاب والبالغ عددهم (٥٠٠) في صورتي الاختبار، حيث كانت احتمالية المطابقة لهم جميعًا أعلى من مستوى الدلالة المتبنى (α=٠.٠١).

وفيما يتعلق بفحص مطابقة الفقرات للنموذج المستخدم، أعيد التحليل باستخدام برنامج (BILOG-MG). وبناءً على مؤشر مربع كآي عند مستوى الدلالة ($\alpha = 0.05$)، الذي يعتبر الفقرة غير مطابقة للنموذج إذا كان مقدار الاحتمالية لهذه الفقرة أقل من (0.05)، حيث أظهرت نتائج التحليل عدم مطابقة الفقرة رقم (٢٢) من الصورة الأولى (الاختبار الإلكتروني)، وعدم مطابقة الفقرة رقم (١٩) من الصورة الثانية (الاختبار الورقي).

وبعد حذف جميع الفقرات غير المطابقة للنموذج في صورتي الاختبار، أصبح الاختبار بصورتيه مكون من (٢٣) فقرة.

للإجابة عن السؤال الثاني: ما أثر طريقة تقديم اختبار الاختيار من متعدد (ورقي، إلكتروني) على معالم فقرات الاختبار وفق نظرية الاستجابة للفقرة؟

تمَّ تقدير قيم معالم فقرات الاختبار (معلمة الصعوبة b، ومعلمة التمييز a، ومعلمة التخمين c)، وِفقاً للنَّموذج اللوجستي ثلاثي المعلمة (3PLM) في نظرية الاستجابة للفقرة لصورتي الاختبار (الإلكتروني والورقي) وعدد فقراتهما (٢٣) فقرة لكل اختبار، وذلك باستخدام برنامج (BILOG-MG3) والجدولين (٤، ٥) يبينان قيم معالم هذه الفقرات والخطأ المعياري لكل منها.

جدول (٤)

القيم المقدَّرة لكلٍّ من معلمة الصعوبة والتمييز والتَّخمين لفقرات الاختبار الإلكتروني وِفقاً للنَّموذج ثلاثي المعلمة

رقم الفقرة	الصعوبة b	الخطأ المعياري للصعوبة	التمييز A	الخطأ المعياري للتمييز	التخمين c	الخطأ المعياري للتخمين
1	−1.274	0.498	1.191	0.363	0.5	0.109
2	−2.547	0.731	1.231	0.456	0.5	0.112
3	−0.984	0.379	1.684	0.552	0.498	0.106
4	0.718	0.274	1.438	0.523	0.339	0.072
5	2.682	0.822	1.438	0.792	0.199	0.036
6	−1.905	0.525	1.322	0.418	0.495	0.111
7	0.541	0.229	1.516	0.502	0.281	0.069
8	−0.274	0.419	1.22	0.352	0.455	0.098

رقم الفقرة	الصعوبة b	الخطأ المعياري للصعوبة	التمييز A	الخطأ المعياري للتمييز	التخمين c	الخطأ المعياري للتخمين
9	−0.782	0.481	1.132	0.336	0.493	0.105
10	−1.202	0.482	1.239	0.392	0.5	0.108
11	−0.885	0.344	1.574	0.387	0.425	0.102
12	−0.548	0.309	1.749	0.591	0.409	0.097
13	0.372	0.241	1.809	0.679	0.378	0.074
14	−0.507	0.299	1.724	0.479	0.397	0.095
15	0.604	0.297	1.483	0.561	0.404	0.075
16	−0.2	0.29	1.833	0.639	0.421	0.089
17	−0.349	0.255	1.904	0.549	0.351	0.087
18	−0.003	0.179	2.141	0.629	0.261	0.068
19	0.263	0.24	1.573	0.484	0.308	0.075
20	−0.291	0.234	2.125	0.738	0.351	0.085
21	0.084	0.311	1.497	0.51	0.406	0.086
22	−0.139	0.318	1.495	0.467	0.401	0.09
23	−0.724	0.225	2.714	0.894	0.353	0.091
الوسط الحسابي	−0.320	0.364	1.610	0.534	0.397	0.089
الانحراف المعياري	1.027	0.162	0.370	0.144	0.084	0.018

جدول (٥)

القيم المقدَّرة لكلٍّ من معلمة الصعوبة والتمييز والتَّخمين لفقرات الاختبار الورقي وِفقاً للنَّموذج ثلاثي المعلمة

رقم الفقرة	الصعوبة b	الخطأ المعياري للصعوبة	التمييز A	الخطأ المعياري للتمييز	التخمين c	الخطأ المعياري للتخمين
1	−2.259	0.823	0.753	0.217	0.5	0.111
2	−2.836	0.972	0.797	0.28	0.5	0.112
3	−0.997	0.466	1.222	0.389	0.488	0.106

رقم الفقرة	الصعوبة b	الخطأ المعياري للصعوبة	التمييز A	الخطأ المعياري للتمييز	التخمين c	الخطأ المعياري للتخمين
4	1.172	0.353	1.219	0.483	0.341	0.069
5	1.786	0.387	1.691	0.845	0.182	0.04
6	−2.314	0.827	0.752	0.216	0.494	0.111
7	1.091	0.313	1.287	0.502	0.31	0.067
8	−0.382	0.51	1.033	0.325	0.5	0.102
9	−0.947	0.526	1.022	0.299	0.487	0.107
10	−1.295	0.402	1.44	0.374	0.436	0.106
11	−1.142	0.47	1.199	0.36	0.474	0.107
12	−0.856	0.346	1.816	0.682	0.471	0.104
13	0.304	0.315	1.428	0.511	0.4	0.084
14	−0.667	0.351	1.52	0.425	0.43	0.1
15	0.284	0.235	2.041	0.736	0.387	0.073
16	−0.338	0.286	1.627	0.387	0.364	0.089
17	−0.262	0.277	1.921	0.635	0.406	0.09
18	−0.08	0.283	1.582	0.494	0.358	0.086
19	0.297	0.29	1.355	0.428	0.335	0.08
20	−0.474	0.257	1.88	0.522	0.347	0.089
21	0.189	0.346	1.23	0.381	0.376	0.087
22	−0.33	0.224	2.331	0.787	0.347	0.084
23	−0.831	0.217	3.336	2.125	0.389	0.095
الوسط الحسابي	−0.473	0.412	1.499	0.539	0.405	0.091
الانحراف المعياري	1.104	0.204	0.576	0.386	0.080	0.018

يلاحظ من الجدول (٤) أنَّ قيم معلمة الصعوبة للاختبار الإلكتروني تراوحت ما بين (2.547−) و(2.682)، وبمتوسط مقداره (٠.٣٢٠−)، في حين أنَّ قيم معلمة التمييز تراوحت ما بين (١.١٣٢)

و(٢.٧١٤)، وبمتوسط مقداره (١.٦١٠)، في حين تراوحت قيم معلمة التَّخمين ما بين (٠.١٩٩) وَ(0.500)، وبمتوسط مقداره (٠.٣٩٧).

بينما يلاحظ من الجدول (٥) أنَّ قيم معلمة الصعوبة للاختبار الورقي تراوحت ما بين (٢.٨٣٦-) و(١.٧٨٦)، وبمتوسط مقداره (٠.٤٧٣-)، في حين أنَّ قيم معلمة التمييز تراوحت ما بين (0.752) و(3.336)، وبمتوسط مقداره (1.499)، في حين تراوحت قيم معلمة التَّخمين ما بين (0.182) وَ(0.500)، وبمتوسط مقداره (0.405).

ويشير هامبلتون (Hambleton, 1994) إلى أنه عندما تتراوح قيم معلم الصعوبة بين ٢- و ٢ والتمييز بين ٠.٤ و٢ والتخمين بين صفر و٠.٢٥، فإن المعالم تتمتع بالاتساق (Consistency). وعند مقارنة هذه القيم بنتائج الدراسة الحالية في الجدول ٢، نجد بأن المعالم متسقة.

يتضح أيضاً أن مدى صعوبة الفقرات كان كبيراً، وأن قيم معاملات التمييز كانت مرتفعة، وهذا ما أشار إليه امبرتسون و رايس (Embretson & Reise، ٢٠٠٠) إلى أن منحنى توزيع العلامات الخاص به منبسط نسبيا (Flat) على مدى متصل السمة، وهذا يعني أن مجموعة من الفقرات ذات تمييز عالي وبمدى واسع من معلم الصعوبة.

وفيما يتعلق بمعلمة التخمين فإن نتائج الدراسة الحالية تتفق مع القيم التي حصل عليها (حمادنة، ٢٠٠٩) حيث بلغت قيمة المتوسط الحسابي لمعلمة التخمين التي حصل عليها ٠.٢٦٤. وبهذا يمكن القول بأن أن تقديرات معالم الفقرات في هذه الدراسة مقبولة ضمن محكات القياس التربوي والنفسي.

وللتحقق من وجود أثر لطريقة تقديم اختبار الاختيار من متعدد (ورقي، إلكتروني) على معالم فقرات الاختبار وفق نظرية الاستجابة للفقرة تم استخدام اختبار (T-test) للمقارنة بين متوسطات قيم معالم الفقرات في الاختبارين الورقي والإلكتروني.

أولاً: معلمة الصعوبة.

بالاستعانة ببرنامج Jasp تم التحقق أولاً من افتراضات اختبار (T-test) للمقارنة بين متوسطات عينتين مستقلتين وهي:

- اعتدالية التوزيع للبيانات: باستخدام اختبار (Shapiro-Wilk) أظهرت النتائج اعتدالية التوزيع، والجدول (٦) يبين نتائج الاختبار.

جدول (٦)

اختبار (Shapiro-Wilk) لاعتدالية التوزيع لقيم معلمة الصعوبة لمجموعتي الاختبار

المجموعة	قيمة الاختبار	الدلالة الإحصائية (.sig)
إلكتروني	0.943	0.205
ورقي	0.972	0.737

يتضح من الجدول (٦) تحقق شرط الاعتدالية عند مستوى الدلالة $(\alpha = 0.05)$.

- شرط تجانس التباين: باستخدام اختبار (Levene) أظهرت النتائج تحقق شرط تجانس التباين للمجموعتين، والجدول (٧) يبين نتائج الاختبار.

جدول (٧)

اختبار (Levene) لتجانس التباين بين قيم معلمة الصعوبة للاختبارين

قيمة الاختبار (F)	درجة الحرية١ (df_1)	درجة الحرية٢ (df_2)	الدلالة الإحصائية (.sig)
0.231	1	44	0.633

يتضح من الجدول (٧) تحقق شرط تجانس التباين عند مستوى الدلالة $(\alpha = 0.05)$. وبعد التحقق من افتراضات الاختبار تم استخدام الاختبار (T-test) للتحقق من وجود فروق بين متوسطات معاملات الصعوبة باختلاف طريقة تقديم الاختبار (ورقي، إلكتروني). والجدول (٨) يبين نتائج الاختبار.

الجدول (٨)

دلالة الفروق بين متوسطات معلمتي الصعوبة باختلاف طريقة تقديم الاختبار (ورقي، إلكتروني)

الاختبار	عدد الفقرات	المتوسط الحسابي	الانحراف المعياري	قيمة (t)	درجة الحرية	الدلالة الإحصائية (.sig)
إلكتروني	23	0.320 -	1.027	0.489	44	0.627
ورقي	23	0.473 -	1.104			

يتضح من الجدول (٨) عدم وجود فروق ذات دلالة إحصائية بين متوسطات معاملات الصعوبة باختلاف طريقة تقديم الاختبار (ورقي، إلكتروني).

ثانياً: معلمة التمييز

تم التحقق أولاً من افتراضات اختبار (T-test) للمقارنة بين متوسطات عينتين مستقلتين وهي:

- اعتدالية التوزيع للبيانات: باستخدام اختبار (Shapiro-Wilk) أظهرت نتائج اختبار اعتدالية التوزيع، والجدول (٩) يبين نتائج الاختبار.

جدول (٩)

اختبار (Shapiro-Wilk) لاعتدالية التوزيع لقيم معلمة التمييز لمجموعتي الاختبار

المجموعة	قيمة الاختبار	الدلالة الإحصائية (.sig)
إلكتروني	0.907	0.035
ورقي	0.894	0.019

يتضح من الجدول (٩) عدم تحقق شرط الاعتدالية عند مستوى الدلالة $(\alpha = 0.05)$. لذا تم استخدم اختبار مان ويتني اللامعلمي للفروق بين متوسطي الرتب لمعاملات التمييز للاختبارين (ورقي، إلكتروني). والجدول (١٠) يبين نتائج الاختبار.

جدول (١٠)

قيم (Z, W, U) ودلالتها للفروق بين متوسطات رتب درجات الاختبارين (ورقي، إلكتروني)

المعلمة	(U)	W	Z	Sig.
التمييز	212	488	−1.153	0.249

يتضح من الجدول (١٠) عدم وجود فروق ذات دلالة عند مستوى الدلالة $(\alpha = 0.05)$. بين متوسطي الرتب لمعاملات التمييز للاختبارين (ورقي، إلكتروني).

ثالثاً: معلمة التخمين:

تم التحقق أولاً من افتراضات اختبار (T-test) للمقارنة بين متوسطات عينتين مستقلتين وهي:

- اعتدالية التوزيع للبيانات: باستخدام اختبار (Shapiro-Wilk) أظهرت النتائج اعتدالية التوزيع، والجدول (١١) يبين نتائج الاختبار.

جدول (١١)

اختبار (Shapiro-Wilk) لاعتدالية التوزيع لقيم معلمة التخمين لمجموعتي الاختبار

المجموعة	قيمة الاختبار	الدلالة الإحصائية (sig.)
إلكتروني	0.932	0.124
ورقي	0.907	0.135

يتضح من الجدول (١١) تحقق شرط الاعتدالية عند مستوى الدلالة ($\alpha = 0.05$).

- شرط تجانس التباين: باستخدام اختبار (Levene) أظهرت النتائج تحقق شرط تجانس التباين للمجموعتين، والجدول (١٢) يبين نتائج الاختبار.

جدول (١٢)

اختبار (Levene) لتجانس التباين بين قيم معلمة التخمين للاختبارين

قيمة الاختبار (F)	درجة الحرية١ (df_1)	درجة الحرية١ (df_2)	الدلالة الإحصائية (sig.)
0.010	1	44	0.922

يتضح من الجدول (١٢) تحقق شرط تجانس التباين عند مستوى الدلالة ($\alpha = 0.05$).

وبعد التحقق من افتراضات الاختبار تم استخدام الاختبار (T-test) للتحقق من وجود فروق بين متوسطات معلمتي التخمين باختلاف طريقة تقديم الاختبار (ورقي، إلكتروني). والجدول (١٣) يبين نتائج الاختبار.

جدول (١٣)

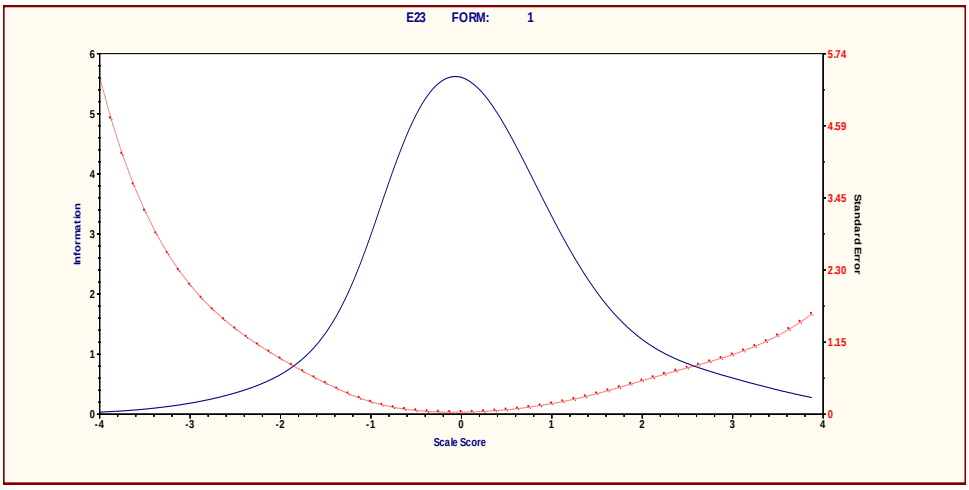
دلالة الفروق بين متوسطات معلمتي التخمين باختلاف طريقة تقديم الاختبار (ورقي، إلكتروني).

الاختبار	عدد الفقرات	المتوسط الحسابي	الانحراف المعياري	قيمة (t)	درجة الحرية	الدلالة الإحصائية (sig.)
إلكتروني	23	0.397	0.084	0.355 -	44	0.724
ورقي	23	0.405	0.080			

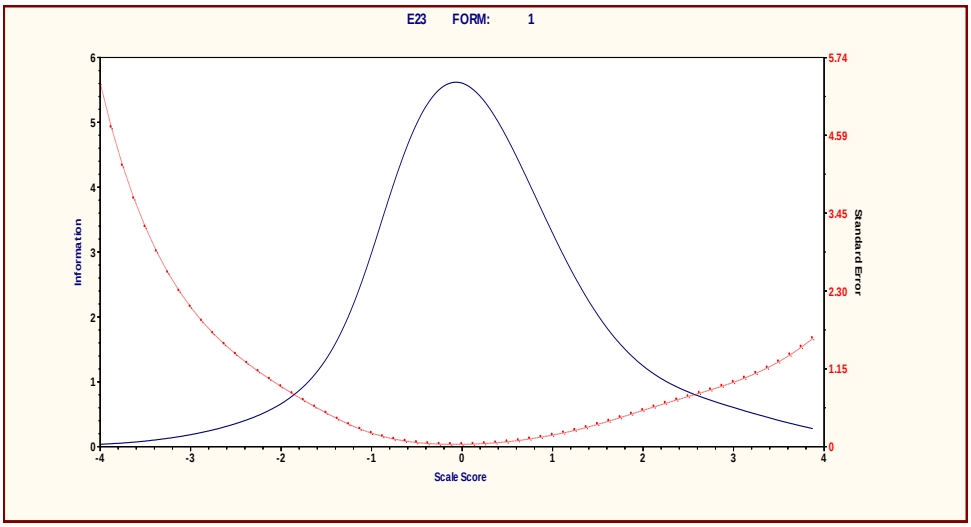
يتضح من الجدول (١٣) عدم وجود فروق ذات دلالة إحصائية بين متوسطات معلمة التخمين باختلاف طريقة تقديم الاختبار (ورقي، إلكتروني).

للإجابة عن السؤال الثالث: ما أثر طريقة تقديم اختبار الاختيار من متعدد (ورقي، إلكتروني) على خصائص الاختبار السيكومترية وفق نظرية الاستجابة للفقرة؟

للاجابة عن هذا السؤال تم رسم المنحنيات الخاصة بدالة المعلومات والقيم العظمى لها باستخدام برمجية BILOG-MG، والأشكال التالية توضح منحنيات دوال المعلومات لصورتي الاختبار.



شكل (٣) منحنى دالة المعلومات للاختبار الإلكتروني



شكل (٤) منحنى دالة المعلومات للاختبار الورقي

يتضح من الشكلين (٣، ٤) بأن كمية دالة المعلومات كانت للاختبار المقدم إلكترونياً أعلى من كمية دالة المعلومات كانت للاختبار المقدم ورقياً، وربما يعود ذلك إلى معلمة التمييز للفقرات، حيث كانت قيم التمييز للفقرات عندما تم تقديم الاختبار إلكترونياً أعلى من قيم التمييز للفقرات عندما تم تقديم الاختبار، فكلما زاد معلم التمييز للفقرة زادت كمية المعلومات التي تسهم بها الفقرة حول معلم الصعوبة.

ثم تم حساب معاملات الثبات الامبريقي لطريقة تقديم الاختبار، ثم تم استخدام الاحصائي M للكشف عن دلالة الفروق بين قيم معاملات الثبات وفقاً لطريقة تقديم الاختبار، ويبين الجدول (١٢) معاملات الثبات لطرق الاختبار الثلاث.

فقد تم إيجاد معاملات الثبات الامبريقي المنبثق من نظرية الاستجابة للفقرة لكل صورة والجدول (١٤) يبين قيم معاملات الثبات الامبريقي لصورتي الاختبار.

جدول (١٤)

قيم معاملات الثبات الامبريقي لكل صورة من صور الاختبار

صورة الاختبار	الثبات الإمبريقي
إلكتروني	0.787
ورقي	0.769

يتضح من النتائج الواردة في الجدول (١٤) قيم معاملات الثبات الامبريقي لصورتي كانت متقاربة؛ ويمكن تفسير ذلك بأن معامل الثبات الإمبريقي يرتبط بدالة المعلومات للاختبار والتي يتم حسابها وفق القانون:

$$I(\theta) = \sum_{g=1}^{n} I_g(\theta)$$

$I_g(\theta)$: دالة المعلومات الفقرة g.

فقد بين ريفي (Reeve, 2004) بأن دالة المعلومات للاختبار ترتبط مع ثبات المقياس من خلال العلاقة التالية:

$$r = 1 - \frac{1}{I(\theta)}$$

وحيث أن قيم دوال المعلومات لصورتي الاختبار كانت متقاربة (٤.٦٩٥ للاختبار الإلكتروني، ٤.٣٢٩ للاختبار الورقي) بالتالي فإن قيم معاملات الثبات الامبريقي لصورتي كانت متقاربة.

ولمعرفة دلالة الفروق في معاملات الثبات الامبريقي المحسوبة من نظرية الاستجابة للفقرة، فقد تم استخدام اختبار (M) المقترح من قبل هاكستين وولن (Hakstin & Whalen, 1976) والذي يتبع توزيع مربع كاي بدرجات حرية (عدد المعاملات – ١)، ويحسب وفق المعادلة التالية:

$$\left(\sum_{k=1}^{k} B_k \; - \; \frac{\left[\sum_{k=1}^{k} B_k (1-r_k)^{\frac{-1}{s}}\right]^2}{\sum_{k=1}^{k} B_k (1-r_k)^{\frac{-2}{s}}} \right) M = \frac{j-1}{18j}$$

حيث: $B_k = \frac{(9n_k - 11)^2}{n_k - 1}$،

J: عدد فقرات الاختبار، r_k: معامل ثبات الطريقة، n_k: عدد أفراد العينة المستخدمة في الطريقة.

حيث كانت قيمة الإحصائي M المحسوبة تساوي (٤.٢٦٦) وهي أقل من القيمة الحرجة لكاي تربيع بدرجات حرية (١) والبالغة (٩.٢١) مما يشير إلى عدم وجود فروق ذات دلالة إحصائية بين قيم معاملات الثبات الامبريقي، فبالرغم من اختلاف قيمها إلا أن قيمها متقاربة.

التوصيات

في ضوء نتائج الدراسة يوصي الباحثان باستخدام طريقة تقديم الاختبار الإلكترونية لما لها من مميزات مثل توفير الوقت والجهد ولما أظهرته من خصائص سيكومترية جيدة للاختبار والفقرات.

المراجع

المراجع باللغة العربيّة:

- أبو وردة، إيمان والشريفين، نضال. (٢٠٢٠). بناء اختبار لقياس القدرة على التفكير المجرد باستخدام نظرية استجابة الفقرة. **المجلة الأردنية في العلوم التربوية**، ١٨(٣)، ٥٧٩-٦١٢.
- بني عطا، زايد. (٢٠١٤). تقصي دقة تقدير النموذج اللوجستي ثلاثي المعلمة لعالم الفقرة وقدرة الأفراد في ضوء تغير طول الاختبار وحجم العينة: دراسة محاكاة، **مجلة جامعة الشارقة للعلوم الإنسانية والاجتماعية**، ١١(٢):١-٣٧.
- حمادنة، إياد. (٢٠٠٩). استخدام نظرية الاستجابة للمفردة في بناء اختبار محكي المرجع في

الرياضيات وفق النموذج اللوجستي ثلاثي المعلم، **مجلة العلوم التربوية والنفسية**، ١٠(٢): ٢١٥-٢٣٨.

- الدلالعة، أسامة، عبابنة، زياد، والزبون، مالك. (٢٠١٩). أثر الاختبارات الإلكترونية ونمط التفكير على التحصيل وقلق الاختبار والكفاءة الذاتية لدى طلبة الجامعات الأردنية. **مجلة دراسات**، ٤٦(٣). ٣٩١-٤١١.
- الزبيدي، محمد. (٢٠١٨). بناء اختبار تحصيلي محكي المرجع في مقرر الاختبارات والمقاييس باستخدام النموذج الأحادي المعلم نموذج راش لطلبة الدبلوم التربوي في كلية التربية بجامعة الطائف، **المجلة الدولية التربوية المتخصصة**، ٧(٢):٤٠-٥٤.
- عبد الوهاب، محمد. (٢٠٢٢). أثر طريقة عرض مفردات الاختبار الإلكتروني وعددها في خصائصه السيكومترية والأداء عليه. **مجلة جامعة الفيوم للعلوم التربوية والنفسية**، ١(١٦). ١-٦٣.
- علام، صلاح الدين. (٢٠٠٥). **الاختبارات والمقاييس التربوية والنفسية**، ط١، القاهرة: دار الفكر العربي.
- العمري، محمد. (٢٠٠٥). الكفايات التعليمية اللازمة لمعلمي مبحث الحاسوب في المرحلة الثانوية ومدى ممارستهم لها من وجهة نظر المعلمين أنفسهم. **مجلة مؤتة للبحوث والدراسات**، 20(7).85- 116.
- محمود، سومية. (٢٠١٨). مقارنة الخصائص السيكومترية بين الاختبارات التحصيلية الإلكرتونية والورقية. **المجلة العلمية لكلية التربية- جامعة أسيوط**، ٢٤(٦). ١١٧-١٤٦.
- المومني، صبحة أحمد والشريفين، نضال كمال محمد. .(٢٠١٢) . **أثر طريقة تقديم اختبار اللغة الإنجليزية لدى طلبة الصف الخامس الأساسي على تقديرات القدرة وخصائص الاختبار وخصائص فقراته السيكومترية**. (رسالة ماجستير). جامعة اليرموك، إربد. مسترجع من http://search.mandumah.com/Record/740028.

المراجع باللغة الإنجليزية:

- Albanese, M. A & ,. Forsyth, R. (1984). The one- tow and modified two parameter latent trait models: An empirical study of relative fit. *Educational and psychological Measurement*, 44, 229-246.

- Bennett, R. E. (1999). Using New Technology to Improve Assessment. E*ducational Measurement: Issues and Practice*, 18 (3),5–12.
- Embretson, S. & Reise, S. (2000). *Item response theory for psychologists*. New Jersey: Lawrence Erlbaum Associates.
- Fraser, C., & McDonald, R. P. (1988). NOHARM*: least squares item factor analysis.Multivariate Behavior Research*, 23: 267-269.
- Hakstain, A.R & Whalen, T.E.(1976). A K-sample significance test for independent alpha coefficients. *Pyschometrika*, *41*,219-231.
- Hambleton, R., K.(1994). Item Response Theory: A broad psychometric framework for measurement advances. *Psicothema*,6, 535-556.
- Hambleton, R., & Swaminathan, H. (1985). *Item Response Theory: principles and Applications*. Boston: Kluwer–Nijhoff publishing.
- Reeve, B. (2004). *Applications of item response theory (IRT) modeling for building and evaluating questionnaires measuring patient-reported outcomes*[On-Line]:http://outcomes.caner.