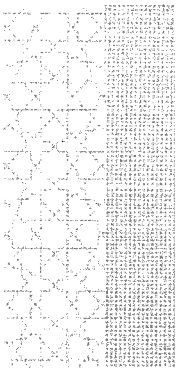
فاعلية استخدام كل من نظرية السمات الكامنة والنظرية الكلاسيكية في قياس التحصيل الدراسي في الجبر بالمرحلة الإعدادية - دراسة مقارنة

د.محمد أنور إبراهيم فراج
أستاذ علم النفس التربوي المشارك –
كلية التربية جامعة الإسكندرية ، كلية
المعلمين جامعة الملك سعود

د.خالد حسن الشريف
مدرس علم النفس التربوي كلية
التربية جامعة الإسكندرية



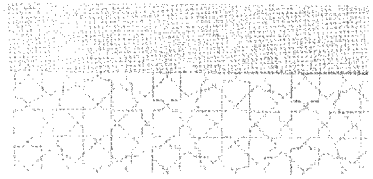
فاعلية استخدام كل من نظرية السمات الكامنة و النظرية الكلاسيكية في قياس التحصيل الدراسي في الجبر بالمرحلة الإعدادية - دراسة مقارنة

د.محمد أنور إبراهيم فراج
أستاذ علم النفس التربوي المشارك – كلية التربية جامعة الإسكندرية ، كلية المعلمين جامعة الملك سعود

د.خالد حسن الشريف
مدرس علم النفس التربوي كلية التربية جامعة الإسكندرية

ملخص الدراسة:

إن دقة النتائج الواردة من الاختبارات التحصيلية المبنية على أساس النظرية الكلاسيكية في القياس كانت مثار جدل بين الباحثين التربويين والمتخصصين في مجال القياس والتقويم في ضوء العوامل التي تؤثر في تقدير الدرجة الحقيقية للمتعلم باستخدام هذا النموذج الكلاسيكي ، وهو ما يدعو إلى استخدام أحد نماذج نظرية القياس الحديثة(السمات الكامنة) وذلك للتحقق من فاعلية كلا النموذجين في تقدير المستوى التحصيلي للمتعلمين بدقة ، وهوهدف أساسي للعديد من الدراسات المقارنة بين النظرية الكلاسيكية ونظرية السمات الكامنة ، وقد قام الباحثان بتصميم اختبار تحصيلي مرجعي الهدف مكون من ٢٥ مفردة في مادة الجبر وقد تم تطبيقه على عينة مكونة من ٨٠ تلميذاً من تلاميذ الصف الأول الإعدادي بمحافظة الإسكندرية ، وقد تم تحليل البيانات بطريقتين مختلفتين: الأولى في ضوء النظرية الكلاسيكية واعتمدت الدرجة الكلية كمحك ، والثانية في ضوء نظرية الاستجابة للمفردة الاختبارية وقد اعتمدت قدرة التلميذ على تحقيق الهدف الذي تقيسه المفردة كمحك. وقد أشارت النتائج إلى تمتع الاختبار التحصيلي مرجعي الهدف بمؤشرات صدق وثبات مناسبة كذلك أكدت النتائج قابلية البيانات للتحليل باستخدام نظرية الاستجابة للمفردة ، وقد كانت البيانات المستمدة من النموذجين قابلة للمقارنة ، وأن نظرية السمات الكامنة كانت أكثر دقة في تحديد قدرة المفحوصين على الأداء على مفردات الاختبار وبأخطاء معيارية أقل في القياس من نظيراتها في النظرية الكلاسيكية، وكذلك اختلف النموذجان في تحديد مؤشرات صعوبة مفردات الاختبار.



مقدمة

يتضمن القياس التربوي عملية المقارنة ، حيث تتم المقارنة بين الأفراد وبعضهم بالطرق الكمية التي تستخدم النظرية الكلاسيكية أو النظرية الحديثة. وتستخدم في ذلك مفردات الاختبارات التي تقدر قدرة الفرد المعرفية وترى نادية محمد عبد السلام(١٩٩٢) أن الاختبارات يمكن أن تستخدم في تقدير المستويات المختلفة لمتغير الأداء أو متغير المهارة (ص ٨٨). حيث يهتم مصممو الاختبارات بكفاءة مفردات الاختبار، وكيف يجيب عليها المفحوصون عند تطبيق هذه الاختبارات عليهم ، وغالباً ما يستخدم المشتغلون بالقياس النفسي مدخلاً من مدخلين أساسيين في تحليل بيانات الاختبارات:- نظرية الاختبار الكلاسيكي Classical Test Theory (CTT) ، ونظرية الاستجابة للمفردة Item Response Theory (IRT). والهدف الأساسي لكل من النظرية الكلاسيكية ونظريةالسمات الكامنة هو تحديد العلاقة بين استجابات الأفراد على اختبار معين والسمة الكامنة وراء هذه الاستجابات ، والأمر الأكثر أهمية في القياس النفسي والتربوي بعامة هو تحديد مقدار السمات الكامنة وراء أداء الأفرادعلى الاختبارات المختلفة والاستفادة منها في تفسير النتائج والتنبؤ بسلوكهم في مواقف مماثلة عن طريق تحديد معاملي صعوبة المفردة ، والقدرة لمن يتم تطبيق الاختبارات عليهم ، كما أن كلا النظريتين تهتم بتحسين معاملات ثبات وصدق الاختبارات من خلال تقديم طرق متنوعة لحساب معاملات الثبات والصدق ، وكل ذلك يساعد على اتخاذ قرارات صائبة بشأن الأفراد الذين تطبق عليهم هذه الاختبارات في ضوء هذا التقديرالكمي للسمات.

و الهدف الأساسي للنظرية الكلاسيكية هو قياس التحصيل الدراسي في المقررات الدراسية للمتعلمين بدقة بحيث يتم من خلاله تقدير المستوى التحصيلي الحقيقي للمتعلم ؛ غير أن دقة نتائج الاختبارات التحصيلية المبنية في ضوء هذه النظرية كانت مثار جدل بين الباحثين التربويين لاسيما في السنوات الأخيرة نظراً لتدخل ذاتية الفاحص في تصميم هذه الاختبارات وتطبيقها وكذلك في تحليل نتائجها ؛ الأمر الذي أثارالشك في دقة تعبير نتائجها عن القدرات التحصيلية الحقيقية للمتعلم. خصوصاً أنه كلما زادت قيم الأخطاء المعيارية للقياس كلما قلت درجة موثوقية الدرجة المقاسة للخاصية أو السمة.

ورغم اجتهاد الباحثين في التحقق من ملاءمة الاختبارات المصممة في ضوء النظرية الكلاسيكية للتطبيق في مستويات عمرية متباينة ، من خلال حساب معاملات الثبات والصدق إلا أن هذه الاختبارات تظل متأثرة بالعينات التي طبقت عليها وخصائصها وهذا يجعلها غير مفيدة إذا أعيد استخدامها في بحوث ودراسات على عينات أخرى دون إعادة تقنين. كما أنه جرت العادة في النظرية الكلاسيكية على تركيز التحليل الإحصائي على الدرجة الكلية للاختبار، وكذلك الاعتماد على تكرارات الإجابات الصحيحة لتقدير معاملات صعوبة المفردات ، كذلك يكون التركيز على معامل ثبات الاختبار وحساب معاملات ارتباط كل مفردة بالدرجة الكلية للاختبار ومعاملات تمييز المفردات ، وهذه الأساليب تستخدم على مدى واسع إلا أنها محدودة لارتباطها بالعينة وخصائصها. أما الاختبارات مرجعية المحك CRTs فتستخدم لتقييم وتشخيص الأداء الراهن. ويتطلب الاختبار التحصيلي المرجع إلى هدف صياغة الهدف أو الأهداف المطلوب قياس تحققها لدى الأفراد صياغة سلوكية ، ويشتمل الاختبار على مجموعة من المفردات التي تقيس الهدف أو الأهداف المحددة. (صلاح علام ،٢٠٠٧:ص٣١) وهذه الاختبارات تدخل في إطار النماذج الحديثة في القياس التي ظهرت بدورها نتيجة للانتقادات الموجهة للنظرية الكلاسيكية ، والتى سيطرت وما زالت تسيطر على ميدان القياس النفسي والتربوى بصفة عامة وعلى بناء الاختبارات وتفسير درجاتها بصفة خاصة ، و من أهم هذه الانتقادات:-

- اعتماد درجات المستجيبين على الاختبار على مفردات ذلك الاختبار، مما يعنى أنه يتم تعريف القدرة فى ضوء الاختبار ذاته، ودرجة الفرد المستجيب على الاختبار تعتمد على الاختبار ذاته.
- اعتماد إحصاءات المفردات على عينة الأفراد المستجيبين على الاختبار،مما يجعل تلك الإحصاءات محدودة الاستخدام ، فبارامترات المفردة المشتقة باستخدام الطرق التقليدية ليست لها صفه الثبات وتعتمد على عينة الأفراد التى اشتقت منها.
- لا تقدم النظرية الكلاسيكية أية تفسيرات تتعلق بكيفية أداء المستجيبين على مفردات الاختباربالرغم من أن هذه التفسيرات تعد ضرورية إذا أردنا التنبؤ

بخصائص الأفراد الحاصلين على درجات معينة ، أو عند إعداد اختبارات تتميز بخصائص تناسب مجتمع من الأفراد، فالنظرية الكلاسيكية تركز على الاختبار ككل أكثر من تركيزها على مفرداته ، فالوحدة الأساسية في تلك النظرية هى الاختبار Test-Oriented.

- تفترض النظرية الكلاسيكية تساوى أخطاء القياس لدى جميع المستجيبين على الاختبار، فى حين أن الأفراد ذوي القدرات المختلفة لا يتم تقدير قدراتهم بنفس مستوى الدقة على الاختبار الواحد.

(Hambleton&swaminthan,:1985; Hambleton&swaminthan, 1991) كما أن النظرية الحديثة للقياس تمدنا بمفهوم جديد يدعى (دالة المعلمة / المعلمات) التي نحصل عليها من تطبيق بند أواختبار. وهذه المعلمة تعني مقدار الثقة في أننا حصلنا على معلومات تقودنا إلى تقدير قدرة فرد معين أو مجموعة من الأفراد في مستوى واحد من القدرة. أي أننا نحصل على قيم متعددة لكمية المعلمات التي حصلنا عليها من البند أو الاختبار. وهي بذلك المعنى تختلف عن الثبات الذي هوعبارة عن مؤشر واحد فقط للاختبار لكل الأفراد الذين طبق عليهم هذا الاختبار.

وتمثل دالة (المعلمة / المعلمات) عادة بمنحنى يبين أعلى وأقل كمية معلومات حصلنا عليها عند مستويات القدرة المختلفة. كل هذا دعا إلى استخدام نظرية السمات الكامنةLatent Traits Theory ، التي يشير إليها الباحثون في القياس النفسي على أنها نظرية الدرجة الحقيقية ، وهي نظرية حديثة بالمقارنة بالنظرية الكلاسيكية ، حيث تعتمد على مجموعة افتراضات قوية ، مفادها أن هناك استقلالية للمفردات الاختبارية في قياس القدرات المختلفة ، وأن كل مفردة اختبارية لها خصائصها السيكومترية الممثلة في منحنى القدرة الخاص بها الذي يصف احتمالية الحصول على الإجابة الصحيحة أو الخاطئة بمعرفة قدرة الفرد.(Magno, 2009:p2)

مشكلة البحث

مما سبق يتضح للباحثين ظهور اتجاهين في القياس النفسي وهما اتجاه النظرية الكلاسيكية التقليدية والتى تركز بدورها على الدرجة الكلية للاختبار دون التركيز على خصائص مفرداته ، بمعنى أن القياس التقليدى يتعامل مع المفردة دون النظر إلى تقديم

أية معلومات متعلقه بأداء المستجيبين وقدرتهم على مفردات الاختبار وبالتالى فإن هذه المعلومات الخاصة بالأداء تعد مهمة في تفسير نتائج القياس الخاص بالمستجيبين وهذا ما حرصت علية النظريات الحديثة في القياس ،ومنها نظرية السمات الكامنة والتى تهتم بإيجاد دالة المعلمات الخاصة بكل مفردة ، كما تهتم بتحديد العلاقة بين استجابات الأفراد على اختبار معين والسمة الكامنة وراء هذه الاستجابات وهذا بطبيعة الحال يؤدي إلى الثقة في قدرة المفردة للحكم على المستجيب نتيجة تقديمها لكمية معلومات كافية وشاملة ودقيقة تتعلق بالفرد المستجيب والمفردة.

ومن ثم تتمثل مشكلة البحث في الأسئلة التالية:

١- ما الخصائص السيكومترية التى تتوفر للاختبار التحصيلي المصمم باستخدام النظرية الكلاسيكية في مقرر الجبر؟

٢- ما الخصائص السيكومترية التى تتوفر للاختبار التحصيلي مرجعي الهدف المصمم باستخدام نظرية الاستجابة للمفردة في نفس المحتوى في مقرر الجبر؟

٣- ما درجة ملاءمة نموذج راش لبيانات الاختبار التحصيلي مرجعي الهدف؟

٤- هل تختلف البيانات المستمدة من النظرية الكلاسيكية عن البيانات المستمدة من نظرية الاستجابة للمفردة ؟

٤- إلى أي مدى تتشابه الخصائص السيكومترية الخاصة بمفردات الاختبار باستخدام النظرية الكلاسيكية، مع الخصائص السيكومترية التى تتوفر للاختبار التحصيلي مرجعي الهدف المصمم باستخدام نظرية الاستجابة للمفردة؟

أهداف البحث

يهدف البحث الحالى إلى:

١. التعرف على الخصائص السيكومترية للاختبار التحصيلي المصمم على نفس المحتوى في الجبر على عينة من تلاميذ الصف الأول الإعدادي باستخدام النظرية الكلاسيكية.

٢. تحديد الخصائص السيكومترية للاختبار التحصيلي مرجعي الهدف في الجبر على عينة من تلاميذ الصف الأول الإعدادي باستخدام نموذج راش.

٣. التحقق من إمكانية التنبؤ من درجات التلاميذ على الاختبار التحصيلي مرجعي الهدف في الجبر المشتقة من نموذج راش بالتحصيل الدراسى لديهم في مادة الجبر.

٤. تفسير الاختلافات بين النظريتين: الكلاسيكي ، والسمات الكامنة في تقدير بارامترات صعوبة المفردات بالاختبار التحصيلي.

٥. مقارنة فاعلية النظريتين في قياس التحصيل الدراسي بالوحدة المختارة في مقرر الجبر بالصف الأول الإعدادي.

أهمية البحث

– الأهمية النظرية

يستمد هذا البحث أهميته من أهمية متغيراته حيث تشير نتائج دراسات Macdonald&Paunonen(2007)، و Silvestre-Tipay(2009)، ودراسة Magno(2009) أن البيانات الإحصائية المستمدة من كلا النظريتين: النظرية الكلاسيكية ، ونظرية الاستجابة للمفردة قابلة للمقارنة إذا استخدمت النماذج الإحصائية المناسبة لهذه المقارنة ، والدراسات المقارنة في بحوث التقويم والقياس التربوي مهمة في التعرف على كفاءة وفاعلية الأساليب المستخدمة في تحليل بيانات الاختبارات بحيث يتسق التحليل مع التصميم والأهداف التي من أجلها صممت وطبقت هذه الاختبارات ، ولا يمكن اتخاذ قرار مناسب بشأن ذلك بدون دراسات كافية عن مميزات وعيوب كل إطار قياسي بحيث يمكن اختيار الإطار المناسب في كل اختبار يتم تصميمه.

كما أن الخصائص السيكومترية للمفردات المكونة للاختبار مهمة في دراسة فعالية نموذج القياس المستخدم ، فخصائص المفردات في ضوء النظرية الكلاسيكية تعتمد على درجة الاتساق فيما بينها ، وارتباط درجة كل مفردة بالدرجة الكلية هو مؤشر على الاتساق الداخلي للاختبار ، أما في نظرية الاستجابة للمفردة فإن المفردات تتمتع بدرجة مناسبة من الاستقلالية التي تجعلها وحدة قياس مستقلة تقيس قدرة معينة ويمكن التنبؤ بمسلكها من خلال منحنى القدرة الخاص بها. غير أنه يمكن القول

أن الدراسات المقارنة لا تهدف إلى إثبات أن نظرية الاستجابة للمفردة دائماً أفضل من النظرية الكلاسيكية بقدر ما تهدف إلى توضيح الاختلافات النظرية والتطبيقية بين النظريتين كأطر قياسية مستخدمة بكثرة في شتى جوانب القياس النفسي والتربوي.

- الأهمية التطبيقية

أما عن الأهمية التطبيقية فالدراسة تسعى إلى تصميم اختبار لتشخيص مدى تحقيق أهداف دراسة الوحدة الثانية من مقرر الجبر للصف الأول الإعدادي ، وهي الوحدة المخصصة للحدود والمقادير الجبرية فقد توجد هناك بعض الصعوبات في بداية المرحلة الإعدادية خاصة في فترات الانتقال من المحسوس (الأعداد) إلى المجرد(الرموز) التي تعبر عن حدود ومقادير جبرية وهذا ما دعا الباحثان إلى اختيار الوحدة الثانية في مقرر الجبر في الفصل الدراسي الأول من الصف الأول الإعدادي.

الإطار النظري للبحث

النظرية الكلاسيكية Classical Test Theory

تعرف نظرية الاختبار الكلاسيكي بأنها "نظرية الدرجة الحقيقية" ، وتقوم هذه النظرية على فرضية أن هناك تأثيرات خارجية على إجابات المفحوصين بخلاف قدراتهم: مثل الظروف المحيطة بتطبيق الاختبار ، والعوامل الداخلية الخاصة بالمفحوصين.

ويشير Magno(2009) إلى أن النموذج المركزي للنظرية الكلاسيكية في تصميم الاختبارات يعتمد على فكرة تقدير الدرجة الملاحظة True Observed Score(TO) وتتكون من الدرجة الحقيقية (T) ودرجة الخطأ (E) ، وهناك استقلالية بين الدرجة الحقيقية ودرجة الخطأ ويمكن توضيحها من خلال الصيغة TO=T+E (p.1)

وتفترض النظرية الكلاسيكية أن لكل فرد درجة حقيقية ، وأنه يمكن الحصول عليها إذا لم يكن هناك أية أخطاء في عملية القياس ، وحيث أن أدوات القياس بطبيعتها غير دقيقة فإن الدرجة الملاحظة قد تختلف عن الدرجة الحقيقية بسبب أخطاء القياس.

لذلك تفترض النظرية الكلاسيكية أنه من النادر أن تكون الدرجة التي يحصل عليها الفرد في اختبار ما تمثل تقديراً دقيقاً لدرجته الحقيقية ، أي أن الدرجة الملاحظة أو

الدرجة المقاسة هي تقريباً الدرجة الحقيقية متأثرة بقدر من الخطأ. وهذا الخطأ قد يجعل الدرجة الملاحظة أكبر أو أقل من الدرجة الحقيقية.

ويشير Magno(2009) إلى أن الخطأ يمثل على أنه متغيراً عشوائياً له توزيع طبيعي، ونظرياً فإن الانحراف المعياري لتوزيع الأخطاء العشوائية لكل فرد يدلنا على حجم أخطاء القياس. ومن المفترض أن يكون توزيع الأخطاء العشوائية متشابهاً لجميع الأفراد. ويمكن حساب الخطأ المعياري للقياس بدلالة الانحراف المعياري للدرجة المقاسة، وثبات الاختبار. (p.2). ومع ظهور نظرية السمات الكامنة أصبحت هناك مرجعية أدق في التعبير عن القدرات التحصيلية الحقيقية للمتعلمين فالافتراضات التي تقوم عليها هذه النظرية من استقلالية بنود المقياس عن السمة أو القدرة المقاسة، ولذلك فإن نظرية السمات الكامنة تشترك في هذا الهدف الأساسي مع النظرية الكلاسيكية بل أن نظرية السمات الكامنة تهدف إلى أكثر من ذلك فهي تهدف إلى تحديد العلاقة بين استجابات الأفراد على اختبار معين والسمة الكامنة وراء هذه الاستجابات، والأمر الأكثر أهمية في القياس النفسي والتربوي بعامة هو تحديد مقدار السمات الكامنة وراء أداء الأفراد على الاختبارات المختلفة والاستفادة منها في تفسير النتائج والتنبؤ بسلوكهم في مواقف مماثلة، واتخاذ قرارات معينة بشأنهم في ضوء هذا التقدير الكمي للسمات.

ويشير صلاح علام (٢٠٠٧) إلى أن النماذج الكلاسيكية ونماذج الاستجابة للمفردة في افتراض وجود متصل للسمة، وأنه يمكن تقدير احتمال إجابة فرد إجابة صحيحة على مفردة اختبار إذا علمنا موقعه على هذا المتصل. وبتعبير رياضي يمكن القول بأن الاختبار يعد بمثابة راسم فوقي (دالة) من عينة الأفراد المختبرين إلى متصل السمة الكامنة. كما تتفق هذه النماذج في أن مقدار هذا الاحتمال يكون دالة متزايدة مطردة لمواقع الأفراد على متصل السمة. وهذا يعني أن احتمال توصل الفرد إلى الإجابة الصحيحة على مفردة اختبار يزداد بازدياد مقدار السمة لديه.

غير أن نماذج الاستجابة للمفردة تختلف عن النماذج الكلاسيكية في تصورها لشكل هذه الدالة. فبينما تفترض النماذج الكلاسيكية أن هذه الدالة تتخذ شكل المنحنى الاعتدالي في الاختبارات مرجعية المعيار، وتتخذ شكلاً ملتوياً التواءاً سالباً في

الاختبارات مرجعية المحك ، تفترض نماذج السمات الكامنة أنها تتخذ شكل المنحنى اللوغاريتمي. وتسمى هذه الدالة في الحالتين بالدالة المميزة للمفردة ، (صلاح علام ، ٢٠٠٧: ص ٢١٤).

نظرية السمات الكامنة Latent Traits Theory

وتمثل نظرية السمات الكامنة الاتجاه الجديد فى القياس النفسي / السلوكى. وتفترض هذه النظرية وجود واحد أو أكثر من المميزات أو السمات الأساسية ، التى تحدد استجابات الفرد الملاحظة لبنود اختبار ما. وقد اصطلح على تسميتها بالسمات الكامنة أو القدرات فى حالة الاختبارات المعرفية نظراً لعدم إمكانية ملاحظتها أو قياسها بصورة مباشرة.

نماذج السمات الكامنة:-

تهدف نماذج السمات الكامنة إلى تحديد علاقة بين أداء الفرد في اختبار معين وبين السمات أو القدرات التي تكمن وراء هذا الأداء وتفسره، لذلك فإن هذه النماذج هي دوال رياضية احتمالية. وتنقسم هذه النماذج إلى مجموعتين إحداهما النماذج الاستاتيكية ، والأخرى النماذج الديناميكية ، وتهتم النماذج الاستاتيكية بالقياس في مدة زمنية واحدة ، كما تهتم بتحديد العمليات التي ينطوي عليها الأداء في الاختبارات السيكولوجية والتربوية. أما النماذج الديناميكية فتهتم بمشكلة قياس التغير في السمات النامية المختلفة في فترات زمنية متباعدة غير أن النماذج الاستاتيكية هي التي نالت كثيراً من الاهتمام في مجال الاختبارات محكية المرجع.

ويحدد نموذج السمات الكامنة العلاقة المتوقعة بين الاستجابات الملاحظة على الاختبار والسمات أو القدرات غير الملاحظة التى يفترض أنها تحدد هذه الاستجابات. والسمة تمثل بعد كمى يمكن أن يحدد عليه مواضع الأفراد ، ولا يصح نظرياً أن يتوقف موضع الفرد على السمة المقاسة على صفات أى من العينات التى ينتمى إليها الفرد فعلى سبيل المثال ، ينبغى أن يستقل وضع الفرد على متصل سمة ما – مثلما يستقل وزنه أو طوله مثلاً – عن اعتبارات العمر و الجنس والشريحة الاجتماعية. بهذا المعنى توفر نماذج السمات الكامنة تقديراً للقدرة مستقلاً عن العينة كما توفر أيضاً مميزات القياس ذى الفئات المتساوية ،ومعنى أن يكون القياس متحرراً من العينة Sample – free ، أن

يعبر عن تقديرات القدرة بوحدات لا تتعلق بصفات أى عينة ، أو مجموعة معينة من الأفراد ، فكما أن تقدير وزن فرد ما لا يتعلق بعمره أو جنسه ، فإن تقدير قدرة الفرد باستخدام أحد نماذج السمات الكامنة لا يتعلق بهذه العوامل أو غيرها من المميزات ويعد نموذج راش من أهم نماذج السمات الكامنة حيث يمكن أن تتوفر متطلبات الموضوعية عندما تستوفى فروض النموذج.

نموذج راش أحادي البعد:-

ونموذج راش هو نموذج استاتيكي أحادي البعد: وتشير أمينة كاظم (١٩٨٨) إلى أن أحادية البعد تعني أن السمة موضع القياس تعرف بواسطة مجموعة من البنود (المفردات) ذات صعوبة أحادية البعد ، أي أن بنود الاختبار لا تختلف فيما بينها إلا من حيث مستوى الصعوبة فقط. كما يكون الأفراد ذوي قدرة أحادية البعد تحدد وحدها مستوى أدائهم على الاختبار.

أما استقلالية القياس فتعني أنه لا يعتمد تقدير صعوبة البند ، على تقديرات صعوبة البنود الأخرى المكونة للاختبار ، ولا على تقديرات قدرة الأفراد الذين يجيبون علي الاختبار.

ولايعتمد تقدير قدرة الفرد ، على تقديرات قدرة أي مجموعة من الأفراد الذين يؤدون الاختبار ، أو على تقديرات صعوبة البنود التي يؤدونها ؛ وتتحقق استقلالية القياس هذه عندما تتحق ثلاث أمور:- الأول ملاءمة بنود الاختبار للنموذج ، وهذا يعتمد على صدق البنود في قياس القدرة موضع القياس ، وعلى صدق تدرجها على مقياس هذه القدرة. الثاني: ملاءمة الأفراد الذين يجيبون على بنود الاختبار للنموذج ، وهذا يعتمد على صدق استجابات هؤلاء الأفراد. الثالث: توافق بنود الاختبار مع قدرات الأفراد ، ويعتمد هذا على تقارب مستوياتها. (أمينة كاظم ، ١٩٨٨، ص ٢٢-٢٣)

وعندما يستخدم نموذج راش في بناء الاختبارات التحصيلية وفى تكوين بنوك الأسئلة ، يمكن عندئذ التغلب على مشكلتين من أهم مشكلات التقويم الأكاديمى وهما:-

-**الأولى:** موضوعية تقدير المستوى الأكاديمى للطالب ، بالرغم من اختلاف الإختبارات المستخدمة وتباين مستواتها.

والثانية: موضوعية تقدير المستوى الأكاديمى للطالب ، بالرغم من اختلاف مستوى المجموعة التى ينتمى إليها.(أنور الشرقاوى وآخرون: ١٩٩٦، ص ٤٣١).

وتؤكد أمينة كاظم أن إستخدام نماذج السمات الكامنة تعمل على تحقيق مايلي:-

- بنود صادقة يمكنها تعريف المتغير موضع القياس تعريفاً إجرائياً.
- صدق التدرج لهذه البنود بحيث يمكنها تمثيل هذا المتغير بواسطة خط مستقيم.
- أنماط استجابات صادقة يمكنها تحديد مواضع الأفراد على متصل المتغير.
- توافق بين درجات الأفراد وخصائص البنود بحيث تؤدى إلى تقديرات لمستويات الأفراد لا تعتمد على إختبار معين ، ويمكن استخدامها لوصف ما يتميز به الأفراد بصورة عامة.
- قياسات خطية يمكن استخدامها لدراسة النمو أو للمقارنة بين الأفراد.

(أمينة كاظم: ١٩٨٨ ب: ص ص ٣٩٠ – ٤١٩)

ويرتكز البحث الحالي على نموذج راش أحادي البارامتر ، ونموذج راش مبني على فرضية أن كلاً من:- التخمين على المفردة ، والفروق في تقديرات معاملات التمييز للمفردات يمكن معرفتهما وأن تحليل بيانات الاختبار كي يكون مفيداً لابد أن يرتكز حول المتعلم (الفرد) مع بارامترات منفصلة للمفردات والأفراد.(Magno,2009:p3)

ويشير Magno(2009) إلى أن نموذج راش من النماذج الاحتمالية لنظرية الاستجابة للمفردة يؤكد على: (١) كلما كان السؤال سهل كلما زادت احتمال الإجابة الصحيحة عليه ، (٢) كلما كانت القدرة أعلى لدى الأفراد كلما أجابوا على السؤال إجابة صحيحة بالمقارنة مع ذوي القدرة الأقل.

ويؤكد أن تصميم الاختبار في ضوء هذه المنهجية العلمية يتطلب استبعاد أية مفردة لا تحقق هذين الافتراضين.(p3)

ولقد اشتق نموذج راش من الصيغة التالية:

$$\varepsilon = \frac{\delta}{\theta}$$

حيث ε هي دالة البارامترات التي تصف قدرة المفحوص، وصعوبة الاختبار، θ تمثل قدرة المفحوص، وδ تمثل صعوبة الاختبار والتي يمكن تقديرها بجمع الإجابات الخاطئة في الاختبار.

ولقد تم تطوير هذا النموذج على فرضية مفادها: أن احتمالية الإجابة الصحيحة من المتعلم هي دالة لوجستية للفرق بين قدرة المتعلم (θ)، وصعوبة المفردة (β) ، أي أنها تمثل القدرة المطلوبة لإجابة سؤال إجابة صحيحة، والدالة اللوجستية لهذا الفرق تعطي الطريقة التي اعتمد عليها نموذج راش. وهذا يعني أن كل استجابة لكل مفردة تعتمد على قدرة المتعلم، وصعوبة المفردة ؛ ولذلك حينما تصلح البيانات الإحصائية المعطاة لتطبيق أحد نماذج الإستجابة للمفردات الإختبارية عليها فإن المفردات تتمتع باستقلالية عن القدرات الخاصة بالمتعلمين، والنتيجة الأكثر اعتماداً على ذلك هي أنه يمكن أن يتحقق إستخدام نموذج الدرجة الكلية.

ونموذج راش هو نموذج ملائم لنمذجة الإستجابات الأحادية وذلك من خلال منحنى خصائص المفردة اللوجستي، ودالة قدرة الفرد حيث يتضح من خلال المساحات أسفل المنحنى احتمالية لكل من:- الإجابة الصحيحة والإجابة الخاطئة، ولذلك فنموذج راش هو نموذج لوجستي أحادي البارامتر يفترض أن معاملات تمييز كل المفردات مساوية للواحد الصحيح.(p4)

الخصائص السيكومترية لمفردات الاختبارات في إطار النماذج الكلاسيكية ونماذج الاستجابة للمفردة:-

يشير (صلاح علام ،٢٠٠٧:ص ٢٢١-٢٢٣) إلى أن الخصائص السيكومترية لمفردات الاختبار، مثل الصعوبة والتمييز هي التي تحدد العلاقة بين السمة الكامنة التي يهدف الاختبار إلى قياسها، ودرجات الأفراد في الاختبار المستخدم ؛ فإذا أجريت موازنة بين النماذج الكلاسيكية، ونماذج الاستجابة للمفردة في هذا الشأن، يتضح أنه على الرغم من اتفاقها حول هذه الخصائص، إلا أن هناك اختلافاً بينهما في كيفية معالجتهما للمفاهيم المختلفة المتعلقة بهذه الخصائص.

تقدير معاملات صعوبة المفردات:-

مفهوم "صعوبة المفردة" في إطار النماذج الكلاسيكية يدل على نسبة عدد أفراد المجموعة الذين أجابوا إجابة صحيحة على هذه المفردة. أو بعبارة أخرى يدل على متوسط احتمالات الإجابة الصحيحة على المفردة لجميع أفراد المجموعة.

ومعرفتنا للقيمة التقديرية لصعوبة مفردة اختبار بالنسبة لمجموعة معينة من الأفراد لا يفيدنا كثيراً في معرفة درجة صعوبة المفردة بالنسبة لواحد من أفراد المجموعة.

ويشير صلاح علام (١٩٨٧) إلى اعتماد قيم معاملات الصعوبة المشتقة من الطرق التقليدية على عينة الأفراد المستجيبين، الأمر الذى يجعل استخدام معاملات الصعوبة المشتقة بهذه الطريقة محدود، حيث تختلف قيم معاملات الصعوبة باختلاف متوسط ومدى قدرة أفراد المجموعة المستخدمة في اشتقاق تلك المعاملات.(صلاح علام، ١٩٨٧، ص ص ١٨ – ٤٣)

كما يشير هشام جاد الرب (٢٠٠٧) إلى اعتماد الطريقة التقليدية فى تحديد صعوبة المفردات على تطبيق المفردات استطلاعياً على عينة من الأفراد، التى يفترض أنها تمثل المجتمع الذى بنى الاختبار لقياس أحد خصائصه ثم حساب معاملات صعوبة المفردات وفقاً لعدد الأفراد الذين يستجيبون على المفردة استجابة صحيحة دون الربط بين تباين مستويات صعوبة المفردات وتنوع العمليات النفسية والمعرفية التى تقع خلف الاستجابة على تلك المفردات. (هشام جاد الرب: ٢٠٠٧، ٢٧٧)

أما مفهوم "تمييز المفردة" فيشير إلى مدى فاعلية مفردة الاختبار في التمييز بين المستويات المختلفة للسمة المراد قياسها لدى المختبرين؛ وتعتمد قيمة معامل التمييز على درجة صعوبة المفردة وعلى خصائص مجموعة الأفراد المختبرين.

ويتخذ هذان المفهومان معنى مختلفاً في إطار نماذج الاستجابة للمفردة: ويرجع ذلك إلى افتراض هذه النماذج وجود دالة مميزة لكل مفردة من مفردات الاختبار تتخذ المنحنى اللوغاريتمي التراكمي. وتتميز هذه الدالة اللوغاريتمية التراكمية بخصائص يمكن الإفادة منها في جعل ميزان درجات الاختبار خطياً، وفي تقدير الخصائص السيكومترية للاختبارات تقديراً مستقلاً عن خصائص عينة الأفراد المختبرين. (صلاح علام، ٢٠٠٧: ص ٢٢١–٢٢٣).

ويشير Progar and Socan(2008) إلى أن نظرية الاستجابة للمفردة IRT تمكننا من التقدير المستقل لبارامترات (المفردة والفرد)، والتقدير الموضعي لأخطاء القياس؛ هذه الخصائص تمثل ميزة نسبية لنظرية الاستجابة للمفردة في المقارنة مع النظرية الكلاسيكية CTT.

لكن الدليل الامبريقي غالباً ما يفشل في كشف الفروق المتسقة بين بارامترات (المفردة والفرد) في كلا النظريتين، وبين القياسات الثابتة لتقديرات بارامترات النظريتين.

وفي الاختبارات التحصيلية مرجعية الهدف فإن المفردة مرتبطة بالهدف الذي تقيسه بحيث أن الاستجابة الصحيحة للمفردة تعني أن المفحوص قد حقق الهدف؛ لذلك فإن معامل صعوبة المفردة مؤشر مهم يدل على صعوبة الهدف.

وقد أكد صلاح علام (٢٠٠٧) على أهمية تحديد الهدف من عملية تحليل المفردات في كل إطار قياسي مستخدم؛ ذلك لأن هناك تفاوت في انتقاء المفردات الاختبارية بين النظرية الكلاسيكية، ونظرية الاستجابة للمفردة فالهدف من تحليل المفردات في النظرية الكلاسيكية هو الكشف عن مفردات الاختبار غير الصالحة، وتقدير البارامترات المتعلقة بالمفردة، واستخدام القيم التقديرية لهذه البارامترات في تقدير الخصائص السيكومترية للاختبار ككل. أما الهدف في حالة نماذج الاستجابة للمفردات الاختبارية فيكون موجهاً نحو الحصول على تدريج مفردات الاختبار Item Calibration، أي تقدير صعوبة مفرداته، وتقييم حسن مطابقة درجات مفردات الاختبار للنموذج المعين باستخدام قيم مربع كاي مثلاً، والإفادة من ذلك في تقدير سمات الأفراد. (صلاح علام، ٢٠٠٧: ص ٢٢٥).

ويشير Magno(2009) إلى أن أهم مايميز مدخل نظرية الاستجابة للمفردة هو الدقة في تحليل المفردات الاختبارية؛ وأن لكل مفردة خصائصها السيكومترية الممثلة في منحنى القدرة الخاص بها الذي يصف إحتمالية الحصول على الإجابة الصحيحة بمعرفة قدرة الفرد.

ونموذج راش بوصفه أحد نماذج نظرية الاستجابة للمفردة فإنه يقوم على فرضية أن معاملات تمييز كل المفردات مساوية للواحد الصحيح. (p.2)

ومن أهم ما يميز نظرية الاستجابة للمفردة أيضاً معالجة موضوع ثبات الاختبار، وتقدير الخطأ خلال دراسة سلوك المفردة من خلال معلومات محسوبة خصيصاً لكل مفردة على حدة وليس الاختبار ككل. وتأخذ دالة معلمات المفردة في اعتبارها كل البارامترات المتعلقة بالمفردات، وكذلك تظهر كفاءة المفردة في قياس مستويات مختلفة من القدرة.

وتوفر نظرية الاستجابة للمفردة ميزة أخرى تتمثل في عدم تباين بارامترات المفردة وهو ما يؤكد استقلالية النتائج عن خصائص العينة. (p5)

وقد تختلف الخصائص السيكومترية للمفردة نتيجة لاختلاف الإطار المرجعي النظري الذي يتم تقديرها في ضوءه ؛ سواء كان النظرية الكلاسيكية في الاختبار أو نظرية الاستجابة للمفردة لذلك كان من الضروري بحث هذه الخصائص السيكومترية لهذه المفردات في أي دراسات مقارنة بين هاتين النظريتين.

ويشير صلاح علام (٢٠٠٧) إلى أن الاختبارات مرجعية الهدف هي تلك الاختبارات التي تبنى على أساس مجموعة من الأهداف التعليمية المصاغة صياغة سلوكية، وتكون هناك مزاوجة بين مفردات الاختبار وهذه الأهداف، غير أن النطاق السلوكي الذي تمثله هذه الأهداف لا يكون محدداً، وبالتالي فإن المفردات التي يشتمل عليها الاختبار تكون قليلة العدد نسبياً لأنها لا تمثل النطاق الشامل للمفردات الممكنة التي تقيس مجموعة الأهداف.

وتطبق هذه الاختبارات عادة عند الانتهاء من دراسة وحدة تعليمية أو وحدة نسقية معينة بهدف تصنيف المختبرين في مجموعتين إحداهما حققت الأهداف والأخرى لم تحققها في ضوء نسبة مئوية محددة من المفردات التي ينبغي الإجابة عليها إجابات صحيحة والتعرف على الأهداف التي لم يستطع كل منهم تحقيقها. كما يمكن تصنيفهم في عدة مجموعات بدلاً عن التصنيف الثنائي وذلك بتحديد مستويات أداء في إطار هذه الأهداف.(صلاح الدين علام، ٢٠٠٧: ص ٢٨)

ويرى الباحثان أن مدخل الاختبار التحصيلي مرجعي الهدف يعد أقرب المداخل إلى أهداف البحث الحالي حيث يهدف الباحثان إلى قياس التحصيل الدراسي في الوحدة الثانية (الحدود والمقادير الجبرية) بالصف الأول الإعدادي حيث يدرس التلميذ الرموز

المجردة بمعاملات عددية كحدود جبرية ويجري عليها العمليات الحسابية المتنوعة ؛ وتعد فترات الانتقال من المحسوس (الأعداد) إلى المجرد (الرموز) فترات صعبة في التعلم تظهر فيها بعض مشكلات التحصيل الدراسي التي يجب بحثها بدقة.

وبعد تحديد المحتوى يجب اختيار الكفايات أو المهارات الرئيسة في هذا المحتوى ، ويشير (صلاح علام ، ٢٠٠٧: ص ٤٤) إلى أنه يمكن التوصل إلى ذلك عن طريق التحليل المتعمق لمحتوى المنهج الدراسي أو المجال التدريبي وما يتضمنه هذا المحتوى من نواتج سلوكية.

مصطلحات البحث:-

١- الكفاية Competency: هي مجموعة متكاملة من المعارف والمهارات الوظيفية المحددة تحديداً دقيقاً والمتعلقة بمجال تعليمي أو تدريبي معين بحيث يمكن تحقيقها وقياسها من خلال البرنامج. (صلاح علام ، ٢٠٠٧: ص ٤٤)

٢- النظرية الكلاسيكية CTT: تعرف بنظرية الدرجة الحقيقية (T) وتقوم على فرضية أن الدرجة الملاحظة (TO) تختلف عن الدرجة الحقيقية بسبب أخطاء القياس وتصف العلاقة بين الدرجتين من خلال الصيغة: TO=T+E حيث E درجة الخطأ (Magno,2009:p1)

٣- نظرية الاستجابة للمفردة IRT: تفترض هذه النظرية أن هناك دالة احتمالية تربط بين بارامترين أو معلمين أحدهما يتعلق بالفرد ، والآخر يتعلق بالمفردة التي يجيب عليها. ونظراً لأن استجابات الفرد لمفردات الاختبار تكون مشوبة بالخطأ ، فإن هذه النظرية تهدف إلى التوصل إلى قيم تقديرية لكل من هذين المعلمين ، ومن ثم استخدام هذه القيم في تقدير احتمال الاستجابة الصحيحة لكل مفردة من مفردات الاختبار. (صلاح علام ، ٢٠٠٧: ص ٢١١)

٤- نموذج راش: هو نموذج لوغاريتمي أحادي البعد يمثل طريقة للحصول على القياسات الإحصائية الأساسية والموضوعية المزودة بمعالجة الأخطاء المعيارية وإحصاءات مواءمة الجودة وقد استنتجها عالم الرياضيات الدانماركي جورج راش George Rash من ملاحظات عشوائية لفئة من

الاستجابات الملاحظة ، وذلك في العام ١٩٥٣ أثناء تحليله للاستجابات في سلاسل من اختبارات القراءة. (Linacre,2006:p12)

ويشير صلاح علام (٢٠٠٧) إلى أن تطبيق هذا النموذج يتطلب تساوي مقادير تمييز جميع مفردات الاختبار أي توازي منحنياتها ، ولكن تختلف عن بعضها فقط في نقطة التقائها بالمحور الأفقي الذي يمثل متصل السمة الكامنة ، أي تختلف عن بعضها في الصعوبة فقط. (صلاح علام ،٢٠٠٧: ص ٢١٧)

٥- قدرة الفرد مقدرة باللوجيت: - هي اللوغاريتم الطبيعي لمميز نجاح الفرد على المفردات التي تعبر نقطة صفر التدريج عن صعوبتها. (أمينة كاظم ،١٩٨٨:ص٢٧)

٦- صعوبة المفردة: مقدرة باللوجيت هي اللوغاريتم الطبيعي لمميز الفشل لدى الأفراد الذين تعبر نقطة صفر التدريج عن قدرتهم. (أمينة كاظم ،١٩٨٨:ص٢٧- ٢٨)

٧- وحدة القياس (اللوجيت) Logit: هي وحدة قياس كل من قدرة الفرد وصعوبة المفردة ، وتعرف باللوغاريتم الطبيعي لمميز نجاح الفرد على المفردات التي تعبر نقطة صفر التدريج عن صعوبتها ، عندما يساوي هذا المميز ثابتاً هو الأساس الطبيعي (e) أي (2.72). (أمينة كاظم ،١٩٨٨:ص٢٨)

البحوث والدراسات السابقة

- **دراسة فان (١٩٩٨)Fan**

بالرغم من الاختلافات بين نظرية الاستجابة للمفردة ، والنظرية التقليدية إلا أنه هناك نقص في المعرفة الإمبريقية عن كيفية ومدى الاختلاف بين النظريتين فيما يخص (إحصاءات الفرد / المفردة).وتبحث هذه الدراسة إمبريقياً المسلك الذي تسلكه إحصاءات (الفرد / المفردة) المشتقة من هذين الإطارين القياسيين المختلفين ، وقد ركزت هذه الدراسة على نقطتين أساسيتين: - الأولى بحث العلاقات الإمبريقية بين الإحصاءات المتمركزة على النظرية التقليدية ، والإحصاءات المتمركزة على نظرية الاستجابة للمفردة، والنقطة الثانية: التعرف على مدى تتشابه إحصاءات المفردة

المستمدة من النظرية الكلاسيكية مع إحصاءات المفردة المستمدة من نظرية الاستجابة للمفردة وذلك باختلاف المفحوصين في العينات.

وتشير النتائج إلى أنه إحصاءات (المفردة / الفرد) المستمدة من كلا النظريتين كانت قابلة للملاحظة، ودرجات ثبات بيانات وإحصاءات المفردة عبر العينات بدا أنها متقاربة في كلا النظريتين (الإطارين القياسيين)، رغم التفوق النظري لنماذج نظرية الاستجابة للمفردة.

- دراسة **Fraley et al.(2000)**

وبحثت هذه الدراسة أساليب تقدير الدرجة بطرق متنوعة لبعض مقاييس التقرير الذاتي (مثل المتوسطات أو الدرجة الكلية)، وقد بحثت الدراسة تحليل تقرير ذاتي لقياس الالتصاق بالكبار باستخدام نظرية الاستجابة للمفردة، وكذلك تحديد ما إذا كانت هناك مشكلات تتعلق بتدريج البنود أو تقدير الدرجة وأجرى المؤلفون تحليلاً في ضوء نظرية الاستجابة للمفردة لأربع استبيانات بنظام التقرير الذاتي، وأكدت النتائج المستمدة من التطبيق على ١٠٨٥ فرد أن المقاييس الأربعة يمكن تطويرها بطرق متنوعة، ويمكن استخدام أساليب نظرية الاستجابة للمفردة لتطوير هذه المقاييس مع مراعاة خصائصها السيكومترية.

- دراسة هوانج **Hwang,D(2002).**

قارنت هذه الدراسة بين نظرية الاختبار الكلاسيكية (CTT)، ونظرية الاستجابة للمفردة (IRT)، في جزئية مسلك إحصاءات وبيانات (المفردة /المفحوص) المستمدة من هذين الإطارين القياسيين، وقد تم تحليل هذه البيانات وفحصها إمبريقياً باستخدام اختبار مكون من ١٥ مفردة، وعينة مكونة من ٦٠٠ مفحوص من تلاميذ الصف الثاني الإعدادي.

وأشارت النتائج الإمبريقية إلى أن إحصاءات (المفردة / المفحوص) المشتقة من هذين الإطارين القياسيين كانت قابلة للمقارنة، وقد استخدمت الدراسة نمط واحد فقط في تصميم المفردات الاختبارية، وقد أوصت الدراسة بضرورة تنوع أساليب تقدير الدرجات المخصصة للمفردات الاختبارية تبعاً لاختلاف خصائص هذه المفردات في البحوث المستقبلية.

- دراسة Prieto; Alonso and Lamarca (2003).

وهدفت الدراسة إلى تصميم استبيان يستخدم في المجال الإكلينيكي لقياس جودة الحياة وفق نظريتين: نظرية الاختبار الكلاسيكي (CTT)، ونظرية الاستجابة للمفردة (IRT) وتحديداً يمثلها نموذج راش أحادي البعد. وقد ركزت الدراسة على بحث أمرين الأول: الافتقار إلى نظام واضح لاتصالية المفردات التي تمثل بناء أحادي البعد، والثاني: الافتقار إلى بيانات واضحة لتدريج المقياس في ظل النظرية الكلاسيكية؛ على النقيض من نموذج راش الذي يقدم طريقة منهجية بديلة في تدريج المقياس تمكن من اختبار البناء الهرمي، وأحادية البعد لبيانات استبيان جودة الحياة.

وقد استخدم الباحثون في المقارنة طريقة الاختزال المتوازي لاستبيان مكون من ٣٨ مفردة من خلال تحليل استجابات عينة مكونة من ٩٤١٩ فرد. وقد أشارت النتائج فيما يخص النظرية الكلاسيكية (٢٠ مفردة في ٤ أبعاد)، بينما نموذج راش (٢٢ مفردة في بعدين) وكلا الأداتين أظهرتا نفس الخصائص السيكومترية، حيث تراوحت معاملات الارتباطات بين المفردة والدرجة الكلية في استبيان العشرين مفردة وفق نظرية (CTT) ، بين (0.45:0.75) بينما تراوحت هذه المعاملات في استبيان ال ٢٢ مفردة المصمم وفق نموذج راش بين (0.46:0.68)، وانحصر الثبات في الاستبيان الأول بين (0.82: 0.93)، بينما تراوح في الاستبيان الثاني بين (0.87: 0.94).

ورغم الاختلافات في المحتوى فإن درجات تحليل التغاير أكدت درجة عالية من الارتباطية بين الاختبارين (0.78: 0.95). وقد نجح نموذج راش في تحليل البيانات تبعاً لمستوى القياس بينما فشلت النظرية الكلاسيكية في مواءمة جودة المعايير التي وفرها نموذج راش.

- دراسة غادة عيد (٢٠٠٤)

هدفت هذه الدراسة إلى: ١- الكشف عن الفروق في الدرجات الخام لليقظة العقلية بين

الذكور والإناث وبين التخصصات العلمية و الأدبية. ٢- الكشف عن الفروق في الدرجات الحقيقية لليقظة العقلية والمقدرة بنموذج راش بين الذكور والإناث، وبين التخصصات الأدبية والعلمية. ٣-الكشف عن الفروق في الدرجات الحقيقية لليقظة

العقلية المقدرة بالنماذج الكلاسيكية بين الذكوروالإناث، وبين التخصصات العلمية والأدبية. ٤- الكشف عن الفروق بين الدرجة الحقيقية المقدرة بنموذج راش وتلك المقدرة بالنظرية الكلاسيكية. وتكونت عينة الدراسة من ٢٥٠ طالب وطالبة من طلاب وطالبات كلية التربية جامعة الكويت، بقسميها العلمي (رياضيات، كيمياء، فيزياء ١٥٠ طالب وطالبة) والأدبي (اللغة العربية، اللغة الإنجليزية) (١٠٠ طالب وطالبة) وطبق عليهم اختبار اليقظة العقلية كمقياس للذكاء العام. وتم التعويض في المعادلة

ت = م + رأ (س – م).

حيث "ت" تعني الدرجة الحقيقية، رأ معامل الثبات، م متوسط الدرجات، س الدرجة الخام ،

للحصول على الدرجات الحقيقية المناظرة للدرجات الخام باستخدام النظرية الكلاسيكية. واستخدمت طريقة PROX للحصول على القدرات الحقيقية للطلاب مقدرة باللوجيت من نموذج راش الأحادي ، بينت النتائج: وجود فروق ذات دلالة إحصائية بين متوسطي الدرجات الخام للطلاب والطالبات في مقياس اليقظة العقلية لصالح الطالبات، وكذلك وجود فروق ذات دلالة إحصائية بين متوسطي التخصصات الأدبية والتخصصات العلمية في مقياس اليقظة العقلية لصالح التخصصات العلمية، وعدم وجود فروق ذات دلالة إحصائية بين متوسطي الدرجات الحقيقية (نموذج راش) للطلاب والطالبات على مقياس اليقظة العقلية. وعدم وجود فروق ذات دلالة إحصائية بين متوسطي الدرجات الحقيقية للتخصصات العلمية والأدبية على مقياس اليقظة العقلية، عدم وجود فروق ذات دلالة إحصائية بين متوسطي الدرجات الحقيقية المقدرة بالنظرية الكلاسيكية (التقليدية) والمقدرة بنظرية السمات الكامنة.

- **دراسة كورفيل .Courville.(2004)**

هدفت هذه الدراسة إلى المقارنة بين الاتجاه التقليدي في بناء المفردات ، ونظرية الاستجابة للمفردة ، وقد فحصت الدراسة إمبريقياً بيانات مستمدة من مقياس شامل مرجع إلى معيار ، وكيف أمكن تتبع احصاءات (المفردة /الشخص) ضمن النظريتين.

وقد تضمنت الدراسة سؤالين أساسيين: الأول هل تختلف مقارنة بيانات (المفردة / الفرد) المشتقة من نظرية الاستجابة للمفردة ، والنظرية الكلاسيكية في الاختبارات؟ والثاني: ما مدى ثبات احصاءات المفردة في كل إطار قياسي عبر عينة المفحوصين؟

وأشارت النتائج إلى أنه رغم اختلاف النظريتين أنتج كل إطار من الإطارين القياسيين إحصاءات (مفردة /فرد) متشابهة. كما لم تختلف إحصاءات نظرية القياس الكلاسيكية (بالنسبة لهذه العينة) وإحصاءات نظرية القياس الحديثة.

- **دراسة السيد أبوهاشم (٢٠٠٦).**

وهدفت إلى معرفة الخصائص السيكومترية التي تتوفر لقائمة مداخل الدراسة باستخدام النظرية التقليدية ، وباستخدام نموذج راش ، والتحقق من إمكانية التنبؤ بالتحصيل الدراسي من خلال درجات الطلاب على قائمة مداخل الدراسة المشتقة من نموذج راش ، وتكونت العينة من ٢٤٤ طالباً وطالبة بالفرقة الرابعة بكلية التربية – جامعة الزقازيق.

وأشارت النتائج إلى توافر درجة مناسبة من الصدق والثبات لقائمة مداخل الدراسة باستخدام النظرية الكلاسيكية. وأشارت النتائج كذلك إلى وجود فروق ذات دلالة إحصائية عند مستوى (٠١.) بين متوسطات درجات مرتفعي ومنخفضي التحصيل الدراسي في مداخل الدراسة وذلك لصالح مرتفعي التحصيل. وأشارت قيم إحصاءات (المفردة / المفحوص) إلى ملاءمة الفقرات للنموذج ، وأشارت الإحصاءات إلى انخفاض قيم الخطأ المعياري للفقرات وللاختبار ككل وهو ما يدل على دقة فقرات القائمة في قياس السمة المطلوبة ، وقد أمكن كذلك الحصول على منحنى يعبر عن العلاقة بين الدرجات الكلية المحتلمة على الاختبار والقيم التقديرية للقدرة المناظرة لها. وأشارت النتائج إلى اختلاف الفقرات المكونة لمقياس مداخل الدراسة باختلاف النموذج الإحصائى المستخدم ، وكذلك وجود فروق بين مرتفعى ومنخفضى التحصيل الدراسى من طلاب الجامعة فى مداخل الدراسة ، لصالح المرتفعين وأشارت النتائج أيضاً إلى أن النسخة المطورة لقائمة مداخل الدراسة باستخدام نموذج راش منبىء جيد بالتحصيل الدراسي لدى طلاب الجامعة.

- دراسة مكدونالد وبونين.Macdonald and Paunonen(2007)

بالرغم من المزايا النظرية المعروفة للجميع عن نماذج نظرية الاستجابة للمفردة (IRT) في مواجهة النظرية الكلاسيكية (CTT). فإن البحث الذي يستهدف خصائصهما الإمبريقية قد أخفق في شرح الاختلافات، ومدى الاتساقية بين البيانات المستمدة من النظريتين.

وباستخدام أساليب دراسة مسلك إحصاءات (المفردة / الفرد) المشتقة من هذين الإطارين القياسيين. فقد أشارت النتائج إلى أن تقديرات معاملات صعوبة المفردات، وقدرة الفرد كانت قابلة للمقارنة، وتتمتع بقدر من الإتساق والثبات ودقيقة في ضوء النظرية الكلاسيكية، ولكن كانت تقديرات معاملات تمييز المفردات المستمدة من نظرية الاستجابة للمفردة أكثر دقة، وكانت معاملات تمييز المفردات المستمدة من النظرية الكلاسيكية دقيقة فقط في بعض ظروف التجريب.

- دراسة Chuan-Ju Lin (2008)

وهدفت هذه الدراسة إلى بحث الفروق بين عدة صيغ من الاختبارات في مقرر إرشادي باستخدام نظريتي:- الاستجابة للمفردة، والكلاسيكية وقد تم استخدام صيغ معدلة كل منها مكون من ٦٠ مفردة منتقاة من بنك أسئلة يضم ٦٠٠ مفردة، وقد صممت ثلاثة صيغ إحداها في ضوء النظرية الكلاسيكية، والصيغتين الآخريتين تم تصميمهما في ضوء نظرية الاستجابة للمفردة.

وقد تمت المقارنة بين الصيغ الاختبارية الثلاث في كل من: الاتساق الداخلي، والتطابق الإحصائي، والتطابق في المحتوى.

وأشارت النتائج إلى أن أداء المفردات في ضوء النظرية الكلاسيكية كان مشابهاً لأداء المفردات في صيغتي نظرية الاستجابة للمفردة.

- دراسة Progar and Socan(2008)

في هذه الدراسة الامبريقية استخدمت مجموعة بيانات حقيقية من الدراسة الدولية للعلوم والرياضيات (١٩٩٥م)، وذلك لبحث ثلاث أسئلة بحثية: الأول هل يمكن مقارنة بارامترات المفردة والفرد القائمة على كلا النظريتين ؟، والثاني: مامدى ثبات بارامترات

المفردة في كلا النظريتين عبر مجموعات مختلفة؟، والثالث ما مدى ثبات بارامترات (المفردة / الفرد) المستمدة من كلا النظريتين عبر مجموعات مختلفة من المفردات؟

أشارت النتائج إلى أن بارامترات (المفردة /الفرد) المستمدة من كلا النظريتين قابلة للمقارنة ، وأن بارامترات المفردة في كلا النظريتين أظهرت نفس خاصية الثبات عندما تم تقديرها في مجموعات مختلفة من المفحوصين ، وأن بارامترات الفرد في نظرية الاستجابة للمفردة كانت أكثر ثباتاً عبر مجموعات مختلفة من المفردات. وأن بارامترات المفردة في النظرية الكلاسيكية كانت ثابتة في مجموعات مختلفة من المفردات مثل نظرية الاستجابة للمفردة.

وأشارت النتائج أيضاً إلى أن بحث خاصية الثبات لم يمنع بحث أن بارامترات المفردة والفرد في ضوء نظرية الاستجابة للمفردة عموماً تفوقت إمبريقياً على بارامترات المفردة والفرد في ضوء النظرية الكلاسيكية ، ولكن هذا في حالة ما إذا استخدم نموذج ملائم من نماذج نظرية الاستجابة للمفردة.

- دراسة **Hernandez(2009)**

وهدفت هذه الدراسة إلى المقارنة بين معامل صعوبة المفردة ، ومعامل تمييز المفردة في اختبار الاستعداد العقلي السريع ، وذلك باستخدام طريقتين: طريقة النظرية الكلاسيكية ، وطريقة نظرية الاستجابة للمفردة IRT عبر ثلاث بارامترات إحصائية ، وقد تم جمع البيانات وتحليلها للعلاقة المحتملة بين خصائص المفردات الاحصائية باستخدام طرق: النظرية الكلاسيكية ، والاستجابة للمفردة.

وتشير النتائج إلى أن بارامترين من الثلاث تكاد تكون متشابهة في كل من نماذج نظرية الاستجابة للمفردة ، والنظرية الكلاسيكية في الاختبار اللفظي وغير اللفظي في حدود معامل تمييز المفردة.

- دراسة **Magno (2009)**

بحثت هذه الدراسة الاختلافات بين النظرية الكلاسيكية (CTT)، ونظرية الاستجابة للمفردة (IRT) باستخدام بيانات اختبار واقعي في الكيمياء لطلاب المرحلة الثانوية ، وقد تمت مقارنة النظريتين: من خلال عينتين ، واختبارين وقد تضمنت أوجه

المقارنة: مؤشرات صعوبة المفردات ، والاتساق الداخلي ، وأخطاء القياس. وكان نموذج الاستجابة للمفردة المستخدم بالتحديد هو نموذج راش أحادي البعد.

وقد تم اشتقاق عينتين متكافئتين من مدرسة ثانوية خاصة في الفلبين ، وقد تمت مقارنة هاتين المجموعتين من البيانات في مؤشرات صعوبة المفردات في الاختبارين ، ومعاملات ثبات التجزئة النصفية ، وطريقة ألفا لكرونباخ ، ومعاملات صعوبة المفردة باستخدام نموذج راش لثبات المفردات /المفحوص ، وتقديرات أخطاء القياس.

وقد وضعت النتائج حدود معينة للنظرية الكلاسيكية ووضحت مزايا استخدام نظرية الاستجابة للمفردة IRT وقد أظهرت النتائج أن مؤشرات الاتساق الداخلي للمفردات باستخدام نظرية الاستجابة للمفردة IRT كانت ثابتةً باختلاف العينات بينما فشلت مؤشرات الاتساق الداخلي الخاصة بالنظرية الكلاسيكية CTT في أن تكون ثابتة عبر العينات ؛ كذلك كان مدخل نظرية الاستجابة للمفردة IRT كان أقل أخطاءاً في القياس من مدخل نظرية الاختبارات الكلاسيكيةCTT.

- **دراسة موراليس (2009) Morales**

تتمحور هذه الدراسة حول بحث تقويم جودة القياس للمحددات التالية: الثبات والصدق ومدى تحيز البند ، وقد طبقت هذه الدراسة باستخدام النظرية الكلاسيكية ، ونظرية الاستجابة للمفردة لتقويم جودة التقييم المبني على قياس تحصيل طلاب الكلية في الرياضيات وتكونت العينة من ٨٠ طالب من طلاب السنوات الأولى والنهائية من كلية المعلمين في شرق الفلبين ، وقد تم تطبيق اختبار تحصيلي في الرياضيات عليهم ، وقد تم تحليل البيانات في اتجاهين الأول: الخصائص السيكومترية وقد تم تحليلها باستخدام النظرية الكلاسيكية ،ونظرية الاستجابة للمفردة ، الثاني وهو تحري مدى تحيز البنود وقد تم بحثه بطريقة الدالة التفاضلية للبنود.

وقد أشارت النتائج إلى أن نظرية الاستجابة للمفردة كانت أكثر دقة في تحليل بيانات الاختبار التحصيلي مقارنة مع النظرية الكلاسيكية.

- **دراسة سلفستر تباي (2009) .Silvestre-Tipay.**

وقد بحثت هذه الدراسة: مسلك إحصاءات (المفردة / الفرد) المستمدة من نظرية الاستجابة للمفردة والنظرية الكلاسيكية. وقد أجابت الدراسة عن الأسئلة التالية:

مامدى الاتساق بين مستويات صعوبة المفردات عبر النظريتين؟ الثاني: كيف يمكن مقارنة بيانات الاتساق الداخلي المستمدة في ضوء نظرية الاستجابة للمفردة مع بيانات الاتساق الداخلي للاختبار المستمدة من النظرية الكلاسيكية؟ الثالث: ما أبعاد قياس المفردات؟ الرابع: كيف يمكن مقارنة التمايز في توظيف المفردات بين النظريتين؟

وأشارت النتائج إلى أن إحصاءات (المفردة / الفرد) المستمدة من الإطارين القياسيين يمكن مقارنتهما. كما يوجد تشابه في بيانات (المفردة / الفرد) الإحصائية عبر العينات متشابهة رغم التفوق النظري لنماذج نظرية الاستجابة للمفردة. كما أنه يوجد مساحات اختلاف بين النظريتين لا يجب إغفالها في البحث والدراسة.

- **دراسة تغريد حجازي وزايد عطا (٢٠١٠)**

وهدفت هذه الدراسة إلى التحقق من مطابقة الاستجابات على فقرات اختبار أوتيس– لينون للقدرة العقلية في صورته المعدلة للبيئة الأردنية (الصورة ز) مع النظرية الحديثة في القياس. ولتحقيق هدف الدراسة تم تطبيق القسم الأول من الاختبار الذي يتكون من (٢٠) فقرة على عينة مؤلفة من (٤٠٠) طالباً وطالبة موزعين على الصفوف الأول والثاني والثالث والرابع الابتدائي تم اختيارهم عشوائياً من مدارس مديرية تربية إربد الأولى. وقد تم تحليل الاستجابات باستخدام النموذج الثنائي التدريج المنبثق عن النظرية الحديثة. أشارت نتائج التحليل إلى مطابقة الاستجابات على (١٩) فقرة من فقرات الاختبار لافتراضات نموذج راش، وحذف فقرة واحدة لم تطابق النموذج، وبلغت قيم معامل الثبات للأفراد والفقرات (٠,٧٥– ٠,٩٨) على التوالي. كما تمتع الاختبار بدلالات صدق متعددة.

- **دراسة جالدين ولورانسل Galdin & Laurencelle(2010)**

قام فيها مركز أبحاث مونت كارلو بدراسة الخواص الإحصائية لتقدير مستوى القدرة (θ) في نظرية الاستجابة للمفردة (IRT) مبنية على درجة تعميم بيانات مستمدة من تصميم متعدد العوامل ذو أربع حالات. و قد تم اختيار المفردات ثنائية التفرع، ونموذج الاستجابة للمفردة اللوجستي ثنائي البارامتر في مجال أحادي البعد. كما تضمنت إجراءات التقدير:– طريقة تقدير بارامترات انحياز البنود الهامشية، وتقدير EAP فيما يخص تقدير مستوى القدرة (θ). وقد نوقشت خاصية الثبات في النتائج.

وأظهرت النتائج أن تقدير مستوى القدرة (θ) بدا منحازاً، ومقيد بعدد المفردات، وتؤدي المفردات بشكل أفضل عندما يزداد عدد المفردات، وعدد المفحوصين. الأكثر من ذلك فإن بارامترات نظرية الاستجابة للمفردة لا تؤدي بشكل أفضل من النظرية الكلاسيكية.

- دراسة **Sharkness and DeAngelo(2011)**

تقارن هذه الدراسة بين تصميم الأداة السيكومترية للاختبارب النظرية الكلاسيكية (CTT) ونظرية الاستجابة للمفردة (IRT) لبناء النطاق مع البيانات المستمدة من معهد البحوث في التعليم العالي جامعة كاليفورنيا، و تشير النتائج إلى أنه بالرغم من أن كلا النظريتين:النظرية الكلاسيكية ،ونظرية الاستجابة للمفردة يمكن استخدامهما للحصول على نفس المعلمات حول مدى الاستفادة من بنود في قياس سمة كامنة، .فإن نظرية الاستجابة للمفردة توفر معلومات أكثر ثراءاً حول دقة القياس، فضلاًعن أنها أكثر وضوحاً في تحسين نطاق محتوى دراسي معين واسع وتدعم النتائج استخدام نظرية الاستجابة للمفردة لبناءالاختبارات وتطوير المحتويات على نطاق الدراسة في التعليم العالي.

التعليق على الدراسات السابقة:

١- فيما يخص قابلية المقارنة بين نظرية الاختبار الكلاسيكية (CTT)، ونظرية الاستجابة للمفردة الاختبارية (IRT). في إحصاءات (المفردة / الفرد) أكدت دراسات مثل:- Fan(1998) و.Hwan (2002)، وMacdonald&Paunonen(2007)، وProgar and Socan(2008) ، وMagno(2009)، وSilvestre-Tipay(2009) أن البيانات المستمدة من هذين الإطارين البحثيين قابلة للمقارنة.

٢- فيما يخص الدقة في القياس أشارت بعض الدراسات مثل دراسة Fan(1998)، ودراسة Macdonald & Paunonen(2007)، وSilvestre-Tipay(2009)، و Progar & Socan(2008) إلى أن قياسات النماذج المستخدمة من نظرية الاستجابة للمفردة كانت أكثر دقة من نظرية الكلاسيكية، وتفوقت عليها في حساب معاملات الصعوبة للمفردات، وكذلك قياسات الاتساق الداخلي.

٣- أكدت بعض الدراسات مثل (Courville(2004 ، و(Lin(2008 إلى التشابه بين نظريتي الاستجابة للمفردة ، والنظرية الكلاسيكي في البيانات الإحصائية الخاصة ب(المفردة / الفرد) وتحديداً معاملات تمييز المفردات في دراسة Hernandez(2009)

٤- فيما يخص عدد مفردات الاختبار، وعدد المفحوصين أكدت نتائج دراسة Galdin Laurencelle(2010) & إلى أن تطبيق نظرية الاستجابة للمفردة يتطلب زيادة عدد المفردات وزيادة عدد المفحوصين في العينة حتى تؤدي المفردات بشكل أفضل.

٥- فيما يخص دراسات المقارنة بين نظريتي الاستجابة للمفردة والنظرية الكلاسيكية باستخدام اختبارات في مادة الرياضيات فلم يجد الباحثان – في حدود علمهما- دراسات حديثة في ذلك باستثناء دراسة (Morales(2009 ، والتي بحثت مدى تحيز المفردات ، وقارن التحليل الإحصائي بين النظريتين.

٦- فيما يخص النماذج المستخدمة من نظرية الاستجابة للمفردة كان نموذج راش هو الأكثر استخداماً في العديد من الدراسات على سبيل المثال: Lee et al(1998) ، و(Prieto(2003 ودراسة السيد أبوهاشم (٢٠٠٦) ، و(Magno(2009.

٧- اقتصرت بعض الدراسات على بحث مدى مطابقة بعض المقاييس والاختبارات لافتراضات النظرية الحديثة في القياس دون توظيف هذه الاختبارات بفعالية في قياس متغيرات أو حل مشكلات تربوية مثل دراسة غادة عيد (٢٠٠٤)، و دراسة تغريد حجازي وزايد عطا(٢٠١٠) وهو ما يؤكد أن الدراسات العربية التي اتجهت للمنهج المقارن قليلة للغاية – في حدود علم الباحثين- وهو ما يؤكد أهمية البحث الحالي.

ويشير صلاح علام (٢٠٠٧) إلى أن إعادة إجراء تحليل المفردات يكون غير ضروري في حالة استخدام أحد نماذج الاستجابة للمفردة ، وذلك لأن المفردات الصالحة المتبقية بعد استبعاد المفردات التي لم تطابق النموذج تكون كافية للحصول على قيم تقديرية للسمة المراد قياسها لدى الأفراد. وهذه القيم تكون مستقلة عن عينة المفردات التي أجاب عليها الفرد.

ورغم أن هذه خاصية تميز هذه النماذج إلا أنها لا تخلو من مشكلات أبرزها أن استقلال بارامترات نماذج السمات الكامنة ، أي عدم تباين Invariance تقديرات قدرة الأفراد بتباين عينة المفردات التي يختبرون بها، وعدم تباين الخصائص السيكومترية للمفردات بتباين عينة الأفراد التي أجرى عليها الاختبار، ليس لها مايبررها في الواقع العملي ، ولكنها ممكنة فقط في إطار نظري إحصائي بحت. لذلك فإنه من المهم التحقق بصورة دورية من عدم حدوث تغيرات ربما تطرأ على قيم بارامترات نماذج السمات الكامنة ، التي يتم تقديرها نتيجة لتغير خصائص أفراد المجتمع المستهدف ، أو تغير طبيعة أو محتوى أو معنى المفردات بالنسبة إليهم بمرور الزمن. (صلاح علام ، ٢٠٠٧: ص ٢٢٥- ٢٢٦).

فروض البحث

١- تتوفر درجة مناسبة من الصدق والثبات للاختبار التحصيلي المرجع إلى هدف باستخدام نظرية القياس الكلاسيكية.

٢- تنطبق البيانات المستمدة من الاختبار التحصيلي المرجع إلى هدف في الجبر على نموذج راش.

٣- لاتختلف إحصاءات (المفردة /الفرد) في الاختبار التحصيلي المرجع إلى هدف وفق نظرية: الاستجابة للمفردة عنها وفق النظرية الكلاسيكية.

٤- لاتختلف الخصائص السيكومترية للاختبار التحصيلي المرجع إلى هدف وفق نظرية الاستجابة للمفردة عن الخصائص السيكومترية للاختبار وفق النظرية الكلاسيكية.

بعد الانتهاء من تصحيح الاستجابات، استخدم الباحثان البرنامج الإحصائي (SPSS 17)، والبرنامج الإحصائي (WINSTEPS)في إجراء التحليلات الإحصائية اللازمة للإجابة عن أسئلة البحث،واختبار صحة الفروض والتي تتضمن:

- معامل ألفا &.
- تم إجراء التحليل العاملي، للتأكد من أحادية البعد للاختبار.
- قدرات الأفراد.
- الخطأ المعياري في تقدير قدرات الأفراد .

- المتوسط الحسابي والانحراف المعياري لتقديرات قدرات الأفراد .
- المتوسط الحسابي والانحراف المعياري للأخطاء المعيارية في القياس لتقديرات قدرات الأفراد.
- قيم معالم الصعوبات للمفردات.
- المتوسط الحسابي والانحراف المعياري لتقديرات قيم صعوبات المفردات .
- المتوسط الحسابي والانحراف المعياري للأخطاء المعيارية في القياس لتقديرات قيم صعوبات المفردات.

أدوات البحث

قام الباحثان بتصميم اختبار تحصيلي مرجع إلى هدف في الوحدة الثانية مقرر الجبر بالصف الأول الإعدادي:-

الصورة (أ) تم تحليل بياناتها في ضوء نظرية الاختبار الكلاسيكي وتكونت من (٢٥) مفردة من نوع الاختيار من متعدد ، وتم تصميم الصورة (أ) باستخدام جدول مواصفات يوضحه الجدول التالي:

جدول (١) مواصفات الاختبار التحصيلي المرجع إلى هدف الصورة (أ) في ضوء النظرية الكلاسيكية

المهارات / الدروس	التذكر	الفهم	التطبيق	التحليل	المجموع	الوزن النسبي *
الحدود والمقادير الجبرية	1	1	1	2	5	%20
الحدود المتشابهة	1	1	1	1	4	%16
ضرب الحدود الجبرية وقسمتها	–	1	1	–	2	%8
جمع الحدود الجبرية وطرحها	1	1	1	1	4	%16
ضرب حد جبري في مقدار جبري	1	1	1	1	4	%16
ضرب مقدار جبري مكون من حدين في مقدار جبري آخر	–	–	1	–	1	%4
قسمة مقدار جبري على حد جبري	–	1	1	–	2	%8
التحليل بإخراج العامل المشترك الأعلى	–	1	1	1	3	%12
المجموع	4	7	8	6	25	%100

*تم تحديد الوزن النسبي لكل موضوع تبعاً لأهميته في الوحدة وآراء السادة المحكمين

يتضح من الجدول (١) أن المفردات تغطي مهارات معرفية في التذكر (٤) مفردة ، و الفهم (٧) مفردة ، والتطبيق (٨) مفردة ، والتحليل (٦) مفردة.

حساب معامل ثبات الصورة (أ) من الاختبار:-

قام الباحثان بتطبيق الاختبار على عينة مكونة من ٨٠ تلميذ من تلاميذ الصف الأول الإعدادي منهم (٤١) تلميذ من مدرسة الرمل الإعدادية بنين بإدارة شرق الإسكندرية التعليمية و(٣٩) تلميذ من مدرسة حمدي عاشور الإعدادية بنين بإدارة المنتزة التعليمية ،وقد بلغت قيمة معامل ثبات كودر ريتشاردسون – Ku R (٨٢٣.) وهو معامل ثبات مرتفع ، وقد قام الباحثان كذلك بحساب مؤشرات ثبات ألفا كرونباخ للاختبار بطريقة الاتساق الداخلي ، باستخدام معادلة كودر ريتشاردسون (٢٠– kr) وقد بلغت قيمته (٨٥.) وهو يدل على تمتع الاختبار بمؤشرات ثبات عالية والجدول (٢) يوضح قيم معامل ثبات ألفا لكل مفردة من مفردات الاختبار وعددها (٢٥) مفردة.

جدول (٢) قيم معامل ثبات ألفا لمفردات الاختبار التحصيلي الصورة (أ)

رقم المفردة	معامل ثبات ألفا للمفردة	رقم المفردة	معامل ثبات ألفا للمفردة
1	.8104	14	.8323
2	.8146	15	.8155
3	.8185	16	.8024
4	.8026	17	.8683
5	.8003	18	.8091
6	.8137	19	.8026
7	.8052	20	.8031
8	.8225	21	.8355
9	.8140	22	.8226
10	.8216	23	.8208
11	.8151	24	.8249
12	.8065	25	.8072
13	.8134	معامل ثبات ألفا الكلي	.8538

يتضح من الجدول (٢) أن معاملات ثبات ألفا للمفردات كلها تقريباً لم تقل عن ٨.، ويشير السيد أبوهاشم (٢٠٠٦) إلى أن طريقة ألفا كرونباك تعد مؤشراً على الاتساق الداخلي لمفردات المقياس مع بعضها البعض ، ومع الدرجة الكلية ، ويؤكد أن الاتساق الداخلي أكثر قرباً لمعنى الثبات.

وفي هذا الإطار تحديداً يؤكد Magno(2009) على أهمية تقدير الاتساقات الداخلية لكل اختبار يتم تحليل نتائجه باستخدام النظرية الكلاسيكية وذلك على العينة الحالية، وذلك لأن تقديرات الاتساق الداخلي مرتبطة بالعينة التي تم تطبيق الاختبار عليها.(P.8-9)

خطوات تصميم الاختبار التحصيلي المرجع إلى هدف في ضوء نظرية الاستجابة للمفردة.

اعتمد الباحثان على طريقة تجمع بين تحليل الإجراءات ، والتحليل البنائي الهرمي في تحليل الكفايات الأساسية التي يقيسها الاختبار.

وقد مر بناء الاختبار بعدة مراحل توضحها الخطوات التالية:-

١- **مرحلة التحليل:-**

أ- تحديد الكفايات الأساسية في الوحدة الثانية بمقرر الجبر.

ب- تحديد النطاق السلوكي للكفايات.

ج- صياغة الأهداف السلوكية وترتيبها في نسق هرمي.

٢- **مرحلة البناء:-** وتضمنت كتابة مفردات الاختبار ، وقام الباحثان بصياغة مفردة واحدة لقياس كل هدف من أهداف الاختبار السلوكية ، وقد روعى أن تكون المفردات من نوع الاختيار من متعدد (خمسة بدائل).

٣- **مرحلة التجريب:-**

أ- التحقق من صدق محتوى الاختبار:

- قام الباحثان بتحديد الكفايات الرئيسة المقاسة في الوحدة الثانية بمقرر الجبر الخاصة بالحدود والمقادير الجبرية والعمليات عليها ؛ وقد تم تحديدها في (٨) كفايات رئيسة وقد تم تحليل هذه الكفايات وفق الطريقة التي حددها صلاح علام (٢٠٠٧)، والطريقة جمعت بين تحليل الإجراءات والتحليل البنائي الهرمي ، وقد تمت صياغة الأهداف السلوكية في ضوء هذا التحليل بطريقة موضوعية قابلة للملاحظة والقياس وعددها (٢٥) هدف سلوكي وكل هدف تم تحويله إلى مفردة في الاختبار.

– قام الباحثان بتحكيم المفردات بواسطة مجموعة من الخبراء في مجال القياس النفسي وتدريس الرياضيات*، وقد طبق الباحثان طريقة التحكيم المعززة بالمعلومات التي اقترحها بابام Popham لتحديد مستويات الأداء حيث استرشد فيها المحكمون ببيانات تتعلق بالأداء الفعلي للأفراد في الاختبار التحصيلي المرجع إلى هدف ؛ مثل إحصاءات خاصة بالمفردات كمعاملات الصعوبة والتمييز، ومتوسط درجات الاختبار بعد تطبيقه على العينة المناسبة من المتعلمين، وبعض تعليقات الخبراء حول مفردات الاختبار والاختبار ككل، واقتراحات المعنيين باستخدام الاختبار حول الحد الأدنى لمستوى الاجتياز في الاختبار. (صلاح علام ، ٢٠٠٧: ص٢٧٤)

وطلب من كل محكم توضيح مدى مناسبة كل مفردة لقياس المجال الذي تقيسه وتم في هذه الخطوة حساب معامل اتفاق المحكمين على مدى مناسبة كل مفردة وقد أكد كل من روفينلي وهامبلتون Rovinelli&Hambleton أن طريقة حساب معامل اتفاق المحكمين تتطلب اختيار عدد من المتخصصين للتحقق من أن كل مفردة تقيس الهدف الإجرائي الذي وضعت لقياسه في الاختبار. (محمود منسي، ٢٠٠٣: ص٢٠٨)

وقد تراوحت نسب الاتفاق بين المحكمين على مفردات الاختبار(٢٥ مفردة)بين (٨٢,٣%)و(١٠٠%) وقد كانت هذه المعلومات عرضة للمناقشات بين الباحثين والمحكمين حتى تم التوصل إلى اتفاق حول الحد الأدنى الأمثل لمستوى الاجتياز المطلوب وهو ٧٠%.

وقد اتفق الخبراء على أهمية تمكن التلاميذ من الكفايات الأساسية الثماني بالوحدة موضع القياس.

ثبات الاختبار التحصيلي المرجع إلى هدف:

قام الباحثان باستخدام طريقة إعادة التطبيق ؛ في حساب قيمة ثبات الاختبار التحصيلي المرجع إلى هدف ،وذلك على عينة مكونة من (٤١) تلميذ من تلاميذ الصف الأول

---

* يشكر الباحثان : أ.د / محمود منسي أستاذ القياس والتقويم النفسي والتربوى، و أ.د / محمود الإبياري استاذ المناهج وطرق تدريس الرياضيات، ود. أيمن مصطفى أستاذ مساعد المناهج وطرق تدريس الرياضيات للمعاونة الصادقة في تصميم وتحكيم أداة البحث.

الإعدادي بمدرسة الرمل الإعدادية بنين، وذلك بفاصل زمني عشرة أيام * وقاما بحساب نسبة الاتفاق بين التطبيقين حيث تم تحديد مستوى المحك (نسبة ٧٠% التي حددها المحكمين) بإجابة ١٧ مفردة إجابة صحيحة من إجمالي مفردات الاختبار البالغ عددها (٢٥) مفردة. وقد بلغت قيمة نسبة الاتفاق (.٣١٣)، وقد تم في ضوءها حساب درجة الاتفاق (كابا) فبلغت ٠١. ويشير محمود منسي (٢٠٠٣) إلى أهمية أن تكون قيمة (ك) موجبة –أكبر من الصفر– لأن القيمة السالبة تشير إلى عدم اتساق بين نتائج التطبيقين، والقيمة الدنيا ليست مهمة في التطبيق العملي لهذه الطريقة، وتمثل القيمة (.٣١٣) نسبة الاتفاق الخالية من عوامل الصدفة، ويشير محمود منسي (٢٠٠٣) إلى أن الفرق بينها وبين قيمة (ك) يرجع إلى أثر عوامل الصدفة.

(محمود منسي، ٢٠٠٣: ص ٢٠٦)

عينة البحث:

تألفت العينة من ٨٠ تلميذ من التلاميذ الذكور بمدرستين من مدارس محافظة الإسكندرية بواقع: (٤١) تلميذ من مدرسة الرمل الإعدادية بنين بإدارة شرق الإسكندرية التعليمية، و(٣٩) تلميذ من مدرسة محمد حمدي عاشور الإعدادية بإدارة المنتزة.

نتائج البحث

تقارن النتائج بين طريقتين منهجيتين في القياس: النظرية الكلاسيكية، ونظرية الاستجابة للمفردة من خلال صورتان من الاختبار التحصيلي: الصورة (أ) في ضوء النظرية الكلاسيكية، والصورة (ب) في ضوء نظرية الاستجابة للمفردة.

وقد تمت مقارنة المنهجيتين عبر العينة، والصورتين (أ،ب) في تقديرات معاملات الصعوبة، والاتساقات الداخلية، وأخطاء القياس.

---

* يشكر الباحثان ا/ سامي دياب المدرس الأول في مدرسة الرمل الإعدادية بنين لمعاونته الصادقة في تطبيق الاختبار.

مناقشة وتفسير النتائج

فيما يخص الفرض الأول تمتع الاختبار التحصيلي المرجع إلى هدف بمؤشرات ثبات مناسبة في الإطارين القياسيين المستخدمان في هذا البحث: النظرية الكلاسيكية، ونظرية الاستجابة للمفردة، كما أن حساب ثبات الاختبار بأكثر من طريقة يجعل الباحثان يطمئنان إلى ثبات الاختبار.

ويشير Magno(2009) إلى أن أغلب الباحثين يلجأون إلى حساب معامل ألفا كرونباخ في النظرية الكلاسيكية، وذلك لحساب معاملات الاتساق الداخلي للمفردات أو حساب معامل ارتباط المفردة بالدرجة الكلية لأنها أكثر سهولة ومتاحة في برامج الحزم الإحصائية.

ولكنه يؤكد على أنه لابد أن يكون هناك تركيز أكبر على دراسة وتفسير كيفية حساب الثبات باستخدام نظرية الاستجابة للمفردات الاختبارية IRT (p9).

بالنسبة لصدق الاختبار فإن التحقق منه وفق النظرية الحديثة أتاح تحليل محتواه بدقة واستخدام طريقة التحليل البنائي الهرمي في تحليل الكفايات الرئيسة بمادة الاختبار أيضاً يجعل الباحثان مطمئنان لصدق محتوى مادته.

فيما يخص التحقق من صحة الفرض الثاني: الذي ينص على أنه:" تنطبق البيانات المستمدة من الاختبار التحصيلي المرجع إلى هدف في الجبر على نموذج راش".

لمعرفة درجة ملاءمة نموذج راش لبيانات الاختبار التحصيلي المرجع إلى هدف قام الباحثان بالتحقق من ذلك من خلال الافتراضات التالية:-

افتراض أحادية البعد:- يكون الاختبار أحادي البعد إذا كانت البيانات المشاهدة لمفردات ذلك الاختبار تقيس سمة كامنة واحدة فقط لا أكثر ويمكن التحقق من افتراض احادية البعد باستخدام طريقة التحليل العاملي:- وقبل الشروع في تنفيذ هذه الطريقة قام الباحثان بالتحقق من الشروط الواجب توافرها في مصفوفة معاملات الارتباط الخاصة بالتحليل العاملي على النحو الآتي:

- يجب أن تكون قيمة محدد مصفوفة معاملات الارتباط لا تساوي صفر $|R| \neq 0$
- قياس تجانس العينة بالنسبة لحجم العينة:- ويتم ذلك من دلالة قيمة كاي تربيع (Chi-Square) لاختبار بارتلت ، والجدول التالي يوضح قيمة كاي تربيع ، ودرجة الحرية ومستوى دلالتها.

جدول (٣) قيمة كاي تربيع لاختبار بارتلت

قيمة كاي تربيع	درجات الحرية	مستوى الدلالة
1067.85	300	0.001

يتضح من الجدول (٣) أن قيمة كاي تربيع تساوي (١٠٦٧,٨٥) ودرجة الحرية (٣٠٠) وهي دالة عند (٠١.).

- قياس تجانس العينة ككل: ويتم ذلك بواسطة حساب قيمة كايزر- ماير- أولكن (K.M.O) والذي يجب أن لا يقل عن (٥.) حسب محك كايزر والجدول (٤) يلخص قيمة (K.M.O)

جدول (٤) قيمة كايزر – ماير – أولكن (K.M.O)

كايزر – ماير – أولكن (K.M.O)	0.672

يتضح من الجدول (٤) أن قيمة كايزر –ماير – أولكن (K.M.O) تساوي (٦٧٢.) وهي أكبر من (٥.) بمعنى أنه دال إحصائياً.وهو يؤكد على ملاءمة البيانات للتحليل العاملي.

بعد تحقق شروط التحليل العاملي قام الباحثان باستخدام التحليل العاملي الاستكشافي بطريقة المكونات الأساسية Principle Components لاستجابات التلاميذ عن مفردات الاختبار وذلك للتحقق ما إذا كان الاختبار يقيس سمة واحدة . كما هو مفترض.

وقد اعتمد الباحثان على عينة مكونة من (٨٠) تلميذ من تلاميذ الصف الأول الإعدادي بمدرستي: الرمل الإعدادية بنين ، وحمدي عاشور الإعدادية بنين ، وحجم العينة يزيد عن الحد الأدنى للقيام بالتحليل العاملي حيث يشير حمزة دودين (٢٠١٠) إلى ضرورة ألا يقل عدد المشاهدات عن (٥٠) لكل سمة مفترضة على الأقل.

وقد تم حساب قيمة الجذر الكامن Eigenvalue ، ونسبة التباين المفسر Explained Variance ، وكذلك التباين المفسر التراكمي لكل عامل من العوامل.

والجدول (٥) يوضح قيم الجذر الكامن ، والتباين المفسر ، والتباين المفسر التراكمي للعوامل المستخلصة.

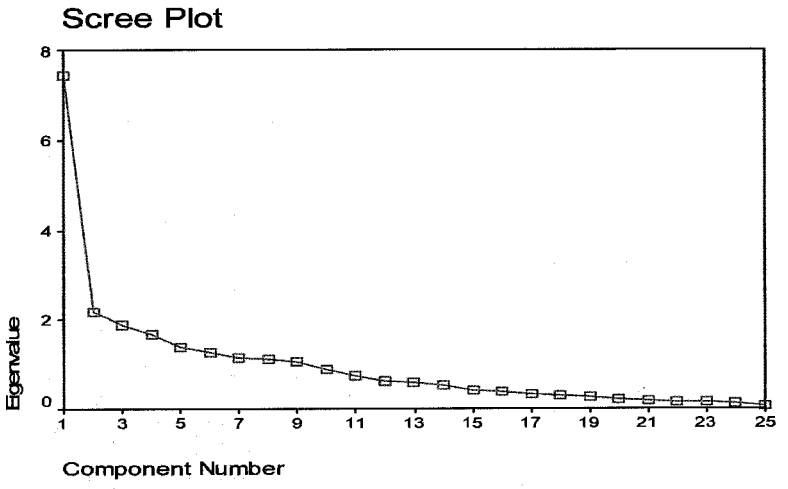
جدول (٥) التباين الكلي المفسر للتحليل العاملي الخاص بالاختبار التحصيلي

الجذور الكامنة الابتدائية Initial Eigenvalues			مجموع مربعات التشبعات المستخلصة (الجذور الكامنة النهائية) Extraction sums of squared Loadings		
العامل	الجذر الكامن	نسبة التباين المفسر	نسبة التباين المفسر التراكمي	الجذر الكامن	نسبة التباين المفسر	نسبة التباين المفسر التراكمي
1	7.45	29.80	29.80	7.45	29.80	29.80
2	2.17	8.70	38.51	2.17	8.70	38.51
3	1.89	7.57	46.09	1.89	7.57	46.09
4	1.66	6.66	52.75	1.66	6.66	52.75
5	1.39	5.55	58.31	1.39	5.55	58.31
6	1.27	5.11	63.43	1.27	5.11	63.43
7	1.14	4.57	68.00	1.14	4.57	68.00
8	1.11	4.44	72.44	1.11	4.44	72.44
9	1.05	4.20	76.64	1.05	4.20	76.64
10	0.87	3.48	80.13			
11	0.72	2.88	83.02			
12	0.60	2.42	85.44			
13	0.58	2.31	87.76			
14	0.52	2.10	89.86			
15	0.41	1.67	91.54			
16	0.38	1.53	93.07			
17	0.33	1.34	94.42			
18	0.28	1.15	95.57			
19	0.27	1.09	96.67			
20	0.21	0.86	97.53			
21	0.17	0.70	98.23			
22	0.14	0.58	98.82			
23	0.14	0.56	99.38			
24	0.10	0.43	99.82			
25	0.004	0.17	100.00			

يتضح من الجدول (٥) أن قيم الجذور الكامنة النهائية للعوامل المستخلصة أكبر من الواحد الصحيح حسب معيار كايزر كما يتضح أيضاً من الجدول (٥) أن العامل الأول

يفسر أكبر نسبة تباين في درجات التلاميذ مقارنة ببقية العوامل؛ حيث بلغت قيمة الجذر الكامن للعامل الأول (٧,٤٥)، بنسبة تباين (٢٩,٨٠%)، مما يعني أن هذا العامل هو المسيطر على تفسير التباين الكلي لدرجات الاختبار، وبالتالي يعد هذا الاختبار أحادي البعد أي أن هناك سمة كامنة واحدة يقيسها الاختبار وهي المسئولة عن تفسير مايحدث من تباين في درجات الاختبار.

طريقة الرسم البياني: تعد أحد مخرجات التحليل العاملي، وممثلاً لقيم الجذور الكامنة للعوامل الممكن استخلاصها من الاختبار، والرسم البياني في الشكل (١) يوضح تمثيل قيم الجذور الكامنة للعوامل المرشحة للاستخلاص من النموذج.



شكل (١) تمثيل قيم الجذور الكامنة للعوامل المرشحة للاستخلاص من النموذج.

يتضح من الشكل (١) ارتفاع قيمة الجذر الكامن للعامل الأول مقارنة ببقية العوامل، وبالتالي هو العامل المسيطر على تفسير التباين الكلي لدرجات الاختبار مقارنة ببقية العوامل التي يمكن استخلاصها.

افتراض الاستقلال المحلي:- ويمكن التحقق منه من خلال فحص معامل الثبات لجميع الدرجات الكلية المختلفة (مستويات القدرة المختلفة)، والذي يعد مؤشراً عن مدى الاتساق الداخلي لمفردات الاختبار، وذلك عند تثبيت الدرجة الكلية المكتسبة من قبل التلاميذ على الاختبار، وإن هذا التثبيت للدرجة الكلية سيكون بمنزلة ضبط تأثير

السمة الكامنة التي يقيسها الاختبار أوما تسمى بالقدرة ، والجدول (٦) يوضح قيم معاملات الثبات بطريقة تحليل التباين للاختبار عند المستويات المختلفة لقدرة التلاميذ.

جدول (٦) قيم معاملات الثبات بطريقة تحليل التباين للاختبار عند المستويات المختلفة لقدرة التلاميذ

الدرجة الكلية	عدد المفردات	عدد الحالات	قيم الثبات F	الدرجة الكلية	عدد المفردات	عدد الحالات	قيم الثبات F
24	25	1	0.00	10	25	7	0.00
22	25	1	0.00	9	25	3	0.00
21	25	6	0.00	8	25	4	0.00
20	25	6	0.00	7	25	4	0.00
19	25	12	0.00	6	25	5	0.00
18	25	7	0.00	5	25	1	0.00
17	25	3	0.00	4	25	1	0.00
16	25	5	0.00	3	25	1	0.00
15	25	2	0.00				
14	25	3	0.00				
13	25	2	0.00				
12	25	4	0.00				
11	25	2	0.00				

يتضح من الجدول (٦) أن قيم معامل الثبات بطريقة تحليل التباين لجميع الدرجات تساوي صفر، وعند مقارنة هذه القيم بمعامل ثبات كودر ريتشاردسون لجميع الحالات (ن = ٨٠)، والذي يساوي (٨٢٣.) نستنتج أن المسئول عن ارتفاع معامل ثبات الاختبار هو قدرة التلاميذ (الدرجة الكلية لكل تلميذ)، بحيث إذا أزيل تأثير هذا العامل انعدمت تقريباً الارتباطات والاتساقات في درجات التلاميذ على مفردات الاختبار، وبالتالي تحقق افتراض الاستقلال المحلي للاختبار.

افتراض تساوي مؤشرات التمييز:- وفقاً لمقترح هاملتون وآخرون Hambleton et al.(1991) فإن فحص توزيع معاملات ارتباط مفردات الاختبار بالدرجة الكلية المكتسبة في الاختبار يعطي تصوراً عن مدى تجانس مؤشرات تمييز مفردات الاختبار.

وللتحقق من تساوي معاملات التمييز قام الباحثان بحساب معاملات الارتباطات ثنائية التسلسل الحقيقية (Point-Biserial Correlations)، وذلك بالطريقة التي وضحها Varma(2011) والجدول(٧) يوضح نتائج هذا الحساب:-

جدول (٧) قيم معاملات الارتباطات ثنائية التسلسل الحقيقية لمفردات الاختبار

رقم المفردة	معامل الارتباط ثنائي التسلسل الحقيقي	رقم المفردة	معامل الارتباط ثنائي التسلسل الحقيقي
1	.5123	14	-.1080
2	.4431	15	.3827
3	.3289	16	.6950
4	.7013	17	.0371
5	.7354	18	.5487
6	.4467	19	.6827
7	.6513	20	.6732
8	.1626	21	-.1178
9	.4241	22	.1800
10	.2102	23	.2496
11	.4088	24	.1110
12	.6085	25	.5881
13	.4353		

يتضح للباحثان من الجدول (٧) تقارب قيم معاملات الارتباطات الحقيقية ثنائية التسلسل للمفردات باستثناء المفردات رقم: ١٠، ١٤،٢١،١٧،٢٢، ٢٤ وهي نفسها المفردات التي تتمتع بمعاملات صعوبة عالية وهو ما يتطلب بحث وضعها ودراسة منحنيات القدرة الخاصة بها بدقة ؛ ومن الجدول السابق يمكن حساب إحصاءات توزيع معاملات الارتباطات ثنائية التسلسل الحقيقية لمفردات الاختبار حسب الجدول (٨):

جدول (٨) قيم الإحصاءات الوصفية لمعاملات الارتباطات ثنائية التسلسل الحقيقية لمفردات الاختبار

معاملات الارتباطات ثنائية التسلسل الحقيقية	العينة الكلية	المدى	المتوسط	الانحراف المعياري
	80	0.85	0.39	0.25

يتضح من الجدول (٨) أن متوسط توزيع معاملات الارتباطات ثنائية التسلسل الحقيقية (Point-Biserial Correlations) مقداره (٣٩.)، وبانحراف معياري قيمته (٢٥.) وأن المدى (٨٥.)

وتشير الإحصاءات السابقة إلى أن توزيع مؤشرات التمييز لجميع مفردات الاختبار إلى حد ما متجانس، بحيث يمكن القبول بافتراض تساوي مؤشرات التمييز، وبالتالي استخدام نموذج راش لتحليل بيانات الاختبار.

افتراض تدني عامل التخمين:- وقد تحقق الباحثان من افتراض تدني عامل التخمين من خلال فحص مؤشرات صعوبة مفردات الاختبار على مستوى العينة ، ومن ثم تحديد المفردات التي يقل معامل صعوبتها عن (٢٥.) والجدول (٩) يوضح قيم مؤشرات المفردات الأكثر صعوبة.

جدول (٩) قيم مؤشرات المفردات الأكثر صعوبة

رقم المفردة	مؤشر الصعوبة
10	0.17
14	0.22
24	0.20

وللتأكد من تضمين نموذج راش لأثر التخمين أم لا ، قام الباحثان بفحص أداء (١٠%) من التلاميذ الأقل حصولاً على الدرجة الكلية المكتسبة – حسب اقتراح هامبلتون وسوامناثان (١٩٨٥)– ودراسة أدائهم على المفردات الأكثر صعوبة.

وذلك لمقارنة نسبة أولئك التلاميذ الأقل قدرة (والذين أجابوا إجابة صحيحة على تلك المفردات الصعبة) بالقيمة النظرية للتخمين العشوائي في حالة البدائل الخمسة ، والتي قيمتها (٢٠.) والجدول (١٠) يوضح نسبة الطلاب منخفضي القدرة الذين أجابوا بطريقة صحيحة على المفردات الصعبة في الاختبار.

جدول (١٠) نسبة التلاميذ منخفضي القدرة الذين أجابوا إجابة صحيحة على المفردات الصعبة في الاختبار

رقم المفردة	عدد التلاميذ 10%	عدد الذين أجابوا إجابة صحيحة على المفردة منهم	نسبة الذين أجابوا إجابة صحيحة على المفردة منهم
10	8	0	0
14	8	1	%12.5
24	8	0	0

يتضح من الجدول (١٠) أن نسبة التلاميذ منخفضي القدرة الذين أجابوا بطريقة صحيحة على المفردات الصعبة في الاختبار هي نسبة منخفضة جداً، وذلك بمقارنتها مع النسبة النظرية (٢٠.) إذا قام التلاميذ بالإجابة بشكل عشوائي على هذه المفردات – وذلك لأن فئات الاستجابة على فقرات اختبار الاختيار من متعدد موضع هذا البحث هي خمسة فئات استجابة وبذلك يستنتج الباحثان أن لا يوجد تأثير تقريباً لعامل التخمين على الأداء.

افتراض أن الاختبار ليس اختبار سرعة:- وأمكن للباحثين التحقق منه من خلال فحص نسبة التلاميذ الذين أكملوا الاختبار وكذلك فحص المفردات التي لم يجب عنها التلاميذ. ووفقاً لمايراه هامبلتون وآخرون (١٩٩١) Hambleton et al فإنه إذا كان (٧٥%) من التلاميذ قد أكملوا الإجابة على الاختبار، وإذا كان (٨٠%) من مفردات الاختبار قد أجاب عليها التلاميذ فإن السرعة لن تعد في هذه الحالة عامل مهم في الأداء على الاختبار.

وحيث أن نسبة التلاميذ الذين أكملوا الاختبار هي (١٠٠%)، وأن جميع مفردات الاختبار قد أجاب عليها التلاميذ فإن هذا يعني أن الاختبار يقيس قدرة وليس سرعة.

هذه النتائج تتفق جزئياً مع نتائج دراسات كل من: السيد أبوهاشم (٢٠٠٦)، و(2007)Macdonald and Paunonen، و(2008)Progar and Socan، و(2009)Hernandez وعلى زكري (٢٠١١).

فيما يخص التحقق من صحة الفرض الثالث*

٥- والذي ينص على أنه "لاتختلف إحصاءات (المفردة /الفرد) في الاختبار التحصيلي المرجع إلى هدف وفق نظرية: الاستجابة للمفردة عنها وفق النظرية الكلاسيكية."

وللتحقق من هذا الفرض قام الباحثان بتقدير صعوبة مفردات الاختبار وذلك باستخدام برنامج Winsteps؛ حيث يقوم البرنامج باستبعاد جميع البيانات التامة والصفرية (بالنسبة للتلاميذ والمفردات على السواء) قبل أن يبدأ أي تحليل، وفي البحث الحالي لم يتم استبعاد أي تلميذ من العينة، أو أي مفردة من مفردات الاختبار.

حيث بدأ التحليل بعدد (٨٠) تلميذ، و(٢٥) مفردة، وهدف التحليل الأول إلى حذف التلاميذ غير الملائمين، وقد قام الباحثان بحساب إحصاءات الملاءمة Fit Statistics وقد حدد الباحثان مدى الملاءمة للبرنامج مابين (+٢)، و(-٢) بناءاً على نتائج الدراسات السابقة.

والجدول (١١) يوضح قيم معاملات الصعوبة لكل مفردة وفق نظرية الاستجابة للمفردة

جدول (١١) التدريج النهائي لمفردات الاختبار مرتباً تبعاً لمستوى الصعوبة

رقم المفردة	تقدير الصعوبة باللوجيت	الخطأ المعياري باللوجيت
1	-16.81	18.33
2	86.03	18.34
3	95.3	18.24
4	49.56	2.74
5	17.53	3.97
6	86.42	18.34
7	66.15	4.68
8	49.46	2.74
9	21.19	3.42
10	66.39	7.51
11	68.35	5.19
12	61.88	4.38
13	32.28	2.63

* يشكر الباحثان د. محمد إبراهيم للمعاونة الصادقة في تحليل بيانات الفرض الثالث

رقم المفردة	تقدير الصعوبة باللوجيت	الخطأ المعياري باللوجيت
14	87.2	18.31
15	74.26	7.23
16	67.12	5.21
17	33.25	2.53
18	47.98	4.44
19	58.54	3.65
20	60.91	4.06
21	32.48	2.77
22	55.38	4.66
23	57.47	3.48
24	49.66	2.88
25	30.15	2.82

يتضح من الجدول (١١) أن أقل قدرة كانت للمفردة التاسعة والخامسةوهي بالتالي تقابل أعلى معامل صعوبة.

هذا بالنسبة للمفردات أما بالنسبة للأفراد فالجدول (١٢) يوضح التحليل على أساس قدرة المفحوص مقدرة باللوجيت:-

جدول (١٢) قدرة كل فرد من أفراد العينة مقدرة باللوجيت

تقدير القدرة باللوجيت	الخطأ المعياري باللوجيت	التكرارات
10.68	11.24	1
15.47	11.26	1
15.95	11.34	1
20.72	8.98	1
26.69	7.81	1
26.8	7.84	1
27.29	9.25	1
28.43	8.18	1
28.83	9.22	1
30.57	8.38	2
31.34	8.59	1
31.47	7.4	2
32.46	7.35	1
32.76	7.38	3
35.03	7.52	1
35.75	7.57	1
37.93	6.96	5

التكرارات	الخطأ المعياري باللوجيت	تقدير القدرة باللوجيت
1	6.97	37.95
1	7.89	38.09
1	7.65	38.53
1	7.09	38.86
36	6.61	39.72
5	7.53	39.96
2	6.95	40.78
1	6.76	42.41
1	6.76	43.25
1	6.63	45.83
3	6.66	46.05
2	5.8	56.13

يتضح من الجدول (١٢) أنه كلما زادت قيمة القدرة قلت قيمة الخطأ المعياري للقياس حيث كانت أقل قيمة للخطأ المعياري (5.8) مقابلة لأعلى قيمة في القدرة وهي (56.13) وهذه النتيجة تتفق مع ما توصل إليه Magno(2009) حيث أكد على انخفاض قيمة الخطأ المعياري في نظرية الاستجابة للمفردة مقارنة بالنظرية الكلاسيكية.

وحيث إن نموذج راش يؤسس تقديراته لقدرات الأفراد على افتراضين متعلقين بالعلاقة بين قدرات الأفراد ، وصعوبات المفردات ؛ حيث يتم أثناء عملية التدريج حذف الأفراد ، ذوي نماذج الاستجابة غير الملائمة لافتراضات نموذج راش ، ينتج عن عملية التدريج جدول قدرة يوضح القدرة المقابلة لكل درجة خام كلية محتملة على المقياس.

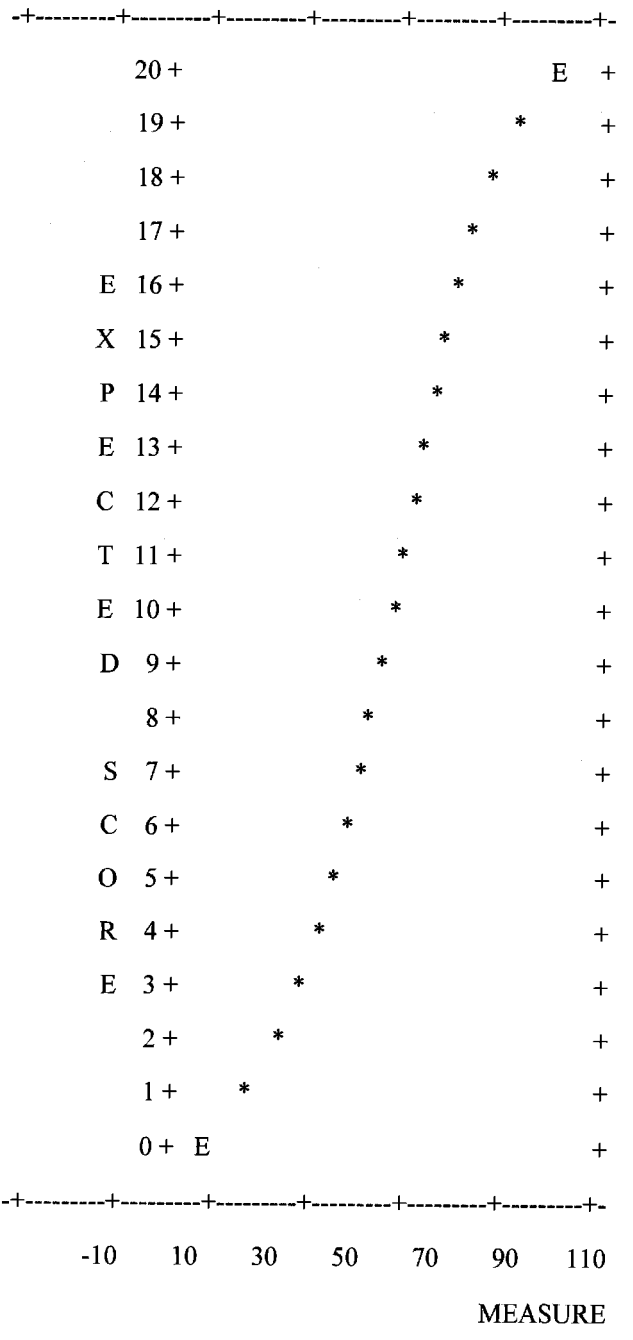
والجدول (١٣) يوضح القدرة المقابلة لكل درجة خام كلية محتملة على الاختبار التحصيلي

جدول (١٣) القدرة المقابلة لكل درجة

خام كلية وفق تدريج بيانات الاختبار على نموذج راش

الدرجة	القدرة المقابلة	الخطأ المعياري
0	-4.62	18.81
1	8.86	11
2	17.91	8.45
3	24.13	7.42
4	29.19	6.85
5	33.62	6.48
6	37.65	6.21
7	41.36	5.99
8	44.84	5.81
9	48.12	5.66
10	51.26	5.56
11	54.31	5.5
12	57.32	5.49
13	60.35	5.53
14	63.46	5.65
15	66.76	5.86
16	70.38	6.2
17	74.57	6.79
18	79.87	7.89
19	88	10.57
20	100.85	18.57

يتضح من الجدول (١٣) أنه كلما زادت الدرجة الكلية زادت القدرة المقابلة لها، وأن أعلى درجة مقابلة لأعلى قدرة. والشكل (٢) يلخص العلاقة بين الدرجة الكلية والقدرة المقابلة لها بالنسبة لتلاميذ العينة:-



شكل (٢) العلاقة بين الدرجات الكلية والقدرات المقابلة لها لأفراد العينة

وعند استخدام المقياس مرة أخرى فيما بعد ، يتم حساب الدرجة الخام الكلية لكل فرد ، ثم الكشف عن القدرة المقابلة لهذه الدرجة ، وذلك يعني حصول جميع الأفراد الحاصلين على نفس الدرجة الخام الكلية على نفس القدرة، بصرف النظر عن المفردات التي أجاب عنها المفحوص بشكل صحيح سواء أكانت صعبة أم سهلة. (وليد القفاص ، ٢٠٠٦:ص ١٥٩)

ويتفق الباحثان مع وليد القفاص (٢٠٠٦) حيث تشير الدرجات الخام الكلية الموضحة في الجدول (١٣) إلى وصول كل التلاميذ الذين يحصلون على نفس الدرجة الخام الكلية إلى نفس المستوى من الأداء على الاختبار المتمثل في القدرة المقابلة لهذه الدرجة.

ولبحث فاعلية النموذج الكلاسيكي في قياس القدرة التحصيلية لتلاميذ الصف الأول الإعدادي في الوحدة الثانية بمقرر الجبر، قام الباحثان بتحويل الدرجة الخام لكل تلميذ إلى درجة حقيقية من خلال المعادلة الموضحة في الإطار النظري حيث تشير غادة عيد (٢٠٠٤) إلى أن الدرجة الحقيقية يمكن الحصول عليها من خلال المعادلة :

$$ت = م + رأ (س - م)$$

حيث ت الدرجة الحقيقية ، وم متوسط الدرجات ، ورأ معامل ثبات الاختبار، وقد قام الباحثان في ضوء تحليل النظرية الكلاسيكية بتحويل الدرجة الملاحظة إلى درجة حقيقية باستخدام برنامج SPSS الإصدار السابع عشر ، والجدول (١٤) يوضح بيان بالدرجة الحقيقية المقابلة للدرجة الملاحظة للتلاميذ في العينة (٨٠ تلميذ) لكل مجموع درجات خام كلي على الاختبار:-

جدول (١٤) الدرجة الحقيقية المقابلة لكل

مجموع درجات خام على الاختبار في ضوء النظرية الكلاسيكية

الدرجة الخام	عدد التلاميذ	الدرجة الحقيقية المقابلة	الدرجة الخام	عدد التلاميذ	الدرجة الحقيقية المقابلة
24	1	22.55	12	4	12.35
22	1	20.85	11	2	11.5
21	6	20	10	7	10.65
20	6	19.15	9	3	9.8
19	12	18.3	8	4	8.95
18	7	17.45	7	4	8.1
17	3	16.6	6	5	7.25
16	5	15.75	5	1	6.4
15	2	14.9	4	1	5.55
14	3	14.05	3	1	4.7
13	2	13.2	المجموع	80	

يتضح من الجدول (١٤) أن الفرق بين الدرجات الكلية الخام ، والدرجات الحقيقية المقابلة لها ليس كبيراً ، وقد وجد الباحثان أن هذا الفرق بين الدرجتين غير دال إحصائياً، وقد يرجع إلى أخطاء القياس ويلاحظ كذلك أن الدرجات الحقيقية المقابلة أقل في قيمتها من الدرجات الخام الأعلى من المتوسط (١٤،٣)، بينما كانت الدرجات الحقيقية المقابلة أعلى في قيمتها من الدرجات الخام الأقل من المتوسط (١٤،٣) وذلك بنسبة قليلة.

ويتضح مما سبق أن هناك تناظر في مستويات قيم إحصاءات المفردات في الإطارين القياسيين: الكلاسيكي والاستجابة للمفردة فيما يخص تقدير القدرة التحصيلية ممثلة بالدرجة الكلية في النموذج الكلاسيكي ، وممثلة بقيمة القدرة في النموذج الحديث.

ويمكن القول أن الفرض الثالث لم يتحقق جزئياً وذلك نظراً لاختلاف تقديرات قيم ومعاملات الصعوبة لمفردات الاختبار بين النظريتين وهو ما تؤكده نتائج الفرض الرابع.

النتائج الخاصة بالفرض الرابع والذي ينص على أنه: – " لاتختلف الخصائص السيكومترية للاختبار التحصيلي المرجع إلى هدف وفق نظرية الاستجابة للمفردة عن الخصائص السيكومترية للاختبار وفق النظرية الكلاسيكية."

وللتحقق من صحة هذا الفرض قام الباحثان بحساب معاملات صعوبة مفردات الاختبار التحصيلي الصورة (أ) التي تم تحليلها وفق النظرية الكلاسيكية وذلك بطرح عدد الإجابات الخاطئة من عدد الإجابات الصحيحة لكل مفردة وقسمة النتيجة على العدد الكلي للإجابات والجدول (١٥) يوضح قيم هذه المعاملات:-

جدول (١٥) معاملات صعوبة مفردات الاختبار التحصيلي وفق النظرية الكلاسيكية

رقم المفردة	معامل الصعوبة	رقم المفردة	معامل الصعوبة
1	0.66	14	0.22
2	0.85	15	0.42
3	0.90	16	0.61
4	0.66	17	0.32
5	0.55	18	0.68
6	0.78	19	0.47
7	0.71	20	0.57
8	0.96	21	0.51
9	0.70	22	0.18
10	0.17	23	0.33
11	0.80	24	0.20
12	0.67	25	0.65
13	0.65		

يتضح من الجدول (١٥) ارتفاع معاملات صعوبة المفردات رقم: ١٠، ١٤، ٢٢، ٢٤

وفي هذا الصدد يشير رجاء أبوعلام (٢٠٠٣) إلى أنه قبل اتخاذ قرار بحذف أي من المفردات التي مؤشرات صعوبتها مرتفعة ، فإنه لابد من دراسة محتواها لنتأكد من أنه لابد فعلاً من حذفها ، إلا أننا إذا وجدنا أن المفردة من حيث بناءها سليمة فلابد من البحث عن عوامل أخرى ربما تكون السبب في ذلك.(رجاء أبوعلام ، ٢٠٠٣:ص ٣٣٤)

ويتفق الباحثان مع رجاء أبوعلام (٢٠٠٣) فعندما نصمم اختبار تحصيلي مرجع إلى هدف-وفق الطريقة الموضحة- يتمتع بصدق محتوى جيد ؛ وتكون معاملات صعوبة بعض مفرداته مرتفعة فإن هذا مرتبط بمستوى القدرة التي تقيسها هذه المفردة وهو ما يتطلب بحث مؤشرات صعوبة هذه المفردات باستخدام نظرية الاستجابة للمفردة للوقوف على حقيقة وضع هذه المفردات بدقة فقد يكون الخلل في طريقة التحليل المتبعة في النظرية الكلاسيكية أو أن أفراد العينة لم يصلوا إلى مستوى القدرات التي

تتطلبها الاستجابة على هذه المفردات وهذا كله يوضح أهمية المقارنة بين النظريتين في التحليل الإحصائي لهذه البيانات.

وبالمقارنة بين جدول (١١)، وجدول(١٥) نجد أن النتائج اختلفت بعض الشيء بين النظريتين في تحديد صعوبة الأداء على كل مفردة ففي حين أكدت النظرية الكلاسيكية على أن أكثر المفردات صعوبة هي المفردات أرقام: (١٠)، (١٤)، (٢٤) فقد تمتعت اثنتان من هذه المفردات بقيم قدرة مناسبة في تحليل نظرية الاستجابة للمفردة- مؤشر على تمتعها بمعاملات صعوبة مناسبة - وهما المفردتين (١٠)، و(٢٤) في حين كان معامل صعوبة المفردة (١٤) منخفضاً، أن أكثر المفردات صعوبة تبعاً لتحليل نظرية الاستجابة للمفردة كانت هي المفردات: (٥)،و (٩)، و (٢٥).

هذا وقد قام الباحثان بحساب مؤشرات التمييز لكل مفردة من مفردات الاختبار التحصيلي وذلك بطريقة المقارنة الطرفية حيث تم حساب متوسط درجة كل مفردة لدى المجموعة التي تمثل أعلى ٢٧% من العينة في المجموع الكلي للدرجات، وتم طرح من كل قيمة متوسط درجة كل مفردة لدى المجموعة التي تمثل أدنى ٢٧% من العينة في المجموع الكلي للدرجات والجدول (١٦) يوضح قيم هذه المؤشرات:-

جدول (١٦) مؤشرات تمييز مفردات الاختبار التحصيلي في ضوء النظرية الكلاسيكية

رقم المفردة	مؤشر التمييز	رقم المفردة	مؤشر التمييز
1	0.6363	14	-0.1363
2	0.4091	15	0.6364
3	0.2727	16	0.8636
4	0.8182	17	0.3182
5	0.8636	18	0.6818
6	0.5	19	0.9546
7	0.7727	20	0.8636
8	0.0909	21	0.1363
9	0.5	22	0.2273
10	0.2273	23	0.5
11	0.4545	24	0.2273
12	0.7273	25	0.6818
13	0.6818		

ويتضح من الجدول (١٦) أن أغلب قيم معامل التمييز كانت موجبة ماعدا المفردة رقم (١٤) حيث تتدخل عوامل مثل التخمين في اختيار البدائل.

يشير محمود منسي (٢٠٠٣) إلى أن مفردة الاختبار الجيدة هي التي يزيد معامل تمييزها عن ٢. ويتضح من الجدول السابق أن كل مفردات الاختبار التحصيلي الصورة (أ) التي تم تحليلها في ضوء النظرية الكلاسيكية قد تميزت بمؤشرات تمييز مناسبة باستثناء المفردات: ٨ ، ١٤، ٢١؛ وهذه المفردات تقيس كفايات رئيسية في المقرر تم تحليلها وتحكيمها بالتالي قبل اتخاذ قرار بشأن استبعادها لابد من بحث وضعها في ضوء تحليل الصورة (ب) باستخدام نظرية الاستجابة للمفردة حيث يتضح من نتائج الفرض الثالث في جدول (١١) أن هذه المفردات قيم قدرات التلاميذ في الأداء عليها مناسبة باستثناء المفردة (٢١) التي ينخفض قيمة قدرة الأداء عليها قليلاً عن المفردتين الآخرتين وهو ما يؤكد صدق تحليل هذه المفردات وفق نموذج راش ؛ فقد يكون تقارب الأداء عليها بين الطرفين-وفق النظرية الكلاسيكية- دليل على توافر القدرات التي تقيسها لدى العينة كلها وهو أمر غير مستبعد ؛ وهذا ما يميز الدراسات المقارنة أو الدراسات في مجال القياس التربوي بصفة عامة – في رأي الباحثان- أن ما يهمنا هو بحث الأسباب الكامنة وراء ضعف المؤشرات أكثر من الاهتمام بمجرد استبعاد المفردات فقد يكون الخلل في طريقة التحليل الإحصائي نفسها.

ويتضح من النتائج السابقة أن الأخطاء المعيارية في القياس لنظرية الاستجابة للمفردة كانت أقل من الأخطاء المعيارية في القياس للنظرية الكلاسيكية ، وأن تحليل قدرة التلاميذ على الأداء على مفردات الاختبار التحصيلي باستخدام نظرية الاستجابة للمفردة كان أكثر تعبيراً عن القدرات التحصيلية للتلاميذ ؛ فكلما ارتفع الأداء على هذه المفردات انخفضت قيمة صعوبتها وهو ما اختلف قليلاً عند المقارنة مع التحليل السيكومتري لنفس المفردات على نفس العينة ولكن باستخدام النظرية الكلاسيكية ؛ فكون البيانات في النظريتين قابلة للمقارنة كما أكد الفرضين الثالث والرابع لا يعني حتمية تطابق بيانات (المفردة / الفرد) في كلا التحليلين الإحصائيين. وقد اتفق الباحثان في ذلك مع نتائج دراسات كل من: Magno(2009)، و Hernandez(2009) ، واتفق الباحثان كذلك مع دراسة Silvestre-Tipay(2009) في إمكانية مقارنة بيانات (المفردة / الفرد) عبر النظريتين وأن الاستجابة للمفردة كان لها التفوق النظري والتطبيقي على النظرية الكلاسيكية في التقدير الدقيق للقدرة التحصيلية المقاسة ، وقد اتفق الباحثان

مع Morales(2009) في أهمية تصميم اختبارات تحصيلية في الرياضيات وفق نظرية السمات الكامنة ، وتحليل البيانات وفق النظريتين وقد استخدم عينة مشابهة لعينة البحث الحالي قوامها ٨٠ طالب.

وبذلك يقبل الباحثان الفرض الرابع وأن بيانات الاختبار التحصيلي المستمدة من النظريتين كانت قابلة للمقارنة رغم الفروق في التقديرات التي ظهرت في قيم معاملات الصعوبة والتي تدعو إلى التأني في اتخاذ قرارات استبعاد المفردات من الاختبار المصمم قبل التحقق من دقة قيم صعوبتها بأكثر من إطار قياسي واحد.

الخلاصة هناك حاجة لمزيد من الدراسات حول المقارنة بين القياس التقليدي الكلاسيكى وأساليب ونماذج القياس الحديثة في مجال القياس النفسي والتربوى واستكمالا للجهد البحثي الحالى توصي الدراسة الحالية بإجراء المزيد من الدراسات والبحوث حول النقاط التالية:-

- المقارنة بين فاعلية نماذج السمات الكامنة والكلاسيكية في قياس صعوبة وسهولة المفردات بالاختبارات التحصيلية في المقررات الدراسية الأخرى.
- استخدام نموذج نظرية السمات الكامنة في تقدير كفاءة مفردات اختبار الكفايات للمعلمين.
- المقارنة بين نتائج نموذج راش أحادي البارامتر ونموذج لورد ثنائي البارامتر ونموذج بيرنبوم ثلاثي البارامتر.
- اختلاف الدرجة الحقيقية للأداء على الاختبارات والمقاييس المعرفية واللامعرفية باختلاف النوع والتخصص الدراسي.

* * *

المراجع

- السيد محمد أبوهاشم (٢٠٠٦). دراسة مقارنة بين النظرية التقليدية ونموذج راش في اختيار قائمة مداخل الدراسة لدى طلاب الجامعة، مجلة كلية التربية بالزقازيق، العدد ٥٢، يناير ٢٠٠٦.
- أمينة محمد كاظم (١٩٨٨). استخدام نموذج راش في بناء اختبار تحصيلي في علم النفس وتحقيق التفسير الموضوعي للنتائج ،دولة الكويت: مطبوعات جامعة الكويت ومتاح الكترونياً على الرابط:

http://al-mostafa.info/data/arabic/depot2/gap.php?file=003404.pdf

- أمينة محمد كاظم (١٩٨٨ ب): مستوى العينة وتدرج بنك الاسئلة باستخدام نموذج راش: دراسة تجريبية. ابحاث مؤتمر علم النفس الرابع، ص ص ٣٩٠ – ٤١٩.
- أنور محمد الشرقاوى و سليمان الخضرى الشيخ و أمينة محمد كاظم و نادية محمد عبد السلام (١٩٩٦): اتجاهات معاصرة في القياس والتقويم النفسي والتربوى، القاهرة، الانجلو المصرية.
- تغريد حجازي وزايد عطا(٢٠١٠). مدى مطابقة استجابات اختبار أوتيس – لينون للقدرة العقلية في صورته المعدلة للبيئة الأردنية (الصورةj) مع النظرية الحديثة للقياس.، مجلة جامعة الشارقة للعلوم الإنسانية والاجتماعية، المجلد ٧ – العدد ٢ جمادى الآخرة ١٤٣١هـ، يوليو ٢٠١٠ م.
- حمزة دودين (٢٠١٠). التحليل الإحصائي المتقدم للبيانات باستخدام SPSS، دار المسيرة للطباعة والنشر والتوزيع، الأردن، عمان.
- رجاء أبو علام (٢٠٠٣). التحليل الإحصائي للبيانات باستخدام برنامج SPSS، دار النشر للجامعات، مصر، القاهرة.
- صلاح الدين محمود علام (١٩٨٧). دراسة موازنة ناقدة لنماذج السمات الكامنة والكلاسيكية فى القياس النفسي والتربوى. المجلة العربية للعلوم الانسانية، الكويت، العدد ٢٧، المجلد ٧، ص ص ١٨ – ٤٣.
- صلاح الدين محمود علام (٢٠٠٧). الاختبارات التشخيصية مرجعية المحك: في المجالات التربوية والنفسية والتدريبية ،القاهرة: دار الفكر العربي.
- علي محمد زكري (٢٠١١). تطوير وتدريج اختبار (أوتيس–لينون) للقدرة العقلية المستوى المتوسط الصورة (J) باستخدام نموذج راش، مجلة كلية التربية جامعة المنصورة – العدد ٧٥ –الجزء الثاني– يناير ٢٠١١

- غادة عيد (٢٠٠٤). الدرجة الحقيقية المقدرة باستخدام نظرية السمات الكامنة والنظرية الكلاسيكية "دراسة سيكومترية"، مجلة جامعة أم القرى للعلوم التربوية والاجتماعية والإنسانية، المجلد السادس عشر، العدد الثاني، جمادى الأولى ١٤٢٥ هجري، يوليو ٢٠٠٤ م.
- محمود عبد الحليم منسي (٢٠٠٣). التقويم التربوي، الإسكندرية: دار المعرفة الجامعية، ط٢
- نادية محمد عبد السلام (١٩٩٢): "ثلاثة محكات رئيسية لتحديد مستوى الاتقان في الاختبارات محكية المرجع – دراسة امبريقية"، المجلة المصرية للدراسات النفسية، العدد الثالث، ديسمبر، ص ص ٨٨ – ١٠٦.
- هشام فتحى جاد الرب (٢٠٠٧). العلاقة بين المكونات المعرفية اللازمة للاجابة الصحيحة على البند ومستوى صعوبته باستخدام تحليل المسار ونموذج راش.المجلة المصرية للدراسات النفسية، المجلد السابع عشر، العدد ٥٥، ص ص ٢٦٦ – ٣١٠.
- وليد كمال عفيفي القفاص (٢٠٠٦). تقدير قدرات الأفراد ذوي نماذج الاستجابة الشاذة والعادية بحسب نموذج راش باستخدام نسب موزونة (W-ratio) للبنود بحسب مستوى صعوباتها، مجلة دراسات نفسية، المجلد ١٦، العدد الاول يناير ٢٠٠٦، ص ص ١٦٠:١٣٧.
- Chuan-Ju Lin(2008). Comparisons between Classical Test Theory and Item Response Theory in Automated Assembly of Parallel Test Forms, The Journal of Technology, Learning, and Assessment, Volume 6, Number 8 · April 2008
- Courville,T.(2004). An Empirical Comparison Of Item Response Theory And Classical Test Theory Item/Person Statistics , Submitted to the Office of Graduate Studies of Texas A&M University ,in partial fulfillment of the requirements for the degree of Doctor Of Philosophy, Educational Psychology, available on line at http://repository.tamu.edu/bitstream/handle/1969.1/1064/etd-tamu-2004B-EPSY-Courville-2.pdf?sequence=1
- Fan ,X.(1998). Item Response Theory And Classical Test Theory: An Empirical Comparison Of Their Item/Person Statistics, Educational And Psychological Measurement, June 1998, Vol58n3,p357(25).
- Fraley, R. Chris; Waller, Niels G.; Brennan, Kelly A.(2000). An item response theory analysis of self-report measures of adult attachment.,Journal of Personality and Social Psychology, Vol 78(2), Feb 2000, 350-365.

- Galdin , M & Laurencelle, L(2010). Assessing parameter invariance in item response theory's logistic two item parameter model: A Monte Carlo investigation , Tutorials in Quantitative Methods for Psychology 2010, Vol. 6(2), p. 39-51.
- Hambleton, R.K.&swaminthan,H.(1985).ITEM- RESPONSE THEORY:principles and applications,Boston:Kluwer-Nijhoff.
- Hambleton, R.K.&Swaminthan,H.,&Rogers,h.j.(1991). Fundamentals Of Item Response Theory.Newbury Park,Ca:Sage. The International Professional Publishers.
- Hambleton, R.K.&Swaminthan,H(1985). Item Response Theory , Principles And Applications. Boston: Kluwer. Nijhoff Publishing amembers of the Kluwer Academic Publishers Group.
- Hernandez,R(2009). Comparison of the Item Discrimination and Item Difficulty of the Quick-Mental Aptitude Test using CTT and IRT Methods, The International Journal of Educational and Psychological Assessment,April 2009, Vol. 1, Issue 1, pp. 12-18.
- Hwang,D(2002). Classical Test Theory And Item Response Theory: Analytical And Empirical Comparisons , Paper Presented At The Annual Meeting Of The Southwest Educational Research Association (Austin, TX, feb. 14-16) available at: www.eric.ed.gov , ERIC no: ED 466779.
- Lee,Y; Izard,J; Yeoh,O.(1998). Teacher knowledge of biological evolution from the perspectives of classical test and item response theory, available at: www.eric.ed.gov , ERIC no: ED 440 110.
- Lin, C.-J. (2008). Comparisons between Classical Test Theory and Item Response Theory in Automated Assembly of Parallel Test Forms. Journal of Technology, Learning, and Assessment, 6(8). Retrieved [24/9/2011] from http://www.jtla.org.
- Linacre,J.M(2006). A user's guide to WINSTEPS MINISTEP: Rasch-Model computer programs, ISBN 0-941938-03-4, available at:www.winsteps.com.
- Macdonald,P and Paunonen, S(2007). A Monte Carlo Comparison of Item and Person Statistics Based on Item Response Theory versus Classical Test Theory, Educational and Psychological Measurement June 1, 2007 67: 394-409.
- Magno, C.(2009). Demonstrating the Difference between Classical Test Theory and Item Response Theory Using Derived Test Data, The International Journal of Educational and Psychological Assessment April 2009, Vol. 1, Issue 1, pp. 1-11.

- Morales,R.(2009). Evaluation of Mathematics Achievement Test: A Comparison between CTT and IRT, The International Journal of Educational and Psychological Assessment,April 2009, Vol. 1, Issue 1, pp. 19-26.
- Prieto,L; Alonso ,J and Lamarca,R (2003). Classical test theory versus Rasch analysis for quality of life questionnaire reduction, Health and Quality of Life Outcomes 2003, 1:27 doi:10.1186/1477-7525-1-27,the electronic version is available at:www.hqlo.com/content/1/1/27
- Progar,S and Socan ,G. (2008). An empirical comparison of Item Response Theory and Classical Test Theory, Horizons of Psychology, 17, 3, 5-24.
- Sharkness,J and DeAngelo,L(2011) Measuring Student Involvement: A Comparison of Classical Test Theory and Item Response Theory in the Construction of Scales from Student Surveys, Res High Educ (2011) 52:480–507Published online: 30 November 2010available at:www.Springerlink.com
- Silvestre-Tipay, J. L. (2009). Item response theory and classical test theory: an empirical comparison of item/person statistics in a biological science test. international Journal of Educational and Psychological Assessment, v.1, no.1, April 2009, p.19-31.
- Varma,S.(2011). Preliminary Item Statistics Using Point-Biserial Correlation and P-Values, educational data systems, Inc. 15850 Concord Circle,Suite A,Morgan Hill CA 95037 available online at:- http://www.eddata.com/resources/publications/EDS_Point_Biserial.pdf

* * *