تقدير معالم بنك أسئلة في ضوء نماذج الاستجابة للفقرة ثنائي وثلاثي المعلم (دراسة مقارنة)

Estimating Item Bank Parameters in Light of the Two-Parameter and Three-Parameter Item Response Models (A Comparative Study)

أ. **فتحي أحمد ناصر الضبيبي:** طالب دكتوراه، قياس وتقويم، كلية التربية، جامعة صنعاء، اليمن.

د. نجاح عبدالرحيم محمد عثمان: أستاذ القياس والتقويم والإحصاء التربوي المشارك، جامعة تعز، اليمن.

Fatehi Ahmad Nasser Adhubibi: Doctoral Researcher, Measurement and Evaluation, Faculty of Education, Sana'a University, Yemen.

Email: fathiahmd1982@gmail.com

Dr. Nagah Abdelrahim Mohammed Othman: Associate Professor of Measurement, Evaluation, and Educational Statistics, Taiz University, Yemen.

Email: adimi.najah.@gmail.com

DOI: https://doi.org/10.56989/benkj.v5i10.1609

تاريخ الاستلام:	2025-07-26	تاريخ القبول:	2025-09-15	تاريخ النشر	2025-10-01

الملخص:

هدفت الدراسة إلى المقارنة بين نماذج الاستجابة للفقرة (ثنائي، ثلاثي) المعلم في تقدير معالم قدرات الأفراد والفقرات، وتضمنت الدراسة عينة مولدة محاكاة لعينة حقيقية ببرنامج (wengen3) تألفت من (3000) استجابة، حيث تم توليد (1000) استجابة لكل صورة اختبارية، وتكونت أداة البحث من بنك أسئلة مكون من ثلاث صور اختبارية متكافئة (**A** ، **B** ، **C**)، كل صورة اختبارية مكونة من (44) فقرة، واستخدم البحث برنامج (Bilog mg3) للتحقق من ملاءمة البيانات للنموذجين (أحادي، ثنائي) المعلم، واستخراج تقدير معالم الأفراد والفقرات ودالة معلومات الاختبار والثبات الإمبريقي، واستخدم البحث برنامج SPSS للتحقق من افتراضات نظرية الاستجابة للفقرة واستخراج الفروق بين نماذج الاستجابة للفقرة (ثنائي، ثلاثي) المعلم، في تقدير معالم الأفراد والفقرات والخطأ المعياري لتقدير القدرة، ومن أبرز النتائج التي توصلت إليها الدراسة، عدم وجود فروق دالة إحصائيًا لبيانات العينة المولدة تعزى لنوع النموذج (ثنائي، ثلاثي) المعلم، في تقدير معلم القدرة للأفراد، ومعالم الفقرات (الصعوبة ـ التمييز)، ووجود فروق دالة إحصائيًا لبيانات العينة المولدة تعزى لنوع النموذج (ثنائي، ثلاثي) المعلم، في الخطأ المعياري لتقدير القدرة لصالح النموذج ثنائي المعلم.

الكلمات المفتاحية: تقدير المعالم، نماذج الاستجابة للفقرة، توليد البيانات، المقارنة بين النماذج، بنك أسئلة.

Abstract:

The study aimed to compare the two-parameter and three-parameter item response models in estimating the parameters of individuals' abilities and item characteristics. A simulated sample, generated to resemble a real sample using the (Wengen3) program, consisted of 3,000 responses, with 1,000 responses generated for each test form. The research instrument comprised an item bank containing three equivalent test forms (A, B, C), with each form consisting of 44 items. The study used the (Bilog-MG3) program to verify the data fit to the one-parameter and two-parameter models, and to estimate item and person parameters, the test information function, and empirical reliability. The SPSS program was used to verify the assumptions of item response theory and to examine the differences between the two-parameter and three-parameter models in estimating individual and item parameters, as well as the standard error of ability estimation. The key findings of the study include: no statistically significant differences in the generated sample data attributed to the type of model (two-parameter vs. three-parameter) in estimating individuals' ability parameters and in estimating both discrimination and difficulty parameters; and statistically significant differences in the standard error of ability estimation in favor of the two-parameter model.

Keywords: Item bank, item response models, estimation parameters, data generation, model comparison.

الإطار المنهجي للدراسة:

المقدمة:

النماذج ثنائية التدريج من أشهر نماذج الاستجابة للفقرة أحادية البعد استخدامًا في بناء الاختبارات والمقاييس إذ يوجد منها ثلاثة نماذج، أولها: النموذج اللوغاريتمي ذو المعلمة الواحدة (One –Parameter Logistic Model))، والمشهور باسم "نموذج راش" (Rasch Model)، ويعد من أكثر النماذج الثنائية استخدامًا، حيث لا يتطلب عمليات حسابية معقدة على الرغم من تشدد افتراضاته، وهي أن الفقرات متساوية في تمييزها ولا يوجد فيها تخمين (Harris, 1989). أما النموذج الثاني فهو النموذج الثنائي اللوغاريتمي ثنائي المعلمة (Tow –Parameter Logistic Model)، وهو يختلف عن النموذج الأحادي بأنه يتيح معلمة التمييز للفقرة إضافة إلى معلمة الصعوبة، ويعد النموذج ثنائي المعلمة أقل عمومية من النموذج الثالث والأخير من بين هذه النماذج وهو النموذج اللوجستي ثلاثي المعلمة (Three – Parameter Logistic Model) (علام، 2005).

ولقد حققت نظرية الاستجابة للفقرة Item Response Theory، والنماذج الرياضية المتعلقة بها تقدمًا حقيقيًا واضحًا وملموسًا في تحقيق الموضوعية المطلوبة منها، كما ظهرت فكرة ضم مفردات من اختبارين في تدريج واحد مشترك، ثم ظهرت فكرة إنشاء بنوك الأسئلة، ومعادلة درجات الاختبارات المسحوبة منها، وبظهور هذه البنوك في مختلف المجالات وخاصة في المجالات النفسية والتربوية أمكن التغلب على أغلب المشكلات وأوجه الضعف والقصور التي ظلت تعاني منها أساليب القياس التقليدي لفترة طويلة، وتحقيق الموضوعية المطلوبة في هذا الصدد (David & Seung, 2010).

واستقلالية القياس إحدى أهم المزايا المرتبطة بنظرية الاستجابة للفقرة، وتعني أن تقدير معالم الفقرات (الصعوبة، والتمييز، والتخمين) يكون مستقلاً عن خصائص الأفراد الذين استخدموا في تقدير هذه المعالم (Sample Free)، وإن تقدير قدرة الفرد يكون مستقلاً عن عينة الفقرات التي تطبق عليه (Item Free)، وبتحقق هذه الخاصية يصبح بالإمكان القيام ببعض التطبيقات مثل: معادلة درجة الاختبار، وبناء بنوك الأسئلة، والقياس التكيفي (Reckase, 2009). ولتحقيق استقلالية القياس، اهتم البحث السيكومتري لنظرية الاستجابة للفقرة بتطوير نماذج احتمالية تقوم على تحديد العلاقة بين أداء الفرد على فقرات الاختبار وبين السمات أو القدرات الكامنة التي تقف وراء هذا الأداء.

وبالرغم من مزايا نظرية الاستجابة للفقرة بنماذجها المختلفة وبخاصة تحرر تقدير معالم الفقرات من تقديرات معالم الأفراد وتحرر معالم الأفراد من تقديرات خصائص الفقرات، بالإضافة إلى إمكانية حساب الخطأ المعياري في التقدير عند كل مستوى قدرة، إلا أنه ينبغي على مطوري الاختبار إعطاء

المزيد من الاهتمام لقضية المطابقة بين بيانات الاختبار ونموذج نظرية الاستجابة للفقرة، وتعبر المطابقة (Fit) عن التوافق بين التكرارات المشاهدة وتلك المتوقعة من النموذج المستخدم لمجموعة من بيانات عينة عدد أفرادها (ن) (Douglas, 1982).

وقبل تطبيق الاختبار الذي يتم إعداده وفقًا لأحد النماذج الرياضية المرتبطة بنظرية الاستجابة للفقرة، من الضروري تقدير معالم كل فقرة داخل هذا الاختبار وفقًا للنموذج الذي تم اختياره، وذلك من خلال تطبيق الاختبارات على عينة كبيرة نسبيًا من الأفراد؛ ثم اختيار نموذج الاستجابة للفقرة المناسب؛ تمهيدًا لاشتقاق معالم: الصعوبة، التمييز، التخمين (Thelma et al., 2010).

إن نماذج الاستجابة للفقرة تقدم لنا مزايا تتعلق بتفسير علامات الاختبار وتقديم تقرير بالنتائج بشكل عملي ومفيد، إلا أن هذه المزايا تتحقق من ناحية عملية فقط عند وجود اتساق بين النموذج المستخدم وبيانات الاختبار، بمعنى أن أهمية اختيار أحد نماذج الاستجابة للفقرة تبرز في قدرته على تفسير البيانات مع افتراضات النموذج، والذي يخدم النموذج، والذي يخدم التطبيق العملي الذي يهتم فيه البحث كمثل بناء الاختبار لأغراض تحديد المسار أو تصنيف الأفراد لغايات معينة، أو حتى اختيار أفضل اختبار من بنك أسئلة فيه عدد من الفقرات المدرجة أو غيرها من التطبيقات (Hambelton & Swaminathan, 1985).

ويعد اختيار النموذج المناسب دالة لكل من القيم المتوقعة لدرجة الفقرة، وعدد السمات الكامنة المؤثرة في الاستجابة، وعدد الفقرات، وحجم العينة، ومجال المحتوى (Panter & Reeve, 2002: 25)، حيث تنقسم نماذج النظرية تبعًا لمستوى الاستجابة (القيم المتوقعة لدرجة الفقرة) إلى نماذج ثنائية وأخرى غير ثنائية، كما تنقسم تبعًا لعدد معالم الفقرة التي يتم تقديرها والافتراضات الإحصائية التي تحكم شكل المنحنى المميز للفقرة، إلى نماذج أحادية المعلم، وثنائية المعلم، وثلاثية المعلم (Bejar, 1983: 31–32) (محمود، 2012).

ويذكر "كروكر وألجينا" أن هناك عدة عوامل تؤثر في اختيار النموذج المناسب، منها مدى واقعية افتراضاته، فيبدو استعمال النموذج ثلاثي المعلم مناسبًا مع فقرات الاختيار من متعدد، والصواب والخطأ، ويبدو النموذج ثنائي المعلم مناسبًا لأنواع أخرى من الفقرات، ومن ناحية أخرى يمكن أن يهمل التخمين في بعض الاختبارات ويحل محله النموذج الثنائي ويكون مناسبًا لتحليل بياناته، وكذلك التمييز يمكن أن يهمل في أي شكل من أشكال الفقرات فيكون النموذج الأحادي هو الأنسب، إذ إن استخدام نموذج معقد في غير موضعه قد يؤدي إلى نتائج أقل دقة (Crocker & Algina, 1986: 361).

ومن أهم العوامل المؤثرة في اختيار النموذج المناسب هو حجم العينة، حيث وجد "لورد" (Lord, 1983) أنه في العينات المكونة من (200) فرد أو أقل، تفوق نموذج راش قليلاً على

النموذج ثنائي المعلم في تقدير الأفراد، ومناسبة عملية تدريج الفقرات وفق النموذجين ثنائي المعلم وثلاثي المعلم، وأن تكون حجم العينة (1000) فرد، كدراسة (العبدالله، 2021) التي هدفت إلى أثر حجم العينة على تقدير بنك الأسئلة، وكانت أحجام العينة (1000، 800، 400) وأشار فيها إلى ارتفاع جودة التدريج كلما زاد حجم العينة. وفي دراسة (Mislevy, 1986) التي قدمت نتائج تشير إلى أن برنامج BILOG يعمل جيدًا في استخراج معالم الفقرات في عينة بحجم (1000).

وفي هذا الصدد يذكر علام (2015) أنه يمكن اختيار نموذج بسيط رغم مخالفة بعض افتراضاته إذا كان ذلك لا يؤثر كثيرًا في دقة التقديرات الناتجة عن استخدام هذا النموذج في تطبيقات معينة، فقد يستخدم "نموذج راش" البسيط بدلاً من النموذج الثلاثي المعلم نظرًا لأن النموذج الثلاثي المعلم يتطلب أن تكون عينة الأفراد لا تقل عن (1000) فرد. ولهذا فإن المحك الأساسي في اختيار نموذج التدريج هو الهدف من الاختبار وتطبيقاته، ومدى تأثير انتهاك افتراضات النموذج على تطبيقات الاختبار، فعلى الرغم من أن النموذج أحادي المعلم يتضمن عددًا كبيرًا من الافتراضات فإنه يؤدي إلى نتائج دقيقة، حتى عندما لا يتحقق افتراض أو أكثر من تلك الافتراضات (Crocker & Algina, 1986: 362)، وفي دراسة (Sinhary & Johson, 2003) نجد أنهما توصلا إلى عدم ملاءمة النموذج الأحادي المعلم بالنسبة للبيانات التي تم الحصول عليها من النموذجين الثنائي والثلاثي المعلم، في حين أن النموذج ثنائي المعلم يتوافق مع تلك البيانات التي تم الحصول عليها من خلال النموذج ثلاثي المعلم، وهذا لا يعني أن النموذج الثلاثي أفضل من النموذج الثنائي أو العكس، أو أن النموذج الأحادي هو الأفضل من النموذجين، ولكن النموذج الأفضل هو النموذج الذي يلائم البيانات بشكل أفضل من النموذجين الآخرين، وعند استخدام النموذج غير المناسب فإن النتائج التي سنحصل عليها من هذا النموذج ستكون موضع شك (إبراهيم، 2009: 296-297).

وهذا ما أثار تساؤلات "كروكر وألجينا" (Crocker & Algina, 1986) حول أي نماذج الاستجابة للفقرة يجب استخدامه في وضع معين، وما هي المعايير التي تبرر استخدام النموذج الثنائي المعلم أو الثلاثي المعلم مثلاً، وفي هذا السياق فقد عرضا فكرة أنه عندما يسمح باستبعاد التخمين في اختبار موضوعي (اختيار من متعدد أو صح وخطأ) فعندها يكون النموذج ثنائي المعلم مناسبًا للاستخدام، بالإضافة إلى ذلك إذا أمكن كتابة الفقرات بطريقة يتم فيها استبعاد التمييز أو التقليل من أثره بحيث يكون لتلك الفقرات نفس مستوى التمييز، فإن النموذج أحادي المعلم كافيًا للاستخدام (حامد، 2008).

ويشير "لورد ونوفيك" (Lord & Novick, 1968) إلى أن أحد أهم القضايا الرئيسة عند استخدام نماذج نظرية الاستجابة للفقرة تكمن في تقدير معالم النموذج، التي تعتمد على أساليب التحليل العددي من خلال استخدام البرامج الحاسوبية المختلفة، وذلك من خلال التقدير الإحصائي

للعلاقات بين احتمال الاستجابة الصحيحة عن فقرة من فقرات الاختبار والقدرة التي يقيسها الاختبار، المشكلة الرئيسة لمستخدم هذه النظرية، حيث بينت ستوكنغ (Stocking, 1990) أن تقدير معالم الفقرة يعد من أهم القضايا التي يعتمد عليها نجاح نظرية الاستجابة للفقرة، خصوصًا في التطبيقات التي تعتمد كثيرًا على تلك المعالم، كالقياس التكيفي، وبنوك الأسئلة وغيرها من التطبيقات يعتمد بشكل كبير على دقة التقديرات لكل من معالم الفقرات ومعلمة القدرة للأفراد، مما جعل البحث السيكومتري يهتم بالبحث عن أفضل أساليب التقدير الإحصائي لمعالم الفقرات وقدرات الأفراد (عطا، 2014) (نصراوين، 2018).

وتعد بنوك الأسئلة أحد أبرز الاتجاهات الحديثة في مجال القياس السلوكي، وبصفة خاصة تلك التي يتم إعدادها في ضوء النماذج اللوغاريتمية المرتبطة بنظرية الاستجابة للفقرة، وعلى الرغم من أن هذه البنوك يعتمد عليها كأحد الوسائل المتطورة في تحقيق الموضوعية المرجوة في هذا المجال، إلا أن هناك بعض المتغيرات التي يمكن أن تؤثر في دقة تدرج فقرات هذه البنوك وكذلك دقة معادلة درجات الاختبارات المختلفة التي يمكن أن تسحب منها (الشافعي، 2014: 431).

2. إشكالية الدراسة:

نظرًا للتوجه المتنامي للباحثين المعاصرين في ميدان القياس التربوي والنفسي إلى نظرية الاستجابة للفقرة لتميزها في تحقيق الموضوعية والدقة في القياس وتلافيها أوجه القصور في نظرية القياس التقليدية، واجه الباحث حيرة في تحديد أي من نماذج الاستجابة للفقرة (أحادي، ثنائي، ثلاثي) المعلم، أكثر ملائمة لبيانات البحث وأكثر دقة في تقدير معالم الأفراد والفقرات؛ بناءً على ما سبق تتحدد مشكلة البحث في الإجابة على التساؤل الآتي: هل توجد فروق ذات دلالة إحصائية في تقدير معالم الأفراد والفقرات بين نماذج الاستجابة للفقرة (ثنائي، ثلاثي) المعلم لبنك الأسئلة.

3. أهداف الدراسة:

هدفت الدراسة إلى المقارنة بين نماذج الاستجابة للفقرة (ثنائي، ثلاثي) المعلم، في تقدير معالم قدرة الأفراد والفقرات لبنك الأسئلة وتدريجه، ويتفرع من هذا الهدف الرئيس الأهداف الفرعية الآتية:

1ـ المقارنة بين نماذج الاستجابة للفقرة (ثنائي، ثلاثي) المعلم، في تقدير معلم القدرة للأفراد، لبنك الأسئلة وفق عينة مولدة.

2ـ المقارنة بين نماذج الاستجابة للفقرة (ثنائي، ثلاثي) المعلم، في تقدير معالم الفقرات (الصعوبة، التمييز)، لبنك الأسئلة وفق عينة مولدة.

3- المقارنة بين نماذج الاستجابة للفقرة (ثنائي، ثلاثي) المعلم، في الخطأ المعياري لتقدير القدرة، لبنك الأسئلة وفق عينة مولدة.

4. أهمية الدراسة:

- تناول بنوك الأسئلة التي تُعد أداة مهمة لضمان تحقيق جودة التقويم والتعليم العالي والاعتماد الأكاديمي، فهي تمثل استثمارًا مهمًا في تطوير العملية التعليمية والارتقاء بمستوى مخرجاتها، كما يساهم في نشر الوعي لمفاهيم بنوك الأسئلة النظرية والتطبيقية.
- يمثل تطبيقًا عمليًا لدراسة جدوى نماذج الاستجابة للفقرة في تحقيق دقة القياس وموضوعيته، مما يشجع الباحثين والمهتمين بالقياس والتقويم استخدام هذه النظرية في أبحاثهم، وفي بناء أدوات القياس وتحليل بياناتها.
- من أوائل الأبحاث على مستوى البيئة اليمنية على حد علم الباحث في تناوله للمقارنة بين نماذج الاستجابة للفقرة في تقدير معالم قدرات الأفراد والفقرات لبنك أسئلة في مقرر القياس والتقويم التربوي وتدريجه، وفيما سيقدمه من إضافة علمية في هذا الاتجاه لميدان القياس والتقويم.
- سلط الضوء على أهمية اختيار نموذج الاستجابة للفقرة الأكثر ملاءمة للبيانات ودقة في تقدير معالم الأفراد والفقرات ودالة المعلومات، لضمان جودة وموضوعية عملية القياس والتقويم.

الدراسات السابقة:

- دراسة ضعضع (2021) والتي هدفت إلى دراسة أثر اختيار أنموذج الاستجابة للمفردة (أحادي، ثنائي، ثلاثي، رباعي) المعلم، وطرائق التقدير (MLP, MAP, EAP)في تقدير معالم المفردة (صعوبة وتمييز وتخمين وعدم اهتمام) والقدرة (للأفراد) ودقتها والكشف عن أثر التفاعل بين هذين المتغيرين، حيث اعتمدت الباحثة على اختبار موضوعي في اللغة الإنجليزية ثنائي الاستجابة (1،0)، وتكونت عينة الدراسة من (1000) من الطلبة المتقدمين لاختبارات القبول من كليات نظرية وتطبيقية، وبعد تأكد الباحثة من تحقق افتراضات نظرية الاستجابة للفقرة، قامت الباحثة من ملائمة المفردات للنماذج باستخدام برنامج Jmetrik، وتوصلت نتائج الدراسة إلى أنه لا يوجد فروق ذات دلالة إحصائية في نسبة المفردات الملائمة بين النماذج ثنائي وثلاثي ورباعي المعلم، في حين أظهرت النتائج دلالة الفروق في نسبة المفردات الملائمة مع الأنموذج أحادي المعلم، ووجود فروق ذات دلالة إحصائية في قيمة معامل صعوبة المفردة (لم تظهر فروق دالة بين النموذج الأحادي المعلم والثنائي المعلم) ودقتها والتمييز ودقته وقيمة معامل التخمين، في حين أظهرت النتائج وجود فروق ذات دلالة إحصائية في دقة تقدير قدرة الأفراد تبعًا للأنموذج المستخدم لصالح الأنموذج الأكثر معالم.
- دراسة صالح (2020) والتي هدفت إلى فحص أثر اختيار نموذج الاستجابة للفقرة ثنائية الاستجابة (أحادي، ثنائي، ثلاثي) المعلم، في تقدير معالم الفقرة (الصعوبة والتمييز والتخمين) ودقتها، حيث طبقت الباحثة اختبار عكس الأرقام لاختبار القدرات المعرفية "وودكوك جونسون"

(البطارية المعيارية) والمكون من 30 فقرة ثنائية الاستجابة، على عينة من طلاب الصف السادس الإعدادي العلمي والبالغ عددها (640) من طلبة مدارس مختلفة في بغداد، وتوصلت الدراسة إلى عدم وجود فروق ذات دلالة إحصائية في تقدير معامل صعوبة الفقرة ودقتها تعزى إلى الأنموذج (أحادي وثنائي وثلاثي) المعلم، كما وتوصلت الدراسة أيضًا إلى عدم وجود فروق ذات دلالة إحصائية في تقدير معامل التمييز ودقته تعزى إلى الأنموذج (ثنائي وثلاثي المعلم)، وأيضًا إلى عدم وجود فروق ذات دلالة إحصائية في تقدير معامل قدرة الأفراد ودقتها تعزى إلى عامل الأنموذج، وتوصلت الدراسة إلى ملائمة اختبار عكس الأرقام لاختبار القدرات المعرفية "وودكوك جونسون" (البطارية المعيارية) للأنموذج أحادي المعلم.

- دراسة محمد (2016) والتي هدفت إلى مدى تأثر الدقة الإحصائية في تقدير معالم النماذج الرياضية للاستجابة للفقرة من خلال مقارنة النماذج واختلاف عينة التحليل، حيث استخدمت الدراسة أسلوب المحاكاة للبيانات لتوليد الاستجابات الثنائية لعدد (70) فقرة، ومن ثم توليد عينة من الاستجابات بلغ كل منها (500) استجابة، وتوصلت الدراسة إلى عدم وجود فروق دالة إحصائيًا في تقدير القدرة تعزى لنوع النموذج، وأن النموذج الثلاثي يقدم خطأ معياريًا أقل لمتوسط الأخطاء المعيارية لتقديرات القدرة، ويقدم أقصى قيمة لدالة معلومات الاختبار.
- دراسة الزاملي (2015) والتي هدفت إلى الكشف عن مدى اختلاف تقديرات القدرة للأفراد باختلاف النموذج المستخدم (أحادي ، وثنائي، وثلاثي) المعلم، والكشف أيضًا عن دقة تقدير قدرات الأفراد عن طريق الخطأ المعياري باختلاف النماذج الثلاثة، والكشف عن مدى اختلاف توزيعات القدرة للأفراد باختلاف النماذج الثلاثة، حيث طبقت الدراسة اختبار رافن غير اللفظي المتحرر من أثر الثقافة المكون من (60 فقرة) على طلبة المرحلة المتوسطة، البالغ عددهم (512) فردا، واستخدمت الدراسة تحليل القياس المتكرر للمعالجات الإحصائية، وتوصلت الدراسة إلى أن أقل قيمة للخطأ المعياري لتقدير القدرة لصالح النموذج ثنائي المعلم.
- دراسة الشافعي (2011) والتي هدفت إلى إجراء مقارنة بين نماذج الاستجابة للفقرة (أحادي، ثنائي، ثلاثي) المعلم، في تقديرات دوال معلومات الفقرة من خلال (أقصى قيمة للمعلومة، نقاط القدرة المقابلة لأقصى قيم للمعلومات، متوسط قيم المعلومات) وعلاقتها بدقة تقديرات القدرة، حيث طبقت الدراسة اختبار المصفوفة المتتابعة "لجون رافن" على عينة مكونة من (858) من طلاب الثانوية في القاهرة، واستخدم الباحث تحليل القياس المتكرر في المعالجات الإحصائية، وتوصلت الدراسة إلى أن النموذج الثنائي المعلمة أفضل النماذج الثلاثة في تقديمه لأعلى تقديرات لقيم دوال معلومات فقرات الاختبار، وهي نفس النتائج التي توصل إليها فيما يتعلق بمتغيري الخطأ المعياري لتقدير القدرة وثبات تقدير دوال المعلومات كمؤشرين لدقة تقدير قدرة الأفراد، كما

توصلت الدراسة إلى وجود علاقة بين دقة تقديرات القدرة ودوال معلومات فقرات الاختبار، مع ازدياد قيم (متوسط) معلومات فقرات الاختبار.

مناقشة الدراسات السابقة:

- تشابهت هذه الدراسة مع الدراسات السابقة في: مقارنة بين نماذج الاستجابة للفقرة (أحادي، ثنائي، ثلاثي) المعلم، في تقدير معالم الأفراد والفقرات لمعرفة تأثير هذه النماذج على تقدير المعالم، كدراسة ضعضع (2020)، ودراسة محمد (2015)، ودراسة الزاملي (2015)، ودراسة صالح (2020)، ودراسة (Hambleton & Trub, 1991).

وتتميز هذه الدراسة عن الدراسات السابقة في:

- توليد استجابات لعينة كبيرة بحجم (3000) استجابة موزعة على ثلاث صور اختبارية مختلفة في تقدير معالمها محاكاة لخصائص عينة حقيقية، لضمان المقارنة بين نماذج الاستجابة للفقرة وفق خصائص سيكومترية دقيقة وعالية، واختلفت مع دراسات لم تحقق هذا الشرط كدراسة محمد (2016)، حيث ولدت عينة بحجم (500) فرد، ودراسة صالح (2020) بعينة تجريبية بحجم (640) فردا، ودراسة الزاملي (2015) بعينة تجريبية بحجم (512) فردا، ودراسة الشافعي (2011) بعينة تجريبية بحجم (858) فردا.
- توليد البيانات بناءً على بيانات محاكاة لخصائص عينة حقيقية بخلاف بعض الدراسات التي ولدت بيانات منتظمة وثابتة.
- استخراج ومقارنة أكبر عدد من معالم نماذج الاستجابة للفقرة، معالم الفقرات (الصعوبة، التمييز، التخمين)، معلم القدرة للأفراد، دالة معلومات الاختبار، ودراسة الشافعي (2011) في دالة المعلومات، ودراسة محمد (2015) في تقدير القدرة ودالة المعلومات.

الدراسة الميدانية:

1) منهج الدراسة:

المنهج الوصفي، والذي يتناسب مع طبيعة متغيرات وأهداف الدراسة من خلال الوصف الإحصائي لبيانات الفقرات والأفراد ومدى التطابق والملائمة مع نماذج الاستجابة للفقرة، وأيضًا المقارنات اللازمة في تقدير المعالم بين الفقرات والأفراد ودالة معلومات الفقرات والاختبار، ومن ثم التدريج وفق النموذج اللوجستي المناسب والمتطابق مع بيانات الدراسة، يهتم بجمع البيانات التي تساعد في اختبار الفرضيات أو الإجابة عن أسئلة محددة في الظاهرة أو الحدث موضوع البحث، واستخدام أدوات البحث لجمع المعلومات والكشف عن جوهر الظاهرة (الخليلي، 2012: 329).

2) عينة الدراسة:

اعتمدت الدراسة على عينة مولدة محاكاة لخصائص عينة حقيقية باستخدام البرنامج الإحصائي (WENGEN3)، الخاص بتوليد استجابات للأفراد وفق التوزيع الطبيعي الذي تم الاعتماد عليه في هذه الدراسة، حيث تم توليد (1000) استجابة لكل صورة اختبارية، بإجمالي (3000) استجابة للثلاث صور الاختبارية، من ضوء بيانات خصائص العينة التجريبية، لمعلم القدرة ولمعالم الفقرات وعدد (44) فقرة لكل صورة اختبارية والتي من خلالها تم إجراء المقارنات الإحصائية لمعرفة الفروق بين متغيرات الدراسة.

3) مراحل توليد البيانات:

توليد (1000) استجابة لكل صورة اختبارية، من الصور الاختبارية المكونة لبنك الأسئلة بإجمالي (3000) استجابة للثلاث الصور الاختبارية.

أ- تم الاعتماد على التوزيع الطبيعي لمعلمة القدرة بمتوسط حسابي وانحراف معياري (0 ،1) على التوالي، لكل الصور الاختبارية الثلاث، في توليد معلمة القدرة.

ب-توليد معلم التمييز للفقرات وفق توزيع (Log normal) وذلك بمتوسط حسابي وانحراف معياري على التوالي (1، 0.29) للصورة الاختبارية A، و(0.63، 2.30) للصورة الاختبارية B، و(0.25، 1.28) للصورة الاختبارية **C**.

ت-توليد معلم الصعوبة للفقرات وفق توزيع (Log normal) وذلك بمتوسط حسابي وانحراف معياري على التوالي (1.203، 0.51)للصورة الاختبارية A، و(0.64، 0.74) للصورة الاختبارية B، و(1.20، 0.64) للصورة الاختبارية C.

ث-توليد معلم التخمين للفقرات وفق توزيع (Log normal) وذلك بمتوسط حسابي وانحراف معياري على التوالي (0.018، 0.13) للصورة الاختبارية A، و(0.05، 0.26) للصورة الاختبارية B و(0.02، 0.18) للصورة الاختبارية C.

4) أدوات الدراسة:

تكونت أداة الدراسة من بنك أسئلة يتألف من ثلاث صور اختبارية متكافئة، وقد تم بناؤها في ضوء الخطوات العلمية المتعارف عليها في بناء بنوك الأسئلة (من إعداد الباحث)، ولقد تأكد الباحث من الإجراءات الخاصة بنماذج الاستجابة للفقرة ضمن خطوات بناء بنك الأسئلة للعينة المولدة وهي:

1- التحقق من افتراضات نظرية الاستجابة للفقرة للصور الاختبارية المكونة لبنك الأسئلة:

حيث قام الباحث بنقل بيانات استجابات الطلبة إلى برنامج EXCEL تمهيدًا لنقلها إلى البرنامج الإحصائي SPSS لاستخراج التحليل العاملي الاستكشافي لكل صورة اختبارية للتأكد من تحقق افتراضات نظرية الاستجابة للفقرة، كما هو موضح في الجدول (1).

أ. أحادية البعد: تعد فرضية أحادية البعد من أهم الفرضيات الواجب التحقق منها، حيث تفترض نظرية الاستجابة للفقرة وجود سمة واحدة تفسر أداء الفرد في الاختبار، ولذلك تسمى بالنماذج أحادية البعد. وللتحقق من هذا الافتراض، استخدم الباحث المؤشرات التي تعتمد على المكونات الأساسية الرئيسية (Indices based on principal components)، وهي مؤشرات تستند إلى استخدام التحليل العاملي (factory analysis)، وقد تم حساب الجذر الكامن ونسبة التباين المفسر، والتباين المفسر التراكمي لكل عامل من العوامل كما في كل صورة من الصور الاختبارية الثلاث المكونة لبنك الأسئلة، كما هو موضح في الجداول (1).

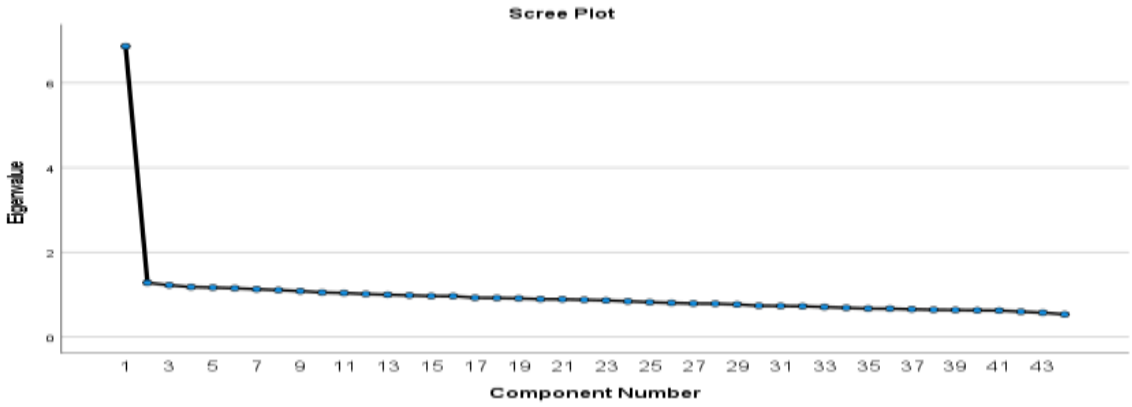
جدول (1) يوضح نتائج التحليل العاملي الاستكشافي للصور الاختبارية

العامل	الصورة الاختبارية A			الصورة الاختبارية B			الصورة الاختبارية C		
	الجذر الكامن	نسبة التباين المفسر%	نسبة التباين المفسر التراكمي %	الجذر الكامن	نسبة التباين المفسر%	نسبة التباين المفسر التراكمي %	الجذر الكامن	نسبة التباين المفسر%	نسبة التباين المفسر التراكمي %
1	6.870	15.614	15.614	7.498	17.041	17.041	6.759	15.361	15.361
2	1.278	2.905	18.519	1.231	2.798	19.839	1.275	2.899	18.260
3	1.220	2.772	21.291	1.193	2.712	22.551	1.187	2.698	20.958

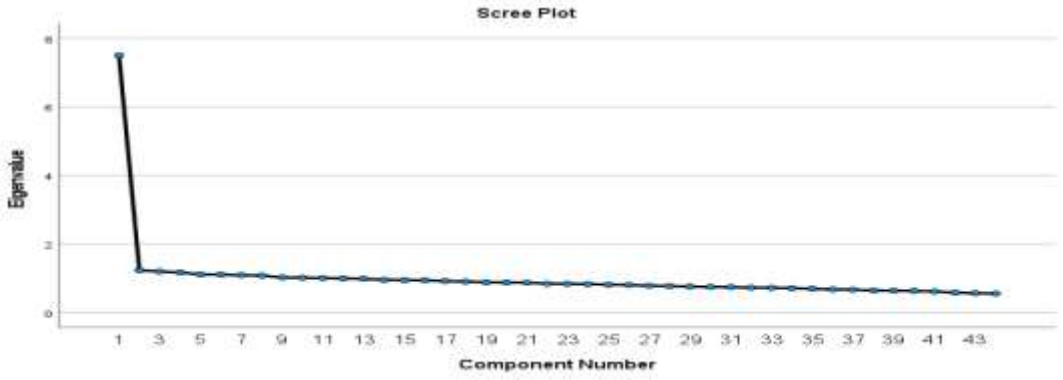
يتضح من الجدول (1) أن قيم الجذر الكامن للعامل الأول في الصور الاختبارية الثلاث مرتفعة مقارنة بقيم العوامل الأخرى التي كانت منخفضة ومتقاربة نسبيًا، فكانت نسبة الجذر الكامن للعامل الأول إلى العامل الثاني للصور الاختبارية الثلاث على الترتيب (5.30 ، 6.09،5.37) وتبعًا لما أورده (Lord, 1980) إن كانت نسبة قيمة الجذر الكامن للعامل الأول إلى قيمة الجذر الكامن للعامل الثاني تزيد على (2) يعد مؤشرًا على وجود عامل مسيطر ومؤشر لأحادية البعد، مما يشير إلى وجود عامل سائد يدعم افتراض أحادية البعد.

كما أظهر منحنى الانحدار (Scree plot)[أشكال (1،2،3)] انكسارًا واضحًا بعد العامل الأول، مما يعد مؤشرًا إضافيًا على وجود بعد واحد مسيطر.

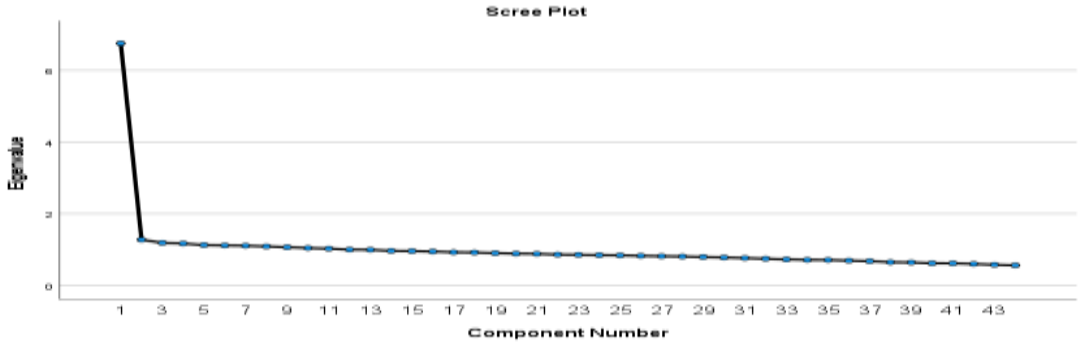
شكل (1): التمثيل البياني لقيم الجذور الكامنة للصورة الاختبارية (A)



شكل (2) يبين التمثيل البياني لقيم الجذور الكامنة للصورة الاختبارية (B)



شكل (3): التمثيل البياني لقيم الجذور الكامنة للصورة الاختبارية (C)



ب. الاستقلال المركزي: ينص هذا الافتراض على أن استجابة الفرد للفقرات الاختبارية يجب أن تكون مستقلة إحصائيًا (Reise & Waller, 2003) ولكي يتحقق هذا الافتراض يجب ألا يتأثر أداء الفرد على الفقرة بالإيجاب أو السلب باستجاباته لأي من الفقرات الأخرى بالاختبار(Hambleton & Swaminathan, 1985)، وهذا يعني أن محتوى الفقرة يجب ألا يعطي مؤشرات أو أدلة على فقرة أخرى، ويعني افتراض الاستقلالية أن السمة أو السمات التي يقيسها الاختبار هي العوامل المؤثرة على استجابة الفرد على فقرات الاختبار، وهذا يعني أن افتراض

استقلالية القياس يتحقق بتحقق افتراض أحادية البعد. (الشافعي، 2008) وما يجب الإشارة إليه أن افتراض استقلال الموضع لا يعني بالضرورة عدم وجود ارتباطات بين فقرات الاختبار من خلال المجموعة الكلية للأفراد، حيث إن الارتباطات الموجبة بين زوج من الفقرات هي في واقع الأمر نتيجة طبيعية ما دامت هناك فروق فردية بين الأفراد على القدرة المقاسة بواسطة فقرات هذا الاختبار، في حين لا تكون درجات الفقرات مرتبطة عند مستوى ثابت من القدرة، بالتالي إذا كان فرض أحادية البعد محققًا فإن فرض الاستقلال المركزي محقق.

2 ـ التأكد من ملائمة البيانات (الفقرات) لنماذج الاستجابة للفقرة (ثنائي، ثلاثي) المعلم، للعينة المولدة:

يتبين أن النموذج ثنائي المعلم أقل النماذج في الفقرات غير الملائمة بواقع (3) فقرات في الثلاث الصور الاختبارية، يليه النموذج أحادي المعلم بواقع (4) فقرات، يأتي النموذج ثلاثي المعلم بواقع (11) فقرة، وبالتالي فإن النموذج ثنائي المعلم هو النموذج الأكثر ملاءمة لبيانات البحث.

ـ التأكد من ملائمة (الأفراد) لتوقعات نماذج الاستجابة للفقرة (أحادي، ثنائي، ثلاثي) المعلم، للعينة المولدة، وتم ذلك باستخدام برنامج Bilog mg3 من خلال مطابقة الأفراد FIT PROB.

ومن خلال الدلالة الإحصائية للمطابقة (FIT PROB) لمربع كاي X^2 (0.01 <)، كشف التحليل الصور الاختبارية الثلاث، في النموذج ثلاثي المعلم تم حذف (4) أفراد لعدم ملائمتهم توقعات النموذج، وفي النموذج أحادي المعلم تم حذف (3) أفراد، وعدم حذف أي فرد في النموذج ثنائي المعلم.

3 ـ التحقق من الثبات الإمبريقي للصور الاختبارية المكونة لبنك الأسئلة للعينة المولدة:

ولقد قام الباحثان بحساب الثبات الإمبريقي للصور الاختبارية (**A، B، C**) وفق نماذج الاستجابة للفقرة (ثنائي، ثلاثي) المعلم، باستخدام البرنامج الإحصائي (**Bilog mg3**).

جدول (2) يوضح نتائج معامل الثبات الإمبريقي للصور الاختبارية (A، B، C) وفق نماذج الاستجابة للفقرة (ثنائي، ثلاثي) المعلم

الصور الاختبارية	النموذج الثنائي	النموذج الثلاثي
A	0.962	0.936
B	0.919	0.872
C	0.887	0.838

يلاحظ من الجدول (2) أن الثبات الإمبريقي للنموذج ثنائي المعلم هو الأعلى في الصور الاختبارية الثلاث (**A، B، C**).

4 ـ التأكد من المتوسط الحسابي والانحراف المعياري لتقدير معالم الفقرات للصور الاختبارية المكونة لبنك الأسئلة للعينة المولدة:

جدول (3) يوضح نتائج المتوسط الحسابي والانحراف المعياري لمعالم الفقرات (الصعوبة، التمييز، التخمين) بعد حذف الفقرات وفق نماذج الاستجابة للفقرة (ثنائي، ثلاثي) المعلم للصورة الاختبارية

الصور الاختبارية		ثنائي المعلم				ثلاثي المعلم					
		الصعوبة	الانحراف المعياري	التمييز	الانحراف المعياري	الصعوبة	الانحراف المعياري	التمييز	الانحراف المعياري	التخمين	الانحراف المعياري
A	أعلى قيمة	3.70	0.92	1.90	0.21	5.54	1.68	1.73	0.34	0.24	0.09
	أدنى قيمة	– 1.59	0.15	0.35	0.06	– 2.33	0.37	0.49	0.09	0.10	0.03
	المتوسط الحسابي	0.12	1.06	1.07	0.34	0.61	1.45	1.16	0.29	0.15	0.05
B	أعلى قيمة	2.87	0.31	4.26	0.30	2.01	0.19	4.90	1.14	0.35	0.05
	أدنى قيمة	– 0.67	0.04	1.15	0.10	0.05	0.13	1.35	0.32	0.10	0.03
	المتوسط الحسابي	0.76	0.69	2.41	0.65	0.89	0.56	2.69	0.87	0.26	0.06
C	أعلى قيمة	5.06	1.51	1.80	0.14	2.92	0.60	3.02	0.74	0.37	0.09
	أدنى قيمة	– 1.32	0.11	0.67	0.08	– 1.76	0.52	0.65	0.10	0.13	0.03
	المتوسط الحسابي	0.65	1.18	1.36	0.24	0.48	1.10	1.50	0.43	0.20	0.04

5 ــ التعرف على دالة معلومات الاختبار للصور الاختبارية المكونة لبنك الأسئلة للعينة المولدة:

بعد عملية توليد بيانات محاكاة لبيانات خصائص العينة وفق النموذج ثلاثي المعلم عن طريق البرنامج **wingen**3 قام الباحث باستخدام برنامج الحاسب الآلي **Bilog mg3** لتحليل البيانات والحصول على دالة معلومات الاختبار للصور الاختبارية **(A B, C,)** باستخدام نماذج الاستجابة للفقرة (ثنائي، ثلاثي) المعلم، لاستخراج أقصى دالة معلومات للاختبار ومتوسط الخطأ المعياري للتقدير.

جدول (4) يوضح نتائج دالة معلومات الاختبار للصور الاختبارية (A B, C,) ومتوسط الخطأ المعياري للتقدير وفق نماذج الاستجابة للفقرة (ثنائي، ثلاثي) المعلم

الصور الاختبارية	خصائص إحصائية	ثنائي المعلم	ثلاثي المعلم
A	دالة معلومات الاختبار	8.526	7.007
	متوسط الخطأ المعياري	0.34	0.37
B	دالة معلومات الاختبار	39.65	24.71
	متوسط الخطأ المعياري	0.16	0.21
C	دالة معلومات الاختبار	12.59	10.42
	متوسط الخطأ المعياري	0.28	0.32

نلاحظ من الجدول (4) أنه بلغت أعلى قيمة لدالة معلومات الاختبار للنموذج ثنائي المعلم في الصور الاختبارية الثلاث **(A B, C,)**، وبلغت أقل قيمة لمتوسط الخطأ المعياري في النموذج ثنائي المعلم للصور الاختبارية الثلاث.

4.7. المعالجة الإحصائية:

تمت معالجة بيانات هذا البحث باستخدام برنامج (Bilog mg3) الخاص بنظرية الاستجابة للفقرة، والحقيبة الإحصائية للعلوم الاجتماعية (SPSS) بعد توليد البيانات من خلال البرنامج (WENGEN3) كالآتي:

ـ برنامج **Bilog mg**3.

ـ تقدير حسن ملاءمة البيانات لنماذج الاستجابة للفقرة (أحادي، ثنائي، ثلاثي) المعلم.

ـ تقدير معالم القدرة للأفراد وللفقرات (الصعوبة، التمييز، التخمين) لنماذج الاستجابة للفقرة (ثنائي، ثلاثي) المعلم.

ـ تقدير الثبات الإمبريقي للتحقق من ثبات الصور الاختبارية لفقرات بنك الأسئلة وفق نماذج الاستجابة للفقرة (ثنائي، ثلاثي) المعلم.

ـ تقدير دوال معلومات الاختبار للصور الاختبارية (**A B, C**,) المكونة لبنك الأسئلة وفق نماذج الاستجابة للفقرة (ثنائي، ثلاثي) المعلم.

ـ برنامج **SPSS**.

ـ التحليل العاملي الاستكشافي باستخدام طريقة المكونات الرئيسة لإثبات أحادية البعد والاستقلال المركزي.

ـ إيجاد قيم الاختبار التائي لعينتين مستقلتين (**t–test**) للتعرف على أثر نماذج الاستجابة للفقرة (ثنائي، ثلاثي) المعلم، في تقدير معلم (الصعوبة، التمييز) ومعلم القدرة للأفراد، والخطأ المعياري لتقدير القدرة.

ـ برنامج **WINGEN**3.

بعد استخراج تقديرات القدرة للأفراد وللفقرات للعينة التجريبية، تم توليد بيانات حقيقية محاكاة والتي على ضوئها تمت عملية المقارنة بين نماذج الاستجابة للفقرة (ثنائي، ثلاثي) المعلم.

عرض نتائج الدراسة ومناقشتها:

الهدف الفرعي الأول:

التعرف على الفروق بين نماذج الاستجابة للفقرة (ثنائي، ثلاثي) المعلم، في تقدير معلم (القدرة للأفراد) للصور الاختبارية المكونة لبنك الأسئلة. تم إجراء الاختبار التائي لعينتين مستقلتين لمقارنة متوسطات معلم القدرة للأفراد بين نماذج الاستجابة للفقرة (ثنائي، ثلاثي) المعلم، والنتائج مبينة في الجدول (5).

جدول (5) يوضح نتائج اختبار"t–test " لتقديرات معلم القدرة للأفراد بين نماذج الاستجابة للفقرة (ثنائي، ثلاثي) المعلم، للصور الاختبارية (A B, C,)

المعلم	الصور الاختبارية	النموذج	المتوسط الحسابي	الانحراف المعياري	درجة الحرية	قيمة "ت"	مستوى الدلالة
القدرة	A	ثنائي المعلم	−0.0036	1.00657	1998	0.870	0.38
		ثلاثي المعلم	−0.0443	1.08433			
	B	ثنائي المعلم	0.0000	1.00050	1998	0.072	0.94
		ثلاثي المعلم	−0.0032	1.00516			
	C	ثنائي المعلم	0.0000	1.00050	1998	1.284	0.19
		ثلاثي المعلم	−0.603	1.09962			

يلاحظ من الجدول (5) أنه لا توجد فروق دالة إحصائيًا عند مستوى دلالة (0.05) في الصورة الاختبارية (A B, C,) بين النموذجين ثنائي وثلاثي المعلم، في دقة تقدير معلم القدرة للأفراد، وهذا يشير إلى أنه لا يوجد اختلاف في تقديرات معلم القدرة للأفراد بين النموذجين، مما يبين إمكانية استخدام النموذجين في تقدير معلم القدرة للأفراد دون أن يؤدي ذلك إلى تأثير على النتائج العامة، وتتفق هذه النتائج مع دراسة صالح (2020) ومحمد (2016)، وتختلف مع دراسة ضعضع (2021) التي خلصت إلى أن الفروق بين النماذج لصالح النموذج الأكثر معالم ودراسة هامبلتون وتريوب (1991) والتي كانت على عينة مولدة منتظمة بحجم (500) فرد، حيث خلصت إلى وجود فروق بين نماذج الاستجابة للفقرة الأحادي والثنائي لصالح الثنائي، وبين الثنائي والثلاثي لصالح الثلاثي، وأعزى الفروق لاختلاف مستوى التمييز.

الهدف الفرعي الثاني:

التعرف على الفروق بين نماذج الاستجابة للفقرة (ثنائي، ثلاثي) المعلم، في تقدير معالم الفقرة (الصعوبة، التمييز) للصور الاختبارية المكونة لبنك الأسئلة.

ـ معلم الصعوبة:

تم إجراء الاختبار التائي لعينتين مستقلتين لمقارنة متوسطات معلم الصعوبة بين نماذج الاستجابة للفقرة (ثنائي، ثلاثي) المعلم، والنتائج مبينة في الجدول (6).

جدول (6) يوضح نتائج اختبار"t-test " لتقديرات معلم الصعوبة بين نماذج الاستجابة للفقرة (ثنائي، ثلاثي) المعلم، للصور الاختبارية (A B, C,)

المعلم	الصور الاختبارية	النموذج	المتوسط الحسابي	الانحراف المعياري	درجة الحرية	قيمة "ت"	مستوى الدلالة
الصعوبة	A	ثنائي المعلم	0.896	1.06314	84	-1.838	0.07
		ثلاثي المعلم	0.5968	1.46425			
	B	ثنائي المعلم	0.7609	0.69473	85	-1.041	0.30
		ثلاثي المعلم	0.9023	0.56313			
	C	ثنائي المعلم	0.6528	1.18848	84	0.665	0.50
		ثلاثي المعلم	0.4885	1.10096			

يلاحظ من الجدول (6) أنه لا توجد فروق دالة إحصائيًا عند مستوى دلالة (0.05) في الصورة الاختبارية (A B, C,) بين النموذجين ثنائي وثلاثي المعلم، في دقة تقدير معلم الصعوبة، وهذا يشير إلى أنه لا يوجد اختلاف في تقديرات معلم التمييز بين النموذجين، مما يبين إمكانية استخدام النموذجين (ثنائي، ثلاثي) المعلم، في تقدير معلم الصعوبة دون أن يؤدي ذلك إلى تأثير على النتائج العامة، وتتفق هذه النتيجة مع دراسة صالح (2000)، وتختلف مع دراسة ضعضع (2021) التي خلصت نتائجها إلى وجود فروق بين النموذج الثنائي والثلاثي، ولا يوجد فروق بين النموذجين أحادي وثنائي المعلم.

- معلم التمييز:

تم إجراء الاختبار التائي لعينتين مستقلتين لمقارنة متوسطات معلم التمييز بين نماذج الاستجابة للفقرة (ثنائي، ثلاثي) المعلم، والنتائج مبينة في الجدول (7).

جدول (7) يوضح نتائج اختبار"t-test " لتقديرات معلم التمييز بين نماذج الاستجابة للفقرة (ثنائي، ثلاثي) المعلم، للصور الاختبارية (A B, C,)

المعلم	الصور الاختبارية	النموذج	المتوسط الحسابي	الانحراف المعياري	درجة الحرية	قيمة "ت"	مستوى الدلالة
التمييز		ثنائي المعلم	1.0775	0.3397	85	-1.347	0.181

	A	ثلاثي المعلم	1.1694	0.2943		
	B	ثنائي المعلم	2.4160	0.6573	85	− 1.698	0.093
		ثلاثي المعلم	2.7005	0.8856		
	C	ثنائي المعلم	1.3617	0.24317	85	−1.926	0.058
		ثلاثي المعلم	1.5075	0.43437		

يلاحظ من الجدول (7) أنه لا توجد فروق دالة إحصائيًا عند مستوى دلالة (0.05) في الصورة الاختبارية (A B, C,) بين النموذجين ثنائي وثلاثي المعلم، في دقة تقدير معلم التمييز، وهذا يشير إلى أنه لا يوجد اختلاف في تقديرات معلم التمييز بين النموذجين، مما يبين إمكانية استخدام النموذجين في تقدير معلم التمييز دون أن يؤدي ذلك إلى تأثير على النتائج العامة، وتتفق هذه النتائج مع دراسة صالح (2020) ودراسة ضعضع (2021).

الهدف الفرعي الثالث:

التعرف على الفروق بين نماذج الاستجابة للفقرة (ثنائي، ثلاثي) المعلم، في الخطأ المعياري لتقدير القدرة للصور الاختبارية المكونة لبنك الأسئلة. تم إجراء الاختبار التائي لعينتين مستقلتين لمقارنة متوسطات الخطأ المعياري لتقدير القدرة بين نماذج الاستجابة للفقرة (ثنائي، ثلاثي) المعلم، والنتائج مبينة في الجدول (8).

جدول (8) يوضح نتائج اختبار"t-test " للخطأ المعياري لتقدير القدرة بين نماذج الاستجابة للفقرة (ثنائي، ثلاثي) المعلم، للصور الاختبارية (A B, C,) المكونة لبنك الأسئلة

المعلم	الصور الاختبارية	النموذج	المتوسط الحسابي	الانحراف المعياري	درجة الحرية	قيمة "ت"	مستوى الدلالة
الخطأ المعياري لتقدير القدرة	A	ثنائي المعلم	0.3425	0.05285	1983	−18.545	0.001
		ثلاثي المعلم	0.4387	0.15502			
	B	ثنائي المعلم	0.2601	0.16158	1997	−10.508	0.001
		ثلاثي المعلم	1.0126	2.25887			
	C	ثنائي المعلم	0.3016	0.09704	1982	−14.728	0.001

		ثلاثي المعلم	0.4083	0.20714	

نلاحظ من الجدول (8) وجود فروق دالة إحصائيًا عند مستوى دلالة (0.01) بين نماذج الاستجابة للفقرة (ثنائي، ثلاثي) المعلم، بين النموذجين لصالح النموذج الثنائي، حيث بلغت قيمة (t) للصورة الاختبارية (A B, C,) على الترتيب (-18.545، -10.508، -14.728)، مما يشير إلى أفضلية النموذج ثنائي المعلم في تقدير معلم القدرة مقارنة بالنموذج ثلاثي المعلم، وتتفق هذه الدراسة مع دراسة الزاملي (2015)، وتختلف مع دراسة محمد (2016) التي خلصت نتائجها إلى أن النموذج ثلاثي المعلم الأقل في الخطأ المعياري لتقدير القدرة.

الخاتمة:

أثبتت نتائج هذه الدراسة أنه لا يوجد فروق بين نماذج الاستجابة للفقرة (ثنائي، ثلاثي) المعلم، في تقدير معالم الأفراد والفقرات، ولهذا على الباحث عند اختيار أي من نماذج الاستجابة للفقرة في تحليل بياناته يفضل اختيار النموذج الأنسب والملائم للبيانات في ضوء افتراضات كل نموذج، وبحسب طبيعة البيانات ونوعيتها وحجمها، والدراسة الحالية أكدت الدقة الإحصائية والخطأ المعياري الأقل في تقدير المعالم والأكثر ملائمة للبيانات للنموذج ثنائي المعلم وأفضليته من النموذج ثلاثي المعلم.

تقدم الدراسة مجموعة من المقترحات على النحو التالي:

1- ضرورة دراسة أثر اختلاف نوع البيانات (تجريبية ـ مولدة) على عينات تتناسب مع افتراضات نماذج الاستجابة للفقرة للمقارنة بين نماذج الاستجابة للفقرة (أحادي، ثنائي، ثلاثي) المعلم، في تقدير معالم الأفراد والفقرات.

2- ضرورة دراسة أثر اختلاف نوع البرنامج الإحصائي (Bilog mg3، Winsteps، Microscle، PARSCALE، Xcalibere.4.2.2) في تقدير معالم الأفراد والفقرات.

3- ضرورة دراسة أثر اختلاف نماذج الاستجابة للفقرة (أحادي، ثنائي، ثلاثي) المعلم، في تقدير معالم الأفراد والفقرات بناء على عينات مولدة باختلاف كلٌ من نوع التوزيع (اعتدالي، منتظم)، وقيم مختلفة ذات التوزيع المنتظم (سالب الالتواء، موجب الالتواء).

قائمة المراجع والمصادر:

أولاً: المراجع العربية:

- إبراهيم، السيد وآخرون (2009): مقارنة بين النماذج اللوغاريتمية لنظرية الاستجابة للمفردة من حيث ملاءمتها للبيانات، مجلة البحث العلمي في التربية، العدد (10).
- حامد، شيرين (2008): أثر نموذج الاستجابة للفقرة وتعدد الأبعاد وطريقة المطابقة في تقدير معالم الأفراد والفقرات، أطروحة دكتوراه غير منشورة، الأردن: جامعة عمان العربية.
- الخليلي، خليل (2012): أساسيات البحث العلمي التربوي. ط1، دبي: دار القلم.
- الدرابسة، رياض (2012): أثر طريقة القدرة وطريقة التعامل مع القيم المفقودة على دقة تقدير معالم الفقرات والأفراد، أطروحة دكتوراه غير منشورة، عمَّان: جامعة اليرموك.
- الزاملي، علي (2015): تأثير نماذج الاستجابة للمفردة على دقة تقديرات القدرة للأفراد باستخدام اختبار المصفوفات المتتابعة لرافن، لارك للفلسفة واللسانيات والعلوم الاجتماعية، ع 8.
- الشافعي، محمد (2011): تأثير نماذج التحليل اللوغاريتمية الأحادية والثنائية والثلاثي البارامتر على تقديرات دوال معلومات المفردات الاختبارية وعلاقتها بدقة تقديرات القدرة: دراسة مقارنة، المجلة المصرية للدراسات النفسية، مج 21، ع73.
- الشافعي، محمد (2014): تأثير انتهاك بعض افتراضات نظرية الاستجابة للمفردة على تدريج بنك الأسئلة ودقة معادلة درجات الاختبارات البنكية المسحوبة . مجلة كلية التربية (جامعة بنها) ، القاهرة: مج 25، ع 98، 437–440.
- صالح، نوال (2020): أثر اختيار نموذج الاستجابة للمفردة في دقة تقدير معالم المفردة ثنائية الاستجابة، لارك للفلسفة واللسانيات والعلوم الاجتماعية، مج38، عدد38.
- طيفور، مصطفى (2007): دراسة مقارنة لنماذج نظرية الاستجابة للمفردة في معادلة درجات الاختبار، أطروحة دكتوراه غير منشورة، جامعة القاهرة.
- العبد الله، زياد (2012): أثر بعض طرق التقدير على دقة تقدير المعالم ضمن نماذج الاستجابة للمفردة متعددة التدريج، أطروحة دكتوراه غير منشورة: معهد الدراسات التربوية، جامعة القاهرة.
- عطا، زايد (2013): تقصي دقة تقدير النموذج اللوجستي ثلاثي المعلمة لمعالم الفقرة وقدر الأفراد في ضوء تغير طول الاختبار وحجم العينة: دراسة محاكاة، مجلة جامعة الشارقة، مج 11، ع2.
- علام، صلاح الدين (2005): نماذج الاستجابة للمفردة الاختبارية أحادية البعد ومتعددة الأبعاد وتطبيقاتها في القياس النفسي والتربوي، القاهرة: دار الفكر العربي.

- محمد، شحته (2016): الدقة الإحصائية لتقدير بارامترات النماذج الرياضية للاستجابة للمفردة، مجلة رسالة التربية وعلم النفس، ع52.
- محمود، سومية (2012): استخدام نموذج راش في بناء بنك أسئلة لقياس التحصيل في مقرر سيكولوجية التعلم لدى طلاب كلية التربية بالمنيا، أطروحة دكتوراه منشورة، القاهرة: جامعة المنيا، كلية التربية.
- نصراوين، معين (2018): دقة تقدير معالم الفقرات عند استخدام أربعة نماذج لوجستية في إطار نظرية الاستجابة للفقرة، مجلة دراسات العلوم التربوية، مج45، ع4.

ثانيًا: المراجع الأجنبية:

- Baker, F. (2001). The Basics of Item Response Theory Eric clearing house on Assessment and Evaluation.
- Crocker, L. & Algina, J. (1986). Introduction to classical and modern test theory. New York: Holt, Rinehart, 83(3), 154-159.
- Davier, A., Holland, P., & Thayer, D. (2004). The Kernel method of test equating y. New York: Springer.
- Embertson, S. E. & Reise, S. P. (2000). Item Response Theory for Psychologiste. New Jjersey: Lawrence Erlbaum Associates, Publisher.
- Hambleton, R. & Swaminathan, H. (1985). Item response theory: Principles and applications. Boston: Kluwer.
- Hambleton, R., Swaminathan, H., and Rogers, H. (1991). Fundamentals of item response Theory. International Educational and Professional. Sage publication, Inc Newbury park, USA.
- Keller, L. & Keller, R. (2011). The Long – term Sustainability of different item response theory Scaling methods Educational and Psychological Measurement, 7/ (2), 362-379.
- Klein, P. (1993). The handbook of psychological testing, First Published, London, Rouledge.
- Kolen, M. & Brennan, R. (2014). Test equating Scaling and Linking (3rd ed). New York: Springer.

- Lord, F. (1980). Applications of item response theory to practical testing problems. Hillsdale: Lawrence Erlbaum Associates.
- Lord, F. M. & Novick, M. R. (1968). Statistical Theories of mental test scores. London, Addison, Wesley Company.
- Mislevy, R. J. & Bock, R. D. (1990). Bilog3: Item analysis and test scoring with binary logistic models (2nd ed), Scientific software, inc.
- Pomplun, M., Omar, H. & Custer, M. (2004). A Comparison of Winsteps and Bilog-mg for vertical scaling with the Rasch model, Educational and Psychological Measurement, 64(4), 600-616.
- Reeve, B. B. (2002). An Introduction to modern measurement theory Division of cancer control and population sciences, National cancer Institute.
- Stevens, P., Lehman, K. & Rodolfo, D. (2005). Item Response Theory, Fundamentals, Applications, and promise in psychological Research, American psychological Society, 14(2), p327-355.
- Thorndike, R. (1997). Measurement and evaluation in psychology and evaluation, London Prentice-Hall.
- Wright, D. & Stone, M. (1979). Best test design: A hand book for Rasch Measurement Chicago: MESA Press.
- Yu-Feng et al. (2013). Item response analysis on in examination anesthesiology for medical students in Taiwan: A comparison of one- and two- parameter logistic models. Journal of the Chinese medical Association 76.
- Zhonghua, Z. (2010). Comparison of different equating methods and an application to Link test – based tests (Unpublished doctoral dissertation), the Chinese University of Hong Kong.