مقارنة مطابقة الفقرات بين نموذج موكن اللابارامتري والنموذج ثنائي المعالم البارامتري

رنا ثاني ضامن المومني*

ملخص

هدفت الدراسة إلى الكشف عن النموذج الأفضل لمطابقة فقرات اختبار تحصيلي وذلك بمقارنة نتائج تحليل البيانات باستخدام النموذج ثنائي المعلمة البارامتري مع نموذج موكن للتجانس الاطرادي اللابارامتري، ولتحقيق أهداف الدراسة تم تطبيق اختبار تحصيلي في الرياضيات (من إعداد الباحثة) تكون من (28) فقرة من نوع الاختيار من متعدد بأربعة بدائل على عينة الدراسة التي تكونت من (834) طالبة من طالبات الصف العاشر الأساسي وذلك في العام الدراسي 2011/2010 تم اختيارها بالطريقة العشوائية البسيطة. بعد ذلك تم التحقق من الافتراضات الأساسية لنظرية استجابة الفقرة باستخدام الحزمة الاحصائية (SPSS) ومن ثم تحليل البيانات باستخدام الحزمة الاحصائية (R package) لاستخلاص النتائج الخاصة بنموذج موكن للتجانس الاطرادي، كما تم استخدام برنامج (Bilog-Mg) لاستخلاص النتائج الخاصة بالنموذج ثنائي المعلمة وأيضا تم استخراج دوال المعلومات لجميع فقرات الاختبار وكذلك للاختبار ككل باستخدام البرنامج الاحصائي (TESTGRAF)، وقد أظهرت نتائج الدراسة تفوق نموذج موكن في مطابقة فقرات الاختبار عند مستوى الدلالة ($\alpha = 0.05$) إذ بلغ عدد الفقرات المطابقة له (23) فقرة مقابل (14) فقرة طابقت النموذج ثنائي المعلمة، كما أظهرت النتائج المتعلقة بدوال معلومات الاختبار بصورتيه البارامترية واللابارامترية أن الاختبار بصورته اللابارامترية كان الأفضل في تقديم المعلومات حول السمة المقيسة، وقد تمت مناقشة النتائج التي توصلت اليها الدراسة في ضوء الأدب النظري والدراسات السابقة، وأوصت الباحثة بدراسة المطابقة على عينات مختلفة الأحجام، ونماذج بارامترية ولابارامترية، وأنواع مختلفة من الاختبارات والفقرات.

الكلمات الدالة: نظرية استجابة الفقرة، النماذج البارامترية، النماذج اللابارامترية، نموذج موكن للتجانس الاطرادي، نماذج موكن، معاملات التدريج.

المقدمة

تعد نظرية استجابة الفقرة بنماذجها البارامترية المختلفة خياراً جيداً لتحليل فقرات الاختبار للخروج بنتائج دقيقة وموثوقة، إلا أن استخدام هذه النماذج يتطلب تحقق مجموعة من الافتراضات الأساسية إضافة الى بعض الشروط الخاصة لكل نموذج الأمر الذي يقلل إمكانية استخدام هذه النماذج نظراً لصعوبة تحقيق هذه الشروط، وتعد مطابقة فقرات الاختبار واحدة من أهم الأمور التي تقف عائقا أمام استخدام نماذج نظرية استجابة الفقرة مما يظهر الحاجة لدراستها ومحاولة توفير بدائل مناسبة لتلافي هذه المشكلة من خلال استخدام نماذج استجابة الفقرة اللابارامترية التي تتطلب افتراضات أقل تشدداً.

توضح نظرية استجابة الفقرة العلاقة بين قدرات المفحوصين واستجاباتهم على فقرات الاختبار بنماذج احصائيةٍ (Crocker & Algina,1986) وتنقسم هذه النماذج الى نوعين رئيسين، الأول يعرف بنماذج استجابة الفقرة البارامترية (Parametric Item Response Theory (PIRT)) حيث يكون شكل دالة استجابة الفقرة (Item Response Function (IRF)) محدداً، بينما يعرف النوع الثاني بنماذج استجابة الفقرة اللابارامترية (Non-parametric Item Response Theory (NIRT)) التي لا تحدد شكل دالة استجابة الفقرة، وبالرغم من الاختلافات بينها إلا أنها تشترك في كثيرٍ من الاستخدامات التطبيقية كتطوير المقاييس المختلفة وتحديد صدق المقياس.

يشترط نجاح استخدام أحد تطبيقات نماذج نظرية استجابة الفقرة والحصول على الفوائد المتوخاة منها حسن مطابقة النموذج المختار للبيانات، إذ يعد حسن المطابقة أحد أهم الموضوعات التي يتناولها الباحثون والدارسون في هذه النظرية فمن الضروري فهم مسبباتها المختلفة (Maydeu, 2005).

إن مطابقة النموذج البارامتري للبيانات ذات المستوى الرتبي من مشكلات المطابقة حيث أساغها البعض في حال توفرت

* قسم علم النفس، جامعة القصيم، السعودية. تاريخ استلام البحث 2014/04/17، وتاريخ قبوله 2015/01/19.

©2017 عمادة البحث العلمي وضمان الجودة/ الجامعة الأردنية. جميع الحقوق محفوظة.

افتراضات النظرية للبيانات الرتبية (Crocker & Algina, 1986)، في حين ركز البعض الآخر على التمييز بين المقاييس الرتبية ومقاييس الفترة وتطبيقاتهما النهائية على الإحصاءات البارامترية مع تطوير الحلول المناسبة لهذا الغرض، وقد كانت التوصية باستخدام النماذج اللابارامترية في تحليل البيانات الرتبية لأن المقاييس التربوية نادراً ما تتحلى بقوة خصائص مقاييس الفترة (Mokken & Lewis, 1982).

تعد دالة استجابة الفقرة ركيزةً أساسيةً للمنحيين البارامتري واللابارامتري وتعرف على أنها العلاقة بين احتمالية حصول المفحوص على الدرجة (1) (الإجابة المفتاحية) لفقرة ثنائية والسمة الكامنة أو القدرة لذلك المفحوص (θ)، وتفترض النماذج البارامترية قيوداً لشكل هذه الدالة إلا أن نظيرتها اللابارامترية لا تفترض شكلاً معيناً لها (Sijtsma, 1998) الأمر الذي يجعل من النماذج اللابارامترية خياراً أفضل في كثير من الأحيان، وذكر كل من جنكر وسيجتسما (Junker & Sijtsma, 2001) ثلاثة أسباب تؤكد أهمية النماذج اللابارامترية، وهي:

- تقدم النماذج اللابارامترية فهماً أعمق لما تقوم به نماذج استجابة الفقرة البارامترية.
- تقدم النماذج اللابارامترية إطاراً أكثر مرونة في حال ضعف المطابقة للنماذج البارامترية.
- تقدم النماذج اللابارامترية طرقاً وإجراءات سهلة للبيانات وتسمح باستخدام عينات أقل من الفقرات والأفراد عن تلك المستخدمة في النماذج البارامترية.

نظرية استجابة الفقرة (Item Reponse Theory):

أشار كلٌ من كليف وكيتس (Cliff & Keats, 2003) إلى أن نماذج نظرية استجابة الفقرة تنقسم بشكل رئيس إلى قسمين أساسيين هما النماذج البارامترية واللابارامترية، إلا أنها جميعاً تقوم على فكرة وجود تدريج كامن لقدرة معينة أو سمة شخصية وأن احتمال الإجابة الصحيحة (الموافقة) على الفقرة تتغير عبر ذلك التدريج حيث تقترب من الواحد بالانتقال من المفحوصين ذوي القدرة المنخفضة إلى ذوي القدرة العالية، وقد تم عرض دوال رياضية مختلفة لتمثيل دالة استجابة الفقرة حيث تم اقتراح الشكل اللوجستي والذي مثل المنحى البارامتري منذ سنوات عدة كخيار جيد.

وقدمت نظرية استجابة الفقرة معايير هامة في عملية اختيار الفقرات لتتميز عن غيرها ضمن هذا المجال فهي تعتمد على دالة استجابة الفقرة بشكلٍ أساسي إضافة إلى الاستفادة مما يعرف بدالة معلومات الفقرة ($I_i(\theta)$) Item Information Function) ودالة معلومات الاختبار ($I(\theta)$) Test Information Function) (Lord, 1977).

ويشير بيكر (Baker, 2001) إلى أن نماذج استجابة الفقرة البارامترية الثلاثة هي النماذج الأكثر شهرة وانتشاراً، والتي يمكن التعبير عنها من خلال المعادلة التالية:

$$Pi(\theta) = c_i + (1 - c_i)\frac{exp[Da_i\ (\theta - b_i)]}{1+exp[Da_i\ (\theta - b_i)]} \quad .....(1)$$

حيث أن،

(θ): قدرة المفحوص.

$Pi(\theta)$: احتمال أن المفحوص الذي يمتلك مستوى قدرة (θ) أجاب عن الفقرة (i) إجابة صحيحة.

b_i: النقطة التي تقابل Pi (θ)= 0.5(c+1) ، وهو عبارة عن معلم صعوبة الفقرة.

a_i: ميل الدالة عند النقطة b على متصل السمة، وهو عبارة عن معلم تمييز الفقرة.

c_i: المماس السفلي للدالة، وهو عبارة عن معلم التخمين.

D: معلم التدريج، وإذا ساوى القيمة (1,702) فان الدالة تتبع (Normal Ogive Distribution).

دقة القياس في نماذج استجابة الفقرة البارامترية.

يذكر بيكر (Baker, 2001) أن تحديد دقة القياس في النماذج البارامترية يتم من خلال دالة معلومات الفقرة، والتي تزود بالمعلومات التي تقدمها الفقرة عند كل مستوى من مستويات القدرة والمعطاة بالمعادلة التالية في حال النموذج البارامتري ثنائي المعلمة:

$$I_i(\theta) = a^2 p_i(\theta) Q_i(\theta) \quad .....(2)$$

كما يضيف أنه وبعد استخراج دوال معلومات جميع الفقرات يتم استخراج دالة معلومات الاختبار التي تحدد حجم المعلومات التي يزودنا بها الاختبار ككل عند مستويات القدرة المختلفة من خلال المعادلة التالية:

$$I(\theta) = \sum_{i=1}^{n} I_i(\theta) \quad .....(3)$$

ويقوم كل من برنامجي (TETGRAF) و(Bilog-MG) باستخراج هاتين الدالتين إلا أن رامسي (Ramsay, 2000) يرى أن برنامج (TETGRAF) يقوم بتقدير هذه الدالة بشكل أسرع بـ (500) ضعف عنه في (Bilog-MG)، إضافة إلى تفوقه في دقة تقدير دالة معلومات الفقرة.

نماذج نظرية استجابة الفقرة اللابارامترية:

لعبت نظرية استجابة الفقرة اللابارامترية دوراً مهماً في تطوير نظرية استجابة الفقرة، فهي تمثل الحالة العامة لنظريات استجابة الفقرة التي تشكل البارامترية حالة خاصة منها (Sijtsma & Hemker, 2000) إلا أن النماذج اللابارامترية لا

تفترض شكلاً محدداً لدالة استجابة الفقرة، فالشرط الوحيد أن لا تكون متناقصةً مع زيادة مستوى القدرة (θ) (Sijtsma, 1998).

وأثبتت النماذج اللابارامترية أنها نماذج إحصائية يمكن استخدامها في القياس النفسي لدراسة جودة وكفاءة المقياس من خلال تحليلات جتمان (Guttman scalogram) التي تفترض أن كل من الفقرة والمفحوص يمتلكان موقعاً على متصل القدرة، حيث يجيب المفحوص على الفقرة إجابة صحيحة فقط في حال كانت قدرته أعلى من صعوبة تلك الفقرة، وقد لا يتحقق ذلك تجريبياً مما أدى إلى ظهور الفكرة القائمة على الاحتمالية لا التحديد، التي نصت على أن احتمالية الإجابة الصحيحة مرتفعة (لكن $\neq 1$) كلما ازدادت قدرة المفحوص، كما أنها قليلة (لكن $\neq 0$) عند انخفاض قدرة المفحوص (Linden & Hambleton, 1997). وتقاس قوة التدريج لموكن والحكم عليه من خلال عدد أخطاء جتمان لكل من فقرات الاختبار والاختبار ككل ودلالتها الاحصائية (Watson & Thompson, 2014).

وتتبع أهمية نماذج نظرية استجابة الفقرة اللابارامترية من قدرتها على استخراج معالم هامة مثل نسبة الإجابات الصحيحة، ومؤشرات الصعوبة، ومعاملات التدريج المختلفة والتي تشير الى القوة التمييزية لفقرات الاختبار والاختبار ككل، وتضمن هذه النماذج ترتيب المفحوصين باستخدام درجاتهم على الاختبار على الرغم من وجود الخطأ العشوائي، إذ أنها تعتبر أن هذه الدرجات تعكس رتب القدرة (θ)، فعند استخدام أحد نماذج استجابة الفقرة لبناء اختبار وكان التصور التطبيقي كافٍ لتحقيق مقياسٍ رتبي التدريج للمفحوصين، فإن النماذج البارامترية تمتلك محدداتٍ غير ملائمةٍ لمثل هذا الهدف مما يؤدي الى رفض الكثير من الفقرات التي لا تتوائم مع الشكل اللوجستي، مما يعني فقدان الكثير من الفقرات التي تفيد في دقة ترتيب الأفراد والتي تمتلك دوال استجابة غير متناقصة مما يؤثر بشكلٍ واضح في معامل الثبات (Sijtsma & Molenaar, 2002). وقد ذكرت ويند (Wind, 2014) أن النماذج اللابارامترية مناسبة في تطوير الاختبارات التي تقع استجاباتها على مقياس رتبي إذ إن النماذج البارامترية في هذه الحالة تعد متشددة بشكل مبالغ به.

وتقوم النماذج اللابارامترية على مجموعة من الافتراضات الأقل تشدداً وهي كالتالي:

1. أحادية البعد (Uni-dimentionality): الاستجابات على الفقرات تتبع متغيركامن وحيد.
2. الاستقلال الموضعي (Local Independence): احتمالية الاجابة على أي فقرة غير مرتبطة بالاستجابة على أي فقرة أخرى في الاختبار.
3. الاطرادية (Monotonicity): بازدياد القدرة تزداد احتمالية الاجابة الصحيحة أو تبقى ثابتة.
4. الاطرادية المضاعفة (Double Monotonicity): هو الافتراض الأصعب لتحقق النموذج والمتضمن امتلاك دوال استجابة غير متقاطعة لفقرات الاختبار التي تشكل التدريج.

وتقسم النماذج اللابارامترية الى قسمين رئيسين وذلك كما يلي:

نموذج التجانس الاطرادي (Monotone Homogeneous model (MHM)):

يعد نموذج التجانس الاطرادياً أو نموذج موكن (Mokken Scaling model) أحد النماذج اللابارامترية للاستجابات الثنائية، وقد وصف كل من مولينار وسيجتسما (Molenaar & Sijtsma, 2000) تحليل التدريج لموكن على أنه نسخة معدلة احتمالية لتحليل التدريج لجتمان، ويتيح هذا النموذج إمكانية ترتيب الأفراد تبعاً لمستوى القدرة $(\boldsymbol{\theta})$ باستخدام الدرجة الكلية للاختبار.

ويذكر كل من مولينار وسيجتسما (Molenaar & Sijtsma, 2000; Reidpath & Ahmadi, 2014) أن نموذج موكن للتجانس الاطرادي يقود إلى ترتيب الأفراد على متصل القدرة $(\boldsymbol{\theta})$ باستخدام درجاتهم الكلية على الاختبار بالرغم من عدم إمكانية الحصول على تقديراتٍ رقميةٍ للقدرة كما في النماذج البارامترية، وتعد هذه النتيجة المفتاحية التي تبرر استخدام هذا النموذج في قياس المفحوصين.

معامل التدريج (scalability coefficient):

يعد نموذج جتمان أصل معاملات التدريج (H, H_i, H_{ij}) فقد تم تبني معامل لوفنجر (Loevinger`s H) من قبل موكن في العام (1971م) لتعريف عائلة جديدة من المعاملات التي تشير إلى التجانس الاطرادي ومدى تحققه والذي يأخذ بعين الاعتبار كلاً من أزواج الفقرات للاختبار، إضافةً إلى كل فقرة تبعاً لباقي الفقرات، وفقرات الاختبار ككل(Linden & Hambelton, 1997).

قابلية التدريج (Scalability):

عرف كلٌ من موكن ولويس (Mokken& Lewis, 1982) التدريج على أنه مجموعة من الفقرات التي ترتبط ايجابياً حيث يكون معامل التدريج (H_i) أكبر أو يساوي قيمة ثابتة موجبة ($0 < c < 1$)، كما يجب أن يكون معامل التدريج الكلي للمقياس أكبر من هذه القيمة الثابتة التي عرفت بأنها ثابت تعريف التدريج (scale-defining constant) وفي حال تحقق التعريف السابق على مجموعة الفقرات يطلق عليها أنها فقرات تتبع تدريج موكن.

ويستخدم نموذج موكن معامل لوفنجر المرتكز على خاصية التجانس الاطرادي والذي يستخرج بدلالة معاملات تدريج الفقرات (H_i) (Mokken & Lewis, 1982)، حيث يعتبر معامل (H) معامل لتدريج الاختبار ككل، ومعامل (H_i) معاملاً لتدريج الفقرة (item-level coefficient) الذي يستخدم لتقييم تدريج الفقرة تبعاً للفقرات الأخرى في المقياس، كما يعد مؤشراً على القوة التمييزية للفقرة (Van Onna, 2003)، ووجد أن قيم (H_i) المرتفعة تتواجد عند المناطق شديدة الانحدار لدالة إستجابة الفقرة (شبه العمودية) بينما تتواجد القيم المنخفضة عند المناطق المستوية لدالة إستجابة الفقرة، وقد بينت الدراسات أن القوة التمييزية للفقرات في مقاييس الاتجاهات كانت أكبر منها في المقاييس العقلية المختلفة (Meijer, 2014)، وتقبل قيمة (H_i) في حال كانت موجبة وفي بعض الدراسات إذا بلغت (0.3) فأكثر كما يتم استخراج معامل تدريج الفقرة (H_i) من خلال معامل تدريج أزواج الفقرات (H_{ij}) الذي يرتكز الى نموذج جتمان في تحديد أخطاء جتمان لأزواج الفقرات المختلفة في المقياس (Mokken & Lewis, 1982). وقد أفاد شينكن وآخرون (Shenkin, Watson, Laidlaw, Starr, Deary, 2014) أن النماذج اللابارمترية نماذج مفيدة في تطوير وبناء المقاييس الهرمية.

وقد أورد سيجتسما (Sijtsma, 1988) معاملات تدريج موكن المختلفة وطرق حسابها، إذ يتم حساب معامل (H_{ij}) من خلال المعادلة:

$$H_{ij} = 1 - \frac{F_{ij}}{E_{ij}} \quad \ldots\ldots\ldots\ldots\ldots (5)$$

حيث أن:

- i , j :فقرتين في الاختبار.
- F_{ij} : عدد أخطاء جتمان.
- E_{ij}: العدد المتوقع من الأخطاء تحت افتراض استقلالية الفقرات.

كما يمكن استخراج معامل تدريج الفقرة (H_i) لكل فقرة من فقرات المقياس المكون من(k) من الفقرات، وهو معامل يناظر معلم التمييز في النماذج البارامترية الثنائية والثلاثية من خلال المعادلة التالية:

$$H_i = 1 - \frac{F_i}{E_i} \quad \ldots\ldots\ldots\ldots\ldots (6)$$

حيث أن: $E_i = \Sigma E_{ij}$, $F_i = \Sigma F_{ij}$

وهي مجموع تكرارات الأخطاء لجميع الفقرات $(i = 1,2,\ldots\ldots\ldots,k\,;i \neq j)$.

ثم يتم استخراج معامل التدريج الكلي للمقياس من خلال جميع قيم (Hi) للفقرات باستخدام المعادلة:

$$\ldots\ldots\ldots\ldots\ldots\ldots (7)$$

حيث أن: $F = \Sigma F_i, E = \Sigma E_i$

فكلما زادت قيمة (H) دل ذلك على موثوقية أكبر في المقياس وقدرته في ترتيب المفحوصين على التدريج باستخدام الدرجة الملاحظة كما في نظرية الاختبار الكلاسيكية (Sijtsma, 1998).

ويضيف سيجتسما (Sijtsma, 1998) أن مجموعة الفقرات توصف بأنها تتبع تدريج موكن إذا كان معامل الارتباط (product-moment correlation)(ρ_{ij}) لأي قيمة ثابتة (c) حيث أن $(0 < c < 1)$ وفقا لمحكين، هما:

- $\rho_{ij} > 0$ أو يكافئها $H_{ij} > 0$ لجميع أزواج الفقرات.
- $Hi > c > 0$ لجميع قيم i.

إذ يقتضي المحك الثاني أن تكون جميع معاملات التدريج موجبة كالتالي:

$0 \leq H_{ij} \leq 1$ لجميع قيم i,j

$0 \leq H_i \leq 1$ لجميع قيم i مع باقي فقرات المقياس.

$0 \leq H \leq 1$

وقد أشار كل من سيجتسما ومولينار (Sijtsma & Molenaar, 2002) أن موكن قدم بعض الاجتهادات بوضع معايير لمعاملات التدريج لضمان جودة التدريج، واقترح حداً أدنى لمعامل تدريج الفقرة ($H_i = 0.3$)، كما اقترح بعض المعايير للحكم على جودة التدريج الكلي وهي كالتالي:

- مقياس غير قابل للتدريج $H < 0.3$.
- مقياس ضعيف $0.3 \leq H \leq 0.4$.
- مقياس متوسط $0.4 \leq H \leq 0.5$.
- – مقياس قوي $H \geq 0.5$.

الانحدارالالابارامتري (Non-parametric regression):

أشار دوجلس (Douglas,1997) إلى أن تقدير دالة استجابة الفقرة ضمن تحليل الانحدار اللابارامتري يتم دون أي قيودٍ لشكلها حيث يوجد على الأقل طريقتان لتقديرها وهما:

- Kernel Smoothing (KS).
- Isotonic Regression Estimation (IRE).

ويضيف أن الأفضلية التطبيقية تعود إلى طريقة (KS) وذلك لوجود برنامج حاسوبي مجاني (TESTGRAF) الذي يمكن من خلاله تقدير دالة احتمالية الإجابة الصحيحة $\boldsymbol{Pi(\theta)}$.

دقة القياس في نماذج استجابة الفقرة اللابارامترية:

يشير رامسي (Ramsay, 2000) إلى أن دقة القياس في النماذج اللابارامترية تقدر من خلال دالة معلومات الفقرة ودالة معلومات باستخدام برنامج (TESTGRAF) حيث يتم استخراج

الرسوم البيانية لمتوسط معلومات الفقرة الثنائية من خلال المعادلة:

$$I_i(\theta) = \left[\frac{dp_i(\theta)}{d\theta}\right]^2 / \left[p_i(\theta)\left(1 - p_i(\theta)\right)\right] \quad \ldots(8)$$

حيث إن $p_i(\theta)I_i(\theta)$: قيمة دالة خصائص الفقرة.

إذ يعمل هذا البرنامج على رسم متوسط لدالة معلومات الفقرة (Mean item information)، وترتبط دالة معلومات الفقرة مع مبدأ تمييز الفقرة وهو الذي يجمع بين نماذج نظرية استجابة الفقرة البارامترية بنظيرتها اللابارامترية من خلال المعاملات (H_i, a_i)، حيث إن الفقرة تقدم معلومات عند قيمة معينة من القدرة في حال كان الميل عالياً (Santor& Ramsay, 1998).

وتستخرج دالة معلومات الاختبار بالاعتماد على دوال معلومات الفقرات للمقياس كاملاً، ثم يتم استخراج دالة متوسط معلومات الاختبار ككل (Average test information function) $(I_i(\theta)/n)$ ثم تقدير الخطأ المعياري لتقدير القدرة الذي يرتبط عكسياً مع كمية المعلومات (Ramsay, 2000).

مشكلة الدراسة وأهميتها:

تنبثق مشكلة الدراسة من الحاجة للكشف عن النموذج الأفضل في مطابقة البيانات المستخرجة من الاختبار التحصيلي، حيث أن هناك بعض الاعتراضات حول إمكانية استخدام الإجراءات البارامترية للبيانات ذات المستوى الرتبي، فأساغها البعض بادعاء أن القياس مستقل في حال تحققت افتراضات النظرية للبيانات ذات المستوى الرتبي (Crocker & Algina, 1986)، إلا أن البعض الآخر الذين يدعون (القياس–الموجه) يعتقدون أن المقاييس ذات المستوى الرتبي نادراً ما تتحلى بخصائص مقاييس الفترة، لذا تكون الإجراءات اللابارامترية أكثر ملاءمة لمثل هذه الحالات عن نظيرتها البارامترية (Crocker & Algina, 1986)، وحيث إن غالبية البيانات في العلوم التربوية لها خصائص رتبية ولا يمكن افتراض أن تكون ذات مستوى فترة، كما أن النتائج التي يخرج بها الباحثون مبنية ومرتكزة على الرتبية في تطبيقاتها، لذلك لابد أن تكون المعالجة الإحصائية رتبية أيضا (Cliff & Donoghue, 1992).

وتتحدد مشكلة الدراسة من خلال الأسئلة التالية:

1. ما عدد الفقرات المطابقة للنموذج ثنائي المعلمة البارامتري للاختبار التحصيلي عند مستوى الدلالة $(\propto = 0.05)$ ؟
2. ما عدد الفقرات المطابقة لنموذج موكن للتجانس الاطرادي اللابارامتري للاختبار التحصيلي عند مستوى الدلالة $(\propto = 0.05)$ ؟
3. أي النموذجين البارامتري واللابارامتري يطابق عدداً أكبر من الفقرات؟
4. أي النموذجين يعطي كميةً أكبر من المعلومات للاختبار وذلك لمستويات القدرة المختلفة عند مستوى الدلالة $(\propto = 0.05)$؟

أهمية الدراسة:

الأهمية النظرية:

تتضح أهمية هذه الدراسة في محاولتها الكشف عن حسن المطابقة باستخدام نموذجين من نماذج نظرية استجابة الفقرة وهما، النموذج البارامتري ثنائي المعلمة ونموذج موكن للتجانس الاطرادي اللابارامتري؛ الأمر الذي قد يساعد في فهم أكبر لمسببات عدم المطابقة بالنظر إلى طبيعة الاختبار ومستوى القياس الذي تصنف تبعه البيانات.

الأهمية العملية:

تقوم هذه الدراسة بتوفير مقياس تحصيلي في الرياضيات يتمتع بالمواصفات التي توفرها نظرية استجابة الفقرة لضمان جودة الاختبار،إضافة إلى تسليط الضوء على النماذج اللابارامترية وفوائدها في وضع وتطوير الاختبار بحيث تزداد فرصة مطابقة الفقرة للنموذج الأمر الذي يقلل من هدر الفقرات ورفضها نتيجة عدم مطابقتها للنماذج البارامترية.

أهداف الدراسة:

هدفت هذه الدراسة إلى تحديد عدد الفقرات المطابقة لكل من النموذج البارامتري الثنائي والنموذج اللابارامتري للتجانس الاطرادي لمعرفة النموذج الذي امتلك قدرةً أكبر على المطابقة، ثم تحديد كمية المعلومات التي امتلكها الاختبار الذي تمخض عن كل نموذج.

محددات الدراسة:

محددات مكانية: تحددت عينة الدراسة ضمن المدارس التابعة لمديرية التربية والتعليم في محافظة إربد.

محددات زمنية: اقتصرت هذه الدراسة على طالبات الصف العاشر الأساسي فقط للعام الدراسي 2011/2010م.

محددات بشرية: اقتصرت عينة الدراسة على طالبات الصف العاشر الأساسي فقط.

محددات موضوعية: اقتصرت متغيرات الدراسة على مؤشرات المطابقة للنموذج الثنائي المعالم ونموذج موكن للتجانس الاطرادي.

محددات التعميم: يعتمد تعميم نتائج هذه الدراسة على درجة تمثيل الاختبار المستخدم في الدراسة وهو اختبار الرياضيات.

مصطلحات الدراسة:

- مطابقة الفقرة: وهي ملاءمة الفقرة للنموذج المختار من خلال محكات للحكم خاصة بكل نموذج فيما يخص شكل دالة استجابة تلك الفقرة ومعالمها.
- النموذج ثنائي المعلمة: وهو أحد نماذج استجابة الفقرة البارامترية، والذي يعبر عن شكل دالة استجابة الفقرة من خلال المعادلة:

$$Pi(\theta) = \frac{\exp\left[Da_i\left(\theta - b_i\right)\right]}{1 + \exp\left[Da_i\left(\theta - b_i\right)\right]}$$

- نموذج التجانس الاطرادي: أحد النماذج اللابارامترية ويفترض أن يكون شكل دالة استجابة الفقرة غير متناقصٍ مع الانتقال للمفحوصين ذوي القدرة المرتفعة.

الدراسات السابقة:

أجرى كل من سيجتسما وايمونز وبوميستر ونكليشك ورودرا (Sijtsma, Emons, Boumeester, Nyklıcek & Rodra, 2007) دراسة تحليلية لنظرية استجابة الفقرة اللابارامترية لأبعاد مقياس جودة الحياة وتطبيقاته، حيث تم تطبيق المقياس من قبل مجموعة من طلبة علم النفس في جامعة (تيلبرغ) على مجموعتين متساويتين من الرجال والنساء تزيد أعمارهم على (30) عاماً من الهولنديين، وبعد تحليل النتائج باستخدام النموذجين البارامتري واللابارامتري أنتج النموذج اللابارامتري تدريجات أحادية البعد لكل بعد من أبعاد المحتوى، كما أظهرت بعض الفقرات معاملات تدريجٍ محدودةٍ تبعاً لباقي الفقرات ضمن نفس التدريج، كما رفضت نتائج تحليل النموذج البارامتري بعض الفقرات، إلا أن النتائج النهائية أظهرت أن النموذج اللابارامتري للتجانس الاطرادي لموكن كان الأنسب والأكثر مطابقة للبيانات ضمن سياق جودة ورفاهية الحياة.

أما ويلس وبولت (Wells & Bolt, 2008) فقد قاما باختبار طريقة لابارامترية سميت بخطأ الجذر التكاملي التربيعي (Root Integral Squared Error (RISE)) لتقييم حسن المطابقة لنظرية استجابة الفقرة، حيث قارنت الدراسة معدلات الخطأمن النوع الأول وقوة احصاءات المطابقة ($S-X^2$, G^2) (RISE باستخدام بيانات مولدة من نوع مونتي كارلو تناولت أعداداً مختلفةً من المفحوصين (250، 500، 1000)، وأعداداً مختلفةً من الفقرات (10، 20، 40، 80)، ونسباً مختلفةً من عدم التطابق، وقد أظهرت نتائج التحليل تحكم الطريقة اللابارامترية (RISE) بمعدلات الخطأ من الدرجة الأولى في جميع الظروف بشكل أفضل من نظيريها كما أنها أظهرت نتائجاً أفضل في وصف طبيعة عدم التطابق ، وموقع حدوثه ، بالإضافة لإمكانية تحديد سبب عدم التطابق.

في حين أجرى لي وآخرون (Lee, Wollak, & Douglas, 2009) دراسة هدفت الى تقييم مطابقة النموذج ثنائي المعالم من خلال تقدير منحنى استجابة الفقرة البارامتري مع إجراءات تقدير دالة خصائص الفقرة اللابارامتري، حيث تم توليد بيانات تتطابق وأغراض الدراسة باستخدام عينات بأحجام مختلفة، واختبارات ذات أطوال مختلفة، ونسب مختلفة من عدم التطابق، ومن ثم تقدير دوال استجابة الفقرات البارامترية واللابارامترية التي تم تقدير اللابارامترية منها باستخدام ثلاث طرق هي: (Isotonic regression: IR, Smoothed isotonic regression: SIR, and Kernel smoothing: KR) حيث تم حساب الفرق بين الدالتين البارامترية واللابارامترية لاستجابة الفقرة المقدرة باستخدام مؤشر ((RISD) (Root Integrated Squared Difference) كمؤشر لتقدير المطابقة من خلال حساب الفرق بين الدالتين وإيجاد دلالة هذا الفرق، وتوصلت الدراسة الى نتائج تفيد بأن الطرق اللابارامترية الثلاث أنتجت دوال استجابة فقرة مشابهة لتلك المستخرجة باستخدام النموذج البارامتري ثنائي المعلمة للفقرات التي تم توليدها لتطابق النموذج البارامتري، أما الفقرات التي لم تتشابه فيها دوال استجابة الفقرات، التي لم تتطابق وافتراضات النموذج اللابارامتري، فقد كانت إجراءات تقدير (KR) الأقوى في الكشف عن عدم التطابق بين النموذجين ثنائي المعلمة والنموذج اللابارامتري.

وقامت ديهاوس (Dyehouse, 2009) بمقارنة مطابقة النموذج للبيانات باستخدام نظرية استجابة الفقرة بمنحييها البارامتري واللابارامتري، إذ قامت بمقارنة نتائج التحليل لنموذج الطي المتدرج العام البارامتري مقابل نتائج التحليل لنموذج موكن اللابارامتري، حيث قامت الباحثة باستخدام معاملات التدريج، وعدد انتهاكات التجانسات الاطرادية، وعددانتهاكات الاطرادية المضاعفة للحكم على مطابقة البيانات لنموذج موكن، بالمقابل تم استخدام احصائيات (χ^2) للحكم على مطابقة البيانات للنموذج البارامتري، وقد تكونت عينة الدراسة من مجموعة من الطلبة ذوي الاحتياجات الخاصة من فئات اضطرابات التوحد، والإعاقة العقلية (المتوسطة– الشديدة)، حيث تم جمع البيانات باستخدام تقديرات المعلمين على مقياس انديانا للكفاءات التعليمية والمكون من (20) فقرة، وأظهرت النتائج تطابقاً أفضل لنموذج (GGUM) لمعظم الفقرات

ومجموعات الاضطرابات المختلفة، بينما لم يتحقق الافتراض الأكثر صعوبة وهو الاطرادية المضاعفة لأي من الفقرات، في حين تحقق افتراض التجانس الاطرادي الأساسي لنموذج موكن لمعظم الفقرات ومجموعات الاضطراب مما يجعل من نموذج موكن نموذجاً مفيداً لأغراض الاختيار.

وقدم ليانغ (Liang, 2010) منحىً لابارامتري جديد لمطابقة النموذج للبيانات سمي بخطأ الجذر التكاملي التربيعي ((Root Integrated Squared Error (RISE)، حيث هدفت دراسته إلى اختبار هذه الطريقة، بالإضافة إلى اختبار نواتج استخدامها، إذ قام الباحث بإجراء دراسته على بيانات افتراضية مولدة وأخرى تجريبية مع إدخال بعض المتغيرات الأخرى للدراسة كحجم العينة، وطول الاختبار، والنموذج المختار، وبينت نتائج الدراسة أن (RISE) تفوقت على إحصائيي (G^2) و($S - X^2$) بالتحكم بالخطأ من النوع الأول كما أثبت قوتها تحت جميع الظروف، فيما كشف كل من (RISE) و($S - X^2$) في الدراسة التجريبية عن أعداد منطقية من الفقرات غير المطابقة، إلا أن احصائي (G^2) لم يطابق معظم الفقرات تقريباً عندما كان حجم العينة كبيراً، وخلصت الدراسة أيضا إلى أنه وبمقارنة إحصائيات (RISE) و($S - X^2$) و(G^2) أظهر الأول صورةً أكثر وضوحاً عن موقع وحجم عدم التطابق للفقرات غير المطابقة.

وقام مساعده (2013) بدراسة أثر اختلاف شكل الفقرة (اختيار من متعدد، الصح والخطأ، التكميل) في مطابقة الفقرات لنموذج موكن للتجانس الاطرادي والنموذج الثنائي باستخدام اختبار تحصيلي في مبحث الرياضيات مكون من (34) فقرة على عينة مكونة من (1841) طالباً وطالبة، وقد أظهرت النتائج عدم وجود فرق في نسبة الفقرات المطابقة بين النموذجين لأشكال الفقرات الثلاث الا أن النموذج الثنائي كان الأفضل في تقديم المعلومات حول السمة المقيسة.

كما قام دلوع (2013) بدراسة مدى التوافق بين النموذج ثلاثي المعلمة ونموذج موكن للتجانس الاطرادي لشكلين من أشكال الفقرة (الاختيار من متعدد و الصح والخطأ) باستخدام اختبار تحصيلي في مبحث الرياضيات مكون من (37) فقرة على عينة مكونة من (1185) طالباً وطالبة، وقد بينت نتائج الدراسة وجود توافق بين النموذجين الثلاثي والتجانس الاطرادي في حال كانت الفقرات على صورة اختيار من متعدد حيث بلغت نسبة التوافق (86,49%) كما بلغت نسبة التوافق في فقرات الصح والخطأ (94,59%)، حيث أوضحت الدراسة أن نموذج موكن كان أكثر تشدداً في مطابقة الفقرات بينما كان النموذج الثلاثي أكثر تشدداً في مطابقة الأفراد فنموذج موكن كان أكثر اهتماماً بالفقرات بينما النموذج الثلاثي كان أكثر اهتماماً بالأفراد، كما بينت النموذج الثلاثي قدم قدراً أكبر من المعلومات لكلا الشكلين من أشكال الفقرة.

التعليق على الدراسات السابقة:

باستعراض الدراسات السابقة التي تم ذكرها، تبين أنه لم يتم التوصل إلى إجماع من قبل الباحثين في النتائج التي توصلت إليها دراساتهم فيما يتعلق بتحديد النموذج الأمثل لمطابقة البيانات باختلاف نوعها الرتبية وغيرها، ففي حين توصلت بعض الدراسات الى نتائج تفيد بأن النماذج اللابارامترية أكثر قدرة على المطابقة من النماذج البارامترية (Sijtsma,et al., 2007)، بينما أكدت بعض منها أفضلية مطابقتها للبيانات الرتبية تحديداً باعتبار البيانات السلوكية التي تقيس سمات شخصية هي بيانات رتبية (Wells & Bolt, 2008) كما أشارت دراسات أخرى إلى أن النماذج البارامترية أكثر قدرة من النماذج اللابارامترية في التعرف على المطابقة ومنها دراسة دايهاوس (Dyehouse, 2009) التي توصلت إلى نتائج تشير إلى أن نموذج الطي المتدرج العام البارامتري (GGUM) أظهر نتائج مطابقة أفضل لمعظم الفقرات، ولم يخلُ البحث من دراسات تشير إلى عدم وجود فرق في القدرة على تحديد المطابقة ما بين النماذج البارامترية والنماذج اللابارامترية مثل دراسة لي وزملاؤه (Lee, Wollak & Douglas, 2009) و دراسة مساعده (2013) التي أشارت إلى تشابه النماذج اللابارامترية مع النموذج البارامتري ثنائي المعلمة في إنتاج دوال استجابة الفقرات اضافةً الى دراسة دلوع (2013) التي أشارت الى وجود توافق كبير في مطابقة الفقرات بين النموذج الثلاثي البارامتري ونموذج موكن، وأشارت دراسة ويلس وبولت (Wells & Bolt, 2008) إلى أن الطرق اللابارامترية تتحكم بالخطأ من النوع الأول بشكل أفضل من النتائج البارامترية، وهي أفضل في وصف عدم التطابق، وموقع حدوثه، وتحديد سبب عدم التطابق.

ومن هذا التضارب في نتائج الدراسات جاءت هذه الدراسة كمحاولة من الباحثة من إلقاء المزيد من الضوء على حقيقة الأمر فيما يتعلق بالنماذج الأكثر قدرة على مطابقة الفقرات للبيانات خاصة الرتبية منها نظراً لشيوعها وانتشارها في العلوم الإنسانية والسلوكية.

الطريقة والاجراءات

مجتمع الدراسة:

تكون مجتمع الدراسة من جميع طالبات الصف العاشر

الأساسي للعام الدراسي (2011/2010م) اللواتي يدرسن في المدارس الحكومية التابعة لمديرية التربية والتعليم لمنطقة إربد الأولى والبالغ عددهن (4053) طالبة توزعن على (46) مدرسة حسب إحصائية وزارة التربية والتعليم لعام (2011/2010م).

عينة الدراسة:

قامت الباحثة باختيار عينة عشوائية بسيطة وتم اختيار خمس مدارس تضمنت (834) طالبة.

أداة الدراسة:

تكونت أداة الدراسة من اختبار لقياس القدرة الرياضية لطالبات الصف العاشر الأساسي بمحتوى وحدة دراسية واحدة من مادة الرياضيات، وهي الوحدة الدراسية الخامسة في الكتاب المدرسي المقرر، وقد تكون الاختبار من (30) فقرة – وهو عدد مناسب للاختبارات التحصيلية وفقاً لما أشار إليه نانالي (Nunnally, 1978) ويلي كل فقرة من فقرات الاختبار أربعة بدائل أحدها فقط إجابة صحيحة، وقد استعانت الباحثة ببعض الفقرات الواردة في دليل معايير ومؤشرات الأداء لمبحث الرياضيات للصف العاشر والمعتمد من قبل مديرية الامتحانات والاختبارات في وزارة التربية والتعليم الأردنية، وقد تم الالتزام بالشروط الفنية لكتابة الفقرة التي تمت الإشارة لها في عودة (2010) لتطوير الاختبارات التحصيلية.

الدراسة الاستطلاعية:

تم تطبيق الاختبار على عينة استطلاعية مكونة من (52) طالبة من طالبات الصف العاشر من خارج عينة الدراسة، وقد ساهمت نتائج هذا التطبيق القبلي في تقديم معلومات هامة حول وضوح الفقرات والبدائل الأمر الذي ساهم في تطوير الاختبار، وقد تم التعديل في صياغة بعض الفقرات إضافة إلى تقدير الزمن اللازم للاختبار وهو ساعة وربع الساعة حيث اكملت ما يزيد على (80%) من الطالبات الاختبار خلال هذا الزمن كما تمت الاستفادة من التطبيق على العينة الاستطلاعية في تقدير مؤشرات صعوبة وتمييز الفقرات، حيث تراوحت قيم صعوبة الفقرات بين (0,21– 0,79) وقيم تمييزها بين (0,21– 0,86) والتي تعتبر فقرات مناسبة تبعاً لما أشار إليه عودة (2010) كما تم تقدير معامل الارتباط المصحح لكل فقرة من فقرات الاختبار مع الدرجة العامة للمقياس الكلي والذي يعد مؤشراً آخراً لقدرة الفقرة التمييزية (Magnusson, 1967)، وبناءً عليه تم حذف فقرتين كان ارتباطهما مع الدرجة الكلية غير دال إحصائياً تبعاً لما أشار له ألن و ين (Allen & Yen, 1979) وكروكر وألجينا (Crocker & Algina, 1986). وبذلك أصبح الاختبار بصورته النهائية مكوناً من (28) فقرة جاهزة للتطبيق على عينة الدراسة.

صدق الاختبار:

أولا: صدق المحتوى:

للتحقق من صدق المحتوى للاختبار المعد تم عرضه بصورته الأولية على (12) محكما من أعضاء هيئة التدريس لأساليب تدريس الرياضيات والرياضيات في الجامعات الأردنية إضافةً الى عددٍ من المشرفين التربويين ومعلمي ومعلمات الرياضيات في وزارة التربية حيث طلب إليهم الحكم على كل فقرة من فقرات الاختبار تبعاً لدرجة تمثيل الفقرة للهدف، ووضوح الصياغة اللغوية كما وضعت الباحثة معيار قبلي للحكم على الفقرة وهو حذف وتغيير الفقرة التي أوصى بحذفها أو تغييرها 30% من المحكمين أو أكثر.

ثانياً: الصدق المحكي التلازمي:

كما تم تقدير معامل الصدق المحكي التلازمي للاختبار من خلال حساب معامل ارتباط بيرسون بين درجات الطالبات على الاختبار المعد من قبل الباحثة ودرجاتهن في مبحث مادة الرياضيات التي تم الحصول عليها من مدرسة المادة في المدرسة التي أخذت منها العينة الاستطلاعية، وكانت قيمة معامل ارتباط بيرسون (0,86) وهي قيمة مقبولة لغايات الدراسة.

ثبات الاختبار:

أولاً: ثبات الاتساق الداخلي:

تم تقدير ثبات الاختبار بطريقة التجزئة النصفية باستخدام معامل ثبات جتمان لنصفي الاختبار (الفقرات من 1–15، والفقرات من 16–30) وقد كانت قيمة معامل الثبات (0,84)، كما تم حساب ثبات الاختبار من خلال معادلة كرونباخ – الفا وبلغت قيمته (0,89) وبناءً على ذلك يمكن اعتبار الاختبار مناسباً لجمع البيانات اللازمة للدراسة، حيث أن دلالات الثبات والصدق تفي لتلك الأغراض.

تصحيح الاختبار:

تم تصحيح الاختبار ضمن تدريج ثنائي حيث تعطى الطالبة عند إجابتها إجابة صحيحة الدرجة (1)، أما عند إجابتها إجابة خاطئة فتعطى الدرجة (0)، وبذلك تكون أدنى درجة يمكن أن تحصل عليها المفحوصة على الاختبار الكلي

هي الدرجة (0) بينما تكون أعلى درجة يمكن أن تحصل عليها هي الدرجة (28).

المعالجات الإحصائية:

تم استخدام المعالجات الإحصائية التالية:

المعالجات الإحصائية الخاصة بفحص مطابقة البيانات للافتراضات الأساسية لنظرية استجابة الفقرة (أحادية البعد والاستقلال الموضعي).

المعالجات الإحصائية لمطابقة البيانات للنموذج البارامتري ثنائي المعلمة.

المعالجات الإحصائية لمطابقة البيانات للنموذج اللابارامتري للتجانس الاطرادي (نموذج موكن).

المعالجات الإحصائية الخاصة بفحص البيانات للافتراضات الأساسية لنظرية استجابة الفقرة (أحادية البعد والاستقلال الموضعي).

نتائج الدراسة:

أولاً: التحقق من الافتراضات الأساسية لنماذج نظرية استجابة الفقرة :

تم التحقق من الافتراضات الأساسية التي تسوّغ استخدام نماذج نظرية استجابة الفقرة المختلفة وهما أحادية البعد والاستقلال الموضعي، وقد تم فحص أحادية البعد للبيانات الناتجة من تطبيق الاختبار باستخدام برنامج (SPSS) حيث أجري التحليل بطريقة المكونات الأساسية (Principal Component Analysis) وملاحظة نسب قيمة الجذر الكامن (Eigen Value) للعامل الأول إلى الجذر الكامن للعامل الثاني إضافةً إلى مقدار التباين الذي فسره العامل الأول والذي بلغ (30,04) من التباين الكلي وذلك كما هو مبين في الجدول (1).

يلاحظ من الجدول (1) التأكيد على أحادية البعد من خلال:

- نسبة ما يفسره العامل الأول من التباين الكلي أكبر من (20%).
- نسبة ما يفسره العامل الأول الى العامل الثاني أكبر من (2).

كما تم تمثيل الجذور الكامنة بيانياً من خلال ما يعرف ب (Scree Plot) حيث يمثل المحور الأفقي العوامل ويمثل المحور العمودي مقدار الجذر الكامن والذي أكد أحادية البعد من خلال تميز العامل الأول عن باقي العوامل كما تم اعتبار الافتراض الثاني وهو الاستقلال الموضعي متحققاً نتيجة تمتع الاختبار بأحادية البعد تبعاً لهامبليتون وسوامينيثان.

ثانياً: الخصائص السيكومترية للاختبار تبعاً للنظرية الكلاسيكية في الاختبار:

للتعرف على الخصائص الكلاسيكية لاختبار الرياضيات كأداة لهذه الدراسة، تم استخراج عدد الاستجابات الصحيحة على كل فقرة وصعوبة الفقرة إضافةً إلى معاملات ارتباط الفقرة مع الاختبار، حيث إن ارتباط الفقرة بالسمة المقيسة يعد كمؤشرٍ على تمييز الفقرة وذلك كما هو مبين في الجدول (2) والمستخرج باستخدام برنامج (bilog-mg).

يلاحظ من الجدول (2) أن الفقرات (2، 4، 5) تعتبر فقرات سهلة، بينما تعد الفقرات (15، 19) فقرات صعبة كما يلاحظ ارتباط جميع فقرات الاختبار مع الاختبار الكلي (السمة الكامنة) غالباً بشكلٍ جيدٍ وهذا ما تشير إليه قيم معاملات الارتباط في العمود الأخير.

ثالثاً: النتائج المتعلقة بالسؤال الأول:

كان السؤال الأول في هذه الدراسة "ما عدد الفقرات المطابقة للنموذج ثنائي المعلمة البارامتري للاختبار التحصيلي عند مستوى الدلالة $(\propto = 0.05)$؟"

للإجابة عن السؤال الأول تم إيجاد معالم الفقرات والأخطاء المعيارية لها وإحصائيات جودة الملائمة للتعرف على عدد فقرات الاختبار المطابقة للنموذج البارامتري ثنائي المعلمة ، وذلك كما هو موضح في الجدول (3).

الجدول (1)
العوامل العشرة الأولى وجذورها الكامنة ونسب تباينها

العامل	قيمة الجذر الكامن	نسبة التباين المفسر للعامل
1	8.41	%30.04
2	1.72	%6.15
3	1.19	%4.23
4	1.09	%3.88
5	1.03	%3.66
6	0.99	%3.55
7	0.97	%3.46
8	0.93	%3.32
9	0.89	%3.19
10	0.82	%2.94

الجدول (2)
عدد الاستجابات الصحيحة ونسبتها ومعامل ارتباط بيسيريال لفقرات الاختبار

ITEM	#RIGHT	PCT	BISERIAL
ITEM0001	595	71.5	0.897
ITEM0002	645	77.5	0.494
ITEM0003	484	58.2	0.888
ITEM0004	645	77.5	0.464
ITEM0005	648	77.9	0.485
ITEM0006	577	69.4	0.535
ITEM0007	517	62.1	0.556
ITEM0008	441	53.0	0.452
ITEM0009	439	52.8	0.531
ITEM0010	382	45.9	0.505
ITEM0011	541	65.0	0.598
ITEM0012	470	56.5	0.956
ITEM0013	469	56.4	0.555
ITEM0014	551	66.2	0.632
ITEM0015	79	9.50	0.590
ITEM0016	436	52.4	0.678
ITEM0017	454	54.6	0.706
ITEM0018	506	60.8	0.921
ITEM0019	157	18.9	0.733
ITEM0020	565	67.9	0.638
ITEM0021	437	52.5	0.753
ITEM0022	410	49.3	0.457
ITEM0023	403	48.4	0.564
ITEM0024	280	33.7	0.679
ITEM0025	383	46.0	0.586
ITEM0026	403	48.4	0.645
ITEM0027	459	55.2	0.615
ITEM0028	175	21.0	0.690

الجدول (3)

قيم صعوبة وتمييز فقرات الاختبار والأخطاء المعيارية وإحصائيات جودة الملائمة

ITEM	SLOPE S.E.	THRESHOLD S.E.	CHISQ (PROB)
ITEM0001	2.773 0.175*	-0.753 0.053*	71.0 (0.000)
ITEM0002	1.051 0.095*	-1.450 0.126*	20.7 (0.008)
ITEM0003	2.644 0.196*	-0.247 0.049*	31.6 (0.000)
ITEM0004	0.959 0.096*	-1.541 0.145*	19.6 (0.012)
ITEM0005	0.985 0.093*	-1.538 0.141*	21.7 (0.005)
ITEM0006	1.075 0.096*	-0.951 0.098*	14.3 (0.074)
ITEM0007	1.077 0.094*	-0.577 0.082*	8.6 (0.379)
ITEM0008	0.819 0.082*	-0.167 0.092*	13.3 (0.101)
ITEM0009	0.977 0.086*	-0.133 0.079*	45.2 (0.000)
ITEM0010	0.941 0.088*	0.216 0.082*	9.4 (0.306)
ITEM0011	1.205 0.102*	-0.675 0.074*	18.0 (0.021)
ITEM0012	3.461 0.270*	-0.167 0.038*	27.7 (0.000)
ITEM0013	1.079 0.092*	-0.294 0.075*	3.4 (0.911)
ITEM0014	1.282 0.100*	-0.708 0.077*	24.9 (0.002)
ITEM0015	2.132 0.264*	1.700 0.110*	4.7 (0.315)
ITEM0016	1.490 0.108*	-0.081 0.058*	11.8 (0.159)
ITEM0017	1.538 0.115*	-0.165 0.058*	5.6 (0.590)
ITEM0018	2.912 0.197*	-0.335 0.043*	13.1 (0.069)
ITEM0019	2.964 0.265*	1.102 0.047*	14.6 (0.006)
ITEM0020	1.381 0.107*	-0.759 0.074*	30.7 (0.000)

ITEM0021	1.756 0.116*	-0.074 0.053*	35.3 (0.000)
ITEM0022	0.863 0.082*	0.044 0.088*	8.9 (0.352)
ITEM0023	1.108 0.092*	0.081 0.072*	9.8 (0.280)
ITEM0024	1.654 0.119*	0.652 0.060*	22.6 (0.002)
ITEM0025	1.176 0.096*	0.188 0.069*	6.5 (0.593)
ITEM0026	1.334 0.103*	0.078 0.062*	3.2 (0.923)
ITEM0027	1.272 0.102*	-0.212 0.066*	13.5 (0.097)
ITEM0028	2.234 0.172*	1.083 0.057*	21.2 (0.001)

يشير الجدول (3) إلى وجود بعض الفقرات السهلة كما هو الحال في الفقرات (2، 4، 5) وبعض الفقرات الصعبة مثل (15، 19، 28) كما يبين الجدول قيم تمييز فقرات الاختبار، إذ تمتعت بعض الفقرات بتمييزٍ عالٍ مثل الفقرات (1، 3، 12، 15، 18، 19) كما يقدم هذا الجدول إحصاءات مطابقة الفقرة إذ تشير القيم الاحتمالية التي تقل عن (0,05) الى الفقرات غير المطابقة للنموذج ثنائي المعلمة (Haley et. al, 2009) وهي الفقرات (1، 2، 3، 4، 5، 9، 11، 12، 14، 19، 20، 21، 24، 28).

كما تم تقدير خصائص الاختبار بصورة عامة المكون من (28) فقرة، حيث تمتع بصعوبةٍ كليةٍ مقدارها (−0,203) وهي صعوبة متوسطة كما تمتع بتمييز كلي مقداره (1,576) وهي قيمة جيدة عملياً تبعاً لما تم ذكره من قبل بيكر (Baker, 2001). كما تم إعادة استخراج خصائص الاختبار بعد حذف الفقرات التي فشلت في مطابقة النموذج البارامتري ثنائي المعلمة والتي بلغت (14) فقرة حيث تبين انخفاض قيم الخطأ المعياري في تقدير إحصاءات الاختبار كما بلغ معامل صعوبة الاختبار (0,92) ومعامل تمييز الاختبار (1,432).

تعود مطابقة هذا العدد القليل نسبياً من الفقرات للنموذج بشكل أساسي إلى صعوبة وتشدد نماذج استجابة الفقرة البارامترية في افتراضاتها لمطابقة الفقرات للنموذج والتي يعد لوجستية شكل دالة استجابة الفقرة أكثرها تشدداً إضافة الى تحقيق الشروط الخاصة بمطابقة الفقرات للنموذج الثنائي المعلمة من جهة أخرى فالنموذج الثنائي يفترض عدم لجوء المفحوصين للتخمين في الإجابة عن فقرات الاختبار وقد يعد تحقق هذا الفرض أمرٌ صعب من الناحية التطبيقية الأمر الذي قد يكون أدى لمثل هذه النتيجة. كما أن قصر طول الاختبار نسبياً وذلك أن الاختبار لم يشمل سوى وحدة دراسية واحدة من كتاب الطالب، حيث تكون من (28) فقرة من نوع اختيار من متعدد وهو عدد قليل نسبياً، الأمر الذي أدى إلى قلة عدد الفقرات المطابقة للنموذج البارامتري ثنائي المعلمة، وهذا ما أشار إليه الكثير من الباحثين مثل سيجتسما ومولينار (Sijtsma & Molenaar, 2002).

كما لوحظ أن الفقرات التي طابقت النموذج الثنائي قد تمتعت بمستوى تمييزٍ جيدٍ حول (1) ومستوى صعوبة متوسط حول (0) في آنٍ واحد معاً، وأن الفقرات التي لم تطابق النموذج كانت تتمتع بمؤشراتٍ مرتفعةٍ أو متدنيةٍ لكل من مؤشرات تمييز وصعوبة الفقرة إحداها أو كلها، وهذا يدعو الى ظهور تساؤلاتٍ جديدةٍ حول ارتباط خصائص الفقرة ومعالمها مع مؤشر المطابقة الخاص بها الأمر الذي قد يساعد في فهم أعمق لمسببات عدم التطابق وإمكانية التحكم بها.

رابعاً: النتائج المتعلقة بالسؤال الثاني:

كان السؤال الثاني في هذه الدراسة "ما عدد الفقرات المطابقة لنموذج موكن للتجانس الاطرادي اللابارامتري للاختبار التحصيلي عند مستوى الدلالة $(\propto = 0.05)$؟"

للإجابة عن السؤال الثاني تم استخدام برنامج (R Package) استخراج كل من:

5. معاملات تدريج أزواج الفقرات (H_{ij}): تم حساب

معاملات تدريج أزواج فقرات الاختبار، والتي يشترط نموذج موكن أن تكون جميعاً موجبة وذلك كما هو مبين في الجدول (4).

يوضح الجدول (4) معاملات تدريج أزواج الفقرات والتي حققت شرط نموذج التجانس الاطرادي اللابارامتري، حيث امتلكت جميعها معاملات تدريج موجبة.

2. معاملات تدريج الفقرات (H_i): تم حساب معاملات تدريج الفقرات والتي يشترط نموذج موكن أن تكون موجبة وتزيد عن (0.3) لضمان جودة التدريج (Sijtsma & Molenaar, 2002). ويبين الجدول (5) معاملات تدريج فقرات الاختبار.

الجدول (4)

معاملات تدريج أزواج الفقرات التي تشكل الاختبار

	i1	i2	i3	i4	i5	i6	i7	i8	i9	i10	i11	i12	i13	i14	i15	i16	i17	i18	i19	i20	i21	i22	i23	i24	i25	i26	i27	i28	tot
i1	1	0.51	0.83	0.36	0.46	0.39	0.42	0.36	0.30	0.50	0.40	0.75	0.47	0.51	0.91	0.58	0.57	0.74	0.84	0.48	0.6	0.36	0.49	0.7	0.52	0.63	0.53	0.76	0.87
i2		1	0.58	0.21	0.25	0.20	0.29	0.13	0.21	0.32	0.32	0.54	0.21	0.27	0.61	0.31	0.18	0.49	0.55	0.43	0.36	0.16	0.34	0.27	0.24	0.3	0.32	0.39	0.53
i3			1	0.48	0.50	0.46	0.37	0.29	0.33	0.36	0.48	0.66	0.34	0.46	0.76	0.46	0.39	0.66	0.82	0.51	0.52	0.29	0.39	0.57	0.43	0.42	0.40	0.78	0.85
i4				1	0.21	0.22	0.15	0.22	0.24	0.21	0.21	0.54	0.17	0.34	0.77	0.28	0.36	0.45	0.66	0.34	0.29	0.31	0.16	0.36	0.30	0.33	0.23	0.49	0.51
i5					1	0.41	0.21	0.10	0.11	0.29	0.16	0.44	0.31	0.37	0.48	0.23	0.32	0.44	0.60	0.35	0.30	0.07	0.33	0.35	0.41	0.50	0.23	0.51	0.53
i6						1	0.26	0.19	0.20	0.29	0.23	0.42	0.21	0.28	0.63	0.35	0.40	0.43	0.63	0.24	0.33	0.27	0.18	0.45	0.34	0.28	0.24	0.55	0.56
i7							1	0.24	0.27	0.20	0.21	0.36	0.35	0.36	0.60	0.40	0.27	0.34	0.58	0.28	0.38	0.18	0.29	0.38	0.29	0.32	0.25	0.49	0.58
i8								1	0.45	0.29	0.26	0.31	0.17	0.20	0.49	0.18	0.22	0.28	0.46	0.18	0.20	0.24	0.19	0.34	0.26	0.17	0.12	0.45	0.49
i9									1	0.29	0.36	0.36	0.21	0.20	0.46	0.23	0.29	0.33	0.54	0.18	0.30	0.25	0.22	0.43	0.31	0.22	0.20	0.55	0.56
i10										1	0.30	0.35	0.23	0.27	0.46	0.28	0.33	0.35	0.48	0.25	0.26	0.26	0.27	0.27	0.23	0.24	0.29	0.37	0.54
i11											1	0.45	0.32	0.21	0.75	0.39	0.29	0.41	0.76	0.24	0.38	0.24	0.34	0.56	0.32	0.40	0.25	0.66	0.61
i12												1	0.39	0.47	0.88	0.49	0.50	0.74	0.94	0.60	0.58	0.38	0.40	0.73	0.47	0.53	0.41	0.86	0.91
i13													1	0.36	0.56	0.25	0.30	0.41	0.45	0.38	0.28	0.09	0.23	0.45	0.30	0.35	0.26	0.46	0.58
i14														1	0.63	0.38	0.43	0.43	0.59	0.30	0.38	0.29	0.34	0.43	0.31	0.43	0.35	0.51	0.64
i15															1	0.81	0.64	0.81	0.52	0.80	0.71	0.45	0.61	0.50	0.48	0.63	0.52	0.50	0.78
i16																1	0.39	0.55	0.68	0.35	0.52	0.24	0.24	0.50	0.29	0.30	0.27	0.68	0.69
i17																	1	0.53	0.73	0.36	0.43	0.29	0.37	0.50	0.39	0.34	0.28	0.67	0.71
i18																		1	0.85	0.53	0.60	0.33	0.42	0.68	0.51	0.48	0.49	0.78	0.87
i19																			1	0.64	0.76	0.57	0.68	0.55	0.47	0.6	0.64	0.49	0.86
i20																				1	0.43	0.22	0.23	0.53	0.38	0.36	0.37	0.63	0.64
i21																					1	0.29	0.34	0.50	0.36	0.33	0.38	0.75	0.74
i22																						1	0.19	0.37	0.11	0.23	0.24	0.36	0.50
i23																							1	0.40	0.28	0.36	0.30	0.52	0.59
i24																								1	0.41	0.42	0.51	0.52	0.72
i25																									1	0.29	0.38	0.48	0.62
i26																										1	0.38	0.56	0.67
i27																											1	0.63	0.64
i28																												1	0.79
tot																													1

الجدول (5)

معاملات تدريج فقرات الاختبار

Item no.	H_i
i1	0.66
i2	0.41
i3	0.61
i4	0.38
i5	0.40
i6	0.41
i7	0.41
i8	0.34
i9	0.39
i10	0.39
i11	0.44
i12	0.65
i13	0.40
i14	0.47
i15	0.69
i16	0.49
i17	0.50
i18	0.64
i19	0.73
i20	0.47
i21	0.53
i22	0.35
i23	0.42
i24	0.57
i25	0.45
i26	0.48
i27	0.44
i28	0.67

يبين الجدول (5) أن قيم معاملات تدريج الفقرات حققت شروط النموذج كاملةً، وقد تمتعت بعض الفقرات بقيم تدريجٍ عاليةٍ كما هو الحال في الفقرات (12، 15، 19، 28) على سبيل المثال.

3. معامل تدريج الاختبار الكلي (H): تم حساب معامل التدريج الكلي للاختبار والذي بلغ (0,66) وحيث إن هذه القيمة تزيد على (0,50) فإن هذا الاختبار يصنف على أنه اختبار قوي تبعاً لما ذكره سيجتسما و مولينار (Sijtsma & Molenaar, 2002).

4. اختبار الاطرادية (Monotonicity): على الرغم من أن قيم (H_i) المرتفعة تدل على فرصة عالية لتحقيق افتراضات نموذج موكن، إلا أن اختبار الاطرادية يعد دليلاً مباشراً على تحقق شرط الاطرادية، ويحدد هذا الاختبار الفقرات التي طابقت نموذج موكن للتجانس الاطرادي إذ يتم اعتبار الفقرة التي تمتلك عدد من الانتهاكات الدالة إحصائياً (zsig#) فقرةً غير مطابقةٍ لنموذج موكن (Ringdal et al., 1999). ويبين الجدول (6) عدد انتهاكات الاطرادية (vi#) بمجموعها الدالة وغير الدالة.

الجدول(6)

معاملات تدريج فقرات الاختبار وعدد انتهاكات الاطرادية ودلالاتها تبعا لنموذج التجانس الاطرادي لموكن.

Item	ItemH	#vi	#zsig	Item		ItemH	#vi	#zsig
i1	0.66	3	0	i15		0.69	0	0
i2	0.41	6	2	i16		0.49	1	1
i3	0.61	1	1	i17		0.50	0	0
i4	0.38	2	0	i18		0.64	0	0
i5	0.40	1	0	i19		0.73	0	0
i6	0.41	2	1	i20		0.47	1	0
i7	0.41	0	0	i21		0.53	3	0
i8	0.34	0	0	i22		0.35	1	0
i9	0.39	4	2	i23		0.42	1	0
i10	0.39	1	0	i24		0.57	1	0
i11	0.44	2	0	i25		0.45	0	0
i12	0.65	0	0	i26		0.48	0	0
i13	0.40	0	0	i27		0.44	0	0
i14	0.47	2	0	i28		0.67	0	0

يتضح من الجدول(6) وجود خمس فقراتٍ غير مطابقةٍ لنموذج موكن للتجانس الاطرادي وهي الفقرات (2، 3، 6، 9، 16) إذ تمتلك قيم انتهاك دالة إحصائياً.

يتبين مما سبق تحقيق جميع فقرات الاختبار المكونة من (28) فقرة لشروط نموذج فيما يخص معاملات التدريج (H) بأنواعها الثلاثة، الأمر الذي يجيز تسمية التدريج الناتج بتدريج موكن، فقد كانت جميع معاملات تدريج أزواج الفقرات موجبة، وجميع معاملات تدريج الفقرات موجبة، إضافةً إلى أن معامل التدريج الكلي للاختبار بلغ قيمة تؤشر على قوة الاختبار، مما يشير الى القبول الأولي لجميع فقرات الاختبار، كما تم اختبار الاطرادية للفقرات من خلال ما يعرف بعدد الانتهاكات الدالة احصائياً للاطرادية، وقد أظهرت النتائج وجود عدد من الانتهاكات الدالة إحصائياً عند مستوى الدلالة ($\propto= 0.05$) لخمس فقرات، بينما طابقت (23) فقرة نموذج التجانس الاطرادي لموكن.

اتفقت هذه النتائج مع النتائج التي خرج بها بعض الباحثين كما في دراسة ميجر وآخرين (Meijer, Sijtsma & Smid, 1990) إذ اتفقت جميع فقرات اختبار الذكاء اللفظي الهولندي مع نموذج موكن للتجانس الاطرادي، كذلك الدراسة التي أجراها زيكار (Zickar, 1997) حيث طابقت معظم الفقرات مقياس هوغان للشخصية.

وتعزى مثل هذه النتيجة للسهولة النسبية لافتراضات نموذج موكن اللابارامتري، إذ لا يفترض هذا النموذج شكلاً معيناً لدالة استجابة الفقرة باستثناء أن تكون دالة استجابة الفقرة غير متناقصة عبر مدى القدرة المقدر بالدرجة الكلية للفرد على الاختبار (Sijsma & Molenaar, 2002). إضافةً الى أنه يمكن عزو مثل هذه النتيجة لقيم معاملات التدريج التي يتبين في هذه الدراسة أنها تحققت بمستوياتٍ عامةٍ أعلى من المحكات المقبولة فقد تمتع معاملات تدريج أزواج الفقرات بقيم موجبة تراوحت بين (0,07 −0,94)، كما تمتعت قيم معاملات تدريج الفقرات بقيم تزيد عن (0,3) حيث تراوحت بين (0,34− 0,73)، إضافةً الى معامل التدريج الكلي الذي بلغ (0,66) وهي تعد قيمةً مرتفعةً تشير الى قوة المقياس، ولعل مثل هذه النتيجة تدعو الى دراسة إمكانية التنبؤ بتطابق الفقرات تبعاً لهذا النموذج من خلال مدى قيم معاملات التدريج المختلفة.

خامساً: النتائج المتعلقة بالسؤال الثالث:

كان السؤال الثالث في هذه الدراسة "أي النموذجين البارامتري واللابارامتري يطابق عدداً أكبر من الفقرات؟"

للإجابة عن السؤال الثاني تمت مقارنة نتائج مطابقة فقرات الاختبار للنموذج ثنائي المعلمة البارامتري والتي مثلت الاختبار بصورته البارامترية، ونتائج مطابقة فقرات الاختبار لنموذج التجانس الاطرادي اللابارامتري والتي مثلت الاختبار بصورته اللابارامترية من خلال مقارنة عدد الفقرات التي طابقت كل نموذج إضافةً إلى استخراج نسبة الفقرات المطابقة للاختبار الأصلي، إذ تعود الأفضلية للنموذج الذي طابق عدداً أكبر من الفقرات، وذلك ما يبينه الجدول (7).

الجدول (7)

فقرات الاختبار المطابقة لكلا النموذجين البارامتريواللابارامتري

رقم الفقرة	مطابقة للنموذج البارامتري	مطابقة للنموذج اللابارامتري
1		√
2		
3		
4		√
5		√
6	√	
7	√	√
8	√	√
9		
10	√	√
11		√
12		√
13	√	√
14		√
15	√	√
16	√	
17	√	√
18	√	√
19		√
20		√
21		√
22	√	√
23	√	√
24		√
25	√	√
26	√	√
27	√	√
28		√

يلاحظ من الجدول (7) أن النموذج البارامتري ثنائي المعلمة طابق (14) فقرة من فقرات الاختبار الكلي المكون من (28) فقرة أي ما نسبته (50%) من الفقرات، بينما طابق نموذج موكن للتجانس الاطرادي (23) فقرة من أصل (28) فقرة أي ما نسبته (82%) من الفقرات مع وجود بعض الفقرات التي لم تطابق أي من النموذجين مثل الفقرات (2، 3، 9) وفقاً لما سبق فإن نموذج موكن كان الأفضل في مطابقة فقرات الاختبار.

توافقت هذه النتيجة مع النتيجة التي توصل إليها ميجر وآخرون (Meijer, Sijtsma & Smid, 1990) في دراستهما، إذ طابقت جميع فقرات اختبار الذكاء اللفظي الهولندي نموذج التجانس الاطرادي لموكن بينما طابق عدد قليل من الفقرات نموذج راش، كما توافقت نتائج هذه الدراسة مع النتائج التي توصل إليها سيجتسما وزملاؤه (Sijtsma et al., 2007) والتي أظهرت تفوق نموذج التجانس الاطرادي اللابارامتري لموكن على نموذج الاستجابة المتدرج البارامتري في مطابقة بيانات تطبيق مقياس جودة ورفاهية الحياة لمنظمة الصحة العالمية، بينما تعارضت نتائج هذه الدراسة مع النتيجة التي خلصت إليها

دراسة ديهاوس (Dyehouse, 2009) والتي أظهرت تطابقاً أفضل لنموذج الطي المتدرج العام البارامتري على النموذج اللابارامتري للتجانس الاطرادي، وقد يعود الاختلاف في نتيجة الدراسة الحالية عن نتيجة دراسة ديهاوس إلى طبيعة العينة التي تم اختيارها من قبل ديهاوس، والتي تكونت من مجموعة من الطلبة ذوي الاحتياجات الخاصة ممن يعانون من أشكال مختلفة من الاضطرابات الأمر الذي قد يدعو الباحثين لدراسة تأثير موقع عينة الدراسة (وسط التوزيع و طرفي التوزيع) على مؤشر مطابقة الفقرة، كما قد تعزى مثل هذه النتيجة إلى استخدام ديهاوس لمقاييس متعددة الأبعاد وليست وحيدة البعد كما في الدراسة الحالية وباقي الدراسات التي توافقت في نتائجها مع نتيجة الدراسة الحالية.

تعزى نتيجة هذه الدراسة في مطابقة النموذج اللابارامتري لعدد أكبر من الفقرات إلى السهولة النسبية لمحددات مطابقة فقرات نموذج موكن إذ لا يفترض شكلاً محدداً لدالة استجابة الفقرة ما عدا أن تكون عدم متناقصة مقارنة مع النموذج الثنائي الذي يفترض أن يكون شكل دالة استجابة الفقرة لوجستياً إضافةً الى أنه يشترط أن تحقق البيانات الشروط الخاصة بالنموذج الثنائي بشكل خاص وهو عدم وجود عامل التخمين. كما قد تشير هذه النتيجة إلى دعم المقولة التي تؤيد وقوع بيانات الاختبارات التحصيلية ضمن مقياس رتبي وأنها لا ترقى لمستوى القياس الفئوي.

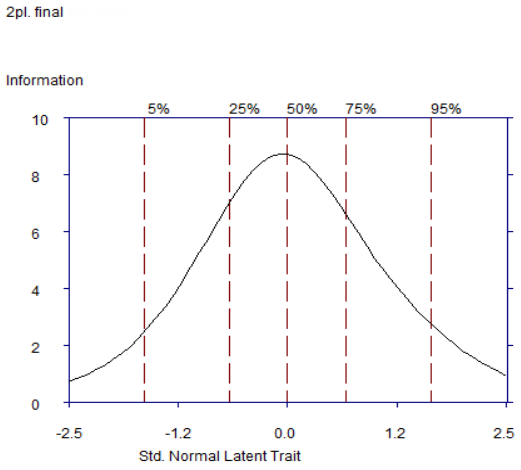
سادساً: النتائج المتعلقة بالسؤال الرابع:

كان السؤال الرابع في هذه الدراسة " أي النموذجين يعطي كمية أكبر من المعلومات للاختبار وذلك لمستويات القدرة المختلفة عند مستوى الدلالة $(\alpha = 0.05)$؟"

للاجابة عن هذا السؤال تم تمثيل دوال استجابة فقرات الاختبار بيانياً لفقرات الاختبار المكون من (28) فقرة والتي تم استخراجها باستخدام برنامج (TESTGRAF).

قدمت دوال استجابة فقرات الاختبار معلومات حول فقرات الاختبار، إذ قدمت انطباعاً أولياً ساعد في الحكم المبدئي على فقرات الاختبار التي قد تتطلب بعضها نماذج أكثر تعقيداً مثل النموذج ثلاثي أو رباعي المعلمة كما هو الحال في الفقرات (2)، (4)، (9)، (10)، (11) والتي قد تكون أقرب للنموذج الثلاثي المعلمة، والفقرات (5)، (19) والتي قد تكون أقرب للنموذج الرباعي المعلمة كما ساعد التمثيل البياني لدوال استجابة الفقرة تحديد المنحنيات التي امتلكت معاملات تمييزٍ عاليةٍ أو معاملات تدريجٍ عاليةٍ للفقرة والتي يمكن تقديرها من خلال شدة انحدار دالة استجابة الفقرة كما هو الحال في الفقرات (1)، (3)، (12)، (18) والتي امتلكت منحنيات استجابةٍ عالية الانحدار.

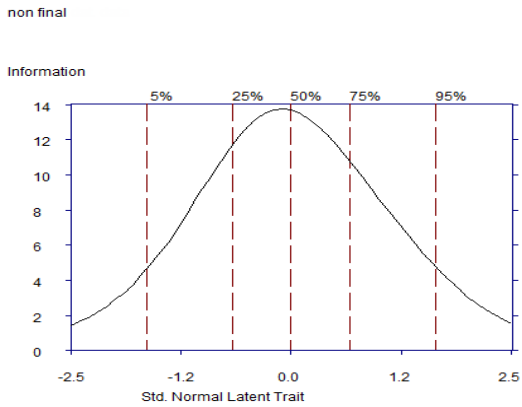
كما تم تمثيل دالة معلومات الاختبار للفقرات التي طابقت النموذج ثنائي المعلمة البارامتري كما هو مبين في الشكل (1)، وأيضاً تم تمثيل دالة معلومات الاختبار التي طابقت نموذج موكن اللابارامتري كما هو مبين في الشكل (2).



الشكل (1)

دالة معلومات الاختبار المؤلف من الفقرات المطابقة للنموذج البارامتري ثنائي المعلمة

يبين الشكل (1) مقدار المعلومات التي قدمها الاختبار بصورته البارامترية والتي بلغت ما يقارب (8.5).



الشكل (2)

دالة معلومات الاختبار المؤلف من الفقرات المطابقة لنموذج موكن اللابارامتري

كما يبين الشكل (2) مقدار المعلومات التي قدمها الاختبار بصورته اللابارامترية والتي بلغت ما يقارب (14).

يتضح من الشكلين (1) و (2) أن كمية المعلومات التي يمنحها النموذج اللابارامتري لموكن والتي بلغت (14) أكبر من المعلومات التي يزود بها النموذج البارامتري ثنائي المعلمة والتي بلغت (8.5).

مما سبق يتضح تفوق نموذج موكن اللابارامتري للتجانس الاطرادي في تقديمه للمعلومات على النموذج البارامتري ثنائي المعلمة، إذ قدم النموذج اللابارامتري قدراً من المعلومات بلغ (14) بينما قدم النموذج البارامتري قدراً من المعلومات بلغ (8.5) حيث تمركزت القيمة العظمى للمعلومات في كلتا الحالتين عند متوسط القدرة للطالبات، وقد تعزى مثل هذه النتيجة لسبب منطقي وهو عدد فقرات الاختبار، فقد بلغ عدد الفقرات لصورة الاختبار البارامترية للفقرات (14) فقرة، بينما بلغ عدد الفقرات لصورة الاختبار اللابارامترية (23) فقرة وهو فرقٌ واضحٌ في طول الاختبار لكلتا الصورتين الأمر الذي ساهم في إغناء صورة الاختبار اللابارامترية بقدرٍ أكبر من المعلومات كما ويمكن القول أن هذه النتيجة قد تعزى إلى ارتباط النماذج البارامترية واللابارامترية بمستويات الخطأ والتي يتم تقديرها من خلال علاقتها العكسية بكمية المعلومات التي يقدمها الاختبار، والتي قد تؤيد أن النماذج اللابارامترية تتمتع بمستويات أقل من الخطأ لتقدير دالة معلومات الاختبار، حيث أثبت كل من ويلس وبولت (Wells & Bolt, 2008) وليانغ (Liang, 2010) في دراستين منفصلتين قارنتا معدلات الخطأ بين طريقة لابارامترية وطريقتين بارامتريتين، من خلال تقدير دوال استجابة الفقرات ودوال المعلومات لابارامترياً باستخدام طرق الانحدار اللابارامتري الموجودة في برنامج (TESTGRAF) وهو البرنامج الذي تم استخدامه في هذه الدراسة حيث أظهرت النتائج تحكم الطريقة اللابارامترية في الخطأ في جميع الظروف التي تم خلقها من قبل الباحثين بشكلٍ أفضل عنها في الطرق البارامترية.

التوصيات:

وفي ضوء النتائج التي توصلت إليها الدراسة تقترح الباحثة ما يلي:

1. توصي الباحثة بتحليل نتائج البيانات الناتجة عن الاختبارات التحصيلية باستخدام نماذج نظرية استجابة الفقرة اللابارامترية المختلفة، إذ إنها توفر للباحثين معلومات جيدة عن الاختبار وفقراته دون فقدان فقرات مناسبة لهذا الغرض.
2. توصي الباحثة بتحليل البيانات الواقعة على مقياسٍ فئوي باستخدام كلا النوعين من النماذج البارامترية واللابارامترية إذ توفر كل منها معلومات مفيدة حول البيانات، أما في حال وقوع البيانات على مقياسٍ رتبي فيجب استخدام نماذج نظرية استجابة الفقرة اللابارامترية.
3. توصي الباحثة بعمل دورات تدريبية ومحاضرات تثقيفية لأساتذة الجامعات والمدارس والباحثين في المجال التربوي والنفسي للتعريف بالنماذج اللابارامترية المختلفة وفوائد استخدامها وكيفية العمل بها.
4. توصي الباحثة باستخدام النماذج اللابارامترية في بناء وتدريج المقاييس خاصة الهرمية منها.
5. توصي الباحثة بدراسة أثر اختلاف حجم العينة وطول الاختبار في نتائج مطابقة النموذجين ثنائي المعلمة والتجانس الاطرادي، اضافةً لدراسة المطابقة باستخدام نماذج بارامتريةٍ ولابارامتريةٍ أخرى.
6. توصي الباحثة بدراسة حسن المطابقة البيانات باستخدام أنواع أخرى من الاختبارات مثل مقاييس الاتجاهات ومقاييس الشخصية.
7. توصي الباحثة بدراسة ارتباط خصائص الفقرة ومعالمها مع مؤشر المطابقة الخاص بها للنماذج البارامترية المختلفة مما يساهم في فهم أعمق لمسببات عدم التطابق وإمكانية التحكم بها.

المصادر والمراجع

دلوع، أ. (2013). مدى التوافق بين نموذج استجابة الفقرة الثلاثي المعلمة البارامتري ونموذج موكن اللابارامتري لنوعين من فقرات اختبار تحصيلي. رسالة دكتوراه غير منشوره. الأردن: جامعة اليرموك. ص 1-73.

عوده، أ. (2010). القياس والتقويم في العملية التدريسية، ط.3 الأردن: دار الأمل للنشر والتوزيع. ص 159-240.

مساعده، م. (2013). أثر شكل فقرات الاختبار في مطابقة الفقرات لنموذج موكن اللابارامتري ونموذج ثنائي المعالم البارامتري. رسالة دكتوراه غير منشوره. الأردن: جامعة اليرموك. ص 1-88.

Allen, M. and Yen, W. (1979). Introduction to Measurement Theory, Montry, California: Brooks/Cole Publishing Company. P 118-127

Baker, (2001). The Basics of Item Response Theory, 2nded USA: ERIC Clearinghouse on Assessment & Evaluation. P21-35

Cliff, N. and Donoghue, S. R. (1992). Ordinal Test Fidelity Estimated by an Item Sampling Model. Psychometrika, (57): 217-236.

Cliff, N. and Keats, J. (2003). Ordinal Measurement in the Behavioral Sciences, NJ: Mahwan Lawrence Elbaum Associates. P 57-85

Crocker, L. and Algina, J. (1986). Introduction to classical and modern test theory, NY Holt: Rinhart & Winston. P 340-361

analysis of multiple - choice tests and questionnaire data [Computer software and manual]. Retrieved from http://www.psych.mcgill.ca/faculty/ramsay/ramsay.htm.

Reidpath,D. and Ahmadi, K. (2014). A novel nonparametric item response theory approach to measuring socioeconomic position: a comparison using household expenditure data from a Vietnam health survey, 2003. Emerging Themes in Epidemiology. 11, (9): 1-9.

Ringdal, K. Ringdal, G. kaasa, S. Bjordal, K. Wisloff, F. Sund storm, S. and Hjermstad, M. (1991). Assessing the consistency of psychometric properties of the HRQoL scales within the EORTC QLQ-C30 across populations by means of the Mokken scaling model. Quality of life research, (8): 25-43.

Santor, D. A. and Ramsay, J. O. (1998). Progress in the Technology of Measurement: Applications of Item Response Models. Psychological Assessment, (6): 255-270.

Shenkin, SD. Watson, R. Laidlaw, K. Starr, JM. Deary, IJ. (2014) The Attitudes to Ageing Questionnaire: Mokken Scaling Analysis. PLOS ONE, 9, (6): e99100, P 1-11.

Sijtsma, K. (1988). Contributions to Mokken's Nonparametric Item Response Theory, Amesterdam: Free University Press. P 83-104.

Sijtsma, K. (1998). Methodology review: Nonparametric IRT approaches to the analysis of dichotomous item scores. Applied Psychological Measurement, (22): 3-31.

Sijtsma, K. and Hemker, T. (2000). A taxonomy of IRT models for ordering of persons and items using simple sum scores. Journal of Educational and Behavioral Statistics, (25): 391-415.

Sijtsma, K. and Molenaar, I. (2002). Introduction to Nonparametric Item Response Theory: London, New Delhi: Sage Publication, International Educational and Professional Publisher Thousand Oaks. P 50-120.

Sijtsma, K. Emons, W. H. M. Bouwmeester, S. Nyklicek, I. and Roorda, L. D. (2007). Nonparametric IRT analysis of quality of life scales and its application to the World Health Organization Quality of Life Scale (WHOQOL-bref). Quality of Life Research, (17): 275-290.

Tendeiro, J. N. and Meijer, R. R. (2013). Detection of invalid test scores: The usefulness of simple nonparametric statistics. LSAC Research Report Series. 13, (5): 1-27.

Van Onna, M. (2003). Estimates of the Sampling Distribution of Scalability Coefficient H. Applied Psychological Measurement, (28): 427-449.

Watson, R.; Wang, W. and Thopson, D. (2014). Violations of local stochastic independence exaggerate scalability in Mokken

Douglas, J. (1997). Joint consistency of nonparametric item characteristic curves and ability estimation. Psychometrika, (62): 7-28.

Dyehouse, M. (2009). A Comparison of Model-Data Fit for Parametric & Nonparametric Item Response Theory Models Using Ordinal-Level Ratings. Dissertation Abstract International. (UMI No. 3379330).P 12-69

Haley, S. Pinkham, M. Dumas, H. Ni, P. Gorton, G. Watson, K. Montpent, K. Bilodeau, N. Hambleton, R. and Tucker, C. (2009). Evaluation of Item Bank for a Computerized Adaptive Test of Activity in Children with Cerebral Palsy. Physical Therapy. 89, (6): 589-600.

Junker, B. and Sijtsma, k. (2001). Cognitive Assessment Models with Few Assumptions &Connections with Nonparametric Item Response Theory, Applied Psychological Measurement, (25): 258-272.

Lee, S. Wollak, J. and Douglas, J. (2009). On the Use of Nonparametric Item Characteristic Curve Estimation Techniques for Checking Parametric Model Fit. Educational & Psychological Measurement. 69, (2): 181-197.

Liang, T. (2010). An Assessment of the Nonparametric Approach of Evaluating the Fit of Item Response Model. Dissertation Abstract International, (UMI No. 3397726). P 8-70>

Linden, W. and Hambleton, R. (1997). Handbook of Modern Item Response Theory, New York Berlin Heidelberg: Springer-Verlag. P 5-50.

Lord, F. (1977). Practical Application of Item Characteristic Curve Theory. Journal of Educational Measurement, (14): 117-138.

Magnusson, d. (1967). Test theory, Addison – Wesley: Reading, MA. P 198-218.

Maydeu, A. (2005). Further empirical results on parametric versus nonparametric IRT modeling of Likert-type personality data. Multivariate Behavioral Research, (40): 261-279.

Meijer, R. Sijtsma, K. and Smid, G. (1990). Theoretical and empirical comparison Of the Mokken and the Rasch approach to IRT. Applied Psychological Measurement, (14): 283-298.

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

Molenaar, I. W. and Sijtsma, K. (2000). MSP5 for Windows. User's manual MSP. Groningen, The Netherlands: iecProGAMMA. P 5-68.

Nunally, J. (1978). Psychometric Theory, USA: McGraw-Hill Book Company. P 256-300

Ramsay, J. O. (2000). TestGraf. A program for the graphical

Rasch Measurement Theory and Mokken Scaling. (Unpublished Doctoral Dissertation). Emory University. Atlanta, Georgia, 110-153.

Zickar, M. (1997). Identifying Untraited Individuals Using Model-based Measurement. Dissertation Abstract International. (UMI No. 9737304). P 8-72.

scaling analysis of the Chinese Mandarin SF-36. Health and Quality of Life Outcomes. 12, (149): 1-10.

Wells, C. S. and Bolt, D. M. (2008). Investigation of a nonparametric procedure for assessing goodness-of-fit in item response theory. Applied Measurement in Education. 21, (1): 22 - 40.

Wind, S. (2014). Evaluating Rater-Mediated Assessment with

A Comparison of Model Data Fit for the Mokken Non-Parametric and the Two Parameters Parametric Item Response Theory Model

Rana Thani Momani*

ABSTRACT

This study aimed at inspecting the optimal model which better fit the items of an achievement test by comparing the results of data analysis using the 2-parameters model and the non-parametric model of Mokken monotonic homogeneity. To achieve the purpose of the study the researcher used an achievement test in mathematics (teacher made) consists of (28) multiple choice items with four alternatives applied to a randomly selected sample for the female students of the 10th grade in Irbid for the academic year 2010/2011 and contsist of (834) students. To verify the basic assumptions of the IRT models SPSS package was used, then the R package used for analysing data fitness for mokken model and Bilog-Mg used for analysing data fitness for 2- parameters model, Finally, TESTGRAF used to estimate items and test information functions.The results showed that the monotonic homogeneity model fitted the items better than the 2-parameters model at the significancy level (α = 0.05), the nonparametric model fitted (23) items while the 2-parameters model fitted only (14) items. Finally, the tests information functions for the two versions of the tests (parametric and non- parametric) were plotted and showed that non-parametric model provides more information than the 2-pl parametric model.It was recommended to study item fit for different sample sizes, models and item types.

Keywords: Item Response Theory, Parametric Models, Non-parametric Model, Monotone Homogeneous Model, Scalability Coefficients.

* Department of Psychology, Al-Qassim University, KSA. Received on 17/04/2014 and Accepted for Publication on 19/01/2015.