مجلة جامعة الأقصى– سلسة العلوم التربوية والنفسية، المجلد السادس، العدد الثاني، أبريل2023، ص:01–18
ISSN 2791 - 3783

دالة معلوماتية الاختبار لنموذجي نظرية الاستجابة للفقرة الثلاثي المعلمي وموكن اللامعلمي في ضوء نوع ونسبة الأداء التفاضلي

أ. إسراء "محمد أنور" الخطيب
الجامعة الاردنية
د. حسان غازي العمري
الجامعة الاردنية

(تاريخ الاستلام 2022/08/19، تاريخ القبول 2022/10/17)

Test information function for the two models of item response theory for parametric model -three-parameter model-and non-parametric model -Mokken- in the light of the type and percentage of differential item parameter

Mr. Israa "Muhammad Anwar" Al-Khatib
University of Jordan
Dr. Hassan Ghazi Al-Omari
University of Jordan

Assistant Professor of Business Administration / Palestine Technical College

(Received 19/08/2022, Accepted 17/10/2022)

أ. إسراء الخطيب – الجامعة الاردنية E-mail address: asra32915@gmail.com

د. حسان العمري – الجامعة الاردنية E-mail address: has.omari@ju.edu.jo

الملخص:

هدفت الدراسة الحالية إلى تقصي قيم دالة معلوماتية الاختبار المقدرة وكذلك فحص التوافق في مستويات دالة معلوماتية الاختبار المقدر باستخدام النموذجين (النموذج ثلاثي المعلمة ونموذج موكن اللامعلمي) في ضوء نوع ونسبة الأداء التفاضلي. ولتحقيق ذلك اتبعت الدراسة المنهج التجريبي باستخدام بيانات مولدة تحاكي ظروف متعلقة بخصائص الفقرة وقدرات الأفراد في ضوء نوع ونسبة الأداء التفاضلي. وقد توصلت الدراسة إلى أن النموذج اللامعلمي أكثر كفاءة عند مستويات القدرة المتوسطة، بينما يكون النموذج المعلمي أكثر كفاءة عند الأطراف حيث القدرة المتطرفة وهذا يعكس دقة أعلى للنموذج المعلمي عند المستويات الدنيا والعليا للقدرة، كما توصلت الدراسة إلى أن دالة معلوماتية الاختبار تقدر بشكل مختلف عبر مستويات القدرة باختلاف النموذج المستخدم ولصالح النموذج اللامعلمي.

الكلمات المفتاحية: دالة المعلومات، النماذج المعلمية، النماذج اللامعلمية، نموذج موكن، الأداء التفاضلي.

Abstract:

The study aimed to investigate the values of the test informatics function estimated. As well, as examine compatibility in the levels of the estimated testing function, with three-parameter model and the non-parametric Moken model. In the light of the type and percentage of differential performance. In order to achieve this, the study followed the experimental method using generated data that simulates conditions related to the characteristics of the paragraph and the abilities of individuals in the light of the type and percentage of Differential Items Functioning. This reflects a higher accuracy of the parametric model at the lower and higher levels of ability. The study also found that the test informatics function is estimated differently across the ability levels according to the model used and in favor of the non-parametric model.

.

Keywords: information function, parametric models, nonparametric models, Moken model, Differential Items Functioning.

مقدمة:

تعتبر نظرية الاستجابة للفقرة Item Response Theory (IRT)، من النظريات الهامة في القياس النفسي، والتي جاءت بهدف التغلب على المشكلات التي عجزت عن مواجهتها نظرية الاختبار الكلاسيكية (Classical Test Theory (CTT، والتي يرى بعض علماء القياس مثل لورد (Lord) ونوفيك (Novick) أنها تُعد امتداداً لنظرية الاختبار الكلاسيكية.

ويعد أهم ما يميز هذه النظرية (IRT) أن تقدير معالم الفقرة (الصعوبة، التمييز، التخمين) مستقل عن قدرة المفحوصين، وكذلك استقلال تقدير قدرة المفحوصين عن معالم الفقرات المستخدمة في عملية التقدير، أي أن نماذج هذه النظرية توفر تقديراً للقدرة مستقلاً عن المفحوص، وأن تقديرات القدرة يعبر عنها بوحدات لاتتأثر بصفات العينة، أو مجموعة معينة من المفحوصين، وإن تقدير قدرته المشتق وفق أي من نماذج نظرية الاستجابة للفقرة لا يتعلق بهذه العوامل أو غيرها (Acar & Kelecioglu, 2010)

من هنا برزت الحاجة إلى بناء اختبارات ومقاييس ذات فقرات تؤدي الغرض من الاختبار في ضوء معالم الفقرات الثلاثة (الصعوبة، التمييز، التخمين)، غير متأثرة بالعوامل الدخيلة الأخرى كالجنس أو اللغة أو الثقافة أو المستوى الاقتصادي وغيرها من العوامل، التي من شأنها أن تظهر نوع التحيز والأداء التفاضلي في قياس السمات الكامنة للمفحوصين (Abbott, 2007).

إذ يستدل على وجود الأداء التفاضلي للفقرة من خلال الفروق في احتمالات الاستجابة الصحيحة للفقرة بين مجموعتين ممن هم في مستوى القدرة نفسه (Gruijter & Kamp, 2005)، كما إن وجود الأداء التفاضلي للفقرة يشير الى وجود عوامل اخرى تؤثر في احتمال إجابة مجموعة معينة دون الأخرى للفقرة، وبالتالي فإنها تهدد مفهوم العدالة في تقييم الأفراد (Gybels, 2004)، وللكشف عن الأداء التفاضلي لا بد من فحص العلاقة بين احتمالات الاستجابة الصحيحة على الفقرة، وبين السمة الكامنة المراد قياسها، ومن ثم يتم مقارنة منحنيات خصائص الفقرة بين المجموعتين اللتين يتم رسمهما وفق النموذج المستخدم (نماذج معلمية أو لا معلمية).

مشكلة الدراسة:

إن اتخاذ القرارات التربوية والتعليمية المختلفة يستند بشكل مباشر على نتائج المقاييس والاختبارات المعدة لقياس السمات الكامنة للأفراد والمجموعات باختلاف ثقافاتهم، لذا كان من الواجب الحصول على مقاييس تتمتع بدقة عالية وثبات مرتفع وهو ما يتمثل في دالة معلوماتية الاختبار والمكونة من مجموع دوال معلوماتية فقراته. ولا بد من تضمين فقرات ذات دالة معلوماتية مرتفعة بالاعتماد على نماذج نظرية الاستجابة للفقرة المعلمية واللامعلمية. والتي تخلو من أي أداء تفاضلي يمكن أن يؤدي إلى تهديد قدرة هذه النماذج على الوصول إلى نتائج دقيقة أو غير صادقة، فوجود الأداء التفاضلي في الاختبار يمكن أن يؤدي إلى آثار ضارة على الأفراد، مثل عدم تكافؤ الفرص، والتشخيص الخاطئ، وعدم المساواة في الفرص التعليمية، والاستنتاج غير الدقيق في البحث (Croudace & Brown, 2012).

من هنا برزت مشكلة الدراسة في التعرف على دالة معلوماتية الاختبار لنموذجي نظرية الاستجابة للفقرة الثلاثي المعلمي وموكن اللامعلمي في ضوء نوع ونسبة الأداء التفاضلي؛ من خلال الإجابة عن الأسئلة الآتية:

أسئلة الدراسة

مما سبق عرضه يبرز تساؤل حول نوع الأداء التفاضلي (المنتظم، غير المنتظم) ونسبته في التأثير على الخصائص السيكومترية للاختبار باستخدام النماذج المعلمية واللامعلمية في نظرية الاستجابة للفقرة، ومن هنا سعت هذه الدراسة إلى التعرف على أثر نسبة ونوع الأداء التفاضلي في الخصائص السيكومترية للاختبار وقدرات الأفراد وفق النماذج المعلمية واللامعلمية لنظرية الاستجابة للفقرة من خلال الإجابة **عن الأسئلة التالية:**

السؤال الأول: ما قيم دالة معلوماتية الاختبار المقدرة باستخدام النموذجين (النموذج ثلاثي المعلمة ونموذج موكن اللامعلمي) في ضوء نوع ونسبة الأداء التفاضلي؟

السؤال الثاني: هل يوجد توافق في مستويات دالة معلوماتية الاختبار المقدرة باستخدام النموذجين (النموذج ثلاثي المعلمة ونموذج موكن اللامعلمي) في ضوء نوع ونسبة الأداء التفاضلي؟

أهداف الدراسة:

تسعى الدراسة الحالية إلى تقصي أثر وجود الأداء التفاضلي في فقرات الاختبار نوعا وكما على ثبات الاختبار والمتمثل في دالة معلوماتية الاختبار وذلك بتحقيق الأهداف الآتية:

1. حساب قيم دالة معلوماتية الاختبار المقدرة باستخدام النموذجين (النموذج ثلاثي المعلمة ونموذج موكن اللامعلمي) في ضوء نوع ونسبة الأداء التفاضلي.

2. فحص التوافق في مستويات دالة معلوماتية الاختبار المقدرة باستخدام النموذجين (النموذج ثلاثي المعلمة ونموذج موكن اللامعلمي) في ضوء نوع ونسبة الأداء التفاضلي.

أهمية الدراسة:

تظهر أهمية الدراسة الحالية من خلال ما يأتي:

الأهمية النظرية: تتلخص الأهمية النظرية للدراسة في إلقاءها الضوء على أحد أهم مهددات الصدق الداخلي للاختبارات والمقاييس التربوية، ألا وهو الأداء التفاضلي بمختلف نسبه وانواعه وما يؤديه من تأثير على كمية المعلومات المستمدة من الفقرات والاختبارات، وهو الأمر الواجب التعامل معه بحذر شديد لتجنب أي هدر للمعلومات أو الوصول إلى استنتاجات خاطئة بشأن المفحوصين.

الأهمية العملية؛ تظهر الأهمية العملية للدراسة من خلال الآتي:

1. قد تسهم نتائج هذه الدراسة في تزويد باني الاختبار بنسبة ونوع الأداء التفاضلي المسموح بها في الاختبارات بحث لا تؤثر على صدق الاختبار وتحقيق أهدافه عبر المجموعات.

2. قد تسهم نتائج هذه الدراسة في المفاضلة بين نماذج نظرية الاستجابة للفقرة المعلمية واللامعلمية في ضوء نوع ونسبة الأداء التفاضلي للحصول على اختبار أكثر حيادية وصلاحية عبر المجموعات المتمايزة.

3. قد تساعد نتائج هذه الدراسة العاملين في المجال التربوي على بناء اختبارات محكية خالية من الفقرات ذات الأداء التفاضلي والتي من شأنها أن تراعي الفروق الفردية بين الطلبة، كخطوة إيجابية في زيادة فاعلية مخرجات العملية التعليمية.

حدود الدراسة:

تقتصر الدراسة الحالية على الحدود الآتية:

- استخدام نموذج نظرية الاستجابة للفقرة ثلاثي المعلم ونموذج موكن اللامعلمي.
- استخدام ثلاث قيم لنسب الفقرات ذات الأداء التفاضلي في الاختبار، وهي (10%، 20%، 30%).
- استخدام فقرات ثنائية التدريج (0، 1).
- استخدام بيانات مولدة (Simulated Data) من خلال برنامج (WinGen)، والذي تم تطويره كبرنامج حاسوبي لتوليد بيانات استجابة ثنائية ومتعددة التدريج للعديد من نماذج نظرية الاستجابة للفقرة (Han & Hambleton, 2007).

مصطلحات الدراسة:

- الأداء التفاضلي: عرفه (Jensen) بانه خطأ منتظم يجعل الاداء على الاختبار لمجموعة افضل من اخرى، أو هو تقدير متدني او مرتفع بانتظام لمعلمة المجتمع اعتماداً على إحصائي العينة (Jensen, 1980)، كما عرفه كل من كروكر والجينا (Crocker & Algina, 1986) بانه عبارة عن تلك الفقرة التي تكون متحيزة إذا كانت لا تتأثر بنفس مصادر التباين عند مستوى القدرة نفسها باختلاف المجموعات التي ينتمي لها الفرد، كما عرفه انجوف (Angoff) أنه: كفقرة تبدي خصائص مختلفة في مجموعات مختلفة ظاهرة للعيان بعد معادلة المجموعات على مقياس القدرة ، ويعرف ايضا بأنه احتمال اختلاف أداء

المجموعات المختلفة من الافراد من نفس مستوى القدرة على الفقرة (Zumbo, 2007)، ويعرفه أيضا الرحيل (2013) على أنه دلالة اشتقت إحصائيا للتعبير عن الفروق في الاستجابة للفقرة بين مجموعتين من الافراد ممن هم في نفس المستوى من القدرة. ويعرف اجرائياً في هذه الدراسة على أنه طريقة للتعبير عن الفروق في الاجابة عن الفقرات بين المفحوصين ممن هم في نفس المستوى من القدرة، والتي يمكن الكشف عنها وفق طرق نظرية الاستجابة للفقرة من خلال المعادلات الاحصائية التي يمكن أن تحدد فيما إذا كانت الفقرة تبدي أداءً تفاضلياً بين الأفراد من عدمه.

- نوع الأداء التفاضلي: نوع الأداء الذي تظهر فيه الفقرة دلالة مشتقة إحصائيا للتعبير عن الفروق في احتمالات الإجابة، إذا كان احتمال الإجابة الصحيحة للفقرة بين مجموعتين ممن هم في مستوى القدرة نفسه اكبر، ويكون منتظماً اذا كان احتمال الاجابة الصحيحة للفقرة اكبر لمجموعة دون مجموعة أخرى عند جميع مستويات القدرة، وغير منتظم إذا كان احتمال الإجابة الصحيحة للفقرة متفاوتا حيث يكون لصالح مجموعة عند مستوى قدرة معين، ولصالح المجموعة الأخرى عند مستوى قدرة آخر (Baker, 2001) ، كما عُرف الأداء التفاضلي المنتظم بوجود علاقة ثابتة بين درجة المفردة وانتماء الفرد لأحد المجموعات قيد الدراسة، في حين يكون الأداء التفاضلي غير المنتظم عندما تكون هذه العلاقة غير ثابتة (Michaelides, 2008).
- نظرية الاستجابة للفقرة (IRT): هي إحدى النظريات في مجال القياس والتقويم التي تفترض أن هناك سمات أو خصائص معينة يشترك فيها جميع الأفراد، ولكنهم يختلفون في مقدارها، وبالرغم من أن هذه السمات غير ملاحظة، إلا أنه يمكن الاستدلال على مقدارها من السلوك الملاحظ للفرد المتمثل في استجاباته على فقرات الاختبار، حيث تعتمد هذه النظرية على نماذج رياضية تستخدم في عمليات التقدير لكل من المعالم التي تصف الفقرة والقدرات الكامنة التي تصف الأفراد (Hambleton, 1990).
- الخصائص السيكومترية للفقرات: تتمثل في تقدير معالم الصعوبة، ومعالم التمييز، ومعالم التخمين، التي تم تقديرها بالاعتماد على نماذج نظرية الاستجابة للفقرة المستخدمة (المعلمية، واللامعلمية)، وهذه الخصائص تتمثل في صعوبة الفقرة (β) (Difficulty Parameter) وهي خاصية تصف الفقرة وتأخذ قيما تتراوح (− ∞ الى

+ ∞)، وعملياً من (−3 الى +3) وهي النقطة الواقعة على متصل القدرة التي تقابل احتمال إجابة صحيحة قدره ((1+2/(c)) في حالة النموذج اللوجستي ثلاثي المعلمة المستخدم في هذه الدراسة، حيث (c) هو معلمة التخمين للفقرة، وتمييز الفقرة (α) (Discrimination) والذي هو عبارة عن خاصية تصف الفقرة، وتأخذ قيما تتراوح (− ∞ الى + ∞)، وعملياً من (−3 الى +3) ويستدل عليها من خلال شدة انحدار أو ميل المماس لمنحنى خصائص الفقرة عند نقاط انعطاف منحنى الخصائص للفقرة، وتخمين الفقرة (c Guessing) وهو خاصية تصف الفقرة، وتأخذ قيما نظرية تتراوح في المدي (0 − 1)، وتمثل احتمال الإجابة عن الفقرة إجابة صحيحة بالتخمين، وتمثل تقاطع الخط التقاربي الأدنى لمنحنى الخصائص للفقرة مع المحور العمودي (McKinley & Reckase, 1983).

- الخصائص السيكومترية للاختبار: تتمثل في تقدير كمية المعلومات التي يقدمها الاختبار من خلال دالة المعلومات للاختبار، وتقدير الثبات للاختبار، التي تم تقديرها بالاعتماد على نماذج نظرية الاستجابة للفقرة المستخدمة (المعلمية، واللامعلمية).

- نماذج نظرية الاستجابة المعلمية: هي النماذج التي يمكن تطبيقها على فقرات ثنائية التدريج ذات شكل استجابة لوجستي محدد، وتفترض أن البيانات فيها واقعة بمستوى القياس الفئوي، ولا يوجد عدد محدد لهذه النماذج التي يمكن أن تتولد في إطار نظرية الاستجابة للفقرة، وفي هذه الدراسة تم استخدام النموذج اللوجستي ثلاثي المعلمة، والنموذج ثلاثي المعلمة هو النموذج الذي يمثل الفقرة فيه بواسطة دالة الترجيح اللوغاريتمي، معتمداً على كل من معلم التميز والتخمين بالإضافة لمعلم الصعوبة (Hambleton; Swaminthan & Rogers 1991).
- نماذج نظرية الاستجابة اللامعلمية: هي النماذج التي يمكن تطبيقها على فقرات ثنائية التدريج ذات شكل استجابة لوجستي غير محدد، وتتميز بإمكانية استخدامها للبيانات الرتبية والفئوية، وفي هذه الدراسة تم استخدام نموذج موكن اللامعلمي (Mokken & Lewis, 1982).
- المجموعة المستهدفة (Group Focal): هي المجموعة التي تشتمل على فقرات ذات أداء تفاضلي (منتظم أو غير منتظم)، ويتم مقارنتها مع المجموعة المرجعية (Zumbo, 1999).

- المجموعة المرجعية (Group Referenced): هي المجموعة التي لا تشتمل على أي فقرة ذات أداء تفاضلي، وهي المجموعة التي تقارن المجموعة المستهدفة معها (Yan, 2005).

الإطار النظري والدراسات السابقة:

أولًا: نظرية الاستجابة للفقرة (Item Response theory)

ظهرت نظرية الاستجابة للفقرة (IRT) نتيجة أوجه القصور والعيوب التي عانت منها النظرية الكلاسيكية (Classical Test Theory)، كتأثر إحصائيات الفقرة بخصائص المفحوصين من حيث الصعوبة والتمييز، والتغير في قدرات المفحوصين تبعًا لمؤشرات الفقرة الإحصائية، إذ يظهر هذا جليًا عندما يتم مقارنة مجموعتين من الأفراد خضعوا لصور متكافئة من الاختبار (Lord & Nonvick, 2008).

ولقد بين هامبلتون وسومثيان (Hambleton & Swaminatham, 1985) ثلاث مزايا رئيسة لنظرية السمات الكامنة أو النظرية الحديثة في القياس هي:

أولاً: بافتراض وجود مجموعة كبيرة نسبياً من فقرات الاختبار التي تقيس نفس السمة، يكون تقدير قدرة الفرد مستقلاً عن عينة الفقرات التي تطبق عليه، أي إن تقدير قدرات الافراد متحررة من الفقرات (Item-free).

ثانياً: بافتراض وجود مجتمع كبير من الافراد تكون تقديرات الخصائص السيكومترية للفقرات (مثل معاملات الصعوبة والتمييز) مستقلة عن عينة الافراد التي استخدمت في تقدير هذه الخصائص اي لا تتأثر معالم الفقرات بالعينة المختارة من المجتمع (Person-free).

ثالثاً: يمكننا الحصول على احصائي مثل (الخطأ المعياري في التقدير) لتقدير درجة الدقة في قياس القدرة لكل فرد، وربما يختلف هذا الاحصائي من فرد إلى آخر.

كما تنقسم النماذج في النظرية الحديثة (IRT) إلى نماذج معلميه (Parameter Models) ونماذج لا معلميه (Non-parameter Models)، وتهدف جميعها الى تحديد العلاقة بين اداء الفرد في الاختبار وهو ما يمكن ملاحظته مباشرة، وبين السمات أو القدرات التي تكمن وراء هذا الاداء وتفسيره.

وتستند نماذج نظرية الاستجابة للفقرة إلى عدد من الافتراضات، والتي لا بد من تحقيقها في البيانات الكمية لكي تؤدي إلى نماذج يمكن الوثوق بها، وهذه الافتراضات هي؛ (أحادية البعد، الاستقلال الموضعي، التحرر من السرعة، ومنحنى خاصية الفقرة).

نماذج الاستجابة للفقرة اللامعلمية (Non Parametric Item Response Theory Models):

وهي تلك النماذج التي لا تحدد شكل محدد لدالة استجابة الفقرة باستثناء ألا تكون متناقصة، كما أنها لا تفترض أي شكل سابق؛ كما يمكن الافتراض أن دوال الاستجابات التي يتم الحصول عليها من خلال النماذج اللامعلمية أقرب إلى الاستجابات الحقيقية التي يتم الحصول عليها من النماذج المعلمية؛ لأنها تعتمد على افتراضات أقل حول النموذج اللوجستي (Van der Linden & Hambleton, 1997).

كما أن النماذج اللامعلمية تستند على مفهوم التراكمية، ويمكن استخدامها في القياس النفسي والتربوي لدراسة جودة المقاييس وكفاءتها من خلال تحليلات جوتمان، التي تفترض أن كل فقرة وكل مفحوص يمتلكان موقعًا على متصل القدرة، إذ يجيب المفحوص إجابة صحيحة على الفقرة إذا كانت القدرة التي يمتلكها أكبر من صعوبة تلك الفقرة، الأمر الذي

يمكننا من التنبؤ بنمط استجابة المفحوص من خلال الدرجة الكلية التي يحصل عليها على الاختبار (Sijtsma & Hemker, 2002).

تقديرات المعالم في نظرية الاستجابة للفقرة:

عندما يتم اختيار نموذج من نماذج استجابة الفقرة، فمن الضروري تحديد قيمة معالم الفقرة: (الصعوبة، التخمين، التمييز) والقدرة والتي تصف كل من الفقرات والمفحوصين، وعلى افتراض أن الافضاء الكامل هو أحادي البعد (unidimensional) فعندها تكون هناك معلمة واحدة مسؤولة عن اداء المفحوص وهي (θ) (والتي تمثل قدرة المفحوص)، فالهدف الأول والأساسي في ضوء نظرية الاستجابة للفقرة هو تحديد موقع المفحوص على مقياس القدرة، ومن ثم معرفة احتمال الإجابة الصحيحة لكل مفحوص، وتقييمه في حدود القدرة التي يمتلكها، أما الهدف الثاني فهو إجراء المقارنات بين المفحوصين بهدف إعطائهم العلامات واتخاذ القرارات المختلفة بحقهم (Fitzpatrick & Wendy, 2001)، لذا ومن هذا المنطلق هناك عدة طرق تستخدم لتقدير قدرة المفحوصين والمعالم اللوجستية للفقرات؛ وهذه الطرق هي؛ (طرق تقدير الأرجحية العظمى، طريقة تقدير بييز).

الأداء التفاضلي للفقرة (Differential Item Functioning: DIF):

يستخدم الأداء التفاضلي للفقرة للتعبير عن الفروق في احتمالات الاستجابة الصحيحة للفقرة بين مجموعتين من نفس القدرة، فهو عبارة عن مشتقة إحصائية يتم من خلالها فحص العلاقة بين الاحتمالات المشروطة للاستجابات الصحيحة على الفقرة وبين السمة الكامنة (القدرة) المراد قياسها، بحيث تتم عملية المقارنة بين منحنيات خصائص الفقرة (ICCS) في المجموعتين الفرعيتين لمجتمع إحصائي معين، لكشف الأداء التفاضلي للفقرة بينهما، والتي يتم تمثيلها لاحقًا وفق النموذج اللوجستي المستخدم (Gierl, Jodoin & Ackerman, 2000).

كما إن الحديث عن مفهوم الأداء التفاضلي يقودنا إلى مصطلح الأداء التفاضلي للاختبار (Differential test Functioning: DTF)، والذي يعرف على أنه خطأ منتظم في القياس يؤثر على صدق الاختبار أكثر من ثباته، والذي يشير إلى الفارق في الأداء على الاختبار بين عينة المفحوصين من مجموعات مختلفة في العرق أو الجنس أو مكان السكن أو المستوى الاقتصادي وغيرها من المتغيرات الديمغرافية (Osterlind, 1983)، إذ يمكن التحقق من هذا الفرق من خلال استخدام مجموعة من الطرق تعتمد على النظرية الحديثة في القياس (IRT)، مثل الفرق بين منحنيات خصائص الاختبار (Test Characteristics Curve) في المجموعات المختلفة، وبهذا يمكن التأكيد على أن الأداء التفاضلي في الاختبار أكثر أهمية منه في الأداء التفاضلي للفقرة، إذ أن القرارات التي يتم اتخاذها بحق المفحوصين تعتمد علامات الاختبار ككل وليس علامة الفقرات بشكل مفرد (Teresi, 2001).

وفي هذا الصدد قام العديد من الباحثين والمختصين بالقياس النفسي والتربوي بأجراء العديد من الدراسات والابحاث المتعلقة بالأداء التفاضلي للفقرات والاختبارات، ومن هذه الدراسات ما يأتي:

حيث أجرى الشريفيين (2018) دراسة هدفت الكشف عن أثر نوع الأداء التفاضلي للفقرات (منتظم، غير منتظم) على الخصائص السيكومترية للفقرات، والاختبار وفق النموذج اللوجستي ثلاثي المعلمة، والنموذج اللوجستي الثلاثي اللامعلمي لنظرية الاستجابة للفقرة. ولتحقيق هدف الدراسة تم استخدام بيانات مولدة لاختبار من نوع الاختيار من متعدد، تكون كل منهما من (50) فقرة، لـ(1000) مفحوص، حيث اظهرت النتائج وجود فرق دال إحصائي بين الوسطين الحسابيين لكل من معالم التميز ومعالم التخمين يعزى

لنموذج نظرية الاستجابة للفقرة المستخدم؛ لصالح النموذج المعلمي مقارنة بالنموذج اللامعلمي، واظهرت النتائج ايضا وجود فرق دال إحصائيا بين الوسطين الحسابيين لمعالم الصعوبة يعزى لنموذج نظرية الاستجابة للفقرة المستخدم؛ لصالح النموذج اللامعلمي مقارنة بالنموذج المعلمي، وأشارت النتائج الى وجود فروق دالة إحصائياً في الوسط الحسابي لكل من الصعوبة والتخمين تعزى للتفاعل بين النموذج المستخدم ونوع الأداء التفاضلي، كما أشارت النتائج إلى وجود فرق دال إحصائي بين الوسطين الحسابيين لكمية معلومات للاختبار يعزى للنموذج، لصالح النموذج اللامعلمي مقارنة بالنموذج المعلمي، وبينت النتائج أن قيم معاملات الثبات المقدرة وفق النموذجين كانت عالية، وأن قيم معاملات الثبات المقدرة وفق النموذج المعلمي كانت الأعلى، سواء كان الأداء التفاضلي للفقرات منتظما أو غير منتظم.

وقام النوافلة (2013) بدراسة هدفت التعرف على أثر نسبة الفقرات ذات الأداء التفاضلي ومستوى الأداء التفاضلي في تقدير معالم الفقرات وقدرات الأفراد وفق النموذج اللوجستي ثلاثي المعلمة، حيث تم استخدام أسلوب المحاكاة في توليد البيانات تحت ظروف اختبارية متباينة، وقد اظهرت النتائج وجود دلالة إحصائية بين متوسطات تقديرات معالم الصعوبة تعزى إلى نسب الفقرات ذات الأداء التفاضلي لصالح المجموعة المستهدفة، وفروقات مختلفة بين متوسطات معالم التمييز والتخمين لصالح المجموعة المرجعية، كما تبين عدم وجود فروق ذات دلالة إحصائية في تقديرات القدرة تعزى إلى متغيري الدراسة، أما كمية المعلومات فإنها تنخفض كلما ازدادت نسبة الفقرات ذات الأداء التفاضلي.

وأجرى مساعدة (2013) دراسة هدفت إلى مقارنة ثلاثة أنواع من أنواع صياغة فقرات الاختبار (الاختيار من متعدد، الصواب والخطأ، التكميل) في مطابقة الفقرات وفق كل من نموذج موكن اللامعلمي ونموذج ثنائي المعلم، وبينت النتائج وجود فروق دال إحصائيًا بين متوسطات كمية المعلومات للاختبارات (اختيار من متعدد، صواب وخطأ، تكميل) تعزى لنموذجي موكن وثنائي المَعلم، ولصالح نموذج موكن لما يقدمه من كمية المعلومات لكلٍّ من الاختبارات مقارنة بما يقدمه النموذج ثُنائي المَعلَم من دالة للمعلومات لكلّ من الاختبارات.

كما قام دلوع (2013) بدراسة هدفت إلى بيان مدى التوافق بين النموذج الثلاثي المعلم البارامتري ونموذج موكن اللابارمتري لنوعين من الفقرات لاختبار تحصيلي (اختيار من متعدد، صواب/خطأ)، وقد اظهرت النتائج أن نموذج موكن اللامعلمي أكثر تشدداً في افتراضاته لمطابقة فقرات الاختبار، كما أن النموذج الثلاثي المعلم أكثر تشدداً في افتراضاته لمطابقة عدد الأفراد، والنموذج الثلاثي المعلم يهتم

أكثر بقدرة الأفراد، وبينت النتائج ايضا أن كمية المعلومات التي يقدمها النموذج الثلاثي المعلم أكبر من كمية المعلومات التي يقدمها نموذج موكن اللامعلمي.

كما أجرى القيسي (2013) دراسة هدفت إلى مقارنة دقة تقدير معالم الفقرة والقدرة باستخدام نماذج نظرية استجابة الفقرة المعلمية واللامعلمية باختلاف حجم العينة وطول الاختبار اعتماداً على مؤشري التحيز، والجذر التربيعي لمتوسط مربعات الخطأ، وذلك باستخدام بيانات مولدة (محاكاة)، وتوصلت الدراسة إلى أن تقديرات معالم الفقرة والقدرة بالطريقة المعلمية أقل منها في الطريقة اللامعلمية، بما يتعلق بمؤشر الدقة في القياس فقد دلت النتائج إلى أن قيم المتوسطات لطريقة التقدير المعلمية كانت اقل من المتوسطات لطريقة التقدير اللامعلمية في كل حالات الدراسة.

كما قام كل من دي كونينغ وسيجتسما وهمرز (De Koning, Sijtsma & Hamers, 2002) بدراسة هدفت الى مقارنة نموذجين معلميين ونموذجين آخرين لامعلميين من نماذج نظرية الاستجابة للفقرة، وذلك بهدف التعرف على فاعليتها في تحليل البيانات التجريبية للاختبار، والعمل على مقارنة نتائج نماذج نظرية الاستجابة للفقرة اللامعلمية مع مثيلاتها نماذج نظرية الاستجابة للفقرة المعلمية، وتوصلت النتائج إلى أفضلية الجمع بين النوعين من النماذج اللامعلمية والنماذج المعلمية، إذ أنه ثبت أن النماذج المعلمية فاعليتها في تقديم معلومات جيدة ومفيدة عن معالم الفقرات، بالإضافة فاعليتها في ثمة جوانب تطبيقية اخرى كمعايرة درجات الاختبار والاختبارات التكيفية، فالنماذج المعلمية واللامعلمية قدمت معلومات متباينة باستخدام إحصائيات مختلفة، واشارت الدراسة الى أن افضلية الجمع بينهما لتحسين نوعية الاختبار وجودة فقراته.

يتبين من خلال الدراسات السابقة وجود الكثير من التباين والاختلافات في النتائج التي تم التوصل إليها في فحص واختبار الأداء التفاضلي، وهو الأمر الذي يؤكد على أهمية متابعة مثل هذا النوع من الدراسات، لتقديم المزيد من الأدلة على النموذج الذي يوفر أكبر قدر من المعلومات لفقرات الاختبار وللاختبار ككل.

الطريقة والإجراءات:

منهجية الدراسة:

اتبعت الدراسة الحالية المنهج التجريبي للإجابة عن تساؤلات الدراسة كونه المنهج المناسب للدراسة، حيث تم استخدام بيانات مولدة تحاكي ظروف متعلقة بخصائص الفقرة وقدرات الأفراد في ضوء نوع ونسبة الأداء التفاضلي.

عينة الدراسة:

تكونت عينة الدراسة من مجموعة من الأفراد الافتراضيين عددهم (2000) موزعين بالتساوي على المجموعة المرجعية والمجموعة المستهدفة بحث تتبع قدرتهم التوزيع الطبيعي بوسط حسابي (0) وانحراف معياري (1). وقد تم استخدام برمجية (WinGen) لإجراء عملية التوليد.

أداة الدراسة:

تكونت أداة الدراسة من اختبار افتراضي بطول (60) فقرة بحيث يطابق النموذج الثلاثي التابع لنظرية الاستجابة للفقرة (3PLM) وتتبع معالم فقراته التوزيعات المبينة في الجدول (1).

الجدول (1): خصائص معالم فقرات الاختبارات المولَّدة.

المعلمة	التوزيع	الخصائص		بالرموز
الصعوبة	الطبيعي	متوسط حسابي = 0	انحراف المعياري =1	~N(0,1)
التمييز	المنتظم	حد أدنى = 0.5	حد اعلى = 1.5	~Uniform(0.5,1.5)
التخمين	المنتظم	حد أدنى = 0.1	حد اعلى = 0.25	~Uniform(0.1,0.25)

وقد تم توليد اختبارات متشابهة في توزيع معالم كل منها ولكنها متفاوتة في نوع الأداء التفاضلي (منتظم وغير منتظم) ونسبة الأداء التفاضلي (10%، 20%، 30%) في كل منها وهي النسب الأكثر استخداما في الدراسات السابقة وهي النسب الأكثر تواجدا في الاختبارات المعتمدة على البيانات الحقيقية كما هو موضح في الجدول (1) الذين يتناول الظروف التجريبية حيث تم محاكاة الأداء التفاضلي المنتظم من خلال عمل إزاحة في معلمة الصعوبة مقدراها (1) وكذلك تم محاكاة الأداء التفاضلي غير المنتظم من خلال عمل إزاحة في معلمة التمييز مقدارها (0.8) اتفاقا مع الدراسات السابقة، وقد تم توليد الصعوبة بتوزيع مشابه لتوزيع القدرة. كما وتم استخدام التوزيع المنتظم لمعلمه التمييز بقيمه صغرى (0.5) وقيمه عليا (1.5) لتشكل توزيع منتظم حول القيمة (1) وهي القيمة المحايدة لمعلمه التمييز. كما تم استخدام التوزيع المنتظم لمعلمه التخمين بقيمه صغرى (0.1) لتكون بحدها الأدنى غير المعلوم لتشير للنموذج الثلاثي المعلمة وبحدها الاعلى (0.25) وهو الحد الاعلى في معظم الدراسات والذي يشير لأربعه بدائل.

بيانات الدراسة:

تم إجراءات توليد استجابات افتراضية، حيث تم استخدام برمجية (WinGen) لتوليد استجابات افتراضية تحاكي قدرات عينة الدراسة ومعالم فقرات الاختبارات الافتراضية حيث نتج (6) مصفوفات من الاستجابات تحاكي ظروف الدراسة التجريبية الموضحة في الجدول (1).

ومن ثم تم تحليل الاستجابات المولدة في الخطوة السابقة من خلال كل من النموذجين (الثلاثي المعلمي، وموكن اللامعلمي) باستخدام برمجيتي (Bilog-MG، TestGraf) على الترتيب وفق النموذج ثلاثي المعلمة بعد ان تم التحقق من افتراضات نظرية الاستجابة للفقرة (أحادية البعد من خلال التحليل العاملي باستخدام برمجية (SPSS)؛ حيث تجاوزت قيمة نسبة الجذر الكامن الأول الى الجذر الكامن الثاني القيمة (2)، كما تم التحقق من الاستقلال الموضعي من خلال نسبة الأزواج المرتبطة الى الأزواج غير المرتبطة باستخدام برمجية (LDID) وقد بلغت أربعة اضعاف، ومنحنى خصائص الفقرة حيث تم تمثيلة باستخدام برمجية (Bilog-MG) وقد تبين من التمثيل البياني للمنحنيات خضوعها لما يسمى بدالة الاحتمالية من خلال اتخاذ الشكل المنحني (S)، ولم يتم التحقق من التحرر من السرعة لان البيانات افتراضية ومدة الاستجابة لا يمكن قياسها). كما تم التحقق من مدى مطابقة الأفراد والفقرات للنموذج ثلاثي المعلمة باستخدام برمجية بايلوج (BILOG-MG) باستجابات من خلال مؤشر مربع كاي لمطابقة الفقرات ومؤشر الاحتمال الهامشي للأفراد.

وقد تم تحليل الاستجابات وفق النموذجين المعلمي واللامعلمي حيث تم تقدير معالم الفقرات والافراد والمؤشرات السيكومترية للاختبار.

النتائج:

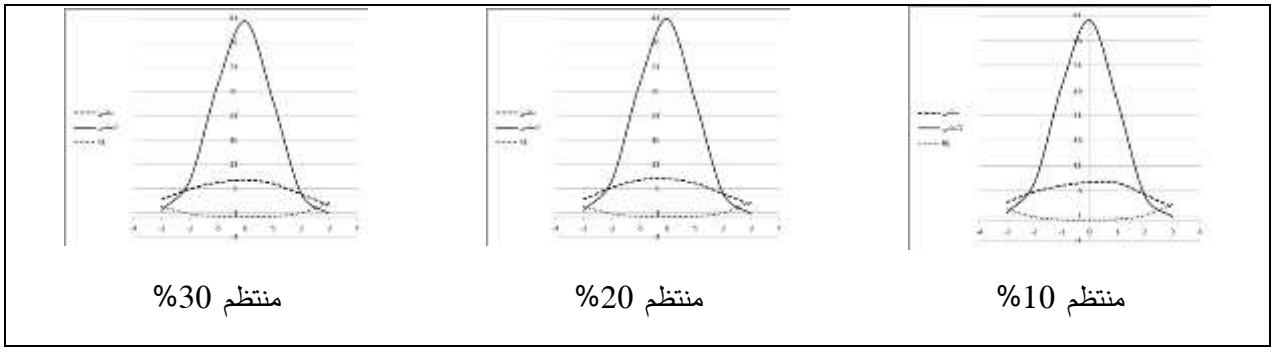
النتائج المتعلقة بالسؤال الأول والذي ينص على "ما قيم دالة معلوماتية الاختبار المقدرة باستخدام النموذجين (النموذج ثلاثي المعلمة ونموذج موكن اللامعلمي) في ضوء نوع ونسبة الأداء التفاضلي؟"

للإجابة عن السؤال تم استخراج قيم دالة معلوماتية الاختبار للنموذج المعلمي باستخدام برمجية (BilogMG) وللنموذج اللامعلمي باستخدام برمجية (TestGraf) ولكل من الظروف التجريبية عند (7) مستويات مختلفة للقدرة كما هو مبين في الجدول (2).

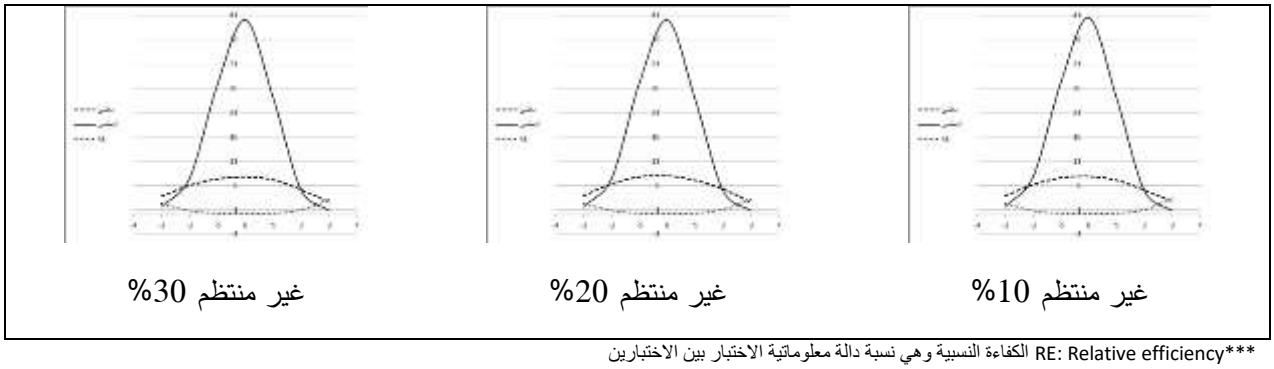
الجدول (2): دالة معلوماتية الاختبار في ضوء نوع ونسبة الأداء التفاضلي.

		منتظم			غير منتظم		
النموذج	القدرة	10%	20%	30%	10%	20%	30%
معلمي	−3	3.7085	3.7754	3.7363	3.907	3.8924	3.791
	−2	5.7996	6.1615	5.902	6.1459	6.3548	6.0271
	−1	7.0405	7.7277	7.2001	7.5797	7.7688	7.3517
	0	7.6915	7.9779	7.772	7.8582	7.9658	7.7675
	1	7.5507	7.0645	6.9891	7.0841	7.0496	7.1773
	2	5.2211	4.8808	4.7962	5.0381	5.2384	5.1104
	3	2.6672	2.6302	2.5812	2.7102	2.5298	2.6181
لا معلمي	−3	1.61069	1.58262	1.5781	1.65497	1.645	1.65818
	−2	7.63869	7.47607	7.5372	7.81811	7.40469	7.46676
	−1	27.1776	27.378	27.1733	27.2548	27.125	26.7614
	0	40.0119	40.7988	40.3103	40.4297	39.9631	40.005
	1	23.8728	24.0315	23.8402	23.93	23.584	23.9518
	2	4.92132	4.93242	4.91213	4.89684	4.86914	5.02033
	3	0.802703	0.81743	0.825632	0.795936	0.804575	0.839419

يلاحظ من قيم دالة معلوماتية الاختبار أنها متفاوتة عبر مستويات القدرة للنموذجين المعلمي واللامعلمي في ضوء نسبة ونوع الأداء التفاضلي وقد تم تمثيل الكفاءة النسبية للنموذجين كما يوضحها الشكل (1).



منتظم 10%　　منتظم 20%　　منتظم 30%



***RE: Relative efficiency الكفاءة النسبية وهي نسبة دالة معلوماتية الاختبار بين الاختبارين

الشكل (1): التمثيل البياني الكفاءة النسبية ودالة معلوماتية الاختبار لكل من النموذجين في ضوء نوع ونسبة الأداء التفاضلي.

يلاحظ من الشكل (1) ما يلي:

القيم القصوى لدالة معلوماتية كل من الاختبارين تكون عند الصفر ويمكن تفسير ذلك من خلال تمركز معلمة الصعوبة والقدرة حول القيمة صفر ومن المعلوم ان القيمة القصوى لدالة معلوماتية الفقرة تحصل عندما مستوى القدرة الذي يساوي صعوبتها. ولان دالة معلوماتية الاختبار عبارة عن مجموع قيم دوال معلوماتية فقراته فان القيم القصوى قد تقاطعت عند الصفر.

كما يلاحظ ان النموذج اللامعلمي أكثر كفاءة عند مستويات القدرة المتوسطة، بينما يكون النموذج المعلمي أكثر كفاءة عند الأطراف حيث القدرة المتطرفة وهذا يعكس دقة اعلى للنموذج المعلمي عند المستويات الدنيا والعليا للقدرة. وهذا يمكن تفسيره من خلال قدرة النموذج المعلمي على التقدير عبر مستويات القدرة بشكل متقارب وهو في ذلك يستند الى الافتراضات التي يشترط تحققها، بينما لا يشترط النموذج اللامعلمي تحقق افتراضات نظرية الاستجابة للفقرة ومنها أحادية البعد ومن هنا يمكن ان نلاحظ ان السمة أي القدرة تعتبر البعد الرئيس المفترض ان يقيسه الاختبار ويعتبر تطرف السمة يعتبر بعد آخر ولذلك اختلفت دقة التقدير. وكذلك يمكن تفسير ذلك من خلال انتهاك افتراض أحادية البعد فيما يتعلق بظهور بعد آخر يقيس أداء المجموعة بسبب احتواء الاختبار على فقرات ذات أداء تفاضلي وهو ما يفسر تفوق النموذج اللامعلمي على النموذج المعلمي على اغلب مستويات القدرة.

وتتفق نتيجة هذا السؤال جزئيا مع نتيجة دراسة مساعده (2013) التي خلصت إلى وجود فرق دال إحصائيا بين الوسطين الحسابيين لكمية معلومات للاختبار يعزى للنموذج، لصالح النموذج اللامعلمي مقارنة بالنموذج المعلمي لأغلب مستويات القدرة.

النتائج المتعلقة بالسؤال الثاني والذي ينص على: "هل يوجد توافق في مستويات دالة معلوماتية الاختبار المقدرة باستخدام النموذجين (النموذج ثلاثي المعلمة ونموذج موكن اللامعلمي) في ضوء نوع ونسبة الأداء التفاضلي؟"

تم تقريب القيم المقدرة لدالة معلوماتية الاختبار عند (7) مستويات للقدرة باستخدام النموذجين الثلاثي المعلمي باستخدام برمجية (BILOG-MG) وموكن اللامعلمي باستخدام برمجية (Test-Graf) وفي ضوء الظروف التجريبية المتمثلة في نوع ونسبة الأداء التفاضلي كما هو مبين في الجدول (3):

الجدول (3): مستوى دالة معلوماتية الاختبار للنموذجين عبر مستويات القدرة في ضوء نوع ونسبة الأداء التفاضلي.																			
	نوع الأداء	منتظم									غير منتظم								
	نسبة الأداء	10%			20%			30%			10%			20%			30%		
	النموذج	معلمي	لا معلمي	المجموع	معلمي	لا معلمي	المجموع	معلمي	لا معلمي	المجموع	معلمي	لا معلمي	المجموع	معلمي	لا معلمي	المجموع	معلمي	لا معلمي	المجموع
القدرة	-3	4	2	6	4	2	6	4	2	6	4	2	6	4	2	6	4	2	6
	-2	6	8	14	6	7	13	6	8	14	6	8	14	6	7	13	6	7	13
	-1	7	27	34	8	27	35	7	27	34	8	27	35	8	27	35	7	27	34
	0	8	40	48	8	41	49	8	40	48	8	40	48	8	40	48	8	40	48
	1	8	24	32	7	24	31	7	24	31	7	24	31	7	24	31	7	24	31
	2	5	5	10	5	5	10	5	5	10	5	5	10	5	5	10	5	5	10
	3	3	1	4	3	1	4	3	1	4	3	1	4	3	1	4	3	1	4
	المجموع	41	107	148	41	107	148	40	107	147	41	107	148	41	106	147	40	106	146

يلاحظ من خلال الجدول (3) وجود فروق ملاحظة في قيم دالة معلوماتية الاختبار لمستويات القدرة وللتحقق من التوافق بين تلك القيم للنموذجين تم استخدام اختبار مربع كاي ولكل من الظروف المختلفة كما هو مبين في الجدول (4):

الجدول (4): التوافق بين النموذجين في ثبات الاختبار (دالة معلوماتية الاختبار).																			
نوع الأداء	منتظم									غير منتظم									
نسبة الأداء	10%			20%			30%			10%			20%			30%		
الاختبار	قيمة الاختبار	درجة الحرية	الاحتمالية*	قيمة الاختبار	درجة الحرية	الاحتمالية*	قيمة الاختبار	درجة الحرية	الاحتمالية*	قيمة الاختبار	درجة الحرية	الاحتمالية*	قيمة الاختبار	درجة الحرية	الاحتمالية*	قيمة الاختبار	درجة الحرية	الاحتمالية*
مربع كاي	16.998	6	0.009	17.691	6	0.007	17.463	6	0.008	16.839	6	0.010	17.368	6	0.008	18.009	6	0.006
نسبة الأرجحية	15.737	6	0.015	16.372	6	0.012	16.024	6	0.014	15.555	6	0.016	16.081	6	0.013	16.563	6	0.011
العدد	148			148			147			148			147			146		

*مستوى الدلالة = 0.05

يتبين من خلال دلالة اختبار كاي في الجدول (4) أن دالة معلوماتية الاختبار تقدر بشكل مختلف عبر مستويات القدرة باختلاف النموذج المستخدم. ويمكن تفسير عدم وجود توافق في مستوى دالة معلوماتية الاختبار للنموذجين ولجميع الظروف ولصالح النموذج اللامعلمي من خلال افتراض التحليل اللامعلمي بعدم التناقص في منحنى الفقرة وهو ما يشير لعدم ظهور قيم سالبة لتمييز الفقرة وهذا بدوره يرفع من قيمة دالة معلوماتية الفقرة حيث أن التمييز بعلاقة إيجابية مع دالة معلوماتية الفقرة، في حين يسمح النموذج الثلاثي بتناقص المنحنى من خلال وجود معلمة التمييز التي تحتمل القيم السالبة وهذا بدوره يقلل من قيم دالة معلوماتية الفقرة، ولان دالة معلوماتية الاختبار ناتجة عن جمع دوال معلوماتية فقراته فانه ومما سبق يمكن تفسير الفروق في دالة معلوماتية الاختبار بالنسبة للنموذجين موكن اللامعلمي والثلاثي المعلمي. أما بالنسبة لعدم التوافق بين النموذجين في قيم معلوماتية الاختبار فيمكن تفسيره من خلال الفروق الواضحة بين النموذجين في قيم دالة معلوماتية الاختبار ولأغلب مستويات القدرة وخصوصا في المستويات المتوسطة من القدرة وهذه نتيجة طبيعية عند انتهاك افتراض أحادية البعد بوجود أداء تفاضلي في الاختبار حتى وان كان في بعض الفقرات حيث ان معلوماتية الاختبار ما هي إلا تجمع لمعلوماتية فقراته.

وتتفق نتيجة هذا السؤال مع نتيجة دراسة الشريفين (2018) حيث خلصت دراسة الشريفين لوجود فرق دال إحصائيا بين الوسطين الحسابيين لكمية معلومات للاختبار يعزى للنموذج، لصالح النموذج اللامعلمي مقارنة بالنموذج المعلمي. وتتفق مع نتيجة دراسة القيسي (2013) التي توصلت إلى أن المعلومات المقدمة من الاختبارات حسب النموذج اللامعلمي اعلى من تلك المقدمة عند استخدام النموذج المعلمي. وكذلك تتفق مع نتيجة دراسة مساعده (2013) حيث توصلت إلى أن كمية المعلومات الناتجة من استخدام نموذج موكن أعلى من تلك الناتجة من استخدام النموذج الثلاثي.

بينما تختلف نتائج هذا السؤال مع دراسة دلوع (2013) التي توصلت إلى أن النموذج الثلاثي ينتج معلومات أعلى من تلك المعلومات الناتجة من نموذج موكن اللامعلمي. وكذلك تختلف من نتائج دراسة دي كونينج وآخرون (De Koning, et al., 2002). التي توصلت لان النماذج المعلمية دائما تقدم معلومات أكثر. ويمكن تفسير الاختلاف من خلال الأداء التفاضلي في فقرات الاختبار لهذه الدراسة والتي أعطت للنموذج اللامعلمي أفضلية في بعض مناطق القدرة التي تتضح الفروق فيها بوجود أداء تفاضلي.

التوصيات والمقترحات:

في ضوء النتائج التي تم التوصل لها فانه يمكن التوصية بما يلي:

- استخدام النموذج المعلمي عند تحقق افتراضات نظرية الاستجابة للفقرة عندما يكون الافراد ضمن الفئات متطرفة القدرة.
- ضرورة التحقق من افتراضات النظرية الحديثة عند توخي أعلى درجات الثبات في الاختبار وفي ضوء ذلك يتم اختيار النموذج الأنسب (معلمي/ لا معلمي) عند تقدير معالم الفقرات.
- إجراء الدراسة باستخدام بيانات حقيقية.

المصادر والمراجع:

دلوع، أحمد عبد العزيز (2013). ***مدى التوافق بين نموذج استجابة الفقرة الثلاثي المعلمة البارامتري ونموذج موكن اللابارمتري لنوعين من فقرات اختبار تحصيلي.*** رسالة دكتوراه غير منشورة، جامعة اليرموك، إربد–الأردن

الرحيل، راتب (2013). ***أثر وجود أداء تفاضلي في الفقرات المرساوية على دقة المعادلة العمودية لاختبار اوتيس لينون للقدرات العقلية.*** المجلة الدولية التربوية المتخصصة، 2(8)، 755–771.

الشريفين، نضال (2018). أثر نوع الأداء التفاضلي للفقرات على الخصائص السيكومترية للفقرات والاختبار وفق النماذج المعلمية والنماذج اللامعلمية لنظرية الاستجابة للفقرة. ***مجلة دراسات العلوم التربوية***، مج45، ملحق (632–605).

القيسي، حسين عبد النبي (2013). ***دقة تقدير معالم الفقرة والقدرة باستخدام نماذج نظرية استجابة الفقرة البارامترية اللابارامترية باختلاف حجم العينة وطول الاختبار.*** رسالة دكتوراه غير منشورة، جامعة اليرموك، إربد–الأردن

مساعدة، محمد (2013). ***أثر شكل فقرات الاختبار في مطابقة الفقرات لنموذج موكن اللابارمتري ونموذج ثنائي المعلمة البارامتري.*** رسالة دكتوراه غير منشورة، جامعة اليرموك، إربد، الأردن.

النوافلة، على (2013). ***أثر نسبة الفقرات ذات الأداء التفاضلي ومستواه في تقدير معالم فقرات الاختبار وقدرات***

الأفراد وفق نموذج الاستجابة ثلاثي المعلمة. أطروحة دكتوراه غير منشوره، جامعة اليرموك، إربد، الأردن.

Abbott, M. (2007). A Confirmatory Approach to differential item functioning on an ESL reading assessment. ***Language Testing***, 24 (1), 7-36.

Acar, T., & Kelecioglu, H. (2010). Comparison of Differential Item Functioning Determination Techniques: HGLM, LR and IRT-LR. ***Educational Sciences: Theory and Practice,*** 10(2), 639-649.

Baker, F. (2001). ***The Basics of item Response Theory***. University of Maryland, college park, MD: ERIC clearinghouse on Assessment and Evaluation.

Crocker, L., & Algina, J. (1986). ***Introduction to classical and modern test theory*****.** Holt, Rinehart and Winston, 6277 Sea Harbor Drive, Orlando, FL 32887.

Croudace, T., & Brown, A. (2012). ***Measurement Invariance and Differential Item Functioning.*** Short Course in Applied Psychometrics-Peter house College, 10-12.

De Koning, E., Sijtsma, K., & Hamers, J. H. (2002). Comparison of Four IRT Models When Analyzing Two Tests for Inductive Reasoning. ***Applied Psychological Measurement***, 26(3), 302-320.

Fitzpatrick, A. & Wendy, M. (2001). The Effects of Test Length and Sample Size on the Reliability and Equating of Tests Composed of Constructed Response Items. ***Applied Measurement In Education***, 14(1), 31–57.

Gierl, M. J., Jodoin, M. G., & Ackerman, T. A. (2000). ***Performance of MantelHaenszel, Simultaneous Item Bias Test, and logistic regression when the proportion of DIF items is large.*** Paper presented at the Annual Meeting of the American Educational Research Association (AERA), New Orleans, LA.

Gruijter, D. & Kamp, L. (2005). ***Statistical Test Theory for Education and Psychology*****.** Retrieved December 30, 2021 from: Www.Leidenuniv.Nl/Gruijterdnmde.

Gybels, J. (2004). The Performance of Some Observed and Unobserved Conditional Invariance Techniques for the Detection of Differential Item Functioning. ***Quality and Quantity***, 38(6), 681-702.

Hambleton, R. K., & Swaminathan, H. (1985). **A Look at Psychometrics in the Netherlands.**

Hambleton, R. (1990). Item response theory: introduction and bibliography. ***Psicothema***, 2(1), 97-107.

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). **Fundamentals of item response theory**. Newbury Park California: stage publications.

Journal of Educational and Behavioral Statistics, 25(2), 391-415.

Teresi, A. (2001). Statistical Methods for Examination of Differential Item Functioning (DIF) with Applications to Cross-Cultural Measurement of Functional, Physical and Mental Health. ***Journal of Mental Health and Aging,*** 7(1):31-40

Van der Linden, W.J, & Hambleton, R. K. (1997). Part IV: Non-Parametric models introduction in W. J. van der Linden & R. K. Hambleton, (1994), ***Handbook of modern item response theory*** (pp.347- 349).New York: Springer- Verlag.

Yan, Song (2005). ***Gender-related Differential Item Functioning in Mathematics Assessment on the Third International Mathematics and Science Study-Repeat (TIMSS-R).*** The University of Toledo, ProQuest Dissertations Publishing, 2005. 3177610.

Zumbo, B (2007). Three Generations of DIF Analyses: Considering Where It Has Been, Where It Is Now, And Where It Is Going. ***Language Assessment Quarter,*** 4(2), 223-233.

Zumbo, B. (1999). ***A Handbook on the Theory and Methods of Differential Item Function (DIF).*** Ottawa، on Directory of Human Resources research And Evaluation Department of National Defense.

Han, T., & Hambleton, K. (2007). ***User's Manual: WinGen (*Center for Educational Assessment Report No. 642).** Amherst, MA: University of Massachusetts, School of Education.

Jensen, R. (1980). ***Bias in Mental Testing***. New York: A Division of Macmillan Publishing Co.

Lord, F. M., & Novick, M. R. (2008). ***Statistical theories of mental test scores*.** IAP.

McKinley, R. Reckase, M. (1983). ***An Extension of the Two-Parameter Logistic Model to the Multidimensional Latent Space***. Iowa City, the American College Testing Program.

Michaelides, M. (2008). An Illustration of a Mantel-Haenszel Procedure to Flag Misbehaving Common Items in Test Equating. ***Practical Assessment, Research & Evaluation***, 13(7). Available online: http://pareonline.net/getvn.asp?v= 13&n=7.

Mokken, R. J. and Lewis, C. (1982). A nonparametric approach to the analysis of Dichotomous item responses**.** ***Applied Psychological Measurement***, (6), 417-430.

Osterlind, S. (1983). ***Test Item Bias***. Beverly Hills; Sage publications.

Sijtsma, K. & Hamker, B. T. (2002). A taxonomy of IRT models for ordering of persons and item using simple sum scores.