

بناء اختبار محكي المرجع في الرياضيات وفق النموذج اللوجستي ثلاثي المعلمة

Constructing a Criterion - Referenced Test in Mathematics Using the Three-Parameter Logistic Model

Mahdi M. Olimat
Ministry of Education\ Jordan
drmahdiolimat@gmail.com

مهدي محمد عليما
وزارة التربية والتعليم/ الأردن

Received: 27/ 8/ 2021, Accepted: 17/ 11/ 2021.

تاريخ الاستلام: 27 / 8 / 2021م، تاريخ القبول: 17 / 11 / 2021م.

DOI: 10.33977/1182-013-038-004

E-ISSN: 2307-4655

<https://journals.qou.edu/index.php/nafsia>

P-ISSN: 2307-4647

والتقويم عملية منهجية تتطلب جمع بيانات موضوعية وصادقة من مصادر متعددة باستخدام أدوات متنوعة في ضوء أهداف محددة بغرض التوصل إلى تقديرات كمية وأدلة صافية يُستند إليها في إصدار أحكام، أو اتخاذ قرارات مناسبة تتعلق بالأفراد، ومما لا شك فيه أن هذه القرارات لها أثر كبير في مستوى أداء المتعلم وكفاءته في القيام بأعمال أو مهام معينة. وتعد الاختبارات من أهم أدوات التقويم، فالاختبار الجيد مقياس لا غنى عنه لمعرفة مدى نجاح المواقف التعليمية المختلفة. مما يبين تزايد الاهتمام بدراسة أساليبها، وأنواعها، وطرق صياغة فقراتها لتطويرها لتؤدي الغرض المطلوب منها.

تتعدد أنواع الاختبارات حسب الأغراض والأهداف التي أعدت لأجلها، لذلك يمكن تصنيف الاختبارات وفقاً لمعايير مختلفة.

وذكر عودة (2010) أن الاختبارات تصنف حسب تفسير للنتائج إلى قسمين: الاختبارات معيارية المرجع (Norm - Ref) (NRT) (erenced - Test) يقارن أداء الطالب في الاختبار بأداء مجموعته المعيارية (Norm - Group)، وهذه المجموعة قد تكون من طلاب صفه، أو من المرحلة العمرية نفسها. ويتركز الاهتمام فيه على ترتيب الفرد لأقرانه في القدرة التي يقيسها الاختبار. وفي هذا النوع من الاختبارات يحصل الفرد على درجة كلية تمثل أداءه بوجه عام محتوى معين، وكثير من الاختبارات التحصيلية المقننة (Stan-dardized Tests) تعد من هذا النوع. ولكن من الصعب استخدام هذه الاختبارات لتحديد وتشخيص مواطن القوة والضعف في التحصيل الدراسي للأفراد (علام، 2001). والاختبارات محكية المرجع (Criterion - referenced Test) (CRT) وفيها يقارن أداء الطالب بمستوى أداء معين يحدد بصرف النظر عن أداء المجموعة (عودة، 2010). ويتطلب الاختبار محكي المرجع تحديد مستويات مسبقة للأداء، كأن يجب الطالب على 80% من أسئلة الاختبار على الأقل، وذلك يفيد في تحديد مستويات إتقان كل فرد لأهداف معينة مرتبطة بمحتوى دراسي أو تدريبي، ويساعد هذا النوع من الاختبارات في اتخاذ قرارات تعليمية تختلف عن القرارات التي تتخذ في حالة الاختبارات المعيارية المرجع، ففي الاختبارات محكية المرجع نهتم باتخاذ قرار على مستوى أداء كل فرد بالنسبة إلى مجموعة من المهام أو المهارات أو المعارف التي تشكل النطاق السلوكي الذي يقيسه الاختبار، وهذا النطاق السلوكي الذي يقيسه الاختبار يكون المحك (Criterion) الذي ينسب إليه أداء الفرد (علام، 2001).

وتختلف الاختبارات محكية المرجع عن اختبارات معيارية المرجع من جوانب عدة، فمن حيث الهدف الرئيس للاختبار، فإن الاختبار محكي المرجع يهدف إلى مقارنة أداء المفحوص بمحك محدد مسبقاً لمعرفة الأهداف التي حققها المفحوص وتلك التي لم يحققها، أما الاختبار معياري المرجع فيهدف إلى مقارنة أداء المفحوص لأفراد مجموعته من خلال تحديد الأسئلة التي أجاب عنها المفحوص إجابة صحيحة. كما أن الدرجة في الاختبار محكي المرجع تفسر بناءً على درجة القطع، بينما تفسر الدرجة في الاختبار معياري المرجع بناءً على درجات معيارية تمثل موقع المفحوص في مجموعته. وتبنى فقرات الاختبار محكي المرجع بالاعتماد على مدى تحقيق المفردة للهدف السلوكي التعليمي، وبالمقابل فإن

المخلص:

هدفت هذه الدراسة إلى بناء اختبار محكي المرجع في الرياضيات وفق النموذج اللوجستي ثلاثي المعلمة، وجرى بناء اختبار تحصيلي مكون من (35) فقرة من نوع الاختيار من متعدد بأربعة بدائل، وطبق الاختبار على عينة الدراسة وعددها (968) طالباً وطالبة. وحذفت فقرة واحدة كان تمييزها سالبا، وأظهرت النتائج عدم مطابقة (3) فقرات للنموذج المستخدم. وبلغت قيمة معامل لفينجستون 0.83. وبلغ معامل الثبات الإمبريقي للاختبار (0.833). كما تحققت افتراضات النموذج المستخدم، وأظهرت نتائج الدراسة أن تقديرات معالم الفقرات (الصعوبة، والتمييز، والتخمين) كانت مقبولة. وحددت درجة القطع بطريقة أنجوف وكانت تساوي (22). وأوصى الباحث باستخدام الاختبار أو أي اختبار جزئي منه من قبل معلمي الرياضيات، ومن قبل الباحثين، وأوصى بإفادة من المنهجية المستخدمة في بناء الاختبار لبناء اختبارات لمقررات أخرى.

الكلمات المفتاحية: اختبار رياضيات، النهايات والاتصال، اختبار محكي المرجع، درجة القطع، نظرية الاستجابة للفقرة، النموذج ثلاثي المعالم.

Abstract:

This study aimed to construct a criterion - referenced test in mathematics using the Three Parameters Logistic Model. To achieve this aim, 35 multiple - choice items with four alternatives were prepared. A sample of 968 students responded to the items. One item with a negative discrimination was deleted. In addition, 3 items did not fit the used model. The value of Livingstone's coefficient was equal to 0.83. Test empirical reliability coefficient was 0.833. The assumptions of the model were verified. The results also indicated that items parameters estimations comprising difficulty, discrimination, and guessing were acceptable within the indexes available in educational measurement literature. Cut off score using the Angoff method was equal to 22. The researcher recommended taking advantage of the methodology used in building the test to build tests for other courses.

Keywords: Math test, limits and continuity, criterion - referenced test, cutoff score, item response theory.

المقدمة:

يعد التقويم أحد أهم عناصر المنهاج، ويشير الحكمي (2007) إلى بروز توجهات عديدة في مختلف أنحاء العالم من أجل إصلاح عمليات التقويم، وجعلها مستمرة مع استخدام أساليب متعددة، وتطبيق اختبارات معدة بإحكام على المستويين الوطني والعالمي.

(جدول المواصفات). يلي ذلك كتابة فقرات الاختبار لتقيس الأهداف السلوكية والمواقف السلوكية المحددة في المجال السلوكي، ويراعى في كتابة فقرات الاختبار بعض القواعد العامة مثل: قياس الكفاية أو المهارة التي حددت بأبعادها كلها. وكتابة عدد أكبر مما هو مطلوب من الفقرات، لأن بعضها سيستبعد. وصياغتها بأسلوب واضح وبسيط، وبلغة سليمة، وبإيجاز غير مخل. وجعلها مستقلة عن بعضها بعضاً، أي لا يعتمد إجابة إحدى الفقرات على إجابة فقرة أخرى. وتحديد تعليمات الإجابة بوضوح للمفوضين. وأخيراً مراجعة الفقرات وتدقيقها بعناية. أما الخطوة التالية فهي تحليل فقرات الاختبار، والتأكد من مطابقتها للمجال لسلكي للاختبار استرشاداً بأحكام الخبراء وتحليلهم المنطقي لمحتوى الفقرات في ضوء مواصفات المجال المحدد. يلي ذلك تحديد مستوى الأداء المقبول (المحك) في الاختبار: والذي يمكن أن يكون وقتاً محدداً، أو درجة أداء، أو عدد أخطاء مسموح به. والذي يمكن تحديده بطرق عديدة، يمكن حصرها في ثلاث مجموعات. والخطوة الأخيرة هي التجريب الميداني للفقرات، ويتطلب ذلك تطبيق الاختبار الذي يتكون من هذه الفقرات على عينة من المفوضين.

وهناك طرق عدة لتحديد درجة القطع (مستوى الإتقان، منها:

الطرق التحكيمية: يكون الاعتماد على آراء المحكمين المختصين في مجال محتوى الاختبار، أو خبراء القياس والتقويم، والاعتماد على آرائهم في تحديد المستوى الأدنى من الكفاية للمفوض. ومن هذه الطرق: طريقة إيبيل Ebel's method، وطريقة أنجوف Angoffs method، وطريقة ندلسكي Nedlesky's Method، وطريقة ندلسكاي - أنجوف Angoff Nedelsky combination method، وطريقة تقدير أهمية الصعوبة difficulty importance method.

والطرق التي تعتمد جزئياً على التحكيم، وتسترشد ببيانات تجريبية: أدرك الباحثون قصوراً ملحوظاً في الطرق التحكيمية، ومن ذلك القصور اعتمادها تحكيم مفردات الاختبار دون النظر إلى الأداء الفعلي للمختبرين مما يؤدي إلى الحصول على معايير غير متميزة بدرجة عالية من الصدق. (علام، 2005).

اقترح علماء القياس العديد من الطرق المختلفة لتحديد مستويات الأداء التي تتضمن طرقاً تجريبية محاولين في ذلك تجاوز عيوب الطرق التحكيمية، والحصول على مستويات أداء دقيقة يمكن الاعتماد عليها، ومن هذه الطرق: الطريقة التحكيمية المعززة بالمعلومات Informed Judgment method، وطريقة أنجوف التجريبية المعدلة Modified Angoff Empirical method، وطريقة توافق بين الطرق المطلقة والطرق النسبية Absolute Relative Compromise method. وأخيراً طرق تعتمد على البيانات التجريبية وتسترشد بالتحكيم Empirical Judgmental: ومن هذه الطرق: طريقة المجموعات المحكية Criterion Groups، وطريقة المجموعات المتضادة Contrasting Groups method، وطريقة المجموعة الحدية Borderline Group، وطريقة المحك معياري 1 - لمرجع Normed Referenced Criterion method، وطريقة الناتج التربوي Educational Consequences method، وطريقة إمريك لتقييم اختبار التمكن Emrick's Mastery testing Evaluation. (علام، 2005).

بناءها في الاختبار معياري المرجع يعتمد على تباين الدرجات ومعامل التمييز، وتجنب الفقرات السهلة جداً، والصعبة جداً. يسعى الاختبار محكي المرجع إلى تحديد جوانب الضعف والقوة في أداء المفوضين والبرامج التعليمية، بينما يسعى الاختبار معياري المرجع إلى تقويم أداء الطالب ومقارنته مع أداء الطلاب الآخرين.

ذكر علام، (2001) مجموعة من المواقف التي تستخدم فيها الاختبارات محكية المرجع هي:

تقويم أداء الفرد (المفوض) كل حسب قدراته واستعداداته.

قياس إتقان (تمكن) المهارات يسعى أو الكفايات الأساسية لمجال معين سواء كان مجالات تعليمية، أو مهنية. وتحديد مستوى إتقان الأفراد (المفوضين) للمتطلبات المحددة مسبقاً للبدء بدرس جديد، أو مادة جديدة. كذلك الحكم على سلوك الفرد (المفوض) بالنسبة لقدراته واستعداداته وإمكاناته الذاتية، لا بالنسبة لقدرات وإمكانات الآخرين، وهذا يمثل قيمة تربوية كبيرة لتعديل وتغيير في بعض البرامج التربوية والتعليمية.

تتميز الاختبارات مرجعية المحك عن الاختبارات معيارية المحك بالمميزات الآتية: (علام، 2001).

الاختبارات مرجعية المحك تعمل على تحديد مستوى أداء الفرد (المفوض) في ضوء مستويات معينة من الإتقان (التمكن) محددة مسبقاً. وأن الاختبارات مرجعية المحك تركز على مجموعة محددة ومعينة من سلوكيات المتعلم مما يجعل المدرس قادراً على أن يحدد إتقان طلبته مهارة أو كفاية، أو عدم إتقانهم كما أن فقرات الاختبارات مرجعية المحك تكون ذات صعوبة تتناسب مع المهام التعليمية المطلوبة والمحددة مسبقاً.

درجات الاختبارات مرجعية المحك تكون على بنسب مئوية من الإجابة الصحيحة، أو على بمؤشر يدل مدى التمكن أو الإتقان من التعلم أو من عدمه.

إضافة إلى أن الاختبارات مرجعية المحك ترتبط بالتعليم الفردي، أو التدريس بمجموعات صغيرة (Small groups). كما أن الاختبارات مرجعية المحك تستعمل في الاختبارات التحصيلية فقط، وليس في الاختبارات النفسية. ومن مميزات أن مدى التباين (الفروق الفردية أو الاختلافات) يصبح أقل ما يمكن أن يكون عليه عندما يتقن (يتمكن) الطلبة جميعاً، العناصر الأساسية للموقف التعليمي. بالإضافة إلى ذلك فإن تفسير درجة الفرد (المفوض) يتحقق بالرجوع إلى الدرجة الفاصلة، أو درجة القطع، وليس بالرجوع إلى درجات الزملاء، أو الأقران في الصف ذاته، لأن أداء كل فرد (مفوض) يُقارن بمستوى أداء محدد ومعين مسبقاً، وليس له علاقة بأداء زملائه في الصف الدراسي الذي ينتمي إليه.

وذكر عودة (2010) خطوات لبناء الاختبار محكي المرجع: فالخطوة الأولى هي تحديد الكفايات الرئيسية المرجو تحققها، وتحديد مجالها السلوكي. يجري بعد ذلك تحليل الكفايات الرئيسية إلى مكوناتها تحليلياً وإجرائياً، وترتيب مكوناتها ترتيباً منطقياً يكشف عن العلاقات بينها (أي تحليل السلوك إلى مكوناتها البسيطة). والخطوة التالية صياغة الأهداف السلوكية في ضوء نواتج تحليل الكفايات، وتحديد السلوك الذي يحتويه المجال لسلكي المراد قياسه تحديداً تفصيلياً. وإعداد المواصفات التفصيلية للاختبار

القياس والتقويم النفسي والتربوي، النموذج اللوغاريتمي أحادي المعلم، والنموذج اللوغاريتمي ثنائي المعلم، والنموذج اللوغاريتمي ثلاثي المعلم.

وهذه النماذج تصف التفاعل بين قدرة الفرد الممتحن ومفردات الاختبار من خلال دالة احتمالية تربط بين معلمين إحداهما يتعلق بالفرد يعرف بمعلم القدرة (0)، والأخر يتعلق بخصائص المفردة الاختبارية المتمثلة في: معلم الصعوبة (b)، ومعلم التمييز (a)، ومعلم التخمين (ع)، ومن ثم تهدف هذه النظرية إلى الوصول إلى قيم تقديرية عديدة واحدة تتعلق بالفرد وتسمى بمعلم القدرة أو السمة المقاسة لدى الفرد "Ability Parameter"، وقيمة تقديرية عديدة واحدة أو أكثر تتعلق بالمفردة الاختبارية، وتسمى بمعلم أو معلمات المفردة "Item Parameter".

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, 1989).

وتفترض نظرية الاستجابة للفقرة مجموعة من الافتراضات وهي: أحادية البعد (Unidimensionality) وتعني أن أداء الفرد المفحوص على الاختبار يمكن أن يُعزى إلى سمة مهيمنة، أو قدرة واحدة فقط، وتفترض بعض نماذج السمات الكامنة (Latent - trait models) وجود سمة واحدة تكمن وراء تفسير أداء الفرد المفحوص على الاختبار، وهذا دعا لتسمية مثل هذه النماذج النظرية بالنماذج أحادية البعد، ويتحقق هذا الافتراض عندما تقيس فقرات الاختبار سمة أو قدرة واحدة فقط، وعندها يمكن القول إن إجابة الفرد المفحوص تُعزى لهذه القدرة التي تقيسها فقرات الاختبار، وهناك العديد من الإجراءات والطرق الإحصائية، التي يجري التحقق من افتراض أحادية البعد لمجموعة من الفقرات في اختبار ما. ومن الجدير بالذكر أنه من الممكن للاختبار أن يكون أحادي البعد في مجتمع ما، ومتعدد الأبعاد في مجتمع آخر؛ وذلك عندما تكون فقراته متحيّزة.

والافتراض الثاني هو الاستقلال الموضعي للفقرات (Local Item Independence): ويعني هذا الافتراض أن استجابات الأفراد المفحوصين على الفقرات المختلفة للاختبار مستقلة إحصائياً (Statistically Independent)؛ بمعنى ألا يؤثر أداء الفرد المفحوص على فقرة ما في الاختبار سلباً أو إيجاباً على استجابته على أي فقرة أخرى من فقرات الاختبار، وعند تحقق الاستقلال الموضعي لفقرات اختبار ثنائي الاستجابة (بحيث تعطى الاستجابة الصحيحة القيمة 1، والاستجابة الخاطئة القيمة 0)؛ يكون احتمال الحصول على نمط معين لاستجابة فرد مفحوص يجري اختياره عشوائياً على فقرات الاختبار مساوياً لحاصل ضرب احتمالات استجابته لكل فقرة من فقرات ذلك الاختبار وفق النمط موضع الاهتمام. على سبيل المثال، هب أن هناك اختباراً مكوناً من خمس فقرات، وحصل فرد مفحوص ما على العلامة (3) من (5)، وكان نمط استجابته (Response Pattern) على الفقرات الخمس كالآتي: (10110)

وتمثل درجات القطع المحور الأساسي في كل الاختبارات محكية المرجع التي تستخدم في اتخاذ قرارات تعليمية تتعلق بدرجة إتقان الفرد المجال السلوكي من المعارف والمهارات لمحددة، ومستوى التمكن المطلوب من محتوى هذا المجال (Do-main). وتستخدم درجات القطع في هذه الاختبارات لتصنيف الأفراد إلى حالتين من التمكن (متمكن، غير متمكن)، واتخاذ قرارات التصنيف ينبغي أن تعتمد على تحديد مسبق لمستويات الأداء في الاختبار (علام، 2005).

طور أنجوف Angoff عام 1971 طريقه عرفت باسمه (طريقه انجوف Angoff Method) لحساب درجة القطع. وتعد من أسهل الطرق في التطبيق وأكثرها شيوعاً، وتناسب أنواع الاختبارات. ففي هذه الطريقة يجري تحديد درجة القطع التي طورها أنجوف Angoff عام 1971، وتعد من أسهل الطرق في التطبيق وأكثرها شيوعاً، وتناسب أنواع الاختبارات. ففي هذه الطريقة يجري تحديد درجة القطع كالتالي: يطلب من كل محكم في هذه الطريقة فحص كل مفردة من مفردات الاختبار، ثم يطلب منهم تصور مجموعة من الأفراد الذين حققوا الحد الأدنى للكفاية التي يقيسها الاختبار، ثم يقدر نسبة عدد الأفراد الذين يحتمل أن يجيبوا إجابة صحيحة عن كل مفردة من مفرداته، ويمثل متوسط هذه النسب الحد الأدنى لمستوى الاجتياز في الاختبار. (الزيلي، 2014).

أما بالنسبة لنظرية الاستجابة للفقرة، فتشير الناغى (2011) أنه على الرغم من انتشار استخدام نظرية الاختبارات التقليدية في بناء وتحليل البيانات المشتقة من الاختبارات التحصيلية المختلفة، إلا أنه تبين قصور هذه النظرية في مواجهة هذه المشكلات، بالإضافة إلى قصورها في الوفاء بمتطلبات كثير من التطبيقات المعاصرة مثل: بناء بنوك الأسئلة، وبناء الاختبارات مرجعية المحك، والكشف عن دالة المفردة المميزة (تحيز المفردة)، وبناء الاختبارات الموائمة بالحاسب، ومعادلة درجات الاختبارات التحصيلية.

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

ويندرج في نظرية الاستجابة للمفردة مجموعة من النماذج الرياضية تعرف باسم نماذج الاستجابة للمفردة Item Response Models التي تحقق القياس الموضوعي، وتعالج كثيراً من مشكلات القياس النظرية والتطبيقية التي عجزت نظرية الاختبارات التقليدية عن مواجهتها، ولعل أكثر هذه النماذج استخداماً وانتشاراً في مجال

■ وأما النموذج اللوجستي ثلاثي المعلمة (3PLM): فتظهر في معادلاته الرياضية المعالم الثلاث الصعوبة (bj)، والتمييز (aj)، والتخمين (cj)، وتكون معادلاته على الصورة:

$$P_i(\theta) = c_i + \frac{1}{1 + e^{-1.7(\theta - b_i)}} \dots \dots \dots (1)$$

دراسة ويلسون وماكغليفييري (Wilson & Macgillivray, 2007) والتي هدفت قياس المهارات الرياضية باستخدام اختبار موضوعي (اختيار من متعدد) تكونت عينة الدراسة من (566) طالباً، ومعظم الطلبة أعطوا مهارات في الجبر والحساب تعتمد على المهارات الرياضية، وجرى استخدام نموذج راش عند تحليل البيانات لاعتماد الاختبار، ووصف مستويات الطلبة في المهارات الرياضية.

أظهرت النتائج أن درجات الطلبة على المقياس تعتمد على الأكبر سناً، ومعرفة بالمهارات الرياضية إضافة إلى مستوياتهم السابقة بالقدرات الرياضية، مفهوم الذات، والعمر، وأظهر النموذج تطابق في البيانات المجمعة من عينة الدراسة.

هدفت دراسة حمادنة (2009) إلى استخدام نظرية الاستجابة للمفردة في بناء اختبار محكي المرجع في الرياضيات، وفق النموذج اللوجستي ثلاثي المعلم. ولتحقيق هذا الهدف جرى بناء اختبار في الرياضيات يقيس تحصيل الطالب في الإحصاء، ومؤلف من (28) فقرة من نوع الاختيار من أربعة بدائل. تكونت عينة الدراسة من 411 طالباً وطالبة، من طلبة الصف الأول الثانوي العلمي. أظهرت نتائج الدراسة تحقق افتراضات نظرية الاستجابة للمفردة في بيانات الدراسة، ومطابقة الاستجابات عن 24 فقرة من فقرات الاختبار التوقعات النموذج اللوجستي ثلاثي المعلم، وحذف 4 فقرات لم تطابق النموذج. وبينت النتائج المتعلقة بتقدير معالم الفقرات (الصعوبة، والتمييز، والتخمين) أنها كانت مقبولة ضمن المحكات التي أوردتها أدبيات القياس التربوي.

وفي دراسة ادواردز والكوك (Alcock & Edwards, 2010) التي هدفت إلى تحليل نتائج اختبار للرياضيات تكون من (11) فقرة في بريطانيا، تكونت عينة الدراسة من (164) طالباً وطالبة من طلبة السنة الأولى في المرحلة الجامعية، طلب منهم ترتيب فقرات الاختبار وفق مستوى صعوبة كل فقرة، وجرى تحليل نتائج الدراسة وفق نموذج راش لتقييم مدى مطابقة الفقرات النموذج راش أحادي المعلمة، ودلت النتائج على مطابقة فقرات الاختبار لستة أشخاص فقط من عينة الدراسة بعد إعادة ترتيب فقرات الاختبار وفق نموذج راش.

وفي دراسة قام بها أون (Onn, 2013) هدفت إلى المقارنة بين النظرية التقليدية والحديثة في القياس من حيث عدد الفقرات المنتقاة ومعامل الثبات. ولتحقيق أهداف الدراسة جرى إعداد اختبار في مادة الفيزياء مكون من (50) فقرة من نوع الاختيار من متعدد جرى تطبيقه على عينة مكونة من (69) طالباً وطالبة من طلبة المدارس في نيجيريا. وجرى تحليل فقرات الاختبار باستخدام برنامج SPSS لتحليل الفقرات وفق النظرية التقليدية في القياس، واستخدام برنامج X - Calibreprogramme لتحليل الفقرات وفق

U، حيث U: تعني نمط الاستجابة على الفقرات الخمس، 1: تعني إجابة صحيحة للفقرة، 0: تعني إجابة خاطئة للفقرة، فإن احتمال الحصول على هذا النمط من الاستجابة للفرد المفحوص المقصود ذو القدرة المقدرة θ هو:

$$P(U|\theta) = P_1 \cdot (1 - P_2) \times P_3 \times P_4 \cdot (1 - P_5)$$

حيث:

الأرقام 1، 2، 3، 4، 5: تشير إلى الفقرات.

P_j احتمالية الإجابة على الفقرة j بشكل صحيح.

$1 - P_j$: احتمالية الإجابة على الفقرة j بشكل خاطئ.

أما الافتراض الثالث فهو: منحني خصائص الفقرة (Item Characteristic Curve، ICC): وهو عبارة عن اقتران رياضي (Mathematical Function) يربط بين احتمال الاستجابة الصحيحة على فقرة ما، وبين قدرة الفرد المفحوص المقياس بمجموعة فقرات الاختبار، وصيغة معادلة هذا الاقتران الرياضية تختلف باختلاف نوع النموذج المستخدم من حيث كونه أحادي المعلمة، أو ثنائي المعلمة، أو ثلاثي المعلمة كما سيتضح لاحقاً.

وعادةً يكون شكل منحني خصائص الفقرة تراكمياً صاعداً (Monotonic Increasing)، وهذا ما يدل على أن احتمال الاستجابة الصحيحة على الفقرة يجب أن يزداد بزيادة القدرة، ويمكن استخدام أشكال مثل هذه المنحنيات في تحديد نوع الأداء التفاضلي الذي يمكن أن تظهره الفقرة، من حيث كونه منتظماً أم غير منتظم، وسيجري الحديث عن أنواع الأداء التفاضلي هذه لاحقاً.

الافتراض الأخير هو: افتراض انتفاء السرعة (Non-Speed-Edness): والمقصود به أن الأفراد المفحوصين ممن فشلوا في إجابة فقرات الاختبار قد فعلوا ذلك بسبب قدرتهم المحدودة، وليس بسبب فشلهم في الوصول لجميع فقرات الاختبار نظراً لضيق الوقت. (Hambleton & Swaminathan, 1985)

تقسم نماذج نظرية استجابة الفقرة ثنائية الإجابة حسب عدد المعالم كما أوردتها أمبرينسون وريز (Embretson & Reise, 2000) إلى ثلاثة نماذج هي:

■ النموذج اللوجستي أحادي المعلمة (1PLM): تظهر في معادلاته الرياضية معلمة الصعوبة (bj) فقط، ويفترض هذا النموذج أن معلمة التخمين تساوي صفراً، وأن الفقرات جميعها لها معلمة التمييز نفسها وتكون معادلاته على الصورة:

$$P_i(\theta) = \frac{1}{1 + e^{-1.7(\theta - b_i)}} \dots \dots \dots (1)$$

■ والنموذج النموذج اللوجستي ثنائي المعلمة (2PLM): وتظهر في معادلاته الرياضية معلمة الصعوبة (bj)، والتمييز (aj) فقط، ويفترض أن معلمة التخمين تساوي صفراً، وتكون معادلاته على الصورة:

$$P_i(\theta) = \frac{1}{1 + e^{-1.7a_j(\theta - b_i)}} \dots \dots \dots (2)$$

التحقق من مطابقة عينة الدراسة وفقرات الاختبار لافتراضات نموذج راش، وحسبت معاملات الصعوبة لفقرات الاختبار باستخدام البرنامج الإحصائي (Bigsteps)، وتكون الاختبار بصورته النهائية من (39) فقرة، وأظهرت النتائج تمتع الاختبار بدلالات صدق وثبات مقبولة.

وهدفت دراسة فريجات وبني يس (2016) إلى تقدير الخصائص السيكومترية لاختبار محكي المرجع في الكيمياء وفق النظرية الحديثة في القياس النفسي والتربوي. ولتحقيق هدف هذه الدراسة بني اختبار تحصيلي في مادة الكيمياء مؤلف من (57) فقرة من نوع الاختيار من متعدد بأربع بدائل، وطبق الاختبار على عينة الدراسة البالغ عددها (390) طالبا وطالبة من طلاب الصف الأول الثانوي العلمي في مديرية التربية والتعليم الأولى في إربد.

أشارت نتائج الدراسة إلى مطابقة الاستجابات عن (47) فقرة من فقرات الاختبار الافتراضات النموذج أحادي المعلمة (نموذج راش)، وحذف (10) فقرات لم تطابق هذا النموذج، وكانت فقرات الاختبار بصورته النهائية ضمن حدود المطابقة بالنسبة لمؤثري متوسطات المربعات الداخلية والخارجية، وبلغ معامل ثبات الأفراد (0.89)، ومعامل ثبات الاختبار (0.97). وجرى التحقق من تمتع الاختبار بصدق المحتوى والصدق التلازمي. وأن الاختبار يقدم أكبر كمية من المعلومات عن أداء المفحوصين عند مستويات القدرة المتوسطة، ويقدم أقل كمية من المعلومات عند مستويات القدرة العالية والمنخفضة.

وفي السياق ذاته أجرى دراسة دبوس (2016) هدفت إلى استخدام نظرية الاستجابة للفقرة في بناء تجمع فقرات اختبار محكي المرجع في الرياضيات وفق النموذج اللوجستي ثنائي المعلم. ولتحقيق هذا الهدف جرى إعداد تجمع من الفقرات ثنائية التدرج مكون من (50) فقرة، وتجمع من الفقرات متعددة التدرج مكون من (10) فقرات. وقد تكونت عينة الدراسة من (502) طالبا وطالبة أجابوا عن الفقرات. وقد أظهرت النتائج تحقق افتراضات نظرية الاستجابة للفقرة. كما بينت نتائج التحليل فيما يتعلق بمطابقة الفقرات الثنائية التدرج للنموذج الثنائي المعلم أن الفقرات كانت مطابقة للنموذج ثنائي المعلم، وأن الفقرات المتعددة التدرج كانت مطابقة لنماذج الاستجابة المتدرجة GRM.

هدفت دراسة صباح (2017) إلى استخدام نظرية الاستجابة للفقرة في بناء اختبار محكي المرجع في اللغة الانجليزية (1) لطلبة جامعة القدس المفتوحة، وفق نموذج راش، ولتحقيق هذا الهدف جرى بناء اختبار تحصيلي في اللغة الانجليزية (1) مؤلفا من (45) فقرة من نوع الاختيار من متعدد من أربعة بدائل، يقيس تحصيل الطالب في اللغة الانجليزية (1)، وحذفت (5) فقرات. وأجمع المحكمون على أن هناك فقرات تؤدي غرضها، فأصبح الاختبار بصورته الأولية مكونا من (40) فقرة. وطبق على أفراد عينة الدراسة البالغ عددهم (956) طالبا وطالبة، من طلبة السنة الأولى في جامعة القدس المفتوحة ومن أبرز النتائج التي توصلت إليها الدراسة: تحقق افتراضات نظرية الاستجابة للفقرة في بيانات الدراسة، ومطابقة الاستجابات عن (38) فقرة من فقرات الاختبار الافتراضات نموذج راش، وحذف فقرتين، ولم تطابق النموذج، وحذف (63) فردا غير مطابقين للنموذج، كما بينت النتائج

النموذج اللوجستي ثنائي المعلمة في نظرية استجابة الفقرة. وقد أظهرت نتائج التحليل مطابقة (29) فقرة للنظرية التقليدية في القياس، ومطابقة (38) فقرة للنموذج اللوجستي ثنائي المعلمة. وبينت النتائج انخفاض معامل ثبات الاختبار في كلتا النظريتين، فبلغت قيمة معامل الثبات في النظرية التقليدية في القياس (0.49) بينما بلغت قيمة معامل الثبات وفقا للنموذج اللوجستي ثنائي المعلمة (0.67).

ومن الدراسات ذات الصلة ببناء الاختبارات محكية المرجع دراسة علي (2012) التي هدفت إلى تعرف فاعلية بناء اختبار محكي المرجع لمقرر القياس والتقويم في التربية وفق نموذج راش؛ وتكون الاختبار من (99) فقرة اختبارية من نوع الاختيار من متعدد وبأربعة بدائل، وتكونت عينة البحث من (426) طالبا وطالبة من طلبة كلية التربية في قسم معلم الصف من جامعة دمشق، الذين درسوا المقرر في العام الدراسي 2009 / 2010. وقسمت العينة إلى قسمين، وطبق الاختبار عليهما في وقت واحد. واستخدم برنامج (WINSTEPS) في تحليل البيانات الناتجة عن تطبيق الاختبار محكي المرجع وفق نموذج راش، وقد نتج عن هذا البرنامج حذف عشر فقرات بسبب وقوعها خارج حدود المطابقة التي حددت بين (0.7 - 1.3)، وأصبح العدد النهائي لفقرات الاختبار (89) فقرة، تقع ضمن حدود المطابقة الداخلية والخارجية. وبينت النتائج أن معامل الثبات للفقرات بلغ (0.89)، ومعامل ثبات الأفراد بلغ (0.94) وكلاهما مرتفع. كما بينت النتائج أن تدرج الاختبار يحقق افتراضات نظرية الاستجابة للفقرة.

وهدفت دراسة حجازي والخطيب (2014) إلى الكشف عن التوافق بين النظرية الكلاسيكية والنموذج ثنائي المعلمة في مطابقة فقرات اختبار محكي المرجع في المادة النظرية لأحكام التلاوة والتجويد. ولتحقيق ذلك جرى بناء اختبار محكي المرجع في المادة النظرية لأحكام التلاوة والتجويد تكون من (41) فقرة من نوع الاختيار من متعدد وبأربعة بدائل، وطبق على عينة من طلبة بلغ عددها (404) طالبا وطالبة اختيروا بالطريقة العشوائية الطبقية من (16) مركزا قرآنيا، (9) للذكور، و7 للإناث في العاصمة عمان للعام الدراسي 2011 / 2012.

أشارت نتائج الدراسة إلى مطابقة (40) فقرة للنظرية الكلاسيكية، و (39) فقرة للنموذج ثنائي المعلمة، كما تطابقت (39) فقرة مع كل من النظرية الكلاسيكية والنموذج ثنائي المعلمة. وقد بلغت قيمة معامل الثبات وفق النظرية الكلاسيكية (0.927)، بينما كانت وفق والنموذج ثنائي المعلمة (0.943). أما معامل الصدق فقد بلغت قيمته وفق النظرية الكلاسيكية (0.73)، بينما كانت وفق والنموذج ثنائي المعلمة (0.65). وأشارت النتائج إلى وجود فروق ذات دلالة إحصائية عند مستوى الدلالة ($\alpha = 0.05$) في تقدير معاملي الثبات وفق النظرية التقليدية والنموذج الثنائي المعلمة، ولصالح النموذج الثنائي المعلمة.

هدفت دراسة الطراونة (2016) إلى بناء اختبار محكي المرجع في الإحصاء التحليلي لطلبة الدراسات العليا في كليات العلوم التربوية في الجامعات الأردنية الحكومية باستخدام نظرية استجابة الفقرة، تكونت عينة الدراسة من (309) طالبا وطالبة من الماجستير في كليات العلوم التربوية في الجامعات الأردنية. جرى

وجرى التطبيق على عينة مكونة من (208) تلميذاً في الصف الأول الإعدادي، منهم (118) تلميذاً طبقت الصورة الاختبارية الأولى عليهم، و (140) تلميذاً، وطبقت الصورة الاختبارية الثانية عليهم، وجرى التحقق من افتراضات نظرية الاستجابة للمفردة، وتدريب الصورتين الاختباريتين، وأسفرت نتائج البحث عن: تحقق افتراضات نظرية الاستجابة للمفردة، وأن المفردات في المهارات الابتكارية في الصورة الاختبارية الأولى جاءت مطابقة، بينما في الصورة الاختبارية الثانية رفضت بعض المفردات، وكانت قيم تقديرات الخطأ المعياري تقل عند المنتصف، وتزداد عند الأطراف في المهارات الابتكارية الثلاثة، وذلك دليل على دقة القياس وثباته، وبينت الباحثة إمكانية سحب صور اختبارية متكافئة ومتعادلة القياس من بنك الأسئلة القياس المهارات الابتكارية في مادة العلوم للصف الأول الإعدادي، وأوصت الباحثة بضرورة تطويع الاختبارات محكية المرجع في قياس المهارات الابتكارية في مواد دراسية أخرى، وتدريبها باستخدام أحد نماذج نظرية الاستجابة للمفردة.

يتضح من الدراسات السابقة أهمية توظيف نظرية الاستجابة للفقرة (IRT) في بناء الاختبارات؛ لما تتمتع به من خصائص كموضوعية القياس واستقلال في القياس؛ إذ إن تقدير معالم المفردات (الفقرات) تكون مستقلة عن خصائص الأفراد (Person Free)، وتقدير قدرات الأفراد مستقلة عن عينة الفقرات (Item Free)؛

وأشارت نتائج دراسة حجازي والخطيب (2014)، إلى وجود فروق ذات دلالة إحصائية في تقدير معاملي الثبات ولصالح النموذج الثنائي المعلمة، مما يؤكد أهمية بناء اختبارات وفق نماذج نظرية الاستجابة للفقرة (IRT). وتبين تركيز أغلب الدراسات على استخدام نموذج واحد من نماذج نظرية الاستجابة للفقرة وهو النموذج الأحادي المعلمة، ما عدا دراسة حمادنة (2009) - وهي قديمة نوعاً ما - ودراسة دبوس (2016) التي استخدم فيها النموذج ثنائي المعلمة. كما يمكن ملاحظة ندرة الدراسات التي تناولت بناء اختبارات في موضوع الرياضيات وفق نماذج نظرية الاستجابة للفقرة، وخاصة موضوع النهايات والاتصال على الرغم من أن هذا الموضوع يعد من أهم موضوعات الرياضيات؛ فهو في الغالب يدرس للطلبة في نهاية المرحلة الثانوية التي غالباً ما تكون اختبارات وطنية، أي أنها اختبارات واسعة النطاق (Extensive tests) تعتمد عليها وزارات التعليم العالي في قبول الطلبة في الجامعات؛ ولذا فإن معلمة التمييز تعد مهمة هنا.

مشكلة الدراسة وأسئلتها:

على الرغم من التطورات المهمة التي حدثت في عملية التقويم، وبناء الاختبارات، والتوجه إلى الاختبارات القائمة على نظرية الاستجابة للفقرة (IRT) والاختبارات محكية المرجع، إلا أننا نجد أن هناك قصوراً في إعداد الاختبارات التحصيلية في الأردن، ولا سيما الاختبارات محكية المرجع والتي تعد الأفضل لتحديد مستوى الأداء. إذ لا تزال علامة القطع ثابتة عند العلامة (50) للنجاح في المواد المدرسية.

ومن خلال اطلاع الباحث على الدراسات السابقة وجد ندرة

المتعلقة بتقديرات معامل الصعوبة أنها كانت مقبولة. 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 $(\theta = b)$ 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.

وأجرى الزبيدي (2018) دراسة هدفت إلى استخدام النموذج الأحادي المعلم (راش) في بناء اختبار تحصيلي محكي المرجع في مقرر الاختبارات والمقاييس لطلبة الدبلوم التربوي في كلية التربية في جامعة الطائف، ولتحقيق هذا الهدف جرى بناء اختبار محكي المرجع في مقرر الاختبارات والمقاييس يقيس تحصيل الطلبة، ويقدم تقويمات موضوعية للمهارات المطلوب إتقانها، وهو مكون من (37) فقرة من نوع الاختيار من متعدد بأربع بدائل، وطبق الاختبار على عينة مكونة من (147) طالباً. وأشارت النتائج على ملائمة الاستجابات عن (37) فقرة وهي فقرات الاختبار بصورته النهائية لافتراضات نموذج راش، وقد بلغ معامل الثبات للأفراد (0.90) أما معامل الثبات للاختبار فبلغ (0.79). وأوصى الباحث باستخدام الاختبار.

وكان الهدف من دراسة العطوي والمسعودي (2019) بناء اختبار محكي المرجع لقياس كفايات الإحصاء لدى طلبة الدراسات العليا وفق نموذج راش، ولتحقيق أهداف البحث اعتمدت الباحثة المنهج الوصفي التحليلي؛ وجرى بناء اختبار محكي المرجع مكون من (47) فقرة، تغطي الكفايات الضرورية للإحصاء. وللتأكد من صدق وثبات الاختبار تم تطبيق الأداة على عينة استطلاعية عددها (30)، وجرى التحقق من صدق المحتوى، وصدق البناء. وبلغ معامل الثبات بطريقة ألفا كرونباخ (0.99)؛ مما يشير إلى أن الاختبار يتمتع بدرجة ثبات عالية، وفحصت الخصائص السيكومترية للفقرات من خلال برنامج (SPSS)، فكانت معاملات الصعوبة متوسطة، ومعاملات التمييز مناسبة. وأشارت نتائج البحث بعد تطبيقه على عينة البحث المكونة من (200) فرد، باستخدام برنامج (Bilog - mg3) إلى مطابقة (47) فقرة النموذج راش، ومطابقة بيانات (200) فرد للنموذج، وحددت درجة القطع بطريقة أنجوف التحكيمية، وبلغت قيمتها (32)، وحاز الاختبار على معامل ثبات مرتفع وفق نظرية القياس الحديثة (0.94)، وهذا مؤشر على أن الاختبار يتمتع بدرجة عالية من الدقة والموضوعية، ومن شأنه التمييز بين المتقنين وغير المتقنين لكفايات الإحصاء من طلبة الدراسات العليا. وخرج البحث بمجموعة من التوصيات؛ منها: استخدام الاختبار الذي جرى بناؤه لقياس كفايات الإحصاء لدى طلبة الدراسات العليا.

وهدف دراسة شحاتة (2020) إلى تطوير بنك أسئلة لقياس المهارات الابتكارية في مادة العلوم للصف الأول الإعدادي مستخدماً تكامل مدخلي القياس محكي المرجع ونظرية الاستجابة للمفردة في بناء وتدريب البنك، ولتحقيق هذا الهدف أعدت قائمة بالمهارات الابتكارية في مادة العلوم في وحدة الطاقة مكونة من (6) مهارات أساسية يندرج أسفلها (30) هدفاً سلوكياً، كما وبنيت صورتان اختباريتان، كل صورة مكونة من (39) مفردة،

للفقرة (IRT) يجعله أكثر دقة وموضوعية واستقلالية في القياس.

■ إمكانية الإفادة من الدراسة الحالية في التعرف على كيفية إعداد الاختبارات محكية المرجع باستخدام النموذج ثلاثي المعلمة.

■ قلة الدراسات المحلية والعربية التي تناولت هذا المجال، مما يجعل هذه الدراسة ذات أهمية.

■ مساعدة المعلم ومتخذي القرار في تحديد مستوى الإتقان الذي وصله المتعلم، وعلى إصدار أحكام موضوعية على مدى نجاح أساليب التدريس في تنظيم العملية التعليمية بتطبيق الاختبارات المبنية وفق نظرية الاستجابة للفقرة.

حدود الدراسة ومحدّداتها:

تقتصر هذه الدراسة على:

■ طلبة الصف الثاني الثانوي العلمي في المدارس الحكومية، والخاصة التابعة لمديرية التربية والتعليم للواء قسبة المفرق للعام الدراسي 2019 / 2020 م.

■ فقرات اختبار من نوع اختيار من متعدد، تعطى العلامة (1) للاستجابة الصحيحة على الفقرة، والعلامة (صفر) للاستجابة الخاطئة.

■ استخدام نموذج واحد من نماذج الاستجابة للفقرة هو النموذج ثلاثي المعلم في بناء اختبار محكي المرجع من نوع الاختيار من متعدد بأربعة بدائل في مبحث الرياضيات، وحدة النهايات والاتصال في كتاب الصف الثاني الثانوي العلمي في الأردن.

التعريفات الاصطلاحية والإجرائية:

تعتمد الدراسة التعريفات الآتية لمصطلحاتها:

◀ الاختبارات التحصيلية: (Achievement Tests) يعرف الاختبار التحصيلي أنه طريقة منظمة لتحديد مستوى تحصيل الطالب لمعلومات ومهارات في مادة دراسية جرى تعلمها مسبقاً بصفة رسمية، من خلال إجاباته عن عينة من الأسئلة (الفقرات) التي تمثل محتوى المادة الدراسية (عودة، 2010).

◀ الاختبار محكي المرجع: (Criterion Referenced Test) الاختبار الذي يقارن أداء طالب بمستوى أداء معين يجري تحديده بصرف النظر عن أداء المجموعة (عودة، 2010).

ويعرف إجرائياً في هذا البحث أنه مجموعة من الفقرات من نوع الاختيار من متعدد من أربعة بدائل (أحدها صحيح والبقية خاطئة)، أعدت وفق نظرية الاستجابة للفقرة، وتحليل نتائج استجابات الطلبة باستخدام النموذج ثلاثي (3PLM) المعلمة بما يحقق دقة وموضوعية قياس تحصيل الطلبة في مقرر موضوع النهايات.

◀ نظرية الاستجابة للفقرة: (Item response theory) نظرية حديثة في القياس النفسي والتربوي تحدد العلاقة بين أداء المفحوص والسمة الكامنة موضع القياس وفق دالة رياضية محددة، وتعتمد هذه النظرية على عدد من النماذج تسمى نماذج لسماات الكامنة (Latent trait models) التي تربط بين الأداء على الفقرة وقدرة المفحوص (Hambelton & Swaminathan, 1985).

في إعداد الاختبارات محكية المرجع للمواد المدرسية التي تبني وفق أسلوب صحيح؛ ويعزى ذلك إلى ندرة الخبرات العربية في مجال تطوير هذا النوع من الاختبارات، والتحليلات الإحصائية المناسبة لكل خطوة، وكيفية تفسيرها، وكيفية استخدامها، والقرارات التي يمكن أن تصدر عنها، فإعداد هذه الاختبارات يحتاج إلى العناية والدقة واتباع خطوات معينة في بنائها، وتضافر جهود فريق من الاختصاصيين في هذا المجال، وتوافر مراجع وأدلة لتصميمها وبنائها.

وجد الباحث إن أغلب الدراسات التي استخدمت نماذج نظرية الاستجابة للفقرة (IRT) قد استخدمت النموذج أحادي المعلمة (نموذج راش)؛ لذا جاءت فكرة استخدام أحد نماذج نظرية الاستجابة للفقرة (IRT) وهو النموذج الثلاثي المعلمة من أجل بناء اختبار تحصيلي في الرياضيات؛ لندرة مثل هذه الاختبارات في بيئتنا العربية عامة، والأردنية خاصة مما شكل دافعاً للباحث لإجراء مثل هذه الدراسة؛ إذ يتميز هذا النموذج عن نموذج راش أحادي المعلمة بأن هذا النموذج يقيس ثلاث معالم للفقرات وهي: معلمة الصعوبة، ومعلمة التمييز، ومعلمة التخمين، بينما يقيس نموذج راش معلمة واحدة للفقرات وهي معلمة الصعوبة. مما يؤكد تفضيل النموذج ثلاثي المعلمة على نموذج راش.

سعت الدراسة للإجابة عن الأسئلة الآتية:

● ما معالم فقرات الاختبار محكي في الرياضيات في وحدة النهايات للصف الثاني الثانوي العلمي (الصعوبة، التمييز، والتخمين) المقدره وفقاً للنموذج الثلاثي المعلمة للفقرات الاختبارية المطابقة للنموذج؟

● ما دلالات صدق وثبات الاختبار محكي في الرياضيات في وحدة النهايات للصف الثاني الثانوي العلمي وفقاً للنموذج الثلاثي المعلمة؟

● ما درجة القطع الفاصلة بين الطلبة المتقنين للكفايات الرياضية في وحدة النهايات للصف الثاني الثانوي العلمي، والذين لا يتقنون هذه الكفايات بناءً على الاختبار؟

● ما مستوى أداء المفحوصين للاختبار مقارنة بمستوى الإتقان؟

أهداف الدراسة:

هدفت هذه الدراسة إلى بناء اختبار محكي المرجع في الرياضيات وفق لنماذج نظرية الفقرة (النموذج ثلاثي المعلمة) يتصف بمعايير الاختبار الجيد من حيث خصائصه السيكومترية من صدق وثبات، وتتميز فقراته بمعاملات صعوبة، وتمييز وتخمين مناسبة لموضوع النهايات؛ وذلك من أجل قياس تحصيل طلبة وتصنيفهم وفق علاماتهم على هذا الاختبار إلى متمكنين أو غير متمكنين.

أهمية الدراسة:

- تعد الاختبارات التحصيلية أهم أدوات قياس التحصيل.
- أهمية الاختبارات محكية المرجع التي تقارن أداء الفرد بمحك أداء محدد مسبقاً.
- إن بناء أي اختبار تحصيلي وفق نماذج نظرية الاستجابة

لوحة النهايات والاتصال في كتاب الرياضيات للصف الثاني الثانوي العلمي، وتحديد الكفايات الرياضية المتضمنة فيها، وهذه الكفايات هي:

- مفهوم النهاية.
- قيمة نهاية اقتران عند عدد بيانياً.
- نظريات النهايات.
- نهايات الاقترانات النسبية والكسرية والمتشعبة والمثلثية.
- مفهوم اتصال الاقتران عند نقطة وعلى فترة.
- كيفية البحث في اتصال الاقتران عند نقطة وعلى فترة.

3. صياغة النتائج التعليمية الخاصة (الأهداف التفصيلية) للوحدة، وقد أعدت قائمة مكونة من خمسة عشر هدفاً تفصيلياً شاملة للموضوع، وهي:

يحدد نهاية اقتران من خلال رسم منحني قيم نقطة الاقتراب.	يحدد نهاية اقتران من خلال جدول يمثل قيم نقطة الاقتراب.
يحدد نهاية اقتران كسري.	يحدد نهاية اقتران من خلال نظريات النهايات.
يحدد نهاية اقتران متشعب.	يحدد نهاية اقتران كسري لاقتران دائري.
يحدد قيم مجاهيل في نهاية معطى قيمتها.	يحدد نهاية اقتران كسري من خلال معرفة نهاية معطاة تحوي الاقتران.
يبحث في اتصال اقتران عند نقطة قاعدته معطاه.	يحدد قيم مجاهيل في نهاية اقتران متشعب.
يستخدم نظريات الاتصال في البحث في اتصال اقتران عند نقطة.	يحدد قيم مجاهيل في اقتران متصل عند نقطة.
يحدد قيم مجاهيل في اقتران متصل على فترة.	يبحث في اتصال اقتران على فترة.
	يبحث في اتصال اقتران عند نقطة من خلال رسم منحني الاقتران.

4. تحديد المستوى العقلي لتلك النتائج بربط مستويات النتائج التعليمية الخاصة بالمحتوى الدراسي موضوع الاختبار، وعرضت الأهداف على مجموعة من المحكمين الميدانيين (مختصين في مجال أساليب تدريس الرياضيات، ومختصين في مجال القياس والتقييم، ومعلمين، ومشرفين تربويين) لإبداء الرأي حول مدى شمول الأهداف موضوع الوحدة الدراسية وطريقة صياغتها، وجرت التعديلات بناءً على ملاحظاتهم. ليتم بعد ذلك إعداد جدول مواصفات الاختبار الملحق رقم (أ) وتحكيه من قبل المحكمين للتحقيق من صدق البناء، والمذكورة أسماؤهم في الملحق رقم (ب).

5. صياغة فقرات الاختبار بصورته الأولية: بالاعتماد على جدول مواصفات الاختبار، وجرت كتابة فقرات الاختبار بصورتها الأولية حسب الأسس العلمية والفنية المتبعة في كتابة فقرات الاختبار، وتكون الاختبار بصورته الأولية من (35) فقرة من نوع الاختيار من متعدد بأربعة بدائل لكل فقرة، بديل واحد منها فقط صحيح. مع مراعاة قواعد صياغة هذا النوع من الفقرات، مع التحقق من أن تقيس كل فقرة هدفاً محدداً، وفقاً لقائمة الأهداف التفصيلية الخاصة بالوحدة الدراسية.

6. عرضت الصورة الأولية للاختبار على (10) محكمين من ذوي الخبرة والاختصاص، أستاذ جامعي، وثلاثة من مشرفي الرياضيات في وزارة التربية والتعليم، وستة مدرّسين ممن يدرون

النموذج ثلاثي المعلمة: هو أحد نماذج نظرية الاستجابة للفقرة ثنائية التدرج، أن يمكن لهذا النموذج تقدير أربع معلمات، هي: قدرة الفرد، ومعامل صعوبة الفقرة، ومعامل التمييز، ومعامل تخمين الفقرة (بني عطا، 2014).

الخصائص السيكومترية: يقصد بالخصائص السيكومترية وفق نظرية الاستجابة للفقرة: معالم الصعوبة والتمييز والتخمين للفقرات، ومعامل الصدق والثبات للاختبار (علام، 2005).

درجة القطع (Cut-off Score): 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 (علام، 2005).

منهجية الدراسة وإجراءاتها

يتبع هذا البحث المنهج الكمي التحليلي في تحليل الاختبار وفقراته وفق النموذج الثلاثي المعلمة من نماذج نظرية الاستجابة للفقرة، ولذا يعد هذا المنهج الأنسب استخداماً في البحث الحالي.

مجتمع الدراسة

تكون مجتمع الدراسة من طلبة الصف الثاني الثانوي العلمي في محافظة المفرق المدارس الحكومية والخاصة للعام الدراسي 2019/2020م.

عينة الدراسة

عينة المحكمين:

تشكّلت عينة المحكمين من (15) محكماً ومحكمة لتحكيم ملائمة أهداف الوحدة الدراسية وجدول مواصفات الاختبار، و (10) محكمين لتحكيم ملائمة فقرات الاختبار لقواعد صياغة فقرات الاختبار من متعدد.

عينة المفحوصين:

تشكّلت عينة المفحوصين من طلاب وطالبات الصف الثاني الثانوي العلمي في المدارس الحكومية والخاصة الثانوية التابعة لمديرية التربية والتعليم للواء قصبه المفرق، للعام الدراسي 2019/2020م. وقد بلغ عددهم (968) طالباً وطالبة، منهم (332) طالباً، و (636) طالبة.

أداة الدراسة

تمّ بناء اختبار تحصيلي في مبحث الرياضيات للصف الثاني الثانوي العلمي في وحدة «النهايات والاتصال» من كتاب الرياضيات المقرر في الفصل الدراسي الأول للعام الدراسي 2019/2020م، وجرى الاستعانة بدليل المعلم في ذلك المبحث، وفيما يأتي وصفاً للإجراءات التي اتبعت في إعداد وبناء الاختبار:

1. تحديد الغرض من الاختبار: والذي تمثّل بقياس تحصيل طلاب وطالبات الصف الثاني الثانوي العلمي للعام الدراسي 2019/2020م، في مبحث الرياضيات في وحدة «النهايات والاتصال». وتحديد درجة القطع للاختبار لتصنيف الطلبة (متقن، غير متقن) للمهارات الرياضية في الوحدة الدراسية.

2. تحديد النطاق السلوكي للاختبار: بعمل تحليل محتوى

بعد أن أصبح الاختبار بصورته الأولى والمكوّن من (35) جاهزاً للتطبيق، جرى تطبيقه على العينة الاستطلاعية المكونة من (35) طالباً، وذلك بعد الزيارة الأولى للمدرسة العينة الاستطلاعية، والتي هدفت إلى تعريف مدرّسي المبحث بالاختبار والإطلاع عليه، وآلية تطبيقه، وما يلزمها من إجراءات ومعلومات يستلزمها الاختبار، والتأكد من وضوح الصياغة اللغوية لل فقرات، وتحريّ الفقرات التي يجري التساؤل حولها للكشف عن وجود غموض فيها أم لا، بالإضافة إلى ضبط الزمن الذي يستغرقه الطلبة في إنهاء الاختبار، كما جرى الاتفاق على موعد محدد باليوم والحصة لتطبيق الاختبار يُعلم به الطلبة مسبقاً. وبناءً على عملية تطبيق الاختبار على العينة الاستطلاعية حدد الزمن المناسب للإجابة على فقرات الاختبار: فوجد أن معدّل (95%) من الطلبة تقريباً أنها الإجابة على فقرات الاختبار بعد مرور (70) دقيقة على بداية الاختبار، وعليه اعتمد زمن الاختبار بفقراته الخمس والثلاثين ليكون (70) دقيقة؛ أي ما معدّله دقيقتان لإجابة كل فقرة. كما تبين وضوح فقرات وتعليمات الاختبار للمفحوصين. وجرى التأكد من الإجابة عن الأسئلة قبل استلام الاختبار من الطلبة، وصححت فقراته. وفرغت استجابات أفراد العينة الاستطلاعية على فقرات الاختبار، وحسبت معاملات الصعوبة، ومعاملات التمييز لفقرات الاختبار وفقاً للنظرية الكلاسيكية، والجدول رقم (1) يمثل هذه القيم.

المبحث، وطلب منهم الإجابة عن استبانة للحكم على مدى ملائمة فقرات الاختبار. والملحق رقم (ج) يبين أسماء المحكمين. وأعدت استبانة للحكم على الاختبار، تكونت من قواعد صياغة فقرات الاختبار من متعدد. وقد وضع أمام كل فقرة مقياس تقدير متدرج من 1 - 3، وطلب منهم اقتراح التعديل الذي يراه مناسباً على الفقرة. والملحق رقم (د) يبين استبانة التحكيم التي قدمت للمحكمين المذكورين في الملحق (ج). وتراوحت قيم الأوساط الحسابية لتقديرات المحكمين لصياغة فقرات الاختبار بين 2.27 و3.0، وهي قيم مرتفعة مما يعني أن فقرات الاختبار مصاغة صياغة جيدة ومقبولة. ونسبة الاتفاق بين المحكمين زادت عن 80% في الفقرات، فكما يرى (عوده، 2010)، لم تحذف أي فقرة، ليبقى الاختبار بصورته الأولى بعد تعديلات المحكمين مكوّناً من (35) فقرة. وفي ضوء ملاحظات المحكمين ومناقشة مقترحاتهم جرت التعديلات اللازمة على فقرات الاختبار. وأعدت صياغة إحدى بدائل الفقرة رقم (17)، وأعدت صياغة متن الفقرتين رقم (1) و(20).

7. إعداد تعليمات الاختبار والإجابة النموذجية: جرى إعداد ورقة تعليمات الاختبار، وورقة الإجابة النموذجية.

8. التجريب الأولي للاختبار:

الجدول (1)

قيم معاملات الصعوبة ومعاملات التمييز لفقرات الاختبار

رقم الفقرة	معامل الصعوبة	معامل التمييز	رقم الفقرة	معامل الصعوبة	معامل التمييز	رقم الفقرة	معامل الصعوبة	معامل التمييز	رقم الفقرة	معامل الصعوبة	معامل التمييز
1	.91	.45	10	.86	.67	19	.74	.27	28	.60	.61
2	.91	.22	11	.80	.58	20	.71	.40	29	.51	.40
3	.80	.22	12	.86	.40	21	.69	.52	30	.29	.23
4	.60	.32	13	.71	.18	22	.80	.61	31	.74	.28
5	.29	.42	14	.83	.47	23	.60	.49	32	.22	.30
6	.86	.18	15	.49	.59	24	.66	.53	33	.23	.23
7	.46	.39	16	.74	.27	25	.86	.23	34	.27	.42
8	.89	.38	17	.86	.31	26	.63	.49	35	.66	.52
9	.83	-.09	18	.54	.475	27	.66	.47			

على مدى اتساق تصنيف الأفراد إلى متمكنين، وغير متمكنين باستخدام درجات الاختبار.

9. تطبيق الاختبار على عينة الدراسة:

لتحقيق أهداف الدراسة طبق الاختبار على عينة الدراسة، وجرى التأكد من الإجابة على الأسئلة قبل استلام الاختبار من الطلبة، وصححت فقراته تمهيداً لاستكمال بناء الاختبار وفق النموذج ثلاثي المعلمة. وفرغت استجابات أفراد العينة على فقرات الاختبار لاستكمال عملية تحليل البيانات والإجابة على أسئلة الدراسة.

التحقق من افتراضات نظرية الاستجابة للفقرة:

يتضح من الجدول رقم (1) أن قيم معاملات الصعوبة لفقرات الاختبار تراوحت بين 0.22 و0.91، بمتوسط حسابي 0.66 وانحراف معياري 0.21. كما يتضح من الجدول رقم (2) أن تمييز الفقرة رقم 9 يساوي (-0.09)، لذلك حذفت الفقرة، وتراوحت قيم معاملات التمييز لبقية الفقرات بين 0.18 و0.67، بمتوسط حسابي 0.4 وانحراف معياري 0.14.

كما وجد أن القيمة التقديرية لمعامل الثبات كودر - ريتشاردسون 0.85. ونظراً لأن هذه الطريقة تناسب الاختبارات معيارية المرجع، وتؤدي عادة إلى معاملات ثبات أقل نسبياً من غيرها من الطرق، فقد عد الباحث أن هذه القيمة مرتفعة ومقبولة.

ووجد أن قيمة معامل لفنجستون 0.83، وهذه القيمة تدل

حيث N: حجم العينة. والقيمة الناتجة (0.011) منخفضة وتكاد تقترب من الصفر، وهذا مؤشّر كاف على تحقق افتراض أحادية البعد (Fraser & McDonald, 1988).

الاستقلال الموضوعي: Local Independence

ينص هذا الافتراض على أن استجابات المفحوصين الذين لديهم القدرة نفسها لفقرة ما مستقلة إحصائياً، أي أن استجابة المفحوص على فقرة ما يجب ألا تؤثر في استجابته على أي فقرة أخرى، بمعنى أن إجابة الفقرة لا تُعطي تلميحات أو معيقات لإجابة فقرة أخرى. وهذا الافتراض يعني أن لا شيء يؤثر في التحصيل في الاختبار سوى قدرة المفحوص وخصائص الفقرة. ويشير هامبلتون وسواميناثان (Swaminathan & Hambleton, 1985) إلى أن هذا الافتراض يكافئ افتراض أحادية البعد، وهذا يعني أنه إذا تحقق افتراض أحادية البعد في الاختبار فإن الاختبار يحقق افتراض الاستقلال لموضوعي. وهذا ما حصل في الدراسة الحالية.

افتراض التحرر من السرعة في الأداء.

أشار الزبيدي، (2018) إلى أنه يمكن التأكد من أن الاختبار ليس اختبار سرعة من خلال فحص نسبة الذين أنهوا الاختبار، ففحص الفقرات التي لم يجب عنها الأفراد. ويرى هامبلتون Ham-bleton, 1994)، إذا كان (75%) من الأفراد أكملوا الإجابة على الاختبار، و (80%) من فقرات الإجابة قد تمت الإجابة عنها، فإن السرعة لن تعد في هذه الحالة عاملاً مهماً في الأداء على الاختبار، إذ إن نسبة الطلاب والطالبات الذين أكملوا هذا الاختبار هي (100%)، وأن نسبة الفقرات التي تمت الإجابة عنها أيضاً تساوي (100%)، فإن هذا الاختبار يقيس القوة، ولا يقيس السرعة.

وجرى التأكد من مطابقة الأفراد والفقرات للنموذج اللوجستي ثلاثي المعلمة (3PLM) باستخدام برنامج (BILOG - MG3) بعد تزويده بالملف الناتج من تحويل البيانات إلى صورة صفر وواحد، وجذب الأفراد الذين لم تتطابق استجاباتهم مع توقعات النموذج المستخدم، ويتحقق ذلك عندما تكون القيمة الاحتمالية لتقدير القدرة لديهم أقل من 0.01 (Fit Proba-0.01 < bility)، أو عند تعذر تقدير الخطأ المعياري لقدراتهم، وقد أظهرت نتائج التحليل عدم مطابقة (19) طالباً وطالبة للنموذج اللوجستي ثلاثي المعلمة كانت القيم الاحتمالية لتقدير القدرة لديهم أقل من (0.01).

وجرى فحص مطابقة فقرات الاختبار للنموذج اللوجستي ثلاثي المعلمة، فأعيد التحليل مرة ثانية بعد حذف استجابات الطلاب والطالبات غير المطابقين، وحذفت (3) فقرات لعدم مطابقتها للنموذج المستخدم، وهي الفقرات ذات الأرقام (15، 34، 35)، ولم تُطابق لأسباب تتعلق بنظرية الاستجابة للفقرة من حيث انخفاض احتمالية الخطأ لهذه الفقرات دون مستوى الدلالة المتبني في البرنامج، وبالباقي قيمته (0.01). وأظهرت نتائج التحليل مطابقة جميع الفقرات للنموذج المستخدم، ليبقى الاختبار بشكله النهائي مكوناً من (31) فقرة.

نتائج الدراسة ومناقشتها:

◀ النتائج المتعلقة بالسؤال الأول والذي ينص على: ما معالم فقرات الاختبار محكي في الرياضيات في وحدة النهايات للصف الثاني الثانوي العلمي (الصعوبة، التمييز، والتخمين) المقدره وفقاً للنموذج الثلاثي المعلمة للفقرات الاختبارية

تم التحقق من افتراضات النموذج اللوجستي ثلاثي المعلمة وهو أحادية البعد لما له من أثر في دقة التقديرات، باستخدام التحليل العاملي للاختبار باستخدام طريقة المكونات الرئيسية. (Analysis Principle Component) ويظهر جدول (2) قيم الجذر للكامن (Eigen Values)، ونسب التباين المفسر (Explained Vari-ance) للعوامل الخمسة الأولى للاختبار.

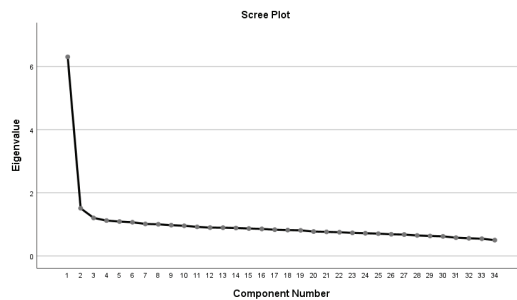
الجدول (2)

قيم الجذر الكامن ونسب التباين المفسر للعوامل الخمسة الأولى للاختبار

رقم العامل	الجذر الكامن	نسبة التباين المفسر %
1	6.306	18.547
2	1.514	4.452
3	1.206	3.548
4	1.123	3.303
5	1.045	3.112

يلاحظ من جدول (2)، أن قيمة الجذر الكامن للعامل (المكون) الأول (6.306)، وفسر ما نسبته (18.547%) من التباين الكلي، وبلغت نسبة قيمة الجذر الكامن للعامل الأول (المكون) إلى قيمة الجذر الكامن للعامل (المكون) الثاني (4.17)، ولأن هذه النسبة أكبر من (2) فإن هذا مؤشّر على أن العامل الأول هو عامل مسيطر، وهذا يؤكد تحقق افتراض أحادية البعد للاختبار، أي إن الاختبار يقيس سمة كامنة واحدة.

تطابقت نتائج اختبار فرز العوامل بالرسم البياني (Scree plot) الشكل (1) مع نتائج التحليل العاملي، فتبين من شكل (1) أن العامل الأول يمثل عاملاً سائداً، وأن المنحنى قد تغير اتجاهه، وانعطف انعطافاً ملحوظاً عند العامل الثاني، مما يرجح افتراض أحادية البعد لأغراض التحليل وفق نظرية الاستجابة للفقرة.



الشكل (1)

التمثيل البياني لقيم الجذور الكامنة للعوامل المكونة للاختبار

مؤشّر الجذر التربيعي للوسط الحسابي لمربعات البواقي: Root Mean Square of Residuals (RMSR)

وقد بلغت قيمته المحسوبة (0.011) تقريباً لدى مجموعة الطلبة، وهي أقل من القيمة الحرجة له البالغة قيمتها (0.1318)، والتي تحسب من خلال المعادلة:

$$C.V. = 4.1/\sqrt{N}$$

المطابقة للنموذج؟

جرى تقدير قيم معالم فقرات الاختبار (معلمة الصعوبة *b*، معلمة التمييز *a*، ومعلمة التخمين *c*) لمجموعة الطلبة، وفقاً للنموذج اللوجستي ثلاثي المعلمة (3PLM) في نظرية الاستجابة للفقرة، باستخدام برنامج (BILOG - MG3) والجدول (3) يبين قيم معالم هذه الفقرات.

الجدول (3)

القيم المقدرة لكل من معلمة الصعوبة والتمييز والتخمين لفقرات الاختبار وفقاً للنموذج ثلاثي المعلمة

رقم الفقرة	الصعوبة <i>b</i>	الخطأ المعياري للصعوبة	التمييز <i>a</i>	الخطأ المعياري للتمييز	التخمين <i>c</i>	الخطأ المعياري للتخمين
1	- 1.984	.505	.904	.158	.233	.107
2	- 1.092	.364	1.243	.232	.226	.095
3	- 1.242	.327	1.526	.278	.245	.099
4	.044	.184	1.241	.206	.248	.060
5	1.914	.242	1.416	.422	.147	.027
6	- 1.976	.418	1.207	.209	.248	.108
7	.344	.188	1.091	.193	.238	.057
8	- .919	.312	1.289	.22	.193	.089
10	- 1.330	.435	.963	.168	.240	.101
11	- 1.068	.383	.950	.160	.188	.094
12	- 1.117	.267	1.420	.223	.228	.088
13	.000	.323	.892	.157	.202	.078
14	- .918	.266	1.191	.176	.222	.082
16	.315	.188	1.451	.319	.249	.057
17	- .397	.158	1.897	.312	.249	.061
18	- .512	.227	1.256	.190	.207	.072
19	- .306	.170	1.244	.169	.213	.058
20	- .636	.197	1.515	.240	.239	.072
21	- .257	.158	1.400	.193	.231	.057
22	- .833	.152	1.998	.298	.241	.060
23	- .963	.119	1.898	.260	.207	.057
24	- .327	.246	1.203	.211	.228	.074
25	- .555	.168	1.770	.275	.209	.066
26	- 1.096	.139	1.901	.366	.241	.069
27	- .4.2	.139	1.727	.238	.242	.057
28	- .451	.130	1.832	.234	.231	.056
29	- .273	.159	1.613	.241	.249	.060
30	- .190	.137	1.622	.233	.232	.054
31	.044	.105	1.542	.236	.177	.039
32	- .862	.179	1.694	.219	.204	.071
33	1.681	.212	1.151	.305	.181	.036

رقم الفقرة	الصعوبة b	الخطأ المعياري للصعوبة	التمييز a	الخطأ المعياري للتمييز	التخمين c	الخطأ المعياري للتخمين
الوسط الحسابي	.009 -	.254	1.421	.245	.222	.073
الانحراف المعياري	.850	.105	.3.18	.062	.025	.020

■ ثانياً: صدق الاتساق الداخلي:

جرى التحقق من الاتساق الداخلي للمقياس من خلال حساب قيم معاملات الارتباط بين فقرات الاختبار بصورته النهائية بعد حذف الفقرات غير المطابقة للنموذج اللوجستي ثلاثي المعلمة والدرجة الكلية عليه، والجدول رقم (4) يبين معاملات الارتباط بين فقرات المقياس والدرجة الكلية عليه، وقد تراوحت قيم معاملات ارتباط فقرات المقياس بالدرجة الكلية عليه بين 0.13 إلى 0.59.

الجدول (4)

معاملات الارتباط بين فقرات المقياس والدرجة الكلية عليه

رقم الفقرة	معامل الارتباط	رقم الفقرة	معامل الارتباط	رقم الفقرة	معامل الارتباط	رقم الفقرة	معامل الارتباط
1	.29	10	.31	19	.47	27	.59
2	.13	11	.36	20	.49	28	.48
3	.38	12	.42	21	.57	29	.52
4	.42	13	.36	22	.59	30	.48
5	.28	14	.42	23	.41	31	.49
6	.27	16	.48	24	.51	32	.30
7	.42	17	.43	25	.57	33	.31
8	.40	18	.47	26	.53		

يتبين من خلال الجدول (4) أن قيم معاملات ارتباط فقرات المقياس بالدرجة الكلية كانت دالة إحصائياً عند مستوى الدلالة (0.01). وبذلك تحقق من صدق الاختبار منطقياً من خلال صدق المحتوى (اتفاق المحكمين) وإحصائياً من خلال الاتساق الداخلي.

يمكن اعتبار التحقق من أحادية البعد (Unidimensionality)، أن الاختبار يقيس سمة واحدة، وكون نسبة الاتفاق بين المحكمين حول قياس الفقرات للأهداف كانت عالية، فإن ذلك يعد مؤشراً آخر على صدق الاختبار، وأنه يقيس السمة التي أعد لها.

ثبات الاختبار:

بعد تطبيق الاختبار على العينة الاستطلاعية وجد أن القيمة التقديرية لمعامل الثبات كودر - ريتشاردسون 0.85. ونظراً لأن هذه الطريقة تناسب الاختبارات معيارية المرجع، وتؤدي عادة إلى معاملات ثبات أقل نسبياً من غيرها من الطرق، فقد عد الباحث أن هذه القيمة مرتفعة ومقبولة.

وعندما طبق الاختبار بصورته النهائية على عينة الدراسة المكونة من 968 طالباً أصبحت قيمة لمعامل الثبات كودر - ريتشاردسون 0.81، وهي قيمة مرتفعة تشير إلى ثبات الاختبار. إن الاختبارات مرجعية المحك تستخدم عادة لأغراض

يلاحظ من الجدول (3) أن قيم معلمة الصعوبة تراوحت ما بين (- 1.984) و (1.914)، وبمتوسط مقداره (- 0.499)، في حين أن قيم معلمة التمييز تراوحت ما بين (0.892) و (1.998)، وبمتوسط مقداره (1.421)، في حين تراوحت قيم معلمة التخمين ما بين (0.147) و (0.249)، وبمتوسط مقداره (0.222).

ويشير هامبلتون (Hambleton, 1989) إلى أنه عندما تتراوح قيم معلم الصعوبة بين 2 - 2، والتمييز بين 0.4 و 2 والتخمين بين صفر و 0.25، فإن المعامل تتمتع بالاتساق (Consistency). وعند مقارنة هذه القيم بنتائج الدراسة الحالية في الجدول 2، نجد بأن المعامل متسقة.

يتضح أيضاً أن مدى صعوبة الفقرات كان كبيراً، وأن قيم معاملات التمييز كانت مرتفعة، وهذا ما أشار إليه امبرتسون و رايس (Embretson & Reise, 2000) إلى أن منحني توزيع العلامات الخاص به منبسط نسبياً (Flat) على مدى متصل السمة، وهذا يعني أن مجموعة من الفقرات ذات تمييز عال، وبمدى واسع من معلم الصعوبة.

وفيما يتعلق بمعلمة التخمين فإن نتائج الدراسة الحالية تتفق مع القيم التي حصل عليها (حمادنة، 2009) فبلغت قيمة المتوسط الحسابي لمعلمة التخمين التي حصل عليها 0.264. وبهذا يمكن القول إن تقديرات معالم الفقرات في هذه الدراسة مقبولة ضمن محكات القياس التربوي والنفسي.

◀ النتائج المتعلقة بالسؤال الثاني والذي ينص على: ما

دلالات صدق وثبات الاختبار وفق النموذج المستخدم؟

يشير عيابنة (2009) إلى أن المبادئ الأساسية في صدق الاختبارات لا تختلف، سواء كان الاختبار معياري المرجع، أو محكي المرجع؛ إذ يشير مصطلح صدق الاختبار (Validity) إلى قياس الاختبار ما صمم لقياسه أصلاً، فالصدق عادة يرتبط بغرض الاختبار وبهدفه، وبالقرارات التي يجري اتخاذها بناء على درجات الطلبة على الاختبار، ولأن الاختبار يستخدم لأغراض مختلفة، فإن دلائل وجود صدق الاختبار تتنوع بتنوع أغراضه. ولقد استخدم المختصون بالقياس والتقويم مصطلحات عدة تتعلق بصدق الاختبار؛ منها: الصدق العاملي، والصدق التجريبي، والصدق التمييزي، إلا أن أحد أشهر أهم التقسيمات المستخدم في حالة الاختبارات معيارية المرجع هي صدق المحتوى، والصدق المرتبط بالمحك، والصدق البنائي.

■ أولاً: صدق المحتوى:

جرى التحقق من أن فقرات الاختبار تقيس المهارات الخاصة بالوحدة الدراسية من خلال حساب الأوساط الحسابية لتقديرات المحكمين التي تخص مدى ارتباط الفقرة بالهدف الذي تقيسه الفقرة، فبلغت قيمته 3 بانحراف معياري صفر، وهذا يدل على اتفاق المحكمين على أن الفقرات تقيس الأهداف التي تقيسها.

أسلوب أنجوف، وتطبيق مثال توضيحي على كيفية استخدام طريقة أنجوف. وجرى إعلام المحكمين بضرورة تخيل مجموعة مكونة من (100) طالب ممن يمتلكون الحد الأدنى من الكفاية (مستوى الأداء المقبول من وجهة نظرك) « كم منهم سيجيب عن المفردة إجابة صحيحة؟ » بمنحهم استمارة خاصة لتدوين تقديراتهم أمام كل مفردة من مفردات الاختبار، وجرى الحصول على مجموع القيم الاحتمالية التي قدرها كل محكم كل مفردة من مفردات الاختبار، وحسب الوسط الحسابي والانحراف المعياري. والجدول رقم (5) يبين نتائج تحديد درجة القطع بطريقة أنجوف.

الجدول (5)

تحديد درجة القطع بطريقة أنجوف

رقم الفقرة في الاختبار	متوسط تقديرات المحكمين	رقم الفقرة في الاختبار	متوسط تقديرات المحكمين	رقم الفقرة في الاختبار	متوسط تقديرات المحكمين
1	.90	13	.78	25	.72
2	.91	14	.69	26	.84
3	.62	16	.75	27	.72
4	.25	17	.73	28	.75
5	.92	18	.65	29	.71
6	.57	19	.64	30	.66
7	.81	20	.76	31	.50
8	.82	21	.61	32	.81
10	.85	22	.86	33	.34
11	.82	23	.81		
12	.86	24	.71		
			المجموع الكلي لمتوسط تقديرات المحكمين		.72
			درجة القطع		22/ 31

يتضح من الجدول (5) أن درجة القطع للاختبار المكون من الفقرات المطابقة للنموذج المستخدم من نماذج نظرية الاستجابة للفقرة تساوي (22)؛ أي أن الطالب الذي يحصل على العلامة (22) فأعلى يكون قد أتقن المهارات المستهدفة في هذه الوحدة الدراسية، وأنه وصل إلى مستوى التمكن من الأهداف التدريسية في الوحدة. وكانت درجة القطع جاءت أعلى من (50%) التي تؤكد النتيجة التي توصل إليها شانج (1999, change) المشار إليه في (أبو جراد، 2017) أن أسلوب أنجوف ينتج درجات قطع عالية وغالبا ما تكون أعلى من (50%)، وقد برر ذلك أن معظم المحكمين غالبا ما يفكرون في الطالب في مستوى المتوسط (أدأوه على الاختبار مقارنة بغيره يكون في وسط التوزيع) أو فوقه، ولا يفكر إلا القليل منهم في الطالب الذي يمتلك الحد الأدنى من الكفاية، كما أنهم يتأثرون بقدرات الطلبة الذين يدرسونهم وتوقعاتهم العالية لهم، مما يجعل تقديراتهم لدرجات القطع تميل نحو التشدد.

تصنيف الطلبة إلى (متقين و غير متقين)، وبيان درجة تمكنهم من المجال أو النطاق السلوكي المراد قياسه والمحدد مسبقا، لذا فان ثبات هذا النوع من الاختبارات يطلق عليه ثبات التصنيف الاختبارات الإتيقان، أو ثبات تقدير درجات الأفراد في نطاق سلوكي معين (Ojer,2010).

جرى التحقق من ثبات الاختبار باستخدام معامل لفينجستون (Livingston Index)، وهي إحدى طرق تقدير ثبات درجات المفحوصين في النطاق السلوكي، ويجري من خلال هذه الطريقة تعديل أو تصحيح معامل الثبات المحسوب بأي طريقة من طرق حساب الثبات المعروفة (كرونيباخ، كودر ريتشاردسون، ...) بما يتناسب مع المحك حسب المعادلة المشار إليها في (عودة، 2010):

$$L_i = \frac{\rho_{xx}\sigma_x^2 + (\bar{X} - c)^2}{\sigma_x^2 + (\bar{X} - c)^2}$$

حيث:

Li: معامل لفينجستون.

P_{xx}: باين الدرجات على الاختبار. \bar{X} : الوسط الحسابي للدرجات على الاختبار.

C: العلامة التي اعتمدت كمحك. (عودة، 2010)

وعند حساب قيمة معامل لفينجستون وجد أنها تساوي 0.83، وهذه القيمة تدل على مدى اتساق تصنيف الأفراد إلى متمكنين وغير متمكنين باستخدام درجات الاختبار.

ولأن مفهوم الثبات في نظرية الاستجابة للفقرة يرتبط بدالة معلومات الفقرة والاختبار، فقد حسبت معامل الثبات الأمبريقي (empirical reliability) المنبثق من نظرية الاستجابة للفقرة، والذي يقدر بإيجاد النسبة بين تباين الدرجة الحقيقية وتباين الدرجة الملاحظة، فتباين الدرجة الملاحظة يتكون من تباين الدرجة الحقيقية وتباين الخطأ، ويقدر تباين الدرجة المباشرة من خلال حساب تباين تقديرات قيم القدرة للمفحوصين، ويقدر تباين الخطأ بإيجاد دقة القياس الذي يسمى الانحراف المعياري البعدي (posterior standard deviation). وبلغت قيمة معامل الثبات الإمبريقي (التجريبي) للاختبار (0.833). وهذه القيمة مرتفعة وتقدم مؤشرا جيدا على توفر دلالة مقبولة لثبات الاختبار.

◀ النتائج المتعلقة بالسؤال الثالث والذي ينص على: ما درجة القطع الفاصلة بين الطلبة المتقنين للكفايات الرياضية في وحدة النهايات للصف الثاني الثانوي العلمي، والذين لا يتقنون هذه الكفايات بناء على الاختبار؟

وللإجابة عن هذا السؤال استخدم الباحث طريقة أنجوف لتحديد درجة القطع، فجرى الالتقاء بعينة المحكمين المكونة من محكمي ملائمة فقرات الاختبار لقواعد صياغة فقرات الاختبار من متعدد، وطلب منهم تحديد درجة القطع للاختبار بعد توضيح هدف الدراسة لهم، وتوضيح الأساس المنطقي لطريقة أنجوف التي يستخدمونها في تحديد الحد الأدنى لمستوى الأداء المقبول، وتطبيق مثال توضيحي على كيفية استخدام طريقة أنجوف. كما جرى تزويدهم بنشرة تفصيلية عن خطوات تحديد مستوى الأداء وفق

محكي المرجع مثل: (أثر عدد فقرات الاختبار محكي المرجع في دالة الاختبار للاختبار، ...).

المصادر والمراجع العربية:

- أبو جراد، حمدي. (2017). فاعلية النموذج اللوجستي ثالثي المعلمة في معايرة مفردات اختبار تحصيلي محكي المرجع في مقرر الرياضيات للصف السابع، مجلة اربد للبحوث والدراسات، 19 (1): 268 - 293.
- بني عطا، زايد. (2014). تقصي دقة تقدير النموذج اللوجستي ثلاثي المعلمة لعالم الفقرة وقدرة الأفراد في ضوء تغير طول الاختبار وحجم العينة: دراسة محاكاة، مجلة جامعة الشارقة للعلوم الإنسانية والاجتماعية، 11 (2): 1 - 37.
- حجازي، تغريد والخطيب، عبد الله. (2014). التوافق بين نظرية الاختبار الكلاسيكي ونموذجي المعلمات اللوجستي في تصنيف العناصر للاختبار المرجعي المعياري في أحكام التلاوة والتجويد، مجلة جامعة النجاح للبحوث - ب (العلوم الإنسانية)، 28 (10): 2270 - 2239.
- الحكمي، علي. (2007). تقويم التعليم وضمان الجودة. ورقة مقدمة الاجتماع السنوي الرابع عشر للجمعية السعودية للعلوم التربوية، جامعة القصيم، القصيم، السعودية 15 - 16 أبريل 2007م.
- حمادنة، إياد. (2009). استخدام نظرية الاستجابة للمفردة في بناء اختبار محكي المرجع في الرياضيات وفق النموذج اللوجستي ثلاثي المعلم، مجلة العلوم التربوية والنفسية، 10 (2): 215 - 238.
- دبوس، محمد. (2016). استخدام نظرية الاستجابة للعناصر في بناء تجمع العناصر في المعيار - الاختبار المرجعي مع العناصر ثنائية التفرع ومتعددة العناصر وفقاً للنموذج اللوجستي ذي المعاملين، مجلة جامعة النجاح للعلوم الإنسانية، 30 (7): 1453 - 1480.
- الزبيدي، محمد. (2018). بناء اختبار تحصيلي محكي المرجع في مقرر الاختبارات والمقاييس باستخدام النموذج الأحادي المعلم نموذج راش لطلبة الدبلوم التربوي في كلية التربية بجامعة الطائف، المجلة الدولية التربوية المتخصصة، 7 (2): 40 - 54.
- الزليعي، محمد. (2014). إنشاء اختبار مرجعي للمعايير التشخيصية لقياس مهارات البحث العلمي لطلبة الدراسات العليا في كلية التربية بجامعة الباحة، (رسالة ماجستير غير منشورة)، جامعة أم القرى، مكة، السعودية.
- شحاته، غادة. (2020). تكامل مدخلي القياس محكي المرجع ونظرية الاستجابة للمفردة في تقييم المهارات الابتكارية في مادة العلوم لدى تلاميذ الصف الأول الإعدادي، المجلة التربوية، جامعة سوهاج، 80 (2): 479 - 583.
- صباح، عبد الهادي. (2017). استخدام نظرية الاستجابة للفقرة في بناء اختبار محكي المرجع في اللغة الإنجليزية (1) لطلبة جامعة القدس المفتوحة وفق نموذج (راش)، مجلة جامعة القدس المفتوحة للأبحاث والدراسات التربوية والنفسية، 6 (20): 133 - 151.
- 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.

الناتج المتعلقة بالسؤال الرابع الذي ينص على: ما مستوى أداء المفحوصين للاختبار مقارنة بمستوى الإتقان؟ استخدم اختبار (t) لعينة واحدة، والجدول رقم (6) يبين نتائج الاختبار.

الجدول (6)

نتائج اختبار t لمقارنة متوسط أداء المفحوصين بدرجة القطع

المتوسط الحسابي	الانحراف المعياري	درجة القطع	قيمة t	درجة الحرية	مستوى الدلالة
22.9	5.94	22	10.17	967	.00

يتضح من الجدول رقم (6) أن هناك فروقا دالة إحصائية بين متوسط أداء الطلبة، ودرجة القطع مما يدل على زيادة نسبة المتمكنين من المهارات المتضمنة في الاختبار.

مستخلص للنتائج

تبين من نتائج تحليل البيانات النتائج الآتية:

1. مطابقة 31 فقرة من فقرات الاختبار الذي جرى بناؤه للنموذج ثلاثي المعلمة.
2. تمتعت تقديرات معالم الفقرات بقيم مقبولة ضمن محكات القياس التربوي والنفسية.
3. تحقق نوعان من الصدق لأداة الدراسة (الاختبار) هما: الصدق المنطقي من خلال صدق المحتوى (اتفاق المحكمين). وإحصائياً من خلال الاتساق الداخلي.
4. أظهرت النتائج أن الاختبار يتمتع بقدر عالٍ من الثبات إذ تبين أن قيمة لمعامل الثبات كودر - ريتشاردسون للصورة النهائية للاختبار كانت 0.81 بينما كانت قيمة معامل لفنجستون تساوي 0.83 وهذه القيمة تدل على مدى اتساق تصنيف الأفراد إلى متمكنين، وغير متمكنين باستخدام درجات الاختبار.
5. تبين من النتائج أن درجة القطع للاختبار المكون من الفقرات المطابقة للنموذج المستخدم من نماذج نظرية الاستجابة للفقرة تساوي (22)؛ أي أن الطالب الذي يحصل على العلامة (22) فأعلى يكون قد أتقن المهارات المستهدفة.
6. كما أظهرت النتائج ارتفاع نسبة المتمكنين من المهارات المتضمنة في الاختبار.

التوصيات:

- في ضوء النتائج التي توصلت إليها الدراسة الحالية، يوصي الباحث ما يلي:
1. استخدام الاختبار أو أي اختبار جزئي منه من قبل معلمي ومعلمات الرياضيات، الذي جرى التأكد من خصائصه السيكومترية.
 2. الاستفادة من المنهجية المستخدمة في بناء الاختبار لبناء اختبارات لمقررات أخرى.
 3. استخدام الاختبار أو أي اختبار جزئي منه من قبل الباحثين في دراسة أثر أي متغير على متغير آخر يتعلق باختبار

- test length and sample size: a simulation study, *University of Sharjah Journal of Humanities and Social Sciences*, 11 (2): 1 - 37.
- Dabous, M. (2016). *Using Item Response Theory in Constructing the Item Pool in Criterion - Referenced Testing with Dichotomous and polychromous Items according to Two - Parameter Logistic Model*, *Najah University Journal of Human Sciences*, 30 (7): 1453 - 1480.
- Fraihat, Ayman and Bani Yassin, Omar (2016). *Psychometric properties of the Criteria - referenced test in chemistry According to the modern theory of Psychological and educational measurement*. *University Studies Journal*, Ammar Thaleji. 42: 70 - 88.
- Al - Hakami, A. (2007). *Education Evaluation and Quality Assurance, Paper presented at the 14th annual meeting of the Saudi Society for Educational Sciences held in Qassim in the period 15 - 16 April 2007*.
- Hamadna, Iyad. (2009). *Using the vocabulary response theory in constructing a reference test in mathematics according to the three - teacher logistic model*, *Journal of Educational and Psychological Sciences*, 10 (2): 215 - 238.
- Hegazy, T. & Al - Khatib, A. (2014). *The compatibility between the theory of the classical test and the two models of logistic parameters in the classification of items for the standard reference test in the provisions of recitation and intonation*, *Najah University Journal of Research - B (Humanities)*, 28 (10), 2239 - 2270.
- Al - Naghi, H. (2011). *The effect of the number of test items on estimating individuals' abilities and the information function of a reference achievement test in mathematics using item response theory models*, *Journal of the Faculty of Education - Port Said University*, 10: 676 - 721.
- Odeh, A. (2010). *Measurement and Evaluation in the teaching process*, 4th ed. , Irbid: Dar Al - Amal for publication and distribution.
- Sabah, A. (2017). *The use of item response theory in constructing a Criteria - reference test in language English (1) for Al - Quds Open University students, according to the (Rush) model*, *Journal of Al - Quds Open University Educational and Psychological Research and Studies*, 6 (20): 133 - 151.
- Shehata, Ghada (2020). *The integration of the two approaches to measurement, the Criteria - reference and the theory of response to the singular in assessing the innovative skills in science for first - grade intermediate students*, *The Educational Journal*, Sohag University, 80 (2): 479 - 583.
- Al - Tarawneh, S. (2016). *Constructing Criteria - reference test in Analytical statistics for Higher Basic Students Classes in Jordan Using Item response theory*, *Educational and Psychological Scientific Journal*, 17 (3): 547 - 568.
- AL - Zayla'I, Mohammad. (2014). *Constructing a Diagnostic Criterion Referenced Test for the Measuring of Scientific Research Skills of the graduate Students in the Collage of Education at Al - Baha University, (Unpublished MA thesis)*, faculty of education, Umm Al - Qura University, Mecca, Saudi Arabia.
- Al Zubaidy, M. (2018). *Building Criterion - Reference Achievement Test for the Tests and Stands Course by Using the Monochrome Monochrome Model (Rash Model) for the Students of Educational Diploma in the Faculty of Education at Taif University*, *International Interdisciplinary Journal of*
- عباينة، عماد. (2009). الاختبارات محكية المرجع: فلسفتها وأسس تطويرها، ط2، عمان: دار المسيرة للنشر والتوزيع والطباعة.
- علام، صلاح الدين. (2001). الاختبارات التشخيصية المرجعية المعيارية في المجالات التربوية والنفسية، ط1، القاهرة: دار الفكر العربي.
- علام، صلاح الدين. (2000). الاختبارات والمقاييس التربوية والنفسية، ط1، القاهرة: دار الفكر العربي.
- علي، نداء. (2012). فاعلية استخدام نموذج راش في بناء اختبار مرجعي معياري في مقرر القياس والتقويم في التعليم (أطروحة دكتوراه غير منشورة)، كلية التربية، جامعة دمشق، دمشق، سوريا.
- العطوي، تهاني والمسعودي، أحمد. (2019). بناء اختبار محكي المرجع لقياس كفايات الإحصاء لدى طلبة الدراسات العليا بجامعة تبوك، المجلة الدولية للدراسات التربوية والنفسية، 1 (7): 97 - 123.
- عودة، أحمد. (2010). القياس والتقويم في العملية التعليمية. إربد: دار الأمل.
- فريحات، أيمن وبني يس، عمر. (2016). الخصائص السيكومترية لاختبار محكي المرجع في الكيمياء وفق النظرية الحديثة في القياس النفسي والتربوي، مجلة دراسات الجامعة عمار تلج، 42: 70 - 88.
- الناعى، هبة (2011). أثر عدد مفردات الاختبار على تقدير قدرات الأفراد ودالة المعلومات لاختبار تحصيلي مرجعي المحك في الرياضيات باستخدام نماذج نظرية الاستجابة للمفردة، مجلة كلية التربية - جامعة بورسعيد، 10: 676 - 721.

المصادر والمراجع العربية مترجمة:

- Abu Jarad, Hamdi. (2017). *The effectiveness of the three - parameter logistic model in calibrating the vocabulary of a criterion achievement test in the seventh grade mathematics course*, *Irbid Journal for Research and Studies*, 19 (1): 268 - 293.
- Ali, Neda'. (2012). *Effectiveness of Using Rasch Model in Constructing a Criterion Reference Test in Course of Measurement and Evaluation in Education (Unpublished Ph. D.)*, Faculty of Education, Damascus University, Damascus, Syria.
- Allam, S. (2001). *Criterion Reference Diagnostic Tests in Educational & Psychological Domains*, first Ed. , Cairo: Dar Al - Fikre Alarabi.
- Allam, S. (2005). *Models of uni and multi - dimensional in Item Response Theory, Principles and Application in Educational and Psychological Measurement*, first Ed. , Cairo: Dar El - fikr El - Arabi for publication and distribution.
- Ababneh, Imad. (2009). *Criterion - referenced tests: their philosophy and the foundations of their development*, 2nd Ed. , Amman: Dar Al Masirah for publication, distribution and printing.
- Al - Atwi, T. & Al - Masaudi, A. (2019). *Building a criterion - referenced test to measure students' statistical competencies Postgraduate studies at the University of Tabuk*, *International Journal of Educational and Psychological Studies*, 1 (7): 97 - 123.
- Bani Atta, Zayed. (2014). *Investigating the accuracy of estimating the three - parameter logistic model of the item world and the ability of individuals in light of the change in*

Education, 7 (2): 40 - 54.

المصادر والمراجع الأجنبية:

- Du Toit, M. (Ed.). 2003. *IRT from SSI. BILOG - MG, MULTILOG, PARSCALE, TESTFACT*, Lincolnwood, IL: Scientific Software International.
- Edwards. H & Alcock. L. (2010). *Using Rasch analysis to identify uncharacteristic responses to undergraduate assessment*, *Teaching Mathematics and Its Applications*, 29: 165 - 175.
- Embretson, S. & Reise, S. (2000). *Item response theory for psychologists*. New Jersey: Lawrence Erlbaum Associates.
- Fraser, C. , & McDonald, R. P. (1988). *NOHARM: least squares item factor analysis*. *Multivariate Behavior Research*, 23: 267 - 269.
- Hambleton, R. (1989). *Principles and selected applications of item response theory*. New York: Macmillan Publishing Company.
- Hambleton, R. (1994). *The rise and fall of criterion - referenced measurement*. *Educational Measurement*, 13 (4): 21 - 26.
- Hambleton, R. , & Swaminathan, H. (1985). *Item Response Theory: principles and Applications*. Boston: Kluwer-Nijhoff publishing.
- Hambleton, R. K. , & Swaminathan, H. (1989). *Item Response Theory Principles and Applications*. Kluwer - Nijh of publishing, Boston, U. S. A.
- Hambleton, K & Swaminathan, H. & Rogers, H. (1999). *Fundamentals of item response theory*. Newbury Park, CA: Sage Publication.
- Lord, F. & Stocking, M. (1989). *Item Response Theory*. New York: Pergamon Press.
- Ojer. L. (2010). *Measurement and Evaluation*. New York: Holt Rinehart and Winston.
- Onn, D. (2013). *Classical test theory versus Item response theory: An evaluation of the comparability of item analysis results*, *Joint Admission and Matriculation Board*. 1 - 23.
- Stocking. M. (1999). *Item response theory*. In: G. Masters, & J. Keevs (Eds). *Advances in Measurement in Educational Research and Assessment*. (322) the Boulevard, UK: Pergamon.
- Wilson, T. & Mac Gillivray, H. (2007) , *Counting on the Basics: Mathematical Skills among Tertiary Entrants*, Retrieved 27/ 10/ 2021 from *International Journal of Mathematical Education in Science & Technology*.