استخدام نموذج (راش) في تكافؤ درجات صور اختبارية مرجعية المحك ومقارنة أداء الطلاب عليها في ضوء بعض المتغيرات

إعـــــداد

د. جاد الله أبو المكارم جاد الله

باحث بالمركز القومي للامتحانات والتقويم التربوي

ملخص الدراسة :

تعد الاختبارات التحصيلية أحد الوسائل الرئيسية لعمليات التقويم التي تعتمد عليها المؤسسات التعليمية ، للتعرف على مقدار التقدم في تحصيل الطلاب وعلى الرغم من استخدامها على نطاق واسع ، إلا أنها ما زالت تعاني بعض المشكلات التي تعترض موضوعيتها . مثل اعتماد الدرجة الكلية للاختبار على عينة المفردات، عدم اطراد درجاته مع قيمة المتغير المقيس ، تغير تكـوين ومعنى مفرداته بمضي الزمن ، بالإضافة إلى مشكلات الصدق والثبات . ولقد حاول علماء القياس التوصل إلى الموضوعية في قياس تقدير السلوك وأسفرت جهودهم عن نظرية الاستجابة للمفردة والتي أصبحت بديلا عن المدخل الكلاسيكي في بناء الاختبارات .

ثم تطورت أساليب تقويم تحصيل المتعلمين وأدواته ، حيث لم تقتصر على تحديـد المكانـة النسبية للمتعلم بين أقرانه ، إنما تحديد ما حققه من مهارات وما أتقنه من كفايات . ومن ثم نشأت فكرة الاختبارات مرجعية المحك كبديل عن الاختبارات مرجعية الجماعة التي وجه إليها العديد من الانتقادات .

ولقد قامت الدراسة الحالية باستخدام نموذج (راش) كأحد نمـاذج الاسـتجابة للمفـردة فـي تكافؤدرجات صور اختبارية مرجعية المحك في مادة (الإحصاء) المقررة على طـلاب المرحلـة الثانوية العامة ومقارنة أداء الطلاب عليها في ضوء متغيري الجنس (ذكر / إناث) والتخصـص (علمي / أدبي) للتحقق من مدى إتقان الطلاب للمهارات المتضمنة هذه المادة .

وقد أظهرت النتائج وجود فروق دالة إحصائيا بين متوسطات درجات الذكور / الإناث لصالح عينة الذكور ، وان كان حجم التأثير ضئيلا ، بينما لم توجد هذه الفروق بين طـلاب (العلمـي / الأدبي) ، كما لم يوجد تأثير لتفاعل متغيري الجنس والتخصص على الأداء في هذه المادة . كمـا أشارت النتائج إلى وجود فروق دالة إحصائيا بين الذكور / الإناث لصالح الـذكور فـي النسـب المئوية لتحقيق مهارتين فقط ، بينما لو توجد هذه الفروق بين النسب المئوية لتحقيق مهارات هذه المادة بين طلاب التخصص (العلمي / الأدبي) ، ولقد تحققت جميع المهارات لدى عينات الدراسة النوعية بنسب تراوحت ما بين (٧١,١٧%) إلى (٩٠,٥٢%)

استخدام نموذج (راش) في تكافؤ درجات صور اختبارية مرجعية المحك ومقارنة أداء الطلاب عليها في ضوء بعض المتغيرات

إعـــداد

د. جاد الله أبو المكارم جاد الله

باحث بالمركز القومي للامتحانات التقويم التربوي

المقدمة :

تعد الاختبارات أحد الوسائل الرئيسية لعمليات التقويم التى تعتمد عليها المؤسسات التعليمية والتربوية للتعرف على مقدار التقدم فى تحصيل الطلاب، ومدى تحقيقهم للأهداف التعليمية، ومن هذا المنطلق تعتبر أيضا أساسا مهما فى تشخيص صعوبات التعلم لدى المتعلمين ، وتحديد مدى كفاءتهم.

وعلى الرغم من استخدام الاختبارات التحصيلية لأغراض متعددة فى عملية التقويم التربوى على نطاق واسع، إلا إنها مازالت تعانى من بعض المشكلات التى تعترض دقتها وموضوعيتها، ومن أهم تلك المشكلات اعتماد الدرجة الكلية للاختبار على عينة المفردات التى تشتمل عليها اعتمادا أساسيا، وعدم إطراد درجات الاختبار مع قيمة المتغير المقاس، وتغير تكوين ومعنى مفردات الاختبار بمضى الزمن بالاضافة إلى المشكلات المتعلقة بصدق وثبات هذه الاختبارات.

(صلاح علام : ١٩٨٧ : ٢١)

ولقد فطن بعض علماء القياس النفسى والتربوى إلى خطورة تلك المشكلات، وما يترتب عليها من أضرار بالغة تتمثل فيما يمكن أن تتوصل إليه تلك الاختبارات من نتائج غير واقعية تؤدى إلى تنبؤات غير صادقة، ومن ثم بذلت محاولات عديدة بهدف التوصل إلى الموضوعية Objectivity المنشودة فى قياس وتقدير السلوك، وقد أسفرت تلك الجهود عن اتجاهات معاصرة في مجال القياس والتقويم منها نظرية السمات الكامنة Latent Trait Theory (LTT) أو نظرية الاستجابة

للمفردة Item Response Theory (IRT) ، وقد حظى هذا المدخل الجديد باهتمام كثير من علماء القياس المعاصر من أمثال هاملتون Hampleton ، رايت Wright، لورد Lord وغيرهم حيث ساعد فى التغلب على كثير من المشكلات الناجمة عن استخدام المدخل الكلاسيكى فى بناء وتصميم الاختبارات، وتحليل وتفسير نتائجها ومواجهة أوجه القصور فيها.

(صلاح علام ١٩٨٥ : ٢٠-٢٤)

مشكلة الدراسة:

تقاس مدى صدق وموضوعية نتائج الاختبارات بمدى دقة الأساليب المتبعة فى تصميمها وانتقاء مفرداتها، وكذلك فى وصفها للقدرة التى يقيسها الاختبار من حيث التدرج فى الصعوبة بما يشمل المستويات المختلفة للقدرة ، وأيضا دقة الأساليب المتبعة فى تفسير نتائج تلك الاختبارات.

وقد واكبت التطورات التى حدثت فى العقود القليلة الماضية فى مجال تصميم الاستراتيجيات التعليمية تطورات مناظرة فى أساليب تقويم تحصيل المتعلمين وأدواته، وكان من أبرز هذه التطورات عدم اقتصار دور تقويم التحصيل على تحديد المكانة النسبية للمتعلم بين أقرانه، وإنما تحديد ما حققه المتعلم من المعارف والمهارات والاتجاهات المتعلقة بالمجالات الدراسية المختلفة، وكذلك عدم الاقتصار على مجرد التحقق من حفظ، ونقل الخبرات السابقة، وإنما التركيز على التحقق من الكفايات التى اتقنها ، الإنجازات التى حققها، وتشخيص مواطن القوة والضعف فى تحصيله، وأصبحت هذه التطورات هدفا أساسيا من أهداف تقويم التحصيل الدراسى للمتعلمين، ومن ثم نشأت فكرة الاختبارات مرجعية المحك Griterion Referenced Tests كاستراتيجية قياس.وتقويم جديدة من أجل تحقيق ذلك.

وتختلف هذه الاستراتيجية اختلافا ملحوظا عن الاستراتيجيات المتعارف عليها، والتى كانت ولا تزال تبنى على أساسها الاختبارات الصفية التقليدية ، والاختبارات مرجعية الجماعة Norm – Referenced Tests والتى وجه إليها العديد من الانتقادات ، التى من أهمها:

- أن التوزيع الاعتدالى Normal Curve الذي يستند إليه في بناء هذه الاختبارات، وانتقاء مفرداتها يؤدى إلى استبعاد المفردات التي يجيب عنها جميع الطلاب ، إما إجابة صحيحة أو إجابة خاطئة أى أنها غير مميزة . حيث يؤدى هذا إلى استبعاد مفردات ذات أهمية في قياس بعض الأهداف المرجو تحقيقها لدى المتعلم. (Popham ;1978 : 84)
- تقيس هذه الاختبارات فى الغالب نواتج تعليمية قد تختلف عن النواتج التعليمية التى يريد المعلم أن يحققها تلاميذه. (صلاح علام ؛ ١٩٨٦ : ١٣-١٥)
- عدم قدرة هذه الاختبارات على تشخيص جوانب القوة، والضعف لدى الطلاب، وتركيزها على الكشف عن الفروق الفردية بينهم فقط، وبذلك لا تسهم فى تحقيق أهداف العملية التعليمية المتمثلة فى إكساب المتعلمين المعارف، والمهارات والاتجاهات المرجوة. (صلاح علام ؛ ١٩٨٤ : ١٥)

وعلى الرغم من أهمية الاختبارات مرجعية المحك إلا إنه لم يوجه إليها الاهتمام الكافي في نظامنا التعليمي ، وما زالت العديد من مؤسساتنا التعليمية تفتقر إلى هذا النوع من الاختبارات حيث جرت العادة على قياس أداء الطلاب باستخدام اختبارات مرجعية الجماعة وفي ظل نظرية القياس الكلاسيكية مما يشكك كثيرا في صدق النتائج وإمكانية تعميمها لتأثر نتيجة القياس بخصائص مجموعة المفردات التي يتكون منها الاختبار وأيضا بخصائص عينة الأفراد التي تتطبق عليها وهذا يؤدى إلى صعوبة مقارنة درجات الأفراد على الاختبارات المختلفة .

ومن هنا برزت مشكلة الدراسة الحالية، والتي تتمثل في استخدام نموذج (راش) فى تكافؤ درجات صور إختبارية مرجعية المحك في مادة (الإحصاء) المقررة على طلاب المرحلة الثانوية العامة ومقارنة أداء الطلاب عليها في ضوء بعض المتغيرات.

هدف الدراسة: تهدف الدراسة الحالية إلى:

- بناء صور إختبارية مرجعية المحك متكافئة القياس في مادة الإحصاء المقررة

على طلاب المرحلة الثانوية العامة في ضوء الكفايات المرجو تحقيقها من جراء تدريس هذه المادة.

- تحليل وانتقاء مفردات هذه الصور الإختبارية باستخدام أحد نماذج الاستجابة للمفردة (نموذج راش) اللوغاريتمي أحادى المعلم لتحقيق استقلالية القياس عن خصائص كلا من الأفراد والمفردات .
- استخدام إحدى طرق تكافؤ درجات الاختبار (طريقة البنود المشتركة) حيث يؤدى ذلك إلى تدريج المفردات تبعا لصعوبتها على متصل بوحدة قياس محددة، ويعد هذا التدريج ميزانا تتحدد على أساسه مستويات التحصيل (القدرة) Ability مقدرة بهذه الوحدة المعرفة، حيث يتم تدريج قدرات الطلاب على الميزان نفسه.
- مقارنة أداء الطلاب على هذه الصور الاختبارية فى ضوء متغيرى الجنس (ذكور / إناث) والتخصص (علمي / أدبي) للتأكد من مدى تحقيق الكفايات والأهداف الخاصة بهذه المادة.

أهمية الدراسة :

الإسهام في إضافة أداة قياس تعتمد في بناءها على موضوعية القياس وخطيته باستخدام أحد نماذج الاستجابة للمفردة نموذج (راش) حيث يمكن من خلال هذه الأداة التحقق من مدى إتقان الطلاب للكفايات الخاصة بمادة (الإحصاء) وتحقيقهم للأهداف المرجوة من تدريسها حيث لم يتوفر ذلك (حسب علم الباحث) لهذه المادة نتيجة لحداثة تدريسها في المرحلة الثانوية العامة.

مصطلحات الدراسة:

١- **الاختبار مرجعى المحك** Griterion – Referenced Test: هو الاختبار الذي يستخدم في تقدير أداء الفرد بالنسبة إلى نطاق سلوكي (أي مجموعة من المعارف، والمهارات المعرفة بطريقة إجرائية) بصرف النظر عن علاقة أدائه بأداء غيره من الأفراد الذين يطبق عليهم الاختبار نفسه. **(Hambleton & Stephen;1997 : 11)**

٢- المحك Criterion: يقصد به مستوى أداء معين يتحدد بمواصفات معينة مصوغة فى صورة أهداف سلوكية بحيث يمكن استخدامه في تحديد مستوى أداء المختبرين أو تصنيفهم بحسب درجة اتقانهم لأهداف محتوى تعليمى معين. (صلاح علام ؛ ١٩٨٦ : ١٧)

٣- موضوعية القياس easurment Objectivity : ويقصد بموضوعية القياس تحرره من تأثير كلا من الأداة المستخدمة (طالما أنها أداة قياس مناسبة) ومن العناصر (الأفراد التي استخدمت هذه الأداة في تقديرهم بالإضافة إلى تدريج هذه الأداء بوحدة قياس مطلقة ثابتة تتوافق مع تدرج مستويات المتغير المقيس (امينة كاظم ، ٢٠٠٠ : ٣٢٧ -٣٢٨)

٤- نماذج الاستجابة للمفردة Item Response Models: هى دوال رياضية احتمالية تهدف لتحديد علاقة بين أداء الفرد فى اختبار معين، وبين السمات أو القدرات التى تكمن وراء هذا الأداء ، وتفسره، وتصميم ميزان لوغاريتمى خطى موحد لكل من صعوبة المفردات ، وقدرات الأفراد.

(صلاح علام ؛ ١٩٨٦ : ١٠٢)

الإطار النظري

أولا - القياس مرجعى المحك -أهميته - استخداماته: نشأت حركة القياس مرجعى المحكCriterion Referenced Measurment مرتبطة بفلسفة تربوية تختلف عن فلسفة القياس مرجعى الجماعة Norm – Referenced Measurment وقد أثرت منهجية القياس مرجعى المحك تأثيرا واضحا فى النظر إلى التحصيل الدراسى من منظور جديد، فأصبح ينظر إليه على أنه تقييم للكفايات. Competence Assessment. أى تقويم القدرة على تنفيذ مهام محددة بنجاح بدلا من اختزان بعض المعلومات، والمعارف المتفرقة ، وفى ضوء ذلك فإن التحصيل يفسر على أنه أداء Performance أو قدرة على أداء شئ. (صلاح علام ؛ ١٩٨٦ : ١٣)

وقد أسهم هذا التوجه فى تطوير تصميمات ، ونماذج تعليمية مبنية على أساس الكفايات Competency – Based Models من أجل التمكن والاتقان ، فى ضوء فلسفة تربوية مؤداها أن كل طالب مهما كانت قدرته يستطيع أن يتعلم تعليما فاعلا ويتقن الأهداف والكفايات المحددة إذا ما اتيح له الوقت الكافى، والأسلوب التعليمى الذى يناسب مستوى قدرته، وهذا يتطلب تنوعا فى أساليب التعليم، وأدوات التقويم، وتقنياتها. (صلاح علام ؛ ١٩٩٥ : ٢٩)

ولعل حركة القياس مرجعى المحك، وما ارتبط بها من تقنيات، وتطبيقات، وما ترتب عليها من منهجية جديدة فى قياس التحصيل، تؤكد التكامل بين عمليتى التعليم، والتقويم ، وهذا ما تسعى إلى تحقيقه الأنظمة التربوية المعاصرة، كما أن الاختبارات مرجعية المحك (CRT) تمثل أسلوبا هاما فى تقويم التحصيل يتناسب مع الأهداف المتطورة للنظم التربوية المعاصرة. حيث أنها تهتم بقياس مدى تحقيق المتعلم لمجموعة متكاملة ، ومحددة من المهام وتزوده بأدلة عن درجة تمكنه منها ، وما يعترضه من صعوبات تعوق هذا التقدم، وتعتمد هذه الاختبارات فى تفسير الدرجات المستمدة من تطبيقها على موازنة أداء الفرد بمحك أداء متوقع يصاغ عادة على صورة كفايات محددة أو نواتج متوقعة أو أهداف سلوكية مرتبة ، بحيث تصف مختلف مستويات الأداء، ولا تستند مرجعية تفسير الدرجة إلى أداء الأقران ، أو معيار جماعة التقنين ، بل إلى الأداء المتوقع أو المرجو تحقيقه ، والذى يحدد تحديدا دقيقا ، وفى هذا الاطار يفترض أن هناك متصلا لاكتساب المعارف، والمهارات يمثل أحد طرفية عدم الكفاءة ، ويمثل الطرف الآخر الكفاءة التامة، ويكون أداء الفرد مناظرا لاحدى طرفى هذا المتصل. (صلاح علام ؛ ١٩٩٨: ١٧)

وعلى نفس المعنى يؤكد انور الشرقاوى واخرون أن القياس الذى يرد إلى محك تحدد نقطة الارتكاز فيه عند طرفيه ، فإن الدرجة التى تكون فى أعلى المقياس تدل على التمكن الكامل أو على أعلى مستوى للأداء الصحيح لبعض القدرات أو المهارات المحددة أم الدرجة التى تكون عند أسفل المقياس فإنها تدل على أدنى

مستوى لهذه القدرات أو المهارات، والتي قد يصل أحيانا إلى درجة الصفر. (أنور الشرقاوى وآخرون ؛ ١٩٩٦ : ٢٧)

ثانيا : نظرية الاستجابة للمفردة: اهتم المشتغلون بميدان القياس النفسي والتربوي بالإفادة من التقدم الهائل فى مجال الحاسب الآلى بمحاولة التوصل إلى وسائل سيكومترية حديثة يمكن من خلالها حل المشكلات الخاصة بأساليب القياس التقليدية ومواجهة أوجه القصور المتعلقة بها، والتي ينتج عنها نتائج مضللة تؤدى إلى تنبؤات غير دقيقة، وما يترتب على ذلك من قرارات غير صائبة.

ومن ثم ظهرت بعض الاتجاهات الحديثة فى القياس، منها نظرية الاستجابة للمفردة (IRT) أو نظرية السمات الكامنة (LTT) وتستند نظرية الاستجابة للمفردة على ركيزتين أساسيتين هما:

١- أداء الفرد على مفردة الاختبار يمكن أن يتنبأ به من خلال مجموعة من العوامل تسمى سمات كامنة أو قدرات .

٢- العلاقة بين أداء الفرد على المفردة ومجموعة السمات الكامنة للأداء على المفردة يمكن أن يوصف من خلال دالة تزايدية مطردة Monatonically, Increasing Function تسمى دالة خاصية المفردة ، وأحيانا تسمى دالة الاستجابة للمفردة وهذه الدالة تصف أنه كلما كانت السمة المقاسة متزايدة فإن احتمال حدوث الاستجابة الصحيحة على المفردة يتزايد تبعا لذلك ومن خلال هذه النظرية الجديدة انبثقت مجموعة من النماذج الرياضية Mathematical Models الهدف منها هو تحديد العلاقة بين أداء الفرد على الاختبار الذى يمكن ملاحظته بصورة مباشرة وبين مجموعة السمات ، أو القدرات التى تكمن وراء هذا الأداء، وتفسره وهذه النماذج الرياضية هى نماذج احتمالية Probabilistic وليست نماذج حتمية Deterministic حيث أن العلاقة التى يحددها كل نموذج تخضع لنظرية الاحتمالات. (Hambleton & Swaminathan, 1985:6)

ولقد تعددت النماذج المتعلقة بهذه النظرية، وتمايزت عن بعضها البعض حيث إن لكل نموذج منها مزاياه وعيوبه، كما تباينت أشكال المنحنيات المميزة للمفردات تبعا لاختلاف عدد بارامترات المفردات، وترتكز نماذج الاستجابة للمفردة على مجموعة من الافتراضات التى يجب أن تتوافر فى البيانات المستمدة من الاختبار، ويعتمد الاختيار الأمثل لأحد تلك النماذج على مدى تحقق هذه الافتراضات فى البيانات لكى يمكن الاستفادة بمميزات النموذج المستخدم، من أهم هذه الافتراضات:

(أ) أحادية البعد Unidimensionality of The Latent Space : تعتمد بعض نماذج السمات الكامنة على افتراض وجود قدرة واحدة فقط يمكن أن تفسر أداء الفرد على اختبار ما وتسمى هذه النماذج بنماذج السمات الكامنة أحادية البعد (صلاح علام: ١٩٨٦ : ١٠٤)

وتتميز بنود الاختبارات التى يتم إعدادها وفقا لهذه النماذج بأنها لا تختلف فيما بينها إلا من حيث مستوى صعوبتها فقط كما أن الأفراد ذوى القدرة الاحادية البعد يتحدد مستوى أدائهم على الاختبار وفقا لهذه القدرة فقط . (أمينه كاظم ؛ ١٩٨٨ : ٤٢).

(ب) الاستقلال المركزى Local Independence : ويعنى أن تقدير صعوبة أى بند لا يعتمد على تقديرات صعوبة البنود الأخرى أى تكون استجابة الفرد للمفردات المختلفة فى الاختبار مستقلة استقلالا إحصائيا (أمينه كاظم : ١٩٨٨ : ١٢)

وهذا يعنى ألا تؤثر استجابة الفرد لإحدى مفردات الاختبار على استجاباته للمفردات الأخرى. أى أن تقدير صعوبة أى مفردة اختبارية لا يعتمد على تقديرات صعوبة المفردات الأخرى المكونة للاختبار، ولا يعتمد على قدرة الأفراد الذين يجيبون عليها، كذلك لا يعتمد تقدير قدرة الأفراد على قدرة أى مجموعة أخرى من الأفراد الذين يطبق عليهم الاختبار ولا يعتمد على قيمة صعوبة المفردات التى يجيبون عليها. (صلاح علام؛ ١٩٨٦: ١٠٤)

(ج) المنحنيات المميزة للمفردة: Item Characteristic Curves: المنحنى

المميز للمفردة هو دالة رياضية تربط بين احتمال نجاح الفرد فى إجابة مفردة اختبارية وبين القدرة التى تقيسها مجموعة المفردات التى يشتمل عليها الاختبار، ولذلك يمكن اعتبار أن المنحنى المميز للمفردة عبارة عن انحدار الدرجة التى يحصل عليها الفرد فى إحدى المفردات على قدرته، ويعد عدم تغير المنحنى المميز للمفردة، وما يتبع ذلك من عدم تغير للبارامترات التى تحدده إحدى المميزات الهامة التى تتميز بها نماذج الاستجابة للمفردة (صلاح علام، ١٩٨٦، ١٠٥، ١٠٦)

(د) التحرر من عامل السرعة Speededness: تفترض نماذج الاستجابة للمفردة أن عامل السرعة لا يلعب دورا فى الإجابة على مفردات الاختبار بمعنى أن اخفاق بعض المختبرين فى إجابة بعض مفردات الإختبار يرجع إلى إنخفاض قدرتهم وليس إلى تأثير عامل السرعة فى إجاباتهم، لأنه إذا ما اعتبرت السرعة عاملا مؤثرا فى أداء الفرد على الاختبار ففى هذه الحالة يكون هناك على الأقل سمتان أو قدرتان تؤثران فى الأداء هما سرعة الأداء، القدرة أو السمة المقاسة من خلال محتوى الاختبار، وهذا لا يتفق مع أحادية البعد التى تمثل افتراضا أساسيا لهذه النماذج (Hambleton, 1985 : 130)

وتوجد العديد من هذه النماذج فى سياق نظرية الاستجابة للمفردة منها النموذج اللوغاريتمى أحادى البارامتر (نموذج راش) One- Parameter Logistic Model (Rasch Model)

والنموذج اللوغاريتمى ذا البارامترين Two- Parameters Logistic Model والذى اقترحه "بيرنبوم" Birnbaum (١٩٥٧، ١٩٥٨، ١٩٦٨) والنموذج اللوغاريتمى ثلاثى البارامتر (نموذج بيرنبوم) Three – Parameter Logistic Model وهو امتداد للنموذج الثنائى البارامتر بعد إجراء بعض العمليات الرياضية عليه، وللتعرف على مزيد من هذه النماذج يمكن الرجوع إلى أدبيات القياس فى هذا الشأن.

ويعرض الباحث فيما يلى بشئ من الايجاز إلى (نموذج راش) اللوغاريتمى أحادى المعلم حيث إنه النموذج الذى ستتبناه الدراسة الحالية.

ثالثا : النموذج اللوغاريتمى أحادى البارامتر (نموذج راش)(Rasch Model)

ارتبط هذا النموذج باسم عالم الرياضيات الدانمركى "جورج راش" G. Rasch والذى نادى بأهمية بناء نظام قياسى موضوعى فى العلوم السلوكية، وكان يهدف إلى تحقيق مفهوم الموضوعية Objectivity بمعنى أن درجة الفرد فى الاختبار لا يجب أن تكون دالة لعينة الأفراد التى استخدمت فى التدريج الأصلى للمفردات Item Calibration التى يشتمل عليها الاختبار كما أنه يجب أن يحصل الفرد على الدرجة نفسها فى كل من اختبارين يقيسان السمة أو القدرة نفسها مهما اختلفت صعوبة مفردات كل منهما، وهذا يعنى أن تكون القياسات مستقلة عن المفردات المستخدمة في القياس كذلك ، وأن تكون أدوات القياس مستقلة عن خصائص عينة تقنينها. (صلاح علام ؛ ١٩٨٦ : ١١٨)

وقد ظل هذا النموذج فى حيز التفكير النظرى الرياضي البحت مدة من الزمن إلى أن بدأ العالم الأمريكي بنجامين رايت Benjamin Wright وزملاؤه بجامعة شيكاغو بعرض أول دراسة عن هذا النموذج عام (١٩٦٧) بعنوان "التدريج الحر لمفردات الاختبار وقياس الأشخاص".وقد أظهر "رايت" في هذه الدراسة كيفية الانتقال من الجانب النظري الرياضي للنموذج إلى الجانب العملي التطبيقي. (محمد الشافعى؛ ١٩٩٦ : ٥٦)

ويستخدم نموذج "راش" Rasch فى تحليل مفردات الاختبارات التى تكون الإجابة عنها أما خطأ أو صواب بمعنى إن درجة كل مفردة هى (صفر ، ١) ، وعندما يستجيب الفرد لمفردة اختبار ما ، فإنه يحدث تفاعل بين قدرة هذا الفرد، وصعوبة هذه المفردة ، ومن خلال هذا التفاعل يتضح الجانب النفسى فى هذا النموذج.

فروض نموذج (راش): يعتمد نموذج "راش" على فرضين أساسيين هما:

١- يزيد احتمال الاجابة الصحيحة عن المفردة الاختبارية بزيادة قدرة الفرد التى تقيسها تلك المفردة.

٢- احتمال أن يجيب فرد إجابة صحيحة عن مفردة سهلة أكبر من احتمال أن يجيب إجابة صحيحة عن مفردة صعبة. **(صلاح علام ؛ ١٩٨٦ : ١٢٤)**

والفرض الأخير يعنى أنه كلما زادت قدرة الفرد بالنسبة إلى صعوبة المفردة ازداد احتمال حدوث الاستجابة الصحيحة للمفردة والعكس صحيح. أما إذا تساوت قدرة الفرد مع صعوبة المفردة ، فإن احتمال حدوث الاستجابة الصحيحة يتساوى مع احتمال حدوث الاستجابة الخطأ. **(أمينه كاظم؛ ١٩٨٨ب : ٤٦)**

وقد يحدث أحيانا أن تزيد قدرة الفرد عن صعوبة المفردة ، ومع هذا يفشل الفرد فى الاجابة عن هذه المفردة السهلة نسبيا، وتكون الاستجابة خاطئة، أو قد يحدث العكس. **(Wright & Ston ; 1979: 12)**

. ويذكر هاملتون Hambleton بعض الفروض الأخرى الخاصة بهذا النموذج مثل افتراض تساوى مفردات الاختبار فى قوة التمييز، أن يكون التخمين أقل ما يمكن، وعلى الرغم من أن اختيار المفردات للنموذج يجب أن يكون على أساس اشتراكها فى بعض الصفات ، والخصائص المتعلقة بالفروض السابقة. إلا أن فرضية تساوى المفردات فى قوة التمييز قد لا تتحقق أحيانا. **(Hambleton 1985 : 46)**

الصورة الرياضية لنموذج (راش): يختلف نموذج (راش) اختلافا جوهريا عن النماذج الأخرى للقياس فى خصائصه السيكومترية، وفى نوعية المفردات التى يستخدم فى تحليلها حيث يستخدم فى تحليل المفردات التى تكون درجتها على الصورة (صفر ، ١). فإذا كانت استجابة فرد (V) قدرته (β) على بند (I) صعوبته (δ_i) استجابة ما (X) دالة للفرق بين قدرة هذا الفرق وصعوبة هذا البند $F(\beta_v - \delta_i)$ ويرمز لهذه الدالة الاحتمالية $P(X_{vi} = X/\beta_v , \delta_i)$ حيث X=١

عندما تكون الإجابة صواب و X= 0 عندما تكون الإجابة خطا وتتمثل هذه الدالة بالمعادلة التالية

$$P(X_{vi} = X/\beta_v , \delta_i) = \frac{EXP\ [X(\beta_v - \delta_i)]}{1 + EXP(\beta_v - \delta_i)}$$

(امينة كاظم ; ١٩٨٨ ب : ٢٥)

مفهوم التكافؤ

يعنى مفهوم التكافؤ بين الاختبارات بأنه العملية الإحصائية التي تحول بها درجات الأفراد على صورة ما للاختبار بحيث تجعلها مكافئة بطريقة ما للدرجات على صورة أخرى للاختبار وذلك حتى يمكن مقارنة تلك الدرجات. (Kolen, M. & Brennan, R. ; 1995 :2)

وهناك نوعان من التكافؤ هما التكافؤ الأفقي Horizental Equating والتكافؤ الرأسى Vertical Equating ويمكن إجراء التكافؤ بنوعيه باستخدام نماذج (IRT) كما انه في كلا النوعين تكون المهمة هي تكافؤ درجات الأفراد على اختبارين أو اكثر ، وعندما تكون الاختبارات مختلفة يمكن عندئذ أن تكون صالحة وملائمة للاستخدام لمختلف الأفراد وبالتالى يمكن معالجة الدرجات المستخلصة منها كما لو كانت مشتقة من اختبار واحد (Holland & Rubin ; 1982 :1)

- شروط تكافؤ اختبارين : يتطلب إجراء عملية التكافؤ بين اختبارين أو اكثر الشروط التالية :

١- أن تقيس الاختبارات نفس المحتويات . أى تضم نفس الخليط من المفردات التي تقيس نفس المحتوى .

٢- تحقيق شرط التساوى . أى يؤدى التكافؤ إلى عدم وجود فرق بالنسبة للفرد إذا أدى أي من الاختبارين .

٣- أن يكون التحويل متماثلا . أي يكون التحويل من درجات الاختبار الأول إلى

درجات الاختبار الثاني هو معكوس التحويل من درجات الاختبار الثاني إلى درجات الاختبار الأول.

٤- أن يكون التحويل مستقلا عن المجتمع الفرعى الذي اجري عليه التكافؤ : بمعنى ألا تختلف الدرجات التي تعتبر متكافئة على الاختبارين باختلاف مجموعة الأفراد التي يجرى عليها التكافؤ

(Wainer, H. et others ; 2000: 144) (Kolen, M. & Brennan, R. ; 1995 :9-12)

- ***طرق إجراء التكافؤ***

أولا:- الطرق التقليدية لتكافؤ اختبارين

١- التكافؤ باستخدام المئينيات Equipercantile Equating حيث تعتبر الدرجات الخام الكلية على اختبارين متكافئة إذا تساوت الرتب المئينية المناظرة لكل منها وذلك عندما يؤدى الاختبارين مجموعة واحدة من الأفراد وتستخدم دالة التكافؤ المئينية إذا تطابق التوزيع التكرارى لدرجات الاختبار الأول والدرجات المحولة للاختبار الثاني في مجتمع ما .

٢- التكافؤ الخطى Linear Equating ويعد حالة خاصة من التكافؤ بتساوى المئينيات ويستخدم عندما تتقارب صعوبة الاختبارات موضع عملية التكافؤ وفي هذه الحالة يوجد تحويل بحيث تعتبر الدرجات على اختبارين متكافئة إذا ابتعدت عن المتوسط بنفس القدر من وحدات الانحراف المعيارى

(Hambleton, et, Others;1991 :123-125)

إلا أن هذه الطرق تعانى من بعض اوجه القصور ،منها إعتمادها على الدرجات الخام وهذا الاتجاه غير مرغوب فيه لاسباب تتعلق بضرورة تساوي عدد أسئلة كل اختبار وأيضا تماثلها وتقارب مجموعات الأفراد التي ينطبق عليها الاختبار بالإضافة إلى أن بعض من هذه الطرق التقليدية يحتاج عينات كبيرة نسبيا من الأفراد وبعضها الآخر يقوم على افتراضات يصعب تحقيقها في الواقع العملى . كما أن التكافؤ بتساوى المئينيات لا يحقق شرط الاستقلالية عن مجموعة الأفراد المستخدمة **(Hambleton& Swaminathan; 1985 :201)**

ثانيا : التكافؤ من خلال نظرية الاستجابة للمفردة (IRT) : إن إجراء تكافؤ الاختبارات من خلال نظرية (IRT) يقضى على العديد من المشكلات السابقة حيث إنه إذا كان هناك ملائمة بين النموذج المختار ومجموعة بيانات الاختبار ذات الاهتمام فإنه يمكن الحصول على المزايا التالية:

١- تقدير متغيرات المفردة (مستوى الصعوبة – مستوى التمييز ...) تكون مستقلة عن مجموعة الممتحنين المسحوبة من المجتمع الأصلى للممتحنين والتى صمم الاختبار بناء عليها.

٢- تقدير متغيرات قدرة الممتحنين تكون مستقلة عن مجموعة مفردات الاختبار الخاصة المختارة، والتى تم استخدامها من المجموعة الكلية للمفردات ، والتى تم تدريجها.

٣- دقة وإحكام تقديرات القدرة. (Hambleton , 1985 : 11)

ويمكن أن تتم عملية التكافؤ بين اختبارين أو اكثر من خلال نظرية الاستجابة للمفردة بعدة طرق منها :

(أ) طريقة الأفراد المشتركين Common Person Equating : Designوتستخدم فى تكافؤ اختبارين عن طريق تطبيق كل من هذين الاختبارين على مجموعتين مختلفتين من الأفراد بينهما مجموعة من الأفراد المشتركين يتلقون نفس الاختبارين، وبالتالي فإن مستويات الصعوبة للاختبارين لا تكون متداخلة مع مستويات القدرة للأفراد.

(ب) طريقة البنود المشتركة Common Items Equating Desing: وتستخدم هذه الطريقة إذا كان المطلوب هو دمج مفردات اختبارين يمثلان مستويين مختلفين فى تدريج واحد مشترك حيث يمكن استخدام مجموعة من المفردات المشتركة بين كل من الاختبارين بحيث تشكل رابطا بينهما وتضاف هذه المفردات إلى كل من الاختبارين، والمدرجين باستخدام نموذج(راش)، ووفقا لهذا النموذج يجب أن تتكافئي تقديرات الصعوبة لهذه المفردات المشتركة، والتي اشتقت من تطبيق كل من الاختبارين على مجموعتين مختلفتين.

(ج) طريقة المجموعة المتكافئة Equivalent Group Design: وفى هذه الطريقة، فإن الاختبارين المطلوب تكافؤهما (فى تدريج واحد) يمكن تطبيقهما على مجموعتين متكافئتين وليس من الضرورى أن يكونا متطابقين بصورة تامة.

وسوف يستخدم الباحث في تكافؤ درجات الصور الإختبارية مرجعية المحك طريقة (البنود المشتركة) من خلال نظرية الاستجابة للمفردة وباستخدام نموذج (راش) اللوغاريتمي.

طرق التحليل المستخدمة في إجراء التكافؤ :

يوجد عدة طرق لإجراء عمليات التكافؤ منها ما يعتمد على العمليات الحسابية اليدوية ومنها ما يقوم به الحاسب الآلي وفيما يلي طريقتين لإجراء عملية التكافؤ سواء عن طريق الأفراد المشتركين أو البنود المشتركة.

- **الطريقة الأولى** : تحليل بنوك الأسئلة Item bank anchoring method : وتتم عن طريق تحليل إحدى الصور الاختبارية ثم تعطى المتغيرات الناتجة من التحليل الأول للبرنامج عند تحليل باقي الصور الأخرى .
- **الطريقة الثانية** : التحليل التلازمى Concurent Scaling حيث يتم التكافؤ بإدخال الاستجابات القديمة والجديدة للإفراد في آن واحد ويقوم البرنامج بإجراء التعادل أثناء تقدير المتغيرات وسوف يستخدم الباحث الطريقة الأولى في إجراء عملية التكافؤ .

الدراسات السابقة

أولا: دراسات تناولت المقارنة بين النظرية التقليدية ونظرية الاستجابة للمفردة فى انتقاء وتحليل مفردات الاختبار

١- **دراسة بيركنز و ميللر** Perkins & Miller 1984: وهدفت إلى مقارنة التحليلات المرتبطة بالنظرية التقليدية للاختبار ونظرية الاستجابة للمفردة، وذلك فى اختبار فهم القراءة فى اللغة الإنجليزية على عينة مكونة من (٨٨ طالبا) كان من نتائجها أن نموذج (راش) ذا البارامتر الواحد يكشف عن

المفردات الصعبة وغير الملائمة Misfiting بصورة أكثر وضوحا من نظرية الاختبارات التقليدية وأنه يسمح للباحث بأن يحدد السمة الكامنة التى تظهر فى سلوك الممتحن الملاحظ. كما أوضحت الدراسة أن الدالة اللوغاريتمية التى يتم تدريجها وفقا لهذا النموذج، والتى تؤدى إلى صعوبة المفردات تكون متحررة من تأثيرات متوسط، وتباين العينة.

٢- دراسة هاملتون ، روجرز (Hembleton & Rogers (1988):وهدفت إلى حل مشكلات القياس المرجع إلى المحك باستخدام نماذج الاستجابة للمفردة وقد أجريت الدراسة على ثلاثة عينات مختلفة الحجم (٧٩ ، ١٧٨ ، ٣٠٧) من خلال برنامج المحاكاه للحاسب الآلى. كما تم استخدام أربع طرق لانتقاء مفردات الاختبار هى العشوائية Random ، الطريقة الأمثل Optimal ، الطريقة الأمثل فى ضوء المحتوى Content – Optimal Method والطريقة التقليدية Classical وخلصت النتائج إلى أن نماذج الاستجابة للمفردة (IRT) تمدنا بأساس جيد لانتقاء المفردات، والأشخاص ، ودرجات القطع على مقياس واحد مشترك وهذا فى حد ذاته من الأهمية لبيان مستوى الدقة عند أى مستوى للقياس.

٣- دراسة أمينة كاظم ١٩٨٨ أو هدفت إلى بناء اختبار تحصيل دراسي موضوعي في مادة علم النفس لطلاب كلية الآداب جامعة (الكويت) وذلك للتحقق من مدى موضوعية التفسير لنتائج الاختبار وذلك على عينة قوامها (٤١٨) طالب وطالبة واستخدم نموذج (راش) Rasch فى تحليل النتائج، وقد انتهت الدراسة إلى إعداد ، وبناء اختبار موضوعى تتوافر فيه متطلبات الموضوعية فى القياس، وهى تحرر القياس من قدرة العينة التى تؤدى الاختبار، تحرر القياس من صعوبة المفردة المستخدمة فى القياس.

٤- دراسة صلاح الدين محمود علام (١٩٩٠) :وهدفت إلى استخدام نموذج (راش) فى بناء مقياس هدفى المرجع لقياس مدى إتقان طلاب الدراسات العليا بكليات التربية للمعارف الخاصة بتصميم خطة البحوث النفسية

والتربوية، وتصميم مقياس لوغاريتمي خطى يمكن من خلاله تحديد مستويات الإتقان التي يمكن توقعها من الطالب فى كل هدف من الأهداف السلوكية التى أعدت مفردات الاختبار لقياسها، وقد طبق الاختبار على عينة (٤٨) طالب واستخدم طريقة Prox فى تحليل مصفوفة البيانات الإختبارية للحصول على تقديرات صعوبة المفردات، وقدرات الطلاب على متصل متغير المعارف الأساسية بوحدات اللوجيت (Logits) وبتقييم مطابقة القيم التقديرية للنموذج استخدمت طريقة (Calfit)، وقد انتهت الدراسة إلى إعداد المقياس الذى تكون من (٣٢) مفردة بعد حذف المفردات التى لم تطابق النموذج.

٥- دراسة (ساير) Sayre 1990: وهدفت إلى تطوير الحد الأدنى للكفاءة من خلال مقارنة الطرق التقليدية بطرق نظرية الاستجابة للمفردة حيث اهتمت الدراسة ببناء اختبارات باستخدام الطريقتين بهدف مقارنة كل من ثبات قيم دوال معلومات الاختبار عند درجات القطع والصدق لهذه الاختبارات وأجريت على عينة (١٩١٥) طالبا وقد توصلت الدراسة إلى العديد من النتائج أهمها أن المفردات الناتجة من خلال نظرية الاستجابة للمفردة لها ثبات وصدق أفضل من تلك التى تنتج من خلال طرق تحليل المفردة بالطرق التقليدية، وأن نظرية الاستجابة للمفردة تعتبر أداة جيدة لبناء اختبارات ذات كفاءة عالية تستطيع أن تميز بين الأفراد عند درجات القطع المحددة.

٦- دراسة شحته عبد المولى ١٩٩٩ : وهدفت إلى تقويم بناء الاختبارات المرجعة إلى (المحك) /(المعيار) فى ضوء نظرية الاستجابة للمفردة والنظرية التقليدية) وذلك عن طريق التحقق من صدق وثبات هذه الاختبارات فى ضوء النظرية المستخدمة، وقد استخدم الباحث أساليب المحاكاة بالكمبيوتر فى توليد المفردات الاختبارية، وكذلك عينات التطبيق واستخدمت الدراسة الطريقة التقليدية فى مقابل ثلاث طرق تنتمى إلى نظرية الاستجابة للمفردة تعتمد على النماذج (الأحادى – الثنائى – الثلاثى) البارامتر ، وقد

أوضحت النتائج أفضلية نماذج الاستجابة للمفردة عن الطريقة التقليدية فى بناء الاختبارات بنوعيها، وكذلك إلى دقة أفضل فى تقدير القدرة وكذلك أقل مستوى من الخطأ المعيارى فى تقديرها كما أدت نماذج الاستجابة للمفردة إلى درجة أعلى من الثبات ودقة القياس.

٧- **دراسة هارفى وها مير** (Harrvey & Hammer (1999) :وهدفت إلى المقارنة بين نظرية الاستجابة للمفردة، والنظرية الكلاسيكية فى بناء الاختبارات، وقد اعتمدت الدراسة على ثلاث نماذج هى (الأحادى – الثنائى – الثلاثى) البارامتر – لتقدير بارامترات المفردة والقدرة، وخلصت النتائج إلى أن نظرية الاستجابة للمفردة تقدم وصفا تفصيليا عن مستوى صعوبة المفردة، والاختبار أفضل من الطريقة الكلاسيكية. كما أن تقدير الدرجة كان أكثر دقة وإحكاما فى حالة نماذج الاستجابة للمفردة، وكذلك تقديرات بارامترات كل من المفردة والأفراد، كما أشارت النتائج إلى أن نموذج (راش) كان الأفضل مطابقة، والأكثر دقة فى تقدير كل من القدرة ، الصعوبة عنها فى (نموذج لورد) وأن نماذج الاستجابة للمفردة حققت النتائج الأفضل فى حالة المجموعات الأكثر اتساعا، والأكبر عددا.

٨- **دراسة اسماعيل حسن الوليلى ٢٠٠٢** :وهدفت إلى استخدام ثلاثة نماذج من نماذج الاستجابة للمفردة هى النموذج (الأحادى البارامتر (نموذج راش) ، الثنائى البارامتر (نموذج لورد) ، الثلاثى البارامتر (نموذج بيرنيوم) فى تصميم وانتقاء مفردات اختبار مرجعى المحك فى مادة الرياضيات فرع الجبر لطلاب الصف الثالث الاعدادى، وقد طبقت الدراسة على عينة قوامها (١٨٠٠) طالبا وطالبة، وقد توصلت الدراسة إلى العديد من النتائج أهمها وجود فروق دالة إحصائيا بين نماذج الاستجابة للمفردة الثلاث فى انتقاء مفردات الاختبار مرجعية المحك سواء فى طول الاختبار، أو فى القوة التمييزية لمفردات الاختبار المنتقاه، أو فى قيم دوال المعلومات وفى قيم ثبات

درجات الاختبارات وذلك لصالح نموذج (بيرنيوم) كما وجد تقارب بين النماذج الثلاثة فى تقديرات قدرة الأفراد.

ثانيا: دراسات تناولت طرق تكافؤ درجات الاختبار

١- دراسة اسكاجز ، ليستز Skaggs & Lissitz 1985 : هدفت هذه الدراسة إلى اختبار مدى استجابة أربعة إجراءات لتكافؤ إجراءات الاختبار ، الطريقة الخطية ، تساوى النسب المئوية ، نموذج راش ، المعالم الثلاثة للمواقف التى يختلف فيها خصائص وصفات الاختبارين التى يجرى تكافؤهما ، وقد استخدم نموذج المعالم الثلاثة لإعداد إختباري الدراسة واللذين يوجد بينهما فرق في الصعوبة والتمييز،وفي كل حالة للتكافؤ تم الحصول على مجموعتين من البيانات المكونة من استجابات (٤٠٠٠) طالب على ٨٥ مفردة اختبارية ، ٣٥ مفردة لكل اختبار ، بالإضافة إلى (١٥) مفردة مشتركة بين الاختبارين وقد أوضحت النتائج أن طريقة تساوى النسب المئوية كانت ثابتة عبر جميع حالات الدراسة ، أما الطريقة الخطية، نموذج (راش) كان لهما حساسية زائدة لانتهاك فروض النماذج الخاصة بها.

٢- دراسة هيلس Hells 1987: وهدفت إلى مقارنة بعض طرق تكافؤ النهاية الصغرى لاختبارات الكفاءة ، وقد استخدمت درجات اختبار STATEWIDE (١٩٨٦) للكفاءة الصغرى المتطلب للصفوف العليا في ولاية فلوريد ، وتم وضع هذه الدرجات على مقياس درجات عام ١٩٨٤ لنفس الاختبار باستخدام خمسة إجراءات مختلفة للتكافؤ هي الطريقة الخطية، نموذج راش ، نظرية (IRT) ذات الثلاثة معالم (الطريقة السائدة) ، وذات الثلاثة معالم (طريقة المتغير الثابت) ، ونظرية (IRT) طريقة الصيغة. وذلك على عينات من الطلاب تتراوح ما بين (٨٠٠)، (١٥٠٠)فرداً وأجريت مقارنات لنتائج طرق التكافؤ المختلفة ، وأيضا مقارنة النتائج المستخلصة من ستة أطوال مختلفة من البنود المشتركة وتوصلت الدراسة

إلى عدة نتائج أهمها : أن طرق التكافؤ المختلفة أدت إلى نتائج متشابهة وان هذه الطرق تكون كافية في حالة الاختبارات المتكافئة في الصعوبة والمحتوى ، كما أن البنود المشتركة التي عددها (١٠) بنود كان لها نفس تأثير البنود التي عددها (٣٠) وذلك باستخدام طريقة (IRT) ذات المعالم الثلاثة (الطريقة السائدة).

٣- **دراسة هاريس ، كولين Harris & Kolen 1990**:.وهدفت إلى مقارنة نتائج طريقتين من طرق تكافؤ درجات اختبارين لتحديد مدى التشابه بين نتائج هاتين الطريقتين وهما طريقتا التقدير التكراري ، البنود المشتركة. واستخدم في هذه الدراسة ثلاثة اختبارات ، اختبار (X) صورة حديثة واختبار (Y) صورة قديمة واختبار (V) مجموعة من البنود المشتركة. وطبق الاختبار X على عينة(٣٩١٤٩)فردا ، والاختبار (Y) على عينة(٣٩٧٩٠) فردا ، وتلقت العينتان اختبار (V) وأشارت النتائج إلى وجود تشابه كبير بين نتائج هاتين الطريقتان كما أشارت إلى وجود بعض المشكلات المتعلقة ببعض متغيرات البنود المشتركة مثل العدد المناسب لها، مستوى صعوبتها.

٤- **دراسة اسنيكاس Snieckus 1992** : وهدفت إلى التعرف على إمكانية استخدام إجراءات التكافؤ أحادى البعد على اختبار ثنائي البعد والتعرف على أخطاء التكافؤ الناتجة عن استخدام ثلاث طرق للتكافؤ وهى الطريقة الخطية، تساوى النسب المئوية ، طريقة (IRT) وتم تطبيق الطرق الثلاث على خمس صور من الاختبارات متعددة البعد وقّدر خطأ التكافؤ بواسطة إعادة التطبيق للصور الأساسية والصور النهائية وتم تطبيق الاختبار الأساسي على عينة عشوائية (٥٠٠) فرد ، تطبيق صور الاختبار على مجموعات أخرى مكافئة لها . وأظهرت النتائج أن إجراء التكافؤ أحادى البعد للاختبار ثنائي البعد يعطى نتائج جيدة برغم انتهاك فردية أحادية البعد وذلك عندما تكون التوزيعات الخاصة بالقدرة غير مختلفة من خلال

مجموعة التطبيقات لصور الاختبار ، كما أظهرت النتائج أيضا أن طرق إجراء التكافؤ الثلاث أدت إلى نتائج متشابه .

٥- دراسة محمد منصور الشافعى ١٩٩٦ :وهدفت إلى التعرف على أهم المتغيرات المطلوبة في إحكام تدريج مفردات بنك الأسئلة باستخدام نموذج (راش) ومن بين هذه المتغيرات. معادلة درجات الاختبار، عدد البنود المشتركة مستوى صعوبة البنود المشتركة، تمثيل البنود المشتركة للمتغير المقيس ، عدد مفردات اختباري الضم ، خصائص العينة المستخدمة، وأجريت الدراسة على (١٤٠٥) من طلاب الصف الثاني الثانوي حيث طبق عليهم اختبار فى مادة الرياضيات وقد أظهرت نتائج الدراسة أفضلية طريقة الأفراد المشتركين وكذلك طريقة البنود المشتركة عن طريقة المجموعة المتكافئة فى كفاءة ضم مفردات اختبارين في تدريج واحد مشترك كما بينت النتائج عدم تأثر إجراء ضم مفردات اختبارين (سهل / صعب) أو متكافئين، بحجم عينة التحليل عند استخدام طريقة البنود المشتركة أو طريقة الأفراد المشتركين.

٦- دراسة شادية عبد العزيز ١٩٩٦: وهدفت الى تدريج بنك للأسئلة الموضوعية لأحد مقررات علم النفس التعليمى باستخدام نموذج(راش) والتوصل إلى موضوعية تقدير وتقويم الطلاب وذلك من خلال تطبيق (١٧) صورة إختبارية تتضمن(٣٤٤) مفردة من نوع الاختيار من متعدد على عينة قوامها (٣٢٣٦) طالبة من الفرقة الرابعة والفرقة الثانية بالدبلومة العامة بكلية البنات – جامعة عين شمس وأشارت النتائج إلى أن الصور الاختبارية المختلفة تصلح لتدريج بنك الأسئلة باستخدام نموذج راش وان هذا البنك يحقق شرط استقلالية القياس.

٧- دراسة اعتدال غازى عبد الله ٢٠٠٠: وهدفت إلى بناء صور اختبارية محكية المرجع متكافئة القياس في مقرر علم النفس التعليمى لقياس مدى

تحقق الأهداف المعرفية المتعلقة بهذا المقرر وذلك من خلال تطبيق صورتى اختبار على مجموعتين متساويتين من أفراد عينة الدراسة البالغ عددها (١٩٤) طالبة من طالبات الفرقة الرابعة بكلية البنات – جامعة عين شمس وأشارت النتائج إلي انه تم تكوين صور اختبارية مرجعة إلى محك، متكافئة القياس في مقرر علم النفس التعليمي ، كما تم تحديد النسب المئوية لعدد الأفراد الذين يقعون في كل مستوى من مستويات الإتقان ، وتم تشخيص نقاط القوة والضعف في تحقيق أهداف القياس .

التعليق على نتائج الدراسات السابقة

- تميزت عينات الدراسات السابقة بكبر حجمها حتى وصلت فى بعض الدراسات إلى عدد (٣٩١٤٩) فردا كما فى دراسة Harris & Kolen 1990 وكان أقل عدد (٤٨) فى دراسة صلاح علام (١٩٩٠).
- أشارت نتائج جميع الدراسات على أن نظرية الاستجابة للمفردة (IRT) تعتبر أكثر ملائمة وفعالية من النظرية التقليدية فى الكشف عن المفردات الصعبة وغير الملائمة. كما أكدت دراسة Sayre 1990 ودراسة Harvey & Hammir (1990) على أن المفردات الناتجة من خلال نظرية (IRT) لها ثبات وصدق أفضل من تلك التى تنتج من خلال طرق تحليل المفردة بالطريقة التقليدية .
- أكدت نتائج غالبية الدراسات على أن نموذج (راش) Rasch اللوغاريتمي يعتبر الأفضل فى الكشف عن المفردات الصعبة وغير الملائمة Perkins & Miller (1984)، Hambleton & Rogers (1988) أمينة كاظم (١٩٨٨ أ) صلاح علام (١٩٩٠)، Sayre 1990 ، شحته عبد المولى (١٩٩٩) اعتدال غازى (٢٠٠٠)إلا أن دراسة إسماعيل الوليلى (٢٠٠٢) اختلفت نتائجها عن باقى الدراسات حيث رأت أن نموذج (بيرنبوم) Birnbaum ثلاثى البارامتر يعتبر الأفضل فى انتقاء مفردات الاختبارات مرجعية المحك.
- تعددت الطرق المستخدمة فى تكافؤ درجات الاختبار حيث استخدمت العديد

منها، طريقة تساوى النسب المئوية، طريقة البنود المشتركة، طريقة الأفراد المشتركين، الطريقة الخطية، نماذج (IRT) وهكذا ومن ثم تباينت نتائج تلك الدراسات تبعا لذلك فنجد أن دراسة Snieckus (1992) شادية عبد العزيز(١٩٩٦)، اعتدال غازى (٢٠٠٠)، قد أشارت نتائجها إلى فعالية نظرية الاستجابة للمفردة (IRT) نجد أن دراسة Skaggs & Lissitz 1985 تشير إلى فعالية طريقة النسب المئوية باعتبارها أكثر ثباتا من الطرق الأخرى، ونجد دراسات أخرى لا تفرق بين طريقة وأخرى كما فى دراسة محمد الشافعى (١٩٩٦) والتى لم تجد فروقا فى استخدام أى من طريقتي البنود المشتركة أو الأفراد المشتركين في إجراء عملية التكافؤ، ودراسة Hills (1987) والتى لم تجد فروقا بين جميع الطرق المستخدمة فى التكافؤ.

ويرى الباحث أن ذلك قد يعود إلى عوامل كثيرة منها عدد المفردات المشتركة وخصائصها ومستويات صعوبتها وأيضا إلى قدرات الأفراد المشتركين وعدد أفراد العينة المستخدمة

تساؤلات الدراسة :

س١: ما المواصفات الإحصائية للاختبار التحصيلى مرجعي المحك في مادة الإحصاء المقررة على طلاب المرحلة الثانوية العامة ؟

س٢: كيف يمكن استخدام نموذج راش في تكافؤ درجات صور اختبارية مرجعية المحك في مادة الإحصاء المقررة على طلاب المرحلة الثانوية العامة ؟

س٣: هل يختلف أداء طلاب المرحلة الثانوية العامة في مادة الإحصاء باختلاف متغيرى الجنس (ذكور / إناث)، التخصص (علمي / أدبي) ؟ وما هو حجم التأثير الناتج عن كلا المتغيرين والتفاعل بينهما على هذا الأداء ؟

س٤ : هل توجد فروق دالة إحصائيا بين النسب المئوية لتحقيق المهارات المتضمنة مادة الإحصاء نتيجة لاختلاف متغيرى الدراسة الجنس (ذكور/ إناث) ، التخصص (علمي / أدبي) ؟

عينة الدراسة :

اشتقت عينة الدراسة الحالية من طلاب المرحلة الثانوية العامة بالمدارس

الحكومية بمحافظة الإسكندرية وبالتحديد من إدارات الجمرك، غرب، وسط، شرق تعليمية بواقع مدرستين من كل إدارة إحداهما للبنين والأخرى للبنات وقد بلغ جملة عدد أفراد العينة (١٣٤٧) طالبا وطالبة [وذلك بعد إجراء عملية حذف الطلاب غير الملائمين لعملية التحليل]. والجدول رقم (١) يوضح التوزيع النهائي لأفراد العينة.

جدول رقم (١)

التوزيع النهائي لأفراد عينة الدراسة تبعا لمتغيرى

(الجنس/ التخصص) على الصور الاختبارية الخمس

الجنس ، التخصص / الصورة الاختبارية	ذكور			إناث			المجموع
	علمى	أدبى	الجملة	علمى	أدبى	الجملة	
الصورة الأولى	٧٦	٥٥	١٣١	٦٢	٦٧	١٢٩	٢٦٠
الصورة الثانية	٦١	٥٩	١٢٠	٨٠	٦٧	١٤٧	٢٦٧
الصورة الثالثة	٨٣	٦٥	١٤٨	٦٣	٤٩	١١٢	٢٦٠
الصورة الرابعة	٧٦	٥٩	١٣٥	٧٧	٦١	١٣٨	٢٧٣
الصورة الخامسة	٥٩	٧٦	١٣٥	٨٦	٦٦	١٥٢	٢٨٧
المجموع	٣٥٥	٣١٤	٦٦٩	٣٦٨	٣١٠	٦٧٨	١٣٤٧

الإجابة عن تساؤلات الدراسة

السؤال الأول: ما المواصفات الإحصائية للاختبار التحصيلى مرجعي المحك في مادة الإحصاء المقررة على طلاب المرحلة الثانوية العامة ؟

وستتم الإجابة عن هذا السؤال من خلال إعداد هذا الاختبار والذي مرت خطوات إعداده بالمراحل التالية حيث تم تحكيم كل خطوة من خطواته بواسطة عدد خمسة من المحكمين المتخصصين من أساتذة طريق تدريس الرياضيات(*) بالجامعات ومراكز البحوث المصرية .

(*) (١) أ.د/ وليم تادرس عبيد . تربية عين شمس (٢) أ.د/ إسماعيل وهبة. المركز القومي للامتحانات
(٣) أ.د/ مصطفي حاتم. المركز القومي للامتحانات (٤) د./ حسن خليل. تربية إسكندرية
(٥) د./ محمد محمد حمادة. تربية حلوان

أ- *تحليل المحتوى:* تم تحليل محتوى مقرر مادة (الإحصاء) إلى المفاهيم والعلاقات والمهارات التي يحتوى عليها حيث بلغت عدد وحدات التحليل (١٤٨) وحدة تضمنت (٣٥) مفهوما (٥١) علاقة ، (٦٢) مهارة. وبعد عرض وحدات التحليل على مجموعة المحكمين تم حذف وحدة التحليل التي لم تحقق نسبة اتفاق (٨٠%) فيما بينهم وجدول رقم (٦) ملحق (١) يوضح بيان وحدات التحليل.

ب- *حساب ثبات تحليل المحتوى:* استخدم الباحث طريقة "أولى" Ole- Method حيث تتطلب الاستعانة بعدد ثلاثة من الباحثين ذوى خبرة في مجال تحليل المحتوى إضافة إلى الباحث الأصلي وباستخدام معادلة معامل ثبات تحليل المحتوى =

$$\frac{٤م}{ن١ + ن٢ + ن٣ + ن٤}$$

حيث : م: عدد الوحدات التى اتفق عليها فـي التحليـل ، ن١ ، ن٢، ن٣ ، ن٤ عدد الوحدات التى حددها الباحث الأصلي إضافة إلى الباحثين الثلاثة على الترتيب. (محمد عباس المغربى ؛ ١٩٩٤ : ١٤٧)

وقد بلغ معامل الثبات بهذه الطريقة (٠,٩٨) وجدول رقم (٢) يوضح ذلك.

جدول رقم (٢)

يوضح معامل ثبات تحليل محتوى مادة الإحصاء بين الباحث الأصلي وعدد ثلاثة من الباحثين

وحدات التحليل / الباحث	المفاهيم	العلاقات	المهارات	عدد وحدات التحليل	معامل الثبات
الأصلي	٣٥	٥١	٦٢	١٤٨	٠,٩٨
الأولى	٣٩	٥٤	٦٣	١٥٦	
الثاني	٣٦	٥١	٦٣	١٥٠	
الثالث	٣٥	٥٢	٦١	١٤٨	

جـ- تحديد المهارات الرئيسة والفرعية : بعد إجراء عملية تحليل المحتوى وفـي ضوء الأهداف العامة لمادة الإحصاء المقررة على طـلاب المرحلـة الثانويـة

العامة: توصل الباحث إلى عدد (٩) مهارات رئيسة ينبغي على الطالب في هذه المرحلة أن يتقنها بعد دراسته لهذا المحتوى. كما تم تحليل كل مهارة من هذه المهارات التسع إلى أنماط مترابطة من المهارات الفرعية التي تتدرج تحتها في شكل هرمي حسب أولوية اسهام كل منها في تحقيق المهارة وذلك باستخدام طريقة تحليل المهام Task Analysis التي اقترحها جاجنيه Gagne 1968 و تم التوصل إلى (٣٧) مهارة فرعية. وبعد عرض المهارات الرئيسة والفرعية على عينة المحكمين واستبعاد أي مهارة لا تحقق نسبة اتفاق (٨٠%) فيما بينهم تم التوصل إلى عدد (٧) مهارات رئيسة تتدرج تحتها (٣٢) مهارة فرعية تعدادها على التوالي (٥، ٣، ٤، ٦، ٧، ٤، ٣). وهذه المهارات السبع الرئيسة هي:

١) يحدد العلاقات بين العمليات على الأحداث (المؤكدة/ المتنافية) ويستخدمها في حل التطبيقات.

٢) يطبق قواعد ومسلمات الاحتمال في حل تمرينات ومشكلات حياتية.

٣) يستخدم قواعد حساب التوزيعات الإحتمالية للمتغيرات العشوائية المتقطعة والمتصلة في حل التطبيقات الإحصائية.

٤) يفسر النتائج التي يحصل عليها من حل تمرينات على الوسط الحسابي ومقاييس التشتت لمتغير عشوائي متقطع.

٥) يحل مشكلات حياتية على التوزيع الطبيعي من خلال تحويله إلى توزيع طبيعي معياري مستخدما جداول المساحات.

٦) يحسب ويفسر معاملات الارتباط المختلفة التي يحصل عليها من حل التمرينات المتنوعة ويبين دلالتها.

٧) يستنتج العلاقة بين معامل الارتباط بين متغيرين ومعاملي انحدار كل منهما على الآخر ويوظف هذه العلاقة في حل التطبيقات الإحصائية.

وشكل (١) ملحق (٢) يوضح المهارات الفرعية التي تتدرج تحت كل مهارة رئيسة.

د- **الأهداف الإجرائية** : صيغت الأهداف الإجرائية في ضوء نواتج التحليل السلوكى للمهارات السابقة إستنادا إلى تقسيم "بلوم" Bloom للأهداف المعرفية.حيث مثلت هذه الأهداف اربع مستويات معرفية هى :(التذكر، الفهم ، التطبيق ، حل المشكلات) وبعد عرضها على عينة المحكمين لبيان مدى ملائمة كل هدف للمهارة التى يقيسها وبعد حذف الأهداف التي لم تحقق نسبة (٨٠%) من اتفاق المحكمين عليها توصل الباحث إلى عدد (٦٥) هدفا إجرائيا وزعت على المهارات السبع الرئيسة كما يلى: ١٩ ، ٣، ٧، ٨ ، ١٧ ، ٨ ، ٣ على الترتيب.

مواصفات الاختبار Test Specifications :-اعد الباحث مواصفات الاختبار في ضوء طريقة بابام (1980) Popham في تصميم مواصفات الاختبارات المرجعة إلى محك والهدف من هذه المواصفات هو تحديد النطاق السلوكى الذي يقيسه الاختبار تحديداً اكثر تفصيلا مما تحدده عبارات الأهداف الإجرائية التي تمت صياغتها سابقا . وتتكون مواصفات الاختبار من ثلاثة عناصر أساسية بالإضافة إلى ملحق المواصفات وهذه العناصر هي :

١- **عبارات الأهداف** : وهى وصف مختصر للسلوك الذي تقيسه مفردات الاختبار .

٢- **عناصر المثيرات** : وهى تتكون من مجموعة من العبارات التي نحاول عن طريقها تحديد ووصف المثيرات التي تقدمها مفردات الاختبار .

٣- **عناصر الاستجابات** : وهى تتكون من مجموعة من العبارات التي نحاول عن طريقها تحديد قواعد تكوين المشتتات والإجابة الصحيحة لكل مفردة اختبارية.

(صلاح علام ، ١٩٨٥ : ١٢١)

هـ- **بناء المفردات** : قام الباحث ببناء (١٢٦) مفردة اختبارية من نوع الاختيار من متعدد لقياس الأهداف الإجرائية مسترشدا بمواصفات الاختبار التي تم

إعداده . وقد ارتبط مستوى صعوبة كل مفردة بمستوى صعوبة الهدف الإجرائي الذي تقيسه ومستواه المعرفي .

- **صدق المفردات:** عرضت المفردات الاختبارية على عينة المحكمين لحساب مدى إتفاقهم على مناسبة كل مفردة للهدف الذى وضعت لقياسه وباستخدام معادلة مؤشر الاتساق التى اقترحها كل من روفينلى، هاملتون (1976) Rovinelli & Hambleton والتي على الصورة.

$$\text{ت ق} = \frac{(\text{م}-١)\ \text{ر هـ} - \text{ر هـ}}{٢\ \text{ن}\ (\text{م} - ١)}$$

حيث (م)عدد الأهداف ، (رهـ) مجموع رتب الهدف (هـ) ، (رّ هـ) مجموع رتب جميع الأهداف عدا الهدف (هـ) ، (ن) عدد المحكمين. وطبقا لهذه الطريقة تم حذف (٢٦) مفردة اختبارية ممن تقل معاملات اتساقها عن (+٠,٧٨) حيث تراوحت قيم معامل الاتساق بين كل مفردة والهدف الذى تقيسه ما بين (+٠,٧٨) ، (+١) وتعتبر قيم هذا المعامل هي المحك الذي يتم في ضوءه الحكم على صدق محتوى المفردة وجدول رقم (٧) ملحق (٣) يوضح مواصفات الاختبار.

صدق الاختبار: اعتمد الباحث على الصدق الوصفى Descriptive Validity للمحتوي حيث يعد مؤشرا جيدا على صدق الاختبار.

ثبات الاختبار: تم الاعتماد على مدخل القرارات Decision – Theoretic Approach متمثلا فى معامل الاتفاق لسابكوفياك ^{s} - Coefficients Subkoviak 'Agrement(1980) وهى من الطرق الاكثر ملائمة للاختبارات مرجعية المحك ، وتتطلب هذه الطريقة تحديد درجة قطع Cut-off Score معينة للاختبار تفصل بين الطلاب المتقنين وغير المتقنين للأهداف التى يقيسها الاختبار. وباستخدام طريقة "انجوف" Angoff^{ss} Method فى تحديد هذه الدرجة والتى تتخلص في عرض مفردات الاختبار عل مجموعة من المحكمين يطلب من كل منهم أن يتصور مجموعة من الطلاب الذين حققوا الحد الادنى للكفاية في المجال

الذى يقيسه الاختبار ثم يطلب من كل محكم تقدير احتمال أن يجيب الفرد اجابة صحيحة عن المفردة دون أن يلجا إلى التخمين وعن طريق حساب مجموع أو القيم الاحتمالية التي قدرها كل محكم لكل مفردة من مفردات الاختبار وحساب متوسطها، انحرافها المعيارى لجميع المحكمين ولجميع مفردات الاختبار تم تحديد درجة القطع للاختبار والتي بلغت(٦٨,٩٨) أي حوالى (٦٩%) بانحراف معيارى (١,٠٩) (اسماعيل الوليلى ، ٢٠٠٣ : ١٧٨) (صلاح علام : ١٩٩٥ : ٢٣٩) .

نظرا لطول الاختبار (١٠٠) مفردة فقد تم تقسيمه إلى جزئين يضم الأول المفردات المتعلقة بالمهارات الأولى والثانية والثالثة وعددها (٥٢) مفردة أما الجزء الثانى فيضم المهارات الرابعة والخامسة والسادسة والسابعة وعددها (٤٨) مفردة وقد طبق الجزأين على التوالى على عينة عشوائية مكونة من (١٢٠) طالبا وطالبة من طلاب المرحلة الثانوية العامة ثم رصدت النتائج فى مصفوفة واحدة.

استخدم الباحث المعادلة التالية لحساب معامل سابكوفياك

$$ر = \frac{ع^{٢} + ر_{ii}(م-جـ)^{٢}}{ع^{٢} + (م-جـ)^{٢}}$$

(اسماعيل الوليلى، ٢٠٠٣ :١٧٩)

حيث(ع٢) تباين درجات مجموعة المفردات وتساوى (١٠,٧) ، (ر ii) معامل ألفا ويساوى (٠,٦٥)، (م) المتوسط الحاسبى لدرجات مجموعة المفردات ويساوى (٦٥,١٦) ،(جـ) درجة القطع وتساوى (٠,٦٩).

وقد تبين أن قيمة هذا المعامل عند درجة القطع السابقة بلغت (٠,٧٩٤) وتدل هذه القيمة على أن تصنيف الطلاب من حيث كونهم متقنين أو غير متقنين في أدائهم على مجموعة المفردات الاختبارية متسق بنسبة ٧٩% وتعد هذه القيمة مقبولة وتعبر عن ثبات درجات مجموعة المفردات ويوضح ملحق (٤) مفردات الاختبار مرجعي المحك لمادة الإحصاء في صورته النهائية . أما جدول رقم (٨) ملحق (٥) فيوضح المستويات المعرفية ومفاتيح التصحيح والمهارات التي تقيسها هذه المفردات .

السؤال الثاني : كيف يمكن استخدام نموذج راش في تكافؤ درجات صور اختبارية مرجعية المحك في مادة الإحصاء المقررة على طلاب المرحلة الثانوية العامة ؟ وللإجابة عن هذا السؤال اتبع الباحث الخطوات التالية :-

١- وزعت مفردات الاختبار المرجعي المحك والبالغ عددها (١٠٠) مفردة على خمس صور اختبارية تضم كل صورة (٢٠) مفردة بالإضافة إلى (٥) مفردات مشتركة بين كل صورة والتى تليها ليبلغ عدد مفردات كل صورة (٢٥) مفردة وجدول رقم (١٣) ملحق (١٠) يوضح أرقام مفردات كل صورة إختبارية وكذا المفردات المشتركة .

٢- طبقت كل صورة من الصور الاختبارية الخمس على العينة الخاصة بها جدول رقم (١) وتم الحصول على استجابات الطلاب من خلال ورقة الإجابة التي تم إعدادها لهذا الغرض ملحق رقم (١١).

٣- بعد أجراء عملية التطبيق قام الباحث بيما يلى :-

أ- تحديد القوة التمييزية للمفردة :

استخدم الباحث طريقة التحليل البيانى للمفردات raphical Item A Nalysis للحكم على القوة التمييزية للمفردة ومدى فاعليتها حيث تقدم هذه الطريقة رسما بيانيا لمحتوى كل مفردة يوضح العلاقة بين الدرجة الكلية والنسب المئوية لكل من البديل الصحيح والبدائل الخاطئة للمفردات كما تقدم معلومات توضيحية سهلة عن مواصفات المفردات ، كما يوفر التحليل البياني للمفردات نظرة مبدئية تساعد على التعرف على المفردات الغير مناسبة على أساس أن نسبة الطلاب الذين يختارون الإجابة الصحيحة من المفترض أن تزداد مع زيادة الدرجة الكلية والعكس صحيح (Batenburg & laros ; 2000 :1-13)

ب- استبعاد كل طالب لم يتمكن من الإجابة (الصواب) على أى مفردة – حيث يعتبر مستوى هذا الطالب أقل من مستوى الاختبار، أيضا استبعاد كل طالب أجاب (صوابا) على جميع المفردات حيث يعتبر مستواه أعلى من مستوى الاختبار وفى كلتا الحالتين يعتبر هذا الطالب خارج نطاق القياس.

جـ- حذف المفردات التى أجاب عنها جميع الطلاب إجابة (صحيحة) حيث تعتبر هذه المفردات أقل من مستوى الطلاب، كما حذفت المفردات التى لم يجب عنها أى من الطلاب حيث تعتبر أعلى من مستواهم وفى كلتا الحالتين تعتبر هذه المفردات غير مميزة. (أمينه كاظم، سليمان الخضري ، ٢٠٠١ : ٢٠)

٤- استخدام البرنامج الاحصائى (بايلوج) Bilog فى إجراء عملية التكافؤ لتدريج مفردات الصور الاختبارية الخمس وتحويلها إلى تدريج واحد مشترك باستخدام طريقة (البنود المشتركة) بحيث تشكل بنكا واحدا متدرجا من حيث (الصعوبة) بوحدة قياس مطلقة هى (اللوجيت) Logit وهى وحدة مشتقة ومعرفة بواسطة معادلات رياضية ولا يختلف المعنى الكمى لها باختلاف المتغير المقيس ويمكن تحويلها إلى وحدات أخرى مناظرة بهدف سهولة وتيسير التعامل معها ومنها التدريج بوحدة (الواط) Watt (المستخدمة في هذه الدراسة) لتقدير كل من متغيرى الصعوبة والقدرة ، ويتم تحويل وحدات اللوجيت إلى وحدات الواط الجديدة باستخدام المعادلات الآتية .

$$D = 50+(15/\ LN4)\ .d \ \&\quad B=50+(15/\ LN4).b$$

حيث (D&B) هما تقديرا الصعوبة والقدرة بوحدة الواط ، (d&b) هما تقديرا الصعوبة والقدرة بوحدة اللوجيت وفي هذا التدريج يكون متوسط صعوبة المفردات = (٥٠)، وتتدرج كل من صعوبات المفردات وقدرات الأفراد من القيمة صفر وحتى القيمة ١٠٠ أمينه كاظم(١٩٨٨)، صلاح علام (١٩٨٦)، محمد الشافعى (١٩٩٦)، اسماعيل الوليلى، (٢٠٠٠)

٥- أجريت عملية التدريج والتكافؤ على ثلاث مراحل :

المرحلة الأولى: حذف الطلاب غير الملائمين لعملية التدريج وفقا لدلالة ($كا^2$) حيث يستبعد كل طالب تكون قيمة ($كا^2$) عنده (دالة إحصائية) عند مستوى (٠,٠٥)

ونتيجة لهذه المرحلة فقد تم حذف (٧١) طالبا ليصبح العدد النهائى للطلاب (١٣٤٧) طالبا .

المرحلة الثانية: تحليل المفردات للعينة الكلية بعد حذف الطلاب غير الملائمين من المرحلة الأولى وذلك لحذف المفردات غير الملائمة والتى تكون قيمة ($كا^2$) عندها (دالة إحصائيا) عند مستوى (٠,٠٥) . مما يعنى عدم استقلال هذه المفردة عن باقى المفردات ونتيجة لهذه الخطوة تم حذف عدد (١٥) مفردة هى أرقام ٣، ٢٢ ، ٢٤ ، ٢٦ ، ٢٩ ، ٣٨ ، ٣٩ ، ٤٨ ، ٦٤ ، ٧٤ ، ٧٥ ، ٨٠ ، ٩٥ ، ٩٧ ، ٩٩ ليصبح عدد المفردات المكونة للبنك فى صورته النهائية (٨٥) مفردة.

المرحلة الثالثة :- إعادة التحليل للحصول على الصورة النهائية لتتدريج صعوبة المفردات المتبقية (٨٥) مفردة بعد حذف كل الأفراد والمفردات غير الملائمة وكذلك الحصول على القدرات المحتملة المقابلة لكل درجة خام من درجات الاختبار والجداول أرقام (٩) ،(١٠) ، (١١) ،(١٢) بالملاحق تشير إلى نتائج الخطوات السابقة .

نتائج الإجابة عن السؤال الثاني :

١- يتضح من الجدول رقم (٩) ملحق (٦) والخاص بترتيب مفردات البنك من حيث الصعوبة بكل من وحدتي اللوجيت ، الواط وكذلك الخطأ المعيارى لكل مفردة بعد حذف المفردات الغير ملائمة لعملية التدريج أن تقديرات الصعوبة تراوحت ما بين (-٣,٢) إلى (٣,١٤) مقدرة بوحدة اللوجيت وهو ما يقابل (١٥,٣٨) إلى (٨٣,٩٣) مقدرة بوحدة الواط .

٢- تشير نتائج الجدول رقم (١٠) ملحق (٧) إلى أن تقديرات القدرة المحتملة المقابلة لكل قدرة خام كلية من درجات بنك الأسئلة وأيضا الخطأ المعيارى مقدرا بكل من وحدتي اللوجيت والواط قد تراوحت ما بين (-٣,٥٠) إلى (٣,٤٤) مقدرة بوحدة اللوجيت وهو ما يقابل (١٢) إلى (٨٧) مقدرة بوحدة الواط .

٣- تشير نتائج الجدول رقم(١١) ملحق (٨) إلى أن تقديرات الصعوبة والخطأ المعيارى لمفردات كل صورة اختبارية من الصور الخمس كانت كما يلى :

- بالنسبة للصورة الأولى : كما بين (-٣,٠١) إلى (٢,٩٥) مقدرة بوحدة اللوجيت وهو ما يقابل (١٧,٤٣) إلى (٨١,٩٣) مقدرة بوحدة الواط .
- بالنسبة للصورة الثانية : ما بين (-٣,١٨) إلى (٢,٨١) مقدرة بوحدة اللوجيت وهو ما يقابل (١٥,٥٩) إلى (٨٠,٤) مقدرة بوحدة الواط .
- بالنسبة للصورة الثالثة : ما بين (-٣,١٥) إلى (٣,١٤) مقدرة بوحدة اللوجيت وهو ما يقابل (١٥,٩٢) إلى(٨٣,٩٣) مقدرة بوحدة الواط .
- بالنسبة للصورة الرابعة : ما بين (-٢,٧٣) إلى (٣,١١) مقدرة بوحدة اللوجيت وهو ما يقابل (٢٠,٤٦) إلى (٨٣,٦٥) مقدرة بوحدة الواط .
- بالنسبة للصورة الخامسة: ما بين (-٣,٠٢) إلى (٣,١١) مقدرة بوحدة اللوجيت وهو ما يقابل (١٥,٣٨) إلى (٨٣,٦٥) مقدرة بوحدة الواط .

٤- تشير نتائج الجدول رقم (١٢) ملحق (٩) إلى أن تقديرات القدرة المحتملة للأفراد والمقابلة لكل درجة خام من درجات كل صورة اختبارية من الصور الخمس كانت كما يلى :

- بالنسبة للصورة الأولى : ما بين (-٢,٦٨) إلى (٢,٧٥)مقدرة بوحدة اللوجيت وهو ما يقابل (٢٠,٩٨) إلى (٩٧,٧٧) مقدرة بوحدة الوط .
- بالنسبة للصورة الثانية: ما بين (-٣,٠٧) إلى (٢,٩١) مقدرة بوحدة اللوجيت وهو ما يقابل (١٦,٧٤) إلى (٨١,٤٩) مقدرة بوحدة الوط .
- بالنسبة للصورة الثالثة: ما بين (-٢,٩٣) إلى (٣,١٢) مقدرة بوحدة اللوجيت وهو ما يقابل (١٨,٣٥) إلى (٨٣,٧٦) مقدرة بوحدة الواط .
- بالنسبة للصورة الرابعة: ما بين (-٣,٣٢) إلى (٢,٥٠) مقدرة بوحدة اللوجيت وهو ما يقابل (١٤,٠٦) إلى (٧٧,٠٥) مقدرة بوحدة الواط .
- بالنسبة للصورة الخامسة: ما بين (-٢,٧١) إلى (٢,٨٦) مقدرة بوحدة اللوجيت وهو ما يقابل (٢٠,٦٣) إلى (٨٠,٩٥) مقدرة بوحدة الواط .

السؤال الثالث: هل يختلف أداء طلاب المرحلة الثانوية العامة في مادة الإحصاء باختلاف متغيرى الجنس (ذكور / إناث)، التخصص (علمي/أدبي)؟ وما هو حجم التأثير الناتج عن كلا المتغيرين والتفاعل بينهما على هذا الأداء ؟

وللإجابة عن هذا السؤال استخدم الباحث تحليل التباين ذي التصميم العاملى ٢×٢ لدراسة تأثير كل من متغيرى الجنس ، التخصص والتفاعل بينهما على أداء طلاب المرحلة الثانوية العامة فى مادة (الإحصاء) والمقدر بوحدة "اللوجيت" كما تم حساب حجم تأثير كل من المتغيرين والتفاعل بينهما باستخدام مربع "ايتا" لكل قيمة من قيم (ف) الدالة إحصائيا مستعينا بالبرنامج الإحصائي SPSS (10) فى الوصف الاحصائى لعينة الدراسة والإجابة عن هذا السؤال. (رجاء أبو علام : ٢٠٠٣)

والجدول رقم (٣) يوضح المتوسط والانحراف المعيارى لتقديرات القدرة المحتملة لأفراد عينة الدراسة النوعية مقدرا بكل من وحدتي اللوجيت والواط أما الجدول رقم (٤) فيشير إلى نتائج تحليل التباين .

جدول رقم (٣)

المتوسط والانحراف المعيارى لتقديرات قدرة عينات الدراسة والخطأ المعيارى بكل من وحدتي اللوجيت والواط "

العينة النوعية	عدد الأفراد	اللوجيت		الواط	
		المتوسط	الانحراف المعيارى	المتوسط	الانحراف المعيارى
بنين	٦٦٩	١,٠٦	٠,٦١٤	٦١	٧
بنات	٦٧٨	٠,٩٤٢	٠,٦٢٥	٦٠	٧
علمى	٧٢٣	١,٠١	٠,٦٠٨	٦١	٧
أدبى	٦٢٤	٠,٩٨٨	٠,٦٣٩	٦١	٧
العينة الكلية	١٣٤٧	١	٠,٦٢٢	٦١	٧

جدول رقم (٤)

يوضح نتائج تحليل التباين ٢×٢ بين متغيرى الدراسة

الجنس والتخصص فى أدائهم لمادة الاحصاء

مصدر التباين	مجموع المربعات	درجات الحرية	متوسط المربعات	قيمة (ف)	مستوى الدلالة	مربع ايتا
الجنس	٤,٧٨	١	٤,٧٨	١٢,٤٤	٠,٠٠١	٠,٠٠٩*
التخصص	٠,٢٢٨	١	٠,٢٢٨	٠,٥٩	غير دالة	
التفاعل	٠,٠٧	١	٠,٠٧	٠,١٨	غير دالة	
الخطأ	٥١٥,٩٤٤	١٣٤٣	٠,٣٨٤			

* حجم التأثير ضئيل ، رشدي فام منصور ؛ ١٩٩٧ : ٥٧ – ٧٥

نتائج الإجابة عن السؤال الثالث : يتضح من الجدول رقم (٤) ما يلى :

أ- بالنسبة لمتغير الجنس: يتضح من قيمة (ف) ومستوى دلالتها الإحصائية وجود فروق دالة إحصائيا عند مستوى (٠,٠٠١) بين الذكور والإناث فى القدرة على الأداء فى مادة الاحصاء وذلك لصالح عينة الذكور إلا أن قيمة مربع (ايتا) = ٠,٠٠٩ مما يعنى أن حجم التأثير الناتج عن اختلاف الجنس يعتبر ضئيلا.

ب- بالنسبة لمتغير التخصص (علمي / أدبي) يتضح من قيمة (ف) ومستوى دلالتها الإحصائية أنها غير دالة ، ومن ثم لا توجد فروق دالة إحصائيا بين طلاب العلمي، الأدبي فى القدرة على الأداء فى مادة الاحصاء.

جـ- بالنسبة لتفاعل كلا من الجنس، التخصص على الأداء فى مادة الاحصاء يتضح من قيمة (ف) ومستوى دلالتها الإحصائية أنها غير دالة، مما يعنى عدم وجود تفاعل بين الجنس والتخصص يمكن أن يؤثر على اداء الطلاب لهذه المادة.

السؤال الرابع :هل توجد فروق دالة إحصائيا بين النسب المئوية لتحقيق كل مهارة من المهارات السبع الرئيسة المتضمنة مادة الإحصاء

نتيجة الاختلاف متغيرى الجنس (ذكور/إناث)، التخصص (علمى/ادبى)؟

وللإجابة عن هذا السؤال تم حساب النسب المئوية للإجابات الصحيحة عن مجموعة المفردات التى تتضمنها المهارة ثم حساب المتوسط الحسابي لهذه النسب لدى كل عينة نوعية،وذلك بعد إجراء عملية التدريج وحذف المفردات والأفراد الغير ملائمين كما حسبت قيمة (Z) لدلالة فروق هذه النسب ، والجدول رقم (٥) يشير إلى النتائج.

جدول (٥)

دلالة فروق النسب المئوية لتحقيق المهارات

الرئيسة لمادة الإحصاء لدى عينات الدراسة النوعية

المهارات الرئيسة	عدد المفردات بعد التدريج	الجنس			التخصص		
		ذكور	إناث	قيمة (Z)	علمى	أدبى	قيمة (Z)
الأولى	١٨	٨٩,٨٧	٨٨,٢٧	*٢,١٢٦	٨٦,٨	٨٨,٠٧	١,٥٩٥
الثانية	١٣	٨٥,٣٣	٨٢,٩٨	١,٨٦٨	٨٣,٦٦	٨٤,٦٣	٠,٧٦٨
الثالثة	١٣	٨١,٤٧	٧٧,٦٢	*٢,٩٩	٧٨,٤٧	٨٠,٧	١,٧٢٥
الرابعة	٦	٨٨,١٩	٩٠,٥٢	١,٨١٥	٩٠,٠٥	٨٨,٩٣	٠,٨٧٥
الخامسة	٢١	٧٢,٦٦	٧١,٢٦	١,٢٥١	٧١,١٧	٧٢,٨٥	١,٤٩٧
السادسة	٩	٨٨,٣٤	٨٨,٦٦	٠,٢٧٩	٨٩,١	٨٧,٨١	١,١٢٤
السابعة	٥	٨٨,٦٦	٨٦,٤٤	١,٢٠٩	٨٧,٣٨	٨٧,٧٩	٠٢٢٣

* دالة عند مستوى ٠,٠١

نتائج الإجابة عن السؤال الرابع : يتضح من الجدول رقم (٥)ما يلى:

- بالنسبة لمتغير الجنس توجد فروق دالة إحصائيا عند مستوى (٠,٠١) بين الذكور والإناث فى تحقيق المهاراتين الأولى والثالثة وذلك لصالح عينة الذكور أما باقي المهارات فلا توجد فروق دالة إحصائيا بين الجنسين فى تحقيقها.
- بالنسبة لمتغير التخصص. لا توجد فروق دالة إحصائيا بين طلاب التخصص العلمي والتخصص الأدبي فى تحقيق مهارات مادة الإحصاء.تراوحت النسب المئوية لتحقيق المهارات الرئيسة كما يلى:

- بالنسبة لعينة الذكور ما بين (٧٢,٦٦) للمهارة الخامسة إلى (٨٩,٨٧) للمهارة الأولى.
- بالنسبة لعينة الإناث ما بين (٧١,٢٦) للمهارة الخامسة إلى (٩٠,٥٢) للمهارة الرابعة.
- بالنسبة لطلاب التخصص العلمي ما بين (٧١,١٧) للمهارة الخامسة إلى (٩٠,٠٥) للمهارة الرابعة.
- بالنسبة لطلاب التخصص الأدبى ما بين (٧٢,٨٥) للمهارة الخامسة إلى (٨٨,٩٣) للمهارة الرابعة.

التعليق على نتائج الدراسة:

- من خلال تطبيق نظرية الاستجابة للمفردة (IRT) وباستخدام نموذج (راش) فى تكافؤ درجات الصور الاختبارية الخمس للاختبار مرجعي المحك تم حذف عدد (١٥) مفردة اختبارية واستنادا إلى نتائج الدراسات السابقة والتى أكدت أفضلية نظرية (IRT) فى انتقاء المفردات أمينه كاظم(١٩٨٨ أ)، صلاح علام(١٩٩٠) ، Hambleton & Rogers (1990) وبذلك تعتبر المفردات المتبقية لهامعاملات صدق وثبات عالية Harvy ، Sayre (1990) & Harrise (1990).
- أحادية القياس واستقلاليته التى يتميز بها نموذج (راش) أتاح الفرصة لتدريج مفردات الصور الاختبارية الخمس على ميزان واحد مشترك له وحدة قياس (اللوجيت) أو ما يعادلها (بالواط) بضم جميع هذه الاختبارات فى اختبار واحد ويتيح هذا تعادل نتائج القياس بمعنى التوصل إلى تقديرات لقدرة الفرد (تحصيله) لا تختلف باختلاف الصورة الاختبارية. (أمينه كاظم ، سليمان الخضرى ، ٢٠٠١ : ١٧)
- على الرغم من اختلاف الدرجة الخام الكلية للفرد تبعا لاختلاف خصائص الصورة الاختبارية المستخدمة إلا أن ترجمتها إلى تقدير لقدرة هذا الفرد يكون متكافئ طالما أن الاختبارات المستخدمة تندرج فيما بينها فى تدريج واحد مشترك.

- يعتبر الباحث أن الصور الاختبارية الخمس متعادلة القياس بعد حذف المفردات الغير ملائمة كما أن المقياس الذى يجمعها فى تدريج واحد يتمتع بالصدق والثبات المطلوبين حيث لا يختلف القياس (سواءا بالنسبة لقدرة الفرد أو لصعوبة المفردة) باختلاف عينة التدريج أو باختلاف الصورة الاختبارية المستخدمة.

وعن نتائج الإجابة عن تساؤلات الدراسة:

- أظهرت النتائج عدم وجود فروق دالة إحصائيا بين طلاب القسم العلمى والقسم الأدبي أو لتفاعل متغيرى الجنس، نوع التخصص، في القدرة على الأداء فى مادة الاحصاء. وإن ظهرت هذه الفروق دالة إحصائيا لصالح الذكور فى متغير الجنس إلا أن حجم التأثير كان ضئيلا مما يعنى عدم تأثر الأداء فى مادة الإحصاء باختلاف التخصص أو الجنس وقد يعود ذلك إلى اهتمام جميع الطلاب ذكورا وإناث فى القسمين العلمي والأدبى بهذه المادة التى تمثل احدى المواد الاختيارية من مواد الثانوية العامة نظرا لما تمثله من درجات (٥٠) درجة مثلها مثل أى مادة أساسية كما أن عملية توزيع الطلاب على القسمين العلمي والأدبي لم تعد قائمة على التفوق فى المواد العلمية أو الأدبية كما كان من ذى قبل بل تداخلت أسس التوزيع.
- وانعكس ذلك أيضا على تحقيق المهارات الرئيسة لمادة الإحصاء حيث لم توجد فروق دالة إحصائيا بين طلاب التخصص العلمي والأدبي فى تحقيق هذه المهارات . وإن وجدت الفروق لصالح الذكور فى تحقيق مهارتين هما الأولى والثالثة فقط إلا أن هذا لا ينفى ارتفاع نسبة تحقيق هاتين المهارتين بين الذكور والإناث على حد سواء حيث بلغت نسبة تحقيق هذه المهارات (٨٩,٨٧%) للذكور، (٨٨,٢٧%) للإناث بالنسبة للمهارة الأولى، ونسبة (٨١,٤٧%) للذكور، (٧٧,٦٢%) للإناث بالنسبة للمهارة الثالثة. وهذه نسبة تفوق حد التمكن.والذى يحدد عادة بنسبة ٠,٧٥ كما لم توجد فروق دالة إحصائيا بين النسب المئوية لتحقيق هذه المهارات الرئيسة لطلاب القسم العلمى أو الأدبى.

المراجـــع

١- أحمد الرفاعى غنيم(١٩٨٥): "تطبيقات على ثبات الاختبارات" ط١ ، القاهرة، مكتبة نهضة الشرق.

٢- اسماعيل حسن الوليلى(١٩٩٦): "أثر استخدام بعض طرق التغذيـــة المرتـــدة على ثبات درجات اختبار هـــدفى المرجـــع وعلـــى اســـتبقاء المعلومات" رسالة ماجستير غير منشورة ، كليـــة التربيـــة – جامعة الأزهر.

٣- اسماعيل حسن الوليلى(٢٠٠٢): "دراسة سيكومترية مقارنة لـــبعض نمـــاذج الاستجابة للمفردة فى انتقاء مفـــردات الاختبـــارات مرجعيـــة المحك" رسالة دكتوراه غير منشورة – كلية التربية – جامعـــة الأزهر.

٤- اعتدال غازى عبد الله(٢٠٠٠) : القياس محكي المرجع لاحد مقررات علم النفس التعليمي – دراسة ميدانية . رســـالة ماجســـتير غيـــر منشورة كلية البنات – جامعة عين شمس .

٥- أمينه محمد كاظم (١٩٨٨ أ) : "استخدام نموذج راش فى بناء اختبار تحصيل فى علم النفس وتحقيق التفسير الموضوعى للنتـــائج"، جامعـــة الكويت – مطبوعات جامعة الكويت.

٦- أمينه محمد كاظم (١٩٨٨ ب): دراسة نظرية نقدية حول التفسير الموضوعى للسلوك (نموذج راش) ، ط١ ، الكويت، مؤسسة الكويت للتقدم العلمى.

٧- أمينه محمد كاظم، سليمان الخضرى الشـــيخ(٢٠٠١): "المشـــروع العربـــى للتقويم المقارن لمستويات التحصيل الدراسى فى التعليم العـــام

– المرحلة الثالثة". تقرير جمهورية مصر العربية – القاهرة – المركز القومى للامتحانات والتقويم التربوى.

٨- أنور الشرقاوى ، سليمان الخضرى الشيخ، أمينه كاظم ، ناديه عبد السلام(١٩٩٦): "إتجاهات معاصرة فى القياس والتقويم النفسى والتربوى" القاهرة –مكتبة الانجلو المصرية.

٩- رجاء محمود أبو علام (٢٠٠٣): "التحليل الاحصائى للبيانات باستخدام برنامج SPSS" ط١ – القاهرة– دار النشر للجامعات

١٠- رشدى فام منصور(١٩٩٧): "حجم التأثير الوجه المكمل للدلالة الاحصائية" المجلة المصرية للدراسات النفسية – القاهرة– المجلد السابع– ع١٦ .

١١- شادية عبد العزيز منتصر(١٩٩٦) : استخدام نموذج (راش) في بناء بنك للأسئلة لمقرر علم النفس التعليمى وتحديد الدرجات الفاصلة المقابلة للتقديرات الجامعية . رسالة ماجستير غير منشورة كلية البنات – جامعة عين شمس.

١٢- شحته عبد المولى عبد الحافظ (١٩٩٩): "تقويم بناء الاختبارات المرجعة إلى (المحك/ المعيار) فى ضوء نظرية الاستجابة للمفردة والنظرية التقليدية" رسالة دكتوراه غير منشورة – كلية التربية –جامعة عين شمس.

١٣- صلاح الدين محمود علام (١٩٨٤): "بعض الاتجاهات العالمية المعاصرة فى مجال تقويم الطلاب، ورقة مقدمة فى ندوة ورشة تطوير أساليب تقويم طلاب التعليم العام لدول الخليج العربى" ، الكويت– المركز العربى للبحوث التربوية لدول الخليج العربى".

١٤- صلاح الدين محمود علام (١٩٨٥): "تحليل بيانات الاختبارات العقلية باستخدام (نموذج راش) اللوغاريتمى الاحتمالى" "دراسة

تجريبية" المجلة العربية للعلوم الإنسانية، الكويت، المجلد الخامس- ع١٧ ،.

١٥- صلاح الدين محمود علام (١٩٨٥) : بناء اختبار هدفي المرجع لقياس مهارات المعلمين في تطوير الاختبارات المدرسية . الكويت مجلة العلوم الاجتماعية ، المجلد الثالث عشر، ع٤ ، ص ١١٧ - ١٢٨ .

١٦- صلاح الدين محمود علام (١٩٨٦): "تطورات معاصرة فى القياس النفسي والتربوي" الكويت- إدارة التأليف والترجمة والنشر.

١٧- صلاح الدين محمود علام (١٩٨٧): "دراسة موازنة ناقدة لنماذج السمات الكامنة والنماذج الكلاسيكية فى القياس النفسي والتربوي" المجلة العربية للعلوم الإنسانية - الكويت، المجلد السابع ع٢٧ ،.

١٨- صلاح الدين محمود علام (١٩٩٠): "استخدام نموذج راش فى بناء مقياس هدفى المرجع للمعارف الأساسية فى إعداد خطة البحوث النفسية والتربوية" مجلة كلية التربية - جامعة الأزهر، القاهرة - ع١٧.

١٩- صلاح الدين محمود علام (١٩٩٥): "التوجهات المستقبلية لتقويم تحصيل الطلاب فى ضوء متطلبات القرن الحادي والعشرين"، مجلة كلية التربية - جامعة الأزهر ، ع٤٩.

٢٠- صلاح الدين محمود علام (١٩٩٥): "الاختبارات التشخيصية مرجعية المحك فى المجالات التربوية والنفسية والتدريبية": ط١ - القاهرة - دار الفكر العربي.

٢١- محمد عباس المغربى (١٩٩٤): "أثر استخدام بعض أساليب العقاب ومستويات الدافع المعرفي فى تحصيل الفيزياء لدى تلاميذ

الصف الثاني الثانوي" رسالة ماجستير غير منشورة – كلية التربية – جامعة الإسكندرية.

٢٢- محمد منصور الشافعى (١٩٩٦): "أثر طرق معادلة درجات الاختبار وضوابط اختيار العينة على تدريج بنود بنك الأسئلة باستخدام (نموذج راش) رسالة دكتوراه غير منشورة – كلية التربية – جامعة المنصورة.

٢٣- محمود عبد الحليم منسى، أحمد شعبان محمد (٢٠٠٠)"التقويم التربوي ومبادئ الإحصاء" كلية التربية- جامعة الإسكندرية .

24- Batenburg , T. & Laros, J(2000): Date Screening , Graphical Analysis of items. University of Groningen, The Netherlands, T.A.T. A. Van Batenburg @ PPSW. RVG. NL.

25- Gagne,R.M(1968): Learning Hierarchies. Educational psychologist, V.6 No.1, November, PP (1-9).

26- Hambleton, R.K & Swaminathan, H. (1985): Algina J. & Coulson, D.B. (1978): Criterion- Referenced testing and Measurment A revew of Technical Issues and Development" Revew of Educational Research, 84, No.1, PP. 1-48.

27- Hambleton, R.K.& Swaminathan, H. Item Response theory, Principles and Applications. Kluwer Academic publisher, Assinppi park, No r well, U.S.A.

28- Hambleton, R.K & Swaminathan, H. & Rogers, H.J. (1991): Fundamentals of Item, Response theory? SAGE publications Inc., Teller Road, Newbury park, California,.

29- Hambleton, R.K. & Stephen, (1997): G Future Directions for Norm- Referenced and Criterion – Referenced Achienvement Testing. Laboratory of psychometric and Evalutive Research Report. International Journal of Educational Research, Vo,27, No.5,

30- Harris, D.J. & Kolen, M.J(1990) :Acomparison of two

Equipercentile Equating methods for common Item Equating. Educational and psychological Measurement, Vo, 50,.

31- Harvey, R.J.& Hammer, A.L. (1999): Item Response theory counseling psychologist, college park, Vo.27, Issues, May PP 353-383.

32- Hills, J.R, et all. (1987): Equating minimum Competency Tests; Comparison of methods Florida state univ, Tallahossee,.

33- Holland, P.W. & Rubin, D.B. (1982) : Test Equating. London, Academic Press,.

34- Kolen, M.J. & Brennan, R.L (1995): Test Equating Methods and Practices. New York, Springer.

35- Perkins, K.& Miller, L.D(1984): comparative Analyses of English as A second Language Reading comprehension Data: classical Test theory and latent trait Measurment. Language Testing.1.. PP. (21-32).

36- Popham, W.J. (1978): Criterion – Referenced Measurment. Englewood Cliffs, Prentice- Hall, Inc, New Jersey.

37- Popham, W.J. (1980): Modern Educational Measurment. Englewood Cliffs, Prentice – Hall, Inc, New Jersey..

38- Popham, W.J. (1988): Educational Evaluation. Second Edition Englewood Cliffs prentic, Hall, , New Jersey..

39- Rovinelli, R. & Hambleton, R.K(1976): on the use of content specialists in Assessment of criterion – Referenced test Items validity. Laboratory of psychometric and Evaluative Research Report, No.24. Amherst Mass; Univ-Massachusetts.

40- Sayre, K.A(1990): Developing. Minimum Competency Tests: conventional vereus. Item Response theory methods. Unpublished, Univ, of Lioncivil U.S.A..

41- Skaggs, G.& Lissitz, R.W. (1985): An Exploration of the Robustness of four Test Equating Models: Paper presented at the Annual meeting of the AERA. Chicago, March , April..

42- Snieckus, A.H. (1992): Stability of the Equated score scale In the presence of Amultidimensional Test structure; Diss, Abs, Int Dec..

43- Wainer, H. & Others (2000): Computerized Adaptive Testing . Aprimer Second Edition, New Jersey Lowrence Elbraum ASSOCIATES . Inc.

44- Wright, B.D, Stone, M.H: Best Test Design, Chicago: Mesapress.(1979).

Abstract

The educational achievement test is considered one of the principle methods in the evaluation process. The educational organizations depend on it to recognize the development of the student achievement. Although achievement test score widely -used in the educational — field they suffer from many problems which hinder their objectivity such as the total score of test which depend on the data'S items. The scores which don't increase monotonically with the measured variable The changability of items form and meaning through time and finally the problems related of reliability and validity. The scientists of measurement tried to achieve the objectivity in measure and behavior evaluation their efforts resulted in developing the Items — response theorem which substituted of the classical approach in test construction and the learner evaluation techniques and it's tools have also been developed so that they don't depend only on determination the learner's place among his peers but also on his practices and competencies . As result of this, The criterion referenced tests appeared instead of norm — referenced tests which faced a lot of criticism. In this study the researcher used the (Rasch) model to equate a criterion —referenced tests scores in (statistic)that set for the secondary schools students and comparing their performances in t light of the variables of sex and science specialization to ensure how well they master the subject skills. The results of the study should showed statistically significant differences between the mean co of males and females in favor of the males and the size of effect was few, But these differences disappeared between the science section student and literary section student There is also no effect due to the interaction between the variables of sex and. Specialization on the performances in this subject . The results also referred to statistically significant differences between males and females in favor of the males concerning percentage of the two competencies only but there are no differences between percentages in science and literary students All competencies are satisfied in the study_sample as a percentage ranging from (71.17% to 90.52%).