فاعلية النموذج اللوجستي ثلاثي المعلمة في معايرة مفردات اختبار تحصيلي محكي المرجع في مقرر الرياضيات للصف السابع

تاريخ الاستلام: 2014/6/8 تاريخ القبول: 2015/1/21

د. حمدي أبو جراد(*)

الملخص

هدفت الدراسة الحالية إلى تعرف مدى فاعلية النموذج اللوجستي ثلاثي المعلمة في معايرة مفردات اختبار تحصيلي محكي المرجع في مقرر الرياضيات للصف السابع ، ولتحقيق هدف الدراسة، تم بناء اختبار تحصيلي في مقرر الرياضيات للصف السابع مؤلف من (56) مفردة من نوع الاختيار من أربعة أبدال، تم تطبيقه على عينة مكونة من (624) طالبا وطالبة، وأشارت النتائج إلى مطابقة استجابات الأفراد عن (44) مفردة للنموذج، وبلغ معامل الثبات للاختبار(0.93)، وتم التحقق من صدق الاختبار بطرق مختلفة، وتم حساب درجة القطع للاختبار وفقاً لأسلوب أنجوف وبلغت(58%)، وبينت النتائج أيضاً أن متوسط أداء الطلبة عن الاختبار أعلى من درجة القطع .

عدد الكلمات (101)

(*)كلية التربية– جامعة القدس المفتوحة

(الكلمات المفتاحية: اختبار محكي المرجع – النموذج اللوجستي ثلاثي المعلمة – درجة القطع – فقرات الاختيار من متعدد - نظرية الاستجابة للمفردة)

Abstract

The present study aimed at identifying the efficiency of Three-Parameter Logistic model in calibrating the items of Criterion-Referenced Achievement Test in 7th grade Math curriculum. To achieve the aim of the study, an achievement test was constructed in 7th grade Math curriculum that consists of 56 multiple choice items with four options for each one. The test was applied on 624 male and female students .The results showed that responses of persons on 44 items fitted the model. The reliability coefficient of the test was 0.93. The validity of the test was proved in different ways. The cut score of the test was calculate according to Angof approach and it was 58%. The results also showed that the performance average of students in the test was higher than the cut score of the test.

(Key words: Criterion- referenced Test- Three-Parameter Logistic model – Cut Score- Multiple –Choice items- Item response theory)

مقدمة

تعتبر الاختبارات من أدوات التقويم المهمة حيث تستخدم نتائجها في اتخاذ القرارات التربوية، وكلما زادت أهمية القرار الذي يتوقع أن يتم اتخاذه، زادت الحاجة إلي أن تكون المعلومات التي نحصل عليها من الاختبار دقيقة وذات صلة بالغرض الذي أعد من أجله .

وتعتبر الدرجات التي يحصل عليها الاختبار درجات خام raw score ،أي ليس لها معنى، وحتى يكون للدرجة معنى على الاختبار يجب أن تفسر بالمقارنة مع بيانات أخرى، كدرجات أداء الطلبة الآخرين في الاختبار نفسه، أو مقارنتها بمعايير أداء محددة(Peers,1996). والأسلوب الذي تفسر في ضوئه العلامات من خلال مقارنة أداء الفرد بأداء أقرانه في المجموعة التي ينتمي إليها يسمى بالقياس معياري المرجع norm reference measurement ،أما إذا تمت مقارنة درجة الفرد بمعايير محددة للأداء حينها يسمى هذا التفسير محكي المرجع criterion reference measurement، وقد أطلق على الاختبارات محكية المرجع عديد من المسميات مثل : اختبارات التمكن proficiency test أو اختبارات الإتقان mastery tests، ومع ظهور مفهوم التعلم من أجل الإتقان لم يعد الهدف هو التركيز على الفروق بين الأفراد، والتمييز بينهم، فظهرت المناداة بالابتعاد عن التوزيع الاعتدالي الذي اعتمدت عليه المقاييس جماعية المرجع في التمييز بين الأفراد في الأداء وذلك لأن النشاط التربوي نشاط هادف ومقصود يبذل بهدف أن يتقن الطلبة ما تعلموه، وبذا تركز الاهتمام حول المستوى الذي يصل إليه أداء الفرد بصرف النظر عن مستوى زملائه . (كاظم ، 1996)

وتمثل الاختبارات محكية المرجع أسلوباً مهماً يساعد في النهوض بمستوى المتعلمين حيث تستخدم درجات الاختبار محكي المرجع الذي يشتمل على عينة عشوائية من

المفردات، التي تمثل النطاق السلوكي في تقدير درجة الفرد في النطاق الشامل لهذه المفردات domain score estimate وتستخدم في تصنيف الأفراد بحسب مستويات تمكنهم أو إتقانهم لهذا النطاق، لذا يتطلب ذلك تحديداً مسبقاً لمستويات الأداء في الاختبار (درجات قطع cut off score) تفصل بين المتقنين وغير المتقنين من الأفراد (علام ،1995) .

ويربط المتخصصون في القياس(Goodwin , 1996 ; Jager,1989 ; Shepard,1984 ; Kane,1987) بين درجات القطع (cut off –scores) ، والمستويات(standards) عند الحديث عن المعايير، حيث يتضح ذلك جلياً عند الحديث عن الاختبارات محكية المرجع مقابل الاختبارات معيارية المرجع، أو عند الحديث عن الاختبارات المفصلة (tailored testing) أو اختبارات التمكن (Mastery tests) .

ويطلق على مستويات الأداء في الاختبارات محكية المرجع تسميات مختلفة مثل: درجات القطع cutting scores ودرجات التمكن Mastery Scores ، ودرجات النجاح Passing Scores(علام،2001؛ Cizek,1996)

ودرجة القطع هي الدرجة التي ينبغي أن يحصل عليها الفرد في المجال الشامل Universe specifications لمفردات الاختبار لكي يكون متقناً لمحتوى أو مهارة معينة ، وأحياناً تشير درجة القطع إلى الحد الأدنى للأداء المقبول لكي يتمكن الفرد من أداء مهام تالية (Guskey, 1988) .

واعتبر كين (Kane, 1998) أن درجة القطع نقطة على مقياس العلامة الملاحظة تفصل بين الأداء المقبول وغير المقبول. ويعرفها هامبلتون(Hambleton, 1982)

بأنها درجة على متصل السمة المقيسة يصنف في ضوئها المفحوصون إلى فئتين: إحداها متمكنة، والأخرى غير متمكنة من الأهداف التي يقيسها الاختبار.

إنَ عملية انتقاء مفردات اختبارات محكية المرجع يجب أن تتم بدقة لكي تكون عينة مفردات الاختبار ممثلة للنطاق السلوكي الشامل، حيث يتم اتخاذ قرار تصنيف الفرد على أساس الدرجة التي يحصل عليها في هذا النطاق السلوكي الشامل والمحدد تحديداً دقيقاً. كما ينبغي أن يضم الاختبار مفردات تميز بشكل مناسب بين الأفراد عند المستوى المعين الذي يتم اختياره محكّاً للأداء، حتى ينجح الاختبار في التمييز بين الأفراد تبعاً لهذا المستوى (Hambelton & Gruijter, 1983). ويعتبر نتيكو (Nitko, 2001) أنَ تفسير تقدير الأفراد على القياس محكي المرجع يفقد صدقه عندما لا يتم تعريف مجال الأداءات Domain of performances الذي ينسب إليه أداء الأفراد بدقة، أو عندما تكون عينة المفردات غير ممثلة لهذا المجال.

إن الاعتماد على المؤشرات التقليدية في انتقاء المفردات يشوبه كثير من الانتقادات لخصها هامبلتون وسومنيثان (Hammbelton & swaminthan , 1985) في النقاط التالية :

1- معاملات الصعوبة والتمييز للمفردات تعتمد على عينة المفحوصين التي تطبق عليها الاختبار.

2- تتأثر إحصائيات المفردة بمعدل ومدى القدرة للمفحوصين؛ فقيمة معامل الصعوبة للمفردة يكون أعلى من قيمتها عندما تكون قدرات المفحوصين ذات مستوى أعلى من متوسط القدرة، ومعامل التمييز يكون مرتفعاً، عندما تكون المجموعة غير متجانسة .

3- ثبات الاختبار يعرف بدلالة الصور المتوازية وهذا أمر يصعب تحقيقه من الناحية العملية

4- تفترض نظرية الاختبار أن تباين الخطأ للقياسات متساوٍ لجميع المفحوصين .

لذا جاءت نظرية الاستجابة للمفردة لتلافي تلك العيوب حيث تفترض هذه النظرية أنه يمكن التنبؤ بأداء الأفراد أو تفسير أدائهم في اختبار نفسي أو تربوي في ضوء خصائص تميز هذا الأداء تسمى مسميات أو قدرات وهذه السمات غير ملاحظة، ولكن يمكن قياسها من خلال الأفراد على مفردات الاختبار، وتتميز نظرية الاستجابة للمفردة بأنها تمكننا من:

1- تقدير قدرة المفحوص بحيث تكون مستقلة إحصائياً عن عينة المفردات الخاصة التي طبقت على المفحوصين طالما أنها مفردات مناسبة، وذلك بشرط أن تكون جميع المفردات مدرجة معاً على نفس المتصل بحيث تقيس نفس السمة .

2- الحصول على مؤشرات للمفردة مثل(معامل الصعوبة ، ومعامل التمييز) تكون مستقلة عن عينة المفحوصين الخاصة المستخدمة في معايرة المفردة طالما أنها عينة مناسبة .

3- الحصول على معامل إحصائي يبين مدى دقة تقدير قدرة كل فرد بواسطة مفردات الاختبار، وهذا المعامل يمكن أن يختلف من مفحوص لآخر.

إن تحقيق هذه المميزات الثلاثة يعتمد على مدى الانسجام بين مجموعة بيانات الاختبار والنموذج المستخدم (أى تحقق شروطه)، وإذا لم يتحقق هذا الانسجام أو تحقق بشكل ضعيف فإن هذه المميزات لا يمكن الحصول عليها .

4- استبدال مفهوم الثبات من خلال الصور المتوازية بمفهوم التقدير الإحصائي والأخطاء المعيارية المترافقة Associated standard errors (Hammbelton & Swaminthan, 1985).وتقوم هذه النظرية على مجموعة من الافتراضات :

أولا : أحادية البعد Unidimensionality

ترى كاظم (1994) أن أحادية البعد تعنى: أن مفردات المقياس تكون متجانسة فيما بينها وتقيس في أساسها الصفة نفسها، وهذا يعنى أن أي مفردة من هذه المفردات المتدرجة الصعوبة تتطلب في حلها النوع نفسه من الإجراءات والعمليات السلوكية، ولكنها تختلف فيما بينها من حيث تدرج صعوبتها فقط . أما إذا كانت البنود تقيس أكثر من متغير"قدرة" فيمكن تجميع مفردات كل قدرة في مجموعة متجانسة باستخدام التحليل العاملي ،ثم استخدام أحد نماذج السمات الكامنة مع كل مجموعة متجانسة من المفردات لتدريجها(مراد، وسليمان،2002). ويعتبر هامبلتون وسواميناثان (Hammbelton & Swaminthan ,1985) أن افتراض أحادية البعد شيء مرغوب فيه عند جميع مطوري الاختبارات وذلك من أجل تطوير عملية التفسير لدرجات الاختبار وتعزيزها، إلا أن هذا الافتراض ينتهك في الأبحاث التربوية، حيث توجد عوامل كثيرة تؤثر على أداء المفحوص مثل الشخصية، وعوامل تتعلق بتطبيق الاختبار مثل مستوى الدافعية، وقلق الاختبار، والقدرة على العمل بسرعة، ومعرفة الاستخدام الصحيح لأوراق الإجابة ...الخ ، ولكن ما هو مقصود بذلك الافتراض أن يكون هناك عامل واحد سائد على العوامل الأخرى بحيث يكون هذا العامل هو السمة التي يقيسها الاختبار.

ثانيا : الاستقلال المحلى Local Independence

يقصد بالاستقلال المحلى Local Independence أن استجابة المفحوص على مفردات الاختبار مستقلة إحصائيا عندما يؤخذ مستوى المفحوص بعين الاعتبار ، أي إنّ استجابة المفحوص لمفردة ما يجب ألا يؤثر على مفردة أخرى ، أي إن تقدير صعوبة أي مفردة لا يعتمد على تقديرات صعوبة المفردات الأخرى، ولا يعتمد على قدرة الأفراد الذين يجيبون عنها، وكذلك لا يعتمد على تقدير قدرة أي مجموعة

أخرى من الأفراد الذين طبق عليهم الاختبار، ولا يعتمد على قيم صعوبة المفردات التي يجيبون عنها .

وهذا الافتراض يوضح أن قدرة المفحوص وخصائص المفردة هما وحدهما اللتان تؤثران في الأداء أو الاستجابة . وبتحقق هذا الافتراض فإن احتمال الحصول على أى تسلسل من العلامات لمجموعة من المفردات هو ببساطة حاصل ضرب الاحتمالات لكل هذه المفردات. (Allen &Yen, 1979) . ويعتبر علام (1995) أن تحقق هذا الشرط يجعل للدرجة الكلية للفرد على الاختبار معنى، بحيث يمكن استخدامها في التقدير الإحصائي للسمة المقيسة تقديرا يتسم بالكفاية Sufficient estimate وعندئذ يمكن اعتبار الاختبار يقيس سمة أحادية البعد.

ثالثاً: السرعة Speededness

إن الافتراض الضمني عند جميع مستخدمي نماذج الاستجابة للمفردة هو أن الاختبارات الملائمة للنموذج لا تطبق تحت شرط السرعة، حيث إنّ المفحوصين الذين أخفقوا في الإجابة عن مفردات الاختبار كانت بسبب القدرة المحدودة وليس بسبب فشلهم في عدم الوصول لمفردات الاختبار، وربما لا يعلن عن هذا الافتراض حيث إنه متضمن ضمن افتراض أحادية البعد . فعندما تؤثر السرعة في الأداء على الاختبار فان هناك سمتين– على الأقل– تقاسان بهذا الاختبار هما: سرعة الأداء، والسمة المقيسة بمحتوى الاختبار (Hammbelton & Swaminthan , 1985) .

لقد انبثق عن نظرية الاستجابة للمفردة مجموعة من النماذج التي استخدمت في بناء المقاييس وتطويرها، والاختبارات التي يمكن من خلالها الحصول على مؤشرات إحصائية للمفردة لا تعتمد على خصائص المفحوصين وتقديراتهم، ولا تعتمد على صعوبة مفردات المقياس، ومن هنا أعطيت أهمية اللاتغير(Invariance) في

تقدير معالم المفردات بين مختلف مجموعات الأفراد، ووصفت هذه الأهمية بأنها أكبر الصفات المهمة في نظرية الاستجابة للمفردة (Lord , 1980).

وتختلف هذه النماذج في عدد معالم المفردة التي تقدرها، فمنها الأحادي، والثنائي، والثلاثي الذي ضم معالم المفردة الثلاثة: الصعوبة، والتمييز، والتخمين، و ينسب هذا النموذج إلى برنبوم Birnbaum وتعطى الصورة الرياضية لهذا النموذج من خلال العلاقة التالية:

$$p_i(\theta) = c_i + (1- c_i)/(1+\exp(-1.7a_i\ (\theta - b_i))$$

حيث إنّ $p_i(\theta)$ هى احتمال الاستجابة الصواب للفرد الذي قدرته θ للمفردة i و(b_i) و(a_i) و(c_i) هى معالم المفردات الثلاثة : الصعوبة ، والتمييز، والتخمين على الترتيب (Suen , 1990)، وأطلق على معلم التخمين بارامتر الخط التقاربي الأدنى للمنحنى المميز للمفردة ويحدد هذا البارامتر احتمال أن يجيب فرد يفترض أن مستوى قدرته منخفض انخفاضا لانهائيا ، ومع هذا يجيب إجابة صحيحة عن بعض مفردات الاختبار التي ترتفع في مستوى صعوبتها عن مستوى قدرة هذا الفرد عن طريق التخمين. لذا تبدو أهمية هذا البارامتر عند مطابقة بيانات مستمدة من مجموعات من مفردات اختبارية من نوع الاختيار من متعدد لهذا النموذج (علام ،2002) .

ويلاحظ أن احتمال الاستجابة الصحيحة من خلال هذا النموذج هو$2/(1+c)$ ، وعندما يكون بارامتر التخمين $c = 0$ فإن احتمال الإجابة الصحيحة يكون مساوياً للقيمة 50% ، أما إذا زادت قيمة بارامتر التخمين على الصفر فإن احتمال الإجابة الصحيحة سوف يزيد على 50% (Hammbelton & Swaminthan , 1985).

أما بالنسبة لقيمة معامل التمييز فتتراوح نظرياً بين (0، 1) ، وعملياً تتراوح قيمته بين (0، 0.35).

ونظراً لفاعلية ومميزات نماذج الاستجابة للمفردة في بناء وتصميم الاختبارات، فقد استخدمت في عديد من الدراسات؛ فقد أجرى حمادنة (2009) دراسة هدفت إلي استخدام نظرية الاستجابة للمفردة في بناء اختبار محكي المرجع في الرياضيات وفق النموذج اللوجستي ثلاثي المعلم، ولتحقيق هدف الدراسة تم بناء اختبار في الرياضيات يقيس تحصيل الطالب في الإحصاء مؤلف من (28) مفردة من نوع الاختيار من متعدد من أربعة أبدال، وتكونت عينة الدراسة من (411) طالباً وطالبة من طلبة الصف الأول الثانوي العلمي، وأظهرت النتائج مطابقة الاستجابات لأربع وعشرين (24) مفردة من مفردات الاختبار لتوقعات النموذج اللوجستي ثلاثي المعلم، وحذف(4) مفردات لم تطابق النموذج، كما بينت نتائج الدراسة أن النتائج المتعلقة بتقديرات معالم المفردات(الصعوبة، والتمييز، والتخمين) كانت مقبولة.

وأجرى الشريفان (2006) دراسة هدفت إلى تقدير الخصائص السيكومترية لاختبار محكي المرجع في القياس والتقويم وفق نظرية الاستجابة للمفردة، ولتحقيق هدف الدراسة تم بناء اختبار محكي المرجع في موضوعات القياس والتقويم مكون من (60) مفردة من نوع الاختيار من متعدد، وتم عرض مفردات الاختبار على خمسة محكمين من أعضاء هيئة التدريس في الجامعات والذين يدرسون مساق مبادئ أساسية في القياس والتقويم ،وفي ضوء آراء المحكمين تم حذف (6) مفردات وبلغ عدد مفردات الاختبار في صورته الأولية (54) مفردة وتم تطبيق الاختبار على عينة من طلبة جامعة آل البيت والذين يدرسون مساق القياس والتقويم التربوي في الفصل الدراسي الأول من العام الجامعي 2005/2004 والبالغ عددهم (222) طالباً وطالبة موزعين

على (4) شعب، وتوصلت النتائج إلى مطابقة (37) مفردة لنموذج راش.كما بينت النتائج أن قيمة معامل الثبات المتعلق بكل من الأفراد والمفردات بلغ (0.9 ، 0.95)على الترتيب ، كما تم حساب معامل الارتباط بين درجات الأفراد على الاختبار ومعدلاتهم الجامعية وتبين أنه متدنٍ حيث بلغ (0.51).

وأجرت المهاجري (2006) دراسة هدفت إلى بناء اختبار محكي المرجع لقياس الكفايات الرياضية في حل المعادلات ، والمتباينات من الدرجة الأولى (بمتغير واحد ومتغيرين) لطالبات المرحلة المتوسطة بمكة المكرمة، واستخدام هذا الاختبار لقياس تحصيل الطالبات للكفايات الرياضية في حل المعادلات والمتباينات من الدرجة الأولى (بمتغير واحد ومتغيرين) بالمرحلة المتوسطة لتعرف مدى إتقانهن لهذه الكفايات، وتكون الاختبار من (50) مفردة وتم التحقق من صدق الاختبار من خلال حساب معاملات الاتساق الداخلي والصدق المرتبط بمحك وصدق المحتوى، كما تم تقدير ثبات الاختبار من خلال طريقة الإعادة وطريقة كرونباخ ألفا وطريقة ليفنجستون وطريقة هاريس. وتوصلت نتائج الدراسة إلى انخفاض مستوى طالبات الصف الثالث في حل المعادلات والمتباينات والمسائل الحسابية والأنظمة بشكل واضح .

كما هدفت دراسة عبد الله (2000) إلى بناء صور اختبارية محكية المرجع متعادلة القياس في أحد مقررات علم النفس التعليمي تقيس مدى تحقق الأهداف المعرفية المتعلقة بهذا المقرر من قبل طالبات الفرقة الرابعة من الأقسام العلمية التربوية بكلية البنات جامعة عين شمس، ولتحقيق أهداف الدراسة تم تكوين بنك للأسئلة وتدريج المفردات التي بلغ عددها (120) مفردة بعد تطبيقها على عينة مكونة من (194) طالبة من طالبات الفرقة الرابعة بكلية البنات، موزعات على ثلاث شعب (الرياضيات – التاريخ الطبيعي – الكيمياء والطبيعة)، وأسفرت نتائج الدراسة عن

تكوين صور اختباريه محكية المرجع متعادلة القياس، كما تم تعيين الدرجات الفاصلة المقابلة للتقديرات الجامعية المعتمدة، كما تم تحديد موقع أداء الفرد على مقياس تحصيل مقرر علم النفس التعليمي عند كل محك من محكات الإتقان.

وهدفت دراسة علام (1991) إلى الموازنة بين أربع طرق لتحديد مستويات الأداء في الاختبار، الذي أعده الباحث: اثنتان منها تعدان من الطرق المطلقة ، أي التي تعتمد على تحكيم محتوى الاختبار ومفرداته ، وهما طريقتا أنجوف وندلسكاي ، والطريقتان الأخريان تعدان من الطرق النسبية، أي التي تعتمد على تحكيم أداء المختبرين وجمع بيانات تجريبية فعلية تتعلق بهذا الأداء وهما طريقتا المجموعة الحدية والمجموعة المتناقضة.وتكونت عينة المحكمين من (26) محكماً من ثلاث مجموعات متباينة في مستوى تمكنها في مجال القياس والتقويم التربوي والنفسي؛ اشتملت المجموعة الأولى على (14) معلماً من معلمي المدارس الثانوية العامة الحاصلين على دبلوم خاص في التربية ، أو على دورة تدريبية في مجال القياس والتقويم، واشتملت المجموعة الثانية على ستة مدرسين مساعدين حاصلين على درجة الدكتوراه في علم النفس التعليمي ولديهم خبرة لا تقل عن ثلاثة أعوام في تدريس مادة الاختبارات المقاييس ، واشتملت المجموعة الثالثة على ستة من أساتذة علم النفس التعليمي؛ أما بالنسبة لعينة الطلاب الذين أُجرِي تحكيم أدائهم في الاختبار فكانت (382) طالباً من طلاب كلية التربية ممن كانوا يدرسون مقرراً سيكولوجياً في الفروق الفردية، ويقوم بتدريسهم عينة المحكمين المكونة من الأساتذة والمدرسين المساعدين الذين اختيروا لهذه الدراسة. وتوصلت نتائج الدراسة إلى اختلاف في درجات القطع المتحصل عليها نتيجة لاختلاف الطرق المستخدمة حيث بلغت متوسطاتها (40، 41،43،44) لكل من طريقة ندلسكاى والمجموعة الحدية، وأنجوف والمجموعات المتناقضة على التوالي.

من خلال عرض الدراسات السابقة يتبين أن معظمها اتفق على فعالية نظرية الاستجابة للمفردة وما انبثق عنها من نماذج في عملية بناء الاختبارات والمقاييس، وتطويرها مقارنة بالنظرية التقليدية، كما قدمت نظرية الاستجابة للمفردة عديداً من المفاهيم الإحصائية التي لعبت دوراً مهما في تحقيق دقة القياس وعدالته مثل: دالة المعلومات للمفردة والاختبار، كما اختلفت الدراسات السابقة في العينات المستخدمة؛ فبعض الدراسات السابقة تناولت المرحلة المتوسطة (المهاجري،2006) وبعضها تناول المرحلة الجامعية(علام ،1991؛ عبد الله ،2000 ؛ الشريفين، 2006) ، كما اختلفت الدراسات السابقة في برنامج الحاسب الآلي المستخدم في تحليل البيانات .

مشكلة الدراسة :

يتطلب تقويم مخرجات التعليم تحديد محكات معينة للأداء تتمثل في مستويات قياسية لها مؤشرات قابلة للتحديد والقياس يعتمد عليها في إعداد أدوات مناسبة لقياس أداء الطالب وتقديره تتمثل في الاختبارات المدرسية بأشكالها المختلفة ، وتعتبر هذه المستويات القياسية أو المعيارية هي الإطار المرجعي الذي يتم مزاولته، وتقدر على أساسه مستويات إتقان الطلبة بحيث يمكن من خلالها وصف ما يستطيع أن يتعلمه الطالب من مهارات ومعارف نتيجة لدراسته لمحتوى مجال دراسي معين، ونظراً لعدم توافر أداة قياس موضوعية يمكن من خلالها التحقق من مدى امتلاك الطلبة للمهارات والمعارف حسب المستويات المنشودة، جاءت الدراسة الحالية لتوفير اختبار محكي المرجع في مقرر الرياضيات للصف السابع يتمتع بدرجة عالية من الصدق والثبات، بحيث يعتمد عليه في قياس مدى تحقق الأهداف المعرفية المتعلقة بالمقرر.

أسئلة الدراسة :-

1- ما درجة مطابقة استجابات أفراد عينة الدراسة لمفردات الاختبار لتوقعات النموذج اللوجستي ثلاثي المعلم؟

2- ما دلالات الصدق والثبات المتوافرة لمفردات اختبار محكي المرجع في مقرر الرياضيات للصف السابع والمتحررة من خصائص الأفراد؟

3- ما مستوى أداء الطلبة لهذا الاختبار مقارنة بمستوى التمكن؟

أهمية الدراسة:

تتمثل أهمية الدراسة الحالية في :

1- استخدام أسلوبين من أساليب القياس السلوكي(القياس محكي المرجع ، والقياس الموضوعي الذي يقوم على نظرية الاستجابة للمفردة (I.R.T).

2- التوصل إلى أداة قياس موضوعية لتحصيل الطلبة في مقرر الرياضيات للفصل الأول للصف السابع الأساسي يمكن الاستفادة منه في تشخيص نقاط القوة والضعف في أداء الطلبة ، كما يمكن استخدامه كنواة لإنشاء بنك أسئلة في المقرر .

3- قد تساعد نتائج هذه الدراسة في إعادة النظر في هذا المقرر وتقويمه.

هدف الدراسة :

تهدف الدراسة الحالية إلى الكشف عن مدى فاعلية النموذج اللوجستي ثلاثي المعلم في معايرة مفردات اختبار تحصيلي محكي المرجع في مقرر الرياضيات للصف السابع، وتحديد مستوى التمكن المتوقع من الطلبة من الاختبار.

مصطلحات الدراسة :-

النموذج اللوجستي ثلاثي المعلم : أحد نماذج نظرية الاستجابة للمفردة ، حيث يمكن لهذا النموذج تقدير ثلاثة معاملات هي: معامل الصعوبة، ومعامل التمييز، ومعامل تخمين المفردة .

الاختبار محكي المرجع : اختبار مستخدم لتحديد مستوى أداء الفرد بالنسبة لمستوى أو محك يحدد حسب الأهداف الموضوعة للقياس وبصرف النظر عن مستوى الأقران .

معامل الصعوبة: مقدار القدرة التي يكون احتمال إجابة المفحوص عندها للسؤال إجابة صحيحة = $(1+c)/2$.

معامل التمييز: مؤشر يفيد بقدرة المفردة على التمييز أو التفريق بين المستويات المختلفة من القدرة المقيسة، ويقاس بظل الزاوية التي يميل بها المماس للمنحنى المميز للمفردة على محور القدرة.

معامل التخمين: يشير إلى احتمال وصول المفحوص إلى الإجابة الصحيحة عن المفردات التي لم تمكنه قدرته من الإجابة عنها عن طريق التخمين العشوائي.

حدود الدراسة :

1- محتوى مقرر الرياضيات للصف السابع الذي يدرس في المدارس الحكومية في الفصل الأول من العام الدراسي(2013-2014)

2- عينة الدراسة حيث طبقت الدراسة على عينة من طلبة المدارس الحكومية في محافظة غزة في الفصل الأول من العام (2013 – 2014)

الطريقة والإجراءات :

مجتمع الدراسة :

تكون مجتمع الدراسة من جميع طلاب وطالبات الصف السابع في المدارس الحكومية التابعة لمديرية غرب غزة خلال العام الدراسي (2013–2014) والبالغ عددهم (2887) طالباً وطالبة . منهم (1197) طالبا و(1690) طالبة، وقد تم الحصول على أعداد الطلبة من دائرة التخطيط في مديرية غرب غزة .

عينة الدراسة :

تكونت عينة الدراسة من (624) طالباً وطالبة تم اختيارهم بطريقة عشوائية عنقودية، حيث تم اختيار عينة من المدارس بطريقة عشوائية بلغ عددها (16) مدرسة منها (7) مدارس للذكور، و(9) مدارس للإناث، واختير من كل مدرسة شعبة بطريقة عشوائية .

أداة الدراسة :

تم بناء اختبار تحصيلي في مقرر الرياضيات للصف السابع الأساسي، الجزء الأول، الذي اشتمل على الوحدات (المجموعات – الأعداد الصحيحة – الأعداد النسبية – التناسب الطردي والعكسي) وفقاً للخطوات التالية:

1- تحديد الغرض من الاختبار وهو قياس مقدار ما اكتسبه الطالب من معارف ومهارات وقوانين في وحدات المقرر من منهاج الصف السابع الأساسي.

2- تحليل محتوى المقرر إلى مفاهيم، ومصطلحات، ورموز، وتعميمات، ومهارات، وتطبيقات.

3- تم صوغ الأهداف العامة للوحدة من خلال الاستعانة بدليل المعلم والكتاب المدرسي ومعلمي المقرر من أصحاب الخبرة، كما تم اشتقاق جميع الأهداف السلوكية الممكنة التي يشتمل عليها المقرر. وللتأكد من صوغ الأهداف، وتحديدها، وشمولها للوحدة الدراسية، وتمثيلها للمستويات المعرفية المختلفة(التذكر، والاستيعاب، والتطبيق،

والتحليل، والتقويم، والإبداع) حسب تصنيف بلوم المعدل للأهداف(Krathwohl , 2002)، تم عرض قائمة الأهداف على مجموعة من المعلمين والموجهين التربويين في الرياضيات، وتم مناقشة ملاحظاتهم بالتفصيل وإجراء التعديلات المناسبة بناءً عليها، وتم الوصول إلى صيغة نهائية لقائمة مكونة من (60) هدفاً سلوكياً بحيث يمثل كل هدف من هذه الأهداف جزءاً محدداً من وحدات المقرر، وتغطي هذه الأهداف النطاق السلوكي للوحدات الدراسية .

4- تم إعداد جدول مواصفات للاختبار، حيث تم ربط مستويات الأهداف السلوكية بمحتوى وحدات المقرر، وتم مراعاة الوزن النسبي للمفاهيم والتعميمات والقوانين الخاصة بالوحدة حسب عدد الحصص التي يقضيها المعلم في التدريس.

5- تم كتابة (70) مفردة من نوع الاختيار من متعدد ولكل مفردة أربعة أبدال أحدها صحيح والأخرى أبدال خطأ (مموهات)، وقد تم كتابة أكثر من مفردة للهدف الواحد في بعض الأحيان بحسب درجة أهمية الهدف ودرجة صعوبة تحقيقه، وحاول الباحث تغطية جميع الجوانب المتعلقة بكل هدف سلوكي بحيث تشمل الفقرات التي تقيس الهدف السلوكي جميع جوانبه. وقد روعي في كتابة المفردات المعايير الفنية الخاصة بمفردات اختبارات الاختيار من متعدد، وأن تكون المفردة مرتبطة بالهدف الذي تقيسه من حيث المحتوى ودرجة الصعوبة، كما روعي تحقيق المفردات لشرط استقلالية القياس حيث لم تعتمد إجابة مفردة معينة على إجابة المفردات الأخرى.

6- تم عرض مفردات الاختبار وقائمة الأهداف على (12) محكم من المعلمين والمشرفين التربويين بالإضافة إلى (5) محكمين من أساتذة الجامعات في تخصص أساليب تدريس الرياضيات، وطلب منهم إبداء الرأي في مدى مطابقة مفردات الاختبار للأهداف المتعلقة بها، ومدى تغطية مفردات الاختبار لمحتوى التعلم واقتراح

ما يرونه مناسباً، وفي ضوء آراء المحكمين تم حذف (14) مفردة لاتفاق المحكمين بنسبة (95%) أنها متشابهة وعليه أصبح عدد المفردات للاختبار في صورته الأولية (56) مفردة.

التجريب الأولي للاختبار:

تم تطبيق الصورة الأولية للاختبار على (75) طالباً وطالبة من طلبة مدرسة عباد الرحمن الخاصة موزعين على ثلاث شعب بغرض تعرف مستوى صعوبة المفردات وتحديد زمن الاختيار والتأكد من سلامة المفردات وقياسها للأهداف بدقة، وتعرف أي مشكلة يمكن أن تواجه الباحث في التطبيق، وبناءً على التجريب الأولي للاختبار تم تحديد زمن الاختبار ب (80) دقيقة، وتم حساب صعوبة المفردات من خلال حساب نسبة الطلبة الذين أجابوا عن المفردة إجابة صحيحة من بين الطلبة الذين حاولوا الإجابة عن هذه المفردة، كما تم حساب معامل التمييز لجميع مفردات الاختبار، و يبين الجدول 1 قيم معاملات الصعوبة والتمييز لكل مفردة.

الجدول 1: قيم معاملات الصعوبة والتمييز لكل مفردة

رقم المفردة	معامل الصعوبه	معامل التمييز	رقم المفردة	معامل الصعوبة	معامل التمييز
1	0.40	0.35	29	0.63	0.40
2	0.32	0.54	30	0.85	0.33
3	0.52	0.44	31	0.55	0.29
4	0.36	0.36	32	0.54	0.32
5	0.53	0.40	33	0.65	0.29
6	0.43	0.60	34	0.52	0.37
7	0.81	0.21	35	0.42	0.46
8	0.66	0.42	36	0.48	0.51
9	0.78	0.55	37	0.65	0.39
10	0.95	0.35	38	0.77	0.53

0.38	0.95	39	0.36	0.74	11
0.60	0.59	40	0.46	0.62	12
0.24	0.47	41	0.50	0.65	13
0.38	0.88	42	0.43	0.74	14
0.19	0.49	43	0.62	0.35	15
0.20	0.74	44	0.52	0.62	16
0.30	0.84	45	0.43	0.56	17
0.29	0.64	46	0.60	0.78	18
0.46	0.38	47	0.46	0.35	19
0.62	0.54	48	0.38	0.46	20
0.25	0.85	49	0.37	0.82	21
0.27	0.43	50	0.44	0.71	22
0.33	0.77	51	0.28	0.70	23
0.45	0.76	52	0.38	0.63	24
0.38	0.58	53	0.30	0.72	25
0.41	0.61	54	0.42	0.58	26
0.20	0.33	55	0.35	0.64	27
0.25	0.38	56	0.38	0.47	28

يتضح من نتائج الجدول (1)أن قيم معاملات الصعوبة للمفردات تراوحت بين (0.32 إلى 0.95) بمتوسط حسابي 0.61 وانحراف معياري 0.17 بينما تراوحت قيم معاملات التمييز بين (0.19 إلى 0.62) بمتوسط حسابي 0.39 وانحراف معياري 0.11 ، مما يعني سلامة المفردات للتطبيق.

إجراءات الدراسة:

تم تطبيق الاختبار بصورته على عينة الدراسة، وتم تصحيح الاختبار بحيث أعطى المفحوص "1" في حالة الإجابة عن المفردة إجابة صحيحة ، و"صفراً" في حالة

الإجابة عن المفردة إجابة خطأ، وتم إدخال البيانات لذاكرة الحاسوب حيث تمت معالجتها وفقاً لأسئلة الدراسة.

تم التحقق من افتراضات نظرية الاستجابة للمفردة على النحو التالي:

أحادية البعد

تم تحديد الصدق البنائي للاختبار من خلال إجراء التحليل العاملي له و باستخدام طريقة المكونات الأساسية Principle Component Analysis حيث بينت النتائج أن هناك عاملاً واحداً رئيسياً تتجمع حوله مفردات الاختبار.ويوضح الجدول 2 معاملات تشبع المفردات على العامل الأول.

جدول2: معاملات تشبع المفردات على العامل الأول

رقم المفردة	معامل تشبع المفردة	رقم المفردة	معامل تشبع المفردة	رقم المفردة	معامل تشبع المفردة	رقم المفردة	معامل تشبع المفردة
1	0.52	15	0.73	29	0.39	43	0.74
2	0.47	16	0.84	30	0.60	44	0.73
3	0.65	17	0.62	31	0.44	45	0.42
4	0.49	18	0.77	32	0.55	46	0.54
5	0.51	19	0.76	33	0.47	47	0.60
6	0.64	20	0.56	34	0.33	48	0.73
7	0.74	21	0.57	35	0.52	49	0.78
8	0.63	22	0.53	36	0.57	50	0.37
9	0.65	23	0.66	37	0.67	51	0.79
10	0.70	24	0.46	38	0.59	52	0.36
11	0.46	25	0.82	39	0.61	53	0.30
12	0.38	26	0.48	40	0.47	54	0.60
13	0.49	27	0.57	41	0.56	55	0.58
14	0.57	28	0.53	42	0.58	56	0.64

يلاحظ من خلال الجدول (2) أن جميع معاملات تشبع المفردات على العامل الأول كانت دالة إحصائياً

وقد فسر هذا العامل حوالي (33%) من تباين درجات الأفراد على الاختبار، أما العامل الثاني فلم يفسر سوى (5%) من التباين، ويبين الجدول 3 قيم الجذر الكامن Eigen Values ونسب التباين المفسر Explained Variance للعامل الأول والثاني وناتج قسمة الجذر الكامن للعامل الأول على الجذر الكامن للعامل الثاني .

جدول 3 : قيم الجذر الكامن ونسب التباين المفسر للعامل الأول والثاني وناتج قسمة الجذر الكامن للعامل الأول على الجذر الكامن للعامل الثاني

	العامل الأول	العامل الثاني	ناتج القسمة
الجذر الكامن	5.95	1.78	2.83
التباين المفسر	32.85	5.03	

يتضح من النتائج الواردة في الجدول 3 بأن نسبة التباين المفسر من العامل الأول كانت أعلى من 20 وأن ناتج قسمة الجذر الكامن للعامل الأول على الجذر الكامن للعامل الثاني أكبر من 2 مما يشير إلى وجود سمة سائدة للاختبار، وهذا يعني تحقق افتراض أحادية البعد للاختبار(Hattie, 1985 ; Reckase , 1997)

الاستقلال الموضعي Local Independence

يقصد بالاستقلال الموضعي عدم وجود مفردات تعتمد في إجابتها على مفردات أخرى في نفس الاختبار لذا تم الحرص عند بناء المفردات المكونة للاختبار على تحقق هذا الافتراض، كما أن تحقق افتراض أحادية البعد يعتبر مؤشرا على تحقق هذا الافتراض

حيث يعتبر هذا الافتراض مكافئاً لافتراض أحادية البعد (Hambleton & Swaminathan ,1985)

عامل السرعة

تفترض نظرية الاستجابة للمفردة تحرر الاختبار من عامل السرعة، بمعنى أن يكون الوقت مناسباً وكافيا لجميع أفراد العينة وتبين من خلال تطبيق الاختبار على العينة الاستطلاعية ملاءمة وقت الاختبار لجميع الطلبة.

المعالجات الإحصائية

تم استخدام برنامج spss لحساب بعض الإحصاءات الوصفية (الوسط الحسابي، والانحراف المعياري، معاملات الارتباط) والتحليل العاملي.

تم استخدام برنامج Bilog- MG3 للحصول على تقديرات الأفراد ومعالم المفردات (الصعوبة– التمييز – التخمين) واشتقاق دوال المعلومات .

تم استخدام اختبار t لمقارنة متوسط عينة بمتوسط مجتمع افتراضي.

نتائج الدراسة ومناقشتها

أولاً: النتائج المتعلقة بالسؤال الأول

ما درجة مطابقة استجابات أفراد عينة الدراسة عن مفردات الاختبار لتوقعات النموذج اللوجستي ثلاثي المعلم؟

– تم إدخال البيانات الخاصة باستجابة أفراد عينة الدراسة عن مفردات الاختبار حيث تمت معالجتها عن طريق برنامج Bilog- MG3 وذلك كما يلي:

تم استبعاد أفراد العينة غير المطابقة لاستجاباتهم للنموذج وذلك حسب إحصائي مربع كاي؛ أي الذين تقل قيمتهم الاحتمالية المقابلة لقيمة مربع كاي عن (0.01) وعليه تم إسقاط (35) فرداً من أفراد العينة لم تتوافق استجاباتهم مع توقعات النموذج.

ويعني ذلك أن أولئك الأفراد قد تجاوزوا الحد المقبول إحصائيا بإجابتهم صواباً عن المفردات التي تزيد في مستوى صعوبتها على مستوى قدراتهم، أو يفشلون في الإجابة صواباً عن مفردات تتطلب مستوى قدرة أقل من قدراتهم؛ مما يعنى أنهم قد اعتمدوا على التخمين، أو عدم جديتهم، أو صدقهم في الاستجابة (Wright& Linacre, 1998).

- تم إعادة التحليل مرة أخرى بهدف حذف المفردات غير المطابقة لتوقعات النموذج وذلك حسب القيمة الاحتمالية لمربع كاي؛ أي تلك التي تقل لها القيمة الاحتمالية عن (0.05) ، وعليه تم حذف (12) مفردة غير مطابقة للنموذج، حيث كانت القيمة الاحتمالية لكل منها أقل من (0.05).
- أعيد التحليل مرة أخرى بهدف إيجاد معالم الصعوبة والتمييز والتخمين لجميع المفردات والجدول (4) يبين ذلك

جدول4: تقديرات معامل الصعوبة والتمييز والتخمين لجميع مفردات الاختبار بصورته النهائية

رقم المفردة	الصعوبة	الخطأ المعياري لمعلم الصعوبة	التمييز	الخطأ المعياري لمعلم التمييز	التخمين	الخطأ المعياري لمعلم التخمين	رقم المفردة	الصعوبة	الخطأ المعياري لمعلم الصعوبة	التمييز	الخطأ المعياري لمعلم التمييز	التخمين	الخطأ المعياري لمعلم التخمين
1	-2.90	0.28	0.50	0.19	0.10	0.03	23	0.11	0.24	1.20	0.21	0.01	0.02
2	-2.65	0.24	0.72	0.22	0.02	0.02	24	0.25	0.26	0.98	0.13	0.24	0.03
3	-2.15	0.21	0.61	0.15	0.01	0.04	25	0.39	0.31	0.85	0.23	0.22	0.01
4	-1.98	0.17	0.54	0.12	0.04	0.06	26	0.47	0.25	1.03	0.23	0.21	0.05
5	-1.77	0.19	0.64	0.42	0.06	0.02	27	0.59	0.18	0.97	0.35	0.13	0.04

0.03	0.17	0.31	1.45	0.14	0.68	28	0.03	0.12	0.23	0.72	0.14	-1.68	6
0.02	0.21	0.22	1.12	0.12	0.87	29	0.04	0.04	0.11	0.66	0.12	-1.62	7
0.01	0.22	0.26	0.78	0.14	1.20	30	0.06	0.08	0.15	0.47	0.13	-1.55	8
0.02	0.20	0.28	0.98	0.16	1.29	31	0.01	0.12	0.22	0.63	0.17	-1.48	9
0.03	0.19	0.32	0.76	0.14	1.31	32	0.05	0.20	0.23	0.49	0.16	-1.35	10
0.04	0.19	0.18	1.25	0.17	1.36	33	0.03	0.08	0.11	0.47	0.15	-1.20	11
0.01	0.22	0.14	1.47	0.15	1.40	34	0.04	0.01	0.14	0.56	0.17	-0.99	12
0.01	0.21	0.20	0.68	0.13	1.43	35	0.03	0.02	0.23	0.69	0.18	-0.88	13
0.02	0.18	0.15	0.79	0.25	1.48	36	0.03	0.12	0.20	1.20	0.19	-0.75	14
0.03	0.21	0.16	1.23	0.28	1.50	37	0.12	0.21	0.15	0.97	0.20	-0.65	15
0.02	0.16	0.17	0.19	0.21	1.53	38	0.13	0.14	0.24	0.78	0.21	-0.52	16
0.04	0.23	0.20	0.45	0.22	1.60	39	0.03	0.17	0.31	0.88	0.22	-0.44	17
0.06	0.12	0.26	0.24	0.24	1.62	40	0.05	0.15	0.13	1.22	0.24	-0.35	18
0.08	0.21	0.15	0.65	0.26	1.68	41	0.12	0.21	0.11	0.87	0.28	-0.27	19
0.11	0.23	0.22	0.78	0.23	1.75	42	0.11	0.16	0.26	1.34	0.25	-0.11	20
0.02	0.23	0.14	0.98	0.34	1.80	43	0.03	0.10	0.21	0.99	0.25	-0.06	21
0.03	0.21	0.18	0.47	0.35	1.89	44	0.11	0.20	0.18	0.77	0.27	0.00	22

يلاحظ من خلال الجدول (3) أن قيم معاملات الصعوبة لمفردات الاختبار في صورته النهائية تراوحت من (2.90−) إلى (1.89)، بمتوسط حسابي مساوٍ للصفر، وقد كانت معظم مفردات المقياس ضمن المدى (−2 ، + 2) وهو المدى الطبيعي، حيث إنّ معظم قدرات الأفراد تقع داخل هذا المدى، كما يتضح أيضاً من الجدول(3) أن معاملات الصعوبة تغطي مدى واسعاً من الصعوبة عند الأخذ في الاعتبار الخطأ المعياري لها، كما يتضح أن الخطأ المعياري للأسئلة الأصعب أكبر من الخطأ المعياري لمعاملات الصعوبة للأسئلة الأسهل وهو أمر منطقي؛ حيث إنّه كلما زادت مستويات الصعوبة قل عدد المستجيبين بشكل صحيح عن المفردة ، وبالتالي يزيد الخطأ المعياري للمفردة ، كما يتبين من خلال الجدول عدم وجود فجوات على متصل السمة التي يقيسها الاختبار حيث إنّ الفرق بين معاملات الصعوبة لأى مفردتين من مفردات الاختبار أقل من مجموع الخطأ المعياري لهما مما يدل على جودة تدريج مفردات الاختبار.

وتراوحت قيم معاملات التمييز من (0.19) إلى (1.47)، ويكون معامل التمييز للمفردة موجباً إذا كان احتمال إجابة هذه المفردة من قبل الطلبة ذوي التحصيل العالي أكبر من ذوي التحصيل المتدني(Richichi , 1996). وكلما زادت قيمة التمييز للمفردة كان ذلك أفضل ودليلاً على جودتها، ويمكن القول إنّه وحسبما أشار شوى (Choi,1992) فإنه إذا كانت قيمة تمييز المفردة أكبر من (1) فهذا دليل على أن المفردة جيدة، أما إذا كانت قيمة معامل التمييز بين (0.70 إلى 1) فتمييز المفردة مقبول، وتقل جودة المفردة إذا قلت قيمة تمييزها عن (0.70) ، وبالاعتماد على ذلك التصنيف تبين من خلال الجدول أن مفردات الاختبار توزعت إلى (10) مفردات (22.8%) ذات تمييز ممتاز، و(17) مفردة (38.6%) ذات تمييز مقبول، و(17) مفردة (38.6%) ذات تمييز قليل .

وتراوحت قيم معاملات التخمين لمفردات الاختبار من (0.01) إلى(0.24) وهي قيم صغيرة وأقل من القيمة النظرية المتوقعة للتخمين وهي (0.25) في حالة وجود أربعة أبدال للمفردة

ثانياً: النتائج المتعلقة بالسؤال الثاني

ما دلالات الصدق والثبات المتوافرة لمفردات اختبار محكي المرجع في مقرر الرياضيات للصف السابع والمتحررة من خصائص الأفراد؟

أولاً: صدق المقياس

عند استخدام أحد نماذج نظرية الاستجابة للمفردة في تطوير وبناء المقاييس التربوية لابد من توفير أدلة على صدق المقياس ، حيث لا يعنى مطابقة البيانات للنموذج المستخدم أنها تقدم دليلا على صدقه، وإنما يشير ذلك إلى أن مفردات المقياس تقيس

سمة عامة (Hambleton & Swaminathan ,1985) لذا تم إيجاد المؤشرات التالية للدلالة على صدق المقياس وهى:

أ- صدق الاتساق الداخلي

لحساب صدق الاتساق الداخلي للمفردات تم إيجاد معاملات الارتباط بين درجات الأفراد على المفردات ودرجاتهم على الدرجة الكلية للاختبار، والجدول (5) يبين ذلك

جدول5: معامل الارتباط بين الدرجة على المفردة والدرجة الكلية على الاختبار

المفردة	معامل الارتباط	المفردة	معامل الارتباط	المفردة	معامل الارتباط	المفردة	معامل الارتباط
1	0.45	12	0.43	23	0.53	34	0.77
2	0.42	13	0.53	24	0.63	35	0.38
3	0.37	14	0.67	25	0.60	36	0.45
4	0.47	15	0.65	26	0.65	37	0.65
5	0.65	16	0.63	27	0.61	38	0.74
6	0.68	17	0.70	28	0.71	39	0.38
7	0.73	18	0.44	29	0.48	40	0.75
8	0.58	19	0.48	30	0.65	41	0.64
9	0.66	20	0.78	31	0.70	42	0.67
10	0.54	21	0.64	32	0.88	43	0.81
11	0.72	22	0.74	33	0.81	44	0.61

يلاحظ من خلال الجدول (5) أن جميع معاملات الارتباط بين درجات الأفراد على المفردات ودرجاتهم على المقياس الكلى كانت عالية و موجبة و دالة إحصائيا عند

مستوى دلالة (0.01) وهو ما يوفر دليلاً على فاعلية مفردات الاختبار، حيث تقيس هذه المفردات ما يقيسه الاختبار.

ب- صدق التدرج

تعتبر (كاظم ،1996) أن المفردات غير الملائمة للنموذج هي تلك التي لا تتسق في تدريجها مع تدريج باقي المفردات على المتصل موضوع القياس، وغالبا ما تتضمن هذه المفردات قياساً لصفة أخرى، غير التي هي موضوع القياس، أو تكون هذه المفردات غامضة، أو هناك عيب في صياغتها .

وبناء على ذلك فإن استبعاد المفردات غير الصادقة، وإبقاء المفردات الصادقة في تعريفها للمتغير موضوع القياس يؤكد صدق مفردات الاختبار في قياسها للسمة المراد قياسها.

وهذا ما تم إجراؤه من خلال استبعاد المفردات غير الملائمة للنموذج، حيث استبعد التحليل (12) مفردة من مفردات الاختبار، ومما تجدر الإشارة إليه أن استبعاد المفردات جاء بعد استبعاد الأفراد غير الملائمين للنموذج، وحذف الأفراد غير الملائمين معناه حذف أولئك الأفراد غير الصادقين في استجاباتهم وهم الذين يختلف نمط استجاباتهم عن معظم الأفراد، وبحذف الأفراد غير الملائمين والمفردات غير الملائمة يتوافر لنا :-

صدق مفردات الاختبار في قياس السمة المراد قياسها وصدق تدرجها.

صدق تدرج قدرات الأفراد على متصل هذا المتغير.

ثبات الاختبار:

يشير الثبات إلى مدى الدقة في تقدير موقع كل من الأفراد والمفردات على متصل السمة المراد قياسها، ويتم تحديد مدى دقة المفردات في تعريف هذا المتصل بإيجاد النسبة بين الانحراف المعياري للقيم التقديرية المتحررة لصعوبة المفردات ومتوسط الخطأ المعياري لهذه القيم، ويطلق على هذه النسبة معامل الفصل بين المفردات Item Separation Index(G_p) حيث يجب أن تكون قيمة هذا المعامل أكبر ما يمكن فإذا لم تزد هذه القيمة عن (2) يصعب قياس المتغير بواسطة هذه المفردات. وقد وجد الباحث قيمة معامل الفصل بين المفردات في الاختبار بعد تدريجه باستخدام النموذج اللوجستي ثلاثي المعلم (3.78)، ونظراً لأن هذه القيمة تزيد على (2) فان مفردات المقياس تعد كافية لتعريف متصل السمة الذي تقيسه. ويمكن الحصول على معامل الثبات من خلال الصيغة الرياضية التالية:

$$R = \frac{G_p^2}{1+G_p^2}$$

حيث(G_p) ترمز إلى معامل الفصل(Wright &Masters ,1982). وقد تم حساب قيمة معامل الثبات لمفردات الاختبار وفقاً للعلاقة السابقة، وقد بلغ معامل الثبات (0.93) ومن المؤشرات الأخرى التي يستدل منها على ثبات المقياس باستخدام نظرية الاستجابة للمفردة دالة المعلومات (Information function) حيث تساعد في تحديد مقدار المعلومات التي تقدمها مفردة ما عن فرد ما وذلك من خلال تحديد أقصر ارتفاع للمنحنى الممثل لدالة المعلومات لتلك المفردة على متصل القدرة، وبالتالي

يمكن تحديد أى المفردات الاختبارية تقيس المتغير المراد قياسه(السمة) بدرجة أفضل عند مستويات محددة للقدرة .

وتعد دالة المعلومات للاختبار (Test information Function) من المؤشرات التي يستدل منها على معامل الثبات للمقياس وفق نظرية الاستجابة للمفردة، وبالتالي فان زيادة كمية المعلومات يؤدى إلى نقصان الخطأ المعياري للقياس، فقد أشار ريفي (Reeve , 2004) إلى أن دالة المعلومات للاختبار ترتبط مع ثبات المقياس من خلال العلاقة التالية:

$$r = 1 - \frac{1}{\sum_{i=1}^{I} I(\theta)}$$

وعليه يتوقع أنه كلما زادت دالة المعلومات للمقياس عند مستوى معين من القدرة فان ذلك يؤدى إلى زيادة الثبات، وبالتالي نقصان الخطأ المعياري للقياس والذي يتيح الفرصة لتقدير الخطأ المعياري عند كل مستوى من مستويات القدرة المختلفة ومعرفة مدى مساهمة كل مفردة من مفردات المقياس في تحديد دقة القياس(Hambleton & Swaminathan,1985).

وتم حساب دالة المعلومات لمفردات الاختبار، والجدول (6) يبين ذلك

جدول 6: دوال المعلومات لمفردات الاختبار

المفردة	دالة المعلومات	المفردة	دالة المعلومات	المفردة	دالة المعلومات	المفردة	دالة المعلومات
1	0.453	12	0.365	23	0.657	34	0.356
2	0.478	13	0.478	24	0.498	35	0.537
3	0.564	14	0.365	25	0.678	36	0.765
4	0.345	15	0.698	26	0.986	37	0.843
5	0.247	16	0.754	27	0.783	38	0.678
6	0.123	17	0.745	28	0.741	39	0.861
7	0.468	18	0.865	29	0.365	40	0.354
8	0.524	19	0.457	30	0.485	41	0.637
9	0.478	20	0.623	31	0.651	42	0.723
10	0.765	21	0.781	32	0.784	43	0.981
11	0.657	22	0.648	33	0.691	44	0.865

يتبين من خلال الجدول (5) أن هناك تفاوتاً في مقدار المعلومات التي تعطيها المفردات في المقياس ، وأن القيم القصوى لدالة المعلومات للمفردات جاءت عالية، وتم حساب دالة المعلومات التي يعطيها الاختبار الكلى، وقد بلغت (26.8) .

ثالثاً : النتائج المتعلقة بالسؤال الثالث

ما مستوى أداء الطلبة على الاختبار مقارنة بمستوى التمكن؟

لتحديد درجة القطع تم الالتقاء بعينة المحكمين وطلب منهم تحديد درجة القطع للاختبار بعد توضيح هدف الدراسة لهم وتوضيح الأساس المنطقي لطريقة أنجوف التي يستخدمونها في تحديد الحد الأدنى لمستوى الأداء المقبول، كما تم تزويدهم بنشرة تفصيلية عن خطوات تحديد مستوى الأداء وفق أسلوب أنجوف وتم إعلام المحكمين بضرورة تخيل مجموعة مكونة من (100) طالب ممن يمتلكون الحد الأدنى من الكفاية (مستوى الأداء المقبول من وجهة نظرك) " كم منهم سيجيب عن المفردة إجابة صحيحة ؟ " وذلك من خلال منحهم استمارة خاصة لتدوين تقديراتهم أمام كل مفردة من مفردات الاختبار، وتم الحصول على مجموع القيم الاحتمالية التي قدرها كل محكم لكل مفردة من مفردات الاختبار وتم حساب الوسط الحسابي والانحراف المعياري

لجميع المحكمين ولجميع مفردات الاختبار، وتراوحت الأوساط الحسابية لدرجات القطع المقدرة بين (47.4) (67.2) بمتوسط حسابي(58.16) وانحراف معياري (0.11) وبذلك تكون درجة القطع (58%) أى إنّ الطالب يجب أن يجيب عن (26) مفردة إجابة صحيحة كي يعد متمكناً . ومن الملاحظ أن درجة القطع جاءت أعلى من (50%) والتي تؤكد النتيجة التي توصل إليها شانج (change ,1999) بأن أسلوب أنجوف ينتج درجات قطع عالية وغالبا ما تكون أعلى من (50%) وقد برر ذلك بأن معظم المحكمين غالبا ما يفكرون في الطالب في مستوى المتوسط (أداؤه على الاختبار مقارنة بغيره يكون في وسط التوزيع) أو فوقه، ولا يفكر إلا القليل منهم في الطالب الذي يمتلك الحد الأدنى من الكفاية، كما أنهم يتأثرون بقدرات الطلبة الذين يدرسونهم وتوقعاتهم العالية لهم، مما يجعل تقديراتهم لدرجات القطع تميل نحو التشدد. ولمعرفة مدى اختلاف متوسط أداء الطلبة الفعلي عن درجة القطع تم استخدام اختبار (t) والجدول (7) يبين ذلك.

جدول7 : نتائج اختبار (t) لحساب دلالة الفروق بين متوسط أداء الطلبة على الاختبار ودرجة القطع

الوسط الحسابي	الانحراف المعياري	درجة القطع	قيمة t	درجات الحرية	مستوى الدلالة
28.45	4.85	26	9.8	*384	0.01

يتضح من الجدول (7) أن هناك فروقاً دالة إحصائياً بين متوسط أداء الطلبة ودرجة القطع مما يدلل على زيادة نسبة المتمكنين من المهارات المتضمنة في الاختبار

الخلاصة والمقترحات

* حجم العينة بعد حذف الأفراد غير الملائمين للنموذج

أمكن للباحث في الدراسة الحالية من الكشف عن مدى فاعلية النموذج اللوجستي ثلاثي المعلم في معايرة مفردات اختبار تحصيلي محكي المرجع في مقرر الرياضيات للصف السابع ، تتمتع مفرداته بدرجة عالية من موضوعية القياس تبرر استخدامه في قياس أداء الطلبة في المقرر، وعليه يوصي الباحث ب:

استخدام النموذج اللوجستي ثلاثي المعلم في بناء اختبارات تحصيلية في مقررات دراسية أخرى.

الاعتماد على مفردات الاختبار كنواة لبناء بنك أسئلة للمقرر متحررة من خصائص عينة المفحوصين.

المراجع

حمادنة، إياد محمد(2009).استخدام نظرية الاستجابة للمفردة في بناء اختبار محكي المرجع في الرياضيات وفق النموذج اللوجستي ثلاثي المعلم. مجلة العلوم التربوية والنفسية، جامعة البحرين ،10 (2)،215- 237.

الشريفين، نضال كمال(2006). الخصائص السيكومترية لاختبار محكي المرجع في القياس والتقويم التربوي وفق النظرية الحديثة في القياس التربوي والنفسي. مجلة العلوم التربوية والنفسية، جامعة البحرين،7 (4)،80 -109.

عبد الله، اعتدال(2000). القياس محكي المرجع لأحد مقررات علم النفس التعليمي، رسالة ماجستير غير منشورة،جامعة عين شمس، القاهرة.

علام، صلاح الدين محمود(2002) .القياس والتقويم التربوي والنفسي. أساسياته وتطبيقاته وتوجهاته المعاصرة.القاهرة : دار الفكر العربي.

علام، صلاح الدين(2001). الاختبارات التشخيصية مرجعية المحك في المجالات التربوية والنفسية والتدريبية .(ط2). القاهرة : دار الفكر العربي .

علام، صلاح الدين محمود (1995). الاختبارات التشخيصية مرجعية المحك في المجالات التربوية والنفسية والتدريبية (ط1). القاهرة : دار الفكر العربي .

علام، صلاح الدين محمود(1991). دراسة مقارنة لبعض طرق تحديد مستويات الأداء في اختبار محكي المرجع، المجلة المصرية للدراسات النفسية،ع77،1-96.

كاظم، أمينة محمد(1996). استخدام نموذج راش في بناء اختبار تحصيلي في علم النفس وتحقيق التفسير الموضوعي للنتائج. أنور الشرقاوي، سليمان الخضري، نادية عبد السلام. اتجاهات معاصرة في القياس والتقويم النفسي والتربوي(ص ص431-545) القاهرة: مكتبة الأنجلو المصرية.

كاظم، أمينة محمد (1994). تدريج ومعايرة المقاييس. محمد عماد الدين إسماعيل وأمينة كاظم وآخرون . معايير نمو طفل ما قبل المدرسة (ص ص114-232). القاهرة: المجلس القومي للأمومة والطفولة.

مراد، صلاح أحمد، وسليمان، أمين على(2002).الاختبارات والمقاييس في العلوم النفسية والتربوية . خطوات إعدادها وخصائصها.القاهرة : دار الكتاب الحديث.

المهاجري، ميرفت بنت عبد الهادي(2006). بناء اختبار محكي المرجع لقياس الكفايات الرياضية في حل المعادلات والمتباينات من الدرجة الأولى (بمتغير واحد ومتغيرين) لطالبات المرحلة المتوسطة بمدارس مكة المكرمة الحكومية. رسالة ماجستير غير منشورة، جامعة أم القرى، مكة المكرمة.

المراجع الأجنبية

Allen , M. & Yen , W.(1979**).** *Introduction To Measurement Theory* . Brooks Cole Publishing Company Monterey, California.

Change , L.(1999).Judgment al item analysis of the Nedelsky and Angoff standard- setting methods. *Applied Measurement in Education*,12 ,151-166.

Choi-Inn- Cul & Bachman, F.(1992). An investigation into the adequacy of three IRT models for data from two EFL reading tests. *Language Testing*, 9,51-87.

Cizek , G.(1996). Setting Passing Scores. *Journal of Educational Measurement*, 15,20-30.

Goodwin , L.(1996). Focus on Quantitative Methods Determining cut- Off Scores. *Research in Nursing & Health* ,19,249-256.

Guskey, T. (1988) .*Implementing Mastery Learning California* : wadsworth Publishing.Co.

Hambleton, R.(1982). *Test Score validity and standards – setting Methods*. In Berk, R.A.(ed),Criterion – referenced Measurement: the state of Art, 2ed. London: The john Hopkins Press Ltd.

Hambleton, R., Gruijter, D. (1983). Application of item response models to Criterion Referenced test item selection. *Journal of Educational Measurement* 20(4), 355-367.

Hambleton , R. & Swaminathan ,H.(1985). I*tem Response Theory "Principles & Applications* ..Kluwer Nijhoff Publishing.

Hattie , J .(1985). Methodology Review : Assessing unidimensionality of tests and items. Applied Psychological Measurement. 9,139-164

Jager, R.(1989).Certification of student competence. In Robert. Linn(ed). *Educational Measurement* . Collier Macmillan Publishers London.

Kane , M .(1987). On the Use of I.R.T. Models with judgmental Standard Setting Procedures. *Journal of Educational Measurement*, 24,333-345.

Kane , M .(1998).Choosing Between Examinee- Centered and Test – Centered Standard – Setting Methods. *Educational Assessment*, 5,129-145

Krathwhl , D.(2002). A Revision of Bloom's Taxonomy: an overview. *Theory into Practice*, 41(4),212-218.

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

Nitko , J. (2001).*Educational Assessment of student* .(3rd ed) .New Jersey. U.S.A. Prentice- Hall.Inc.

Peers, I. (1996). *Statistical Analysis For Education And Psychology* , London: The Fajmer Press.

Reckase , M .(1997). Unifactor latent trait models applied to multi-factor tests: Results and Implications .*Journal of educational statistics*,4(3),207-230.

Reeve, B. (2004). Applications of item response theory (I.R.T.) modeling for building and evaluating questionnaires measuring patient-reported outcomes. Retrieved on October 12 *2013,from: http://outcomes.cancer.gov/conference/irt/reeve.pdf*.

Richichi ,R. (1996). *An analysis of test bank multiple-choice items using item response theory*.(ERIC Document Reproduction Service NO,ED 367-405.

Shepard, L. (1984). Setting performance standards In Berk,R.(ed). *Guide to Criterion – referenced Test construction*. Hopkins University Press. Texas A&m University, Corpus Christi.

Suen , H. (1990). *Principles Of Test theories* . Hillsdale,New Jersey.

Wright, B.& Linacre, J.(1998). *A User's Guide to Winsteps /Bigsteps, Version2.88,* Chicago,U.S.A :MES ApPress

Wright , B. & Masters, G. (1982*). Rating Scales Analysis* .Chicago: META press.