مدى التوافق بين النموذج ثلاثي المعلمة والنظرية التقليدية في بناء اختبار محكي المرجع للثقافة الإسلامية لطلبة جامعة القصيم

د. أحمد علي مُحَمَّد الشريم

وحدة القياس والتقويم – كلية الشريعة والدراسات الإسلامية– جامعة القصيم – المملكة العربية السعودية

المُلخَص

هدفت الدراسة إلى التحقق من مدى التوافق بين النموذج ثلاثي المعلمة كأحد نماذج نظرية السمات الكامنة والنظرية التقليدية في اختيار فقرات اختبار محكي المرجع لمقرر الثقافة الإسلامية لطلبة جامعة القصيم، وتكونت عينة الدراسة النهائية من (١٠٨٦) طالبا، وتم بناء اختبارا محكي المرجع للمقرر تكون من (٧٠) فقرة من نوع الاختيار من متعدد، وتم تحليل نتائج الطلاب باستخدام برنامج (SPSS) للنظرية التقليدية، وبرنامج (BILOG-Mg3) للنموذج ثلاثي المعلمة، وأظهرت نتائج الدراسة: تحقق افتراضات النموذج ثلاثي المعلمة في بيانات الدراسة، وأن النموذج ثلاثي المعلمة أكثر تشدداً من النظرية التقليدية في اختيار الفقرات فقد استبعدت النظرية التقليدية ست فقرات من العدد الأصلي (٧٠) فقرة حيث انخفضت قيم معاملات التمييز لها عن المستوى المقبول، بينما النموذج الثلاثي استبعد تسع فقرات من (٦٤) فقرة لم تطابق النموذج الثلاثي ، وقد بلغ عدد الفقرات التي تلائم كلا من النموذج الثلاثي المعلمة والنظرية التقليدية (٥٥) فقرة ، حيث كانت قيم معالمها (الصعوبة والتمييز والتخمين) مقبولة ضمن المحكات التي أوردتها أدبيات القياس التربوي، وهي تشكل العدد النهائي لفقرات الاختبار محكي المرجع، كما تبيَّن أن هناك توافقاً كبيراً بين النظرية التقليدية والنموذج ثلاثي المعلمة في تقدير معامل ثبات الاختبار

الكلمات المفتاحية: النموذج ثلاثي المعلمة، نظرية الاستجابة للمفردة، الاختبار محكي المرجع، النظرية التقليدية في القياس

مُقَدِمَةٌ :

تُعد الاختبارات من أهم الوسائل التي يعوّل عليها في قياس وتقويم قدرات الطلاب، ومعرفة مدى مستواهم التحصيلي، ويتم بوساطتها الوقوف على مدى تحقق الأهداف السلوكية، أو النواتج التعليمية، وما يقدمه المدرس من نشاطات تعليمية مختلفة تساعد على رفع الكفايات التحصيلية لدى الطلاب، ومع تزايد الحاجة للاختبارات في مختلف مجالات العلوم الانسانية والتربوية والتعليمية فلا بد أن تكون هذه الاختبارات ذات كفاءة عالية في عملية القياس والتقويم. وقد وُجهت لأساليب النظرية التقليدية في القياس Classical Test Theory (CTT) العديد من الانتقادات انصبت أغلبها على موضوعية الاختبارات، وما تتسم به من قصور في الصدق والثبات، وقد تبلورت هذه الانتقادات من خلال ظهور بعض المشكلات منها اعتماد الدرجة الكلية للاختبار على عدد فقرات الاختبار، وعدم وجود خطية للقياس، وكذلك وحدات القياس غير ثابتة من اختبار لآخر، بالإضافة إلى القياس في أكثر من بُعد (Crocker & Algina, 1986).

وقد بُذلت جهوداً عديدة للقضاء على هذه المشكلات والتي أسفرت أخيراً عن ظهور بعض الاتجاهات الحديثة للقياس والتي أظهرت تقدماً ملموساً في تحقيق الموضوعية المطلوبة، ومن بين هذه الاتجاهات ظهور ما يسمى نظرية استجابة المفردة (IRT) Item Response Theory والنماذج المتعلقة بها.

وباستخدام نظرية استجابة المفردة (IRT) في تحليل بيانات الاختبارات النفسية والتحصيلية والتربوية، تمّ إيجاد حلول لمعظم المشكلات التي تعاني منها النظرية التقليدية في القياس ومن أهم هذه الحلول الحصول على تقديرات لمعالم المفردة متحررة من قدرات المفحوصين، وكذلك الحصول على تقديرات لمعالم القدرة للفرد متحررة من خصائص الفقرات (Hambleton & Swaminathan,1985).

وقد ظهرت فكرة تطوير اختبارات محكية المرجع وإنشاء بنوك الأسئلة في مختلف المجالات وذلك من أجل التغلب على معظم المشكلات ونواحي القصور التي ظلت تعاني منها الاختبارات

التقليدية لفترة طويلة، ويبدأ الأسلوب الرئيس في إنشاء بنك الأسئلة بإعداد فقرات الاختبار وتجريبها ثم اختبار مدى ملاءمتها لاستبعاد غير الملائم منها أو تعديلها، والإبقاء على الملائم منها وترميزها وتجميعها وحفظها لاستخدامها وقت الحاجة، ويشير مراد (١٩٩٦) أنه نتيجة انتشار فكرة إنشاء بنوك الأسئلة وفق النماذج الحديثة في القياس كوسيلة لتحقيق الموضوعية المنشودة للقياس فإنه تم ضبط الاختبارات المسحوبة فقراتها من هذه البنوك وقد حققت دقة وكفاءة تقترب من الدقة والكفاءة المعهودة في أدوات القياس الفيزيقية.

مشكلة الدراسة وتساؤلاتها:

تتحدد مشكلة الدراسة الحالية في عدم وجود اختبار تحصيلي محكي المرجع في مادة المدخل للثقافة الإسلامية لطلبة جامعة القصيم (متطلب لجميع التخصصات في الجامعة)، مبني وفق نظرية السمات الكامنة ونماذج القياس المنبثقة عنها، وخاصة النموذج الأكثر تطوراً وهو النموذج ثلاثي المعلمة لما يتمتع به هذا النموذج تحديداً من دقة في قياس قدرة الأفراد، ومعالم المفردة الثلاث (الصعوبة والتمييز والتخمين)، حيث التحرر من أثر قدرة المفحوصين على معالم فقرات الاختبار، وبنفس الوقت التحرر من أثر معالم الفقرات على قدرات الأفراد المفحوصين الذين يجيبون عنها. وبالتحديد تحاول الدراسة الحالية الإجابة عن التساؤلات التالية:

١. ما مدى تحقق افتراضات النموذج ثلاثي المعلمة في بيانات الدراسة؟
٢. ما مدى التوافق في تقدير معالم فقرات الاختبار المحكي المرجع وفق كل من النموذج الثلاثي المعلمة والنظرية التقليدية اعتماداً على ما يمكن استخلاصه من نسب اتفاق الأسلوبين في تصنيف الفقرات اعتماداً على معالم الفقرات؟
٣. ما مدى التوافق بين النظرية التقليدية في القياس باستخدام (Spearman-Brown; Guttman ; Livingston Alfa ;) والنموذج الثلاثي المعلمة (التجريبي) في تقدير قيمة معامل الثبات للاختبار ؟

أهداف الدراسة:

سعت الدراسة إلى تحقيق الأهداف التالية:

١. الاستفادة من ميزات النموذج ثلاثي المعلمة كأحد نماذج السمات الكامنة في بناء اختبار محكي المرجع في مقرر الثقافة الإسمية لطلبة جامعة القصيم كون هذا النموذج يقدر جميع معالم المفردة (الصعوبة والتمييز والتخمين) بالإضافة إلى قدرة الأفراد.
٢. الكشف عن مدى التوافق بين النظرية التقليدية في القياس والنموذج الثلاثي المعلمة في تقدير معالم فقرات الاختبار، وتقدير قيمة معامل ثبات الاختبار.

أهمية الدراسة:

تحاول الدراسة الحالية التوصل إلى أداة قياس موضوعية لقياس تحصيل طلبة جامعة القصيم في مقرر المدخل للثقافة الإسلامية وفقا لفلسفة القياس محكي المرجع حيث المحك هو الأهداف السلوكية للمقرر، وذلك باستخدام أهم نماذج استجابة المفردة وهو النموذج الثلاثي المعلمة، مما يساهم في إلقاء مزيداً من الضوء على أهمية الاختبارات المحكية المرجع في تشخيص مستوى تحصيل المتعلم ومدى إتقانه للأهداف الدراسية بالمقرر المقصود.

كما تحاول الدراسة إلقاء مزيدا من الضوء على أحد البرامج الإحصائية المتقدمة والمستخدمة في تحليل البيانات الاختبارية في ضوء نماذج استجابة المفردة، للحصول على تقديرات معالم الفقرات والأفراد بدقة وكفاءة عالية وهو برنامج بيلوج (BILOG-MG3).

مصطلحات الدراسة:

١. النموذج اللوغاريتمي ثلاثي المعلمة Three-Parameter Logistic Model: هو أحد نماذج الاستجابة للمفردة أحادية البعد، والتي تستخدم في تحليل الفقرات ثنائية الاستجابة (الدرجة إما واحد أو صفر) ويفترض أن الفقرات الاختبارية تتفاوت في صعوبتها وقوتها التمييزية، وكذلك في احتمالية التخمين، وهذا يسمح بتقاطع المنحنيات المميزة للفقرات، ومن ثم يستخدم هذا النموذج في صيغته الرياضية ثلاثة معلمات للمفردة هي: الصعوبة والتمييز والتخمين، بالإضافة إلى معلمة القدرة للأفراد المختبرين (Hambleton,1989).

٢. معلم صعوبة المفردة Item Difficulty (bi):هي القيمة على متصل السمة الكامنة المقابلة لاحتمال أن تكون إجابة الفرد عن المفردة إجابة صحيحة يساوي (٠.٥٠)كاظم (١٩٨٨).

٣. معلم القدرة Ability Parameter: هو مستوى السمة المقيسة لدى الأفراد المختبرين الذين يستجيبون على فقرات الاختبار المعد لقياس هذه السمة (Hambleton,1989).

٤. **اللوجيت Logit**: وحدة قياس كل من قدرة الفرد وصعوبة المفردة، واللوجيت هي قدرة الفرد على النجاح على الفقرات التي يعبر صفر التدريج عن صعوبتها، وصفر التدريج لكل من قدرة االفرد وصعوبة المفردة هو متوسط صعوبة الفقرات المستخدمة، ويؤدي ذلك إلى تقديرات موجبة وسالبة، مما دعى الباحثين إلى وضع وحدات مناسبة للتخلص من القيم السالبة مثل **وحدة (النيت) الجماعية** وهي تساوي (٥٠+١٠ × اللوجيت)، بمتوسط مقداره (٥٠)، و انحراف معياري يساوي(١٠)، وهي تحويل خطي لوحدة اللوجيت، كاظم (١٩٨٨).

وقد أشارت كاظم (١٩٨٨) خلال تعريفها لصعوبة المفردة وقدرة الفرد ووحدة قياسها (اللوجيت) بأنه عندما يجيب الفرد (المفحوص) (v) على المفردة (i) فإن احتمال حدوث أي من

الاستجابتين (نجاح/فشل) يعتمد فقط على قدرة الفرد (θ v) وصعوبة المفردة (bi)، ويحددها المقابل اللوغاريتمي للفرق بين هاتين المعلمتين، حيث يعتبر المقدار (θ v - bi) exp مرجحاً للنجاح في حال كانت قدرة الفرد أكبر من صعوبة المفردة (bi < θ v)، ويكون احتمال الإجابة صحيحة أكبر من(50%)، أما إذا كان (bi > θ v) فإن المقدار (θ v - bi) exp مرجحاً للفشل ويكون احتمال الإجابة الخطأ أكبر من(50%).

٥. دالة معلومات الاختبار Test Information Function : تمثل هذه الدالة بعلاقة منحنية بين متغيرين هما مستويات القدرة التي يمثلها المحور الأفقي، والمعلومات المقدمة من الاختبار ككل والتي يمثلها المحور الرأسي، وهي تعبر عن كمية المعلومات المقدمة من المجموع الكلي لفقرات الاختبار (كمية المعلومات المقدمة بواسطة الاختبار بالكامل عند أي مستوى للقدرة) لكل فرد (علام، ٢٠٠٥).

٦. الاختبار محكي المرجع Criterion Reference Test : هو ذلك الاختبار الذي يستخدم في تقدير أداء الفرد بالنسبة إلى محك (Criterion) أو مستوى أداء محدد مسبقاً، دون الحاجة إلى مقارنة أدائه بأداء زملائه، أي أن هذه الاختبارات تقيس مستويات يمكن تفسيرها تفسيراً مباشراً في ضوء مستويات محددة تتطلب تحديداً دقيقاً للمجال السلوكي الذي يقيسه الاختبار، ويقصد بالمجال السلوكي مجموعة المعارف والمهارات المحددة تحديداً دقيقاً، ويتكون الاختبار من عينة عشوائية من فقرات أو أسئلة ممثلة للمجال السلوكي المطلوب بحيث تمكن من معرفة ما يستطيع الفرد أداؤه وما لا يستطيع.

٧. النظرية التقليدية في القياس: Classical Theory of Management

تقوم النظرية التقليدية للقياس بشكل أساسي على تفسير العلامة المشاهدة للمفحوص مفاده: أن العلامة الملاحظة (X) لاختبار معين تتكون من جزأين الأول هو العلامة الحقيقية (T) ، والثاني هو الخطأ العشوائي للقياس (E) والذي توضحه العلاقة : X= T + E

محددات الدراسة:

تتحدد نتائج هذه الدراسة بالاختبار التحصيلي المستخدم، وبقدر ما يتمتع به من خصائص سيكومترية مقبولة.

كما تتحدد النتائج بقدر تحقق افتراضات النموذج المستخدم في الدراسة وهو النموذج ثلاثي المعلمة.

البرامج المستخدمة في الدراسة:

١. برنامج (BILOG-MG3): وهو برنامج يستخدم لتحليل بيانات تحتوي على فقرات للاختبارات ذات التصحيح الثنائي ويستخدم لتقدير كلٍ من معالم الفقرات (الصعوبة والتمييز والتخمين)، والقدرة للمفحوصين).

٢. برنامج (SPSS): وهو برنامج يتم تقدير إحصائيات الفقرات حسب النظرية التقليدية في القياس مثل معامل صعوبة المفردة وهو نسبة الاستجابات الصحيحة على المفردة، ومعامل تمييز المفردة وهو معامل ارتباط بوينت بيسيريال بين علامة المفردة والعلامة الكلية على الاختبار.

حيث تتضمن عملية التحليل ثلاث مراحل، وهي:

١. المرحلة الأولى: وفي هذه المرحلة يتم تقدير إحصائيات الفقرات حسب النظرية التقليدية في القياس مثل معامل صعوبة المفردة وهو نسبة الاستجابات الصحيحة على المفردة، ومعامل تمييز المفردة وهو معامل ارتباط (بوينت بيسيريل) بين علامة المفردة والعلامة الكلية على الاختبار.

٢. المرحلة الثانية وفي هذه المرحلة يتم تقدير معالم المفردة، كما يتم تحديد الفقرات الملائمة للنموذج.

٣. المرحلة الثالثة وفي هذه المرحلة يتم تقدير القدرة للأفراد و معالم المفردة حسب النموذج المستخدم في التحليل. وقد استخدم في هذه الدراسة النموذج اللوجستي الثلاثي المعلمة، والذي يتم من خلاله تقدير معالم المفردة (الصعوبة، والتمييز، والتخمين) وتقدير معلمة القدرة للمفحوصين.

الأدب النظري والدراسات السابقة:

تشير كثير من الدراسات (Hambleton & ; Hambleton Swaminathan,1985; Hambleton,1989 & Rogers,1991 ; مراد، ١٩٩٦ ؛ علام، ٢٠٠٥؛ كاظم، ١٩٨٨) إلى أنه وبالرغم من النجاح والتقدم الواضح الذي حققته بنوك الاسئلة في مجال القياس السلوكي؛ إلا أن هذه الاختبارات تحتاج لمزيد من الدراسات والبحوث لدراسة المتغيرات المختلفة التي قد تؤثر على دقة القياس والخصائص البارمترية للفقرات التي تتألف منها هذه الاختبارات.

والهدف الأساس لكل من النظرية التقليدية ونظرية استجابة المفردة هو تحديد العلاقة بين استجابات الأفراد على اختبار معين والسمة الكامنة وراء هذه الاستجابات، والأمر الأكثر أهمية في القياس النفسي والتربوي بشكل عام هو تحديد مقدار السمات الكامنة وراء أداء الأفراد على الاختبارات المختلفة، والاستفادة منها في تفسير النتائج، والتنبؤ بسلوكهم في مواقف مماثلة، واتخاذ قرارات معينة بشأنهم في ضوء هذا التقدير الكمي للسمات (Mislevy & Bock, 1990).

النظرية التقليدية تتحدث عن الدرجة الحقيقية للفرد، بينما النظرية الحديثة تتحدث عن القدرة للفرد، والواقع أن الدرجة الحقيقية والقدرة يرمزان للشيء ذاته ولكن بمقاييس مختلفة، إلا أن الفارق بينهما هو أن الدرجة الحقيقية تعتمد على الفقرات التي يتكون منها الاختبار، في حين أن مقياس " القدرة " مستقل عن فقرات الاختبار. (Hambleton & Cook 1983)

وللتفريق بين النظريتين فإن نظرية استجابة المفردة توفر مفهوماً جديداً يختلف عن الثبات يسمى دالة المعلومات لكل من الفقرات

والاختبار كاملاً، ونحصل عليها من تطبيق فقرة أو اختبار، وهذه المعلومة تعني مقدار الثقة في أننا حصلنا على معلومات تقودنا إلى تقدير قدرة فرد معين أو مجموعة من الأفراد في مستوى محدد من القدرة، أي أن كمية المعلومة نحصل عليها عند كل مستوى على متصل مقياس القدرة، بمعنى آخر أننا نحصل على قيم متعددة لكمية المعلومات التي حصلنا عليها من المفردة أو الاختبار، وهي بالتالي تختلف عن الثبات الذي هو عبارة عن مؤشر واحد فقط للاختبار لكل الأفراد المفحوصين. وتمثَّل دالة المعلومة عادة بمنحنى يبين أعلى وأقل كمية معلومات حصلنا عليها عند مستويات القدرة المختلفة. وتعتبر دالة المعلومة مهمة في نظرية استجابة المفردة مثلما الثبات في النظرية التقليدية، إلا أن الثبات يعتمد على خصائص الاختبار وخصائص المجموعة التي طُبّق عليها الاختبار، ولهذا يصعب تعميمها، أما دالة المعلومة فهي تعتمد فقط على مقياس القدرة وعلى منحنى استجابة المفردة، كذلك فإن الثبات مؤشر للخطأ المعياري في القياس وهذا المؤشر ثابت على جميع مستويات الدرجة الحقيقية، بينما دالة المعلومة مؤشر للخطأ المعياري في التقدير، وهذا الخطأ يحسب عند كل مستوى من مستويات قدرة المفحوص، وتُعرّف بالتالي دقة تقدير قدرة الفرد على كل مستوى من المستويات على متصل القدرة، بعبارة أخرى إن مؤشرات الاختبار في النظرية التقليدية تتغير بتغير المجموعة التي يطبق عليها الاختبار، أما مؤشرات الاختبار في نظرية استجابة المفردة فهي متحررة من خصائص المجموعة التي طُبّق عليها الاختبار، كما أن قدرة الفرد متحررة من خصائص الفقرات والاختبار، وعليه، فإن بناء اختبار اعتمادا على معايرة الفقرات باستخدام نماذج نظرية الاستجابة للمفردة أكثر فاعلية من مجرد الاعتماد على إحصائيات المفردة التقليدية (Lee,2004).

إن نظرية استجابة المفردة وفرت مجالاً خصباً لتحسين تطبيقات المقاييس النفسية والتربوية، بالإضافة إلى مجالات تطبيق أخرى مثل تقنين وتحليل الاختبارات، حيث أصبح من الممكن مقارنة قيم القدرات للأفراد عند التقنين بدلاً من الاعتماد على التوزيع التكراري للعينة التي طُبق عليها الاختبار، أي أنه أصبح من الممكن إعداد بنوك الأسئلة التي حددت مؤشراتها والاستفادة منها في مختلف المجالات الإنسانية(Mislevy & Bock, 1990) .

ونماذج استجابة المفردة نماذج رياضية تعرّف العلاقة بين أداء المفحوص الملاحظ على الاختبار وبين السمة أو القدرة غير الملاحظة للمفحوص والتي تكون مسؤولة عن أداء المفحوص على الاختبار، ويعد التقدير الإحصائي للعلاقة بين احتمال الاستجابة الصحيحة لفقرة اختبارية والسمة التي يفترض أن الاختبار يقيسها هو المشكلة الرئيسة التي تواجه من يستخدم نماذج هذه النظرية في بناء الاختبارات وتحليل فقراتها، وتمثل هذه العلاقة لكل فقرة من فقرات الاختبار بمنحنى يطلق عليه منحنى خصائص المفردة (ICC) ويتخذ شكل المنحنى اللوغاريتمي التراكمي، وهذا المنحنى يمثل الانحدار غير الخطي لدرجات المختبرين في فقرة من فقرات الاختبار على القيم التقديرية للسمة أو القدرة التي يهدف الاختبار لقياسها، ومن منحنى خصائص المفردة تتضح كيفية عمل الفقرات، كما تتضح قدرة المفحوص المناظرة لاحتمال معين للأداء على فقرة ما(علام، ٢٠٠٥).

وترتكز نماذج نظرية استجابة المفردة مثل كثير من النماذج الرياضية على افتراضات عدة يجب تحققها في البيانات التي تخضع للتحليل باستخدام هذه النماذج، وهي أحادية البعد للسمة التي يقيسها الاختبار، والاستقلال الموضعي للفقرات، والمطابقة لمنحنى خصائص المفردة، والتحرر من السرعة (Hambleton,1989).

وتشتمل نماذج استجابة المفردة على مجموعة من النماذج الرياضية تُعرف باسم نماذج الاستجابة للمفردة Item Response Models وهذه النماذج تحقق القياس الموضوعي، وتعالج كثير من المشكلات في القياس التي عجزت عن مواجهتها النظرية التقليدية في القياس، ولعل أكثر النماذج استخداماً وانتشاراً في مجال القياس والتقويم النفسي والتربوي ثلاثة نماذج هي: النموذج اللوغاريتمي أحادي المعلمة، والنموذج ثنائي المعلمة، والنموذج ثلاثي المعلمة.

وتعد النماذج الأحادية والثنائية والثلاثية من أهم النماذج لنظرية استجابة المفردة، وهذه النماذج ملائمة للفقرات ثنائية التدريج.

النموذج أحادي المعلمة والمعروف بنموذج راش يعتمد على تدريج الفقرات وفقا لصعوبتها فقط، أي أن فقرات الاختبار تتفاوت في صعوبتها فقط، بينما تتساوى في قدرتها التمييزية، ولا تسمح بالتخمين، وبالتالي فإن احتمال استجابة الفرد على المفردة يعتمد على قدرة الفرد وصعوبة المفردة. وقد تم إضافة معلم ثاني للنموذج الأحادي وهو معلم التمييز لكل فقرة، وبالتالي فإن منحنيات الفقرات لن تتوازى بل ستتقاطع، ويفترض النموذج الثنائي عدم وجود أثر للتخمين، ويقوم النموذج بتقدير صعوبة الفقرات والتمييز لها بالإضافة إلى معلم القدرة للأفراد فقط.

أما النموذج الثلاثي المعلمة - والذي ستستخدمه الدراسة الحالية- فيعود الفضل إلى بيرنباوم Birrnbaum الذي أضاف له معلم ثالث وهو التخمين لكل فقرة، ليفسر الأداء غير الصفري للأفراد منخفضي القدرة على فقرات الاختبار من نوع الاختيار من متعدد، وهذا الأداء غير الصفري يرجع إلى احتمال تخمين الإجابات الصحيحة لفقرات الاختبار عندما يكون الفرد غير متأكد من الإجابة الصحيحة، وبالتالي يفترض النموذج ثلاثي المعلمة أن فقرات الاختبار تتفاوت في صعوبتها وفي قوة تمييزها بين المفحوصين، واحتمالية تخمين الإجابة لها، وقد سمي هذا النموذج بثلاثي المعلمة لأن احتمال

الاستجابة هي دالة لقدرة المفحوص وثلاثة من معلمات المفردة هي: الصعوبة والتمييز والتخمين (Hambleton & Rogers,1991).

إن تقييم حُسن المطابقة للبيانات الاختبارية للفقرات للنموذج في نظرية استجابة المفردة يتطلب التحقق من افتراضات النموذج التالية:

١. أحادية البعد Unidimentionality: تكون مجموعة الفقرات أحادية البعد إذا كانت تقيس قدرة أو سمة واحدة، وتعني أن فقرات الاختبار لا تختلف فيما بينها إلا من حيث الصعوبة فقط.

٢. الاستقلال الموضعي للمفردة Local Item Independence: أشار هامبلتون وسواميناثان (Hambleton & Swaminathan,1985) إلى أن افتراض الاستقلال الموضعي يعني أن استجابات المفحوصين على فقرات الاختبار تكون مستقلة إحصائيا عند مستوى قدرة معينة، أي أن استجابة المفحوص لفقرة ما يجب أن لا تؤثر سلبا أو إيجابا على استجابته لفقرة أخرى؛ بمعنى أن إجابة المفردة لا تعطي أية تلميحات أو معيقات لإجابة فقرة أخرى، وهذا الافتراض يوضح أن قدرة المفحوص وخصائص المفردة هي وحدها التي تؤثر في استجابته على فقرات الاختبار. ويتحقق افتراض الاستقلال الموضعي للمفردة إذا كانت مجموعة الفقرات أحادية البعد، أي إذا كانت تقيس قدرة أو سمة واحدة، فإذا لم يتحقق افتراض أحادية البعد فان درجات فقرات الاختبار ستكون مرتبطة عند مستوى قدرة معين لأن فقرات الاختبار تقيس سمة ثانية أو أكثر. وإذا كانت الفقرات تقيس أكثر من بُعد أو(قدرة) فيمكن تجميع الفقرات لكل بعد في مجموعة متجانسة باستخدام التحليل العاملي، ثم استخدام النموذج مع كل مجموعة متجانسة من الفقرات لتدريجها
(كاظم، ١٩٨٨).

٣. التحرر من السرعة Non-Speediness: تفترض نماذج استجابة المفردة أن الاختبار ليس اختبار سرعة، بمعنى أن الزمن ليس له دور في الإجابة على فقرات الاختبار.

وقد اهتم العديد من الباحثين والمتخصصين في القياس النفسي والتربوي في الجامعات والمراكز البحثية أمثال (Crocker ; Algina, Anderson; Hambleton; Wright; Brooks) وغيرهم بإجراء العديد من الدراسات على نماذج السمات الكامنة، وتم تطوير برامج تحليل حاسوبية خاصة تتلاءم مع هذه النماذج مثل برنامج (بيكال، وديكوت، وميكروسكال، وبيلوج، وروم، وميني ستيبس، وبغ ستيبس...) وغيرها العديد من البرامج، كما تم تطبيق النماذج في تدريج فقرات اختبارات ومقاييس عالمية. وقد أجريت العديد من الدراسات لمقارنة الخصائص السيكومترية للاختبار حسب النظرية التقليدية ونظرية السمات الكامنة، كما جرت محاولات عديدة لبناء أو تدريج اختبارات ومقاييس مختلفة باستخدام نماذج نظرية السمات الكامنة وفيما يلي عرض لأهم هذه الدراسات العربية والأجنبية مرتبة حسب تاريخ نشرها.

فقد هدفت دراسة علام (١٩٩٠) إلى استخدام نماذج استجابة المفردة في بناء وتصميم مقياس هدفي المرجع لقياس مدى تمكن طلاب الدراسات العليا بكلية التربية بجامعة الأزهر من المعارف الأساسية المتعلقة بإعداد خطة البحوث النفسية التربوية وتصميم ميزان لوغاريتمي خطي لتحديد مستويات التمكن المتوقعة من الطالب في كل هدف من الأهداف السلوكية التي تقيسها الفقرات الخاصة بالمقياس، وقد حدد الأهداف في ضوء التحليل السلوكي للمكونات التي ينبغي توافرها في خطة البحوث النفسية التربوية، ثم قام بكتابة فقرة لقياس كل هدف من قائمة الأهداف التي أعدها، ثم أجرى دراسته على عينة من (٤٨) طالبا واستخدم طريقة بروكس (PROX) في تحليل بياناته لتقدير صعوبة الفقرات وقدرات الطلاب على المقياس بوحدة اللوجت حسب نموذج أحادي المعلمة، وناقش نتائجه في ضوء التقديرات التي توصل اليها باستخدام النموذج.

وهدفت دراسة جمحاوي (٢٠٠٠) إلى المقارنة بين النظرية التقليدية ونظرية السمات الكامنة من حيث مدى اتفاقها في انتقاء فقرات متوافقة من حيث الصعوبة والتمييز في مقياس للقدرة الرياضية للصفوف الثامن والتاسع والعاشر، وقد قارنت خصائص (٣٩) فقرة وفق الاسلوبين على عينة من (١٠٦١) طالبا وطالبة، وحللت البيانات باستخدام برنامج راسكال لنظرية استجابة المفردة وبرنامج SPSS للنظرية التقليدية، وأشارت نتائجها إلى وجود اتفاق عالي بين النظرية التقليدية والنموذج الثنائي المعلمة في تقدير الصعوبة والتمييز للفقرات، وأشارت كذلك إلى أن خصائص الفقرات المنسجمة مع النموذج تزداد بزيادة عدد معالمه، وأشارت النتائج كذلك إلى عدم وجود فرق بين معاملي ثبات المقياس الناتجين من النظرية التقليدية والنموذج الثنائي المعلمة، بينما يوجد فرق بين قيم الثبات الناتجة من النظرية التقليدية والنموذجين الأحادي والثلاثي المعلمة.

وفي دراسة قام بها كيم (Kim, 2001) هدفت إلى مقارنة دقة التقديرات التي تم الحصول عليها باستخدام طريقة الأرجحية العظمى الشرطية وطريقة الأرجحية العظمى الهاشمية وطريقة الأرجحية العظمى المشتركة وطريقة بييز التي تعتمد على توقع التوزيع البعدي وذلك من خلال استخدام أسلوب المحاكاة، إذ توصلت الدراسة إلى أن نتائج تقدير معلمات المفردة ومعلمة القدرة كانت متشابهة للطرق الأربعة.

وقد أجرى بيلتون (Pelton, 2002) دراسة هدفت إلى مقارنة الدقة والاستقرار في تقدير معلمة الصعوبة ومعلمة القدرة باستخدام النظرية التقليدية في القياس والنماذج اللوجستية في نظرية الاستجابة للمفردة وذلك من خلال استخدام أسلوب المحاكاة. إذ توصلت الدراسة إلى مجموعة من النتائج منها أن تقديرات القدرة يمكن مقارنتها عبر النظرية التقليدية ونظرية الاستجابة للمفردة، حيث تتباين تقديرات القدرة حسب كمية المعلومات المتوفرة في مجموعة البيانات المولدة، التي تتأثر بدورها بأحادية البعد، ودرجة التخمين ومدى التباين في صعوبة الفقرات مقارنة بقدرات الأفراد، كما توصلت الدراسة إلى أنه بوجود حجم عينة معتدل (٩٩٩) مفحوص وعدد

فقرات ملائم (٣٣) فقرة، فإن النموذج اللوجستي الثنائي يقدم تقديرات أكثر دقة لمعلمة الصعوبة من نموذج راش والنموذج اللوجستي الثلاثي والنظرية التقليدية في القياس وذلك في ظل وجود تخمين قليل في الإجابة على فقرات الاختبار.

أما واجنر وهارفي (Wagner & Harvey,2003) وبهدف تطوير أداة جديدة لأوجه التفكير الناقد لاختبار واتسون- جليسر باستخدام نظرية استجابة المفردة، فقد قام الباحثان بزيادة عدد البدائل لتكوين اختبار جديد(WAT) وتكونت عينة الدراسة من (٤٠٧) طالبا وطالبة بالمستوى الجامعي، وأظهرت نتائج التخمين عن انخفاض نسبة معلم التخمين (C) للفقرات، وتحسن في دالة معلومات المفردة في الاختبار المطور، وكانت اخطاء القياس في اختبار واتسون – جليسر أكبر من أخطاء القياس للاختبار الجديد بمقدار(٥٠%) على الأقل.

أما الدراسة التي قام بها عبابنه (٢٠٠٤) فهدفت لمعرفة أثر حجم العينة وطريقة انتقائها وعدد الفقرات وطريقة انتقائها على دقة تقدير معالم المفردة والقدرة لاختبار قدرة عقلية باستخدام نظرية الاستجابة للمفردة توصلت الدراسة إلى عدة نتائج منها أن الدقة في تقدير معلمة القدرة تزداد عند زيادة عدد فقرات الاختبار، كما أن تقديرات معلمة القدرة تكون أكثر دقة عند استخدام الطريقة المعيارية في اختيار الفقرات مقارنة بالطريقة العشوائية.

وبهدف الكشف عن الفروق في الدرجة الحقيقية لليقظة العقلية المقدرة باستخدام نظرية السمات الكامنة والنظرية التقليدية أجرت عيد (٢٠٠٤) دراسة على عينة مكونة من (٢٥٠) طالبا وطالبة من جامعة الكويت وطبقت عليهم اختبار اليقظة العقلية كمقياس للذكاء العام، وقد بينت النتائج وجود فروق ذات دلالة إحصائية بين الدرجات الخام للطلبة تعزى للجنس، وكذلك وجود فروق ذات دلالة إحصائية بين الدرجات الخام للطلبة تعزى للتخصص العلمي، بينما لا يوجد فروق ذات دلالة إحصائية بين متوسطي الدرجات الحقيقية (نموذج راش) للطلبة على مقياس اليقظة العقلية تعزى للجنس، وكذلك عدم وجود فروق ذات دلالة إحصائية بين متوسطي الدرجات الحقيقية (نموذج راش) تعزى للتخصص، وكذلك الحال بالنسبة للدرجة الحقيقية المقدرة لليقظة العقلية المقدرة حسب النظرية التقليدية فلا يوجد فروق بين المتوسطات تعزى للجنس أو للتخصص العلمي، كما بينت نتائج دراستها عدم وجود فروق ذات دلالة إحصائية بين متوسطي الدرجات الحقيقية المقدرة بنموذج راش وتلك المقدرة بالنظرية التقليدية.

وقد أجرى النجار وعيلبوني (٢٠٠٨) دراسة بهدف بناء بنك أسئلة في الثقافة الحاسوبية للمرحلة الثانوية في الأردن باستخدام نظرية استجابة المفردة والنظرية التقليدية، وقد قاما بكتابة قائمة من الفقرات (٣٣٠) فقرة من نوع الاختيار من متعدد مقسمة إلى صورتين، وتم إيجاد الخصائص السيكومترية للفقرات وفق النظرية التقليدية، ثم إيجادها وفق نظرية السمات الكامنة (نموذج راش) باستخدام برنامج BILOG-Mg3، وقد تم استبعاد الفقرات حسب معايير النظرية التقليدية حيث استبعدت الفقرات التي يقل تمييزها عن (٠.٢٥)، والفقرات التي يقع معامل صعوبتها خارج المدى (٠.٣٥ – ٠.٨٩). وحسب نظرية استجابة المفردة تم حذف الفقرات التي لم تطابق النموذج من خلال اختبار كاي تربيع عند مستوى الدلالة الإحصائية (٠.٠٠١)، وأسفرت النتائج عن حذف ٦% من الفقرات حسب النظرية التقليدية، بينما حذفت ٢٤% من الفقرات حسب نموذج راش، وبلغ معامل الارتباط بين قيم الصعوبة للفقرات حسب النظرية التقليدية وقيم الصعوبة حسب نظرية السمات الكامنة (- ٠.٩٩) وهو ارتباط عكسي عالي.

وأجرى تيان (Tian, ٢٠٠٩) دراسة بهدف الحصول على نماذج متكافئة من اختبار الكفاءة في اللغة الإنجليزية لطلبة جامعة ميشغان Michigan، باستخدام النموذج ثلاثي المعلمة مقارنة بالنظرية التقليدية في القياس، وقد استخدم نتائج الطلبة في الجامعة على فقرات الاختبار المكون من (١٥٠) فقرة على عينة مكونة من (٣٣٠٠) طالبا وطالبة، وأظهرت النتائج أن النموذج ثلاثي المعلمة لم يختلف عن النظرية التقليدية في القياس بما يخص قسم الاستماع فقط، بينما كانت الأفضلية للنموذج الثلاثي لبقية أقسام الاختبار (القراءة، الكتابة، القواعد) وذلك في تقدير معالم القدرة للأفراد ومعالم الفرات اعتمادا على قيم الخطأ المعياري في التقدير ودالة المعلومات للاختبار.

وهدفت دراسة أبو خليفة (٢٠٠٩) إلى توظيف نموذج راش في انتقاء فقرات مقياس تقدير لتقييم اتجاهات طلبة كلية العلوم التربوية، ولتقييم اتجاهات عينة الدراسة المكونة من (٢٥٠) طالبا وطالبة، تم اعداد مقياس تقدير يتكون من (٣٨) فقرة مدرجة تدريجا خماسياً، وحللت البيانات الناتجة عن تقديرات الطلبة لاتجاهاتهم نحو التخصص، باستخدام برنامج SPSS وبرنامج (RUMM2020) وتم تقدير معالم الفقرات ومعالم الأفراد والخطأ المعياري في التقدير للمعالم كما تم حساب إحصائي مطابقة كل فقرة لنموذج سلم التقدير، وتم استبعاد (١٠) فقرات وذلك لأن إحصائي حسن المطابقة (Fit Residuals) لكل منها خارج نطاق المدى (- ٢.٥ - ٢.٥).

وهدفت دراسة عرفان (٢٠٠٩) إلى تطوير مقياس دافعية الإنجاز باستخدام أسلوب القياس الحديث القائم على نظرية الاستجابة للمفردة وبغرض الحصول على صورة مختصرة متعادلة القياس من المقياس المطور، بالإضافة إلى دراسة أثر متغير الجنس على دافعية الانجاز في ظل القياس الموضوعي القائم على نظرية الاستجابة للمفردة، وطبق المقياس على عينة مكونة من (٧٩٥) طالباً وطالبة من طلبة الصف الأول ثانوي، وباستخدام برنامج التحليل

الإحصائي Win-steps أسفرت نتائج الدراسة عن الحصول على مقياس خطي أحادي البعد مكون من (١٢٩) فقرة بعد استبعاد (١٥) فقرة لعدم ملاءمتها للنموذج، وقد تم سحب صورتين من المقياس تتكون كل صورة من (٥٥) فقرة تغطي كافة جوانب النموذج النظري للمقياس، كما أسفرت النتائج عن عدم وجود أثر لمتغير الجنس على دافعية الانجاز.

وأجرت الناغي (٢٠١١) دراسة بهدف الكشف عن أثر اختلاف عدد فقرات الاختبار على تقدير قدرات الأفراد على الاختبار التحصيلي مرجعي المحك في الرياضيات باستخدام النموذج اللوغاريتمي ثلاثي المعلمة، وكذلك أثر اختلاف عدد فقرات الاختبار على دالة معلومات الاختبار التحصيلي باستخدام النموذج الثلاثي المعلمة، وقد قامت الباحثة ببناء اختبار تحصيلي في وحدة المصفوفات مكون من (٧٨) فقرة اختبارية،وقد جربت الاختبار على عينة استطلاعية مكونة من (٥٠٠) طالبا وطالبة من طلاب الصف الأول الثانوي، واعتماداً على قيم معامل ثبات ألفا تم حذف ثمان فقرات حيث كانت تؤثر سلبا على ثبات الاختبار، وبلغ معامل الثبات (كيودر وريتشاردسون ٢٠) الكلي للاختبار (٠.٨٧) ونفس القيمة لمعامل ثبات ألفا. وتم تقدير مؤشرات صدق للاختبار مثل الصدق العاملي وصدق المحتوى عن طريق المحكمين، وفي نهاية الاجراءات للصدق والثبات أصبح عدد فقرات الاختبار النهائي (٦٦) فقرة، واستخدمت برنامج BILOG-MG3 لتحليل البيانات الاختبارية حسب النموذج الثلاثي المعلمة، وأجرت التحليلات بحيث تم حذف البيانات الصفرية والتامة، واعادت التحليل بهدف استبعاد الأفراد غير الملائمين للنموذج، وتم اشتقاق ثلاث أطوال للاختبار (٢٢، ٤٤، ٦٦) فقرة على التوالي، تم إعادة التحليل مرة ثالثة لتحديد الفقرات غير الملائمة وفقا لمعامل الارتباط الثنائي الأصيل، ومرة رابعة للحصول على تقديرات القدرة للأفراد، وأظهرت النتائج أن أعلى قيمة لمتوسط تقدير قدرة الافراد كان في حالة الاختبار المكون من (٤٤) فقرة، وتنخفض قيمة الجذر التربيعي لمتوسط مربعات الأخطاء بزيادة عدد فقرات الاختبار.

وهدفت دراسة الحموري (٢٠١١) إلى استقصاء إمكان استخدام نماذج السمات الكامنة في بناء اختبار تحصيلي من إعداد المعلمة، وتم بناء اختباراً تحصيلياً مكون من (٢٣) فقرة من نمط الاختيار من أربعة بدائل، تقيس العدد نفسه من الأهداف السلوكية المقصودة، وطبق الاختبار على عينة من (٢٨٤) طالباً وطالبة، وحللت البيانات الناتجة من الاستجابة للفقرات وفق النماذج الثلاثة لنظرية استجابة المفردة (الأحادي والثنائي والثلاثي)، وأظهرت نتائج دراستها عدم ملاءمة نماذج الاستجابة للمفردة الثلاثة للاختبارات ذات العدد القليل من الفقرات.

وهدفت دراسة هادي ومراد (٢٠١٣) إلى إعداد وتقنين اختبار في اللغة العربية لقبول الطلاب ببرنامج الدراسات العليا بكليات العلوم الإنسانية بجامعة الكويت، وتم إعداد (١٣٠) فقرة وتجريبها على (٨٦) طالبا وطالبة بالدراسات العليا، وأسفر التجريب عن حذف (٣٠) فقرة وتعديل (١٦)، ثم تم تطبيق الاختبار على عينة حجمها (٣٣٥)، وتم تحليل البيانات بالطريقة التقليدية وباستخدام النموذج الأحادي المعلمة (راش) وأسفرت نتائج النظرية التقليدية في القياس عن حذف (١٧) فقرة وبلغ معامل الثبات كرونباخ ألفا (٠.٨٦)، أما التحليل باستخدام نموذج راش فقد نتج عنه حذف (٢٦) فرداً غير ملائمين للنموذج، وحذف (١٩) فقرة غير ملائمة للنموذج، وبلغ ثبات الاختبار (٠.٨١) للفقرات، و(٠.٩٩) لقدرات الأفراد وتراوح مدى القدرة بين (-٢ - ٢) بمتوسط يساوي (٠.٥٠) لوجيت.

تعليق على الدراسات السابقة:

تعددت الدراسات التي تناولت نماذج نظرية السمات الكامنة والنظرية التقليدية في القياس وخاصة في بناء الاختبارات والمقاييس وانتقاء الفقرات، وركزت أغلب الدراسات على استخدام نموذج راش (أحادي المعلمة) باعتباره النموذج الأسهل كونه ينتج معلمة القدرة للأفراد ومعلمة واحدة فقط للمفردة وهي الصعوبة، ويفترض تساوي الفقرات جميعها في معامل التمييز، وهذا الافتراض يصعب الدفاع عنه أو تحقيقه كون الفقرات لا تتساوى في القدرة التمييزية بين الأفراد، كما أن نموذج (راش) لا يقدّر احتمال التخمين، ومن المعلوم أن الفقرات من نوع الاختيار من متعدد قد يلجأ المفحوص إلى التخمين في حالة عدم قدرته على الإجابة الصحيحة، والدراسات التي استخدمت النموذج ثلاثي المعلمة في انتقاء فقرات المقاييس والاختبارات التحصيلية قليلة جداً -في حدود علم الباحث- مع أن هذا النموذج هو أكثر نماذج نظرية السمات الكامنة تطوراً فقد جاء تطويراً للنموذج الأحادي والثنائي المعلمة. إن استخدام النموذج ثلاثي المعلمة في بناء الاختبارات التحصيلية الهامة كاختبار المقررات العامة التي يستفيد منها عدد كبير من الطلبة لم يحظ بالقدر الكافي من الدراسات على الرغم من الفوائد الكبيرة التي أشار لها كثير من الباحثين وعلماء القياس والتقويم، ولذلك تأتي هذه الدراسة كمحاولة للاستفادة من ميزات هذا النموذج في بناء اختبار هام يستفيد منه آلاف الطلبة في جامعة القصيم كون هذا النموذج يقدر جميع معلم المفردة (الصعوبة والتمييز والتخمين) بالإضافة إلى قدرة الأفراد.

الطريقة والإجراءات:

عينة الدراسة:

تم اشتقاق عينة الدراسة الاستطلاعية من الطلاب الذكور الذين يدرسون مقرر المدخل للثقافة الإسلامية للفصل الدراسي الثاني من العام الدراسي ١٤٣٣/١٤٣٤هـ في كلية البكيرية وكلية مجتمع بريدة وهما كليتان تدرّسان نفس المقرر، وقد اختار الباحث هاتين الكليتين كونهما بعيدتان عن كليات المقر الرئيسي للجامعة لضمان عدم تسرب الأسئلة لعينة الطلاب الرئيسية، ومع أن تطبيق الاختبار على العينة الرئيسية سيكون في الفصل الدراسي الأول للعام التالي وليس بنفس الوقت للعينة الاستطلاعية، وتكونت العينة

الاستطلاعية من (٢٦٠) طالباً، وقد اختار الباحث طلاب العينة الاستطلاعية للتحقق من الخصائص السيكومترية لأداة الدراسة الحالية أثناء بنائها، بينما تكونت العينة النهائية من (١٠٨٦) طالباً من الطلاب الذكور الذين يدرسون مقرر المدخل للثقافة الإسلامية للفصل الدراسي الأول من العام الدراسي ١٤٣٤/١٤٣٥هـ في مختلف كليات جامعة القصيم (المقر الرئيسي) باستثناء كلية البكيرية وكلية مجتمع بريدة.

أداة الدراسة:

لتحقيق أهداف الدراسة قام الباحث ببناء اختبار تحصيلي محكي المرجع لمقرر المدخل للثقافة الإسلامية بالاتفاق مع عمادة كلية الشريعة بجامعة القصيم ليكون هو الاختبار النهائي الفصلي للمقرر، والذي يتم تصحيحه آلياً، حيث الأهداف السلوكية في المقرر هي المحك، ووفقاً لخطوات بناء الاختبار محكي المرجع التالية (علام، ٢٠٠٠):

أولاً: تحديد المجال السلوكي الذي يقيسه الاختبار، وتحليله إلى مكوناته من معارف ومهارات. وقد قام الباحث بتحليل بنية المحتوى لمقرر المدخل للثقافة الإسلامية الذي تدرّسه جامعة القصيم.

ثانياً: صياغة الأهداف السلوكية المتعلقة بنواتج تحليل المجال السلوكي، حيث قام الباحث بتحليل مقرر الثقافة الإسلامية تحليلاً دقيقاً ومفصلاً، وقد أعد قائمة مكونة من (١٠١) هدفاً شاملة لجميع مفردات المقرر الدراسي (ملحق رقم (١)). وللتأكد من شمول الأهداف للوحدة الدراسية وتمثيلها للمستويات المعرفية المختلفة، قام الباحث بعرض قائمة الأهداف على أربعة من أعضاء هيئة التدريس وأصحاب الخبرة في تدريس المقرر في كلية الشريعة بجامعة القصيم لإبداء الرأي حول مدى شمول الأهداف للمقرر الدراسي وطريقة صياغتها، وقد تمت مناقشة ملاحظاتهم بالتفصيل وإجراء التعديلات المناسبة بناءً عليها.

ثالثاً: بناء جدول المواصفات للاختبار بغرض أن ترتبط فقرات الاختبار ارتباطاً مباشراً بالأهداف السلوكية، وبالتالي نضمن صدق المحتوى للاختبار ككل، ويسهل عملية تحكيم الاختبار للتحقق من صدق البناء له.

رابعاً: كتابة فقرات الاختبار: فقد قام الباحث بكتابة جميع الفقرات الممكنة التي تقيس جميع الأهداف السلوكية من نوع الاختيار من متعدد، ولكل فقرة أربعة بدائل واحد منها يمثل الإجابة الصحيحة، وقد راعى الباحث الشروط الواجب توافرها في مثل هذا النوع من الفقرات، وقد حرص الباحث على أن تقيس كل فقرة هدفاً محدداً وفقاً لقائمة الأهداف التي أعدها وبلغ عدد الفقرات الأولي (١١٢) فقرة.

خامساً: تحكيم فقرات الاختبار، وقد أعد الباحث استبانة للحكم على الاختبار من حيث:

١. مدى ارتباط المفردة بالهدف الذي تقيسه من حيث المحتوى والمستوى المعرفي.
٢. صياغة الفقرة بلغة واضحة وسليمة ومفهومة.
٣. خلو الفقرة من أية إشارات لفظية للاجابة الصحيحة.
٤. استقلال الفقرة عن غيرها من فقرات الاختبار.
٥. ابراز الفقرة لمشكلة واضحة ومحددة.
٦. جاذبية وملاءمة البدائل غير الصحيحة للفقرة للطلبة مناسبة.

وقد وضع أمام كل فقرة مقياس تقدير متدرج من (١- ٥) ، ثم عُرضت هذه الاستبانة مع فقرات الاختبار وجدول المواصفات وقائمة الأهداف السلوكية للمقرر على ستة أساتذة من أعضاء هيئة التدريس في كلة الشريعة ([1]). وبالاضافة إلى وضع التقديرات ، فقد طلب منهم التعديلات المناسبة، ومن خلال هذه الملاحظات ومناقشاتهم قام الباحث بإجراء التعديلات المناسبة، حيث تراوحت متوسطات قيم التقديرات للفقرات بين (٢.١١ – ٥.٠) وقد قام الباحث باستبعاد الفقرات التي يقل متوسط اتفاق المحكمين عليها عن (٣) وبالتالي فقد تم حذف (١٨) فقرة، وهذه الفقرات إما كانت سهلة جداً أو مكررة في الفكرة مع فقرة أخرى، أو تقيس معلومة اختلف العلماء في الاتفاق عليها، وبقي (٩٤) فقرة تم تعديل عددا منها حسب ملاحظات المحكمين. وقد أعيدت الفقرات المتبقية والتي تم تعديلها إلى المحكمين مرة أخرى ومن خلال جلستي نقاش مشتركة بين الباحث والمحكمين تم الموافقة بالإجماع على الصياغة النهائية للفقرات ومناسبتها للمقرر والطلبة.

وبعد ذلك قام الباحث باختيار (٧٠) فقرة، (ملحق رقم (٢)) من بين الفقرات حسب جدول المواصفات الذي أعده الباحث لهذا الاختبار وبشكل يتطابق مع الأوزان النسبية لفصول ومفردات المقرر، لتكون هذه المجموعة هي الفقرات التي يتكون منها الاختبار النهائي المعتمد للمقرر وجاء اختيار هذا العدد ليتناسب مع الدرجة الكلية للاختبار النهائي المعمول به في جامعة القصيم وهي (٧٠) درجة، وبواقع درجة واحدة لكل فقرة، وكذلك ليناسب الوقت المخصص للاختبار حسب نظام الجامعة وهو ساعتان.

---

[1]) يتقدم الباحث بالشكر الجزيل لفريق التحكيم أعضاء هيئة التدريس: د. محمد عبد الرحمن الدخيل، د. عمر عبد الله العمر، د. إبراهيم المرشد، د. نائل قرقز، د. محمود العلوان، د. حسن العمري، على جهودهم الكبيرة وأوقاتهم الثمينة التي قدموها لإخراج الاختبار بصورته النهائية.

❖ **صدق وثبات الأداة:**

أ – الصدق:

أولاً: للتحقق من أن فقرات الاختبار تشكل عينة ممثلة لمجتمع الفقرات التي تقيس المهارات الخاصة بالمقرر، أو بمعنى آخر للتحقق من صدق المحتوى للاختبار ؛ فقد تم حساب متوسطات تقديرات المحكمين للمفردة الخاصة بمدى ارتباط كل فقرة بالهدف الذي تقيسه بشكل منفرد (من الاستبانة الخاصة بتقديرات المحكمين لصياغة فقرات الاختبار)، وقد وجد أن متوسط تقديرات المحكمين تراوحت بين (٣.٢٢ - ٥) للفقرات التي تم الابقاء عليها بعد حذف الفقرات ذات التقديرات المنخفضة، وتعديل الفقرات حسب ملاحظات المحكمين حيث تم الاجتماع بالمحكمين بواقع جلستي عمل مطولتين وتم مناقشة كل فقرة منفردة والاتفاق بالإجماع على صياغتها ومناسبتها ومناسبة جميع بدائلها، مما يدل على اتفاق المحكمين فيما يتعلق بمطابقة كل هدف للمفردة التي تقيسه، ومناسبتها للاختبار النهائي للمقرر.

ثانياً: للتحقق من قدرة الاختبار على التمييز بين المجموعات المتمايزة في السمة (موضوع الاختبار)، فقد قام الباحث بتجريب الاختبار على عينة استطلاعية مكونة من (٤٨) طالباً لم يدرسوا المقرر، ومقارنة متوسطاتهم على الفقرات مع متوسطات عينة الدراسة الاستطلاعية

(٢٦٠) بعد أن درسوا المقرر.

ولاختبار دلالة الفرق بين المتوسطين تم استخدام اختبار (t) للبيانات المستقلة، والجدول التالي يبين الدلالة الاحصائية لاختبار (t).

جدول رقم (١): الدلالة الإحصائية لاختبار (t) للمجموعات المتمايزة

المجموعات	المتوسط /٧٠	الانحراف المعياري	قيمة ت	الدلالة الإحصائية
قبل التعليم	١٤.٤٥	٤.٧٨	١٣.٠٧	٠.٠٠٠
بعد التعليم	٤٤.٨٢	١٢.٧١		

يوضح الجدول (١) أن هناك فرقا دالا إحصائياً عند مستوى دلالة احصائية يقل عن (٠.٠٠١) بين أداء المجموعتين، مما يدل على أن الاختبار يميز بين المجموعات المتمايزة على السمة، أي أن الاختبار يقيس الأداء المتعلق بمجال سلوكي محدد من المعارف والمهارات التي تعلمها الطلبة في مقرر الثقافة الإسلامية.

ثالثاً: الصدق العاملي، للتحقق من الصدق العاملي للاختبار استخدم الباحث التحليل العاملي Factor Analysis باستخدام برنامج SPSS، وباستخدام طريقة المكونات الأساسية Principal Components، وحساب الجذر الكامن لكل عامل، ونسبة التباين المفسر الذي تفسره العوامل الناتجة من التحليل، وقد كشف التحليل عن وجود عاملا واحداً مسيطراً، وسيأتي التفصيل عن هذا التحليل في عرض نتائج الدراسة.

ب. ثبات الاختبار:

للتأكد من ثبات الاختبار والمكون من (٧٠) فقرة فقد تم تطبيق الاختبار على العينة الاستطلاعية (٢٦٠) طالباً، وتم تحليل بياناتهم على مستوى المفردة الواحدة، وقد تم حساب الثبات بعدة طرق هي كما يأتي:

١. طريقة الاتساق الداخلي: تم حساب قيم معاملات الارتباط بين درجة المفردة والدرجة الكلية للاختبار، وقد تراوحت قيم معاملات الارتباط بين (٠.١٧ و ٠.٥٣) وجميعها ذات دلالة إحصائية عند مستوى دلالة إحصائية أقل من (٠.٠٠١)، ويتضح من ذلك أن قيم معاملات الارتباط بين الفقرات والدرجة الكلية للاختبار مقبولة مما يدل على درجة عالية من الاتساق الداخلي لفقرات الاختبار.

٢. طريقة ألفا Alfa: تم تقدير قيمة معامل الثبات الكلي للاختبار بطريقة ألفا وكانت تساوي (٠.٩٢٢) في حين تراوحت قيم ثبات ألفا العام للاختبار في حال حذف كل فقرة من الاختبار بين (٠.٩٢٠و٠.٩٢٤)، وهذا مؤشر آخر على ثبات الاختبار ومناسبة الفقرات جميعها، ولم يكن هناك حاجة لحذف أي فقرة.

٣. طريقة التجزئة النصفية: تم تقدير ثبات الاختبار بطريقة Spearman-Brown للأطوال المتساوية وقد بلغ (٠.٩٧٢)، كما تم حسابه بطريقة Guttman للتجزئة النصفية وبلغ (٠.٩٦٨)، وهي قيم مرتفعة تدل على ثبات الاختبار التحصيلي.

٤. حساب معامل ثبات الاختبار بطريقة ليفنجستون([٢]):Livingston: إن ثبات الاختبارات محكية المرجع والتي تطبق مرة واحدة تعتمد إلى حد كبير على الفرق بين متوسط درجات المفحوصين ودرجة القطع للاختبار، وقد تم حساب معامل الثبات حسب معادلة ليفنجستون وباعتبار درجة القطع (٥٠%) وهي درجة شائعة الاستخدام في اختبارات التحصيل، وبلغ معامل الثبات

---

[٢]) معادلة ليفنجستون هي: ر ٢١ ع^{٢} + (س$^{-}$ -د)^{٢} / ع^{٢} + (س$^{-}$ -د)^{٢}

حيث: ر ٢١:معامل كيودر ريتشاردسون (او أي معامل ثبات).

ع^{٢}: التباين.

س$^{-}$: الوسط الحسابي للدرجات.

د : درجة القطع.

(٠.٩١)، وهي قيمة مرتفعة تدل على ثبات الاختبار محكي المرجع.

ويتضح من نتائج التحليلات السابقة وإجراءات الثبات والصدق أن الاختبار التحصيلي محكي المرجع لمقرر الثقافة الاسلامية في جامعة القصيم له مؤشرات صدق وثبات مقبولة وصالحاً للتطبيق في الدراسة الحالية.

قيم الصعوبة والتمييز لفقرات الاختبار للعينة الاستطلاعية:

يوضح الجدول (٢) قيم معاملات الصعوبة والتمييز للفقرات للعينة الاستطلاعية

جدول رقم (٢): قيم معاملات الصعوبة والتمييز للفقرات للعينة الاستطلاعية

المفردة	الصعوبة	التمييز	المفردة	الصعوبة	التمييز	المفردة	الصعوبة	التمييز
1	.61	0.28	25	0.49	0.42	49	0.35	0.44
2	.71	0.36	26	0.66	0.34	50	0.57	0.59
3	.87	0.34	27	0.51	0.12	51	0.87	0.26
4	.90	0.28	28	0.81	0.33	52	0.57	0.59
5	.81	0.33	29	0.40	0.43	53	0.73	0.29
6	.64	0.36	30	0.43	0.54	54	0.50	0.12
7	.54	0.28	31	0.83	0.33	55	0.54	0.29
8	.79	0.33	32	0.35	0.44	56	0.77	0.24
9	.64	0.33	33	0.68	0.34	57	0.49	0.42
10	.78	0.39	34	0.44	0.14	58	0.66	0.34
11	.57	0.59	35	0.86	0.27	59	0.51	0.13
12	.68	0.44	36	0.43	0.54	60	0.81	0.33
13	.57	0.59	37	0.83	0.33	61	0.40	0.43
14	.79	0.33	38	0.35	0.44	62	0.43	0.54
15	.68	0.34	39	0.81	0.33	63	0.61	0.28
16	.70	0.42	40	0.40	0.43	64	0.71	0.36
17	.78	0.39	41	0.43	0.54	65	0.64	0.33
18	.57	0.59	42	0.83	0.33	66	0.78	0.39
19	.87	0.26	43	0.35	0.44	67	0.91	0.17
20	.78	0.39	44	0.43	0.54	68	0.68	0.44
21	.73	0.29	45	0.57	0.59	69	0.57	0.59
22	.50	0.12	46	0.79	0.33	70	0.79	0.33
23	.54	0.29	47	0.68	0.34			
24	.77	0.24	48	0.70	0.42			

يوضح الجدول رقم (٢) أن قيم معاملات الصعوبة لفقرات الاختبار تراوحت بين (٠.٣٥ – ٠.٩١)، ويمكن اعتبار أن جميع قيم الصعوبة مقبولة باستثناء المفردة رقم (٦٧) إذ يمكن اعتبارها سهلة جداً، وتراوحت قيم معاملات التمييز للفقرات بين (٠.١٢ - ٠.٥٩) ، وقد تم اعتبار أدنى قيمة مقبولة لمعامل التمييز هي

(٠.٢٤)، وبالتالي فإن الفقرات التي قل معامل تميزها عن هذا المستوى هي ست فقرات (٢٢، ٢٧، ٣٤،٥٤، ٥٩، ٦٧)
وبالتالي وحسب النظرية التقليدية في القياس فإن هذه الفقرات يمكن حذفها او إعادة صياغتها من جديد، ولكن الباحث قام بحذفها من التطبيق النهائي ليبقى (٦٤) فقرة صالحة للتطبيق.

إجراءات الدراسة لتطبيق الاختبار على العينة النهائية:

بعد انتهاء إجراءات هذه المرحلة (الأولى) مرحلة بناء وتجريب الاختبار على العينة الاستطلاعية، والتأكد من ملاءمة فقرات الاختبار في صورته النهائية للتطبيق يتم الانتقال إلى المرحلة الثانية حيث تتلخص الإجراءات بالخطوات التالية:

- قام الباحث وبالتعاون مع كلية الشريعة في جامعة القصيم بتطبيق الاختبار على العينة الرئيسة والمكونة من (١٠٨٦) طالبا درسوا مقرر الثقافة الإسلامية في كليات جامعة القصيم، وليس من بينهم أي فرد من العينة الاستطلاعية. وقد تمت الإجابة على نماذج إجابة خاصة تصحح بجهاز تصحيح آلي، وبعد تصحيح جميع الأوراق تم تفريغ البيانات بطريقة ثنائية (٠ ، ١) إلى برنامج SPSS حيث تكونت صفحة الادخال من (٦٥) عمودا، يعبر العمود الأول عن تسلسل المفحوصين، والبقية (٦٤) عمود تعبر عن الفقرات ويعاملها البرنامج كمتغيرات، و(١٠٣٥) صفا تمثل الأفراد (بعد حذف البيانات التامة للأفراد وعددهم ٥١ طالباً حصلوا على درجة كاملة). ويتم حفظ الملف بامتداد (DAT) لكي يتم قراءته من قبل برنامج BILOG-MG3.

عرض نتائج الدراسة

أولاً: النتائج المتعلقة بالسؤال الأول: "ما مدى تحقق افتراضات النموذج ثلاثي المعلمة في بيانات الدراسة؟"

أ. افتراض أحادية البعد: الصدق العاملي، للتحقق من الصدق العاملي للاختبار استخدم الباحث التحليل العاملي Factor Analysis باستخدام برنامج SPSS وباستخدام طريقة المكونات الأساسية Principal Components، ويبين الجدول رقم (٣) قيم الجذر الكامن ونسبة التباين المفسر لكل عامل، وكذلك نسبة التباين المفسر التراكمية.

الجدول رقم(٣): نتائج التحليل العاملي للاختبار

رقم العامل	الجذر الكامن	نسبة التباين المفسر	نسبة التباين المفسر التراكمي
١	١٣.١٤٥	٢٠.٥٣٨	٢٠.٥٣٨
٢	٦.٠٨٦	٩.٥٠٩	٣٠.٠٤٨
٣	٥.٠٥٢	٧.٨٩٤	٣٧.٩٤٢
٤	٣.٨١٦	٥.٩٦٢	٤٣.٩٠٤
٥	٣.٤٥٨	٥.٤٠٣	٤٩.٣٠٨
٦	٢.٩٧٦	٤.٦٤٩	٥٣.٩٥٧
٧	٢.٩١٩	٤.٥٦١	٥٨.٥١٨
٨	٢.٥٦٧	٤.٠١١	٦٢.٥٢٩
٩	٢.٤٣٦	٣.٨٠٦	٦٦.٣٣٥
١٠	٢.٢٣٥	٣.٤٩٢	٦٩.٨٢٧
١١	٢.٠٩٦	٣.٢٧٤	٧٣.١٠١
١٢	١.٨٧٩	٢.٩٣٦	٧٦.٠٣٧
١٣	١.٧٤٣	٢.٧٢٣	٧٨.٧٦٠
١٤	١.٦٣٧	٢.٥٥٨	٨١.٣١٨
١٥	١.٥٢٨	٢.٣٨٧	٨٣.٧٠٤
١٦	١.٣٩٨	٢.١٨٤	٨٥.٨٨٩
١٧	١.٣٧٩	٢.١٥٤	٨٨.٠٤٣
١٨	١.٣١٥	٢.٠٥٤	٩٠.٠٩٧
١٩	١.٢٩١	٢.٠١٨	٩٢.١١٤
٢٠	١.١٦٥	١.٨٢٠	٩٣.٩٣٤

ويوضح الجدول رقم (٣) أن التحليل كشف عن وجود (٢٠) عاملاً، وقد فسرت جميع العوامل ما نسبته(٩٣.٩٣٤%) من التباين الكلي، وتبين أن العامل الأول فسر لوحده (٢٠.٥٤%) من التباين، ويشير هامبلتون ,Hambleton (2004) إلى أنه إذا فسر العامل الأول (٢٠%) فأكثر من التباين فيُعد عاملا مسيطراً، وأن الاختبار أحادي البعد، وبلغت قيمة الجذر الكامن الأول (١٣.١٤٥)، وبلغت قيمة الجذر الكامن الأول إلى الجذر الكامن الثاني (٧.٠٩٥)، وهذه المؤشرات تدل على تحقق افتراض أحادية البعد للاختبار ووجود عاملا واحدا سائدا (مسيطرا) وأن الاختبار يقيس سمة واحدة.

ب. افتراض الاستقلال الموضعي للفقرات Local Item Independence: ويقصد به عدم اعتماد استجابة أي فقرة على الاجابة الصحيحة عن فقرة أخرى في الاختبار، ولذلك فقد حرص الباحث وأثناء بناء الاختبار على عدم وجود فقرات تعتمد في اجابتها على فقرات أخرى، وتم

التأكد من ذلك أثناء تحكيم الاختبار وعرضه على مجموعة الخبراء. ومما يؤكد تحقق افتراض الاستقلال الموضعي للفقرات كذلك هو ان الاختبار يحقق أحادية البعد فتحقق أحادية البعد للفقرات دليل على الاستقلال الموضعي للفقرات. وللتحقق من تحقق الاستقلال الموضعي للفقرات كذلك قام الباحث بحساب معاملات الارتباط الثنائية بين الفقرات أي بين كل فقرة والمفردة التي تليها، وأظهر هذا التحليل أن جميع قيم الارتباطات الثنائية بين الفقرات منخفضة جداً ولم يقترب أي معامل ارتباط من الواحد الصحيح، ومن هنا يتبين أنه لا يوجد فقرات متطابقة او تعتمد على بعض في الإجابة، ومما سبق يمكن القول انه قد تحقق افتراض الاستقلال الموضعي لفقرات الاختبار.

ج. افتراض التحرر من السرعة: ويعني عدم اعتماد أداء الأفراد في الاختبار على السرعة في الإجابة، وبالتالي من الضروري أن يكون أداء الطالب يعبر عن القدرة الحقيقية للطالب، أي يبذل الطالب أقصى أداء ممكن للإجابة عن فقرات الاختبار، وفي حال أن يجيب الطالب إجابة خاطئة عن المفردة فلأنه بسبب عدم توافر القدرة أو السمة المقيسة لديه وليس بسبب عدم تمكنه من الوصول إليها بسبب قلة الوقت المتاح، وفي هذا الاختبار لم يكن عامل الزمن له أي تأثير على أداء الطلاب المفحوصين، وتأكد الباحث من ذلك في مرحلة تجريب الاختبار على العينة الاستطلاعية، حيث تم حساب زمن اداء الطلاب على الاختبار حيث كان متوسط زمن أداء الطلاب (٧٥) دقيقة، وقد خُصص زمن (١٢٠) دقيقة لتطبيق الاختبار، وهو وقت كافٍ حيث لم ترد أي ملاحظة من قاعات الاختبار للمطالبة بوقت اطول من ذلك، وبالتالي لم يكن لعامل السرعة أي تأثير على الاطلاق على أداء الطلاب على الاختبار. ومما يؤكد كذلك ان الاختبار هو اختبار قدرة هو جعل الاختبار هو الاختبار النهائي للمقرر حيث يعامله الطلاب بجدية كبيرة لأنه يؤثر بشكل مباشر على نتائج الطلاب، وبالتالي من المتوقع أن يبذل الطلبة أقصى ما لديهم للحصول على أعلى الدرجات.

ثانياً: النتائج المتعلقة بالسؤال الثاني " ما مدى التوافق في تقدير معالم فقرات الاختبار المحكي المرجع وفق كل من النموذج الثلاثي المعلمة والنظرية التقليدية اعتماداً على ما يمكن استخلاصه من نسب اتفاق الأسلوبين في تصنيف الفقرات اعتماداً على معالم الفقرات؟"

قبل الشروع بالتحليل باستخدام برنامج BILOG-MG3 تم استبعاد البيانات التامة والصفرية وقد تم فحص جميع بيانات العينة النهائية للاختبار المكون من (٦٤) فقرة والمفحوصين البالغ عددهم (١٠٨٦) طالباً لحذف البيانات التامة والصفرية، وقد تبين وجود (٥١) طالباً حصلوا على الدرجة النهائية (٦٤) وبالتالي تم حذفهم، ولا يوجد أي طالب حصل على صفر، كما تبين عدم وجود أي فقرة حصلت على الدرجة التامة، وكذلك عدم وجود أي فقرة حصلت على الدرجة الصفرية، وبالتالي لم يتم حذف أي فقرة.

بعد ذلك تم إدخال بيانات استجابات العينة النهائية (١٠٣٥) طالبا عن فقرات الاختبار النهائي المكون من (٦٤) فقرة على برنامج BILOG-MG3 وتم استخراج النتائج وفق المراحل التالية:

- **النتائج المتعلقة بملاءمة الأفراد والفقرات للنموذج الثلاثي المعلمة:**

أظهرت نتائج تحليل المرحلة الأولى على البيانات ومن خلال إحصائي مربع كاي(χ^2) likelihood-ratio chi-squared عند مستوى الدلالة الإحصائية (٠.٠١) عدم مطابقة استجابات (٥٧) طالباً لتوقعات النموذج الثلاثي المعلمة، حيث كان مقدار الاحتمالية Fit probability لبعضهم أقل من (٠.٠١)، وكان مقدار الخطأ في تقدير القدرة لبعضهم الآخر كبيرا جدا، لذا تم حذف بياناتهم والابقاء على بيانات (٩٧٨) طالباً.

ويفسر الباحث أسباب عدم تطابق استجابات الطلاب الـ(٥٧) إلى احتمال ميل بعض الطلاب من ذوي القدرات المتدنية إلى التخمين في الإجابة عن الإجابة الأسئلة، وخاصة تلك الأسئلة التي تفوق قيم معلم الصعوبة لها عن مستوى القدرة لديهم، وفيما يتعلق بالطلاب من ذوي القدرات العالية فقد يعود عدم مطابقتهم لتوقعات النموذج الثلاثي المعلمة إلى إخفاقهم في التعامل مع الفقرات التي دون مستوى قدراتهم، فقد أشار هامبلتون وسواميناثان (Hambleton & Swaminathan,1985) بأن الأفراد ذوي القدرات العالية جداً عادة ما يخفقون في الإجابة عن الفقرات السهلة، وذلك لأنهم ربما يفكرون بأبعد مما تعنيه المفردة، أو ربما يميلون إلى تفسير مدلولات المفردة والبدائل بأكثر مما تحتمل.

وبشكل عام فإن نسبة الأفراد الذين استبعدهم النموذج ثلاثي المعلمة تُعد قليلة (٥٧/ ١٠٣٥) أي ما يقارب (٠.٠٥٥) فقط، وبالتالي فإنه يمكن القول أن غالبية أفراد العينة قد طابقت استجاباتهم توقعات النموذج ثلاثي المعلمة بنسبة (٩٤.٥%).

ج. بعد حذف الطلاب غير الملائمين للنموذج تم إعادة التحليل مرة أخرى باستخدام برنامج BILOG-MG3 لاختبار مدى ملاءمة الفقرات للنموذج ثلاثي المعلمة، وقد أظهرت نتائج التحليل في المرة الثانية ومن خلال إحصائي مربع كاي(χ^2) عند مستوى الدلالة الإحصائية (٠.٠٥) عدم مطابقة تسع فقرات للنموذج هي(٣، ٤، ٢٨، ٣٢، ٣٥، ٣٨، ٤٢، ٥١، ٦٠)، حيث كانت قيمة احتمالية المطابقة Fit probability أقل من (٠.٠٥).

ومن خلال ما سبق يتبين لنا أن النظرية التقليدية للقياس أفرزت لنا (٦٤) فقرة مناسبة للتطبيق من أصل (٧٠) فقرة أي ما نسبته (٩١%) من فقرات الاختبار، أما النموذج ثلاثي المعلمة فقد استبعد تسع فقرات غير ملائمة من أصل (٦٤) فقرة ليتبقى (٥٥) فقرة صالحة للتطبيق، تشكل الفقرات النهائية للاختبار محكي المرجع أي ما نسبته (٨٦%) من الفقرات التي أخضعت بياناتها للتحليل.
ومن ذلك نستنتج أن النموذج ثلاثي المعلمة يُعد أكثر صرامة في انتقاء الفقرات، خاصة وأنه أضاف معلم مهم وهو معلم التخمين. وتتفق هذه النتيجة مع نتائج دراسة كيلكار ووايتمان وليشت (Kelkar ; Wightman & Leucht, 2000) في أن إضافة بارامترات إضافية للنموذج تُحسّن مطابقته للبيانات الناتجة عن الاستجابة لفقرات الاختبار، خاصة إذا لم تكن هناك عوامل غير القدرة للأفراد تؤثر على استجاباتهم مثل الغش أو عدم الاهتمام أو اللامبالاة، وهذه العوامل -وحسب اعتقاد الباحث- لم يكن لها أي تأثير في إجراءات هذا الاختبار، حيث تم إجراء الاختبار في ظروف مناسبة من حيث التنظيم والضبط والمراقبة، كونه أحد الاختبارات الرسمية للجامعة، ويفترض أن الطلاب بذلوا أقصى درجة من الجهد للإجابة عن الأسئلة بشكل صحيح كون نتائجه مرتبطة بنجاح أو رسوب كل منهم في المقرر المذكور.

وبشكل عام نلاحظ اتفاق عالي بين النظرية التقليدية والنموذج ثلاثي المعلمة في تقدير الصعوبة والتمييز للفقرات وهذه النتيجة تتفق مع دراسة جمحاوي (٢٠٠٠) حيث أشارت نتائج دراستها إلى وجود اتفاق عالي بين النظرية التقليدية ونماذج الاستجابة للمفردة وخاصة النموذج الثنائي المعلمة في تقدير الصعوبة والتمييز للفقرات، وأشارت كذلك إلى أن خصائص الفقرات المنسجمة مع النموذج تزداد بزيادة عدد معالمه.

د. مرحلة ثالثة من التحليل أجريت للبيانات بعد حذف الفقرات غير الملائمة للنموذج حيث تبقى (٥٥) فقرة تلائم النموذج الثلاثي المعلمة، وفي هذا التحليل الأخير تم الحصول على تقديرات نهائية لكل من معالم الفقرات (الصعوبة، والتمييز، والتخمين)، ومعلم القدرة للطلاب، والخطأ المعياري للتقدير بطريقة الأرجحية العظمى الهامشية (Maximum likelihood method) وهي طريقة تعطي تقديرات عالية الدقة من خلال إعادة متعاقبة لعمليات التقدير (Kyung, 2012 ; Cuervo & Andrade, 2004 ; Hambleton, & Swaminatha, 1985).

ويبين الجدول رقم (٤) قيم معالم الفقرات: الصعوبة والتمييز والتخمين والخطأ المعياري لتقدير كل معلم، بعد حذف الفقرات غير المطابقة للنموذج الثلاثي المعلمة:

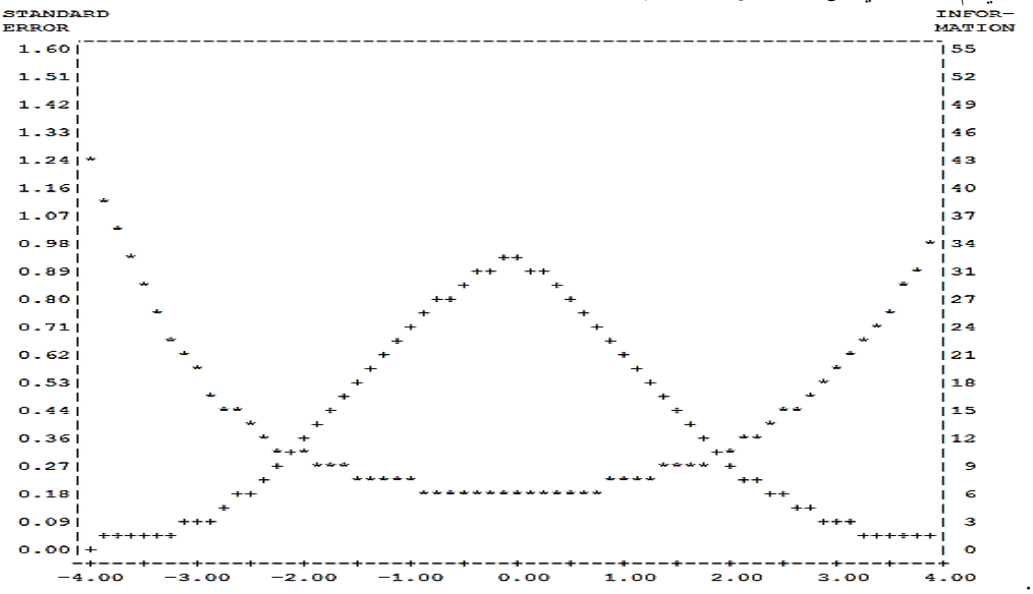
الجدول (٤): قيم تقديرات معالم المفردة الصعوبة والتمييز والتخمين والخطأ المعياري للتقدير لفقرات الاختبار بصورته النهائية حسب النموذج ثلاثي المعلمة

المفردة i	الصعوبة b_i	الخطأ المعياري لتقدير b_i	التمييز a_i	الخطأ المعياري لتقدير a_i	التخمين c_i	الخطأ المعياري لتقدير c_i	المفردة i	الصعوبة b_i	الخطأ المعياري لتقدير b_i	التمييز a_i	الخطأ المعياري لتقدير a_i	التخمين c_i	الخطأ المعياري لتقدير c_i
1	-0.419	0.136	1.064	0.291	0.217	0.091	33	1.564	0.365	1.888	0.201	0.204	0.052
2	0.338	0.312	1.110	0.241	0.123	0.088	34	1.216	0.220	1.392	0.253	0.214	0.035
5	1.138	0.066	1.312	0.242	0.148	0.093	36	-1.265	0.182	1.348	0.176	0.150	0.057
6	-0.364	0.108	1.494	0.110	0.225	0.039	37	1.568	0.226	1.740	0.247	0.142	0.041
7	-0.572	0.227	1.045	0.087	0.271	0.089	39	-1.931	0.314	1.398	0.082	0.204	0.085
8	0.835	0.127	1.212	0.091	0.176	0.098	40	-0.834	0.305	1.797	0.276	0.160	0.071
9	-0.348	0.171	1.204	0.119	0.342	0.087	41	1.047	0.057	1.944	0.098	0.142	0.096
10	0.745	0.148	1.537	0.102	0.241	0.093	43	0.155	0.309	1.269	0.111	0.294	0.074
11	-0.561	0.106	2.037	0.113	0.136	0.017	44	0.280	0.228	1.477	0.243	0.112	0.041
12	0.026	0.102	1.850	0.082	0.118	0.092	45	-1.432	0.103	1.650	0.101	0.147	0.025
13	-0.536	0.111	2.056	0.102	0.148	0.021	46	-0.461	0.117	1.848	0.288	0.135	0.041
14	0.945	0.124	1.248	0.091	0.180	0.039	47	1.772	0.199	1.891	0.189	0.144	0.054
15	0.080	0.281	1.444	0.109	0.135	0.042	48	-0.454	0.106	1.012	0.097	0.257	0.014
16	0.277	0.031	1.608	0.112	0.233	0.091	49	0.534	0.365	1.970	0.109	0.181	0.026
17	0.747	0.045	1.546	0.109	0.189	0.091	50	0.549	0.351	1.148	0.096	0.117	0.032
18	-0.491	0.314	2.064	0.201	0.212	0.054	52	-0.732	0.096	0.398	0.192	0.281	0.066
19	1.873	0.136	0.736	0.296	0.254	0.042	53	0.009	0.174	1.578	0.112	0.189	0.096
20	0.755	0.370	1.568	0.121	0.174	0.054	54	1.278	0.125	1.429	0.102	0.354	0.071
21	0.498	0.068	1.122	0.176	0.145	0.048	55	-0.949	0.103	1.370	0.090	0.191	0.055
22	-0.564	0.160	1.104	0.177	0.145	0.059	56	-0.762	0.322	1.770	0.102	0.201	0.014
23	0.548	0.108	1.338	0.242	0.165	0.050	57	-0.384	0.097	0.970	0.196	0.145	0.035
24	-0.761	0.305	1.570	0.246	0.201	0.064	58	0.338	0.200	1.064	0.287	0.247	0.053
25	-0.221	0.123	1.417	0.242	0.188	0.058	59	0.778	0.283	0.531	0.091	0.175	0.040
26	1.207	0.055	1.337	0.197	0.189	0.075	61	0.158	0.102	0.568	0.219	0.218	0.041
27	-1.301	0.097	1.670	0.136	0.118	0.081	62	-0.450	0.210	0.875	0.102	0.185	0.049
29	1.534	0.169	1.386	0.212	0.203	0.091	63	1.055	0.088	0.575	0.113	0.314	0.061
30	-2.021	0.285	1.780	0.291	0.251	0.092	64	1.564	0.121	1.291	0.082	0.147	0.048
31	0.102	0.313	1.475	0.241	0.295	0.064	المتوسط	**0.140**		**1.391**		**0.194**	

يوضح الجدول رقم (٤) أن قيم التقديرات لمَعلم الصعوبة قدر تراوحت بين (-٢.٠٢١ - ١.٨٧٣)، وبمتوسط حسابي مقداره (٠.١٤)، وتراوحت قيم الخطأ المعياري لتقدير معلم الصعوبة بين (٠.٠٣١ – ٠.٣٧)، أما قيم تقديرات معلم التمييز فتراوحت بين (٠.٣٩٨ – ٢.٠٦٤)، وبمتوسط حسابي مقداره (١.٣٩١)، وتراوحت قيم الخطأ المعياري لتقدير معلم التمييز بين (٠.٢٩١ – ٠.٢٩٦)، وقد تراوحت قيم التقديرات لمعلم التخمين بين(٠.١١٢ – ٠.٣٥٤)، وبمتوسط حسابي مقداره (٠.١٩٤)، وتراوحت قيم الخطأ المعياري لتقدير معلم التخمين بين(٠.٠١٤ – ٠.٠٩٨). وبالنظر إلى قيم تقديرات الصعوبة للفقرات يلاحظ اتساع المدى لهذه التقديرات، وهذا الاتساع للمدى من خصائص النموذج اللوجستي ثلاثي المعلمة، كما يتبين كذلك ارتفاع قيم التمييز للفقرات، وفي نفس الوقت يلاحظ انخفاضاً لقيم معلم التخمين، وقد أشار هامبلتون وسوامينيثان Hambleton, & Swaminathan,

(1985) إلى أنه عندما تكون قيم التقديرات لمعلم الصعوبة بين (-٢ – ٢)، وقيم معلم التمييز بين (٠.٠ -٢)، وقيم معلم التخمين بين (٠.٠ – ٠.٢٥) فإن المعالم تكون متسقة، وبالنظر إلى النتائج المتعلقة بقيم الصعوبة للفقرات في الجدول رقم (٤) نجد أن فقرة واحدة فقط تجاوزت المدى المناسب الذي اقترحه هامبلتون وسواميثان (Hambleton & Swaminathan,1985) لقيم الصعوبة وهي المفردة رقم (٣٠) حيث بلغت قيمة الصعوبة لها (-٢.٠٢١)، بينما بقية الفقرات فجاءت قيم الصعوبة لها ضمن المدى (-٢ - ٢)، أما قيم معلم التمييز فقد جاءت ثلاث فقرات خارج المدى مع أنها قريبة جداً من الحد الأعلى له وهي الفقرات (١١، ١٣، ١٨) وقيم التمييز لها على الترتيب (٢.٠٣٧، ٢.٠٥٦، ٢.٠٦٤)، أما قيم التخمين للفقرات فقد جاءت ثمان فقرات خارج المدى هي الفقرات (٧، ٩، ٣١، ٤٨، ٥٢، ٥٤، ٦٣) وقد تراوحت قيم التخمين لها بين (٠.٢٥٧ – ٠.٣٥٤)، ونلاحظ أن النموذج الثلاثي لم يستبعد أي من هذه الفقرات بل اعتبرها مناسبة خاصة وأن قيم الخطأ المعياري في تقدير أي من المعالم الثلاثة كانت قيم منخفضة ومقبولة.

لقد تم تقدير قيم معلم القدرة للطلاب المفحوصين في المرحلة الأخيرة من التحليل باستخدام طريقة الأرجحية العظمى الهامشية (Maximum likelihood method) الذي يوفرها برنامج بيلوج (BILOG-MG3)،وأظهرت نتائج التحليل أن قيم القدرة قد تراوحت بين (-٤ - ٤)، وأظهرت النتائج كذلك أن قيم الخطأ المعياري تتناقص كلما ارتفعت قيم القدرة، وبحساب دالة المعلومات للاختبار Test Information Function (TIF) بصورته النهائية المكون من (٥٥) فقرة وهي من المؤشرات التي يستدل منها على معامل ثبات الاختبار في نظرية الاستجابة للمفردة، وبالتالي فإن زيادة كمية المعلومات يؤدي إلى نقصان الخطأ المعياري للقياس، ويوضح الشكل رقم (١) دالة معلومات الاختبار والخطأ المعياري للقياس لتوزيع القدرة للطلاب المفحوصين



شكل(١): دالة معلومات الاختبار والخطأ المعياري للقياس لتوزيع القدرة للطلاب المفحوصين

يوضح الشكل رقم (١) أن أكبر قيمة لدالة المعلومات بلغت (٣٢) عند مستوى (٠.٠) من القدرة، وهي تقابل أقل قيمة للخطأ المعياري (٠.١٨) عند مستوى القدرة بين (-١.٠ - ٠.٩)، وتعد فقرات المقياس الأكثر دقة في هذا المدى من القدرة، وبذلك فهي تعطي أقصى قيمة للمعلومات في هذا المدى لارتباط الخطأ المعياري ارتباطاً عكسياً مع دالة المعلومات، أما أقل قيمة لدالة المعلومات فقد بلغت (٢) عند مستوى القدرة (٤، -٤).

ويمكن تفسير هذا المستوى من كمية المعلومات التي وفرها الاختبار إلى ارتفاع قيم معاملات التمييز للفقرات، وهذا مؤشر على أن الاختبار متوسط الصعوبة بشكل عام، وهذا يتفق ما أشار له هامبلتون و وسواميثان (Hambleton & Swaminathan,1985) إلى أن الاختبار السهل من المتوقع أن يقدم معلومات أكبر عند مستويات القدرة المنخفضة، بينما الاختبار الصعب فمن المتوقع ان يقدم معلومات أكبر عند مستويات القدرة العالية. وتتفق هذه النتيجة مع ما توصلت إليه دراسة بني عطا والشريفين(٢٠١٢) عند مقارنة كميات المعلومات لمقاييس عند مستويات مختلفة من القدرة أشارت نتائج دراستها أن أكبر كمية للمعلومات قدمها المقياس المعاير من التوزيع الملتوي نحو اليمين وكانت عند مستويات القدرة العالية، بينما كانت كمية المعلومات التي يسهم بها المقياس عندما تمت معايرته من التوزيع الملتوي نحو اليسار هي الأقصى عند مستويات القدرة المنخفضة .

النتائج المتعلقة بسؤال الدراسة الثالث: " ما مدى التوافق بين النظرية التقليدية في القياس باستخدام (Spearman-Brown; Livingston ; Alfa ; Guttman) والنموذج الثلاثي المعلمة (التجريبي) في تقدير قيمة معامل الثبات للاختبار ؟"

ولأن مفهوم ثبات الاختبار في نظرية الاستجابة للمفردة يرتبط بدالة المعلومات للمفردة وللاختبار فقد تم إيجاد معامل الثبات التجريبي المنبثق من نظرية الاستجابة للمفردة، فقد بين ريفي (Reeve, 2004 المشار له في بني عطا والشريفين، ٢٠١٢) أن دالة المعلومات للاختبار ترتبط مع ثبات الاختبار من خلال العلاقة التالية: $r = 1 - \frac{1}{1\sum_{i=I}^{I} I(\theta)}$ ،

أما معاملات ثبات الاختبار وفق النظرية التقليدية للقياس فقد تم حسابه كما سبق تفصيله في الحديث عن ثبات الاختبار، والجدول رقم (٥) التالي قيم الثبات للاختبار وفق النظريتين:

الجدول رقم (٥): قيم معاملات الثبات التقليدية والتجريبية للاختبار

معاملات الثبات للاختبار حسب النظرية التقليدية		معامل الثبات التجريبي حسب النموذج ثلاثي المعلمة
Alfa	0.922	0.950
Guttman Split-half /	0.968	
Split- half / Spearman-Brown	0.972	
Livingston	0.910	

يوضح الجدول رقم (٥) لقيم معاملات الثبات للاختبار حسب النظريتين (التقليدية، ونظرية الاستجابة للمفردة) أن جميع قيم معاملات الثبات كانت مرتفعة ومتقاربة وقد تراوحت قيمها حسب النظرية التقليدية بين (٠.٩١ - ٠.٩٧٢)، أما قيمته حسب النموذج ثلاثي المعلمة فقد بلغت (٠.٩٥)، وهي قيم مرتفعة وتُعد مؤشراً على ثبات الاختبار.

ومن هذه النتيجة نستنتج أن كلا النظريتين (التقليدية، ونظرية الاستجابة للمفردة (النموذج ثلاثي المعلمة)) متوافقتين من حيث قيم معاملات الثبات للاختبار.

وبشكل عام ومن خلال جميع النتائج السابقة يمكن القول أن جميع الفقرات الـ(٥٥) التي تم مطابقتها للنموذج في صورتها النهائية مناسبة لتشكل اختباراً مناسباً محكي المرجع، حيث المحك هو الأهداف السلوكية للمقرر لقياس تحصيل الطلاب في مقرر الثقافة الإسلامية، ويُعد طول الاختبار الناتج مناسب إلى حدٍ ما ويقترب من الطول المثالي لاختبارات التحصيل للمرحلة الجامعية (٦٠) فقرة الذي اقترحه دراسجو (Drasgow, 1982)، وأشار كذلك العديد من الباحثين والمختصين في مجال القياس والتقويم (Crocker & Algina, 1986 ؛علام،٢٠٠٥ ؛ عودة،٢٠٠٥ ؛ بني عطا والشريفين، ٢٠١٢) إلى أن الطول المناسب للاختبار التحصيلي لطلاب المدارس والجامعات يتراوح بين (٥٠ – ٦٠) فقرة، فهذا الطول للاختبار يمكن أن يغطي المجال السلوكي بشكل مقبول، ومناسب كذلك للوقت المتاح للاختبار فلا يتأثر بعامل السرعة.

لقد كان الهدف من هذه الدراسة بناء اختبار تحصيلي صادق وثابت وموثوق، ومناسب من حيث الطول ليغطي المجال السلوكي للمقرر المقصود، وكانت النتيجة الحصول على اختبار مكون من (٥٥) فقرة جميعها تم التأكد من صلاحيتها ومناسبتها لقياس تحصيل الطلاب في مقرر الثقافة الإسلامية لمرحلة البكالوريوس في جامعة القصيم، وهذه الفقرات حققت جميع افتراضات النموذج الثلاثي والذي يعد من أنسب النماذج لبناء الاختبارات، وكذلك جاءت مطابقة للمؤشرات الإحصائية التي تفترضها النظرية التقليدية في القياس. وأن الفقرات التي تم استبعادها من خلال إجراءات النظرية التقليدية في القياس، والنموذج الثلاثي المعلمة وهو أحد نماذج نظرية الاستجابة للمفردة وكان عددها (١٥) يمكن استبعادها أو مراجعتها وإجراء التعديلات المناسبة على صياغتها لتعديل قيم الصعوبة والتمييز لها، أو استبدالها بفقرات أخرى تقيس نفس الأهداف السلوكية، وقد أشار علام (١٩٨٥) إلى أنه ليس بالضرورة استبعاد عدد قليل من الفقرات إذا أردنا قياس قدرة الطلاب قياساً موضوعيا، وإنما يعني الاعتماد عليها فقط يؤدي إلى تقديرات غير دقيقة للقدرة المقيسة، أي زيادة الأخطاء المعيارية لهذه التقديرات، في حين أنه يمكن اختبار الطالب بأي عدد من الفقرات المتبقية للحصول على تقدير دقيق لقدرته.

وعلى الرغم من التفوق الملحوظ لنماذج السمات الكامنة وخاصة النموذج ثلاثي المعلم على النظرية التقليدية في القياس في بناء الاختبارات التحصيلية ؛ إلا أن هناك مشكلة لا يمكن تجاهلها وهي أن هذه النماذج تحتاج إلى متخصصين يعرفون افتراضاتها، ولديهم القدرة على استخدام البرمجيات المختلفة الخاصة بها، ولديهم معرفة بالاختبارات الاحصائية المناسبة لاختبار أي من الافتراضات التي تقوم عليها نظرية السمات الكامنة، وكذلك قدرة على قراءة النتائج والتوصل إلى قرارات في ضوئها، كما أن هذه النماذج لنظرية السمات الكامنة تحتاج إلى عينات كبيرة جداً للتجريب مقارنة مع النظرية التقليدية، ولذلك فإن نماذج السمات الكامنة قد لا تناسب الاختبارات الصفية التي يقوم بها المدرسون ذات العدد القليل من الفقرات أو المفحوصين، وفي هذه الحالة يوكل أمر تطوير الاختبارات وبناء بنوك الأسئلة إلى متخصصين في بناء وتطوير الاختبارات لأغراض مختلفة وفق نماذج هذه النظرية ويكون دور المدرسين استخدام هذه الاختبارات الموثوقة.

التوصيات:

في ضوء النتائج التي توصلت إليها الدراسة الحالية يوصي الباحث بما يلي:

١. استخدام الاختبار محكي المرجع الذي تم بنائه في صيغته النهائية المكون من (٥٥) فقرة لقياس التحصيل في مقرر الثقافة الإسلامية بجامعة القصيم نظراً لما يتمتع به الاختبار من صدق وثبات ولما تتمتع به جميع فقراته من مستويات مناسبة للصعوبة والتمييز .

٢. استخدام فقرات الاختبار في بناء بنك أسئلة للمقرر، أو لاستخدامها كاختبارات فرعية لقياس التحصيل في مقرر الثقافة الإسلامية.

٣. استخدام نماذج السمات الكامنة وخاصة النموذج ثلاثي المعلمة في بناء ومعايرة الاختبارات وبنوك الأسئلة، لما يتمتع به النموذج من دقة كبيرة في انتقاء الفقرات .

إجراء مزيد من الدراسات على نماذج السمات الكامنة لتوفير أدب نظري كاف باللغة العربية في المكتبات والدوريات العربية لتشجيع الباحثين على استخدام نماذجها والاستفادة من ميزاتها في بناء بنوك الأسئلة حيث أن افتراضات نماذج نظيرة السمات الكامنة ما زالت بحاجة لإثباتها بشكل عملي ومن خلال دراسات تجريبية على عينات حقيقية من الفقرات والأفراد وفي بيئات مختلفة ومستويات من القدرة مختلفة كذلك.

المراجع

أولاً: المراجع العربية.

أبو خليفة، ابتسام توفيق. (٢٠٠٩). توظيف نموذج راش في انتقاء فقرات مقياس تقدير لتقييم
اتجاهات طلبة كلية العلوم التربوية نحو تخصص معلم الصف. **مجلة العلوم التربوية**
والنفسية،ج١٠،ع٤. الأونروا – الاردن.

بني عطا، زايد، والشريفين، نضال.(٢٠١٢). أثر اختلاف شكل توزيع القدرة على معالم الفقرة
ودالة المعلومات للاختبار. **المجلة الأردنية في العلوم التربوية .مجلد ٨،عدد ٢، (١٥١-**
١٦٦).

جمحاوي، إيناس محمود (٢٠٠٠). **مقارنة خصائص الفقرات وفق النظرية التقليدية ونظرية**
استجابة المفردة في مقياس القدرة الرياضية، رسالة ماجستير غير منشورة، جامعة اليرموك، الاردن.

الحموري، هند عبد المجيد (٢٠١١). دراسة استكشافية لملاءمة نماذج نظرية الاستجابة للمفردة
في بناء اختبار تحصيلي من إعداد المعلمة. **مجلة العلوم التربوية والنفسية، مجلد١٢، ع٢ الجامعة الهاشمية- الاردن.**

عبابنة، عماد. (٢٠٠٤). **أثر حجم العينة وطريقة انتقائها وعدد الفقرات وطريقة انتقائها على**
دقة تقدير معالم المفردة والقدرة لاختبار قدرة عقلية باستخدام نظرية الاستجابة للمفردة.
رسالة دكتوراه، جامعة عمان العربية للدراسات العليا، عمّان.

عرفان، أسماء عبد المنعم أحمد.(٢٠٠٩). دراسة سيكومترية حول قياس دافعية الإنجاز
باستخدام نظرية الاستجابة للمفردة. **مجلة البحث العلمي في التربية،ج٢ ع١٠**، جامعة عين
شمس، مصر.

علام صلاح الدين محمود (١٩٨٥). تحليل بيانات الاختبارات العقلية باستخدام نموذج راش
اللوغرتمي الاحتمالي (دراسة تجريبية) .**المجلة العربية للعلوم النفسية، المجلد (٥) عدد**
(١٧) ص. ص.(١٠٠- ١٢١).

علام، صلاح الدين محمود (١٩٩٠). استخدام نموذج راش في بناء مقياس هدفي المرجع
للمعارف الأساسية في إعداد خطة البحوث النفسية والتربوية، **مجلة كلية التربية، جامعة**
الأزهر،١٧، ص.ص.(١٢٩-١٨٢) القاهرة.

علام ، صلاح الدين محمود .(٢٠٠٠) **.القياس والتقويم التربوي والنفس ي أساسياته**
وتطبيقاته وتوجهاته المعاصرة.القاهرة:دار الفكر العربي.

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

عودة، أحمد (٢٠٠٥)**.القياس والتقويم في العملية التدريسية، ط. (٣)،** إربد: دار الأمل،
الأردن.

عيد، غادة خالد.(٢٠٠٤) الدرجة الحقيقية المقدرة باستخدام نظرية السمات الكامنة والنظرية
التقليدية. "دراسة سيكومترية" **مجلة جامعة أم القرى التربوية والاجتماعية والإنسانية.**
مجلد١٦-عدد٢ص.ص.(٢٢٩- ٢٨٤).

كاظم، أمينة.(١٩٨٨). استخدام نموذج راش في بناء اختبار تحصيلي في علم النفس وتحقيق
التفسير الموضوعي للنتائج. **مؤسسة الكويت للتقدم العلمي ع١٩: جامعة الكويت.**

مراد، صلاح أحمد. (١٩٩٦) **اتجاهات معاصرة في القياس النفسي.** كلية التربية: جامعة
الكويت.

الناغي، هبه إبراهيم مُحَمَّد علي(٢٠١١). أثر عدد مفردات الاختبار على تقدير الأفراد ودالة
المعلومات لاختبار تحصيلي مرجعي المحك في الرياضيات باستخدام نماذج نظرية
الاستجابة للمفردة. **مجلة كلية التربية جامعة بور سعيد ،ع١٠،** ص.ص. (٦٧٦ – ٧٢١).

النجار، نبيل جمعة؛ عيلبوني، سمير فؤاد.(٢٠٠٨).بناء بنك أسئلة في الثقافة الحاسوبية للمرحلة
الثانوية في الأردن باستخدام نظرية استجابة المفردة "نموذج راش" والنظرية التقليدية.
مجلة كلية التربية- جامعة عين شمس،ج١،ع٣٢، ص.ص.(٤٩٣ – ٥٢٧).

هادي، فوزية؛ مراد، صلاح.(٢٠١٣). استخدام نموذج راش في إعداد وتقنين اختبار في اللغة
العربية لقبول طلاب الدراسات العليا بكليات العلوم الإنسانية بجامعة الكويت. **المجلة
التربوية ج١،ع١٠٦**- الكويت.

المراجع الأجنبية

Crocker, L. & Algina, J. (1986). **Introduction to classical and modern test theory.** New York: Holt, Rinehart and Winston, Inc.

Cuervo, E. C. & Andrade, J. M (2004). Modeling Abilities in 3-IRT Models. **Revista Columbiana de Statistical, Volume 27 No 1. Pages. 27 – 41**.

Drasgow, F.(1982). Biased test items and differential validity. **Psychological Bulletin, 92, 526-531.**

Garre. G. & Vermunt. K. (2006). Avoiding Boundary Estimation in Latent Class Analysis by Bayesian Posterior Mode Estimation. **Behavior Mentrika, Vol. 33. No. 1.**

Hambleton, R. K. (1989). **Principles and Selected Application of Item Response Theory. In Linn, R. T.(Ed.), Educational Measurement.(3rd) ed**. New York: Macmillan Publishing Company.

Hambelton, R. & Rogers, H. J. (1991). **Fundamentals of item response theory**. SAGE publications, Newbury Park: The international professional publishers.

Hambleton, R. (2004). **Personal communication. Umeå University.** Umeå: Department of Educational Measurement.

Hambleton, R. K, & Cook. L. L. (1983). Robustness of Item Response Models and Effects of Test Length and Sample Size in the Precision of Ability Estimates. New York. In D. J. Weiss (Ed), **New Horizons in Testing**. Pp.31-49.

Hambleton, R. K. & Swaminathan, H. (1985). **Item Response Theory: Principles and Applications**, Boston, Kluwer, Nijh off Publishing.

Huang, C. Y, Lohss, W. E, Lin, C. Shin, D. (2001). Item Calibration of Licensure Test with Multiple Specialty Components. Submitted to Division DI: **Educational Measurements, Psychometrics and Assessment**. Enabled Tiger, a web-based Manuscripts Processing System. Michigan State University.

Kelkar, V., Wightman, L. & Leucht, R. (2000). Evaluation of the IRT Parameter Invariance Property for the MCAT. Paper presented at the annual meeting of the **National Council on Measurement in Education,** New Orleans, LA, April 25 - 27.

Pelton, W. (2002). **The Accuracy of Unidimensional Measurement Models in the Presence of Deviations form the Underlying Assumptions**. Unpublished Doctoral Dissertation. Brigham Young University. USA.

Tian, Song (2009). Investigating Different Item Response Models in Equating the Examination for the Certificate of Proficiency in English (ECPE). **English Language Institute Vol 7.** PP.(85–98): University of Michigan

Wagner, A., & Harvey, R. (2003). **Developing a new critical thinking test using item response theory.** Paper presented at the Annual conference of the Society for Industrial Organizational Psychology. Orlando. www.criticalthinking.org.

Kim, S. (2001). An Evaluation of A Markov Chain Monte Carlo Method for the Rasch Model. **Applied Psychological Measurement**, Vol. 25, No. 2, 163-176.

Kyung T. Han.(2012). Fixing the c Parameter in the Three-Parameter Logistic Model. **Practical Assessment, Research & Evaluation. Vol. 17, Num. 1, ISSN:1531-7714, P.1-24.**

Lee, G., (2004). Comparison of dichotomous and polychromous item response models in equating scores from tests composed of testes. **Applied Psychological Measurement, 25(4), (357-372).**

Mislevy,R.; Bock R.(1990). **Bilog 3 Item analysis and test Scoring With Binary Logistic Models, computer Program** Mooresvilly in Scientific Software Inc.

The compatibility between the three-parameter model and the classical theory in the construction of criterion-referenced test of Islamic culture for the students of the University of Qassim

Dr. Ahmad Ali Al Shriem

Measurement and Evaluation - College of Sharia University of Qassim Dr. Ali

Abstract

This study aimed at verifying the compatibility between the three-parameter model as one of the features underlying the latent response theory and the classical theory in the selection of the items of a criterion-referenced test of the Islamic culture for the students of the University of Qassim. The final study sample consisted of (1086) students, a criterion-referenced test of (70) items of the type of multiple choice was built. Results of students has been analyzed using the(SPSS) program for the classical theory, and (BILOG-Mg3) program for the three parameter model, The results showed: Checking model the assumptions of the three- parameter model in the study data, and three-parameter model was more stringent than the classical theory in the selection of items. The classical theory has ruled out six items of the original number (70). The discrimination coefficients values of items were less than the distinguish it from the acceptable level, whereas the three-parameter model ruled out nine items (64), items did not match the model, and the number of items that fit both form three parameter and the classical theory was (55) items, where the values of its parameters (difficulty, discrimination and guessing) are acceptable within criterions reported by the literature of educational measurement, which constitute the final number of criterion-referenced test, it turns out that there is a broad consensus between the classical theory and the three-parameter model in estimating the reliability coefficient of the test.

Key words: Three-parameter model, Item response theory, Criterion-referenced test, Classical theory of measurement.