

بناء اختبار تحصيلي في الرياضيات وفق نظرية الاستجابة للفقرة (النموذج ثنائي المعلمة)

Constructing an achievement test in mathematics using item response theory (two parametric model)

مهدي عليمات

Mahdi Olimat

وزارة التربية والتعليم، المفرق، الأردن

Ministry of Education, Mafraq, Jordan

الباحث المراسل: mahdi.olimat@yahoo.com

تاريخ التسليم: (2019/1/29)، تاريخ القبول: (2019/8/5)

ملخص

هدفت هذه الدراسة إلى بناء اختبار في الرياضيات (وحدة التكامل) للصف الثاني الثانوي العلمي نظرية الاستجابة للفقرة (النموذج ثنائي المعلمة) حيث تمّ بناء اختبار تحصيلي مكوّن من (40) فقرة من نوع الاختيار من متعدد بأربعة بدائل، وطُبق الاختبار على عينة الدراسة الكلية البالغ عددها (679) طالباً وطالبة من طلبة الصف الثاني الثانوي العلمي في المدارس الحكومية الثانوية التابعة لمديرية تربية وتعليم لواء قصبه المفرق في الفصل الدراسي الثاني للعام الدراسي 2018/2017م. وحللت الفقرات باستعمال البرنامج المحوسب BILOG-MG3، حيث أظهرت نتائج التحليل عدم مطابقة (5) فقرات للنموذج المستخدم، ليصبح الاختبار يشكله النهائي مكوّنًا من (35) فقرة فقط. وبلغت قيمة معامل الثبات الإيميريقي للاختبار (0.82). كما تمّ التحقق من افتراضات النموذج المستخدم وفقاً لنظرية الاستجابة للفقرة، وهي: أحادية البعد (Unidimensionality)، باستخدام مؤشرين: مؤشر تانكا Tanaka's Index ومؤشر الجذر التربيعي للوسط الحسابي لمربعات البواقي (Root Mean Square of Residuals, RMSR). وأما افتراض الاستقلال الموضعي (Local Independence) فقد تمّ التحقق منه باستخدام مؤشر (Z_{Q_3}) باستخدام برنامج LDID (Local Dependence Indices for Dichotomous Items). وأوصى الباحث بأن تقوم وزارة التربية والتعليم بإنشاء بنك يحوي مثل هذه الاختبارات المقننة وتطبيقها من خلال المدارس. وأن يستخدم الاختبار من قبل معلمي ومعلمات الرياضيات ومن قبل الباحثين حيث تمّ التأكد من خصائصه السيكومترية، كما أوصى بالاستفادة من المنهجية المستخدمة في بناء الاختبار لبناء اختبارات لمقررات أخرى والتأكد من خصائصها السيكومترية.

الكلمات المفتاحية: التكامل، نظرية الاستجابة للفقرة، أحادية البعد، الاستقلال الموضعي.

Abstract

The objective of this study was to construct a test in mathematics (the integration unit) for the second secondary (science stream). The theory of response to the paragraph (the double parameter model), where the achievement test consisted of (40) paragraphs of the multiple choice type with four alternatives, the number of (679) students from the second secondary education in the secondary government schools belonging to the Mafraq Directorate in the second semester of the academic year 2017/2018. The results were analyzed using BILOG-MG3, where the results of the analysis showed that there was no match between the (5) paragraphs in the used model. The value of the empirical stability coefficient was (0.82). The model assumptions used in response to the paragraph were also verified: unidimensional; using two indicators: The Tanaka's Index and the square root index of the Root Mean Square of Residuals (RMSR). The Local Independence hypothesis was verified using the (Z_{Q_3}) indicator using the Local Dependence Indications for Dichotomous Items (LDID) program. The researcher recommended that the Ministry of Education establish a bank that includes such standardized tests and apply them through schools. The test should be used by the teachers and the mathematics teachers and by the researchers. The psychometric characteristics of the test should be used and the methodology used to construct the test should be used to construct tests for other courses and to confirm their psychometric characteristics.

Keywords: Integration, Item Response Theory, Unidimensionality, Local Dependence.

مقدمة

تعد الرياضيات من أصعب المواد الدراسية تعلماً وتعلماً لما تتصف به من تسلسل منطقي تجريد في المفاهيم والعلاقات وتراكم موضوعاتها ذات البنية المحكمة، أي إنه يصعب الوصول إلى مستوى دون المرور بالمستويات التي تسبقه، ومما يزيد من صعوبة تعليم وتعلم الرياضيات الاختلاف في القدرات ومستويات الإدراك لدى المتعلمين. وحيث إن التقدم الحضاري يواكب التقدم العلمي، ويعتمد عليه اعتماداً مباشراً يمكننا إدراك الأثر الفعال والمباشر الذي تقوم به الرياضيات من أجل تحقيق الرفاهية والرخاء للبشرية، إذ تعد الأداة المباشرة التي مهدت الطريق لتطور الفكر

البشري (Majdee, 2006). لذا فقد أخذ تعليم وتعلم الرياضيات اهتماماً عالمياً، ولم تعد عملية التقييم التي تقدمها اختبارات المعلم داخل غرفة الصف كافية للحكم على نوعية التعليم، وبالتالي أخفقت في تقديم صورة حقيقية عن نوعية الأنظمة التربوية، ومع التطور العلمي والتكنولوجي الهائل وسطوة مفاهيم العولمة، لم تعد المؤشرات التي تقدمها دراسات التقييم الوطني كافية أيضاً لمحاكمة نوعية التعليم في الأنظمة التربوية التي تنشأ تلك النوعية (Mostfa, 2007). ويشير الحكمي (Alhakami, 2007) إلى أن وجود نظم تقييم فعالة يؤثر بشكل إيجابي ومباشر في سلوك مديري المدارس والمعلمين وجميع المعنيين بالعملية التربوية، ويوضح ما المطلوب منهم، وما المعايير والمواصفات التي ينبغي أن يستندوا إليها في ممارساتهم التقييمية. كما يشير إلى بروز توجهات عديدة في مختلف أنحاء العالم من أجل إصلاح عمليات التقييم، وجعلها مستمرة مع استخدام أساليب متعددة، وتطبيق اختبارات معدة بإحكام على المستويين الوطني والعالمي. يُعد التقييم 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 على مستوى أداء المتعلم وكفاءته في القيام بأعمال أو مهام معينة. ومن أدوات جمع البيانات الإختبارات، فالإختبار الجيد مقياس لا غنى عنه لمعرفة مدى نجاح المواقف التعليمية المختلفة. مما يبين تزايد الإهتمام بدراسة أساليبها وأنواعها وطرق صياغة فقراتها لتطويرها لتؤدي الغرض المطلوب منها. وفي الواقع فإنّه كلما زاد اعتماد النظام التربوي على علامات الإختبار لغايات التقييم، والتشخيص، والحكم، والتصنيف، زاد اهتمام الباحثين والعاملين في بناء وتطوير الإختبارات والمقاييس بإيجاد الوسائل والسبل التي تمكّنهم من بناء إختبارات ومقاييس قادرة على تحقيق مبدأ العدالة والمساواة بين الأفراد فيما يخص القرارات التي يتّم اتخاذها بشأنهم، وفي المحصلة تحقيق إختبارات ومقاييس متحررة من التحيز (Osterlind, 1983). شهد العقد الأخير تطورات في مجال مناهج الرياضيات تمثلت بتوجهات جديدة في تعلم الرياضيات وتعليمها وتقييم تعلم الطلبة وقد تم توثيق تلك التوجهات في وثائق عالمية وكان من أبرزها وثيقة مبادئ ومعايير مناهج الرياضيات الصادرة عام (2000) عن المجلس الوطني الأمريكي لمعلمي الرياضيات (NCTM) National Council of Teachers of Mathematic وقد ركزت معظم الوثائق على مناهج الرياضيات ومواصفاته ومدى تفاعله مع البيئة والتطور التكنولوجي ومرونته في استيعاب هذه التكنولوجيات المتطورة باستمرار (Alabsi, 2002). لقد تم وضع منهج الرياضيات للمدارس في الأردن على أساس أفضل المعايير لمناهج الرياضيات في العالم، ويوظف النهج الجديد للمناهج النظرية التعليمية الحديثة كالتعليم الاستدلالي والتعليم التشاركي بهدف تحقيق مستويات عالية من المعرفة والمهارة في الرياضيات والإقبال على التعلم، وسعى المنهج الجديد بشكل خاص لتقوية مهارات الطلاب في تطبيق الرياضيات على نطاق واسع في الحالات العملية في المواضيع الأخرى، انطلاقاً من إدراك أهمية هذه المهارات لمستقبل الأردن كإقتصاد معرفي (Mohana, 2007). وللتحقق من أن المنهج قد حقق أهدافه فلا بد من إيجاد أدوات لقياس مدى المعرفة التي حصل عليها الطلبة في هذا المنهج، وخصوصاً مناهج الرياضيات. ومن أهم هذه الأدوات الإختبارات التحصيلية. ويمكن إعداد الإختبارات وفق نماذج إحدى نظريتي القياس وهما:

النظرية التقليدية في القياس، والنظرية الحديثة في القياس، حيث تُعدُّ النظرية الحديثة في القياس ثورة في مجال القياس النفسي والتربوي، وذلك كما يعتبرها كثير من العاملين في مجال علم القياس (Allam, 1986). إذ تُفترض هذه النظرية إمكانية التنبؤ بأداء الأفراد المفحوصين في خِصَمٍ خاصية أو خصائص مميزة، ومفسرة لهذا الأداء، وهي ما يُمكن أن يطلق عليه السِّمة أو السِّمات الكامنة (Latent Traits)، ولأنَّ من الصعب ملاحظة هذه السِّمة أو السِّمات الكامنة مباشرة، فإنه يتمُّ تقدير قدرات الأفراد المفحوصين على السِّمة موضع القياس، من خلال الأداء الملاحظ للأفراد المفحوصين على مجموعة من الفقرات الاختبارية (Hambleton & Swaminathan, 1985). يغلب على الاختبارات والمقاييس المستخدمة في البيئة العربية اعتمادها عند بنائها وتقنيها على نماذج النظرية التقليدية في القياس، على الرغم من ظهور العديد العيوب المرتبطة بهذا النموذج، وعلى الرغم من وجود نظرية الاستجابة للفقرة (IRT) والتي حظيت باهتمام الباحثين حيث أنها تغلبت على كثير من مشكلات القياس التقليدية. 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; Lord & Stocking, 1989). كما أن أغلب الدراسات التي بُنيت وفق نظرية الاستجابة للفقرة استخدمت النموذج أحادي المعلمة (نموذج راش)، ومن هنا فقد ارتأى الباحث استخدام هذا النموذج في بناء اختبار تحصيلي في مادة الرياضيات باستخدام النموذج ثنائي المعلمة.

مشكلة الدراسة وأسئلتها

تعتبر نماذج نظرية الاستجابة للفقرة (IRT) من النماذج السيكمترية المعاصرة التي تثبتت منفعتها في التغلب على كثير من مشكلات القياس النفسي والتربوي التي واجهت النظرية التقليدية، ويرى المختصون أن مبادئ نظرية الفقرة تختلف اختلافاً جوهرياً عن مبادئ النموذج التقليدي. فنظرية الفقرة تُعد تطوراً حديثاً ومهماً في القياس النفسي؛ إذ أنها تحرر الأفراد من عينة الأفراد المستخدمة وعينة الفقرات، كما تمكن المختص من الإجابة عن أي سؤال عن الفقرة أو الاختبار أو المفحوص (Embretson & Reise, 2000).

وكما ذُكر سابقاً فإن 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 الثنائي المعلمة من أجل بناء اختبار تحصيلي في الرياضيات؛ وذلك لندرة مثل هذه الاختبارات في بيئتنا العربية بشكل عام والأردنية بشكل خاص مما شكل دافعاً للباحثين لإجراء مثل هذه الدراسة؛ حيث أن يتميز هذا النموذج عن نموذج راش أحادي المعلمة بأن هذا النموذج يقيس معلمتين للفقرات وهما: معلمة الصعوبة ومعلمة التمييز، بينما يقيس نموذج راش معلمة واحدة للفقرات وهي معلمة الصعوبة. مما يؤكد أن تفضيل النموذج ثنائي المعلمة على نموذج راش.

وتحديداً، فإن الدراسة سعت للإجابة عن الأسئلة الآتية:

1. ما دلالات صدق وثبات الاختبار المعدّ لطلبة الصف الثاني الثانوي العلمي في وحدة التكامل وفقاً للنموذج الثنائي المعلمة؟
2. ما هي معالم فقرات الاختبار (الصعوبة، والتمييز) المقدرة وفقاً وفقاً للنموذج الثنائي المعلمة؟
3. ما مدى تحقق افتراضات النموذج المستخدم وفقاً لنظرية الاستجابة للفقرة؟

أهمية الدراسة

تتمثل أهمية الدراسة النظرية الحالية في:

- تعد الاختبارات التحصيلية إحدى أهم أدوات قياس التحصيل.
- إن بناء أي اختبار تحصيلي وفق نماذج نظرية الاستجابة للفقرة (IRT) يجعله أكثر دقة وموضوعية واستقلالية في القياس.
- إن بناء الاختبار التحصيلي لأي مادة دراسية يعد مساهمة نوعية تضاف إلى مكتبة القياس النفسي والتربوي.

بينما تتمثل الأهمية العملية لهذه الدراسة في:

- مساعدة المعلم على إصدار أحكام موضوعية على مدى نجاح أساليب التدريس في تنظيم العملية التعليمية التعلمية من خلال تطبيق الاختبارات المبنية وفق نظرية الاستجابة للفقرة.
- تحديد الجوانب الايجابية في أداء المتعلم والعمل على تعزيزه، فضلاً عن تشخيص جوانب الضعف والقوة في تحصيل الطلبة.
- أن بناء الاختبارات التحصيلية المقننة يؤدي إلى توفير الجهد على المعلم.

أهداف الدراسة

هدفت هذه الدراسة إلى بناء اختبار تحصيلي في الرياضيات وفق لنماذج نظرية الفقرة (النموذج ثنائي المعلمة) يتصف بمعايير الاختبار الجيد من حيث خصائصه السيكومترية من صدق وثبات، وتتميز فقراته بمعاملات صعوبة وتمييز مناسبة لموضوع التكامل.

محدّدات الدراسة

تقتصر هذه الدراسة على

1. طلبة الصف الثاني الثانوي العلمي في المدارس الحكومية التابعة لمديرية التربية والتعليم للواء قصبه المفرق للعام الدراسي 2017/2018م.

2. فقرات اختبار من نوع اختيار من متعدد ، بحيث تعطى العلامة (1) للاستجابة الصحيحة على الفقرة، والعلامة (صفر) للاستجابة الخاطئة.
3. استخدام نموذج واحد من نماذج الاستجابة للفقرة هو النموذج ثنائي المعلمة (2PLM) في بناء اختبار تحصيلي موضوعي من نوع الاختيار من متعدد بأربعة بدائل فقط في مبحث الرياضيات/الفصل الدراسي الثاني/وحدة التكامل.

مصطلحات الدراسة: تعتمد الدراسة التعريفات الآتية لمصطلحاتها

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

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

البارامتر أو المعلمة Parameter: هي قيمة عددية تصف وتلخص لنا جميع بيانات المجتمع الذي تطبق عليه الدراسة (Allam, 2005).

اللوجيت (Log Unit): يعرف اللوجيت رياضياً باعتباره وحدة لقياس قدرة الفرد بأنه اللوغاريتم الطبيعي لمقدّر نجاح الفرد في الإجابة على الفقرات، وباعتباره وحدة لقياس صعوبة الفقرة بأنه اللوغاريتم الطبيعي لمقدّر فشل الأفراد في الإجابة على الفقرة. (Ludlow, & Haley, 1995)

الدراسات السابقة

لقد أجريت العديد من الدراسات التي استخدمت نظرية الاستجابة للفقرة في بناء الاختبارات وتحليلها، ولكن أغلب هذه الدراسات ركزت على استخدام نموذج واحد وهو نموذج راش أحادي المعلمة، ومن هذه الدراسات دراسة علام (Allam, 1990) والتي هدفت إلى استعمال نموذج راش في بناء اختبار تشخيصي محكي المرجح لقياس المعارف الأساسية في إعداد خطة البحوث التربوية والنفسية لدى طلبة الدراسات العليا بكلية التربية في مصر. تألف الاختبار بصورته الأولية من (32) فقرة من نوع الاختيار من متعدد لكل فقرة خمسة بدائل، ويقاس الاختبار (30) هدفاً سلوكياً تتعلق بالمعارف الأساسية في إعداد خطة البحوث التربوية والنفسية، وقد تألفت عينة التحليل الإحصائي من (48) طالباً ممن أنهموا مقرر مناهج البحث ويعدون خطة رسالة الماجستير في أحد فروع العلوم التربوية والنفسية. تم تحليل البيانات وفق نموذج راش وباستعمال الطريقة (Prox) التي اقترحها كل من رايت Wright ودوكلاس Douglas (1979). أظهرت النتائج عدم ملائمة (4) فقرات لافتراضات النموذج وأيضاً عدم ملائمة بيانات فردين من العينة لمتطلبات

النموذج وتراوحت القيم التقديرية لصعوبة الفقرات بين (-2.7) لوجيت و(2.3) لوجيت بمتوسط مقداره (صفر)، وانحراف معياري مقداره (1.23) لوجيت.

وفي دراسة دعنا (Da'ana, 2005) التي هدفت لبناء اختبار مكيف هرمي في الرياضيات للصف الثامن الأساسي وفقاً لنموذج راش، وتم بناء اختبارين من نوع الاختيار من متعدد بأربعة بدائل لكل فقرة تقيس فقراتها المستويات الخمسة الأولى من تصنيف بلوم في المجال المعرفي، الأول يتعلق بمحتوى الفصل الدراسي الأول عدد فقراته (35)، والثاني يتعلق بمحتوى الفصل الدراسي الثاني عدد فقراته (44)، وطبق الاختباران على عينة التحليل الإحصائي المؤلف من 376 مفحوصاً اختيروا بطريقة عشوائية مرحلية ذات مرحلتين، وأظهرت النتائج حذف (7) فقرات من الاختبار الأول وحذف (8) فقرات من الاختبار الثاني وذلك لأن قيم مربع كاي فيها دالة إحصائية عند مستوى دلالة (0.05)، وأشارت النتائج إلى أن الاختبارين يقيسان عاملاً واحداً (نسبة التباين المفسر لم تقل عن 20%)، وقد تم ترتيب الفقرات تصاعدياً وفق صعوبتها.

وهدفت دراسة العيثاوي (Ithawi, 2009) إلى استعمال نموذج راش أحادي المعلمة وهو أحد نماذج نظرية السمات الكامنة في بناء اختبار تحصيلي في مادة علم نفس الفروق الفردية، ولتحقيق أهداف الدراسة شرع الباحث في بناء الاختبار مستنداً إلى المفردات المقررة من قبل الهيئة القطاعية لوزارة التعليم العالي. أعد الباحث (173) فقرة من نوع الاختيار من متعدد تغطي مستويات الأهداف المعرفية الثلاث الأولى بحسب تصنيف بلوم وهي: (المعرفة والفهم والتطبيق)، واشتملت عينة البحث على (336) طالباً وطالبة من طلبة كلية التربية للجامعات العراقية تم اختيارهم بالطريقة الطبقيّة العشوائية أدخلت البيانات لبرنامج الحاسب الآلي راسكال (Rascal) ومن خلال معالجة البيانات تضمنت الصورة النهائية للاختبار (108) فقرة مطابقة لافتراضات النموذج ومتطلباته فقد تم حذف (28) فقرة لا تحقق أحادية البعد من خلال التحليل ألعاملي لفقرات الاختبار بطريقة المكونات الأساسية، فيما حذفت (29) فقرة لدلالة قيمة مربع كاي، واقتربت قيمة معامل التميز من (1) عدد صحيح حيث كانت القيمة ضمن المدى المقبول (0.629)، كما تم استبعاد (8) فقرات لا تحقق استقلالية القياس. دُرّجت الـ(108) فقرات بحسب صعوبة الفقرات وقدرات الأفراد للحصول على الاختبار بصيغته النهائية والذي بلغت قيمة معامل ثباته (0.952) ومعامل تمييزه (0.713).

وأجرى ادواردز والكوك (Edwards & Alcock, 2010) دراسة هدفت إلى تحليل نتائج اختبار في الرياضيات في بريطانيا وفق أنموذج راش، ومن أجل تحقيق هدف الدراسة قام الباحثان ببناء اختبار مؤلف من (11) فقرة، وقد تم توزيعه على عينة الدراسة المؤلفة من (164) طالباً وطالبة من طلبة السنة الأولى في المرحلة الجامعية، وقد طلب من أفراد العينة ترتيب فقرات الاختبار حسب مستوى صعوبتها، وبعد ذلك تم تحليل البيانات، التي تم الحصول عليها، باستخدام أنموذج راش، وأشارت النتائج إلى عدم مطابقة فقرات الاختبار لسنة أشخاص فقط من عينة الدراسة، بعد إعادة ترتيب فقرات الاختبار حسب أنموذج راش.

وهدفت دراسة الخياط (Alkhaiat, 2012) إلى التحقق من فاعلية النموذج اللوغاريتمي ذي المعلمة الواحدة (نموذج راش)، ومطابقة البيانات للنموذج لاختبار TIMSS الذي يقيس المهارات الرياضية لمستوى الصف الثامن الأساسي، ولتحقيق هدف الدراسة تم تطبيق فقرات الاختبار على عينة من طلبة الصف الثامن الأساسي بلغ عددها (599) طالبًا وطالبة؛ وأشارت نتائج الدراسة إلى تمتع الاختبار المقدم بالصدق والثبات، حيث بلغ معامل الثبات للاختبار (0.9)، وأما معامل الصدق بدلالة المحك بين درجات الطلبة على الاختبار ودرجاتهم في مادة الرياضيات فقد بلغت قيمته (0.87)، وأشارت النتائج إلى أن الاختبار يصنف بأنه متوسط الصعوبة، حيث تراوح مستوى صعوبة الفقرات بين (0.305 - 0.782) لوجيت. ومن ثم أوجد الباحث قيمة مربع كاي بحيث دلت قيمة مربع كاي لعدم مطابقة 10 فقرات لنموذج راش. وأشارت النتائج أيضا إلى مناسبة (25) فقرة من فقرات الاختبار والبالغ عددها (35) فقرة للبيئة المحلية.

أما علي (Ali, 2012) فهدفت دراسته إلى تعرف فاعلية بناء اختبار محكي المرجع لمقرر القياس والتقويم في التربية وفق نموذج راش؛ بحيث تكون الاختبار من (99) فقرة اختبارية من نوع الاختيار من متعدد وبأربعة بدائل، وتكونت عينة البحث من (426) طالبًا وطالبة من طلبة كلية التربية في قسم معلم الصف من جامعة دمشق، الذين درسوا المقرر في العام الدراسي 2010/2009. وتم تقسيم العينة إلى قسمين، وتطبيق الاختبار عليهما في وقت واحد. وتم استخدام 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)، وبالتالي أصبح العدد النهائي لفقرات الاختبار (89) فقرة، جميعها 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 تدريج الاختبار يحقق افتراضات نظرية الاستجابة للفقرة.

وهدفت دراسة أديوين و موكوبي (Adedoyin & Mokobi, 2013) إلى تحليل فقرات الاختبار من متعدد في اختبار بوسطن في الرياضيات الورقة الأولى للعام 2010. وتكون الاختبار من (40) فقرة تم بناءها من منهاج الرياضيات للسنة الثالثة. تكون مجتمع الدراسة من جميع الطلبة الذين تقدموا للاختبار في العام 2010 وعددهم (36940) طالبًا وطالبة، وبلغ حجم العينة (10000) طالبًا وطالبة تم اختيارهم عشوائيًا باستخدام برمجية (SPSS). وتم تحليل استجابات المفحوصين باستخدام أحد نماذج نظرية الاستجابة للفقرة وهو النموذج ثلاثي المعلمة، حيث تم حساب معالم الفقرات وهي: صعوبة الفقرات، وتمييزها، وتخمينها. كما تم رسم منحنيات معالم الفقرات المطابقة للنموذج ثلاثي المعلمة وعددها (23) فقرة من أصل (40) فقرة، وأظهرت نتائج الدراسة أن (12) من الـ (23) فقرة صنفت كفقرات ضعيفة، (10) فقرات صنفت كفقرات جيدة ويمكن تنقيحها أو تحسينها، وفقرة واحدة تعتبر فقرة جيدة يمكن اعتمادها دون أي تعديل. وتمت التوصية بأن يتم تحسين جودة فقرات الاختبار لتكون مطابقة لنماذج نظرية الاستجابة للفقرة (IRT).

وقام أون (Onn, 2013) بدراسة هدفت إلى مقارنة بين النظرية التقليدية ونظرية الاستجابة للفقرة، وذلك من حيث عدد الفقرات المنتقاة، ومعلمة الثبات، ومن أجل تحقيق هدف الدراسة، قام الباحث بإعداد اختبار تحصيلي في مادة الفيزياء تكون من (50) فقرة من نوع الاختبار من متعدد، قام بتوزيعه على عينة مؤلفة من (69) طالبا وطالبة من طلبة المدارس في نيجيريا، وقد أشارت نتائج الدراسة إلى مطابقة (38) فقرة لنظرية الاستجابة للفقرة، وحذف (12) فقرة، كما أشارت النتائج إلى تدني معامل الثبات للاختبار وذلك وفق نظرية الاستجابة للفقرة، فقد بلغ معامل الثبات (0.67).

وهدفت دراسة الزليعي (AL-Zayla'I, 2014) إلى بناء اختبار تشخيصي المرجع لقياس محكي مهارات البحث العلمي لدى طلاب الدراسات العليا بكلية التربية بجامعة الباحة، واستخدم هذا الاختبار للتعرف على مدى تمكن طلبة الدراسات العليا من هذه المهارات، ومعرفة الفروق في مدى تمكن الطلاب في ضوء بعض المتغيرات، وتكون مجتمع الدراسة من طلاب وطالبات مرحلة الماجستير كلية التربية بجامعة الباحة. ولتحقيق أهداف هذه الدراسة قام الباحث ببناء اختبار تشخيصي محكي المرجع لقياس تمكن الطلبة من مهارات البحث العلمي مروراً بثلاث مراحل هي: (مرحلة التحليل- مرحلة البناء- مرحلة التجريب). واستخدم الباحث المنهج الوصفي المسحي، وكان من أهم نتائج الدراسة أن الاختبار الذي تم بناؤه في هذه الدراسة يتمتع بتوافر خصائص الاختبار الجيد، حيث يتمتع بدلالات صدق وثبات جيدة، كما أظهرت النتائج انخفاضاً واضحاً في مدى التمكن لدى طلبة الدراسات العليا بكلية التربية بجامعة الباحة سواء في نتائج الاختبار الكلي أو نتائج الاختبارات الفرعية.

وهدفت دراسة حجازي والخطيب (Hijazi, T. & Alkhateeb, A. 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 تم بناء اختبار محكي المرجع في المادة النظرية لأحكام التلاوة والتجويد تكون من (41) فقرة من نوع الاختبار من متعدد وبأربعة بدائل، وطبق على عينة من طلبة بلغ عددها (404) طالباً وطالبة تم اختيارهم بالطريقة العشوائية الطبقية من (16) مركزاً قرآنياً (9 للذكور، و 7 للإناث) في العاصمة عمان للعام الدراسي 2012/2011 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.

وفي دراسة علوان (Olwan, 2015) التي هدفت الدراسة الى استعمال انموذج راش للاحتمال اللوغارتمي في تطوير اختبار الاستعداد المكاني، طبق الاختبار على عينة بلغت (500) طالبا وطالبة من طلبة المرحلة الاولى في جامعات في مدينة بغداد و حللت النتائج وفق انموذج

راش فقد استبعدت (8) فقرات من المكونات واستبعدت (3) فقرات لعدم تحقيقها استقلالية القياس، وبذلك يكون عدد الفقرات المتبقية (88) فقرة في الاختبار الكلي وتحققت موضوعية لقياس كما تمثل بأنموذج راش.

وهدفت دراسة دبوس (Dabous, 2016) إلى استخدام نظرية الاستجابة للفقرة في بناء تجمع فقرات اختبار محكي المرجع في الرياضيات وفق النموذج اللوجستي ثنائي المعلم. ولتحقيق هذا الهدف تم إعداد تجمع من الفقرات ثنائية التدرج مكون من (50) فقرة، وتجمع من الفقرات متعددة التدرج مكون من (10) فقرات. وروعي في بناء هذه الفقرات أن تقيس تحصيل الطالب في الرياضيات في المعرفة المفاهيمية والمعرفة الإجرائية وحل المشكلات. وقد تكونت عينة الدراسة من (502) طالباً وطالبة أجابوا عن جميع الفقرات. وقد أظهرت النتائج تحقق افتراضات نظرية الاستجابة للفقرة. كما بينت نتائج التحليل فيما يتعلق بمطابقة الفقرات الثنائية التدرج للنموذج الثنائي المعلم أن جميع الفقرات كانت مطابقة للنموذج ثنائي المعلم، وأن الفقرات المتعددة التدرج كانت مطابقة لنماذج الاستجابة المتدرجة GRM. وفي ضوء نتائج الدراسة أوصى الباحث بمجموعة من التوصيات.

وهدفت دراسة محمود وصباح (Mahmoud, L. & Sabah, S. 2016) إلى بناء اختبار تحصيلي في الهندسة لطلبة الصف الخامس باستخدام نموذج راش، وتشخيص فهم الطلبة للمفاهيم الهندسية. تم بناء اختبار تحصيلي مكون من (30) فقرة من نوع الاختيار من متعدد. طبق الاختبار على (216) طالباً وطالبة. وأظهرت النتائج أن الاختبار ثابت، حيث بلغ معامل الثبات للفقرات (0.98) ومعامل الثبات للأفراد (0.78) وأن فقرات الاختبار غطت مدى الصعوبة بشكل جيد، فتراوحت قيمة الصعوبة للفقرات ما بين (-3.29 – 2.69) لوجيت، وظهر أن معظم الفقرات تقع ضمن إحصائيات الملاءمة لنموذج راش. وأظهرت النتائج أيضاً وجود أخطاء متنوعة في المفاهيم الهندسية منها: خصائص متوازي الأضلاع، مفهوم المحيط، مفهوم الشبكات، والربط بين الأشكال الهندسية، وفي المقابل هناك مجموعة من المفاهيم الهندسية التي أجاب عليها الطلبة بنسبة أكبر من (88%) منها: مفهوم المضلع، مفهوم المضلع المنتظم، ومفهوم الزاوية وقياسها باستخدام المنقلة.

وهدفت دراسة محمود (Mahmoud, 2016) لاستخدام نموذج راش في بناء اختبار تحصيلي محكي المرجع في مقرر علم النفس التربوي. ولتحقيق هذا الهدف صممت الباحثة اختباراً مكوناً من (50) فقرة من نوع الاختيار من متعدد في ذلك المقرر، بلغت قيمة معامل ثباته كرونباخ ألفا (0.87). طبقت الباحثة الاختبار على عينة من (246) طالبة من طالبات جامعة أم القرى، من المسجلات في مقرر علم النفس التربوي، خلال الفصل الدراسي الثاني (1434/1433 هـ) الموافق (2013/2012 م). واستخدمت برنامج (Bilog-MG3) في فحص مطابقة بيانات الاختبار لنموذج راش وفي حساب معالم الفقرات. أظهرت نتائج الدراسة تحقق افتراضات نموذج راش، ومطابقة استجابات عن (25) فقرة من فقرات الاختبار للنموذج، بينما حذفت (25) فقرة لم تطابق النموذج. بلغت قيمة معامل ثبات الاختبار وفق نظرية استجابة الفقرة (0.92)، كما تراوحت قيم معالم صعوبة فقراته بين (-0.698) و(0.570)، مما يؤكد أن الاختبار متوسط الصعوبة. وتكمن أهمية هذه الدراسة بأنها توضح الآلية التي يتم من خلالها تطوير وبناء الاختبارات التحصيلية وفق

النظرية الحديثة في القياس، مما يمكن الآخرين من الاستفادة منها عند بناء هذا النوع من الاختبارات.

هدفت دراسة صباح (Sabah, 2016) إلى استخدام نظرية الاستجابة للفقرة في بناء اختبار محكي المرجع في اللغة الإنجليزية لطلبة جامعة القدس المفتوحة، وفق نموذج (راش)، ولتحقيق هذا الهدف تم بناء اختبار تحصيلي في اللغة الإنجليزية مؤلفاً من (45) فقرة من نوع الاختبار من متعدد من أربعة بدائل، يقيس تحصيل الطالب في اللغة الإنجليزية، تم حذف (5) فقرات أجمع المحكمون على أن هناك فقرات تؤدي غرضها، فأصبح الاختبار مكوناً بصورته الأولى 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 استجابات (63) فرداً غير مطابقين للنموذج، كما بينت النتائج المتعلقة بتقديرات معاملات الصعوبة أنها كانت مقبولة ضمن المحكات التي أوردتها أدبيات القياس التربوي. وأسفر التحليل عن اختبار جيد يتمتع بمظاهر متعددة من الصدق، وثبتت (0.99) لفقرات الاختبار، (0.80) لقرات الأفراد. كما بينت النتائج أن قيمة دالة المعلومات كانت أقصى ما يمكن عندما كانت $(\hat{\theta} = b)$ 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.

هدفت دراسة الطراونة (Al-Tarawneh, 2016) إلى بناء مقياس للاستواء لطلبة الصفوف الأساسية العليا في الأردن باستخدام نموذج سلم التقدير، وهو أحد نماذج نظرية الاستجابة للفقرة متعدد التدرج. تكونت عينة الدراسة من (2862) طالباً وطالبة من طلبة الصفوف الأساسية العليا في الأردن، وتم التحقق من مطابقة عينة الدراسة وفقرات المقياس لافتراضات نموذج سلم التقدير. وتم التحقق من افتراضات نظرية الاستجابة للفقرة للمقياس، وأظهرت النتائج دلالات صدق وثبات مقبولة للمقياس.

هدفت دراسة سوايمة والزعيبي (Sawalmeh & Alzoabi, 2017) إلى بناء اختبار لقياس مهارات التفكير الناقد باستخدام نظرية استجابة الفقرة. ولتحقيق هذا الهدف تم بناء 228 من نوع الاختبار من متعدد، تغطي مهارات التفكير الناقد. ومن خلال التحكيم والتجريب تم الاحتفاظ بـ 89 فقرة. وطبق الاختبار على 444 طالباً وطالبة في المرحلة الأساسية العليا، و610 طالباً وطالبة في المرحلة الثانوية، و120 طالباً وطالبة في المرحلة الجامعية. وتم التحقق من افتراض أحادية البعد والصدق والثبات للاختبار. كما تم التحقق من مطابقة الاستجابات عن فقرات الاختبار لنماذج نظرية استجابة الفقرة، وقد اعتمدت الدراسة النموذج ثلاثي المعلمة؛ لكون تقديرات القدرة وفقاً له أكثر دقة منها للنموذجين أحادي وثنائي المعلمة، خاصة عند مستويات القدرة العليا، واتصفت معالم الفقرات ومعلم القدرة بخاصية اللاتغير، وقد قدرت معالم الفقرات ودالة المعلومات لها، كما حسبت بعض إحصائيات الاختبار.

هدفت دراسة الزبيدي (Al Zubaidy, 2018) إلى استخدام النموذج الأحادي المعلمة (راش) في بناء اختبار تحصيلي محكي المرجع في مقرر الاختبارات والمقاييس لطلبة الدبلوم التربوي في كلية التربية بجامعة الطائف ولتحقيق هذا الهدف تم بناء اختبار محكي المرجع في مقرر الاختبارات والمقاييس يقيس تحصيل الطلبة ويقدم تقويماً موضوعياً للمهارات المطلوب إتقانها 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 جميع فقرات الاختبار بصورته النهائية لافتراضات نموذج راش، وقد بلغ معامل الثبات للأفراد (0.9) أما معامل الثبات للاختبار فبلغ (0.79). وخلصت الدراسة إلى مجموعة من التوصيات.

التعليق على الدراسات السابقة

يتضح من خلال الدراسات السابقة تركيزها على استخدام نموذج واحد من نماذج نظرية الاستجابة للفقرة وهو النموذج الأحادي المعلمة، ما عدا دراسة دبوس (Dabous, 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 التكامل وتطبيقاته على الرغم من أن هذا الموضوع يعتبر من أهم مواضيع الرياضيات؛ حيث أنه في الغالب يدرس للطلبة في نهاية المرحلة الثانوية والتي غالباً ما تكون إختبارات وطنية، أي أنها اختبارات واسعة النطاق (spread wide tests) تعتمد عليها وزارات التعليم العالي في قبول الطلبة في الجامعات، ولذا فإن معلمة التمييز تعتبر مهمة هنا. كما تبين من شارت نتائج دراسة حجازي والخطيب (Hijazi, T. & Alkhateeb, A. 2014)، 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، مما يؤكد أهمية بناء اختبارات وفق النموذج الثنائي المعلمة.

منهجية البحث

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

مجتمع الدراسة

تم اختيار طلبة الصف الثاني الثانوي العلمي في محافظة المفرق المدارس الحكومية للعام الدراسي 2018/2017م كمجتمع ممثل لطلبة هذا الصف في الأردن.

عينة الدراسة

تشكَّلت عينة الدراسة من جميع طلاب وطالبات الصف الثاني الثانوي العلمي في المدارس الحكومية الثانوية التابعة لمديرية التربية والتعليم للواء قصبه المفرق، للعام الدراسي 2018/2017م.

وبيين الجدول (1) توزيع أفراد عينة الدّراسة وفقاً للجنس، وعدد المدارس، وعدد الطلبة الذين تمّ اختيارهم من مديرية التربية والتعليم للواء قصبه المفرق.

جدول (1): توزيع أفراد عينة الدّراسة وفقاً للجنس، وعدد المدارس، وعدد الطلبة التّابعين لمديرية التربية والتعليم للواء قصبه المفرق.

عدد الطلبة	عدد المدارس	الجنس
354	11	ذكور
325	10	إناث
679	21	المجموع

أمّا عينة الدّراسة الاستطلاعيّة فقد تشكّلت من طلاب وطالبات مدرستين من مدارس مدينة المفرق التابعة لمديرية التربية والتعليم لواء قصبه المفرق الذين خضعوا لأداة الدّراسة بصورتها الأولى. وبيّن الجدول (2) توزيع أفراد عينة الدّراسة الاستطلاعيّة وفقاً للجنس، وعدد المدارس، وعدد الطلبة الذين تمّ اختيارهم من مديرية التربية والتعليم للواء قصبه المفرق.

جدول (2): توزيع أفراد عينة الدّراسة الاستطلاعيّة وفقاً للجنس والمدارس وعدد الطلبة.

عدد الطلبة	الجنس	المدرسة
33	ذكور	الملك عبدالله الثاني للتميز
27	إناث	حي العليّات الثانوية
57		المجموع

وبهدف الحصول على بيانات نقيّة، واقعيّة تعكس مستوى أداء الطلبة الحقيقي، فقد تمّ الاتفاق مع المعلمين والمعلمات، وبالتنسيق مع مديري المدارس التي طبّقت فيها أداة الدّراسة على موعد لتطبيق أداة الدّراسة، بحيث يُعَلَّم به الطلبة مسبقاً، على أن تكون العلامة على أداة الدّراسة هي بمثابة علامة معتمدة لدى المدرسة ضمن تحصيل الطلبة في مبحث الرياضيات للمصف الثاني الثانوي العلمي، حيث أن الطالب الذي لا ينجح مدرسياً لا يتقدم لامتحان شهادة الدراسة الثانوية.

أداة الدراسة

تمّ بناء اختبار تحصيلي في مبحث الرياضيات للمصف الثاني الثانوي العلمي في وحدة "التكامل" من كتاب الرياضيات المقرّر في الفصل الدّراسي الثاني للعام الدّراسي 2018/2017م، حيث تمّت الاستعانة بدليل المعلم في ذلك المبحث، وفيما يأتي وصفاً للإجراءات التي اتبعت في إعداد وبناء الاختبار:

تحديد الغرض من الاختبار: والذي تمثّل بقياس تحصيل طلاب وطالبات المصف الثاني الثانوي العلمي للعام الدّراسي 2018/2017م، في مبحث الرياضيات في وحدة التكامل من كتاب الرياضيات المقرّر.

تحديد الموضوعات الداخلة في الاختبار: وذلك من خلال القيام بصياغة النتائج التعليمية الخاصة (الأهداف التفصيلية) لوحدة التكامل، وعمل تحليل محتوى للوحدة، ومن ثم تحديد المستوى العقلي لتلك النتائج من خلال ربط مستويات النتائج التعليمية الخاصة بمحتوى المادة الدراسية موضوع الاختبار، لينتج بعد ذلك إعداد جدول مواصفات الاختبار ببعدين: البعد الأول، ويمثل محتوى المادة الدراسية (موضوعات الوحدة)، والبعد الثاني، ويمثل مستويات الأهداف حسب تصنيف بلوم للأهداف التدريسية في المجال المعرفي الذي يتضمن: المعرفة، والاستيعاب، والمهارات العقلية العليا (تطبيق، تحليل، تركيب، تقويم)، ومن ثم تم تحديد الأهمية النسبية لكل موضوع في الوحدة موضع الاهتمام بناءً على معيار عدد الأهداف التفصيلية لكل درس، مع الأخذ بعين الاعتبار الأهمية العلمية لكل موضوع، ومدى ارتباطه بحاجات المتعلم، وعدد الحصص المخصصة لتدريس كل موضوع في الوحدة.

صياغة فقرات الاختبار بصورته الأولية: بالاعتماد على جدول مواصفات الاختبار، والأهداف التفصيلية لكل درس في الوحدة موضع الاهتمام تمت كتابة فقرات الاختبار بصورته الأولية حسب الأسس العلمية والفنية المتبعة في كتابة فقرات الاختبار، مع مراعاة ألا تعتمد إجابة إحدى هذه الفقرات على إجابة الفقرات الأخرى لضمان تحقق افتراض الاستقلال الموضوعي الذي تمت الإشارة إليه في معرض الحديث عنه في الفصل الثاني، حيث تكوّن الاختبار بصورته الأولية من (40) فقرة من نوع الاختيار من متعدد بأربعة بدائل لكل فقرة، بديل واحدة منها فقط يحتمل الإجابة الصحيحة.

التأكد من صدق محتوى الاختبار: تم عرض الصورة الأولية للاختبار، وجدول المواصفات، والنتائج التعليمية التفصيلية، على (10) محكمين ممن هم من ذوي الخبرة والاختصاص، ثلاثة منهم من ذوي تخصص القياس والتقويم، وواحد من ذوي تخصص أساليب تدريس الرياضيات، وثلاثة من مشرفي الرياضيات في وزارة التربية والتعليم، ومن ذوي تخصص أساليب تدريس الرياضيات أيضاً، وثلاثة مدرّسين ممن يدرّسون المبحث، حيث طلب منهم الإجابة على استبانة للحكم على مدى ملائمة جدول المواصفات، وفقرات الاختبار.

وفي ضوء تقديرات المحكمين ومناقشة مقترحاتهم تم إجراء التعديلات اللازمة على فقرات الاختبار، حيث تمت إعادة صياغة إحدى بدائل الفقرة رقم (17)، كما تمت إعادة صياغة متن الفقرتين رقم (1) و (20)، ولم تحذف أي فقرة، ليبقى الاختبار بصورته الأولية بعد أخذ تعديلات المحكمين بعين الاعتبار مكوناً من (40) فقرة.

التجريب الأولي للاختبار: بعد أن أصبح الاختبار بصورته الأولية جاهزاً للتطبيق، تم تطبيق الاختبار بصورته الأولية المكوّن من (40) فقرة على العينة الاستطلاعية، وذلك بعد الزيارة الأولى لمدارس العينة المقصودة، والتي كان الهدف منها تعريف مدرّسي ومدّرات مبحث الرياضيات للصف الثاني الثانوي العلمي في المدارس المعنية بالاختبار والإطلاع عليه، وآلية تطبيقه، وما يلزمها من إجراءات ومعلومات يستلزمها الاختبار ويتطلب تسجيلها من قبل المعلمين والمعلمات خلال عملية تطبيق الاختبار كوضوح تعليمات الاختبار، والتأكد من وضوح الصياغة اللغوية للفقرات، وتحريّ الفقرات التي يتمّ التساؤل حولها للكشف عمّا إذا كان هنالك غموض فيها أم لا،

بالإضافة إلى ضبط الزمن الذي يستغرقه الطلبة في إنهاء الاختبار، كما تمّ الاتفاق على موعد محدد باليوم والحصّة لتطبيق الاختبار يُعلم به الطلبة مسبقاً، على أن تكون العلامة على الاختبار بصورته الأولى بمثابة علامة الطالب أو الطالبة على الاختبار الشهري الأول في مبحث الرياضيات، إذ تزامن تطبيق الاختبار بصورته الأولى مع الاختبار الشهري الأول في تلك المدارس.

وبناءً على عملية تطبيق الاختبار على العينة الاستطلاعية تمّ تحديد الزمن المناسب للإجابة على فقرات الاختبار: حيث وجد أنّ معدّل (95%) من الطلبة تقريباً قد توفّقوا عن الإجابة على فقرات الاختبار بعد مرور (80) دقيقة على بداية الاختبار، وعليه تمّ اعتماد زمن الاختبار بفقراته الأربعين ليكون (80) دقيقة؛ أي ما معدّله دقيقتان لإجابة كل فقرة. كما تبين وضوح فقرات وتعليمات الاختبار للمفحوصين.

تطبيق الاختبار على عينة الدراسة: طبق الاختبار على عينة الدراسة والتي كان عددها (679) طالبا وطالبة، وقد تم التأكد من الإجابة على جميع الأسئلة قبل استلام الاختبار من الطلبة، وصححت فقراته تمهيدا لاستكمال بناء الاختبار وفق النموذج ثنائي المعلمة.

استعمال الحاسوب في إدخال العلامات: تمّ تفريغ استجابات جميع أفراد العينة والبالغ عددهم (679) طالباً وطالبة على فقرات الاختبار تمهيداً لاستكمال عملية تحليل البيانات.

مطابقة الأفراد والفقرات للنموذج اللوجستي ثنائي المعلمة (2PLM) وفق نظرية الاستجابة للفقرة: لتقدير معالم فقرات الاختبار (معلمة الصعوبة b ، ومعلمة التمييز a) فقد تمّ التأكد أولاً من مطابقة الأفراد والفقرات للنموذج اللوجستي ثنائي المعلمة (2PLM) باستخدام برنامج (BILOG-MG3) بعد تزويده بالملف الناتج من تحويل البيانات إلى صورة صفر وواحد، حيث تمّ تشغيل البرنامج ثلاث مرات:

التشغيل الأولى وهدفت إلى حذف الأفراد الذين لم تتطابق استجاباتهم مع توقعات النموذج المستخدم، وقد أظهرت نتائج التحليل عدم مطابقة (18) طالباً وطالبة للنموذج اللوجستي ثنائي المعلمة منهم (4) طلاب وطالبات كانت القيم الاحتمالية لتقدير القدرة لديهم أقل من (0.01)، وهم ممن يحملون الأرقام 105، 176، 311، 501 و (14) طالباً وطالبة تعذر تقدير الخطأ المعياري لقدراتهم، وهم ممن يحملون الأرقام 45، 64، 99، 111، 120، 137، 145، 151، 153، 207، 371، 399، 652، 655.

والتشغيل الثانية هدفت إلى فحص مطابقة فقرات الاختبار للنموذج اللوجستي ثنائي المعلمة، حيث أُعيد التحليل مرة ثانية بعد حذف استجابات الطلاب والطالبات غير المطابقين الناتجين من التشغيل الأولى للبرنامج، حيث أظهرت نتائج التحليل عدم مطابقة (5) فقرات للنموذج المستخدم، منها (3) فقرات، وهي الفقرات ذوات الأرقام 21، 30، 40، لم تكن مطابقة وفقاً للنظرية الكلاسيكية في القياس، و (2) من الفقرات، وهما الفقرتان ذواتا الرقمين 17، 29، لم تكونا مطابقتين نظراً لأنّ القيم الاحتمالية لهما وفقاً لمؤشر (χ^2) كانت أقل من 0.01، ليصبح الاختبار بشكله النهائي مكوناً من (35) فقرة فقط.

أما التشغيلية الثالثة وهدفت إلى تقدير معالم فقرات الاختبار (معلمة الصعوبة b ، ومعلمة التمييز a) بشكل نهائي، حيث أُعيد التحليل للمرة الثالثة بعد حذف استجابات الطلاب والطالبات غير المطابقين للنموذج المستخدم، الناتجين من التشغيلية الأولى للبرنامج، بالإضافة لحذف الفقرات غير المطابقة أيضاً، والناتجة من التشغيلية الثانية للبرنامج.

نتائج الدراسة ومناقشتها

النتائج المتعلقة بالسؤال الأول والذي ينص على: ما دلالات صدق وثبات الاختبار المعد لطلبة الصف الثاني الثانوي العلمي في وحدة التكامل وفقاً للنموذج الثنائي المعلمة؟

صدق المحتوى: تمّ عرض الصُّورة الأولى للاختبار، وجدول المواصفات، والنتائج التعليمية التفصيلية، على (10) محكمين ممن هم من ذوي الخبرة والاختصاص، ثلاثة منهم من ذوي تخصص القياس والتقويم، وواحد من ذوي تخصص أساليب تدريس الرياضيات، وثلاثة من مشرفي الرياضيات في وزارة التربية والتعليم، ومن ذوي تخصص أساليب تدريس الرياضيات أيضاً، وثلاثة مدرّسين ممن يدرّسون المبحث، حيث طلب منهم الإجابة على استبانة للحكم على مدى ملائمة جدول المواصفات، وفقرات الاختبار.

وفي ضوء تقديرات المحكمين ومناقشة مقترحاتهم تمّ إجراء التعديلات اللازمة على فقرات الاختبار، حيث تمّت إعادة صياغة إحدى بدائل الفقرة رقم (17)، كما تمّت إعادة صياغة متن الفقرتين رقم (1) و (20)، ولم تحذف أي فقرة، ليبقى الاختبار بصورته الأولى بعد أخذ تعديلات المحكمين بعين الاعتبار مكوّناً من (40) فقرة. وبذلك تم التحقق من صدق الاختبار، حيث تم اختيار الفقرات التي تم الاتفاق عليها من قبل المحكمين.

تقدير ثبات الاختبار: بلغت قيمة معامل الثبات الإمبريقي للاختبار (0.82)، في حين بلغت قيمة معامل الثبات النظري (0.81)، والتي تحسب من خلال أخذ مقلوب القيمة $(1 + \sigma^2_{S.E_{\hat{\theta}}})$ ، حيث $\sigma^2_{S.E_{\hat{\theta}}}$ هو تباين الخطأ لتقدير قدرات الأفراد، والتي يوفرها برنامج (BILOG-MG3) في الملف ذو الامتداد PH3*. وهاتان القيمتان مرتفعتان و تقدمان مؤشراً جيداً على توفر دلالة مقبولة لثبات الاختبار.

النتائج المتعلقة بالسؤال الثاني والذي ينص على: ما هي معالم فقرات الاختبار (الصعوبة، والتمييز) المقدرة وفقاً وفقاً للنموذج الثنائي المعلمة؟

يبين الجدول (3) القيم المقدّرة لكلّ من معلمة الصعوبة b ، ومعلمة التمييز a الناتجة من التشغيلية الثالثة لبرنامج (BILOG-MG3)، لكلّ فقرة من فقرات الاختبار بصورته النهائية، والبالغ عددها (35) فقرة.

جدول (3): القيم المقدّرة لكلّ من معلمة الصعوبة والتمييز لفقرات الاختبار بصورتها النهائي والبالغ عددها (35) فقرة.

الخَطأ المعياري للتمييز	معامل التمييز A	الخَطأ المعياري للصعوبة	معامل الصعوبة b	رقم الفقرة في الاختبار
0.13	0.55	1.29	-5.46	1
0.27	0.75	0.84	3.05	2
0.50	2.91	0.08	0.65	3
0.29	1.75	0.11	0.50	4
0.12	0.74	0.55	-3.18	5
0.13	0.90	0.34	-1.13	6
0.13	1.04	0.28	-1.66	8
0.15	1.04	0.23	-0.42	9
0.12	0.89	0.35	-2.22	10
0.40	2.25	0.09	1.24	11
0.17	1.09	0.23	-0.30	12
0.43	2.53	0.08	1.26	13
0.40	1.80	0.16	1.79	14
0.13	0.45	0.71	1.92	15
0.17	1.07	0.17	0.33	16
0.36	1.95	0.11	1.54	17
0.26	0.88	0.41	2.43	19
0.28	1.29	0.17	1.62	20
0.12	0.78	0.25	0.98	21
0.30	0.87	1.70	2.29	22
0.13	0.83	0.19	-0.39	24
0.62	0.79	0.67	5.01	25
0.39	1.02	0.07	-0.29	26
0.37	1.53	0.30	2.99	27
0.44	2.61	0.23	0.71	28
0.20	1.25	0.19	2.06	29
0.48	1.55	0.21	1.92	30
0.15	1.26	0.22	1.64	33

...تابع جدول رقم (3)

رقم الفقرة في الاختبار	معامل الصعوبة b	الخطأ المعياري للصعوبة	معامل التمييز A	الخطأ المعياري للتمييز
34	2.06	0.21	1.87	0.17
35	-0.44	0.48	1.06	0.12
36	-1.68	0.26	1.47	0.13
37	1.45	0.17	0.51	0.21
38	-0.89	0.07	0.99	0.38
39	0.52	0.26	1.15	0.12
40	1.05	1.70	2.45	0.30
الوسط الحسابي	0.60	0.38	1.31	0.26
الانحراف المعياري	1.96	0.41	0.64	0.14

يلاحظ من الجدول (3) أنَّ قيم معلمة الصعوبة لل فقرات قد تراوحت من (-5.46) لوجيت للفقرة رقم (1) إلى (5.01) لوجيت للفقرة رقم 25، وبمعدّل مقداره (0.60) تقريباً، بينما تقع باقي القيم ضمن الفترة (-3، 3) وهذا مؤشر على تجانس فقرات الاختبار. تراوحت قيم معلمة التمييز من (0.45) للفقرة رقم 15 إلى (2.91) للفقرة رقم 3، وبمعدّل مقداره (1.31).

كما يلاحظ من الجدول (3) انخفاض قيم الانحراف المعياري وهذا مؤشر على تقارب البيانات من النموذج، فزيادة تقارب البيانات من النموذج يؤدي إلى انخفاض قيم الانحراف المعياري، وهذا يعد مؤشراً للثبات.

النتائج المتعلقة بالسؤال الثالث والذي ينص على: ما مدى تحقق افتراضات النموذج المستخدم وفقاً لنظرية الاستجابة للفقرة؟

أحادية البعد: Unidimensionality

تمّ التحقق من افتراض أحادية البعد لبيانات عينة الدراسة الاستطلاعية باستخدام برنامج *NOHARM* (Normal Ogive Harmonic Analysis Robust Method)، الموضوع من قبل فراسر (Fraser, 1988)، حيث تمّت معالجة البيانات المطابقة للنموذج اللوجستي ثنائي المعلمة، والمتعلقة باستجابات (661) طالباً وطالبة على الاختبار المكوّن من (35) فقرة، وتمّ الكشف عن أحادية البعد باستخدام البرنامج المذكور باستخدام مؤشرين:

مؤشر تاناكا Tanaka's Index: وهو مؤشر على حسن المطابقة ما بين النموذج المستخدم والبيانات، ويعمل كمعامل تحديد وتلخيص لنسبة التباين المفسّر بواسطة النموذج، ومعادلته هي:

$$Y_{ULS} = 1 - \frac{Tr(R^2)}{Tr(C^2)} \dots \dots \dots (30)$$

حيث:

\underline{R} : مصفوفة التباين المشترك للبواقي (Residual Covariance Matrix).

\underline{C} : مصفوفة التباين المشترك للعينة (Sample Covariance Matrix).

Tr : مجموع القيم الواقعة على القطر الرئيسي للمصفوفة (Matrix Trace).

ويعتبر مؤشر تاناكا دليلاً على المستوى المقبول من المطابقة ما بين النموذج والبيانات إذا كانت قيمته أكبر من أو تساوي 0.90، في حين يعتبر دليلاً على المستوى الجيد من المطابقة ما بين النموذج والبيانات إذا بلغت قيمته 0.95 أو أكثر، أما المطابقة التامة بين النموذج والبيانات فتحصل عندما تبلغ قيمته 1.0 (Jasper, 2010).

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

وبهدف تقدير قيمة هذا المؤشر فقد أوصى فراسر و ماك دونالد (Fraser & McDonald, 1988) بمقارنة قيمة هذا المؤشر مع قيمة الخطأ المعياري للبواقي كمعيار، والتي يتم حسابها من خلال قسمة القيمة 4.1 على الجذر التربيعي لحجم العينة؛ أي:

$$RMSR \leq \frac{4.1}{\sqrt{N}} \dots \dots \dots (31)$$

ويعرض الجدول 4 إحصائيات أحادية البعد للاختبار بصورته النهائية المكوّن من (35) فقرة، والمطبّق على العينة الاستطلاعية البالغ عددها (661) طالباً وطالبة.

جدول (4): إحصائيات أحادية البعد للاختبار بصورته النهائية المكوّن من (35) فقرة، والمطبّق على العينة الاستطلاعية.

العينة الاستطلاعية		إحصائيات أحادية البعد
نقطة القطع الاحتمالية قبول قيمة الإحصائي	القيمة	
	0.072	مجموع مربعات البواقي (دون القطر الرئيسي لمصفوفة البواقي)
0.146803549	0.011	الجذر التربيعي للوسط الحسابي لمربعات البواقي (دون القطر الرئيسي)
	0.951	مؤشر Tanaka لحسن المطابقة

يتبين من الجدول (4) أنّ مؤشر تاناكا قد تحقق فيه المستوى الجيد من المطابقة ما بين النموذج والبيانات إذ زادت قيمته عن (0.95)، كما يتبين من الجدول (4) أيضاً أنّ قيمة الجذر التربيعي للوسط الحسابي لمربعات البواقي والبالغة (0.011) تقريباً منخفضة وتكاد تقترب من الصفر، وهي أيضاً كما

يلاحظ أقل من المعيار أو نقطة القطع الموضوعية لاحتمال قبول قيمة الإحصائي، وهذا دليل كاف على تأهيل بيانات العينة الاستطلاعية لتحقيق افتراض أحادية البعد.

الاستقلال الموضوعي: Local Independence

تمّ التحقق من افتراض الاستقلال الموضوعي لبيانات العينة الاستطلاعية من خلال مؤشر Z_{Q_3} ، والذي يتمّ حسابه وفقاً للخطوات الآتية:

حساب ما يسمى مؤشر ين (Yen's Index) المعروف بـ Q_3 ، ويُعرّف على أنه معامل ارتباط بيرسون للبواقي الناتجة من النموذج المتعلق بنظرية الاستجابة للفقرة بين زوج من الفقرات بعد ضبط السمة المقدّرة، ويتمّ حساب قيمة الباقي لاستجابة المفحوص i على الفقرة j وفقاً للمعادلة:

$$d_{ij} = u_{ij} - T_j (\hat{\theta}_i) \dots \dots \dots (32)$$

حيث:

u_{ij} : استجابة المفحوص i على الفقرة j .

$T_j (\hat{\theta}_i)$: العلامة الحقيقية على الفقرة j عند مستوى القدرة المقدّرة $\hat{\theta}$ للمفحوص i .

وعند إعطاء الوزن صفر للإجابة الخاطئة، والوزن 1 للإجابة الصحيحة في حالة البيانات ثنائية الاستجابة كما هي الحالة في الدراسة الحالية تكون:

$$T_j (\hat{\theta}_i) = P_j (\hat{\theta}_i) \dots \dots \dots (33)$$

حيث: $P_j (\hat{\theta}_i)$ احتمال الاستجابة الصحيحة عند مستوى القدرة المقدّرة $\hat{\theta}$ للمفحوص i على الفقرة j .

وبحساب جميع قيم البواقي لاستجابات جميع المفحوصين عند كل مستوى قدرة مقدّرة على فقرتين من فقرات الاختبار مثل j, j' يكون:

$$Q_{3jj'} = r(d_j, d_{j'}) \dots \dots \dots (34)$$

حيث r هو معامل ارتباط بيرسون بين متغير البواقي على الفقرة j أي (d_j) ، و متغير البواقي على الفقرة j' أي $(d_{j'})$.

– حساب قيم Q_3 لجميع أزواج الفقرات والبالغ عددها $\frac{n(n-1)}{2}$ حيث n عدد فقرات الاختبار.

– حساب قيم مؤشر Z_{Q_3} من خلال عمل تحويل فيشري Fisher Transforming لقيم Q_3 وفقاً للمعادلة:

$$Z_{Q_3} = \frac{1}{2} \text{Ln} \frac{1+Q_3}{1-Q_3} \dots \dots \dots (35)$$

حيث: Ln هو اللوغاريتم الطبيعي (أساسه العدد النيبيري $e = 1.718$).

– حساب الانحراف المعياري لقيم Z_{Q_3} من خلال المعادلة:

$$\sigma_{Z_{Q_3}} = \sqrt{\frac{1}{N-3}} \dots \dots \dots (36)$$

حيث: N عدد المفحوصين أو عدد أفراد العينة (Kim & Cohen, 2005).

– للحكم على تحقق الاستقلال الموضعي لفقرتين يجب أن تقع قيمة Z_{Q_3} المحسوبة لهاتين الفقرتين ضمن فترة ثقة بانحرافين معياريين عن الوسط الحسابي لقيم Z_{Q_3} المحسوبة.

– إذا كانت أزواج الفقرات التي تتحقق بها الاستقلالية الموضعية أكبر منها لأزواج الفقرات التي تتحقق بها الاعتمادية أو التبعية الموضعية فيعتبر هذا مؤشراً على تحقق الاستقلال الموضعي للاختبار ككل.

تم استخدام برنامج (Local Dependence Indices for Dichotomous Items) **LDID** الذي وضعه كل من كيم و كوهن (Kim & Cohen, 2005)، لحساب مؤشر Z_{Q_3} لاستجابات العينة الاستطلاعية البالغ عددها (780) طالباً وطالبة على فقرات الاختبار البالغ عددها (35) فقرة، ويعرض الجدول (5) نتائج افتراض الاستقلال الموضعي لفقرات الاختبار لدى بيانات عينة الدراسة الاستطلاعية.

جدول (5): نتائج التحقق من افتراض الاستقلال الموضعي لفقرات الاختبار لدى بيانات عينة الدراسة.

العينة	الإحصائي	
595	عدد الأزواج	
-0.214	القيمة الدنيا	مدى القيم المشاهدة
0.199	القيمة العظمى	
-0.020	الوسط الحسابي المشاهد	
0.060	الانحراف المعياري المشاهد	
-0.029	الوسط الحسابي النظري	
0.001	التباين لـ Z_{Q_3}	
0.036	الانحراف المعياري لـ Z_{Q_3}	
0.072	الانحراف المعياري لـ $2 \times Z_{Q_3}$	
-0.091255	الحد الأدنى	فترة الثقة
0.052244	الحد الأعلى	

يلاحظ من الجدول (5) أن قيمة الوسط الحسابي المُشاهد لمؤثر Z_{Q_3} الخاصة بأزواج الفقرات البالغ عددها (595) لعينة الدراسة الاستطلاعية، قد بلغت (-0.020) والتي تقع بين الحد النظري الأدنى لفترة الثقة لقيم Z_{Q_3} البالغة قيمته (-0.091) وبين الحد النظري الأعلى لفترة الثقة البالغة قيمته (0.052).

وبالاعتماد على نتائج فترة الثقة لقيم Q_3 المحوَّلة إلى قيم Z الفُشرية المناظرة لها والمعروضة في الجدول (5) السَّابق، فقد تمَّ رصد عدد أزواج الفقرات التي وقعت قيمة Z_{Q_3} لها ضمن فترة الثقة المحقَّقة لشرط الاستقلال الموضوعي على أنَّها إما معتمِدة أو مستقلَّة، والجدول (6) يبين التكرارات والنسب المئوية لحالتي الاستقلال الموضوعي لدى بيانات عينة الدراسة الاستطلاعية.

جدول (6): التكرارات والنسب المئوية لحالتي الاستقلال الموضوعي لدى بيانات عينة الدراسة الاستطلاعية

العينة		حالة الاستقلال الموضوعي
النسبة المئوية	عدد الأزواج	
20.8	124	معتمد
79.2	471	مستقل
100.0	595	الكلي

يلاحظ من الجدول (6) أنَّ عدد أزواج الفقرات التي وقعت خارج مدى فترة الثقة (124)؛ أي ما نسبته (20.8%) من عدد الأزواج الكلي، بينما كان عدد أزواج الفقرات التي وقعت ضمن مدى فترة الثقة (471)؛ أي ما نسبته (79.2%) من عدد الأزواج الكلي، وهذا يبين أنَّ عدد أزواج الفقرات التي حققت الاستقلالية الموضوعية أعلى بأربعة أضعاف تقريباً من عدد أزواج الفقرات التي حققت التَّبعية الموضوعية، وهذا مؤشِّر كافٍ على تحقُّق افتراض الاستقلال الموضوعي لدى بيانات عينة الدراسة الاستطلاعية.

التوصيات

- يوصي الباحث بأن تقوم وزارة التربية والتعليم بإنشاء بنك يحوي مثل هذه الاختبارات المقننة وتطبيقها من خلال المدارس.
- استخدام الاختبار من قِبل معلمي ومعلمات الرياضيات ومن قِبل الباحثين حيث تم التأكد من خصائصه السيكومترية.
- الاستفادة من المنهجية المستخدمة في بناء الاختبار لبناء اختبارات لمقررات أخرى والتأكد من خصائصها السيكومترية.

References (Arabic & English)

- Adas, M. (2001). *Entrance to the kindergarten*, 2nd ed., Dar EIFeker AlArabi, Amman, Jordan.
- Adedoyin, O. & Mokobi, T. (2013). Using IRT Psychometric Analysis in Examining the Quality of Junior Certificate Mathematics Multiple Choice Examination Test Items. *International Journal of Asian Social Science*. 3(4), 992-1011.
- Al-Absi, M. (2002). *Development of a model based on (NCTM) standards and measurement of its impact on achievement, thinking and attitudes of students in the basic stage*. Unpublished MA Thesis, faculty of Education, Amman Arab University: Amman, Jordan.
- Ithawi, Mohammed Menem Salman. (2009). *Using Rasch Model according to Latent Traits Theory in building achievement test in a course of psychology of individual differences*. Unpublished MA Thesis, Baghdad University, College of Education-Ibn Rushd.
- Al-Aythawi, M. (2009), *Using Rasch Model in the Latent Trait Theory for Constructing an Achievement Test in Psychology of Individual Differences*, Unpublished MA Thesis, faculty of education, Baghdad University, Iraq.
- Al-Hakami, Ali. (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.
- Ali, Neda'. (2012). *The Effectiveness of Using Rasch Model in Constructing a Criterion Reference Test in Course of Measurement and Evaluation in Education*. Unpublished Ph.D. Damascus university, Syria.
- Alkhaiat, M. (2012) Fitness of Achievement Test in Mathematical Skills for Ninth Grade to Rash Model, *Alaqsa Educational Journal*. 16 (1), 87-111.

- Allam, S. (1990). *Constructing an objective-referenced scale of a basic knowledge for preparing psychological and educational research plan according to Rasch model*. Journal of the Faculty of Education, University of Al-Azhar 17, 129- 182.
- Allam, S. (2005). *Models of uni and multi-dimensional in Item-Response Theory, Principles and Application in Educational and Psychological Measurement*, 1st ed., Dar El-fikr El-arabi for publication and distribution, Cairo.
- Al-Tarawneh, S. (2016). *Constructing Bullying Scale for Higher Basic Students Classes in Jordan Using Rating Scale Model*. Educationan 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, 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 Education. 7(2), 40-54.
- Birnbaum, A. (1968). *Some latent trait models and their use in inferring an examinee's ability*. In F. M. Lord & M. R. Novick, *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.
- Da'ana, Z.(2005) *The Construction of a Tailored Test in Mathematics for Grade 8 Students Using the Pyramidal Strategy According to Rasch Model in the Latent Trait Theory*. Derasat for Educaional Research,32, 42-61.
- Dabous, M.(2016) *Using Item Response Theory in Constructing the Item Pool in Criterion- Referenced Testing with Dichotomous and*

- polytomous Items according to Two – Parameter Logistic Model.* Najah University Journal of Human Sciences, 30 (7), 1453-1480.
- Edwards, H. and ASlcock, I. (2010). *Using Rasch analysis to identify uncharacteristic responses to undergraduate assessment.* Teacher Mathematics and its Application, 1(29), 165-175.
 - Embrteson, Susan, Reise, Steven. (2000). *Item Response Theory for Psychologists*, Lawrence Erlbaum Associates.
 - Hamblton, R. & Swaminathan, H. (1985). *Item response theory: principles and applications*, Kluwer–Nijhoff publishing, Boston.
 - Hamblton, R. K., & Swaminathan, H. (1989). *Item Response Theory Principles And Applications.* Kluwer-Nijh of publishing, Boston, U.S.A.
 - Hijazi, Taghreed. & Alkhateeb, Abdallah. (2014). *The Compatibility between the Classical Test Theory and the two Parameter Logistic models in Item Stratification of CriterionReferenced Test in Provisions of Recitation and Tajweed.* An-Najah University Journal for Research - B (Humanities), 28 (10). 2239-2270.
 - Jasper, F. (2010). *Applied Dimensionality and Test Structure Assessment with The START-M Mathematics test.* The International Journal of Educational and Psychological Assessment 6(1), 109-126.
 - Lord, F. & Stocking, M. (1989). *Item Response Theory.* Pergamon Press, New York.
 - Ludlow, L. Haley, S. (1995). *Rasch model Logits: interpretation, use, and transformation.* Educational and Psychological Measurement, 55, 967 - 975.
 - Mahmoud F. (2016). *The psychometric characteristics of an achievement criterion-referenced test in educational psychology course, constructed according to Rasch Model.* Educational Journal, 30(120), 171-189.

- Mahmoud, L. & Sabah, S.(2016). *Developing an Achievement Test of Geometry for the Fifth Graders Using Rasch Model*. Derasat for Educational Research,43, 1353-1367.
- Majdee, I. (2006). *Teaching Mathematics for Students having learning difficulties*.Ciro: Alam Alkotob.
- Mohana, W. (2007). *Evaluation of Computerized mathematics course on the net in Jordan*. Sixth International Internet Education Conference and Exhibition, September 2-4-2007 ICT Learn 2007.
- Mostfa, Mohammad. (2007). *Guidelines suggestion for strategy to extend and Active Results of International Study TIMSS*, confidence of quality of teaching in Palastine: Reality effect and confrontation, Ministry of High Education in Palastine, Palastine.
- Odeh, A. (2010). *Measurement and Evaluation in the teaching process*, 4Th ed., Dar Al-Amal for publication and distribution, Irbid, Jordan.
- Olwan, Belal, (2015). *Using Rash Model to construct place preparation for preparatory students*, (Unpublished MA Thesis). Faculty of Education, Baghdad University, Bagdad.
- Onn, D. (2013). *Classical test theory versus item response theory: An evaluation comparability of item analysis results*. Joint Admissions and Matriculation Board, 1-23.
- Osterlind, S. (1983). *Test item bias*. Beverly Hills: Sage Publications.
- Sabah, A. (2016). *Using Item Response Theory for Constructing a Criterion- Referenced Test in English Language (1) for Al-Quds Open University Students According to Rasch Model*. Journal of Al-Quds Open University for Educational and Psychological Research and Studies. 6(20), 133-151.
- Sawalmeh, Y. & Alzoabi, A. (2017). *Constructing a Test to Measure Critical Thinking Skills Using Item Response Theory*. Journal of the Union of Arab Universities for Education and Psychology. 15(3), 55-92.