أثر عدد مفردات الاختبار على تقدير قدرات الأفراد ودالة المعلومات لاختبار تحصيلي مرجعي المحك في الرياضيات باستخدام نماذج نظرية الاستجابة للمفردة

إعداد

هبه إبراهيم محمد على الناغى

إشراف

الأستاذ الدكتور
أحمد عبد الرحمن إبراهيم عثمان
أستاذ علم النفس التربوي
كلية التربية – جامعة الزقازيق

الدكتورة
شيرين محمد أحمد دسوقي
أستاذ علم النفس التربوي المساعد ورئيس
قسم علم النفس
كلية التربية – جامعة بورسعيد

المقدمة

لقد حظي مجال تطوير أدوات وأساليب القياس والتقويم النفسي والتربوي باهتمام كبير سواء من قبل الهيئات والمؤسسات البحثية والتربوية، أو من قبل الباحثين والعلماء سواء في المجتمعات المتقدمة أو النامية.

وقد أكدت الدراسات والبحوث التى أجريت في مجال القياس والتقويم النفسي والتربوي أنه على الرغم من أن الاختبارات التحصيلية تعد أقدم النظم المعروفة في التربية والتعليم كوسائل للقياس والتقويم، كما أنها تعد أداة جيدة للحصول على البيانات والمعلومات عن صفات أو سمات تربوية معينة وهامة كالتحصيل الدراسي، إلا أن هذه الاختبارات قد عانت من العديد من المشكلات(مثل: تقيد درجة الفرد على الاختبار بمفرداته، وتقيد الخصائص السيكومترية للمفردات الاختبارية بخصائص أفراد عينة التقنين، وعدم توافر خطية القياس، وتعدد بعدية القياس، وتغير تكوين ومعنى مفردات الاختبار بمرور الزمن.....) التى أثرت تأثيراً سلبياً على دقتها وموضوعيتها، الأمر الذى أفقدها قيمتها لدى الكثير من أولياء الأمور والطلاب.

وعلى الرغم من انتشار استخدام نظرية الاختبارات التقليدية في بناء وتحليل البيانات المشتقة من الاختبارات التحصيلية المختلفة، إلا أنه تبين قصور هذه النظرية في مواجهة هذه المشكلات، بالإضافة إلى قصورها في الوفاء بمتطلبات الكثير من التطبيقات المعاصرة مثل: بناء بنوك الأسئلة، وبناء الاختبارات مرجعية المحك، والكشف عن دالة المفردة المميزة(تحيز المفردة)، وبناء الاختبارات الموائمة بالحاسب، بالإضافة إلى معادلة درجات الاختبارات التحصيلية.

وقد فطن علماء القياس النفسي والتربوي إلى خطورة تلك المشكلات، وما يترتب عليها من أضرار بالغة بناء على نتائج غير دقيقة نتجت عن تلك الاختبارات التحصيلية المستخدمة بشكل واسع في مؤسساتنا التربوية والتعليمية، الأمر الذى يؤدى إلى اتخاذ قرارات غير صائبة. وبناء على ذلك، تضافرت الجهود البحثية للتغلب على تلك المشكلات إلى حد ما بغرض الوصول إلى الموضوعية المنشودة في القياس النفسي والتربوي، الأمر الذى أدى إلى ظهور العديد من التطورات المعاصرة من أهمها نظرية الاستجابة للمفردة Item Response Theory لمواجهة قصور نظرية الاختبارات التقليدية في معالجة الكثير من المشكلات السيكومترية، ومواجهة الانتقادات العديدة التى وجهت لبناء الاختبارات التحصيلية وتحليلها وتفسير نتائجها، بالإضافة إلى قصورها في تلبية متطلبات التطبيقات المعاصرة في مجال القياس والتقويم.

ويندرج تحت نظرية الاستجابة للمفردة مجموعة من النماذج الرياضية تعرف باسم نماذج الاستجابة للمفردة Item Response Models وهذه النماذج تحقق القياس الموضوعي، وتعالج

كثير من مشكلات القياس النظرية والتطبيقية التى عجزت نظرية الاختبارات التقليدية عن مواجهتها، ولعل أكثر هذه النماذج استخداماً وانتشاراً في مجال القياس والتقويم النفسي والتربوي، النموذج اللوغاريتمى أحادى المعلم، والنموذج اللوغاريتمى ثنائي المعلم، والنموذج اللوغاريتمى ثلاثي المعلم.

وهذه النماذج تصف التفاعل بين قدرة الفرد الممتحن ومفردات الاختبار من خلال دالة احتمالية تربط بين معلمين إحداهما يتعلق بالفرد يعرف بمعلم القدرة(θ)، والأخر يتعلق بخصائص المفردة الاختبارية المتمثلة في: معلم الصعوبة(b)، ومعلم التمييز(a)، ومعلم التخمين(C)، ومن ثم تهدف هذه النظرية إلى الوصول إلى قيم تقديرية عددية واحدة تتعلق بالفرد وتسمي بمعلم القدرة أو السمة المقاسة لدى الفرد "**Ability Parameter**"، وقيمة تقديرية عددية واحدة أو أكثر تتعلق بالمفردة الاختبارية وتسمي بمعلم أو معلمات المفردة "**Item Parameter**".

وهناك من العوامل المتعلقة بأداة الدراسة التى ربما تؤثر في استجابة الأفراد للمفردة، ومنها عدد المفردات المكونة لأداة الدراسة، وعدد المفردات كما يمكن أن يكون لها تأثير في استجابة الأفراد لهذه المفردات، فلها تأثير أيضاً على نتائج عملية القياس والخصائص السيكومترية للاختبار كالصدق والثبات، والخصائص السيكومترية لمفردات الاختبار كالصعوبة والتمييز والتخمين، وهذا التأثير ربما يتباين تبعاً لنظرية القياس المستخدمة في عملية تحليل البيانات المشتقة عن هذا الاختبار، والتي يتبناها الباحث في بناء وتفسير نتائج الأدوات المستخدمة في بحثه.

وقد وجد تناقض بالنسبة لنتائج الدراسات والبحوث السابقة التى اهتمت ببحث أثر عدد مفردات الاختبار على تقديرات قدرات الأفراد في ضوء نماذج الاستجابة للمفردة، حيث أشارت بعض الدراسات والبحوث إلى أن تقديرات قدرات الأفراد تكون أكثر دقة في حالة الاختبارات الطويلة، بينما خلصت بعض الدراسات والبحوث إلى أن تقديرات معلمات المفردات وقدرات الأفراد أكثر دقة في حالة الاختبارات القصيرة. في حين أظهرت بعض الدراسات عدم وجود تأثير لاختلاف عدد المفردات المكونة للاختبار على تقدير معلمات قدرات الأفراد.

ويعد مفهوم دوال المعلومات لبنة أساسية في إطار عمل نظرية الاستجابة للمفردة، وتمثل دالة معلومات الاختبار مجموع دوال المعلومات لكل مفردة من المفردات المكونة للاختبار، بمعنى أن كل مفردة من المفردات الاختبارية تسهم بصورة مستقلة في دالة معلومات الاختبار. وهذا لا يعني أنه بزيادة عدد المفردات المكونة للاختبار ترتفع قيمة المعلومات المقدمة من هذا الاختبار عند مستوي معين من القدرة، ولكن يشترط لهذا أن تتمتع هذه المفردات بخصائص سيكومترية جيدة، وهذا يشير إلى وجود تأثير محتمل لعدد المفردات الاختبارية على دالة معلومات الاختبار.

وبناء على ذلك، هدفت الدراسة الحالية إلى بحث أثر عدد مفردات الاختبار على تقدير قدرات الأفراد، ودالة المعلومات لاختبار تحصيلي مرجعي المحك في وحدة "المصفوفات" المقررة ضمن منهج الجبر لطلاب الصف الأول الثانوي العام باستخدام النموذج اللوغاريتمى ثلاثي المعلم، وذلك من أجل الإفادة من خصائص ومزايا نظرية الاستجابة للمفردة في تحقيق القياس الموضوعي المنشود للأهداف السلوكية المتعلقة بالوحدة الدراسية المختارة.

مشكلة الدراسة:

وفي ضوء ما سبق، تتحدد مشكلة الدراسة الحالية في الأسئلة التالية:

١- هل تختلف تقديرات قدرات الأفراد على الاختبار التحصيلي مرجعى المحك باختلاف عدد مفردات الاختبار باستخدام النموذج اللوغاريتمي ثلاثي المعلم؟

٢- هل تختلف دالة معلومات الاختبار التحصيلي مرجعى المحك باختلاف عدد مفردات الاختبار باستخدام النموذج اللوغاريتمي ثلاثي المعلم؟

أهداف الدراسة:

تهدف الدراسة الحالية على وجه التحديد إلى:

١- الكشف عن أثر اختلاف عدد مفردات الاختبار على تقدير قدرات الأفراد على الاختبار التحصيلي مرجعي المحك باستخدام النموذج اللوغاريتمي ثلاثي المعلم.

٢- دراسة أثر اختلاف عدد مفردات الاختبار على دالة معلومات الاختبار التحصيلي مرجعي المحك باستخدام النموذج اللوغاريتمي ثلاثي المعلم.

أهمية الدراسة:

يمكن تحديد أهمية الدراسة الحالية في جانبين، هما:

أولاً: الأهمية النظرية:

تكمن الأهمية النظرية للدراسة الحالية في:

١- تناول اتجاه حديث ومعاصر في القياس النفسي والتربوي يتمثل في نظرية الاستجابة للمفردة، واستخدام واحد من أكثر نماذج الاستجابة للمفردة واقعية، هو "النموذج اللوغاريتمي الثلاثي المعلم"، في تحليل البيانات المشتقة من الاختبار التحصيلي مرجعي المحك، في محاولة لإلقاء الضوء على أهمية تحقق متطلبات القياس الموضوعي بدلاً من استخدام نظرية الاختبار الكلاسيكية(التقليدية).

٢- تحاول الدراسة الحالية التوصل إلى أداة قياس موضوعي لقياس تحصيل طلاب الصف الأول الثانوي العام في وحدة "المصفوفات" المقررة ضمن منهج الجبر، وفقاً لفلسفة القياس مرجعي المحك باستخدام نماذج الاستجابة للمفردة، مما يساهم في إلقاء مزيد من الضوء على أهمية الاختبارات مرجعية المحك في تشخيص نقاط القوة والضعف في أداء المتعلم ومدى اتقانه للأهداف الدراسية بالوحدة الدراسية المختارة.

٣- إلقاء الضوء علي البرامج الإحصائية المتقدمة والمستخدمة في تحليل البيانات الاختبارية في ضوء نماذج الاستجابة للمفردة، للحصول على تقديرات معلمات المفردات الاختبارية وقدرات الأفراد، والمنحنيات المميزة للمفردة، ودوال المعلومات وغيرها من التحليلات الإحصائية بدقة وكفاءة عالية.

ثانياً: الأهمية التطبيقية:

تتمثل الأهمية التطبيقية للدراسة الحالية فيما يلي:

١- تسلط الدراسة الحالية الأضواء على ضرورة العناية باستخدام الاختبارات مرجعية المحك في الدراسات والبحوث التربوية، وداخل المؤسسات التربوية والتعليمية.

٢- تزويد معلمي مادة الرياضيات بمرحلة التعليم الثانوي العام باختبار تحصيلي مرجعي المحك في وحدة "المصفوفات" المقررة ضمن منهج الجبر على طلبة الصف الأول الثانوي العام، ذات خصائص سيكومترية جيدة وفقاً لنظرية الاستجابة للمفردة، ويمكن الاستفادة منه في تشخيص نقاط القوة والضعف في أداء الطلاب، كما يمكن استخدامه كنواة لإنشاء "بنك أسئلة" Item Bank في الرياضيات يسمح بسحب مفردات، وتكوين اختبارات موائمة تبعاً لمستوى قدرة كل طالب.

٣- تزود نتائج الدراسة الحالية مصممي الاختبارات التحصيلية ببيانات ونتائج عن أفضل طول للاختبار المستخدم، بما يؤدي إلى دقة النتائج المستمدة منه أكثر.

مصطلحات الدراسة:

١- نماذج نظرية الاستجابة للمفردة Item Response Theory Models:هي دوال رياضية تتصف بأنها احتمالية وليست حتمية، تهدف لتحديد علاقة رياضية بين أداء الفرد على الاختبار، وبين السمات أو القدرات التى تكمن وراء هذا الأداء وتفسره، أي أن العلاقة التى تحددها الدالة في ضوء هذه النماذج تسير وفق نظرية الاحتمالات، وتمثل هذه الدالة بعائلة من منحنيات الترجيح اللوغاريتمي Logistic Curves(صلاح الدين محمود علام، ٢٠٠٥: ٦٥).

٢- النموذج اللوغاريتمي ثلاثي المعلم Three-Parameter Logistic Model:هو أحد نماذج الاستجابة للمفردة أحادية البعد، والتي تستخدم في تحليل المفردات ثنائية الاستجابة(الدرجة)، ويفترض أن المفردات الاختبارية تتفاوت في صعوبتها وقوتها التمييزية، وكذلك في احتمالية التخمين، وهذا يسمح بتقاطع المنحنيات المميزة للمفردات، ومن ثم يستخدم هذا النموذج في صيغته الرياضية ثلاثة معلمات للمفردة هي الصعوبة والتمييز والتخمين، بالإضافة إلى معلم تقدير القدرة للفرد الممتحن.

٣- معلم القدرة Ability Parameter:هو مستوى القدرة أو السمة المقاسة لدى الأفراد المختبرين الذين يستجيبون على مفردات الاختبار المعد لقياس هذه السمة (Reeve, 2002: 3).

٤- دالة معلومات الاختبار Test Information Function (TIF):تمثل هذه الدالة بعلاقة منحنية بين متغيرين هما مستويات القدرة التى يمثلها المحور الأفقي، والمعلومات المقدمة من الاختبار ككل والتي يمثلها المحور الرأسي، وهى تعبر عن كمية المعلومات المقدمة من المجموع الكلي لمفردات الاختبار(كمية المعلومات المقدمة بواسطة الاختبار بالكامل عند أي مستوى للقدرة) لكل فرد(صلاح الدين محمود علام، ٢٠٠٥: ١١٧).

٥- الاختبار مرجعي المحك Criterion – Referenced Test:تعرف الباحثة الاختبار مرجعي المحك إجرائياً في الدراسة الحالية بأنه "الاختبار الذى يعتمد في بناء مفرداته على محك سلوكي متوقع ومحدد(أهداف، كفايات، مهارات، معارف.....)، ويتم تحديد مستوى تمكن الفرد بناء على هذا المحك دون الرجوع إلى أداء أقرانه في نفس الجماعة المعيارية".

الإطار النظري والدراسات السابقة

أولاً: نظرية الاستجابة للمفردة Item Response Theory(IRT):

تعد نظرية الاستجابة للمفردة من النظريات السيكومترية المعاصرة التى ظهرت في الثلث الأخير من القرن الماضي، والتى اتضحت فائدتها في التغلب على العديد من المشكلات التى عجزت النظرية الكلاسيكية للاختبار عن مواجهتها(Reeve, 2002: 5).

ويرجع الفضل في ظهور نظرية الاستجابة للمفردة إلى "ثورستون" Thurstone، أما نظرية الاستجابة للمفردة الحديثة فقد تطورت على يد كل من "لورد" Lord "وبيرنبوم"Birnbaum، وقد قام "راش"Rasch بتقديم اتجاهاً جديداً في تطوير هذه النظرية، ثم أتى من بعده "رايت" Wright وقام بتوسيع منظور "راش" بشكل إضافي فيما يتعلق بتقدير القدرة الكامنة والقياس الموضوعي(Bock, 1997: 31).

وتفترض نظرية الاستجابة للمفردة أن هناك سمات أوخصائص معينة يشترك فيها جميع الأفراد، ولكنهم يختلفون في مقدراها، وعلى الرغم من أن هذه السمات غير منظورة، إلا أن هذه السمات تكمن خلف أداء الفرد على الاختبار، حيث يمكن التنبؤ بقدرة الفرد من خلال أدائه على الاختبار عن طريق درجاته على تلك السمات، ونظراً إلى أن هذه السمات لا يمكن ملاحظتها أو قياسها مباشرة لذا يطلق عليها السمات الكامنة(Schatschneider & et al..., 1999: 441).

وتمتلك نظرية الاستجابة للمفردة أساس نظري قوى ونماذج توزيع احتمالية لنجاح الأفراد على مستوى المفردة، وفي ضوء مسمى نظرية الاستجابة للمفردة لا تجرى تحليلات نماذج الاستجابة للمفردة على مستوى درجات الاختبار فقط، ولكن على مستوى المفردة أيضاً، بمعنى أن نظرية الاستجابة للمفردة تركز أولاً على المعلومات على مستوى المفردة، وفي المقابل تركز نظرية الاختبار الكلاسيكية أولاً على المعلومات على مستوى الاختبار(Fan, 1998: 358).

وتعتمد هذه النظرية على فرضية أساسية مفادها أن القيمة الاحتمالية لاستجابة فرد لمفردة اختبارية تكون دالة لكل من السمة أو القدرة(Trait) التى يفترض أن الاختبار يقيسها لدى الفرد وخصائص تلك المفردة التى يحاول الإجابة عنها، وأنه يمكن التنبؤ بأداء الفرد أو تفسير أداءه على الاختبار في ضوء خاصية أو خصائص مميزة لهذا الأداء تسمى السمات، وبذلك فإن هذه النظرية تهدف إلى محاولة تقدير درجات الأفراد في هذه السمات، واستخدام هذه الدرجات في التنبؤ بأداء الفرد على الاختبار والمفردة، حيث أنه من الصعب ملاحظة هذه السمات مباشرة، لذلك يلجأ لتقديرها أو الاستدلال عليها من أداء الأفراد – الذى يمكن ملاحظته– في مجموعة مفردات الاختبار ولذا يطلق عليها السمات الكامنة(عبد الناصر سند العكايلة، ٢٠٠٧: ٥٢).

وفي ضوء هذا يندرج تحت نظرية الاستجابة للمفردة مجموعة من النماذج الرياضية تعرف باسم "نماذج الاستجابة للمفردة الاختبارية Item Response Models"، وهذه النماذج تسمح بالقياس الموضوعى، كما أنها تستخدم في مواجهة كثير من مشكلات القياس التى عجزت نظرية الاختبارات التقليدية عن مواجهتها، وتقوم هذه النماذج على فكرة رئيسة هي أنه يمكن التنبؤ بأداء الفرد على مفردة ما باستخدام خواص هذه المفردة والسمات التى يقيسها الاختبار، أي أن هذه النماذج تحدد العلاقة بين الاستجابة على المفردة الاختبارية والسمة الكامنة التى تقيسها هذه المفردة، بمعنى أنه يمكن اشتقاق تقديرات لمعلمات المفردات والفرد التى تنطوي عليها مجموعة من الاستجابات لمجموعة من المفردات الاختبارية(Reeve, 2002:7).

ويعد التقدير الإحصائي للعلاقة بين احتمال الاستجابة الصحيحة لمفردة اختبارية والسمة التى يفترض أن الاختبار يقيسها هو المشكلة الرئيسة التى تواجه من يستخدم هذه النظرية في بناء الاختبارات وتحليل مفرداتها، وتمثل هذه العلاقة لكل مفردة من مفردات الاختبار بمنحنى يطلق عليه "المنحنى المميز للمفردة"(ICC)، ويتخذ شكل المنحنى اللوغاريتمي التراكمي Logistic Curve، أي أن هذا المنحنى يمثل الانحدار غير الخطى للدرجات التى يحصل عليها المختبرين في مفردة من مفردات الاختبار على القيم التقديرية للسمة أو القدرة أو الخاصة التى يهدف الاختبار لقياسها(صلاح الدين محمود علام، ٢٠٠٧: ٢١٢).

وتقوم نماذج الاستجابة للمفردة على مجموعة من الافتراضات القوية، التى ينبغي التأكد من تحقيقها في البيانات الاختبارية لكي تؤدي إلى نتائج يمكن الوثوق بها، ومن هذه الافتراضات: أحادية البعد، الاستقلال الموضعي(المركزي)، المنحنيات المميزة للمفردات الاختبارية (Johnson, 2007: 4178)

ويعد النموذج اللوغاريتمي أحادي المعلم المعروف بنموذج "راش" Rasch Model أشهر نماذج الاستجابة للمفردة، ويعتمد هذا النموذج على تدريج مفردات الاختبار وفقاً لصعوبتها فقط، أي أن هذه المفردات تتفاوت في مستوى صعوبتها وتتساوى في قوتها التمييزية، كما أنها لا تسمح بالتخمين عند الإجابة عليها. وبناء على ذلك، فإن احتمال استجابة الفرد(j) على المفردة(i) تعتمد على قدرة الفرد (θ) وصعوبة المفردة (b_i) (Van der Linden & Hambleton, 1997: 225).

وفي النموذج اللوغاريتمي ثنائي المعلم يتم إضافة معلم جديد إلى نموذج "راش" وهو معلم التمييز لكل مفردة، وهذا يؤدي إلى تقاطع المنحنيات المميزة لمفردات الاختبار التى تصمم وفقاً لهذا النموذج، أي أن هذا النموذج يفترض انعدام التخمين فقط، ويقوم بتقدير صعوبة المفردات (b_i)، ومعلمات التمييز (a_i)، ومن ثم فإن احتمال الاستجابة (θ) P_{ij} في إطار هذا النموذج هى دالة للقدرة واثنين من معلمات المفردة هى الصعوبة والتمييز(Bock, 1997: 24).

بينما أضاف "بيرنبوم" Birnbaum معلم ثالث إلى النموذج اللوغاريتمي ثنائي المعلم ليفسر الأداء غير الصفري للأفراد منخفضي القدرة على مفردات الاختيار من متعدد، وهذا الأداء اللاصفري يرجع إلى احتمال تخمين الإجابات الصحيحة لمفردات الاختبار، وبالتالي يفترض النموذج اللوغاريتمي ثلاثي المعلم أن مفردات الاختبار تتفاوت في صعوبتها وفي قوتها التمييزية، واحتمالية تخمين الإجابات لتلك المفردات أيضاً، وبالتالي فإن احتمال الاستجابة $P_{ij}(\theta)$ في هذا النموذج هى دالة للقدرة وثلاثة من معلمات المفردة هى الصعوبة والتمييز والتخمين(Partchev, 2004: 34).

وبناء على ذلك، يتضح أن النموذج اللوغاريتمي الأحادي المعلم من أبسط نماذج الاستجابة للمفردة، وأكثرها انتشاراً في الأوساط التربوية إلا أنه أقل فاعلية ومرونة من النماذج الأخرى، فالنموذج اللوغاريتمي الأحادي المعلم يفترض أن المفردات الاختبارية تتباين فقط في صعوبتها وتتساوى في قوتها التمييزية ولا تسمح بالتخمين، وهذا أمر غير واقعي نظراً لصعوبة إيجاد تلك المفردات التى تتباين في معلم الصعوبة فقط، وعلى الرغم من ذلك يعد هذا النموذج أكثر دقة وفاعلية في بناء وتطوير الاختبارات والمقاييس من نظرية الاختبار التقليدية، وفي المقابل نجد أن النموذج اللوغاريتمي ثلاثي المعلم أكثر مرونة وفاعلية ودقة من النموذج اللوغاريتمي(الأحادي–والثنائي) المعلم، وذلك لأنه أكثر النماذج الثلاثة واقعية لاشتراطه أن تكون المفردات الاختبارية متفاوتة في صعوبتها وقوتها التمييزية، كما أنها تعالج قضية التخمين عند الإجابة عن تلك المفردات، أي أن هذا النموذج اللوغاريتمي ثلاثي المعلم يتعامل مع ثلاث معلمات خاصة بالمفردة هي: الصعوبة، والتمييز، والتخمين، بالإضافة إلى معلم تقدير قدرة الفرد.

ويعتمد اختيار النموذج المناسب على مجموعة من المحكات منها: واقعية افتراضات النموذج، تحمل النموذج لمخالفة افتراضاته، أنواع البيانات الاختبارية المراد تحليلها، علاوة على اعتبارات عملية هامة، مثل: خبرة الباحث باستخدام نموذج معين، وتوافر برامج حاسوب لتنفيذ ما يتطلبه النموذج من عمليات إحصائية وحسابية(صلاح الدين محمود علام، ٢٠٠٥: ٨٤ –٨٧).

بالإضافة إلى أن طبيعة المفردات وعددها، وحجم العينة المستخدمة لإجراء التحليل تعد من العوامل الهامة التى ينبغي أن تضع في الاعتبار عند اختيار النموذج المستخدم. فالنموذج الذين يتضمن معلمات أقل يتطلب عدد مفردات أقل وحجم عينة أصغر، وفي المقابل تتطلب النماذج التى تتضمن عدد معلمات أكبر عدد مفردات أكبر وحجم عينة أكبر للحصول على دقة تقديرات المعلمات الخاصة بالمفردة والفرد بأخطاء معيارية طفيفة(Barnes & Wise, 1991: 143).

وقد أصبحت نظرية الاستجابة للمفردة وسيلة أساسية وشائعة في بناء وتطوير الاختبارات، وقد شجع توافر برامج الكمبيوتر على تطبيق النماذج المختلفة لهذه النظرية في مجال الاختبارات والمقاييس، وتتلخص فكرة هذه النماذج في ربطها خصائص المفردات بمعلم أو أكثر، فهى تقدم بديلاً لنظرية الاختبارات التقليدية، في تقديرها لمعلمات المفردة والفرد بأقل قدر من الخطأ، ودن اللجوء إلى عينة مفردات للاختبار عشوائية من نطاق موضوع القياس أو الحصول على عينة كبيرة جداً من المفردات الممثلة لهذا النطاق(Van der Linden & Hambleton, 1997:8).

وحيث إن الاختبارات مرجعية المحك Criterion-Referenced Tests تمثل أسلوباً مهماً يساعد في النهوض بمستوى المتعلمين، حيث تستخدم درجات الاختبار مرجعي المحك الذى يشتمل على عينة عشوائية Random Sample (أو عشوائية طبقية Stratified Random) من المفردات التى تمثل النطاق السلوكي في تقدير درجة الفرد في النطاق الشامل لهذه المفردات Domain Score Estimate، وتستخدم في تصنيف الأفراد بحسب مستويات تمكنهم أو إتقانهم لهذا النطاق، وهذا يتطلب تحديداً مسبقاً لمستويات الأداء في الاختبار Performance Standard(درجات القطع Cut – off Score تفصل بين المتقنين Masters وغير المتقنين Non–Masters من الأفراد)(صلاح الدين محمود علام، ٢٠٠٧: ١٥١ – ١٥٢).

وقد اتفقت معظم الدراسات السابقة التى أجريت في هذا المجال مثل دراسة "هامبلتون وآخرين"(1983)...Hambleton &et al، و دراسة "هالادينا ورويد" Haladyna & Roid (1983) على أن الأساليب والطرق الكلاسيكية غير ملاءمة لتحليل وانتقاء مفردات الاختبار التحصيلي مرجعي المحك، وأن استخدام نماذج الاستجابة للمفردة يمثل أسلوباً جيداً وأكثر دقة لانتقاء مفردات هذا النوع من الاختبارات حتى تكون ممثلة للنطاق السلوكي الشامل، حتى يمكن اتخاذ قرارات تصنيف الأفراد على أساس الدرجة التى يحصل عليها كل فرد في هذا النطاق السلوكي الشامل والمحدد تحديداً دقيقاً.

كما أن المنحنى المميز لكل مفردة من مفردات الاختبار Item Characteristic Curve (ICC) يساعد مصمم الاختبار على فحص المفردات الاختبارية فحصاً دقيقاً للتعرف على الفاعلية النسبية لكل مفردة منها على حدة في دقة القياس وما تضيفه من معلومات إلى بقية مفردات الاختبار في قياس السمة التى يقيسها الاختبار، وبهذا يمكن تفسير درجات الاختبارات التحصيلية التى تصمم وفقاً لتلك النماذج تفسيراً محكي المرجع(صلاح الدين محمود علام، ٢٠٠٥: ٥٩ – ٦٢).

بالإضافة إلى أن مفهوم دوال المعلومات مثل مفهوم دالة الاستجابة للمفردة، فهو مفهوم أساسي في نظرية الاستجابة للمفردة، والمعلومات في القياس النفسي أو في المفهوم الإحصائي تمثل مقدار الدقة في الشئ المقاس، وهى بديل للدقة مع أي معلم يكون مقدراً، وفي نظرية الاستجابة للمفردة من الممكن تحديد كمية المعلومات(أقصى ارتفاع لمنحنى دالة المعلومات) عند مستوى قدرة (θ) معين وتكون أكثر دقة للقياس عند هذا المستوى، ولذلك تظهر أهمية هذه النظرية في قدرتها على تحديد أى المفردات تقيس المتغير بدرجة أفضل عند مستويات محددة للقدرة (θ). (Van der Linden & Hambleton, 1997: 13)

وتمثل دالة معلومات اختبار ما مجموع دوال المعلومات لجميع المفردات المشتملة في الاختبار، حيث إن كل مفردة من مفردات الاختبار تسهم بصورة مستقلة في دالة معلومات الاختبار، ويعتمد مصمموا الاختبارات على مقارنة دوال معلومات الاختبارات المختلفة لتحديد مدى كفاءة وفاعلية الاختبار في قياس القدرة المحددة التى يقيسها، مما يساعد في تقويم تلك الاختبارات وانتقاء الأفضل والأكفء منها، كما أن دوال معلومات الاختبار تمثل وسيلة قوية وفعالة تساعد مصمم الاختبار في تقويم خواص الاختبار وتعديلها بما يتناسب مع أفضل قياس يمكن أن يحققه الاختبار على متصل القدرة، كما أنها تساعد في تقييم مدى الدقة التى يقيس بها الاختبار عند المستويات المختلفة للسمة.(صلاح الدين محمود علام، ٢٠٠٥: ١١٧)

ونظراً لفاعلية ومميزات نماذج الاستجابة للمفردة في بناء وتصميم الاختبارات مرجعية المحك، وكذلك في تحليل وتفسير نتائجها بدقة عالية، فقد استخدمت العديد من الدراسات والبحوث السابقة نماذج الاستجابة للمفردة في عملية معايرة وانتقاء مفردات الاختبارات مرجعية المحك وتقدير قدرات الطلاب، وتحديد درجات القطع سعياً لتحديد مستويات الإتقان المتعلقة بالنطاق السلوكي الذى تقيسه مفردات الاختبار لاتخاذ قرارات تصنيف الأفراد بدقة عالية وذلك سعياً لتحقيق أهدافها مثل دراسة كل من: "جونيوس" Junius(1984)، و"بوزيل" Pozehl(1990)، وصلاح الدين محمود علام(١٩٩٠)، و"تونان وآخرين"Noonan & et al...(1992)، وإسماعيل حسن الوليلي(٢٠٠٢)، و"والكير وبيريتفاس"Walker & Beretvas(2003)، و"وانج"Wang (2003)، و"هيندراون وآخرين" Hendrawn & et al...(2005)، ومنار أحمد طومان(٢٠٠٦)، ومصطفي أحمد طيفور(٢٠٠٧)، وإياد محمد حمادنه(٢٠٠٩)، ومحمد إبراهيم محمد(٢٠١٠)، وعماد عبد المسيح الصفطى(د. ت).

ثانياً: النموذج اللوغاريتمي ثلاثي المعلمThree-Parameters Logistic Model:

إحدى الحقائق الثابتة في الحياة فيما يتعلق بالاختبارات، أنه إذا كان الاختبار يمثل قيمة مرتفعة للأفراد الممتحنين، وأنهم لا يملكون المعرفة الكافية لكيفية الإجابة عن المفردة إجابة صحيحة، فإن معظمهم سيحاولون اللجوء إلى التخمين، والتخمين واحد من أكثر الأسباب الواضحة للاختلاف بين نموذج الاستجابة للمفردة والبيانات الفعلية. وبيانات الحياة – الواقعية من المحتمل أن تكون خليط بين بيانات نماذج نظرية الاستجابة للمفردات وبيانات لبعض نماذج التخمين، لذا عدل "بيرنبوم" Birnbaum(1968) النموذج اللوغاريتمي ثنائي المعلم ليضم معلم جديد يشرح ويوضح إسهام التخمين في احتمال الاستجابة بشكل صحيح على المفردة، والنموذج اللوغاريتمي ثلاثي المعلم

هو محاولة لضبط عامل التخمين، لكن بدون تقديم بعض الصعوبات الإضافية.(Baker, 2001: 28; Partchev, 2004: 46)

ويعد النموذج اللوغاريتمي ثلاثي المعلم امتداداً للنموذج اللوغاريتمي ثنائي المعلم، حيث يمكن الحصول على النموذج اللوغاريتمي ثلاثي المعلم من النموذج اللوغاريتمي ثنائي المعلم، وذلك بإضافة معلم ثالث يرمز له بالرمز(C_i)، وعلى ذلك فيمكن كتابة الدالة الرياضية التى تمثل المنحنى اللوغاريتمي ثلاثي المعلم في الصيغة الرياضة التالية:

$$P_{ij}(\theta)= C_i+(1-C_i)\frac{e^{Dai(\theta \quad j - bi)}}{1 + {}^{\theta} e^{Dai(\ j - bi)}} \quad (i=1,2, \ldots, \quad -١) \quad (١$$

n)

حيث$P_{ij}(\theta)$ تشير إلى احتمال أن يجيب الفرد(j) ذي القدرة(θ) عن المفردة(i) إجابة صحيحة،

b_i هي معلم صعوبة المفردة(i)،

a_i هي معلم تمييز المفردة(i)،

C_i هي معلم الخط التقاربي الأدنى(التخمين) للمفردة(i)،

D = ١،٧ عامل ضبط تداريج القياس(Li & Lissitz, 2004: 89).

ويشير هذا المعلم(C_i) إلى احتمال الوصول إلى الإجابة الصحيحة باستخدام التخمين فقط، ويسمى أيضاً بـ"معلم المستوى الزائف للصدفة" Pseudo – Chance – Level Parameter. ويلاحظ من الصيغة(١-١) أن قيمة(C_i) لا تختلف كدالة لمستوى القدرة، وبالتالي فإن التوصل للإجابة الصحيحة بالتخمين لا يرتبط بارتفاع أو انخفاض مستوى قدرة الأفراد الممتحنين، والمدى النظري للمعلم C هو(١،٠ = > C = > صفر)، ولكن في الممارسة العملية القيمة التى تفوق(٠،٣٥) لا تعد مقبولة، لذا فإن المدى(٠،٣٥ > C = > صفر) هو المستخدم عملياً(Baker, 2001: 30).

والمنحنى المميز للمفردة في إطار النموذج اللوغاريتمي ثلاثي المعلم يوصف بواسطة ثلاثة معلمات تشتق من بيانات تطبيقية رياضية، هي: معلم الصعوبة للمفردة(b_i)، ومعلم تمييز المفردة (a_i)، ومعلم التخمين(C_i)، والذي يأخذ في الاعتبار احتمال أن الأفراد من ذوى القدرة المنخفضة سيجيبون عن المفردة إجابة صحيحة، ويشير معلم تمييز المفردة(a_i) إلى ميل المنحنى ويرتبط عكسياً بالمسافة التى يقطعها الفرد على طول متصل القدرة(θ) بهدف زيادة $P_i(\theta)$، وكلما زادت قيمة(a_i) كلما زاد الميل. أما معلم صعوبة المفردة(b_i) فيشير إلى نقطة على موضع متصل القدرة

عندما يكون احتمال الاستجابة الصحيحة $P_i(\theta)$ مساوياً لـ(٠،٥٠) (Anastasi & Urbina, 1997: 189).

وفي ضوء ما سبق، فإن النموذج اللوغاريتمي ثلاثي المعلم يساعد في الكشف عن مشكلة تخمين الإجابة الصحيحة في حالة المفردات ذات الاختيار من متعدد ومفردات الصواب والخطأ، فإذا كان احتمال إجابة فرد ذا مستوي معين من القدرة المقاسة عن مفردة ما تتجاوز صعوبتها مستوي قدرته إجابة صحيحة، فيمكن اعتبار أن هذا الفرد قد لجأ إلى التخمين في إجابته مع تجنب عامل الغش، ومن ثم نستنتج أن هذه المفردة تعد غير مناسبة لقياس قدرة هذا الفرد.

ثالثاً: عدد مفردات الاختبار The Number of Test Item:

من أهم المتغيرات المرتبطة بأداة القياس، والتى ربما تؤثر على نتائج عملية القياس، والخصائص السيكومترية للاختبار كالصدق والثبات، وعلى الخصائص السيكومترية للمفردات الاختبارية، هي عدد المفردات التى تتضمنها أداة القياس، وهذا التأثير ربما يتباين تبعاً لنظرية القياس المستخدمة في عملية تحليل البيانات الاختبارية، والتى يتبناها الباحث أيضاً في بناء وتفسير نتائج أداة القياس المستخدمة في بحثه.

ويعرف طول الاختبار المرجع إلى المحك بأنه الاختبار الذى يحدد طوله بعدد الأهداف التى يقيسها الاختبار، وكذا بعدد المفردات التى تقيس كل هدف مستقل(محمود محمد عبد الله،١١: ١٩٩٠).

وترى النظرية الكلاسيكية للقياس أنه كلما زاد عدد مفردات الاختبار ارتفع ثبات درجاته، ويمكن التنبؤ بقيمة الثبات عندما يزداد أو يقل عدد مفردات الاختبار باستخدام الصيغة الرياضية التى اقترحها "سبيرمان وبراون" Spearman –Brown formula. أما نظرية القياس المعاصرة فإنها تؤكد أن الاختبارات القصيرة التى تشتمل على نوعية جيدة من المفردات يمكن أن تكون أكثر ثباتاً إذا افترضنا أن تمييز مفرداتها ثابت، كما تؤكد أن الدرجات المتطرفة تكون أخطاؤها المعيارية أكبر. ويمكن استخدام نظرية الاستجابة للمفردات الاختبارية(IRT) بطريقة مباشرة لتحديد العدد الأمثل لمفردات الاختبار التى تفصل بين الأفراد المتمكنين وغير المتمكنين في المنطقة المحيطة بدرجة القطع المحددة(صلاح الدين محمود علام، ٢٠٠٥: ٥٦، ٢٠٠٧: ٣٣٠).

وقد أكدت دراسة "ويلكوكس"(1980)Wilcox، ودراسة "هامبلتون وآخرين" Hambleton & et al...(1983) على أهمية استخدام نظرية الاستجابة للمفردة في تحديد أطوال الاختبارات مرجعية المحك.

وأظهرت دراسة "هولين وآخرين" (1982)...Hulin & et al أن تقديرات معلمات المفردة وقدرات الأفراد أقل دقة في حالة الاختبارات القصيرة، وأكثر دقة في حالة الاختبارات الطويلة باستخدام كل من النموذج(الثنائي– والثلاثي) المعلم. بينما أظهرت نتائج دراسة "سميث" Smith(1988) أن التكرارات النسبية لقيم عنصر البواقي المعيارية لتقدير معلمات المفردة والفرد لا تعتمد على عدد المفردات، وذلك على طول متصل توزيع معلمات المفردة والفرد.

في حين اتفقت نتائج دراسة كل من "بوزيل" Pozehl(1990)، و"تونان وآخرين" Noonan & et al...(1992) على أن طول الاختبار لا يؤثر في دقة نتائج الاختبارات، وأن تأثير طول الاختبار على قيم درجات القطع كان قليلاً في ضوء نماذج الاستجابة للمفردة.

وأوضحت نتائج دراسة "هامبلتون وجونيس" Hambleton & Jones(1994) أن طول الاختبار يؤثر على دقة دوال معلومات الاختبار، حيث ظهرت مشكلات خطيرة فيما يتعلق بالدقة خاصة عندما يتم معايرة مفردات الاختبار عند حجم العينة المتوسط (ن= ٥٠٠ فرداً)، وعندما كانت بنوك مفردات الاختبار كبيرة فيما يتعلق بأطوال الاختبار.

وبينت نتائج دراسة "ميجير وآخرين"(1994)...Meijer & et al أن النسبة المئوية للأفراد المستجبين استجابة شاذة تزداد بزيادة طول الاختبار، كما أن النتائج تكون أقل استقراراً عندما يتناقص طول الاختبار.

وقد خلصت نتائج دراسة "كوتش ودودد" Koch & Dodd(1995) إلى أن تقديرات السمة المقاسة تكون أكثر دقة عندما يكون حجم مستودع المفردات أقصر باستخدام تقديرات الأرجحية القصوى من حجم مستودع المفردات الأطول في ضوء نموذج الفترات المتتابعة. وفي المقابل، أكدت دراسة "ستون وهانسين" Stone & Hansen(2000) على أن الاختبارات الطويلة يتم عن طريقها تقدير القدرة بشكل أدق باستخدام نموذج الاستجابة المتدرجة.

وأشارت منى ربيع متيرد(٢٠٠٠) إلى أن تقدير قدرة الأفراد على اختبار المصفوفات لـ"رافن" بعد تدريجه باستخدام نموذج "راش"، لا تتأثر بعدد المفردات المستخدمة في تقدير هذه القدرة، كما يمكن تقدير القدرة بأحد الاختبارات الفرعية فقط التى يتألف منها الاختبار الكلي بدلاً من استخدام الاختبار ككل بفروعه المختلفة. إلا أن دراسة الشيماء عبد الحميد طنطاوي(٢٠٠٧) أكدت على زيادة تقدير قدرة الأفراد مقدرة باللوجيت بزيادة طول الاختبار باستخدام النموذج اللوغاريتمي أحادي المعلم، كما حقق طول الاختبار(٣٠) مفردة أفضل تمييزاً بين الأفراد مقارنة بكل من طول الاختبار(٦٠) مفردة، و(٩٠) مفردة، وخلصت النتائج إلى أن طول الاختبار يؤثر في دقة نتائج

الاختبارات التحصيلية مرجعية المحك في الرياضيات باستخدام النموذج اللوغاريتمي أحادي المعلم(نموذج راش).

وخلصت دراسة "ماجيز وآخرين" Magis & et al...(2011) إلى أنه عن طريق تصحيح طول الاختبار يمكن معالجة التحيز في تقدير مستويات الكفاءة(القدرة) القصوي العظمي والصغري الناتجة عن الأسلوب البعدي الأقصي كأحد الأساليب البيزية، في ضوء نموذج "راش" كأحد نماذج نظرية الاستجابة.

ويتضح من الدراسات السابقة التى تناولت فحص تأثير طول الاختبار على تقدير قدرات الأفراد، وتحليل نتائج الاختبارات التحصيلية مرجعية المحك، أهمية تحديد طول الاختبار للوصول إلى الدقة المنشودة في عملية المعايرة وتقدير قدرات الأفراد، وكذلك في تحليل البيانات الاختبارية المستمدة من الاختبار التحصيلي مرجعي المحك، وذلك لتحقيق الهدف من استخدام هذا النوع من الاختبارات بكفاءة وفاعلية مما يساعد في اتخاذ القرارات التربوية والتعليمية الصائبة.

رابعاً: تعقيب عام على الدراسات السابقة واشتقاق فروض الدراسة الحالية:

من خلال إطلاع الباحثة على الدراسات والبحوث السابقة التى أجريت في مجال الدراسة، وفحص وتحليل ما جاءت به من أهداف ونتائج، قد استخلصت الباحثة ما يلي من استنتاجات:

١- قلة الدراسات والبحوث العربية – وذلك في حدود علم الباحثة- التى اهتمت بنماذج نظرية الاستجابة للمفردة مقارنة بالدراسات والبحوث التى أجريت في البيئة الأجنبية، سعياً منها للإفادة من إمكانات ومميزات هذه النماذج في مجال القياس والتقويم.

٢- اتفقت معظم الدراسات والبحوث السابقة على فعالية نظرية الاستجابة للمفردة وما انبثق عنها من نماذج في عملية بناء وتطوير الاختبارات والمقاييس مقارنة بنماذج القياس الكلاسيكية(التقليدية)، حيث أنها تستند على أساس نظري وافتراضات قوية، تجعل من هذه النماذج أسلوب متناهي الدقة يساعد في الوصول إلى البناء الأفضل لتصميم وبناء الاختبارات الموضوعية المنشودة في عملية القياس النفسي والتربوي.

٣- ظهر جلياً من خلال بعض الدراسات والبحوث السابقة أهمية استخدام نماذج الاستجابة للمفردة في بناء وتصميم الاختبارات مرجعية المحك وتحليل مفرداتها الاختبارية، خاصة النموذج اللوغاريتمي الثلاثي المعلم، حيث يقدم صورة دقيقة للقياس واستقلاليته في مجال القياس النفسي والتربوي لا تتحقق باستخدام النماذج السيكومترية الكلاسيكية(التقليدية)، ويتضمن معلم جديد يفيد في التمييز بين الأفراد المتقنين وغير المتقنين لمحتوي النطاق السلوكى موضع

القياس، كما يعالج النموذج اللوغاريتمي ثلاثي المعلم مشكلة التخمين عند الإجابة على مفردات الاختيار من متعدد، ويحقق هذا النموذج مطابقة أفضل للمفردات الاختبارية مقارنة بالنموذجين اللوغاريتميين الأحادي والثنائي المعلم، ومن ثم فهو أفضل هذه النماذج لتمثيل النطاق السلوكي المقاس بشكل أوسع مشتملاً على درجة أكبر من عناصر هذا النطاق، وبالإضافة إلى هذا قدمت نظرية الاستجابة للمفردة مجموعة من المفاهيم الاحصائية التى تلعب دوراً هاماً في تحقيق دقة القياس المنشودة مثل دوال معلومات المفردة والاختبار والتى تفيد في تقويم دقة القياس بوجه عام، ودقة تقدير القدرة(θ) من خلال الخطأ المعيارى لتقدير هذه القدرة، كما أنها تساعد في تحديد الكفاءة النسبية للاختبارات سعياً لانتقاء أفضل اختبار يمكن أن يقدم أقصى قيمة من المعلومات عن مستوى قدرة معين وبأدني قيمة من الخطأ المعياري. ومن هنا ظهرت الحاجة إلى استخدام هذه النماذج الرياضية الاحتمالية والقائمة على نظرية الاستجابة للمفردة، وخصوصاً النموذج اللوغاريتمي ثلاثي المعلم، لتحقيق أكبر قدر من الموضوعية والدقة المنشودة في أدوات القياس المستخدمة، مما دفع العديد من الباحثين إلى استخدام هذه النماذج في بناء وتحليل الاختبارات مرجعية المحك.

٤- اختلفت الدراسات والبحوث السابقة في طريقة أو أسلوب الدراسة، فبعض الدراسات اعتمدت على المحاكاة لتوليد بيانات اختبارية تحاكى الواقع باستخدام الحاسب الآلى، وسوف تعتمد الباحثة في دراستها للمتغيرات موضع البحث والدراسة الحالية على بيانات مشتقة من الدراسة الميدانية للواقع الفعلي. وهذا لا يعنى أننا لا نحتاج إلى أسلوب المحاكاة في الدراسات والبحوث، أو أن هذا يقلل من قيمة مثل هذه الدراسات، والتى يكون لها فائدة كبيرة في العديد من التطبيقات والميادين خاصة عند الحاجة إلى استخدام عينات من الأفراد ذات أحجام كبيرة جداً قد يصعب توفيرها.

٥- تباينت العينات المستخدمة في الدراسات السابقة من حيث المرحلة الدراسية، فبعض هذه الدراسات تناولت تلاميذ المرحلة الابتدائية وما قبلها، وبعضها تناول تلاميذ المرحلة الاعدادية، في حين اهتمت بعض الدراسات بطلاب المرحلة الثانوية، كما تناولت بعض هذه الدراسات طلاب المرحلة الجامعية وما بعدها. وهذا يؤكد أن نماذج نظرية الاستجابة للمفردة لا تعتمد على مرحلة عمرية معينة، ومن ثم لا يقتصر استخدامها على مرحلة دراسية معينة دون أخرى.

٦- اختلفت الدراسات السابقة من حيث برامج الحاسب الآلى المستخدمة في تحليل البيانات الاختبارية، وسوف تعتمد الباحثة على برنامج(BILOG-MG V.3.0)لتحليل بيانات الاختبار

في ضوء النموذج اللوغاريتمي ثلاثي المعلم، وذلك نظراً لما يتمتع به هذا البرنامج من مميزات كثيرة.

٧- اختلفت نتائج الدراسات والبحوث التى تناولت فحص تأثير عدد مفردات الاختبار على تقدير قدرات الأفراد في ضوء نماذج نظرية الاستجابة للمفردة، حيث توصلت بعض الدراسات والبحوث إلى أن تقديرات قدرات الأفراد تكون أكثر دقة في حالة الاختبارات الطويلة، بينما خلصت بعض الدراسات والبحوث إلى أن تقديرات قدرات الأفراد أكثر دقة في حالة الاختبارات القصيرة، في حين أظهرت بعض الدراسات عدم وجود تأثير لعدد المفردات المكونة للاختبار على تقدير قدرات الأفراد.

فروض الدراسة:

في ضوء تحليل الإطار النظري والدراسات السابقة وما توصلت إليه من نتائج، يمكن صياغة فروض الدراسة الحالية، على النحو التالي:

الفرض الأول: "تختلف تقديرات قدرات الأفراد على الاختبار التحصيلي مرجعى المحك باختلاف عدد مفردات الاختبار باستخدام النموذج اللوغاريتمي ثلاثي المعلم اختلافاً دالاً إحصائياً."

الفرض الثانى: "تختلف دالة معلومات الاختبار التحصيلي مرجعى المحك باختلاف عدد مفردات الاختبار باستخدام النموذج اللوغاريتمي ثلاثي المعلم."

الطريقة والإجراءات:

أولاً: عينة الدراسة:

تم اشتقاق عينة الدراسة الحالية من طلاب وطالبات الصف الأول الثانوي بجميع مدارس التعليم الثانوي العام بمحافظة بورسعيد، في الفصل الدراسي الأول من العام الدراسي(٢٠٠٩/٢٠١٠م)، حيث تكونت العينة الاستطلاعية للدراسة الحالية من(٥٠٠) طالباً وطالبة من طلاب الصف الأول الثانوي بمدارس التعليم الثانوي العام بمحافظة بورسعيد موزعين تبعاً للجنس كالآتي: (٢٠٠) طالباً، و(٣٠٠) طالبة. وقد اختارت الباحثة طلاب وطالبات العينة الاستطلاعية للتحقق من الخصائص السيكومترية لأدوات الدراسة الحالية، وقد روعي أن تمثل العينة الاستطلاعية العينة النهائية المستخدمة في هذه الدراسة تمثيلاً تاماً. بينما تكونت العينة النهائية من(١٩٣٢) طالباً وطالبة من طلاب الصف الأول الثانوي بمدارس التعليم الثانوي العام موزعين على(٦٨٢) طالباً، و(١٢٥٠) طالبة، وذلك بعد استبعاد(٦٩) طالباً، و(٦٦) طالبة للأسباب الآتية: الطلاب الذين لم يكملوا الاستجابة على أدوات الدراسة الحالية، الطلاب الذين لم يحضروا أحد جلسات التطبيق الميداني لأدوات الدراسة، ازدياد أو نقصان السن عن النطاق العمري المحدد للعينة المستهدفة. وقد

تراوح العمر الزمني لأفراد عينة الدراسة الحالية بين(١٤،٧) إلى(١٦) سنة بمتوسط قدره(١٥،٢٨١٦)، وانحراف معياري قدره(٠،٤٦٣٨)، حيث اقتصرت عينة الدراسة الحالية على طلاب وطالبات الصف الأول الثانوي العام الجدد، وقد استخدمت بيانات هذه العينة في اختبار صحة فروض الدراسة الحالية.

ثانياً: أدوات الدراسة:

قامت الباحثة من أجل التحقق من صحة فروض الدراسة الحالية، وفي سبيل تحقيق أهدافها، ببناء اختبار تحصيلي مرجعي المحك في وحدة المصفوفات. وسوف تعرض الباحثة فيما يلي الخطوات المتبعة في بناء هذا الاختبار التحصيلي وفقاً لفلسفة القياس مرجعي المحك في الوحدة الدراسية المختارة "المصفوفات"، وتقدير الخصائص السيكومترية للاختبار ولمفرداته الاختبارية.

الاختبار تحصيلي مرجعي المحك في وحدة المصفوفات:

تتطلب الدراسة الحالية بناء مجموعة من المفردات الاختبارية Item Pool في الجبر، وقد اختارت الباحثة وحدة "المصفوفات" ضمن منهج الجبر المقرر على طلبة الصف الأول الثانوي العام لتمثل النطاق السلوكي الذى تنتمي إليه مجموعة المفردات الاختبارية، وقد بلغ العدد النهائي للمفردات الاختبارية(٦٦) مفردة من نوع الاختيار من متعدد. وقد اتبعت الباحثة عدد من الخطوات في بناء هذه المجموعة من المفردات الاختبارية، وكانت على النحو التالي:

(١)تحديد النطاق السلوكي المراد قياسه وتحليله إلى مكوناته:

أ- تحديد النطاق السلوكي المراد قياسه:

يعرف النطاق السلوكي بمحتوي الوحدة الدراسية أو الموضوعات المقررة على صف دراسي معين، وهذا المحتوي السلوكي يتضمن المفاهيم، والتعميمات، والمهارات المحددة تحديداً دقيقاً بما يمكننا من تحديد ما يستطيع الفرد أداءه وما لا يستطيع أداءه دون الرجوع إلى مستوي أداء أقرانه في نفس الجماعة المعيارية.

وقد حددت الباحثة وحدة "المصفوفات" المقررة ضمن منهج "الجبر" على طلبة الصف الأول الثانوي العام لتمثل النطاق السلوكي المختار.

ب-تحليل النطاق السلوكي إلى مكوناته:-

استعانت الباحثة في عملية تحليل هذا النطاق السلوكي، بثلاث كتب دراسية، تتمثل في:-

١-كتاب وزارة التربية والتعليم للعام الدراسي٢٠٠٩/٢٠١٠م.

٢-كتاب المعاصر في الرياضيات للعام الدراسي٢٠٠٩/٢٠١٠م.

(عزيز إسحق سرجيوس وآخرون، ٢٠٠٩/٢٠١٠)

٣- كتاب الفائز في الرياضيات للعام الدراسي٢٠٠٩/٢٠١٠م.

(خيري سند، محمد سعيد، ٢٠٠٩/٢٠١٠)

وتضمنت وحدة "المصفوفات" المقررة ضمن منهج الجبر للصف الأول الثانوي العام ثلاثة موضوعات رئيسة، وذلك حسب ورودها في الكتاب المقرر من قبل ورزارة التربية والتعليم، كما يلي: المصفوفات، العمليات على المصفوفات(جمع – طرح)، ضرب المصفوفات.

وقد قامت الباحثة بتحليل محتوي الوحدة الدراسية المختارة في ضوء تصنيف ثلاثي لجوانب التعلم في مادة الرياضيات، والتي تتمثل في العناصر الثلاثة التالية: المفاهيم، والتعميمات، والمهارات.

وبعد عرض تحليل محتوي الوحدة الدراسية المختارة الذى قامت به الباحثة على مجموعة من المحكمين تتكون من(٣٨) محكماً، بلغ عدد المفاهيم(٣٠) مفهوم بنسبة(٢٥،٢١%)، وعدد التعميمات وصل إلى(٢٨) تعميم بنسبة(٢٣،٥٣%)، بينما بلغ عدد المهارات المتضمنة في وحدة "المصفوفات"(٦١) مهارة بنسبة(٥١،٢٦%).

(٢)تحديد المهارات الرئيسة للوحدة الدراسية وتحليلها إلى مكوناتها وتقييمها:

قامت الباحثة في هذه الخطوة بتحديد المهارات الرئيسة المرجو من الطلاب اكتسابها بعد دراستهم لوحدة "المصفوفات"، مسترشداً في ذلك بنتائج تحليل محتوي الوحدة الدراسية وآراء السادة المحكمين المشار إليهم سابقاً، بالإضافة إلى خبرة الباحثة التى اكتسبتها من خلال الاشراف على طلبة التربية العملية بشعبة الرياضيات بمدارس التعليم الثانوي العام للعام الدراسي(٢٠٠٨/٢٠٠٩م)، و(٢٠٠٩/٢٠١٠م)، وتوصلت الباحثة بناء على ذلك، وبعد التحكيم هذه المهارات إلى اثنتي عشرة مهارة ينبغي أن يتمكن منها كل طالب بعد دراسة الوحدة الدراسية، هم على النحو التالي:

١- مهارة التعبير عن المصفوفة وتحديد نظمها وتوظيفها في حل تمرينات تطبيقية متنوعة.
٢- مهارة التعبير عن العنصر داخل المصفوفة وتوظيفها في حل تمرينات تطبيقية متنوعة.
٣- مهارة استخدام المصفوفات في حل مسائل حياتية.
٤- مهارة تعيين بعض المصفوفات الخاصة والتعبير عنها بنظم مختلفة.
٥- مهارة إيجاد مدور المصفوفة واستخدامه في حل تمرينات تطبيقية متنوعة.
٦- مهارة استخدام تساوي مصفوفتين في حل تمرينات تطبيقية متنوعة.
٧- مهارة جمع المصفوفات.
٨- مهارة ضرب عدد حقيقي في مصفوفة وتوظيفها في حل تمرينات تطبيقية متنوعة.

٩- مهارة استخدام خواص عملية جمع المصفوفات في حل تمرينات تطبيقية متنوعة.

١٠- مهارة طرح المصفوفات.

١١- مهارة ضرب المصفوفات.

١٢- مهارة استخدام ضرب المصفوفات وخواصها في حل تمرينات تطبيقية متنوعة.

ونظراً لأن هذه المهارات الرئيسة مهارات مركبة، فقد قامت الباحثة بإجراء تحليل سلوكي لهذه المهارات الرئيسة، ثم بتنظيم هذه المهارات الفرعية والمعارف التى تندرج تحتها في شكل هرمي بحسب أولوية إسهام كل منها في تحقيق المهارة الرئيسة المتعلقة بها.

(٣)صياغة الأهداف السلوكية في ضوء نواتج التحليل السلوكي للمهارات الرئيسة:

قامت الباحثة بعد التحليل السلوكي لكل مهارة رئيسة بصياغة نواتج هذا التحليل صياغة سلوكية، وترتيبها منطقياً حسب أهمية كل هدف سلوكي منها في الإسهام في تحقيق المهارة الرئيسة المتعلقة به، بحيث يمكن الاستدلال منه على تحقق تلك المهارة. وقد مثلت هذه الأهداف السلوكية خمسة مستويات معرفية مختلفة هي: "التذكر، الفهم، التطبيق، التحليل، التركيب". وقد بلغ العدد الإجمالي لهذه الأهداف بعد عرضها على المحكمين(٣٤) هدفاً سلوكياً.

(٤)تقييم الأهداف السلوكية:

قامت الباحثة بعرض الأهداف السلوكية المتعلقة بوحدة "المصفوفات" والبالغ عددها(٣٤) هدفاً على مجموعة من المحكمين، وذلك من أجل التحقق من مدي وضوحها وسلامة صياغتها اللغوية، وملاءمتها للهدف الأساسي من الاختبار الحالي، بالإضافة إلى التحقق من كفايتها في تحقيق المهارات الرئيسة الاثنتي عشرة الخاصة بوحدة "المصفوفات" والموضحة سابقاً. وبناء على ذلك أجرت الباحثة التعديلات التى اقترحها المحكمون، والتي لم تتعدي إجراء بعض التعديلات المتعلقة بالصياغة، وقد بلغ عدد الأهداف السلوكية المتعلقة بوحدة "المصفوفات" في صورتها النهائية(٣٤) هدفاً سلوكياً.

(٥)بناء المفردات الاختبارية:

تعد مجموعة المفردات الاختبارية الترجمة الفعلية والواقعية للهدف السلوكي، فالمفردة الاختبارية تمثل التعريف الإجرائي للهدف السلوكي. لذا فقد قامت الباحثة بدراسة كل هدف سلوكي كل على حدة دراسة متأنية دقيقة، وتحديد كافة الأفكار والطرق التى يمكن من خلالها بناء مفردات اختبارية تقيس هذا الهدف السلوكي. وفي ضوء هذا لم يكن عدد الأسئلة التى تقيس هدفاً سلوكياً معيناً مساوياً للعدد الذى يقيس هدفاً آخر، حيث اعتمد عدد الاسئلة التى تقيس هدفاً سلوكياً معيناً على كم الأفكار والجوانب المتعلقة بهذا الهدف والتي يمكن استخدامها في بناء أسئلة يعتمد عليها

لقياس هذا الهدف من جانب، وعلى صعوبة الهدف ومدي أهميته من جانب آخر. وقد حاولت الباحثة أن تغطي جميع الجوانب المتعلقة بكل هدف سلوكي بحيث تشمل المفردات التى تقيس الهدف السلوكي جميع الأفكار والجوانب الخاصة به.

وقد اعتمدت الباحثة في بناء مجموعة المفردات الاختبارية على المفردات من نوع الاختيار من متعدد ذات البدائل الثلاثة إحداها هي الإجابة الصحيحة للمفردة، وذلك لقياس كل هدف من الأهداف السلوكية التى تمت صياغتها ومراجعتها في الخطوة السابقة. وقد بلغ عدد المفردات الاختبارية التى قامت الباحثة ببنائها(٨٥) مفردة اختبارية وذلك قبل عرضها على السادة المحكمين، وقد تباينت في صعوبتها نظراً لارتباط مستوي صعوبة المفردة الاختبارية بمستوي صعوبة الهدف السلوكي الذى تقيسه ومستواه المعرفي "التذكر، الفهم، التطبيق، التحليل، التركيب".

(٦)تقييم المفردات الاختبارية:

قامت الباحثة بعد بناء مجموعة المفردات الاختبارية، بإعداد استمارة لتقييم مجموعة المفردات الاختبارية، حيث اشتملت على الأهداف السلوكية والمفردات الاختبارية التى يقيسها، حيث يقابل كل هدف سلوكي مجموعة من المفردات الاختبارية التى تقيسه، ثم قامت بتقديمها لمجموعة من المحكمين البالغ عددهم(١٥) محكماً. وقد اقترح المحكمون حذف المفردات الاختبارية ذات الأفكار المكررة أو المتشابهة، والمفردات التى لا تقيس الهدف السلوكي مباشرة، بالإضافة إلى حذف المفردات الصعبة جداً والتي لا تناسب طلاب الصف الأول الثانوي العام. وفي ضوء ذلك، أصبح الاختبار مرجعي المحك في وحدة المصفوفات مكون في صورته الأولية من(٧٨) مفردة، ثم قامت الباحثة بتجميع المفردات الاختبارية في صورتها الأولية، وصياغة تعليمات التطبيق متضمنة الهدف من التطبيق، وعدد المفردات الاختبارية المطبقة، وتعليمات الإجابة عنها، وطريقة تسجيل الإجابات.

(٧)تقدير الخصائص السيكومترية لمجموعة المفردات الاختبارية:

أولاً: ثبات الاختبار التحصيلي مرجعي المحك في وحدة "المصفوفات":

(أ) ثبات المفردات الاختبارية:

للتأكد من ثبات مفردات الاختبار التحصيلي مرجعي المحك في وحدة "المصفوفات" المقررة ضمن منهج الجبر علي طلبة الصف الأول الثانوي، قامت الباحثة بتطبيق الاختبار التحصيلي الحالي المكون من(٧٨) مفردة اختبارية على عينة استطلاعية مكونة من(٥٠٠) طالباً وطالبة، تمهيداً لتحليل استجاباتهم على مستوي المفردة الواحدة، وذلك باتباع الطرق الأتية:

(١) باستخدام "طريقة الاحتمال المنوالي": قد تبين أن جميع قيم الاحتمال المنوالي تشير إلى أن قيم ثبات المفردات تتراوح ما بين(٠،١٥٧) إلى(١)، وجميع قيم معاملات ثبات المفردات هذه دالة

عند مستوي دلالة(٠،٠١). ويلاحظ أن هذه القيم تدل على درجة مرتفعة من ثبات المفردات الاختبارية.

(٢) باستخدام "طريقة الاتساق الداخلي": قد اتضح أن قيم معاملات الارتباط بين درجة المفردة الاختبارية والدرجة الكلية للاختبار تراوحت بين(-٠،٠١١) إلى (٠،٣٥٩)، ووجد أن قيم معاملات الارتباط هذه دالة احصائياً عند مستوي دلالة (٠،٠١)، باستثناء قيم معاملات الارتباط المتعلقة بالمفردتين رقم(١)، ورقم(٣٤) وهي على الترتيب(٠،٠٧٤)، (-٠،٠١١) وهي غير دالة احصائياً، مما يشير إلى ضرورة حذف هاتان المفردتان. ويتضح مما سبق أن قيم معاملات الارتباط بين درجة كل مفردة اختبارية والدرجة الكلية للاختبار، قيم مرتفعة مما يدل على درجة عالية من الاتساق الداخلي للمفردات الاختبارية.

(٣) باستخدام "طريقة ألفا كرونباخ": وقد تبين أن قيمة معامل ثبات ألفا العام للاختبار ككل دون حذف أي مفردة يساوي(٠،٨٦١١)، في حين تراوحت قيم معامل ثبات ألفا العام للاختبار ككل في حالة حذف درجة كل مفردة على حدة ما بين(٠،٨٥٦٠) إلى(٠،٨٦٦١). وبمقارنة قيمة ثبات ألفا العام للاختبار ككل دون حذف أي مفردة بقيمة ثبات ألفا العام للاختبار ككل في حالة حذف درجة كل مفردة، وجد أن هناك بعض المفردات غير الثابتة، حيث أن قيمة معامل ثبات ألفا العام في حالة حذف هذه المفردة يكون أكبر منه في حالة وجودها، مما يؤثر سلباً على ثبات الاختبار، وهذه المفردات تقابل الأرقام(١، ١٦، ٣٣، ٣٤، ٦٦، ٧٥، ٧٧، ٧٨) وعددها(٨) مفردات ليصبح الاختبار التحصيلي مرجعي المحك في وحدة "المصفوفات" مكون من(٧٠) مفردة اختبارية، وذلك بعد حذف المفردات غير الثابتة. وبعد حذف المفردات غير الثابتة تم حساب معامل ثبات ألفا العام للاختبار ككل في حالة حذف درجة كل المفردة على حدة، للتأكد من أن جميع المفردات المتبقية ثابتة. ووجد أن جميع المفردات ثابتة.

ويلاحظ من جميع ما سبق أن قيم ثبات المفردات الاختبارية في الحالات الثلاث السابقة تدل على درجة عالية من الاتساق الداخلي للمفردات الاختبارية.

(ب) الثبات الكلي للاختبار:

تم حساب الثبات الكلي للاختبار التحصيلي مرجعي المحك في وحدة المصفوفات بعد حذف المفردات غير الثابتة بواسطة أربع طرق، على النحو التالي:

(١) حساب معامل الثبات بطريقة التجزئة النصفية: واتضح أن معامل ثبات الاختبار التحصيلي مرجعي المحك باستخدام معادلتي سبيرمان/ براون يساوي(٠،٨٠٧٨)، وباستخدام معادلة "جتمان"

للتجزئة النصفية يساوي(٠،٧٤٥٨)، وهي معاملات ثبات مرتفعة مما يدل على ثبات الاختبار التحصيلي مرجعي المحك ككل.

(٢) حساب معامل التجانس لكيودر وريتشاردسون: وبتطبيق الصيغة(٢٠) لكيودر وريتشاردسون، كانت قيمة معامل التجانس(الثبات الكلي) للاختبار مساوياً (٠،٨٧٨٠٤١٦)، وتشير هذه القيمة إلى معامل ثبات مرتفع للاختبار، ويعد هذا دليلاً على تجانس الاختبار التحصيلي الحالي ككل.

(٣) حساب معامل الثبات بطريقة ألفا كرونباخ: وقد تبين أن معامل ثبات ألفا العام للاختبار ككل بعد حذف المفردات غير الثابتة والتي يجب حذفها يساوي (٠،٨٧٧٤)، وهذه القيمة مقبولة وتعبر عن ثبات درجات مجموعة المفردات الاختبارية ككل بدرجة معقولة،مما يؤكد الثقة في نتائج الاختبار.

(٤) حساب معامل الثبات بطريقة "ليفنجستون" :_وتم حساب قيمة معامل ثبات الاختبار المرجعي المحك في وحدة المصفوفات باستخدام معادلة "ليفنجستون" عند درجات القطع(٤٠%، ٥٠%، ٦٠%، ٧٠%، ٨٠%، ٩٠%)، وهذه هي درجات القطع الشائع استخدامها في الدراسات السابقة لتحديد مستويات الاتقان، وتعبر درجة القطع(٤٠%)عن درجة الاجتياز أو الاتقان في مادة الرياضيات المعتمدة من قبل وزارة التربية والتعليم، وقد كانت قيم معاملات الثبات(٠،٩٩٣٤٩٢٦، ٠،٩٨٩٦٣٢٨، ٠،٩٨١٢٩٦٧، ٠،٩٥٩٥٠٦٩، ٠،٩٠١٤٣٤١، ٠،٨٩٨١٢٩٥)على الترتيب.

ويتضح من نتائج التحليلات السابقة المتعلقة بحساب معامل الثبات الكلي للاختبار، أن قيم الثبات الكلي لمجموعة المفردات الاختبارية مرتفعة، مما يؤكد الثقة في نتائج المفردات الاختبارية للاختبار المستخدم.

ثانياً: صدق الاختبار التحصيلي مرجعي المحك في وحدة "المصفوفات":

(أ) صدق المفردات الاختبارية:

للتأكد من صدق مفردات الاختبار التحصيلي مرجعي المحك في وحدة المصفوفات، تم استخدام استجابات أفراد العينة الاستطلاعية وتحليلها على مستوي المفردة الواحدة، وذلك باتباع الطرق التالية:

(١) عن طريق حساب معاملات ارتباط درجة المفردة بالدرجة الكلية للاختبار في حالة حذف درجة المفردة: وقد تراوحت قيم معاملات الارتباطات هذه ما بين(-٠،٠٧٤٣) إلى(٠،٤٤٣٢)، وجميع هذه القيم دالة احصائياً عند مستوي دلالة(٠،٠١) ماعدا قيم المفردات ذات الأرقام(٤، ١٦، ٢٦، ٦٦، ٧٥) فقد كانت دالة إحصائياً عند مستوي دلالة(٠،٠٥)، وباستثناء قيمتين كانتا غير دالة

إحصائياً وهما(٠،٠٥٢٦) وتقابل المفردة رقم(١)، والقيمة(-٠،٠٧٤٣) وتقابل المفردة رقم(٣٤)، لذا ينبغي استبعاد هاتان المفردتان من التحليل. ويلاحظ أن نتائج هذا التحليل تتفق مع نتائج التحليلات الخاصة بالثبات، ومن ثم ينبغي حذف المفردات الاختبارية غير الثابتة وغير الصادقة المتمثلة في المفردات ذات الأرقام(١، ١٦، ٣٣، ٣٤، ٦٦، ٧٥، ٧٧، ٧٨) ليصبح الاختبار التحصيلي مرجعي المحك مكون من(٧٠) مفردة اختبارية.

(٢) الصدق العاملي: استخدمت الباحثة التحليل العاملي Factor Analysis الذى تم انجازه بواسطة برنامج SPSS(10) باستخدام طريقة المكونات الأساسية Principal Component باعتباره أقوي الأساليب الإحصائية لحساب الصدق.

ووجد أن هناك(١٥) عامل تشبع بها الاختبار، وقد فسرت هذه العوامل نسبة(٤٠،٣٧%) من التباين الكلي. واتضح أن (١٣،٩٧١%) من حجم التباين الكلي المفسر تم استخلاصه بواسطة العامل الأول فقط، والنسبة بين قيمة الجذر الكامن للعامل الأول لقيمته للعامل الثاني(٣،٩٥٨)، بينما النسبة بين قيمة الجذر الكامن للعامل الثاني لقيمته للعامل الثالث(١،٥٧٢)، كما يتضح اقتراب قيمة الجذر الكامن للعوامل بعد العامل الأول، أي العامل الثاني والثالث.....والخامس عشر، مما يؤكد وجود عامل رئيس Dominate first factor ووحيد يقع خلف الأداء على الاختبار التحصيلي. وقد تشبع على العامل العام الأول للاختبار وحده(٦٦) مفردة تشبعاً دالاً احصائياً، وقد تراوحت تشبعات المفردة على هذا العامل ما بين(٠،٣٠١) إلى(٠،٥١٤)، ما عدا المفردات(٤٣، ٦٢، ٦٤، ٧٦). مما يشير إلى تشبع هذا العامل العام الأول للاختبار بحوالي(٩٤،٢٨٦%) من العدد الكلي لمفردات الاختبار، مما يؤكد وجود قدرة واحدة فقط هي التى تقف خلف أداء الفرد في الاختبار التحصيلي.

مما سبق يتضح أن جميع مفردات الاختبار التحصيلي صادقة في قياس أداء الأفراد على هذا الاختبار ما عدا المفردات ذات الأرقام(٤٣، ٦٢، ٦٤، ٧٦)، ليصبح الاختبار في صورته النهائية مكون من(٦٦) مفردة، لقياس مستوي تحصيل طلاب الصف الأول الثانوي العام في موضوع "المصفوفات" المقرر دراسته في فرع "الجبر".

(ب) صدق الاختبار التحصيلي مرجعي المحك ككل:

للتأكد من صدق الاختبار مرجعي المحك في وحدة المصفوفات ككل، والمكون في صورته النهائية من(٦٦) مفردة، بعد التحقق من صدق مفرداته تم استخدام أسلوبين،على النحو التالي:

(١) الصدق الوصفي للاختبار:

يعد الاختبار المرجعي المحك صادقاً إذا أمكن باستخدامه وصف أداء الفرد بالنسبة للنطاق السلوكي الذى يقيسه الاختبار والذي يكون محدداً تحديداً دقيقاً، لذلك يعد أسلوب صدق المحتوي

Content Validity أو ما يسمي بالصدق الوصفي Descriptive Validity مؤشراً جيداً لصدق الاختبارات مرجعية المحك(نادية محمد عبد السلام، ١٩٩٦: ٧٣).

وقد استفادت الباحثة من ملاحظات ومقترحات المحكمين في اجراء بعض التعديلات اللازمة لبعض المفردات الاختبارية التى استلزمت ذلك، مما أدي إلى اتفاقهم على عناصر التحكيم بنسبة(١٠٠%)، وبذلك تعد مجموعة المفردات الاختبارية صادقة وصالحة لقياس ما وضعت لقياسه.

(٢) الصدق التمييزي للاختبار:

تم حساب الصدق الكلي للاختبار التحصيلي مرجعي المحك – وذلك بعد استبعاد المفردات غير الثابتة وغير الصادقة الناتجة عن التحليلات السابقة– عن طريق حساب الصدق التمييزي لدي العينة الاستطلاعية(ن = ٥٠٠ طالباً وطالبة)، وتم استخدام اختبار "ت" لحساب دلالة الفروق بين متوسطي الفئة العليا والفئة الدنيا على الدرجة الكلية للاختبار التحصيلي مرجعي المحك في وحدة المصفوفات، والنتائج موضحة بالجدول رقم(١):

جدول(١)

نتائج اختبار "ت" لدراسة الفروق بين متوسطي الفئتين الطرفتين على الاختبار التحصيلي

الفئة	العدد	المتوسط	الانحراف المعياري	درجات الحرية	قيمة (ت)	مستوي الدلالة
الفئة العليا	١٣٥	٦٤،٠١٤٨	١،٠٩٩٤	١٤٣،٤٧٥	٣١،٣٨١	٠،٠٠٠١
الفئة الدنيا	١٣٥	٤٧،٩٥٥٦	٥،٨٤٣٦			

ويتضح من الجدول السابق وجود فرق دال احصائياً عند مستوي دلالة(٠،٠٠٠١) بين متوسط درجات طلاب الفئة العليا ومتوسط درجات طلاب الفئة الدنيا على الاختبار التحصيلي مرجعي المحك، لصالح مجموعة طلاب الفئة العليا، مما يدل على الصدق التمييزي للاختبار. وهذا يؤكد قدرة الاختبار على التمييز بين المرتفعين والمنخفضين على الاختبار الحالي، مما يشير إلى تحقق الصدق الكلي للاختبار التحصيلي مرجعي المحك في وحدة المصفوفات لدي طلاب الصف الأول الثانوي العام.

وتبين من إجراءات الثبات والصدق – السابق ذكرها– أنه بعد حذف المفردات غير الثابتة وغير الصادقة يصبح الاختبار التحصيلي مرجعي المحك في وحدة المصفوفات مكوناً في صورته النهائية من(٦٦) مفردة اختبارية، لقياس مستوي تحصيل طلاب الصف الأول الثانوي العام في موضوع المصفوفات المقرر دراسته ضمن منهج الجبر.

ثالثاً: إجراءات الدراسة:

تتلخص إجراءات الدراسة الحالية في الخطوات التالية:

١- قامت الباحثة بتطبيق الاختبار التحصيلي مرجعي المحك في وحدة المصفوفات على عينة الدراسة الاستطلاعية المكونة من(٥٠٠) طالباً وطالبة من طلاب الصف الأول الثانوي العام، بهدف التعرف على المفردات التى تحتاج إلى مراجعة وتعديل أو تبديل، بالإضافة إلى التحقق من توافر الخصائص السيكومترية الضرورية في مجموعة المفردات الاختبارية، ثم قامت الباحثة بتطبيق هذه الاختبار على عينة الدراسة في صورتها النهائية والتي بلغت(١٩٣٢) طالباً وطالبة.

٢- صححت الباحثة إجابات طلاب وطالبات العينة النهائية على أداة الدراسة الحالية باستخدام مفتاح تصحيح خاص أعد لهذا الغرض، حيث كانت درجات أي مفردة(سؤال) بالاختبار التحصيلي مرجعي المحك في وحدة المصفوفات إما(صفر) أو (واحد).

٣- تم رصد درجات طلبة العينة النهائية في جدول علي هيئة مصفوفة تتكون من (٦٦) عموداً تمثل المفردات المتعلقة بمفردات الاختبار التحصيلي مرجعي المحك، و(١٩٣٢) صفاً يمثلون العدد الإجمالي لأفراد العينة حيث يمثل(٦٨٢) صفاً في حالة الطلاب،و(١٢٥٠)صفاً في حالة الطالبات.

٤- قامت الباحثة بإدخال نتائج تصحيح أدوات الدراسة على هيئة مصفوفة في الحاسب الآلي باستخدام برنامج SPSS(10) تتكون من(٦٦) عموداً تمثل مفردات كل من الاختبار التحصيلي مرجعي المحك، و(٦٨٢) صفاً في حالة الطلاب، و(١٢٥٠) صفاً في حالة الطالبات.

٥- التحقق من مدي تحقق افتراضات نماذج نظرية الاستجابة للمفردة المتمثلة في: أحادية البعد، الاستقلال الموضعي (المركزي)، المنحنيات المميزة للمفردة، التحرر من تدخل عامل السرعة كعامل في التطبيق. وذلك بالنسبة للبيانات الاختبارية المستمدة من مفردات الاختبار التحصيلي الحالي.

٦- استخدمت الباحثة برنامج الحاسب الآلي "بايلوج – إم جي BILOG-MG V.3.0" في تحليل البيانات الاختبارية المشتقة من تطبيق الاختبار التحصيلي مرجعي المحك في وحدة المصفوفات في ضوء النموذج اللوغاريتمي ثلاثي المعلم، باستخدام مجموعة من الأوامر الخاصة ببرنامج "BILOG-MG V.3.0"، مع إجراء التعديلات المطلوبة في الأوامر حسب كل حالة وفقاً للنموذج المستخدم في التحليل.

٧- تم إنشاء مجلد جديد New Folder ثم نسخ الأيقونة الخاصة ببرنامج "BILOG-MG V.3.0" ووضعها في هذا المجلد، ثم إنشاء ملف جديد للبيانات الخاصة بالاختبار بصيغة Note Pad حيث تم إدخال البيانات في صيغة Text Document بحيث تشغل البيانات الخاصة بكل متغير عدداً متساوياً من الأعمدة دون ترك أية مسافات، ويتم حفظ الملف مع معرفة أسماء المتغيرات التى تشغل كل عمود(وتم تكرار هذه الخطوة لكل اختبار من الأطوال الثلاثة للاختبار التحصيلي الحالي، وذلك بعد حذف الأفراد غير الملائمين للنموذج اللوغاريتمي ثلاثي المعلم).

٨- تم تحليل مفردات الاختبار التحصيلي مرجعي المحك في وحدة المصفوفات بهدف حذف البيانات التامة والصفرية.

٩- إعادة التحليل مرة ثانية بهدف استبعاد الأفراد غير الملائمين في ضوء النموذج ثلاثي المعلم.

١٠- ونظراً لأن عدد مفردات الاختبار يمثل المتغير المستقل في الدراسة الحالية، لذا تم معالجته علي النحو التالي، حيث تم اشتقاق ثلاثة أطوال للاختبار التحصيلي مرجعي المحك بحيث تغطي جميع الموضوعات الدراسية التى تضمنتها الوحدة الدراسية المختارة(المصفوفات) وأهدافها السلوكية، وموزعة على النحو التالي: الاختبار التحصيلي الأول والمكون من(٢٢) مفردة اختبارية، والاختبار التحصيلي الثاني والمكون من(٤٤) مفردة اختبارية، والاختبار التحصيلي الثالث والمكون من(٦٦) مفردة اختبارية، وذلك باستخدام برنامج(10)SPSS.

١١- ولأن عدد مفردات الاختبار يمثل المتغير المستقل في الدراسة الحالية – كما أشرنا من قبل، تم تحليل البيانات الاختبارية المشتقة من الأطوال الثلاثة للاختبار التحصيلي مرجعي المحك في وحدة المصفوفات(٢٢، ٤٤، ٦٦) مفردة اختبارية باستخدام برنامج "BILOG-MG V.3.0"في ضوء النموذج اللوغاريتمي الثلاثي المعلم، وذلك وفقاً لأداء عينة الدراسة الكلية المكونة من(١٩٠٠) طالباً وطالبة بعد استبعاد الأفراد غير الملائمين للنموذج اللوغاريتمي المستخدم.

١٢- إعادة التحليل مرة ثالثة بهدف تحديد المفردات غير الملاءمة للنموذج المستخدم، لدي العينة عينة الدراسة الكلية على الأطوال الثلاثة(٢٢، ٤٤، ٦٦مفردة) للاختبار التحصيلي مرجعي المحك وفقاً لمعامل الارتباط الثنائي الأصيل.

١٣- إعادة التحليل مرة رابعة بهدف الحصول على تقديرات قدرات الأفراد، واشتقاق دالة المعلومات لكل اختبار من الأطوال الثلاثة(٢٢، ٤٤، ٦٦مفردة) للاختبار التحصيلي مرجعي المحك في ضوء النموذج اللوغاريتمي ثلاثي المعلم.

١٤- اختبار صحة فروض الدراسة الحالية باستخدام الأساليب الاحصائية المناسبة لبحث دلالة الفروق بين الأطوال الثلاثة للاختبار التحصيلي مرجعي المحك في وحدة المصفوفات، ومناقشة النتائج وتفسيرها في ضوء الإطار النظري والدراسات السابقة وثقافة المجتمع السائدة.

١٥- صياغة التوصيات والمقترحات في ضوء نتائج الدراسة.

رابعاً: المعالجات والأساليب الاحصائية المستخدمة في الدراسة:

وقد استخدمت الباحثة في اختبار فروض الدراسة الحالية سعياً لتحقيق أهدافها بعض برامج الحاسب الآلي الإحصائية، وهي: برنامج (10)SPSS وذلك لحساب بعض الاحصاءات الوصفية،

والتحليل العاملي للمفردات الاختبارية، واشتقاق تقديرات الثبات المختلفة، واشتقاق معاملات الصدق المختلفة، واشتقاق الاختبارات الفرعية المختلفة المستخدمة في الدراسة،وإجراء تحليل التباين. وبرنامج "BILOG-MG V.3.0" وذلك للحصول على: الأفراد والمفردات غير الملاءمة، وفحص المنحنيات المميزة للمفردات الاختبارية، وتقدير معلمات المفردة(الصعوبة – التمييز – التخمين)، وتقدير قدرات الأفراد، واشتقاق دوال معلومات المفردة ومنحني المعلومات الخاصة بها، واشتقاق دوال معلومات الاختبار ومنحني المعلومات الخاصة بها. وفي الدراسة الحالية استخدمت الباحثة أسلوب الأرجحية القصوى لتقدير معلمات المفردة وقدرات الأفراد في ضوء النموذج اللوغاريتمي ثلاثي المعلم باستخدام البرنامج الإحصائي "BILOG-MG V.3.0" وذلك سعياً للتحقق من صحة فروض الدراسة الحالية.

واعتمدت الباحثة على عدد من الأساليب الاحصائية لوصف عينة الدراسة(الاستطلاعية والنهائية)، والتحقق من ثبات وصدق أدوات الدراسة، وتحليل نتائجها، تمهيداً لاختبار صحة فروض الدراسة الحالية، وذلك في سبيل تحقيق أهدافها، وهي: معامل الارتباط لـ"بيرسون"، معامل الثبات بطريقة ألفا كرونباخ، معامل الثبات بطريقة التجزئة النصفية لـ سبيرمان/ براون – وجتمان، معامل الثبات بطريقة "ليفنجستون"، طريقة الاحتمال المنوالي، أسلوب التحليل العاملي، بعض الاحصاءات الوصفية(المتوسط، الوسيط، الانحراف المعياري، التباين، معامل الالتواء، معامل التفلطح.....)، تحليل التباين.

نتائج الدراسة ومناقشتها:

أولاً: النتائج المتعلقة بالتحقق من توافر افتراضات النموذج اللوغاريتمي ثلاثي المعلم:-

قامت الباحثة بالتحقق من مدى توافر افتراضات نماذج نظرية الاستجابة للمفردة في البيانات الاختبارية المشتقة من الاختبار التحصيلي مرجعي المحك في وحدة المصفوفات، وذلك من خلال النتائج التالية التى تم الوصول إليها.

١- افتراض أحادية البعد: قامت الباحثة باختبار مدى تحقق افتراض أحادية البعد في بيانات الاختبار التحصيلي مرجعي المحك المستخدم في الدراسة الحالية، وذلك باستخدام التحليل العاملي Factor Analysis بطريقة المكونات الأساسية Principle Components الذى تم انجازه بواسطة برنامج(10)SPSS- كما هو موضح سابقاً.

وأسفرت النتائج عن وجود عامل واحد هو العامل العام، وقد تشبع عليه عدد(٦٦) مفردة اختبارية من إجمالي(٧٠) مفردة بتشبعات دالة إحصائياً، أي بنسبة(٩٤،٢٨٦ %) وهذا يحقق شرط أحادية البعد، ووجد أربع مفردات غير مشبعة بالعامل العام ومن ثم يجب حذفها لضمان أن بقية

المفردات تقيس قدرة واحدة فقط وهي "التحصيل الدراسي في وحدة المصفوفات"، والمفردات المحذوفة هى المفردات ذات الأرقام(٤٣، ٦٢، ٦٤، ٧٦).

واستكمالاً للتحقق من مدى افتراض أحادية البعد في البيانات الاختبارية الحالية، قامت الباحثة بحساب درجة الاتساق الداخلي ككل بعد حذف المفردات غير المشبعة بالعامل العام، ووجد أن قيم معاملات ارتباط درجة كل مفردة بالدرجة الكلية للاختبار تتراوح ما بين(٠،٢٣٩) إلى(٠،٥١٤)، وجميع هذه القيم دالة إحصائياً عند مستوى دلالة(٠،٠١). كما تم حساب قيمة معامل ألفا للمقياس في حالة حذف درجة كل مفردة على حدة، وتبين أن قيمة معامل ثبات ألفا العام للاختبار ككل تساوي(٠،٩٠١٩)، بينما تراوحت قيم معاملات ثبات ألفا للاختبار في حالة حذف درجة كل مفردة على حدة ما بين(٠،٨٩٩٢) إلى(٠،٩٠١٨)، وهذه القيم تعد دليلاً على أن بيانات الاختبار الحالي تتحقق فيها خاصية أحادية البعد **Unidimensionality** التى تمثل أحد الافتراضات الرئيسة للنماذج أحادية البعد المنبثقة عن نظرية الاستجابة للمفردة.

٢- افتراض الاستقلال الموضعي(المركزى): ويقصد به عدم وجود مفردات تعتمد في إجاباتها على مفردات أخرى في الاختبار نفسه، لذلك فقد حرصت الباحثة عند بناء المفردات الاختبارية المكونة للاختبار التحصيلي الحالي أن تحقق هذا الافتراض، وذلك من خلال تحرى الدقة في بنائها بحيث لا تعتمد إجابة أي مفردة اختبارية على إجابة مفردة اختبارية أخرى، ومما يؤكد تحقق هذا الافتراض في البيانات الاختبارية في الدراسة الحالية، هو أن الاختبار التحصيلي المستخدم يقيس سمة أحادية البعد، حيث أن تحقق افتراض أحادية البعد في البيانات الاختبارية كاف لتحقيق افتراض الاستقلال الموضعي(المركزى) ودليل على تحققه. واستكمالاً للتحقق من من مدى افتراض الاستقلال الموضعي(المركزى) في البيانات الاختبارية الحالية، قامت الباحثة بحساب معاملات الارتباط بين درجات المفردات الاختبارية، ووجد أن جميع قيم معاملات ارتباطات المفردات لم تصل إلى الواحد الصحيح الذي يدل على التطابق التام في إجابة مفردتين، مما يشير إلى عدم اعتماد الإجابة على أحد المفردات الاختبارية على الإجابة عن مفردة أخرى.

٣- افتراض المنحنيات المميزة للمفردة: وللتحقق من هذا الافتراض استخدمت الباحثة برنامج "**BILOG–MG V.3.0**" في تحليل بيانات الدراسة حيث يضمن تحقيق هذا الافتراض المتعلق بوجود منحنى مميز لكل مفردة اختبارية، حيث إن مخرجات التحليل باستخدام هذا البرنامج تشتمل على ثلاثة جوانب يمثل أحدها جانباً هاماً من جوانب التحليل وهو التحليل البياني لمفردات الاختبار التحصيلي مرجعي المحك، والثاني يوضح دالة معلومات المفردة والاختبار، والخطأ المعياري المتعلق بالقياس باستخدامها، وبذلك فإن البيانات الاختبارية تم فحصها والتأكد من تحقق الافتراض المتعلق بوجود منحنى مميز لكل مفردة من المفردات الاختبارية.

٤- افتراض التحرر من عامل السرعة: أحد افتراضات نظرية الاستجابة للمفردة هو ألا يعتمد أداء الأفراد أو الطلاب في الاختبارات على السرعة، أي أن عامل السرعة ليس عاملاً حاسماً في الاستجابة على المفردات الاختبارية، فمن الضروري أن يكون أداء الطالب هو أقصى أداء يمكن أن يقوم به للإجابة على مفردات الاختبار، ومن ثم ففشل الطالب في الإجابة الصحيحة عن مفردة اختبارية تكون بسبب عدم توافر القدرة أو السمة المقاسة لديه، وليس بسبب عدم تمكنه من الوصول إليها بسبب عامل السرعة. وبالتالي فإن هذا الافتراض يعد من الافتراضات الضمنية، وتحقيق شرط أحادية البعد، يعنى عدم وجود تأثير لعامل السرعة في الأداء على الاختبار، حيث أن السرعة إن كانت عاملاً مؤثراً في الأداء على الاختبار فهذا يعنى أن التحليل العاملي للمفردات المكونة للاختبار يجب أن يصل إلى عامل آخر بخلاف العامل الرئيس المراد قياسه من الاختبار. إلا أن النتائج التى تم التوصل إليها عند اختبار افتراض أحادية البعد، تؤكد أن هناك عامل واحد هو المسيطر على أداء الأفراد في الاختبار وهو السمة أو القدرة المقاسة "التحصيل الدراسي في وحدة المصفوفات"، وأن السرعة ليست عامل مؤثراً أو حاسماً في الاستجابة على هذه المفردات الاختبارية الحالية. واستكمالاً للتأكد من أن الاختبار المستخدم اختبار قوة وليس اختبار سرعة، تم فحص نسبة الطلاب الذين أجابوا على جميع المفردات المكونة للاختبار بدون ترك مفردات، وقد بلغت النسبة المئوية للطلاب المجيبين على جميع مفردات الاختبار(٩٦،٥٦%) وتعبر هذه النسبة المرتفعة عن عدم اعتماد الإجابة على هذه المفردات الاختبارية على عامل السرعة، وقد استبعدت الباحثة كل أوراق الإجابة للطلاب الذين لم يكملوا الإجابة على مفردات الاختبار الحالي لضمان تحقق هذا الافتراض بشكل تام. وقد روعي عند تطبيق الاختبار الحالي أن يكون زمن التطبيق مفتوح، حتى يكون أداء الطالب على هذا الاختبار، أقصى أداء يمكن بذله لمحاولة الإجابة عن جميع المفردات المكونة لهذا الاختبار. وبناء على ذلك، فإن بيانات الاختبار التحصيلي مرجعي المحك في الدراسة الحالية تحقق افتراضات نظرية الاستجابة للمفردة مما يؤكد ملائمة النموذج اللوغاريتمي ثلاثي المعلم لتحليل البيانات الاختبارية الحالية، وبالتالي فإن نتائج التحليل في ضوء هذه النموذج تتسم بدرجة كبيرة من الدقة والموضوعية.

ثانياً: النتائج المتعلقة بملائمة(الأفراد والمفردات) للنموذج اللوغاريتمي ثلاثي المعلم:

قبل البدء في إجراء التحليلات هذه تم فحص البيانات الاختبارية، من أجل استبعاد البيانات التامة والصفرية، وقد أسفرت هذه الخطوة عن عدم وجود بيانات تامة أو صفرية سواء بالنسبة للأفراد أو المفردات.

- **استبعاد الأفراد غير الملائمين للنموذج اللوغاريتمي ثلاثي المعلم: قامت الباحثة بعد فحص البيانات**

الاختبارية للتأكد من عدم وجود بيانات تامة أو صفرية بتحليل بيانات الأطوال الثلاثة للاختبار التحصيلي على عينة الدراسة النهائية(١٩٣٢ طالبا وطالبة) مرة ثانية بهدف استبعاد الأفراد غير الملائمين للنموذج اللوغاريتمي ثلاثي المعلم المستخدم في الدراسة، وذلك باستخدام برنامج "BILOG–MG V.3.0" تبعاً إلى: مؤشر الملائمة التقاربي In Fit، ومؤشر الملائمة التباعدي Out Fit. وقد قامت الباحثة بحذف الأفراد غير الملائمين وعددهم(٣٢ طالباً وطالبة) وهم الأفراد الذين تجاوزوا الحدود المقبولة لهذين المؤشرين وهى(± ٢،٥) لكل منهما.

- **استبعاد المفردات غير الملائمة للنموذج اللوغاريتمي ثلاثي المعلم: وباستخدام البرنامج الإحصائي"BILOG –MG V.3.0" تم إعاد تحليل نتائج الأطوال الثلاثة للاختبار التحصيلي مرجعي المحك في ضوء النموذج اللوغاريتمي ثلاثي المعلم، بناء على استجابات أفراد عينة الدراسة الكلية. وذلك بعد حذف الأفراد غير الملائمين بهدف تحديد المفردات غير الملائمة للنموذج، وذلك وفقاً لمعامل الارتباط الثنائي الأصيل. وتبين أن جميع المفردات الاختبارية المكونة للأطوال الثلاثة(٢٢، ٤٤، ٦٦ مفردة) للاختبار التحصيلي مرجعي المحك مطابقة للنموذج اللوغاريتمي ثلاثي المعلم لدى عينة الدراسة الكلية، وذلك طبقاً لمعامل الارتباط الثنائي الأصيل.**

ثالثاً: النتائج المتعلقة باختبار فروض الدراسة:

تعرض الباحثة فيما يلي النتائج التى أسفرت عنها الدراسة والمتعلقة بكل فرض من فروض الدراسة الحالية:

النتائج المتعلقة باختبار الفرض الأول:

ينص الفرض الأول عن أنه "تختلف تقديرات قدرات الأفراد على الاختبار التحصيلي مرجعي المحك باختلاف عدد مفردات الاختبار باستخدام النموذج اللوغاريتمي ثلاثي المعلم اختلافاً دالاً إحصائياً".

قامت الباحثة في هذا الجزء بتحديد قدرة كل فرد من أفراد العينة المكونة من(١٩٠٠) طالباً وطالبة من خلال مخرجات التحليل لنتائج كل اختبار من الاختبارات التحصيلية الثلاثة المكونة من(٢٢، ٤٤، ٦٦) مفردة على الترتيب في ضوء النموذج اللوغاريتمي الثلاثي المعلم باستخدام برنامج"BILOG–MG V.3.0"، ويبين الجدول التالي ملخصاً لنتائج التحليل للقيم المتحررة لقدرات أفراد عينة الدراسة الكلية على الأطوال الثلاثة للاختبار التحصيلي مرجعي المحك:

جدول(٢)

نتائج تحليل القيم المتحررة لقدرة أفراد عينة الدراسة الكلية على الأطوال الثلاثة للاختبار التحصيلي مرجعي المحك في ضوء النموذج اللوغاريتمي الثلاثي المعلم

الوصف الاحصائي لتقديرات القدرة على	العدد	متوسط تقدير القدرة باللوجيت	الانحراف المعياري	الجذر التربيعي لمتوسط مربعات الأخطاء المعيارية	تباين الأخطاء المعيارية في التقدير	معامل الثبات التجريبي
الاختبار الأول(ن=٢٢)	١٩٠٠	٠،٠٤٨٣	١،٠٦٩٧	٠،٥٥٠٩	٠،٣٠٣٥	٠،٧٣٤٨
الاختبار الثاني(ن=٤٤)	١٩٠٠	٠،٠٨٦٧	١،٠٤٦٨	٠،٣٦٩٩	٠،١٣٦٨	٠،٨٧٥٢
الاختبار الثالث(ن=٦٦)	١٩٠٠	٠،٠٨٦٣	٠،٩٨٩٢	٠،٣٠١١	٠،٠٩٠٧	٠،٩٠٧٣

ويتضح من الجدول السابق أن قيم متوسط تقدير قدرة أفراد عينة الدراسة الكلية(ن=١٩٠٠طالباً وطالبة) على الأطوال الثلاثة(٢٢، ٤٤، ٦٦ مفردة) للاختبار التحصيلي مرجعي المحك في وحدة المصفوفات باستخدام النموذج اللوغاريتمي ثلاثي المعلم بلغت(٠،٠٤٨٣، ٠،٠٨٦٧، ٠،٠٨٦٣) لوجيت على الترتيب، بتباين قدره(١،١٤٤٣، ١،٠٩٥٨، ٠،٩٧٨٤) على الترتيب، بينما بلغت قيم الجذر التربيعي لمتوسط مربعات الأخطاء المعيارية في التقدير لهذه الأطوال الثلاثة من الاختبار التحصيلي(٠،٥٥٠٩، ٠،٣٦٩٩، ٠،٣٠١١)، بتباين للأخطاء المعيارية في تقدير الأرجحية القصوى للقدرة بلغ(٠،٣٠٣٥، ٠،١٣٦٨، ٠،٠٩٠٧) على الترتيب، وعن القيم المقدرة لمعاملات الثبات التجريبي للأطوال الثلاثة للاختبار التحصيلي فقد بلغت(٠،٧٣٤٨، ٠،٨٧٥٢، ٠،٩٠٧٣) على الترتيب باستخدام النموذج اللوغاريتمي ثلاثي المعلم.

ويتضح من الجدول السابق أيضاً أن أعلى قيمة لمتوسط تقدير قدرة أفراد عينة الدراسة الكلية كان لصالح الاختبار التحصيلي مرجعي المحك الثاني المكون من(ن= ٤٤) مفردة، بينما أقل قيمة لمتوسط تقدير قدرة الأفراد كان في حالة الاختبار التحصيلي مرجعي المحك الأول المكون من(ن=

٢٢) مفردة. وكما يلاحظ أنه بزيادة عدد مفردات الاختبار التحصيلي تنخفض قيمة الجذر التربيعي لمتوسط مربعات الأخطاء المعيارية للتقدير، وتزداد قيمة معامل الثبات التجريبي للاختبار.

ومن أجل بحث دلالة الفروق بين متوسطات تقديرات قدرة الأفراد على الأطوال الثلاثة للاختبار التحصيلي مرجعي المحك في وحدة المصفوفات باستخدام النموذج اللوغاريتمي ثلاثي المعلم، قامت الباحثة بدراسة البيانات الواردة بالجدول السابق والمتعلقة بتقديرات قدرة الأفراد على الأطوال الثلاثة للاختبار التحصيلي الحالي، فتبين أن توزيع هذه البيانات يقترب من الاعتدالية، بالإضافة إلى أن كبر حجم عينة الأفراد الممتحنين يضمن إلى حد كبير أن هذه البيانات تتوزع توزيعاً يقترب من الاعتدالية. ومن ثم قامت الباحثة بإجراء تحليل التباين لهذه البيانات للتعرف على دلالة الفروق بينهما، والجدول التالي يوضح نتائج تحليل التباين:

جدول(٣)

نتائج تحليل التباين لتقديرات قدرات أفراد عينة الدراسة الكلية(ن=١٩٠٠طالباً وطالبة) على الأطوال الثلاثة للاختبار التحصيلي مرجعي المحك باستخدام النموذج اللوغاريتمي الثلاثي المعلم

المتغير	مصدر التباين	مجموع المربعات	د.ح	متوسط المربعات	قيمة(ف)	مستوى الدلالة
تقديرات القدرة	بين المجموعات	١،٨٤٨٥٢٢٦	٢	٠،٩٢٤٢٦١٣	٠،٨٦١	٠،٥١٧
	داخل المجموعات	٦١١٥،١٥	٥٦٩٧	١،٠٧٣٣٩٨٣		
	المجموع الكلي	٦١١٦،٩٩٨٥٢٢٦	٥٦٩٩			

ويتبين من الجدول السابق أن الفروق بين الأطوال الثلاثة(٢٢، ٤٤، ٦٦ مفردة) للاختبار التحصيلي مرجعي المحك في تقدير قدرات الأفراد باستخدام النموذج اللوغاريتمي الثلاثي المعلم غير دالة إحصائياً، مما يدل على عدم وجود فروق جوهرية بين الاختبارات التحصيلية الثلاثة في تقدير قدرات الأفراد في وحدة المصفوفات، وهذا يؤكد أن الأطوال الثلاثة للاختبار التحصيلي المستخدمة حالياً متسقة مع بعضها إلى درجة كبيرة في تقديراتها لقدرات الأفراد باستخدام النموذج اللوغاريتمي ثلاثي المعلم. وبناءً على ذلك، لا تختلف تقديرات قدرات الأفراد على الاختبار التحصيلي مرجعي المحك باختلاف عدد مفردات الاختبار باستخدام النموذج اللوغاريتمي الثلاثي المعلم.

وتتفق نتيجة الدراسة الحالية مع نتيجة دراسة كل من "سميث" (1988)Smith، ومنى ربيع متيرد(٢٠٠٠)، ويمكن تفسير ذلك بأن الأطوال الثلاثة للاختبار التحصيلي مرجعي المحك تعد متكافئة

في تمثيل النطاق السلوكي(وحدة المصفوفات) المختار في الدراسة الحالية، الأمر الذى أدى إلى وصف متقارب ومتكافئ لأداء كل فرد في كل اختبار، وبالتالي عدم وجود فروق جوهرية بين الأطوال الثلاثة للاختبار التحصيلي مرجعي المحك في تقدير قدرات الأفراد باستخدام النموذج اللوغاريتمى(الأحادي – الثنائي – الثلاثي) المعلم. بينما اختلفت هذه النتيجة مع نتائج دراسة كل من: "هولين وآخرين" (1982)...Hulin & et al، و"جونيوس" (1984)Junius، و"كوتش ودود" (1995)Koch & Dodd، و"ستون وهانسين" (2000)Stone & Hansen، والشيماء عبد الحميد طنطاوي(٢٠٠٧)، وقد يرجع ذلك إلى اختلاف عينة الدراسة من حيث المرحلة العمرية وحجمها، ومجال الدراسة، وطبيعة البيانات الاختبارية(اصطناعية – حقيقية)، وأداة الدراسة، بالإضافة إلى اختلاف النماذج المستخدمة في تلك الدراسات، ونوع البرنامج الإحصائي المستخدم في تحليل البيانات الاختبارية في ضوء النموذج المستخدم، ومدى التحقق من توافر افتراضات النموذج في البيانات، أو مدى تحقق المطابقة الكافية للبيانات المتعلقة بتلك الدراسات للنموذج المستخدم.

وهذه النتيجة تؤكد أهم مميزات نماذج نظرية الاستجابة للمفردة، والخاصة باستقلال تقديرات قدرات الأفراد عن خصائص عينة المفردات المستخدمة ومنها عدد المفردات أو طول الاختبار، وهذا دليل هام على تحقق افتراض هام من الافتراضات الأساسية لنظرية الاستجابة للمفردة هو استقلالية القياس، مما يتيح إمكانية المقارنة بين الأفراد رغم اختلاف الاختبار المستخدم في عملية القياس، وهذا الأمر غير متوفر في إطار نظرية الاختبارات الكلاسيكية(التقليدية).

وفي ضوء ما سبق من نتائج تم التوصل إلى عدم تحقق صحة الفرض الأول والذي ينص على "تختلف تقديرات قدرات الأفراد على الاختبار التحصيلي مرجعي المحك باختلاف عدد مفردات الاختبار باستخدام النموذج اللوغاريتمي ثلاثي المعلم اختلافاً دالاً إحصائياً".

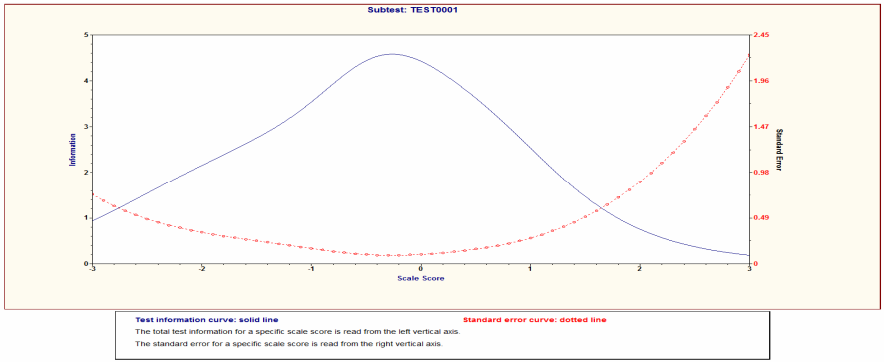
النتائج المتعلقة باختبار الفرض الثاني:

ينص الفرض الثاني عن أنه "تختلف دالة معلومات الاختبار التحصيلي مرجعى المحك باختلاف عدد مفردات الاختبار باستخدام النموذج اللوغاريتمي ثلاثي المعلم".

ولاختبار صحة هذا الفرض فقد استخدمت الباحثة برنامج "BILOG–MG V.3.0"، ومن خلال مجموعة الأوامر الخاصة بدوال معلومات الاختبار، حصلت الباحثة على الأشكال الممثلة لدوال معلومات كل اختبار من الاختبارات الثلاثة الممثلة للأطوال(٢٢، ٤٤، ٦٦) مفردة كل على حدة، وفقاً للنموذج اللوغاريتمي ثلاثي المعلم، وذلك باستخدام الاستجابات الخاصة بالعينة الكلية، ثم تم تعيين أقصى ارتفاع لدالة المعلومات تمهيداً للتعرف على مدى القدرة الذى يقدم عنده كل طول من الأطوال الثلاثة للاختبار أفضل قياس بأعلى دقة وأعلى ثبات، بالإضافة إلى الحصول على الخطأ المعياري

للقياس مباشرة من خلال الشكل البياني لدالة معلومات الاختبار، حيث أن الخطأ المعياري يمثل بيانياً بمقلوب المنحنى الممثل لدالة معلومات الاختبار.

ويعبر الشكل التالي عن دالة معلومات الاختبار التحصيلي مرجعي المحك المكون من(ن= ٢٢) مفردة، والخطأ المعياري للقياس، باستخدام النموذج اللوغاريتمي الثلاثي المعلم:

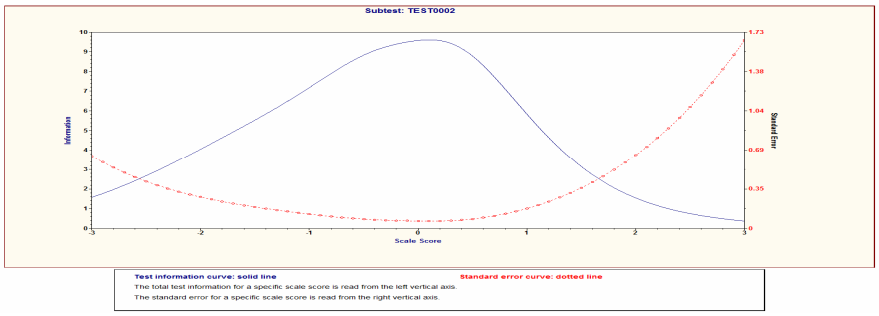


شكل(١)

منحنى دالة المعلومات والخطأ المعياري للاختبار التحصيلي مرجعي المحك المكون من(ن=٢٢) مفردة لدى عينة الدراسة الكلية في ضوء النموذج اللوغاريتمي الثلاثي المعلم

يتضح من الشكل السابق أنه وفقاً للنموذج اللوغاريتمي ثلاثي المعلم بلغت أقصى قيمة للمعلومات التى يمكن أن يقدمها الاختبار التحصيلي مرجعي المحك المكون من(ن=٢٢) مفردة(٤،٦) عند مستوى القدرة المتوسطة(- ٠،٢) بوحدة لوجيت، وبأقل قيمة للخطأ المعياري بلغت(٠،١٢)، مما يشير إلى أن الاختبار يمكن أن يوفر أكبر قدر من المعلومات عند مستوى القدرة المتوسطة، أي أن المفردات المكونة للاختبار تتلاءم أكثر مع الأفراد متوسطي القدرة بدقة أكبر. وأفضل ارتفاع لدالة معلومات الاختبار امتد على مدى القدرة من القيمة(- ١،٢) إلى القيمة (+٠،٧)، ويلاحظ أنه كلما اقتربنا من طرفى متصل القدرة تزداد قيم الخطأ المعياري للقياس، وبالطبع تقل قيم دالة معلومات الاختبار، لتصل قيمة الخطأ المعياري عند طرف القدرة المنخفضة ليبلغ(٠،٧٠) تقريباً، بينما وصلت قيمة هذا الخطأ عند طرف القدرة المرتفعة إلى(٢،٠) تقريباً.

ويمثل الشكل التالي دالة معلومات الاختبار التحصيلي مرجعي المحك المكون من(ن= ٤٤) مفردة، والخطأ المعياري للقياس، باستخدام النموذج اللوغاريتمي الثلاثي المعلم:

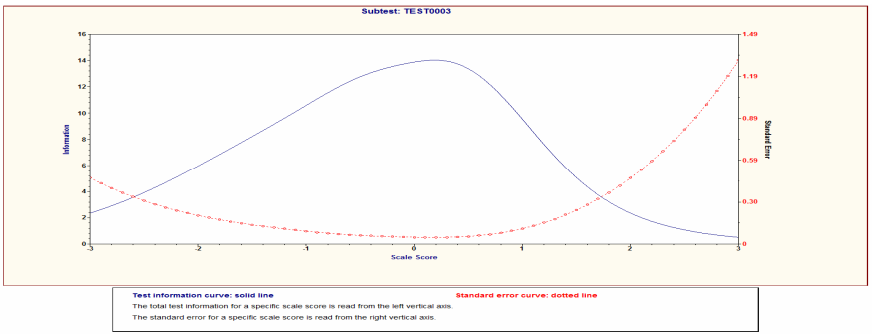


شكل(٢)

منحنى دالة المعلومات والخطأ المعياري للاختبار التحصيلي مرجعي المحك المكون من(ن=٤٤) مفردة لدى عينة الدراسة الكلية في ضوء النموذج اللوغاريتمي الثلاثي المعلم

ويتضح من الشكل السابق أن دالة معلومات الاختبار التحصيلي مرجعي المحك المكون من(ن=٤٤) مفردة، قد أظهرت قيمة مرتفعة من المعلومات التى يقدمها الاختبار بلغت(٩،٦) عند مستوى القدرة المتوسطة(صفر) وحدة لوجيت، وبأدنى قيمة للخطأ المعياري وصلت إلى(٠،٠٨٧) وفقاً للنموذج اللوغاريتمي ثلاثي المعلم، مما يشير إلى أن الاختبار يوفر أكبر قدر من المعلومات عند مستوى القدرة المتوسطة، وبأكبر درجة من الدقة، وبأعلى درجة من الثبات، وهذا يدل على أن المفردات المكونة للاختبار تتلاءم أكثر مع الأفراد متوسطي القدرة بدقة أكبر. وقد غطى الاختبار أفضل ارتفاع لدالة المعلومات على مدى القدرة الذى يمتد من القيمة(−١) وتشير إلى المستوى المنخفض من القدرة إلى القيمة(+ ٠،٨) وتشير إلى المستوى المرتفع من القدرة ، بينما يلاحظ أنه كلما اقتربنا من طرفى متصل القدرة تقل قيم معلومات الاختبار، وفي المقابل تزداد قيم الخطأ المعياري لتصل قيمة الخطأ المعياري لهذا الاختبار عند طرف القدرة المنخفضة حوالي(٠،٦٠)، بينما وصلت هذه القيمة إلى(١،٧٣) تقريباً عند طرف القدرة المرتفعة.

بينما يمثل الشكل التالي دالة معلومات الاختبار التحصيلي مرجعي المحك المكون من(ن= ٦٦) مفردة، والخطأ المعياري للقياس باستخدام النموذج اللوغاريتمي الثلاثي المعلم:



شكل(٣)

منحنى دالة المعلومات والخطأ المعياري للاختبار التحصيلي مرجعي المحك المكون

من(ن=٦٦) مفردة لدى عينة الدراسة الكلية في ضوء النموذج اللوغاريتمي الثلاثي المعلم

ويظهر من خلال الشكل السابق أن أقصى معلومات يمكن أن يقدمها الاختبار عن قدرات الأفراد الممتحنين بلغت(١٤) عند مستوى(+٠،١٨) من القدرة بوحدة اللوجيت، بأعلى دقة وأدنى خطأ للقياس والتي بلغت(٠،٠٧). وهذا يعنى أن الاختبار يمكن أن يوفر أكبر قدر من المعلومات عند متوسط القدرة، وبالتالي فإن المفردات المكونة للصورة الكاملة للاختبار التحصيلي مرجعي المحك(ن= ٦٦ مفردة) تتلاءم أكثر مع الفرد متوسط القدرة حيث يعطى دقة أكبر للقياس بأقل خطأ معياري، باستخدام النموذج اللوغاريتمي ثلاثي المعلم. وبالنظر إلى شكل دالة معلومات الصورة الكاملة للاختبار التحصيلي مرجعي المحك وفقاً للنموذج اللوغاريتمي ثلاثي المعلم نجد أن أفضل ارتفاع لدالة المعلومات كان على مدى القدرة الذى امتد من القيمة(−١،٢) إلى القيمة(+ ١،٢)، أى من المستوى المنخفض إلى المستوى المرتفع من القدرة، ثم انخفض ارتفاع هذه الدالة حيث بلغ أقل ارتفاع عند طرفى متصل القدرة المقاسة. وفي المقابل، ونظراً لأن الخطأ المعياري يمثل بيانياً بمقلوب المنحنى الممثل لدالة معلومات الاختبار، نجد أن أدنى قيمة للخطأ المعياري للقياس امتد من القيمة(− ٠،٨) إلى(+ ٠،٦) على متصل القدرة لتزداد هذه القيمة ارتفاعاً لتصل إلى أقصى ما يمكن عند طرفى متصل القدرة، حيث بلغت قيمة الخطأ المعياري عند طرف القدرة المنخفضة(٠،٣٩) تقريباً، بينما بلغت هذه القيمة(١،٢٩) تقريباً عند طرف القدرة المرتفعة.

ويمكن أن نستنتج من الأشكال السابقة أن المعلومات المقدمة من الصورة الكاملة للاختبار التحصيلي مرجعي المحك المكون من(ن= ٦٦) مفردة اختبارية في ضوء النموذج اللوغاريتمي ثلاثي المعلم أكبر منها في حالة الاختبارين الآخرين، كما أن المعلومات المقدمة من الاختبار التحصلي مرجعي المحك المكون من(ن=٤٤) مفردة اختبارية أكبر منها في حالة الصورة الأقصر للاختبار التحصيلي مرجعي المحك المكونة من(ن=٢٢) مفردة اختبارية، وذلك وفقاً للنموذج اللوغاريتمي ثلاثي المعلم. كما اتضح أن قيمة الخطأ المعياري للقياس كان أكبر ما يمكن في حالة الصورة القصيرة من الاختبار التحصيلي مرجعي المحك المكون من(ن=٢٢) مفردة، في حين كانت أقل قيمة منها في حالة الاختبار التحصيلي مرجعي المحك المكون من(ن= ٤٤) مفردة، بينما كانت أقل ما يمكن في حالة الصورة الكاملة من الاختبار التحصيلي مرجعي المحك المكونة من(ن= ٦٦) مفردة، وفقاً للنموذج اللوغاريتمي الثلاثي المعلم. ويلاحظ أيضاً أن الأطوال الثلاثة(٢٢، ٤٤، ٦٦ مفردة) من الاختبار التحصيلي مرجعي المحك قدمت أقصى قيمة من المعلومات عند مستويات القدرة المتوسطة بدرجات متفاوتة.

ويلخص الجدول التالي قيم أقصى المعلومات للأطوال الثلاثة للاختبار التحصيلي، وتقدير القدرة المقابلة لها، وكذلك الخطأ المعياري للقياس بالنسبة لكل اختبار في ضوء النموذج اللوغاريتمي ثلاثي المعلم:-

جدول(٤)

قيمة أقصى المعلومات والقدرة المقابلة لها وكذلك الخطأ المعياري للأطوال الثلاثة من الاختبار التحصيلي مرجعي المحك لدى عينة الدراسة الكلية في ضوء النموذج اللوغاريتمي الثلاثي المعلم

الاختبارات / المتغيرات	الاختبار الأول (ن= ٢٢ مفردة)	الاختبار الثاني (ن= ٤٤ مفردة)	الاختبار الثالث (ن= ٦٦ مفردة)
أقصى قيمة معلومات للاختبار	٤،٦	٩،٦	١٤
قيمة القدرة المقابلة	-٠،٢	صفر	+٠،١٨
الخطأ المعياري للقياس	٠،١٢	٠،٠٨٧	٠،٠٧

ويتبين من الجدول السابق أن الاختبار التحصيلي مرجعي المحك المكون من(ن= ٦٦) مفردة – والذي يمثل الصورة الكاملة للاختبار– في ضوء النموذج اللوغاريتمي ثلاثي المعلم قدم أفضل القيم من حيث أقصى قيمة لمعلومات الاختبار، وأدنى قيمة للخطأ المعياري للقياس، ومن الملاحظ أن الأطوال الثلاثة للاختبار التحصيلي مرجعي المحك قدمت أقصى قيمة للمعلومات عند مستوى القدرة المتوسطة.

وهذا يعنى أنه كلما زاد عدد المفردات المكونة للاختبار التحصيلي مرجعي المحك كلما ارتفعت قيمة المعلومات المقدمة من الاختبار عن قدرات الأفراد الممتحنين، وبالتالي انخفضت قيمة الخطأ المعياري للقياس وفقاً لنماذج الاستجابة للمفردة أحادية البعد المستخدمة في الدراسة الحالية.

كما اتضح بمطالعة جميع دوال معلومات الأطوال الثلاثة للاختبار التحصيلي مرجعي المحك أن مقدار المعلومات المقدمة من كل اختبار ليست متساوية على طول الاختبار وليست بنفس القيمة خلال المستويات المختلفة للقدرة، وذلك يرجع إلى أن المفردات الاختبارية غير متساوية فيما بينها في مقدار المعلومات المقدمة من كل منها، فمقدار المعلومات يختلف من مفردة إلى أخرى، كما أن قيم الخطأ المعياري ليست متساوية على طول الاختبار، وإنما اختلفت على طول متصل القدرة حسب قيمة الخطأ المعياري للقياس الخاصة بكل مفردة من مفردات الاختبار.

وتتفق هذه النتيجة مع نتائج دراسة "هامبلتون وجونيس" (1994)Hambleton & Jones، ودراسة "مارك" (2007)Mark، ودراسة نضال كمال الشريفين(٢٠٠٩)، ويمكن تفسير ذلك في ضوء أن الاختبار الأطول أكثر تمثيلاً للخاصية أو السلوك المراد قياسه، مما يتيح الفرصة للأفراد

الممتحنين بالتعبير عن قدراتهم،الأمر الذى يزيد من قدرة الاختبار لتقديم معلومات أكثر عن الأفراد المجيبين عليه.

ونظراً لأن المستوى العام لدالة معلومات الاختبار تعتمد على عدد مفردات الاختبار، فبصفة عامة كلما كانت هناك مفردات أكثر بخصائص سيكومترية جيدة وليست ضعيفة كلما كان مقدار المعلومات أكبر، ومن ثم ستقوم الاختبارات الأطول بقياس قدرة الممتحن بدقة أكبر من الاختبارات الموجزة(Baker, 2001: 110).

وقد ترجع هذه النتيجة إلى أن المفردات ذات الخصائص السيكومترية الجيدة سوف تمثل النطاق السلوكي بشكل أفضل، ومن ثم فكلما زاد عدد مفردات الاختبار فإنه سيتم تمثيل عناصر النطاق السلوكي بدرجة أفضل وبشكل أوسع، حيث أن زيادة عدد المفردات تزيد من عدد الأفكار المعبرة عن كل هدف من الأهداف السلوكية لقياسه، ومن ثم يغطى نقاط متنوعة من متصل السمة المقاسة، وبالتالي فالاختبار الأطول يقدم كمية من المعلومات أكبر عن تقديرات قدرات الأفراد الممتحنين.

وحيث أن دوال معلومات الاختبار هي عبارة عن مجموع دوال معلومات المفردات، والتي تقدم أقصى معلومات لكل مفردة عند مستوى محدد من القدرة، فإن دوال معلومات الاختبار الأطول سوف تكون بالفعل أعلى، ومن ثم أدق. وبناء على ذلك فالاختبار الطويل يعطى فرصة لتقديم معلومات عن قدرة الأفراد أكبر مما يقدمه الاختبار القصير مما جعل الخطأ المعياري للقياس أقل مما يمكن باستخدام الاختبار الأطول، ومن ثم فإن الاختبار الأطول يؤدى إلى أعلى دقة في تصنيف الطلاب حسب إتقانهم لأهداف الوحدة الدراسية ومدى إتقانهم للمهارات الرئيسة لهذه الوحدة، ومن ثم اتخاذ قرارات صائبة فيما يتعلق بتصنيف الطلاب، كما أن نتائجه تكون أكثر اتساقاً واستقراراً.

وبذلك يتحقق صحة الفرض الثاني الذى ينص على أنه "تختلف دالة معلومات الاختبار التحصيلي مرجعي المحك باختلاف عدد مفردات الاختبار باستخدام النموذج اللوغاريتمي ثلاثي المعلم".

توصيات الدراسة:

بناء على ما أسفر عنه التحليل النظري للمفاهيم والمتغيرات الأساسية للدراسة الحالية، وما أسفرت عنه من نتائج يمكن أن توجه أنظار المربين والمشاركين في عملية التقويم التربوي والباحثين في مجال القياس والتقويم، بالإضافة إلى المهتمين بتطوير أساليب التقويم والامتحانات إلى العديد من التوصيات التى يمكن تلخيصها فيما يلي:

١- يمكن الاستفادة من مجموعة المفردات الاختبارية المستخدمة في أداة الدراسة الحالية كنواة لبناء بنك أسئلة في مقرر الجبر للصف الأول الثانوي العام.

٢- عقد دورات تدريبية للباحثين في مجال القياس النفسي والتربوي والمشاركين في عملية التقويم التربوي، يتم خلالها تدريبهم كيفية استخدام نماذج نظرية الاستجابة للمفردة المعلم في بناء الاختبارات التحصيلية في المواد الدراسية المختلفة، لرفع مستوى خصائصها السيكومترية من أجل تحقيق موضوعية القياس المنشودة، وعلى كيفية استخدام برامج الحاسب الآلى اللازمة لتحليل البيانات ونتائج الاختبارات باستخدام نماذج نظرية الاستجابة للمفردة أحادية البعد، مثل برنامج"BILOG-MG V.3.0"، وبرنامج"QUEST"، وبرنامج "MICROSCALE"، وبرنامج"MULTILOG"، وبرنامج"WINSTEPS"، بالإضافة إلى توفير البرامج الإحصائية الهامة واللازمة لتطبيق هذه النماذج مثل برنامج"Excel"، وبرنامج"SPSS".....

٣- أسفرت نتائج الدراسة الحالية عن عدم اختلاف تقديرات قدرات الأفراد باختلاف عدد مفردات الاختبار باستخدام النموذج اللوغاريتمي ثلاثي المعلم، الأمر الذى يؤكد أهمية نماذج نظرية الاستجابة للمفردة، وما توفره من مميزات عديدة جعلها تتغلب على العديد من مشكلات القياس النفسي والتربوي، مثل عدم استقلالية القياس عن خصائص الأفراد والمفردات. وهذه النتيجة تؤكد استقلالية القياس عن عدد المفردات التى يؤديها الفرد في إطار عمل هذه النماذج، مما يساعد في تحقيق موضوعية القياس المنشودة في مجال القياس النفسي والتربوي، بالإضافة إلى إمكانية المقارنة بين الأفراد رغم اختلاف الاختبارات التى يختبرون بها وتباين طولها، لذا توصى الباحثة باستخدام نماذج نظرية الاستجابة للمفردة لتحليل وتطوير الاختبارات التحصيلية في مختلف المواد الدراسية.

٤- بناءً على نتائج الدراسة الحالية التى أوضحت أن الأطوال الثلاثة للاختبار التحصيلي مرجعي المحك والمستخدمة في الدراسة الحالية متكافئة في تقدير قدرات الأفراد في ضوء النموذج اللوغاريتمي ثلاثي المعلم، فإنه يمكن استخدام أحد هذه الأطوال الثلاثة في تشخيص مواطن القوة والضعف في تحصيل الطلاب في وحدة "المصفوفات" بدقة متكافئة، والتحقق من إتقانهم للمهارات الرئيسة المنشودة من دراستهم لهذه الوحدة الدراسية، وهذا من شأنه تحسين عملية تقويم طلاب الصف الأول الثانوي العام في هذه الوحدة الدراسية المختارة.

٥- بناءً على ما توصلت إليه الدراسة الحالية من اختلاف دوال معلومات الاختبار التحصيلي مرجعي المحك باختلاف عدد مفردات الاختبار باستخدام النموذج اللوغاريتمي ثلاثي المعلم، وأن الصورة الكاملة(الأطول) للاختبار التحصيلي تقدم معلومات أكبر عن قدرات الأفراد الممتحنين، حيث أن دالة معلومات الاختبار هي عبارة عن محصلة مجموع دوال معلومات المفردات المكونة للاختبار

عبر جميع مستويات القدرة. وفى ضوء ذلك توصى الباحثة بضرورة انتقاء المفردات التى تتمتع بخصائص سيكومترية جيدة، وليس مجرد زيادة لعدد المفردات الاختبارية فقط للحصول على قيمة معلومات أكبر، فالمفردات الأكثر تمييزاً بين قدرات الأفراد أفضل في تقديم معلومات أكبر عن قدرات هؤلاء الأفراد.

الدراسات والبحوث المقترحة:

في ضوء الإطار النظري للدراسة الحالية، وما أسفرت عنه من نتائج، يمكن أن تقترح الباحثة عدداً من النقاط البحثية التالية كدراسات مستقبلية، تتمثل فيما يلي:

١- دراسة أثر عدد مفردات الاختبار على اتساق قرارات تصنيف الطلاب باستخدام اختبار تحصيلي مرجعي المحك في ضوء نماذج نظرية الاستجابة للمفردة أحادية البعد.
٢- دراسة أثر طول الاختبار علي تقدير الخطأ المعياري في اختبارات القدرات العقلية أو اختبارات الذكاء في ضوء نماذج نظرية الاستجابة للمفردة.
٣- دراسة أثر مخالفة افتراضات نماذج نظرية الاستجابة للمفردة أحادية البعد على دقة نتائج الاختبارات مرجعية المعيار/ المحك.
٤- دراسة أثر اختلاف البرنامج الاحصائى المستخدم في تحليل البيانات الاختبارية على دقة نتائج الاختبارات التحصيلية في ضوء نماذج نظرية الاستجابة للمفردة.

المراجع

أولاً:المراجع العربية:

١- إسماعيل حسن فهيم اسماعيل الوليلى(٢٠٠٢). دراسة سيكومترية مقارنة لبعض نماذج الاستجابة للمفردة في انتقاء مفردات الاختبارات مرجعية المحك، رسالة دكتوراة غير منشورة، كلية التربية، جامعة الأزهر.

٢- الشيماء عبد الحميد محمد طنطاوي(٢٠٠٧). أثر كل من حجم العينة وطول الاختبار على دقة نتائج الاختبارات التحصيلية مرجعية المحك في الرياضيات باستخدام نموذج راش، رسالة ماجستير غير منشورة، كلية التربية، جامعة الزقازيق.

٣- إياد محمد حمادنه(٢٠٠٩). استخدام نظرية الاستجابة للمفردة في بناء اختبار محكي المرجع في الرياضيات وفق النموذج اللوجستي ثلاثي المعلم، مجلة العلوم التربوية والنفسية، كلية التربية، جامعة البحرين، المجلد(١٠)، العدد(٢)، يونيو، ص ص ٢١٥ – ٢٣٨.

٤- خيري سند، محمد سعيد(٢٠٠٩/٢٠١٠). الفائز في الرياضيات. القاهرة: مؤسسة الفائز للنشر والتوزيع.

٥- شحتة عبد المولى عبد الحافظ محمد(١٩٩٩). تقويم بناء الاختبارات المرجعة إلى المحك/ المعيار في ضوء نظرية الاستجابة للمفردة والنظرية التقليدية، رسالة دكتوراة غير منشورة، كلية البنات، جامعة عين شمس.

٦- صلاح الدين محمود علام(١٩٩٠). استخدام نموذج راش في بناء اختبار تشخيصي مرجعي المحك للمعارف الأساسية في اعداد خطة البحوث التربوية والنفسية، مجلة كلية التربية، جامعة الأزهر، العدد(١٧)، ص ص ١٢٧ – ١٨٥.

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

٨- صلاح الدين محمود علام(٢٠٠٧). الاختبارات التشخيصية مرجعية المحك في المجالات التربوية والنفسية والتدريبية. القاهرة: دار الفكر العربي.

٩- عبد الناصر سند العكايلة(٢٠٠٧). دراسة مقارنة بين النظرية الكلاسيكية والنظرية الحديثة فى القياس فى كشف وتقدير الخطأ المعيارى فى اختبارات القدرات المعرفية لطلبة المرحلة الأساسية فى الأردن، رسالة دكتوراة غير منشورة، معهد الدراسات والبحوث التربوية، جامعة القاهرة.

١٠- عزيز إسحق سرجيوس، حسين جاويش، محمد سمير القرشى(٢٠٠٩/٢٠١٠). المعاصر في الرياضيات للصف الأول الثانوى العام. الفصل الدراسي الأول. القاهرة: مكتبة طلبة.

١١- عماد عبد المسيح يوسف الصفطى(د.ت). استخدام نموذج راش اللوغاريتمى الاحتمالي في بناء اختبار تشخيصى هدفى المرجع لقياس اتقان المعلمين والباحثين للمهارات الأساسية اللازمة لبناء الاختبارات التحصيلية، بحث مقدم إلى اللجنة العلمية للترقية إلى درجة الأستاذية بقسم علم النفس التربوي، ص ص ١ – ٥٤.

١٢- محمد إبراهيم محمد محمد(٢٠١٠). استخدام نموذج التقدير الجزئى لتحليل مستويات بنية ناتج التعلم للاختبارات العنقودية ذات الفقرات المرتبة هرميا لدى عينة من طلاب كلية التربية بالمنيا، رسالة دكتوراة غير منشورة، كلية التربية، جامعة المنيا.

١٣- محمود محمد إبراهيم عبد الله(١٩٩٠). دراسة سيكومترية مقارنة – لطرق حساب معامل ثبات الاختبارات المرجعة إلى المحك، رسالة ماجستير غير منشورة، معهد الدراسات والبحوث التربوية، جامعة القاهرة.

١٤- مصطفى أحمد طيفور(٢٠٠٧). دراسة مقارنة بين النظرية الكلاسيكية والنظرية الحديثة في القياس في كشف وتقدير الخطأ المعيارى في اختبارات القدرات المعرفية لطلبة المرحلة الأساسية في الأردن، رسالة دكتوراة غير منشورة، معهد الدراسات والبحوث التربوية، جامعة القاهرة.

١٥- منار أحمد طومان(٢٠٠٦). فاعلية استخدام نظرية الاستجابة للمفردة (IRT) في عملية معادلة درجات الاختبارات متعددة الأبعاد والمتغيرات المؤثرة فيها، رسالة دكتوراة غير منشورة، معهد الدراسات والبحوث التربوية، جامعة القاهرة.

١٦- منى ربيع الطنطاوى متيرد(٢٠٠٠). دراسة سيكومترية حول تطوير اختبار المصفوفات المتتابعة لرافن باستخدام نموذج (راش)، رسالة ماجستير غير منشورة، كلية البنات للآداب والعلوم والتربية، جامعة عين شمس.

١٧- نادية محمد عبد السلام(١٩٩٦). ثبات وصدق القياس محكى المرجع، في: أنور محمد الشرقاوي، سليمان الخضري الشيخ، أمينة محمد كاظم، نادية محمد عبد السلام(١٩٩٦). محررون: اتجاهات معاصرة في القياس والتقويم النفسي والتربوي، القاهرة: مكتبة الأنجلو المصرية، ص ص ٦٣ – ٨٥.

١٨- نضال كمال الشريفين(٢٠٠٩). أثر صياغة الفقرة في مقياس اتجاه في الخصائص السيكومترية للفقرات وللمقياس وتقديرات القدرة للأفراد وفق نظرية الاستجابة للفقرة، مجلة العلوم التربوية والنفسية، كلية التربية، جامعة البحرين، المجلد(١٠)، العدد(٤)، ديسمير، ص ص ١٣ – ٤٠.

١٩- وزارة التربية والتعليم(٢٠٠٩/٢٠١٠). الجبر وحساب المثلثات للصف الأول الثانوى العـــام، كتاب الشرح والتدريبات، الفصل الدراسي الأول.

ثانياً:المراجع الأجنبية:

20- Anastasi, A. & Urbina, S. (1997). Psychological Testing. Prentice Hall, 7th Ed, Upper Saddle River, New Jersey.

21- Baker, F. (2001). The Basics of Item Response Theory. ERIC Clearinghouse on Assessment and Evaluation, United States of America.

22- Barnes, L.B. & Wise, S.L. (1991). The Utility of a Modified One-Parameter IRT Model with Small Samples. Applied Measurement in Education, Vol.4 (2), Pp. 143–157.

23- Bock, D.R. (1997). A Brief History of Item Response Theory. Educational Measurement: Issues And Practice, Vol.16 (4), Pp. 21–32.

24- Fan, X. (1998). Item Response Theory and Classical Test Theory: An Empirical Comparison of Their Item\Person Statistics. Educational and Psychological Measurement, Vol.58 (3), Pp. 357–381.

25- Haladyna, T.M. & Roid, G.H. (1983). A Comparison of Two Approaches to Criterion-Referenced Test Construction. Journal of Educational Measurement, Vol.20 (3), Pp. 271–282.

26- Hambleton, R.K. & Jones, R.W. (1994). Item Parameter Estimation Errors and Their Influence on Test Information Functions. Applied Measurement in Education, Vol.7 (3), Pp. 171-186.

27- Hambleton, R.K.; Mills, C.N. & Simon, R. (1983). Determining the Lengths for Criterion-Referenced Tests. Journal of Educational Measurement, Vol.20 (1), Pp. 27-38.

28- Hendrawn, I.; Glas, C.A.W. & Meijer, R.R. (2005). The Effect of Person Misfit on Classification Decisions. Applied Psychological Measurement, Vol.29 (1), Pp. 26-44.

29- Hulin, C.L.; Lissak, R.I; Drasgow, F. (1982). Recovery of Two and Three-Parameter Logistic Item Characteristic Curves: A Monte Carlo Study. Applied Psychological Measurement, Vol. 6 (3), Pp. 249-260.

30- Johnson, M.S. (2007). Modeling Dichotomous Item Response with Free-Knot Splines. Computational Statistics Data Analysis, Vol.51 (9), Pp. 4178-4192.

31- Junius, T. (1984). An Investigation of The Effect Length on Short-Form Basic Skills Competency Tests Developed by Using The One Parameter Item Response Model. Educat. D., Virginia Polytechnic Institute and State University, 149 Pages, AAT8421987.

32- Koch, W. & Dodd, B. (1995). An Investigation of Procedures for Computerized Adaptive Testing Using the Successive Intervals Rasch

Model. Educational and Psychological Measurement, Vol.55 (6), Pp. 976-990.

33- Li, H.Y. & Lissitz, R.W. (2004). Applications of the Analytically Derived Asymptotic Standard Errors of Item Response Theory Item Parameter Estimates. Journal of Educational Measurement, Vol.41 (2), Pp.85-117.

34- Magis, D., Béland, S. & Raîche, G. (2011). A Test-Length Correction to the Estimation of Extreme Proficiency Levels, Applied Psychological Measurement, Vol.35 (2), Pp. 91-109.

35- Mark, G.S. (2007). Application of Item Response Theory Based Computerized Adaptive Testing to The Strong Interest Inventory. Ph.D., University of Minnesota, AAT3285664.

36- Meijer, R.R.; Molenaar, I.W. & Sijtsma, K. (1994). Influence of Test and Person Characteristics on Nonparametric appropriateness Measurement. Applied Psychological Measurement, Vol.18 (2), Pp. 111-120.

37- Noonan, D.W.; Boss, M.W. & Gessarol, M.E. (1992). The Effect of Test Length and IRT Model on The Distribution and Stability of Three appropriateness Indexes. Applied Psychological Measurement, Vol.16 (4), Pp. 345-352.

38- Partchev, I. (2004). A Visual Guide to Item Response Theory. Friedrich-Schiller University at Jena, Pp.1-61.

39- Pozehal, B.J. (1990). Application of Item Response Theory to Criterion-Referenced Measurement: an Investigation of the Effects of Model Choice, Sample Size and Test Length on Reliability and Estimation Accuracy. Ph.D., University of Nebraska, Lincoln, Nebraska, May.

40- Reeve, B.B. (2002). An Introduction to Modern Measurement Theory. Applied Research Program, Division of Cancer Control and Population Sciences, Pp.1-67.

41- Schatschneider, C.; Francis, D.J.; Foorman, B.R.; Fletcher, J. & Mehta, P. (1999). The Dimensionality of Phonological Awareness: An Application of Item Response Theory. Journal of Educational Psychology, Vol. 91 (3), Pp. 439-449.

42- Smith, R.M. (1988). The Distributional Properties of Rasch Standardized Residuals. Educational and Psychological Measurement, Vol.48 (3), Pp. 657-667.

43- Stone, C.A. & Hansen, M.A. (2000). The Effect of Errors in Estimating Ability on Goodness-of-Fit Tests for IRT Models. Educational and Psychological Measurement, Vol.60 (6), Pp. 974-991.

44- Van der Linden, W.J. & Hambleton, R. H. (1997).Item Response Theory. In: Handbook of Modern Item Response Theory. Springer- Verlag, New York, Inc.

45- Walker, C.M. & Beretvas, N.S. (2003). Comparing Multidimensional and Unidimensional Proficiency Classifications: Multidimensional IRT as a

Diagnostic Aid. Journal of Educational Measurement, Vol.40 (3), Pp. 255-275.

46- Wang, N. (2003). Use of the Rasch IRT Model in Standard Setting: An Item-Mapping Method. Journal of Educational Measurement, Vol.40 (3), Pp. 231-253.

47- Wilcox, R.R. (1980). Determining The Length of A Criterion-Referenced Test. Applied Psychological Measurement, Vol.4 (4), Pp. 425-446.