التوافق بين النظرية التقليدية في القياس ونظرية استجابة الفقرة في مطابقة فقرات اختبار محكي المرجع في وحدة الهندسة التحليلية.

د/ باسل خميس سالم أبوفودة

- **مستخلص البحث** :

هدفت الدراسة إلى الكشف عن التوافق بين الخصائص السيكومترية للاختبار وفقراته وفق النظرية التقليدية في القياس والنموذج اللوجستي ثنائي المعلمة في مطابقة فقرات اختبار محكي المرجع. ولتحقيق أهداف الدراسة تم بناء اختبار محكي المرجع في مادة الرياضيات في وحدة الهندسة التحليلية لطلبة الصف العاشر الأساسي. وقد تكون الاختبار بصورته النهائية من(٣٠) فقرة من نوع الاختيار من متعدد من أربعة بدائل جرى التحقق من صدقها وثباتها. وجرى تطبيق الاختبار على عينة مكونة من(١٤٠) طالبا تم اختيارهم بالطريقة العشوائية العنقودية من مدارس تربية محافظة جرش خلال العام ٢٠١٢/٢٠١٣. وقد أظهرت نتائج الدراسة مطابقة (٢٩) فقرة للنظرية التقليدية في القياس، ومطابقة(٢٨) فقرة للنموذج اللوجستي ثنائي المعلمة، وقد طابقت (٢٨) فقرة لكلا النظريتين. وقد بلغت قيمة معامل الصدق وفقا للنموذج اللوجستي ثنائي المعلمة (٠,٩٢٥)، بينما بلغت قيمة معامل الصدق وفقا للنظرية التقليدية في القياس(٠,٩٣٦)، وأشارت النتائج إلى وجود فرق ذي دلالة إحصائية بين معاملي الصدق لصالح النظرية التقليدية في القياس. وبلغت قيمة معامل الثبات التجريبي وفقا للنموذج اللوجستي ثنائي المعلمة(٠,٩٦٥٨)، بينما بلغت قيمة معامل الثبات وفقا للنظرية التقليدية في القياس (٠,٩٥٤٩)، وأشارت النتائج إلى وجود فرق ذي دلالة إحصائية في تقدير معاملي الثبات ولصالح النموذج اللوجستي ثنائي المعلمة.

الكلمات المفتاحية: النظرية التقليدية في القياس، النموذج اللوجستي ثنائي المعلمة، اختبار محكي المرجع، احصائيات الفقرة، احصائيات الاختبار.

The Compatibility between the Traditional Test Theory and the item response theory in item stratification of Criterion Referenced Test.

Dr. Basel Khamis Abu Foudeh

Abstract:

This study aimed at detecting the the Compatibility between the traditional test theory and the item response theory in item stratification of criterion referenced test. To achieve the goals of the study, the researcher preparing a criterion referenced test in mathematics in the unit analytical engineering students tenth grade.The final version of the test consisted of (30) items with four options. The test was administered on sample consisting of (140). The study used statistical programs (BILOG – MG3, SPSS) to analyze the data. The results of the study showed that (29) items fit the traditional test theory, (28) items fit the two parameter logistic model, and (28) items fit both the two theories. The results indicated that the criterion related validity according to the traditional test theory was(0.936), While the validity coefficient according to the item test theory was(0.925). The result of the criterion related validity showed that there were significant differences in favor of the traditional theory of measurement. The results

also indicated that the reliability coefficient according to the traditional test theory was (0.9549), While the empirical reliability coefficient according to the item test theory was(0.9658). The result of the reliability showed that there were significant differences in favor of the two parameter logistic model.

Key Words:Traditional Test Theory,The two - parameter logistic model, Criterion Reference Test, Item parameters, Test Parameters.

- **المقدمة** :

تعّد عمليتي القياس النفسي والتقويم التربوي من أهم عناصر العملية التعليمية، حيث يرتبط نجاح المؤسسة التعليمية بقدرتها على القياس الدقيق والتقويم لمخرجات التعلم ومستوياته المتعددة، بحيث يؤدي إلى إصدار أحكام صحيحة تؤدى إلى اتخاذ قرارات سليمة تحسن وتطور من أداء المؤسسة التعليمية ، وبالرغم من أهمية القياس النفسي والتقويم التربوي إلا أنهما يعانيان من عدة تحديات شكلت أزمة حقيقية يشعر بها المتخصصون في القياس النفسي والمعلمون والطلاب، ومن بين التحديات نذكر منها: اختلاف نظم تدريج وتحليل المفردات الاختبارية حيث ظهر مدخلان هما: النظرية التقليدية في القياس، ونظرية استجابة الفقرة.

ويسعى علماء القياس والمعنيون بالعلوم التربوية والنفسية للتوصل إلى الموضوعية في قياس سلوك الأفراد واستجاباتهم، فقد بات بناء الأداة التي تمكن الباحث من الوصول إلى التقدير الموضوعي للسلوك من الأهداف المهمة التي ينشدها علماء التربية وعلم النفس(كاظم، ١٩٨٨م، ٧).

وعليه فقد ركز أصحاب الاتجاه المعاصر في القياس والتقويم التربوي والنفسي جهودهم للتوصل إلى أعلى مستويات الدقة، والموضوعية في القياس بحيث تتحقق أدق علاقة بين أداة القياس والسمة المراد قياسها، وبدأ اهتمامهم ينصب نحو نوع من الاختبارات التحصيلية أطلق عليها الاختبارات محكية المرجع (علام، ٢٠٠١م، ٢١). وقد ارتبط هذا النوع من الاختبارات بمدخل معاصر من مداخل القياس التربوي والنفسي أطلق عليه نماذج نظرية استجابة الفقرة، حيث جاءت بمنطلقات تتصدى للكثير من السلبيات التي تشوب النظرية الكلاسيكية في القياس(عيد، ٢٠٠٤م، ٢٣٥).

- **خلفية الدراسة** :

يميز علماء القياس بين مدخلين رئيسين لتحليل وتدريج المفردات هما: النظرية التقليدية في القياس والتي يطلق عليها نظرية الدرجة الحقيقية والخاطئة، حيث تستخدم هذه النظرية بغرض تحديد العوامل التى تؤثر على الدرجة التى يحصل عليها الفرد فى الاختبار، والمدخل الثاني: نظرية الاستجابة للمفردة التي ظهرت نتيجة للانتقادات التي وجهت لنظرية القياس التقليدية.

ولقد قدمت النظرية التقليدية في القياس حلولاً لبعض المشكلات التي تواجه الباحثين في بناء الاختبارات وتطويرها، إلا أنها عجزت عن حل مشكلات أخرى، حيث تفترض أن الخطأ المعياري في القياس متساو لكل المفحوصين، وهذا الافتراض يفتقر إلى الدقة، كما أن التعبير عن قدرة الفرد يتم من خلال الدرجة الحقيقية التي تتضح من خلال أدائه على الاختبار ككل، وليس على مستوى الفقرة، وبالتالي سيتغير وضع قدرة الفرد حسب تغير مستوى الاختبار، وعلاوة على ذلك فإنّ الاختبار والفقرات تتغير خصائصها بتغير خصائص الأفراد، كما أن خصائص الأفراد تتغير بتغير خصائص الاختبار من حيث الصعوبة والسهولة، ولا تصلح تلك النظرية لبناء الاختبارات مرجعية المحك (Hambleton,SwaminathanandRogers,1999;HambletonandSwaminathan ;, 1985; Randall, 1998, p.4؛ علاّم، ٢٠٠٥م، ٥٣ ـ ٥٦).

وحيث إن النظرية التقليدية في القياس تؤدي إلى بناء اختبارات غير مرنة، فقد وجه المتخصصون فى القياس جهودهم لوجود نظام قياس أكثر موضوعية يركز على انتقاء المفردات الاختبارية بشكل أفضل، ويسمح بإضافة أو حذف مفردات إلى الاختبار دون أن يتأثر الاختبار ككل، وقد أدت الجهود إلى ظهور النظرية الحديثة في القياس، إذ تُعدّ هذه النظرية ثورة في القياس النفسي والتربوي.

لقد ساعدت النظرية الحديثة في القياس في تقديم الكثير من الحلول لمشاكل تتعلق ببناء الاختبارات وتطويرها، وخاصة فيما يتعلق بتكافؤ الاختبارات ومعادلتها، وبناء الاختبارات المحكية المرجع، وبناء بنوك الأسئلة، والكشف عن تحيز الفقرات ونحو ذلك (Embretson and Reise, 2000؛ كاظم، ١٩٨٨م، ٤١)، كما لعبت النظرية دوراً هاماً في تحليل فقرات الاختبارات، وبالتالي المساهمة في تقييم جودة هذه الاختبارات، والميزة الكبرى لنظرية استجابة الفقرة، أنها تقود إلى فقرات تتسم معالمها باللاتغير عند تغيير العينة، إضافة إلى التعقيدات الرياضية في الأساليب المستخدمة للحصول على هذه التقديرات.

لقد نتجت عن النظرية الحديثة في القياس مجموعة من النماذج تعرف باسم نماذج السمات الكامنة، وتحدد جميع هذه النماذج العلاقة بين الأداء الملاحظ للفرد على الاختبار وبين السمة أو القدرة التي تكمن وراء هذا الأداء وتفسره. وترتكز نماذج النظرية الحديثة في القياس على مجموعة من الافتراضات، الواجب توافرها في البيانات المستمدة من الاختبار، إذ يعتمد الاختيار المناسب للنموذج على تحقيق هذه الافتراضات في البيانات، وتُعدّ الافتراضات القائمة عليها نماذج نظرية استجابة الفقرة قوية وصارمة. وتستند هذه النماذج إلى افتراضين أساسيين هما: افتراض أحادية البعد Unidimensionality، ويعني:

وجود قدرة واحدة تفسر أداء الفرد في الاختبار. وافتراض الاستقلال الموضعي Local Independence، ويعني: أن استجابات المفحوص للفقرات المختلفة في الاختبار مستقلة عن بعضها إحصائياً عن مستوى قدرة معين، أي أنّ أداء المفحوص في فقرة ما لا يتأثر سلباً ولا إيجاباً بأدائه عن الفقرات الأخرى(Hambleton and Swaminathan, 1985). والفارق الأساسي بين هذه النماذج هو عدد المعالم التي توصف بها الفقرة(Dato and Leo, 2005).

ومن أشهر النماذج المستخدمة باختلاف عدد المعالم: النموذج اللوجستي أحادي المعلمة (نموذج راش)، ويفترض هذا النموذج أن جميع الفقرات تميز بنفس القدر بين المفحوصين، لكنها تتباين فقط في صعوبتها. وهنا النموذج اللوجستي ثنائي المعلمة والنموذج اللوجستي ثلاثي المعلمة والذي يُعدّ امتدادا للنموذج اللوجستي ثنائي المعلمة، إذ يضيف معلماً جديداً للفقرة وهو معلم التخمين (ci)(Hambleton and Swaminathan, 1985, P.49- 55).

وقد اهتمت الدراسة الحالية بالنموذج اللوجستي ثنائي المعلمة Two Parameter Logistic Model، حيث يقوم هذا النموذج على افتراض اختلاف الفقرات في صعوبتها وتمييزها، وغياب عامل التخمين، ويعد النموذج أكثر واقعية من النموذج اللوجستي أحادي المعلمة؛ لأنه من الصعب إيجاد مجموعة من الفقرات لها نفس القدرة التمييزية على مستويات مختلفة من القدرة، والمعادلة الرياضية لهذا النموذج هي:

$$Pi(\Theta) = [1 + e^{-Dai(\Theta - bi)}]^{-1}$$

حيث:

Pi(Θ) : احتمال أن يجيب المفحوص ذو القدرة (Θ) عن الفقرة (i) إجابة صحيحة.

D : عامل التدريج Scaling Factor، وهو ثابت لجميع فقرات الاختبار ويساوي (1.70).

ai : معامل تمييز الفقرة (i): وهي قيمة متغيرة في هذا النموذج، ويشير إلى ميل منحنى خاصية الفقرة عند انعطاف المنحنى.

bi : معامل صعوبة الفقرة (i): وهي النقطة التي تقع على متصل القدرة عندما يكون ميل منحنى خاصية الفقرة أكبر ما يمكن.

e : هي الأساس اللوغاريتمي الطبيعي وتساوي (2.718).

وقد جرى الاعتماد على النموذج اللوجستي ثنائي المعلمة في هذه الدراسة بسبب أنّ المؤشرات الإحصائية المستخرجة من هذا النموذج أكثر اتفاقاً مع

المؤشرات التقليدية مقارنةً بالنماذج الأخرى(Hernandez, 2009)، بالإضافة إلى أفضلية هذا النموذج في انتقاء الفقرات(Pelton, 2002).

وفيما يتعلق بتفسير درجات الطلبة فيوجد في أدبيات القياس النفسي مدخلين متباينين لتفسير أداء الطلاب هما: المدخل مرجعي المعيار والآخر مرجعي المحك، ويتم في المدخل مرجعى المعيار تصنيف الطالب في ضوء بعض الإحصاءات المشتقة من درجات عينة الاختبار الذي يقيس السمة المراد تصنيفه فيها بهدف قياس الفروق الفردية بين الطلاب، وفي هذه الحالة يجب أن يكون توزيع الدرجات اعتدالي ، ومن ثم يقوم معد الاختبار بحذف المفردات المتطرفة بالصعوبة مهما كانت الأهداف التي تقيسها مما يؤثر على صدق الاختبار، ومن ثم فإن تفسير الدرجات يعتمد على طبيعة الجماعة المرجعية التي تستخدم في الحصول على المعيار، ولا تفيد تلك الدرجات في تحديد موضع الفرد من السمة المقاسه(علاّم، ٢٠٠١م، ١٥ – ١٧).

ونظراً للانتقادات التي وجهت للاختبارات مرجعية المعيار فقد ظهرت الاختبارات مرجعية المحك التي تهدف إلى تقدير أداء الطالب بالنسبة لمجموعة من الأهداف المحددة مسبقا بصرف النظر عن أداء أقرانه(الشرقاوي وآخرون، ١٩٩٦م، ٢٠). ويعّرف ثورندايك (Thorndaike, 1997, p:56) الاختبارات المحكية بأنها: الاختبارات التي تغطي نطاقاً ضيقاً وتستعمل لأجل قرارات التمكن. وغالباً ما تهدف الاختبارات محكية المرجع في الجانب المعرفي إلى التحقق من مدى إتقان تعلم الطلبة لمادة دراسية معينة، وتكمن أهمية هذه الأنواع من الاختبارات في مدى ما يستطيع الطالب تحقيقه ، أو ما لا يستطيع بغض النظر عن موقعه بالنسبة لباقي الطلبة (عودة، ٢٠١٠م، ٦٤؛ علاّم، ٢٠٠٦م، ١٣٠).

إنّ مسألة التمييز بين الاختبارات محكية المرجع والاختبارات معيارية المرجع ليست سهلة، حيث يمكن أن يستخدم نفس الاختبار في كلا الحالتين، ويكمن الفرق الرئيس بينهما في قضية تفسير النتائج. وفي هذا الصدد يشير عودة (٢٠١٠م، ١٤٨ - ١٥٠) إلى أن أوجه الشبه بينهما ربما كانت أوضح من أوجه الاختلاف، وقد بين ذلك من خلال عدة نقاط أهمها: أنّ كلا النوعين أعّد لقياس أهداف تدريسية معينة، إلاّ أنّ الاختبار محكي المرجع يتطلب درجة عالية من الانسجام بين الهدف والفقرة التي تقيسه.

ويرى الشريفين(٢٠٠٦م، ٧) إلى أن واقع التعليم لم يعّد مقتصراً على مجرد التمييز بين الطلبة في السمة المقيسة، وإنما يجب أن يركز على اكتسابهم مهارات معينة وتحقيق أهداف محددة، بل وإتقان تلك المهارات والمعلومات، وبالتالي مقارنة أداء الفرد بمستوى أداء معين في مجال من السلوك، وهذا ما تهدف إليه الاختبارات محكية المرجع.

لقد ذكر عودة(٢٠١٠م، ١٤٩) إلى أنه وفي الاختبارات مرجعية المحك يتم تحديد أداء الطالب في ضوء إحصاء يعرف بدرجة القطع التي تمثل الدرجة التي ينبغي للفرد الحصول عليها لكي يعد متقناً للسمة المقاسه. ونظراً لإدراك المختصين في التربية لأهمية تحديد درجات القطع؛ اجتهد عدد كبير منهم في تطوير أساليب ونماذج (Jager,1989; Sizmur,1997). وترتكز أغلب الطرق في تحديد درجة القطع على تقديرات المحكمين؛ لذا فلابّد أن يمتاز هؤلاء المحكمون حسب جيجر (Jager,1989) بالخبرة والمعرفة في مجال الاختبار، وأن يتفهم الخبراء أنماطاً واسعة من الممارسات في المجال المقصود.

وقد استخدمت الدراسة الحالية طريقة ندلسكي لمناسبتها لفقرات الاختبار من نوع الاختيار من متعدد، بالإضافة إلى سهولة تطبيقها. وتعتمد طريقة ندلسكي على أحكام الخبراء، بحيث يحدد كل محكم من بين بدائل كل فقرة من فقرات الاختيار من متعدد تلك التي يمكن استبعادها من قبل المفحوصين ذوي القدرة المتدنية؛ لأنها لا تمثل في نظرهم الإجابة الصحيحة للفقرة، ويكون الحد الأدنى لمستوى اجتياز الطالب لفقرة الاختبار مقلوب عدد البدائل المتبقية والذي يمثل احتمال الإجابة الصحيحة للفقرة. وبعد الانتهاء من التحكيم يتم جمع القيم لجميع فقرات الاختبار وجميع المحكمين، ومن ثمّ نقسم الناتج على عدد المحكمين، وبالتالي نحصل على متوسط الحد الأدنى لمستوى الاجتياز في الاختبار والذي يمثل درجة القطع(Berk, 1980).

- **الدراسات السابقة** :

ومن الدراسات التي تتعلق بموضوع الدراسة الحالية دراسة أجراها جمحاوي(٢٠٠٠م) حيث هدفت الدراسة إلى المقارنة بين النظرية التقليدية ونظرية الاستجابة للفقرة. ولتحقيق الهدف من الدراسة استخدمت اختبار للقدرة العقلية، وطبقت الاختبار على عينة مؤلفة من 1061 طالبا وطالبة من طلبة الصف التاسع الأساسي في مديرتي اربد الأولى والثانية .أشارت نتائج الدراسة إلى وجود اتفاق عالي بين النظريتين في تقدير صعوبة الفقرات، وأشارت كذلك إلى عدم وجود فروق ذات دلالة إحصائية بين معاملي ثبات الاختبار الناتج من النظرية التقليدية والنموذج ثنائي المعلمة، بينما وجدت فروق ذات دلالة إحصائية بين معامل الثبات للاختبار الناتج من النظرية التقليدية والناتج من نموذج راش والنموذج ثلاثي المعلمة.

وقام ياسين(٢٠٠٤م) بدراسة هدفت الدراسة إلى البحث في الخصائص السيكومترية لاختبار محكي المرجع في الكيمياء لطلاب الصف الأول الثانوي العلمي مقدرة وفق النظريتين الكلاسيكية والحديثة في القياس. وقد تكون الاختبار من(٥٢) فقرة من نوع الاختيار من متعدد من أربعة بدائل. وتكونت عينة الدراسة من(٤٨١) طالباً وطالبة موزعين على (١٤) شعبة، تم اختيارهم بالطريقة

العشوائية العنقودية. وجرى استخدام البرامج الإحصائية (Minsteps, Bigsteps,Microcat,SPSS) وذلك للحصول المؤشرات الإحصائية وفق النظريتين التقليدية والحديثة في القياس.

وقد أظهرت نتائج الدراسة مطابقة (٤٥) فقرة للنظرية التقليدية في القياس، ومطابقة (٤٣) فقرة وفق النموذج اللوجستي أحادي المعلمة، ومطابقة (٤٨) فقرة للنموذج اللوجستي ثلاثي المعلمة، وبلغ معامل الصدق المرتبط بمحك(0.84)، ومعامل الاتساق الداخلي بطريقة كرونباخ ألفا(0.90)، وبلغت قيمة معامل الثبات للاختبار باستخدام النموذج اللوجستي أحادي المعلمة (0.99).

وأجرى أبو هاشم(٢٠٠٦م) دراسة هدفت إلى المقارنة بين النظرية التقليدية ونموذج راش في اختيار فقرات قائمة مداخل الدراسة لدى طلاب الجامعة. حيث أظهرت نتائج الدراسة وجود اختلاف بين الفقرات المكونة لقائمة مداخل الدراسة باختلاف النموذج الإحصائي المستخدم، حيث أفرزت المؤشرات التقليدية للصدق والثبات قائمة مكون من (٢٨) بندا، ولم تحقق فقرتين خصائص جيدة حيث جرى استبعادهم، بينما أفرز نموذج راش قائمة مكون من (٢١) بندا، ولم تحقق (٩) فقرات ملائمة جيدة مع النموذج، وبذلك أصبحت القائمة فى صورته النهائية مكون من (٢١) فقرة لقياس مداخل الدراسة. وأظهرت النتائج أن تدريج فقرات قائمة مداخل الدراسة لا يتغير بتغير مستوى قدرة الأفراد المستخدمة فى الحصول على هذا التدريج، وأن مداخل الدراسة متنبئ جيد بمستوى التحصيل الدراسي لدى طلاب الجامعة على أساس ما يحققه نموذج راش من خطية التدرج وأحادية القياس واستقلاليته.

ومن الدراسات الأجنبية في هذا المجال دراسة أجراها أوينز(Owens, 2000) حيث هدفت هذه الدراسة إلى المقارنة بين معايير اختيار الفقرات وفق النظرية الكلاسيكية ونظرية استجابة الفقرة، من خلال بناء خمسة مقاييس مختلفة في عدد فقراتها وفي طرق انتقاء فقراتها. وأشارت نتائج الدراسة إلى وجود فروق ذات دلالة احصائية في معامل الثبات لصالح نظرية الاستجابة للفقرة، وقد كان الخطأ المعياري في التقدير لفقرات المقاييس التي تم انتقاؤها وفق نظرية استجابة الفقرة أقل مقارنة بالنظرية الكلاسيكية في القياس.

وقام هوانج(Hwang,2002) بدراسة هدفت إلى المقارنة بين النظرية الكلاسيكية ونظرية الاستجابة للفقرة بنماذجها اللوجستية الثلاثة من حيث تقديرات مؤشرات صعوبة الفقرات، وتمييزها، ومستويات القدرة للطلاب، وكذلك من حيث استقرار تقديرات الفقرات الاختبارية باختلاف مستويات قدرة الطلاب. وقد بلغ حجم عينة الدراسة(٦٠٠) طالب من طلاب الصف الثامن، وقد استخدم اختبار تحصيلي في الرياضيات والذي يشتمل على(١٥) فقرة. وقد

أظهرت نتائج الدراسة إلى أنّ معاملات الارتباط كانت مرتفعة حيث كان معظمها أكبر من ٠,٩٠ بين المؤشرات الإحصائية المستمدة من النظريتين وهذا يشير إلى وجود تشابه كبير بينهما، وأظهرت النتائج وجود تشابه واضح بين النظريتين في تقدير مؤشرات صعوبة الفقرات، إلا أن هذا التشابه يقل فيما يتعلق بتقدير مؤشرات تمييز الفقرات.

وأجرى ستيج(Stage,2003) دراسة هدفت إلى معرفة ما إذا كان التحول من النظرية الكلاسيكية في القياس إلى نظرية الاستجابة الفقرة هو عملية تطوير للفقرة. وقد قامت الدراسة بإجراء المقارنة بين صعوبة الفقرة وفق النظرية الكلاسيكية في القياس وصعوبة الفقرة وفق نظرية استجابة الفقرة. وتوصلت النتائج إلى أنّ النموذج ثنائي المعلمة يكون غير مناسب للبيانات، ويختلف تقدير صعوبة الفقرات لكل من النظرية الكلاسيكية ونظرية الاستجابة للفقرة، وأظهرت النتائج أن النظرية الكلاسيكية في القياس لا تعطي تدعيمًا قوياً مقارنة بالنموذج ثلاثي المعلمة لكل الاختبارات الفرعية الخمسة، وخلصت الدراسة إلى أنّ تحليل الفقرة في ضوء نظرية الاستجابة للفقرة أفضل مقارنة بالنظرية الكلاسيكية في القياس.

وأجرى كان (Kan, 2006) دراسة هدفت إلى المقارنة بين الخصائص السيكومترية ممثلة بالصعوبة والتمييز من خلال النظرية الكلاسيكية والنموذج اللوجستي ثنائي المعلمة. وجرى تحليل بيانات (٢٥) فقرة من فقرات تحديد المستوى الوطني من نوع الاختيار من متعدد. وقد جرى استخدام برنامج SPSS لتحليل الفقرات وفق النظرية التقليدية في القياس وبرنامج BILOG – MG3 وبرنامج Iteman لتحليل البيانات وفق النموذج اللوجستي ثنائي المعلمة. وأظهرت نتائج التحليل العاملي عن تحقق افتراض أحادية البعد لفقرات الاختبار، ووجود فروق ذات دلالة احصائية في متوسط معلمة صعوبة الفقرة ومتوسط معلمة تمييزها بين النظريتين ولصالح النموذج اللوجستي ثنائي المعلمة، وقد كانت معاملات الارتباط بين تقدير معلمة الصعوبة وفق النظريتين قوية، كما أظهرت النتائج كذلك وجود معاملات ارتباط قوية بين تقديرات معلمة التمييز وفق النظريتين.

وقام جايميلو وسلفستر(Jimelo and Silvestre, 2009) بدراسة هدفت إلى المقارنة بين النظريتين التقليدية والحديثة في القياس من خلال بناء اختبار محكي المرجع في مادة الأحياء يتكون من(٦٠) فقرة، وتم تحديد درجة قطع(٢٦) باستخدام طريقة أنجوف. وقد جرى تطبيق الاختبار على عينة مكونة من(٣٢٦) طالباً وطالبة في جامعة أريلونا. وقد جرى استخدام برنامج SPSS لتحليل الفقرات وفق النظرية التقليدية في القياس، واستخدام برنامج WINSTEPS لتحليل البيانات وفق النظرية الحديثة في القياس. أشارت نتائج الدراسة إلى

أنّ(٣٣) فقرة من فقرات الاختبار طابقت النظريتين التقليدية والحديثة في القياس، وقد تقاربت قيمة معامل الثبات وفقاً للنظريتين، حيث بلغت قيمة معامل الثبات وفقاً للنظرية التقليدية في القياس(٠,٧٢) وبلغت قيمة معامل الثبات وفقاً للنظرية الحديثة في القياس(٠,٧٠). وأظهرت نتائج التحليل عدم وجود فروق ذات دلالة احصائية(α = ٠,٠٥) في متوسط صعوبة الفقرات بين النظريتين.

وأجرى أون(Onn, 2013) دراسة هدفت إلى المقارنة بين النظريتين التقليدية والحديثة في القياس وذلك من حيث عدد الفقرات المنتقاه ومعامل الثبات. ولتحقيق أهداف الدراسة جرى إعداد اختبار في مادة الفيزياء مكون من(٥٠) فقرة من نوع الاختيار من متعدد جرى تطبيقه على عينة مكونة من (٦٩) طالبا وطالبة من طلبة المدارس في نيجيريا. وجرى تحليل فقرات الاختبار باستخدام برنامج SPSS لتحليل الفقرات وفق النظرية التقليدية في القياس، واستخدام برنامج وX-Calibreprograme لتحليل الفقرات وفق النموذج اللوجستي ثنائي المعلمة في نظرية استجابة الفقرة. وقد أظهرت نتائج التحليل عن مطابقة (٢٩) فقرة للنظرية التقليدية في القياس، ومطابقة(٣٨) فقرة للنموذج اللوجستي ثنائي المعلمة. وبينت النتائج انخفاض معامل ثبات الاختبار في كلا النظريتين، حيث بلغت قيمة معامل الثبات في النظرية التقليدية في القياس(٠,٤٩)، بينما بلغت قيمة معامل الثبات وفقاً للنموذج اللوجستي ثنائي المعلمة(٠,٦٧).

ويلاحظ من نتائج البحوث والدراسات السابقة أن نتائجها متناقضة، فقد أظهرت نتائج دراسة (أبو هاشم، ٢٠٠٦م؛ ياسين، ٢٠٠٤م) إلى وجود اختلاف واضح في عدد الفقرات التي تم استبعادها في كلا النظريتين، وتتشابه نتائج الدراستين مع نتائج دراسة أون(Onn, 2013) التي أظهرت وجود اختلاف في عدد الفقرات المطابقة في كلا النظريتين، وبينت نتائج دراسة جايميلو وسلفستر(Jimelo and Silvestre, 2009) وجود (٣٣) فقرة من فقرات الاختبار طابقت النظريتين.

وأظهرت نتائج دراسة (Owens, 2000; Stage, 2003) تفوق مؤشرات نظرية استجابة الفقرة على مؤشرات النظرية التقليدية في القياس، بينما أشارت دراسة هوانج(Hwang,2002) إلى وجود تشابه بين النظريتين في تقدير مؤشرات صعوبة الفقرات، إلا أن هذا التشابه يقل فيما يتعلق بتقدير مؤشرات تمييز الفقرات، وتتناقض هذه النتيجة مع نتائج دراسة(Stage, 2003). وأشارت دراسة كان (Kan, 2006) إلى وجود فروق ذات دلالة إحصائية في متوسط معلمة صعوبة الفقرة وتمييزها بين النظريتين ولصالح النموذج اللوجستي ثنائي المعلمة، وأنّ معاملات الارتباط بين تقديرات معلمة الصعوبة والتمييز وفق النظريتين قوية. وتتناقض نتيجة هذه الدراسة فيما يتعلق بمعلمة الصعوبة مع دراسة جايميلو وسلفستر(Jimelo and Silvestre, 2009) التي أظهرت عدم وجود فروق ذات

دلالة إحصائية في متوسط صعوبة الفقرات بين النظريتين، وقد تقاربت قيمة معامل الثبات وفقاً للنظريتين، كما بينت دراسة جمحاوي(٢٠٠٠) وجود اتفاق عالي بين النظريتين في تقدير صعوبة الفقرات. وفيما يتعلق بمعاملات الثبات فقد أظهرت نتائج دراسة أون(Onn, 2013) إلى انخفاض معامل ثبات الاختبار في كلا النظريتين، بينما أظهرت دراسة جمحاوي(٢٠٠٠م) إلى عدم وجود فروق ذات دلالة إحصائية بين معاملي ثبات الاختبار الناتج من النظرية التقليدية والنموذج ثنائي المعلمة.

- **مشكلة الدراسة وأسئلتها** :

اهتم العديد من الباحثين بالمقارنة بين المؤشرات الإحصائية المستخرجة وفق النظريتين التقليدية والحديثة في القياس، وذلك من أجل الحصول على اختبار يتمتع بخصائص سيكومترية جيدة. وقد تباينت نتائج الدراسات السابقة في هذا المجال فقد أشارت بعض الدراسات إلى تفوق مؤشرات النظرية الحديثة في القياس على المؤشرات المستخرجة وفق النظرية التقليدية في القياس مثل دراسة كل من (Kan, 2006; Owens, 2000)، في حين أشارت بعض الدراسات إلى عدم وجود فروق بين المؤشرات الإحصائية في كلا النظريتين مثل دراسة (Jimelo and Silvestre, 2009)، كذلك التناقض الواضح في نتائج الدراسات فيما يتعلق بالخصائص السيكومترية للاختبار ممثلة الصدق والثبات. وقد اختلفت الدراسات السابقة كذلك في نوع النموذج اللوجستي المستخدم. وعلى الرغم من وجود اختلاف واضح بين النظريتين الحديثة والتقليدية في القياس فيما يتعلق بطرق انتقاءها لفقرات الاختبار، إلا أنه لا توجد مبررات كافية على تفضيل نظرية دون الأخرى، وبالتالي جاءت فكرة هذه الدراسة باستخدام أحد النماذج الحديثة في القياس وهو النموذج اللوجستي ثنائي المعلمة والذي يعّد من أكثر النماذج اتفاقاً مع المؤشرات الإحصائية للنظرية التقليدية في القياس مقارنة بالنماذج الأخرى، وكذلك أفضلية هذا النموذج في انتقاء فقرات الاختبار، من أجل الكشف عن التوافق بين النظريتين الحديثة والتقليدية في القياس في مطابقة فقرات اختبار محكي المرجع. وبالتحديد فإنّ هذه الدراسة حاولت الإجابة عن الأسئلة الآتية:

- ما مدى مطابقة فقرات اختبار محكي المرجع مع النظرية التقليدية في القياس والنموذج اللوجستي ثنائي المعلمة؟
- ما مدى التوافق بين النظرية التقليدية في القياس والنموذج اللوجستي ثنائي المعلمة في انتقاء فقرات اختبار محكي المرجع؟
- ما الخصائص السيكومترية للاختبار محكي المرجع وفق النظرية التقليدية في القياس والنموذج اللوجستي ثنائي المعلمة؟
- هل يوجد فرق ذو دلالة إحصائية عند مستوى الدلالة الإحصائية(α = ٠,٠٥) بين معاملي الثبات المقدّرين باستخدام النظرية التقليدية في القياس والنموذج اللوجستي ثنائي المعلمة؟

- هل يوجد فرق ذو دلالة إحصائية عند مستوى الدلالة الإحصائية(α = ٠,٠٥) بين معاملي الصدق المقدّرين باستخدام النظرية التقليدية في القياس والنموذج اللوجستي ثنائي المعلمة؟

- **أهداف الدراسة** :

هدفت الدراسة الحالية إلى تحديد مدى التوافق بين النظرية التقليدية في القياس والنموذج اللوجستي ثنائي المعلمة في انتقاء فقرات اختبار محكي المرجع، من خلال تحديد أوجه التشابه أو الاختلاف بين النظرتين في انتقاء الفقرات، من حيث عدد الفقرات ومستوى صعوبتها وتمييزها، وكذلك صدق الاختبار وثباته.

- **أهمية الدراسة** :

نظراً لاعتماد العديد من الباحثين على تحليل فقرات الاختبار باستخدام النظرية التقليدية في القياس، وكذلك أهمية النماذج المستخدمة وفق النظرية الحديثة في القياس في تحليل فقرات الاختبار، فقد أمكن تحديد أهمية الدراسة من خلال إعطاء صورة واضحة لمعدي اختبارات التحصيل بأفضل الطرق في انتقاء فقرات الاختبار والتي تقدم خصائص سيكومترية للاختبار وفقراته مقبولة، كما أن الاعتماد على المقارنة وفق المؤشرات الإحصائية المستخرجة وفق النموذج اللوجستي ثنائي المعلمة في النظرية الحديثة في القياس ممثلة بالصعوبة والتمييز، وهي نفس المؤشرات المقابلة لها في النظرية التقليدية في القياس ليضيف أهمية أخرى لهذه الدراسة من حيث تحقق الموضوعية أثناء إجراء المقارنة. وتتضح أهمية الدراسة كذلك في تقديمها لأداة قياس تتمتع بخصائص سيكومترية جيدة في مادة الهندسة التحليلية على مستوى مرحلة التعليم الأساسية في الأردن.

- **مصطلحات الدراسة** :

- **الاختبار محكي المرجع**:

هو الاختبار الذي يستخدم لتفسير أداء الفرد بالنسبة لمجال سلوكي محدد تحديداً جيداً وليس التأكد من المركز النسبي للفرد مقارنة بمجموعته(علام، ٢٠٠١م، ٢٤). ويعرفّ الاختبار المحكي المرجع إجرائياً في هذه الدراسة بالاختبار الذي أعده الباحث لقياس الكفايات الرياضية في وحدة الهندسة التحليلية لطلبة الصف العاشر والمكون من(٣٠) فقرة من نوع الاختيار من متعدد.

- **معامل صعوبة الفقرة من منظور النظرية التقليدية في القياس**:

يشير إلى نسبة الطلبة الذين اجابوا عن الفقرة إجابة صحيحة من بين المفحوصين الذين حاولوا الإجابة عن هذه الفقرة.

- **معامل تمييز الفقرة من منظور النظرية الكلاسيكية في القياس**:

يشير إلى قدرة الفقرة على التمييز بين الفئة العليا والفئة الدنيا، بمعنى: أن ينسجم تمييز الفقرة مع تمييز الاختبار كله، وجرى حسابه من خلال إيجاد

معامل الارتباط بين نتائج المفحوصين على كل فقرة من فقرات الاختبار ونتائجهم على الاختبار الكلي (Corrected item total correlation, rpbis) (عودة، ٢٠١٠م، ٢٨١)

- **معلمة صعوبة الفقرة**(bi) **حسب النموذج اللوجستي ثنائي المعلمة**:

تشير إلى نقطة على متصل السمة الكامنة يكون عندها احتمال الإجابة الصحيحة عن الفقرة(i) يساوي (٠,٥٠).

- **معلمة تمييز الفقرة**(ai) **حسب النموذج اللوجستي ثنائي المعلمة**:

وهي ميل منحنى خصائص الفقرة الذي يحدث عنده تغيير في اتجاه المنحنى (inflexion) الذي يقابل الصعوبة على متصل القدرة والتي يكون عندها احتمال الإجابة الصحيحة عن الفقرة(i) يساوي (٠,٥٠).

- **النموذج اللوجستي ثنائي المعلمة**(2PLM: Two Parameter Logistic Model):

وهو أحد النماذج المستخدمة في النظرية الحديثة في القياس أحادية البعد، إذ يقوم على افتراض اختلاف الفقرات في صعوبتها وتمييزها، وغياب عامل التخمين (Hambleton & Swaminathan, 1985, P.49).

- **الخصائص السيكومترية للاختبار**:

ويقصد بها معاملي الصدق والثبات للاختبار محكي المرجع المعّد في هذه الدراسة.

- **درجة القطع**:

وهي الدرجة التي على الطالب اجتيازها حتى يعدّ متمكناً(علام، ٢٠٠١م). وتشير في هذه الدراسة إلى الدرجة التي على الطالب اجتيازها حتى يعّد متمكناً في وحدة الهندسة التحليلية، وقد بلغت هذه الدرجة وفق طريقة ندلسكي(١٥).

- **حدود الدراسة** :

اقتصرت الدراسة على مادة الرياضيات في وحدة الهندسة التحليلية والتي تدّرس لطلبة الصف العاشر الأساسي في تربية محافظة جرش بالأدرن خلال العام ٢٠١٢/٢٠١٣م. واقتصرت على استخدام النموذج اللوجستي ثنائي المعلمة، وكذلك على طريقة ندلسكي لتحديد درجة القطع.

- **منهج الدراسة** :

استخدمت الدراسة الحالية المنهج الوصفي في البحث.

- **مجتمع الدراسة** :

تكون مجتمع الدراسة من جميع طلبة الصف العاشر الأساسي الذكور والبالغ عددهم(١٠٤٨) طالباً موزعين على (٨٣) مدرسة وذلك حسب الإحصائيات التي تم تزويدها للباحث من مديرية تربية محافظة جرش بالأردن خلال الفصل الدراسي الأول لعام ٢٠١٢/٢٠١٣م.

- **عينة الدراسة** :

جرى اختيار عينة الطلبة بالطريقة العشوائية العنقودية من مدارس الذكور، حيث كانت الشعبة هي وحدة الاختيار، وجرى اختيار أربع مدارس بطريقة عشوائية من بين المدارس التي تحوي الصف العاشر الأساسي، واختيرت شعبة واحدة عشوائيا من كل مدرسة، وقد بلغت عينة الدراسة (١٤٠) طالبا. وقام الباحث كذلك باختيار(٧) محكمين بالطريقة العشوائية البسيطة من مجتمع المعلمين والمشرفين التربويين وذلك لتحديد درجة القطع.

- **أداة الدراسة** :

ولتحقيق أهداف الدراسة قام الباحث ببناء اختبار محكي المرجع في وحدة الهندسة التحليلية من منهاج الرياضيات لطلبة الصف العاشر الأساسي، حيث راعى الباحث الأسس العلمية في بناء الاختبارات محكية المرجع(علاّم، ٢٠٠٠م). وتكون الاختبار من(٣٠) فقرة من نوع الاختيار من متعدد لكل فقرة أربعة بدائل واحد منها يمثل الإجابة الصحيحة. وقد حرص الباحث على أن تقيس كل فقرة هدفا محددا، وفقا لقائمة الأهداف التي أعدها الباحث لهذه الوحدة. وجرى تحليل محتوى الوحدة تحليلا دقيقا ومفصلا، وأعدت قائمة مكونة من ثلاثين هدفا تفصيليا شاملا لمحتوى الاختبار.

وللتأكد من شمول الأهداف للوحدة الدراسية وتمثيلها للمستويات المعرفية الثلاثة(التذكر، والفهم، والتطبيق)، قام الباحث بعرض قائمة الأهداف على مجموعة من المختصين في تدريس مادة الرياضيات لطلبة الصف العاشر الأساسي، وكذلك مجموعة من المشرفين التربويين ممن يحملون درجة الماجستير في أساليب تدريس الرياضيات؛ وذلك من أجل إبداء الرأي حول مدى شمول الأهداف لموضوع الوحدة الدراسية وطريقة صياغتها. وبعد إقرار جميع الأهداف التفصيلية للوحدة الدراسية المعنية قام الباحث بكتابة مجموعة من الأسئلة المحتملة على كل هدف بشكل منفصل. وقد تراوح عدد الفقرات المحتملة على لكل هدف بين خمس إلى ست فقرات، ومن ثمّ قام الباحث باختيار فقرة عشوائيا لكل هدف من الأهداف.

وجرى التحقق من فعالية الفقرات في الاختبار المعّد وذلك من خلال جمع آراء المحكمين والمختصين في تدريس مادة الرياضيات لطلبة الصف العاشر الأساسي، وذلك لبيان درجة التوافق بين الفقرة والهدف الذي يقيسه، حيث تم الاعتماد على درجة الاهتمام بصياغة الأهداف وتحديد مستوى الهدف ودورها في اختيار أنسب فقرة لقياس الهدف الواحد.

وقام الباحث بعرض مقياس تقدير متدرج من(١ ـ ٥) على مجموعة من المحكمين والمختصين في مجال تدريس الرياضيات؛ وذلك لأخذ وجهات نظرهم

في مدى صدق الفقرة الاختبارية في قياس الهدف المحدد، واقتراح ما يرونه من تعديل في ضوء استجابتهم على فقرات المقياس من حيث: متن الفقرة، وجاذبية الموهات، استقلالية الفقرات، خلوها من الأخطاء، مناسبة اللغة، ارتباط الفقرة بالهدف. وقد تراوحت قيم المتوسطات الحسابية لتقديرات المحكمين لصياغة فقرات الاختبار بين(٤,٧٥ ـ ٥)، وهي قيم مرتفعة مما يعني أن فقرات الاختبار مصاغة بشكل سليم. وقد وجد أن متوسط تقديرات المحكمين يساوي(٥) وهو أعلى تقدير بانحراف معياري مقداره صفر، مما يدل على اتفاق المحكمين فيما يتعلق بمطابقة كل هدف للفقرة التي تقيسه. وبالتالي يتمتع الاختبار بدرجة مرتفعة من صدق المحتوى بمعنى: أنّ فقرات الاختبار تشكل عينة ممثلة من مجتمع الفقرات التي تقيس المهارات المتعلقة بالوحدة المعنية.

وللتحقق من قدرة الاختبار على التمييز بين المجموعات المتمايز في السمة موضع الاختبار، قام الباحث بتطبيق الاختبار على عينة استطلاعية مكونة من(٣٩) طالباً قبل أن يتعلموا الوحدة الدراسية، وجرى مقارنة متوسطاتهم الحسابية على الفقرات مع المتوسطات الحسابية لعينة الدراسة(١٤٠) طالباً بعد أن تعلموا تلك الوحدة. وقد جرى اختبار دلالة الفرق باستخدام اختبار "ت" لعينيتن مستقلتين. ويبين الجدول رقم(١) نتائج التحليل:

جدول رقم(١)

نتائج اختبار (t) لاختبار الفرق بين المجموعتين والدلالة الاحصائية.

المجموعة	العدد	المتوسط الحسابي	الانحراف المعياري	قيمة (t) المحسوبة	درجات الحرية	قيمة الاحتمال
قبل التدريس	٣٩	١١,٨٧	٧,٣٨	١٨,٥٤	١٧٧	٠,٠٠ ❖❖
بعد التدريس	١٤٠	٢٦,٩٢	٣,٢٧			

تعني: ذات دلالة احصائية عند مستوى الدلالة الاحصائية($\alpha = 0.01$)

ويلاحظ من نتائج الجدول رقم(١) وجود فرق ذي دلالة إحصائية بين أداء المجموعتين حيث بلغت قيمة (ت) المحسوبة(١٨,٥٤) باحتمال (٠,٠٠) وهذه القيمة دالة إحصائياً عند مستوى الدلالة ($\alpha = 0.01$)، مما يدل على أن الاختبار يميز بين المجموعات المتمايزة على السمة، أي أن الاختبار يقيس الأداء المتعلق بمجال سلوكي محدد من المعارف والمهارات التي تعلمها الطلبة في وحدة الهندسة التحليلية.

وجرى التحقق من صدق أداة الدراسة من خلال الصدق المرتبط بمحك التحصيل في الرياضيات، إذ بلغت قيمته (٠,٨٧). ولتقدير ثبات الاتساق الداخلي للاختبار، قام الباحث بتطبيق الاختبار على عينة استطلاعية مكونة من(٢٧) طالباً من طلبة الصف العاشر الأساسي بعد أن تعلموا الوحدة الدراسية، وقد بلغت قيمة معامل ثبات كرونباخ ألفا للاختبار (٠,٨٩). وتجدر الإشارة في هذا

المجال إلى أن طريقة تقدير ثبات كرنباخ ألفا تناسب الاختبارات معيارية المرجع، وعند استخدامها مع الاختبارات محكية المرجع فإنها تعطي تقديراً منخفضاً للثبات، الأمر الذي يدعونا أن نثق بثبات الاختبار الحالي والمستخدم في هذه الدراسة. وقد تراوحت قيم معاملات الصعوبة لفقرات الاختبار بصورته الاستطلاعية بين (٠,٣٣ ـ ٠,٧٨)، وتراوحت قيم معاملات التمييز بين (٠,٢٧ ـ ٠,٥٧)، وبالتالي تم قبول جميع فقرات الاختبار وعددها (٣٠) فقرة.

وفي الاختبار محكي المرجع المعّد في الدراسة الحالية جرى تحديد أداء الطالب في ضوء إحصاء يعرف بدرجة القطع والتي تمثل الدرجة التي ينبغي للطالب الحصول عليها لكي يعّد متقناً للسمة المقاسه. حيث قام الباحث بتحديد درجة القطع من خلال اختيار عينة عشوائية بسيطة قوامها (٧) محكمين؛ (٥) معلمين يدرسون مبحث الرياضيات للصف العاشر الأساسي، (٢) من مشرفي الرياضيات تخصص مناهج وأساليب التدريس، حيث طلب من كل محكم تحديد من بين بدائل كل فقرة من فقرات الاختيار من متعدد تلك التي يمكن استبعادها من قبل المفحوصين ذوي القدرة المتدنية؛ لأنها لا تمثل في نظرهم الإجابة الصحيحة للفقرة، ويكون الحد الأدنى لمستوى اجتياز الطالب لفقرة الاختبار مقلوب عدد البدائل المتبقية والذي يمثل احتمال الإجابة الصحيحة للفقرة. ويبين الملحق رقم(١) آراء المحكمين لتحديد درجة القطع حسب طريقة ندلسكي. وبعد الانتهاء من التحكيم جرى جمع القيم لجميع فقرات الاختبار وجميع المحكمين، ومن ثمّ قسمة الناتج على عدد المحكمين، وتمثل القيمة الناتجة متوسط الحد الأدنى لمستوى الاجتياز في الاختبار المعّد والذي يمثل درجة القطع، وقد بلغت درجة القطع بطريقة ندلسكي (١٥).

وبعد الخروج بالصورة النهائية للإختبار جرى تطبيقه بشكل نهائي على أفراد عينة الدراسة المكونة من(١٤٠) طالباً من طلبة الصف العاشر الأساسي كل في مدرسته، وبشكل جماعي داخل غرفة الصف، حيث جرى إبلاغ الطلبة بموعد الاختبار، وقام الباحث بالاستعانة بمعلمي مبحث الرياضيات في تلك المدارس لتطبيق الاختبار.

- **المعالجات الإحصائية** :

جرى إدخال البيانات إلى ذاكرة الحاسوب، واستخدم البرنامج الإحصائي (SPSS) وبرنامج (MG3ـ BILOG) لإجراء التحليلات الإحصائية اللازمة للإجابة عن أسئلة الدراسة.

- **نتائج الدراسة ومناقشتها** :

النتائج المتعلقة بالسؤال الأول ومناقشتها: ما مدى مطابقة فقرات اختبار محكي المرجع مع النظرية التقليدية في القياس والنموذج اللوجستي ثنائي المعلمة؟

قام الباحث أولاً بالتحقق من افتراضات نظرية استجابة الفقرة في فقرات الاختبار الكلي البالغة (٣٠) فقرة، حيث جرى التحقق من افتراض أحادية البعد Unidimensionality باستخدام المؤشرات التي اعتمدت التحليل العاملي للمكونات الرئيسة Principal Components Analysis من خلال استخدام التحليل العاملي، حيث أفرز التحليل(١٠) عوامل قيمة الجذر الكامن لكل منها تزيد عن واحد، ويبين الجدول (٢) قيم الجذور الكامنة ونسبة التباين المفسر لكل عامل، وكذلك نسبة التباين المفسر التراكمية.

الجدول رقم(٢) نتائج التحليل العاملي للاختبار الكلي

رقم العامل	الجذر الكامن	نسبة التباين المفسر ٪	نسبة التباين المفسر التراكمية٪
١	٩,١١٥	٣٠,٣٨٤	٣٠,٣٨٤
٢	٢,٥٣٥	٨,٤٥٠	٣٨,٨٣٤
٣	٢,١٢٧	٧,٠٨٨	٤٥,٩٢٢
٤	٢,٠٥٢	٦,٨٤٠	٥٢,٧٦٢
٥	١,٦٤٠	٥,٤٦٨	٥٨,٢٣٠
٦	١,٥٠١	٥,٠٠٤	٦٣,٢٣٤
٧	١,٣٨٣	٤,٦٠٩	٦٧,٨٤٣
٨	١,٢٧٧	٤,٢٥٨	٧٢,١٠١
٩	١,١٣٣	٣,٧٧٥	٧٥,٨٧٦
١٠	١,٠٥٢	٣,٥٠٦	٧٩,٣٨٢

ويلاحظ من نتائج الجدول(٢) أنّ قيمة الجذر الكامن للعامل الأول (٩,١١٥) ويفسر ما نسبته (٣٠,٣٨٤٪) من التباين الكلي، ويعدّ الاختبار أحادي البعد إذا كان نسبة ما يفسره العامل الأول أكثر من (20٪) تقريباً (Reckase, 1979: 230 - 207). وتُعتَمد في التحليل العاملي أحادية البعد كذلك من خلال نسبة الجذر الكامن للعامل الأول إلى الجذر الكامن للعامل الثاني، وتكون نسبة كبيرة لا تقل عن (٢) (Hambleton and Swaminathan, 1985: 157)، ويتضح من الجدول(٢) سابق الذكر أنّ ناتج قسمة قيمة الجذر الكامن للعامل الأول على قيمة الجذر الكامن للعامل الثاني تساوي (٣,٦٠) وهي نسبة تزيد عن المعيار (٢)، وعند النظر إلى نسبة الفرق بين الجذر الكامن للعامل الأول والجذر الكامن للعامل الثاني إلى الفرق بين الجذر الكامن للعامل الثاني والجذر الكامن للعامل الثالث، تبين أنّ النسبة كبيرة وتساوي (١٦,١٣)، وهذا يوضح أنّ الاختبار المعّد يحقق افتراض أحادي البعد، وبالتالي يمكن الاعتماد على فقراته في الوصول الى قرارات سليمة. وقد بلغت قيمة معامل ثبات كرونباخ ألفا (٠,٩٠٣) وهذه القيمة مرتفعة. وقد اكتفى الباحث بالتحقق من افتراض أحادية البعد للإشارة إلى تحقق افتراض الاستقلال الموضعي Local Independence (Hambleton and Swaminathan, 1985: 22 - 25). وفيما يتعلق بافتراض خاصية المنحنى المميز للفقرة أو المفردة Item Characteristic Curve (ICC) فمن المعروف أنّ احتمال إجابة الطالب لفقرة الاختبار إجابة صحيحة تزداد بازدياد قدرته. وتمّ التحقق

من أنّ عامل السرعة لم يلعب دوراً في إجابة الطلبة عن فقرات الاختبار، حيث أعطي الطلبة الوقت الكافي للإجابة عن فقرات الاختبار.

وبعد التحقق من الافترضات السابقة جرى استخدام برنامج(BILOG - MG3)، لفحص حسن المطابقة للبيانات حسب النموذج اللوجستي ثنائي المعلمة؛ وذلك للكشف عن الأفراد غير المطابقين للنموذج اللوجستي ثنائي المعلمة. وقد أفرزت نتائج التحليل باستخدام اختبار مربع كآي عند مستوى الدلالة($0.01 = \propto$) عدم مطابقة(٤) أفراد للنموذج كانت قيمة الاحتمالية أقل من (0.01). وجرى إعادة التحليل وذلك لاختيار الفقرات المطابقة للنموذج اللوجستي المستخدم، وأظهرت النتائج مطابقة(٢٨) فقرة للنموذج المستخدم، وعدم مطابقة الفقرتين رقم(١٠، ١٧) في نموذج الاختبار كانت قيمة الاحتمالية لها على الترتيب(٠,٠٠٢ ، ٠,٠٠٤١) وهاتان القيمتان أقل من (0.01). كما قام الباحث بتحليل البيانات باستخدام برنامج (SPSS) وذلك للحصول على معاملات الصعوبة والتمييز لفقرات الاختبار من منظور النظرية الكلاسيكية في القياس. ويبين الجدول رقم(٣) المؤشرات الإحصائية لفقرات الاختبار قبل حذف الفقرات غير المطابقة وفقاً للنظريتين الحديثة(IRT) والتقليدية(CTT) في القياس:

الجدول رقم(٣)معاملات الصعوبة والتمييز لفقرات الاختبار وفقاً للنظريتين الحديثة والتقليدية في القياس

رقم الفقرة	التحليل باستخدام CTT		التحليل باستخدام 2PLM		رقم الفقرة	التحليل باستخدام CTT		التحليل باستخدام 2PLM	
	معامل الصعوبة (p)	معامل التمييز(r_{pbi})	معلمة الصعوبة(b)	معلمة التمييز(a)		معامل الصعوبة(p)	معامل التمييز(r_{pbi})	معلمة الصعوبة(b)	معلمة التمييز(a)
١	٠,٤٧	٠,٣٨	٠,٥٤	٠,٨٤٩	١٦	٠,٥٩	٠,٤٠	٠,٠٩	١,٣١
٢	٠,٥٢	٠,٦٢	- ١,٢١	٠,٧٦	١٧	٠,٥٥	٠,٦٠	٠,١٢	٠,٣٠
٣	٠,٦٩	٠,٤٠	- ١,٢٤	١,٩٧	١٨	٠,٥٢	٠,٤٨	٠,٥٣	١,٢٩
٤	٠,٥٧	٠,٣٩	- ١,٤٠	١,٣١	١٩	٠,٤٠	٠,٦٩	- ١,٤٩	١,٥٠
٥	٠,٨١	٠,٤١	- ٠,٥٩	١,٠١	٢٠	٠,٤٨	٠,٧١	- ٠,١٩	٢,٥٧
٦	٠,٤٨	٠,٥٥	- ١,١٧	١,٢٩	٢١	٠,٥٩	٠,٣٩	- ١,٣٧	٢,٨٣
٧	٠,٥٢	٠,٤٤	- ٠,٩٦	١,١٩	٢٢	٠,٥٣	٠,٤٥	٠,٠١٧	١,٦٩
٨	٠,٤٦	٠,٧٠	٠,١٨	٠,٩	٢٣	٠,٧٧	٠,٣٩	- ٠,٤١	١,١٩
٩	٠,٤٥	٠,٤٧	- ١,٢٠	١,١٢	٢٤	٠,٦٥	٠,٤٢	- ٠,٩٣	١,٢٤
١٠	٠,٣٨	٠,١٢	- ٠,٨٣	١,١٠	٢٥	٠,٥٩	٠,٤٣	- ٠,٧٧	١,٦٥
١١	٠,٥٦	٠,٥٧	٠,٣٣	١,٥٠	٢٦	٠,٧١	٠,٤٠	- ١,٠١	١,٠٩
١٢	٠,٧١	٠,٤٩	- ١,٠٦	١,٢٤	٢٧	٠,٥٩	٠,٥٩	- ٠,٦١	١,٢٤
١٣	٠,٤٢	٠,٤١	- ٠,٥٤	١,٦٢	٢٨	٠,٨٣	٠,٤٧	- ١,٣٠	١,٣٣
١٤	٠,٤٢	٠,٦٦	- ٠,٣٣	١,٠٨	٢٩	٠,٤٨	٠,٥٣	- ١,١٣	١,٤٨
١٥	٠,٥١	٠,٣٩	- ٠,٩٨	١,٢٣	٣٠	٠,٧٧	٠,٤١	- ١,٠٣	٠,٧٢

ويلاحظ من نتائج الجدول رقم(٣) والمتعلق بالخصائص السيكومترية للفقرات في ضوء النظرية الكلاسيكية في القياس أنّ قيم معاملات الصعوبة قد تراوحت من(٠,٣٨ ـ ٠,٨٣) بمتوسط حسابي مقداره(٠,٥٧) وانحراف معياري(٠,١٢). وقد تراوحت قيم معاملات التمييز بين(٠,١٢ ــ ٠,٧١) بمتوسط حسابي مقداره(٠,٤٧) وانحراف معياري(٠,١١). وبعد النظر بالفقرات التي تحقق الإحصائيات المتبعة في هذه الدراسة، وهي الإحصائيات المقترحة من قبل(عودة، ٢٠١٠م، ص: ٢٨١ ـ ٢٨٥) والتي تتلخص بأن المدى المقبول لصعوبة وتمييز الفقرة يتراوح بين(٠,٢٠ـ ٠,٨٠)، وأنّ أي فقرة معامل تمييزها أعلى من(٠,٣٩) تعتبر فقرة جيدة، وأنّ الفقرات التي معامل تمييزها أقل من(٠,٢٠) تعتبر فقرة ضعيفة. وفي ضوء المعايير السابقة فإن الفقرات التي يزيد معامل صعوبتها على (٠,٨٠) عدت سهلة في هذا الاختبار وهي الفقرات (٥، ٢٨) وقد تم الاحتفاظ بهاتين الفقرتين لأن معاملات تمييزها جيدة وضمن المدى المقبول، بالإضافة إلى ضرورة المحافظة على صدق محتوى الاختبار. وقد جرى حذف الفقرة رقم(١٠) من نموذج الاختبار كون معامل تمييزها يقل عن(٠,٢٠) ويساوي(٠,١٢). وبذلك تم الإبقاء على (٢٩) فقرة أي ما نسبته (٩٦,٦٧٪)، في حين كانت هناك فقرة واحدة غير مطابقة وهي الفقرة رقم(١٠) من فقرات الاختبار بنسبة بلغت(٣,٣٣٪).

ويلاحظ في نِسبة الفقرات المطابقة للنظرية التقليدية في القياس في الدراسة الحالية مقارنة بالدراسات السابقة، وجود اختلاف واضح حيث بلغت عدد الفقرات المطابقة في دراسة أون (Onn, 2013) (٢٩) فقرة من بين (٥٠) فقرة أي ما نسبته(٠,٥٨)، وأشارت نتائج دراسة ياسين(٢٠٠٤م) إلى مطابقة (٤٥) فقرة من بين(٥٢) فقرة أي ما نسبته (٨٦,٥٤٪)، وأظهرت نتائج دراسة أبوهاشم(٢٠٠٦م) إلى مطابقة (٢٦) فقرة من بين(٢٨) فقرة أي ما نسبته (٩٢,٢٩٪). كما وتختلف النسبة بالدراسة الحالية مع نتائج دراسة جايميلو وسلفستر(Jimelo and Silvestre, 2009) حيث أشارت إلى أنّ(٣٣) فقرة أي ما نسبته(٥٥٪) من بين(٦٠) فقرة طابقت النظرية التقليدية في القياس. ويفسر الباحث سبب ذلك إلى اختلاف حجوم عينات التحليل، وعدد الفقرات المستخدمة في هذه الدراسات، كذلك إلى قصور النظرية التقليدية في القياس حيث تتأثر خصائص فقرات الاختبار بعينة الأفراد، حيث تختلف قيم معاملات الصعوبة والتمييز لفقرات الاختبار باختلاف حجم العينة.

ويلاحظ من نتائج الجدول رقم(٣) سابق الذكر والمتعلقة بالخصائص السيكومترية لفقرات الاختبار في ضوء النظرية الحديثة في القياس باستخدام النموذج اللوجستي ثنائي المعلمة أن قيم معاملات الصعوبة قد تراوحت من(ـ ١,٤٩ـ ٠,٥٤) بمتوسط حسابي مقداره(ـ٠,٦٦) وانحراف معياري(٠,٦٠)، وتراوحت قيم الأخطاء المعياري لتقديرات معلمة الصعوبة بين(٠,٠٤٤ ـ ٠,٥٠) بمتوسط حسابي مقداره(٠,١٩) وانحراف معياري مقداره(٠,١٢). وقد تراوحت قيم معاملات التمييز لفقرات الاختبار بين(٠,٣٠ ــ ٢,٨٣) بمتوسط حسابي مقداره(١,٣٢)

وانحراف معياري(٠,٥٠)، وتراوحت قيم الأخطاء المعياري لتقديرات معلمة التمييز بين(٠,٢١ ــ ٢,٧٦) بمتوسط حسابي مقداره(٠,٨٣) وانحراف معياري مقداره(٠,٧٩). وقد لوحظ وباستخدام اختبار مربع كآي لحسن المطابقة عند مستوى الدلالة(α =0.01) أن الفقرتين(١٠، ١٧) لم تكونا متطابقتين مع النموذج اللوجستي ثنائي المعلمة حيث كانت قيمة الاحتمالية لها أقل من (0.01)، وتشكل هاتان الفقرتان ما نسبته(٦,٦٧٪) من بين فقرات الاختبار، وجرى استبعاد هاتين الفقرتين من الاختبار، بينما تم الإبقاء على الفقرات الأخرى التي طابقت النموذج المستخدم والبالغ عددها(٢٨) فقرة أي ما نسبته(٩٣,٣٣٪).

وبالاعتماد على النظرية الحديثة في القياس فإنّ الفقرات التي طابقت النموذج اللوجستي ثنائي المعلمة تعّد ذات استقلال إحصائي عن بعضها البعض وعن عينة المفحوصين. ويلاحظ من نتائج التحليل وجود اختلاف واضح في نسبة الفقرات المطابقة للنموذج اللوجستي ثنائي المعلمة في هذه الدراسة مقارنة بنسبة الفقرات المطابقة في بعض الدراسات السابقة والتي استخدمت النموذج اللوجستي ثنائي المعلمة، حيث كانت عدد الفقرات المطابقة في دراسة أون (Onn, 2013) (٣٨) فقرة من بين(٥٠) فقرة أي ما نسبته(٧٦٪)، كما أظهرت دراسة جايميلو وسلفستر(Jimelo and Silvestre, 2009) إلى أنّ(٣٣) فقرة أي ما نسبته(٥٥٪) من بين(٦٠) فقرة طابقت النظرية الحديثة في القياس.وقد يعود السبب إلى اختلاف حجم عينة التحليل وعدد الفقرات المستخدمة في هذه الدراسات، واختلاف صيغة فقرات الاختبار ومحتوى مادة الاختبار، بالإضافة إلى اختلاف برامج التحليل المستخدمة في هذه الدراسات حيث استخدمت الدراستين برنامج X-Calibreprograme وWINSTEPS، بينما استخدمت الدراسة الحالية برنامج MG3-BILOG .

وبالرغم من اختلاف نوعية وأهداف الاختبارات المستخدمة في الدراسات السابقة والدراسة الحالية، إلا أن فقرات الاختبار التي يتم انتقاؤها وفقا لنوعي النماذج تختلف إذا كان مدى تمييز فقرات الاختبار متسعا، فالنظرية التقليدية في القياس تستبعد فقط الفقرات ذات التمييز المنخفض، أما نماذج الاستجابة للمفردة فتستبعد المفردات ذات التمييز المتطرف(المنخفض أو المرتفع)، وهذا التفاوت فى انتقاء الفقرات ربما يرجع إلى اختلاف الهدف من تحليل الفقرات فى كلٍ من الحالتين، فالهدف فى حالة النظرية الكلاسيكية في القياس يكون موجها نحو الكشف عن الفقرات الاختبارية غير الصالحة، وتقدير البارامترات المتعلقة بكل مفردة واستخدامها فى تقدير الخصائص السيكومترية للاختبار ككل، أما في حالة نماذج الاستجابة للمفردة فيكون الهدف موجها نحو الحصول على تدريج لمفردات الاختبار Items Calibration ، أي تقدير صعوبة المفردات Items Difficulty وتقييم حسن مطابقة درجاتها للنموذج المستخدم والإفادة من ذلك في تقدير سمات الأفراد.

النتائج المتعلقة بالسؤال الثاني ومناقشتها: ما مدى التوافق بين النظرية التقليدية في القياس والنموذج اللوجستي ثنائي المعلمة في انتقاء فقرات اختبار محكي المرجع؟

قام الباحث بإجراء مقارنة بين المؤشرات الإحصائية المستخرجة وفقاً للنظرية الحديثة في القياس ممثلة بالنموذج اللوجستي ثنائي المعلمة وبين المؤشرات الإحصائية للنظرية التقليدية في القياس. وقد تم استخراج عدد الفقرات ونسبتها في كلا النظريتين. ويبين الجدول رقم(٤) عدد الفقرات ونسبتها حسب مطابقتها لكلا النظريتين:

الجدول رقم(٤)عدد الفقرات ونسبتها حسب مطابقتها وفقاً للنظريتين.

النظرية الحديثة في القياس/ النموذج اللوجستي ثنائي المعلمة	النظرية التقليدية في القياس						النسبة
	حالة الفقرة	مطابقة		غير مطابقة		عدد الفقرات	
		عدد الفقرات	نسبتها	عدد الفقرات	نسبتها		
	مطابقة	٢٨	٩٣.٣٣٪	صفر	٠٪	٢٨	٩٣.٣٣٪
	غير مطابقة	١	٣.٣٣٪	١	٣.٣٣٪	٢	٦.٦٧٪
الكلي		٢٩	٩٦.٦٦٪	١	٣.٣٣٪	٣٠	١٠٠٪

ويلاحظ من الجدول رقم(٤) النتائج الآتية:

- **يوجد اتفاق بين المؤشرات الإحصائية المستخرجة وفق النموذج اللوجستي ثنائي المعلمة والنظرية التقليدية في القياس من حيث استبعاد كلا النظريتين لفقرة واحدة فقط من نموذج الاختبار محكي المرجع.**
- **يوجد اتفاق بين المؤشرات التقليدية وبين النموذج اللوجستي ثنائي المعلمة في استبقاء (٢٨) فقرة من فقرات الاختبار محكي المرجع.**
- **لا يوجد اتفاق بين المؤشرات الإحصائية للنظرية التقليدية وبين النموذج اللوجستي ثنائي المعلمة لفقرة واحدة، حيث تم استبعادها وفقا للنموذج اللوجستي ثنائي المعلمة بينما تم استبقاؤها وفقاً للنظرية التقليدية في القياس وهي الفقرة رقم(١٧) في نموذج الاختبار محكي المرجع.**
- **طابقت(٢٩) فقرة للنظرية التقليدية في القياس.**

وبالرغم من اختلاف النموذج المستخدم في هذه الدراسة وكذلك عدد الفقرات المستخدمه مقارنة مع بعض الدراسات السابقة مثل دراسة جايميلو وسلفستر(Jimelo and Silvestre, 2009) التي استخدمت نموذج راش في التحليل، إلا أن نتائج الدراستين تتفقان من وجود عدد من الفقرات المطابقة في كلا النظريتين، وأظهرت نتائج دراسة ستيج(Stage,2003) أنّ تحليل الفقرة في ضوء نظرية الاستجابة للفقرة أفضل مقارنة بالنظرية الكلاسيكية في القياس.

النتائج المتعلقة بالسؤال الثالث ومناقشتها: ما الخصائص السيكومترية للاختبار محكي المرجع وفق النظرية التقليدية في القياس والنموذج اللوجستي ثنائي المعلمة؟

- **معامل الصدق** :

جرى تقدير معلمة القدرة باستخدام طريقة الأرجحية العظمى الهامشية (MML) بالاعتماد على برنامج (BILOG – MG)، وتراوحت قيم معلمة القدرة الناتجة من تطبيق نموذج الاختبار محكي المرجع بين (0.902 - 2.2901)، وقد تراوحت قيم الأخطاء المعيارية لتقديرات معلمة القدرة بين (0.1387–3.9343)، بمتوسط حسابي مقداره (0.8963) بانحراف معياري(1.839). وجرى استخراج معاملٍ ارتباطٍ بيرسون بين درجات الطلبة المدرسية في مادة الرياضيات والذي يعدّ محكاً خارجياً وبين نتائجهم على الاختبار المحكي المستخدم في هذه الدراسة. كما جرى حساب معامل ارتباط بيرسون بين معلمة القدرة والعلامات الخام، نظراً لأنّ برنامج (BILOG – MG) يعطينا تقديرات للقدرة بوسط حسابي مقداره صفراً، وانحرف معياري مقداره واحد، وهذا يجعل اللجوء إلى المقارنة اعتماداً على المتوسطات الحسابية للقدرة يعطينا دائماً صفراً؛ لذا تمّ اللجوء إلى حساب معامل ارتباط بيرسون. ويبين الجدول رقم(٥) قيم معاملات الصدق المحكي للاختبار قبل حذف الفقرات غير المطابقة وبعد حذفها وفقاً للنظريتين الحديثة والتقليدية في القياس.

الجدول رقم(٥)قيم معاملات الصدق المحكي للاختبار قبل وبعد حذف الفقرات غير المطابقة للنظريتين

التحليل الإحصائي وفق	حالة فقرات الاختبار محكي المرجع				عدد الطلبة المطابقين
	قبل حذف الفقرات		بعد حذف الفقرات		
	قيمة معامل الصدق	عدد الفقرات	قيمة معامل الصدق	عدد الفقرات	
النظرية الحديثة في القياس/ النموذج اللوجستي ثنائي المعلمة.	٠,٩٠٥	٣٠	٠,٩٢٥	٢٨	١٣٦
النظرية التقليدية في القياس.	٠,٩٢١	٣٠	٠,٩٣٦	٢٩	١٤٠

ويلاحظ من نتائج الجدول رقم(٥) أن قيمة معامل الصدق المحكي للاختبار قد ارتفعت بعد حذف الفقرات غير المطابقة وذلك وفقاً للنظريتين الحديثة والتقليدية في القياس. وقد بلغت قيمة معامل الصدق المحكي للاختبار قبل حذف الفقرات غير المطابقة وفقاً للنظرية الحديثة في القياس(٠,٩٠٥) وأصبحت (٠,٩٢٥) بعد حذف الفقرات غير المطابقة للنموذج اللوجستي ثنائي المعلمة. وبلغت قيمة معامل الصدق المحكي للاختبار قبل حذف الفقرات غير المطابقة وفقاً للنظرية التقليدية في القياس(٠,٩٢١) وأصبحت (٠,٩٣٦) بعد حذف الفقرات غير المطابقة لها. ويفسر الباحث سبب ارتفاع قيمة معامل الصدق في كلا النظريتين بعد حذف الفقرات غير المطابقة إلى انخفاض الخطأ المعياري في التقدير.

- **معامل الثبات** :

جرى استخراج معامل ثبات الاختبار بإستخدام معادلة كيودر– ريتشاردسون ٢٠(KR-20) إذ يعد هذا المؤشر متطلباً لمعادلة الثبات لمعامل ليفنجستون

الخاصة بالاختبارات محكية المرجع. كما جرى استخراج قيمة معامل الثبات الإمبريقي(التجريبي) باستخدام برنامج BILOG-MG3. ويبين الجدول رقم(٦) قيم معاملات الثبات قبل حذف الفقرات غير المطابقة وبعد حذفها وفقا للنظريتين الحديثة والتقليدية في القياس.

الجدول رقم(٦) قيم معاملات الثبات للاختبار قبل وبعد حذف الفقرات غير المطابقة للنظريتين

التحليل الإحصائي وفق	معامل الثبات	حالة فقرات الاختبار محكي المرجع				عدد الطلبة المطابقين
		قبل حذف الفقرات		بعد حذف الفقرات		
		قيمة معامل الثبات	عدد الفقرات	قيمة معامل الثبات	عدد الفقرات	
النظرية الحديثة في القياس/ النموذج اللوجستي ثنائي المعلمة	التجريبي	٠,٩٤٥٥	٣٠	٠,٩٦٥٨	٢٨	١٣٦
التقليدية في القياس	KR-20	٠,٩٣١٠	٣٠	٠,٩٥٤٩	٢٩	١٤٠

ويلاحظ من الجدول رقم(٦) سابق الذكر أن قيمة معامل ثبات الاختبار قبل حذف الفقرات غير المطابقة وفق النموذج اللوجستي ثنائي المعلمة أعلى وتساوي (٠,٩٤٥٥) مقارنة بالنظرية التقليدية في القياس والذي بلغت قيمته(٠,٩٣١٠)، وكانت قيمة معامل ثبات الاختبار بعد حذف الفقرات غير المطابقة وفق النموذج اللوجستي ثنائي المعلمة أعلى وتساوي (٠,٩٦٥٨) مقارنة بالنظرية التقليدية في القياس والذي بلغت قيمته(٠,٩٥٤٩). وتتناقض هذه النتيجة مع نتيجة دراسة أون(Onn, 2013) التي أظهرت تدني معامل الثبات في كلا النظريتين، ويفسر الباحث ذلك نظرا لأن عدد الفقرات التي طابقت في دراسة أون كانت (٢٩) فقرة وفق النظرية التقليدية و(٣٨) فقرة طابقت للنموذج اللوجستي ثنائي المعلمة من بين (٥٠) فقرة، بينما في الدراسة الحالية طابقت (٢٨) فقرة للنموذج اللوجستي ثنائي المعلمة، و(٢٩) فقرة طابقت للنظرية التقليدية في القياس من بين (٣٠) فقرة، حيث أن زيادة عدد الفقرات تؤدي إلى ارتفاع في قيمة معامل الثبات. وعند النظر في قيم معاملات التمييز للفقرات وفق النظرية التقليدية في القياس(جدول رقم ٣) نجد أنها كانت لجميع فقرات الاختبار محكي المرجع أكبر من (0.20)، وتدل هذه النتيجة وفقا لمحك نانلي (Nunnally, 1970)، أنّ الفقرات في نموذج الاختبار تشترك جميعها في قياس بعد واحد تعبر عنه الدرجة الكلية. وبالرغم من اختلاف عدد الفقرات المطابقة في الدراسة الحالية مع دراسة جايميلو وسلفستر(Jimelo and Silvestre, 2009)، إلا أنه يلاحظ وجود تقارب في قيمة معامل الثبات المستخرجة في كلا النظريتين.

النتائج المتعلقة بالسؤال الرابع ومناقشتها: هل يوجد فرق ذو دلالة إحصائية عند مستوى الدلالة الاحصائية(α = ٠,٠٥) بين معاملي الثبات المقدّرين باستخدام النظرية التقليدية في القياس والنموذج اللوجستي ثنائي المعلمة؟

وللكشف عن دالة الفرق بين معاملي الثبات لعينتين مترابطتين استخدم الباحث معادلة السوالمة وفلدت (Alsawalmeh & Feldt, 1994). ويبين الجدول رقم(٧) نتائج الإحصائي(W):

الجدول رقم (٧) نتائج الإحصائي(F) لفحص مدى الاختلاف بين معاملي الثبات وفق النظريتين

التحليل الإحصائي وفق	معامل الثبات	عدد الفقرات	عدد الطلبة	قيمة(W) المحسوبة	قيمة(F) الجدولية
النموذج اللوجستي ثنائي المعلمة	٠,٩٦٥٨	٢٨	١٣٦	١,٣٦	١,١١
التقليدية في القياس	٠,٩٥٤٩	٢٩	١٤٠		

ويلاحظ من الجدول رقم (٧) وجود فرق ذي دلالة إحصائية عند مستوى الدلالة الاحصائية(α = ٠,٠٥) في تقدير معاملي الثبات وفق النظريتين ولصالح النموذج اللوجستي ثنائي المعلمة حيث بلغت قيمة اختبار(W) المحسوبة(١,٣٦) وهذه القيمة أكبر من قيمة (F) الجدولية(١,١١) عند مستوى الدلالة الاحصائية(α = ٠,٠٥). ويمكن تفسير هذه النتيجة من خلال طريقة استخراج الثبات التجريبي، إذ تعتمد طريقة الحساب بإيجاد ناتج قسمة تباين العلامة الحقيقية على تباين العلامة الكلية، حيث تباين العلامة هو تباين العلامات المستخرج من خلال الأرجحية العظمى، أما التباين الحقيقي فيتم تقديره عن طريق طرح تباين الخطأ من تباين العلامة الكلية، وبالتالي يمكن عزو ارتفاع معامل الثبات التجريبي مقارنة بالنظرية التقليدية في القياس إلى انخفاض تباين الخطأ. وتتناقض هذه النتيجة مع دراسة جمحاوي(٢٠٠٠م) التي أظهرت عدم وجود فروق ذات دلالة إحصائية بين معاملي ثبات الاختبار الناتج من النظرية التقليدية والنموذج ثنائي المعلمة.

النتائج المتعلقة بالسؤال الخامس ومناقشتها: هل يوجد فرق ذو دلالة إحصائية عند مستوى الدلالة الاحصائية(α = ٠,٠٥) بين معاملي الصدق المقدّرين باستخدام النظرية التقليدية في القياس والنموذج اللوجستي ثنائي المعلمة؟

وللكشف عن دالة الفرق بين معاملي الصدق لعينتين مترابطتين، استخدم الباحث معادلة السوالمة وفلدت (Alsawalmeh & Feldt, 1994) سابقة الذكر، إذ تصلح هذه المعادلة للمقارنة كذلك بين معاملي الصدق. ويبين الجدول رقم(٨) نتائج التحليل.

ويلاحظ من نتائج الجدول رقم(٨) وجود فرق ذي دلالةٍ إحصائية عند مستوى الدلالة الاحصائية(α = ٠,٠٥) بين معاملي الصدق وفقاً للنظريتين حيث بلغت قيمة(W) المحسوبة (١,٢١) وهذه القيمة أكبر من قيمة(F) الجدولية (١,١١)،

وقد كان الفرق لصالح معامل الصدق المحسوب باستخدام النظرية التقليدية في القياس ويساوي(٠,٩٣٦) وهو أكبر من معامل الصدق المحكي المحسوب باستخدام النموذج اللوجستي ثنائي المعلمة ويساوي(٠,٩٢٥). وبالرغم من ارتفاع قيمة معاملي الصدق إلا أنّ قيمة معامل الصدق في النظرية التقليدية أعلى مقارنة بالنموذج اللوجستي ثنائية المعلمة ويفسر الباحث هذه النتيجة إلى وجود الارتباطات العالية بين فقرات الاختبار، والارتباطات العالية بين الفقرات وهو مؤشر على أن ما تقيسه كل من هذه الفقرات هو ما يقيسه الاختبار، وبالتالي يسهم في رفع قيمة معامل الارتباط، كما إن الفقرات المتجانسة تزيد من الارتباطات الداخلية لفقرات الاختبار، أي أنّ الفقرات المتجانسة تشترك جميعها في قياس سمة واحدة وهي تحصيل الطلبة في وحدة الهندسة التحليلية.

الجدول رقم(٨)نتائج اختبار (سوالمة وفلدت) للكشف عن الفروق الثنائية بين معاملي الصدق المحكي وفق النظريتين.

التحليل الإحصائي وفق	معامل الصدق المحكي	عدد الفقرات	عدد الطلبة	قيمة(W) المحسوبة	قيمة(F) الجدولية
النموذج اللوجستي ثنائي المعلمة	٠,٩٢٥	٢٨	١٣٦	١,٢١	١,١١
التقليدية في القياس	٠,٩٣٦	٢٩	١٤٠		

وبالرغم من وجود العديد من المشكلات التي تعاني منها النظرية التقليدية في القياس، إلا أن نتائج الدراسة أظهرت وجود اتفاق بين النظرية التقليدية في القياس والنموذج اللوجستي ثنائي المعلمة فيما يتعلق باستبعاد فقرة مشتركة واحدة من نموذج الاختبار محكي المرجع هي الفقرة رقم(١٠)، وبالتالي فإن تحليل الفقرة في ضوء النموذج اللوجستي ثنائي المعلمة يتقارب مع النظرية التقليدية في القياس، حيث بلغت عدد الفقرات المطابقة وفق النظريتين (٢٨) فقرة وتشكل ما نسبته (٩٣,٣٣٪) من بين(٣٠) فقرة وهذه النسبة مرتفعة. وحيث وجد هناك فرق ذو دلالة إحصائية في تقدير معاملي الثبات بعد حذف الفقرات غير المطابقة وفق النظريتين ولصالح النموذج اللوجستي ثنائي المعلمة، إلا أن معامل الثبات المحسوب وفق النظرية التقليدية كان مرتفع أيضا. وبالرغم من ارتفاع قيم معاملي الصدق في كلا النظريتين إلا أن النتائج أظهرت وجود فرق ذي دلالة إحصائية بين معاملي الصدق وفقا للنظريتين لصالح النظرية التقليدية في القياس. وبالتالي لايوجد مبرر كاف للاعتماد على نظرية قياس دون الأخرى، وبالتالي فإن الاعتماد على النموذج اللوجستي ثنائي المعلمة فقط في انتقاء الفقرات لا يعطي تدعيما قويا مقارنة بالنظرية التقليدية في القياس.

وعند النظر بمتطلبات نماذج النظرية الحديثة في القياس ممثلة بالنموذج اللوجستي ثنائي المعلمة نرى أنها تعتمد على مجموعة من الافتراضات لابّد من تحقيقها، ومن الصعب في بعض الأحيان تحقيقها، كما أن نماذج النظرية

الحديثة في القياس المختلفة والتي تعتمد على اقترانات رياضية لوغاريتمية معقدة وصعبة الفهم عند البعض، كذلك برامجها الإحصائية المختلفة، لتقود البعض في ضوء نتائج هذه الدراسة إلى الاعتماد على النظرية التقليدية في القياس، وبالتالي يوصي الباحث في ضوء نتائج الدراسة الحالية إلى إمكانية الاعتماد على النظرية التقليدية في القياس نظراً لسهولة التعامل معها، حيث تتوافر جوانب اتفاق بين النظرية التقليدية والنموذج اللوجستي ثنائي المعلمة. كما توصي الدراسة الباحثين بضرورة دراسة التوافق بين النظرية التقليدية في القياس والنماذج المختلفة لنظرية استجابة الفقرة.

- **المراجع** :

- أبو هاشم، السيد محمد.(٢٠٠٦م). دراسة مقارنة بين النظرية التقليدية ونموذج راش فى اختيار فقرات قائمة مداخل الدراسة لدى طلاب الجامعة. مجلة كلية التربية، جامعة الزقازيق، العدد(٥٢) يناير.
- جمحاوي، ايناس.(٢٠٠٠م). مقارنة خصائص الفقرات وفق النظرية التقليدية ونظرية استجابة الفقرة في مقياس للقدرة العقلية. رسالة ماجستير غير منشورة، جامعة اليرموك، الأردن.
- الشريفين، نضال. (٢٠٠٦). الخصائص السيكومترية لاختبار محكي المرجع في القياس والتقويم التربوي وفق النظرية الحديثة في القياس والتقويم التربوي والنفسي. مجلة العلوم التربوية والنفسية– جامعة البحرين، ٧(٤): ٨ – ١٠٩
- علام ، صلاح الدين محمود.(٢٠٠١م). الاختبارات التشخيصية مرجعية المحك في المجالات التربوية والنفسية(الطبعة الثانية). القاهرة: دار الفكر العربي.
- علام، صلاح الدين محمود.(٢٠٠٠م). القياس والتقويم التربوى والنفسى– أساسياته وتطبيقاته وتوجهاته المعاصرة. القاهرة: دار الفكر العربي.
- علام ، صلاح الدين محمود.(٢٠٠٥م). نماذج الاستجابة للمفردة الاختبارية أحادية البعد ومتعددة الأبعاد. القاهرة: دار الفكر العربي.
- علام ، صلاح الدين محمود.(٢٠٠٦م). الاختبارات والمقاييس التربوية والنفسية(الطبعة الأولى). القاهرة: دار الفكر العربي.
- عيد، غادة خالد.(٢٠٠٤م). الدرجة الحقيقية باستخدام نظرية السمات الكامنة والنظرية الكلاسيكية: دراسة سيكومترية. مجلة جامعة أم القرى للعلوم التربوية والاجتماعية والإنسانية، العدد(٢)، المجلد(١٦)، ص ٢٣٠ - ٢٨٧.
- عودة، أحمد سليمان. (٢٠١٠م). القياس والتقويم في العملية التدريسية. اربد: دار الأمل.
- كاظم، أمينة محمد.(١٩٨٨م). استخدام نموذج راش في بناء اختبار تحصيلي في علم النفس وتحقيق التفسير الموضوعي للنتائج. الكويت: مؤسسة الكويت للتقدم العلمي.
- ياسين، عمر صالح مفضي.(٢٠٠٤م). الخصائص السيكومترية لاختبار محكي المرجع في الكيمياء لطلاب الصف الأول الثانوي العلمي مقدرة وفق النظريتين الكلاسيكية والحديثة في القياس. رسالة ماجستير غير منشورة، جامعة عمان العربية للدراسات العليا، عمان.

- Alsawalmeh, Y. & Feldt, L. S. (1994). A Modification Of Feldt's Test Of The Equality Of Two Dependent Alpha Coefficients. Psychometrica, 59(1), 49-57.

- Berk, A.(1980). Criterion – Referenced Measurement the state of the Art. USA, Originally.

- Dato N. M. de Gruijter & Leo J. Th. van der Kamp (2005). Statistical Test Theory for Education and Psychology, Retrieved Nov., 2012. from: http://irt.com.ne.kr/data/test_theory.pdf
- Embretson, S. E. and Reise, S. P. (2000). Item response theory for psychologists. New Jersey: Lawrence Erlbaum Associates, Inc.
- Hambleton, R., and Swaminathan, H. (1985). Item Response Theory: principles and applications. Boston:Kluwer Nijhoff publishing.
- Hambleton, K., Swaminathan, H. and Rogers, H. (1999). Fundamentals of Item Response Theory. Sage Publication, Newbury park, CA.
- Hernandez, R. (2009). Comparison of the Item Discrimination and Item Difficulty of theQuick-Mental Aptitude Test using CTT and IRT Methods. The International Journal of Educational Employment and Psychological Assessment, 1 (1), 12 – 18.
- Hwang, D.Y. (2002). Classical Test Theory and Item Response Theory: Analytical and Empirical Comparison. Paper presented at the Annual Meeting of the Southwest Educational Research Association, Austin, TX. (ERIC Document Reproduction Service No. ED 466779).
- Jager, R. (1989). Certification of Student Competence. In Robert L. Linn (ed.). Educational Measurement. Collier Macmillan Publishers London.
- Jimelo, L. and Silvestre, T.(2009). Item Response Theory and classical Test Theory: An Empirical Comparison of Item/Person statistics in A Biological Science Test. The International Journal of Educational and Psychological Assessment, 1(1), 19 – 31.
- Kan, A.(2006). Classical Test Theory and Latent Trait Theory Predicted by Empirical Working on Item Parameters. Mersin Universitesi Egitim Fakultesi Dergisi, 2(2), 227 – 235.
- 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.
- Owens, J. (2000). Using a Brand Loyalty Scale to Compare Item Selection Criteria and Scale Evaluation Techniques of Classical Test

and Item Response Theory. PHD Degree, Southern Illinois University.

- Pelton, W. (2002). The accuracy of unidimensional measurement models in the presence of deviations from the underlying assumptions. Unpublished doctoral dissertation, Brigham Young University,U.S.A.
- Randall ,S.(1998). Comparing Measurement Theories . Paper Present at the Annual Meeting of the American Educational Research Association (San Diego , CA , April 13-17).
- Reckase. D. (1979). Unifactor latent trait models applied to multifactor tests: Results and implications. Journal of Educational Statistics, 4, 207 – 230.
- Stage, C. (2003). Classical Test Theory or Item Response Theory: The Swedish Experience. Umea University,pp: 1 - 30.
- Sizmur, S. (1997). Look Back in Angoff : A Cautionary Tale. British Educational Research Journal, 23, 3-11. Retrieved: 20th, Feb.2007 from: EBSCOhost Research database.
- Thorendike, M. (1997). Measurement and Evaluation In Psychology And Education(6rd.ed.). Upper Saddle River, New Jersey, Merrill.

****