استخدام نظرية الاستجابة للفقرة في بناء اختبار محكي المرجع في اللغة الانجليزية (1) لطلبة جامعة القدس المفتوحة، وفق نموذج (راش) *

د. عبد الهادي وجيه صبّاح **

* تاريخ التسليم: 2016/8/9م، تاريخ القبول: 2016/9/25م.
** أستاذ مشارك/ جامعة القدس المفتوحة/ فلسطين.

ملخص:

هدفت هذه الدراسة إلى استخدام نظرية الاستجابة للفقرة في بناء اختبار محكي المرجع في اللغة الانجليزية(1) لطلبة جامعة القدس المفتوحة، وفق نموذج) راش(، ولتحقيق هذا الهدف تم بناء اختبار تحصيلي في اللغة الانجليزية(1) مؤلفاً من (45) فقرة من نوع الاختيار من متعدد من أربعة بدائل، يقيس تحصيل الطالب في اللغة الانجليزية(1)، تم حذف (5) فقرات أجمع المحكمون على أن هناك فقرات تؤدي غرضها، فأصبح الاختبار بصورته الأولية مكوناً من (40) فقرة. طبق على أفراد عينة الدراسة البالغ عددهم (956) طالباً وطالبة، من طلبة السنة الأولى في جامعة القدس المفتوحة.

ومن أبرز النتائج التي توصلت إليها الدراسة: تحقق افتراضات نظرية الاستجابة للفقرة في بيانات الدراسة، وكذلك مطابقة الاستجابات عن (38) فقرة من فقرات الاختبار لافتراضات نموذج (راش)، وحذف فقرتين، لم تطابق النموذج، وحذف (63) فرداً غير مطابقين للنموذج، كما بينت النتائج المتعلقة بتقديرات معامل الصعوبة أنها كانت مقبولة ضمن المحكات التي أوردتها أدبيات القياس التربوي.

وأسفر التحليل عن اختبار جيد يتمتع بمظاهر متعددة من الصدق، وبثبات (0.99) لفقرات الاختبار، (0.80) لقدرات الأفراد. كما بينت النتائج أن قيمة دالة المعلومات كانت أقصى ما يمكن عندما كانت (Θ=b) تقريباً، وذلك كما هو متوقع من نموذج (راش). وأن الاختبار يقدم أكبر كمية من المعلومات عند مستويات القدرة المتوسطة، ويقدم أقل كمية من المعلومات عند مستويات القدرة العالية والمتدنية.

الكلمات المفتاحية: بناء الاختبارات، الاختبارات محكية المرجع، نظرية الاستجابة للفقرة، نموذج راش، اللغة الانجليزية(1).

Using Item Response Theory for Constructing a Criterion- Referenced Test in English Language (1) for Al-Quds Open University Students According to Rasch Model

Abstract:

The present study aimed at using the item response theory for constructing a criterion-referenced test in English Language (1) for Al-Quds Open University (QOU) Students according to Rasch Model. To achieve this objective, a test of 45-multiple-choice items with four choices was constructed to measure students' achievement in English Language (1). Later and upon the referees' recommendations, five items were excluded because they were found to be similar to other items which reduced the number of items to 40. The 40- item test was administered among a sample of 956 male and female freshmen students at QOU. The most significant results were shown in verifying the assumptions of item response theory in this study in addition to having 38 items that were in consistent with Rasch Model while 2 items were excluded as they found to be inconsistent with Rasch Model. Furthermore, 63 individuals were excluded from the sample because they were found inconsistent with the model. Additionally item difficulty of the test was found to be acceptable and consistent with the criteria given in the review of literature. The analysis also produced a good test that has different forms of validity in addition to (0.99) reliability of items and (0.80) reliability of individuals. The results also revealed that the maximum value of information was achieved at (Θ=b) which agrees with Rasch Model and that the test offers the maximum information at medium ability levels while the minimum information was at both lowest and highest ability levels.

Keywords: *Test construction, Criterion Referenced Test (CRT), Item Response Theory (IRT), Rasch model, English Language (1).*

مقدمة:

وجه العديد من علماء القياس بعض الانتقادات إلى الاختبارات معيارية المرجع، التي ارتبط ظهورها بفلسفة التربية التي كانت سائدة في الولايات المتحدة الأمريكية في أوائل القرن العشرين، وكان هدفها تصنيف الأفراد حسب مراكزهم النسبية بين أقرانهم في القدرات المختلفة، واهتمت الاختبارات المعيارية المرجع بإبراز الفروق الفردية بين الأفراد بغض النظر عن مدى إتقان الفرد للمهارات والمعلومات المراد قياسها، وكان لهذه الانتقادات الدور الأكبر في ظهور الاختبارات محكية المرجع، والتي لم يعد هدفها مقتصراً على مجرد التمييز بين الطلبة في القدرة أو المستوى، وإنما تركز على اكتسابهم مهارات معينة، وتحقيق أهداف محددة، بل وإتقان تلك المهارات وتلك المعلومات، كما وتهدف الاختبارات محكية المرجع إلى مقارنة أداء الفرد بمستوى أداء مطلق ومحدد في مجال من السلوك أو بمستوى كفاية معين، دون الحاجة إلى مقارنة أداء الفرد بأداء الآخرين، كما هو الحال في الاختبارات معيارية المرجع (علام، 2001). وبهذا فإن الاختبار المحكي المرجع يعتمد في جوهره على تقدير مستوى أداء الفرد بالنسبة لمحك أو مستوى أداء أو درجة قطع (Popham, 1978). فعندما يود المعلم تحقيق مستوى إتقان معين لدى طلبته فينبغي عليه أن يجيب كل منهم إجابة صحيحة عن نسبة مئوية من فقرات الاختبار، يتم تحديدها مسبقاً وتسمى بدرجة القطع؛ وهي الدرجة التي ينبغي أن يحصل عليها الفرد في المجال الشامل لفقرات الاختبار، لكي يكون متقناً لمحتوى أو مهارة معينة. وأحياناً تشير درجة القطع إلى الحد الأدنى للأداء المقبول، لكي يتمكن الفرد من أداء مهام تالية (Guskey، 1988).

وقد اهتم العاملون في القياس والتقويم التربوي بعملية تصنيف الطلبة، حسب مستوى التمكن من الأهداف التدريسية. إذ إن مسألة تصنيف الطلبة تساعد المربين على اتخاذ قرارات تعليمية مهمة، تمس حاضر الطلبة ومستقبلهم، ويتطلب تصنيف الطلبة كفايات ومهارات محددة يجب أن يتقنها المعلم، أو يتمكن منها، ويحللها، ويعيد صياغتها في صورة أهداف تعليمية إجرائية قابلة للقياس، ويقدم هذه الأهداف إلى الطالب قبل بدء العملية التعليمية، ويقدر مدى تحققها باستخدام ما يسمى بالاختبارات محكية المرجع (Criterion Referenced Tests) التي يُعتمد على نتائجها في تصنيف الطلبة إلى مجموعتين: إحداهما متمكنة(Masters)، والأخرى غير متمكنة(Hambleton,1994).

وفي معظم الاختبارات كان يُعتمد على النظرية الكلاسيكية في بناء الاختبارات وتقنينها،والتي تعرضت لمجموعة من الانتقادات وفق ما أشار هامبلتون وسوامينثان(Hambleton & Swaminathan, 1985)، بالإضافة لمعاناتها عيوب وجوانب ضعف محددة (Stocking, 1999). ومن أهم تلك الجوانب أن الخصائص السيكومترية للاختبار تحدد بالنسبة لمجموعة من الأفراد، كما إن خصائص الأفراد، تحدد بالنسبة إلى اختبار معين، وليس بشكل مطلق. وتفترض أيضاً أن الخطأ المعياري في القياس متساوٍ لكل الأفراد وهذا يفتقر إلى الدقة؛ زيادة على إنها لا تبين مدى تحصيل الفرد على مستوى الفقرة، وإنما على الاختبار ككل (Hambleton Swaminathan & Rogers،1999). لذلك،جاءت فكرة استخدام نماذج نظرية الاستجابة للفقرة في بناء الاختبارات، والتي يُفترض فيها تلافي عيوب النظرية الكلاسيكية، كما إنها تعد ثمرة محاولات المهتمين في القياس النفسي والتربوي لتطوير مقاييس أكثر دقة في قياسها للسمات النفسية والتربوية، وكمحاولة لحل المشكلات التي تعانيها النظرية الكلاسيكية (عودة، 1992).

وقد اعتمدت هذه الدراسة في بناء الاختبار وتقدير خصائصه السيكومترية على النظرية الحديثة في القياس، أو ما يعرف بنظرية السمات الكامنة، أو نظرية الاستجابة للفقرة (Item Response Theory, IRT) بوصفها تشكل إطاراً للتوجه الحالي والمستقبل المنظور في اختيار الفقرات (Anastasi, 1982)، وكذلك على اعتبار أنها نظرية بديلة للنظرية الكلاسيكية.

ومن أهم مميزات نماذج نظرية الاستجابة للفقرة أن استخدامها يجعل خصائص الفقرة (الصعوبة، والتمييز، والتخمين) مستقلة عن عينة الأفراد المستخدمة في تقدير هذه الخصائص، وكذلك تجعل تقديرات قدرات الأفراد مستقلة عن فقرات الاختبار المستخدمة في الحصول على هذه التقديرات، وكذلك تحسين دقة النتائج وثباتها من خلال تحديد وحذف الأفراد والفقرات غير المطابقة للنموذج المستخدم (Hambleton, 1988; Lord & Stocking, 1989).

ومن أهم نماذج نظرية الاستجابة للفقرة، نموذج المعلم الواحد، أو ما يسمى بنموذج راش، بوصفه يحقق خاصية الموضوعية في القياس النفسي والتربوي (Wright & Stone, 1979)، والذي ارتبط باسم عالم الرياضيات الدنمركي جورج راش (George Rasch)، الذي اهتم بإرساء دعائم الموضوعية في سلوك الإنسان وقياس الظواهر السلوكية، حيث تبنى الأساس الاحتمالي في تفسير الأداء. بمعنى أن يحصل الفرد على العلامة نفسها في كل الاختبارات التي تقيس السمة نفسها مهما اختلفت صعوبة فقرات هذه الاختبارات، مما يعني استقلال القياسات عن الفقرات المستخدمة في القياس، وأن تكون أدوات القياس مستقلة عن خصائص العينة التي يطبق عليها الاختبار (علام، 1985)، ويهتم النموذج كذلك بتحديد موقع الفقرة على ميزان صعوبة جميع فقرات الاختبار، وتدريج مستويات القدرة للفرد على اختبار معين على نفس ميزان تعيير الفقرات (علام، 2006).

ويفترض نموذج راش إن زيادة قدرة الفرد عن صعوبة الفقرة يؤدي إلى زيادة في احتمال حدوث الاستجابة الصحيحة. لكن العلاقة بين قدرة الفرد وصعوبة الفقرة ليست علاقة قاطعة، فقد تكون قدرة الفرد أكبر من صعوبة الفقرة لكنه يجيب إجابة خاطئة وقد تزداد صعوبة الفقرة عن قدرة الفرد لكنه يجيب إجابة صحيحة. ولذلك توصل راش إلى أن الفرق بين القدرة والصعوبة هو الذي يحدد احتمال الإجابة الصحيحة (كاظم، 1988).

وتتمثل الصيغة الرياضية لهذا النموذج في المعادلة التالية:

$$P_i(0) = \frac{e^{D(e-b_i)}}{1+e^{D(e-b_i)}} \qquad \textbf{(i=1,2,3,.....n)}$$

حيث ترمز P_i (Θ): إلى احتمال الإجابة الصحيحة لفرد قدرته (Θ) عن الفقرة (i،e) إلى الأساس اللوغاريتمي الطبيعي وهو يساوي (2.718) تقريباً، D إلى معامل القياس أو التدريج وهو مقدار ثابت يساوي (1.7) ويستخدم هذا العامل ليؤكد أن هذه الدالة اللوغاريتمية هي تقريب حقيقي للدالة التجميعية الطبيعية لهذا النموذج، b_i إلى معلم الصعوبة للفقرة i(Hambleton & Swaminathan, 1985).

وحتى تتوافر متطلبات الموضوعية في القياس يجب أن تستوفى افتراضات نموذج راش (McCamey, 2002; Kim & Hong, 2004)ومنها:

- **أحادية البعد**: إن أحادية البعد تعني أن يُعرف المتغير (السمة) بوساطة مجموعة من الفقرات ذات الصعوبة فقط، وأن يكون للأفراد قدرة أحادية البعد تحدد وحدها أداءهم على الاختبار (الشرقاوي وآخرون، 1996)، أي تتدرج فقرات الاختبار بحيث تقيس متغيراً واحداً وتشتمل هذه الفقرات على التعريف الإجرائي للمتغير موضوع القياس، وكما يقضي هذا الشرط أيضاً أن يتدرج الأفراد على متصل هذا المتغير وفقاً لمستويات أدائهم في الإجابة على فقرات الاختبار بحيث يكون لهؤلاء الأفراد قدرة أحادية البعد (منسي، 1998).
- **الاستقلال الموضعي**: ويُقصد به إن استجابات الفرد للفقرات المختلفة مستقلة استقلالاً إحصائياً، وهذا يعني أن أداء الفرد على الفقرة لا يتأثر بأدائه على الفقرات الأخرى في الاختبار، ولا يتأثر أيضاً بأداء الأفراد الآخرين الذين يؤدون هذا الاختبار، أي أن تقدير معالم الفقرة لا يعتمد على تقديرات معالم الفقرات المكونة للاختبار، ولا يعتمد على تقديرات قدرات الأفراد الذين يجيبون عن هذه الفقرات، وكذلك تقدير القدرة (السمة) للفرد لا

يعتمد على قدرة الأفراد الآخرين الذين يؤدون الاختبار، ولا على تقديرات معالم الفقرات التي يؤدونها، أي يُتوقع أن نحصل على نفس تقدير القدرة وذلك من خلال أي مجموعة جزئية من الفقرات طالما أنها ملائمة، وكذلك من خلال أي عينة من الأفراد، يُتوقع أن تكون تقديرات معالم الفقرة نفسها، وبالتالي فإن الاستقلال الموضعي يتطلب أن يكون احتمال الإجابة الصحيحة للفرد على فقرة ما مستقلاً عن أي فقرة أخرى عند مستوى معين من القدرة، والعوامل المؤثرة في احتمال استجابة الفرد على فقرة معينة هي فقط السمة الكامنة التي تقيسها الفقرة، ومعالم الفقرة التي يجيب عليها الفرد (Hambleton, 1987).

ويذكر (كيوننجهام، 1986) إن افتراض الاستقلالية يعني ببساطة أن إحدى الفقرات لا يكون لها تأثير على احتمالية توصل الأفراد إلى الإجابة الصحيحة عن الفقرات الأخرى، ويتحقق افتراض الاستقلال الموضعي إذا كانت مجموعة الفقرات تقيس قدرة أو سمة واحدة (Cunningham,1986).

- **خطية القياس**: وتعني أن يكون هناك معدل ثابت لتدريج المقاييس، حيث يتدرج كل من الأفراد والفقرات على نفس المتصل تدريجاً خطياً بوحدة قياس واحدة.

- **منحنى خصائص الفقرة**: أي الحد الذي تميز فيه الفقرات بين الأفراد ذوي المستويات المختلفة في القدرة، وجميع هذه الفقرات يجب أن يكون لها قيم تمييز متساوية،ويمكن وصف العلاقة بين تحصيل الفرد على الفقرة والسمة أو القدرة التي تقيسها مجموعة فقرات، أو يقيسها اختبار يحوي هذه الفقرة، من خلال اقتران تراكمي صاعد يعرف باسم منحنى خصائص الفقرة. إذ يوفر هذا المنحنى احتمالات إجابة الأفراد عن الفقرة في مستويات القدرة المختلفة إجابة صحيحة. وكون المنحنى تراكمياً صاعداً، يشير بوضوح إلى أن احتمال إجابة الفقرة إجابة صحيحة يزداد بازدياد قدرة الفرد.

- **انعدام أثر التخمين**: حيث يفترض أنه لا يوجد فرد يخمن الإجابة الصحيحة، ويتفق هذا مع أحادية البعد.

- افتراض عامل السرعة: تفترض نظرية الاستجابة للفقرة أن عامل السرعة لا يؤدي دوراً في الإجابة عن فقرات الاختبار، بمعنى إن إخفاق الأفراد في الإجابة عن فقرات الاختبار يرجع إلى انخفاض قدرتهم وليس إلى تأثير عامل السرعة في إجاباتهم، ويمكن تقدير ما إذا كان عامل السرعة قد أدى دوراً في الإجابة عن طريق معرفة عدد المفحوصين الذين لم يتمكنوا من الإجابة عن جميع فقرات الاختبار الذي أجري عليهم، وعندما تكون السرعة أحد العوامل المؤثرة في الأداء على الاختبار، فإن هناك قدرتين تؤثران في هذا الأداء وهما القدرة التي يقيسها الاختبار، والقدرة الخاصة بسرعة الأداء، وهذا انتهاك لافتراض أحادية البعد (Hambleton & Swaminathan, 1989).

ويشترط لاستخدام نموذج (راش) توفر أداة مصممة لقياس سمة واحدة، ببيانات ثنائية التدريج، أي من نوع الاختيار من متعدد (Bowles, 2005)، وعندما يستجيب الفرد لفقرة في اختبار ما، فإنه يحدث تفاعل بين قدرة هذا الفرد وصعوبة تلك الفقرة، ومن خلال هذا التفاعل يتضح الجانب النفسي في نموذج راش في اعتماده على أساسين منطقيين، وهما: احتمال أن يجيب الفرد إجابة صحيحة عن فقرة سهلة أكبر من احتمال أن يجيب إجابة صحيحة عن فقرة صعبة، ويزداد احتمال الإجابة الصحيحة للفرد عن فقرة ما بزيادة مستوى قدرته، ويكون هذا الاحتمال محكوم بالفرق بين قدرة الفرد وصعوبة تلك الفقرة (Gyagenda & Engelhard, 1998; Eyres et al.,2005).

ويتطلب بناء الاختبارات محكية المرجع إعداد خطة تفصيلية دقيقة قبل البدء الفعلي في كتابة الفقرات التي يشتمل عليها الاختبار، **ولبناء الاختبار محكي المرجع لا بد من إتباع مجموعة من الخطوات كما حددها** (علام، **2006**):

1. تحديد النطاق السلوكي الذي يقيسه الاختبار وتحليله إلى مكوناته من معارف ومهارات:حيث يُعد تحديد النطاق السلوكي بمثابة الإطار المرجعي الذي تنسب إليه علامة الفرد في الاختبار.
2. صياغة الأهداف السلوكية المتعلقة بنواتج تحليل النطاق السلوكي: حيث ينبغي صياغة المهارات الناتجة عن تحليل النطاق السلوكي صياغة سلوكية ليمكن قياسها بفقرات اختبارية، وتشير هذه الأهداف إلى المحتوى المرجعي الذي سوف يتعامل معه الفرد أثناء اختباره.
3. تكوين المواصفات التفصيلية للاختبار: حيث يتم تحديد مواصفات الاختبار المحكي المرجع بشكل مفصل ودقيق.
4. كتابة فقرات الاختبار، وتتطلب كتابة فقرات الاختبار مهارة كبيرة وتمكن باني الاختبار من محتوى النطاق السلوكي الذي يقيسه الاختبار.
5. تجميع فقرات الاختبار:حيث يتم تنسيق الفقرات في ضوء الأهداف التي تقيسها لتكوين الاختبار، وهذا يتطلب مراعاة التسلسل المنطقي للفقرات وتجميع الفقرات التي تكون من نوع واحد معاً.
6. تجريب الاختبار ميدانياً وتحليل فقراته: يتم تطبيق الاختبار على عينة من الأفراد للتعرف على مدى ملاءَمة فقراته للمختبرين ودرجة تمييزها بين الأفراد الذين حققوا هدفاً معيناً والأفراد الذين لم يتمكنوا من تحقيقه، وذلك لتقييم صدق الفقرات في قياس النطاق السلوكي المتعلق بالأهداف المحددة.
7. التحقق من صدق الاختبار وثبات درجاته: ويتم التحقق من ذلك بالطرق الخاصة بالاختبارات محكية المرجع، فمفهوم الصدق في الاختبارات محكية المرجع يتعلق بمنطقية تفسير الدرجات وما يترتب عليها من قرارات صائبة، أما مفهوم الثبات فيتعلق باتساق الدرجات.

صدق الاختبارات محكية المرجع:

يرى عدد من المختصين بتمييز أنواع الصدق في الاختبارات محكية المرجع عنها في الاختبارات معيارية المرجع فيذكرون الصدق الوصفي كبديل عن صدق المحتوى، والصدق الوظيفي كبديل عن الصدق التجريبي، وصدق انتقاء المجال السلوكي كبديل عن صدق البناء (عبابنة، 2009).

أما في نظرية الاستجابة للفقرة فتتحقق موضوعية تفسير نتائج الاختبار وموضوعية القياس من خلال تحقق افتراضات نموذج راش على بيانات الاختبار بصورته النهائية، وذلك بعد حذف الأفراد والفقرات غير المطابقين لافتراضات النموذج، ويعد التحقق من افتراضات النموذج تقويماً لمدى صدق النموذج في تحقيق موضوعية القياس (Stenner, 1990)، إن استخدام نموذج راش في بناء اختبار ما يدل على توفر متطلبات الموضوعية في قياس متغير القدرة موضوع القياس، مما يعني التحقق من توفر شرطي الصدق والثبات لتقديرات كل من قدرات الأفراد وفقرات الاختبار (كاظم، 1988)، وإن حذف الأفراد الذين لا تتطابق استجاباتهم مع افتراضات النموذج والإبقاء على الأفراد الذين تتطابق استجاباتهم مع افتراضات النموذج (أي الذين تتسق استجاباتهم الملاحظة على الفقرة واحتمال نجاحهم فيها)، يعني وجود اتساق بين استجاباتهم على الفقرة ودرجاتهم الكلية على الاختبار، مما يشير إلى الاتفاق بين السمة التي تقيسها الفقرة والسمة التي تقيسها باقي فقرات الاختبار، وكذلك حذف الفقرات غير المطابقة لافتراضات النموذج يدل على أن كل فقرة مستبقاة في الاختبار تتفق في تعريفها للمتغير موضوع القياس مع ذلك الذي تعرفه باقي الفقرات في الاختبار، أي أن هناك اتفاقاً بين ما تقيسه الفقرة وما تقيسه بقية فقرات الاختبار، ويُعد هذا مؤشراً لصدق القياس (عوض الله، 2010).

ثبات الاختبارات محكية المرجع:

يعتمد ثبات الاختبارات محكية المرجع على خصائص وقدرات عينة المفحوصين بالإضافة إلى اعتماده على شروط تطبيق الاختبار، وهناك عدة طرق لتقدير الثبات، ولكل طريقة أساليبها الخاصة في تقدير الثبات، ويمكن عرض طرق تقدير الثبات من خلال الصيغ التالية:

ثبات الاتساق في قرارات الإتقان (التصنيف): ومنها معامل (كابا) الذي يحسب من خلال المعادلة التالية:

$k = \frac{P_o - P_e}{1 - P_e}$ حيث K: معامل كابا، P_o: مستوى الاتفاق في التصنيف المشاهد، P_e مستوى الاتفاق المتوقع.

ثبات تقدير علامة المجال: إن إجراءات تقدير ثبات علامة المجال تستلزم أن يتم تعريف المجال بصورة واضحة وأن تكون فقرات الاختبار ممثلة عشوائياً للمجال، إن التقدير الأفضل لعلامة المفحوص على المجال يكون من خلال ايجاد فترة ثقة بحيث يكون حدها الأعلى نسبة الإجابة الصحيحة مضافاً لها خطأ معياري واحد، وحدها الأدنى نسبة الإجابة الصحيحة مطروحاً منها خطأ معياري واحد.والمعادلة التالية تبين ذلك:

$$\hat{P} - S_e < \overline{P} < \hat{P} + S_e$$

حيث $\hat{P}$: علامة الطالب الملاحظة على الاختبار، $\overline{P}$: تقدير علامة الطالب على كامل المجال، S_e : الخطأ المعياري.

ثبات الدرجات من خلال معادلات (ليفنجستون وبرينان وكين): طور (ليفنجستون، 1972) معادلة لتقدير ثبات الاختبار حسب الصيغة التالية:

$$k^2(X,T) = \frac{\sigma_x^2 (KR-20) + (\mu_x - nc)^2}{\sigma_x^2 + (\mu_x - nc)^2}$$

حيث X: العلامة المشاهدة، T: العلامة الحقيقية، $K^2(X,T)$: معامل ليفنجستون، σ_x^2 : تباين درجات الاختبار، μ_x : متوسط علامات الاختبار، (KR - 20): معامل ثبات كودر ريتشاردسون – 20، n: عدد فقرات الاختبار، C: علامة القطع.

كما طور (برينان وكين، 1977) مؤشراً لتقدير ثبات الاختبارات محكية المرجع حسب المعادلة التالية:

$$M(C) = \frac{\sigma_p^2 + (\mu - c)^2}{\sigma_p^2 + (\mu - c)^2 + (\sigma_i^2 + \sigma_e^2)/n}$$

حيث M(C): معامل برينان وكين، σ_p^2 : تباين درجات المفحوصين على الاختبار، σ_i^2 : تباين صعوبة فقرات الاختبار، σ_e^2 : تباين أخطاء القياس، μ : متوسط علامات الاختبار.

أما في نظرية الاستجابة للفقرة فيتحقق ثبات الاختبار بتحقق افتراض الاستقلالية، أي استقلالية القياس عن فقرات الاختبار المطبق على العينة، واستقلالية القياس عن العينة المطبق عليها الاختبار بعد حذف الفقرات التي لا تحقق استقلالية القياس، وهذا يعتبر مؤشراً على تحقيق ثبات القياس، ويتم حساب معامل ثبات خاص بالأفراد ومعامل ثبات خاص بالفقرات (كاظم، 1988).

استخدام نموذج(راش) في بناء الاختبارات محكية المرجع

إن استخدام نموذج (راش) في بناء اختبار محكي المرجع يعني توافر الموضوعية في قياس القدرة، ويدل هذا على توافر شرطي الصدق والثبات لتقديرات قدرات الأفراد وصعوبة الفقرات. وتتحدد مطالب الموضوعية في تعريف فقرات الاختبار لمتغير واحد، واستقلال تقديرات الأفراد عن الفقرات، واستقلال صعوبة الفقرات عن أفراد العينة (كاظم، 1988).

ويقوم برنامج (Winsteps) المستخدم لتحليل بيانات الاختبار بحذف الأفراد غير المطابقين للنموذج وهم غير الصادقين في استجاباتهم على فقرات الاختبار وفقاً لقدراتهم، وحذف الفقرات غير المطابقة للنموذج وهي التي لا تتسق مع الفقرات الأخرى على المتغير موضوع القياس. وتدريج فقرات الاختبار مستقلة عن الأفراد، وتقدير أداء الفرد مستقلاً عن الفقرات. وينتج عن ذلك تدريج فقرات الاختبار بوحدات متساوية تسمى اللوجيت (الشافعي، 1996).

وقد حظيت عملية بناء الاختبارات باهتمام الباحثين والدارسين، وسيتم في هذه الدراسة عرض بعض الدراسات التي أجريت حول بناء الاختبارات محكية المرجع وتقدير خصائصها السيكومترية وفق نظرية الاستجابة للفقرة:

قام (ياسين، 2004) بدراسة هدفت إلى تقدير الخصائص السيكومترية لاختبار محكي المرجع في الكيمياء للصف الأول الثانوي العلمي وفق النظريتين الكلاسيكية والحديثة في القياس، تكونت عينة الدراسة من (481) طالباً وطالبة، تكون الاختبار من (52) فقرة من نوع الاختيار من متعدد ذي الأربعة بدائل،

تم تقدير الخصائص السيكومترية للاختبار، الصدق والثبات وفق النظرية الكلاسيكية، وتقدير الثبات وفق النظرية الحديثة باستخدام نموذج راش، حيث بلغ معامل الثبات للاختبار (0.99)، ومعامل الثبات للأفراد (0.88)، وتم تقدير قيم معالم الفقرات لكل فقرة من فقرات الاختبار وفق النظريتين الكلاسيكية والحديثة.

وقام (الشريفين، 2006) بدراسة هدفت إلى تقدير الخصائص السيكومترية لاختبار محكي المرجع في القياس والتقويم التربوي وفق النظرية الحديثة في القياس، تكونت عينة الدراسة من (222) طالباً وطالبة من طلاب جامعة آل البيت، تم بناء اختبار تحصيلي في القياس والتقويم التربوي مكون من (50) فقرة من نوع الاختيار من متعدد ذي الأربعة بدائل، تم حساب معاملات الصعوبة والتمييز لكل فقرة، وبلغ معامل الثبات للأفراد (0.90)، ومعامل الثبات للاختبار (0.95)، وتمتع الاختبار بمظاهر متعددة من الصدق.

وقام (حمادنة، 2009) بدراسة هدفت إلى استخدام نظرية الاستجابة للفقرة في بناء اختبار محكي المرجع في الرياضيات، وفق النموذج اللوجستي ثلاثي المعلم، تم بناء اختبار في الرياضيات، يقيس تحصيل الطالب في الإحصاء، تكون من (28) فقرة من نوع الاختيار من متعدد من أربعة بدائل، وتكونت عينة الدراسة من (411) طالباً وطالبة من طلبة الصف الأول الثانوي العلمي، أظهرت النتائج تحقق افتراضات نظرية الاستجابة للفقرة في بيانات الدراسة، وكذلك مطابقة الاستجابات عن (24) فقرة من فقرات الاختبار لتوقعات النموذج اللوجستي ثلاثي المعلم، كما بينت النتائج المتعلقة بتقدير معالم الفقرات (الصعوبة، والتمييز، والتخمين) أنها كانت مقبولة ضمن المحكات التي أوردتها أدبيات القياس التربوي.

وأجرت (علي، 2012) دراسة هدفت إلى بناء اختبار محكي المرجع وفق نموذج راش لمقرر القياس والتقويم في التربية الذي يدرس لطلبة السنة الرابعة لتخصص معلم صف في كلية التربية بجامعة دمشق، تكون الاختبار من (99) فقرة من نوع الاختيار من متعدد من أربعة بدائل، تكونت عينة الدراسة من (426) طالباً وطالبة من طلبة كلية التربية تخصص معلم صف، بينت النتائج أن معامل ثبات المفردات بلغ (0.98)، وبلغ معامل ثبات الأفراد (0.94)، كما بينت النتائج أن تدريج الاختبار المحكي المرجع في القياس والتقويم في التربية بنموذج راش يحقق افتراضات نظرية الاستجابة للفقرة.

كما أجرتا(هادي ومراد، 2013) دراسة هدفت إلى إعداد وتقنين اختبار في اللغة العربية لقبول الطلاب في برنامج الدراسات العليا بكليات العلوم الإنسانية بجامعة الكويت باستخدام نموذج راش، وتم إعداد (130) فقرة تم تجريبها على (86) طالباً وطالبة بالدراسات العليا، أسفر التجريب عن حذف (30) فقرة وتعديل (16). وتم تطبيق الصورة النهائية من الاختبار على عينة مكونة من (335) طالباً وطالبة، وتم تحليل البيانات باستخدام نموذج راش، وأسفرت النتائج عن حذف (19) فقرة غير ملائمة للنموذج وحذف (26) فرداً غير ملائمين للنموذج، وأسفر التحليل عن اختبار جيد يتمتع بثبات (0.81) لفقرات الاختبار، (0.99) لقدرات الأفراد.

كما قام المصري (2015) بدراسة هدفت إلى استخدام نظرية الاستجابة للفقرة في بناء اختبار محكي المرجع في البحث التربوي وفق النموذج اللوجستي أحادي المعلم، تكون الاختبار من (48) فقرة من نوع الاختيار من متعدد من أربعة بدائل، وبلغت عينة الدراسة (220) طالباً وطالبة من طلبة كلية التربية المحويت باليمن، وأظهرت النتائج تحقق افتراضات النموذج أحادي المعلم، ومطابقة (42) فقرة من فقرات الاختبار لتوقعات النموذج كما بينت النتائج أن تقديرات معلمي الصعوبة والتمييز ضمن المحكات المقبولة في أدبيات القياس التربوي، كما أظهرت النتائج حذف (7) أفراد غير مطابقين للنموذج، وتراوحت معالم القدرة لأفراد عينة الدراسة بين (2.659–) إلى (2.812) لوجيت.

وقام (الطراونة، 2016) بدراسة هدفت إلى بناء اختبار محكي المرجع في الإحصاء التحليلي لطلبة الدراسات العليا في كليات العلوم التربوية في الجامعات الأردنية الحكومية باستخدام نظرية الاستجابة للفقرة، تكونت عينة الدراسة من (319) طالباً وطالبة من الماجستير في كليات العلوم التربوية في الجامعات الأردنية. تم التحقق من مطابقة عينة الدراسة وفقرات الاختبار لافتراضات نموذج راش، وتم التحقق من افتراضات نظرية الاستجابة للفقرة، وتم حساب معاملات الصعوبة لفقرات الاختبار باستخدام برنامج (Bigsteps)، وتكون الاختبار بصورته النهائية من (26) فقرة، وأظهرت النتائج تمتع الاختبار بدلالات صدق وثبات مقبولة.

ومن استعراض الدراسات السابقة تبين أن بعضها هدفت إلى بناء اختبارات محكية المرجع في مباحث مختلفة ومساقات مختلفة، على مستوى المدرسة وعلى مستوى الجامعة، وبعضها استخدم النظرية الكلاسيكية في القياس، والبعض الآخر استخدم النظرية الحديثة في القياس، ولم توجد أية دراسة هدفت إلى بناء اختبار تحصيلي محكي المرجع في اللغة الانجليزية لطلبة الجامعة وتقدير خصائصه السيكومترية باستخدام نظرية الاستجابة للفقرة (نموذج راش)، لذلك هدفت هذه الدراسة إلى بناء اختبار محكي المرجع في مادة اللغة الانجليزية(1) لطلبة جامعة القدس المفتوحة والتحقق من خصائصه السيكومترية وفق نظرية الاستجابة للفقرة، نموذج (راش).

مشكلة الدراسة

يعد مقرر اللغة الانجليزية(1) متطلب جامعة إجباري في جامعة القدس المفتوحة، يدرسه الطالب في السنة الدراسية الأولى، ويتضمن عدداً من المهارات والمعارف التي سبق وأن درسها في المرحلة الثانوية، ويُشترط نجاح الطالب في هذا المقرر (حصوله على 50 % فأعلى) ليتسنى له التسجيل للمقرر اللاحق اللغة الانجليزية(2)، ولوحظ أن كثيراً من الطلبة ذوي التحصيل المرتفع يشكون بأن المهارات التي يدرسونها في هذا المقرر لا تضيف إلى معارفهم الشيء الكثير،على الرغم من الوقت الذي يقضيه الطالب في دراسته، ومن أجل الاستغلال الأمثل لوقت الطالب وجهده،فقد أدى ذلك بالمسؤولين في الجامعة إلى التفكير

في وضع الحلول المناسبة لهذه المشكلة،ونظراً لعدم توافر أداة قياس تتمتع بخصائص سيكومترية مقبولة، لقياس تحصيل الطلبة في مقرر اللغة الانجليزية(1)، وللحاجة الملحة لتوافر أداة قياس موضوعية، يمكن من خلالها التحقق من مدى تمكن أو إتقان الطلبة للمهارات والمعارف المحددة في هذا المقرر. وذلك نظراً لأهمية إتقان مثل هذه المهارات في قياس تحصيل الطلبة في الجامعات وغيرها من مؤسسات التعليم، كما أنه يمكن أن يكون هذا الاختبار أداة مناسبة لتشخيص ومعرفة الأهداف التي تحققت، والأهداف التي لم تتحقق، واقتراح الأساليب التعليمية والتدريبية في ضوء نتائج عملية التشخيص.

وبناء على ما سبق كان التفكير في وضع معيار موضوعي لقياس تحصيل الطلبة في اللغة الانجليزية (1)، وهو إعداد اختبار محكي المرجع في اللغة الانجليزية(1) من نوع الاختيار من متعدد من أربعة بدائل، يتمتع بخصائص سيكومترية مقبولة، يشتمل على المهارات والمعارف الأساسية التي يجب أن يتقنها الطالب بعد انتهائه من دراسة هذا المقرر، وتطبيقه على الطلبة المسجلين في المقرر خلال الأسبوع الأول من بدء الفصل الدراسي، يتم تصنيفهم بناء على نتائج الاختبار إلى مجموعتين: الأولى متمكنة وهم الطلبة الذين يجيبون إجابة صحيحة على (50 %) على الأقل من فقرات الاختبار، وأخرى غير متمكنة. وبناء على نتائج الاختبار يعفى الطلبة الذين يتم تصنيفهم إلى متمكنين من دراسة هذا المقرر ويسمح لهم بدراسة المتطلب اللاحق في نفس الفصل الدراسي الذي يتقدم به للاختبار، أما الطلبة غير المتمكنين فيدرسون المقرر كالمعتاد حتى نهاية الفصل الدراسي.

لذا جاء تكليف الباحث بإعداد هذا الاختبار من المجلس الأكاديمي في الجامعة، لدراسة المقترح وبحث إمكانية إنجاز المطلوب خلال فصل دراسي واحد، واستعان الباحث بثلاثة من أساتذة الجامعة من حملة الدكتوراه في اللغة الانجليزية،واللغة الانجليزية وأساليب تدريسها لمساعدته في إعداد الاختبار وفق المواصفات المطلوبة.

وقد سعت هذه الدراسة إلى بناء اختبار محكي المرجع في اللغة الانجليزية(1) يكون صالحاً للتطبيق في جامعة القدس المفتوحة وفي الجامعات الفلسطينية الأخرى التي تدرس المقرر نفسه وذلك بتوفير دلالات عن صدق الاختبار وثباته وفق نظرية الاستجابة للفقرة (نموذج راش).

أسئلة الدراسة

سعت الدراسة للإجابة عن الأسئلة الآتية:

- السؤال الأول: ما مدى مطابقة الاستجابات على فقرات الاختبار لافتراضات نموذج راش؟
- السؤال الثاني: ما قيم تقديرات معلم الصعوبة لفقرات الاختبار بصورته النهائية وفق نموذج راش؟
- السؤال الثالث: ما الخصائص السيكومترية للاختبار (دلالات الصدق والثبات)؟
- السؤال الرابع: ما هي كمية المعلومات التي يقدمها الاختبار عند مستويات القدرة المختلفة؟

أهداف الدراسة

هدفت هذه الدراسة إلى استخدام نظرية الاستجابة للفقرة في إعداد اختبار محكي المرجع في اللغة الانجليزية (1) وفق النموذج اللوجستي أحادي المعلم (نموذج راش)، بحيث يتمتع هذا الاختبار بدلالات صدق وثبات مقبولة، وبمعايير تتفق مع متطلبات نموذج راش؛ من أجل قياس تحصيل طلبة السنة الأولى في مقرر اللغة الانجليزية(1) في جامعة القدس المفتوحة، وتصنيفهم وفق علاماتهم على هذا الاختبار إلى متمكنين (أجابوا إجابة صحيحة على 50 % من الفقرات على الأقل) أو غير متمكنين.

أهمية الدراسة

يُعد تحصيل الطلبة وإكسابهم معارف ومهارات معينة من أولويات المؤسسات التربوية. وقياس التحصيل يُعد أحد المصادر المهمة لتزويد القائمين على العملية التعليمية بالمعلومات الضرورية اللازمة لاتخاذ القرارات المتعلقة بالنجاح والرسوب، وتصنيف الطلبة وفق مستوياتهم التحصيلية، وتؤدي أساليب القياس والتقويم دوراً مهماً في صنع القرارات التربوية، فمن خلال هذه الأساليب يتم التعرف إلى فاعلية البرامج التربوية.

وتتجلى أهمية هذه الدراسة في بيان آلية بناء اختبار محكي المرجع وفق نظرية الاستجابة للفقرة في أحد المقررات التي تدرس في جامعة القدس المفتوحة، والذي يدرس أيضاً في بعض الجامعات الفلسطينية الأخرى، بحيث يتمتع بدرجة معقولة من الصدق والثبات، لاستخدامه لاحقاً، والذي يمكن اعتباره بمثابة نواة لبنك أسئلة للمقرر، وتنبع الأهمية العملية للدراسة من إنها ستسهم في فحص افتراضات نظرية الاستجابة للفقرة، وبالتحديد نموذج راش فيما يتعلق بتوضيح أثر خصائص الفقرات وخصائص الأفراد على دقة تقدير معالم الفقرات والأفراد، فضلاً على إنها توفر اختباراً يقيس المهارات والمعارف في اللغة الانجليزية (1) بعد تدريجه وفق نموذج راش يصلح للاستخدام في جامعة القدس المفتوحة بشكل خاص وفي الجامعات الفلسطينية بشكل عام، بحيث يمكن من خلاله سحب عينات عشوائية من فقرات المجال الكلي للاختبار، وفحص كل فرد بإحدى هذه العينات، وتقدير قدرته على تدريج مشترك، دون الحاجة إلى مقارنة أداء الطالب بأداء زملائه.

تعريف المصطلحات

◄ **نموذج (راش)**: أحد نماذج نظرية الاستجابة للفقرة الذي يفترض تساوي جميع فقرات الاختبار في التمييز بين مستويات القدرة المقاسة وإن الفرد لا يلجأ إلى التخمين العشوائي في إجابته عن فقرات الاختبار (Wright & Stone,1979).

◄ **الخصائص السيكومترية**: يقصد بالخصائص السيكومترية وفق نظرية الاستجابة للفقرة: معالم الصعوبة والتمييز والتخمين للفقرة، ومعاملي الصدق والثبات للاختبار. وفي هذه الدراسة اقتصرت الخصائص السيكومترية على معلم الصعوبة، ومعاملي الصدق والثبات (علام، 2001).

◄ **الاختبار المحكي المرجع**: هو الاختبار الذي يستخدم لتقدير أداء الطالب بالنسبة لمجال سلوكي محدد تحديداً مسبقاً، أو هو الاختبار الذي يقارن فيه أداء الطالب مع مستوى محدد من الإتقان أو مع مستوى تحقيقه لمجموعة من الأهداف.

ويعرف إجرائياً في هذا البحث على أنه مجموعة من الفقرات من نوع الاختيار من متعدد من أربعة بدائل (أحدها صحيح والأخرى خاطئة) أعدت وفق نظرية الاستجابة للفقرة وتحليل نتائج استجابات الطلبة باستخدام نموذج راش بما يحقق دقة وموضوعية قياس تحصيل طلبة السنة الأولى في مقرر اللغة الانجليزية (1).

محددات الدراسة:

- اقتصرت عينة الدراسة على طلبة السنة الأولى في جامعة القدس المفتوحة المسجلين في مقرر اللغة الانجليزية(1) في الفصل الدراسي الثاني من العام الدراسي 2015/ 2016. مما سيحد من تعميم نتائج الدراسة خارج مجتمعها.
- اقتصرت أداة الدراسة على اختبار محكي المرجع في مقرر اللغة الانجليزية(1)، الذي يتم تدريسه في جامعة القدس المفتوحة، وكانت الفقرات من نوع الاختيار من متعدد من أربعة بدائل.
- اقتصرت الدراسة على استخدام نموذج راش، وما تم استخدامه من برمجيات إحصائية متخصصة ومتوفرة.

منهجية الدراسة وإجراءاتها:

منهجية الدراسة:

اعتمد الباحث في بناء أداة هذه الدراسة على خطوات بناء الاختبارات محكية المرجع، وفقاً لافتراضات نموذج (راش).

مجتمع الدراسة:

تكون مجتمع الدراسة من جميع طلبة السنة الأولى في جامعة القدس المفتوحة المسجلين في مقرر اللغة الانجليزية(1) في كافة فروع الجامعة وعددها (18) فرعاً في الفصل الدراسي الثاني من العام الأكاديمي 2015/ 2016 والبالغ عددهم (7000) طالب وطالبة منهم (3000) ذكور و (4000) إناث.

عينة الدراسة:

تكونت عينة الدراسة من (956) طالباً وطالبة أي ما نسبته (13.66 %) من المجتمع، تم اختيارهم بطريقة العينة الطبقية العشوائية، من طلبة السنة الأولى المسجلين في مقرر اللغة الانجليزية(1)، في الفصل الدراسي الثاني من العام الأكاديمي 2015/ 2016، حيث تم اختيار عينة عشوائية من كل فرع بحسب عدد الطلبة في الفرع، وبلغ عدد الذكور في العينة (410) طالب وعدد الإناث (546) طالبة.

أداة الدراسة:

تم بناء الاختبار المحكي المرجع في موضوعات اللغة الانجليزية (1) وفق الخطوات الآتية:

- تم تحديد الغرض من الاختبار، وهو قياس تحصيل الطلبة في موضوعات اللغة الانجليزية(1).
- تم تحليل محتوى كتاب مقرر اللغة الانجليزية(1) في جامعة القدس المفتوحة، لتحديد محتوى الاختبار والأهداف السلوكية المتعلقة به، بمساعدة أساتذة اللغة الانجليزية،إذ تم استخلاص خمسة موضوعات يفترض أن تتحقق عند المتعلم في مستوى الإتقان التام، واشتملت على الموضوعات الآتية: القراءة والاستيعاب بوزن (24 %)، القواعد بوزن (40 %)، المفردات بوزن (16 %)، استخدام اللغة بوزن (10 %)، والإملاء وعلامات الترقيم بوزن (10 %).
- تم صياغة الأهداف السلوكية التي تضمنتها هذه الموضوعات كنتاجات تعلم يفترض أن تتحقق لدى المتعلمين في مستوى الإتقان، وتم مراعاة مستويات بلوم للأهداف المعرفية (30 % تذكر، 40 % فهم واستيعاب، 30 % مهارات تفكير عليا).
- تم إعداد جدول مواصفات للاختبار جرى فيه ربط مستويات الأهداف بمحتوى المقرر موضوع الاختبار.
- تم عرض جدول المواصفات على سبعة من المحكين، خمسة من أعضاء هيئة التدريس الذين يدرسون المقرر في الجامعة والذين يحملون درجة الدكتوراه في اللغة الانجليزية وأساليب تدريس اللغة الانجليزية، واثنان من حملة درجة الدكتوراه في القياس والتقويم، وكان الهدف من الاحتكام إلى المحكمين إعطاء حكم على نتائج تحليل المحتوى والأهداف كما وردت في جدول المواصفات ومدى تغطيتها للمجال السلوكي المحدد في الكتاب المقرر. وقد تم تعديل بعض الأهداف المصاغة بناء على رأي المحكمين واستقر عدد الأهداف السلوكية على (30) هدف.
- قامت اللجنة المكلفة بإعداد الاختبار بكتابة فقرات على كل هدف تم تحديده،واختلف عدد الفقرات حسب حجم المادة المقررة لكل هدف، وبلغ عدد هذه الفقرات (45) فقرة من نوع الاختيار من متعدد من أربعة بدائل، وقد روعي أن تطابق الفقرة الهدف السلوكي الذي تقيسه من حيث المحتوى والمستوى المعرفي، وقد تطلب قياس بعض الأهداف أكثر من فقرة واحدة، كما روعي أن لا تعتمد إجابة إحدى الفقرات على إجابة الفقرات الأخرى، وهو ما يتطلبه نموذج راش كأحد نماذج نظرية الاستجابة للفقرة، وتم الاسترشاد بأسس بناء الاختبارات المحكية المرجع في صياغة هذه الفقرات، من خلال الاستعانة بالدراسات ذات الصلة التي تم ذكرها سابقاً.
- تم التحقق من صدق المحتوى للاختبار من خلال عرضه على نفس المحكين السبعة، إضافة إلى الثلاثة الذين تم الاستعانة بهم في تحليل المحتوى وكتابة فقرات الاختبار،لإبداء رأيهم حول الفقرات والاختبار ككل واقتراح ما يرونه من تعديل مناسب، وطلب من كل منهم أن يبدي رأيه في: درجة الاتفاق بين الأهداف السلوكية والهدف الأساسي للاختبار وتمثيلها له، ومدى التوافق بين الفقرة والهدف الذي يفترض أنها تقيسه، ومدى

وضوح الصياغة اللغوية للفقرة، ومدى شمول الاختبار للمجال السلوكي، ومدى اتساق الاختبار ككل مع الهدف الأساسي منه. وبناء على ملاحظاتهم تم تعديل بعض الفقرات (بعضها تم التعديل على المتن والبعض الآخر تم التعديل على البدائل)، وتم حذف (5) فقرات أجمع المحكمون على أن هناك فقرات تؤدي غرضها.وبعد حذف هذه الفقرات أصبح الاختبار مكوناً من (40) فقرة تمثل المحتوى المطلوب والأهداف السلوكية التي تغطي المجال السلوكي.

التجريب الأولي للاختبار

تم تطبيق الاختبار المكون من (40) فقرة على عينة استطلاعية مكونة من (100) طالب وطالبة، من طلبة السنة الأولى المسجلين في مقرر اللغة الانجليزية(1) في نهاية الفصل الدراسي الأول من العام الأكاديمي 2015/ 2016.وقد انحصر الغرض من التطبيق على: تحديد الزمن الذي يحتاج إليه المفحوص للإجابة على الاختبار، والتأكد من وضوح الصياغة اللغوية للفقرات، وكذلك التحقق من دلالات صدق الاختبار وثباته، وقد تم تحديد الزمن اللازم للاختبار بساعة ونصف.

التحليل الإحصائي لبيانات التجريب الأولي

أجريت عملية التحليل لاستجابات عينة التجريب باستخدام الحزمة الإحصائية (SPSS) حيث تم التعرف إلى دلالات صدق الاختبار وثباته من خلال الإجراءات التالية:

تم التحقق من دلالات صدق الاختبار بطريقتين:

- الطريقة الأولى: تم التحقق من صدق المحتوى بالاعتماد على تحليل المحتوى وبناء جدول المواصفات، وتأكد ذلك من خلال حكم المحكمين من أعضاء هيئة التدريس كما سبق توضيحه في إجراءات إعداد الصورة الأولية للاختبار.
- الطريقة الثانية:تم التحقق من صدق الاختبار من خلال الصدق التمييزي، إذ قام الباحث بترتيب الطلبة تنازلياً حسب علاماتهم الكلية، وتم تقسيم أفراد عينة الدراسة حسب علاماتهم إلى مجموعتين: مجموعة المتمكنين (الناجحين) (وهم الطلبة الذين حصلوا في الاختبار على علامة 20 فأكثر) ومجموعة غير المتمكنين (الراسبين) (وهم الطلبة الذين حصلوا في الاختبار على علامة أقل من 20)، على اعتبار أن (20) هي علامة النجاح، وأن العلامة الكاملة هي (40)، ثم قام الباحث بالمقارنة بين متوسطي علامات المجموعتين باستخدام اختبار (ت) للعينات المستقلة، وكانت النتائج كما هي في الجدول (1) التالي:

جدول (1)

نتائج اختبار (ت) للعينات المستقلة للمقارنة بين متوسطي علامات الطلبة المتمكنين وغير المتمكنين

المجموعة	العدد	المتوسط الحسابي	الانحراف المعياري	ت المحسوبة	درجات الحرية	مستوى الدلالة
المتمكنين	347	24.42	3.91	42.42	954	0.00
غير المتمكنين	609	13.41	3.83			

يلاحظ من نتائج الجدول(1) أن هناك فروقا واضحة بين متوسطي علامات المجموعتين على الاختبار، ولاختبار فيما إذا كانت هذه الفروق ذات دلالة إحصائية تم استخدام اختبار ت للعينات المستقلة، وقد بلغت قيمة (ت) المحسوبة (42.42) وهي قيمة دالة إحصائياً عند مستوى دلالة ($\alpha \leq 0.01$) لأن قيمة مستوى الدلالة المحسوب من العينة قد بلغت (0.00) وهي أصغر من مستوى الدلالة ($\alpha=0.01$). أي إن هناك فروقاً ذات دلالة إحصائية بين المتوسطين لصالح مجموعة المتمكنين،مما يشير إلى صدق الاختبار في التمييز بين الطلبة المتمكنين وغير المتمكنين على الاختبار الحالي، الأمر الذي يؤكد على قدرة الاختبار على تصنيف الطلبة إلى متمكنين (ناجحين) وغير متمكنين (راسبين).

وتم التحقق من دلالات الثبات للاختبار:من خلال حساب معامل ثبات الاتساق الداخلي باستخدام معادلة (كودر– ريتشاردسون) رقم (20) وباستخدام معادلة (ليفنجستون) بدرجة قطع (50 %)،فكانت قيم معاملات الثبات للاختبار على الترتيب (0.857)، (0.868) وهي قيم مقبولة.

إجراءات جمع البيانات

- بعد أن تم إعداد الاختبار المكون من (40) فقرة، ووضع التعليمات المناسبة لتطبيقه، قام الباحث بالتنسيق المسبق مع عمادة القبول والتسجيل والامتحانات وإدارات الفروع التعليمية المسؤولة عن تطبيق الاختبار، للاتفاق على موعد وآلية تطبيق الاختبار، وقد تم الاتفاق على تطبيقه على جميع الطلبة المسجلين في المقرر وعددهم (7000) طالب وطالبة في اليوم الأخير من الأسبوع الأول من الفصل الدراسي الثاني من العام الأكاديمي 2015/ 2016 الساعة الحادية عشرة صباحاً في كافة الفروع التعليمية.

- انتظر الباحث حتى بداية الفصل الدراسي الثاني من العام الأكاديمي 2015/ 2016 حيث تم إرسال أوراق الأسئلة إلى الفروع التعليمية وتم تطبيق الاختبار على جميع أفراد مجتمع الدراسة في نفس اليوم وفي نفس الوقت الذي تم الاتفاق عليه، وطلب الباحث من الجهات المسؤولة عن تطبيق الاختبار في كافة الفروع التعليمية بالطلب من الطلبة قراءة التعليمات وتعبئة البيانات الشخصية قبل البدء في الإجابة، ورغم وضوح تعليمات الاختبار فقد تمت الإجابة عن جميع استفساراتهم من قبل الجهات المسؤولة. وبعد انتهاء مدة الاختبار المحددة في تعليمات الاختبار قامت الجهات المنظمة بجمع أوراق استجابات الطلبة عن الاختبار وإرسالها إلى الباحث.

- قام الباحث باختيار عينة طبقية عشوائية من أوراق إجابات الطلبة من كل فرع حسب نسبة الذكور والإناث بالفرع الدراسي وحسب نسبة عدد الطلاب في الفرع إلى عدد الطلاب الكلي، وقد بلغ عدد الأوراق التي تم اختيارها من كافة الفروع (956) ورقة إجابة (حجم العينة الكلية)، ثم قام بمراجعة أوراق الطلبة– أفراد العينة الكلية– جميعها للكشف عن الاستجابات ذات النمطية الواحدة. وبناء على ذلك لم يستثن الباحث أية ورقة من التحليل.

- تم تصحيح أوراق إجابات العينة على الاختبار بواسطة برنامج تصحيح الكتروني باستخدام مفتاح الإجابة وتفريغ الإجابات في الحاسوب على صورة (0) للإجابة الخاطئة، (1) للإجابة الصحيحة تمهيداً لمعالجتها الإحصائية.

عرض نتائج الدراسة ومناقشتها

أولاً: نتائج السؤال الأول ومناقشتها

◄ نص السؤال الأول على: **ما مدى مطابقة الاستجابات على فقرات الاختبار لافتراضات نموذج راش؟**

للإجابة عن السؤال الأول، حللت استجابات الطلبة على فقرات الاختبار وذلك باستخدام بعض البرامج الإحصائية المناسبة وتشمل (SPSS) و (Winsteps) المصمم لذلك وفقاً لنموذج راش.وفيما يلي خطوات تحليل استجابات أفراد عينة الدراسة البالغة (956) فرداً على فقرات الإختبار (40) فقرة، وهو عدد مناسب وفق ما أشار إليه فان (Van, 1998) من إن الحد الأدنى لعدد الأفراد وفق نموذج راش يجب أن لا يقل عن (200) فرد.ولاستخدام نموذج راش يجب التحقق من افتراضات النموذج، وهي على النحو التالي:

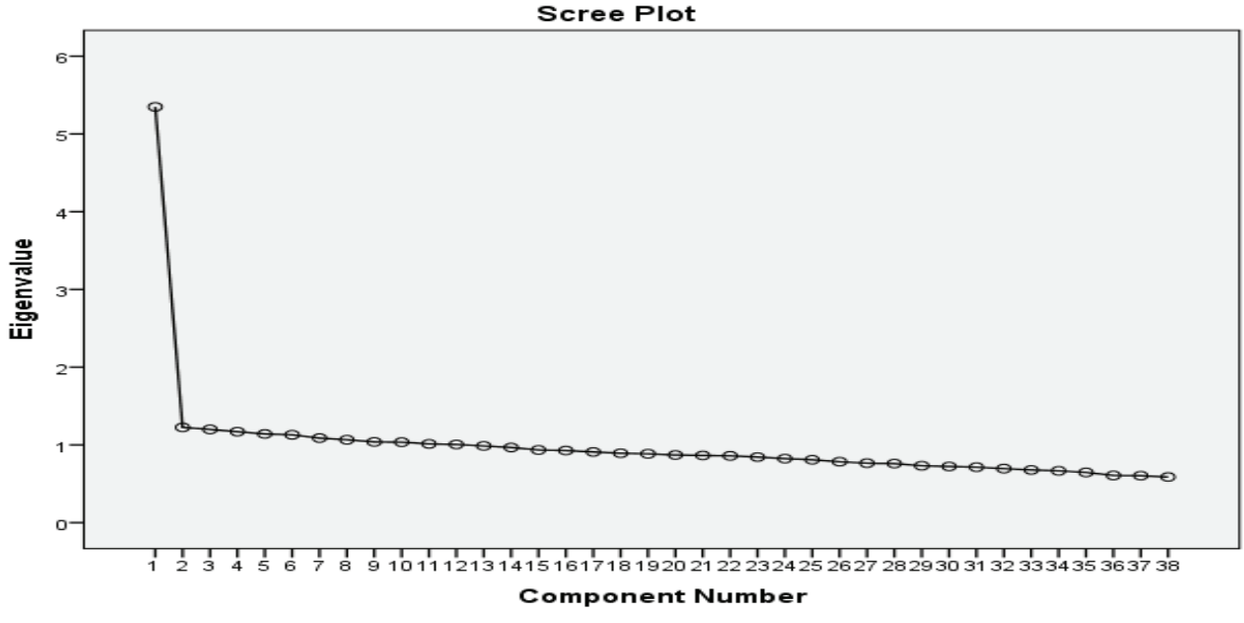
- افتراض أحادية البعد: نظراً لأن أحادية البعد (وجود عامل واحد رئيسي) أحد الشروط الأساسية لنموذج راش وتعني أن فقرات الاختبار تكون متجانسة فيما بينها وتقيس في أساسها نفس السمة، ويمكن التحقق من هذا الافتراض باستخدام التحليل العاملي (Wiberg، 2004)، فقد حسبت مصفوفة الارتباط الرباعي بين فقرات الاختبار وإخضاعها للتحليل العاملي الاستكشافي بطريقة المكونات الأساسية (لهوتلنج) والتدوير المتعامد للمحاور بطريقة (الفاريماكس). إذ أفرز التحليل (12) عاملاً فسرت مجتمعة (45.94 %) من التباين الكلي، وفسر العامل الأول (14.07 %) من التباين. ويبين الجدول(4) هذه العوامل، ونسبة التباين المفسر لكل منها، ونسبة التباين المفسر التراكمية.

جدول (2)

خلاصة نتائج التحليل ألعاملي لاستجابة أفراد عينة الدراسة على فقرات الاختبار

رقم العامل	الجذر الكامن	نسبة التباين المفسر	نسبة التباين المفسر التراكمية	رقم العامل	الجذر الكامن	نسبة التباين المفسر	نسبة التباين المفسر التراكمية
1	5.35	14.07	14.07	7	1.09	2.86	32.36
2	1.23	3.22	17.29	8	1.07	2.81	35.17
3	1.2	3.16	20.45	9	1.04	2.73	37.90
4	1.17	3.08	23.53	10	1.03	2.72	40.62
5	1.14	3.00	26.53	11	1.01	2.67	43.29
6	1.13	2.97	29.5	12	1.0	2.65	45.94

يلاحظ من الجدول (2)، أن قيمة الجذر الكامن للعامل الأول (5.35) وفسر ما نسبته (14.07%) من التباين الكلي، وبلغت نسبة قيمة الجذر الكامن للعامل الأول إلى قيمة الجذر الكامن للعامل الثاني (4.35)، وبما أن هذه النسبة أكبر من (2) فإن هذا مؤشر على إن العامل الأول هو عامل مسيطر مما يؤكد تحقق افتراض أحادية البعد للاختبار، أي إن الاختبار يقيس سمة كامنة واحدة. وقد تشبعت جميع الفقرات على العامل الأول، إذ كانت درجة تشبعها أكثر من (0.30) وهذا دليل آخر على إن الاختبار أحادي البعد. ويتعزز افتراض أحادية البعد من خلال تمثيل الجذور الكامنة بيانياً باستخدام اختبار فحص العوامل (Scree Plot) الذي يظهر في الشكل (1) الآتي:



الشكل (1):

التمثيل البياني لقيم الجذور الكامنة للعوامل المكونة للاختبار على البيانات الكلية

يتبين من الشكل (1) إن العامل الأول يمثل عاملاً سائداً وإن المنحنى قد تغير اتجاهه وانعطف انعطافاً ملحوظاً عند العامل الثاني، مما يرجح افتراض أحادية البعد لأغراض التحليل وفق نظرية نظرية الاستجابة للفقرة.

- افتراض الاستقلال الموضعي: أشار (هامبلتون وسوامناثان) (Hambleton & Swaminathan, 1985) إلى إن افتراض الاستقلال الموضعي يكافئ افتراض أحادية البعد في الاختبار، ويعني هذا أنه إذا تحقق افتراض أحادية البعد في الاختبار فإن الاختبار يحقق افتراض الاستقلال الموضعي.

- افتراض توازي المنحنيات المميزة للفقرات(تساوي معاملات التمييز):للتحقق من توازي المنحنيات المميزة للفقرات تم حساب معاملات الارتباط الثنائي النقطي بين أداء الأفراد على الفقرة وأدائهم على الاختبار ككل، والذي يعتبر مؤشراً لدرجة تمييز الفقرة والجدول (3) يبين نتائج ذلك.

جدول (3)

معاملات تمييز فقرات الاختبار باستخدام معامل الارتباط الثنائي النقطي

رقم الفقرة	معامل التمييز	رقم الفقرة	معامل التمييز	رقم الفقرة	معامل التمييز	رقم الفقرة	معامل التمييز	رقم الفقرة	معامل التمييز
1	0.36	9	0.40	17	0.39	25	0.36	33	0.42
2	0.40	10	0.39	18	0.36	26	0.35	34	0.37
3	0.39	11	0.38	19	0.39	27	0.35	35	0.38
4	0.30	12	0.39	20	0.37	28	0.36	36	0.38
5	0.33	13	0.35	21	0.35	29	0.40	37	0.36
6	0.43	14	0.38	22	0.33	30	0.41	38	0.43
7	0.38	15	0.44	23	0.40	31	0.37	39	0.38
8	0.41	16	0.43	24	0.33	32	0.38	40	0.38

يتبين من الجدول (3) أن معاملات التمييز لفقرات الاختبار تراوحت بين (0.30) وبين (0.44)، أي أن مدى معامل التمييز الذي تقع فيه الفقرات هو (0.14)، وحتى يتحقق افتراض تساوي معاملات التمييز ومطابقتها للنموذج يجب أن تكون جميع قيم معاملات التمييز واقعة ضمن المدى (متوسط معاملات التمييز ± 0.15)، وكان متوسط معاملات التمييز يساوي 0.38، وتراوح مدى المطابقة بين (0.23) وبين (0.53)، وكما تبين من جدول (3) فإن جميع قيم معاملات التمييز للفقرات وقعت ضمن هذا المدى، وهذا مؤشر على تحقق افتراض تساوي معاملات التمييز وبالتالي توازي المنحنيات المميزة للفقرات(Hambleton & Swaminathan, 1985). وهذا افتراض يدل على موضوعية القياس، أي استقلال قدرة الفرد عن الفقرة المستخدمة، واستقلال صعوبة الفقرة عن قدرات الأفراد.

- افتراض انعدام التخمين: تم التحقق من عدم التخمين أو انخفاض مؤشر التخمين من خلال ملاحظة أداء المفحوصين من ذوي القدرة المتدنية على الفقرات الصعبة، حيث كان أداؤهم عليها بشكل متدن وبدرجة كبيرة.

- افتراض التحرر من السرعة: تأكد الباحث أن إخفاق الأفراد في الإجابة عن فقرات الاختبار يعود إلى انخفاض قدراتهم، وليس إلى تأثير عامل السرعة في الإجابة وذلك عن طريق إعطاء الوقت الكافي لهم، للإجابة عن فقرات الاختبار، بالإضافة إلى أنه لم يشتك أي طالب في أثناء التطبيق من ضيق الوقت، وعدم كفايته.

بعد أن تم التحقق من افتراضات نموذج راش تم تحليل استجابات الطلاب على فقرات الاختبار باستخدام برنامج (Winsteps) على ثلاثة مراحل، في المرحلة الأولى تم التحقق من مطابقة الأفراد للنموذج وحذف الأفراد غير المطابقين، وفي المرحلة الثانية تم التحقق من مطابقة الفقرات للنموذج وحذف الفقرات غير المطابقة، وذلك بعد حذف الأفراد غير المطابقين للنموذج، ويعد هذا مؤشراً لصدق القياس، وفي المرحلة الثالثة وبعد حذف الأفراد غير المطابقين للنموذج والفقرات غير المطابقة تم تدريج فقرات الاختبار بصورته النهائية وفق نموذج راش وحساب معامل الثبات،وهذا يعتبر دليلاً على موضوعية تفسير نتائج الاختبار وموضوعية القياس وفيما يلي نتائج التحليل باستخدام نموذج راش:

- يقوم البرنامج بحذف استجابات الأفراد الذين يحصلون على العلامة الكاملة على الاختبار أي (40)، وكذلك بحذف استجابات الأفراد الذين يحصلون على العلامة (صفر) على الاختبار، ولم يحصل أي فرد على العلامة الكاملة (40) بينما حصل فرد واحد على العلامة (صفر)، وبالتالي قام البرنامج بحذف استجابات هذا الفرد، وأصبح عدد أفراد عينة الدراسة التي دخلت في التحليل الأولي (955) فرداً.

- يقوم البرنامج بحذف الفقرات التي يجيب عنها جميع الأفراد إجابة صحيحة، وكذلك جميع الفقرات التي يجيب عنها جميع الأفراد إجابة خاطئة، ولم يكن هناك أي فقرة أجاب عنها جميع الأفراد إجابة صحيحة، وكذلك لم يكن هناك أي فقرة أجاب

عنها جميع الأفراد إجابة خاطئة، وبالتالي لم يحذف البرنامج أي فقرة ودخلت جميع فقرات الاختبار (40) في التحليل الأولي. وقد يعزى ذلك لحرص الباحث على أن لا يحتوي الاختبار على فقرات صعبة جداً يخفق جميع الأفراد في الإجابة عنها إجابة صحيحة، ولا فقرات سهلة جداً يجيب عنها كل الأفراد إجابة صحيحة.

- تم التحقق من مطابقة الأفراد للنموذج: قبل البدء بتحليل الاستجابات عن الفقرات لا بد من حذف الأفراد غير المطابقين للنموذج، ولمعرفة مؤشرات المطابقة الخاصة بالأفراد، تم تقدير قدرة كل فرد بالإضافة إلى الخطأ المعياري في قياس هذه القدرة، وكذلك قيم إحصائي المطابقة الكلي وله مؤشران: مؤشر المطابقة الداخلية التقاربي (ZSTD) ويعبر عنه كذلك بإحصائي متوسط المربعات التقاربي (2)، وهو مؤشر إحصائي للسلوكيات غير المتوقعة التي تؤثر في استجابات الأفراد عن الفقرات التي تكون قريبة من مستوى قدراتهم. وكذلك مؤشر المطابقة الخارجية التباعدي (ZSTD) ويعبر عنه كذلك بإحصائي متوسط المربعات التباعدي (MNSQ) وهو مؤشر إحصائي يعد بديلاً عن السابق، وله صفات مقاربة ومشابهة، ولكنه أكثر حساسية للسلوكيات غير المتوقعة في استجابات الأفراد عن الفقرات التي تكون بعيدة عن مستوى قدراتهم، وذلك لكل تقدير من تقديرات القدرة، والجدول (4) يبين المتوسط الحسابي والانحراف المعياري لكل تقدير من تقديرات القدرة، والخطأ المعياري في قياس هذه القدرة، وقيم إحصاءات المطابقة الداخلية والخارجية ومتوسطات المربعات للمطابقة الداخلية والخارجية.

جدول (4)

المتوسط الحسابي والانحراف المعياري لمعلم القدرة والخطأ المعياري في قياس القدرة وإحصاءات المطابقة الداخلية والخارجية

	القدرة	الخطأ المعياري	إحصائي المطابقة الداخلية INFIT		إحصائي المطابقة الخارجية OUFIT	
			متوسط المربعات MNSQ	قيمة الإحصائي ZSTD	متوسط المربعات MNSQ	قيمة الإحصائي ZSTD
المتوسط الحسابي	-0.32	0.36	1	0	1	0
الانحراف المعياري	0.85	0.05	0.12	0.9	0.18	0.9

يتبين من الجدول(4) أن متوسط قدرات أفراد العينة الكلية قد بلغت (0.32−) لوجيت بانحراف معياري (0.85)، ويتضح أيضاً أن المتوسط الحسابي لمتوسط المربعات للمطابقة الداخلية يساوي (1) والانحراف المعياري لمتوسط المربعات للمطابقة الداخلية تساوي (0.12) وهذه القيم تقترب من القيم التي يفترضها النموذج وهي (1) للوسط الحسابي و (0) للانحراف المعياري، وكانت قيمة المتوسط الحسابي لمتوسط المربعات للمطابقة الخارجية (1) والانحراف المعياري لمتوسط المربعات للمطابقة الخارجية تساوي (0.18) وهذه القيم تقترب من القيم التي يفترضها النموذج وهي (1) للوسط الحسابي و (0) للانحراف المعياري، كما يلاحظ أن المتوسط الحسابي لقيم إحصائي المطابقة الداخلية قد بلغ (0) والانحراف المعياري لها (0.9) وهي أيضاً تقترب من القيم المثالية التي يتوقعها النموذج وهي (0) للوسط الحسابي و (1) للانحراف المعياري، كما يلاحظ أن المتوسط الحسابي لقيم إحصائي المطابقة الخارجية قد بلغ (0) والانحراف المعياري لها (0.9) وهي أيضاً تقترب من القيم المثالية التي يتوقعها النموذج وهي (0) للوسط الحسابي و (1) للانحراف المعياري. وتم حذف (63) فرداً من ذوي القدرات المتدنية وذوي القدرات المرتفعة من العينة، لأن قيمة إحصائي المطابقة الخارجية أو الداخلية المقابلة لعلاماتهم على الاختبار أكبر من (2) أو أقل من (2−)، ويفترض النموذج أن استجابة المفحوص الملاحظة تبتعد عن استجاباته المتوقعة إذا زادت قيمة إحصائي المطابقة الخارجية المقابلة لعلاماته عن (2)، وهذا يعني أن المفحوصين غير مطابقين للنموذج لأن استجاباتهم الملاحظة تبتعد عن توقعات النموذج، وهذا يعني إما أن الأفراد يجيبون صواباً عن الفقرات التي تزيد درجة صعوبتها عن مستوى قدراتهم أو يفشلون في الإجابة عن فقرات درجة صعوبتها أدنى من مستوى قدراتهم وقد يعزى ذلك إلى ميل بعض المفحوصين من ذوي القدرات الدنيا إلى التخمين عند الإجابة على الفقرات التي قيم معامل صعوبتها أكبر من قدراتهم، أما بالنسبة للأفراد من ذوي القدرات العليا الذين تم حذفهم، فقد يعزى إخفاقهم في الإجابة عن الفقرات التي تقل درجة عن مستوى قدراتهم إلى أنهم يميلون إلى تفسير ما تقيسه الفقرة بأكثر مما تحتمل وقد يعزى إلى الإهمال أو التسرع.أما إذا كانت قيمة هذا الإحصائي أقل من (2−) فقد يعني هذا تشابه إجابات الأفراد. وكذلك يتم الحكم على دلالة قيمة المطابقة الداخلية للأفراد من خلال إحصائي المطابقة الداخلية (ZSTD) والذي يتراوح بين (2−) و (2). فإذا كانت قيمة هذا الإحصائي موجبة بحيث تتجاوز القيمة (2) فإن هذا يعني أن الأفراد يجيبون بصورة خاطئة عن فقرات تقترب في مستوى صعوبتها من مستوى قدراتهم. أما إذا كانت قيمة هذا الإحصائي أقل من (2−) فقد يعني هذا تشابه إجابات الأفراد.

إن حذف هؤلاء الأفراد (63) الذين لم تتطابق استجاباتهم مع افتراضات النموذج، يعني الإبقاء على الأفراد الذين تطابقت استجاباتهم مع افتراضات النموذج، أي الذين يكون هناك اتساقاً بين استجاباتهم على الفقرة واحتمال نجاحهم فيها، وهذا يعني وجود اتساق بين استجاباتهم على الفقرة ودرجاتهم الكلية على الاختبار، مما يدل على الاتفاق بين السمة التي تقيسها الفقرة والسمة التي تقيسها باقي فقرات الاختبار، ويُعد هذا مؤشراً لصدق القياس.

- تم التحقق من مطابقة الفقرات للنموذج: بعد استبعاد الأفراد الذين لم تتطابق استجاباتهم الملاحظة مع توقعات النموذج، تم إعادة التحليل لحذف الفقرات غير المطابقة لنموذج راش، حيث تم تقدير معلم الصعوبة لكل فقرة، بالإضافة إلى الخطأ المعياري في قياس هذا المعلم، وقيم إحصائي المطابقة الداخلية (ZSTD) للفقرات ومتوسط المربعات الداخلية (MNSQ) للفقرات، وكذلك قيم إحصائي المطابقة الخارجية (ZSTD) للفقرات وقيم متوسطات المربعات الخارجية (MNSQ) لكل

معلم من معالم الصعوبة. والجدول (5) يبين المتوسط الحسابي والانحراف المعياري لكل من تقديرات معالم الصعوبة للفقرات، والخطأ المعياري في قياس هذه الصعوبة، وكذلك قيم إحصاءات المطابقة الداخلية والخارجية لهذه المعالم.

جدول(5)

المتوسط الحسابي والانحراف المعياري لمعلم الصعوبة والخطأ المعياري في قياس القدرة وإحصاءات المطابقة الداخلية والخارجية

	القدرة	الخطأ المعياري	إحصائي المطابقة الداخلية INFIT		إحصائي المطابقة الخارجية OUFIT	
			متوسط المربعات MNSQ	قيمة الإحصائي ZSTD	متوسط المربعات MNSQ	قيمة الإحصائي ZSTD
المتوسط الحسابي	0	0.08	1	0	0.99	-0.1
الانحراف المعياري	0.77	0	0.02	0.9	0.03	0.8

يتضح من الجدول (5) أن المتوسط الحسابي لصعوبة الفقرات يساوي (0) وهي القيمة التي يفترضها النموذج وأن الانحراف المعياري للصعوبة يساوي (0.77) وهي تقترب من القيمة التي يفترضها النموذج وهي (1)، وكانت قيمة المتوسط الحسابي لمتوسط المربعات للمطابقة الداخلية يساوي (1) والانحراف المعياري لمتوسط المربعات للمطابقة الداخلية تساوي (0.02) وهذه القيم تقترب من القيم التي يفترضها النموذج وهي (1) للوسط الحسابي و (0) للانحراف المعياري، وكانت قيمة المتوسط الحسابي لمتوسط المربعات للمطابقة الخارجية (0.99) والانحراف المعياري لمتوسط المربعات للمطابقة الخارجية تساوي (0.03) وهذه القيم تقترب من القيم التي يفترضها النموذج وهي (1) للوسط الحسابي و (0) للانحراف المعياري، كما يلاحظ أن المتوسط الحسابي لقيم إحصائي المطابقة الداخلية قد بلغ (0) والانحراف المعياري لها (0.9) وهي أيضاً تقترب من القيم المثالية التي يتوقعها النموذج وهي (0) للوسط الحسابي و (1) للانحراف المعياري، كما يلاحظ أن المتوسط الحسابي لقيم إحصائي المطابقة الخارجية قد بلغ (0.1–) والانحراف المعياري لها (0.8) وهي أيضاً تقترب من القيم المثالية التي يتوقعها النموذج وهي (0) للوسط الحسابي و (1) للانحراف المعياري.وفي ضوء محكات المطابقة للفقرات المستخدمة، تبين وجود فقرتين فقط (1، 16) تعدت قيم إحصاءات المطابقة الداخلية والخارجية (ZSTD) المدى (2–، 2)، وهذا مؤشر إلى إنها فقرات غير مطابقة وبعيدة عن توقعات النموذج وإنها قيم مضطربة وتشوش البيانات، فالفقرة رقم (1) تجاوزت قيم إحصاءات المطابقة الداخلية والخارجية الخاصة بها القيمة (2)، وقد يعزى ذلك إلى عيب في صياغتها أو عدم صدقها في قياس ما تقيسه بقية الفقرات، أما الفقرة رقم (16) فقد كانت إحصاءات المطابقة لها تقل عن القيمة (2–) وهذا يعني أن الفقرة شديدة التشابه أو الاعتماد على بعض فقرات المقياس؛ أي لا تكون مستقلة عنها.

إن حذف الفقرات غير المطابقة لافتراضات النموذج وعددها (2) يدل على أن كل فقرة من الفقرات المتبقية في الاختبار وعددها (38) فقرة مطابقة لافتراضات النموذج تتفق في تعريفها للمتغير مع ذلك الذي تعرفه بقية فقرات الاختبار، وهذا يعني أن هناك اتفاقاً بين السمة التي تقيسها الفقرة والسمة التي تقيسها باقي الفقرات، وهذا يُعد مؤشراً لصدق القياس.

- تم التحقق من جودة مطابقة البيانات الناتجة عن استجابات أفراد العينة (892) على الاختبار، والتحقق من موضوعية تفسير نتائج الاختبار بصورته النهائية (38) فقرة؛ أي بعد حذف الأفراد غير المطابقين والفقرات غير المطابقة للنموذج، وكذلك التحقق من الافتراضات الخاصة بنموذج راش إذ يُعد ذلك تحقيقاً لمدى صدق النموذج في تحقيقه لموضوعية القياس، أعيد التحليل للمرة الثالثة للحصول على تقديرات نهائية متحررة لكل من صعوبة الفقرات وقدرات الأفراد، والجدول (6) يبين نتائج التحليل للقيم المتحررة لقدرة الأفراد.

جدول(6)

نتائج التحليل للقيم المتحررة لقدرة الأفراد

	القدرة	الخطأ المعياري	المدى	إحصائي المطابقة الداخلية INFIT		إحصائي المطابقة الخارجية OUFIT		أدنى علامة خام	أقصى علامة خام
				متوسط المربعات MNSQ	قيمة الإحصائي ZSTD	متوسط المربعات MNSQ	قيمة الإحصائي ZSTD		
المتوسط الحسابي	-0.33	0.37	-3.91	1	0	0.99	0		
الانحراف المعياري	0.85	0.05	إلى 2.69	0.11	0.8	0.17	0.8	1	35

يلاحظ من الجدول (6) أن التقديرات النهائية المتحررة لقدرات الأفراد قد تراوحت بين العلامة الكلية (1) كحد أدنى والعلامة الكلية (35) كحد أقصى، وقد بلغ متوسط توزيع القدرة (0.33–) وحدة لوجيت بانحراف معياري قدره(0.85) وحدة لوجيت، أي بمدى يتراوح بين (3.91–) وحدة لوجيت للمستوى المنخفض من القدرة و(2.69) وحدة لوجيت للمستوى المرتفع من القدرة، وبلغ متوسط الأخطاء المعيارية لتقديرات قدرة الأفراد (0.37) وهي قيمة متدنية، مما يشير إلى دقة تحديد مواقع الأفراد على متصل السمة. ويلخص الجدول (7) نتائج التحليل للقيم المتحررة لصعوبة الفقرات.

جدول (7)

نتائج التحليل للقيم المتحررة لصعوبة الفقرات

	الصعوبة	الخطأ المعياري	المدى	إحصائي المطابقة الداخلية INFIT		إحصائي المطابقة الخارجية OUFIT	
				متوسط المربعات MNSQ	قيمة الإحصائي ZSTD	متوسط المربعات MNSQ	قيمة الإحصائي ZSTD
المتوسط الحسابي	0	0.08	-1.78 إلى 1.32	1	0	0.99	0
الانحراف المعياري	0.78	0		0.02	0.7	0.03	0.6

يتضح من الجدول (7) إن قيم التقديرات المتحررة لصعوبة الفقرات قد تراوحت ما بين (1.78–) إلى (1.32) وحدة لوجيت، وتتوزع بمتوسط قدره (0) لوجيت وانحراف معياري قدره (0.78) لوجيت، وقد بلغ متوسط الأخطاء المعيارية للفقرات (0.08) وهي قيمة متدنية، مما يشير إلى دقة تقديرات الصعوبة للفقرات، ولدى مقارنة محوري تدريج الفقرات والقدرة لوحظ تطابقهما إلى حد كبير؛ أي إن هناك نوعاً من الاتساق في تدرج صعوبة الفقرات وإن الاختبار يقيس مدى واسعاً من القدرة، ويقيس متغيراً أحادي البعد، الأمر الذي يعكس صدق البناء للاختبار.

من النتائج أعلاه تم التحقق من موضوعية تفسير نتائج الاختبار وموضوعية القياس من خلال التحقق من افتراضات نموذج راش على بيانات الاختبار بصورته النهائية، وذلك بعد حذف الأفراد غير المطابقين لافتراضات النموذج، وحذف الفقرات غير المطابقة لافتراضات النموذج على المحكات المختلفة، ويُعد التحقق من افتراضات النموذج في جوهره تقويماً لمدى صدق النموذج في تحقيق موضوعية القياس.

ثانياً: نتائج السؤال الثاني ومناقشتها

◄ نص السؤال الثاني على: **ما قيم تقديرات معلم الصعوبة لفقرات الاختبار بصورته النهائية وفق نموذج (راش)؟**

للإجابة عن السؤال الثاني، حسبت قيم معلم الصعوبة لفقرات الاختبار بعد تدريجها باستخدام نموذج راش والخطأ المعياري لكل معلم، والجدول (8) يبين نتائج ذلك.

جدول (8)

معلم الصعوبة لفقرات الاختبار باللوجيت والخطأ المعياري لها مرتبة تصاعدياً

ترتيب الفقرات	الصعوبة باللوجيت	الخطأ المعياري	ترتيب الفقرات	الصعوبة باللوجيت	الخطأ المعياري	ترتيب الفقرات	الصعوبة باللوجيت	الخطأ المعياري
5	-1.78	0.09	22	-0.13	0.07	13	0.53	0.08
37	-1.41	0.08	10	-0.11	0.07	29	0.56	0.08
17	-1.21	0.08	9	-0.09	0.07	26	0.60	0.08
36	-1.09	0.08	20	0.01	0.07	38	0.64	0.08
21	-1.06	0.07	40	0.09	0.07	28	0.65	0.08
14	-1.03	0.07	6	0.10	0.07	34	0.82	0.08
39	-0.99	0.07	7	0.14	0.07	23	0.85	0.08
3	-0.88	0.07	15	0.16	0.07	30	0.88	0.08
33	-0.81	0.07	2	0.20	0.07	27	0.96	0.08
8	-0.63	0.07	32	0.21	0.07	4	1.19	0.09
35	-0.47	0.07	31	0.27	0.07	24	1.25	0.09
19	-0.32	0.07	11	0.33	0.07	25	1.32	0.09
12	-0.21	0.07	18	0.47	0.08			

يتضح من الجدول (8) أن مدى قيم معلم الصعوبة للفقرات تراوح من (1.78–) للفقرة (5) وهي فقرة سهلة إلى (1.32) للفقرة (25) وهي فقرة صعبة نوعاً ما، مما يعني أن معظم الفقرات متوسطة الصعوبة، وهي فقرات مقبولة حيث إن معظم الطلبة تقع قدراتهم ضمن المدى (2–) إلى (2). كما أن قيم الخطأ المعياري منخفضة لجميع الفقرات حيث كانت جميعها دون (0.09).

وللتعرف على جودة تدريج فقرات الاختبار، تم حساب الفروق بين مستوى صعوبة كل فقرتين متتاليتين في التدريج ومقارنته بمجموع الخطأ المعياري لهما لمعرفة فيما إذا كانت هناك فراغات بين الفقرات، فإذا كان الفرق بين الصعوبتين أكبر من مجموع الخطأ المعياري لهما فإن هذا يعني وجود فراغ بين الفقرتين على متصل القدرة المقاسة مما يستدعي إضافة فقرات جديدة لتغطية الفراغ وقياس القدرة. والجدول (9) يبين ترتيب الفقرات والفروق بين معلم صعوبة كل فقرتين متتاليتين، ومجموع الخطأ المعياري لهما والقرار المتخذ بشأنهما.

جدول (9)

الفروق بين مستوى صعوبة كل فقرتين متتاليتين ومجموع الخطأ المعياري لهما والقرار المتخذ بشأنهما.

ترتيب الفقرات	فروق الصعوبة باللوجيت	مجموع الخطأ المعياري	القرار	ترتيب الفقرات	فروق الصعوبة باللوجيت	مجموع الخطأ المعياري	القرار	ترتيب الفقرات	فروق الصعوبة باللوجيت	مجموع الخطأ المعياري	القرار
5	0.37	0.17	فراغ	22	0.02	0.14		13	0.03	0.16	
37	0.20	0.16	فراغ	10	0.02	0.14		29	0.04	0.16	
17	0.12	0.16		9	0.08	0.14		26	0.04	0.16	
36	0.03	0.15		20	0.08	0.14		38	0.01	0.16	
21	0.03	0.14		40	0.01	0.14		28	0.017	0.16	فراغ
14	0.04	0.14		6	0.04	0.14		34	0.03	0.16	
39	0.11	0.14		7	0.02	0.14		23	0.03	0.16	
3	0.07	0.14		15	0.04	0.14		30	0.08	0.16	
33	0.18	0.14	فراغ	2	0.01	0.14		27	0.23	0.17	فراغ
8	0.16	0.14	فراغ	32	0.06	0.14		4	0.06	0.18	
35	0.15	0.14	فراغ	31	0.06	0.14		24	0.07	0.18	
19	0.11	0.14		11	0.14	0.15		25			
12	0.08	0.14		18	0.06	0.16					

يتضح من الجدول (9) أن الفروق بين الصعوبات أقل من مجموع الخطأ المعياري ما عدا في (7) حالات،حالتان في بداية متصل القدرة بين الفقرتين (5) و (37)، وبين الفقرتين (37) و (17)، وحالتان في نهاية المتصل بين الفقرتين (28) و (34)، وبين الفقرتين (27) و (4)، وثلاث حالات في الثلث الأول من الاختبار بين الفقرتين (33) و (8)، وبين الفقرتين (33) و (8)، وبين الفقرتين (8) و (35)، وهذا يدل على جودة تدريج فقرات الاختبار، ويمكن إضافة فقرات إلى الاختبار لتغطية هذه الفراغات، ويكون ذلك من خلال إعداد بنك الأسئلة اختبار اللغة الانجليزية (1) بما يساعد في توفير صور متكافئة من الاختبار عند الاستخدام.

ثالثاً: نتائج السؤال الثالث ومناقشتها

نص السؤال الثالث على: **ما الخصائص السيكومترية للاختبار (دلالات الصدق والثبات)؟**

أ. صدق الاختبار:تتحقق موضوعية تفسير نتائج الاختبار وموضوعية القياس بتحقق افتراضات نموذج راش على بيانات الاختبار بصيغته النهائية، وذلك بعد حذف الأفراد غير المطابقين وحذف الفقرات غير المطابقة لافتراضات النموذج، والتحقق من افتراضات النموذج يعتبر تقويم لمدى صدق النموذج في تحقيق موضوعية القياس (Stenner, 1990)، وتم إجراء ذلك، وتم التحقق من صدق الاختبار كذلك من خلال:

- صدق المحكمين: من خلال الإجراءات التي اتبعت في تحديد مجال التحصيل في موضوعات اللغة الانجليزية (1)، وتحليل المحتوى وكتابة الأهداف السلوكية، وكتابة فقرات الاختبار، والتي تمثل الأهداف السلوكية كنتاج تعلمي يراد تحقيقه، والوصول به إلى مستوى الإتقان، والتأكد من صحة هذه الإجراءات، من خلال الأحكام التي قدمها المحكمون والمختصون في تدريس مقرر اللغة الانجليزية (1) عن مدى ملاءمة نتائج التحليل لمواصفات مجال التحصيل، ومدى التوافق بين مضمون الفقرات، والأهداف التي تحققها تلك الفقرات، وقد سبق تفصيل ذلك عند الحديث عن أداة الدراسة.
- الصدق التنبؤي: تم التحقق من الصدق التنبؤي من خلال حساب معامل ارتباط (بيرسون) بين علامات أفراد عينة الدراسة الذين نجحوا في الاختبار وعلامات هم في مقرر اللغة الانجليزية (2) الذي يدرسه الطالب في نفس الفصل الذي ينجح

به في الاختبار، أي في نهاية الفصل الدراسي الثاني من العام الأكاديمي 2015/ 2016، وكانت قيمة معامل الارتباط تساوي (0.88)، وهي قيمة دالة إحصائياً عند مستوى دلالة (0.01)، وهذا يعني أن (77 %) من التباين في علامات الطلبة في اللغة الانجليزية (2) يعزى إلى اختلاف علاماتهم على الاختبار. مما يؤكد على القدرة التنبؤية للاختبار بالتحصيل في اللغة الانجليزية (2)؛ أي إن الاختبار يتمتع بالصدق التنبؤي.

ب. ثبات الاختبار: بعد التحقق من جودة مطابقة فقرات الاختبار لنموذج (راش)، واستبعاد الفقرات التي لم تحقق هذه المطابقة، وكذلك الأفراد غير المطابقين للنموذج، وتحديد القيم التقديرية لكل من صعوبة الفقرات وقدرات الأفراد، ينبغي الحصول على نوعين من المعاملات، يطلق عليهما الثبات المتعلق بالأفراد، والثبات المتعلق بفقرات الاختبار. فمفهوم الثبات في إطار نموذج (راش)، يشير إلى مدى الدقة في تقدير موقع كل من الفقرات والأفراد على متصل السمة، الذي نهدف إلى قياسه.

ويمكن تحديد مدى دقة الفقرات في تعريف هذا المتصل، بإيجاد النسبة بين الانحراف المعياري للقيم التقديرية المتحررة لصعوبة الفقرات، ومتوسط الخطأ المعياري لهذه القيم. ويطلق على هذه النسبة معامل الفصل بين الفقرات (Gi) والذي يجب أن تكون قيمته أكبر ما يمكن. فإذا لم تزد هذه القيمة عن (2) فإنه يصعب قياس المتغير بوساطة هذه الفقرات، وقد وجد الباحث أن قيمة معامل الفصل بين الفقرات في الاختبار بصورته النهائية (10.14). وبما أن هذه القيمة تزيد عن (2) فإن فقرات الاختبار تعتبر كافية لتعريف متصل السمة الذي تقيسه. وكذلك وجد الباحث أن قيمة معامل الفصل بين الأفراد (Gp) تساوي (2.01)، وترمز إلى النسبة بين الانحراف المعياري للقيم التقديرية المتحررة لقدرات الأفراد، ومتوسط الخطأ المعياري لهذه القيم، ونظراً لأن هذه القيمة أيضاً تزيد عن (2)، فإن عينة الأفراد تعد كافية للفصل بين الفقرات.

وتم حساب معامل الثبات المتعلق بالأفراد من خلال العلاقة ($G_p^2 / (1 + G_p^2)$) فكان يساوي (0.80)، وهي قيمة مقبولة وتدل على إن عينة الأفراد كافية للفصل بين الفقرات والتمييز بين مستويات الصعوبة للفقرات، وبمعنى أوضح فإن الأفراد يتوزعون بشكل مناسب على متصل السمة الممثل بفقرات الاختبار التي بنيت وفق نموذج راش، وتم حساب معامل الثبات للفقرات من خلال العلاقة ($G_i^2 / (1 + G_i^2)$) فكان يساوي (0.99)، وهي قيمة مرتفعة جداً وتدل على إن عينة الفقرات كافية للفصل بين الأفراد وبالتالي التمييز بين مستويات التحصيل لدى هؤلاء الأفراد.

ويلاحظ أن قيمة معامل الثبات المتعلق بالأفراد لم تكن كبيرة بالدرجة الكافية، إذ إنها لم تكن قادرة على إظهار سوى (3) طبقات إحصائية متمايزة للأفراد؛ أي ثلاثة مستويات للتحصيل لدى هؤلاء الأفراد وفق المعادلة ($H_p = (4G_p + 1)/3$) حيث H_p ترمز إلى عدد الطبقات الإحصائية للأفراد، وقد يعزى ذلك إلى طول الاختبار بصورته النهائية (38) فقرة، إذ كان مدى صعوبة الفقرات بين (1.78–) و (1.32)، ويمكن زيادة معامل الثبات المتعلق بالأفراد، بإضافة فقرات سهلة للطلبة الذين تقل قدراتهم عن (1.78–)، وأخرى صعبة للطلبة الذين تزيد قدراتهم عن (1.32).

في حين إن قيمة معامل الثبات المتعلق بالفقرات كانت كبيرة جداً، وبتطبيق المعادلة أعلاه على الفقرات، فقد تبين أن الفقرات كانت قادرة على إظهار أربع عشرة طبقة إحصائية متمايزة تقريباً لصعوبة الفقرات لدى أفراد عينة الدراسة؛ أي إظهار الفروق الفردية بين الأفراد بشكل كبير. وقد يعزى ذلك إلى جودة الفقرات المكونة للاختبار بصورته النهائية.

وبعد أن تم بناء الاختبار عبر هذه المراحل من استبعاد الفقرات غير المطابقة لنموذج راش. واستبعاد الأفراد غير المطابقين. وبعد تحديد الخصائص السيكومترية للاختبار وفقراته، فإن هذا الاختبار المؤلف من (38) فقرة يمثل المجال بشكل متوازن مع الأهداف، وبالتالي يمكن تكوين اختبارات فرعية من هذا الاختبار؛ وتحديد علامة المجال من هذه الاختبارات. كما يمكن تشكيل اختبارات فرعية من هذا الاختبار، حسب درجة الصعوبة أو اختيار فقرات ذات تمييز مرتفع تصف الأفراد عند نقطة قطع محددة لأغراض محددة.

◄ رابعاً: نتائج السؤال الرابع ومناقشتها

نص السؤال الرابع على:**ما كمية المعلومات التي يقدمها الاختبار عند مستويات القدرة المختلفة؟**

استخدم الباحث البرنامج الإحصائي (Winsteps) لإيجاد دالة المعلومات للاختبار الكلي، وقدرت دالة المعلومات للاختبار عند كل مستوى من مستويات القدرة المحددة، والجدول (10) يبين تقديرات قدرات الأفراد بوحدة اللوجيت وقيم دالة المعلومات عند المستويات المختلفة للقدرة.

جدول (10)

تقديرات القدرة لأفراد العينة وقيم دالة المعلومات التي يقدمها الاختبار عند مستويات مختلفة من القدرة للاختبار ثنائي التدريج بصورته النهائية (38) فقرة

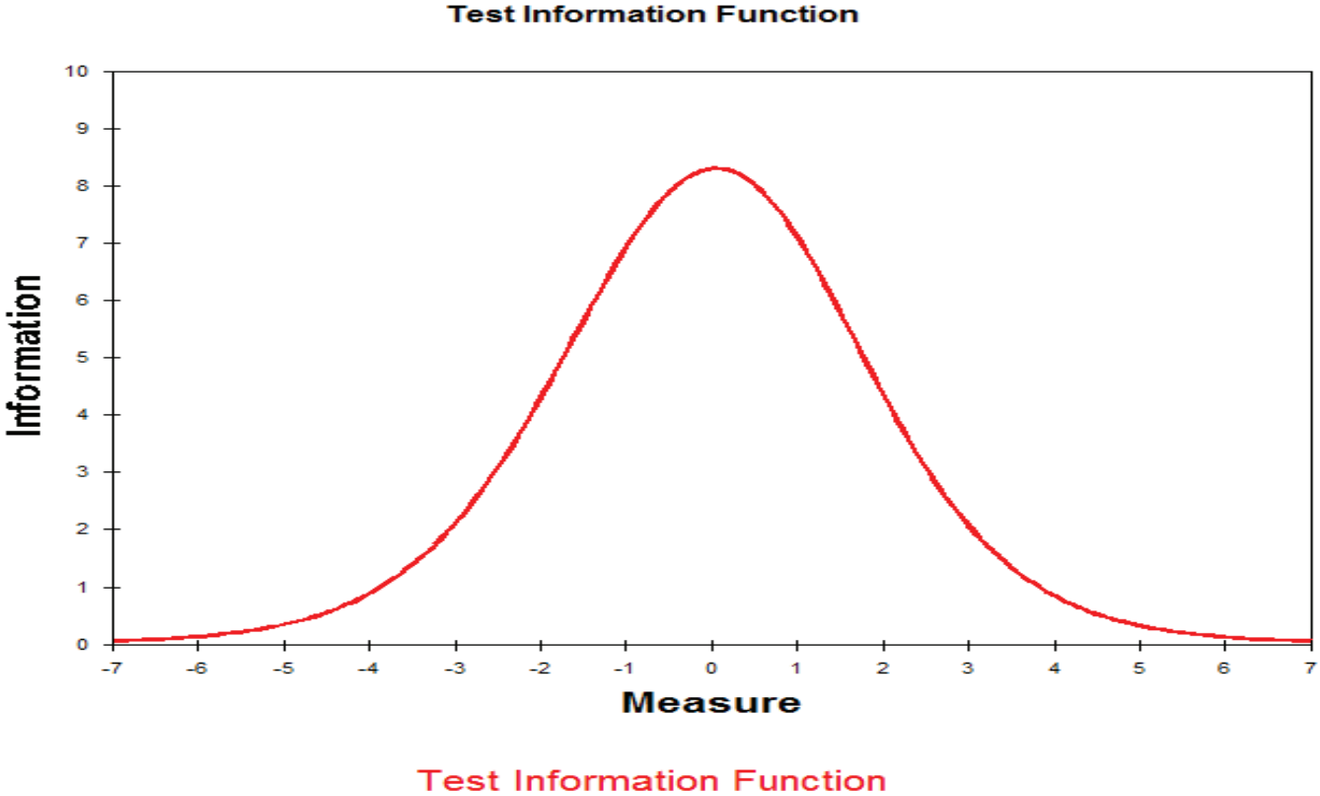
القدرة	الخطأ المعياري	دالة المعلومات	القدرة	الخطأ المعياري	دالة المعلومات	القدرة	الخطأ المعياري	دالة المعلومات
-5.15	1.84	0.30	-0.74	0.37	7.49	0.89	0.37	7.37
-3.91	1.02	0.95	-0.60	0.36	7.73	1.03	0.38	7.06
-3.17	0.74	1.83	-0.48	0.36	7.93	1.17	0.39	6.70
-2.72	0.62	2.62	-0.35	0.35	8.08	1.32	0.40	6.29

القدرة	الخطأ المعياري	دالة المعلومات	القدرة	الخطأ المعياري	دالة المعلومات	القدرة	الخطأ المعياري	دالة المعلومات
-2.38	0.55	3.35	-0.23	0.35	8.20	1.49	0.41	5.82
-2.11	0.50	4.01	-0.11	0.35	8.27	1.67	0.43	5.31
-1.88	0.47	4.62	0.01	0.35	8.31	1.87	0.46	4.74
-1.67	0.44	5.17	0.13	0.35	8.30	2.90	0.49	4.11
-1.49	0.42	5.67	0.25	0.35	8.26	2.36	0.54	3.42
-1.32	0.40	6.13	0.38	0.35	8.17	2.69	0.61	2.67
-1.16	0.39	6.53	0.50	0.36	8.03	3.14	0.7	1.85
-1.01	0.38	6.90	0.62	0.36	7.86	3.87	1.02	0.96
-0.87	0.37	7.22	0.75	0.36	7.64	5.10	1.84	0.30

يتضح من النتائج في الجدول (10) إن قيمة دالة المعلومات للاختبار تتزايد تدريجياً بتزايد القدرة (Θ) حتى تصل إلى أقصى قيمة ممكنة لها عندما تكون قيمة القدرة (0.01) لوجيت، وهي تقابل متوسط الصعوبة للفقرات (0) لوجيت. وهذا ما يتطابق مع ما هو متوقع من نموذج راش، إذ إن دالة المعلومات تكون أقصى ما يمكن عندما تكون قيمة (Θ=b) وذلك لكل من النموذجين الأحادي والثنائي المعلم. وهذا ما تؤكده قيمة الخطأ المعياري للتقدير التي كانت أقل ما يمكن عند متوسط الصعوبة لفقرات الاختبار، إذ بلغت قيمته (0.35). إذ يكون الخطأ المعياري صغيراً كلما اقترب مستوى القدرة للأفراد من مستوى الصعوبة للفقرات. وبالتالي تزداد كمية المعلومات بنقصان الخطأ المعياري حسب العلاقة $1(0) = \frac{1}{\sqrt{S.E\,(0)}}$ حيث إن I (Θ): كمية المعلومات عند مستوى القدرة (Θ) و S.E (Θ) : الخطأ المعياري في تقدير القدرة (Θ). والشكل (2) يبين منحنى دالة معلومات الاختبار

الشكل (2):

منحنى معلومات الاختبار



يتضح من الشكل (2) إن منحنى معلومات الاختبار يأخذ الشكل الاعتدالي المتوقع لمثل هذه المنحنيات، كما كانت أكبر قيمة للمعلومات عند مستوى قدرات الأفراد وفقاً للاختبار بدقة مرتفعة داخل مدى القدرة (–1.32، 1.32)، وهو ما كان متوقعاً؛ لأن معظم فقرات الاختبار غطت هذا المدى. وقد كانت أقصى قيمة لمعلومات الاختبار (8.31) وبأقل خطأ معياري مقداره (0.35) عند مستوى القدرة (0.01). وانخفضت قيمة المعلومات التي يمكن الحصول عليها من الاختبار بشكل كبير مع زيادة مستوى القدرة عن (1.32) أو انخفاض مستوى القدرة عن (–1.32)، وهو ما يؤكد ضرورة إضافة فقرات تغطي تلك المستويات من القدرة.

التوصيات:

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

1. الاستفادة من الاختبار الذي تم بناؤه وفق نظرية الاستجابة للفقرة، والمكون بصورته النهائية من (38) فقرة، أو أي اختبار مستل منه، لقياس تحصيل الطلبة في جامعة القدس المفتوحة في مقرر اللغة الانجليزية(1)؛ نظراً لتمتع الاختبار المبني بدلالات صدق وثبات مقبولة، وبمعايير تتفق مع متطلبات النموذج أحادي المعلم (نموذج راش).
2. التحقق من الخصائص السيكومترية للاختبار الذي تم بناؤه، باستخدام النموذج ثنائي المعلم والنموذج ثلاثي المعلم لنظرية الاستجابة للفقرة، قبل اعتماده كأداة لقياس تحصيل الطالب الجامعي في اللغة الانجليزية(1).
3. الاستفادة من المنهجية المستخدمة في بناء الاختبار لبناء اختبارات لمقررات جامعية أخرى والتأكد من خصائصها السيكومترية.

المصادر والمراجع:

أولاً- المراجع العربية:

1. حمادنة، إياد (2009). استخدام نظرية الاستجابة للفقرة في بناء اختبار محكي المرجع في الرياضيات وفق النموذج اللوجستي ثلاثي المعلم.مجلة العلوم التربوية والنفسية.10(2)، 216 – 230.
2. الشافعي، محمد منصور (1996). أثر طرق معادلة درجات الاختبار وضوابط اختيار العينة على تدريج بنوك الأسئلة باستخدام نموذج راش. رسالة دكتوراه غير منشورة، كلية التربية، جامعة المنصورة، المنصورة، مصر.
3. الشرقاوي، أنور والشيخ، سليمان وكاظم، أمينة وعبدالسلام، نادية (1996).اتجاهات معاصرة في القياس والتقويم النفسي والتربوي. القاهرة: مكتبة الأنجلو المصرية.
4. الشريفين، نضال (2006). الخصائص السيكومترية لاختبار محكي المرجع في القياس والتقويم وفق النظرية الحديثة في القياس التربوي والنفسي.مجلة العلوم التربوية والنفسية.7(4)، 81 – 109.
5. الطراونة، صبري (2016). بناء اختبار محكي المرجع في الإحصاء التحليلي لطلبة الدراسات العليا في كليات العلوم التربوية في الجامعات الأردنية الحكومية باستخدام نظرية استجابة الفقرة. مجلة العلوم التربوية والنفسية. 17(1)، 515 – 542.
6. عبابنة، عماد (2009). الاختبارات محكية المرجع: فلسفتها وأسس تطويرها. عمان: دار المسيرة.
7. علام، صلاح الدين (1985). تحليل بيانات الاختبارات العقلية باستخدام نموذج راش اللوغاريتمي الاحتمالي، دراسة تجريبية. المجلة العربية للعلوم الإنسانية. 5(17)، 101 – 122.
8. علام، صلاح الدين (2001).الاختبارات التشخيصية مرجعية المحك في المجالات التربوية والنفسية والتدريبية. القاهرة: دار الفكر العربي.
9. علام، صلاح الدين (2006). القياس والتقويم التربوي والنفسي أساسياته وتطبيقاته وتوجهاته المعاصرة. القاهرة: دار الفكر العربي.
10. علي، نداء (2012).فاعلية استخدام نموذج راش في بناء اختبار تحصيلي محكي المرجع لمقرر القياس والتقويم في التربية.رسالة دكتوراه غير منشورة، جامعة دمشق، دمشق، سوريا.
11. عودة، أحمد (1992). مدى التوافق بين نموذج راش والمؤشرات التقليدية في اختيار فقرات مقياس اتجاه سباعي التدريج.مجلة كلية التربية، جامعة الإمارات، (8)، 153 – 179.
12. عوض الله، عبدالرحيم (2010). مقارنة بين أسلوبي أنموذج راش والطريقة التقليدية في بناء اختبارات الذكاء باستخدام محك التنبؤ بالتحصيل الدراسي. رسالة دكتوراه غير منشورة، جامعة بغداد، بغداد، العراق.
13. كاظم، أمينة (1988).دراسة نظرية نقدية حول القياس الموضوع للسلوك (نموذج راش) سلسلة الكتب المتخصصة. الكويت: مؤسسة الكويت للتقدم العلمي.
14. المصري، محمد (2015). الخصائص السيكومترية لاختبار محكي المرجع في البحث التربوي وفق نظرية الاستجابة

للفقرة.مجلة كلية التربية الأساسية.21(89)، 701 – 730.

15. منسي، محمود (1998).التقويم التربوي. الاسكندرية: دار المعرفة الجامعية.

16. هادي، فوزية ومراد، صلاح (2013). استخدام نموذج راش في إعداد وتقنين اختبار في اللغة العربية لقبول طلاب الدراسات العليا لكليات العلوم الانسانية في جامعة الكويت. المجلة التربوية. 1(106)، 15 – 53.

17. ياسين، عمر (2004).الخصائص السيكومترية لاختبار محكي المرجع في الكيمياء لطلاب الصف الأول الثانوي العلمي مقدرة وفق النظريتين الكلاسيكية والحديثة في القياس.رسالة دكتوراه غير منشورة، جامعة عمان العربية، عمان، الأردن.

ثانياً المراجع الأجنبية:

1. *Anastasi, A. (1982). Psychological testing (5th ed). New York: McMillan Publishing Co.*
2. *Bowles, R. (2005). Item response theory. Retrieved from http://www.yahoo.com.*
3. *Cunningham, G. K. (1986). Educational And Psychological Measurement.Mac Milan publishers Company, Network.*
4. *Eyres, S., Carey, A., Gilworth, G., Neumann, V. & Tennant, A. (2005). Construct validity and reliability of the rivermeadpost. Concussion Symptoms Questionnaire Clinical Rehabilitation. 19, 818 – 887.*
5. *Guskey, T. (1988). Implementing Mastery Learning. California:Wadsworth Publishing , Co.*
6. *Gyagenda, I. & Engelhard, G. (1998). Applying the Rasch model to explore rater influence on the assessed quality of student's writing ability. paper presented at the annual meeting of the American Educational Research Association. San Diego, pp.1 – 30, ED. 422 350.*
7. *Hambleton ,R. K., & Swaminathan, H. (1989). Item Response Theory Principles And Applications. Kluwer-Nijh of publishing, Boston, U.S.A.*
8. *Hambleton, K & Swaminathan, H. & Rogers , H. (1999). Fundamentals of item response theory. Newbury park, CA: Sage Publication.*
9. *Hambleton, R. (1988). principles and selected application of item response theory. New York: Macmillan publishing company.*
10. *Hambleton, R. (1994). The rise and fall of criterion – referenced measurement. Educational Measurement. 13(4), 21 – 26.*
11. *Hambleton, R. K & Swaminathan, H. (1985). Item response theory principles and applications. Boston: Nijhoff Publishing.*
12. *Hambleton, R. K. (1987). Determining Optimal Test Lengths With A Fixed Total Testing Time. Educational and Psychological Measurement. Vol(47), NO (2), P (339- 338).*
13. *Kim, B. & Hong, S. (2004). A Psychometric revision of the Asian values scale using the Rasch model. Measurement an Evaluation in Counseling and Development.37, 15 –92.*
14. *Lord, F. & Stocking, M. (1989). Item Response Theory. New York: Pergamon Press.*
15. *McCamey, R. (2002). A primer on the one-parameter Rasch model. Paper presented at the annual meeting of the south west educational research association austin, pp. 1 – 18, ED. 462 440.*
16. *Popham, w. J. (1978). Criterion– Referenced measurement. Englewood Cliffs, NJ: Prentice – Hall, Inc. Retrived December 30. 2005 from: www.leidenuniv. N1/gruigterdnmde.*
17. *Stenner, G. (1990). Objectivity specific and general. Rasch Measurement Tran actions, 4(3), 111-118.*
18. *Stocking. M. (1999). Item response theory. In: G. Masters, & J. Keevs (Eds). Advances in Measurement in Educational Research and Assessment. (pp.322) The Boulevard, UK: Pergamon.*
19. *Van, X. (1998). Item response theory and classical theory: An empirical comparison of their item/ person statistics. Educational and psychological Measurement. 58(3), 357 – 381.*
20. *Wiberge, M. (2004). Classical test theory vs. Item response theory An Evaluation of the theory test in the Swedish driving- license test. Umea University, department of educational measurement, EM, 50, ISSN: 1103-2685.*
21. *Wright, D., & Stone, M. (1979). Best test design: A handbook for research measurement. Chicago: MESA Press.*