دراسة سيكومترية لبنوك الأسئلة المحوسبة في الجامعة العربية المفتوحة في الأردن.

الدكتور سمير عبد السلام الصوص أستاذ مساعد في المناهج وطرق التدريس كلية التربية - الجامعة العربية المفتوحة، الأردن
باسل خميس سالم أبوفودة أستاذ مشارك في القياس والتقويم التربوي كلية التربية - الجامعة العربية المفتوحة، الأردن

الملخص

هدفت الدراسة إلى استخدام النموذج اللوجستي ثلاثي المَعلمة في تحليل فقرات بنك أسئلة محوسب في الجامعة العربية المفتوحة في الأردن، ولتحقيق أهداف الدراسة تم الاعتماد على إجابات الطلبة لبنك أسئلة الامتحان النهائي المحوسب لمادة (اللغة العربية لمعلمي المرحلة الابتدائية 2+1) (Arabic Language for Teachers of Elementary School (I+II
خلال الفصل الأول للعام الجامعي 2024/2023 والمكون من(862) فقرة من نوع الاختيار من متعدد، وصحح وخطأ وتكونت عينة الدراسة من استجابة (125) طالبًا وطالبة، جرى اختيارهم بطريقة عشوائية من مجتمع الدراسة، وأظهرت النتائج تحقق افتراضات نظرية استجابة الفقرة في القياس في اختبار (اللغة العربية لمعلمي المرحلة الابتدائية 2+1)، وبلغت قيمة معامل الثبات الإمبريقي (التجريبي) (0.91)، وتراوحت قيم معلمة الصعوبة للفقرات بين (2.66 - -0.33)، بمتوسط حسابي (0.87)، وانحراف معياري (0.69)، وتراوحت قيم الأخطاء المعيارية لتقديرات معلمة الصعوبة بين (0.88 – 0.00)، بمتوسط حسابي مقداره (0.28)، بانحراف معياري مقداره (0.22)، وتراوحت قيم معلمة التمييز للفقرات بين(3.81-0.59)، بمتوسط حسابي (1.83)، وانحراف معياري (0.91)، وبلغت قيمة الأخطاء المعيارية لتقديرات معلمة التمييز بين (2.90 – 0.17)، بمتوسط حسابي (0.81)، وانحراف معياري مقداره (0.62)، وتراوحت قيم معاملات التخمين للفقرات بين (0.49 - 0.13)، بمتوسط حسابي (0.28)، وانحراف معياري (0.09)، وبلغت قيم الأخطاء المعيارية لتقديرات معلمة التخمين بين (0.09–0.00)، بمتوسط حسابي (0.035)، وانحراف معياري مقداره (0.022). وتراوحت قيم معلمة القدرة بين (2.31 - -2.33)، بينما تراوحت قيم

الأخطاء المعيارية لتقديرات قدرات الطلبة بين(4.191– 0.15)، بمتوسط حسابي (0.89)، وانحراف معياري مقداره (1.67).

الكلمات المفتاحيّة: النموذج اللوجستي ثلاثي المعلمة، النموذج اللوجستي أحادي المعلمة، النظرية الحديثة في القياس، بنوك الأسئلة، معالم الفقرات، قدرات المفحوصين

Psychometric study of computerized question banks at the Arab Open University in Jordan

Dr. Sameer Abdul Salam Alsous College of Education, Arab Open University, Amman, Jordan.

Dr. Basel Khamis Salem Abu-Foudeh.Associate Professor Measurement and Evaluation College of Education, Arab Open University, Jordan

Abstract

The study aimed to use the three-parameter logistical model in analyzing the items of a computerized question bank at the Arab Open University in Jordan. To achieve the objectives of the study, students' answers to the computerized final exam questions bank were relied on for the subject of (Arabic Language for Primary School Teachers).

During the first semester of the academic year 2023/2024, which consists of (862) items of the multiple-choice questions, correct and error, and the study sample consisted of a response of (125) students, randomly selected from the study community.

The results showed the realization of the assumptions of the item response theory in the measurement in the (Arabic language for primary teachers 1 +2) test, and the value of the empirical reliability coefficient

was (0.91), and the values of the difficulty parameter for the items ranged between (2.66 - -0.33), with an mean (0.87), and a standard deviation (0.69), and the values of the standard errors of the estimates of the difficulty parameter ranged between (0.88 – 0.00), with an mean of (0.28), with a standard deviation of (0.22), and the values of the discrimination parameter for the items ranged between (3.81-0.59), with an mean of (0.83), and the standard deviation (0.29), and the values of the standard deviation between (0.09-0.00), and the standard deviation of (0.035). The values of the ability parameter ranged between (2.31 -2.33), while the values of the standard errors of the estimates of students' abilities ranged between (4.191-0.15), with an mean (0.89), and a standard deviation of.(1.67)

Key words: Three-parameter logistic model, One-parameter logistic model, Item Response Theory, question banks, Item parameters, Ability of the examiners.

المقدمة

يُلاحظ المتتبع لحركة القياس والتقويم بشكل عام والامتحانات في المؤسسات التعليمية بشكل خاص تزايد الاهتمام باختبارات الاختيار من متعدد لما لها من استخدامات واسعة في جميع المجالات التعليمية والمهنية، وقد أشاد بها كثير من الباحثين؛ منهم جيه(Gay, 1985) ، الذي قال إن اختبارات الاختيار من متعدد تستطيع أن تقيس نتاجات تعلم في جميع المستويات العقلية، ويمكن أن تغطي مجالًا واسعًا من السلوكيات، وبالتالي تكون أكثر صدقًا وثباتًا من بقية الاختبارات، ويشير عودة (2010)، أن اختبارات الاختيار من متعدد أكثر الاختبارات شيوعًا وانتشارًا خاصةً في اختبارات التحصيل المقننة، فهي قادرة على قياس نواتج تعلم في المستويات العليا من المجال المعرفي.

مشكلة البحث

قدمت النظرية الحديثة في القياس العديد من الحلول الخاصة بالمشاكل التي تتعلق ببناء الاختبارات وتطويرها، خصوصًا في ما يتعلق ببناء بنوك الأسئلة، وانطلاقًا من أهمية الدور الذي تلعبه نماذج النظرية الحديثة في القياس في تحليل فقرات الاختبار، جاءت هذه الدراسة لتلقي الضوء على استخدام النموذج اللوجستي ثلاثي المعلمة في تحليل فقرات بنك أسئلة محوسب في الجامعة العربية المفتوحة في الأردن.

أسئلة البحث

1. ما مدى تحقق افتراضات نظرية استجابة الفقرة في بنك الأسئلة المحوسب لمادة (اللغة العربية لمعلمي المرحلة الابتدائية 1+2)؟
2. ما معامل الثبات التجريبي لبنك الأسئلة المحوسب لمادة (اللغة العربية لمعلمي المرحلة الابتدائية 1+2)؟
3. ما الخصائص الإحصائية لفقرات بنك الأسئلة المحوسب لمادة (اللغة العربية لمعلمي المرحلة الابتدائية 1+2) في ضوء نظرية استجابة الفقرة؟
4. ما الخصائص الإحصائية لتقديرات قدرات طلبة الجامعة العربية المفتوحة والأخطاء المعيارية في تقديرها في بنك الأسئلة المحوسب لمادة (اللغة العربية لمعلمي المرحلة الابتدائية 1+2)؟

أهداف البحث

هدفت الدراسة إلى استخدام النموذج اللوجستي ثلاثي المعلمة في تحليل فقرات بنك أسئلة محوسب مادة(اللغة العربية لمعلمي المرحلة الابتدائية 1+2) في الجامعة العربية المفتوحة في الأردن في الاختبار النهائي خلال الفصل الدراسي الأول (2024/2023)، من خلال التحقق من افتراضات النظرية الحديثة في القياس في بنك الأسئلة المحوسب، وبيان معالم الفقرات وقدرات المفحوصين في بنك الأسئلة المحوسب.

أهمية البحث

تكمن أهمية الدراسة النظرية من خلال تناولها لموضوع استخدام النموذج اللوجستي ثلاثي المَعلَمة في تحليل فقرات بنك أسئلة محوسب لمادة (اللغة العربية لمعلمي المرحلة الابتدائية

2+1)، وذلك من خلال بيان مؤشرات الصعوبة والتمييز والتخمين لفقرات بنك الأسئلة المحوسب، أما أهمية الدراسة العملية فتكمن من خلال دراستها التطبيقية لتحليل فقرات بنك الأسئلة المحوسب في الجامعة العربية المفتوحة في الأردن من منظور واقعي، ومحاولتها التوصل إلى بنك أسئلة محوسب يتمتع بخصائص سيكومترية جيدة، وفق منظور نظرية استجابة الفقرة يمكن استخدامه في البيئة الجامعية، ومما يزيد في أهمية الدراسة أنها اعتمدت على بيانات حقيقية، تم الحصول عليها من مواقف امتحانية واقعية في البيئة الجامعية.

حدود البحث

الحدود البشرية: طلبة الجامعة العربية المفتوحة في الاردن.
الحدود الموضوعية: استخدام النموذج اللوجستي ثلاثي المَعْلَمَة في تحليل فقرات بنوك الأسئلة المحوسبة في الجامعة العربية المفتوحة في الأردن
الحدود الزمانية: العام الجامعي 2024/2023.
الحدود المكانية: الأردن.
الإطار النظري

يتعرض الطلبة في المراحل التعليمية المختلفة إلى عدد كبير من الاختبارات باختلاف أشكالها، وتتضمن هذه الاختبارات مجموعة من الأسئلة الموضوعية مثل: أسئلة الاختيار من متعدد، وأسئلة المطابقة، وأسئلة الصواب والخطأ، وأسئلة التكميل، وهنالك الأسئلة المقالية. وتُعدّ فقرات الاختيار من متعدد من أفضل أنواع الفقرات الموضوعية، وأكثرها شيوعًا وانتشارًا في الاختبارات التحصيلية (عودة، 2010).
ويتكون اختبار الاختيار من متعدد من مجموعة من الفقرات، والتي تشكل عينة عشوائية من مجموع الفقرات التي يمكن صياغتها لقياس البناء النفسي أو العقلي المراد قياسه، ويفترض أن تقيس الفقرات بمجموعها سمة واحدة، وهذا ما يسمى بأحادية البعد (Unidimensionality) (Hambleton and Swaminathan, 1985)، ولكن من الصعب الوصول إلى هذه الفقرات من الناحية العملية، فبعض الاختبارات تقيس أكثر من سمة، فامتلاك بعض الأفراد لسمة غير السمة المراد قياسها بواسطة الاختبار، قد تساعدهم في الإجابة عن بعض الفقرات ولا تساعد غيرهم، والفقرة التي تتصف بهذه الخاصية هي فقرة متحيزة.(Lord, 1980)
ويتوقف مستوى أداء المفحوص في الاختبار، على خصائص فقرات الاختبار وقدرة المفحوص(Hambleton,1983) ، وتتأثر علامة المفحوص على الاختبار بعوامل كثيرة

غير مفهوم السمة أو المحتوى الذي يقيسه الاختبار، كحكمة المفحوص في إجابة الفقرة، إضافة إلى عوامل أخرى قد يكون لها تأثير في مستوى الأداء في الفقرة (Plake, Thompson and Lowry, 1981). ومن غير الممكن تجنب مصادر التباين التي ليس لها صلة بالسمة المقيسة، ولكن من المهم أن لا تُعطى هذه المصادر أية امتيازات لمجموعة من المفحوصين دون المجموعة الأخرى، ويمكن أن تحدث مثل هذه الامتيازات عندما تتوزع آثار مصادر التباين هذه بصورة غير متماثلة بين مجموعتين فرعيتين، على الرغم من تساويهما في القدرة أو السمة التي يقيسها الاختبار.(Crocker and Algina, 1986)
وتتعدد مصادر أخطاء القياس، ومنها ما يأتي (عودة، 2010):

1. أخطاء في الاختبار نفسه، ومنها:
- الخطأ العيني، ويحدث هذا الخطأ عندما لا تكون هذه العينة ممثلة لجميع الفقرات المحتملة في الاختبار.
- خطأ التكافؤ، حيث إنه من الصعب أن تكون من مجتمع الفقرات نفسه التي تقيس السمة نفسها في صورتين متكافئتين تمامًا.
- خطأ النقاوة. أي أنّ الفقرات الداخلة في الاختبار ذات صلة بسمة أخرى غير السمة التي أُعدّ لقياسها.
2. أخطاء ناتجة عن إدارة الاختبار، ومنها: ظروف التطبيق والتصحيح.
3. أخطاء ناتجة عن المفحوص نفسه، ومنها:
- قلق الامتحان.
- الحكمة في الإجابة، ونعني بذلك درجة اهتمام الطالب وفهمه للتعليمات قبل الامتحان، ودرجة المخاطرة في التخمين، وتوزيع زمن الإجابة على الفقرات، ونحو ذلك.
- التعب الجسمي والإعياء الفكري، وخاصة إذا كان الاختبار طويلًا نسبيًّا.
- الدافعية نحو تقديم أفضل أداء.
- درجة التأثر بالعوامل الخارجية.

لقد أسفرت جهود علماء القياس والتقويم التربوي والنفسي عن ظهور اتجاهات حديثة في مجال تحليل المفردات الاختبارية، ومن بين هذه الاتجاهات النظرية الحديثة في القياس أو نظرية الاستجابة للمفرد(Item Response Theory (IRT)، أو نظرية السمات الكامنة (Latent Traits Theory (LTT، وحَظِيَ هذا المدخل الجديد باهتمام العديد من الباحثين؛ كونه يتغلب على كثير من مشكلات القياس في النظرية الكلاسيكية في القياس.

وتُعدّ النظرية الحديثة في القياس ثورة في ميدان القياس التربوي والنفسي، حيث ساعدت هذه النظرية في تقديم الكثير من الحلول لمشاكل تتعلق ببناء الاختبارات وتطويرها، وخاصة في ما يتعلق ببناء بنوك الأسئلة وتحليل فقرات الاختبارات، حيث أسهمت هذه النظرية في تقييم جودة هذه الاختبار.(Embretson and Reise, 2000)

ولقد نتجت عن النظرية الحديثة في القياس مجموعة من النماذج تعرف باسم نماذج السمات الكامنة، حيث تختلف هذه النماذج في الصورة الرياضية (Hambleton and Swaminathan, 1985:49).

ومن هذه النماذج والمستخدمة في هذه الدراسة: النموذج اللوجستي ثلاثي المعلمة (Three Parameter Logistic Model)، حيث يُعدّ هذا النموذج امتدادًا للنموذج اللوجستي ثنائي المعلمة؛ إذ يضيف معلمًا جديدًا للفقرة، وهو معلم التخمين(ci) ، ويشير إلى احتمال إجابة الفقرة إجابة صحيحة من المفحوصين ذوي القدرة المتدنية، ويسمى (Pseudo Chance Level)، والمعادلة الرياضية لهذا النموذج هي:

$$Pi\ (\Theta) = ci + (1 - ci)[1 + e^{-Dai(\Theta - bi)}]^{-1} \qquad ; i = 1, 2, 3, \ldots, n$$

حيث:

Pi(Θ):احتمال أن يجيب المفحوص ذو القدرة(Θ)عن الفقرة (i)إجابة صحيحة.

D:عامل التدريج(Scaling Factor) ، وهو ثابت لجميع فقرات الاختبار، ويساوي.(1.7)

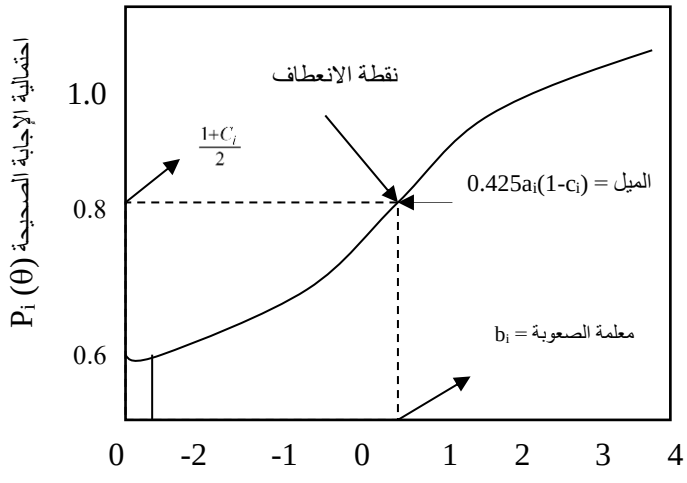
ai:معامل تمييز الفقرة.

bi:معامل صعوبة الفقرة.(i)

e:هي الأساس اللوغاريتمي الطبيعي وتساوي.(2.718)

Ci:معلمة التخمين للفقرة.(i)

ويبين الشكل رقم (1) المنحنى اللوجستي للنموذج ثلاثي المعلمة(3PLM) ، والذي يتبيَّن من خلاله معالم الفقرة: (الصعوبة، التمييز، التخمين):



.رسم المنحنى اللوجستي للنموذج ثلاثي المعلمة (الصعوبة، والتمييز، والتخمين)

وترتكز نماذج النظرية الحديثة في القياس على مجموعة من الافتراضات، الواجب توافرها في البيانات المستمدة من الاختبار، وتقوم هذه النماذج على الافتراضات الآتية: (Hambleton and Swaminathan, 1985):

1. افتراض أحادية البعد(Unidimensionality)، ويعني وجود قدرة واحدة تفسر أداء الفرد في الاختبار.
2. افتراض الاستقلال الموضعي (Local Independence)، بمعنى أنّ استجابات المفحوص للفقرات المختلفة في الاختبار مستقلة عن بعضها إحصائيًّا عند مستوى قدرة معين.
3. العلاقة الوتيرية بين القدرة (السمة) والأداء في الفقرة(Item Characteristic Curve (ICC)، وهو اقتران رياضي، ويشير بوضوح إلى أنّ احتمال إجابة الفقرة إجابة صحيحة يزداد بازدياد قدرة المفحوص.

4. عامل السـرعة(Non-Speededness)، بمعنى أنّ إخفاق الفرد في الإجابة عن فقرات الاختبار يعود إلى انخفاض قدرته، وليس إلى تأثير عامل السرعة في الإجابة.

وبنك الأسئلة هو عبارة عن مجموعة ضخمة من الأسئلة المصممة باحترافية علمية وتقنية ولغوية ومقياسية، أي أن يكون لكل سؤال وزن ومجال من المعرفة يستهدفه السؤال، حيث يتم إدخال بيانات جميع الأسئلة المنتقاة في برنامج خاص ومتطور على الحاسب الآلي الذي بدوره وبشكل آلي يخرج أسئلة متوازنة تختبر أشياء كثيرة في المنهج وتغطي المنهج كله، ويتميز هذا البنك في حالة طلب المستخدم أسئلة أخرى سيعطي أسئلة مختلفة تغطي المنهج ومتوازنة تشمل كل المقرر وتتدرج من السهولة إلى الصعوبة، بحيث تراعي مستويات الطلبة المختلفة من المتميز والمتوسط والعادي، وتتلافى أي أخطاء لغوية أو مصطلحات غير واضحة، إضافة إلى أن بنك الأسئلة هو مكان آمن توضع فيه مجموعات من الأسئلة ذات مستويات مختلفة، ويسهل عن طريقه سحب أو إضافة مجموعة أو عدد من الأسئلة المختلفة المقننة التي لها خصائص مميزة ومعلومة، مثل: معامل السهولة والصعوبة ومعامل التمييز مع إمكانية حفظ واسترجاع الأسئلة المخزونة فيها بسهولة ويسر، وكذلك صدق وثبات المفردات المصنفة وفق وحدات المقرر الدراسي وحسب المستويات العقلية المعرفية المطلوب أداؤها في أثناء الإجابة عنها، وذلك بطريقة تشبه -إلى حد ما- تنظيم وفهرسة الكتب المقرره لكل مقرر. (حسنين، 2017)

ومن فوائد بنك الأسئلة أنه بإمكان المدرس إعداد اختبار تجريبي وبثه على الموقع للطلبة بحيث يقوم الطلبة بتأديته ومعرفة علاماتهم وأخذ التغذية الراجعة، حيث يتاح للطلبة الفرصة للتدريب على الأسئلة بمستوياتها المختلفة بالأسلوب الذي يمكنهم من استيعاب مفردات كل مقرر من المقررات التي يدرسونها، وبالصورة التي تساعدهم في أداء الاختبارات بسهولة ويسر، بعد أن يكون قد تم تدريبهم على أسئلة كل مقرر من المقررات التي يدرسونها. (الصوص،2022)

ولقد تعددت الدراسات في هذا المجال ففي دراسة أجراها القضاة (2023) والتي هدفت إلى تحديد أثر طول الاختبار على معلمتي التمييز والصعوبة للمفردة باستخدام النموذج الثنائي البارميتر، وتكونت عينة الدراسة من(900)طالب، حيث قام الباحث بتطبيق ثلاثة اختبارات رياضيات للصف الثامن الأساسي مكونة من (60 ،45، 30) مفردة اختيار من متعدد، مختزلة من الاختبار الطويل(60) مفردة، وتوصلت النتائج إلى وجود فروق ذات دلالة إحصائية عند تقدير متوسطات معلمة تمييز المفردة تعزى لاختلاف طول الاختبار؛ أي أن

متوســطات قيم معلمة تمييز المفردة قد تغيرت بتغير طول الاختبار، ولمعرفة الاختلاف بين متوســطات تقدير معلمة تمييز المفردة باختلاف طول الاختبار تبين بعد إجراء المقارنات البعدية أن الفروق لصـــالح الاختبار الطويل (60) مفردة، والاختبار متوســط الطول (45) مفردة، وذلك لصالح الاختبار الطويل، و توصلت نتائج الدراسة –أيضًا- إلى وجود فروق في المتوسطات الحسابية لتقديرات معلمة صعوبة المفردة ذات دلالة إحصائية تعزى لتغير طول الاختبار؛ أي أن متوسطات تقديرات معلمة صعوبة المفردة قد تغيرت بتغير طول الاختبار، ولمعرفة الاختلاف بين متوسطات تقديرات معلمة صعوبة المفردة باختلاف طول الاختبار تبين بعد إجراء المقارنات البعدية أن الفروق لصـــالح الاختبار الطويل (60) مفردة، مقابل الاختبار المتوسط (45) مفردة، والقصير (30) مفردة.

وفي دراســـة الوديان وآخرون(2019) والتي هدفت إلى قياس أثر اختلاف طريقة عرض الاختبار الإلكتروني على معاملات الصـعوبة والتمييز وتحصـيل طلبة الصـف العاشـر في مبحث الحاسـوب وفق نظام معالجة المعلومات، ولتحقيق أهداف هذه الدّراسـة تمّ بناء اختبار تحصـيليّ إلكترونيّ من نوع الاختيار من متعدّد في مبحث الحاسـوب لطلبة الصـف العاشـر الأساسي، حيث تكوّن الاختبار بصورته النّهائيّة من (20) فقرة من نوع الاختيار من متعدّد، حيث تمّ تصميم أربع طرق لعرض هذا الاختبار تبعًا لنظريّة التعلم بالوسائط المتعددة، فكانت الطريقة الأولى نَصًّا، والطريقة الثانية نَصًّا وصورة، والطريقة الثالثة نصًّا وصورة وصوتًا، وآخرها طريقة صورة وصوت، وتكوّنت عينة الدراسة من (355) طالبًا وطالبة، طُبّق عليهم الاختبار الإلكتروني بطرق عرضـه الأربعة (نَصٌّ، نَصٌّ وصـورة، نَصٌّ وصـورة وصـوت، وصـورة وصـوت)، وبعد ذلك تم جمع البيانات وتحليلها للوصـول إلى النتائج النهائية، والتي أظهرت أنّ معاملات صـــعوبة فقرات الاختبار تراوحت بين (0.40-0.61) لطريقة عرض النَّصّ، وتراوحـت بين (0.40-0.62) لطريقـة عرض النَّصّ والصـــورة، وتراوحـت بين (0.32-0.59) لطريقة عرض النَّصّ والصـــورة والصـــوت، وتراوحت بين (0.41-0.72) لطريقة عرض الصـورة والصـوت، أما بالنسـبة لمعاملات تمييز فقرات الاختبار فتراوحت بين (0.41-0.63) للنَّصّ، وتراوحت بين (0.41-0.74) للنَّصّ والصـــورة، وتراوحت بين (0.63-0.71) للنَّصّ والصورة والصوت، وتراوحت بين (0.48-0.75) للصوت والصورة، وأظهرت النتـائج –أيضًـــا-وجود فروق ذات دلالـة إحصـــائيـة عنـد مســتوى الـدلالـة (α=0.05)تعزى لطريقة العرض بين متوسّـــطات معاملات الصـــعوبة بلغت فيها قيمة (F=4.787) وبدلالة إحصـائية(p=0.04)، ووجود فروق ذات دلالة إحصـائية عند مسـتوى الدلالة (α=0.05) تعزى لطريقة العرض بين متوسّـطات معاملات التمييز، حيث بلغت قيمة

(F=4.235) وبدلالة إحصائية (p=0.008).واستُخدمت طريقة المقارنات البعدية (LSD) لبيان الفروق الزوجية الدالة إحصائيًّا بين المتوسّطات الحسابية لكلّ من معاملات الصعوبة والتمييز، وأظهرت النتائج -ايضًا- وجود فروق ذات دلالة إحصائيّة(α=0.05)في التحصيل بين طريقة عرض الصّورة والصّوت من جهة، وكلٍّ من طرق العرض (النَّصّ ، و النَّصّ والصورة، و النَّصّ والصورة والصوت)، حيث جاءت الفروق لصالح طريقة عرض الصوت والصورة.

وفي دراسة حسنين(2017) والتي هدفت إلى معرفة مدى قدرة الاختبارات الإلكترونية على أن تكون بديلًا مناسبًا للاختبارات الورقية في تقويم الطلبة في التعلّم المفتوح، وما اتجاهات الطلبة نحو الاختبارات الإلكترونية، وما العقبات التي تواجه الطلبة عند استخدام الاختبارات الإلكترونية، وانبثق عن هذه الأهداف أربعة أسئلة، هي: ما مدى قدرة الاختبارات الإلكترونية على أن تكون بديلًا مناسبًا للاختبارات الورقية في التعلّم المفتوح؟ ما مدى ميل طلبة التعلّم المفتوح نحو استخدام الاختبارات الإلكترونية، ومدى قناعتهم بها؟ ما مدى قناعة طلبة التعلّم المفتوح بقدرة الاختبارات الإلكترونية على تقييم أدائهم بعدالة بمستوى الاختبارات الورقية؟ ما العقبات التي يمكن أن تواجه الطلبة عند استخدام الاختبارات الإلكترونية؟ واستخدم الباحث المنهج التجريبي في الجزء الأول من البحث، والمنهج الوصفي التحليلي في الجزء الثاني، واستخدم الباحث أداتين للدراسة، الأولى هي أداة الاختبار (أحدهما ورقي، والآخر إلكتروني) على عينتين متطابقتين من الطلبة، عدد كل منهما (63) طالبًا من برنامج التربية في الجامعة العربية المفتوحة في السودان، والأداة الثانية هي الاستبانة، وتم تطبيقها على عينة بحثية مكونة من (283) طالبًا من برنامج التربية، وتوصل الباحث إلى تكافؤ الاختبارات الإلكترونية والورقية في قدرتهما على قياس التحصيل الدراسي للطلبة، وميل اتجاهات الطلبة بشكل أكبر إلى استخدام الاختبارات الإلكترونية، مع وجود بعض العقبات التي تواجههم في ذلك، وقد أوصت الدراسة باستخدام الاختبارات الإلكترونية على نطاق واسع، وخصوصًا في التعلّم المفتوح عن بعد، على أن يتم تهيئة البيئة المناسبة لإجراء تلك الاختبارات، وتوفير الأجهزة والبرمجيات المناسبة لذلك، وتأهيل الطلبة وأعضاء هيئة التدريس للتعامل مع الاختبارات بكفاءة.

أما دراسة الشمراني(2016) فقد تناولت فحص أثر حجم العينة وطول الاختبار على تقدير معالم المفردة والخطأ المعياري باستخدام النموذج الثلاثي المعلمة، ولتحقيق هدف الدراسة تم توليد بيانات ثلاث عينات بأحجام مختلفة (200، 600، 1000)، وثلاثة مستويات لطول الاختبار (30، 40، 60) مفردة، باستخدام برنامج (Wengen3.) وتوصلت نتائج الدراسة

إلى وجود فروق ذات دلالة إحصائية عند تقدير متوسطات معلم الصعوبة تعزى لاختلاف حجم العينة، ولا توجد فروق تعزى لطول الاختبار أو للتفاعل بين حجم العينة وطول الاختبار، وأظهرت النتائج وجود فروق عند تقدير مَعلَم التمييز تعزى إلى اختلاف أحجام العينات، وكانت الفروق لصالح حجم العينة المتوسطة والكبيرة، وعدم وجود فروق في متوسطات معلم التمييز تعزى إلى طول الاختبار، ولا يوجد أثر للتفاعل بين طول الاختبار وحجم العينة، وأشارت النتائج – أيضًا- إلى أنَّ تقدير معلم التمييز يزداد بزيادة حجم العينة، وكذلك بزيادة طول الاختبار، إضافة إلى وجود فروق عند تقدير معلمة التخمين تعزى إلى حجم العينة، وإلى طول الاختبار وكانت لصالح حجم العينة الصغيرة، وهذا يعني أن مُعامل التخمين يرتفع كلما قل حجم العينة، علاوة على ذلك فقد أشارت النتائج أن معامل التخمين يرتفع كلما قل عدد مفردات الاختبار أيضًا.

وفي دراسة بني عطا (2014) التي هدفت إلى تقصي دقة تقدير النموذج اللوجستي ثلاثي المعلمة لمعالم المفردة وقدرة الافراد، في ضوء تغير طول الاختبار وحجم العينة، ولتحقيق الهدف من الدراسة وُلِّدت بيانات ثنائية الاستجابة بواقع(50) مرة لستة مستويات من طول الاختبار(10، 25، 50، 75، 100) مفردة من خلال استخدام برنامج(WINGEN) وباستخدام برنامج(Bilog-Mg)حُلِّلَت البيانات المولدة، كشفت نتائج الدراسة عن وجود أثر ذي دلالة إحصائية لطول الاختبار في دقة تقديرات معالم المفردات وقدرة الأفراد، وكشفت النتائج –أيضًا- أن الوسط الحسابي لقيم(RMAS) لمَعْالَم المفردات وقدرة الأفراد أخذ بالتناقص عندما زاد طول الاختبار عن(50) مفردة.

وأجرى السامرائي والخفاجي (2012) دراسة بعنوان: بناء اختبار تحصيلي محكي المرجع في مادة علم نفس الخواص لطلبة أقسام العلوم التربوية والنفسية؛ إذ هدف البحث إلى بناء اختبار تحصيلي محكي المرجع في مادة علم نفس الخواص، وهي مادة دارسية مقررة لطلبة الصف الثالث في أقسام العلوم التربوية والنفسية لكليات التربية في الجامعات العراقية، وتألف الاختبار من(160) فقرة اختبارية من نوع الاختيار من متعدد، وذلك لقياس(160) هدفًا سلوكيًّا تغطي مفردات المادة.

وقد جرى تطبيق الاختبار على عينة طبقية عشوائية بلغت(349) طالبًا وطالبة، وقد اعتمد الباحث على أنموذج(راش) أحد نماذج السمات الكامنة في تحليل بيانات الاختبار، وأُخضعت بيانات فقرات الاختبار إلى التحليل العاملي بطريقة المكونات الأساسية، الذي أظهر أن فقرات الاختبار تقيس عاملًا عامًّا واحدًا عدا (22) فقرة تم استبعادها من الاختبار، وأظهرت النتائج وجود(13) فقرة، كانت قيمة مربع كاي لعشر فقرات منها دالة إحصائيًّا، لذا تم

استبعادها من الاختبار، في ما استبعد برنامج راسكال (Rascal) ثلاث فقرات في الصفحة الأولى من التحليل الإحصـــائي لعدم مطابقتها لأنموذج راش، وقد اقترب معامل التمييز للاختبار من(1) إذ كانت قيمة معامل تمييز الاختبار ضـــمن المدى المقبول والتي بلغت(1 – 1,60). وقد بلغت قيمة ثبات الاختبار(0,957).

وأجرى العيثاوي(2009) دراســـة هدفت إلى اســـتعمال أنموذج راش وفق نظرية الســـمات الكامنة في بناء اختبار تحصـيلي في مادة علم نفس الفروق الفردية، ولتحقيق أهداف الدراسـة تم بناء اختبار مســـتندٍإلى المفردات المقررة من قبل الهيئة القطاعية لوزارة التعليم العالي، وأعد الباحث(173) فقرة من نوع الاختيار من متعدد تغطي مســـتويات الأهداف المعرفية الثلاث الأولى بحسـب تصـنيف بلوم، وهي: المعرفة والفهم والتطبيق، واشـتملت عينة البحث على(336) طالبًا وطالبة من طلبة كلية التربية للجامعات العراقية جرى اختيارهم بالطريقة الطبقية العشوائية، وقد أدخلت البيانات لبرنامج الحاسب الآلي راسكال (Rascal) ومن خلال معالجة البيانات تضمنت الصورة النهائية للاختبار (108) فقرة مطابقة لافتراضات الأنموذج ومتطلباته.

وجرى حذف(28) فقرة لا تحقق أحاديـة البعد من خلال التحليل العـاملي لفقرات الاختبـار بطريقة المكونات الأساسية، في ما حذفت(29) فقرة لدلالة قيمة مربع كآي، وقد اقتربت قيمة معامل التمييز من (1) صـــحيح حيث كانت القيمة ضـــمن المدى المقبول(0,629)، وتم اســتبعاد(8) فقرات لا تحقق اســتقلالية القياس، ودُرّجت ال (108) فقرات الاختبار بحســب صـــعوبة الفقرات، وكذلك قدرات الأفراد للحصـــول على الاختبار بصـــيغته النهائية، والذي بلغت قيمة معامل ثباته(0.952)، ومعامل تمييزه(0.713).

وقام الشـريفين (2005) بإجراء دراسـة هدفت إلى تقدير الخصـائص السـيكومترية لاختبار محكي المرجع في القياس والتقويم التربوي وفق النظرية الحديثة في القياس، ولتحقيق هدف الدراسـة قام الباحث ببناء اختبار تحصـيلي في القياس والتقويم التربوي من نوع الاختيار من متعدد.

وقد جرى تطبيق الاختبار التحصـــيلي على أفراد عينة الدراســـة البالغ عددهم (222) طالبًا وطالبة، وأشـارت نتائج الدراسـة إلى مطابقة الاسـتجابات عن (37) فقرة من فقرات الاختبار لافتراضـــات نموذج راش، وتم حذف (13) فقرة لعدم مطابقتها للنموذج، وبلغ معامل الثبات للأفراد(0.90) ومعامل الثبات للاختبار (0.95) وتم التأكد من تمتع الاختبار بمظاهر متعددة من الصدق، وبينت النتائج أن قيمة دالة المعلومات كانت أقصى ما يمكن عندما كانت (= □

b)، وأن الاختبار يقدم أكبر كمية من المعلومات عند مستويات القدرة المتوسطة، وأقل كمية من المعلومات عند مستويات القدرة العالية والمتدنية.

وأجرى ياسين(2004) بدراسة بعنوان: الخصائص السيكومترية لاختبار محكي المرجع في الكيمياء لطلبة الصف الأول الثانوي العلي مقدرة وفق النظريتين الكلاسيكية والحديثة في القياس، وقد تكون الاختبار من(52) فقرة من نوع الاختيار من متعدد من أربعة بدائل.

وتكون مجتمع الدراسة من طلاب وطالبات الصف الأول الثانوي العلمي للعام الدراسي(2004م) في مدارس مديرية التربية والتعليم لمنطقة إربد الأولى، وعددهم(2462) طالبًا وطالبة موزعين على(74) شعبة.

وتكونت عينة الدراسة من(481) طالبًا وطالبة موزعين على (14) شعبة، تم اختيارهم بالطريقة العشوائية العنقودية، وبعد تطبيق الاختبار التحصيلي المعدّ جرى تصحيح أوراق الاختبار وذلك بإعطاء العلامة واحد للإجابة الصحيحة، والعلامة صفر للإجابة الخطأ.

وجرى استخدام البرامج الإحصائية(Minsteps, Bigsteps, Microcat, SPSS)، وذلك للحصول على تقديرات لقدرات الأفراد، وقيم صعوبة الفقرات والأخطاء المعيارية في تقديرها وإحصائيات المطابقة الداخلية والخارجية، ومتوسطات المربعات الداخلية والخارجية، وقيم معاملات الثبات للأفراد والفقرات وفق النموذج اللوجستي أحادي المعلمة، وقيم معاملات الارتباط الثنائي النقطي، وقيم الصعوبة والتمييز والتخمين للفقرات وأخطائها المعيارية وفق النموذج اللوجستي ثلاثي المعلمة، وقيم معاملات الصدق والثبات وفق النظرية الكلاسيكية في القياس.

وقد أظهرت نتائج الدراسة مطابقة (43) فقرة وفق النموذج اللوجستي أحادي المعلمة، ومطابقة (48) فقرة للنموذج اللوجستي ثلاثي المعلمة، وبلغ معامل الصدق المرتبط بمحك(0.84)، ومعامل الاتساق الداخلي بطريقة كرونباخ ألفا(0.90). وتم تقدير معامل الثبات للاختبار وفق النظرية الحديثة في القياس باستخدام النموذج اللوجستي أحادي المعلمة، حيث بلغت قيمة معامل الثبات للاختبار(0.99).

وأظهرت النتائج الخاصة بنظرية استجابة الفقرة أن متوسط صعوبة الفقرات وفق النموذج اللوجستي أحادي المعلمة صفرَ لوجيت، ومتوسط الخطأ المعياري(0.12)، وفي النموذج ثلاثي المعلمة بلغ متوسط صعوبة الفقرات(0.15) لوجيت، ومتوسط الخطأ المعياري(0.119)، ومتوسط تمييز الفقرات(0.77)، ومتوسط الخطأ المعياري لمعاملات التمييز(0.125)، ومتوسط تخمين الفقرات(0.24).

وقام علاّم (1985) بإجراء دراسة بعنوان: تحليل بيانات الاختبارات العقلية باستخدام نموذج راش اللوغاريتمي الاحتمالي(دراسة تجريبية). وهدفت الدراسة إلى بحث استخدام نموذج راش اللوغاريتمي الاحتمالي في تحليل بيانات مستمدة من اختبار اليقذة العقلية الذي يقيس الذكاء العام.
ويتكون الاختبار من(22) مفردة، وهو أحد الاختبارات الفرعية التي تشتمل عليها بطارية الاستعداد العقلي للمرحلة الثانوية والجامعات، وقد اختار الباحث هذا الاختبار لأنه يتفق وفروض نموذج راش.
وقد طبق الاختبار على عينة تشمل على (126) طالبًا من بين الملتحقين الجدد في السنة الأولى في مخلف أقسام كلية التربية في جامعة الأزهر بالقاهرة، واستخدم الباحث الطريقة التي يطلق عليها (PROX) للتحقق من صحة الفرضيتين الآتيتين:

1. تنطبق البيانات المستمدة من هذا الاختبار على نموذج راش.
2. لا يتغير تدريج الاختبار(Test Calibration) بتغير مستوى قدرات عينة الطلبة المستخدمة في الحصول على هذا التدريج.

وتستخدم طريقة (PROX) في إيجاد القيم التقديرية لصعوبة مفردات الاختبار متحررة من أثر تباين قدرة الطلبة، وكذلك في إيجاد القيم التقديرية لقدرة الطلبة متحررة من أثر تباين صعوبة المفردات.
وقد بينت نتائج تحليل البواقي المعيارية لكل من المفردات والطلبة أنّ متوسطات مجموع هذه البواقي ليس لها دلالة إحصائية عند مستوى الدلالة الإحصائية (= α0.05) مما يدل على أنّ القيم التقديرية لكل من صعوبة المفردات وقدرة الطلبة تطابق النموذج (Fit the Model)، وبذلك تحققت صحة الفرضية الأولى في هذه الدراسة.
ثم قسم الباحث العينة إلى مجموعتين بحيث اشتملت المجموعة الأولى على الطلبة الذين حصلوا على أعلى الدرجات في الاختبار، واشتملت المجموعة الثانية على الذين حصلوا على أقل الدرجات، واستخدم الباحث الطريقة السابقة نفسها في الحصول على القيم التقديرية للقدرة المناظرة لدرجات كل طالب في المجموعتين.
وبينت نتائج التحليل أنّ توزيعي هذه القيم ينطبق إلى حد كبير، وبذلك تحققت الفرضية الثانية في الدراسة، وهي أن القيم التقديرية لصعوبة مفردات الاختبار مستقلة احصائيًّا عن قدرة عينة الطلبة المستخدمة في الحصول على هذه القيم، وبهذا أكدت النتائج استخدام نموذج راش في القياس الموضوعي لقدرات الطلبة.

وأجرى ستيج(Stage,2003) دراسة هدفت إلى معرفة ما إذا كان التحول من النظرية الكلاسيكية للاختبار إلى نظرية الاستجابة للفقرة هو عملية تطوير للفقرة.

وقد تكونت الدراسة من ثلاثة أجزاء حسب الآتي:

- الجزء الأول: تطبيق نماذج الاستجابة للفقرة في اختبار الاستعداد المدرسي السويدي وللاختبارات الفرعية.
- الجزء الثاني: مقارنة بين تحليل الفقرة المعتمد على نظرية الاستجابة للفقرة والنظرية الكلاسيكية للاختبار، والتي بها المقارنات بين صعوبة الفقرة للنظرية الكلاسيكية للاختبار وصعوبة الفقرة لنظرية الاستجابة للفقرة.
- الجزء الثالث: تطبيق نظرية الاستجابة للفقرة على الاختبار الكلي لاختبار الاستعداد المدرسي السويدي.

وتمت معالجة البيانات باستخدام برنامج(BILOG – MG3)، وتوصلت النتائج إلى أنّ النموذج ثنائي المعلم يكون غير مناسب للبيانات، ويختلف تقدير صعوبة الفقرات لكل من النظرية الكلاسيكية للاختبار ونظرية الاستجابة للفقرة، وأظهرت النتائج أن النظرية الكلاسيكية للاختبار والتحليل المبدئي لا يعطيان تدعيمًا للنموذج ثلاثي المعلمة لكل الاختبارات الفرعية الخمسة، وأنّ تحليل الفقرة في ضوء نظرية الاستجابة للفقرة أفضل من النظرية الكلاسيكية للاختبار . (Stage,2003)

وأجرى سواميناثان وزملاؤه((Swaminathan et al, 2003 دراسة بعنوان: تحديد أثر الحكم المسبق على دقة تقديرات معلمات الفقرة في الاختبارات ثنائية التدريج باستخدام نظرية الاستجابة للفقرة، وهدفت الدراسة إلى تفحّص أثر طريقة(Bayesian Estimation) في تقدير معلمات الفقرة، بالإضافة إلى استقصاء العلاقة بين حجم العينة ومواصفات المعلومات القبلية على دقة تقدير معلمات الفقرة.

وقد أظهرت نتائج الدراسة إلى أن التحسن في تقدير معلمات الفقرة يختلف باختلاف النموذج اللوجستي، حيث تبين بأن النموذج اللوجستي أحادي المعلمة كان الأكثر تحسنًا عند زيادة حجم العينة مقارنة بالنموذجين الثنائي والثلاثي.

وفي منحى آخر من الدراسات قام هوانج(Hwang,2002) بدراسة هدفت إلى المقارنة بين النظرية الكلاسيكية ونظرية الاستجابة للفقرة بنماذجها الثلاثة(Parameter 3 &2&1 Logistic Models) من حيث تقديرات مؤشرات صعوبة الفقرات، وتمييزها، ومستويات القدرة للطلبة، ومن حيث استقرار تقديرات الفقرا ت الاختبارية باختلاف مستويات قدرة الطلبة.

وقد بلغ حجم عينة الدراســـة (600) طالب من طلبة الصـــف الثامن، وقد اســـتُخدم اختبار تحصيلي في الرياضيات اشتمل على(15) فقرة.
وقد توصلت نتائج هذه الدراسة إلى أنّ معاملات الارتباط المرتفعة معظمها أكبر من (0,90(بين المؤشـرات الإحصـائية المسـتمدة من النظريتين تشـير إلى تشـابه كبير بينهما، كان هناك تشابه واضح بين النظريتين في تقدير مؤشرات صعوبة الفقرات، إلا أن هذا التشابه يقل في ما يتعلق بتقدير مؤشرات تمييز الفقرات.
وقام كيكار واويتمان ولوتيش Kelkar, Wightman & Lueech, (2000) بدراسة هدفت إلى التحقق من خاصـية اللاتغاير في معالم الفقرات باسـتخدام النموذج أحادي وثنائي وثلاثي المعلمة. ولتحقيق الهدف من الدراسـة تم اسـتخدام اختبار القبول لكلية الطب، وأظهرت نتائج الدراسة بأن صعوبة الفقرات استقرت باختلاف عينات الطلبة، وأنَّ تقديرات مستويات قدرة الطلبة قد استقرت رغم اختلاف صعوبة الفقرات ونسخ الاختبار.

وفي دراســـة أجراها آدمز((Adams,1988 هدفت إلى تطبيق نموذج التقدير الجزئي في تحليل بيانات تم الحصول عليها بعد تطبيق اختبار تشخيصي في مادة الرياضيات على(560) طالبًا من فئة(9 – 10) سنوات في استراليا.
وقد تكون الاختبار من فقرات من نوع الاختيار من متعدد بحيث أعطيت البدائل علامات تتناســـب وقربها من الإجابة الصـــحيحة، وقد عوملت هذه البدائل على أنها مســـتويات الإجابة(خطوات).
وأشارت نتائج الدراسة إلى فائدة نموذج التقدير الجزئي في إمكانية تفسير درجة الطالبة على الاختبار بدلالة مسـتويات المهارة والفهم التي يعبّر عنها الطالب في اسـتجابته على مجموعة المهمات(الخطوات).
وفي مجال المقارنة بين نماذج نظرية اسـتجابة الفقرة فقد أجرى فان ((Fan, 1998دراســة هدفت إلى مقارنة النظرية التقليدية في القياس مع النماذج ثنائية التدريج من حيث معالم الفقرات وقدرات المفحوصـين، بالإضـافة إلى التحقق من مدى اسـتقرار معالم الفقرات عبر العينات المختلفة.
وقد أظهرت نتائج الدراســـة الخاصـــة بالمقارنة بين نماذج نظرية اســـتجابة الفقرة من أنه وبالرغم من أن النموذج أحادي المعلمة أقل النماذج مطابقة للبيانات، إلا أن تقديرات هذا النموذج كانت أكثر استقرارًا من تقديرات النموذجين الثنائي والثلاثي المعلمة. ()

وقام فان دي فيجفر (Van de Vijver, 1986) بدراسة هدفت إلى التحقق من فعالية نموذج راش تحت ظروف توليد البيانات عن طريق المحاكاة. تم استخدام ثلاثة مستويات لطول الاختبار(10، 25، 50) فقرة، وثلاثة مستويات لحجم العينة(25، 50، 500) فرد.
وتم تقييم دقة التقدير لنموذج راش باستخدام المتوسط الحسابي لمعاملات ارتباط قدرة الفرد الحقيقية والمقدرة، وكذلك بين معلمة الفقرة الحقيقية والمقدرة، إضافة إلى إيجاد الانحراف المعياري للبيانات المقدرة الخاصة بالأفراد والفقرات.
وقد أشارت نتائج الدراسة إلى أنّ نموذج راش قد قدم تقديرات أفضل لقدرة الفرد ومعامل صعوبة الفقرة عند استخدام حجم عينة قليل(25، 50) فردًا، وعددًا قليلًا من الفقرات.
أجرى هامبلتون وتروب (Hambleton and Traub,1971) دراسة هدفت إلى مقارنة النموذج اللوغاريتمي ذي المعلمة الواحدة مع النموذج ذي المعلمتين والنموذج ذي الثلاث معلمات، وذلك باستخدام بيانات تم توليدها بطريقة المحاكاة(Simulation Data). وقد تمت المقارنة بالاعتماد على منحنى خصائص الفقرة، لتحديد دقة التقدير. وأشارت نتائج الدراسة إلى أنّ النموذج ذا الثلاث معلمات كان الأفضل في تقدير قدرة الفرد؛ وذلك لاحتوائه على معلمة التخمين خاصة مع الأفراد ذوي القدرات المتدنية.
يلاحظ من خلال العرض السابق للدراسات السابقة أنها تناولت العديد من الجوانب، منها ما تعلق بتحليل البيانات والحصول على الموضوعية والدقة في تقدير قدرات الأفراد، وكذلك تقديرات معالم فقرات الاختبار، ودالة معلومات الاختبار، باستخدام النماذج المختلفة لنظرية استجابة الفقرة، وكذلك المقارنة بين النماذج اللوجستية المختلفة في نظرية استجابة الفقرة، مستخدمةً بذلك العديد من البرامج الإحصائية المختلفة لتحليل البيانات، سواء أكانت بيانات حقيقية أو مُولّدة، ولكن ما زالت هناك جوانب أخرى تحتاج إلى مزيد من البحث والدراسة المعمقة، وهذا ما انفردت به الدراسة الحالية والتي تتناول استخدام النموذج اللوجستي ثلاثي المعلمة أحادي البعد في تحليل فقرات اختبار التقييم والتقويم لدى عينة من طلبة الجامعة العربية المفتوحة في الأردن، مستخدمةً بذلك برنامج(Bilog – Mg3) في الحصول على تقديرات لمعالم قدرات الطلبة والأخطاء المعيارية في تقديرها، وكذلك إيجاد تقديرات معالم فقرات الاختبار(الصعوبة، والتمييز، والتخمين) والأخطاء المعيارية في تقديرها، بالإضافة إلى استخدام برنامج(SPSS) وذلك للحصول على بعض المؤشرات الإحصائية.

منهج البحث

استخدم أسلوب المنهج الوصفي التحليلي، من خلال التحقق تجريبيًّا من مدى تحقق افتراضات نظرية استجابة الفقرة، وكذلك تحليل فقرات بنك الأسئلة المحوسب (اللغة العربية

لمعلمي المرحلة الابتدائية 1+2) باستخدام النموذج اللوجستي ثلاثي المعلمة، واستخدام المؤشرات الإحصائية المختلفة لتقديم أدلة حول صلاحية بنك الأسئلة، وكذلك إجراء مقارنة بين النموذج اللوجستي أحادي وثلاثي المعلمة، من حيث تقديرات معالم الفقرات ومعالم قدرات المفحوصين والأخطاء المعيارية في تقديرها.

مجتمع البحث وعينته

تكون مجتمع الدراسة من جميع طلبة الجامعة العربية المفتوحة والذين تقدموا للامتحان النهائي لمادة (اللغة العربية لمعلمي المرحلة الابتدائية 1+2) والبالغ عددهم (125) طالبًا وطالبة والذين يمثلون عينة التحليل.

أدوات البحث

ولتحقيق أهداف الدراسة جرى الاعتماد على فقرات بنك أسئلة الامتحان النهائي المحوسب لمادة (اللغة العربية لمعلمي المرحلة الابتدائية 1+2) والمكون من(862) فقرة من نوع الاختيار من متعدد وصح وخطأ.

جمع البيانات.

جرى تطبيق بنك أسئلة الامتحان النهائي المحوسب خلال الفصل الأول للعام الجامعي 2024/2023. حيث جرى تطبيق بنك الأسئلة بحيث يستجيب كل طالب على (50) فقرة من فقراته والتي توزع عشوائيًّا على الطلبة من بنك الأسئلة.

وبعد تطبيق بنك الأسئلة على أفراد عينة الدراسة، جرى تصحيح فقراته إلكترونيًّا، وجرى إدخال البيانات إلى ذاكرة الحاسوب، واستخدم البرنامج الإحصائي (SPSS)، وبرنامج(BILOG –MG3) لإجراء التحليلات الإحصائية اللازمة للإجابة عن أسئلة الدراسة.

الأساليب الإحصائية

جرى استخدام التحليل العاملي بطريقة المكونات الرئيسة (Principal Components Analysis) ؛ وذلك للكشف عن مدى تحقق افتراض أحادية البعد في بنك الأسئلة، وجرى تقدير معالم فقرات بنك الأسئلة باستخدام برنامج (BILOG – MG3).

نتائج البحث ومناقشتها

النتائج المتعلقة بسؤال الدراسة الأول ومناقشتها: ما مدى تحقق افتراضات نظرية استجابة الفقرة في بنك الأسئلة المحوسب لمادة (اللغة العربية لمعلمي المرحلة الابتدائية 1+2)؟

أولاً: النتائج المتعلقة بالتحقق من افتراض أحادية البعد(Unidimensionality):

تمّ التحقق من الافتراض باستخدام المؤشرين الآتيين:

- المؤشرات التي اعتمدت التحليل العاملي للمكونات الرئيسة. (Principal Components Analysis)

تمّ إجراء تحليل عاملي استكشافي للبيانات، وأفرز التحليل (12) عاملًا قيمة الجذر الكامن لكل منها تزيد عن واحد، ويبين الجدول (1) قيم الجذور الكامنة، ونسبة التباين المفسر لكل عامل، وكذلك نسبة التباين المفسر التراكمية.

جدول (1): نتائج التحليل العاملي لاستجابات الطلبة عن فقرات بنك الأسئلة.

رقم العامل	الجذر الكامن	نسبة التباين المفسر %	نسبة التباين المفسر التراكمية %
1	20.339	40.679	40.679
2	2.570	5.139	45.818
3	2.257	4.514	50.332
4	2.092	4.184	54.516
5	1.859	3.718	58.234
6	1.723	3.446	61.680
7	1.579	3.157	64.838
8	1.476	2.952	67.790
9	1.369	2.739	70.528
10	1.188	2.376	72.905
11	1.162	2.324	75.228
12	1.105	2.210	77.439

ويلاحظ من نتائج الجدول (1) أنّ قيمة الجذر الكامن للعامل الأول (20.339)، ويفسر ما نسبته (40.679%) من التباين الكلي، ويعدّ الاختبار أحادي البعد إذا كان نسبة ما يفسره العامل الأول أكثر من (20%) تقريبًا (Reckase, 1979: 207-230)، وعندما نتكلم عن نسبة التباين، فإنّ نسبة التباين الذي يفسره العامل الأول، هي (40.679) من التباين الكلي، وللعامل الثاني (5.139) وهكذا لبقية العوامل، وبهذا المعيار يتبين أنّ بنك الأسئلة يحقق افتراض أحادية البعد.

وتُعتَمد في التحليل العاملي أحادية البعد، من خلال نسبة الجذر الكامن للعامل الأول إلى الجذر الكامن للعامل الثاني، وتكون نسبة كبيرة لا تقل عن (2) (Hambleton and Swaminathan, 1985: 157)، ويتضح من الجدول (1) أنّ ناتج قسمة قيمة الجذر الكامن للعامل الأول، على قيمة الجذر الكامن للعامل الثاني تساوي (7.91)، وهي نسبة تزيد عن المعيار (2)، وعند النظر إلى نسبة الفرق بين الجذر الكامن للعامل الأول والجذر الكامن للعامل الثاني، إلى الفرق بين الجذر الكامن للعامل الثاني والجذر الكامن للعامل الثالث، تبين أنّ النسبة كبيرة وتساوي (56.77)، وهذا مؤشر آخر على أن فقرات بنك الأسئلة تحقق افتراض أحادي البعد، وبالتالي يمكن الاعتماد على فقراته في الوصول إلى قرارات سليمة.

- **المؤشرات التي اعتمدت على قيم معاملات الارتباط.**

جرى حساب معامل الارتباط بين الأداء على الفقرة والأداء الكلي على فقرات بنك الأسئلة بعد حذف الفقرة Corrected Item Total Correlation (rit))، وقد تراوحت القيم بين(0.33 – 0.69)، ويلاحظ أنّ جميع قيم معاملات الارتباط تزيد عن (0.20)، وتدل هذه القيم لمعاملات الارتباط لفقرات الاختبار وفق محك نناليNunnally, 1970, cited in)) Hattie, 1985: 141 أنّ الفقرات تقيس بعدًا واحدًا.

وأظهرت القيمة العالية لمعامل ثبات الاتساق الداخلي (KR – 20) (0.93) لبنك الأسئلة، مؤشرًا آخرَ على تحقق افتراض أحادية البعد وفقًا لرأي كرونباخ (Cronbach, 1951)، حيث أشار كرونباخ إلى أنّ قيمة معامل الثبات تعدّ مؤشرًا جيدًا للتحقق من أحادية البعد؛ لأنه يمثل متوسط كل المعاملات النصفية الممكنة، إضافةً إلى أنّ قيمة معامل الثبات المحسوبة تمثل القيمة المتوقعة لنسبة التباين المفسر من العوامل المشتركة بين الفقرات، وذلك عند ارتباط عينتين عشوائيتن من تجمع فقرات بنك الأسئلة، حيث أشار كرونباخ على أنه حتى يكون بنك الأسئلة قابلًا للتفسير، لابُدّ أنّ يُفسِّر العامل الأول نسبة كبيرة من التباين، وأوضح

إلى أنّ تقدير قيمة معامل الثبات تشير إلى تقدير نسبة التباين العائدة إلى العوامل المشتركة بين فقرات بنك الأسئلة؛ إذ يمثل الحد الأدنى لنسبة التباين لبنك الأسئلة، فقيمة هذا المعامل المرتفعة، تقدم مؤشرًا جيدًا لتحقق افتراض أحادية البعد لبنك الأسئلة.

ثانيًا: النتائج المتعلقة بالتحقق من افتراض الاستقلال الموضعي(Local Independence).
تم الاكتفاء بالتحقق من افتراض أحادية البعد للإشارة إلى تحقق افتراض الاستقلال الموضعي لبنك الأسئلة(Hambleton and Swaminathan, 1985: 22 - 25).

ثالثًا: النتائج المتعلقة بالتحقق من افتراض خاصية المنحنى المميز للفقرة أو المفردة Item Characteristic Curve (ICC).
إنّ احتمال إجابة الطالب على فقرات بنك الأسئلة إجابة صحيحة يزداد بازدياد قدرة الطالب.

رابعًا: النتائج المتعلقة بالتحقق من افتراض عامل السرعة(Non Speededness).
جرى التحقق من أنّ عامل السرعة لم يلعب دورًا في إجابة الطلبة عن فقرات بنك الأسئلة، بمعنى أنّ إخفاق الطالب في الإجابة عن فقرات بنك الأسئلة، يعود إلى انخفاض قدرته، وليس إلى تأثير عامل السرعة في الإجابة، وكان ذلك عن طريق إعطاء الطلبة الوقت الكافي للإجابة عن فقرات بنك الأسئلة والذي قدر بـ(60) دقيقة.

نتائج السؤال الثاني ومناقشتها وتفسيرها: ما معامل الثبات التجريبي لبنك الأسئلة المحوسب لمادة (اللغة العربية لمعلمي المرحلة الابتدائية 1+2)؟
تمّ تقدير معامل الثبات باستخدام النظرية التقليدية في القياس، ونظرية استجابة الفقرة، وقد بلغت قيمة معامل الثبات عند تقديره باستخدام معامل كودر – ريتشاردسون (KR – 20) (0.81). ويمكن تفسير ذلك إلى ارتفاع قيم معاملات التمييز لفقرات بنك الأسئلة، حيث تمتعت فقرات بنك الأسئلة بمعاملات تمييز جيدة.

وباستخدام برمجية (BILOG–MG) لتقدير الثبات الإمبريقي (التجريبي) (Empirical Reliability) ، فكانت قيمة معامل الثبات (0.90)، وهذا يشير إلى أنّ بنك الأسئلة تمتع بدرجة ثبات عالية في كلا النظريتين. ويمكن القول أنّ بنك الأسئلة المُعدّ يمتاز بخصائص

سـيكومترية جيدة، وأنّ جودته تزيد بعد مراجعة بعض الفقرات، وبالتالي يمكن الاعتماد على بنك الأسئلة في الحصول على بيانات دقيقة لاتخاذ قرارات تتعلق بأداء الطلبة.
وفي ما يتعلق بقيمة معامل الثبات التجريبي، فقد كانت القيمة متحيزة إلى أعلى في بنك الأسئلة، ويمكن تفسير ذلك من خلال طريقة استخراج الثبات التجريبي لبنك الأسئلة، إذ تعتمد طريقة الحساب بإيجاد ناتج قسمة تباين العلامة الحقيقية على تباين العلامة الكلية، حيث تباين العلامة هو تباين العلامات المستخرج من خلال الأرجحية العظمى، أما التباين الحقيقي فيتم تقديره عن طريق طرح تباين الخطأ من تباين العلامة الكلية، وبالتالي يمكن عزو ارتفاع معامل الثبات التجريبي لبيانات بنك الأسئلة؛ إلى ارتفاع تباين الخطأ مقارنةً بتباين العلامة.

نتائج السؤال الثالث ومناقشتها وتفسيرها: ما الخصائص الإحصائية لفقرات بنك الأسئلة المحوسب لمادة (اللغة العربية لمعلمي المرحلة الابتدائية 1+2) في ضوء نظرية استجابة الفقرة؟

جرى فحص حسن المطابقة للبيانات حسب النموذج اللوجستي ثلاثي المعلمة؛ بهدف اختيار الطلبة المطابقين للنموذج اللوجستي ثلاثي المعلمة، وكذلك اختيار فقرات بنك الأسئلة المطابقة لهذا النموذج، وقد أفرزت نتائج التحليل لأول مرة على البيانات الخام، باستخدام اختبار مربع كآي عند مستوى الدلالة(=0.01 □)، عدم مطابقة استجابات مجموعة من الطلبة، حيث كانت قيمة الاحتمالية أقل من P(0.01) إذ تمّ حذف استجابة (8) طلاب لم تتطابق استجابتهم للنموذج اللوجستي ثلاثي المعلمة.
وبعد ذلك جرى إعادة التحليل لفقرات بنك الأسئلة لاختبار مدى مطابقة الفقرات للنموذج اللوجستي ثلاثي المعلمة، وقد لوحظ أنّ جميع فقرات الاختبار المختارة وعددها (43) فقرة تتطابق مع النموذج اللوجستي ثلاثي المعلمة.

وبناءً على ذلك تم اعتماد عينة الطلبة(117) طالبًا، وعينة الفقرات (42) فقرة، كأساس لإجراء التحليلات الإحصائية المتعلقة بالمؤشرات الإحصائية للفقرات.
ويبين الجدول (2) الخصائص الإحصائية للفقرات بعد حذف الفقرات غير المطابقة للنموذج اللوجستي ثلاثي المعلمة ممثلة بمعلمة الصعوبة والتمييز والتخمين والأخطاء المعيارية في ضوء النظرية الحديثة في القياس.

جدول (2): قيم معلمة الصعوبة والتمييز والتخمين والاخطاء المعيارية في تقديرها للفقرات بصورتها النهائية.

رقم الفقرة	معامل الصعوبة	الخطأ المعياري	معامل التمييز	الخطأ المعياري	معامل التخمين	الخطأ المعياري	رقم الفقرة	معامل الصعوبة	الخطأ المعياري	معامل التمييز	الخطأ المعياري	معامل التخمين	الخطأ المعياري
1	-0.28	0.641	0.67	0.212	0.47	0.053	22	1.601	0.346	1.511	0.737	0.3	0.046
2	1.211	0.407	1.073	0.424	0.388	0.024	23	0.561	0.292	1.389	0.51	0.44	0.071
3	1.69	0.407	1.315	0.61	0.364	0.011	24	0.201	0.281	1.425	0.438	0.316	0.076
4	1.31	0.195	1.93	0.906	0.274	0.003	25	0.311	0.238	1.59	0.521	0.28	0.069
5	0.24	0.373	1.186	0.406	0.403	0.043	26	0.451	0.092	2.87	1.052	0.202	0.043
6	1.28	0.272	1.531	0.699	0.33	0.012	27	0.451	0.062	3.81	1.855	0.188	0.039
7	1.54	0.352	1.318	0.586	0.312	0.011	28	0.501	0.216	1.418	0.392	0.235	0.09
8	1.74	-0.01	1.247	0.579	0.361	0.01	29	0.571	0.121	2.19	0.955	0.18	0.05
9	1.00	0.432	0.966	0.354	0.385	0.031	30	0.353	0.067	3.477	1.59	0.13	0.039
10	1.00	0.441	0.898	0.31	0.358	0.032	31	0.516	0.109	2.713	1.386	0.26	0.05
11	0.61	0.513	0.778	0.235	0.385	0.041	32	0.67	0.076	2.951	1.288	0.171	0.038
12	0.53	0.175	1.677	0.481	0.213	0.014	33	0.542	0.099	2.605	0.958	0.195	0.043
13	0.57	0.197	1.555	0.488	0.24	0.019	34	0.731	0.105	2.3	0.811	0.18	0.041
14	0.35	0.124	2.507	1.004	0.211	0.009	35	1.362	0.207	2.228	1.25	0.316	0.042
15	0.245	0.127	2.865	1.26	0.212	0.01	36	1.102	0.205	1.64	0.681	0.265	0.051
16	0.427	0.094	2.805	1.268	0.182	0.004	37	1.541	0.4	1.046	0.388	0.225	0.01
17	0.264	0.083	3.79	2.86	0.16	0.000	38	1.491	0.48	1.004	0.396	0.263	0.054
18	0.276	0.109	2.709	1.036	0.157	0.004	39	1.576	0.277	1.388	0.545	0.161	0.04
19	0.452	0.321	1.21	0.387	0.336	0.035	40	2.665	0.847	1.013	0.436	0.168	0.037
20	0.5	0.259	1.42	0.475	0.29	0.029	41	1.975	0.802	0.63	0.17	0.364	0.065
21	-0.33	0.041	3.77	0.18	0.19	0.055	42	2.616	0.880	0.59	2.90	0.49	0.019
المتوسط الحسابي								0.87	0.28	1.83	0.81	0.28	0.035
الانحراف المعياري								0.69	0.22	0.91	0.62	0.09	0.022

- **ما قيم معاملات صعوبة فقرات بنك الأسئلة والأخطاء المعيارية في تقديرها؟**

لا يوجد معيار محدد لصعوبة الفقرة في نظرية استجابة الفقرة، فالصعوبة قيمة نسبية ربما تتغير من اختبار إلى آخر، إلا أنّ الفقرات التي يزيد معامل صعوبتها عن (1.5+) تعّد صعبة، بينما الفقرات التي يقل معامل صعوبتها عن (1.5-) تعد سهلة.

ويلاحظ من الجدول (2) أن قيم معلمة الصعوبة للفقرات تراوحت بين (2.66 - -0.30)، بمتوسط حسابي (0.87)، وانحراف معياري (0.69) وتشير هذه النتيجة إلى أنّ بعض فقرات بنك الأسئلة كانت صعبة.

ويمكن تفسير صعوبة بعض فقرات بنك الأسئلة، إلى وجود بعض الغموض في بنية هذه الفقرات، والتي تضلل الطلبة في الوصول إلى الإجابة الصحيحة، إضافةً إلى افتقار بعض الطلبة إلى المعرفة الكافية للإجابة عن هذه الفقرات، مما جعل بعض الفقرات في نموذج بنك الأسئلة تمتلك معالم صعوبة عالية؛ أي أن فقراتها صعبة.

وفي ما يتعلق بالأخطاء المعيارية لتقديرات معلمة الصعوبة لفقرات بنك الأسئلة المختارة فقد تراوحت قيم الأخطاء المعيارية بين (0.88–0.00)، إذ بلغ قيمة المتوسط الحسابي للأخطاء المعيارية لتقديرات معلمة الصعوبة (0.28)، بانحراف معياري (0.22).
ويلاحظ أنّ قيم الأخطاء المعيارية لتقديرات معلمة الصعوبة لفقرات بنك الأسئلة قد كانت منخفضة، وهذا يدل على أنّ فقرات بنك الأسئلة أكثر دقة في قياس تحصيل الطلبة، علاوةً على أنّ متوسط الأخطاء المعيارية لتقديرات معلمة الصعوبة القليل، يشير إلى دقة تقدير صعوبة الفقرة.

- **ما قيم معاملات تمييز فقرات بنك الأسئلة والأخطاء المعيارية في تقديرها؟**

تشير النتائج الخاصة بمعلمة التمييز لفقرات بنك الأسئلة إلى أنها كانت جميعها موجبة، ويكون معامل تمييز الفقرة موجبًا، إذا كان احتمال إجابة هذه الفقرة من قبل الطلبة ذوات التحصيل العالي أكبر من احتمال إجابة الطلبة ذوات التحصيل المنخفض (Richichi, 1996)، وكلما زادت قيمة تمييز الفقرة، كان ذلك أفضل ودليلًا على جودتها.
وتراوحت قيم التمييز للفقرات بين (0.59-3.81)، بوسط حسابي قدره (1.83)، وانحراف معياري مقدره (0.91)، وأشار كوي وبكمان (Choi and Bachman, 1992) إلى أنه إذا كانت قيمة تمييز الفقرة أكبر من (1.0)، فهذا دليل على أنّ الفقرة جيدة، أما إذا كانت قيمة التمييز للفقرة بين(1.0–0.70)، فتمييز الفقرة مقبول، وتقل جودة الفقرة إذا قلت قيمة تمييزها عن (0.70)، وباستخدام هذا التصنيف يمكن القول أنّ جميع فقرات بنك الأسئلة المختارة عشوائيًّا تتمتع بمعاملات تمييز جيد.
وفي ما يتعلق بالأخطاء المعيارية لتقديرات معلمة التمييز، فقد بلغت قيمة المتوسط الحسابي للأخطاء المعيارية لتقديرات معلمة التمييز للفقرات (0.81)، بانحراف معياري مقداره (0.62)، وتراوحت قيم الأخطاء المعيارية لتقديرات معلمة التمييز بين (2.90–0.17).

- **ما قيم معاملات التخمين لفقرات بنك الأسئلة والأخطاء المعيارية في تقديرها؟**

تراوحت قيم معاملات التخمين للفقرات بين (0.13-0.49)، بوسط حسابي قدره (0.28)، وانحراف معياري قدره (0.09).
ويلاحظ أن بعض قيم معلمة التخمين لبعض الفقرات كانت أعلى من (0.25)، وهي بمثابة احتمال إجابة الفقرة إجابة صحيحة من قبل المفحوصين ذوي القدرة المتدنية، ويشير هاريس (Harris, 1972) إلى أنّ قيم معامل التخمين تتراوح بين (0.0 –1) نظريًّا، إلا أنّ القيم العملية يفضل أن تكون أقل من (0.30)؛ ويمكن تفسير ارتفاع قيم معلمة التخمين لبعض الفقرات عن (0.30)، على أساس النزعة للمجازفة بالتخمين عند بعض الطلبة.

ويمكن تفسير ارتفاع معلمة التخمين لبعض فقرات بنك الأسئلة إلى خصوصية بعض أسئلة فقرات الاختبار، من حيث طبيعة المعرفة والقدرة التي يستوعبها، والتي قد تختلف من سؤال إلى آخر، من حيث التراكيب أو العبارات الخاصة، والتي تعتمد على نوع من المهارات لدى كاتب الأسئلة، والتي تجعل المموه على درجة كبيرة أو قليلة من الوضوح، وهذا يختلف من مفحوص إلى آخر، ومن فقرة إلى فقرة أخرى، الأمر الذي أدى إلى ارتفاع معلمة التخمين لبعض فقرات الاختبار.

وفي ما يتعلق بالأخطاء المعيارية لتقديرات معلمة التخمين، فقد بلغت قيمة المتوسط الحسابي للأخطاء المعيارية لتقديرات معلمة التخمين لفقرات نموذج الاختبار (0.035)، بانحراف معياري مقداره (0.022)، وقد تراوحت القيم بين (0.09–0.00). وتعدّ مثل هذه القيم منخفضة.

نتائج السؤال الرابع ومناقشتها وتفسيرها: ما الخصائص الإحصائية لتقديرات قدرات طلبة الجامعة العربية المفتوحة والأخطاء المعيارية في تقديرها في بنك الأسئلة المحوسب (اللغة العربية لمعلمي المرحلة الابتدائية 1+2)؟

جرى تقدير معلمة القدرة باستخدام طريقة الأرجحية العظمى الهامشية (MML) بالاعتماد على برنامج (BILOG–MG)، ويبين الجدول (3) جدول تكراري لقيم قدرات الطلبة:

جدول (3): قيم تقديرات قدرات الطلبة وفق نظرية استجابة الفقرة

مدى قدرات الطلبة	التكرار
-1.30 - -2.30	23
0.00 - -1.27	49
1.30 – 0.01	38
2.30 – 1.31	15
عدد الطلبة	**125**

ويلاحظ من الجدول (3) أن قيم تقديرات قدرة الطلبة الناتجة من تطبيق بنك الأسئلة قد تراوحت بين (2.30 - -2.30).
وفي ما يتعلق بالأخطاء المعيارية لتقديرات قدرات الطلبة، فقد تراوحت قيم الأخطاء المعيارية بين (0.150–4.191)، وبلغت قيمة المتوسط الحسابي لتقديرات الأخطاء المعيارية لمعلمة القدرة الناتجة من تطبيق بنك الأسئلة (0.89) بانحراف معياري (1.67)، وهذا يشير إلى أنّ قيمة الأخطاء المعيارية لتقديرات القدرة كانت مرتفعة.
وتجدر الإشارة في هذا المجال إلى أنّ النتائج الخاصة بمعالم فقرات بنك الأسئلة وقدرات الطلبة تدعم ما أشارت إليه دراسة كل من((Hambleton and Traub, 1971، و(ياسين، 2004) من حيث إنّ النموذج ثلاثي المعلمة أفضل في تقدير قدرة الفرد؛ وذلك لاحتوائه على معلمة التخمين، بالإضافة إلى تقديرات معالم الفقرة.

التوصيات

إنّ النتائج المتعلقة بفعالية النموذج اللوجستي ثلاثي المعلمة في تحليل فقرات بنك الأسئلة، تقدم مبررًا قويًّا لمستخدمي بنوك الأسئلة باستخدام هذا النموذج عند تحليل بنوك الأسئلة، وكذلك استخدام بنك أسئلة امتحان (اللغة العربية لمعلمي المرحلة الابتدائية 1+2) في الحصول على نتائج صادقة ودقيقة.
ولكن يبقى هناك مجالًا كبيرًا للبحث في هذا المجال، وبالتحديد أمكن تقديم التوصيات الآتية:

1. إجراء دراسة مقارنة بين النماذج المختلفة لنظرية استجابة الفقرة أحادية البعد، وذلك باختلاف حجم العينة وطول الاختبار المقدم للطالب.
2. دراسة دقة تقديرات قدرة الأفراد ومعالم فقرات بنوك الأسئلة باختلاف النموذج اللوجستي المستخدم.

المصادر والمراجع:

المراجع العربية

بني عطا، زايد (2014). تقصي دقة تقدير النموذج اللوجستي ثلاثي المعلمة لمعالم الفقرة وقدرة الأفراد في ضوء تغير طول الاختبار وحجم العينة. مجلة جامعة الشارقة للعلوم الانسانية والاجتماعية، 11(7).

حسنين، خالد (2017) ملاءمة استخدام الاختبارات الإلكترونية في التعلّم المفتوح. رسالة دكتوراه غير منشورة، جامعة السودان للعلوم والتكنولوجيا، السودان.
السامرائي، محمد والخفاجي، أحمد.(2012). بناء اختبار تحصيلي محكي المرجع في مادة علم نفس الخواص لطلبة أقسام العلوم التربوية والنفسية. مجلة الاستاذ، 3(20)964 – 1002.
الشريفين، نضال.2005م. تقدير الخصائص السيكومترية لاختبار محكي المرجع في القياس والتقويم التربوي وفق النظرية الحديثة في القياس. مجلة أبحاث اليرموك، جامعة اليرموك، إربد، الاردن.
الشمراني، محمد. (2016). أثر حجم العينة وطول الاختبار والتفاعل بينهما على تقدير معلمة الصعوبة والتمييز والتخمين والخطأ المعياري باستخدام نظرية استجابة الفقرة. مجلة كلية التربية، جامعة طنطا. 64(4).
الصوص، سمير. (2022). اتجاهات الطلبة نحو الاختبارات الإلكترونية وآراؤهم فيها. المجلة التربوية، جامعة الكويت، 36 (144)، 255-291.
العيثاوي، محمد. (2009). استخدام أنموذج ا رش على وفق نظرية السمات الكامنة في بناء اختبار تحصيلي في مادة علم النفس الفروق الفردية. رسالة ماجستير غير منشورة، جامعة بغداد، كلية التربية، ابن رشد.
عودة، أحمد سليمان. (2010). القياس والتقويم في العملية التدريسية. اربد، الأردن: دار الأمل.
علاّم، صلاح الدين. (1985). تحليل بيانات الاختبارات العقلية باستخدام نموذج راس اللوغاريتمي الاحتمالي: دراسة تجريبية. المجلة العربية للعلوم الإنسانية، 17، 100 – 124.
علاّم، صلاح الدين. (1987). دراسة موازنة ناقدة لنماذج السمات الكامنة والنماذج الكلاسيكية فى القياس النفسى والتربوى. جامعة الكويت، المجلة العربية للعلوم الإنسانية، 7(22).
القضاة، محمد. (2023). تأثير مستوى طول الاختبار في دقة تقدير معلمتي الصعوبة والتمييز باستخدام النموذج ثنائي المعلمة في مبحث الرياضيات للصف الثامن الأساسي، مجلة كلية التربية، جامعة أسيوط،2(39).
ياسين، عمر صالح مفضي. (2004). الخصائص السيكومترية لاختبار محكي المرجع في الكيمياء لطلاب الصف الأول الثانوي العلمي مقدرة وفق النظريتين الكلاسيكية والحديثة في القياس. رسالة ماجستير غير منشورة، جامعة عمان العربية للدراسات العليا، عمان.
الوديان، ناريمان؛ الدلالعة، أسامة؛ عبابنة، زياد. (2019). أثر اختلاف طريقة عرض الاختبار الإلكتروني على معاملات الصعوبة والتمييز وتحصيل طلبة الصف العاشر في مبحث الحاسوب، مجلة الجامعة الإسلامية للدراسات التربوية والنفسية. 5 (27).

المراجع الأجنبية:

Adams, R.(1988). Applying the partial credit model to educational diagnosis. Applied Measurement in Education.1(4),347 - 361.
Choi, I., Bachman, F. (1992). An investigation into the adequacy of three IRT models for data from two EFL reading tests. Language Testing, 9(1), 51 – 78.
Crocker, L., and Algina, J. (1986). Introduction to classical and modern test theory. Harcourt Brace Jovanovich: Fort Worth, TX.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. Psychometrika, 16(3), 297 – 334.
Embretson, S. E. and Reise, S. P. (2000). Item response theory for psychologists. Lawrence Erlbaum Associates, Inc. New Jersey.
Fan, X.(1998). Item response theory and classical test theory. An empirical comparison of their item /person statistics. Educational and Psychological Measurement, 58(3), 257-381 .
Gay, L. (1985). Educational Evaluation and Measurement, 2nd ed.. Chales: E. Merrill Pub., Colombous.
Gronlund, N. and Linn, R. (1990). Measurement and Evaluation in Teaching, 6th ed.Macmillan Publishing Company: New York; PP. 166 - 189.
Hambleton, R. and traub, R.(1971). Information Curves and Efficiency of Three Logistic Test models. British Journal of Mathematical and Statistical Psychology, 24(2), 273 - 281 .
Hambleton, R. K. (1983). Item response theory. Kluwer Nijhoff publishing: Boston.
Hambleton, R., and Swaminathan, H. (1985). Item Response Theory: principles and applications. Kluwer Nijhoff publishing: Boston .
Harris, C. W. (1972). An index of efficiency for fixed length mastery test. Paper presented to the American Educational Research Asociation. Chicago, Illinois, April 4, 1972.(ERIC Document Reproduction Service No. ED 064349); 1 – 10.
Hattie, J. (1985). Methodology Review: Assessing Unidimensionality of Tests and Items. Applied Psychological Measurement. 9 (2) 139 – 164.
Hwang, D.Y. (2002). Classical Test Theory and Item Response Theory: Analytical and Empirical Comparison. Paper presented at the Annual Meeting of the Southwest Educational Research Association, Austin, TX. (ERIC Document Reproduction Service No. ED 466779).
Kelkar, V., Wightman, L.P. and Luecht, R.M. (2000). Evaluation of the IRT parameter invariance property for the MCAT. Paper presented at the Annual Meeting of National Council on Measurement in Education, New Orleans, LA (ERIC Document Reproduction Service No ED. 443830.
Lord, F. M. (1980). Applications of item response theory to practical testing problems. Erlbaum Associates: New York.
Nunnally, J. C . (1970). Introduction to Psychological measurement. McGraw – Hill : New York.

Plake, B., Thompson, A., and Lowry, S. (1981). Effects of item arrangement and test anxiety on two scoring methods. The Journal of Experimental Education, 49(4), 214 – 219.
Reckase. D. (1979). Unifactor latent trait models applied to multifactor tests: Results and implications. Journal of Educational Statistics, 4(3), 207 – 230.
Richichi, R. (1996). An analysis of test bank multiple choice items using item response theory. Author; SLD. (ERIC Document Reproduction Service NO. ED 405367).
Stage, C. (2003). Classical Test Theory or Item Response Theory : The Swedish Experience, Umea University,pp: 1 - 30.
Swaminathan, H.; Hambleton, Ronald; Sireci, Stephen and Xing, Dehui.(2003). Small sample estimation in dichotomous item response models: Effect of priors based on judgmental information on the accuracy of item parameter estimates. Journal of Applied Psychological Measurement. 27 (1) 27-51.
Van de Vijver. F.J.R.(1986). The Robustness of Rasch Estimates. Applied Psychological Measurement. 10 (1) 45 – 57.