215
المجلد 10 العدد 2 يونيو 2009

استخدام نظرية الاستجابة للمفردة في بناء اختبار محكي المرجع في الرياضيات وفق النموذج اللوجستي ثلاثي المعلم

د. إياد محمد حمادنه
كلية العلوم التربوية
جامعة آل البيت

استخدام نظرية الاستجابة للمفردة في بناء اختبار محكي المرجع في الرياضيات وفق النموذج اللوجستي ثلاثي المعلم

د. إياد محمد حمادنه
كلية العلوم التربوية
جامعة آل البيت

الملخص

هدفت هذه الدراسة إلى استخدام نظرية الاستجابة للمفردة 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 (٢٨) فقرة من نوع الاختيار من أربعة بدائل. تكونت عينة الدراسة من ٤١١ طالباً وطالبة، من طلبة الصف الأول الثانوي العلمي. أظهرت نتائج الدراسة ما يأتي: تحقق افتراضات نظرية الاستجابة للمفردة في بيانات الدراسة، وكذلك مطابقة الاستجابات عن ٢٤ فقرة من فقرات الاختبار لتوقعات النموذج اللوجستي ثلاثي المعلم، وحذف ٤ فقرات لم تطابق النموذج. كما بينت النتائج المتعلقة بتقديرات معالم الفقرات (الصعوبة، والتمييز، والتخمين) أنها كانت مقبولة ضمن المحكات التي أوردتها أدبيات القياس التربوي.

الكلمات المفتاحية: نظرية الاستجابة للمفردة، أحادية البعد، الاختبارات محكية المرجع.

* تاريخ تسلم البحث: ٢٠٠٧/٩/١م
* تاريخ قبوله للنشر: ٢٠٠٨/٥/٢٢م

Using Item Response Theory in Constructing a Criterion-Referenced Test in Math According to the 3–Parameters Logistic Model

Dr. Iyad M. Hamadneh
Faculty of Educational Sciences
Al-albyt University

Abstract

This study aimed at using Item Response Theory (IRT) in constructing a Criterion-Referenced Test in math according to the 3–Parameters Logistic Model (3-PLM). To achieve this, a test in Math was designed to measure the student's achievement in statistics. The test consisted of (28) four choices items. The sample of the study consisted of (411) male and female from the first secondary class - science stream. Results of the study indicated that: (IRT) assumptions were achieved in the study data. Also, students' responses for 24 items of the test fit (3-PLM), whereas 4 items did not. The results also indicated that items parameters estimations (Difficulty, discrimination and guessing) were acceptable within the indexes available in the literature of educational measurement.

Key Words: item response theory, unidimensionality, criterion referenced tests.

استخدام نظرية الاستجابة للمفردة في بناء اختبار محكي المرجع في الرياضيات وفق النموذج اللوجستي ثلاثي المعلم

د. إياد محمد حمادنه
كلية العلوم التربوية
جامعة آل البيت

المقدمة

حظيت عملية تصنيف الطلبة، حسب مستوى التمكن من الأهداف التدريسية، باهتمام كبير في مجال القياس و التقويم التربوي والنفسي. حيث إنّ مسألة تصنيف الطلبة تساعد المربين على اتخاذ قرارات تعليمية مهمة، تمس حاضر الطلبة ومستقبلهم، وتصنيف الطلبة يتطلب كفايات ومهارات محددة يجب أن يتقنها المعلم، أو يتمكن منها، ويحللها، ويعيد صياغتها في صورة أهداف تعليمية إجرائية قابلة للقياس، ويقدم هذه الأهداف إلى الطالب قبل بدء العملية التعليمية، ويقدر مدى تحققها باستخدام ما يسمى بالاختبارات محكية المرجع (Criterion Referenced Tests) التي يعتمد على نتائجها في تصنيف الطلبة إلى مجموعتين: إحداهما متقنة (Masters)، والأخرى غير متقنة (Hambleton, 1994).

وتستخدم الاختبارات محكية المرجع في تقديرات أداء الفرد بالنسبة إلى محك، أو مستوى أداء محدد مسبقاً، دون الحاجة إلى مقارنة أدائه بأداء زملائه، أي أن هذه الاختبارات تقيس مستويات، يمكن تفسيرها في ضوء مستويات تتطلب تحديداً دقيقاً للمجال السلوكي، الذي يقيسه الاختبار، وليس مجرد النظر إلى أداء الفرد، مقارنة بأداء مجموعته، كما هو الحال في الاختبارات معيارية المرجع (علام، ٢٠٠٠). أضف إلى ذلك، أنه عند بناء الاختبار محكي المرجع نحتاج إلى درجة عالية من الاتساق بين الهدف والسؤال الذي يقيسه، كما أن نسبة حجم العينة إلى العدد الكلي المحتمل من الفقرات التي تغطي المادة الدراسية تكون في الاختيارات محكية المرجع أعلى منها في الاختبارات معيارية المرجع (عودة، ٢٠٠٥؛ النبهان، ٢٠٠٤).

والهدف الأول لأي نظرية في القياس، هو تقديم أساس لعمل تنبؤات حول السمات، أو القدرات التي يتم قياسها بواسطة فقرات الاختبار (عودة، ٢٠٠٥). وقد كانت النظرية الكلاسيكية في القياس، ولمدة طويلة مستخدمة للوصول إلى هذا الهدف، وفي إطار هذه النظرية، فإن مفهوم القدرة يتم التعبير عنه بالعلامة الحقيقية، وتعرف بأنها توقع العلامة المشاهدة، التي تم الحصول عليها على أساس تطبيق الاختبار على المفحوص عدداً

كبيراً من المرات (Crocker & Algina, 1986)، وهنا تعد النظرية الحديثة في القياس أو نظرية الاستجابة للمفردة (Item Response Theory, IRT) ثورة في القياس، النفسي والتربوي، وذلك كما يعدّها كثير من علماء النفس (علام، ١٩٨٦؛ كاظم والشرقاوي والشيخ وعبد السلام، ١٩٩٦). حيث تقدم هذه النظرية أساساً مختلفاً، إذ تفترض أنه يمكن التنبؤ بأداء الأفراد، أو يمكن تفسير أدائهم في اختبار نفسي، أو عقلي معين في ضوء خاصية، أو خصائص مميزة لهذا الأداء تسمى السمات (Traits)، وبالطبع، يصعب ملاحظة هذه السمات مباشرة، لذلك يجب تقديرها أو الاستدلال عليها من أداء الأفراد على مجموعة من فقرات الاختبار (Embretson & Reise, 2000; Sheng ,2005). وأكد كل من هامبلتون وجونز (Hambleton & Jones, 1994) من خلال دراسة بعنوان: "مقارنة تطبيقات النظرية الكلاسيكية ونظرية الاستجابة للمفردة في تطوير الاختبارات" أنه بشكل عام يوجد أسلوبان إحصائيان لمواجهة مشكلات القياس، مثل: بناء الاختبارات، ومعايرتها، وتحديد الفقرات المتحيزة، وهما: النظرية الكلاسيكية في القياس، والنظرية الحديثة في القياس، أو ما يسمى بنظرية الاستجابة للمفردة (IRT). وأشارت الدراسة إلى التفوق الواضح الذي تظهره نظرية الاستجابة للمفردة في حل المشكلات، التي تعجز عن حلها النظرية الكلاسيكية، مثل القياس التكيفي (Adaptive Measurement)، وبنوك الأسئلة في مفهومها الجديد، وتقديرات القدرة المستقلة عن عينة الفقرات للمفحوصين.

وتقوم نظرية الاستجابة للمفردة على مجموعة من الافتراضات، وهي: أحادية البعد (Unidimensionality)، والاستقلال الموضعي للفقرات (Local Items independence)، والوتيرية (Monotonicity)، والتحرر من السرعة (Hambleton & Swaminathan, 1985). ونتج عن هذه النظرية مجموعة من النماذج، تعرف باسم نماذج السمات الكامنة (Latent Trait Models)، تهدف إلى تحديد العلاقات بين أداء الفرد في الاختبار، و السمة التي تكمن وراء هذا الأداء وتفسّره. وتعد النماذج اللوجستية ذات المعلم الواحد (نموذج راش) والمعلمين، والثلاثة معالم، من أهم النماذج واسعة الانتشار. وتختلف هذه النماذج في عدد معالم الفقرة التي تقدرها، إذ يمثل النموذج اللوجستي، ثلاثي المعلم الشكل العام للنماذج اللوجستية؛ لأنه يضم المعالم الثلاثة الممكنة للفقرة، وهي: الصعوبة، والتمييز، والتخمين(c_i, a_i,b_i) على الترتيب، ويعبر عنه بالعلاقة العامة التالية، التي تقيس احتمال إجابة المفحوص ذي القدرة θ على الفقرة i.

$$P_i(\theta) = C_i + \frac{(1 - C_i)}{1 + \exp[-1.7a_i(\theta - b_i)]} \qquad (i = 1,2,\ldots,n) \qquad (1)$$

حيث: $P_i(\theta)$ احتمال إجابة المفحوص الذي اختير عشوائيا من مستوى القدرة(θ) على الفقرة(i) إجابة صحيحة.

Ci معلم التخمين، و ai معلم تمييز الفقرة، و bi معلم صعوبة الفقرة، و θ قدرة المفحوص.

أما في النموذج اللوجستي ثنائي المعلم: فيفترض أن جميع قيم التخمين للفقرات تقترب من الصفر، لذا يعبر عنه بالعلاقة الآتية:

$$P_i(\theta) = \frac{1}{1 + \exp[-1.7a_i(\theta - b_i)]} \qquad (i = 1,2,\ldots,n) \qquad (2)$$

وفي النموذج اللوجستي ذي المعلم الواحد، يفترض أن جميع قيم التخمين للفقرات تقترب من الصفر، وأن جميع الفقرات تميز بالقدر نفسه بين الأفراد، لكنها تتباين فقط في صعوبتها، فتأخذ العلاقة الرياضية الآتية:

$$P_i(\theta) = \frac{1}{1 + \exp[-1.7(\theta - b_i)]} \qquad (i = 1,2,\ldots,n) \qquad (3)$$

ويعد النموذج اللوجستي، ذو المعلم الواحد أكثر نماذج السمات الكامنة في عدد الافتراضات اللازم توافرها في البيانات (Camilli & Shepared, 1994; Gruijter & kamp, 2005).

وتؤدي دالة معلومات الاختبار دوراً مهماً في نظرية الاستجابة للمفردة، إذ يمكن من خلالها تحديد الخطأ المعياري في التقدير، الذي يرتبط عكسيا مع الجذر التربيعي لدالة المعلومات، كما أن استخدام دالة معلومات الاختبار يتيح لباني الاختبار إنتاج اختبار يتسم بالدقة، من خلال الحصول على تقديرات دقيقة لمعلم القدرة، وخصوصاً بالنسبة للاختبارات محكية المرجع (Gruijter & kamp, 2005). ويشير كل من كروكر، وألجينا، وكذلك هاتي (Crocker & Algina, 1986; Hattie, 1984) إلى أن استخدام عينة كبيرة من المفحوصين يضمن الحصول على دقة كبيرة في تقديرات معلم التمييز.

من ناحية أخرى، يتم تقييم الفقرات وفق نظرية الاستجابة للمفردة، حسب مطابقتها لنموذج، يصف العلاقة بين الاستجابة والقدرة، وذلك باستخدام اختبار إحصائي، أو بتحليل البواقي (Residuals). ويمكن التعرف إلى الفقرات الضعيفة عن طريق مؤشرات التمييز (قيمة معلم التمييز (ai) تكون سالبة أو موجبة ضعيفة) وتقديرات مؤشرات الصعوبة (أن لا تكون الفقرات سهلة جداً، أو صعبة جداً لدى مجموعة من المفحوصين)، ومن الأسباب الأخرىً التي تجعل الفقرة ضعيفة عدم تطابق النموذج الذي تم اختياره مع البيانات (Chernyshenko, Stark, Chan, Drasgow & Williams, 2001).

ويشير لورد (Lord, 1977) إلى أنه عند بناء الاختبارات باستخدام نظرية الاستجابة للمفردة، يجب تحديد الخطوط العريضة لعملية بناء الاختبارات، كالتالي: تحديد شكل دالة المعلومات المستهدفة في الدراسة، اختيار فقرات تغطي دوال معلومات مستويات الصعوبة المناظرة لدالة المعلومات المستهدفة، بعد إضافة كل فقرة إلى الاختبار، يتم حساب دالة

معلومات الاختبار التي تم اختيارها، الاستمرار في اختيار الفقرات، وإضافتها حتى تقترب دالة معلومات الاختبار من دالة المعلومات المستهدفة.

وبناء على ما تقدم، يتضح التقدم الكبير الذي أحرزته النظرية الحديثة في القياس، في التحرر من أثر خصائص الأفراد على معالم الفقرات، ومن أثر معالم الفقرات على خصائص الأفراد عند بناء الاختبارات، أي أن الاختبارات المبنية وفق أسس هذه النظرية، والمفاهيم السيكومترية المرتبطة، بها مثل معالم الصعوبة، ومعالم التمييز، ومعالم التخمين، لا تختلف باختلاف خصائص أفراد العينة المستخدمة في حساب هذه المعالم. وتعددت الدراسات التي أجريت حول ذلك (Chernyshenko, et al., 2001; Fan, 1998; Jacobs, 1995).

فقد قام ألبانيز وفورسيث (Albanese & forsyth, 1984) بدراسة كان الغرض منها التحقق فيما إذا كانت الفقرات المستبعدة نتيجة لعدم مطابقتها لأحد نماذج نظرية الاستجابة للمفردة ستؤثر بشكل مختلف في تمثيل الفقرات المتبقية للمجال السلوكي المحدد مسبقاً. تكونت عينة الدراسة من ٩٤٤ مفحوصاً من مستوى الصف التاسع، و ٦٥٠ مفحوصاً من مستوى الصف الثاني عشر، وتم تحليل استجاباتهم على اختبار مكون من خمسة اختبارات فرعية من اختبارات (IOWA) في المهارات الأساسية للقدرات المعرفية. واستخدم الباحثان التحليل العاملي، ومن ثم محك نسبة التباين المفسر للتحقّق من أحادية البعد. كشفت النتائج أن النموذج أحادي المعلم قد أفرز عدداً من الفقرات غير المطابقة للنموذج أكبر من تلك التي افرزها كل من النموذجين الآخرين، حيث كانت نسبة الفقرات غير المطابقة له ٦٣٪، مقابل ٣٩٪ للنموذج ثنائي المعلم، و٢٨٪ للنموذج ثلاثي المعلم. ويعود السبب في ذلك إلى أن النموذج أحادي المعلم يفترض افتراضات يصعب تحقيقها، ومن هذه الافتراضات أن الفقرات متقاربة، أو متساوية في معالم التمييز، وأن تخمين الفقرات يقترب من الصفر.

كما قام التقي (١٩٩٢) بدراسة لفحص مدى تحقق افتراضين أساسيين من افتراضات نظرية الاستجابة للمفردة في أسئلة المقال متعدد الخطوات، وهذان الافتراضان هما: افتراض وجود (لا تغير) في تقدير درجات صعوبة الأسئلة من خلال استعمال عينات مختلفة، والافتراض الثاني هو وجود (لا تغير) في قدرات الأفراد من خلال التقدم لاختبارات تتضمن أسئلة متفاوتة في درجات الصعوبة، وكانت أداة الدراسة عبارة عن اختبار في الرياضيات مكون من عشرين سؤالاً، يقيس قدرة طلاب الصف الأول الثانوي العلمي على حل المعادلات الجبرية والمثلثية في مجهول واحد، وفي مجهولين. أظهرت نتائج الدراسة وجود (لا تغير) (ثبات) في تقديرات درجات صعوبة الأسئلة عند استعمال عينات مختلفة من الأفراد، وكذلك وجود (لا تغير) في تقديرات قدرات الأفراد من خلال التقدم لاختبارات، تتضمن أسئلة متفاوتة في درجات صعوبتها، عند استعمال نموذج التقدير الجزئي، في حين لم يتحقق الافتراضان عند استعمال نموذج سلم التقدير.

وأجرى جاكوبس (Jacobs, 1995) دراسة هدفت إلى دراسة خصائص اختبار كاليفورنيا، (النموذجين A, B). تكونت عينة الدراسة من ١٦٤٦ طالباً وطالبة من طلبة السنة الأولى بالجامعة للعام ١٩٩٣. أظهرت نتائج الدراسة أن هناك اختلافات في صعوبة الفقرة بين النموذجين، وضعفاً في الاتساق الداخلي بين الفقرات، وعدم تحقق افتراض أحادية البعد لفقرات الاختبار، وأن العلامة الكلية على الاختبار تبدو مناسبة وكافية لأغراض بحثية فقط، وليس لأغراض صنع القرارات.

ومن الدراسات التي قارنت بين النظريتين الكلاسيكية والحديثة في القياس، دراسة ثيسن وزيموسكي (Thissen & Zimowski, 1997) التي هدفت إلى مقارنة دقة نظرية الاستجابة للمفردة، مع دقة النظرية الكلاسيكية في القياس في تقديرات علامة المجال لاختبار متعدد الابعاد (في كتابة الكلمات) في الفيزياء، والكيمياء، والأحياء. حيث كانت الكلمة تقرأ ثم تستعمل في جملة، ثم تقرأ ثانية، ويعطى بعدها المفحوص ١٠ ثوان لكتابتها. تكونت عينة الدراسة من تسجيل استجابات ١٠٠٠ طالب من طلبة الدراسات العليا في جامعة كنساس في الولايات المتحدة الامريكية. دلت نتائج الدراسة على تفوق نظرية الاستجابة للمفردة على النظرية الكلاسكية في الدقة في تقدير علامة المجال السلوكي (مجال المحتوى).

كذلك أجرى فان (Fan, 1998) دراسة هدفت إلى مقارنة الخصائص السيكومترية لفقرات اختبار محكي المرجع، وفق النظريتين الكلاسكية، والحديثة في القياس. وتكون الاختبار من اختبارين فرعيين: أحدهما في القراءة، والآخر في الرياضيات، وتألّف اختبار القراءة من ٤٨ فقرة، واختبار الرياضيات من ٦٠ فقرة. تكونت عينة الدراسة من ٨٠ عينة فرعية، سُحبت عشوائياً من مجتمع الدراسة المكون من أكثر من (١٩٣٠٠٠) 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 أحادي المعلم، وكذلك كانت معاملات الارتباط بين تقديرات تمييز الفقرة في النظريتين عالية.

وأجرت الجمحاوي (٢٠٠٠) دراسة هدفت إلى المقارنة بين النظرية الكلاسكية، والنظرية الحديثة في القياس. وكانت أداة الدراسة مقياساً في القدرة الرياضية، مكوناً من ٣٩ فقرة، حيث تمّت مقارنة الخصائص السيكومترية لفقرات هذا المقياس. تكونت عينة الدراسة من (١٠٦١) طالبا وطالبة، من طلبة الصف التاسع في مديريتي إربد الأولى والثانية. أظهرت نتائج الدراسة اتفاقاً عالياً في تقدير صعوبة الفقرات وتمييزها، كما ظهر اتفاقٌ عال بين النظريتين في توزيع قدرة المفحوصين، وكذلك بينت النتائج أن عدد الفقرات التي تمّ اختيارها وفق النظرية الكلاسيكية بلغ ٣٣ فقرة، في حين طابقت ٢٠ فقرة النموذج أحادي المعلم، بينما طابقت ٣٥ فقرة نموذج المعلمين، و٣٨ فقرة نموذج المعالم الثلاثة.

وأجرى بورتون (Burton, 2001) دراسة بعنوان "هل يمكن حقيقة أن تساعدنا مؤشرات

التمييز في تحسين اختباراتنا"؟ قارن فيها بين مؤشرين لإيجاد معامل تمييز الفقرة، وهما مؤشر (U-L): الذي يعتمد على الفرق في نسبة الإجابة الصحيحة على الفقرة في المجموعة العليا والمجموعة الدنيا، ومؤشر (r_{ik}): الذي يمكن إيجاده من خلال معامل الارتباط بين العلامة على الفقرة، والعلامة الكلية على الاختبار. وقد توصل الباحث إلى أن مؤشر التميز (r_{ik}) أكثر ثباتاً من المؤشر (U-L)، في حين أن مؤشر التمييز (U-L) يمكن أن يكون غير ثابت، إذا لم يحسب لعينات كبيرة من المفحوصين، كما أكد الباحث أنه لا بديل عن الحذر، والصياغة الدقيقة لفقرات الاختبار للتقليل ما أمكن من ضعف الثبات لهذه المؤشرات.

أمّا دراسة واغنر وهارفي (Wagner & Harvey, 2003) فقد هدفت إلى تطوير أداة جديدة لأوجه التفكير الناقد لاختبار واتسون–جيلسر باستخدام نظرية الاستجابة للمفردة. حيث قام الباحثان بزيادة عدد البدائل، لتكوين اختبار جديد (WAT). تكونت عينة الدراسة من ٤٠٧ طالب وطالبة في المستوى الجامعي. وقد دلت نتائج التحليل على نسبة تخمين (c_i) أقل، وتحسين في دالة معلومات الفقرة للاختبار المطور، وكانت أخطاء القياس في اختبار واتسون–جليسر أكبر من أخطاء القياس للاختبار الجديد بمقدار ٥٠٪ على الأقل.

ويتبين من خلال عرض الدراسات السابقة، التقدم الكبير الذي أحرزته النظرية الحديثة في القياس عند بناء الاختبارات، حيث أظهرت نتائجها دقة وموضوعية نماذج السمات الكامنة في إعداد الاختبارات وقياس مستوى أداء الأفراد، على الرغم من اختلاف نوعية الاختبارات المستخدمة وأهدافها، دون تسليط الضوء على الاختبارات محكية المرجع في موضوع الإحصاء بالتحديد، ودون التركيز أيضاً على أهمية النموذج اللوجستي ثلاثي المعلم. لذا تتميز الدراسة الحالية عن غيرها من الدراسات، بسعيها لبناء اختبار محكي المرجع في الرياضيات في موضوع الإحصاء، لدى طلبة الصف الأول الثانوي العلمي، من خلال انتقاء فقرات ممثلة لمجال سلوكي محدد، وبخصائص سيكومترية مقبولة، وفق النموذج اللوجستي ثلاثي المعلم لنظرية الاستجابة للمفردة، فهذه المواضيع لم تغطّها الدراسات السابقة بشكل واضح ومفصل، وهو ما حاولت هذه الدراسة القيام به.

مشكلة الدراسة

برز اهتمام واضح من قبل الباحثين بنظرية الاستجابة للمفردة في الآونة الأخيرة، حيث جذبت انتباه العاملين في مجال الاختبارات، واستخدمت في تصميم مختلف أنواع الاختبارات، والمقاييس العقلية، والتحصيلية، ومقاييس الاتجاهات، ومعادلة الاختبارات. ونظراً لأن معالم الفقرات لا تتغير بتغير مجموعة الأفراد التي تختبر بها، فإن بناء الاختبارات وفق نظرية الاستجابة للمفردة (IRT) تمكننا من التمييز بين الأفراد عند مختلف مستويات القدرة (Gruijter & Kamp, 2005; Hambleton & Swaminathan, 1985).

وتعمل المؤسسات المسئولة عن تطوير الاختبارات وبنائها على انتقاء فقرات مناسبة لغرض الاختبار المعدّ، إذ ينبغي أن تغطي هذه الفقرات المحتوى، أو المجال المراد قياسه، وتعمل على إيجاد تقديرات دقيقة لقدرات المفحوصين (Hambleton & Jones, 1994; Thissen & Zimowski, 1997). والواقع الذي يجب أن لا نغفله، هو أن التعليم لم يعد مقتصراً على مجرد التمييز بين الطلاب في القدرات أو المستوى، وإنما يجب أن نركز على مهارات معينة، وتحقيق أهداف محددة، وإتقان تلك المهارات في ظل الحركة المتزايدة في التربية حول التركيز على نتاجات التعلم (النبهان، ٢٠٠٤). من هنا سعت الدراسة الحالية إلى بناء اختبار محكي المرجع في الرياضيات، وفق النموذج اللوجستي، ثلاثي المعلم لنظرية الاستجابة للمفردة، بحيث يتمتع بخصائص سيكومترية مقبولة، ويساعد المربين على اتخاذ قرارات تعليمية تمس حاضر طلبة الصف الأول الثانوي العلمي في موضوع الإحصاء، دون الحاجة إلى مقارنة أداء كل طالب بأداء زملائه.

أسئلة الدراسة

تسعى الدراسة الحالية للإجابة عن الأسئلة الآتية :

١. ما مدى تحقق افتراضات نظرية الاستجابة للمفردة في بيانات الدراسة، وفق النموذج اللوجستي ثلاثي المعلم؟

٢. ما درجة مطابقة الاستجابات عن فقرات الاختبار لتوقعات النموذج اللوجستي ثلاثي المعلم لنظرية الاستجابة للمفردة؟

٣. ما قيم تقديرات معالم فقرات الاختبار (الصعوبة، التمييز، التخمين) وفق النموذج اللوجستي ثلاثي المعلم لنظرية الاستجابة للمفردة؟

أهداف الدراسة

هدفت هذه الدراسة إلى استخدام نظرية الاستجابة للمفردة في بناء اختبار محكي المرجع في الرياضيات وفق النموذج اللوجستي ثلاثي المعلم، بحيث يتمتع هذا الاختبار بدلالات صدق وثبات مقبولة، وبمعايير تتفق مع متطلبات النموذج اللوجستي ثلاثي المعلم؛ من أجل قياس تحصيل طلبة الصف الأول الثانوي العلمي / المستوى الثاني في موضوع الإحصاء.

أهمية الدراسة

تتجلى أهمية الدراسة الحالية في الجوانب الآتية:

١. تبحث في التحقق من التقدم الكبير الذي أحرزته نظرية الاستجابة للمفردة في التحرر من أثر خصائص الأفراد على معالم الفقرات، ومن أثر معالم الفقرات على خصائص الأفراد عند بناء الاختبارات محكية المرجع.

٢. تساهم في تسليط الضوء على ضرورة الاهتمام بالاختبارات محكية المرجع في البحوث التربوية، من خلال بناء اختبار محكي المرجع في الرياضيات في موضوع الإحصاء، لدى طلبة الصف الأول الثانوي العلمي، بانتقاء فقرات ممثلة لمجال سلوكي محدد، وبخصائص سيكومترية مقبولة.

٣. يتوقع أن تزوّد هذه الدراسة معلمي مادّة الرياضيات في وزارة التربية والتعليم باختبار في موضوع الإحصاء، بحيث يمكن من خلاله سحب عينات عشوائية من فقرات المجال الكلي للاختبار، وفحص كل فرد بإحدى هذه العينات، وتقدير قدرته على تدريج مشترك، دون الحاجة إلى مقارنة أداء كل طالب بأداء زملائه.

التعريفات الإجرائية

الاختبار محكي المرجع: هو اختبار يستخدم لتحديد أداء الفرد بالنسبة لمجال سلوكي محدد مسبقاً، ومعرف تعريفاً جيداً.

أحادية البعد (Unidimensionality): وجود قدرة واحدة تفسر أداء الفرد في الاختبار.

الاستقلال الموضعي (Local Independence): أن تكون علامات الفرد مستقلة إحصائياً الواحدة عن الأخرى، عند أي نقطة على متصل السمة.

التحرر من السرعة (Non – Speededness): تفترض نماذج (IRT) أن إخفاق الأفراد في الإجابة عن فقرات الاختبار يرجع إلى انخفاض قدرتهم، وليس إلى تأثير عامل السرعة في الاجابة.

معلم الصعوبة (bi): هو مقدار القدرة التي يكون عندها $P_i(\theta) = \frac{1+C_i}{2}$ حيث (Ci) قيمة معلم التخمين، وتأخذ صعوبة الفقرة في نظرية الاستجابة للمفردة قيماً بين (−∞، +∞) ولكن بشكل عام، فإن القيم التي تاخذها تكون بين (−٣ ، +٣).

معلم التمييز (ai): هو مؤشر يربط بين التغير في القدرات، والتغير في احتمال الإجابة الصحيحة. و(ai) هي ليست الميل، وإنما نسبة منه، ولكن كلما زاد الميل يزيد (ai).

معلم التخمين (ci): هو احتمال إجابة الفقرة إجابة صحيحة من الأفراد ذوي القدرة المتدنية.

محددات الدراسة

تتحدد الدراسة الحالية بما يأتي:

١. اقتصرت عينة الدراسة على طلبة الصف الأول الثانوي العلمي في المدارس الحكومية، التابعة لمديرية التربية والتعليم للواء قصبة المفرق، للعام الدراسي ٢٠٠٦/٢٠٠٧. مما سيحد من تعميم نتائج الدراسة خارج مجتمعها.

٢. اقتصرت أداة الدراسة على منهاج الرياضيات، المقرر تدريسه للصف الأول الثانوي

العلمي (المستوى الثاني) في الأردن، وكانت الفقرات من نوع الاختبار من متعدد بأربعة بدائل في موضوع الإحصاء.

٣. اقتصرت الدراسة على استخدام النموذج اللوجستي، ثلاثي المعلم لنظرية الاستجابة للمفردة، وما تم استخدامه من برمجيات إحصائية متخصصة ومتوفرة.

منهجية الدراسة وإجراءاتها:

منهجية الدراسة

اعتمد الباحث في بناء اختبار هذه الدراسة على خطوات بناء الاختبارات محكية المرجع، وفقاً لافتراضات النموذج اللوجستي، ثلاثي المعلم لنظرية الاستجابة للمفردة.

مجتمع الدراسة

يتكون مجتمع الدراسة من جميع طلبة الصف الأول الثانوي العلمي من الذكور والإناث الملتحقين بالمدارس الحكومية، التابعة لمديرية التربية والتعليم للواء قصبة المفرق في العام الدراسي ٢٠٠٦/٢٠٠٧، والبالغ عددهم ٥٩٣ طالباً وطالبة، منهم ٣٠٤ طالب و ٢٨٩ طالبة، موزعين على ١١ شعبة للذكور، و٩ شعب للإناث، وذلك من خلال الاستعانة بالكراس الإحصائي السنوي في مديرية تربية قصبة المفرق.

عينة الدراسة

تكونت عينة الدراسة من (٤١١) طالباً وطالبة، وقد تم اختيارهم بطريقة عشوائية من طلبة الصف الأول الثانوي العلمي من الذكور والإناث، المنتظمين في المدارس الحكومية التابعة لمديرية التربية والتعليم للواء قصبة المفرق، للعام الدراسي ٢٠٠٦/٢٠٠٧، حيث تم اختيار أربع مدارس، منها مدرستان للذكور، ومدرستان للإناث، وفي كل مدرسة شعبتان على الأقل، وتشكل عينة الدراسة ما نسبته ٦٩٪ تقريباً من المجتمع. ويوضح الجدول رقم (١) توزيع أفراد عينة الدراسة وفق المدرسة، وعدد الشعب، والجنس:

الجدول رقم (١)
توزيع أفراد عينة الدراسة الكلية وفق المدرسة و عدد الشعب و الجنس

اسم المدرسة	عدد الشعب	الجنس	عدد الأفراد
الثانوية الأولى للبنين	٥	ذكور	١٦٣
بلعما الثانوية للبنين	٢	ذكور	٥٥
راية بنت الحسين	٣	إناث	١١٧
الثانوية الأولى للبنات	٢	إناث	٧٦
المجموع	١٢		٤١١

أداة الدراسة

تم بناء اختبار محكي المرجع في مادة الرياضيات، من أجل قياس تحصيل طلبة الصف الأول الثانوي العلمي/المستوى الثاني في موضوع الإحصاء، حيث قام الباحث بصياغة ٤٤ فقرة من نوع الاختيار من متعدد، ولكل فقرة أربعة بدائل، واحد منها يمثل الإجابة الصحيحة. وقد روعيت الشروط الواجب توفراها في مثل هذا النوع من الفقرات، وحرص الباحث على أن تقيس كل فقرة هدفاً محدداً، وفقاً لقائمة الأهداف الخاصة بالوحدة الدراسية، إذ تم تحليلها تحليلاً دقيقاً ومفصلاً، حيث أُعدت قائمة مكونة من ٢٢ هدفاً تفصيلياً شاملة لموضوع الإحصاء، ثم كتابة فقرات الاختبار، وبعد عرضه على هيئة المحكمين، والأخذ باقتراحاتهم، أصبح عدد فقرات الاختبار بصورته الأولية ٣٣ فقرة، وبعد دراسة إجابات العينة الاستطلاعية أثناء عملية التجريب، ودراسة معاملات التمييز للفقرات، حُذفَتْ ٥ فقرات ليتكون الاختبار في صورته النهائية من (٢٨) فقرة، تغطي المجال السلوكي المراد قياسه، والتي تم تطبيقها على العينة المستهدفة في الدراسة الحالية.

صدق الاختبار

تم التأكد من صدق الاختبار باستقصاء مؤشرين من دلالات الصدق، يتعلق المؤشر الأول بصدق المحتوى، وذلك من خلال عرض نتائج تحليل الأهداف، والمحتوى، وجدول المواصفات، والأسئلة على هيئة من المحكمين، ذوي الاختصاص والخبرة، وفي ضوء آرائهم واقتراحاتهم، قام الباحث بإجراء التعديلات اللازمة على فقرات الاختبار إلى أن أصبح جاهزاً للتطبيق. أما المؤشر الثاني فيتعلق بالصدق العاملي على بيانات الاختبار، باستخدام تحليل المكونات الأساسية (Principal Component Analysis)، وحساب قيم الجذر الكامن (Eigen value)، وملاحظة نسبة التباين المفسر(Explained Variance) لكل عامل من العوامل.

ثبات الاختبار

تم تقدير ثبات الاختبار النظري (Theoretical Reliability)، والثبات الإمبريقي/ التجريبي (Empirical Reliability) للاختبار بصورته النهائية وفق النموذج اللوجستي ثلاثي المعلم باستخدام برمجية Bilog–MG3 z، وكذلك تم تقدير ثبات الاتساق الداخلي (كودرد رتشاردسون (KR_20 20))– للاختبار باستخدام برمجية SPSS كما هو مبين في الجدول رقم (٢).

الجدول رقم (٢)
معاملات الثبات النظري والتجريبي لفقرات الاختبار الاتساق الداخلي وفق النموذج اللوجستي ثلاثي المعلم لنظرية الاستجابة للمفردة.

معامل الثبات		
النظري	التجريبي	الاتساق الداخلي
٠,٨٦٥	٠,٩٠٧	٠,٨٥٩

يتبين من الجدول رقم (٢)، أن معامل الثبات التجريبي، الذي يعتمد على كيفية تقدير علامات القدرة، أعلى من معامل الثبات النظري، الذي يعتمد على تقديرات معالم الفقرة للاختبار، وفق النموذج اللوجستي ثلاثي المعلم. كما يلاحظ أن هناك شبه تطابق بين قيمتي معامل الثبات النظري للاختبار، ومعامل الاتساق الداخلي للاختبار نفسه من جهة أخرى. وتشير هذه القيم إلى أن الاختبار يتمتع بدرجة مقبولة في قياسه للسمة موضع الاهتمام.

إجراءات التنفيذ

قام الباحث باتباع الاجراءات الآتية في دراسته:

أ – صياغة الفقرات

– بعد الرجوع الى مقرر الرياضيات (المستوى الثاني) للصف الأول الثانوي العلمي (وحدة الإحصاء)، تم إعداد ٤٤ فقرة من نوع الاختيار من متعدد، بأربعة بدائل، منها بديل واحد فقط صحيح، وقد اقتضى إعداد فقرات الاختبار الاسترشاد بالأسس العامة، المتبعة في بناء الاختبارات محكية المرجع.

– حرص الباحث على أن تقيس كل فقرة هدفاً محدداً، وفقاً لقائمة الأهداف الخاصة بالوحدة الدراسية، إذ تم تحليلها تحليلاً دقيقاً ومفصلاً، وقد أعدت قائمة مكونة من ٢٢ هدفاً تفصيلياً شاملة لموضوع الإحصاء، كما هو مبين في الملحق رقم (١).

– للتأكد من شمول الأهداف للوحدة الدراسية، وتمثيلهما للمستويات المعرفية الثلاثة (معرفة – فهم – تطبيق)، فقد تم عرضهما على مجموعة من المتخصصين في مجال تدريس الرياضيات، وأصحاب الخبرة من المعلمين، لإبداء الرأي حول شمول الأهداف لموضوع الوحدة الدراسية، وطريقة صياغتها.

– بعد تحديد جميع الأهداف التفصيلية للوحدة، تم كتابة مجموعة من الفقرات لكل هدف.

– تم إعداد استبانه للحكم على الاختبار، وطلب من المحكمين إبداء رأيهم في مدى ارتباط الفقرة بالهدف الذي تقيسه، من حيث المحتوى، والمستوى المعرفي، وصياغة الفقرة بلغة واضحة وسليمة ومفهومة، وتناسب لغة الفقرةو مستوى طلبة الصف الأول الثانوي العلمي، وإجابة الفقرة لا تؤثر في إجابة غيرها من فقرات الاختبار، والفقرة تخلو من أية إشارات لفظية للإجابة الصحيحة، ومموهات الفقرة جذّابة ومناسبة للطلبة.

– تم وضع مقياس تدريج أمام كل فقرة، يتدرج من ١–٥ ثم عرضت هذه الاستبانة، مع فقرات الاختبار على هيئة تحكيم من أصحاب الخبرة، تكونت من ٤ أساتذة جامعيين من حملة درجة الدكتوراه، منهم ٢ في مجال تدريس مناهج الرياضيات، و٢ في مجال تدريس القياس والتقويم التربوي، كما شملت هيئة التحكيم ٤ معلمين في مجال تدريس مبحث الرياضيات للمرحلة الثانوية.

– تم الأخذ بجميع آراء الخبراء والمحكمين، حيث قدموا مجموعة من الاقتراحات، ساعدت الباحث في إعادة النظر لعدد من الفقرات التي كانت تحتاج إلى الدقة، أو الوضوح في الصياغة، وبناءً على ملحوظاتهم، تم تغير الكلمات الغامضة في متن السؤال، وتبديلها بحيث يكون لها مدلول واضح حيث تم حذف (١١) فقرة، أجمع المحكمون على أن هناك فقرات أخرى تؤدي غرضها، والإبقاء على (٣٣) فقرة تغطي المجال السلوكي المراد قياسه.

ب – التجريب الأولي

– تمَّ تطبيق الاختبار بصورته الأولية (٣٣) فقرة على عينة استطلاعية من خارج عينة الدراسة بلغ عددها (٨٦) طالباً وطالبة، وقد انحصر الغرض من التطبيق على: تحديد الزمن الذي يستغرقه الطالب في الإجابة عن فقرات الاختبار، والتعرف إلى مستوى صعوبة الفقرات، وقدرتها التمييزية، ومدى وضوح الصياغة اللغوية للفقرات، إضافة إلى دراسة فاعلية المموهات.

– بعد دراسة إجابات العينة الاستطلاعية اثناء عملية التجريب، ودراسة معاملات التمييز للفقرات، تم حذف الفقرات ذات التمييز المنخفض (أقل من ٠,٢)، وبلغ عدد الفقرات المحذوفة (٥) فقرات، وبذلك أصبح عدد فقرات الاختبار بصورته النهائية (٢٨) فقرة، بدلاً من (٤٤) فقرة، وكانت معاملات الصعوبة لها تتراوح بين ٠,٢٢، و٠,٧١ ومعاملات التمييز تتراوح بين ٠,٢١، و٠,٦٧.

ج – جمع البيانات

– بعد أن تمَّ إعداد الاختبار بصورته النهائية، والمكون من ٢٨ فقرة، كما هو مبين في الملحق رقم (٢)، ووضع التعليمات المناسبة لتطبيقه، قام الباحث بالتنسيق المسبق مع مديري ومديرات المدارس للاتفاق على موعد وآليته تطبيق الاختبار.

– تمَّ تطبيق الاختبار على أفراد عينة الدراسة البالغ عددهم (٤١١) طالباً وطالبة في المدارس التي تم تحديدها لعينة الدراسة.

– تمَّ تصحيح الإجابات، وتفريغ البيانات في ذاكرة الحاسوب تمهيداً لمعالجتها الإحصائية.

المعالجة الإحصائية

للإجابة عن أسئلة الدراسة، تم التحقق من افتراض أحادية البعد للاختبار من خلال إجراء تحليل عاملي للبيانات، باستخدام تحليل المكونات الرئيسية بواسطة برمجية (SPSS)، كما

تم التحقق من افتراض حسن المطابقة (goodness–of– Fit–test) لبيانات الدراسة، وفق النموذج اللوجستي ثلاثي المعلم من خلال برمجية (Bilog – MG3) وباستخدام اختبار مربع كاي (χ^2) عند مستوى الدلالة ($\alpha = ٠,٠١$)، كما تم تقدير معالم الفقرات (الصعوبة، التمييز، والتخمين) باستخدام طريقة الأرجحية العظمى الهامشية (MML) من خلال نفس البرمجية.

عرض نتائج الدراسة

أولا: النتائج المتعلقة بالسؤال الأول

نص هذا السؤال على "ما مدى تحقق افتراضات نظرية الاستجابة للمفردة في بيانات الدراسة، وفق النموذج اللوجستي ثلاثي المعلم؟"

أ- افتراض أحادية البعد :(Unidimensionality) تم إجراء تحليل عاملي لبيانات الاختبار النهائي المتعلقة باستجابات (٤١١) فرداً عن (٢٨) فقرة، تمثل الاختبار النهائي، باستخدام البرمجية الإحصائية SPSS، حيث أفرز التحليل ستة عوامل، فسر العامل الأول منها ٢٤,٣٦٪ من التباين، وفسرت جميع العوامل ٤٤,٧٨٪ من التباين الكلي. ويبين الجدول رقم (٣) قيم الجذر الكامن، ونسبة التباين المفسر لكل عامل، وكذلك نسبة التباين المفسر التراكمية.

الجدول رقم (٣)
خلاصة نتائج التحليل العاملي لاستجابة (٤١١) فرداً على الاختبار بصورته النهائية (٢٨) فقرة

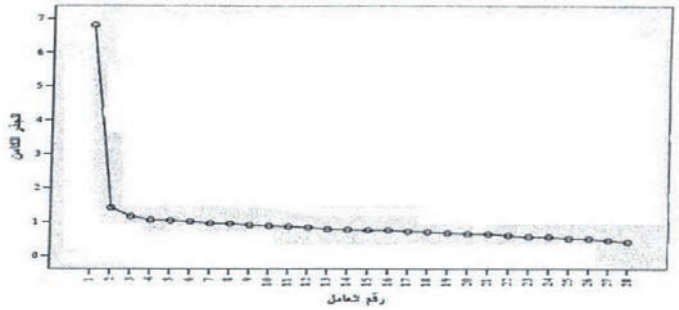
رقم العامل	الجذر الكامن	نسبة التباين المفسر٪	نسبة التباين المفسر التراكمي ٪
١	٦,٨٣	٢٤,٣٦	٢٤,٣٦
٢	١,٤٤	٥,٠٨	٢٩,٤٤
٣	١,١٩	٤,١٨	٣٣,٦٢
٤	١,٠٨	٣,٨٠	٣٧,٤٢
٥	١,٠٦	٣,٧٢	٤١,١٤
٦	١,٠٣	٣,٦٤	٤٤,٧٨

يتبين من الجدول رقم (٣)، أن قيم الجذر الكامن للعامل الأول ٦,٨٣ ويفسر ما نسبته ٢٤,٣٦٪ من التباين الكلي، وبلغت قيمة الجذر الكامن الأول إلى الجذر الكامن الثاني ٤,٧٤. وهذان المؤشران يدلان على تحقق افتراض أحادية البعد للاختبار بصورته النهائية، أي أن الاختبار يقيس سمة واحدة. ويتعزز افتراض أحادية البعد من خلال تمثيل الجذور الكامنة بيانياً باستخدام ما يعرف باختبار فحص العوامل (Scree plot) الذي يظهر في الشكل الآتي.

يتبين من الشكل رقم (١)، أن الجذر الكامن للعامل الأول يتميز بشكل واضح (طاغي) عن الجذور الكامنة لبقية العوامل، وهذا مؤشر أيضاً على أحادية البعد لبيانات الاختبار.

ب- افتراض الاستقلال الموضعي(local Independence): نظراً لأن افتراض أحادية البعد يكافئ افتراض الاستقلال الموضعي، فإن الباحث على -وجه الدقة- اكتفى بالتحقق من

افتراض أحادية البعد، للاستدلال على تحقق افتراض الاستقلال الوضعي(Hambleton & Swaminathan, 1985; Hulin, Drasgow & parson 1983)



الشكل رقم (١)
التمثيل البياني لقيم الجذور الكامنة للعوامل المكونة للاختبارعلى البيانات الكلية

ج – افتراض التحرر من السرعة (Non – Speededness): تأكد الباحث أن إخفاق الأفراد في الإجابة عن فقرات الاختبار يعود إلى انخفاض قدراتهم، وليس إلى تأثير عامل السرعة في الإجابة وذلك عن طريق إعطاء الوقت الكافي لهم، للإجابة عن فقرات الاختبار، بالإضافة إلى أنه لم يشتك أي طالب في أثناء التطبيق من ضيق الوقت، وعدم كفايته.

ثانياً: النتائج المتعلقة بالسؤال الثاني

نص هذا السؤال على "ما درجة مطابقة الاستجابات عن فقرات الاختبار لتوقعات النموذج اللوجستي، ثلاثي المعلم لنظرية الاستجابة للمفردة؟:

تم إدخال البيانات الخاصة باستجابة أفراد عينة الدراسة (٤١١) طالباً وطالبة عن فقرات الاختبار النهائي (٢٨) فقرة، على برمجية (Bilog – MG3) وتم استخراج النتائج وفق المراحل الآتية:

– أفرزت نتائج التحليل عند استخدام البرنامج للمرة الأولى على البيانات الخام، ومن خلال إحصائي مربع كاي (χ^2) عند مستوى الدلالة ($\alpha = ٠,٠١$) عدم مطابقة استجابات (٢٤) فرداً لتوقعات النموذج اللوجستي ثلاثي المعلم، حيث كان مقدار الاحتمالية لبعضهم أقل من (0.01 > 0.01 Fit probability). في حين كان مقدار الخطأ في تقدير القدرة لبعضهم الآخر كبيراً، وهذا ما دلل عليه تعذر البرنامج في حساب الخطأ المعياري لقدراتهم من خلال إعطاء القيمة (٩٩٩,٠٠٠*) كمؤشر على ذلك، لذا تم حذف استجاباتهم والإبقاء على اسجابات (٣٨٧) فرداً.

– ثم أعيد التحليل لاختبار مدى مطابقة فقرات الاختبار للنموذج اللوجستي، ثلاثي المعلم المستخدم في الدراسة الحالية، حيث أظهرت نتائج التحليل في المرة الثانية الخاصة بمطابقة

الفقرات من خلال اختبار مربع كاي، عند مستوى الدلالة (α = ٠,٠١) عدم مطابقة (٤) فقرات للنموذج، حيث كانت قيمة احتمالية المطابقة لكل منها أقل من ٠,٠١ وهي الفقرات ذات الأرقام التسلسلية (٥، ١٧، ٢٣، ٢٤).

– بعد حذف الفقرات الأربع غير المطابقة للنموذج اللوجستي ثلاثي المعلم في المرحلة السابقة، ثم إعادة التحليل للمرة الثالثة؛ للحصول على تقديرات نهائية لكل من معالم الفقرات وقدرات الأفراد على النموذج المستخدم. وهكذا تم الحصول على اختبار محكي المرجع لدى طلبة الصف الأول الثانوي العلمي في الرياضيات (وحدة الإحصاء) وفق النموذج اللوجستي ثلاثي المعلم لنظرية الاستجابة للمفردة، ويتكون بصورته النهائية المعتمدة من (٢٤) فقرة، حيث شكلت فقرات الاختبار المعتمدة ٨٦٪ من فقرات الاختبار النهائي (٢٨) فقرة.

ثالثاً: النتائج المتعلقة بالسؤال الثالث

نص هذا السؤال على:"ما قيم تقديرات معالم فقرات الاختبار (الصعوبة، التمييز، التخمين) وفق النموذج اللوجستي ثلاثي المعلم لنظرية الاستجابة للمفردة؟"

– تم تقدير معالم الفقرة (الصعوبة، التمييز، التخمين) والخطأ المعياري لتقدير المعالم بطريقة الأرجحية العظمى الهامشية (MML) من خلال برمجية Bilog – MG3 والتي تعطي تقديرات دقيقة عن طريق إعادة متعاقبة لعمليات التقدير. ويبين الجدول رقم (٤) قيم تقديرات هذه المعالم، والخطأ المعياري للتقدير لفقرات الاختبار بصورته النهائية المعتمدة، ٢٤ فقرة (بعد حذف الفقرات غير المطابقة)، وفق النموذج اللوجستي ثلاثي المعلم لنظرية الاستجابة للمفردة.

الجدول رقم (٤)
قيم تقديرات معالم الفقرة، والخطأ المعياري لتقدير المعالم للاختبار بصورته النهائية المعتمدة (٢٤) فقرة، وفق النموذج اللوجستي ثلاثي المعلم

رقم الفقرة في الاختبار	معلم الصعوبة	الخطأ المعياري لمعلم الصعوبة	معلم التمييز	الخطأ المعياري لمعلم التمييز	معلم التخمين	الخطأ المعياري لمعلم التخمين	رقم الفقرة في الاختبار	معلم الصعوبة	الخطأ المعياري لمعلم الصعوبة	معلم التمييز	الخطأ المعياري لمعلم التمييز	معلم التخمين	الخطأ المعياري لمعلم التخمين
١	-٠,١٢٦	٠,٢٩١	٠,٥٥٠	٠,٠٨٦	٠,٢٤١	٠,٠٨٣	١٤	٠,١٦٠	٠,٢٥١	٠,٦٢٣	٠,١١٨	٠,٤٢٩	٠,٠٦٥
٢	١,١١٦	٠,١٠٢	١,٢٣٠	٠,٢٤٣	٠,٣٣٧	٠,٠٢٧	١٥	٠,٠٨١	٠,٢٧٠	٠,٤٢٣	٠,٠٧٢	٠,٢٤٤	٠,٠٨٤
٣	١,٢٨٢	٠,١٠٣	١,١٧٧	٠,٢٢٨	٠,٢٧٩	٠,٠٢٦	١٦	١,٠٠٥	٠,١٥٨	٠,٨٤٢	٠,١٩٤	٠,٣٥٥	٠,٠٤٣
٤	٠,١٥٠	٠,٢٣٦	٠,٥٥٨	٠,٠٨٠	٠,١٩١	٠,٠٦٩	١٨	٠,٦٧٤	٠,١٠٣	١,٠٥٣	٠,١٧٥	٠,٢٠٧	٠,٠٣٧
٦	١,٣٦١	٠,١١٧	٠,٩٩٣	٠,٢١٤	٠,٢٤٤	٠,٠٢٩	١٩	٠,٣١٩	٠,٢٥٨	٠,٥٥٢	٠,٠٩١	٠,٢١٦	٠,٠٧٣
٧	١,١٢١	٠,١٥٥	٠,٧١٤	٠,١٤٢	٠,٢٢٠	٠,٠٤٥	٢٠	٠,٩٧٣	٠,١١٨	٠,٨٩١	٠,١٥٢	٠,٢١٤	٠,٠٣٧
٨	١,٥٠٠	٠,١٢٣	٠,٨٧٤	٠,١٧٤	٠,١٧٨	٠,٠٢٨	٢١	-٠,٠٢٢	٠,٣٩٧	٠,٤١٢	٠,٠٦٨	٠,٢٥٩	٠,٠٨٨
٩	٠,٤٤٨	٠,١٦٢	٠,٦٥٦	٠,٠٩٠	٠,١٤٨	٠,٠٥٢	٢٢	٠,٧٠٧	٠,٢٥١	٠,٥١٨	٠,٠٩٣	٠,٢٠٨	٠,٠٦٧
١٠	٠,٩٠٤	٠,١٦١	٠,٧٥١	٠,١٤٢	٠,٢٥٦	٠,٠٤٧	٢٥	٠,٩١٢	٠,١١٧	٠,٩٢٠	٠,١٦٧	٠,٢٣٤	٠,٠٣٨
١١	٠,٢٧٢	٠,١٤٨	٠,٩٧٣	٠,١٥٧	٠,٢٥٥	٠,٠٥٣	٢٦	١,٥٠٨	٠,١٠٥	١,٢٩٣	٠,٢٨٨	٠,٢٥٥	٠,٠٢٢
١٢	١,١١١	٠,١٣٥	٠,٨٤٧	٠,١٩١	٠,٢٩٢	٠,٠٣٥	٢٧	١,٥٥٦	٠,١٤٨	٠,٨٩٨	٠,٢١٣	٠,٢٩١	٠,٠٣٠
١٣	١,٠٧٤	٠,٠٧٩	١,٢٨٩	٠,٢٩٣	٠,٤٧٣	٠,٠٢٧	٢٨	٠,٦١٣	٠,٢٠٦	٠,٧٣٨	٠,١٥٦	٠,٣١٨	٠,٠٦١

يتبين من الجدول رقم (٤)، أن قيم معلم الصعوبة تتراوح بين ٠,١٢٦-و١,٥٥٦ بمتوسط حسابي ٠,٧٩١، وتراوحت قيم الخطأ المعياري لتقدير معلم الصعوبة بين ٠,٠٧٩ و٠,٣٩٧، في حين تراوحت قيم معلم التمييز بين ٠,٤١٢، و١,٢٩٣ بمتوسط حسابي ٠,٨٢٤، وتراوحت قيم الخطأ المعياري لتقدير معلم التمييز بين ٠,٠٦٨ و٠,٢٩٣، وتراوحت قيم معلم التخمين بين ٠,١٤٨ و٠,٤٧٣ بمتوسط حسابي ٠,٢٦٤، وتراوحت قيم الخطأ المعياري لتقدير معلم التخمين بين ٠,٠٢٢ و٠,٠٨٨. ويتبين من قيم المدى لمعلم الصعوبة لفقرات الاختبار أن النموذج اللوجستي، ثلاثي المعلم يتمتع بمدى كبير، حيث يتسم هذا الاختبار، حسب ما ورد عن امبرتسون و رايس (Embretson & Reise ,2000) بأن منحنى توزيع العلامات الخاص به منبسط نسبياً (Flat) على مدى متصل السمة، وهذا يعني أن مجموعة من الفقرات ذات التمييز العالي وبمدى واسع من معلم الصعوبة. وأشار ري (Reee) (Cited in Hattie, 1984) إلى أن القيم النموذجية لمعلم التمييز، وفق النموذج اللوجستي ثلاثي المعلم، تتراوح بين ٠,٥ و٢,٥ بمتوسط حسابي يساوي (واحد صحيح)، ويضيف ري أن القيم التي تكون أكبر من القيمة ٢ من النادر الحصول عليها. ويمكن القول وبشكل تقريبي إنّ معالم النموذج اللوجستي، ثلاثي المعلم تتمتع باتساق (Consistency) معقول، وفق ما ذكر هامبلتون (Hambleton, 1989) بأنه عندما تكون قيم معلم الصعوبة تتراوح بين -٢ و٢ والتمييز بين ٠,٤ و٢ والتخمين بين ٠ و٠,٢٥ فإن المعالم متسقة. وتم تقدير معلم القدرة أيضاً باستخدام طريقة الأرجحية العظمى الهامشية (MML) من خلال برمجية Bilog – MG3، وأظهرت نتائج التحليل تناقص قيم الخطأ المعياري كلما انتقلنا من مستويات القدرة الدنيا إلى المستويات العليا، وكانت أول قيمة لدالة المعلومات قدمها الاختبار وفق النموذج اللوجستي، ثلاثي المعلم هي ٠,٢٨٤ عند القدرة ٢,٢١، بدأت بعد ذلك هذه الكمية بالارتفاع التدريجي، إلى أن وصلت أعلى قيمة ١٣,٦٥١ المقابلة للقدرة ١,٨٥، ثم بدأت القيم بالانخفاض التدريجي إلى أن وصلت ٧,١٨٠ عند مستوى قدرة ٢,٢٣، كما تركزت التكرارات تقريباً في وسط التوزيع للاختبار بصورته النهائية (٢٤) فقرة.

مناقشة النتائج

بالنسبة لمناقشة النتائج المتعلقة بمدى تحقق افتراضات نظرية الاستجابة للمفردة في بيانات الدراسة، وفق النموذج اللوجستي، ثلاثي المعلم، بينت النتائج في الجدول رقم (٣)، والمتعلقة بالتحليل العاملي للبيانات الكلية لاستجابات (٤١١) فرداً عن ٢٨ فقرة، ظهور ستة عوامل، فسر العامل الأول منها (٢٤,٣٦٪) من التباين، وفسرت جميع العوامل ما قيمته (٤٤,٧٨٪) من التباين الكلي، وهذا مؤشر على تحقق افتراض أحادية البعد، وفق المؤشرات المعتمدة على المكونات الأساسية (principal component)، حيث أشار عدد من الباحثين،

منهم ريكيس (Cited in Hattie, 1985) إلى أنه إذا فسر العامل الأول أكبر تباين، فإن هذا مؤشر على أحادية البعد، وحدد ريكيس تلك القيمة بأن تكون على الأقل ٢٠٪ من التباين.

وبالاعتماد على مؤشر لورد (Lord, 1980) الخاص بأحادية، البعد والذي ينص على أن الفقرات تحقق أحادية البعد: إذا كانت قيمة الجذر الكامن الأول كبيرة، بالمقارنة مع قيمة الجذر الكامن الثاني، وتكون قيمة الجذر الكامن الثاني قريبة من قيمة بقية الجذور الكامنة التي تليه، وهذا ما تم التوصل إليه وملاحظته في الجدول رقم (٣).

وأكد استخدام فحص العوامل البياني (Scree plot) لعدد العوامل الناتجة مع قيمة الجذر الكامن للعوامل افتراض أحادية البعد لفقرات الاختبار، من خلال وجود عامل واحد تميز بشكل واضح (طاغي) عن الجذور الأخرى، والمبين في الشكل رقم (١). وبذلك يكون الباحث قد تأكد من تحقق افتراض أحادية البعد لاستجابات (٤١١) فردا، عن ٢٨ فقرة، والعائدة لبيانات الاختبار بصورته النهائية.

كما أظهرت القيمة العالية لمعامل ثبات الاتساق الداخلي (KR_20) للاختبار بصورته النهائية، مؤشرا لتحقق افتراض أحادية البعد، وفقا لرأي كرونباخ (Cronbach, cited in Hattie, 1985) إذ يرى أن معامل الثبات يعد مؤشراً جيداً لتحقق أحادية البعد، لأنه يمثل متوسط كل من المعاملات النصفية الممكنة (Spilt-half Coefficients)، بالإضافة إلى أن (KR_20) تمثل القيمة المتوقعة بالنسبة للتباين المفسر من العوامل المشتركة (Common Factors) بين الفقرات عند ارتباط عينتين عشوائيتين من تجمع الفقرات الكلي.

أما فيما يتعلق بمناقشة النتائج المتعلقة بمطابقة الاستجابات عن فقرات الاختبار لتوقعات النموذج اللوجستي ثلاثي المعلم، وفق نظرية الاستجابة للمفردة، وفي ضوء نتائج تحليل استجابات (٤١١) فرد عن (٢٨) فقرة باستخدام برمجية Bilog – MG3، فقد تم حذف استجابات ٢٤ فرداً من أفراد العينة الكلية، لم تتطابق استجاباتهم وتوقعات النموذج. ويفسر الباحث عدم هذا التطابق إلى عدم جدية بعض الطلبة، وميلهم إلى الإجابة عن الفقرات بشكل عشوائي، ويمكن أن يكون السبب هو ميل بعض الأفراد من ذوي القدرات المتدنية إلى التخمين عن الأسئلة، التي قيم معلم الصعوبة لها يفوق قدراتهم (Jacobs, 1995). وفيما يتعلق بالطلبة ذوي العلامات العالية، ففإن الباحث يفسر سبب حذف البرنامج لاستجاباتهم إلى إخفاقهم في التعامل مع الفقرات، التي هي دون مستوى قدراتهم، وهذا ما أشار إليه ميجر (Meijer, 1996) بأن الأفراد ذوي القدرات العالية جداً عادة ما يخفقون في الإجابة عن الفقرات السهلة، وذلك لأنهم يفكرون بأبعد مما تعنيه الفقرة، أو أنهم عادة يميلون إلى تفسير مدلولات الفقرة بأكثر مما تحتمل.

وعلى العموم، يمكن القول إنّ غالبية أفراد الدراسة الحالية طابقت استجاباتهم توقعات النموذج اللوجستي ثلاثي المعلم بنسبة (٣٨٧/٤١١) = ٩٤٪ من أفراد العينة الكلية. وتتفق

نتائج الدراسة الحالية مع نتائج دراسة كلكار، ووايت مان، و لويخت (Kelkar, Wightman & Luecht, 2000) الذين أشاروا إلى أن المعلمين الإضافيين (التمييز والتخمين) في النموذج اللوجستي ثلاثي المعلم، يتيحان للمفحوص فرصة أعلى لمطابقة البيانات الخاصة للاستجابات لفقرات اختبار من نوع الاختيار من المتعدد.

أما بالنسبة للفقرات الأربع التي لم تطابق توقعات النموذج اللوجستي ثلاثي المعلم، فعند دراسة معالم هذه الفقرات، التي حذفت، تبين أن معلم التخمين لكل منها كان أعلى من القيمة المطلوبة في القياس التربوي، حيث تراوحت هذه القيم بين ٠,٣٢٤، و٠,٧٤٥، علماً بأن المتوسط الحسابي لمعلم التخمين للفقرات المتبقية هو ٠,٢٦٤، وهو يحدد الحد الأعلى لمعلم التخمين لقيم معلم التخمين المقبول لفقرات اختبار من نوع الاختيار من متعدد بأربعة بدائل، لكل منها بديل واحد فقط صحيح (Baker, 2001). ويعزز تلك القيم أيضاً قيم معامل الصعوبة للفقرات المحذوفة لهذا النموذج، حيث تجاوزت قيم معلم الصعوبة الواحد الصحيح. وامتازت بعض الفقرات الأربع المحذوفة أيضاً، بقيم عالية نوعاً ما لمعالم التمييز الخاصة بها، حيث كانت قيمة معلم التمييز لبعض الفقرات المحذوفة ١,٧٦٢، ويشير بيكر (Baker, 2001) إلى أن قيم التمييز التي تتجاوز ١,٧ تعد عالية جداً.

وبالنسبة لمناقشة النتائج المتعلقة بتقديرات معالم فقرات الاختبار (الصعوبة، التمييز، التخمين)، وفق النموذج اللوجستي، ثلاثي المعلم لنظرية الاستجابة للمفردة، وإذا ما ربطنا بين دقة تقدير معلم الصعوبة مع طول الاختبار وعدد المفحوصين، فيمكن القول إنّ هذا المعلم يتمتع بدقة مقبولة، وذلك وفقاً لما ورد عن سواميناثان وجيفورد (Swaminathan & Gifford) (Cited in Hulin, et al., 1983)حيث أشارت نتائج دراستهما إلى وجود ارتباطات عالية بين قيم الصعوبة الحقيقية، والتقديرية لتوزيعات ملتوية القدرة بلغت ٠,٨ في حالة اختبار مكون من عشر فقرات، وعدد المفحوصين ٥٠، ووصلت هذه القيمة إلى ٠,٩٥ في حالة عينة كبيرة من المفحوصين عددها ١٠٠٠، واختبار طوله ٨٠ فقرة. وعند مقارنة نتائج الدراسة الحالية للنموذج اللوجستي، ثلاثي المعلم مع المعايير التي أوردها كل من امبرتسون و رايس (Embretson & Reise, 2000)، و ري (Ree, cited in Hattie, 1984)، وكذلك هامبلتون (Hambleton, 1989) يتبين بأن هناك اتساقاً بين النتائج، إذ إنّ قيم معلم الصعوبة تتراوح بين – ٠,١٢٦ و١,٥٥٦ وقيم معلم التمييز بين ٠,٤١٢ و١,٢٩٣ و قيم معلم التخمين بين ٠,١٤٨ و٠,٤٧٣، كما تبين عند تتبع الأخطاء المعيارية في التقدير لمعلم الصعوبة، والمبينة في الجدول رقم (٤) أن الخطأ الأعلى ارتبط مع الفقرات متوسطة الصعوبة، فيما ارتبط الخطأ الأدنى مع الفقرات الصعبة. وعند مقارنة نتائج سواميناثان وجيفورد مع نتائج الدراسة الحالية، تبين أن توزيع القدرة يتبع التوزيع الاعتدالي، مما يشير الى أن القيم التي أفرزتها بيانات الدراسة الحالية تتمتع بدقة مقبولة .

وتتفق نتائج الدراسة الحالية لقيم معلم التخمين مع القيم التي حصل عليها واغنر، وهارفي (Wagner & Harvey, 2003) حيث حصلا على تقديرات لمعلم التخمين، بلغت قيمة المتوسط الحسابي لها التقديرات ٠,٢٦٢ وهكذا يرى الباحث أن تقديرات معالم الفقرات التي حصل عليها للنموذج اللوجستي ثلاثي المعلم في هذه الدراسة مقبولة ضمن محكات القياس التربوي والنفسي.

وبشكل عام، فقد خرج الباحث باختبار تحصيلي في مبحث الرياضيات لطلبة الصف الأول الثانوي العلمي في موضوع الإحصاء، مبني وفق نظرية الاستجابة للمفردة، ومكوّن في صورته النهائية من ٢٤ فقرة (بعد حذف الفقرات الأربع غير المطابقة للنموذج) تقيس المهارات الأساسية، التي بنى الباحث الاختبار على أساسها، وقد طابقت هذه الفقرات النموذج اللوجستي ثلاثي المعلم لنظرية الاستجابة للمفردة، وتتمتع بخصائص سيكومترية مقبولة ضمن المحكات التي أوردتها أدبيات القياس التربوي.

التوصيات

في ضوء النتائج التي توصلت إليها الدراسة الحالية، يوصي الباحث بما يلي:

١. استخدام الاختبار الذي تم بناؤه وفق نظرية الاستجابة للمفردة، والمكون بصورته النهائية المعتمدة من (٢٤) فقرة، أو أي اختبار مستل منه، لقياس تحصيل الطالب في الصف الأول الثانوي العلمي في الإحصاء؛ نظراً لتمتع الاختبار المبني بدلالات صدق وثبات مقبولة، وبمعايير تتفق مع متطلبات النموذج اللوجستي ثلاثي المعلم لنظرية الاستجابة للمفردة.

٢. إجراء المزيد من الدراسات حول الخصائص السيكومترية للاختبار الذي تم بناؤه، باستخدام نماذج السمات الكامنة الأخرى لنظرية الاستجابة للمفردة، قبل اعتماده كأداة لقياس تحصيل الطالب في الإحصاء.

المراجع

التقي، أحمد. (١٩٩٢). **اللاتغير في تقدير معالم قدرات الأفراد ودرجات صعوبة أسئلة المقال من خلال نموذجي التقدير الجزئي وسلم التقدير كحالتين خاصتين من نماذج راش**. رسالة دكتوراه غير منشورة، الجامعة الأردنية، عمان: الأردن.

جمحاوي، إيناس. (٢٠٠٠). **مقارنة خصائص الفقرات وفق النظرية التقليدية ونظرية الاستجابة للمفردة في مقياس للقدرة الرياضية**. رسالة ماجستير غير منشورة، جامعة اليرموك، اربد: الأردن.

علام، صلاح الدين. (١٩٨٦). **تطورات معاصرة في القياس النفسي والتربوي**. الكويت: كلية الآداب

علام، صلاح الدين. (٢٠٠٠). **القياس والتقويم التربوي والنفسي**. القاهرة: دار الفكر العربي.

عودة، أحمد. (٢٠٠٥). **القياس والتقويم في العملية التدريسية**. عمان: دار الأمل.

كاظم، أمينة والشرقاوي، أنور والشيخ، سليمان وعبدالسلام، نادية. (١٩٩٦). **اتجاهات معاصرة في القياس والتقويم النفسي والتربوي**. القاهرة: مكتبة الأنجلو المصرية.

النبهان، موسى. (٢٠٠٤). **أساسيات القياس في العلوم السلوكية**. عمان: دار الشروق للنشر والتوزيع.

Albanese, M., & Forsyth, R. (1984). The one-, two- and modified two parameter latent trait models: An empirical study of relative fit. **Educational and Psychological Measurement, 44**(2), 229-246.

Baker, F. (2001). **The basics of item response theory** (2nd Ed). Retrieved October 10, 2002, from: www.ericate.net/irt.

Burton, R. (2001). Do item-discrimination indices really help us to improve our tests? **Assessments & Evaluation in Higher Education, 26**(3), 213-220.

Camilli, G., & Shepared, L. (1994). **Methods for identifying biased test items.** Thousand Oaks, CA: Sage Publications.

Chernyshenko, O., Stark, S., Chan, K., Drasgow, F., & Williams, B. (2001). Fitting item response theory models to two personality inventories: Issues and insights. **Multivariate Behavioral Research, 36**(4), 523-562.

Crocker, L., & Algina, J. (1986). **Introduction to classical and modern test theory.** New York: CBS College Publishing.

Embretson, S., & Reise, S. (2000). **Item response theory for psychologists.** New Jersey: 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), 357 – 381.

Gruijter, D., & Kamp, L. (2005). **Statistical test theory for education and psychology.** Retrieved December 30, 2005 from: www.leidenuniv.nl /~ gruijterdnmde.

Hambleton, R. (1989). **Principles and selected applications of item response theory.** New York: Macmillan Publishing Company.

Hambleton, R. (1994). The rise and fall of criterion – referenced measurement. **Educational Measurement, 13**(4), 21 – 26.

Hambleton, R., & Swaminathan, H. (1985). **Item response theory: Principles and applications.** Boston: Kluwer Nijhoff Publishing.

Hambleton, R., & Jones, R. (1994). Comparison of classical test theory and item response theory and their applications to test development. **Educational Measurement, 31**(6), 38 – 47.

Hattie, J. (1984). An empirical study of various indices for determining unidimensionality. **Multivariate Behavioral Research, 19**(1), 49-78.

Hattie, J. (1985). Methodology review: Assessing unidimensionality of tests and items. **Applied Psychological Measurement, 9**(2), 139 – 164.

Hulin, C., Drasgow, F., & Parson, K. (1983). **Item response theory: Applications to psychological measurement.** Homewood, Illinois: Dow Jones – Irwin.

Jacobs, S. (1995). Technical characteristics and some correlates of the California Critical Thinking Skills test, forms A and B. **Research in higher education, 36**(1), 89-108.

Kelkar, V., Wightman, L., & Luecht, R. (2000). **Evaluation of the IRT parameter invariance property for the MCAT.** Paper presented at the annual meeting of the National Council on measurement in education. New Orleans: Medical College Admission Test.

Lord, F. (1977). Practical applications of item characteristic curve theory. **Journal of Educational measurement, 14,** 117 -138.

Lord, F. (1980). **Applications of item response theory to practical testing problems.** Hillsdale, N .J: Lawrence Erlbaum.

Meijer, R. (1996). Person – Fit research: An introduction. **Applied Measurement in Education, 9**(1), 3–8.

Sheng, Y. (2005). **Bayesian analysis of hierarchical IRT models: Comparing and combining the unidimensional & multi-unidimensional IRT models.** Unpublished doctoral dissertation, University of Missouri, Columbia.

Thissen, D., & Zimowski, M. (1997). IRT estimation of domain scores .**Journal of Educational Measurement, 34**(3), 179–211.

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. Orlando. Retrieved May,4,2005, from www.criticalthinking.org.