مجلة البحث في التربية وعلم النفس
كلية التربية – جامعة المنيا

أثر التحليل بالنماذج اللوغاريتمية (الأحادية والثنائية والثلاثية) البارمتر على دقة تدرج بنك أسئلة في مادة الرياضيات

إعداد

د. منى ربيع الطنطاوي متيرد
مدرس القياس النفسي والتربوي
بكلية البنات جامعة عين شمس

أ. محمد جاد محمد محمد
باحث مساعد بالمركز القومي
للبحوث التربوية والتنمية

ملخص الدراسة :

تهدف هذه الدراسة إلى المقارنة بين النماذج اللوغاريتمية الثلاث لنظرية الاستجابة للمفردة (لأحادي/ الثنائي/ الثلاثي) البارامتر في تدريج بنك للأسئلة الموضوعية في مادة الرياضيات للصف الخامس الابتدائي وذلك من حيث دقة القياس التي تتمثل في انخفاض قيمة الخطأ المعياري لكل من صعوبة المفردات، وقدرة الأفراد، وارتفاع مؤشرات الثبات لبنك الأسئلة، وأيضا ارتفاع قيم دوال المعلومات. أجريت الدراسة على عينة مكونة من ١٥٠٠ تلميذاً وتلميذة من طلاب الصف الخامس الابتدائي من مرحلة التعليم الأساسي بمحافظة السويس ، واستخدمت الدراسة برنامج التحليل الإحصائي Parscale **الإصدار** 4.1 **في إجراء المعالجات الإحصائية لبيانات أداء أفراد العينة على بنك أسئلة وفقاً لنماذج الاستجابة للمفردة اللوغاريتمية الثلاث (الأحادية والثنائية والثلاثية) البارامتر. وقد توصلت الدراسة إلى أن النموذج ثلاثي البارمتر أفضل النماذج اللوغاريتمية الثلاث دقة في القياس حيث تبين أن النموذج الثلاثي البارمتر أدت تحليلاته إلى أقل تقدير للأخطاء المعيارية لتقديرات صعوبة المفردات وتقديرات قدرة الأفراد كما قدم النموذج الثلاثي أعلى التقديرات لقيم دوال معلومات مفردات بنك الأسئلة . وفي ضوء ذلك تمكنت الدراسة من تدريج وتقدير صعوبة المفردات وقدرات الأفراد في بنك الأسئلة وفقاً للنموذج اللوغاريتمي ثلاثي البارامتر الأكثر دقة في القياس. كما تم حساب صدق وثبات بنك الأسئلة في صورته النهائية وذلك بعد حذف الأفراد غير الملائمين والمفردات غير الملائمة لأسس القياس الموضوعي باستخدام الطرق المتعددة في حساب الصدق والثبات وفقاً للنموذج اللوغاريتمي ثلاثي البارامتر الأكثر دقة في القياس.**

مقدمة :

تسعى كافة العلوم إلي تطوير أساليب موضوعية ودقيقة لقياس الظواهر المتعلقة بها وذلك لتحقيق الدقة في فهم تلك الظواهر وما يستتبع ذلك من دقة في التنبؤ بالعلاقات القائمة بين تلك المتغيرات ومن ثم ضبطها والتحكم فيها، حيث يبقى تقدم أي علم من العلوم مرهونا بمدى الدقة والموضوعية التي توفرها الأدوات المستخدمة في قياس وتكميم ظواهره ومتغيراته. لهذا بذلت محاولات عديدة من قبل المهتمين بالقياس النفسي والتربوي بهدف التوصل إلى الموضوعية في قياس السلوك ، وقد أسفرت تلك الجهود عن اتجاهات معاصرة في مجال القياس والتقويم ، ومن بين هذه الاتجاهات نظرية الاستجابة للمفردة.

لقد حققت نظرية الاستجابة للمفردة كأحد الاتجاهات الحديثة للقياس وما ارتبط بها من نماذج رياضية لوغاريتمية طفرة هائلة ومتطورة في مجال القياس النفسي والتربوي ، حيث أثبتت العديد من الدراسات الأجنبية والعربية تفوق هذه النظرية في تحقيق الموضوعية المأمولة منها في هذا المجال إذا ما قورنت بالنظرية التقليدية للاختبار (De Champlain,2009:109) ، كما أن هذه النظرية كما يرى كل من (Uttaro&LeHman,1994:41; Rotou,2002:131) تعتمد على افتراضات وشروط أقوى من الافتراضات والشروط ذات الصلة بالنظرية التقليدية التي كانت لسنوات عديدة مضت هي الأساس النظري والعملي للقياس في العلوم السلوكية والتي ركزت على الدرجة الخام الملاحظة وتضمنت نوعين أساسيين من الاختبارات وهما "الاختبارات جماعية المرجع" التي تقتصر على التمييز بين مستويات الأفراد ، و" الاختبارات محكية المرجع " التي تقتصر على الكشف عن مدى تحقق الأهداف في عملية القياس (نقلا عن : محمد الشافعي ، ٢٠١١ : ٥٥٠) .

هذا بالإضافة إلى ما أثبتته العديد من الدراسات الأجنبية والعربية بهذا المجال من حيث فاعلية استخدام نظرية الاستجابة للمفردة والنماذج الرياضية المرتبطة في بناء وتطوير مختلف أنواع أدوات القياس التي تستخدم لأغراض جمع أنواع مختلفة من البيانات إذا ما قورنت بالنظرية التقليدية للقياس (Loken & Kelly, 2010). كما أن الدراسات التي استخدمت نفس النظرية في تقدير بارامترات المفردات الاختبارية قد أثبتت كذلك تميز هذه الأدوات من حيث تمتعها بخصائص الموضوعية

التي تتسم بها تلك الأدوات المستخدمة في مجال القياس الفيزيائي.

ونتيجة لتميز هذه النظرية ونماذجها الرياضية فقد تنامت التطبيقات الخاصة بها في الآونة الأخيرة على الصعيدين العالمي والعربي في مختلف مجالات القياس التي تهتم ببناء وإعداد أدوات جديدة متنوعة لجمع البيانات ، أو إعادة تطوير بعض الاختبارات العالمية شائعة الاستخدام (Edward,2010) ، وكذلك إنشاء بنوك الأسئلة .(Njtiu & Romanoski , 2007).

وتعتبر بنوك الأسئلة بصورتها المطورة من أهم التطبيقات العملية لنماذج نظرية الاستجابة للمفردة فقد بدأ بهذا المفهوم منذ منتصف الستينيات ، وذلك عندما بدأت دراسة مشروع بناء بنك أسئلة في انجلترا 1966 وقد تعددت المسميات التي أطلقت على بنك الأسئلة ومنها بنك الأسئلة/مستودع المفردات ، تجمع المفردات ، وذخيرة المفردات. ولكن مع التطور ارتبطت مفاهيم (مستودع المفردات ، تجمع المفردات ، ذخيرة المفردات) بالنظرية التقليدية للقياس ، بينما ارتبط مفهوم بنك الأسئلة بنظرية الاستجابة للمفردة حيث تطورت بنوك الأسئلة من مجرد مستودع يضم عدد كبير من المفردات الاختبارية إلى مجموعة كبيرة من المفردات المدرجة بوحدة قياس معرفة بحيث تمثل مقياس يغطى مدى واسع من المتغير موضوع القياس.

وتستخدم نظرية الاستجابة للمفردة في تدريج بنوك الأسئلة بنماذجها اللوغاريتمية الثلاث ، وقد اتجهت معظم الدراسات التي أجريت في البيئة العربية إلى استخدام النموذج الأحادي المعلم (نموذج راش) ، كدراسة محمد الشافعي (١٩٩٦) ، ودراسة شادية عبد العزيز (١٩٩٦) في بناء بنك للأسئلة لمقرر في علم النفس التعليمي وتحديد الدرجات الفاصلة المقابلة للتقديرات الجامعية، ودراسة جواهر الزيد (٢٠٠٨) التي درست فاعلية نظرية الاستجابة للمفردة في إعداد بنك للأسئلة في مقرر علم النفس التربوي. بينما اتجهت دراسات أخري في هذا المجال إلى استخدام أكثر من نموذج في تحليل بياناتها بهدف التوصل إلى النموذج الأكثر دقة والأكثر ملاءمة للبيانات ، ولقد تباينت المحكات التي حددتها هذه الدراسات لتحديد النموذج الأدق في القياس ومن هذه المحكات :عدد المفردات غير الملائمة للنموذج المستخدم ، دالة معلومات الاختبار ، انخفاض قيمة الخطأ المعياري لصعوبة المفردات وقدرة الأفراد ، ارتفاع مؤشرات الثبات. ومن هذه الدراسات دراسة علاء الدين أيوب (٢٠٠١) التي توصلت إلى تفوق النموذج اللوغاريتمي

ثلاثي البارامتر على النموذجين الأحادي والثنائي البارامتر في دقة تقدير صعوبة المفردات وقدرة الأفراد ودوال المعلومات ومؤشرات الثبات على أبعاد مقياس المهارات المعرفية. ودراسة إسماعيل الوليلي (٢٠٠٢) التي توصلت إلى عدم وجود فروق ذات دلالة إحصائية في دقة تقديرات قدرات الأفراد بين النماذج الثلاث (الأحادي ،الثنائي ، الثلاثي) البارمتر، كذلك توصل إلى اختلاف واضح في قيم ثبات الدرجات للنماذج الثلاث عند درجات القطع المتعددة وتوصل أيضا على عدم وجود فروق ذات دلالة إحصائية في دقة النماذج الثلاث وكفايتها في قياس السمة المحددة ، وأن الأخطاء المعيارية للقياس المتعلقة بهذه النماذج متقاربة مما يدل على الدقة التي تتسم بها هذه النماذج. ودراسة نسرين إسماعيل (٢٠١٠) التي هدفت إلى دراسة بعض المتغيرات التي تؤثر على دقة القياس باستخدام نظرية الاستجابة للمفردة .

وعلى الرغم من توافر الدراسات الأجنبية التي اهتمت بالمقارنة بين النماذج الثلاث (الأحادي ، الثنائي ، الثلاثي) البارمتر في مجالات متعددة ومنها الاختبارات التحصيلية المختلفة ، إلا أن هناك ندرة لمثل هذه الدراسات في البيئة العربية بالرغم من أهمية الدور الذي تؤديه الاختبارات في العملية التعليمية، حيث تعتبر الاختبارات التحصيلية من أكثر وسائل التقويم المستخدمة في التعليم، الأمر الذى يدعو إلى ضرورة الاهتمام بعملية بناء الاختبارات وتطويرها من قبل التربويين. حيث تعتمد دقة القرارات التربوية في مجال التعليم على صحة المعلومات المستمدة من الاختبارات. لذا فإن هذه الدراسة تهدف إلى المقارنة بين النماذج اللوغاريتمية الثلاث لنظرية الاستجابة للمفردة في تدريج مفردات بنك للأسئلة الموضوعية في مادة الرياضيات للصف الخامس الابتدائي وذلك من حيث دقة القياس كما تتمثل في :

- عدد الاستجابات غير الملائمة.
- انخفاض قيمة الخطأ المعياري لصعوبة المفردات وقدرات الأفراد.
- ارتفاع مؤشرات الثبات لبنك الأسئلة.
- ارتفاع قيم دوال المعلومات.

مشكلة الدراسة :

في ضوء ما سبق ، وبمراجعة أدبيات البحث الخاصة بنظرية الاستجابة للمفردة والنماذج اللوغاريتمية المرتبطة بها نجد أن دقة التقديرات الإحصائية لبارامترات المفردات والأفراد اختلفت من

نموذج لآخر وذلك على النحو التالي:

- أثبتت بعض الدراسات تفوق النموذج الثلاثي البارامتر على النموذجين الأحادي والثنائي البارامتر مثل دراسة (Saba &Ying , 2004) التي توصلت إلى ارتفاع دقة النموذج الثلاثي وتفوقه على النموذج الأحادي البارامتر حيث ارتفعت قيم معاملات الثبات للنموذج ، وانخفضت قيم الأخطاء المعيارية الخاصة به ، كما أثبتت دراسة (Way & Reese 1991), تفوق النموذج اللوغاريتمي الثلاثي البارامتر على النموذجين الأحادي والثنائي البارامتر في قياس ومعادلة اختبارات التويفل فى اللغة الانجليزية.
- في حين أثبتت دراسة (Van Der Mass, Han L. J.,2011) تفوق النموذج الثنائي البارامتر على النماذج الأخرى في اتخاذ القرارات فيما يتعلق بالفروق الفردية ، كما أن هناك بعض الدراسات كدراسة (محمد الشافعي ،٢٠١١) توصلت إلى أن النموذج الثنائي البارامتر أفضل النماذج الثلاثة في تقديمه لأعلى تقديرات لقيم دوال معلومات فقرات الاختبار وهى نفس النتائج التي توصل إليها فيما يتعلق بمتغيري الخطأ المعياري لتقدير القدرة وثبات تقدير دوال المعلومات وذلك كمؤشرين لدقة تقديرات قدرات المتحنين ، وتوصلت كذلك دراسة (Courville,2004) إلى أن النموذج الثنائي كان أفضل النماذج الثلاث وأقلهم من حيث معدلات الخطأ.
- وهناك دراسات أخرى أثبتت أن النموذج الأحادي البارامتر (نموذج راش) هو النموذج الأكثر دقة في ضبط إحصاءات المفردة مقارنة بالنموذجين الثنائي والثلاثي البارامتر. كدراسة كل من (Robert,2010)، (Fan,2009) ، (Chang et al., 2001) .

وهكذا يتبين أن هناك تعارض في النتائج فيما يتعلق بدقة التقديرات الإحصائية لمعالم المفردات الاختبارية وكذلك معالم الأفراد (متمثلة في تقديرات القدرة) باختلاف النموذج اللوغاريتمي المستخدم ،

وعلى الرغم من أن هذه الدراسة لا تتجه نحو حسم هذا الخلاف فيما يتعلق بدقة تقدير معالم المفردات الاختبارية (الصعوبة والتمييز) ، إلا أن الدراسة الحالية تستهدف محاولة الكشف عن أدق النماذج اللوغاريتمية الثلاث لكي يمكن استخدامها في بناء وتدريج بنك أسئلة في الرياضيات.

وفى ضوء ذلك يمكن أن تتحدد مشكلة الدراسة الحالية في الأسئلة التالية :

١- هل تختلف ملاءمة بيانات بنك الأسئلة باختلاف نموذج التحليل المستخدم (أحادي ، ثنائي ، ثلاثي) البارمتر ؟

٢- هل تختلف دقة تقديرات صعوبة المفردات وقدرات الأفراد على بنك الأسئلة باختلاف نموذج التحليل المستخدم (أحادي ، ثنائي ، ثلاثي) البارمتر ؟

٣- ما تدريج وتقدير صعوبة المفردات وقدرات الأفراد في بنك الأسئلة وفقا للنموذج الاكثر دقة؟

٤- ما صدق وثبات بنك الأسئلة في صورته النهائية؟

أهمية الدراسة :

تبدو أهمية الدراسة من خلال النقاط التالية :

١- إن المقارنة بين ثلاثة نماذج لنظرية الاستجابة للمفردة في تدريج بنك للأسئلة يساعد على تحديد أدق هذه النماذج في القياس ، وهذا بدوره له قيمة علمية في مجال التطبيقات التي تستخدم بها تلك النماذج وبصفة خاصة فى مجالات القياس المختلفة.

٢- إن دراسة بعض العوامل التي تؤثر في كفاءة تدريج بنك الاسئلة ، يؤدى إلى التغلب على بعض المشكلات المتعلقة بتدريج بنوك الاسئلة.

٣- إن استخدام أحد الأساليب الحديثة في القياس السلوكي في بناء بنك للأسئلة الموضوعية يساعد في نشر فكرة بنوك الأسئلة وتبسيطها.

E-mail: gamel_abdo59@yahoo.com ☎ 01002702115

أهداف الدراسة :

تهدف الدراسة الحالية إلى المقارنة بين النماذج اللوغاريتمية الثلاث لنظرية الاستجابة للمفردة (الأحادي/الثنائي/الثلاثي) البارمتر في تدريج بنك للأسئلة الموضوعية في مادة الرياضيات للصف الخامس الابتدائي وذلك من حيث دقة القياس التي تتمثل في:

- **انخفاض قيمة الخطأ المعياري لصعوبة المفردات.**
- **انخفاض قيمة الخطأ المعياري لتقديرات قدرة الأفراد.**
- **ارتفاع مؤشرات الثبات لبنك الاسئلة.**
- **ارتفاع قيم دوال المعلومات**

محددات الدراسة :

اقتصرت الدراسة على ما يلى :

١- **ثلاثة نماذج لوغاريتمية من نماذج نظرية الاستجابة للمفردة (الأحادي ، الثنائي ، الثلاثي) البارامتر.**

٢- **وحدة الكسور العشرية والعمليات المختلفة بمقرر الرياضيات للصف الخامس الابتدائي.**

٣- **حقل واحد أو نوعية واحدة من المفردات وهى أسئلة الاختيار من متعدد.**

٤- **عينة من طلاب الصف الخامس الابتدائي من مرحلة التعليم الأساسي تم اختيارها من ٨ مدارس كعينة ممثلة لمحافظة السويس.**

مصطلحات الدراسة :

- **<u>نظرية الاستجابة للمفردة</u> :**

تفترض هذه النظرية أنه يمكن التنبؤ بأداء الأفراد أو يمكن تفسير أدائهم على اختبار نفسي أو تربوي معين، في ضوء خاصية أو خصائص مميزة لهذا الأداء تسمى سمات Traits **أو بمعنى وجود واحد أو أكثر من المميزات أو السمات الأساسية التي تحدد استجابات الفرد الملاحظة لمفردات اختبار ما ، وهذه السمات لا تلاحظ بصورة مباشرة ، ولكن تظهر من خلال أداء الأفراد والذي يمكن ملاحظته وقياسه بصورة**

مباشرة من خلال استجاباتهم على مجموعة من مفردات الاختبار (Hambleton & Swaminthan , 1985 : 6).

- **نماذج الاستجابة للمفردة اللوغاريتمية:**

هي مجموعة من النماذج الرياضية تستخدم لتحديد العلاقة بين أداء الفرد على الاختبار ، الذى يمكن ملاحظته وبين مجموعات السمات أو القدرات التي لا يمكن ملاحظتها والتي يفترض أنها تكمن خلف الأداء على الاختبار ، وتعتمد هذه النماذج اللوغاريتمية على مجموعة من الافتراضات القوية طبقا لنظرية الاحتمالات وهى : أحادية البعد ، استقلالية الموضع أو المحل ، المنحنى المميز للمفردة ، والتحرر من عامل السرعة(Hambleton & Swaminthan , 1985 :10 ; Stage , 2003 : 3) .

- **بنك الأسئلة:**

تعرف أمينة كاظم (١٩٨٨: ١٠٨) بنك الأسئلة بأنه " تجمع ضخم من بنود الأسئلة المثلة لمتغير ما وتدرج هذه البنود على تدريج واحد مشترك بصفر واحد مشترك وبوحدة قياس معرفة رياضيا. وفى الدراسة الحالية يقصد به " مجموعة من المفردات المثلة لمتغير التحصيل في محتوى وحدة الكسور العشرية والعمليات عليها من محتوى منهج مادة الرياضيات المقرر على تلاميذ الصف الخامس الابتدائي من مرحلة التعليم الأساسي بجمهورية مصر العربية والتي لا تختلف فيما بينها إلا من حيث الصعوبة والمدرجة على تدريج خطى بوحدة قياس معرفة بحيث تمثل مقياسا واحدا يغطى مدى واسعا من المتغير موضوع القياس (التحصيل). ويمكن استخدام هذه المفردات في سحب مجموعات من المفردات متعادلة القياس وفقا لمواصفات امتحانية معينة ، كما يمكن استخدام أي من هذه المفردات مرات متعددة ، كما يمكن إضافة مفردات جديدة يتم معايرتها وتدريجها بنفس وحدة القياس وعلى نفس تدريج البنك.

الإطار النظري والدراسات السابقة:

أولا: نظرية الاستجابة للمفردة Item Response Theory:

ساهمت نظرية الاستجابة للمفردة في إيجاد حلول لبعض مشكلات القياس الكلاسيكي وقصوره ، وحققت هذه النظرية إلى حد بعيد الموضوعية المنشودة والتي تتصف بها أدوات القياس المستخدمة في المجال الفيزيائي حيث تفترض النظرية أنه يمكن تفسير الأداء الملاحظ للأفراد على الاختبار بسمات أو

قدرات تميز الأفراد وهذه السمات لا يمكن قياسها بشكل مباشر . وينبثق تحت مظلة هذه النظرية العديد من النماذج الرياضية التي تصف هذه العلاقة بين الأداء الملاحظ للأفراد على الاختبار وتلك السمات الكامنة (Edward,2010 ; Courville,2004). وتقوم نظرية الاستجابة للمفردة على مسلمتين أساسيتين:

- أنه يمكن التنبؤ أو تفسير بأداء أي فرد في أي اختبار بواسطة مجموعة من العوامل يطلق عليها سمات أو قدرات كامنة .
- أنه يمكن وصف العلاقة بين أداء الأفراد على أي مفردة اختبارية ومجموعة السمات أو القدرات الكامنة التي يفترض أنها تؤثر فى أدائه على هذه المفردة بدالة طردية (مضطردة الازدياد) Monotonically increasing Function يطلق عليها الدالة المميزة للمفردة Item Characteristic Function أو المنحنى المميز للمفردة، ومؤدى هذه الدالة أنه كلما ازداد مستوى السمة الكامنة ، يزداد بالتبعية احتمال الإجابة الصحيحة على المفردات. ويحقق استخدام نماذج نظرية الاستجابة للمفردة الاختبارية الكثير من المزايا في مقابل أساليب القياس الكلاسيكية، من أهمها:

١- استقلالية القياس ، وتعنى:

أ. استقلال تقديرات الأفراد Person Measures المشتقة من عينة كبيرة ومناسبة من المفردات الاختبارية عن خصائص تلك المفردات الاختبارية المستخدمة.

ب. استقلال تقديرات المفردات Item Measures المشتقة من عينة كبيرة ومناسبة من المفحوصين عن خصائص هؤلاء الأفراد المستخدمين فى عملية التدريج.

٢- إن عملية التدريج التي تشتمل عليها نظرية الاستجابة للمفردة تسفر عن وضع تقديرات الأفراد وتقديرات المفردات على مقياس خطى واحد ، بما يتيح عمل مقارنات مستقلة عبر المجموعات المختلفة من الأفراد والمجموعات المختلفة من المفردات.

٣- تعطى نظرية الاستجابة للمفردة إحصاءات لدقة القياس عند كل نقطة على طول متصل السمة الكامنة المقاسة ، مما يساعد على تحديد الدقة التي يقيس بها الاختبار كافة مستويات السمة ، وهو ما قد يتيح:

- تحديد الكيفية التي تتباين بها دقة القياس عبر مستويات القدرة المختلفة بالنسبة لاختبار ما.
- إنشاء اختبار يتسم بثبات دقة القياس عبر مدى واسع من مستويات قدرة المفحوصين.
- إنشاء اختبار يتعاظم تمييزه فى نطاق ضيق جدا من قدرات الأفراد(مثلا ،بالقرب من الحد الأدنى من الدرجات المطلوب لاجتياز اختبار إجازة معين أو منح شهادة ما).

٤- تقدم نماذج نظرية الاستجابة للمفردة مجموعة من الآليات للمساعدة في تقييم فئات الاستجابة في نموذج مقياس التقدير ، وخاصة الفئة المحايدة Neutral Category ، ومن ثم تحديد الكيفية المثلى للاستجابات ، فالفئة المحايدة قد تصبح فئة غير وظيفية عندما لا تقع على نفس ميزان تدرج باقي الفئات - تبعا لما تعكسه من مستويات السمة المقاسة - إذ تعكس عدم اكتراث المفحوص او عدم وجود رأى لديه ، وهذا يدل على عدم صدق تلك الاستجابات فى التعبير عن مستوى السمة لدى هؤلاء الأفراد.

٥- تتيح نماذج النظرية—من خلال إحصاءات ملاءمة الأفراد- الكشف عن الأفراد غير الصادقين فى استجاباتهم على الاختبار (بسبب عوامل الإهمال ، أو التخمين، أو الغش ، ... إلخ)، حيث أن مثل تلك العوامل قد تؤدى إلى سوء تقدير قدرات الأفراد (تقديرات غير حقيقية).

٦- تقدم النظرية حلا شافيا لمشكلة الاختبارات الموائمة ، حيث تتيح النظرية استخدام مجموعات مختلفة من المفردات لتقدير الأفراد المختلفين في مستوى السمة ، ويقدم لكل فرد مجموعة من المفردات تتناسب صعوبتها مع مستوى قدرته مما يمكن من تقدير قدرة كل فرد بشكل أكثر دقة.

٧- وإضافة إلى ما سبق ، فإنه يمكن استخدام أداة القياس المعدة وفقا لنظرية الاستجابة للمفردة

للجمع بين أهداف القياس جماعي ومحكي المرجع ، وتحقيق غير ذلك من أهداف.

نماذج نظرية الاستجابة للمفردة:

توجد العديد من النماذج المختلفة لنظرية الاستجابة للمفردة ، وتلك النماذج تستخدم مجموعة دوال رياضية لتحديد العلاقة المتوقعة بين الأداء الملاحظ للمفحوصين على الاختبار والقدرات الكامنة التي يفترض أنها تكمن وراء الأداء على الاختبار. وفى أي تطبيق عملي لتلك النماذج يجب تعريف الصيغة الرياضية للمنحنيات المميزة للمفردة والحصول على تقديرات بارامترات المفردة المطلوبة لوصف تلك المنحنيات، وتوجد نماذج تقيس قدرة واحدة يطلق عليها "النماذج أحادية البعد" ونماذج أخرى تفترض وجود أكثر من قدرة لحساب الأداء على الاختبار وهى "النماذج متعددة الأبعاد" وتلك النماذج أكثر تعقيدا ولم تتطور مثلما تطورت النماذج أحادية البعد (Stage,2003:3-26)، وهناك طرق عديدة لتصنيف النماذج المتعددة لنظرية الاستجابة للمفردة من أهمها:

١- النماذج أحادية البعد Unidimentional IRT Models مقابل النماذج متعددة الأبعاد Multi*dimentional IRT Models*.

٢- النماذج التجميعية الطبيعية Normal Ogive IRT Models مقابل النماذج اللوغاريتمية Logistic IRT Models.

٣- نماذج الاستجابات الثنائية Binary IRT Models مقابل النماذج متعدد الاستجابة Polytomous IRT Models ، ونماذج الاستجابات المتصلة Continuous IRT Models .

٤- النماذج الاستاتيكية Static IRT Models مقابل النماذج الديناميكية Dynamic IRT Models.

٥- التصنيف تبعا لعدد معالم أو بارامترات المفردة وهو التصنيف الأكثر شيوعا ، حيث تنقسم نماذج نظرية الاستجابة تبعا لعدد المعالم التي تصف المفردة إلى ثلاث نماذج أساسية هي :

١- **النموذج أحادي البارامتر:** One-Parameter Logistic Model

يُعد النموذج الأحادي البارامتر One Parameter أبسط نماذج الاستجابة للمفردة وأكثرها

استخداماً في بناء الاختبارات وبنوك الأسئلة وتحليل مفرداتها ، ويرجع الفضل في بناء هذا النموذج إلى عالم الرياضيات الدنماركي جورج راش G. Rasch. وقد طوعه للتطبيق العملي العالم الأمريكي بنجامين رايت B. Wright ، ويرى راش أن استجابة الفرد (v) على المفردة (i) وهى (Xvi) تحدد وفقا لمتغيرين أساسيين هما متغير قدرة الفرد ، ومتغير صعوبة المفردة ، وأنه إذا تزايدت قدرة الفرد عن مستوى صعوبة المفردة فإن الاحتمال الأكبر حينئذ هو حدوث الاستجابة الصواب ، أما إذا حدث العكس فمن المتوقع حدوث الاستجابة الخطأ ، وحيث أن الظواهر السلوكية تتسم بعدم الثبات ، فإن هناك صعوبة للتنبؤ بشكل قاطع في شكل الاستجابة الصادرة عن الفرد ، وبالتالي فإن العلاقة بين قدرة الفرد وصعوبة المفردة يجب ألا تكون علاقة قاطعة، بل هي علاقة احتمالية (Azevedo,2009). وتتمثل الصيغة الرياضية لهذا النموذج كما في (Mislevy,1990:195-197) كما يلى:

$$P_i(\theta)=\frac{1}{1+e^{-D(\theta-b_i)}}, i=1,2,3,...n$$

- $P(\theta)$ احتمال الاستجابة الصواب للفرد الذى قدرته (θ) على المفردة i.
- θ مقدار السمة لدى الفرد كما يقيسها الاختبار.
- b_i صعوبة المفردة i.
- e الأساس اللوغاريتمي الطبيعي وهو يساوى ٢,٧١٨ تقريباً.
- n عدد المفردات في الاختبار.
- D معامل القياس أو التدريج وهو مقدار ثابت يساوى ١,٧ أو ١,٧٠٢ ، ويستخدم هذا العامل ليؤكد أن هذه الدالة اللوغاريتمية هي تقريب حقيقي للدالة التجميعية الطبيعية لهذا النموذج، وبدون هذا العامل تصبح الدالة اللوغاريتمية تحويلاً خطياً للدالة التجميعية الطبيعية

ومن خصائص النموذج أحادي البارامتر أنه يتعامل مع عدد أقل من متغيرات المفردة ، كما أنه يقوم على اعتبار أن جميع مفردات الاختبار تتساوى تقريباً من حيث قوة التمييز، وهو بذلك يفترض أن

E-mail: gamel_abdo59@yahoo.com ☎ 01002702115

التمييز يساوى واحداً صحيحاً (a=1) في جميع المفردات مع اعتبار الخطأ المعياري ، ومن ثم فإنه يخرجه من حسابه عند تقدير صعوبات المفردات وكذلك الأمر بخصوص التخمين حيث يعتبر النموذج أن التخمين في حالته هذه أقل ما يكون. وهو بذلك يضع المسئولية على الباحث في أن يضع مفردات اختباره بحيث يراعى فيها على قدر المستطاع قلة التخمين كما يجب عليه أيضاً عند التطبيق أن يتأكد من ضبط الموقف الاختباري الذى قد يسهم كذلك في ضمان قلة التخمين أو الغش الذى قد يحاوله المتحنون / الأفراد.

٢- **النموذج الثنائي البارامتر:** Two - Parameter Logistic Model

يعتبر لورد (١٩٥٢) أول من قدم النموذج ثنائي البارامتر (الصعوبة والتمييز) بالاعتماد على المنحنى التجميعي الطبيعي ؛ بينما قام بيرنبوم (١٩٦٨) بتطويره ، ويفترض هذا النموذج عدم تأثر الاستجابات على المفردات بعامل التخمين . وتتقاطع - في هذا النموذج - المنحنيات المميزة لمفردات الاختبار التي تختلف في صعوبتها وتمييزها عبر المستويات المختلفة للسمة (De Ayala, 2009 : 99-109)٠ وعلى هذا يقوم هذا النموذج بالتنبؤ باحتمال الاستجابة الصواب للفرد على مفردة اختبارية ما كدالة للتفاعل بين مستوى خاصية واحدة لدى الفرد (معلم واحد للفرد) وخاصيتين أو معلمين للمفردة هما معلم الصعوبة ومعلم التمييز، في حين يفترض انعدام معامل التخمين c أي انه يساوى صفراً. وبناء على ذلك يقوم النموذج بتقدير صعوبة المفردة b ومعامل التمييز a، وتتمثل الصيغة الرياضية لهذا النموذج كما في (Mislevy,1990 :195-197) كما يلى:

E-mail: gamel_abdo59@yahoo.com ☎ 01002702115

$$P_i(\theta)=\frac{1}{1+e^{-Da_i(\theta-b_i)}}, i=1,2,3,...n$$

P (θ) احتمال الاستجابة الصواب للفرد الذى قدرته (θ) على المفردة i.

θ مقدار السمة لدى الفرد كما يقيسها الاختبار.

b_i صعوبة المفردة i.

a المنحنى المميز للمفردة ICC عند نقطة انقلابه أو معلم التمييز.

e الأساس اللوغاريتمي الطبيعي وهو يساوى ٢,٧١٨ تقريباً.

n عدد المفردات في الاختبار.

D معامل القياس أو التدريج وهو مقدار ثابت يساوى ١,٧ أو١,٧٠٢، ويستخدم هذا العامل ليؤكد أن هذه الدالة اللوغاريتمية هي تقريب حقيقي للدالة التجميعية الطبيعية لهذا النموذج، وبدون هذا العامل تصبح الدالة اللوغاريتمية تحويلاً خطياً للدالة التجميعية الطبيعية

ويذكر محمد الشافعي (٢٠١١ : ٥٥٩) أن هناك دراسات أثبتت دقة النموذج ثنائي البارمتر في تقدير الخصائص الإحصائية للمفردات الاختبارية كدراسة (Daniel,2010)**، ودراسة** (Raykov , Dimitrov , Asparouhov, 2010) **التي توصلت إلى تفوق النموذج ثنائي البارمتر في الوصول إلى أعلى تقدير للثبات إذا ما قورن بالنموذج الأحادي البارمتر للبيانات الاختبارية ثنائية التقسيم كما أعطى النموذج أقل تقدير للأخطاء المعيارية من نظائرها بالنموذج أحادي البارمتر، ودراسة** (Jaio,2003) **التي استهدفت تحديد معدلات الخطأ في تقديرات نماذج الاستجابة للمفردة الثلاثة (الأحادي والثنائي والثلاثي) البارمتر ، وفقا للاستجابات ثنائية التقسيم أن النموذج الثنائي كان أفضل النماذج الثلاث وأقلهم من حيث معدلات الخطأ.**

٣- **النموذج الثلاثي البارامتر:** Three - Parameter Logistic Model

ينسب هذا النموذج إلى بيرنبوم (١٩٦٨) الذي أضاف بارامتر التخمين Pseudo Chance level parameter **والذي يعبر عن احتمال توصل الأفراد ذوى السمة أو القدرة المنخفضة جداً إلى**

الاستجابة المناسبة على المفردة. لذا يقوم هذا النموذج بالتنبؤ باحتمال الاستجابة الصواب للفرد على مفردة اختبارية ما كدالة للتفاعل بين مستوى خاصية أو سمة واحدة للفرد (معلم واحد للفرد) ، وثلاث خصائص أو معالم للمفردة هي معلم الصعوبة (b) ،معلم التمييز (a) ، ومعلم التخمين (c) أو ما يطلق عليه معلم الحد الأدنى التقاربي. وعلى هذا يمكن النظر إلى النموذج اللوغاريتمي ثلاثي المعلم على أنه تعميم من النموذج اللوغاريتمي أحادي المعلم يسمح بتباين قوة تمييز المفردات في الاختبار ، كما يأخذ في الاعتبار إمكانية نجاح بعض الأفراد من منخفضي القدرة في الإجابة الصواب على بعض مفردات الاختبار التي تفوق قدرتهم عن طريق التخمين أو الصدفة العشوائية. وتتمثل الصيغة الرياضية لهذا النموذج كما في (Mislevy,1990:195-197) كما يلى:

$$P_i(\theta) = c_i + (1-c_i)\frac{1}{\left[1+\exp\left\{-Da_i(\theta-b_i)\right\}\right]}, i = 1,2,3,...n$$

- $P(\theta)$ احتمال الاستجابة الصواب للفرد الذى قدرته (θ) على المفردة i.
- θ مقدار السمة لدى الفرد كما يقيسها الاختبار.
- b_i صعوبة المفردة i.
- a المنحنى المميز للمفردة ICC عند نقطة انقلابه أو معلم التمييز.
- c الخط التقاربي الأسفل للمنحنى المميز للمفردة أو معلم التخمين.
- e الأساس اللوغاريتمي الطبيعي وهو يساوى ٢,٧١٨ تقريباً.
- n عدد المفردات في الاختبار.
- D معامل القياس أو التدريج وهو مقدار ثابت يساوى ١,٧ أو١,٧٠٢، ويستخدم هذا العامل ليؤكد أن هذه الدالة اللوغاريتمية هي تقريب حقيقي للدالة التجميعية الطبيعية لهذا النموذج، وبدون هذا العامل تصبح الدالة اللوغاريتمية تحويلاً خطياً للدالة التجميعية الطبيعية

ويعد النموذج الثلاثي من أكثر النماذج شهرة وانتشارا في الآونة الأخيرة ، ويعتقد البعض أنه أفضل نماذج نظرية الاستجابة للمفردة على الإطلاق حيث يتيح أكبر عدد من البارامترات. إلا أن (Harris,1989:159) يرى أنه من الخطأ الاعتقاد في ذلك ، حيث أن كثرة عدد البارامترات يضعف من دقة تقدير البارامترات نفسها ، بينما يرى (Wright,1992:196-200) أن هذا النموذج يتمركز حول البيانات Data –Centered فهو مصمم ليتفق مع أي شكل من أشكال البيانات حيث يقترب منها محاولا تطويع نفسه وفقا لطبيعة بيانات الاختبار. كما توصلت العديد من الدراسات التي استهدفت المقارنة بين نظرية الاستجابة للمفردة (باستخدام النموذج الثلاثي البارمتر) والنظرية الكلاسيكية للاختبارات إلى فاعلية هذا النموذج ودقته في مجال تقدير الخصائص الإحصائية للمفردات وفى مجال تقديرات القدرة للأفراد كدراسة (Aimee et.al ,2008) التي توصلت إلى أهمية استخدام النموذج اللوغاريتمي ثلاثي البارامتر نظرا لدقة هذا النموذج إذا ما قورن بكل من النموذجين الأحادي والثنائي البارامتر في انتقاء مفردات اختبار القبول لكليات الطب بالجامعات الأمريكية، كما أثبتت دراسة عبد الناصر العكايلة (٢٠٠٧) أن متوسط تقديرات الخطأ المعياري للقياس المقدر باستخدام النموذج الثلاثي البارامتر أدنى من متوسط الخطأ المعياري للقياس المقدر بالنظرية الكلاسيكية ، مما يشير إلى الدقة العالية لتقدير النموذج الثلاثي البارامتر بالمقارنة بدقة تقديرات الطرق الكلاسيكية ، كما أثبتت دراسة (De Mars,2001) أن هناك أفضلية لاستخدام النموذج الثلاثي البارامتر لمعرفة الفروقات الجماعية المبنية على نظرية الاستجابة للمفردة إذا ما قورن بكل من النموذجين (الأحادي والثنائي) .

مؤشرات دقة القياس في ضوء نظرية الاستجابة للمفردة :

حيث أن الهدف الرئيسي لهذه الدراسة هو المقارنة بين النماذج اللوغاريتمية الثلاث (الأحادية، الثنائية ، الثلاثية) البارمتر من حيث دقة القياس في تدريج بنك للأسئلة الموضوعية في مادة الرياضيات ، فقد تبين من خلال مسح الأدبيات وجود عدد من المؤشرات التي تستخدم لقياس دقة تقدير الخصائص الإحصائية للمفردات والأفراد ، وأصبحت هذه المؤشرات بمثابة الدلالات المرجعية لتفضيل أسلوب عن الآخر، وتتمثل هذه المؤشرات في:

أ) الخطأ المعياري للقياس:

يستخدم الخطأ المعياري في وصف النطاق الذي تتذبذب فيه الدرجات المتوقعة والناتج عن الخطأ ، وتستخدم فترات الثقة المعرّفة بواسطة الخطأ المعياري في تفسير الدرجات. ومن خصائص الخطأ المعياري ما يلي:

- تتغير قيمة الخطأ المعياري لتقدير قدرات أفراد العينة الواحدة.
- تكون الأخطاء المعيارية أقل ما يمكن عندما تكون مجموعة المفردات المستخدمة ملائمة للقياس، كما يكون تمييز المفردات ملائماً أيضا. ويعرف الخطأ المعياري للقياس لاختبار Sm بأنه الانحراف المعياري لدرجات الخطأ للاختبار، ويعرف أيضاً بالانحراف المعياري لتوزيع درجات الخطأ، وتستخدم الصيغة الإحصائية الآتية لتقدير الخطأ المعياري:

$$Sm = SD\sqrt{1-r}$$

حيث r : معامل ثبات الاختبار ، SD : الانحراف المعياري للاختبار

ويلاحظ أنه كلما زاد الثبات كلما قل الخطأ المعياري للقياس عندئذ تقترب الدرجة الملاحظة التي حصل عليها الطالب من الدرجة الحقيقية له ، وكلما قل الثبات كلما زاد الخطأ المعياري للقياس عندئذ تبتعد الدرجة الملاحظة عن الدرجة الحقيقية وبالتالي يمكن القول بأن الخطأ المعياري للقياس هو قياس دقة درجات الاختبار فكلما زادت قيمة الخطأ المعياري للقياس كلما قلت دقة الدرجات والعكس صحيح (Kubiszyn & Borich , 2003 : 124-129). ويستخدم الخطأ المعياري للقياس في حساب ثبات أو تحديد مدى دقة كل درجة، ويتوقف خطأ القياس على طول فقرات الاختبار فكلما زاد طول الاختبار قل الخطأ وكذلك على نوعية الفقرات وعلى تباين المجموعة التي طبق عليها الاختبار،ويقل خطأ القياس إذا وجد تباين بين الأفراد (تيسير كوافحة،٢٠٠٣ :٨٨).

ويرى بابام Popham (٢٠٠٥ : ٤٢) أن الخطأ المعياري للقياس هو مؤشر يعكس اتساق الدرجات التي يمكن أن يحصل عليها الأفراد من خلال اختبارات تم تطبيقها أكثر من مرة ، فمن الناحية العملية يكون من الصعب إعادة نفس الاختبار على نفس الطلبة لمرات عديدة ،ومن هنا تظهر الحاجة إلى تفسير الاختلاف في درجات الاختبار التي يحصل عليها الطلاب وأهمية الخطأ المعياري للقياس. ومن وجهة نظر أخرى يرى رجاء أبو علام (١٩٩٨ :٤٣١) أن الخطأ المعياري ومعامل الثبات خاصيتان تكملان بعضهما البعض

في التعبير عن ثبات المقياس والدقة التي يمكن بها تحديد الدرجة الحقيقية للفرد والتي ترجع إلى حجم الخطأ المعياري الذي يتأثر بمعامل ثبات المقياس وعندما نبحث عن اختبار ما لاستخدامه في جمع البيانات يجب أن يكون معامل الثبات لهذا الاختبار مرتفعاً مع انخفاض الخطأ المعياري له .

ب) دوال معلومات المفردات والاختبار Item & Test Information Function :

تلعب دوال المعلومات دورا بارزا في نظرية الاستجابة للمفردة، وذلك انطلاقا من ارتباطها بدقة تقدير قدرة المفحوصين ومن ثم النجاح في إنجاز المهمة الأساسية لعملية القياس النفسي (Partchev,2004:12). ويرجع الفضل في إرساء المعنى الإحصائي لدالة المعلومات إلى Fisher ، الذى عرفها على أنها مقلوب قياس الدقة التي تتم بها تقدير معالم النموذج. ويعنى ذلك أنه كلما زادت دقة تقدير معلم ما ، زاد بالتبعية قدر المعلومات التي يقدمها هذا القياس عن المعلم أي قلت أخطاء القياس.

- دوال معلومات المفردة :

وتقدر دالة المعلومات للمفردة في النموذج أحادي البارامتر باستخدام الصيغة الرياضية التالية (Parchev,2004: 12 ; Hambleton , 2000) :

$$Ii(\theta) = Pi(\theta)\, Qi(\theta)$$

حيث:

$Pi(\theta)$ هو احتمال حدوث الإجابة الصواب على المفردة (i) للفرد الذى له مستوى القدرة (θ).

$Qi(\theta)$ ، $Qi(\theta) = 1 - Pi(\theta)$ هو احتمال الاستجابة الخطأ حيث (θ) هى مستوى القدرة للفرد موضع الاهتمام. أما في النموذج ثنائي البارمتر فتقدر دالة المعلومات لمفردات الاختبار بالصيغة التالية :

$$Ii(\theta) = ai^2\, Pi(\theta)\, Qi(\theta)$$

حيث : ai تشير إلى متغير تمييز المفردة (i) .

وفى حالة النموذج الثلاثي البارمتر يزيد في تقدير دالة المعلومات لمفردات الاختبار تقدير متغير التخمين (C) على المفردة ، ومن ثم تصبح الصيغة الرياضية لدالة المعلومات كالتالي :

E-mail: gamel_abdo59@yahoo.com ☎ 01002702115

$$Ii(\theta) = ai^2 \left[\frac{Qi(\theta)}{Pi(\theta)}\right]\left[\frac{Pi(\theta) - c^2}{1 - c^2}\right]$$

ويمكن تلخيص أبعاد هذه الصيغ في القواعد التالية:

١- **في النموذجين أحادي وثنائي المعلم نلاحظ أن مقدار المعلومات المستمدة من مفردة اختبارية يكون أكبر ما يمكن حول قيمة بارامتر صعوبة المفردة ولذلك فإنه لكى تسهم المفردات بأكبر قدر من المعلومات ينبغي أن تتمركز صعوبة كل مفردة حول قدرة الفرد التي يختبر بها.**

٢- **في النموذج ثلاثي البارامتر يتضح أن أكبر قدر من المعلومات التي تسهم به مفردة يحدث عند مستوى كامن أقل قليلا من قيمة معلم صعوبة المفردة وذلك بسبب تأثير معلم التخمين (Ci) في الاستجابات.**

٣- **مقدار المعلومات المستمدة من مفردة يتحدد بقيمة معلم تمييز المفردة. فكلما زادت هذه القيمة زاد مقدار المعلومات السيكومترية التي تسهم به المفردة حول معلم الصعوبة. بذلك تقدم منحنيات المعلومات قدر كبير من المعلومات في مدى ضيق لقيم السمة بينما القيمة المنخفضة لمعلم التمييز تجعل منحنى المعلومات أكثر انتشارا مما يقلل من المعلومات التي تسهم بها المفردة** (Embretson&Rise,2000:184).

- **<u>دوال معلومات الاختبار</u>:**

يمكن الحصول على مقدار المعلومات التي يقدمها الاختبار عند مستوى محدد من مستويات القدرة كمجموع لدوال معلومات مفرداته عند هذا المستوى من القدرة ؛ لذا يمكن تعريف دالة معلومات الاختبار رياضيا كما يلى:

$$I(\theta) = \sum_{i=1}^{N} Ii(\theta)$$

حيث:

E-mail: gamel_abdo59@yahoo.com ☎ 01002702115

$I(\theta)$: قدر المعلومات التي يقدمها الاختبار عن قدرة المفحوصين عند مستوى معين للقدرة θ .

$\sum_{i=1}^{N} Ii(\theta)$ ، : مجموع دوال معلومات المفردات من المفردة الأولى إلى المفردة الأخيرة N عند مستوى معين للقدرة θ .

وتبين المعادلة السابقة أن مفردات الاختبار تسهم بصورة مستقلة في دالة معلومات الاختبار ، وبناء على ذلك يمكن تحديد المساهمة الفردية لمفردات الاختبار بدون معرفة بقية مفردات الاختبار الأخرى وهذه الميزة غير متوفرة في النظرية الكلاسيكية للاختبار.

ثانيا : بنوك الأسئلة **Item Banks :**

تعتبر بنوك الأسئلة من أهم التطبيقات العملية لنماذج نظرية الاستجابة للمفردة. وتعتبر بنوك الأسئلة من المفاهيم التربوية المستحدثة منذ منتصف الستينيات وقد تطور مفهوم بنك الأسئلة بتطور أساليب القياس من مجرد مستودع لعدد كبير من الأسئلة Item Pool إلى مجموعة كبيرة من الأسئلة الجيدة المدرجة على تدريج واحد مشترك بصفر واحد مشترك. وتتعدد مسميات بنوك الأسئلة مثل : بنوك المفردات Item Banks ، أو مخزون الأسئلة question Reservoirs ، أو مجموعة مفردات Item Collections، أو قائمة أو ملف الأسئلة Item Pools ، أو مكتبة مفردات اختبارية Test Items Library، أو بنك المهام Tasks Bank ، وغيرها (محمد زيدان،٢٠٠٩ : ٢٤).

وتشير أدبيات القياس المعاصر إلى تعريفات متعددة لمفهوم بنوك الأسئلة إلا أنه لا يوجد اتفاق تام على تعريف بنوك الأسئلة نظرا للاختلاف حول الغرض من استخدامها. والمعايير المستخدمة في انتقاء المفردات أو الأسئلة التي ينبغي تخزينها في البنك، أو محكات الحكم على جودتها، إلا أنه يمكن تصنيف التعريفات المختلفة في مستويات ثلاثة كالتالي:

المستوى الأول : تخزين مفردات أو أسئلة يتم بناؤها بطريقة مناسبة، والتحقق من صدق محتواها.

المستوى الثاني : تخزين مفردات أو أسئلة منظمة ومفهرسة بحسب محتواها، وخصائصها السيكومترية التي يتم تحديدها استنادا إلى نماذج القياس الكلاسيكية.

المستوى الثالث (المستهدف في الدراسة الحالية): تخزين مفردات أو أسئلة بعد تعييرها وتدريجها باستخدام أحد نماذج الاستجابة المفردة Calibrated Item Bank.

E-mail: gamel_abdo59@yahoo.com ☎ 01002702115

وتعرف بنوك الأسئلة في ضوء نظرية الاستجابة للمفردة بأنها "مجموعة من مفردات الاختبار يكون لها خصائص سيكومترية مميزة ومعلومة، حيث يضم بنك الأسئلة عدة اختبارات ، تتدرج بنودها جميعاً في تدريج واحد مشترك بوحدة قياس معرفة، وصفر واحد مشترك ، بحيث تعرف مدى واسعاً من مستويات المتغير موضوع القياس (صلاح علام، ١٩٨٦: ١٥٧ ؛ أمينة كاظم ، ١٩٨٨: ١٠٨). كما تعرفها أمينة كاظم (١٩٨٨: ٨٩) على أنها : "تجمع ضخم من المفردات والأسئلة الممثلة لمتغير ما والتي لا تختلف فيما بينها إلا من حيث مستوى صعوبتها فقط ، وتتدرج هذه البنود من حيث مستوى صعوبتها بوحدة قياس معرفة على تدرج واحد مشترك بصفر واحد". ويعرف ماهر إسماعيل ومحب الرافعي (٢٠٠٨ :٣٧١) بنك الأسئلة بأنه :" دمج بنود اختبارين في تدريج واحد للحصول على شبكة من عدة اختبارات تغطي مدى واسع من المتغير المراد قياسه وتدرج المفردات بوحدة قياس معرفة وتقنن بخطوات معينة". وقد أوضحت أمينة كاظم (٢٠٠٠ :٣٢١- ٣٢٢) أهم أهداف واستخدامات بنوك الأسئلة فيما يأتي:

- تحقيق الموضوعية في تقويم التحصيل الدراسي للطلاب، ويعتبر هذا هو الهدف الأساسي لبنوك الأسئلة. ويعتمد تحقيق هذا الهدف على مجموعة من العوامل التي يمكن أن تؤثر على بناء بنك الأسئلة مثل فلسفة القياس المستخدمة ، ومدى الدقة في صياغة المفردات وتحليلها وتحديد مواصفاتها الإحصائية ومدى تغطيتها لمجال القياس من حيث المحتوى والمستوى وضبط الموقف الاختباري وغير ذلك من عوامل.
- إتاحة الفرصة لقياس المستويات المعرفية العليا التي طالما أغفلت في تقدير تحصيل الطلاب، وذلك نظراً لاتساع المدى الذي يمكن أن يغطيه من حيث محتوى المقرر أو من حيث المستويات المعرفية.
- توفير وقت وجهد المعلم في بناء الاختبارات التحصيلية وذلك بحصوله على اختبارات جيدة من بنوك الأسئلة طبقاً لمواصفات وأهداف المقرر.
- تخفيض عوامل القلق والرهبة والتوتر المصاحبة لموقف الامتحان وذلك باستخدام عينة من مفردات البنك للتقويم المستمر أثناء العام الدراسي حتى يألف الطلاب هذا النوع من الأسئلة الموضوعية.

- مرونة القياس حيث يسهل تشكيل أي اختبار في أي وقت بسرعة وسهولة مما يسهل عملية التقويم المستمر.
- سرعة الحصول على صور متكافئة ومتعادلة من الاختبارات تبعاً لمواصفات المفردات المكونة للبنك مما يساعد على التغلب على عوامل الألفة بالاختبار عند تقويم البرامج السيكولوجية والتربوية وغيرها.
- إمكانية المقارنة الموضوعية بين مستوى تحصيل الطلاب مهما اختلفت الاختبارات المستخدمة ومهما اختلفت مستويات المجموعات التي ينتمون إليها ومهما اختلفت الأعوام ، طالما أن المفردات قد تم تدريجها ومعايرتها ووضعها في بنك واحد وذلك تبعاً لما يسمى بعملية التعادل Equating.
- المساعدة في تطوير العملية التعليمية وذلك لما تحظى به بنوك الأسئلة من مميزات متعددة.

وقد تعددت استخدامات بنوك الأسئلة – كأحد التوجهات المستقبلية في التقويم – وأصبحت أهدافها متسعة في الميدان التربوي؛ لما لها من مزايا . لعل أهمها ما يلي:

١. تكمن الميزة الأولية لبنوك الأسئلة في تطوير الاختبارات، حيث يمكن وضع مفردات الاختيار من متعدد على ميزان عام وذلك باستخدام نماذج الاستجابة للمفردة مثل نموذج "راش" ويشير الميزان إلى الصعوبة النسبية للمفردات ويمكن تكوين الاختبارات والاختبارات الفرعية (الجزئية) المتعادلة القياس عن طريق سحب مجموعات من المفردات المكونة للبنك.
٢. السماح بإيداع مفردات إضافية جديدة لإثراء بنك الأسئلة ، ويمكن عمل إيداع ضخم بدمج بنك الأسئلة مع آخر، كذلك يمكن عمل إيداع صغير بإضافة القليل من المفردات. ويتضمن النوع الأول إجراء تعادل بين موازين المفردات المختلفة ، بينما يتضمن النوع الثاني عدد أقل من المفردات للمفحوصين في مستويات الصفوف المختلفة.
٣. تعمل بنوك الأسئلة على توفير الوقت والجهد في إعداد الاختبارات التحصيلية بصورتها التقليدية، ويمكن تحديد الصعوبة النسبية للمفردات تبعاً لمستويات الصفوف ، وعند التقويم لا يحتاج الفرد إلا للسحب من مفردات البنك.

٤. تساعد بنوك الأسئلة في بناء لغة لمناقشة غايات المنهج والأهداف التعليمية وتصف المفردات المهام الفردية للطلاب، كذلك إن موقع المفردات على ميزان تم تعييره يسمح للفرد بأن يحدد الصعوبة النسبية لمهام محددة وهذا يمهد طرق بناء مناهج جيدة (Rudner & Lawrence,1998:240).

وبذلك يسهم هذا النظام البنكي للأسئلة في تطوير عمليات قياس وتقويم التحصيل الدراسي وجعلها أكثر مرونة وفاعلية ودقة، وكذلك جعلها أكثر تقنينا مما هي عليه في إطار الممارسات الحالية، كما يسهم في تزويد المربين ببيانات تتعلق بالتقدم الدراسي للطلاب في وقت قصير. وبذلك يكون البنك مكونة رئيسية من مكونات المنظومة التعليمية، ويسهم إسهاما فاعلا في تطويرها، كما تصبح عمليتا التعليم والتقويم عملية واحدة متكاملة، وهو ما يسعى إليه خبراء القياس المعاصر. ونظرا لأهمية إنشاء نظام بنك أسئلة يسهم في تطوير أساليب قياس وتقويم التحصيل الدراسي، حدد (صلاح الدين علام ، ٢٠٠٠) المراحل التي يمر بها إنشاء هذا النظام البنكي المقترح فيما يلي:

١. مرحلة التخطيط والتنسيق.
٢. مرحلة التدريب والإعداد.
٣. مرحلة تصميم الأسئلة وبنائها.
٤. مرحلة التجريب والتعيير.
٥. مرحلة التصنيف والتخزين.
٦. مرحلة الاستخدام والتقييم للبنك.

هذا وتختلف طريقة بناء بنك الأسئلة باختلاف أسلوب القياس المتبع في بنائه ، ويمكن تلخيص الخطوات الإجرائية لبناء بنك الأسئلة في ظل أساليب القياس المختلفة فيما يلي :

١- تحديد محتوى المقرر الدراسي الذي يخدمه البنك وتحليله.

٢- صياغة الأهداف التعليمية وتصنيفها في جدول مواصفات تبعاً لمحتوى المقرر الدراسي وتبعاً للمستويات المعرفية المختلفة.

٣- صياغة أو تجميع المفردات أو الأسئلة التي تغطي جميع جوانب المقرر وأهدافه التقويمية، كما تغطي المستويات المختلفة للأهداف المعرفية ويشترك في ذلك مجموعة من المتخصصين في المادة الدراسية.

٤- مراجعة المفردات وتكوين مجموعات منها، ويقوم بذلك لجنة من الخبراء المتخصصين.

٥- تطبيق مجموعات المفردات على عينات مناسبة من الأفراد.

٦- تحليل البيانات وانتقاء المفردات الصالحة للقياس، وتختلف الأساليب التقليدية للقياس عن نماذج الاستجابة للمفردة في ذلك، وإن كانت الفكرة في الحالتين واحدة وهى انتقاء أجود المفردات إحصائياً في حدود صدق المحتوى وخصائص القياس المطلوبة.

٧- تحديد الخواص الإحصائية اللازمة ، تبعاً لأسلوب القياس المتبع التقليدي (جماعي/ محكي المرجع) ونماذج الاستجابة للمفردة.

٨- تكوين بنك الأسئلة من مجموعات المفردات الصالحة للقياس، ويتم ذلك تبعاً لأسلوب القياس المتبع.

٩- تخزين وإدارة البنك.

١٠- متابعة وتطوير البنك وإثرائه (صلاح علام ، 1989 : 7-14 ؛ Green et al. , 1989 : 189-199).

ولا تختلف خطوات بناء بنك الأسئلة باستخدام نظرية الاستجابة للمفردة عن بناء مستودع الأسئلة بالطرق التقليدية إلا في خطوة تكوين الصور الاختبارية وتطبيقها ويبدأ الاختلاف في :

١) تحليل البيانات وانتقاء المفردات الصالحة للقياس وذلك بعد حذف الأفراد والمفردات غير الملائمة (لأسس القياس الموضوعي) واستبقاء تلك التي تحقق شروط النموذج. وجدير بالذكر أن هناك طرقاً حديثة للتعرف على خصائص المفردات ومحاولة الحكم على جودتها قبل إجراء التحليل ومنها طريقة التحليل البياني للمفردات (Graphical Item Analysis(GIA والتي استعان بها الباحثان في التعرف على خصائص المفردات الاختبارية قبل إجراء التحليل.

E-mail: gamel_abdo59@yahoo.com ☎ 01002702115

٢) تدريج كل مجموعة من المفردات تدريجاً خطياً على ميزان تدريج مستقل بوحدة القياس الخاصة بالنموذج (اللوجيت) بحيث يكون صفر تدريج كل مجموعة من المفردات هو متوسط صعوبتها. وباستخدام إما مجموعة من المفردات المشتركة أو الأفراد المشتركين تتم معادلة تدريجي المجموعتين لدمجهما في تدريج واحد مشترك وبتكرار هذه العملية يتم توحيد ميزان تدريج جميع المفردات (أمينة كاظم ، ١٩٩٥ : ٥٧- ٥٨ ؛ Baker , 2001 : 157). وتوفر برامج الحاسب الآلي Rumm , Microscale , Winsteps, Parscale طرقاً سهلة وبسيطة لتدريج المفردات.

٣) تخزين وإدارة البنك وتستخدم في ذلك بعض البرامج الخاصة مثل برنامج Fast.

٤) متابعة وتطوير البنك.

يتضح مما سبق أن بنك الأسئلة الذي يتم بناؤه باستخدام نماذج الاستجابة للمفردة يتميز بنفس ما تتميز به النماذج من استقلالية القياس ، ويتيح ذلك انتقاء مجموعات من المفردات تناسب المستويات المختلفة، كما يتيح انتقاء مجموعات متكافئة من المفردات تناسب مجموعات الأفراد المتشابهة في مستوى قدرتها. وتذكر أمينة كاظم(١٩٩٥: ٢٥٨- ٢٥٩) أن الاختبار يكون مناسباً للفرد أو مجموعة الأفراد الذين يؤدونه عندما تقترب الخواص الإحصائية للاختبار متمثلة في طول الاختبار ومتوسطه ومدى صعوبته من الخواص التي يتصف بها الأفراد.

وأخيرا نشير إلى أنه "على الرغم من تعدد التساؤلات المتعلقة بتصميم بنوك الأسئلة ، ومنهجياتها ، ومشكلاتها، إلا أن كثيرا من خبراء القياس المعاصر يؤكدون أهميتها، بل وضرورتها، والاستمرار في إنشائها، والاستفادة منها في تطوير نظم وأساليب القياس والتقويم والامتحانات. غير أن هذا يتطلب مزيدا من البحث للتغلب على كثير من مشكلات هذه البنوك، وبخاصة البنوك المحوسبة التي يتم تعيير مفرداتها استنادا إلى أحد نماذج الاستجابة للمفردة.

الدراسات السابقة :

تنوعت الدراسات الأجنبية والعربية التي تناولت نماذج نظرية الاستجابة للمفردة ويمكن تصنيفها وفقا لمتغيرات الدراسة الحالية على النحو التالي:

أولاً : دراسات تناولت المقارنة بين نماذج الاستجابة للمفردة المختلفة.
ثانياً : دراسات استخدمت نظرية الاستجابة للمفردة في بناء الاختبارات التحصيلية وبنوك الأسئلة.

أولا : <u>دراسات تناولت المقارنة بين نماذج الاستجابة للمفردة المختلفة</u> :

ومن هذه الدراسات دراسة (Way & Reese,1991) **التي هدفت إلى استخدام نموذجي الاستجابة للمفردة الأحادي والثنائي البارمتر في تدريج ومعادلة اختبارات التويفل ومقارنة نتائج التعادل والتدريج الخاصة بهذين النموذجين بالنتائج الخاصة بالنموذج ثلاثي البارمتر المستخدم بصورة أساسية في تدريج اختبارات التويفل. وقد اعتمدت الدراسة على بيانات أربعة عينات مختلفة الأحجام هي على الترتيب** 600 ، 900 ، 1200 ، 1500 **فرد ، بينما وصل حجم العينة الكلية للبيانات إلى** 2400 ، 3600 ، 4800 ،6000 **فرد. وتم استخدام برنامج** Logist **فى التحليل لعمل المقارنات بين النماذج باستخدام أحجام مختلفة لكل قسم من أقسام التويفل الثلاث. وأشارت أهم نتائج الدراسة إلى أن النموذج الثلاثي البارمتر أدى بصورة أفضل من النموذجين الأحادي والثنائي البارمتر فيما يتعلق بملاءمة البيانات للنموذج طبقا لإحصاء البواقي المعيارية المطلقة** (ASRs)**، حيث اختلفت نتائج تلك الإحصاءات بين النماذج الثلاث طبقا لحجم العينة ، وتأثرت ملاءمة النموذجين الأحادي والثنائي البارامتر بزيادة حجم العينة باستخدام عينة مكونة من** 600 **مفحوص ، في حين لم تتأثر جودة الملاءمة للنموذج الثلاثي البارامتر بحجم العينة.**

وبحثت دراسة (Tam Hak & Li Yuan ,1997) **ما إذا كانت فروق أرجحية إحصاءات مربع كا**^{2} **لمقارنة نماذج نظرية الاستجابة للمفردة تتوزع وفقا لتوزيع مربع كا**^{2}**، وبحث معدل دقة فروق أرجحية إحصاءات مربع كا**^{2} **في اختيار نماذج الاستجابة للمفردة الثلاث (الأحادي والثنائي والثلاثي) البارمتر. وقد استخدمت الدراسة برنامج** Bilog **في تقدير بارامترات المفردات من خلال تحليل استجابات عينة الدراسة التي تمثلت في** 200 **استجابة لبيانات اختبار تم توليدهم من بنك أسئلة. ومن أهم النتائج التي أسفرت عنها الدراسة أن بارامترات الصعوبة والتخمين لم يتم تقديرها بصورة دقيقة وفقا للنموذج ثلاثي البارمتر حيث نتج عن استخدامه وجود فروق سلبية لإحصاءات مربع كا**^{2} **وتلك الفروق لم**

تحدث عند طول اختبار = 50 مفردة لأن بارامترات المفردة تم تقديرها بصورة دقيقة ، كما أعطى النموذج الثلاثي البارمتر معدل خطأ أعلى من مثيله عند تقديره بالنموذجين الثنائي والأحادي.

في حين هدفت دراسة (Chang , Hanson & Harris, 2001) إلى كشف طرق التقنين ومقارنة فعالية استخدام النموذجين الأحادي والثلاثي البارمتر في ضبط إحصاءات المفردة الأولية مع العينات الصغيرة. تمثلت عينة الدراسة في 10 مجموعات من المفحوصين تم توليدهم من 1000 & 2500 مفحوص خلال كل المجموعات بمعدل 100 & 250 مفحوص لكل مجموعة ، وتم بناء 10 صور اختبارية متوازية في المحتوى والمواصفات الإحصائية بحيث تكونت كل صورة من 24 مفردة و12 مفردة مشتركة، وتم استخدام برنامج Bilog في عملية التحليل .وكانت من بين النتائج التي توصلت إليها النتائج أن النموذج أحادي البارمتر أدى بصورة جيدة مع البيانات الحقيقية للاختبار الأولى لضبط إحصاءات المفردة.

وكان من أهداف دراسة (علاء الدين أيوب ،٢٠٠١) اختبار دقة نماذج الاستجابة للمفردة (أحادي ، ثنائي ، ثلاثي) البارمتر كما يتمثل في الخطأ المعياري لتقدير الصعوبة والقدرة – دوال المعلومات- مؤشرات الثبات في تقدير وتدريج صعوبة مفردات اختبارات المهارات المعرفية وتقدير قدرة الأفراد. تكونت عينة الدراسة الأساسية من 1874 طالب وطالبة من الفرق (الأولي ، الثانية ، الثالثة ، الرابعة) من طلاب الجامعة بمحافظتي أسوان وقنا ، وتم استخدام برنامج Bilog في عملية التحليل. وقد بينت نتائج الدراسة تفوق النموذج ثلاثي البارامتر على النموذجين (أحادي ، ثنائي) البارامتر فى دقة تقدير صعوبة المفردات وقدرة الأفراد ، كما يتمثل ذلك في قلة الخطأ المعياري لتقدير الصعوبة وتقدير القدرة على أبعاد مقياس المهارات المعرفية (المتسلسلات ، المتماثلات ، التذكر، الاستدلال اللفظي). وهى نفس النتائج التي توصل إليها فيما يتعلق بتقدير دوال المعلومات وتقدير مؤشرات الثبات ، حيث تبين أن النموذج ثلاثي البارامتر أدت تحليلاته إلي ارتفاع تقدير دوال المعلومات وارتفاع تقدير مؤشرات الثبات على أبعاد المقياس.

وبحثت دراسة (أمين نور الدين ،٢٠٠٢) فاعلية الاختبار التكيفي المحوسب في تقدير قدرة الأفراد ، كما تتمثل في الخصائص السيكومترية للقياس مقارنة بأساليب أخرى للاختبار. وقد بلغ حجم

العينة الكلية 5049 طالبا وطالبة من طلاب قسم علم النفس التربوي بكلية التربية جامعة عين شمس، في فصلين دراسيين متتابعين من العام الدراسي 2000/2001 ، وفى مختلف التخصصات والفرق الدراسية. ولقد قام الباحث بإعداد 138 مفردة من مختلف إصدارات اختبارات المصفوفات المتتابعة لرافن (الملونة - العادية – المتقدمة) قسمت إلى اختبارين بحيث طبق كل اختبار على مجموعة مختلفة من المفحوصين ثم إجراء ربط بين الاختبارين ليصبح اختبارا واحدا. ولقد قام الباحث باختبار مطابقة استجابات المفحوصين لأحد نماذج نظرية الاستجابة للمفردة (الأحادي ، الثنائي ، الثلاثي) البارامتر من خلال عدد من المؤشرات (مقدار للمعلومات - معامل الثبات - عدد المفردات غير الملائمة –قيم معامل التمييز) . وكان من بين النتائج التي بينتها الدراسة تفوق النموذج الثلاثي البارامتر على النموذجين الأحادي والثنائي البارمتر في ارتفاع تقدير كل من : دوال المعلومات ، ومؤشرات الثبات بالنسبة للاختبارين الأول والثاني ، كما كان متوسط معلم التمييز للنموذج الثلاثي البارمتر أعلي منها في حالة النموذج الثنائي البارمتر.

كما هدفت دراسة (إسماعيل الوليلي،٢٠٠٢) إلى التعرف على مدى فاعلية ثلاثة من نماذج الاستجابة للمفردة (الأحادي ، الثنائي ، الثلاثي) البارمتر والمقارنة بينها في دقة تصميم وانتقاء مفردات اختبار مرجعي المحك في وحدة الأعداد الحقيقية المقررة ضمن منهج الجبر لطلاب الصف الثالث الإعدادي وكذلك تحليل وتفسير نتائجه باستخدام هذه النماذج ، وتم استخدام برنامج Bilog في عملية التحليل. تكونت عينة الدراسة من 1800 طالبا وطالبة من طلاب الصف الثالث الإعدادي تم اختيارهم من 10 مدارس إعدادية عامة ، و10 معاهد إعدادية أزهرية بمحافظتي الشرقية والقاهرة في الفصل الدراسي الأول من العام الدراسي 2001 / 2000 . وكان من أهم ما توصلت إليه الدراسة هو أن النموذج الثلاثي البارمتر هو الأكثر دقة في انتقاء البيانات مقارنة بالنموذجين الآخرين، حيث أعطى النموذج الثلاثي البارامتر أكبر مقدار من المعلومات عن المفردات ، كما تفوق تمييز المفردات في حالة النموذجين الآخرين .

كذلك هدفت دراسة (Saba Rizavi &Ying Ly,2004) إلى تقويم النماذج اللوغاريتمية (الأحادية ، الثنائية ، الثلاثية) البارمتر المستخدمة في بناء الاختبارات الموائمة المحوسبة

CAT لتكون موازية مع اختبارات الورقة والقلم والخاصة بالتفكير اللفظي ، وقد استخدمت الدراسة مستودع مفردات يتكون من 8 صور اختبارية بالورقة والقلم للاختبارات الموائمة المحوسبة للتفكير المنطقي ، وكل صورة تم تعييرها باستخدام النماذج اللوغاريتمية الأحادية ، والثنائية ، والثلاثية البارمتر وتم استخدام برنامج التحليل Parscale لتحليل استجابات 500 فرد تم توليدهم في 20 درجة حقيقية صحيحة. وقد بينت أهم نتائج الدراسة أن النموذج أحادي البارمتر كان أقل ثباتا نسبيا من ثبات النموذجين الثنائي والثلاثي البارمتر ، كما جاءت قيم الخطأ المعياري للقياس وفقا للنموذج أحادي البارمتر أعلى منه في حالة النموذجين الثنائي والثلاثي البارمتر. كذلك أدى استخدام النموذج الأحادي البارمتر إلى عدم تكافؤ توزيعات القدرة ومعالم الصعوبة بينما كان الفرق بين تقدير معالم الصعوبة وتوزيع القدرة بالنسبة للنموذج الثلاثي البارمتر أصغر من النموذجين الثنائي والأحادي. كما وجدت الدراسة أن النموذج الثنائي البارمتر أبسط استخداما من النموذج الثلاثي البارمتر.

وقام (نبيل النجار ، ٢٠٠٦) ببناء بنك أسئلة في الثقافة الحاسوبية للمرحلة الثانوية بالأردن باستخدام نماذج نظرية الاستجابة للمفردة وقد صمم ست نماذج اختبارية كل نموذج يتكون من(60) مفردة بينها (10) مفردات مشتركة. تكونت عينة الدراسة من (1800) طالبا وطالبة .وقد بينت نتائج الدراسة عدم وجود اختلاف جوهري بين الخصائص السيكومترية للفقرات (معالم الصعوبة والتمييز) التي تم اختيارها عند بناء البنك باستخدام النموذج أحادي البارمتر عنها في النموذج الثنائي البارمتر ، بينما وجدت فروق جوهرية من حيث الخصائص السيكومترية للفقرات التي تم استبعادها عند بناء البنك باستخدام النموذجين الأحادي والثنائي حيث بلغت نسبة الاختلاف بين النموذجين (20%) بمتوسط (12) من حيث عدد الفقرات التي تم رفضها. كما وجدت الدراسة تشابه جوهري في الخصائص السيكومترية للفقرات المشتركة التي تم اختيارها عند بناء بنك الأسئلة باستخدام النموذجين.

وهدفت دراسة (Custer , Sharairi & Yamazaki, 2008) إلى المقارنة بين النماذج اللوغاريتمية الأحادي والثلاثي البارمتر في ثبات المفردات وتقديرات القدرة ومدي ملاءمة البيانات وذلك خلال مجموعة مستويات مختلفة من القدرة (منخفضة ، متوسطة ، مرتفعة). وتمثلت بيانات الدراسة في الصورة S لاختبار الكلمات لبوابة MacGinitie (GMRT) في اختبارات القراءة

واشتملت العينة على (2070) من طلاب الصف الثاني إلى طلاب الصف السادس بحيث يغطي طلاب كل صف المستويات المختلفة للقدرة. وباستخدام برنامج التحليل الإحصائي Parscale حللت البيانات وفقاً للنماذج اللوغاريتمية الأحادي والثلاثي البارمتر. وبينت أهم نتائج الدراسة وجود فروق في حجم التأثير بين مجموعات القدرة الثلاث في كل صف. كما توصلت إلى أن النموذج الثلاثي البارمتر أكثر ملاءمة للبيانات من النموذج الأحادي البارمتر.

وهدفت دراسة (نسرين إسماعيل ،٢٠١٠) إلى اكتشاف العلاقة بين بعض العوامل المتعلقة بنظرية الاستجابة للمفردة والتي تتمثل في (النموذج المستخدم في التدريج - حجم العينة المستخدمة فى التدريج – عدد المفردات المشتركة الرابطة – مستوى صعوبة المفردات المشتركة الرابطة) وبين دقة القياس كما تتمثل في (ملاءمة البيانات للنموذج المستخدم – مؤشرات الثبات – دالة معلومات الاختبار- الخطأ المعياري لتقدير صعوبة المفردات وقدرة الأفراد). وتكونت عينة الدراسة من 1045 تلميذا وتلميذة من تلاميذ 31 فصل من فصول الصف الثاني الإعدادي ، والتي أخذت من 8 مدارس تابعة لإدارة مدينة نصر التعليمية بالقاهرة ، وتمثلت أداة الدراسة فى بنك أسئلة فى موضوع المعادلات الجبرية المتضمن فى مقرر جبر الصف الثاني الإعدادي ، وقد تكون الاختبار فى صورته النهائية من 79 مفردة، وتم تحليل البيانات باستخدام البارامتر برنامج Parscale. وكان من بين ما توصلت إليه الدراسة تفوق النموذج الثلاثي البارامتر على النموذجين الأحادي والثنائي في معظم مؤشرات الدقة المحددة في الدراسة.

واستهدفت دراسة (محمد الشافعي ،٢٠١١) مقارنة تقديرات دوال معلومات مفردات اختبار المصفوفات المتتابعة (العادية) "لجون رافن " وهى (أقصى قيمة للمعلومات – نقاط القدرة المقابلة لأقصى قيم للمعلومات – متوسط قيم المعلومات) الناتجة عن تحليلات نماذج مرتبطة بنظرية الاستجابة للمفردة (الأحادي- الثنائي – الثلاثي) البارمتر، وعلاقة تقديرات دوال معلومات مفردات الاختبار بدقة تقديرات القدرة الناتجة عن التحليل باستخدام النماذج الثلاثة موضع المقارنة. وقد استخدمت الدراسة في التحليلات الخاصة عينة من طلاب خمس مدارس ثانوية تابعة لمنطقة جنوب القاهرة التعليمية بلغ حجمها (858) طالبا من جميع الصفوف الدراسية (الأول – حتى الثالث)، واستخدمت الدراسة أسلوب تحليل القياس المتكرر في المعالجات الإحصائية الخاصة باختبار فرضيات الدراسة. وقد توصلت الدراسة

إلى أن النموذج ثنائي البارامتر أفضل النماذج الثلاثة في تقديمه لأعلى تقديرات لقيم دوال معلومات فقرات الاختبار يليه النموذج ثلاثي البارامتر ثم أحادي البارامتر ، وهى نفس النتائج التي توصل إليها فيما يتعلق بمتغيري الخطأ المعياري لتقدير القدرة وثبات تقدير دوال المعلومات وذلك كمؤشرين لدقة تقدير قدرات المتحنين. حيث تبين أن النموذج ثنائي البارامتر أدت تحليلاته إلى أقل تقدير للأخطاء لمعيارية لتقديرات القدرة كما أن تقديراته لثبات قيم معلومات المفردات كانت أعلى من مثيلاتها المقدرة بالنموذجين الآخرين.

ثانيا : <u>دراسات استخدمت نظرية الاستجابة للمفردة في بناء الاختبارات التحصيلية وبنوك الأسئلة</u> :

ومنها دراسة (محمد الشافعي ، ١٩٩٦) التي هدفت إلى التعرف على مدى تأثير بعض المتغيرات على إحكام تدريج بنك للأسئلة وهذه المتغيرات هي: طرق المعادلة ، عدد المفردات المشتركة ، مستوى صعوبة المفردات للمتغير المقاس، تمثيل المفردات المشتركة للمتغير المقاس ، حجم وتمثيل العينة للمجتمع الأصل ، عدد مفردات الاختبارين موضوع التعادل (تساوي أو عدم تساوي طولي الاختبارين)، وتقارب وتباعد قدرة أفراد العينة من مستوى صعوبة الاختبار. كما هدفت الدراسة أيضاً إلى تقديم برنامج مقترح للتغلب على بعض المشكلات المتعلقة بتدريج بنوك الأسئلة. وتكونت عينة الدراسة من ١٤٠٥ طالب يمثلون مجموع الطلاب الدارسين لمقرر جبر الصف الثاني الثانوي في بعض محافظات مصر، وتمثلت أدوات الدراسة في اختبار كلي يتكون في صورته النهائية من ١٥٠ مفردة ، كما تم تكوين مجموعات من المفردات ذات خصائص مختلفة من مفردات الاختبار الكلي لاستخدامها في اختبار صحة فروض الدراسة. ومن أهم النتائج التي أسفرت عنها الدراسة عدم تأثر تدريج بنك الأسئلة بطرق المعادلة المختلفة والمتعلقة بنظرية الاستجابة للمفردة. عدم تأثر تدريج بنك الأسئلة (باستخدام طريقة المفردات المشتركة) بمتغيرات (عدد المفردات المشتركة – نسبة تمثيلها للمتغير المقاس– المستوى المتوسط لصعوبة المفردات المشتركة)، و(حجم عينات التحليل المستخدمة في طريقة المفردات المشتركة أو الأفراد المشتركين). كما تحققت موضوعية تقدير القدرة لنفس مجموعة الأفراد المقابلة لدرجاتهم الكلية على اختبارين مختلفي الصعوبة ، يقيسان نفس المتغير (سحبت مفرداتهما من نفس البنك) حيث لم تتأثر تقديرات القدرة المتوقعة من هذين الاختبارين لنفس مجموعة الطلاب أي أن تقديرات القدرة لهذه المجموعة للاختبار السهل تتكافأ مع تقدير القدرة لنفس المجموعة للاختبار الصعب.

ودراسة (Lawson,2006) التي بحث مدى إمكانية تطبيق نظرية الاستجابة للمفردة على

الاختبارات الصغيرة في الحجرة الدراسية. تمثلت عينة الدراسة في ١٠٠ طالب تم اختبارهم في اختبارين يتكونا من ١٠٠ مفردة ، ويتضمن الاختبار الثاني ٢٠ مفردة مشتركة مع الاختبار الأول، وتم تحليل الاختبارين باستخدام النموذج الأحادي البارامتر(نموذج راش) بصورة فردية وبعد ذلك تم تحليلهم بعد دمجهم معاً. وبينت أهم نتائج الدراسة أن درجات الطلاب تمتعت بدرجة عالية من الثبات وكانت أكثر ارتباطاً لأكثر من ٤٠٪ من المفردات التي تم حذفها من الاختبار، وبينت النتائج أيضاً وجود ارتباط عالي بين تحليل نموذج راش ودرجات الطلاب عند دمج الاختبارين معاً وأنه يمكن وضع الأفراد على نفس ميزان القياس عندما يحدث تداخل بنسبة ٢٠٪ من المفردات بين الاختبارات.

كذلك هدفت دراسة (عاشم عثمان ، ٢٠٠٦) إلى بناء بنك أسئلة لمبحث الرياضيات للصف الثاني الثانوي باستخدام نظرية الاستجابة للمفردة وفقا للنموذج ثلاثي المعلم وتكونت عينة الدراسة من (٨٠٠) طالبا وطالبة موزعين على مدارس عمان وتم تطبيق ثماني نماذج متكافئة احتوى كل نموذج منها على (٤٠) مفردة بواقع عشر مفردات مشتركة بينها. وقد أظهرت نتائج التحليل عدم مطابقة مفردتين من مفردات الاختبار للنموذج الثلاثي المعلم .كما تحققت الدراسة من تحرر تقديرات القدرة من معالم المفردات

كما هدفت دراسة (جواهر الزيد ، ٢٠٠٨) إلى الكشف عن فاعلية نظرية الاستجابة للمفردة في بناء بنك أسئلة في مقرر علم النفس التربوي وفقا لمؤشر الصعوبة في نموذج راش. طبقت أداة الدراسة على عينة مكونة من (١٧٤٠) طالبة من طالبات الكليات التربوية بجامعة الملك سعود والامام محمد بن سعود وكليات البنات التابعة لوزارة التربية والتعليم. واستخدمت الدراسة برنامج Microscale لتحليل المفردات الاختبارية باستخدام نموذج راش وتوصلت الدراسة إلى تحقق بعض فرضيات النظرية كأحادية البعد واستقلالية القياس وفق مؤشرات الملاءمة التقاربية والتباعدية . في حين انخفضت بعض الدلالات على تحقق تلك الفروض وفقا لنتائج التحليل العاملي .كما توصلت الدراسة إلى التدريج النهائي لمفردات بنك الأسئلة وفق مؤشرات الصعوبة لنموذج راش.

وهدفت دراسة (حازم بقاعين ،٢٠١٠) إلى بناء بنك أسئلة في الرياضيات للصف الثاني الثانوي للفروع :الأدبي، الإدارة المعلوماتية، الصحي، والشرعي في الأردن، ولتحقيق هذا الهدف تم إعداد تجمع من الفقرات بلغ عدده (٢٠٠) فقرة تم توزيعهم إلى (٤) صور اختبارية طبقت على عينة استطلاعية مكونة من (١٠٠) طالبا وطالبة. تم حساب الصدق باستخدام معدلات الطلبة الفصلية للرياضيات كمحك، وثبات الاختبار بطريقة إعادة الاختبار، ومعاملات الصعوبة والتمييز وفقا للنظرية الكلاسيكية، ثم

طبقت الصور الاختبارية على عينة مكونة من (٤٠٠) طالب وطالبة، وأجري التحليل العاملي للتأكد من أحادية البعد وتشبع الفقرات غلى العامل الرئيسي، حيث حذف عدد من الفقرات وأصبح الاختبار في صورته النهائية مكونا من (١٣٢) فقرة. وتم استخراج معالم الصعوبة، التمييز، التخمين للفقرات، دالة معلومات الفقرة والانحراف المعياري لخطأ تقدير كل منها وفقا لنظرية الاستجابة للمفردة وخلصت النتائج إلى بناء بنك أسئلة في الرياضيات تم حوسبته للاستفادة منه.

أما دراسة (هند الحموري ، ٢٠١١) فهدفت إلي استقصاء إمكان استخدام نماذج نظرية الاستجابة للمفردة في بناء اختبار تحصيلي من إعداد المعلم . تم بناء اختبار تحصيلي مكون من (٢٣) فقرة من نمط الاختيار من أربعة بدائل تقيس العدد نفسه من الأهداف المقصود. طبق الاختبار على (٢٨٤) طالبا وطالبة . وحللت البيانات الناتجة من الاستجابة للفقرات وفق نظرية الاستجابة للمفردة وكشفت نتائج الدراسة إلي عدم إمكان استخدام هذه النظرية لبناء الاختبارات ذات العدد القليل من الفقرات .

وقد استخدمت دراسة (تقى العديلات ، ٢٠١٢) النموذج الثلاثي المعلم في بناء بنك للأسئلة في مقرر الرياضيات لطلبة الصف الرابع. حيث صممت ثلاثة اختبارات من نوع الاختيار من متعدد وتكون كل منها من (٤٢) فقرة بواقع (١٢) فقرة مشتركة بين النماذج الثلاث. وتم تطبيق الاختبارات على عينة مكونة من (١٢٠٠) طالبا وطالبة. وحللت البيانات باستخدام برنامج Bilog-MG3. وقد بينت نتائج التحليل أم (٢٥) مفردة لم تطابق النموذج الثلاثي المعلم ، كما تم تقدير معالم المفردات وحساب دالة المعلومات للمفردات والخطأ المعياري للقياس.

وهدفت دراسة (حمدي أبو جراد ، ٢٠١٢) إلى المقارنة بين الخصائص السيكومترية لاختبار الاختبار من متعدد واختبار التكميل في ضوء نظرية الاستجابة للمفردة ولتحقيق غرض الدراسة تم إعداد اختبارين في مادة العلوم لطلبة الصف السادس الابتدائي أحدهما اختيار من متعدد والآخر من نوع التكميل وكلاهما مكون من (٣٠) مفردة. وباستخدام نموذج راش تم تدريج ومعايرة مفردات كلا الاختبارين على عينة مكونة من (٨٠) طالبا وطالبة والحصول على تقديرات الصعوبة مقدرة بوحدة اللوجيت. وكشفت نتائج الدراسة عن حذف (٧) مفردات من اختبار التكميل و(٦) مفردات من اختبار الاختيار من متعدد. وأشارت النتائج أيضا إلى أن معاملات ثبات تقديرات الصعوبة للمفردات والأفراد في اختبار التكميل أعلى منها لاختبار الاختبار من متعدد.

وهدفت دراسة (هالة مندور، محمد سيد أحمد & نسرين عاشور،٢٠١٣) إلى بناء بنك للأسئلة

لقياس تحصيل طلاب كلية التربية الرياضية لمقرر مادة الهوكي باستخدام نظرية الاستجابة للمفردة. وتمثلت أداة الدراسة في مقياس مكون من (٢٨٥) مفردة من نوع الاختيار من متعدد، وبعد حذف المفردات غير الملائمة بلغ عدد المفردات (٢٤٦) للأبعاد التالية: التطور التاريخي (١٨)، الاعداد البدني (٣٣)، الإعداد المهارى (٩)، المهارات الحركية (٤٥)، طرق التدريس (١٧)، خطط اللعب (٩) ، القانون (١١٥). وتكونت عينة التدريج من (٢٥٤٠) طالب وطالبة من الفرقة الثالثة والرابعة لكليات التربية الرياضية الذين أتموا دراسة مقرر الهوكي تم اختيارهم بطريقة عشوائية من بعض الجامعات المصرية الحكومية (الاسكندرية- المنصورة- الزقازيق- طنطا- بورسعيد). وقد توصلت الدراسة الي أداة قياس موضوعية لقياس وتقويم التحصيل الدراسي لطلاب وطالبات كليات التربية الرياضية في مقرر الهوكي من خلال استخدام نظرية الاستجابة للمفردة تستخدم في تصنيف الطلاب بحسب إتقانهم للمهارات المحددة بما يسمح باتخاذ قرارات تعليمية مناسبة بشأنهم . كما توافر عدة صور اختبارية متكافئة متعادلة القياس تستخدم في القياسات المتعددة تتمتع بدرجة عالية من الصدق والثبات كما تغطي مدى مناسبا من تقديرات الأفراد والمفردات، وتستخدم في تقدير تحصيل الطلاب في مقرر الهوكي على مدار أعوام دراسية متتالية مع ضمان تعادل وموضوعية القياس وتكافؤ الفرص من عام لأخر كون تلك الصور مسحوبة من نفس البنك .

وهدفت دراسة (أحمد الزغيلات ، ٢٠١٤) إلى التعرف على أثر عدد البدائل على دقة تقديرات مؤشرات القدرة للأفراد ودالة المعلومات للاختبار، حيث تم استخدام نظرية الاستجابة للمفردة في تحليل البيانات المستمدة من تطبيق ثلاث صور اختبارية متباينة في عدد البدائل، طبقت الدراسة على عينة عشوائية عنقودية على مستوى الصف من طلبة الصف الثاني الثانوي العلمي في محافظة الكرك، حيث تكونت عينة الدراسة من (١٠٥٠) طالباً وطالبة، تم اختيارهم من خلال تحديد المدارس الثانوية التي يتوفر فيها الصف الثاني الثانوي العلمي والمختارين لمادة الكيمياء، وقد بلغت عينة كل نموذج (٣٥٠) طالباً وطالبة. تمكنت الدراسة من تقدير مؤشرات القدرة للأفراد بطريقة مستقلة عن قدرات الأفراد، كما وفرت أداة قياس يمكن الاعتماد عليها في قياس تحصيل الطلبة في مادة الكيمياء للصف الثاني الثانوي العلمي في الأردن من قبل المعلمين والجهات الرسمية القائمة والمهتمة في هذا المجال.

واستخدمت دراسة (صبري الطراونة ، ٢٠١٦) نظرية الاستجابة للمفردة في بناء اختبار محكي المرجع في الإحصاء التحليلي على عينة مكونة من (٣١٩) طالبا من طلبة الدراسات العليا في كليات العلوم التربوية في الجامعات الأردنية الحكومية . وقد تم التحقق من مطابقة عينة الدراسة وفقرات الاختبار لافتراضات نموذج راش ، كما تم التحقق من افتراضات نظرية الاستجابة للمفردة، وتم حساب معاملات الصعوبة لفقرات الاختبار باستخدام برنامج Bigsteps، وتكون الاختبار في صورته النهائية من (٣٦) مفردة وأظهرت النتائج تمتع الاختبار بدلالات صدق وثبات مقبولة.

التعليق على الدراسات السابقة:

من العرض السابق للدراسات السابقة يتضح ما يلي:

- قامت العديد من الدراسات الأجنبية بالمقارنة بين النماذج اللوغاريتمية لنظرية الاستجابة للمفردة (الأحادي/الثنائي/الثلاثي) البارامتر لتحديد النموذج الأكثر دقة والأكثر ملاءمة لبيانات هذه الدراسات، ولقد تباينت المحكات التي حددتها هذه الدراسات لتحديد النموذج الأدق في القياس ومن هذه المحكات: عدد المفردات غير الملائمة للنموذج المستخدم، دالة معلومات الاختبار، انخفاض قيمة الخطأ المعياري لصعوبة المفردات وقدرة الأفراد، ارتفاع مؤشرات الثبات. وتوصلت معظم نتائج هذه الدراسات إلى تفوق النموذج اللوغاريتمي الثلاثي البارامتر على النماذج اللوغاريتمية (الأحادي والثنائي البارامتر) بوجه عام طبقاً للمحكات السابق ذكرها.
- على الرغم من توافر العديد من الدراسات الأجنبية التي اهتمت بالمقارنة بين نماذج نظرية الاستجابة للمفردة المختلفة في مجالات متعددة ومنها الاختبارات التحصيلية المختلفة إلا أن هناك نُدرة في الدراسات العربية المختلفة المرتبطة بالمقارنة بين نماذج نظرية الاستجابة للمفردة المختلفة.
- تضاربت نتائج الدراسات فيما يتعلق بدقة التقديرات الإحصائية لمعالم المفردات الاختبارية وكذلك معالم الأفراد (متمثلة في تقديرات القدرة) باختلاف النموذج اللوغاريتمي المستخدم (الأحادي ، الثنائي ، والثلاثي). فهناك دراسات أثبتت تفوق النموذج الثلاثي البارمتر على النموذجين الثنائي والأحادي البارمتر مثل دراسة (Saba &Ying , 2004) ، ودراسة (Wilberrg,2004) ، ودراسة (نسرين اسماعيل ، ٢٠١٠). في حين تفوق النموذج الثنائي

البارامتر على النماذج الأخرى في دراسة (Courville , 2004) ، ودراسة (Van Der Mass & Han L. J., 2011) ، ودراسة (محمد الشافعي ،٢٠١١). وهناك دراسات أخرى أثبتت أن النموذج الأحادي البارامتر (نموذج راش) هو النموذج الأكثر دقة في ضبط إحصاءات المفردة مقارنة بالنموذجين الثنائي والثلاثي البارامتر كدراسة كل من (Robert , 2010) ، (Fan , 2009) ، (Chang et al., 2001) .

- على الرغم من توافر العديد من الدراسات الأجنبية والعربية التي اهتمت باستخدام وتطبيق نماذج الاستجابة للمفردة في بناء بنوك الأسئلة والاختبارات المختلفة بأحجام مختلفة من العينات إلا أنه توجد نُدرة في الدراسات العربية المرتبطة ببنوك الأسئلة والتي تناولت دقة نماذج الاستجابة للمفردة (أحادي ، ثنائي ، ثلاثي) البارمتر.
- أشارت الدراسات السابقة المتعلقة باستخدام نظرية الاستجابة للمفردة بصفة عامة ونموذج راش بصفة خاصة إلى أهمية نموذج راش في تكوين بنوك الأسئلة المختلفة وتدريج وبناء الاختبارات التحصيلية المختلفة واستخدامه في مجالات متباينة وتحقيق موضوعية القياس واستقلاليته عن كل من مستوى العينة التي تؤدى الاختبار ومستوى مجموعة المفردات المتضمنة في الاختبار، ومن ثم إمكانية المقارنة بين مستوى تحصيل الطلاب بغض النظر عن الصورة الاختبارية التي أختبر بها الطالب طالما أن هذه الصور الاختبارية مسحوبة من بنك أسئلة واحد وتم تدريجها على ميزان واحد مشترك وبصفر واحد مشترك ،كما يمكن أيضاً مقارنة أداء الطالب اللاحق بأدائه السابق وغيرها من أهداف القياس.

خطة وإجراءات الدراسة :

أولاً : <u>منهج الدراسة ومتغيراتها</u> :

تعتمد الدراسة الحالية على المنهج شبه التجريبي حيث يتمثل المتغير المستقل في نوع النموذج المستخدم من ضمن ثلاث نماذج لوغاريتمية : نموذج راش Rasch Model (أحادي البارامتر)، ونموذج لورد Lord Model (ثنائي البارامتر)، ونموذج بيرنبوم Birnbaum Model (ثلاثي البارامتر)، ويتمثل المتغير التابع في دقة قياس التحصيل الدراسي في مادة الرياضيات كما تتمثل في :

- عدد الاستجابات غير الملائمة.

E-mail: gamel_abdo59@yahoo.com ☎ 01002702115

- انخفاض قيمة الخطأ المعياري لتقديرات صعوبة المفردات وقدرة الأفراد.
- ارتفاع مؤشرات الثبات لبنك الأسئلة.
- ارتفاع قيم دوال المعلومات.

ثانياً : مجتمع وعينة الدراسة :

اختار الباحثان أفراد دراستهما من تلاميذ وتلميذات الصف الخامس الابتدائي من مرحلة التعليم الأساسي بمدارس اللغات التجريبية الرسمية ومدارس اللغات الخاصة في محافظة السويس في الفصل الدراسي الأول من العام الدراسي (٢٠٠٧- ٢٠٠٨) حيث بلغت عينة الدراسة ١٥٠٠ تلميذاً وتلميذة حيث تم سحب مجتمع العينة الدارسين للغات. وقد روعي في سحب عينة الدراسة أن تكون ممثلة لكافة المستويات الدراسية وكذلك كافة المستويات الثقافية والاجتماعية والاقتصادية التي ربما تؤثر في مستوى التحصيل الدراسي ، كما روعي أيضاً أن تتمثل أنواع التعليم في مدارس اللغات بنوعيه (رسمي- خاص) حتى تتنوع المستويات التعليمية وكذلك أيضاً أن تشمل الجنسين (البنين والبنات) وذلك لضرورة توافر قدر من التباين في القدرة بين الأفراد. ويوضح جدول (١) نوع التعليم وأعداد التلاميذ والتلميذات وفقاً لنوع التعليم الذين ينتمون إليه.

جدول (١)
أعداد البنين والبنات في عينة الدراسة ونوع التعليم الذي ينتمون إليه

نوع التعليم / الطلاب	بنين	بنات
لغات رسمي	438	517
لغات خاص	301	244
المجموع	739	761

وكان العمر الزمني لأفراد عينة الدراسة ١٠سنوات، ويرجع سبب اختيار الباحثان لتلاميذ الصف الخامس الابتدائي ليمثلوا عينة الدراسة الحالية إلى أنهم يمثلون مرحلة انتقالية في التعليم الأساسي التي يدرسون خلالها موضوعات ووحدات متكاملة تكسبهم مهارات ضرورية وأساسية تمثل أساساً هاماً للنجاح في الدراسة المستقبلية لكافة فروع الرياضيات ومن الضروري إتقانها والتمكن منها، ومن هذه الوحدات المهمة وحدة "الكسور العشرية والعمليات المختلفة عليها "، ويمكن من خلال الدراسة الحالية

بناء بنك أسئلة في تلك الوحدة الهامة في الرياضيات وتدريجه بنماذج الاستجابة للمفردة الثلاث (الأحادي والثنائي والثلاثي) البارمتر والتعرف على مدى دقتها في تدريج البنك والكشف عن الاستجابات غير الملائمة وتحديد قدرات أفراد العينة.

ثالثاً : أدوات الدراسة :

استخدم الباحثان في دراستهما الحالية 10 صور اختبارية موضوعية تتكون كل صورة من ١٥ مفردة اختبارية من نوع الاختيار من متعدد في مادة الرياضيات للصف الخامس الابتدائي في الوحدة الأولى في الفصل الدراسي الأول، وقد راعى الباحثان في إعداد هذه الصور وجود عدد من المفردات المشتركة (الرابطة) حيث بلغ عددها ٨ مفردات، بواقع ١٥٠ فرد لكل صورة اختبارية. وقد وضع الباحثان خطة لبناء تلك الصور الاختبارية التي تكون بنك الأسئلة Item Bank تتضمن الخطوات التالية:

١- تحديد الهدف من الصور الاختبارية (تحديد مجال القياس).
٢- تحليل محتوى الوحدة الدراسية وتشمل الخطوات الفرعية التالية :
   - تحديد الأهمية النسبية للموضوعات.
   - تحديد الأهداف الإجرائية الخاصة بكل موضوع من الموضوعات.
   - تحديد الأهمية النسبية لمستويات الأهداف.
٣- تحديد الأهداف الإجرائية التي تقيسها الصور.
٤- إعداد جدول المواصفات.
٥- صياغة عدد كبير من المفردات تقيس الأهداف الإجرائية وتصنيفها في ضوء جدول المواصفات تبعاً لأجزاء المحتوى والمستويات المعرفية (تذكر/فهم/تطبيق).
٦- تحكيم المفردات للوقوف على مدى صلاحيتها من الناحية الفنية والعلمية.
٧- تجريب المفردات استطلاعياً.
٨- التحليل التقليدي لمفردات الصور الاختبارية لحساب معاملات الصعوبة والتمييز وقوة تشتيت البدائل.
٩- حذف أو تعديل المفردات في ضوء نتيجة التحليل وتحديد الزمن المناسب للصور الاختبارية.
١٠- التحقق من صدق وثبات الاختبار في صورته الأولية.

رابعا : الخطوات الإجرائية للدراسة :

١- تحليل محتوى الوحدة الأولى في مقرر الرياضيات للصف الخامس الإبتدائى في الفصل الدراسي

الأول ، وتحديد الأهداف الإجرائية ومن ثم تفريغها في جدول المواصفات.

٢- إعداد وصياغة أكبر عدد ممكن من المفردات الموضوعية التي تصلح لقياس تلك الأهداف، وقد تمثلت هذه المفردات في ٨٢ مفردة اختبارية مستقلة أصبح عددها بعد التجريب الأولي ٧٨ مفردة اختبارية.

٣- تكوين ١٠ صور اختبارية وفقاً لجدول المواصفات بحيث تتألف كل صورة من ١٥ مفردة اختبارية من بينها ٨ مفردات مشتركة.

٤- تطبيق الصور الاختبارية على عينة الدراسة المكونة من ١٥٠٠ تلميذاً وتلميذة وكان زمن التطبيق حصة واحدة (٤٥) دقيقة.

٥- تصحيح استجابات التلاميذ باستخدام مفتاح تصحيح حيث كانت درجة أي مفردة إما (١) للاستجابة الصحيحة أو (صفر) للاستجابة الخاطئة .وقد تم ذلك باستخدام برامج الحزمة الإحصائية SPSS .

٦- حساب تشتت البدائل ومراجعة هذه القيم ، وتحديد بعض المفردات التي قد تحذف أثناء التدريج بسبب عيوب المشتتات.

٧- عمل التحليل البياني لمفردات بنك الأسئلة (GIA) بهدف التحقق من جودتها وتشخيص عيوب بعضها.

٨- تجهيز ملفات البيانات الخاصة بالتحليل الإحصائي وفقاً للنماذج اللوغاريتمية الثلاث مع مراعاة المفردات المشتركة في جميع الصور الاختبارية.

٩- التحليل المبدئي لبنك الأسئلة باستخدام برنامج بارسكيل Parscale وتم ذلك بهدف :

- حذف البيانات التامة والصفرية والأفراد غير الصادقين آلياً بواسطة برنامج التحليل علماً بأن هذا البرنامج صمم ليعمل على أساس فلسفة تؤمن بأهمية بقاء جميع الأفراد أثناء التدريج (بقدر الإمكان) حيث أن هذا هو واقع البيانات الذي لا يجب إغفاله.
- الكشف عن مدى ملاءمة البيانات لكل نموذج من النماذج الثلاث وذلك باستخدام دلالة إحصاءات كا^{2} ، حيث يتم حذف المفردات غير الملائمة والتي كانت قيمة كا^{2} الخاصة بها دالة إحصائياً.
- الكشف عن أكثر النماذج دقة في تدريج بنك الأسئلة وقد تم ذلك كالتالي :
- مقارنة قيم الأخطاء المعيارية لتقديرات صعوبة مفردات بنك الأسئلة الناتجة من كل نموذج من النماذج الثلاث.
- مقارنة قيم الأخطاء المعيارية لتقديرات قدرة الأفراد ففي بنك الأسئلة الناتجة من كل نموذج من النماذج الثلاث.

- الكشف عن مؤشرات الثبات لكل نموذج من النماذج الثلاث.
- حساب قيم دالة معلومات الاختبار(TIF)والكشف عن شكلها وفقاً لكل نموذج من النماذج الثلاث.

١٠- التدريج النهائي لمفردات بنك الأسئلة (الصور الاختبارية) بعد حذف الأفراد غير الصادقين والمفردات غير الملائمة وفقاً للنموذج الأكثر دقة ، وذلك بهدف :

- الحصول على تقدير بارامترات صعوبة وتمييز المفردات والخطأ المعياري لتقدير كل منها.
- الحصول على تقدير قدرات الأفراد علماً بأن تقدير قدرة الفرد دالة لعدد المفردات التي أجاب عليها صواباً وصعوبة وتمييز هذه المفردات.

١١- استخدام برنامج التحليل الإحصائي ونستبس Winsteps في رسم خريطة تعريف المتغير وذلك بعد تغذية البرنامج بصعوبات مفردات بنك الأسئلة.

خامساً: خطة المعالجة الإحصائية :

قام الباحثان باستخدام برنامج التحليل الإحصائي بارسكيل Parscale الإصدار 4.1 في تدريج الصور الاختبارية (بنك الأسئلة) وفقاً لنماذج الاستجابة للمفردة اللوغاريتمية (الأحادي والثنائي والثلاثي البارامتر)، وبرنامج SPSS الإصدار (10) في إجراء المعالجات الإحصائية اللازمة ، كما استخدمت الأساليب الإحصائية التالية :

- المتوسطات والانحرافات المعيارية لتقديرات صعوبة المفردات وقدرات الأفراد Means and Standard Deviation.
- تحليل التباين أحادي الاتجاه One Way ANOVA.
- اختبار " ت " لمقارنة فروق المتوسطات بين مجموعتين T.Test.
- اختبار حجم التأثير (مربع إيتا) Effect Size .
- اختبار توكى Tukey Test.

عرض نتائج الدراسة ومناقشتها:

أولاً- <u>النتائج الخاصة بالإجابة على السؤال الأول من أسئلة الدراسة</u>:

ينص السؤال الأول للدراسة على: " هل تختلف ملاءمة بيانات بنك الأسئلة باختلاف نموذج التحليل المستخدم (أحادي ، ثنائي ، ثلاثي) البارمتر" ؟ وتتضمن نتائج الإجابة عن هذا السؤال ما يلي:

- **الكشف عن مدى ملاءمة الأفراد والمفردات المكونة لبنك الأسئلة لكل نموذج من النماذج اللوغاريتمية الأحادي والثنائي والثلاثي البارمتر.**

قام الباحثان بتحليل البيانات الخاصة باستجابات أفراد عينة الدراسة على المفردات الاختبارية المكونة لبنك الأسئلة وفقاً للنموذج اللوغاريتمي أحادي البارامتر، والنموذج اللوغاريتمي ثنائي البارامتر، والنموذج اللوغاريتمي ثلاثي البارامتر وذلك باستخدام برنامج التحليل الإحصائي Parscale " بارسكيل " ، وذلك بهدف اختبار ملاءمة البيانات لكل نموذج من النماذج الثلاث كما تتمثل في ملاءمة (الأفراد والمفردات) ، وذلك تبعاً لإحصاء الملاءمة كا٢ الذي يوفره برنامج التحليل. ويوضح جدول (٢) ملخصاً لنتائج تحليل بنك الأسئلة. حيث يبين عدد مفردات الاختبار وعدد أفراد عينة التدريج وكذا متوسط تقديرات كل من صعوبات المفردات وقدرات الأفراد وفقاً لكل نموذج من النماذج اللوغاريتمية الثلاث (الأحادي والثنائي والثلاثي البارمتر).

جدول (٢) ملخص نتائج تحليل البيانات الخاصة بتدريج بنك الأسئلة وفقاً لكل نموذج من النماذج اللوغاريتمية الثلاث

التحليل / النموذج	التحليل الأول				التحليل الثاني			
	عدد الأفراد	عدد المفردات	متوسط التقديرات		عدد الأفراد	عدد المفردات	متوسط التقديرات	
			للصعوبة	للقدرة			للصعوبة	للقدرة
الأحادي البارمتر	1211	78	-1.419	-0.366	1033	67	-1.456	-0.610
الثنائي البارمتر	1211	78	-1.367	-0.378	955	66	-1.417	-0.492
الثلاثي البارمتر	١١٣٨	78	-0.643	-0.525	941	69	-0.671	-0.529

وقد أسفرت نتائج التحليل الأول وفقاً لكل نموذج من النماذج اللوغاريتمية الثلاث الأحادي

والثنائي والثلاثي البارمتر بواسطة برنامج بارسكيل Parscale كما يتضح من جدول (٢) عن ما يلي :

- بلغ عدد الأفراد غير الملائمين للقياس وفقاً للنموذج الأحادي البارمتر (٤٦٧) فرد بنسبة مئوية تصل إلى ٣١.١٣٪ من العدد الكلي للأفراد (١٥٠٠) فرد.
- وبلغ عدد الأفراد غير الملائمين للقياس وفقاً للنموذج الثنائي البارمتر (٥٤٥) فرد بنسبة مئوية تصل إلى ٣٦.٣٣٪ من العدد الكلي للأفراد (١٥٠٠) فرد.
- كما بلغ عدد الأفراد غير الملائمين للقياس وفقاً للنموذج الثلاثي البارمتر (٥٥٩) فرد بنسبة مئوية تصل إلى ٣٧.٣٧٪ من العدد الكلي للأفراد (١٥٠٠) فرد.

وفي مرحلة التحليل الثانية فقد استخدم اختبار حسن المطابقة كا^{2} ، ومستويات الاحتمال الخاصة به لتحديد وحذف المفردات غير الملائمة للقياس وفقا لكل نموذج من النماذج الثلاث، حيث تعتبر المفردة غير ملائمة للقياس وبالتالي ينبغي حذفها إذا كانت قيمة اختبار كا ^{2} الخاصة بها دالة عند مستوى (0.05) على الأقل. ويلاحظ من جدول (٢) ما يلي :

١- بلغ عدد المفردات غير الملائمة للقياس وفقاً للنموذج الأحادي البارمتر (١١) مفردة بنسبة مئوية تصل إلى ١٤.١٠٪ من العدد الكلي للمفردات (٧٨) مفردة.

٢- بلغ عدد المفردات غير الملائمة للقياس وفقاً للنموذج الثنائي البارامتر (١٢) مفردة بنسبة مئوية تصل إلى ١٥.٣٨٪ من العدد الكلي للمفردات (٧٨) مفردة.

٣- بلغ عدد المفردات غير الملائمة للقياس وفقاً للنموذج الثلاثي البارامتر (٩) مفردة بنسبة مئوية تصل إلى ١١.٥٤٪ من العدد الكلي للمفردات (٧٨) مفردة.

ويرى الباحثان أن البيانات كانت أكثر ملاءمة للنموذج اللوغاريتمي ثلاثي البارمتر حيث يستوعب النسبة الأكبر من المفردات الملائمة استناداً إلى مؤشر النسبة المئوية لعدد المفردات المحذوفة والتي تصل إلى ١١.٥٤٪ من المجموع الكلي لمفردات بنك الأسئلة البالغ عدده ٧٨ مفردة اختبارية حيث تم حذف ٩ مفردات فقط وذلك في المرحلة الأولى أثناء عملية التحليل ، حيث اشتمل بنك الأسئلة وفقاً للنموذج أحادي البارمتر على ٦٧ مفردة، بينما اشتمل بنك الأسئلة وفقاً للنموذج ثنائي البارامتر على ٦٦ مفردة ، في حين اشتمل بنك الأسئلة وفقاً للنموذج ثلاثي البارامتر على ٦٩ مفردة. وهذا يعني أن النموذج اللوغاريتمي ثلاثي البارامتر يتمتع بقدر من المرونة بما يكفي ليتلاءم مع عدد من المفردات يفي بتمثيل النطاق السلوكي على أفضل وجه ، وبما يفي بتحقيق الخصائص السيكومترية في بنك الأسئلة المؤلف من

هذه المفـردات، بينما جاء النموذج اللوغاريتمي أحادي البارامتر في المرتبة الثانية من حيث تقبل المفردات الاختبارية، في حين جاء النموذج اللوغاريتمي ثنائي البارامتر في الترتيب الأخير حيث تبين أن خصائصه أكثر جموداً في تقبل المفردات التي يمكن استخدامها في تكوين بنك الأسئلة.

وتتفق هذه النتيجة مع دراسة كل من (١٩٨٦) Waller ، إسماعيل الوليلى (٢٠٠٢)، أمين نور الدين (٢٠٠٢) ، (2008) Custer , Sharairi &Yamazaki ويمكن تفسير ذلك بأنه كلما زاد عدد بارامترات النموذج كلما كان النموذج أكثـر مرونة وأكثر اقتراباً للواقع التعليمي مما يتيح الفرصـة لاحتواء أكبر عدد من المفـردات الاختبارية المثلة لهذا الواقع التعليمي بما يفي لتمثيله والتعبير عنه وتعريفه. ويرى الباحثان أن كون البيانات كانت أكثر ملاءمة للنموذج اللوغاريتمي ثلاثي البارامتر هي نتيجة تتعلق بطبيعة البيانات وبحجم العينة المستخدمة في التحليل في الدراسة الحالية ولا يمكن تعميمها.

ثانيا- النتائج الخاصة بالإجابة على السؤال الثاني من أسئلة الدراسة :

ينص السؤال الثاني للدراسة على: "هل تختلف دقة تقديرات صعوبة المفردات وقدرات الأفراد على بنك الأسئلة باختلاف نموذج التحليل المستخدم (أحادي، ثنائي، ثلاثي) البارمتر ؟ وتتضمن نتائج الإجابة على هذا السؤال ما يلي :

أولا : مقارنة قيم الأخطاء المعيارية لتقديرات الصعوبة والقدرة :

١- مقارنة قيم الأخطاء المعيارية لتقديرات الصعوبة :

تم حساب تحليل التباين أحادي الاتجاه One Way Anova للتأكد من دقة قياس النماذج الثلاث عن طريق دلالة الفروق بين متوسطات الأخطاء المعيارية لصعوبة المفردات واتجاه هذه الفروق. ويوضح جدول (٣) المتوسطات والانحرافات المعيارية لتقدير الصعوبة والأخطاء المعيارية لتقديرات الصعوبة لكل من النماذج الثلاث.

جدول (٣) المتوسطات والانحرافات المعيارية لتقدير صعوبة المفردات
والأخطاء المعيارية لكل من النماذج الثلاث

	النموذج	عدد المفردات	المتوسط	الانحراف المعياري
تقدير الصعوبة باللوجيت	الأحادي	67	1.559	0.038
	الثنائي	66	1.585	0.137
	الثلاثي	69	1.587	0.102
الخطأ المعياري باللوجيت	الأحادي	67	0.084	0.685
	الثنائي	66	0.062	0.515
	الثلاثي	69	0.024	0.194

يلاحظ من جدول (٣) السابق أن أعلى متوسط في تقدير الصعوبة للنموذج الثلاثي البارامتر 1.587 **، في حين وجد أن أقل متوسط** 1.559 **للنموذج الأحادي البارامتر ، حيث يلاحظ أن الفروق بين قيم متوسطات تقدير الصعوبة تكاد تكون طفيفة . ويلاحظ كذلك أن أعلى متوسط في تقدير الخطأ المعياري لصعوبة المفردات للنموذج الأحادي البارامتر** 0.084 **، في حين وجد أن أقل متوسط في تقدير الخطأ المعياري لصعوبة المفردات للنموذج الثلاثي البارامتر** 0.024 .

ولاختبار دلالة الفرق بين متوسطات الأخطاء المعيارية لصعوبة المفردات للنماذج الثلاث تم استخدام تحليل التباين أحادي الاتجاه One Way Anova **، كما استخدم اختبار حجم التأثير (مربع إيتا) لاختبار حجم الفرق بين متوسطات الخطأ المعياري للتقديرات كما يتضح من جدول (٤) التالي**

جدول (٤) تحليل التباين أحادي الاتجاه لدلالة الفرق بين متوسطات
النماذج الثلاث فى تقدير الخطأ المعياري للصعوبة

المتغير	مصدر التباين	مجموع المربعات	د.ح	متوسط المربعات	النسبة الفائية	الدلالة	حجم التأثير η^2
الخطأ المعياري للصعوبة	بين المجموعات	0.33	2	0.016	2.278	0.001	0.01
	داخل المجموعات	51.031	199	0.256			
	الكلى	51.361	201	–			

يتضح من جدول (٤) السابق أن قيمة "ف" دالة إحصائيا عند مستوى 0.001 للفروق بين متوسطات النماذج الثلاث في تقدير الخطأ المعياري لصعوبة المفردات. أي أنه توجد فروق دالة إحصائيا بين متوسطات الأخطاء المعيارية لتقديرات صعوبة المفردات الناتجة من التحليل بالنماذج الثلاث تعود إلى النموذج المستخدم. ولمعرفة اتجاه الفروق التي ظهرت في تقدير الخطأ المعياري لصعوبة المفردات تم إجراء المقارنات المتعددة بين متوسطات الأخطاء المعيارية باستخدام طريقة (LSD) Least Significant Difference أقل فرق دال . ويوضح جدول (٥) المقارنات المتعددة لفرق متوسطات الأخطاء المعيارية بين النماذج اللوغاريتمية الثلاث LSD.

جدول (٥) المقارنات المتعددة لفرق متوسطات الأخطاء
المعيارية بين النماذج الثلاث (LSD)

النموذج (I)	النموذج (J)	فرق المتوسطات
أحادي	ثنائي	-0.238
	ثلاثي	-0.984*
ثنائي	أحادي	-------
	ثلاثي	-0.946*

ويتضح من جدول (٥) أن هناك فروقاً ذات دلالة إحصائية بين متوسطي الخطأ المعياري

للنموذجين الأحادي والثنائي لصالح النموذج الثلاثي البارامتر، وأن هناك فروقاً ذات دلالة إحصائية بين متوسطي الخطأ المعياري للنموذجين الثنائي والثلاثي لصالح الأقل في قيمة متوسط الأخطاء المعيارية للصعوبة وهو النموذج الثلاثي البارامتر، أي أن النموذج الثلاثي البارامتر هو الذى أعطى قيما للأخطاء المعيارية للصعوبة أقل منها في حالة النموذج الأحادي والثنائي البارامتر، ومن ذلك يتضح أن النموذج اللوغاريتمي ثلاثي البارامتر 3PL هو أكثر النماذج اللوغاريتمية دقة في تقدير صعوبة المفردات.

وتتفق نتيجة الدراسة الحالية مع نتائج بعض الدراسات السابقة التي تشير إلى تفوق النموذج ثلاثي البارامتر على النموذجين أحادي وثنائي البارامتر فيما يتعلق بمدى دقة تقدير صعوبة المفردات ومن تلك الدراسات التي تتفق والدراسة الحالية دراسة كل من : (Way & Reese , (1991) ; Saba Rizavi , way & ying Lu (2004) ; Custer : sharairi : yamazaki, (2008), **علاء الدين أيوب** (٢٠٠١)، **نسرين اسماعيل** (٢٠١٠). **بينما اختلفت نتيجة الدراسة الحالية مع نتائج بعض الدراسات السابقة الأخرى التي تشير إلى تفوق النموذج أحادي البارامتر على النموذج ثلاثي البارامتر ، فيما يتعلق بمدى دقة تقدير صعوبة المفردات ومن تلك الدراسات دراسة:** (Chang & Hanson , (2001) .**وقد يرجع السبب وراء زيادة الدقة في تقدير صعوبة المفردات الاختبارية للنموذج ثلاثي البارامتر عن النموذجين أحادي وثنائي البارامتر هو حسن مطابقة (ملاءمة) البيانات للنموذج ثلاثي البارامتر مما يزيد التفاعل بين قدرة الفرد وصعوبة المفردة ، ومن ثم ينقص بارامتر التخمين الذي يظهر من الأداء عن الأسئلة وبخاصة أسئلة الاختيار من متعدد ، وهذا من شأنه أن يجعل تقديرات الأفراد مستقرة مما يقلل من متوسط الأخطاء المعيارية لصعوبة المفردات.**

٢- **<u>مقارنة قيم الأخطاء المعيارية لتقديرات القدرة</u> :**

باستخدام تحليل التباين أحادي الاتجاه One Way Anova **تم التأكد من دقة قياس النماذج الثلاث عن طريق دلالة الفروق بين متوسطات الأخطاء المعيارية لقدرة الأفراد واتجاه هذه الفروق . ويوضح جدول (٦) المتوسطات والانحرافات المعيارية لتقديرات القدرة والأخطاء المعيارية لتقديرات القدرة لكل من النماذج الثلاث .**

جدول (٦) المتوسطات والانحرافات المعيارية لتقدير قدرة الأفراد
والأخطاء المعيارية لكل من النماذج الثلاث

	النموذج	عدد المفردات	المتوسط	الانحراف المعياري
تقدير القدرة باللوجيت	الأحادي	1033	-0.610	0.826
	الثنائي	955	-0.529	0.790
	الثلاثي	941	-0.491	0.970
الخطأ المعياري باللوجيت	الأحادي	1033	0.717	0.115
	الثنائي	955	0.689	0.235
	الثلاثي	941	0.615	0.670

يلاحظ من جدول (٦) السابق أن أعلى متوسط في تقدير قدرة الأفراد للنموذج ثلاثي البارامتر – 0.491، **في حين وجد أن أقل متوسط في تقدير قدرة الأفراد للنموذج أحادي البارامتر** – 0.610، **كما يلاحظ أن أعلى متوسط في تقدير الخطأ المعياري** 0.717 **للنموذج أحادي البارامتر في حين وجد أن أقل متوسط له** 0.615 **للنموذج ثلاثي البارامتر. ويوضح جدول (٧) نتائج تحليل التباين الأحادي لدلالة الفرق بين متوسطات النماذج الثلاث في تقدير الخطأ المعياري للقدرة،كما يوضح دلالة حجم التأثير (مربع إيتا) لاختبار حجم الفرق بين متوسطات الخطأ المعياري للتقديرات .**

جدول (٧) تحليل التباين أحادي الاتجاه لدلالة الفرق بين متوسطات
النماذج الثلاث في تقدير الخطأ المعياري للقدرة

المتغير	مصدر التباين	مجموع المربعات	د.ح	متوسط المربعات	النسبة الفائية	الدلالة	حجم التأثير η^2
الخطأ المعياري للقدرة	بين المجموعات	5.385	2	2.692	16.110	0.001	0.01 دلالته صغيرة
	داخل المجموعات	488.988	2926	0.167			
	الكلى	494.373	2928	-			

E-mail: gamel_abdo59@yahoo.com ☎ 01002702115

ويتضح من جدول (٧) أن قيمة "ف" دالة إحصائياً عند مستوى 0.001 للفروق بين متوسطات النماذج الثلاث في تقدير الخطأ المعياري لقدرة الأفراد. أي أنه توجد فروق ذات دلالة إحصائية بين النماذج الثلاث في تقدير الخطأ المعياري لقدرة الأفراد. ولمعرفة اتجاه الفروق التي ظهرت في تقدير الخطأ المعياري لقدرة الأفراد بين النماذج الثلاث استخدم اختبار توكي Tukey. **ويوضح جدول (٨) نتائج اختبار** Tukey **لبيان اتجاه الفروق التي ظهرت في تقدير الخطأ المعياري لقدرة الأفراد بين النماذج الثلاث.**

جدول (٨) نتائج اختبار توكى لبيان اتجاه الفروق في تقدير الخطأ المعياري لقدرة الأفراد بين النماذج الثلاث

النموذج (I)	النموذج (J)	فرق المتوسطات
أحادي	ثنائي	0.042
	ثلاثي	0.351*
ثنائي	أحادي	-0.042
	ثلاثي	0.308*
أحادي	ثلاثي	-0.308*
ثنائى		-0.351*

ويتضح من جدول (٨) أن هناك فروقا ذات دلالة إحصائية بين متوسطي الخطأ المعياري للنموذجين الأحادي والثلاثي لصالح النموذج ثلاثي البارامتر ، وأن هناك فروقاً ذات دلالة إحصائية بين متوسطي الخطأ المعياري للنموذجين الثنائي والثلاثي لصالح النموذج ثلاثي البارامتر الأقل في متوسط الأخطاء المعيارية ، هذا بدوره يدل على أن النموذج اللوغاريتمي ثلاثي البارامتر هو النموذج الذي أعطى أقل متوسط للأخطاء المعيارية لتقدير القدرة ، ومن ثم فهو أكثر النماذج اللوغاريتمية دقة في تقدير قدرة الأفراد

وتتفق نتيجة الدراسة الحالية مع نتائج الدراسات السابقة التي تشير إلى تفوق النموذج ثلاثي البارامتر على النموذجين أحادي وثنائي البارامتر وذلك فيما يتعلق بمدى دقة تقدير الخطأ المعياري لقدرة الأفراد ومن تلك الدراسات دراسة كل من (Waller,1986 ; Way & Reese , (1991) ;

Saba Rizavi , way & ying Lu (2004) ; Shu Jing Yen& Leah,Walker,2007 ; Custer , sharairi & yamazaki , 2008) **وأيضا دراسة نسرين إسماعيل(٢٠١٠) ، بينما اختلفت نتيجة الدراسة الحالية مع نتائج بعض الدراسات السابقة كدراسة** Chang &Hanson (2001) **التي أشارت إلى تفوق النموذج أحادي البارامتر على النموذج ثلاثي البارامتر فيما يتعلق بمدى دقة تقدير قدرة الأفراد، ودراسة محمد الشافعي (٢٠١١) التي توصلت إلى أن النموذج الثنائي البارامتر أفضل النماذج الثلاثة من حيث دقته في تقدير قدرات الأفراد.**

وقد يرجع السبب وراء زيادة الدقة في تقدير قدرة الأفراد للنموذج ثلاثي البارامتر عن النموذجين أحادي وثنائي البارامتر في الدراسة الحالية إلى مطابقة البيانات للنموذج ثلاثي البارامتر ، فكلما تقاربت صعوبة المفردات وقدرات الطلاب يؤثر ذلك بشكل جوهري على الأخطاء المعيارية لتقديرات القدرة حيث يؤدي التقارب إلى انخفاض الخطأ المعياري لتقدير قدرة الطلاب بسبب حدوث تفاعل بين قدرات الطلاب وصعوبة المفردات ، وتصدي النموذج ثلاثي البارامتر لمشكلة التخمين التي تمثل مشكلة للطلاب منخفضي القدرة عند الإجابة على أسئلة الاختيار من متعدد ، حيث يلجأ الطلاب ذوي القدرة المنخفضة على بعض مفردات الاختبار وبخاصة ذات مستوى الصعوبة المرتفع إلى حد ما عن قدرتهم ، وهذا من شأنه أن يزيد من متوسط الأخطاء المعيارية لتقديرات القدرة .

ثانيا : <u>الكشف عن مؤشرات الثبات لكل نموذج من النماذج الثلاث</u> :

نظرا لأن برنامج التحليل الإحصائي بارسكيل لا يقوم بحساب ثبات تقديرات بارامترات مفردات بنك الأسئلة (الصعوبة والتمييز) وثبات تقديرات قدرة الأفراد - ويعد ذلك من أوجه قصور البرنامج - لذلك تم حساب مؤشرات ثبات بارامترات مفردات الصور الاختبارية (بنك الأسئلة) وكذلك تقديرات قدرة الأفراد لكل نموذج من النماذج الثلاث الأحادي والثنائي والثلاثي البارامتر باستخدام معامل الثبات الهامشي أو الحدي Marginal Reliability **ويستخدم هذا المعامل في تحديد ثبات الاختبارات المدرجة على نظرية الاستجابة للمفردة حيث يعتمد في حسابه على الخطأ المعياري لتقدير بارامترات المفردات أو القدرة ، ويمكن تفسيره بنفس الطريقة التي يفسر بها معامل الثبات التقليدي إلا أن الأخير لا يأخذ في اعتباره اختلاف الخطأ المعياري للتقدير. والصورة الرياضية له كالتالى :**

معامل الثبات الهامشي (الحدي) = ١- مربع متوسط الخطأ المعياري (FCAT , 2003 : 42)

ويوضح جدول (٩) مؤشرات ثبات بارامترات مفردات بنك الأسئلة وتقديرات قدرة الأفراد لكل نموذج من النماذج الثلاث الأحادي والثنائي والثلاثي البارمتر.

جدول (٩) مؤشرات ثبات بارامترات مفردات بنك الأسئلة وتقديرات قدرة الأفراد لكل نموذج من النماذج الثلاث

النموذج	معامل الثبات الهامشي		
	ثبات الصعوبة	ثبات التمييز	ثبات القدرة
أحادي البارامتر	**0.85**	**0.82**	**0.49**
ثنائي البارامتر	**0.77**	**0.69**	**0.53**
ثلاثي البارامتر	**0.90**	**0.95**	**0.62**

ويتبين من جدول(٩) السابق ارتفاع قيم مؤشرات ثبات بارامترات مفردات بنك الأسئلة وتقديرات قدرة الأفراد وفقا للنموذج اللوغاريتمي ثلاث البارمتر حيث بلغ ثبات صعوبة المفردات 0.90 **، وبلغ ثبات التمييز الحد** 0.95 **، وبلغ ثبات قدرة الأفراد للمستوى** 0.62 **، وبذلك يتفوق النموذج اللوغاريتمي ثلاثي البارامتر على النموذجين أحادي وثنائي البارمتر بمقارنة قيم الثبات الخاصة بصعوبة وتمييز المفردات وقدرة الأفراد ، ومن ثم يمكن القول أن النموذج اللوغاريتمي ثلاثي البارامتر هو أكثر النماذج اللوغاريتمية ارتفاعاً في مؤشرات ثبات بارامترات مفردات بنك الأسئلة وتقديرات قدرة الأفراد. كما بمكن أن نستنتج أن النموذج اللوغاريتمي ثلاثي البارامتر هو أكثر النماذج اللوغاريتمية الثلاثة دقة في تدريج وبناء بنك الأسئلة . وتتفق نتيجة الدراسة الحالية مع دراسة كل من علاء الدين أيوب (٢٠٠٤) ، إسماعيل حسن الوليلي (٢٠٠٢) ، نسرين إسماعيل (٢٠١٠) ،** Saba Rizavi , Way & Ying Lu(2004) .

وقد يرجع السبب وراء زيادة الدقة في مؤشرات ثبات بارامترات مفردات بنك الأسئلة وتقديرات قدرة الأفراد للنموذج ثلاثي البارمتر على النموذجين أحادي وثنائي البارمتر هو تحقيق التفاعل بين قدرة الفرد وصعوبة المفردة بدرجة أفضل من النموذجين الآخرين ومن ثم ينقص بارامتر التخمين الذي يظهر من خلال الأداء على أسئلة الاختيار من متعدد ، أما في النموذجين الآخرين يلجأ الطلاب ذوي القدرة المنخفضة

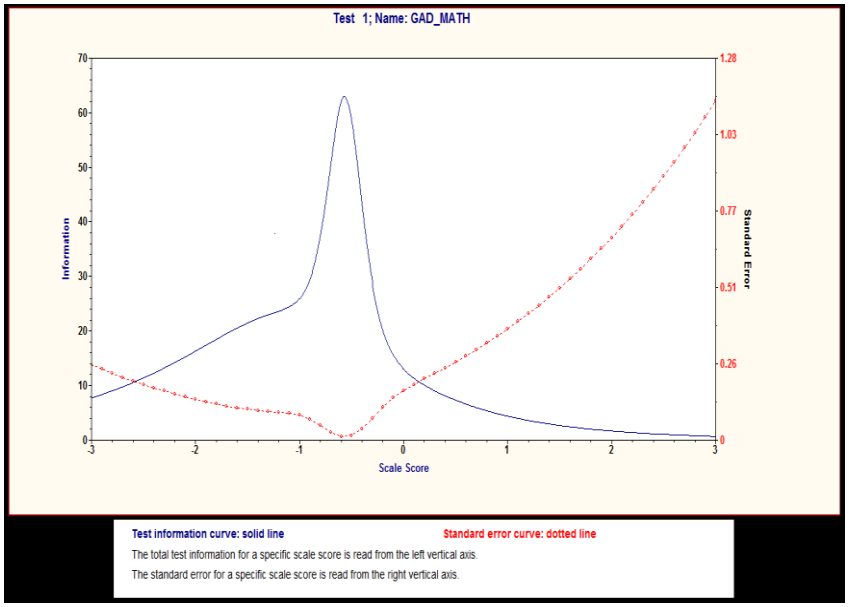
إلى التخمين على بعض مفردات الاختبار وبخاصة المفردات ذات مستوى الصعوبة المرتفع إلى حد ما عن قدرتهم ، وهذا من شأنه أن يجعل تقديرات مؤشرات الثبات غير مستقرة .

ثالثا : الكشف عن دالة معلومات الاختبار وفقاً لكل نموذج من النماذج الثلاث :

تم الكشف عن دالة معلومات الاختبار للنماذج الثلاث أحادي وثنائي وثلاثي البارمتر، وفيما يلي عرض دالة معلومات الاختبار لكل نموذج كلاً على حده :

١- **دالة معلومات الاختبار وفقاً للنموذج اللوغاريتمي الثلاثي البارامتر :**

ضح شكل (١)* دالة معلومات اختبار النموذج اللوغاريتمي ثلاثي البارمتر ومنحنى الخطأ المعياري لمفردات بنك الأسئلة.



شكل (١)
دالة معلومات الاختبار وفقاً للنموذج اللوغاريتمي الثلاثي البارامتر

ويلاحظ من شكل (١) السابق أن قيم منحنى دالة المعلومات تتزايد تزايد مطرد عند مستوى قدرة باللوجيت من 3− إلى 1− ، ويزداد هذا الاطراد بصورة حادة من 1− إلى 0.6− ، ويصل إلى قمته عند

مستوى معلومات قدرة 63 ، ويطرد في الانحدار من 0 إلى 3 ثم يهبط عند مستوى قدرة 0.6- إلى 0 ، ويلاحظ أن كثافة دالة المعلومات تتركز حول المحور عند مستوى قدرة 0.6 - ، وأن مدى منحنى دالة المعلومات يبدأ من 1.5 إلى 63 . كما يلاحظ أن منحنى الخطأ المعياري يبدأ في الانخفاض من أعلى قيمة له وهي 1.12 عند مستوى قدرة 3 ، ثم يصل إلى أدنى قيمة عند مستوى قدرة 0.6- ثم يبدأ في زيادة مطردة من 0.6- إلى 3- ، كذلك يلاحظ أن مدى قيم منحنى الخطأ المعياري للقياس تتراوح من 0.04 إلى 1.17 .

ويبين جدول (١٠) قيم دوال المعلومات الخاصة بكل مفردة من مفردات بنك الأسئلة والتي تم حسابها وفقاً للنموذج الثلاثي البارامتر ، وكذلك دالة معلومات الاختبار.

جدول (١٠) قيم دوال المعلومات لمفردات بنك الأسئلة وفقاً للنموذج الثلاثي البارمتر

المسلسل	كود المفردة	دالة معلومات المفردة	المسلسل	كود المفردة	دالة معلومات المفردة
١	CA0003	٠.١٣٥	١٩	CA0064	٠.٠٦٨
٢	CA0004	٠.٢٤١	٢٠	CA0066	٠.١١٨
٣	CA0005	٠.٣٢٧	٢١	CA0067	٠.١١٤
٤	CA0006	٠.٣٠٦	٢٢	CA0068	٠.١٢١
٥	CA0008	١.٧٥٢	٢٣	CA0070	٠.١٨
٦	CA0016	٠.٨٢٣	٢٤	CA0071	٠.٢٧
٧	CA0019	٠.٩٥	٢٥	CA0072	٠.١٩٨
٨	CA0020	٠.١٢٦	٢٦	CA0076	٠.١٦١
٩	CA0021	٠.٥١١	٢٧	CA0077	٠.١٤٤
١٠	CA0022	٠.٦٨٣	٢٨	CA0078	١٥.٥٣
١١	CA0023	٠.٨٠٢	٢٩	CU0010	٠.٣٠٧
١٢	CA0024	٠.٦	٣٠	CU0011	٠.٧٨١
١٣	CA0026	٠.٢٣٩	٣١	CU0017	٠.٤٢٣
١٤	CA0027	١.٣٦٥	٣٢	CU0018	٠.٥٦٤
١٥	CA0031	٠.١٦٢	٣٣	CU0028	٠.٣٣٥
١٦	CA0033	٠.٤٦٧	٣٤	CU0029	٠.١١
١٧	CA0034	٠.٨٥٢	٣٥	CU0030	٠.٤
١٨	CA0035	٠.٨١٨	٣٦	CU0032	٠.٣٦٦

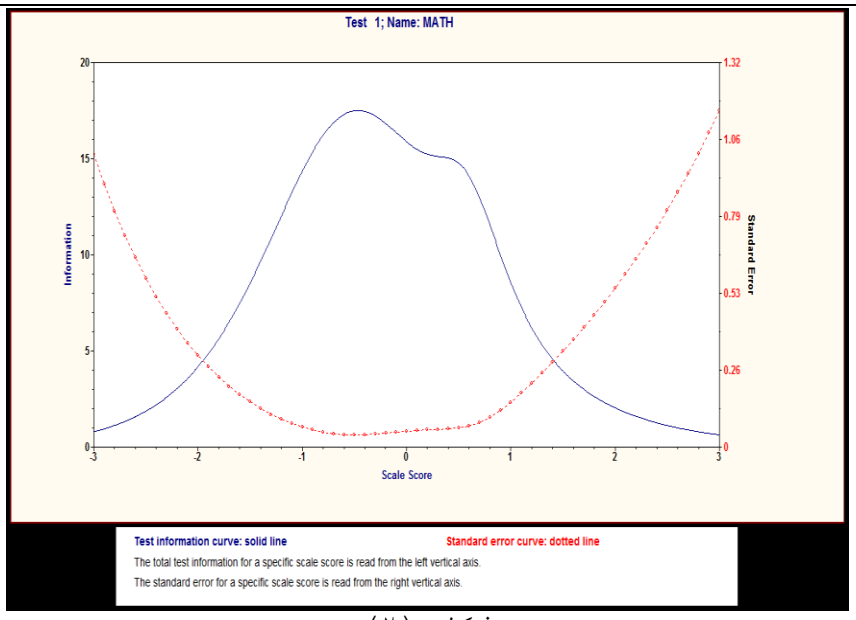
E-mail: gamel_abdo59@yahoo.com ☎ 01002702115

تابع جدول (١٠) قيم دوال المعلومات لمفردات بنك الأسئلة وفقاً للنموذج الثلاثي البارمتر

المسلسل	كود المفردة	دالة معلومات المفردة	المسلسل	كود المفردة	دالة معلومات المفردة
٣٧	CA0036	٠.١٨٤	٥٥	CU0038	٠.٢٢٣
٣٨	CA0037	٠.١٧	٥٦	CU0042	٠.١٥٤
٣٩	CA0039	٠.٣١١	٥٧	CU0047	٠.٢٤١
٤٠	CA0040	٠.٢٩٧	٥٨	CU0050	٠.٦٧٩
٤١	CA0041	٠.٢٤٣	٥٩	CU0053	٠.٢٢٥
٤٢	CA0043	٠.٣٦٩	٦٠	CU0057	٠.٢٨٥
٤٣	CA0044	٠.١١٩	٦١	CU0058	٠.٢٦٧
٤٤	CA0046	٠.٢٥٢	٦٢	CU0063	٠.١٩٨
٤٥	CA0048	٠.١٥٢	٦٣	CU0065	١٥.٥٣
٤٦	CA0049	٠.٢	٦٤	CU0069	٠.١٥٦
٤٧	CA0051	٠.٩٣١	٦٥	CU0074	٠.١٥٤
٤٨	CA0052	٠.١١٨	٦٦	CU0075	٠.٠٨٦
٤٩	CA0054	٠.١٣٨	٦٧	EA0012	٠.١٤٦
٥٠	CA0055	٠.١٢	٦٨	EA0014	٠.٣٢٧
٥١	CA0059	٠.١٩	٦٩	EA0015	٠.٢١٩
٥٢	CA0060	٠.٣٤٧			
٥٣	CA0061	٠.٨٠٦			
٥٤	CA0062	٠.٤١٤			
دالة معلومات الاختبار=		مجموع دوال معلومات مفردات الاختبار = 55.6			

٢- دالة معلومات الاختبار وفقا للنموذج اللوغاريتمى الثنائي البارمتر:

يوضح شكل (٢) دالة معلومات الاختبار وفقا للنموذج اللوغاريتمى ثنائي البارامتر ومنحنى الخطأ المعياري المقابل لمفردات بنك الأسئلة.



شكل (٢)
دالة معلومات الاختبار وفقاً للنموذج اللوغاريتمي الثنائي البارامتر

ويلاحظ من شكل (٢) السابق أن قيم منحنى دالة المعلومات تتزايد عند مستوى قدرة من 3- إلى 0.6-بوحدة اللوجيت، ثم يبدأ المنحنى في الانخفاض عند مستوى قدرة بداية من 0.6 - إلى أن وصل 3بوحدة اللوجيت ، ويلاحظ أن كثافة دالة المعلومات حول المحور عند مستوى قدرة 0.6- لوجيت ، وأن مدى منحنى دالة المعلومات يبدأ من 0.8 إلى 16.48. كذلك يلاحظ أن منحنى الخطأ المعياري للقياس ينخفض عند مستوى قدرة من 3- إلى 0.6- لوجيت ، ويبدأ في التزايد بصورة طفيفة من 0.6- إلى 0.6 ويبدأ في التزايد بصورة ملحوظة من 0.6 حتى يصل إلى قمته عند مستوى قدرة للفرد يصل إلى 3 لوجيت ، والتي كانت قيمة الخطأ المعياري عندها 1.17، وكانت أقل قيمة للخطأ المعياري عند مستوى قدرة للفرد يصل إلى 0.6- ، وبذلك يكون شكل المنحنى متماثل تقريباً عند مستوى قدرة للفرد يصل إلى 0.3- ، ويلاحظ أيضاً أن مدى قيم منحنى الخطأ المعياري للقياس يتراوح من 0.02 إلى 1.12 .

ويبين جدول (١١) قيم دوال المعلومات الخاصة بكل مفردة من مفردات بنك الأسئلة والتي تم حسابها وفقاً للنموذج الثنائي البارامتر ، وكذلك دالة معلومات الاختبار .

جدول (١١)

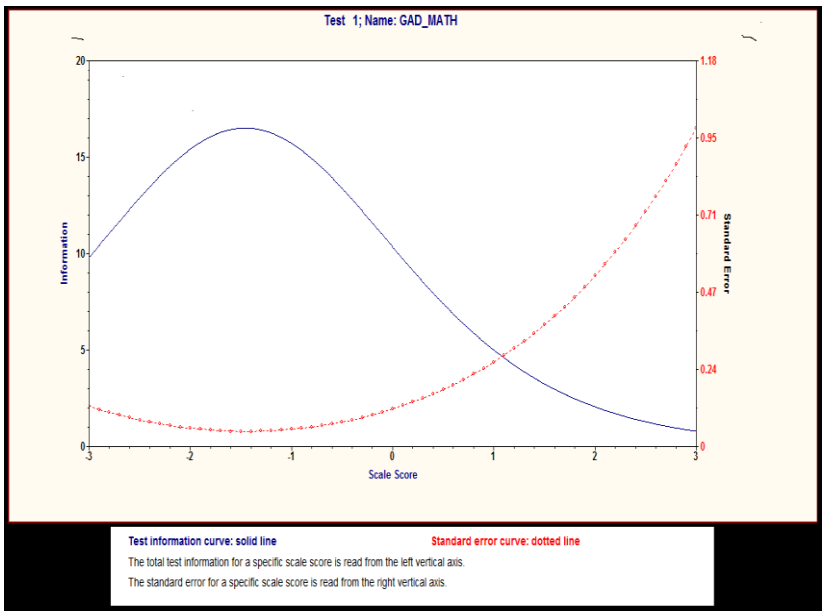
قيم دوال المعلومات لمفردات بنك الأسئلة وفقاً للنموذج الثنائي البارمتر

المسلسل	كود المفردة	دالة معلومات المفردة	المسلسل	كود المفردة	دالة معلومات المفردة
١	CA0003	٠.٢٥١	٣٤	CA0056	٠.٢٥١
٢	CA0004	٠.٢٥١	٣٥	CA0059	٠.٢٥١
٣	CA0005	٠.٢٥١	٣٦	CA0060	٠.٢٥٠
٤	CA0006	٠.٢٥١	٣٧	CA0061	٠.٢٥١
٥	CA0008	٠.٢٥١	٣٨	CA0062	٠.٢٥١
٦	CA0016	٠.٢٥١	٣٩	CA0064	٠.٢٥١
٧	CA0019	٠.٢٥١	٤٠	CA0066	٠.٢٥٢
٨	CA0020	٠.٢٥١	٤١	CA0067	٠.٢٥١
٩	CA0021	٠.٢٥١	٤٢	CA0068	٠.٢٥١
١٠	CA0022	٠.٢٥٢	٤٣	CA0070	٠.٢٥١
١١	CA0023	٠.٢٥١	٤٤	CA0071	٠.٢٥١
١٢	CA0024	٠.٢٥١	٤٥	CA0072	٠.٢٥٠
١٣	CA0025	٠.٢٥١	٤٦	CA0073	٠.٢٥١
١٤	CA0026	٠.٢٥١	٤٧	CA0076	٠.٢٥١
١٥	CA0027	٠.٢٥١	٤٨	CA0077	٠.٢٥٢
١٦	CA0031	٠.٢٥١	٤٩	CA0078	٠.٢٥١
١٧	CA0033	٠.٢٥١	٥٠	CU0010	٠.٢٥١
١٨	CA0034	٠.٢٥١	٥١	CU0011	٠.٢٥١
١٩	CA0035	٠.٢٥١	٥٢	CU0017	٠.٢٥١
٢٠	CA0036	٠.٢٥١	٥٣	CU0018	٠.٢٥١
٢١	CA0037	٠.٢٥١	٥٤	CU0030	٠.٢٥٠
٢٢	CA0039	٠.٢٥١	٥٥	CU0032	٠.٢٥١
٢٣	CA0040	٠.٢٥١	٥٦	CU0038	٠.٢٥٠
٢٤	CA0041	٠.٢٥١	٥٧	CU0042	٠.٢٥١
٢٥	CA0043	٠.٢٥١	٥٨	CU0047	٠.٢٥٢
٢٦	CA0044	٠.٢٥١	٥٩	CU0050	٠.٢٥١
٢٧	CA0046	٠.٢٥١	٦٠	CU0053	٠.٢٥١

المسلسل	كود المفردة	دالة معلومات المفردة	المسلسل	كود المفردة	دالة معلومات المفردة
٢٨	CA0048	٠.٢٥٠	٦١	CU0057	٠.٢٥١
٢٩	CA0049	٠.٢٥١	٦٢	CU0058	٠.٢٥٠
٣٠	CA0051	٠.٢٥١	٦٣	CU0063	٠.٢٥١
٣١	CA0052	٠.٢٥١	٦٤	CU0065	٠.٢٥٠
٣٢	CA0054	٠.٢٥١	٦٥	CU0069	٠.٢٥١
٣٣	CA0055	٠.٢٥١	٦٦	CU0074	٠.٢٥١
دالة معلومات الاختبار=		مجموع دوال معلومات مفردات الاختبار = 16.573			

٣- <u>دالة معلومات الاختبار وفقا للنموذج اللوغاريتمي الأحادي البارمتر</u>:

يوضح شكل (٣) دالة معلومات الاختبار وفقا للنموذج اللوغاريتمي أحادي البارامتر ومنحنى الخطأ المعياري المقابل لمفردات بنك الأسئلة.



شكل (٣)
دالة معلومات الاختبار وفقاً للنموذج اللوغاريتمي الأحادي البارامتر

يلاحظ من شكل (٣) السابق أن قيم منحنى دالة المعلومات تتزايد من مستوى قدرة 3- لوجيت

حتى وصل إلي أعلى قيمة له في دالة المعلومات قدرها 17.46 عند مستوى قدرة 0.5- لوجيت ثم يتناقص منحنى دالة المعلومات من مستوى قدرة 0.5- حتى يصل إلى أدنى قيمة له عند مستوى قدرة 3 لوجيت ، ويلاحظ أن كثافة دالة المعلومات تتركز حول المحور عند مستوى قدرة 0.8- لوجيت وهو يكاد يكون قريب الشكل من منحنى التوزيع الاعتدالي، وأن مدى منحنى دالة المعلومات يبدأ من 3 إلى 17.46 ، كذلك يلاحظ أن منحنى الخطأ المعياري للقياس يبدأ من أعلى قيمة له وهي 1.16 عند مستوى قدرة 3 لوجيت ، ثم يبدأ في الانخفاض إلى أن يصل إلى أدنى قيمة له وهي 0.03 عند مستوى قدرة 0.6- ويبدأ في التزايد مرة أخرى إلى أن يصل منحنى دالة معلومات قيمته 15.3 عند مستوى قدرة 3- ، ويلاحظ أيضاً أن مدى منحنى الخطأ المعياري من 0.03 إلى 1.16 . ويبين جدول (١٢) قيم دوال المعلومات الخاصة بكل مفردة من مفردات بنك الأسئلة والتي تم حسابها وفقاً للنموذج الأحادي البارامتر ، وكذلك دالة معلومات الاختبار.

جدول (١٢)
قيم دوال المعلومات لمفردات بنك الأسئلة وفقاً للنموذج الأحادي البارمتر

المسلسل	كود المفردة	دالة معلومات المفردة	المسلسل	كود المفردة	دالة معلومات المفردة
١	CA0003	0.267	٣١	CA0056	0.385
٢	CA0004	0.238	٣٢	CA0059	0.286
٣	CA0005	0.252	٣٣	CA0060	0.212
٤	CA0006	0.281	٣٤	CA0061	0.253
٥	CA0008	0.838	٣٥	CA0062	0.567
٦	CA0019	0.572	٣٦	CA0064	0.163
٧	CA0020	0.609	٣٧	CA0066	0.348
٨	CA0021	0.522	٣٨	CA0067	0.202
٩	CA0022	0.465	٣٩	CA0068	0.193
١٠	CA0023	0.723	٤٠	CA0070	0.370
١١	CA0024	0.385	٤١	CA0071	0.145
١٢	CA0025	0.305	٤٢	CA0072	0.189
١٣	CA0026	0.436	٤٣	CA0073	0.060

المسلسل	كود المفردة	دالة معلومات المفردة	المسلسل	كود المفردة	دالة معلومات المفردة
١٤	CA0027	0.402	٤٤	CA0076	0.127
١٥	CA0031	0.315	٤٥	CA0077	0.122
١٦	CA0033	0.229	٤٦	CU0010	0.229
١٧	CA0034	0.429	٤٧	CU0011	0.399
١٨	CA0035	0.596	٤٨	CU0017	0.557
١٩	CA0037	0.144	٤٩	CU0018	0.616
٢٠	CA0040	0.290	٥٠	CU0028	0.274
٢١	CA0041	0.277	٥١	CU0030	0.177
٢٢	CA0043	0.136	٥٢	CU0038	0.216
٢٣	CA0044	0.153	٥٣	CU0042	0.211
٢٤	CA0046	0.151	٥٤	CU0047	0.178
٢٥	CA0048	0.176	٥٥	CU0050	0.135
٢٦	CA0049	0.173	٥٦	CU0053	0.135
٢٧	CA0051	0.133	٥٧	CU0057	0.126
٢٨	CA0052	0.258	٥٨	CU0058	0.283
٢٩	CA0054	0.167	٥٩	CU0063	0.314
٣٠	CA0055	0.199	٦٠	CU0065	0.104
٦١	CU0069	0.145	٦٥	EA0007	1.180
٦٢	CU0074	0.131	٦٦	EA0012	0.383
٦٣	CU0075	0.121	٦٧	EA0014	0.127
٦٤	EA0001	0.901			
دالة معلومات الاختبار=		مجموع دوال معلومات مفردات الاختبار = 20.685			

ويلخص جدول (١٣) التالي الجداول السابقة كما يلي :

E-mail: gamel_abdo59@yahoo.com ☎ 01002702115

جدول (١٣)
قيم دوال معلومات الاختبار للنماذج اللوغاريتمية الثلاث

النموذج	أقصى قيمة لدالة معلومات الاختبار	دالة معلومات الاختبار
الأحادي البارمتر	17.46	20.685
الثنائي البارمتر	16.48	16.573
الثلاثي البارامتر	63	55.6

ويتضح من النتائج المتضمنة في جدول (١٣) أن النموذج الثلاثي البارامتر قد قدم أقصى قيمة لدالة المعلومات وأعلى قيمة لدالة المعلومات عن النموذجين الأحادي والثنائي البارمتر، فنجد أن أقصى قيمة لدالة معلومات الاختبار وفقاً للنموذج الثلاثي البارامتر تساوي 63 ، في مقابل 17.46 و 16.48 للنموذجين الأحادي والثنائي البارمتر على الترتيب ، كما أن قيمة دالة معلومات الاختبار وفقاً للنموذج الثلاثي البارمتر هي أعلى من قيمتها في كل من النموذجين الأحادي والثنائي البارمتر. ومن ثم فإن النموذج اللوغاريتمي ثلاثي البارمتر هو أكثر النماذج اللوغاريتمية الثلاث ارتفاعاً في تقدير دوال معلومات الاختبار.

وتتفق نتيجة الدراسة الحالية مع دراسة علاء أيوب (2001) ، ودراسة نسرين إسماعيل (2010) ، حيث وجدت الدراستين أن النموذج الثلاثي البارمتر يزيد من دقة القياس حيث يعطي معلومات إضافية من خلال دوال المعلومات بدرجة أفضل من النموذجين الأحادي والثنائي البارمتر ، وهذه النتيجة تتسق كذلك مع ما توصلت إليه دراسة أمين نور الدين (2002) حيث أشارت إلى أن أعلى متوسط لدالة المعلومات كان في النموذج الثلاثي البارامتر عنه في النموذجين الآخرين.

بينما اختلفت نتيجة الدراسة الحالية مع دراسة محمد الشافعي (2011) التي أشارت إلى أن أعلى متوسط لقيم دوال المعلومات كانت ناتجة من استخدام النموذج الثنائي البارامتر يليه النموذج الثلاثي البارامتر ، بينما أدنى قيمة لمتوسط دوال معلومات مفردات الاختبار تلك التي نتجت من استخدام النموذج أحادي البارمتر. كذلك اختلفت نتيجة الدراسة الحالية مع دراسة Wiberg (2004) التي بينت أن النموذج الثنائي البارامتر يقدم أكثر قدر من المعلومات عن النموذجين الأحادي والثلاثي البارمتر بينما أعطى النموذج الثلاثي البارامتر مزيداً من المعلومات عند مستوى القدرة الذى يناظر الدرجة الفاصلة ، ولذلك فضلت استخدام النموذج الثلاثي البارمتر عن النموذج الثنائي البارمتر.

E-mail: gamel_abdo59@yahoo.com ☎ 01002702115

ثالثاً : النتائج الخاصة بالإجابة عل السؤال الثالث من أسئلة الدراسة :

ينص السؤال الثالث للدراسة على : "ما تدريج وتقدير صعوبة المفردات وقدرات الأفراد في بنك الأسئلة وفقا للنموذج الأكثر دقة ؟ ". وللإجابة عن هذا السؤال حللت البيانات وفقاً للنموذج الأكثر دقة وملاءمة وهو النموذج ثلاثي البارمتر للمرة الثالثة بعد حذف الأفراد والمفردات غير الملائمة للقياس بهدف الحصول على التدريج النهائي لمفردات بنك الأسئلة تبعاً لبارمتر الصعوبة مقدرة بوحدة اللوجيت ، وتدريج قدرات الأفراد مقدرة بوحدة اللوجيت والتي أمكن تحويلها إلى وحدة المنف. وقد أسفرت نتائج تحليل البيانات تبعا للخطوات التي اتبعها الباحثان في تدرج مفردات بنك الأسئلة وفقاً للنموذج اللوغاريتمي ثلاثي البارمتر عن ما يلي :

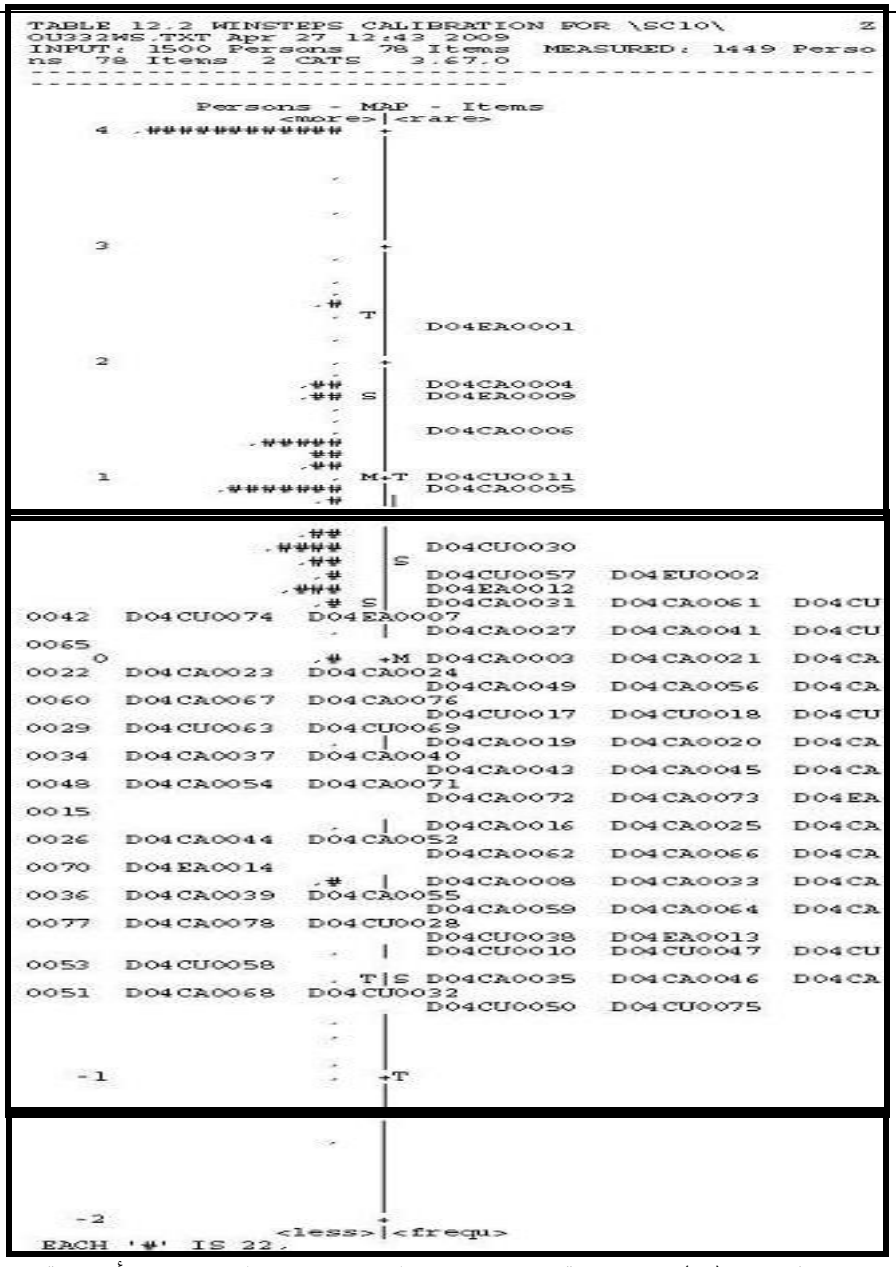
١- **بلغ عدد مفردات البنك بعد حذف المفردات غير الملائمة لأسس القياس الموضوعي (69) مفردة بمدى صعوبة يمتد من 1.196- إلى 0.545 لوجيت أي من 44 إلى 53 منف ، ويعد مدى الصعوبة الذي تغطيه مفردات بنك الأسئلة مدى صغير نسبياً وقد يرجع ذلك إلى السهولة النسبية في صياغة العديد من المفردات ، كما يرجع أيضاً إلى تقارب مستويات أفراد عينة التدريج .**

٢- **تمتد الأخطاء المعيارية لتقديرات الصعوبة لمفردات بنك الأسئلة من 0.064 إلى 0.634 لوجيت أي من 0.318 إلى 3.169 منف .**

٣- **تمتد تقديرات قدرات الأفراد لبنك الأسئلة من 8.465- إلى 1.022 لوجيت ، أي من 8 إلى 55 منف ، بأخطاء معيارية من 0 إلى 54.5 منف .**

٤- **تعرف مفردات البنك جميع مستويات المتغير موضوع القياس (التحصيل الدراسي) بشكل جيد ، ويظهر ذلك في خريطة توزيع المفردات لبنك الأسئلة. حيث استخدم برنامج التحليل الإحصائي ونستبس Winsteps بعد تغذيته بصعوبات المفردات في صورتها النهائية لرسم هذه الخريطة والتي يوضحها شكل (٤).**

E-mail: gamel_abdo59@yahoo.com ☎ 01002702115



شكل (٤) خريطة توزيع المفردات لبنك الأسئلة

وكما يتضح من شكل (٤) أن مفردات بنك الأسئلة تعرف جميع مستويات المتغير حيث لا توجد فجوات في الوسط ، ولكن ظهرت بعض فجوات على الأطراف. ويوضح جدول (١٤) الإجراءات التي استخدمت

للتعرف على حقيقة تلك الفجوات .

جدول (١٤)
علاقة الفرق بين صعوبة مفردتين متتاليتين بالخطأ المعيارى لهما (بالمنف) لبنك الأسئلة

الفجوة	المفردتين المتتاليتين	الوصف الإحصائي المستخدم لتحديد العلاقة				
		الصعوبة	الخطأ المعيارى	فرق الصعوبة	مج الخطأ المعيارى	الحكم
الفجوة الأولى	D04EA0001	52	0	5	2	1
	D04CA0004	47	2			
الفجوة الثانية	D04EA0009	44	3	2	5	0
	D04CA0006	46	2			
الفجوة الثالثة	D04CA0006	46	2	1	3	0
	D04CU0011	45	1			
الفجوة الرابعة	D04CA0005	46	2	3	4	0
	D04CU0030	49	2			
الفجوة الخامسة	D04EU0002	52	0	3	2	1
	D04CU0030	49	2			

يعبر جدول (١٤) عن العلاقة بين مفردتين أ و ب متتاليتين والخطأ المعياري لهما (بالمنف) ، ويعبر عمود الحكم في الجدول عن دلالة المسافة بين المفردتين (الفرق في الصعوبة) حيث يعني (0) أن مسافة الفراغ أصغر من مجموع الخطأ المعياري ، و (1) أن مسافة الفراغ أكبر من مجموع الخطأ المعياري. ويوضح الجدول عدم وجود فجوات جوهرية في وسط المقياس بينما توجد فجوتين عند طرفي المتصل كما بدى من خريطة المتغير. ويدل عدم وجود فجوات جوهرية على متصل صعوبات المفردات وخاصة في وسط المتصل على انتظام تدرج المفردات على ميزان القياس ، مما يعني تعريف مفردات المقياس لجميع مستويات المتغير الذي يمثله هذا المتصل .

رابعا : <u>النتائج الخاصة بالإجابة على السؤال الرابع من أسئلة الدراسة</u> :

ينص السؤال الرابع على : "ما صدق وثبات بنك الأسئلة في صورته النهائية ؟" وتتضمن نتائج الإجابة على هذا السؤال ما يلي :

١- نتائج حساب صدق القياس :

اعتمدت الدراسة الحالية على عدة طرق للتأكد من صدق الصور الاختبارية (بنك الأسئلة) وذلك كما يلي :

- صدق المحتوى :

يبدو صدق المحتوي والذي سبق واتضح في خطوات بناء الصور الاختبارية التي تكون بنك الأسئلة من خلال تحليل المحتوي الذي اعتدت الصور الاختبارية لقياسه وصياغة المفردات الاختبارية بما يتفق مع المهارات التي يقيسها وتكوين الصور الاختبارية بما يتفق مع جدول المواصفات .

- صدق المحكمين :

اعتمد الباحثان على صدق المحكمين كما سبق واتضح في خطوات بناء الصور الاختبارية (بنك الأسئلة) والذي تمثل في تحكيم جميع مراحل بناء الصور الاختبارية (بنك الأسئلة) بدءاً من تحديد المهارات الفرعية وانتهاء إلى تكوين الصور الاختبارية .

- صدق قياس وتعريف المتغير :ويتضمن :

أ) صدق التدرج (كما توفره نظرية الاستجابة للمفردة) :

إن تدرج مفردات مجموعة من الاختبارات – يفترض أنها تقيس نفس السمة ، على ميزان تدرج واحد مشترك باستخدام نظرية الاستجابة للمفردة يعني أن هذه المفردات تحقق شرط أحادية البعد أي أنها تعرف متغيراً واحداً ، وتتحقق أحادية البعد بتحقق صدق المحتوى وملاءمة كل من الأفراد والمفردات (لأسس القياس الموضوعي) تبعاً لمحكات الملاءمة الخاصة بالبرنامج المستخدم في التحليل – برنامج بارسكيل في الدراسة الحالية – حيث توضح هذه المحكات مدى قياس المفردة لما تقيسه باقي المفردات على متصل المتغير موضوع القياس ، كما تبين مدى اتساق تدرج قدرة الفرد مع تدرج قدرات باقي الأفراد على هذا المتصل (أمينة كاظم ، ١٩٩٦ : ٣٦٦- ٣٦٧). وكما سبق فقد تم حذف الأفراد غير الصادقين آلياً

بواسطة برنامج التحليل بارسكيل كما حذفت المفردات غير الملائمة وفقاً لمحكات الملاءمة التي يستخدمها البرنامج ، كما تم تدرج جميع الصور الاختبارية لمقياس التحصيل الدراسي على ميزان تدرج واحد .

ب) **عدم وجود فجوات بين مفردات بنك الأسئلة :**

يتطلب هذا الصدق التحقق من عدم وجود فجوات بين مفردات البنك على متصل السمة المقاسة وهو ما تحقق بالفعل في الدراسة الحالية كما اتضح من شكل (٤) وجدول (١٤) السابقين.

- **صدق التمييز (حساسية المقياس) :**

ويعني مدى حساسية المقياس (بنك الأسئلة) للتمييز بين مستويات الأفراد وتم التحقق من ذلك باستخدام الصورة النهائية بعد تدريج البنك وتقسيم الأفراد الذين أجابوا عنها إلى مجموعتين باستخدام طرفي القدرة، الأولى : مرتفعة المستوى (أعلى 48٪)، والثانية : منخفضة المستوى (أقل 48٪) . وتم حساب متوسط تقدير القدرة لكل عينة ، حيث يوضح جدول (١٥) التالي نتائج اختبار (ت) لدلالة الفرق بين متوسطي قدرة العينتين المرتفعة والمنخفضة لبنك الأسئلة واختبار حجم التأثير (مربع إيتا) .

جدول (١٥)
اختبار "ت" لدلالة الفرق بين متوسطي قدرة العينتين المرتفعة والمنخفضة المستوى لبنك الأسئلة

الفئة	المتوسط	الانحراف المعيارى	درجات الحرية	ت	الدلالة	حجم التأثير
مرتفعي القدرة	51.337	1.311	313.052	27.527	0.001	0.71
منخفضي القدرة	42.099	5.481				

ويتضح من جدول (١٥) السابق وجود فروق دالة بين مستوى أداء المجموعتين المرتفعة والمنخفضة المستوى على بنك الأسئلة لصالح المجموعة المرتفعة المستوى ، وهذا يعني أن هذا البنك يميز بين مستويات الأفراد ، كما يعد ذلك مؤشراً على حساسية البنك للتمييز بين الأفراد ، كما كان حجم الفرق كبيراً مما يؤكد وجود فروق جوهرية بين المجموعتين المرتفعة والمنخفضة المستوى وهو ما يعني أيضاً بدوره أن بنك الأسئلة يميز بين مستويات الأفراد.

٢- **نتائج حساب ثبات بنك الأسئلة في صورته النهائية :**

اعتمد حساب تقدير ثبات الصور الاختبارية (بنك الأسئلة) على ثلاث أساليب كما يلي :

أ- **ثبات تقديرات بارامترات المفردات :**

كما تم الاشارة سابقاً استخدم معامل الثبات الهامشي أو الحدي لتحديد مدى ثبات بارامترات مفردات بنك الأسئلة (الصور الاختبارية) وفقاً للنموذج الأكثر دقة وملاءمة وهو النموذج اللوغاريتمي ثلاثي البارامتر ، ويستخدم هذا المعامل في تحديد ثبات الاختبارات المدرجة وفقاً لنظرية الاستجابة للمفردة حيث يعتمد في حسابه على الخطأ المعياري لتقدير بارامترات المفردات أو القدرة ، ويمكن تفسيره بنفس الطريقة التي يفسر بها معامل الثبات التقليدي إلا أن الأخير لا يأخذ في اعتباره اختلاف الخطأ المعياري للتقدير . ويوضح جدول (١٦) ارتفاع ثبات تقديرات بارامترات مفردات الصور الاختبارية (بنك الأسئلة) في صورته النهائية كالتالي :

جدول (١٦)
ثبات تقديرات بارامترات مفردات بنك الأسئلة في صورته النهائية

البيان	معامل الثبات الهامشي
ثبات الصعوبة	0.90
ثبات التمييز	0.95
ثبات التخمين	0.97

ب- **ثبات تقديرات قدرة الأفراد :**

استخدم أيضا معامل الثبات الهامشي أو الحدي لتحديد مدى ثبات قدرة الأفراد باستخدام النموذج اللوغاريتمي ثلاثي البارامتر ويوضح جدول (١٧) قيمة معامل الثبات لتقديرات قدرة الأفراد على بنك الأسئلة في الصورة النهائية مما يؤكد ثبات تقديرات قدرة الأفراد.

جدول (١٧)
ثبات تقديرات قدرة الأفراد على بنك الأسئلة فى صورته النهائية

البيان	معامل الثبات الهامشي
ثبات تقديرات القدرة	0.62

ج- **ثبات تقديرات القدرة باستخدام دالة معلومات الاختبار :**

تستخدم دالة معلومات لاختبار Test information Function **والمنحنى الدال عليها كمؤشر لثبات تقديرات القدرة على اختبار ما ، حيث أنها مقلوب مربع الخطأ المعياري والذي يشير إلى معامل الخطأ المعياري في تقدير قدرات العينة** (Vukmirovic , 2009). **ومن ثم فإن كمية المعلومات المقدمة بواسطة الاختبار عند مستوى القدرة (θ) تكون مرتبطة عكسياً مع الدقة عند هذه النقطة على متصل القدرة (θ)** (Hambleton et al. , 1991 : 93). **لذلك فقد استخدمت دالة معلومات الاختبار في تحديد ثبات تقديرات القدرة على الصور الاختبارية (بنك الأسئلة). حيث يبين شكل (١) السابق دالة معلومات الاختبار وفقاً للنموذج اللوغاريتمي ثلاثي البارمتر ، ويتضح أنها جيدة بشكل عام حيث بلغت أعلى قيمة لدالة المعلومات** 63 . **ويلاحظ أن منحنى الخطأ المعياري يبدأ في الانخفاض من أعلى قيمة له وهي** 1.12 **عند مستوى قدرة** 3 **، ثم يصل إلى أدنى قيمة عند مستوى قدرة** 0.6− **، ثم يبدأ في زيادة مطردة من** 0.6− **إلى** 3− **، كذلك يلاحظ أن مدى قيم منحنى الخطأ المعياري للقياس تتراوح من** 0.04 **إلى**1.17. **وبذلك يتضح أن منحنى الخطأ المعياري معتدل بشكل لا يؤثر على دقة القياس وهذا يشير إلى أن تقدير قدرات الأفراد على الصور الاختبارية كان ثابتاً بالقدر الكافي الذي يسمح بالاعتماد عليه في المقارنة بين الأفراد.**

الـخـــلاصة :

لقد تبين من نتائج الدراسة الحالية أن النموذج اللوغاريتمي ثلاثي البارمتر أكثر النماذج الثلاثة ملاءمة للبيانات وفقاً لمتطلبات برنامج التحليل الإحصائي بارسكيل وذلك طبقاً للكشف عن مدى ملاءمة الأفراد والمفردات المكونة لبنك الأسئلة لكل نموذج من النماذج اللوغاريتمية الأحادي ، والثنائي ، والثلاثي البارامتر . كما كان أفضل من النموذجين الثنائي والأحادي في دقة تقدير صعوبة المفردات وقدرة الأفراد كما يتمثل ذلك في قلة الخطأ المعياري لتقديري الصعوبة والقدرة لمفردات بنك الأسئلة وهو نفس النتائج التي تم التوصل إليها فيما يتعلق بدوال المعلومات وتقدير مؤشرات الثبات.

ومن ثم فقد أمكن للدراسة تدريج صعوبة مفردات بنك الأسئلة والحصول على تقديرات قدرات الأفراد وفقاً للنموذج اللوغاريتمي ثلاثي البارمتر ، وحساب صدق وثبات بنك الأسئلة في صورته النهائية

بعد حذف استجابات الأفراد غير الملائمة وكذا المفردات غير الملائمة لأسس القياس الموضوعي باستخدام الطرق المتعددة في حساب الصدق والثبات .

بحوث مقترحة :

١- يمكن إجراء دراسات أخرى لاختبار مدى إمكانية تدريج بنوك للأسئلة تتضمن أنواع مختلفة من الأسئلة الموضوعية أو أسئلة المقال باستخدام نماذج الاستجابة للمفردة ، وباستخدام مواد تعليمية مختلفة.

٢- إجراء دراسات أخرى تستخدم في تحليل بياناتها برامج أخرى غير برنامج التحليل الإحصائي باركسيل المستخدم فى الدراسة الحالية.

٣- إجراء دراسات أخرى حول إمكانية بناء بنوك الأسئلة ، والبحث في متغير دقة القياس باستخدام نماذج أخرى غير النماذج اللوغاريتمية الأحادية والثنائية والثلاثية البارمتر .

التوصيات :

في ضوء النتائج التي انتهت إليها الدراسة الراهنة ، يمكن اقتراح التوصيات التالية :

١. التدريب على بناء بنوك الأسئلة وكيفية الاستخدام الأمثل لها من أجل تنمية قدرة الأفراد على حل المشكلات واتخاذ القرارات في المواقف الأكاديمية المختلفة ، والتدريب على برامج التحليل المختلفة لبنوك الأسئلة والاختبارات التحصيلية مثل برامج Parscale , Winsteps , Bilog ، وغيرها من البرامج الهامة في مختلف مراحل التعليم العام والجامعي .

٢. الاستفادة من بنوك الأسئلة (تجمع ضخم من بنود الأسئلة الممثلة لمتغير ما وتتدرج هذه البنود على تدريج واحد مشترك بصفر واحد مشترك) في تشخيص صعوبات التعلم لدى الأفراد والتحصيل المنخفض .

٣. إعادة بناء وتدريج أدوات القياس النفسي والتربوي باستخدام نماذج الاستجابة للمفردة المختلفة .

٤. استخدام النموذج اللوغاريتمي ثلاثي البارامتر مع أسئلة الاختيار من متعدد وذلك لدوره الفعال في علاج قضية التخمين والتي تؤثر على استجابات الأفراد لهذا النوع من الأسئلة ، وبصفة خاصة الأفراد متوسطي ومنخفضي القدرة .

E-mail: gamel_abdo59@yahoo.com ☎ 01002702115

٥. كانت من المشكلات التي واجهت الباحثان عند بداية الدراسة العينات المناسبة للتحليل وعلاقاتها بعدد المفردات المشتركة وعدد الأفراد المشتركين ، لذلك توصي الدراسة الحالية في التوسع في الدراسات الخاصة بشروط التحليل باستخدام الأفراد المشتركين

٦. كما توصي الدراسة الحالية بأن يتم تطوير برنامج التحليل الإحصائي بارسكيل وذلك للتغلب على أوجه القصور ومنها عدم قدرته على حساب الثبات لصعوبة المفردات وقدرة الأفراد والحصول على قدرة الفرد المقابلة لكل درجة كلية محتملة في تدريج بنك الأسئلة .

المراجع

١- أحمد عبد الحافظ عطا الله الزغيلات (٢٠١٤). أثر عدد البدائل على دقة تقديرات مؤشرات الصعوبة والتمييز في اختبار اختيار من متعدد وفق النموذج ثنائي المعلم. عالم التربية ، ١٥(٢): ٢٥٣- ٢٧٩.

٢- إسماعيل حسن الوليلي (٢٠٠٢). دراسة سيكومترية مقارنة لبعض نماذج الاستجابة للمفردة في انتقاء مفردات الاختبارات مرجعية المحك. رسالة دكتوراة غير منشورة ، القاهرة: كلية التربية ، جامعة الأزهر.

٣- اعتدال غازي عبد الله (٢٠٠٠). القياس محكي – المرجع لأحد مقررات علم النفس التعليمي ، دراسة ميدانية. رسالة ماجستير غير منشورة ، القاهرة: كلية البنات ، جامعة عين شمس.

٤- اعتدال غازي عبد الله (٢٠٠٣). استخدام نموذج راش في تدريج مقياس للقدرة العقلية لدراسة بعض العوامل المؤثرة على دقة القياس. رسالة دكتوراة غير منشورة ، القاهرة: كلية البنات ، جامعة عين شمس.

٥- أمين محمد صبري نور الدين (٢٠٠٢). فاعلية استخدام الاختبار الموائم باستخدام الحاسب في تقدير قدرة الأفراد وتحديد الخصائص السيكومترية للمقياس. رسالة دكتوراة غير منشورة ، القاهرة: كلية التربية، جامعة عين شمس.

٦- أمينة محمد كاظم (١٩٨٨).دراسة نظرية نقدية حول القياس الموضوعي للسلوك ، ط١. الكويت: مؤسسة الكويت للتقدم العلمي.

٧- أمينة محمد كاظم (١٩٩٥). اتجاهات معاصرة في بناء بنوك الأسئلة. في : الأسس التربوية لإعداد المعلم الجامعي. القاهرة، جامعة عين شمس، ط٢: ٢٤٧-٢٦٥.

٨- أمينة محمد كاظم (١٩٩٦)، مستوى العينة وتدرج بنك الأسئلة باستخدام نموذج (راش) (دراسة تجريبية). في: أنور الشرقاوي ، وأمينة كاظم ، وسليمان الخضري، ونادية عبد السلام. القياس والتقويم النفسي والتربوي. القاهرة: مكتبة الأنجلو المصرية ، ٥٤٧-٥٨٣.

٩- أمينة محمد كاظم (٢٠٠٠). اتجاهات معاصرة في بناء بنوك الأسئلة. في الأسس التربوية لإعداد

المعلم الجامعي، القاهرة: جامعة عين شمس ، الطبعة الثالثة ، ٣٢١- ٣٤٠.

١٠- أنور محمد الشرقاوي (١٩٩٦). الاختبارات المرجعة إلى محك: وسائل جديدة في القياس النفسي والتربوي. في: أنور الشرقاوي وأمينة كاظم وسليمان الخضري ونادية عبد السلام. اتجاهات معاصرة في القياس والتقويم النفسي والتربوي . القاهرة : مكتبة الأنجلو المصرية.

١١- تقي أحمد العديلات (٢٠١٢). بناء بنك أسئلة في الرياضيات لطلبة الصف الرابع وفقا لنماذج نظرية الاستجابة للفقرة. رسالة ماجستير. كلية التربية. جامعة اليرموك.

١٢- تيسير مفلح كوافحة (٢٠٠٣). القياس والتقييم وأساليب القياس والتشخيص في التربية الخاصة ، الطبعة الأولى. عمان : دار المسيرة للطباعة والنشر.

١٣- جابر عبد الحميد جابر (٢٠٠٦). اتجاهات وتجارب معاصرة في تقويم أداء التلميذ والمدرس. القاهرة : دار الفكر العربي.

١٤- جواهر محمد الزيد (٢٠٠٨). فاعلية نظرية الاستجابة للمفردة في بناء الاختبارات "إعداد بنك أسئلة في مقرر علم النفس التربوي". رسالة دكتوراه غير منشورة. كلية العلوم الاجتماعية . جامعة الإمام محمد بن سعود الإسلامية.

١٥- حازم جريس البقاعين (٢٠١٠). بناء بنك أسئلة في الرياضيات للصف الثاني الثانوي للفروع :الأدبي، الإدارة المعلوماتية، الصحي والشرعي في الأردن في ضوء نظرية استجابة الفقرة باستخدام النموذج اللوغاريتمي ثلاثي المعلمة. رسالة ماجستير. عمادة الدراسات العليا. جامعة مؤتة.

١٦- حمدي يونس أبو جراد (٢٠١٢). الخصائص السيكومترية لاختبارات الاختيار من متعدد والتكميل دراسة مقارنة باستخدام نظرية الاستجابة للمفردة . مجلة العلوم التربوية والنفسية ، ١٣ :٤٠٧- ٤٣٤.

١٧- رانيا ماهر محمد وهدي (٢٠٠٥). دراسة سيكومترية مقارنة بين نموذج راش والقياس التقليدي حول دقة التنبؤ بحالة القلق من سمة القلق ، رسالة ماجستير غير منشورة ، القاهرة : كلية البنات ، جامعة عين شمس.

١٨- رجاء محمود أبو علام (١٩٨٧). قياس وتقويم التحصيل الدراسي، الكويت: دار القلم.

١٩- رجاء محمود أبو علام (١٩٩٨) . مناهج البحث في العلوم النفسية والتربوية ، القاهرة : دار النشر للجامعات.

٢٠- رشدي فام منصور(١٩٩٧). حجم التأثير الوجه المكمل للدلالة الإحصائية . المجلة المصرية للدراسات النفسية ، القاهرة : مكتبة الأنجلو المصرية ، العدد ١٦ : ٥٧- ٧٥.

٢١- رشدي فام منصور(١٩٨٧) . تعدد زوايا الرؤية عند تفسير درجات الاختبار وعند بنائه – ضرورة تربوية. مجلة علم النفس ، القاهرة : الهيئة المصرية العامة للكتاب ، العدد الأول ، السنة الأولى :٢١- ٢٣.

٢٢- رضا عبد القادر الصاوي (٢٠٠٤). قياس مهارات الاتصال لدى طلاب الجامعة في ضوء نظرية الاستجابة للمفردة ، رسالة ماجستير غير منشورة : كلية التربية بالسويس، جامعة قناة السويس.

٢٣- زيد الهويدي(٢٠٠٤). أساسيات القياس والتقويم التربوي ، الطبعة الأولى . العين : دار الكتاب الجامعي.

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

٢٥- صبري حسن الطراونة (٢٠١٦). بناء اختبار محكي المرجع في الإحصاء التحليلي لطلبة الدراسات العليا في كليات العلوم التربوية في الجامعات الأردنية الحكومية باستخدام نظرية استجابة الفقرة. مجلة العلوم التربوية والنفسية،١(١٧)،٥١٥- ٥٤٢.

٢٦- صلاح أحمد مراد، محمد منصور الشافعي (١٩٩٨). أثر حجم العينة في دقة وكفاءة ضم اختبارين في تدريج مشترك . مجلة البحوث النفسية والتربوية ، كلية التربية ، جامعة المنصورة ، العدد الثاني ، السنة الثالثة عشر: ٩٦- ١٤٤.

٢٧- صلاح الدين علام (١٩٨٦). تطورات معاصرة في القياس النفسي والتربوي . الكويت : جامعة الكويت.

٢٨- صلاح الدين علام (١٩٨٩). بعض الاتجاهات المعاصرة في مجال إعداد وبناء نظم بنوك الأسئلة. حلقة إقليمية حول تطور نظم الامتحانات وبناء بنوك الأسئلة في الدول العربية . القاهرة : المركز القومي للبحوث التربوية.

٢٩- صلاح الدين علام (١٩٩١). استخدام نموذج راش في بناء اختبار تشخيصي مرجعي المحك للمعارف الأساسية في إعداد خطة البحوث التربوية والنفسية. في: صلاح الدين علام. الاختبارات التشخيصية مرجعية المحك في المجالات التربوية والنفسية والتدريبية ، القاهرة : دار الفكر العربي ، ٤٥٣- ٤٥٨.

٣٠- صلاح الدين علام (٢٠٠١). الاختبارات التشخيصية مرجعية المحك في المجالات التربوية والنفسية والتدريبية ، الطبعة الأولى . القاهرة : دار الفكر العربي.

٣١- صلاح الدين علام (٢٠٠٢). القياس والتقويم التربوي والنفسي أساسياته وتطبيقاته وتوجهاته المعاصرة ، الطبعة الأولى . القاهرة : دار الفكر العربي.

٣٢- صلاح الدين علام (٢٠٠٥). نماذج الاستجابة للمفردة الاختبارية أحادية البعد ومتعددة الأبعاد وتطبيقاتها في القياس النفسي والتربوي ، الطبعة الأولى . القاهرة : دار الفكر العربي.

٣٣- عاشم فلاح عثمان (٢٠٠٦). بناء بنك أسئلة لمبحث الرياضيات للصف الثاني الثانوي العلمي باستخدام نظرية الاستجابة إلى المفردة. رسالة ماجستير غير منشورة : كلية الدراسات العليا. الجامعة الأردنية.

٣٤- عبد الناصر سند العكايلة (٢٠٠٧) . دراسة مقارنة بين النظرية الكلاسيكية والنظرية الحديثة في القياس في كشف وتقدير الخطأ المعياري في اختبارات القدرات المعرفية لطلبة المرحلة الأساسية في الأردن . رسالة دكتوراة غير منشورة : جامعة القاهرة ، مصر.

٣٥- علاء الدين عبد الحميد أيوب (٢٠٠١). دراسة سيكومترية للمهارات المعرفية باستخدام نظرية الاستجابة للمفردة IRT . رسالة ماجستير غير منشورة : كلية التربية أسوان ، جامعة جنوب الوادي.

٣٦- علي ماهر خطاب (٢٠٠١). القياس والتقويم في العلوم النفسية والتربوية والاجتماعية ، الطبعة الثانية . القاهرة : مكتبة الأنجلو المصرية.

٣٧- عماد عبد المسيح يوسف (١٩٩١). استخدام نموذج (راش) اللوغاريتمي أحادي البارامتر في تحليل مفردات الاختبارات المعرفية مرجعية المعيار ثنائية القطب (دراسة تجريبية) . مجلة البحث في التربية وعلم النفس ، كلية التربية ، جامعة المنيا، العدد الرابع ، ٤٤٣- ٤٧٨.

٣٨- ماهر إسماعيل صبري ، محب محمود الرافعي (٢٠٠٨) . التقويم التربوي "أسسه وإجراءاته". المملكة العربية السعودية : مكتبة الرشد.

٣٩- محمد سعيد زيدان (٢٠٠٩). تقويم بنك الأسئلة في المواد الفلسفية بالمرحلة الثانوية. مجلة الجمعية التربوية للدراسات الاجتماعية ، العدد ٢٠ ، ١٥- ٧٣.

٤٠- محمد منصور الشافعي (١٩٩٦).أثر طرق معادلة درجات الاختبار وضوابط اختيار العينة علي تدريج بنود بنك الأسئلة باستخدام نموذج راش. رسالة دكتوراه غير منشورة: كلية التربية، جامعة المنصورة.

٤١- محمد منصور الشافعي (٢٠١١). تأثير نماذج التحليل اللوغاريتمية (الأحادية والثنائية والثلاثية) البارمتر علي تقديرات دوال معلومات المفردات الاختبارية وعلاقتها بدقة تقديرات القدرة (دراسة مقارنة). المجلة المصرية للدراسات النفسية ، المجلد ٢١، العدد٧٣ ، ٥٤٩ – ٦١٨.

٤٢- منى ربيع الطنطاوي (٢٠٠٠). دراسة سيكومترية حول تطوير اختبار المصفوفات المتتابعة لرافن باستخدام نموذج راش. رسالة ماجستير غير منشورة : كلية البنات ، جامعة عين شمس.

٤٣- منى ربيع الطنطاوي (٢٠٠٨).استخدام نظرية الاستجابة للمفردة في تقييم فاعلية برنامج لتنمية الاستدلال الاستقرائي لدى عينة من تلاميذ المرحلة الابتدائية. رسالة دكتوراه غير منشورة : كلية البنات ، جامعة عين شمس.

٤٤- نادية محمد عبد السلام (١٩٩٦). بناء اختبار هدفي المرجع وجماعي المرجع دراسة مقارنة بين النموذجين. في: أنور الشرقاوي وأمينة كاظم وسليمان الخضري ونادية عبد السلام. اتجاهات معاصرة في القياس والتقويم النفسي والتربوي . القاهرة : مكتبة الأنجلو المصرية.

٤٥- نبيل جمعة النجار (٢٠٠٦). بناء بنك أسئلة في الثقافة الحاسوبية للمرحلة الثانوية في الأردن باستخدام نماذج نظرية الاستجابة للمفردة "دراسة مقارنة بمعلمة ومعلمتين". رسالة دكتوراه غير منشورة : كلية التربية ، جامعة اليرموك.

٤٦- نسرين إسماعيل إبراهيم (٢٠١٠). تأثير بعض المتغيرات على دقة القياس باستخدام نظرية الاستجابة للمفردة . رسالة دكتوراه غير منشورة : كلية البنات، جامعة عين شمس.

٤٧- هالة يوسف مندور ، محمد فتح الله أحمد، نسرين عبد الحميد عاشور(٢٠١٣). بناء وتدريج بنك

أسئلة لمقرر الهوكي باستخدام نظرية الاستجابة للمفردة IRT في ضوء معايير الجودة. المجلة المصرية للدراسات النفسية ، المجلد ٢٣، ٣٦٥–٤١٣.

٤٨- هشام فتحي محمد جاد الرب (١٩٩٩). تطوير اختبار كاتل للذكاء باستخدام نماذج السمات الكامنة وأثر ذلك على قدرة الاختبار على التنبؤ بالتحصيل الدراسي. رسالة ماجستير غير منشورة : كلية التربية ، جامعة المنصورة.

٤٩- هند عبد الحميد الحموري (٢٠١١). دراسة استكشافية لملاءمة نماذج نظرية الاستجابة للمفردة في بناء اختبار تحصيلي من إعداد المعلم . مجلة العلوم التربوية والنفسية ، (١٢) ، ٤٧- ٨١.

٥٠- وليد أحمد مسعود (٢٠٠٤). دراسة سيكومترية لتطوير اختبار رسم الرجل باستخدام نموذج راش . رسالة ماجستير غير منشورة : كلية البنات ، جامعة عين شمس.

٥١- وليد أحمد مسعود (٢٠١٠). دراسة أثر القياس الدينامي على فاعلية برنامج لتنمية الفهم القرائي باستخدام نظرية الاستجابة للمفردة . رسالة دكتوراه غير منشورة : كلية البنات ، جامعة عين شمس.

52- Adedoyin , O. O. (2010) . Using IRT Approach to Detect Gender Biased Items in Public Examinations : A case Study for the Botswana Junior Certificate Examination in Mathematics . Educational Research and Reviews . 5(7) , 385-399.

53- Aiken , L. R. (1982) . Writing Multiple - Choice Items to easure High- only Educational Objective . Educational and Psychological Measurement . 42 , 803-806.

54- Aime , M. (2008) . A Comparison of (CAT) Scores based on Different Measurement Models for test lets . Educational and Psychological Measurement . 31(3) , 371-398.

55- Anastasi , A. (1982). Psychological Testing, ٥th edition . New York : Macmillan.

56- Anastasi, A. (1990) . Psychological Testing, 6th edition . Macmillan Publishing Company. New York .

57- Anastasi , A. & Urbina , S. (1997) . Psychological Testing 7th edition . New Jersey : Prentic- Hall.

58- Anderson , E. B. (1996) . The Rating Scale Model. in Vander

E-mail: gamel_abdo59@yahoo.com ☎ 01002702115

Linden & Hambleton , R.K. (Eds.) Handbook of Item Response Theory . New York , Springer , 67-84.

59- Andrich , D. (1998) . Rasch Models for Measurement . Beverly Hills , London: Stage Publication.

60- Azevedo , C. L. N. (2009) . Some Observation on the Identification and Interpretation of the 3PL IRT Models . Applied Psychological Measurement . 33(7) , 89-114.

61- Bachman , L. F. (1997) .Fundamentals Consideration in Language Testing , 4th edition . London: Oxford University Press.

62- Baker, F. (2001). The Basic of Item Response Theory. University of Maryland: College Park , MD. Eric Clearing House 0n Assessment and Evaluation.

63- Bay , Luz; Loomis & Susan Cooper (1995) .Scoring With Nominal Missing – Response Parameter .43p paper Presented at the Annual Meeting of the American Educational Research Association (San Francisco, Apr.18-22, 1995).

64- Bishop, N. ; Scott Omar & Mid Hafidz (2002). Comparing Scales Derived from Dichotomous and Polytomous IRT Models for a Test Composed of Test lets. 43 paper Presented at the Annual Meeting of the National Council 0n Educational Measurement (New Orleans ,Apr. 2-4 , 2002).

65- Bloom ,B. B. (1974) .Taxonomy of Educational Objective : The Classification of Educational Goals . Handbook I '' Cognitive Domain ''. New York : David Mckay.

66- Brennan , R. L. (2004) . Some Perspectives on Inconsistencies Among Measurement Models in CASMA. Research Report No. 8 , College of Education, University of Iowa , Iowa City.

67- Brown , J. D. (1997) . Computers in Language Testing . present Research and Some Future Direction. Language Learning & Technology . v(1) , 44-59.

68- Chang , Shun-wen; Hanson, Bradley A.; Harris; Deborah, J. (2001. A Comparison of the Standardization and I. R.T Methods of Adjusting Pretest Item Statistics Using Realistic Data . 26p. paper Presented at the Annual Meeting of the American Educational Research Association (Seattle, Apr.10 - 14,2001).

69- Childs, R. A.(1993) . Applying Item Response to the analysis of Educational Test Items with External Criterion .Dissertation

Abstract International . 53(9).p5000.

70- Cook , J. L. (2000) . Educational Psychology: Effective Teaching , Effective Learning, 3th edition. University of Wisconsin-WhiteWater, Mc Graw Hill.

71- Courville, T. G. (2004). An Empirical Comparison of Item Response Theory and Classical Test Theory Item Person Statistics.in Partial Fulfillment of The Requirement for the Degree of Doctor of Philosophy, Texas A&M University.

72- Daniel , R. C . & Embretson, S. E. (2010) . Designing Cognitive Complexity in Mathematical Problem - Solving Items. Applied Psychological Measurement . 34(5) ,348-364.

73- De Ayala , R. J. & Others (1988) . A Comparison of the Nominal and Graded Response Models in Computerized Testing . 33p.; paper Presented at the Annual Meeting of the American Educational Research Association (New Orleans , Apr.5-9,1988).

74- De Ayala , R. J. & Others (1989). A Comparison of the Graded Response and Partial Credit Models for Assessing Writing Ability. 33p.; paper Presented at the Annual Meeting of the National Council on Measurement in Educational (San Francisco, Mar. 28-30,1989).

75- De Ayala, R. J. & Others (1990). A Comparison of the Partial Credit and Graded Response Models in Computerized Adaptive Testing. 17p.; paper Presented at the Annual Meeting of the American Educational Research Association (Boston ,Apr.16-20,1990).

76- De Ayala, R. J. (2009). The Theory and Practical of Item Response Theory. NY. The Guilford press.

77- De Champlain, A. (2009) .A Primer on Classical Test Theory And Item Response Theory for Assessments in Medical Educational . Blackwell Publishing Ltd , Medical Education, 44(1) ,109-117.

78- De Mars, C. (2001). Group Differences based on IRT Scores Does the Model Mater? Educational and Psychological Measurement, 61 (1): 60-70.

79- Du Toit ,M. (2003). IRT from SSI: Bilog-Mg, Multilog, Parscale, Test Fact. U.S.A : Scientific Software International Inc.

80- Edward , H. (2010). Interpretation of the Three –Parameter Testlet Response Model and Information Function. Applied Psychological Measurement,34(7): 467-482.
81- Embretson, S. E. & Reise , S. P. (2000) . Item Response Theory for Psychologists . London, Lawrence Erllaum Associates.
82- Fan , A. (2009). Item Response Theory: an Empirical Comparison of Their Item Statistics . Educational and Psychological Measurement , 44 (2): 256-270.
83- Good , C.V. (1973) . Dictionary of Education. New York:Mc Graw Hill.
84- Green , D., R.; Yen, W., M. & Burket, G. (1989). Experiences In the Application of Item Response Theory, in Test Construction. Applied Measurement in Education 2(4) , 189-199.
85- Gronlund, N. E. & Linn, R. L. (1994). Measurement and Evaluation Teaching, 6th Edition . London: Collier Macmillan Publishers.
86- Gronlund , N. E. (1998) . Assessment of Student Achievement , 6th Edition . University of Illinois : Allyn & Bacon.
87- Hambleton , R. K. &Swaminathan, H. (1985). Item Response Theory Principles and Application. Boston, Kluwer Nijhoff Publishing.
88- Hambleton, R. K . (1989). Principles and Selected Application of Item Response Theory . in Linn, R. T. (ed) Educational Measurement . 3rd Edition . New York; Macmillan Publishing Company.
89- Hambleton, R. K.; Swaminthan, H. & Rogers, H. J.(1991). Fundamentals of Item Response Theory. Newbury Park, California, U.S.A.: Sage Publications.
90- Harris , D. (1989). An NCME Instructional Module : Comparison of 1, 2 & 3 Parameters IRT Models. Instructional Topics in Education Measurement from National Council of Measurement in Education. 157-163. Retrieved on January 8, 2010 from www .ncme. org / pubs / items.13.pdf.
91- Hashway, R. M. (1998). Assessment and Evaluation of Developmental Learning Qualitative Individual Assessment And Evaluation Models. U.S.A.: Library of Congress.
92- Henson , R. k. (1999). Understanding the One -Parameter Rasch

E-mail: gamel_abdo59@yahoo.com ☎ 01002702115

Model of Item Response Theory. paper Presented at the Annual Meeting of the South west Educational Research Association. San Antonio, Texas , January 21-23.

93- Hills, J. R. (1981). Measurement and Evaluation in The Classroom , 2nd edition .Columbus: Bell & Howell Company.

94- Hulin, C.L., Drasgow, F. & Parson , C. K. (1983) . Item Response Theory Application to Psychological Measurement. U.S.A.: Dow Jones – Irwin, the Dorsey Professional Series.

95- Jaccard, J. & Becker, M. A. (2002). Statistics for the Behavioral Sciences, 4th Edition. U.S.A : Wadsworth Thomson Learning.

96- Jason , L. , Mayers, G. , Edwared , M. & Walter, D. (2009) . Item Position and Item Difficulty Change in An IRT-Based Common Item Equating Design. Applied Measurement in Education , 22(3) , 38-60.

97- Jing , Y. (2007) . A comparison between the vertical scaling of test Sensitive to multiple dimensions using common - item and common - group designs. PHD. Dissertation , Michigan State University, United States.

98- Kline , P. (1993). The Handbook of Psychological Testing. London : Rout ledge.

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

100- Kubiszyn , T. & Borich, G. (2003). Educational Testing and Measurement: Classroom Application and Practice, 7th Edition . University of Taxas at Austin , John Wiley & Sons,Inc.

101- Lawson , D. M. (2006) . Applying the Item Response Theory to Classroom Examinations.Journal of Manipulative and Physiological Therapeutics, V (29) , 393-397.

102- Lee , Yong Won (2004) . Examining Passage Related Local Item Dependence (LID) and Measurement Construct Using Q3 Statistics in an EFL Reading Comprehension Test. Language Testing , 21(1), 74-100.University of Alberta , U.S.A : Wadsworth Thomson Learning.

103- Lehman , I. J. & Mehrens, W.A. (2003). Measurement and Assessment in Education and Psychology , 3rd Edition . New York: Holt.

104- Loken , E. & Kelly, L. (2010). Estimation OF A Four - Parameter

E-mail: gamel_abdo59@yahoo.com ☎ 01002702115

Item Response Theory Model . British Journal of Mathematical and Statistical Psychology , 63(2),509-525.

105- Lord , F. M. & Stocking , M. L. (1988) . Item Response Theory . In Walberg , H. J. & Hartel , G. D. The International Encyclopedia of Educational Evaluation. New York : Pergamon press, 281-284.

106- Mckinley , R . L. & Reckase , M. D.(1983) . An Evaluation of One - and Three – Parameter Logistic Tailored Testing Procedures for use with Small Item Pools . American Coll. Testing Program , Iowa City, Iowa, Report-Research / Technical (143) Aug; 1983.

107- Michael Custer And Others (2008) .A Paradox Between IRT Invariance and Model- Data Fit When Utilizing the One - Parameter and Three - Parameter Models, paper Presented at The Annual Meeting of the American Educational Research Association, New York City , New York , March 28,2008.

108- Mislevy , R. J. (1990). Modeling Item Response When Different Subjects Employ Different Solutions Strategies. Netherlands: Psychometrika, 55(2), June, p.p.195-215.

109- Murphy , K. R. & Davidshofer, C. O. (1998) . Psychological Testing :Principles and Application, 4th Edition. USA: Colorado State University, Prentice- Hall.

110- Nitko , J. Anthony (2001) .Educational Assessment of Students, 3rd Edition. New Jersey , U.S.A: Prentice-Hall Inc.

111- Njtiu, J. & Romanoski , J. (2007). Development and Calibration of Physics Items to Create an Item Bank, Using Rasch Measurement Model. The Fourteenth International Conference on Learning at the University of Witwtersrand, Johannesburg,South Africa.

112- O' brien &Hampilos (1988). The Feasibility of Creating An Item Bank. From a Teacher Made Test , Using The Rasch Model. Educational and Psychological Measurement ,48(1),201-212.

113- Parscale for Windows (2003) . (Computer program) . Version 4.1.2328.4. Scientific Software International Inc.

114- Partchev, I. (2004). A visual guide to Item Response Theory. Friedrich –Schiller-Universitat Iena. Retrieved from

http://www.metheval.uni-jena.de/irt/VisualIRT.PDF
115- Popham , W. J. (2005) . Classroom Assessment : What Teachers Need to Know , 3rd Edition. The Johns Hopkins University Press. Boltimore.
116- Reeve , B. B. (2002) . An Introduction to Modern Measurement Theory .Outcomes Research Branch , Applied Research Program , Division of Cancer Control and Population Science, Science ,National Cancer Institute.
117- Ropert, B. (2010). Comparison Between Item Response Theory Models. Counseling Psychologist , College park , 21(3), 234-260.
118- Rudner , Lawrence . (1998) . Item Banking , ERIC /AE Digest , ERIC Clearinghouse on Assessment and Evaluation Washington DC ,Aug; 1998 , ED 423310.
119- Saba Rizavi ,Way ,W. D. &Ying Lu (2004) . Evaluating 1-, 2- and 3- Parameter Logistic Models using Model - Based and Empirically - Based Simulations under Homogeneous and Heterogeneous Set Conditions . paper Presented at the Graduate Student Research Panel of American Association of Medical Colleges , Washington, DC, October 11, 2002.
120- Shu Jing Yen & Leah Walker (2007) .Multidimensional IRT models for Composite Scores. paper Presented at the2007 Annual Meeting of the National Council of Measurement in Education, Chicago.
121- Skaggs , G. & Lissity, R.W. (1990) . To Match or Not Samples on Ability for Equating . A Discussion of Five Articles, Applied Measurement in Education, 3(1), 105-113.
122- Smith, R. M. (1996). A Comparison of the Rasch separate Calibration and between fit methods of detecting item bias. Educational and Psychological Measurement, V (56) , 403- 418.
123- SPSS 15.0 for Windows (2006). Results Coach (Computer Program). SPSS Inc.
124- Stage , C. (2003). Classical Test Theory or Item Response Theory: The Swedish Experience . Paper Published in Spanish by Centro de Estudios Publicos , Santiago , Chile. Available at www. cepchile.cl .
125- Suen , H. K. (1990). Principles of Test Theories . New Jersey,

U.S.A.: Laurence Erlbaum Associates Publishers.
126- Tam , H. P. & Li , Y. H. (1997). Is the Use of the Difference Likelihood Ratio Chi - Square Statistic for Comparing Nested I RT Models Justifiable ? 27p; paper Presented at the Annual Meeting of the American Educational Research Association (Chicago , Mar. 24- 28,1997).
127- Tammiee , S. D. & Brian , T. H. (2006). Rasch and 3pl Ability estimates when the true model Is Multidimensional. paper Presented at the Annual Meeting of the American Educational Research Association , San Francisco CA , April.
128- Thissen, D. & McLeod, L. D. (2001). Test Scoring: Logistic Response Models for Items with More Than Two Scoring Categories . New Jersey: Lawrence Erlbaum Associates.
129- Thissen, D. & Orlando, M. (2001).Test Scoring :Item Response Theory for Items Scored in Two Categories. New Jersey: Lawrence Erlbaum Associates.
130- Thissen, D. And Others (2001). Test Scoring: Logistic Response Models for Items with More Than Two Scoring Categories. New Jersey: Lawrence Erlbaum Associates.
131- Thissen, D. M. , Steinberg, L. & Fitzpatrick , A.R. (1989). Multiple - Choice Items: The Distractor are also Part of The Item. Journal of Educational Measurement, 26, 161-176.
132- Thorndike, R. M. (1997). Measurement and Evaluation in Psychology and Education , 6th Edition. New Jersey, Prentice- Hall.
133- Thorndike, R. M. And Others (1991). Measurement and Evaluation in Psychology and Education , 5th Edition .New York: Macmillan.
134- Van der Lenden, W. J. & Hambleton, R. k. (1996**).** Handbook of Item Response Theory . New York**,** Springer.
135- Vukmirovic , Z. (October,2009). Item Response Theory. Conceptual Introduction. Training Material for Psychometric Workshop . Presented on Psychometric Workshop for NCEEE new Staff. Funded by World Bank -PPMU and provided by American Institutes for Research. International Development Division. USA.
136- Van der Mass, Han, L.(2011). Cognitive Psychology Meets Psychometric Theory :On the Relation Between Process Models

E-mail: gamel_abdo59@yahoo.com ☎ 01002702115

for Decision Making and Latent Variable Models for Individual Differences. Psychological Review, 118(2) , 339-356.

137- Waller, M. I. (1986). Latent Trait Models and Random Guessing: A Comparison of Two Logistic Latent Trait Models. ERIC Clearinghouse on Assessment and Evaluation, Washington DC, Apr;1986, ED278706.

138- Way, W.D. & Reese, C. M. (1991). An Investigation of the Use of Simplified IRT Models for Scaling and Equating The TOEFL Test . Educational Testing Services, Princeton, N. J., Feb.1991.

139- Weiss , D. J. & Yoes , M. E. (1990) . Item Response Theory. In Hambleton , R .K. & Zaal , J. N. (Eds) Advances in Educational Testing . Boston , Kluwer Academic Publisher, pp.69-95.

140- Westfall, P. J. (1989). Re- estimation of Student Ability in Foreign languages Using the Rasch Model. Dissertation Abstract International , 49(9) .p. 2631.

141- Whiston ,S. C. (2000). Principles and Application of Assessment in Counseling , 4th Edition .USA: Brookcole.

142- Wiberg , M. (2004). Classical Test Theory Vs. Item Response Theory: An Evaluation of The Theory Test in The Swedish Driving- License Test . Umea University.

143- Winsteps (2008) .Rasch Model Computer program. Version 3.6.7.0. Scientific Software International Inc.

144- Wright, B.D. (1992). IRT in the 1990s: Which Models work best. Rasch Measurement Transactions, 6(1), 196-200.

145- Wright , B.D. (1988). Rasch Measurement Models. in Walberg, H. J. & Hartel ,G. D. (Eds). The International Encyclopedia of Educational Evaluation .New York : Pergamon Press, 289-295.

146- Wright, B. D. & Linacre, J. M. (1985). Microscale Manual , Version 2, Mediax Interactive Technology.

147- Wright, B.D. & Stone , M. H. (1979). Best Test Design: A handbook for Rasch Measurement . Chicago :Meta Press.

E-mail: gamel_abdo59@yahoo.com ☎ 01002702115

مجلة البحث في التربية وعلم النفس
كلية التربية – جامعة المنيا

The Effect Of Analysis By Logarithmic Models (One , Two And Three) Parameter On Accurately Calibrating Items Bank In Mathematics

Abstract

This study aims to compare among the three models of Item Response Theory (one, two, three) parameter logistic models on calibrating objective questions Items bank in mathematics according to accuracy. Accuracy can be represented through: Decreasing standard error of difficulty and ability estimation, Increase of reliability indexes, and Increase of items information functions. The Sample of the study consisted of 1500 fifth-grade primary stage pupils collected from Suez governorate , Egypt. The computer programs, Parscale version 4.1 and SPSS version 10 were used in making the required statistical data treatment according to (one, two and three parameter) logistic models. The study found that the form of three - parameter logistic model is the best than other models in the light of discovering the fit of items and person's which forming Items bank, The process of calibration involved eliminating 9 items from the Item bank that showed statistical misfit to the three parameter logistic model, so the calibrated Item bank consisted of 69 items. The same results reached with respect to three variables: the standard errors, the reliability indexes, and the values of test information function and therefore as indicators of the accuracy calibrating the items bank, it was found that the form of three - parameter model led his analysis to the lower estimate of the standard errors for items difficulty and persons's ability estimation and the estimates for the reliability and values of test information function were higher than those estimated including its form of others. Finally it was possible to Calibrate and estimate items difficulty and person's ability of items bank according to the three parameter logistic model 3PL , which is the most accurate IRT model, and calculate the validity and the reliability of items bank in its final form after deleting invalid person's and unfit items on the bases of the objective measurement using various methods in Calculating the validity and the reliability according to the three parameter logistic model , which is the most accurate IRT model.

E-mail: gamel_abdo59@yahoo.com ☎ 01002702115