

استخدام نظرية الاستجابة للفقرة في بناء فقرات اختبار محكي المرجع في الرياضيات بفقرات ثنائية التدرج ومتعددة التدرج وفق النموذج اللوجستي ثنائي المعلم

Using Item Response Theory in Constructing the Item Pool in Criterion- Referenced Testing with Dichotomous and Polytomous Items according to Two – Parameter Logistic Model

محمد دبوس

Mohammad Dabous

جامعة الاستقلال، أريحا، فلسطين

بريد الكتروني: mddabous@hotmail.com

تاريخ التسليم: (2015/7/2)، تاريخ القبول: (2016/3/13)

ملخص

هدفت الدراسة إلى استخدام نظرية الاستجابة للفقرة في بناء تجمع فقرات اختبار محكي المرجع في الرياضيات وفق النموذج اللوجستي ثنائي المعلم. ولتحقيق هذا الهدف تم إعداد تجمع من الفقرات ثنائية التدرج مكون من (50) فقرة، وتجمع من الفقرات متعددة التدرج مكون من (10) فقرات. وروعي في بناء هذه الفقرات أن تقيس تحصيل الطالب في الرياضيات في المعرفة المفاهيمية والمعرفة الإجرائية وحل المشكلات. وقد تكونت عينة الدراسة من (502) طالبا وطالبة أجابوا عن جميع الفقرات. وقد أظهرت النتائج تحقق افتراضات نظرية الاستجابة للفقرة. كما بينت نتائج التحليل فيما يتعلق بمطابقة الفقرات الثنائية التدرج للنموذج الثنائي المعلم أن جميع الفقرات كانت مطابقة للنموذج الثنائي المعلم، وأن الفقرات المتعددة التدرج كانت مطابقة لنماذج الاستجابة المتدرجة GRM. وفي ضوء نتائج الدراسة أوصى الباحث بمجموعة من التوصيات.

الكلمات المفتاحية: اختبار محكي المرجع، نموذج لوجستي ثنائي المعلم، نظرية الاستجابة للفقرة، التحصيل، الرياضيات.

Abstract

This study aimed at using the Item Response theory in constructing a Criterion-Referenced Test in Mathematics according to the Two-Parameters Logistic Model. To achieve this aim, the researcher prepared

a pool Dichotomous items consisted from (50) items and a pool of polytomous items consisted of (10) items. The items were constructed to measure the conceptual understanding, procedural knowledge, and problem solving. The study sample consisted of (502) students and all of them responded to all items. Results of the study indicated that (IRT) assumptions were achieved in the study data. Also, the results showed all dichotomous item achieved a goodness of fit according to the 2-parameters model, and all Polytomous items showed goodness of fit according to the Graded response Model. In light of the results of the study, the researcher recommended search a set of recommendations

Keywords: Criterion- Referenced Testing, Two – Parameter Logistic Model, Item Response Theory, Achievement, Mathematics.

المقدمة

تعد الاختبارات محكية المرجع Criterion Referenced Testing من الاتجاهات الحديثة نسبيًا في علم القياس والتقويم التربوي، فقد شهدت الاختبارات محكية المرجع تطورات ملحوظة منذ السبعينيات من القرن الماضي سواء من الناحية النظرية أو الناحية التطبيقية. فخبراء القياس التربوي والنفسي في كثير من الدول المتطورة وبعض الدول النامية يعملون جاهدين على تطوير أساليب تصميم الاختبارات التشخيصية مرجعية المحك وتقنيات بنائها، وابتكار أساليب إمبريقية ونماذج إحصائية مستحدثة يمكن الاسترشاد بها في تحليل فقرات هذا النوع من الاختبارات وتقييم جودتها.

وعلى الرغم من هذه الجهود البحثية والتطويرية، إلا أن هذا المدخل الجديد في القياس التربوي والنفسي لا يزال في مراحل المبكرة إذا ما قورن بمدخل القياس معياري المرجع، وما يرتبط من اختبارات ومقاييس مقننة مما يجعل هذا المدخل الجديد مجالًا خصبا للدراسات والبحوث المستقبلية.

ويمكن إرجاع مصطلح (محكية المرجع) إلى موضوع كتبه العالم الأمريكي روبرت جليسر في عام (1962) بعنوان (بعض التساؤلات عن تكنولوجيا التعليم وقياس نواتج التعلم)، وقد أثار هذا المقال كثير من الجدل بين علماء القياس عامة والمتخصصين في تطبيقات تكنولوجيا التعليم بخاصة، ولكن لم يحدث نشاط ملحوظ نحو تحقيق ما نادى به جليسر إلا حوالي عام (1969) (Ibrahim, 1991).

عندما تستخدم الاختبارات محكية المرجع فإنه لا يُهتم بالمركز النسبي للفرد بين أقرانه وإنما يُهتم بمقارنة علامته في الاختبار بمستوى أداء (Performance Standard) يكون بمثابة محك يدل على الاختبار بمستوى أداء المستوى المقبول لسلوك الفرد أو أدائه، فإذا كانت

علامة الفرد تعادل أو تفوق مستوى الأداء المحدد فإن الفرد يكون قد حقق المستوى المطلوب، أما إذا كانت علامته أقل من المستوى المحدد فإن أداءه يعد ضعيفا (Allam, 2001).

فبينما يُقارن في الاختبارات معيارية المرجع أداء الفرد بأداء أقرانه فإن الاختبارات المحكية المرجع يقارن فيها أداء الفرد بمستوى أداء محدد، دون الرجوع الى أداء الأفراد الآخرين. فالغرض من الاختبارات المعيارية المرجع تسهيل المقارنة بين الأفراد في المجال الذي يقيسه الاختبار، أما في الاختبارات محكية المرجع فهو تقويم أداء الفرد في ضوء كفايات معينة (Al-Dosari, 1998).

لقد أدت الانتقادات الموجهة للاختبارات معيارية المرجع إلى ظهور اتجاهات حديثة في القياس والتقويم التربوي، فأعلن بابهام Popham أن هناك كثيراً من الأسباب التي تجعل الاختبارات معيارية المرجع غير مفيدة في مجال التقويم التربوي، حيث أشارت الى المقاييس محكية المرجع كتجاه جديد أكثر ملائمة لمتطلبات التقويم التربوي (Kazim, 1981).

لقد ركز أصحاب الاتجاه المعاصر في القياس والتقويم التربوي جهودهم للتوصل الى أعلى مستويات الدقة والموضوعية في القياس بحيث تتحقق أدق علاقة بين أداة القياس والسمة المراد قياسها (Nunn ally, 1978). وبدأ اهتمامهم ينصب نحو نوع معين من الاختبارات التحصيلية أطلق عليها الاختبارات محكية المرجع لتحل كإقتراح بديل عن الاختبارات التحصيلية التي سادت وما زالت في المؤسسات التعليمية والتي يطلق عليها الاختبارات معيارية المرجع إذ تعرض هذا النوع من الاختبارات لمجموعة من الانتقادات منها قصورها في قياس فاعلية البرامج التعليمية وتقويم الطلبة تقويماً تشخيصياً يحدد في ضوءه جوانب التميز في العملية التعليمية والعوامل المؤثرة فيها إيجاباً، وسلباً بما لا يتيح إمكانية تحسين تلك النواتج، أو تعديل البرامج التعليمية لأجل تحقيق الأهداف التربوية المرجوة (Allam, 2001).

الاختبارات محكية المرجع تمتاز بمميزات كثيرة منها أنها لا تقتصر على تقويم أداء الطالب بشكل فردي بل تستعمل لأغراض التقويم التربوي عند تقويم أداء المدرسة ككل وحتى في اختيار الطلبة للدراسات العليا، كما أنها تعتمد أساساً في تحديد مجموعة من المخرجات التعليمية المرغوب فيها ومستويات انجازها المحددة، لذا فإن هذا النوع من الاختبارات يكون مفيداً عندما تختلف المؤسسات التربوية في تقديم محتويات مختلفة لطلبتها لأنها ستفرض على هذه المؤسسات وجوب أن يبلغ الطلبة مستويات إتقان محددة في اكتساب معلومات ومهارات محددة (Wikstrom, 2005).

وقد ارتبط هذا النوع من الاختبارات بمدخل معاصر من مداخل القياس والتقويم التربوي أطلق عليه نظرية الاستجابة للفقرة (Item Response Theory IRT)، إذ جاءت هذه النظرية للتغلب على كثير من جوانب القصور في النظرية الكلاسيكية، وقدمت أساليب موثوق فيها في معالجة قضايا أساسية في القياس التربوي مثل معايرة الفقرات وبناء بنوك الأسئلة وبناء الاختبارات محكية المرجع (Hambelton & Swaminathan, 1985).

وتفترض هذه النظرية أنه يمكن التنبؤ بأداء الأفراد أو يمكن تفسير أدائهم في اختبار نفسي أو تربوي معين في ضوء خاصية أو خصائص مميزة لهذا الأداء تسمى السمات (Traits)، أي أن هذه النظرية تفترض وجود سمة أو قدرة هي التي تحدد استجابة الفرد لل فقرات، وهذه السمة يشترك فيها جميع الأفراد (Molenaar & Hoijtink, 1996).

تستند نظرية الاستجابة للفقرة إلى افتراضات قوية ينبغي تحققها في البيانات لكي تؤدي إلى نتائج يمكن الوثوق بها، وفيما يلي توضيح لهذه الافتراضات الأساسية كما ذكرها هامبلتون وسوامنثيان (Hambelton & Swaminathan, 1985) :

1. افتراض أحادية البعد: Unidimensionality ويقصد به أن فقرات الاختبار تقيس سمة أو قدرة واحدة فقط تفسر أداء الفرد على الفقرة، بمعنى أن جميع الفقرات تقيس بعدا واحدا.
2. افتراض الاستقلال الموضعي: Local Independence ويقصد به أن استجابة المفحوص على فقرة ما لا تؤثر إيجابا أو سلبا على استجابته على أي فقرة أخرى (Hambelton & Swaminathan, 1985). وبمعنى آخر أن استجابات المفحوص لفقرات الاختبار مستقلة إحصائيا.
3. افتراض منحنى خصائص الفقرة: Item Characteristic Curve – ICC يمثل منحنى خصائص الفقرة دالة رياضية تربط بين احتمال الإجابة الصحيحة عن الفقرة $P(\theta)$ و قدرة المفحوص (θ) التي يتم قياسها بمجموعة من الفقرات في الاختبار الذي تم بناؤه لتلك الغاية.
4. افتراض سرعة الأداء: Speededness أي أن المفحوصين الذين يفشلون في الإجابة عن فقرات الاختبار يفشلون بسبب محدودية قدرتهم وليس بسبب أن الوقت غير كافٍ للوصول إلى الفقرة والإجابة عنها

تعد الصيغة الرياضية لنموذج الاستجابة للفقرة فارقا أساسيا بين نماذج الاستجابة. ويمكن تقسيم هذه النماذج إلى قسمين وهما:

أولاً: النماذج التي تتعلق بالفقرات ثنائية التدرج

1. النموذج أحادي المعلم (نموذج راش) One – Parameter Logistic Model

حيث يفترض هذا النموذج أن جميع الفقرات لها قيمة تمييزية واحدة وأن التخمين يكون في حده الأدنى، أي يفترض أن التخمين يساوي صفرا تقريبا وأن التمييز ثابت لكل فقرات الاختبار وأن صعوبة الفقرة تأخذ قيمة متغيرة. وتتخذ معادلته الشكل الآتي:

$$P_i(\Theta) = \frac{e^{D(\Theta - b_i)}}{1 + e^{D(\Theta - b_i)}} \quad i=1, 2, 3, \dots, n$$

حيث:

$P_i(\Theta)$: احتمال أن يجيب فرد قدرته Θ إجابة صحيحة عن الفقرة i .

b_i : معلمة صعوبة الفقرة i .

Θ : قدرة الفرد.

e : الأساس اللوغاريتمي الطبيعي ويساوي 2.718.

D : عامل التدرج ويساوي 1.7 وهي القيمة التي تجعل شكل المنحنى اللوغاريتمي يقترب من شكل اقتران المنحنى الطبيعي.

2. النموذج ثنائي المعلم (نموذج بيرنوم) Two – Parameter Logistic Model

يفترض هذا النموذج أن كلا من معاملي الصعوبة والتمييز متغيران وأن التخمين لجميع الفقرات يساوي صفراً. وتمتاز عملياته الحسابية بأنها أكثر صعوبة من نموذج راش، وتخذ معادلته الشكل التالي (Hambelton & Swaminathan, 1985):

$$P_i(\Theta) = \frac{e^{D a_i(\Theta - b_i)}}{1 + e^{D a_i(\Theta - b_i)}} \quad i=1, 2, 3, \dots, n$$

حيث: a_i معلمة تمييز الفقرة.

3. النموذج ثلاثي المعلم (نموذج لورد) Three – Parameter Logistic Model

ويقوم هذا النموذج على ثلاثة معالم وهي الصعوبة والتمييز والتخمين، ويتميز هذا النموذج عن النموذج ثنائي المعلم بأنه أضاف معلمة التخمين التي هي عبارة عن الخط التقاربي السفلي لمنحنى خصائص الفقرة. والمعادلة الرياضية لهذا النموذج هي:

$$P_i(\Theta) = c_i + (1 - c_i) \frac{e^{D a_i(\Theta - b_i)}}{1 + e^{D a_i(\Theta - b_i)}} \quad i=1, 2, 3, \dots, n$$

حيث: c_i هي معلمة تخمين الفقرة.

ثانياً: النماذج التي تتعلق بالفقرات متعددة التدرج ومن أشهرها

1. نموذج الاستجابات المتدرجة (GRM) Graded Response Model

قدم هذا النموذج سميجميا (Samejima)، وبعد هذا النموذج تعميماً للنموذج ثنائي المعلم، ويستخدم فقرات ليس بالضرورة أن تكون متساوية في عدد فئات الاستجابة عليها، حيث إن عدم تحقق هذا الشرط لا ينشأ عنه أي تعقيدات في تقدير معالم الفقرة أو تفسيرها (Embretson & Reise, 2000).

2. نموذج التقدير الجزئي (PCM) Partial Credit Model

يعد نموذج التقدير الجزئي توسيعاً لنموذج راش المتعلق بالفقرات ثنائية التدرج، حيث قام بتطويره ماسترز (Masters, 1982) في استراليا. ليصبح بالإمكان تقدير معالم الفقرات التي يمكن أن تأخذ في عملية التصحيح أكثر من قيمتين، وبشكل يعكس القدرة التي تقيسها الفقرة.

مما سبق يتضح التقدم الكبير الذي أحرزته نظرية الاستجابة للفقرة في التحرر من أثر خصائص الأفراد على معالم الفقرات، ومن أثر معالم الفقرات على خصائص الأفراد عند بناء الاختبارات.

لقد تعددت الدراسات التي تناولت استخدام نظرية الاستجابة للفقرة في بناء الاختبارات المحكية المرجع. فلقد أجرى الزليعي (AL-Zayla'I, 2014) دراسة هدفت إلى بناء اختبار تشخيصي المرجع لقياس محكي مهارات البحث العلمي لدى طلاب الدراسات العليا بكلية التربية بجامعة الباحة، واستخدام هذا الاختبار للتعرف على مدى تمكن طلبة الدراسات العليا من هذه المهارات، ومعرفة الفروق في مدى تمكن الطلاب في ضوء بعض المتغيرات، وتكون مجتمع الدراسة من طلاب وطالبات مرحلة الماجستير كلية التربية بجامعة الباحة. ولتحقيق أهداف هذه الدراسة قام الباحث ببناء اختبار تشخيصي محكي المرجع لقياس تمكن الطلبة في مهارات البحث العلمي مروراً بثلاث مراحل هي (مرحلة التحليل -مرحلة البناء -مرحلة التجريب). واستخدم الباحث المنهج الوصفي المسحي، وكان من أهم نتائج هذه الدراسة أن الاختبار الذي تم بناؤه في هذه الدراسة يتمتع بتوافر خصائص الاختبار الجيد، حيث يتمتع بدلالات صدق وثبات جيدة، كما أظهرت النتائج انخفاضاً واضحاً في مدى التمكن لدى طلبة الدراسات العليا بكلية التربية بجامعة الباحة سواءً في نتائج الاختبار الكلي أو نتائج الاختبارات الفرعية.

كما أجرى السامرائي والخفاجي (Samarrai & Al-Khafaji, 2012) دراسة هدفت إلى بناء اختبار تحصيلي محكي المرجع في مادة علم نفس الخواص وهي مادة دراسية مقررة لطلبة الصف الثالث في أقسام العلوم التربوية والنفسية لكليات التربية في الجامعات العراقية. إذ تم تحديد فقرات المادة من خلال كتاب الهيئة القطاعية لوزارة التعليم العالي والبحث العلمي واتفاق مدرسي المادة على المنهاج المقرر. تألف الاختبار من (160) فقرة اختبارية من نوع الاختيار من متعدد لقياس (160) هدفاً سلوكياً تغطي مفردات المادة. طبق الاختبار على عينة عشوائية بلغت (349) طالبا وطالبة وقد اعتمد الباحث على أنموذج راش في تحليل بيانات الاختبار. وبعد تحليل بيانات الاختبار ظهر أن فقرات الاختبار تقيس عاملاً واحداً عدا 22 فقرة تم استبعادها من الاختبار. كما أظهرت النتائج وجود (13) فقرة كانت قيمة مربع كاي لعشرة فقرات منها دالة إحصائياً لذا تم استبعادها من الاختبار فيما استبعد برنامج راسكال (3) فقرات لعدم مطابقتها لأنموذج راش.

في حين هدفت دراسة علي (Ali, 2012) إلى دراسة فاعلية استخدام نموذج راش أحادي المعلم في بناء اختبار محكي المرجع لمقرر القياس والتقويم في التربية ولتحقيق أهداف البحث

قامت الباحثة ببناء اختبار محكي المرجع للمقرر المذكور، وتكون من (99) مفردة اختبارية من نوع الاختيار من متعدد، وأربعة بدائل، أحدها يمثل الإجابة الصحيحة. واعتمد البحث المنهج التكاملي الذي يضم أنواعاً متعددة من المناهج حيث اعتمد في البحث المنهج الوصفي التحليلي لتحليل مقرر القياس والتقويم في التربية كما اعتمد المنهج التجريبي في تطبيق الاختبار المحكي المرجع الذي تم بناؤه. ولقد بينت النتائج أن تدريج الإختبار محكي المرجع لمقرر القياس والتقويم في التربية بنموذج راش يحقق افتراضات نظرية الاستجابة للمفردة، وذلك من خلال التحقق من افتراضات النظرية وهي: افتراض أحادية البعد، افتراض الاستقلال الموضوعي، افتراض خاصية المنحني المميز للمفردة، وافتراض التحرر من السرعة.

وأجرى حمادنه (Hamadneh, 2009) دراسة هدفت إلى استخدام نظرية الاستجابة للفقرة في بناء اختبار محكي المرجع في الرياضيات وفق النموذج اللوجستي ثلاثي المعلم. ولتحقيق هذا الهدف تم بناء اختبار في الرياضيات يقيس تحصيل الطالب في الإحصاء مؤلف من (28) فقرة من نوع الاختيار من أربعة بدائل. تكونت عينة الدراسة من (411) طالب وطالبة من طلبة الصف الأول الثانوي العلمي. أظهرت نتائج الدراسة ما يأتي: تحقق افتراضات نظرية الاستجابة للفقرة في بيانات الدراسة، وكذلك مطابقة الاستجابات عن 24 فقرة من فقرات الاختبار لتوقعات النموذج اللوجستي ثلاثي المعلم وحذف 4 فقرات لم تطابق النموذج. كم بينت النتائج المتعلقة بتقديرات معالم الفقرات (الصعوبة، والتميز، والتخمين) أنها كانت مقبولة ضمن المحكات التي أوردتها أدبيات القياس التربوي.

في حين أجرى عيثاوي (Ithawi, 2009) دراسة هدفت إلى استعمال أنموذج راش وفق نظرية الاستجابة للفقرة في بناء اختبار تحصيلي في مادة علم نفس الفروق الفردية، ولتحقيق أهداف الدراسة شرع الباحث في بناء الاختبار مستندا إلى المفردات المقررة من قبل الهيئة القطاعية لوزارة التعليم العالي. أعد الباحث (173) فقرة من نوع الاختيار من متعدد تغطي مستويات الأهداف المعرفية الثلاث الأولى بحسب تصنيف بلوم وهي: (المعرفة والفهم والتطبيق) واشتملت عينة البحث على (336) طالبا وطالبة من طلبة كلية التربية للجامعات العراقية تم اختيارهم بالطريقة الطباقية العشوائية أدخلت البيانات لبرنامج الحاسب الآلي راسكال من خلال معالجة البيانات تضمنت الصورة النهائية للاختبار (108) فقرات مطابقة لافتراضات الانموذج ومتطلباته، وقد تم حذف (28) فقرة لا تحقق أحادية البعد من خلال التحليل العاملي لفقرات الاختبار بطريقة المكونات الأساسية، فيما حذفت (29) فقرة لدلالة قيمة مربع كاي، واقتربت قيمة معامل التمييز من واحد صحيح، وكما تم استبعاد (8) فقرات لا تحقق استقلالية القياس. درجت 108 فقرات بحسب صعوبة الفقرات وقدرات الأفراد للحصول على الاختبار بصيغته النهائية والذي بلغت قيمة معامل ثباته (0.952) ومعامل تمييزه (0.713).

كما أجرى واغرن وهارفي (Wagner & Harvey, 2003) دراسة هدفت إلى تطوير أداة جديدة لأوجه التفكير الناقد لاختبار واتسون-جيلسر باستخدام نظرية الاستجابة للفقرة، حيث قام الباحثان بزيادة عدد البدائل لتكوين اختبار جديد (WAT). تكونت عينة الدراسة من (407) طالب وطالبة في المستوى الجامعي. وقد دلت نتائج التحليل على نسبة تخمين أقل وتحسين في

دالة معلومات الفقرة للاختبار المطور، وكانت أخطاء القياس في اختبار واتسون-جليسر أكبر من أخطاء القياس للاختبار الجديد بمقدار 50% على الأقل.

وأجرى فان (Fan, 1998) دراسة هدفت الى مقارنة الخصائص السيكومترية ل فقرات اختبار محكي المرجع وفق النظريتين الكلاسيكية والحديثة في القياس. وتكون الاختبار من اختبارين فرعيين: أحدهما في القراءة والآخر في الرياضيات، وتآلف اختبار القراءة من (48) فقرة، واختبار الرياضيات من (60) فقرة. وتكونت عينة الدراسة من (80) عينة فرعية، سحبت عشوائيا من مجتمع الدراسة المكون من أكثر من (1930000) طالب في الصف الحادي عشر في ولاية تكساس. أظهرت نتائج الدراسة أن معاملات الارتباط بين تقديرات صعوبة الفقرة في النظريتين عالية، وكانت أفضل الارتباطات بين النظرية الكلاسيكية والنموذج احادي المعلم، وكذلك كانت معاملات الارتباطات بين تقديرات تمييز الفقرة في النظريتين عالية.

مشكلة الدراسة وأسئلتها

تتحدد مشكلة الدراسة بالأسئلة الآتية:

1. ما مدى تحقق افتراضات نظرية الاستجابة للفقرة في بيانات الدراسة وفق النموذج اللوجستي ثنائي المعلم؟
2. ما قيم تقديرات معالم فقرات الاختبار (الصعوبة والتمييز) وفق النموذج اللوجستي ثنائي المعلم لنظرية الاستجابة للفقرة؟
3. ما درجة مطابقة الاستجابات عن فقرات الاختبار لتوقعات النموذج اللوجستي ثنائي المعلم لنظرية الاستجابة للفقرة ؟

أهداف الدراسة

هدفت الدراسة إلى استخدام نظرية الاستجابة للفقرة في بناء تجمع فقرات اختبار تحصيلي محكي المرجع في الرياضيات للصف الثامن الأساسي وفق النموذج اللوجستي ثنائي المعلم.

أهمية الدراسة

يمكن إيجاز أهمية هذه الدراسة بالنقاط الآتية:

1. توفير تجمع فقرات لبناء اختبار تحصيلي يمكن من قياس تحصيل طلبة الصف الثامن الأساسي في الرياضيات بمستوى عال من الدقة والموضوعية.
2. تقدم هذه الدراسة فقرات لبناء اختبار محكي المرجع يمكن الاعتماد عليه في قياس تحصيل الطلبة الرياضي بدلا من الاعتماد على اختبارات غير مقننة.
3. تسليط الضوء على ضرورة الاهتمام ببناء الاختبارات محكية المرجع في المواضيع المختلفة وفق نظرية الاستجابة للفقرة.

حدود الدراسة

1. **الحد المكاني:** تقتصر هذه الدراسة على مديرية التربية والتعليم- نابلس.
2. **الحد الزمني:** تم إجراء هذه الدراسة في نهاية الفصل الدراسي الاول من العام الدراسي 2013/2014م.
3. **الحد البشري:** تقتصر هذه الدراسة على عينة من طلبة الصف الثامن الأساسي للعام الدراسي 2013/2014م.

مصطلحات الدراسة:

اختبار محكي المرجع Criterion – Referenced Test: ذلك الاختبار الذي يستخدم في تقييم أداء الفرد بالنسبة إلى محك مستوى أداء مطلق دون الحاجة إلى مقارنة أدائه بأداء الأفراد الآخرين (Allam, 2005).

ويعرف إجرائياً: مجموعة من الفقرات من نوع الاختيار من متعدد بواقع أربعة بدائل (أحدها صحيح والثلاثة الباقية خاطئة) وأخرى مقالية أعدت وفق افتراضات نظرية الاستجابة للفقرة، وتحليل نتائج استجابات الطلبة باستخدام النموذج ثنائي المعلم بما يحقق دقة وموضوعية قياس تحصيل طلبة الصف الثامن في الرياضيات.

الفقرات ثنائية التدرج Dichotomous items

هي الفقرات الموضوعية والتي تتدرج فيها الاستجابة ثنائياً بحيث يعطى المفحوص العلامة (1) عند استجابته لها استجابة صحيحة والعلامة (صفر) عند استجابته لها استجابة خطأ (Croker & Algina, 1986).

الفقرات متعددة التدرج Polytomous Items

هي الفقرات المقالية والتي تتدرج فيها الإجابة في عدد من الخطوات، ويعطى عليها الفرد علامة بناءً على عدد الخطوات التي اجتازها بشكل صحيح وذلك حسب مستويات الاستجابة على هذه الفقرة (Croker & Algina, 1986).

نظرية الاستجابة للفقرة Item response theory

نظرية حديثة في القياس النفسي والتربوي يتم فيها تحديد العلاقة بين أداء المفحوص والسمة الكامنة موضع القياس وفق دالة رياضية محددة وتعتمد هذه النظرية على عدد من النماذج تسمى نماذج لسّمات الكامنة (Latent trait models) التي يتم من خلالها الربط بين الأداء على الفقرة وقدرة المفحوص (Hambelton & Swaminathan, 1985).

النموذج ثنائي المعلم (نموذج بيرنوم) **Tow – Parameter Logistic Model**

هو أحد نماذج نظرية الاستجابة للفقرة والذي يفترض أن كلا من معاملي الصعوبة والتمييز متغيران، وأن التخمين لجميع الفقرات يساوي صفراً (Allam, 2001).

الاختبار التحصيلي **Achievement Test**

أداة قياس لمدى تحصيل الطالب لما اكتسبه من معرفة أو مهارة معينة نتيجة التعليم أو التدريب.

الطريقة والإجراءات

يتضمن هذا القسم تعريفاً بمجتمع البحث وكيفية اختيار العينة ووصفها، كما يتضمن وصفاً لأداة البحث وإجراءات التطبيق ومن ثم جمع البيانات وتحليلها، والتحقق من ثبات الاختبار وصدقه وفق النموذج الثنائي المعلم واستعراض الوسائل الإحصائية التي أستخدمها الباحث لتحقيق أهداف البحث، وفيما يأتي وصفاً لهذه الإجراءات:

منهج الدراسة

بما أن هذه الدراسة تهدف إلى بناء اختبار محكي المرجع في الرياضيات، فقد توصل الباحث بعد تحديد مشكلة الدراسة إلى أن المنهج الملائم للدراسة الحالية هو المنهج الوصفي المسحي لأنه يمد الباحث ببيانات ومعلومات تسهم بشكل كبير بوصف ما هو كائن أثناء البحث، ويتضمن تفسيراً لهذه البيانات.

مجتمع الدراسة

تكون مجتمع الدراسة من جميع طلبة الصف الثامن الأساسي في المدارس الحكومية والخاصة للعام الدراسي 2013/2014م التابعين لمديرية التربية والتعليم لمنطقة نابلس في فلسطين والبالغ عددهم (4462) طالبا وطالبة منهم (2265) طالبا موزعين في (34) مدرسة للذكور و(2197) طالبة موزعات في (35) مدرسة للإناث وذلك بالرجوع إلى التقرير الإحصائي السنوي المعد من قبل قسم التخطيط في المديرية للعام الدراسي 2013/2014م.

ويبين الجدول (1) توزيع مجتمع الدراسة من حيث عدد المدارس والطلبة والجنس للصف الثامن الأساسي في مديرية تربية وتعليم نابلس.

جدول (1): توزيع أعداد المدارس والطلبة والشعب للصف الثامن الأساسي حسب الجنس في مجتمع الدراسة.

المجموع	الإناث	الذكور	
69	35	34	عدد المدارس
4462	2197	2265	عدد الطلبة
127	64	63	عدد الشعب

عينة الدراسة

تكونت عينة الدراسة من (502) طالبا وطالبة من طلبة الصف الثامن الأساسي في مديرية تربية وتعليم نابلس للعام الدراسي 2013/2014م، منهم (252) طالبا موزعين على (9) مدارس للذكور و(250) طالبة موزعات على (10) مدارس للإناث.

تم اختيار العينة بالطريقة القصدية، وذلك بمساعدة من مشرفي الرياضيات في مديرية التربية والتعليم في نابلس وبعض معلمي الرياضيات للصف الثامن، وكذلك مديري المدارس المتعاونة. وقد روعي عند اختيار الشعب وجود مستويات تحصيلية متباينة بين الطلبة من أجل الحصول على تباين في استجابات المفحوصين عن فقرات الاختبار. لأن غرض الدراسة ليس قياس تحصيل الطلبة وإنما الحصول على استجابات الطلبة من أجل بناء فقرات اختبار محكي المرجح. ويبين الجدول (2) توزيع أفراد عينة الدراسة حسب الجنس.

جدول (2): عدد أفراد عينة الدراسة حسب الجنس.

الجنس	عدد الطلبة
ذكور	252
إناث	250
المجموع	502

أداة الدراسة

من أجل تحقيق هدف الدراسة تم بناء تجمع من فقرات اختبارات محكية المرجع في الرياضيات من نوع أسئلة الاختيار من متعدد مؤلف من (50) فقرة، لكل فقرة (4) بدائل واحد منها صحيح. كما تم اعداد تجمع من الفقرات مؤلف من (10) فقرات متعددة التدرج من نوع الأسئلة محددة الاستجابة، كل فقرة منها مكونة من عدة خطوات متدرجة، حيث تعطى علامات جزئية على هذه الخطوات، وهذه العلامات الجزئية تختلف من سؤال لآخر، لتشكل في مجموعها العلامة الكلية للاختبار وهي (40).

صدق الاختبار

تم التحقق من صدق الاختبار ثنائي التدرج والاختبار متعدد التدرج من خلال نوعين من دلالات الصدق: الأول يتعلق بصدق المحتوى وذلك بناءً على نتائج تحليل المحتوى والأهداف وجدول المواصفات والفقرات. فتم التحقق منه من خلال أحكام المحكمين المختصين على مدى ملاءمة نتائج تحليل المحتوى والأهداف لجدول المواصفات، ومدى ارتباط كل فقرة بالهدف الذي تقيسه. وقام المحكمون بإبداء ملاحظاتهم من خلال الاستبانة التي وزعت عليهم، وتم الأخذ بملاحظاتهم وجرى التعديل بناءً عليها.

أما النوع الثاني من دلالات الصدق فهو الصدق العملي حيث استُخدم التحليل العملي لاستجابات الطلبة للتعرف على العوامل التي يتم قياسها في كل اختبار. فتم إدخال البيانات

باستخدام برنامج SPSS.15 وتم إجراء التحليل العاملي من الدرجة الأولى First order لل فقرات وبتدوير مائل Promax oblique rotation لاستخلاص العوامل المسؤولة عن الأداء في نماذج الاختبارات، وقد استخرجت قيمة الجذر الكامن (Eigen Value) ونسبة التباين المفسر لكل عامل من العوامل المستخلصة والنسبة التراكمية للتباين المفسر. ونظراً لتعدد العوامل التي أفرزها التحليل فقد تم إجراء تحليل عاملي من الدرجة الثانية Second order Factor Analysis لعوامل الدرجة الأولى مع التدوير المتعامد Varimax Rotation من أجل اختزال التعدد في العوامل إلى أبسط صورة ممكنة. ولقد أشارت نتائج التحليل العاملي في نموذج الاختبار ثنائي التدرج ونموذج اختبار متعدد التدرج إلى وجود عامل سائد يرجح أحادية البعد لأغراض التحليل باستخدام نماذج نظرية الاستجابة للفقرة. وبتعبير آخر فإن نموذجي الاختبارات كانا يقيسان قدرة عامة واحدة معبراً عنها في العامل السائد وهي القدرة الرياضية (على الرغم من أنها تتكون من ثلاث قدرات فرعية وهي المعرفة المفاهيمية، والمعرفة الإجرائية، وحل المشكلات).

ثبات الاختبار

تم الاستدلال على قيم الثبات لكل من نموذجي الاختبارات الثنائية التدرج والمتعددة التدرج في بايجاد معامل كرونباخ ألفا للاتساق الداخلي. ويبين الجدول (3) معاملات الثبات لنموذجي الاختبارات.

جدول (3): قيم معاملات الثبات بطريقة الاتساق الداخلي لنموذجي الاختبارات.

الفصل الدراسي	نوع الفقرات	معامل الثبات/ كرونباخ ألفا
الأول	ثنائية التدرج	0.888
	متعددة التدرج	0.908

يتبين من الجدول السابق أن قيم معاملات الثبات لنموذج الاختبار ثنائي التدرج بلغت (0.888) وهي ذات دلالة إحصائية، في حين أن معاملات الثبات لنموذج الاختبار متعدد التدرج بلغت (0.908). وتعبر هذه القيم عن معاملات ثبات عالية ومقبولة لأغراض الدراسة، مما يقدم مؤشراً جيداً على توافر دلالة الثبات.

المعالجات الإحصائية

من أجل معالجة البيانات إحصائياً حسب نظرية الاستجابة للفقرة تم استخدام البرمجيات الإحصائية: MULTILOG.7، SPSS.15، MINITAB.15، EXCEL، في تحليل البيانات للعينات الرئيسية وحساب ما يلي:

1. تقدير معلم الصعوبة والخطأ المعياري في تقدير قيم الصعوبة لل فقرات ثنائية التدرج ومتعددة التدرج.

2. تقدير معلم التمييز والخطأ المعياري في تقدير قيم التمييز لل فقرات ثنائية التدرج ومتعددة التدرج.
3. معامل الثبات بإيجاد معامل كرونباخ الفا للاتساق الداخلي.
4. إجراء التحليل العاملي.
5. مطابقة الفقرات الثنائية التدرج للنموذج ثنائي المعلم.
6. مطابقة الفقرات متعددة التدرج لنموذج الاستجابة المترتبة.

عرض نتائج الدراسة ومناقشتها

النتائج المتعلقة بسؤال الدراسة الأول ومناقشتها: ما مدى تحقق افتراضات نظرية الاستجابة للفقرة في بيانات الدراسة وفق النموذج اللوجستي ثنائي المعلم؟
للإجابة على هذا السؤال تم التحقق من افتراضات نظرية الاستجابة للفقرة وهي:

أولاً: افتراض احادية البعد (Unidimensionality):

الاختبار ذو الفقرات ثنائية التدرج

للتحقق من افتراض احادية البعد لبيانات الاختبار ثنائية التدرج تم إجراء التحليل العاملي من الدرجة الأولى First order للفقرات وبتدوير مائل Promax oblique rotation لاستخلاص العوامل المسؤولة عن الأداء في نماذج الاختبارات، وقد استخرجت قيمة الجذر الكامن (Eigen Value)، ونسبة التباين المفسر لكل عامل من العوامل المستخلصة والنسبة التراكمية للتباين المفسر. ويبين الجدول (4) نتائج التحليل العاملي من الدرجة الأولى.

جدول (4): نتائج التحليل العاملي من الدرجة الأولى لاستجابات العينة على الفقرات ثنائية التدرج.

العامل	الجذر الكامن	نسبة التباين المفسر	النسبة التراكمية للتباين المفسر %
1	8.036	16.071	16.071
2	1.751	3.502	19.573
3	1.551	3.102	22.675
4	1.443	2.887	25.398
5	1.418	2.837	28.398
6	1.335	2.670	31.068
7	1.330	2.659	33.727
8	1.271	2.543	36.270
9	1.230	2.460	38.730

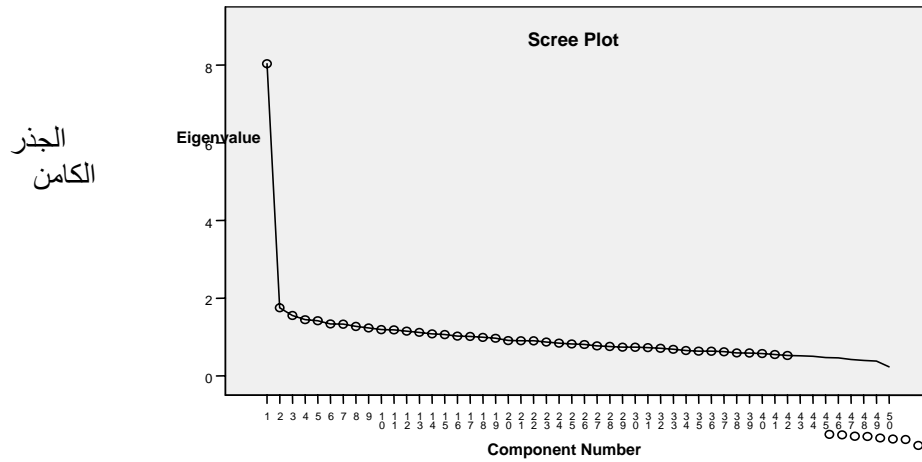
...تابع جدول رقم (10)

العامل	الجذر الكامن	نسبة التباين المفسر	النسبة التراكمية للتباين المفسر %
10	1.188	2.376	41.106
11	1.182	2.364	43.469
12	1.150	2.299	45.769
13	1.117	2.235	48.003
14	1.080	2.159	50.163
15	1.060	2.120	52.283
16	1.022	2.044	54.327
17	1.012	2.025	56.351

يتبين من الجدول (4) أن نتائج التحليل العاملي من الدرجة الأولى للنموذج ثنائي التدرج بينت وجود (17) عاملاً فسرت ما مجموعه (56.352%) من التباين الكلي.

كما تشير نتائج التحليل في الجدول (4) إلى أن قيمة الجذر الكامن للعامل الأول (8.036) وهي مرتفعة مقارنة مع بقية العوامل، وأن بقية الجذور الكامنة للعوامل الأخرى منخفضة ومتقاربة وهذا يعد مؤشراً على وجود عامل سائد يرجح أحادية البعد لأغراض التحليل باستخدام نماذج الاستجابة للفقرة. أما بالنسبة للتباين الذي فسره كل عامل نجد أن نسبة التباين المفسر بالعامل الأول (16.071%) من التباين الكلي، بينما نسبة التباين المفسرة بالعامل الثاني (3.502%) من التباين الكلي. ويلاحظ هنا أن نسبة التباين المفسرة بالعامل الأول عالية بالمقارنة مع نسبة التباين المفسر للعامل الثاني، كما يلاحظ أن الفروق في نسب التباين المفسرة لكل العوامل المتبقية قليلة ومتقاربة جداً، أي أن هناك تماثلاً نسبياً وشبه استقرار في نسب التباين المفسرة لجميع العوامل باستثناء العامل الأول وهذا يرجح تحقق أحادية البعد في بيانات الاختبار، وتتفق هذه النتيجة مع نتائج كل من دراسة السامرائي والخفاجي (Samarrai & Al-Khafaji, 2012)، وعلي (Ali, 2012) وحجازي والخطيب (Hijazi & Alkhateeb, 2014) في تحقق أحادية البعد.

وبملاحظة التمثيل البياني Scree plot للعوامل مع الجذور الكامنة لها في الشكل (1) يتبين فيه أن هناك تحولاً في ميل المنحنى عند العامل الثاني ويبقى الميل متقارباً لبقية العوامل مما يرجح أيضاً وجود عامل سائد يمكن أن يستدل منه على أحادية البعد لأغراض تقدير معالم الفقرات والأشخاص وذلك حسب لورد (Lord, 1980).



رقم العامل

شكل (1): التمثيل البياني لقيم الجذور الكامنة للعوامل من الدرجة الأولى لفقرات نموذج الثنائي التدرج.

ونظراً لتعدد العوامل التي أفرزها التحليل فقد تم إجراء تحليل عاملي من الدرجة الثانية Second order Factor Analysis لعوامل الدرجة الأولى مع التدوير المتعامد Varimax rotation من أجل اختزال التعدد في العوامل إلى أبسط صورة ممكنة. ولقد أشارت نتائج التحليل العاملي في فقرات نموذج ثنائي التدرج إلى وجود عامل سائد يرجح أحادية البعد. ويبين الجدول (5) نتائج التحليل العاملي من الدرجة الثانية

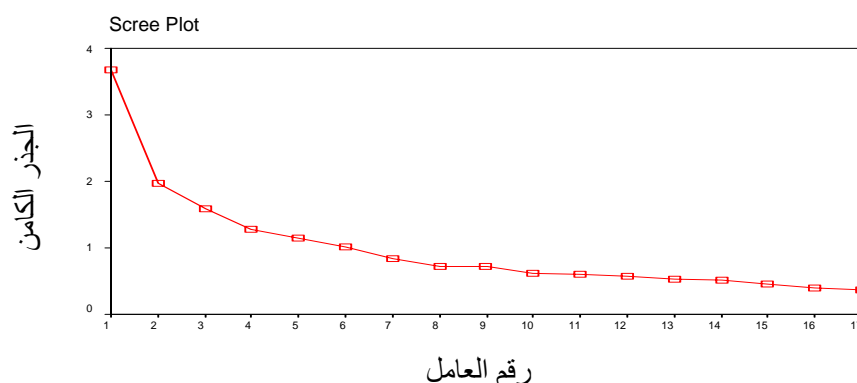
جدول (5): نتائج التحليل العاملي من الدرجة الثانية لاستجابات العينة على الفقرات ثنائية التدرج.

العامل	الجذر الكامن	نسبة التباين المفسر	النسبة التراكمية للتباين المفسر %
1	3.671	21.595	21.595
2	1.965	11.556	33.152
3	1.594	9.377	42.828
4	1.285	7.561	50.089
5	1.150	6.766	56.856
6	1.017	5.980	62.835

يظهر الجدول (5) أن عوامل الدرجة الثانية الستة فسرت ما مجموعه (62.834%) من التباين الكلي، وأن قيمة الجذر الكامن الأول كبيرة بالمقارنة مع بقية العوامل وأن بقية الجذور الكامنة للعوامل الأخرى منخفضة ومتقاربة وهذا مؤشر على وجود عامل سائد يرجح أحادية البعد.

ويوضح الشكل (2) التمثيل البياني لقيم الجذور الكامنة المكونة للعوامل المستخلصة في التحليل العامل من الدرجة الثانية، حيث أن هناك تحولا في المنحنى يبدأ عند العامل الثاني ويكون التغير بعده في ميل المنحنى طفيفا مما يرجح أحادية البعد.

وتعتبر هذه النتيجة متوافقة مع الكثير من الدراسات التي تناولت الخصائص السيكومترية لبناء الاختبارات ثنائية التدرج مثل دراسة الأسمر (Al-Asmar,2007) ودراسة عبد العزيز (Abd- Al Aziz, 2006).



شكل (2): التمثيل البياني لقيم الجذور الكامنة للعوامل من الدرجة الثانية ل فقرات ثنائية التدرج.

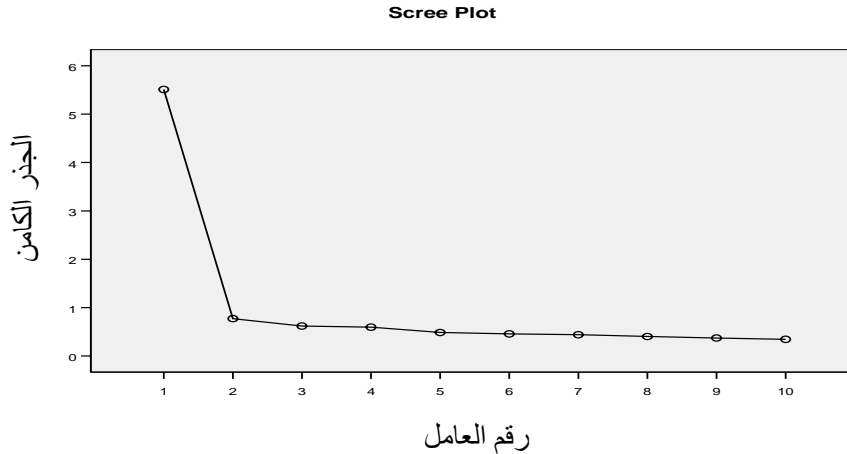
الاختبار ذو الفقرات متعددة التدرج

بينت نتائج التحليل العامل من الدرجة الأولى لنموذج الفقرات المتعددة التدرج وجود عامل واحد فسرت ما مجموعه (55.105%) من التباين الكلي. ويظهر الجدول (6) قيمة الجذر الكامن ونسبة التباين المفسر والنسبة التراكمية للتباين المفسر

جدول (6): نتائج التحليل العامل من الدرجة الأولى لفقرات متعددة التدرج.

العامل	الجذر الكامن	نسبة التباين المفسر	النسبة التراكمية للتباين المفسر %
1	5.510	55.105	55.105
2	0.773	7.727	62.831

وتشير النتائج في الجدول (6) إلى أنه يوجد جذر كامن واحد قيمته أكبر من واحد وأن بقية العوامل أقل من واحد مما يمكن القول أن هناك عاملاً سائداً يرجح أحادية البعد في الاختبار المتعدد التدرج. ويوضح الشكل (3) التمثيل البياني لقيم الجذور الكامنة المكونة للعوامل المستخلصة في التحليل العاملي للنموذج متعدد التدرج حيث إن قيمة الجذر الكامن للعامل الأول كبيرة جداً مقارنة بالعامل الثاني.



شكل (3): التمثيل البياني لقيم الجذور الكامنة للعوامل المكونة للاختبار متعدد التدرج.

ويستدل من هذه النتائج على التحقق من بعدية الفقرات، وهذا يتوافق مع ما ذكره كل من وورم و هامبلتون وسوامنثان (Warm, 1978; Hambleton & Swaminathan, 1985). لذا فإن فقرات نماذج الاختبارات المختلفة تقيس سمة واحدة هي القدرة الرياضية.

ثانياً: افتراض الاستقلال الموضوعي (Local Independence)

نظراً لأن افتراض أحادية البعد يكافئ افتراض الاستقلال الموضوعي، لذا تم الاكتفاء بالتحقق من افتراض أحادية البعد للاستدلال على تحقق افتراض الاستقلال الموضوعي. ومن الجدير ذكره أن هامبلتون وسوامنثان (Hambleton & Swaminathan, 1985) قد أوردوا أن تحقق افتراض أحادية البعد يكافئ الافتراض الثاني وهو افتراض الاستقلال الموضوعي Local Independence. وكذلك يرى وورم (Warm, 1978) أن افتراض أحادية البعد يتضمن افتراض الاستقلال الموضوعي وليس العكس.

ثالثاً: افتراض سرعة الأداء Speededness

تأكد الباحث أن المفحوصين الذين يفشلون في الإجابة عن فقرات الاختبار يفشلون بسبب محدودية قدرتهم وليس بسبب أن الوقت غير كافٍ للوصول إلى الفقرة والإجابة عنها.

النتائج المتعلقة بسؤال الدراسة الثاني ومناقشتها: ما قيم تقديرات معالم فقرات الاختبار (الصعوبة والتمييز) والقدرة للأفراد وفق النموذج اللوجستي ثنائي المعلم لنظرية الاستجابة للفقرة؟

لمعرفة قيم معالم الفقرات لكل فقرة من فقرات نماذج الاختبارات الثنائية التدرج حسب نظرية الاستجابة للفقرة تم اعتماد النموذج ثنائي المعلم، حيث استخدم البرنامج MULTILOG.7 في حساب معالم الصعوبة والتمييز والخطأ المعياري في قياس هذه القيم، كذلك تم حساب القيمة العظمى لدالة المعلومات $I(\theta)_{max}$ والقدرة المقابلة لها $(\theta)_{max}$ ، وذلك لكل فقرة من الفقرات الثنائية التدرج. ويبين الجدول (7) نتائج تحليل الفقرات الثنائية التدرج.

جدول (7): معالم الصعوبة والتمييز و $I(\theta)_{max}$ و $(\theta)_{max}$ للنموذج ثنائي التدرج

القدرة المقابلة للقيمة العظمى لدالة المعلومات $(\theta)_{max}$	القيمة العظمى لدالة المعلومات $I(\theta)_{max}$	الخطأ المعياري لمعلم الصعوبة	معلم الصعوبة	الخطأ المعياري لمعلم التمييز	معلم التمييز	الفقرة
0.8-	0.070	0.34	0.98-	0.14	0.53	1
0.4-	0.061	0.27	0.31-	0.14	0.49	2
0.3-	0.118	0.19	0.26-	0.15	0.69	3
0.4-	0.174	0.17	0.45-	0.15	0.83	4
0.9-	0.111	0.27	0.94-	0.16	0.67	5
0.4-	0.257	0.15	0.66-	0.18	1.02	6
0.2-	0.245	0.14	0.13-	0.16	0.99	7
0.6-	0.430	0.12	0.58-	0.22	1.31	8
0.8-	0.346	0.16	0.99-	0.20	1.18	9
0.3-	0.146	0.18	0.32-	0.16	0.77	10
0.8-	0.152	0.21	0.82-	0.17	0.78	11
0.5-	0.193	0.16	0.47-	0.16	0.88	12
0.5-	0.450	0.11	0.49-	0.20	1.34	13
0.5-	0.141	0.18	0.43-	0.15	0.75	14
0.2-	0.518	0.10	0.28-	0.19	1.44	15

...تابع جدول رقم (7)

القدرة المقابلة للقيمة العظمى لدالة المعلومات	القيمة العظمى لدالة المعلومات	الخطأ المعياري لمعلم الصعوبة	معلم الصعوبة	الخطأ المعياري لمعلم التمييز	معلم التمييز	الفقرة
$(\theta)_{\max}$	$I(\theta)_{\max}$					
0.2	0.186	0.16	0.15	0.15	0.86	16
0.8	0.249	0.18	0.86	0.16	1	17
0.2-	0.488	0.10	0.27-	0.20	1.4	18
0.2-	0.155	0.17	0.25-	0.16	0.79	19
0.6-	0.449	0.12	0.67-	0.20	1.34	20
0.2-	0.192	0.15	0.20-	0.16	0.88	21
0.2-	0.148	0.17	0.24-	0.15	0.77	22
0.4-	0.324	0.13	0.41-	0.18	1.14	23
0.0	0.265	0.13	0.05-	0.17	1.03	24
0.2-	0.309	0.12	0.19-	0.18	1.11	25
0.46	0.048	0.44	1.20	0.13	0.44	26
0.0	0.410	0.11	0.09-	0.18	1.28	27
0.5-	0.321	0.13	0.5-	0.19	1.013	28
0.0	0.228	0.14	0.07	0.16	0.96	29
0.6	0.261	0.16	0.66	0.16	1.02	30
0.2-	0.229	0.14	0.24-	0.16	0.96	31
0.8-	0.135	0.23	0.83-	0.15	0.73	32
0.6	0.252	0.16	0.54	0.16	1	33
1.1-	0.063	0.39	1.11-	0.14	0.5	34
0.4-	0.267	0.14	0.43	0.17	1.03	35
0.5	0.204	0.17	0.48	0.15	0.9	36
0.7-	0.096	0.25	0.68	0.15	0.62	37
1.4	0.036	0.57	1.41	0.13	0.38	38
0.4	0.185	0.17	0.43	0.15	0.86	39
1.1-	0.108	0.30	1.11-	0.16	0.66	40
0.6-	0.152	0.19	0.59-	0.16	0.78	41
0.7-	0.248	0.16	0.67-	0.18	1	42
0.3-	0.168	0.16	0.27-	0.15	0.82	43

...تابع جدول رقم (7)

القدرة المقابلة للقيمة العظمى لدالة المعلومات $(\theta)_{max}$	القيمة العظمى لدالة المعلومات $I(\theta)_{max}$	الخطأ المعياري لمعلم الصعوبة	معلم الصعوبة	الخطأ المعياري لمعلم التمييز	معلم التمييز	الفقرة
0.5	0.141	0.21	0.56	0.15	0.75	44
0.9-	0.252	0.17	0.88-	0.19	1.01	45
0.1-	0.102	0.20	0.18-	0.15	0.64	46
0.4-	0.284	0.13	0.34-	0.18	1.07	47
0.6-	0.253	0.15	0.60-	0.17	1.01	48
0.6-	1.178	0.08	0.60-	0.29	2.17	49
0.2-	0.670	0.09	0.26-	0.20	1.64	50
		0.184	0.224-	0.168	0.945	المتوسط الحسابي

يبين الجدول رقم (7) أن مدى قيم معالم صعوبة الفقرات في النموذج الثنائي التدرج تراوح ما بين (-1.11) و (+1.41). وبمتوسط مقداره (-0.224)، وبخطأ معياري يتراوح بين (0.08) و (0.57) وبمتوسط لهذا الخطأ المعياري قدره (0.184) وهي قيمة منخفضة نوعاً ما، مما يعطي مؤشراً على دقة تقدير معلم الصعوبة للفقرات الثنائية التدرج.

وقد حصلت الفقرة (34) على أدنى قيمة لمعلم الصعوبة حسب نماذج نظرية الاستجابة للفقرة وكانت (-1.11)، في حين حصلت الفقرة (38) على أعلى قيمة لمعلم الصعوبة وبلغ مقدارها (1.41).

وقد أشارت دراسة جمحاوي (2000) أن الفقرات التي تقع قيم معامل صعوبتها بين (-1.5) لوجيت إلى (+1.5) لوجيت تعد ضمن المدى المتوسط لمعاملات الصعوبة، وبناء على ذلك فإن جميع فقرات النموذج ثنائي التدرج تعد فقرات متوسطة الصعوبة.

كما يلاحظ من الجدول (7) أن مدى تمييز الفقرات للنموذج الثنائي التدرج تراوح بين (0.38) و (2.17) بمتوسط مقداره (0.945) وبخطأ معياري تراوح بين (0.13) و (0.29) ومتوسط للخطأ المعياري مقداره (0.168)، وهذا يعطي مؤشراً على أن معاملات تمييز الفقرات كان مناسباً. وقد حصلت الفقرة (49) على أعلى قيمة لمعلم التمييز وكانت (2.17) في حين حصلت الفقرة (38) على أدنى قيمة لمعلم التمييز وكانت (0.38).

ولمعرفة قيم معالم الفقرات لكل فقرة من فقرات نماذج الاختبارات المتعددة التدرج حسب نظرية الاستجابة للفقرة تم اعتماد نموذج الاستجابة المتدرجة Graded Response Model (GRM) حيث استخدم البرنامج MULTILOG.7 في حساب معلم الصعوبة والخطأ

المعياري لمعلم الصعوبة، وحساب معلم التمييز والخطأ المعياري لمعلم التمييز، والقيمة العظمى لدالة المعلومات $I(\theta)_{max}$ والقدرة المقابلة لها $(\theta)_{max}$ ، وذلك لكل فقرة من الفقرات المتعددة التدرج. ويبين الجدول (8) نتائج تحليل الفقرات المتعددة التدرج.

جدول (8): معالم الصعوبة والتمييز و $I(\theta)_{max}$ و $(\theta)_{max}$ للنموذج متعدد التدرج.

القدرة المقابلة للقيمة	القيمة العظمى لدالة المعلومات $I(\theta)_{max}$	صعوبة الخطوة				التمييز الخطأ المعياري للتمييز	الفقرة
		الخطأ المعياري للصعوبة					
		خطوة 4	خطوة 3	خطوة 2	خطوة 1		
1.0	1.292		1.89	0.76	0.43-	1.50	1
			0.19	0.13	0.11	0.18	
0.2-	0.899		0.81	0.04-	0.95-	2.51	2
			0.08	0.07	0.08	0.24	
0.3-	0.376		0.57	0.33-	1.08-	2.86	3
			0.08	0.06	0.08	0.26	
1.2-	0.515	0.83	0.35	0.43-	1.039-	2.90	4
		0.08	0.07	0.06	0.10	0.23	
0.8-	0.583	2.62	1.57	0.50	0.61-	1.60	5
		0.26	0.16	0.11	0.11	0.17	
0.0	1.066			0.08-	1.08-	1.91	6
				0.08	0.11	0.20	
0.2-	0.579			0.01-	1.12-	2.01	7
				0.10	0.15	0.24	
1.0-	0.540			0.09-	1.18-	2	8
				0.09	0.13	0.24	
0.4	0.930		0.63	0.24-	1.21-	2.21	9
			0.09	0.07	0.12	0.20	
1.2-	0.418	2.34	1.54	0.18	0.81-	1.97	10
		0.20	0.13	0.09	0.10	0.19	

يبين الجدول (8) أن مدى قيم معالم الصعوبة لفقرات النموذج المتعدد التدرج يمتد من (0.09-) إلى (2.62) بمتوسط مقداره (0.951)، وبخطأ معياري يتراوح بين (0.08) و(0.26) ومتوسط مقداره (0.125)، وهي قيم منخفضة نوعاً ما مما يعطي مؤشراً على دقة معلم

الصعوبة لل فقرات متعددة التدرج. وقد حصلت الفقرة (8) على أدنى معلم صعوبة و قيمته (-0.09) في حين حصلت الفقرة (5) على أعلى معلم صعوبة و قيمته (2.62). كما تعتبر الفقرات (1، 10) صعبة لأن معلم الصعوبة لها (1.89، 2.34) على الترتيب، وبقية الفقرات ذات صعوبة متوسطة.

أما معلم التمييز فتراوح بين (1.50) و (2.90) بمتوسط مقداره (2.147)، وبخطأ معياري تراوح بين (0.17) و (0.26) ومتوسط قدره (0.215)، وهذا يعطي مؤشرا على أن تمييز الفقرات كان مناسباً.

وقد حصلت الفقرة (1) على أقل معلم تمييز و قيمته (1.5) في حين حصلت الفقرة (4) على أعلى قيمة تمييز و قيمته (2.90).

ويلاحظ من قيم صعوبة الفقرات وفق نظرية الاستجابة للفقرة وجود تباين في صعوبة الفقرات، الأمر الذي نحتاجه في بناء الاختبارات محكية المرجع. كما ويلاحظ أن تمييز الفقرات المتعددة التدرج حسب نظرية الاستجابة للفقرة أفضل من تمييز الفقرات ثنائية التدرج.

ولقد تم حساب تقديرات معلم القدرة لنموذجي الاختبارات ثنائية التدرج ومتعددة التدرج، ويبين الجدول (9) القيم الدنيا والعليا والمتوسط الحسابي لتقديرات القدرة في عينات الدراسة لنموذجي الاختبارات

جدول (9): القيم الدنيا والعليا والمتوسط الحسابي لتقديرات القدرة في عينة الدراسة لنموذجي الاختبارات.

نموذج الاختبار	القيمة العليا للقدرة	القيمة الدنيا للقدرة	المتوسط الحسابي
نموذج ثنائي التدرج	2.818	-1.575	0.248
نموذج متعدد التدرج	2.296	-1.757	0.071

يظهر الجدول (9) أن تقديرات معلم قدرة الأفراد على نموذج الاختبارات ثنائية التدرج تراوحت ما بين (-1.575 لوجيت) إلى (2.818 لوجيت)، وبمتوسط حسابي (0.248 لوجيت). وهنا يلاحظ أن متوسط تقديرات القدرة للأفراد تزيد على متوسط تقديرات صعوبة الفقرات الثنائية التدرج والتي كانت (-0.224 لوجيت) مما يمكن القول إن الفقرات الثنائية التدرج بوجه عام كانت في متناول قدرات أفراد العينة، وتتفق هذه النتيجة مع دراسة الأسمر (2007, AI-Asmar) بالنسبة للفقرات ثنائية التدرج.

أما بالنسبة إلى تقديرات معلم القدرة لنماذج الاختبارات متعددة التدرج، فقد أشارت النتائج إلى أن متوسط تقديرات قدرات الأفراد تراوح ما بين (-1.757 لوجيت) إلى (2.296 لوجيت) وبمتوسط حسابي (0.071 لوجيت)، ويلاحظ هنا أن متوسطات صعوبة الفقرات والتي كانت (0.951) لوجيت كانت أعلى من متوسطات قدرة الأفراد، مما يمكن القول إن الفقرات متعددة التدرج كانت صعبة إلى حد ما، وتختلف هذه النتيجة مع دراسة الأسمر (2007, AI-Asmar)

بالنسبة لل فقرات متعددة التدرج والتي أظهرت أن متوسط قيم قدرات الأفراد أعلى من متوسط صعوبة الفقرات للفقرات متعددة التدرج.

النتائج المتعلقة بسؤال الدراسة الثالث ومناقشتها: ما درجة مطابقة الاستجابات عن فقرات الاختبار لتوقعات النموذج اللوجستي ثنائي المعلم لنظرية الاستجابة للفقرة ؟

للكشف عن درجة مطابقة استجابات المفحوصين على الفقرات ثنائية التدرج مع النموذج ثنائي المعلم تم استخدام البرمجيتين MULTILOG.7 و MINITAB.15 فتم تحليل استجابات الطلبة على الفقرات ثنائية التدرج، حيث أظهرت النتائج تطابق جميع الفقرات لجميع النماذج ثنائية التدرج مع نموذج ثنائي المعلم، وهذا ما أكدته قيمة مربع كاي لفحص المطابقة عند مستوى الدلالة $\alpha = 0.05$. والجدول (10) يبين هذه النتائج:

جدول (10): مربع كاي ومستوى الدلالة لفحص مطابقة الفقرات ثنائية التدرج للنموذج ثنائي المعلم.

النموذج ثنائي التدرج		
الفقرة	قيمة χ^2	مستوى الدلالة
1	0.0000224	0.996
2	0.0000195	0.996
3	0.0000124	0.997
4	0.0000080	0.998
5	0.0000229	0.996
6	0.0000052	0.998
7	0.0000002	1.000
8	0.0000028	0.999
9	0.0000284	0.996
10	0.0000104	0.997
11	0.0000010	0.999
12	0.0000050	0.998
13	0.0000123	0.997
14	0.0000118	0.997
15	0.0000621	0.994
16	0.0000014	0.999
17	0.0000022	0.999
18	0.0000529	0.994

...تابع جدول رقم (10)

النموذج ثنائي التدرج		
مستوى الدلالة	قيمة χ^2	الفقرة
0.998	0.0000068	19
1	0.0000002	20
0.998	0.0000020	21
0.998	0.0000079	22
0.999	0.0000020	23
0.999	0.0000014	24
0.998	0.0000058	25
0.997	0.0000188	26
0.995	0.0000385	27
1	0.00000002	28
1	0	29
1	0	30
1	0.0000002	31
0.996	0.0000228	32
1	0.0000002	33
0.996	0.0000247	34
1	0	35
0.999	0.0000010	36
0.996	0.0000201	37
0.997	0.0000171	38
0.999	0.0000020	39
0.996	0.0000279	40
0.997	0.0000135	41
0.998	0.0000062	42
0.998	0.0000058	43
0.998	0.0000093	44
0.997	0.0000165	45
0.997	0.0000160	46
0.999	0.0000007	47

...تابع جدول رقم (10)

النموذج ثنائي التدرج		
مستوى الدلالة	قيمة χ^2	الفقرة
0.999	0.0000034	48
0.991	0.0001149	49
0.990	0.0001470	50

وللكشف عن درجة مطابقة استجابات الفقرات متعددة التدرج مع نموذج الاستجابة المتدرجة (GRM) Graded Response Model فقد تم تحليل استجابات الطلبة عن الفقرات متعددة التدرج باستخدام البرمجيتين MULILOG.7 و MINITAB 15 عند مستوى الدلالة $\alpha = 0.05$ حيث أظهرت النتائج تطابق جميع الفقرات المتعددة التدرج مع نموذج الاستجابة المتدرجة (GRM) ، وهذا ما أكدته قيمة مربع كاي لفحص المطابقة حيث كان مستوى الدلالة لاختبار مربع كاي أكبر من 0.05. والجدول (11) يبين هذه النتائج.

جدول (11): مربع كاي ومستوى الدلالة لفحص مطابقة الفقرات متعددة التدرج مع نموذج الاستجابة المتدرجة.

نموذج (AIP)		
مستوى الدلالة	قيمة χ^2	الفقرة
1	0.0063386	1
1	0.0069756	2
0.999	0.0210148	3
1	0.0125046	4
1	0.0100901	5
0.999	0.0026211	6
0.991	0.0186342	7
0.994	0.0113106	8
1	0.0103622	9
1	0.0143177	10

يلاحظ مما سبق بالنسبة لمطابقة الفقرات الثنائية التدرج والفقرات متعددة التدرج للنموذج الثنائي المعلم وجود اختلاف بين نسبة الفقرات المطابقة للنموذج الثنائي المعلم في هذه الدراسة والتي كانت 100% ونسبة الفقرات المطابقة للنموذج الثنائي المعلم الواردة في كل من دراسة حجازي والخطيب (Hijazi & Alkhateeb, 2014)، واون (Onn, 2013) وجمحاوي (Jomhawi, 2000) والتي بلغت نسبة الفقرات المطابقة فيها على التوالي (95%، 89%،

76%) وقد يُعزى ذلك إلى اختلاف حجم العينات وعدد الفقرات المستخدم في هذه الدراسات وإلى طبيعة صيغ الفقرات.

كما ويعزو الباحث السبب في مطابقة جميع الفقرات للنموذج المستخدم إلى أن معاملات التمييز للفقرات كانت موجبة، ولا يوجد أي فقرة كان معامل تمييزها سالبا أو متدنيا جدا. كذلك كان لجدية الطلبة في الإجابة عن فقرات النماذج أثر في مطابقة جميع الفقرات للنموذج المستخدم. بناءً على النتائج السابقة فإن الباحث يوصي بما يلي:

1. نظرا لتمتع الاختبار بدلالات صدق وثبات مقبولة وبمعايير تتفق مع متطلبات النموذج اللوجستي ثنائي المعلم فإن الباحث يوصي باستخدام تجمع الاختبار الذي تم بناؤه أو أي اختبار مستل منه لقياس تحصيل طلبة الصف الثامن الأساسي.
2. اجراء المزيد من الابحاث والدراسات على تجمع الفقرات للاختبارات التي تم بناؤها وذلك باستخدام نموذج الثلاثي المعلم.

References (Arabic & English)

- Abdul Aziz, Nabil. (2006). *The Efficiency of Item Characteristic Curve Approaches in Detecting a Gender Related Differential Item Functioning of Polytomous and Dichotomous Mathematics Test Items*. Unpublished Ph.D. Amman Arab University.
- Al-Asmar, Mustafa. (2007). *The Psychometric Characteristics of Two Forms of Mathematics Test, with Dichotomous Items in one Form and Polytomous Items in the Other, According to The Classical and Modern Test Theory*. Unpublished Ph.D. Amman Arab University.
- Al-Dosari, Ibrahim Mubarak. (1999). A reference frame in the educational assessment of Arabic Gulf states. *Journal of Education*, 9(31). 138 - 158.
- Ali, Neda'. (2012). *The Effectiveness of Using Rasch Model in Constructing a Criterion Reference Test in Course of Measurement and Evaluation in Education*. Unpublished Ph.D. Damascus university.
- Allam, Salah al-Deen. (2001). *Criterion Reference Diagnostic Tests in Educational & Psychological Domains*. Cairo: Dar Al -Fikre Alarabi.

- Allam, Salah al-Deen. (2005). *Educational and Psychological Measurement and Evaluation, Fundamentals and Applications of it's Contemporary Trends*. Cairo: Dar Al -Fikr Alarabi.
- AL-Zayla'I, Mohammad. (2014). *Constructing a Diagnostic Criterion-Referenced Test for the Measuring of Scientific Research Skills of the graduate Students in the Collage of Education at Al-Baha University*, Unpublished master's thesis UMM AL- QURA University.
- Crocker, L. L. & Algina, J. (1986). *Introduction to Classical and Modern Test Theory*. New York: Holt, Rinehart & Winston.
- Embretson, Susan, Reise, Steven. (2000). *Item Response Theory for Psychologists*, Lawrence Erlbaum Associates.
- Fan, X. (1998). Item response theory and classical theory: An empirical comparison of their item/ person statistics. *Educational and Psychological Measurement*, 58 (3). 257-381.
- Hamadneh, Iyad Mohammed. (2009). Using Item Response Theory in Constructing a Criterion Referenced Test in Math According to the Three-Parameters Logistic Model. *Journal of Educational and Psychological Sciences*, 10(2). 215-238.
- Hambleton, R.K. Swaminthan, H. (1985). *Item Response Theory: rinciples and Application*, Kluwer. Nijhoff Publishing, Boston.
- Hijazi, Taghreed. & Alkhateeb, Abdallah. (2014). The Compatibility between the Classical Test Theory and the two Parameter Logistic models in Item Stratification of CriterionReferenced Test in Provisions of Recitation and Tajweed. *An-Najah University Journal for Research - B (Humanities)*, 28 (10). 2239-2270
- Ibrahim, Mahmoud Mohammed. (1991). *Contemporary trends in the psychological, educational measurement. RISALAT TARBIA*, Educational Research Department of the Directorate for Educational Development, Ministry of Education, Muscat, Oman.
- Ithawi, Mohammed Menem Salman. (2009). *Using Rasch Model according to Latent Traits Theory in building achievement test in a*

- course of psychology of individual differences*. Unpublished Master Thesis, Baghdad University, College of Education-Ibn Rushd.
- Jmhawi, Inas. (2000). *Comparing Between the Classical Test Theory and Item Response Theory characteristics in Scale of Mathematical Ability*, Unpublished master's thesis, Yarmouk University, Irbid, Jordan.
 - Kazim, Amina Mohamed. (1981). About differing interpretations of the results of the tests, *Journal of Social Sciences*, 9(3). 150 - 183.
 - Lord, F. m. (1980). *Application of Item Response Theory To Practical Testing Problem*. Hillsdal, New Jersey.
 - Masters. N.G. (1982). A Rasch Model for Partial Credit Scoring. *Psychometrika*, 47,149-174.
 - Molenaar, I. & Hoijsink, H. (1996). Person-fit and the rasch model, with an Nakamura.Y. (2001). Rasch Measurement and Item Banking, Theory and Practice. *Education Resources Information*. 21=(1).
 - Onn, D. (2013, May). *Classical test theory versus item response theory: An evaluation of the comparability of item analysis results*. Joint Admissions and Matriculation Board, 1-23.
 - Samarrai, Muhammad Anwar. & Al-Khafaji, Ahmad. (2012). *Building Criterion Reference Test in Psychology of properties among students of Educational and Psychological Sciences Sections*. Al-Ostath magazine, 203: 964 - 1002.
 - Wagner, A., & Harvey, R. (2003). *Developing a new critical thinking test using item response theory*. Paper presented at the Annual conference of the Society for Industrial Organizational Psychology. Orando. Retrieved May, 4, 2005, from: www.criticalthinking.org
 - Wikstrom, Christina. (2005). *Criterion-Referenced Measurement for Educational Evaluation and Selection*, Umea University.