

The Compatibility between the Two-Parameter Logistic Model and Classical Theory in Constructing an Achievement Test in General Science for 6th Grade Students

مدى التوافق بين النموذج ثنائي المعلمة والنظرية التقليدية في بناء اختبار تحصيلي في مبحث العلوم العامة للصف السادس الأساسي

Husni Salem^{1*}, Nedal Al-Shraifin¹.

¹Yarmouk University, Irbid, Jordan.

ARTICLE INFO

Article history:

Received 30 Oct 2017

Accepted 01 Mar 2018

Published 01 Jan 2020

*Corresponding author:

Yarmouk University, Irbid, Jordan.

Email: husni_salem@yahoo.com.

Abstract

This study aimed to assess the compatibility between the two-parameter logistic model and classical theory in constructing an achievement test in science for sixth grade students. To achieve this purpose, a science achievement test consisting of 50 multiple-choice items was constructed and applied to a sample of 567 sixth graders from eight randomly selected public schools in Jarash. The responses were analyzed using SPSS, BILOG-MG 3, NOHARM, and LDID programs to obtain statistical indicators for both examinees and items. The results indicate that the items of the sixth-grade science test can be classified as either matching or non-matching according to item response theory using the two-parameter logistic model and classical theory. There is a notable similarity between the two theories in classifying the items, with an agreement rate of 92% between the conventional indicators and the two-parameter logistic model indicators. This means that 92% of the 50 test items were judged similarly by both theories, while the remaining 8% showed discrepancies. Additionally, the results revealed a significant statistical difference ($\alpha = 0.05$) in reliability between the two theories. Using the modified Feldt equation for non-independent samples, the analysis showed $F = 1.99$ in favor of the item response theory using the two-parameter logistic model. Furthermore, there was a significant statistical difference ($\alpha = 0.05$) in criterion validity between the two theories, with a t-test for non-independent samples yielding $t = 2.16$ in favor of criterion validity according to item response theory. Finally, the results indicated a strong correlation between the capabilities of test examinees and their raw scores or standardized scores, which were positively correlated and statistically significant ($\alpha = 0.05$), classified as a very high correlation.

Keywords: Classical Theory in Measurement, Item Response Theory, Achievement Test, Two-Parameters Logistic Theory.

هدفت الدراسة إلى الكشف عن مدى التوافق بين النموذج ثنائي المعلمة والنظرية التقليدية في بناء اختبار تحصيلي في مبحث العلوم العامة للصف السادس الأساسي. ولتحقيق هدف الدراسة تم بناء اختبار تحصيلي في العلوم العامة تكون في صورته النهائية من (٥٠) فقرة من نوع الاختيار من أربعة بدائل. طبق على أفراد عينة الدراسة المكونة من (٥٦٧) طالباً وطالبة من طلبة الصف السادس الأساسي في جرش موزعين على (٨) مدارس تم اختيارها بالطريقة العشوائية العنقودية. حلت استجابات أفراد عينة الدراسة باستخدام البرامج الآتية: SPSS، BILOG-MG3، NOHARM، وبرنامج LDID، وذلك للحصول على المؤشرات الإحصائية الخاصة بالأفراد والفقرات وفقاً للنموذج ثنائي المعلمة والنظرية التقليدية في القياس.

أظهرت النتائج الخاصة في تصنيف الفقرات للاختبار المعد في مبحث العلوم للصف السادس الأساسي إلى (مطابقة أو غير مطابقة) وفقاً لـ: (نظرية استجابة الفقرة باستخدام النموذج الثنائي المعلمة، النظرية التقليدية) وجود تشابه كبير بين النظريتين في تصنيف الفقرات، حيث بلغت نسبة الاتفاق بين نتائج المؤشرات التقليدية ونتائج مؤشرات النموذج ثنائي المعلمة في تصنيف الفقرات إلى (مطابقة أو غير مطابقة) (٩٢٪)، أي أن (٩٢٪) من فقرات الاختبار والبالغ عددها (٥٠ فقرة) اتفقت النظريتان في الحكم عليها، أما النسبة المتبقية (٨٪)؛ فقد اختلفتا في الحكم عليها.

كما بينت النتائج وجود فرق دال إحصائياً عند مستوى الدلالة الإحصائية ($\alpha = 0.05$) بين معاملي الثبات لنموذج الاختبار وفقاً للنظريتين، باستخدام معادلة فلدت المعدلة Modified Feldt للعينات المترابطة في النظريتين، إذ بلغت قيمة F (١,٩٩٨)؛ لصالح نظرية الاستجابة الفقرة باستخدام النموذج ثنائي المعلمة. وبينت النتائج وجود فرق دال إحصائياً عند مستوى الدلالة الإحصائية ($\alpha = 0.05$) باستخدام معادلة اختبار (t) لمعاملي ارتباط للعينات المترابطة في النظريتين بين معاملي الصدق المحكي للاختبار وفقاً للنظريتين إذ بلغت قيمة t (٢,٠١٦)؛ لصالح نظرية استجابة الفقرة باستخدام النموذج ثنائي المعلمة.

وفيما يخص نتائج تدريج قدرة المفحوصين على الاختبار وفقاً للنظريتين (التقليدية، استجابة الفقرة للنموذج الثنائي المعلمة)؛ فقد أشارت أن العلاقة الارتباطية بين القدرة من جهة وبين (العلامة الخام أو العلامة الخام المحولة معيارياً) من جهة أخرى؛ كانت طردية وذات دلالة إحصائية عند مستوى الدلالة الإحصائية ($\alpha = 0.05$) مُصنفة على أنها علاقة عالية. وفي ضوء النتائج السابقة خلصت الدراسة إلى مجموعة من التوصيات.

الكلمات المفتاحية: النظرية الكلاسيكية في القياس، نظرية استجابة الفقرة، الاختبار التحصيلي

١. المقدمة

١,١ الإطار النظري والدراسات السابقة

لقد شهد علم القياس والتقويم تغيرات ملحوظة خلال العقود الماضية في مفاهيمه ومبادئه وأساليبه، فقد اهتم الخبراء بابتكار أساليب وطرق تُرشد عمليات القياس والتقويم المعاصر وتوجهها، وقد واكب ذلك تطورات في أسس بناء أدوات القياس المتنوعة، وطرق جمع وتحليل وتفسير البيانات المستمدة من خلاله، ويرجع الاهتمام بطرق القياس والتقويم التربوي والنفسي - وأساليبيها إلى الدور البالغ الأهمية للتقويم في صنع القرارات التربوية المختلفة. فمن خلال التقويم يمكن التعرف على فاعلية البرامج التربوية، وتوجيه مسارها في جميع المراحل، كما يمكن من خلالها تقويم المناهج الدراسية من منظور يتسم بالشمولية (علام، ٢٠٠٧).

إن التقويم يُعد أحد أهم مكونات العملية التعليمية التعلمية إذ يقع على عاتق المعلم النوعية المتميزة من التقويم الصفي، ويُعتمد على نتائجه في تحسين العملية التعليمية التعلمية من حيث أهدافها ووسائلها وتقنياتها، كما يمثل التقويم أحد أهم المداخل الحديثة لتطوير التعليم، فمن خلاله يتم التعرف إلى أثر كل ما تم التخطيط له وتنفيذه من عمليات التعليم والتعلم، ونقاط القوة والضعف فيها، ومن ثم اقتراح الحلول التي تساهم في التأكيد على نقاط القوة وتدعيمها، وتلافي مواطن الضعف وعلاجها (علام، ٢٠٠٧).

فالمتبع لحركة القياس والتقويم النفسي- والتربوي بشكل عام، والاختبارات التحصيلية في المؤسسات التعليمية بشكل خاص يلاحظ تزايد الاهتمام بالاختبارات التحصيلية لدورها الهام في العملية التعليمية التعلمية، وقد أشاد بذلك جي (Gay, 1980)، والذي قال إن الاختبارات تستطيع قياس التعلم في جميع مستويات الأهداف، كما أنها تغطي مجالاً واسعاً

من السلوكيات، فمن خلالها يتمكن المعلم من معرفة المستوى التحصيلي للمتعلمين، ومدى فهمهم واستيعابهم، كما أن نتائج الاختبارات تقدم تغذية راجعة للمعلم والمتعلم على حد سواء، ولذلك ينبغي أن يكون الاختبار جزءاً متمماً للعملية التعليمية التعلمية، وأن يراعى ذلك أثناء التخطيط للتعليم؛ حتى تتحقق إمكانية جعل الاختبارات التحصيلية مُعينات تعليمية.

إن تعدد أغراض التقويم وتنوع أدواته وتعددتها في العملية التعليمية التعلمية، وكذلك اتساع مفهوم التحصيل يتطلب تنوعاً في هذه الأدوات والأساليب، بحيث تقيس مختلف الأهداف المتوقعة للمجالات الدراسية المختلفة، بيد أن العديد من المعلمين ما زالوا يركزون على الاختبارات التحصيلية في تقويم أداء طلبتهم (علام، ٢٠٠٧).

وقد يستخدم التقويم أدوات عدة منها: الاختبارات، أو أدوات الملاحظة، أو قوائم التقدير، أو مقاييس الاتجاه أو وسائل التقدير الذاتية. وتمثل الاختبارات التحصيلية إحدى أهم أدوات قياس التحصيل المعرفي للمتعلم، وقد لخص الجلي (٢٠٠٥) نقاط الضعف في إعداد الاختبارات التحصيلية التي أسفرت عنها نتائج الدراسات بالآتي:

- تدني قدرات المعلمين على صياغة الفقرات الموضوعية، وتوجيههم إلى الاختبارات المقالية.
- عدم تغطية فقرات الاختبار لجوانب المقرر الدراسي الذي أعد لقياس تحصيل المتعلمين.
- عدم اهتمام المعلمين بإعداد جداول المواصفات للاختبارات التي يعدونها.
- انخفاض المستويات المعرفية التي تستهدف فقرات الاختبار قياسها لدى المتعلمين.
- عدم قيام المعلم بإجراء عمليات الضبط الإحصائي للاختبارات؛ للتعرف إلى مدى صدقها وثباتها.

وتعد الاختبارات واحدة من وسائل التقويم المتنوعة التي يُستند إليها في اتخاذ القرارات المهمة التي تخص الفرد والمجتمع، وانتشر استخدام الاختبارات انتشاراً واسعاً في العديد من المجالات، حيث تصمم الاختبارات لأهداف متنوعة منها: اختيار شخص لوظيفة ما، أو لأغراض التصنيف كتحديد مسار المتعلمين بما يتناسب مع قدراتهم ومهاراتهم، وفي تقويم تحصيل المتعلمين من خلال الدرجات التي يحصلون عليها في الاختبارات الصفية، وبذلك يمكن العمل على تحسين وتطوير العملية التربوية والتعليمية والسير بها إلى الأفضل عن طريق تطوير هذه الاختبارات وتحسين قدرتها لقياس نواتج التعلم، تحريرية أو أدائية (Allen & Yen, 1979).

فالاختبار إجراء منظم لقياس عينة من سلوك الأفراد، وتكتسب عملية تقويم الأفراد أهمية كبيرة بقدر أهمية القرارات المبينة عليها، وبقدر خطورة القرارات الخاطئة التي يمكن أن تترتب على ذلك في المواقف والمجالات المتعددة على مستوى الفرد والمجتمع إلى حد قد يصعب معالجته، أو قد يحتاج علاجه إلى وقت طويل مما يعيق عملية التنمية ومواكبة تطور المجتمعات الأخرى، وللحصول على قرارات أكثر دقة يجب توفر معلومات صادقة ودقيقة من خلال التخطيط والإعداد الجيد للاختبارات (عوذة، ٢٠١٠).

وهناك أشكال عدة للاختبارات منها: الاختبارات ذات الإجابة المختارة، والاختبارات ذات الإجابة المنتجة. ولل فقرات ذات الإجابة المختارة أشكال متعددة تستخدم لقياس تحصيل المتعلمين، ولكن فقرات الاختبار من متعدد تتمتع بانتشار واسع، وتستخدم لأغراض متعددة ومتنوعة. لذا فإن المتعلم يتعرض في المراحل التعليمية المختلفة إلى عدد كبير من الاختبارات باختلاف أشكالها؛ إذ تتعدد أشكال فقرات الاختبارات التحصيلية من حيث صياغتها، فمنها ما يتطلب من المفحوص أن يعطي الإجابة بلغته الخاصة مثل: الأسئلة المقالية، والأسئلة التي تتطلب الإجابة القصيرة وأسئلة التكميل. ومنها ما يتطلب من المفحوص أن يختار الإجابة الصحيحة من بين مجموعة من البدائل كفقرات الاختبار من متعدد، وفقرات الصواب والخطأ، وفقرات المزاوجة، ويتوقف استخدام أحد هذه الأنواع من الفقرات على الهدف من عملية التقويم؛ إذ تستخدم الأسئلة المقالية في كثير من الأحيان، حينما يستهدف المعلم قياس القدرات التنظيمية، وقدرة الطالب على التعبير وإبداء الرأي، بينما يشيع استخدام الأسئلة الموضوعية في قياس مدى امتلاك الطالب لمهارات التذكر والفهم والتطبيق، ضمن عينة أوسع من محتوى المادة المراد معرفة درجة إتقان الطالب لها، ويتحكم في اختيار الشكل المناسب للفقرات مجموعة من العوامل منها: طبيعة المادة الدراسية، وعمر المفحوص، وغرض الاختبار، وعدد الطلبة، وظروف التطبيق، ومهارة المعلم في كتابة الأسئلة (Gronlund & Linn, 1990).

ويرى براون (Brown, 1976) أن اختيار شكل الفقرات المناسب يعتمد على خصائص الفقرة، ومهارة المعلم في كتابة الفقرات، وفلسفة التربية، فمثلاً إذا رأى المعلم أن طلابه يجب أن يكونوا قادرين على تذكر المادة التعليمية بدلاً من

التعرف إلى الاستجابات الصحيحة، فعليه أن يستخدم اختبارات التكميل والإجابة القصيرة، خاصة أن بناء فقرات التكميل أو الإجابة القصيرة أسهل من بناء فقرات الاختيار من متعدد جيدة المستوى، وفي كثير من الأحيان يمكن قياس الهدف بأكثر من نوع من الفقرات؛ إذ يمكن ترجمة الفقرة من نوع إلى آخر، ومع ذلك فغالباً ما يميل المعلمون إلى استخدام نوع من الفقرات دون الآخر.

ويرى كل من: جي (Gay, 1980) وعودة (٢٠١٠) أن فقرات الاختيار من متعدد تعد من أفضل أنواع الفقرات الموضوعية، وأكثرها شيوعاً وانتشاراً في الاختبارات التحصيلية؛ فهي تتميز بسهولة تصحيحها وموضوعيتها، وتوفر تغطية جيدة للمادة الدراسية، وتمتاز علامة الطالب عليها بدرجة عالية من الثبات، إضافة إلى أنها تحدد نتائج التعلم المقصودة بدرجة عالية، على الرغم من أن إعدادها يتطلب وقتاً طويلاً، وجهداً عالياً، ومهارة فائقة من قبل واضعها، كما أن فقرات الاختيار من متعدد قادرة على قياس نواتج التعلم في المستويات العقلية من المجال المعرفي، بدرجة تفوق فقرات المطابقة، وفقرات الصواب والخطأ، وفقرات التكميل، والإجابة القصيرة.

كما أشار أيكين (Aiken, 1987) إلى أن اختبارات الاختيار من متعدد أصبحت الأكثر انتشاراً من بين كل الاختبارات في التربية؛ وذلك بسبب كفاءتها، وتعدد استخداماتها، فبواسطتها يمكن قياس أهداف بسيطة، وأخرى مركبة في مختلف المواضيع الدراسية.

ولكي يكون الاختبار الذي يتم تقديمه إلى الطلبة جيد التكوين- يعكس مدى الأهمية التي من أجلها وضع من حيث كونه أداة حكم على تحصيل الطلبة وتصنيفهم إلى مستويات، أو من حيث كونه أساساً لتطوير المنهج الدراسي، أو تطوير نوعية الأسئلة نفسها - يراعى أن يتوفر فيه مجموعة من المعايير التي تحدد صلاحيته للاستخدام، وتعد هذه المعايير بمثابة الصفات الأساسية التي تحدد صلاحيته، ولذلك يراعى أن تتوفر في أي اختبار كي يحظى باحترام جميع المختصين، وبالتالي يتم الوثوق بنتائجه والاستفادة منها. إن اتصاف الاختبار بجميع هذه المعايير يعني أنه صالح لقياس السمة أو الظاهرة المراد قياسها، أما إذا فقد الاختبار هذه الشروط والمعايير فإنه يفقد صلاحيته، ويقلل من ثقة الفاحص والمتعلم به، لذلك لا بد أن يتصف الاختبار الجيد بمجموعة من الخصائص، منها: الموضوعية، الصدق والثبات (عمر وفخرو والسبيعي وتركي، ٢٠١٠). كما أن هناك صفات ثانوية يراعى أن تتوفر في الاختبار تتمثل بسهولة تطبيقه، وسهولة تصحيحه، والاقتصاد أو التكلفة المادية. ومما يجعل الاختبارات المقننة جيدة، وجود صفات خاصة بها، تسهل عملية تطبيقها، وتصحيحها وتحليل نتائجها بسهولة ودون تحيز، ودون إرباك للمتعلم (أبو غريبه، ٢٠٠٨).

ولما كانت الاختبارات وسيلة فعالة في قياس التحصيل العلمي عند المتعلمين في مختلف مستوياتهم البسيطة والمعقدة، فإنه يتطلب الأمر ضرورة الاهتمام بها؛ لمساعدة المعلمين في تحسين فاعليتهم عبر اكتساب مفاهيم ومعلومات ومهارات، تمكنهم من تصميم وإعداد اختبارات تتناسب وقدرات المتعلمين من أجل تقويم تحصيلهم العلمي، والحكم على مدى استعدادهم ومقدورهم. كما أن إعداد اختبارات جيدة التكوين يتطلب مهارات تتوافر في المعلم تتطلب بدورها توجيهها من المسؤولين عن العملية التعليمية التعلمية منها: قيام المعلم بتوجيه أسئلة تتعلق بموضوعات المادة الدراسية، وامتلاكه لأسس كتابة الفقرات بأنواعها المختلفة. ويتضح الاهتمام بالاختبارات التحصيلية من خلال التطور الكبير في الاستراتيجيات المستخدمة في بنائها، وتحليل الفقرات، ونتائج الاختبار وفق النظريتين: التقليدية والحديثة في القياس (كاظم، ٢٠٠١).

ويذكر مجيد (٢٠٠٧) أن إعداد اختبار التحصيل يمر في سلسلة من الخطوات تسير بترتيب معين، بحيث تمهد كل خطوة لما بعدها، وتعتمد على ما قبلها، وهي: (تحديد الغرض من الاختبار التحصيلي من خلال صياغة الأهداف التعليمية، وتحديد نواتج التعلم التي سيقاس بها الاختبار أو الأهداف المتوقعة تحقيقها، وتحليل المحتوى الدراسي، وبناء جدول المواصفات أو لائحة المواصفات، كتابة فقرات الاختبار، التجريب الميداني للاختبار، إخراج كراسة الاختبار).

إلا أن الباحثين في بناء الاختبارات وتطويرها يواجهون مشكلات متعددة ومعقدة، قدمت النظرية التقليدية حلولاً لبعض هذه المشكلات، إلا أنها عجزت عن حل مشكلات أخرى معاصرة، ولعل من أهم المشكلات التي واجهتها هذه النظرية تمثلت بصورة أساسية في أن جميع الخصائص السيكونية للاختبارات تعتمد على خصائص الأفراد المطبق عليهم الاختبار، وأن نتائج الاختبار وما يُبنى عليها من تفسيرات يعتمد على خصائص عينة الفقرات التي يشتمل عليها الاختبار (Allen & Yen, 1979).

ويرى النجار (٢٠١٠) أن النمو المتزايد في الوقت الحاضر لتفريد التعليم وقياس نتائج البرامج التعليمية بصورة صادقة وموضوعية ودقيقة، وتمكين المعلمين والمربين من تقدير المستوى الفعلي لأداء المتعلمين والاستفادة منه في عملية التوجيه والإرشاد النفسي- والتربوي والتعليمي والمهني، أدى إلى زيادة الاهتمام من جانب علماء القياس وأخصائيي بناء

الاختبارات والمقاييس النفسية والتربوية في العقدين الأخيرين حول البحث عن وسائل أكثر ملاءمة لتيسير عملية القياس ومقابلة المطالب التعليمية، ومواجهة الكثير من المشكلات السيكومترية ومعالجة أوجه القصور المختلفة التي عجزت عن حلها النظرية السيكومترية التقليدية أو النظرية التقليدية التي سادت وسيطرت على الفكر التربوي فترة طويلة من الزمن تزيد عن نصف قرن، ومن أبرز هذه العيوب والمشكلات التي أشار إليها عودة (٢٠١٠):

- حذف الفقرات التي لا تميز بين أفراد العينة، بغض النظر عن قيمة تلك الفقرات أو جودة تمثيلها للأهداف المراد قياسها.
 - يتم اتخاذ الجماعة كمعيار أو حد لتفسير درجة الاختبار الأمر الذي يفقد البصيرة بموقع الفرد وجماعته التي ينتمي إليها على السواء بالنسبة لأهداف التعلم وأسباب ذلك، وطرق مواجهته.
 - تختلف جميع الخصائص السيكومترية لفقرات الاختبارات (معاملات الصعوبة، ومعاملات التمييز والثبات والصدق) باختلاف طبيعة وخصائص عينة الأفراد المستخدمة، وهذا يشير إلى أن معاملات الصعوبة للفقرات تتذبذب وتختلف باختلاف مستوى درجات هذه السمات أو القدرات، كما تختلف باختلاف انتشار درجات هذه السمات أو القدرات، بالإضافة إلى أن الدرجات الخام المستمدة من هذه الاختبارات ليس لها معنى أو دلالة في حد ذاتها، حيث يختلف معناها ودلالاتها باختلاف مدى صعوبة أو سهولة فقرات الاختبار، فإذا اشتمل اختباران على نفس العدد من الفقرات التي تقيس سمة أو قدرة معينة لدى مجموعة من الأفراد، فإن درجة كل فرد ستختلف باختلاف صعوبة الفقرات أو سهولتها في كل منها، وهذا يجعل الاستفادة من نتائج الاختبارات تقتصر على العينة المستخدمة التي استمدت منها هذه النتائج أو عينة مماثلة بقدر الإمكان.
 - هيمنة المنحنى الاعتدالي المعياري، فأى ابتعاد عنه يعد انحرافاً سواء في مدى صعوبة الفقرات أو سهولتها، إذ يؤدي ذلك إلى التواء توزيع اعتدالها، حيث ينبغي أن يكون توزيع الأداء مختلفاً تماماً عن المنحنى الاعتدالي المعياري، إذ أن الجهود التربوية تعدّ فاشلة بقدر اقتراب توزيع درجات الأفراد من المنحنى الاعتدالي (يوسف، ١٩٩١؛ Popham, 1980).
 - لا نستطيع الموازنة بين مستويات القدرة إذا أجاب الأفراد على فقرات مختلفة ومتباينة في صعوبتها، كون الموازنة بين الأفراد في السمة أو القدرة التي يقيسها الاختبار تقتصر على تطبيق فقرات الاختبار نفسها أو فقرات مكافئة لها على كل فرد من الأفراد (علام، ١٩٨٧).
- كما أظهرت النظرية التقليدية في القياس عجزاً واضحاً أمام متطلبات معاصرة مثل: بناء بنوك الأسئلة، وبناء الاختبارات المفصلة أو التكيفية (Adaptive Testing)؛ ولذلك فقد كانت جهود كثير من علماء القياس النفسي- والتربوي في السنوات الأخيرة كما بيّنوها في دراساتهم أمثال رايت، ميد وبيبل (Wright et al., 1980)؛ ولورد (Lord, 1980) متجه إلى تطوير نماذج سيكومترية جديدة غير تقليدية تفي وتحقق الموضوعية في القياس سميت باسم نماذج السمات الكامنة (Latent Trait Theory (LTT أو نماذج الاستجابة للفقرة (Item Response Theory (IRT) ولهذه النماذج مزاياها المفيدة ويفترض في هذه النماذج الحديثة (LTT) و (IRT) أنه يمكن التنبؤ بأداء الأفراد أو يمكن تفسير أدائهم في اختبار نفسي أو تربوي معين بناء على خاصية أو خصائص معينة مميزة لهذا الأداء تسمى السمات Trait وتحاول هذه النظرية تقدير درجات الأفراد في هذه السمات، ونظراً لصعوبة ملاحظة هذه السمات بطريقة مباشرة أو قياسها بصورة مباشرة، فإنه يجب تقديرها أو الاستدلال عليها من خلال استجابات الأفراد الملاحظة على فقرات الاختبار الذي يقيس السمة أو القدرة المطلوبة ولهذا أطلق عليها أسماء عدة منها نظرية السمات الكامنة (LTT) أو نظرية استجابة الفقرة (IRT) أو نظرية منحنى خصائص الفقرة (Item Characteristics Carve Theory) (ICC) واندرج تحت هذه النظرية مجموعة من النماذج (Models) وفقاً لتعدد معالم الفقرة (Hambelton & Swaminathan, 1985؛ علام، ١٩٨٢؛ الشراوي وآخرون، ١٩٩٧).
- وتعد نماذج السمات الكامنة نماذج أو دوال رياضية تحدد العلاقة بين أداء الفرد في فقرات الاختبار، وهو ما يمكن ملاحظته ملاحظة مباشرة وبين السمات أو القدرات التي يفترض أنها تكمن وراء هذا الأداء، ولهذا تعدّ هذه الدوال احتمالية وليست حتمية، وذلك يعني أن العلاقة التي تحددها الدالة تسير وفق نظرية الاحتمالات، والقيمة الاحتمالية لاستجابة فرد على فقرة اختبار تكون دالة لكل من السمة أو القدرة أو الخاصية التي يفترض أن الاختبار يقيسها لدى الفرد وخصائص الفقرة التي يحاول الإجابة عنها، أي أنها تفترض أن هناك دالة احتمالية تربط بين بارامترين أو معلمين أحدهما يتعلق بالفرد والآخر يتعلق بالفقرة التي يختبر بها (علام، ١٩٨٥؛ ١٩٩٠).

ولكل من هذه النماذج مميزات وعيوب، غير أن النموذج اللوجستي ثنائي المعلمة يعد من أفضل هذه النماذج وأكثرها استخداماً في التطبيقات العملية وبخاصة في مجال بناء الاختبارات العقلية التحصيلية لنتائجه المستقرة، حيث بينت نتائج دراسة رونالد وداتو (Ronald & Dato, 1983) - (المشار إليها في عيد، ٢٠٠٤) - والتي هدفت إلى انتقاء فقرات اختبار محكي المرجع باستخدام نماذج السمات الكامنة (أحادي البارامتر- ثنائي البارامتر) وذلك باستخدام طرق المحاكاة بالكمبيوتر - أفضلية النموذج ثنائي المعلمة (نموذج لورد) في الانتقاء للفقرات. كما أظهرت العديد من الدراسات التي هدفت إلى المقارنة بين النظرية التقليدية والنظرية الحديثة في القياس من خلال بناء المقاييس كدراسة كل من: (جمحاوي، ٢٠٠٠؛ Hernandez, 2009) أن سلوك نموذج المعلمتين كان الأكثر اتفاقاً دائماً مع المؤشرات التقليدية من النماذج الأخرى؛ لأنه الأكثر شبيهاً بالمؤشرات التقليدية في اختيار الفقرات. كما يعد أكثر النماذج قرباً في خصائصه السيكو مترية من النظرية التقليدية في القياس (علام، ١٩٩٠).

ولقد حاز نموذج " لورد" على اهتمام معظم الباحثين وعلماء القياس المعاصر وبخاصة في مجال بناء الاختبارات والمقاييس النفسية والتربوية المختلفة في مجالات أخرى كبنوك الأسئلة على سبيل المثال. وأصبح هذا النموذج يسود ويسيطر على الفكر النفسي- والتربوي في مجال القياس تمشيًا مع الحركة التطورية العالمية السريعة التي انتشرت في كثير من دول العالم ككندا وأستراليا والسويد وغيرها من الدول الأخرى التي جعلت من هذا النموذج واقعاً فعلياً في كافة مجالات القياس المختلفة وبخاصة في مجال العمليات التعليمية والتربوية (يوسف، ١٩٩١).

لقد أضافت النظرية الحديثة في القياس أبعاداً جديدة في مجال بناء الاختبارات وانتقاء الفقرات بشكل خاص بعد أن كان مقتصرًا في النظرية التقليدية اختيار الفقرات بناءً على خصائصها مثل: الصعوبة والتمييز، وارتباطها بالعلامة الكلية. كما ساهمت النظرية الحديثة في إدخال بعداً آخر يعتمد على ما يسمى منحى خصائص الفقرة، ومنحى معلومات الفقرة، ومنحى معلومات الاختبار، فهي بذلك ساعدت الباحثين في حل الكثير من مشاكل بناء الاختبارات وتطويرها لما تقدمه هذه المنحنيات من معلومات تساعد الباحثين في التنبؤ بنتائج الاختبار مُسبقاً خصوصاً في قضايا تصنيف الأفراد (Croker & Algina, 1986).

ويشير علام (١٩٨٦) إلى أن الهدف الأساسي للنظرية الحديثة والنظرية التقليدية في القياس هو تحديد العلاقة بين استجابات الأفراد لاختبار معين والسمة الكامنة وراء هذه الاستجابات، للإفادة في التنبؤ بسلوك الأفراد في مواقف مماثلة لاتخاذ قرارات معينة بشأنهم.

كما يرى عودة (٢٠١٠) أن هناك أسلوبين لاختيار الفقرات: أحدهما يعتمد على النظرية التقليدية (Classical Test Theory)، والآخر يعتمد على النظرية الحديثة في القياس (Modern Test Theory)، فالنظرية التقليدية تفرز فقرات تتمتع بخصائص معينة وكذلك النظرية الحديثة في القياس، وقد تتفق النظريتان في قبولهما أو رفضهما لهذه الفقرات وفقاً لخصائصها ومعاييرها.

وتأسيساً على ما سبق فقد جاءت فكرة هذه الدراسة وهي المقارنة بين خصائص الفقرات وفق النظرية التقليدية والنموذج الثنائي المعلمة في بناء اختبار تحصيلي في مبحث العلوم العامة للصف السادس الأساسي لبيان مدى التوافق بينهما. والسؤال الرئيس هنا ما مدى توافق النموذج الثنائي المعلمة والنظرية التقليدية في انتقائهما لفقرات الاختبار من حيث عددها وصعوبتها وتميزها وثباتها وصدقها وتدرج المفحوصين.

١,١,١ الدراسات السابقة

يتناول هذا الجزء عرضاً للدراسات العربية والأجنبية ذات الصلة بمشكلة البحث، وقد تم عرضها مرتبةً حسب تسلسلها الزمني من الأقدم إلى الأحدث، على النحو الآتي:

دراسة فان (Fan, 1998) التي هدفت إلى المقارنة بين النظرية التقليدية والنماذج الثلاثة لنظرية الاستجابة للفقرة (1, 2 & 3 Parameter Logistic Models)، من حيث خصائص فقرات الاختبار ومستويات القدرة للطلبة، ومن حيث مدى استقرار صعوبة الفقرات وتميزها عبر العينات المختلفة. وقد استخدم لهذا الغرض بيانات ثلاث عينات عشوائية مختلفة وكبيرة الحجم، حيث بلغ حجم العينة الأساسية (١٩٣٢٤٠) طالباً من طلاب الصف الحادي عشر- في العام (١٩٩٢) من الذين تقدموا لاختبار تكساس للمهارات الأكاديمية (Texas Assessment of Academic Skills (TAAS) وهو بطارية اختبارات جماعية المرجع. ويتكون الاختبار من ثلاث اختبارات فرعية هي: اختبار القراءة (٤٨) فقرة، واختبار الرياضيات (٦٠) فقرة، إضافة إلى اختبار الكتابة الذي استبعد من الدراسة لكونه من الاختبارات المقالية متعددة الاستجابة. أشارت متوسطات معاملات الارتباط المرتفعة أن معظمها أكبر من (٠,٨٠)، وبينت المؤشرات

الإحصائية المستمدة من الإطارين النظريين تشابها كبيرا بين الإطارين النظريين، كما لم يتضح تفوق أي من الإطارين النظريين على الآخر من حيث استقرار مؤشرات صعوبة الفقرات وتمييزها. إلا أن مؤشرات تمييز فقرات الاختبار المستمدة من كل نظرية كانت أقل استقرارا مقارنة بمؤشرات صعوبة الفقرات، كما أشارت النتائج إلى أن درجة استقرار تقديرات النظرية التقليدية ونظرية الاستجابة للفقرات يقل كلما زادت درجة اختلاف العينات، أي أن تقديرات العينات العشوائية البسيطة كانت أكثر استقرارا تليها عينات الجنس، وأقلها استقرارا عينات القدرات المرتفعة / المتدنية ومع أن النموذج اللوغاريتمي أحادي البارامتر كان أقل النماذج ملائمة لبيانات الاختبار، إلا أن تقديرات هذا النموذج كانت أكثر استقرار من تقديرات النموذجين ثنائي وثلاثي البارامتر.

كما قامت جمحاوي (٢٠٠٠) بدراسة هدفت إلى المقارنة بين النظرية التقليدية والنظرية الحديثة في القياس من خلال مقياس للقدرة الرياضية المطور وفق المؤشرات التقليدية ليوثم الطلبة في البيئة الأردنية وتحديدًا طلبة الصفوف: السادس والتاسع والعاشر، حيث قورنت خصائص فقرات هذا المقياس المكون من (٣٩) فقرة وفقا للنظريتين التقليدية والحديثة. وتكون مجتمع الدراسة من طلبة الصف التاسع في منطقة اربد، أما عينة الدراسة فقد تكونت من (١٠٦١) طالبا وطالبة من طلبة الصف التاسع في مديرتي اربد الأولى والثانية. وقد حلت إجابات أفراد الدراسة باستخدام برنامج (ASCAL) و (RASCAL) للتحليل بنماذج الفقرة الحديثة، وبرنامجي (ITEMAN) و (SPSS) للتحليل بالمؤشرات التقليدية، حيث بينت النتائج وجود اتفاق عالٍ بين الأسلوبين في تقدير صعوبة الفقرات وتمييزها وكذلك في توزيع القدرة للمفحوصين، كما بيّنت النتائج أن عدد الفقرات المنسجمة مع النموذج يزداد بزيادة عدد معالمه، وأظهرت النتائج أيضا عدم وجود فرق ذي دلالة إحصائية بين معاملي ثبات المقياسين الناتجين في المؤشرات التقليدية ونموذج المعلمتين، ووجود فرق دال إحصائي بين معامل ثبات المقياس الناتج من المؤشرات التقليدية ونموذج المعالم الثلاثة وكذلك نموذج راش. كما أظهرت النتائج أن سلوك نموذج المعلمتين كان الأكثر اتفاقا دائما مع المؤشرات التقليدية من النماذج الأخرى؛ لأنه الأكثر شباها بالمؤشرات التقليدية في اختيار الفقرات.

كما قام بيلتون (pelton, 2002) بدراسة هدفت إلى مقارنة الدقة والاستقرار في تقدير معلم الصعوبة ومعلم القدرة باستخدام النظرية التقليدية في القياس والنماذج اللوجستية في نظرية الاستجابة للفقرات وذلك من خلال استخدام أسلوب المحاكاة. وتوصلت الدراسة إلى مجموعة من النتائج منها: أن تقديرات القدرة يمكن مقارنتها عبر النظرية التقليدية ونظرية الاستجابة للمفردة، إذ تتباين تقديرات القدرة حسب كمية المعلومات المتوفرة في مجموعة البيانات المولدة التي تتأثر بدورها بأحادية البعد، ودرجة التخمين ومدى التباين في صعوبة الفقرات مقارنة بقدرات الأفراد، كما توصلت الدراسة إلى أن النموذج اللوجستي الثنائي يقدم تقديرات أكثر دقة لمعلمة الصعوبة من نموذج راش والنموذج اللوجستي الثلاثي والنظرية التقليدية في القياس وذلك في ظل وجود تخمين قليل في الإجابة على فقرات الاختبار وبوجود حجم عينة معتدل (٩٩٩) مفحوص وعدد فقرات مناسب (٣٣) فقرة.

وفي دراسة هوانج (Hwang, 2002) هدفت إلى المقارنة بين النظرية التقليدية ونظرية الاستجابة بنماذجها الثلاثة (1, 2 & 3 Parameter Logistic Models) من حيث تقديرات مؤشرات صعوبة الفقرات، وتمييزها، ومستويات القدرة للطلاب، ومن حيث استقرار تقديرات الفقرات الاختبارية باختلاف مستويات قدرة الطلاب. وقد بلغ حجم عينة الدراسة (٦٠٠) طالب من طلاب الصف الثامن في أوستن (Austin). وقد استخدم اختبار تحصيلي في الرياضيات اشتمل على (١٥) فقرة. وقد توصلت هذه الدراسة إلى النتائج الآتية: تشير معاملات الارتباط المرتفعة - معظمها أكبر من ٠,٩٠ - بين المؤشرات الإحصائية المستمدة من النظريتين إلى تشابه كبير بينهما، كما أن هناك تشابه واضح بين النظريتين في تقدير مؤشرات صعوبة الفقرات، إلا أن هذا التشابه يقل فيما يتعلق بتقدير مؤشرات تمييز الفقرات.

وقام كورفيلي (Courville, 2004) بدراسة هدفت إلى مقارنة النظرية التقليدية ونظرية الاستجابة للفقرات بنماذجها الثلاثة من حيث صعوبة الفقرات وتمييزها ومستويات قدرة الطلاب، ومدى استقرار تقديرات صعوبة الفقرات وتمييزها عبر العينات المختلفة في كلا النظريتين. وقد استخدم لهذا الغرض بيانات لثلاث عينات عشوائية مختلفة كبيرة الحجم، حيث بلغ حجم العينة الأساسية (٨٠٠٠) طالب، والعينات العشوائية الثلاث هي: عينة عشوائية بسيطة تتكون من (١٠٠) عينة عشوائية لكل اختبار من الاختبارات الأربعة، وعينة الجنس وتتكون من (١٠٠) عينة ذكور و(١٠٠) عينة إناث، وعينة القدرات المرتفعة - المتدنية، التي تكونت من (١٠٠) عينة لمجموعة القدرات المرتفعة و(١٠٠) عينة أخرى لمجموعة القدرات المتدنية. وقد استخدم بيانات اختبار القبول للجامعات الأمريكية (ACT) American College Test والمطبّق على طلاب الصفين الحادي عشر - والثاني عشر - ويتكون الاختبار من أربع اختبارات فرعية هي: اختبار اللغة الإنجليزية (٧٥) فقرة، واختبار القراءة (٤٠) فقرة، واختبار الرياضيات (٦٠) فقرة، واختبار العلوم (٤٠) فقرة. وقد توصلت

هذه الدراسة إلى النتائج الآتية: تشير معاملات الارتباط المرتفعة معظمها أكبر من (٠,٨٥)، بين المؤشرات الإحصائية المستمدة من النظريتين إلى تشابه كبير بين النظريتين في تقدير صعوبة الفقرات وفي تقدير مستويات قدرة الطلاب. أما تمييز الفقرات فكانت أقل ارتباطاً مقارنة بصعوبة الفقرات وبمستويات قدرة الطلاب، على الرغم أنها ظلت مرتفعة. كما أشارت إلى استقرار التقديرات، فكلتا النظريتين تشيران إلى استقرار تقديرات صعوبة الفقرات، وتمييزها وذلك عبر العينات الثلاثة، إلا أن تقديرات تمييز الفقرات كانت أكثر استقراراً في النظرية التقليدية مقارنة بتقديرات نماذج الاستجابة للفقرة.

وأجرت الحكماني (٢٠٠٧) دراسة هدفت إلى مقارنة النظرية التقليدية للاختبار ونظرية الاستجابة للفقرة متمثلة في النموذج اللوغاريتمي ثنائي البارامتر من حيث تقدير مستويات قدرة الطلبة. كما هدفت إلى مقارنة النظريتين من حيث مدى استقرار المؤشرات الإحصائية للفقرات الاختبارية والمتمثلة في صعوبة الفقرات وتمييزها وذلك باختلاف الصفوف الدراسية. وقد تكونت عينة الدراسة من (٣٠٨٢) طالب وطالبة من طلاب صفوف السابع والثامن والتاسع بمدارس محافظة مسقط بسطنة عُمان. وتم جمع بيانات هذه الدراسة باستخدام اختبار المصفوفات المتتابعة العادي لرافن. ولمعالجة البيانات وتحليلها استخدمت البرامج الإحصائية (SPSS, BILOG-MG, ITEMAN32, EXCEL)، وقد توصلت هذه الدراسة إلى النتائج الآتية:

- أن هناك تشابهاً بين درجات الطلاب المقدرة من النظريتين في الترتيب، إلا أن قيم درجات الطلاب وفق النظرية التقليدية تختلف بشكل كبير نسبياً عن تقديرات درجاتهم المستمدة من النموذج اللوغاريتمي ثنائي البارامتر.
- أن المؤشرات الإحصائية للفقرات الاختبارية المقدرة بالنموذج اللوغاريتمي ثنائي البارامتر أكثر استقراراً من تلك المقدرة باستخدام النظرية التقليدية.
- أن مؤشرات الصعوبة أكثر استقراراً من مؤشرات التمييز وذلك في كلا الإطارين النظريين.

كما أجرى أنيل (Anil, 2008) دراسة بعنوان "التنبؤ بمعالم الفقرات اعتماداً على نظرية الاختبار التقليدية والنموذج اللوجستي ثنائي المعلمة كممثل لنظرية الاستجابة للفقرة". حيث هدفت هذه الدراسة إلى الكشف عن قوة التنبؤ بخصائص الفقرة بناءً على آراء الخبراء التي تم حسابها بالاعتماد على نظرية الاختبار التقليدية والنموذج اللوجستي ثنائي المعلمة كممثل لنظرية الاستجابة للفقرة في الظروف التي لا يمكن فيها تطبيق مثل هذه الممارسات. وقد أجريت الدراسة في تركيا على عينة تألفت من (٩٩١٤) طالباً من طلبة المرحلة الثانوية تم اختيارها عشوائياً كانوا قد خضعوا لاختبارات الاختيار المطبقة في تركيا من قبل وزارة التربية والتعليم والتي لا يمكن فيها تطبيق مثل هذه الممارسات التجريبية. حيث تم الحصول على تقديرات الخبراء وفقاً لتطبيقات قياسية مُجهزة بطرق مختلفة طبقاً لنظرية الاختبار التقليدية ونظرية الاستجابة للفقرة والتي تم تطبيقها على (١٦) مدرس رياضيات استناداً على النظريتين التقليدية ونظرية الاستجابة للفقرة. وقد حللت البيانات باستخدام برنامج (window 3.1 program). وأظهرت نتائج الدراسة أن صعوبة الفقرة المقدرة في اختبار الرياضيات عبر المجموعات الفرعية المختارة واختبارات التوظيف استناداً لآراء الخبراء يمكن التنبؤ بها استناداً على النظرية التقليدية في القياس. علاوة على ذلك يمكن التنبؤ بمعالم صعوبة الفقرة اعتماداً على النموذج اللوجستي ثنائي المعلمة كممثل لنظرية استجابة الفقرة. كما أظهرت النتائج أنه لا يمكن تقدير معلمة التمييز اعتماداً على نظرية استجابة الفقرة والنظرية التقليدية.

وقام هيرنانديز (Hernandez, 2009) بدراسة هدفت إلى مقارنة صعوبة الفقرة وتمييزها لاختبار القدرات العقلية السريعة (Q-MAT) (Quick-Mental Aptitude Test) باستخدام نماذج نظرية استجابة الفقرة (IRT) والنظرية التقليدية (CTT). وتكونت عينة الدراسة من (٢٢٩) طالباً من طلبة كلية (De La Salle-College of Saint Benilde)، حيث تم تحليل البيانات الخاصة باستجابات الأفراد باستخدام المؤشرات الإحصائية وفقاً لنماذج نظرية استجابة الفقرة والنظرية التقليدية في القياس، وأشارت النتائج إلى عدم وجود فروق ذات دلالة إحصائية بين النموذج اللوجستي ثنائي المعلمة والنظرية التقليدية في القياس عند مستوى الدلالة الإحصائية (٠,٠١).

وفي دراسة أجراها راج (Raj, 2010) هدفت إلى تقصي تقديرات اللاتغير (Invariance) لمعلمة الفرد على أساس النظرية التقليدية ونظرية الاستجابة للفقرة. ومن أجل تحقيق هذا الهدف تم التحقق من مطابقة الفقرات للنموذج الثنائي المعلمة، حيث وجدت (١١) فقرة من أصل (٤٠) فقرة من الورقة الأولى لاختبار بوتسوانا في الرياضيات الخاص بالمرحلة الثانوية مطابقة للنموذج. وتكونت عينة الدراسة من (٥٠٠) طالباً تم اختيارهم عشوائياً من بين (٣٥٦٢) طالباً، حيث تم تطبيق اختبار بوتسوانا في الرياضيات للمرحلة الثانوية عليهم. وقد تم جمع البيانات وتحليل النتائج لتحديد تقديرات معلمة الفرد باستخدام تحليل التباين الأحادي (ANOVA) للقياسات المتكررة عند مستوى الدلالة (٠,٠٥ = α). وقد أشارت

النتائج إلى وجود قصور في النظرية التقليدية مقارنة مع نظرية استجابة الفقرة التي اتسمت باللاتغير عبر عينات الأفراد المختلفة، ودعت هذه الدراسة إلى استخدام (IRT) في بناء الاختبارات والمقاييس.

١,٢,٣ التعقيب على الدراسات السابقة

بمطالعة الدراسات السابقة التي تم عرضها يُلاحظ أنها تناولت أوجهًا مختلفة حيث استخلص الباحثان من خلال نتائجها ما يلي:

- تناولت بعض الدراسات مقارنة الخصائص السيكومترية بين التقديرات المُستمدة من النظرية التقليدية وتلك المُستمدة من نماذج الاستجابة للفقرة، والمتعلقة بصعوبة الفقرات، وتمييزها، وبمستويات قدرة الطلبة. مثل دراسات (Fan, 1998؛ جمحاوي، ٢٠٠٠؛ Hwang, 2002؛ الحكامي، ٢٠٠٧).
- تباينت العينات في الدراسات السابقة من حيث المرحلة الدراسية والعُمريّة، فبعضها كانت من المرحلة المتوسطة مثل دراسات (جمحاوي، ٢٠٠٠؛ Hwang, 2002؛ الحكامي، ٢٠٠٧). وبعضها الآخر كانت من المرحلة الثانوية مثل دراسات فان (Fan, 1998). وهكذا يتضح أن الدراسات ليست قاصرة على مرحلة دراسية دون أخرى.
- اختلاف الدراسات السابقة في البرامج الحاسوبية المستخدمة في تحليل فقرات الاختبار وتدريبها.

أما أوجه الاستفادة من الدراسات السابقة فإن الباحثين يرون أن هذه الدراسات بمثابة المصدر الأساسي لكثير من المعلومات المهمة؛ التي وجهتهما في دراستهما الحالية من حيث اختيارها وتحديد مشكلتها ومنهجيتها والإجراءات الملائمة لتحقيق أهدافها. هذا بالإضافة إلى أن تلك الدراسات وجهتهما نحو العديد من البحوث والدراسات والمراجع المناسبة ومكنتهما من تكوين تصور شامل عن الأطر النظرية التي ينبغي أن تشملها الدراسة الحالية.

ولذلك فإن الدراسة الحالية، امتداد لهذا المجال، والتي من المؤمل أن تضيف إلى نتائج الدراسات السابقة. وقد تميزت الدراسة الحالية عن غيرها من الدراسات السابقة في غرضها ومجتمعها، حيث جاءت هذه الدراسة بغرض التحقق من مدى التوافق بين النظريتين (التقليدية والحديثة)، فيما يخص المؤشرات الإحصائية لفقرات والأفراد؛ للمساهمة في التخفيف من الجدل القائم في ضوء نتائج الدراسات السابقة؛ وذلك من خلال الإجابة على التساؤل: "ما مدى توافق النموذج الثنائي المعلمة والنظرية التقليدية في انتقائهما لفقرات الاختبار من حيث عددها وصعوبتها، وتمييزها وثباتها وصدقها وتدريب القدرة". كما أن معظم الدراسات التي قارنت بين النظريتين (النظرية التقليدية والنظرية الحديثة) كانت تستخدم نموذج راش كأحد نماذج النظرية الحديثة في القياس، في حين أنه تم استخدام النموذج ثنائي المعلمة في هذه الدراسة لتشابه معالمه مع مؤشرات النظرية التقليدية.

٢. مشكلة الدراسة وأسئلتها

تعدّ الاختبارات التحصيلية ذات أهمية كبيرة في العملية التعليمية التعلمية، إذ يتم في ضوءها قياس وتقويم تحصيل المتعلمين، ويعتمد نجاح تلك الاختبارات أو فشلها على مدى امتلاك المعلمين المهارات اللازمة لبنائها ابتداءً من إعدادها وانتهاءً بتصحيحها وتحليل نتائجها. وفي العادة يتم تحليل نتائج الاختبارات وفق النظرية التقليدية، رغم وجود العديد من الدراسات التي بينت أن هناك قصوراً في تلك النظرية، فقد أظهرت دراسة هامبلتون وجونز (Hambelton & Jonse, 1993) وجود قصور في النظرية التقليدية، يتمثل في أن جميع الخصائص السيكومترية للاختبارات تعتمد على خصائص الأفراد المُطبق عليهم الاختبار، وأن نتائج الاختبار وما يُبنى عليها من تفسيرات يعتمد على خصائص عينة الفقرات التي يشتمل عليها الاختبار، حيث أن النظرية الحديثة قد عالجت أوجه القصور في النظرية التقليدية.

في حين أظهرت نتائج دراسة عودة (١٩٩٢) أفضلية للمؤشرات التقليدية على حساب المؤشرات الحديثة، بينما أظهرت نتائج دراسة كل من (Fan, 1998؛ جمحاوي، ٢٠٠٠). اتفاق النظريتين في النتائج.

كما أشارت نتائج العديد من الدراسات كدراسة كل من: بيلتون (pelton, 2002; Hernandez, 2009) مدى استقرار النتائج التي تم الحصول عليها وفقاً للنموذج ثنائي المعلمة حيث أنه يقدم تقديرات أكثر دقة لمعالم الفقرة من نموذج راش والنموذج اللوجستي الثلاثي والنظرية التقليدية في القياس وذلك في ظل وجود تخمين قليل في الإجابة على فقرات الاختبار، علاوة على تشابه المؤشرات الإحصائية للنموذج الثنائي المعلمة مع المؤشرات الإحصائية التقليدية، حيث يتناول كل منها معلمة الصعوبة، والتمييز، مع تثبيت معلمة التخمين، في كلا الأسلوبين (النموذج ثنائي المعلمة والنظرية التقليدية).

وتأسيساً على ما سبق؛ وفي ظل الجدل القائم بين النظريتين (التقليدية والحديثة) جاءت هذه الدراسة لتكشف عن مدى التوافق بين النظرية التقليدية والنموذج ثنائي المعلمة كممثل للنظرية الحديثة، من خلال مقارنة تحليل نتائج الاختبار المُعد وفقاً لنظرية استجابة الفقرة باستخدام النموذج الثنائي المعلمة والنظرية التقليدية.

٢,١ أسئلة الدراسة

وبشكل أكثر تحديداً ستحاول الدراسة الإجابة عن الأسئلة الآتية:

١. ما مدى التوافق في نتائج تصنيف الفقرات الخاصة باختبار مبحث العلوم للصف السادس الأساسي إلى (مطابقة أو غير مطابقة) وفقاً لـ: (نظرية استجابة الفقرة باستخدام النموذج الثنائي المعلمة، النظرية التقليدية)؟
٢. هل يوجد اختلاف بين قيمة معامل ثبات الاختبار وفقاً للنظريتين (النظرية التقليدية، نظرية استجابة الفقرة للنموذج الثنائي المعلمة) عند مستوى الدلالة الإحصائية ($\alpha = 0.05$)؟
٣. هل يوجد اختلاف بين قيمة معامل الصدق المحكي للاختبار وفق النظريتين (النظرية التقليدية ونظرية استجابة الفقرة للنموذج الثنائي المعلمة) عند مستوى الدلالة الإحصائية ($\alpha = 0.05$)؟
٤. ما مدى التوافق في تدرج قدرة المفحوصين على الاختبار في مبحث العلوم للصف السادس الأساسي وفقاً للنظرية (التقليدية، استجابة الفقرة ذات النموذج الثنائي المعلمة)؟

٣. أهداف الدراسة وأهميتها

٣,١ أهداف الدراسة

هدفت هذه الدراسة من خلال نتائجها إلى الكشف عن مدى التوافق بين النموذج ثنائي المعلمة والنظرية التقليدية في بناء اختبار تحصيلي في مبحث العلوم لطلبة الصف السادس الأساسي.

٣,٢ أهمية الدراسة

تكمُن أهمية الدراسة الحالية من الناحية النظرية في التوصل إلى بناء اختبار يتمتع بخصائص سيكو مترية جيدة استناداً إلى النموذج ثنائي المعلمة والنظرية التقليدية في تحليل وتفسير نتائج الاختبارات والمقاييس، كما ستساهم الدراسة في دعم القاعدة النظرية للبحوث المتعلقة بالمقارنة بين النظريتين التقليدية ونظرية استجابة الفقرة في انتقائهما للفقرات. وربما يستخدم الاختبار الوارد في هذه الدراسة كاختبار تشخيصي للطلبة يكشف عن جوانب القوة والضعف في تحصيلهم لمبحث العلوم؛ ليتم تعزيز جوانب القوة، والعمل على تلافي نقاط الضعف لديهم والعمل على تحسين تحصيلهم.

أما من الناحية التطبيقية فتبرز أهمية الدراسة في نتائجها المتوقعة التي يُؤمل أن يستفيد منها متخذو القرار وراسمو السياسات التربوية في الأردن على تطوير مهارات تدريس مبحث العلوم ومقارنة المستوى الوطني عالمياً وتحسين العملية التعليمية التعلمية من حيث أهدافها ووسائلها وتقنياتها بعقد الدورات المتخصصة في مجال بناء الاختبارات التحصيلية وتحليل وتفسير نتائجها وتطوير النظام التربوي والارتقاء بنوعية مخرجاته.

٤. مصطلحات الدراسة

التوافق بين النظريتين (التقليدية والحديثة)

يعني درجة الاتفاق ما بين النظرية التقليدية والنظرية الحديثة في القياس في نتائج التحليل التي أفضت إليها كل منهما عند تحليل نتائج الاختبار، وفي هذه الدراسة تم حصر النسبة المئوية لعدد فقرات الاختبار التي اتفقت النظريتان (النظرية التقليدية والنظرية الحديثة للنموذج ثنائي المعلمة) على بقائها وفقاً لمعاري الصعوبة والتمييز.

النظرية التقليدية في القياس: تمثل هذه النظرية نموذجاً مبسطاً لوصف الكيفية التي تؤثر بها أخطاء القياس (الأخطاء العشوائية) على العلامات الملاحظة (المشاهدة أو الظاهرة أو الكلية)، ولهذه النظرية مجموعة من المسميات، حيث يطلق عليها عدة أسماء منها: النظرية التقليدية في القياس، أو النظرية التقليدية في القياس (Hambelton & Jonse, 1993).

- النموذج ثنائي المعلمة أو البارامتر (2PL: Lord Model): يعتبر هذا "النموذج" نموذج ثنائي البعد أو البارامتر، ويفترض هذا النموذج أن الفقرات تختلف في صعوبتها وتمييزها بين المستويات مختلفة القدرة، ويرمز لمعامل التمييز (a) ومعامل الصعوبة (b) (التقي، ٢٠٠٧).
- الاختبار: أداة قياس يتم إعدادها وفق طريقة منظمة من عدة خطوات، تتضمن مجموعة من الإجراءات التي تخضع لشروط وقواعد محددة بغرض تحديد درجة امتلاك الفرد لسمة، أو قدرة معينة من خلال إجابته عن عينة من المثيرات (الفقرات) التي تمثل السمة أو القدرة المرغوب قياسها (عودة، ٢٠١٠).
- الاختبار التحصيلي في العلوم: اختبار مصمم للكشف عن مدى تحقق الأهداف التعليمية في مبحث العلوم من خلال إجابات الفرد عن عينة من المثيرات (الفقرات) في مبحث العلوم العامة للصف السادس الأساسي في الوحدة السابعة بعنوان علوم الأرض والبيئة.
- التحصيل: الدرجة أو العلامة التي يحصل عليها المتعلم نتيجة خضوعه لاختبار العلوم العامة المُعد لطلبة الصف السادس الأساسي.
- حدود الدراسة: اقتصرت عينة الدراسة على طلبة الصف السادس الأساسي في المدارس الحكومية في محافظة جرش في الفصل الدراسي الثاني للعام الدراسي ٢٠١٠ / ٢٠١١ م.
- محددات الدراسة: تتحدد نتائج هذه الدراسة بالاختبار المستخدم ويقدر ما يتمتع به من خصائص سيكومترية مقبولة؛ حيث اقتصر محتوى الاختبار التحصيلي على وحدة واحدة (علوم الأرض والبيئة) من كتاب العلوم المقرر من وزارة التربية والتعليم للصف السادس الأساسي للعام الدراسي ٢٠١٠ / ٢٠١١ م. كما اقتصرت هذه الدراسة على استخدام النموذج اللوجستي ثنائي المعلمة كممثل لنظرية استجابة الفقرة.

٥. الطريقة والإجراءات

٥,١ منهجية الدراسة

اعتمدت الدراسة على المنهج الوصفي وفقاً للنظرية التقليدية ولنظرية استجابة الفقرة بأسلوب كمي، للكشف عن مدى التوافق بين النظريتين (التقليدية واستجابة الفقرة) في معلّمتي (الصعوبة والتمييز) للاختبار ومدى التوافق بين النظريتين في تدريج المفحوصين والكشف عن وجود فرق في معاملات الثبات ومعامل الصدق وفقاً للنظريتين.

مجتمع الدراسة وعينتها: تكون مجتمع الدراسة من جميع طلبة الصف السادس الأساسي في المدارس الحكومية التابعة لمديرية التربية والتعليم في محافظة جرش وعددهم (٣٢٥١) طالباً وطالبة موزعين على (١٦٦) مدرسة منها (٧٧) ذكور تحتوي (١٦٧٣) طالباً موزعين على (٧١) شعبة صفية و(٨٩) مدرسة للإناث تحتوي (١٥٧٨) طالبة موزعات على (٦٧) شعبة صفية. تكونت عينة الدراسة من طلبة الصف السادس الأساسي في مدارس محافظة جرش الحكومية، وعددهم (٥٦٧) طالباً وطالبة، تم اختيارهم بالطريقة الطباقية العنقودية العشوائية، حيث كان متغير الجنس هو المتغير الطبقي وكانت (المدرسة) هي المتغير العنقودي، وذلك عن طريق حصر مدارس الذكور في قائمة ومدارس الإناث في قائمة أخرى ومن قائمة مدارس الذكور؛ تم عشوائياً اختيار (٤) مدارس كان عدد طلابها (٢٨٥) طالباً، ومن قائمة مدارس الإناث؛ فقد تم عشوائياً اختيار (٤) مدارس منها تضمنت (٢٨٢) طالبة. وقد روعي في اختيار عينة الدراسة التوزيع الجغرافي وظروف الباحثين العلمية. والجدول (١) يبين توزيع أفراد عينة الدراسة حسب الجنس والمدرسة.

جدول (١): توزيع أفراد عينة الدراسة حسب المدرسة والجنس

أسماء مدارس الذكور	العدد	أسماء مدارس الإناث	العدد	المجموع
سالك الأساسية للبنين	١٠٢	آمنة بنت وهب الأساسية للبنات	١١٢	٢١٤
جرش الأساسية للبنين	١١٠	أسماء بنت يزيد الأساسية للبنات	٦٥	١٧٥
أبو ذر الغفاري الأساسية للبنين	٤٣	جبل الشيخ مصلح الأساسية للبنات	٨٠	١٢٣
باب عمان الأساسية للبنين	٣٠	بليلا الثانوية الشاملة للبنات	٢٥	٥٥
المجموع	٢٨٥	المجموع	٢٨٢	٥٦٧

أداة الدراسة: ولأغراض إنجاز هذه الدراسة قام الباحثان بإعداد اختبار تحصيلي في مبحث العلوم، لطلبة الصف السادس الأساسي في وحدة "علوم الأرض والبيئة"، وقد اقتضى بناء أسئلة الاختبار ووضعه في صيغته النهائية، الاسترشاد

بالأسس المتبعة في تصميم اختبارات التحصيل الصفية (عدس، ٢٠٠٢؛ Gronlund & Linn, 1990)، وفيما يلي الإجراءات التي اتبعها الباحثان:

أولاً- تحديد الغرض من الاختبار التحصيلي: وهو قياس التحصيل المعرفي في وحدة علوم الأرض والبيئة في مبحث العلوم العامة لطلبة الصف السادس الأساسي في المملكة الأردنية الهاشمية.

ثانياً- نتائج التعلم: تم تحديد نواتج التعلم التي يقيسها الاختبار أو الأهداف المتوقع تحقيقها، حيث المقصود هنا كل النواتج التي يقيسها الاختبار، بحيث أن هذا الاختبار الذي تم بناؤه هو عينة من هذه النواتج اعتماداً على خبرة الباحثان التدريسية لهذا المبحث، التي في الجدول (٢).

جدول (٢): تصنيف الأهداف حسب مستواها

الرقم	الهدف	مستوى الهدف
١	يذكر الأساس المعتمد في بناء سلم الزمن الجيولوجي	معرفة
٢	يبني سلم زمني لحياة الإنسان	تركيب
٣	يربط بين تقسيمات الزمن (ساعة، يوم،، إلخ)، وتقسيمات سلم الزمن الجيولوجي (دهر حقبة، ...، إلخ)	تطبيق
٤	يربط بين سلم الزمن الجيولوجي، وسلم الإنسان الزمني	تطبيق
٥	يبني نموذجاً على مقياس زمني يوضح تدريج الحياة على سطح الأرض	تركيب
٦	يصف رُقي الحياة على سطح الأرض	فهم واستيعاب
٧	يعدد أهم الكائنات الحية التي تميز كل حقبة	معرفة
٨	يقارن بين دهر الحياة المستترة ودهر الحياة الظاهرة	فهم واستيعاب
٩	يستنتج النسب الخاصة بتوزيع المياه في الكرة الأرضية	تحليل
١٠	يعدد مصادر المياه العذبة	معرفة
١١	يميز بين مصادر المياه بنوعيتها؛ العذبة والمالحة	فهم واستيعاب
١٢	يذكر مصادر المياه الجوفية	معرفة
١٣	يصف الخزان الجوفي من حيث تركيبه، وكيفية وجود الماء فيه	فهم واستيعاب
١٤	يعرف مفهوم كل من: البئر والينبوع	معرفة
١٥	يفرق بين الينابيع والآبار	فهم واستيعاب
١٦	يتعرف بعض المشكلات المائية في الأردن	معرفة
١٧	يذكر مصادر تلوث المياه في الأردن	معرفة
١٨	يصف بعض الأمراض الناشئة عن تلوث المياه	فهم واستيعاب
١٩	يعدد طرائق ترشيد استهلاك المياه في الأردن	معرفة
٢٠	يوضح مفهوم مشكلة شح المياه مبيناً أسبابها	معرفة
٢١	يقترح حلولاً لمشكلة شح المياه أو تلوثها	تقويم
٢٢	يكتب تقريراً يوضح أهمية إعادة استخدام المياه في المنزل والمدرسة	تقويم

ثم تم استشارة عدد من ذوي الخبرة التربوية بوصفهم محكمين للأهداف؛ للتحقق من صحة صياغة الأهداف، ومدى ملاءمتها للمقرر الدراسي، وفي أي مجال تقع من مجالات الأهداف وفي أي مستوى تقع من المستويات من المجال المعرفي: (المعرفة والتذكر، الفهم، التطبيق، التحليل، التركيب، التقويم).

ثالثاً- تحليل المحتوى الدراسي: تم تحليل محتوى مادة الاختبار بإعداد قائمة بالموضوعات المتضمنة للمادة الدراسية بأكبر قدر من التفصيل؛ بهدف تمثيل الاختبار التحصيلي للمادة التعليمية تمثيلاً صادقاً، لذلك وفي صورة مبسطة، تم تحديد الوزن النسبي لكل موضوع على أساس عدد صفحاته، كما حددت الأوزان النسبية للأهداف التعليمية في المادة الدراسية، ويعتمد ذلك بالتحديد على خبرة الباحثين في تدريسها للمادة، وتحليلهما لمحتواها، حيث قام الباحثان بوضع قائمة بأهم الموضوعات المتناولة بالوحدة الدراسية، تم تصنيفها تبعاً للأهداف المراد قياسها. ولم يكتفِ الباحثان بخبرتهما الشخصية، فلجأ إلى المتخصصين والخبراء من أساتذة الجامعات ومن المعلمين والمشرفين العاملين بوزارة التربية والتعليم لاستشارتهم في تحديد الأوزان النسبية للموضوعات المشمولة بالاختبار، وطلب منهم مراجعة الموضوعات التي تضمنها التخطيط للاختبار، وتصنيفاتها تبعاً للأهداف المراد قياسها، واقتراح إضافة فقرات جديدة أو حذف فقرات أو تعديلها.

رابعاً- بناء جدول المواصفات أو لائحة المواصفات: انطلاقاً من فكرة أنه لا يمكن أن يتم وضع كل الأسئلة المُراد الاستفسار عنها في أي اختبار، لذا تتوخى الاختبارات التحصيلية قياس عينة ممثلة لكلاً من: (النواتج التعليمية، ومحتوى المادة التعليمية المتضمنة)، ذلك أنه هناك الكثير من الحقائق والمصطلحات المتوقع من المتعلمين معرفتها، ولكن أمام فكرة ضيق الوقت المسموح به للاختبار؛ يتم إدراج عدد مناسب من الفقرات في الاختبار لقياس جزء من هذه الحقائق، أو بعض من التطبيقات.

لقد تم بناء جدول مواصفات للاختبار من خلال محورين: أحدهما للأهداف بوصفه (محوراً سلوكياً)، والآخر للمحتوى الدراسي (كمحور معرفي)، وفي هذا المحور أخذ بعين الاعتبار عدد الصفحات لكل فصل من فصول هذه الوحدة الدراسية المأخوذة من كتاب العلوم العامة للصف السادس الأساسي المقرر، كما قسّمها الباحثان.

جدول (٣): لائحة مواصفات الاختبار

المحتوى	معرفي	فهم واستيعاب	تطبيق	تحليل	تركيب	تقويم	المجموع
النسبة المئوية	٤٠,٩١	٢٧,٢٧	٩,٠٩	٤,٥٥	٩,٠٩	٩,٠٩	٪١٠٠
المسمى	٦,٥٧	٤,٣٨	١,٤٦	٠,٧٣	١,٤٦	١,٤٦	١٦,١
تأريخ الأرض	٧	٤	١	١	١	١	١٥,٠
جيولوجية المياه	٥,٨٤	٣,٩٠	١,٣٠	٠,٦٥	١,٣٠	١,٣٠	١٤,٣
المحافظة على المياه	٨,٠٤	٥,٣٦	١,٧٩	٠,٨٩	١,٧٩	١,٧٩	١٩,٦
المجموع	٢٠,٥	١٣,٦	٤,٥	٢,٣	٤,٥	٤,٥	٥٠,٠

يلاحظ من الجدول (٣)، أنه تم تحديد وزن كل موضوع أو أهميته بالنسبة للموضوعات الأخرى اعتماداً على عدد صفحات ذلك الموضوع بالنسبة لعدد صفحات الكتاب المدرسي، حيث تم احتساب الوزن النسبي للفصل بالنسبة للوحدة الدراسية التي تم تناولها وفقاً للعلاقة الرياضية الآتية:

$$\text{الوزن النسبي للفصل} = (\text{عدد صفحات الفصل} \div \text{عدد صفحات الوحدة}) \times ١٠٠ \%$$

$$\text{الوزن النسبي للفصل الأول} = (٢٨ \div ٩) \times ١٠٠ \% = ٣٢,١٤ \%$$

$$\text{الوزن النسبي للفصل الثاني} = (٢٨ \div ٨) \times ١٠٠ \% = ٣٥,٠٠ \%$$

$$\text{الوزن النسبي للفصل الثالث} = (٢٨ \div ١١) \times ١٠٠ \% = ٢٥,٤٥ \%$$

كما تم تحديد أهمية أهداف كل موضوع (الوزن النسبي لمستوى الهدف): وذلك من خلال قسمة عدد أهداف المستوى على عدد الأهداف الكلي مضروباً في ١٠٠٪، فعلى سبيل المثال يمكن تحديد الوزن النسبي للأهداف في مستوى المجال المعرفي من خلال المعادلة الآتية:

$$\text{الوزن النسبي لمستوى الأهداف في المجال المعرفي} = (\text{عدد أهداف المعرفة} \div \text{عدد الأهداف الكلي}) \times ١٠٠ \%$$

$$\text{الوزن النسبي لمستوى الأهداف في المجال المعرفي} = (٢٢ \div ٩) \times ١٠٠ \% = ٢٤٤,٤٤ \%$$

وهكذا يمكن تحديد الأهمية أو الوزن لكل موضوع/مستوى.

سيكون عدد فقرات الاختبار هنا (٥٠) فقرة، وبناء عليه سيتم تحديد عدد الفقرات لكل موضوع/مستوى، فعلى سبيل المثال لتحديد عدد الفقرات للموضوع الأول (تأريخ الأرض) في مستوى المعرفة = $٥٠ \times (٣٢,١٤ \times ٤٠,٩١ \%) = ٥٠ \times ١٣,٢٦ = ٦٦٣$ أي تقريباً (٧) فقرات، وهكذا يمكن تحديد عدد الفقرات لكل موضوع/مستوى كما هي مدونة في الجدول السابق.

خامساً- كتابة فقرات الاختبار: تعد من أهم المراحل التي يمر فيها بناء الاختبار، ذلك أن أول خطوة فيها تتمثل في انتقاء شكل الفقرات المناسبة لقياس النواتج، إذ تم هنا اعتماد الفقرات ذات الإجابة المنتقاة (من أربعة بدائل) على أن يتم

كتابة الفقرات ضمن القواعد السائدة لها، حيث قام الباحثان بصياغة (٥٠) فقرة من نوع الاختيار من متعدد، بأربعة بدائل أحدها يمثل الإجابة الصحيحة؛ لقياس كل هدف من الأهداف السلوكية، وقد تطلب قياس بعضها أكثر من فقرة واحدة.

سادساً: صدق المحتوى: للتحقق من الصدق الظاهري للاختبار؛ قام الباحثان بعرض الصورة الأولية للاختبار، وجدول المواصفات، وتحليل المحتوى، والنتائج التعليمية، على مجموعة من المختصين على النحو الآتي: مشرفو العلوم العامة في تربية محافظة جرش، ومشرف التربية العملية في تدريس العلوم في جامعة اليرموك، وستة مدرسين يدرسون المادة ضمن خبرات تدريسية متفاوتة، بعضهم تخصص قياس وتقويم، والبعض الآخر من حملة ماجستير العلوم وأساليب تدريس العلوم؛ وذلك لأخذ وجهات نظرهم في مدى صدق الفقرة الاختبارية في قياس الهدف السلوكي المحدد، واقتراح ما يروونه من تعديل، من خلال الإجابة عن استبيان تم إعداده لتحقيق هذا الغرض كما في الملحق (١) و(٢). حيث قام الباحثان بإيجاد نسبة الاتفاق الكلية بين آراء المحكمين على ملحق (١) استبيان مناسبة الصياغة اللغوية، وذلك على النحو الآتي:

$$\frac{346}{346 + 154} \times 100\% = 69.2\%$$

نسبة الاتفاق بين المحكمين

أي أن (٦٩,٢٪) من فقرات الاختبار اتفق المحكمين في الحكم عليها كونها مناسبة الصياغة، كما قام الباحثان بإيجاد نسبة الاتفاق الكلية بين آراء المحكمين على ملحق (٢) استبيان ملائمة الفقرة للهدف، وذلك على النحو الآتي:

$$\frac{380}{380 + 120} \times 100\% = 76\%$$

نسبة الاتفاق بين المحكمين

أي أن (٧٦٪) من فقرات الاختبار اتفق المحكمين في الحكم عليها كونها ملائمة للهدف المراد قياسه.

وفي ضوء تقديرات المحكمين، ومناقشة الباحثين مع المحكمين، تم تعديل بعض أسئلة الاختبار خصوصاً الأسئلة التطبيقية، إضافة إلى بعض البدائل الغامضة، وذلك للخروج بصورة نهائية للاختبار التحصيلي.

كما تم التحقق من الصدق المحكي للاختبار من خلال إيجاد معامل الصدق التجريبي بين علامات الطلبة في المدرسة وعلاماتهم على الاختبار المعد في هذه الدراسة باستخدام معامل ارتباط بيرسون Pearson عن طريق تجريب صورة الاختبار الأولية، على عينة استطلاعية مكونة من (٢١٤) طالباً وطالبة من غير مدارس عينة الدراسة، حيث بلغت قيمته (٠,٧٣)، وهذا يدل على أن الاختبار يتمتع بصدق عالٍ.

هذا وقد تم حساب النسب المئوية للطلبة المفحوصين من أفراد العينة الاستطلاعية ضمن كل بديل من بدائل الاختبار كمؤشر على جودة جاذبية المموهات، وذلك كما في الجدول (٤).

جدول (٤): النسب المئوية للطلبة المستجيبين ضمن بدائل الاختبار (فاعلية المموهات)

رقم الفقرة	أ	ب	ج	د	البديل
١	١٨,٥	*٣٥,٦	٢٢,٢	٢٣,٧	
٢	*٢٧,٦	٢٢,٣	٢٥,١	٢٥,٠	
٣	*٢٩,١	٢٢,٥	٢٥,٠	٢٣,٤	
٤	٢٣,٢	*٣٦,٦	٢١,٣	١٨,٩	
٥	*٤٢,١	١٨,٤	١٨,٩	٢٠,٦	
٦	٢٣,٦	٢٣,٨	٢٠,٥	*٣٢,١	
٧	٢١,٥	١٧,١	١٨,٨	*٤٢,٦	
٨	١٨,٠	٢٠,٣	٢١,٣	*٤٠,٤	
٩	*٣٥,٤	١٨,٧	٢١,٣	٢٤,٦	
١٠	*٤٠,٣	٢٤,٨	١٧,٤	١٧,٥	
١١	٢٢,٢	*٣٢,٩	٢١,١	٢٣,٨	
١٢	٢٣,٦	*٢٩,٠	٢٣,١	٢٤,٣	
١٣	*٤٣,٢	١٨,٨	١٩,٤	١٨,٦	

٢٥,٠	٢٢,٠	*٢٨,٦	٢٤,٤	١٤
١٨,٣	١٩,٩	*٤٣,٥	١٨,٣	١٥
٢١,٨	٢١,٧	٢٢,٣	*٣٤,٢	١٦
١٧,٩	١٩,٣	٢٥,١	*٣٧,٧	١٧
٢٢,٩	٢٣,٩	*٣٥,٣	١٧,٩	١٨
١٨,١	*٣٧,٧	١٩,٧	٢٤,٥	١٩
٢١,٦	*٣٦,٩	٢١,٣	٢٠,٢	٢٠
٢٠,١	٢٢,٣	*٣٨,٨	١٨,٨	٢١
١٧,٩	*٤٢,٧	١٩,٥	١٩,٩	٢٢
٢٢,٦	١٨,٨	١٧,٩	*٤٠,٧	٢٣
*٣١,٥	٢٤,٣	٢٢,٢	٢٢,٠	٢٤
٢٠,١	١٨,٧	١٩,٣	*٤١,٩	٢٥
٢٢,٤	٢٣,٧	٢٠,٣	*٣٣,٦	٢٦
١٨,٢	١٧,٩	٢١,٢	*٤٢,٧	٢٧
*٢٩,٥	٢٠,٩	٢٤,٦	٢٥,٠	٢٨
٢٠,٨	١٩,٨	٢٥,٠	*٣٤,٤	٢٩
٢٠,٥	٢٤,٠	*٣٠,٩	٢٤,٦	٣٠
٢٤,٨	١٧,٤	*٣٥,٩	٢١,٩	٣١
٢١,٠	١٨,٢	١٧,١	*٤٣,٧	٣٢
*٤٢,٠	١٩,٨	١٨,٧	١٩,٥	٣٣
١٩,٤	١٨,٦	١٧,٦	*٤٤,٤	٣٤
*٣٧,١	١٨,٩	١٩,٠	٢٥,٠	٣٥
*٣١,٦	٢٤,٩	١٨,٦	٢٤,٩	٣٦
٢٠,١	٢٥,٠	٢١,٥	*٣٣,٤	٣٧
٢٣,٥	١٩,٧	١٨,٧	*٣٨,١	٣٨
*٣٦,٣	٢٣,٣	١٩,٩	٢٠,٥	٣٩
٢٥,١	٢٢,٦	١٧,٨	*٣٤,٥	٤٠
٢٤,٤	٢١,٦	*٣٠,٦	٢٣,٤	٤١
٢١,٥	*٣٩,٩	١٨,٧	١٩,٩	٤٢
٢٢,٢	*٣٥,٧	٢٠,٦	٢١,٥	٤٣
١٧,٦	٢٢,٧	٢١,٣	*٣٨,٤	٤٤
١٧,٦	١٨,٨	٢١,٩	*٤١,٧	٤٥
٢٤,٣	١٩,٦	٢٢,٠	*٣٤,١	٤٦
٢١,٥	٢٢,٣	*٣١,٠	٢٥,٢	٤٧
١٩,٣	*٤١,٠	١٨,٦	٢١,١	٤٨
٢٤,١	*٣١,٦	٢٤,٤	١٩,٩	٤٩
٢٥,٢	٢٠,٢	*٣٥,٤	١٩,٢	٥٠
البديل الصحيح للفقرة				*

يلاحظ من الجدول (٤)، وجود تناسب في توزيع النسب المئوية للطلبة المستجيبين على بدائل فقرات الاختبار، حيث أنها كانت قريبة من النسبة المئوية الخاصة بكل بديل نظرياً ألا وهي (٢٥٪) كون الاختبار رباعي البدائل؛ مما يشير إلى جودة انتقاء بدائل كل فقرة من فقرات الاختبار، مع ملاحظة أن أعلى نسبة مئوية للطلبة المستجيبين كانت ضمن البديل الصحيح لكل فقرة من فقرات الاختبار.

سابعاً: التجريب الأولي للاختبار: تم تجريب صورة الاختبار الأولية، على الطلبة أفراد العينة الاستطلاعية السالفة الذكر، إذ أُبلغ معلمي مادة العلوم في مدارس العينة الاستطلاعية عن موعد تطبيق الاختبار التحصيلي، وعن المادة التعليمية

التي سيطبق عليها الاختبار، وكانت تعليمات الاختبار أنّ لكل فقرة علامة واحدة، أي أن علامة الطالب كانت هي عدد الإجابات الصحيحة عن فقرات الاختبار، وأعطى الطلبة الوقت الكافي للإجابة عن فقرات الاختبار بصورته الأولية، بحيث تم تسجيل الوقت الذي استغرقه كل طالب في الإجابة عن فقرات الاختبار، وقد تم إيجاد المتوسط الحسابي للوقت الذي استغرقه طلبة العينة الاستطلاعية في الإجابة عن فقرات الاختبار حيث بلغ مقداره (ساعة ونصف)، وبناءً على ذلك فقد قُدِّر وقت الاختبار بساعة ونصف، بناءً على الوقت الذي استغرقه طلبة العينة الاستطلاعية في الإجابة عن الاختبار.

ثبات الاختبار: تم تقدير معامل ثبات الاختبار المُعدّ، باستخدام النظرية التقليدية في القياس، ونظرية استجابة الفقرة، حيث تم تناوله في جزء النتائج لكونه غرض من أغراض الدراسة.

وبهذا يمكن القول أنّ الاختبار المعدّ يمتاز بخصائص سيكو مترية جيدة، وأنّ جودته تزيد بعد أن تم مراجعة بعض الفقرات، وبالتالي يمكن الاعتماد على هذا الاختبار في الحصول على بيانات دقيقة لتنفيذ إجراءات الدراسة.

ثامناً، إجراءات جمع البيانات: تم تطبيق صورة الاختبار النهائية على عينة الدراسة النهائية كل في مدرسته، وبشكل جماعي داخل غرفة الصف، وفقاً للبرنامج الزمني المتفق عليه مع معلمي مبحث العلوم، ثم تم جمع استجابات الطلبة على الاختبار وتصحيحها، ثم أدخلت البيانات إلى ذاكرة الحاسوب باستخدام البرنامج الإحصائي (SPSS)، وبرنامج (BILOG 3)، وبرنامج (NOHARM)، وبرنامج (LDID) لإجراء التحليلات الإحصائية اللازمة للإجابة عن أسئلة الدراسة.

٥,٢ إجراءات الدراسة

١. تم بناء الاختبار التحصيلي المكوّن من (٥٠) فقرة في صورته النهائية.
٢. تم تحديد المدارس المعنية بالدراسة.
٣. تم زيارة المدارس المشمولة بالدراسة حيث تم توضيح تعليمات الاختبار وأهميته لجميع الطلبة بالتعاون مع المدرسين قبل بدء الطلبة بالإجابة عن فقرات الاختبار وضرورة الحرص والجدية في الإجابة.
٤. تم تطبيق الاختبار وجمع استجابات الطلبة وبلغ عددها (٥٦٧) ورقة.
٥. تم التحقق من افتراضات نظرية الاستجابة للفقرة، حيث قام الباحثان بالتحقق من افتراضات نظرية استجابة الفقرة، في فقرات الاختبار الكلي البالغة (٥٠) فقرة على النحو الآتي:

افتراض أحادية البعد One-dimensionality: تم التحقق من الافتراض باستخدام المؤشرات الآتية:

أولاً: المؤشرات المعتمدة على برنامج NOHARM:

أدخلت البيانات في ذاكرة الحاسوب حيث استخدم برنامج (NOHARM: Normal Ogive Harmonic Analysis Robust Method) لمعالجة هذه البيانات، المتعلقة باستجابات (٥٦٧) طالباً عن (٥٠) فقرة تمثل الاختبار الكلي، وأفرز التحليل أن قيمة مؤشر (TANAKA) تزيد عن (٠,٩٤)، وهذا مؤشر على تحقق افتراض أحادية البعد، علماً بأن أدنى قيمة مقبولة لمؤشر (TANAKA) تخص أحادية البعد هي (٠,٨) (Jasper, 2010). كما تم اعتماد مؤشر الجذر التربيعي لمتوسط مربعات البواقي (Root Mean Square Residual) (RMSR)، ذلك أنه يفضل أن تكون قيمته قريبة من (٠,٠)، وعند مقارنة القيمة الناتجة بالتحليل مع أعلى قيمة لهذا المؤشر؛ البالغة (٠,١٧) وفقاً للمعادلة $4.1/\sqrt{567}$ ، تبين أن القيمة الناتجة تقترب من الصفر وهذا مؤشر آخر على تحقق افتراض أحادية البعد (Jasper, 2010). ويبين الجدول (٥) قيم مؤشر TANAKA و RMSA التي أفرزها التحليل.

جدول (٥): مؤشرات أحادية البعد باستخدام برنامج NOHARM

مؤشرات أحادية البعد	القيمة	عدد الفقرات	عدد المفحوصين
Tanaka Index لمطابقة البيانات لأحادية البعد	٠,٩٤٨٣١٦٠٧٥٩٤٥٣٨٣	٥٠	٥٦٧
RMSR لمصفوفة البواقي	٠,٠١١٦٠٩٧٥٦٠٦١٩٠٦٦		

ثانياً: المؤشرات المعتمدة على التحليل العاملي للمكونات الرئيسية (Principal Components Analysis):

أدخلت البيانات في ذاكرة الحاسوب، حيث استخدم برنامج (SPSS) لمعالجة هذه البيانات، ومعرفة عدد العوامل المستخلصة، إذ تم إجراء تحليل عاملي لبيانات الاختبار الكلي، المتعلقة باستجابات (٥٦٧) طالباً على (٥٠) فقرة تمثل الاختبار الكلي، باستخدام طريقة المكونات الرئيسية (Principal Components Analysis). بحيث أفرز التحليل (١٨) عاملاً قيمة الجذر الكامن لكل منها تزيد عن واحد، إذ كانت قيمة الجذر الكامن للعامل الأول (١٢,١٥٢) وللعامل الثاني (٢,٦٠٧)، وبين الجدول (٦) قيم الجذور الكامنة، ونسبة التباين المُفسَّر لكل عامل، وكذلك نسبة التباين المُفسَّر التراكمية.

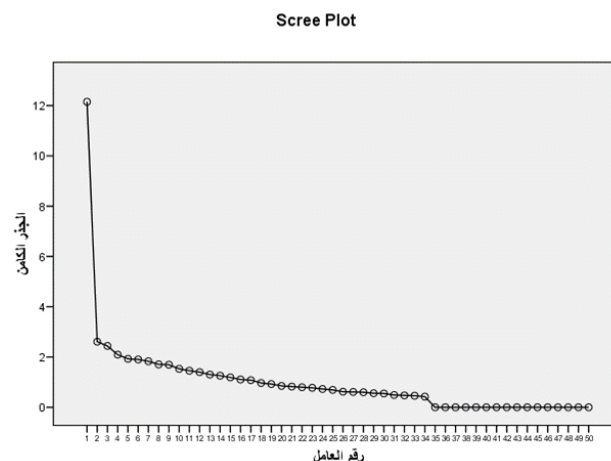
جدول (٦): نتائج التحليل العاملي للاختبار

العامل	الكلي	التباين المُفسَّر التراكمي	استخلاص مجموع التشعبات المربعة			تدوير مجموع التشعبات المربعة		
			الكلي	التباين المُفسَّر التراكمي	التباين المُفسَّر التراكمي	الكلي	التباين المُفسَّر التراكمي	التباين المُفسَّر التراكمي
١	١٢,١٥٢	٢٤,٣٠٣	١٢,١٥٢	٢٤,٣٠٣	٢٤,٣٠٣	٢,٨٨٩	٥,٧٧٨	٥,٧٧٨
٢	٢,٦٠٧	٥,٢١٥	٢,٦٠٧	٥,٢١٥	٢٩,٥١٨	٢,٤٩١	٤,٩٨٢	١٠,٧٦٠
٣	٢,٤٤٠	٤,٨٨٠	٢,٤٤٠	٤,٨٨٠	٣٤,٣٩٨	٢,٤٤٢	٤,٨٨٤	١٥,٦٤٥
٤	٢,٠٩٤	٤,١٨٧	٢,٠٩٤	٤,١٨٧	٣٨,٥٨٥	٢,٢٩٧	٤,٥٩٤	٢٠,٢٣٨
٥	١,٩٢٦	٣,٨٥٣	١,٩٢٦	٣,٨٥٣	٤٢,٤٣٨	٢,٢٩٢	٤,٥٨٥	٢٤,٨٢٣
٦	١,٩٠٣	٣,٨٠٥	١,٩٠٣	٣,٨٠٥	٤٦,٢٤٣	٢,٢٨٨	٤,٥٧٧	٢٩,٤٠٠
٧	١,٨٢٩	٣,٦٥٨	١,٨٢٩	٣,٦٥٨	٤٩,٩٠١	٢,٢٧٨	٤,٥٥٥	٣٣,٩٥٥
٨	١,٧٠٦	٣,٤١٢	١,٧٠٦	٣,٤١٢	٥٣,٣١٣	٢,٢٧٧	٤,٥٥٣	٣٨,٥٠٨
٩	١,٦٩٢	٣,٣٨٣	١,٦٩٢	٣,٣٨٣	٥٦,٦٩٦	٢,٢٣٨	٤,٤٧٦	٤٢,٩٨٥
١٠	١,٥٣٠	٣,٠٥٩	١,٥٣٠	٣,٠٥٩	٥٩,٧٥٥	٢,٢٣٥	٤,٤٧١	٤٧,٤٥٦
١١	١,٤٥٥	٢,٩١١	١,٤٥٥	٢,٩١١	٦٢,٦٦٦	٢,٢١٠	٤,٤٢٠	٥١,٨٧٥
١٢	١,٣٩٨	٢,٧٩٥	١,٣٩٨	٢,٧٩٥	٦٥,٤٦٢	٢,٢٠٤	٤,٤٠٩	٥٦,٢٨٤
١٣	١,٣٠٣	٢,٦٠٦	١,٣٠٣	٢,٦٠٦	٦٨,٠٦٧	٢,١٩٤	٤,٣٨٨	٦٠,٦٧٢
١٤	١,٢٥٧	٢,٥١٥	١,٢٥٧	٢,٥١٥	٧٠,٥٨٢	٢,١٨٠	٤,٣٦١	٦٥,٠٣٣
١٥	١,١٨٨	٢,٣٧٧	١,١٨٨	٢,٣٧٧	٧٢,٩٥٩	٢,١٢٦	٤,٢٥٢	٦٩,٢٨٤
١٦	١,١٠٦	٢,٢١٢	١,١٠٦	٢,٢١٢	٧٥,١٧٠	٢,٠١٣	٤,٠٢٥	٧٣,٣٠٩
١٧	١,٠٧٩	٢,١٥٨	١,٠٧٩	٢,١٥٨	٧٧,٣٢٨	٢,٠٠٩	٤,٠١٩	٧٧,٣٢٨
١٨	٠,٩٦٧	١,٩٣٤	٧٩,٢٦٢					

يتضح من الجدول (٦)، أنَّ قيمة الجذر الكامن للعامل الأول (١٢,١٢٥) بالمُفسَّر للعامل (٠,٣٪) من التباين الكلي، ويعد الاختبار أحادي البعد -وفقاً لأحد المؤشرات- إذا كان نسبة ما يفسره العامل الأول أكثر من (٢٠٪) تقريباً (Reckase, 1979, a cited in Lee, 2004)، وبهذا المعيار يتبين أنَّ الاختبار يحقق افتراض أحادية البعد، إذا كانت نسبة التباين المُفسَّر من العامل الأول هي نسبة عالية، مقارنةً مع ما يفسره العامل الثاني، إذ أنَّ الفجوة الكبيرة بين نسبة التباين المُفسَّر للعامل الأول، ونسبة التباين المُفسَّر للعامل الثاني، يعطي مؤشراً على تحقق افتراض أحادية البعد، ويعد وورم (Warm, 1978) أنَّ الفقرات تكون مطابقة لأحادية البعد، إذا كان الجذر الكامن للعامل الأول كبيراً بالمقارنة مع الجذر الكامن للعامل الثاني، والنسبة بين بقية الجذور الكامنة المتتالية الأخرى متقاربة، وهذا ما تحقق في بيانات الاختبار.

وكمؤشر آخر؛ تُعتمد في التحليل العاملي أحادية البعد، من خلال نسبة الجذر الكامن للعامل الأول إلى الجذر الكامن للعامل الثاني، عندما تكون نسبة كبيرة لا تقل عن (٢) (Hambelton & Swaminathan, 1985). ويتضح من الجدول (٨) أنَّ ناتج قسمة قيمة الجذر الكامن للعامل الأول، على قيمة الجذر الكامن للعامل الثاني كانت (٤,٦٦) وهي نسبة تزيد عن المعيار (٢). وكذلك كمؤشر آخر على أحادية البعد؛ عند النظر إلى نسبة الفرق بين الجذر الكامن للعامل الأول والجذر الكامن للعامل الثاني، إلى الفرق بين الجذر الكامن للعامل الثاني والجذر الكامن للعامل الثالث، تبين أنَّ النسبة كبيرة وتساوي (٥٧,٠١٦)، وهذا يبين أنَّ الاختبار أحادي البعد، علاوة على ما تقدم؛ يلاحظ من الجدول (٨)، أنَّ قيمة التباين المُفسَّر التراكمي الكلي للاختبار قد بلغت (٧٧,٣٢٨)؛ مما يشير إلى جودة تمثيل فقرات الاختبار لمحتوى المادة الدراسية الموضوعة لها (Hattie, 1985).

كما يظهر الشكل (١)، مؤشر آخر من مؤشرات أحادية البعد (Scree Plot)، إذ تم تمثيل الجذور الكامنة للعوامل جميعها بيانياً باستخدام برنامج (SPSS). حيث يلاحظ من الشكل أنّ الجذر الكامن للعامل الأول يتميز بشكل واضح عن الجذور الكامنة لبقية العوامل؛ وذلك لأنّ نسبة الجذر الكامن للعامل الأول إلى الجذر الكامن للعامل الثاني هي: (١,١٢٥:٢,٦٠٧)؛ أي أنّ الجذر الكامن للعامل الأول يعادل تقريباً خمسة أضعاف الجذر الكامن للعامل الثاني (١:٤,٦٦)، علاوةً على التغير الملحوظ في ميل المنحنى، وهذا يوضح أحادية البعد للاختبار.



الشكل (١): التمثيل البياني لقيم الجذور الكامنة للعوامل المُكوِّنة للاختبار في البيانات الكَلِيّة

❖ افتراض الاستقلال الموضوعي Local Independence:

تم إدخال البيانات في ذاكرة الحاسوب لمعالجتها؛ بهدف التحقق من افتراض الاستقلال الموضوعي باستخدام برنامج (LDID: Local Dependence Indices for Dichotomous Items)، وهو برنامج مكتوب بلغة فورتران ٩٠، يتم من خلاله حساب كل من: (ZQ3، Q3، G2) ويحتاج هذا البرنامج الى ثلاث ملفات هي: ملف قدرات المفحوصين المقدره، وملف البيانات الخام المولدة، وملف معالم الفقرات المقدره. وبالاعتماد على برنامج (LDID) تم اعتماد المؤشر (ZQ3) كحد فاصل بين أزواج الفقرات التي بينها انتهاك لافتراض الاستقلال الموضوعي من غيرها؛ فإذا كان المتوسط الحسابي لمؤشرات ZQ3 المشاهدة الخاص بأزواج الفقرات واقعاً بين الحد النظري الأدنى لمؤشر (ZQ3) الخاص بأزواج الفقرات والحد النظري الأعلى لمؤشر (ZQ3) الخاص بأزواج الفقرات، فهذا مؤشر على تحقق افتراض الاستقلال الموضوعي (Ho Kim, Cohen, & Lin, 2005). حيث تبين من التحليل أن قيمة الوسط الحسابي لمؤشر (ZQ3) الخاصة بأزواج الفقرات تقع بين الحد النظري الأدنى وبين الحد النظري الأعلى لمؤشر (ZQ3) الخاصة بأزواج الفقرات البالغ عددها (١٢٢٥)، ويبين الجدول (٧) مؤشرات الاستقلال الموضوعي وفقاً لنظرية استجابة الفقرة.

جدول (٧): مؤشرات الاستقلالية الموضوعية وفقاً لنظرية استجابة الفقرة

القيمة	الإحصائي
٥٠	عدد الفقرات
١٢٢٥	عدد أزواج الفقرات
-٠,٢٧٤٢٩	القيمة الصغرى المشاهدة لمؤشر ZQ3
٠,٥٥٤١٣	القيمة العظمى المشاهدة لمؤشر ZQ3
٠,١١٦٢٣٥٢٠٢	الانحراف المعياري المحسوب لمؤشرات ZQ3 المشاهدة الخاص بأزواج الفقرات
٠,٠٠١٧٧٣٠٥	التباين النظري لمؤشر ZQ3 الخاص بأزواج الفقرات
٠,٠٤٢١٠٧٥٩٦	الانحراف المعياري النظري لمؤشر ZQ3 الخاص بأزواج الفقرات
٠,٠٨٤٢١٥١٩٢	الانحراف المعياري النظري لمؤشر ZQ3 الخاص بأزواج الفقرات ٢x
٠,٠١٦٥٢٥٧٥	المتوسط الحسابي لمؤشرات ZQ3 المشاهدة الخاص بأزواج الفقرات
-٠,٠٦٧٦٨٩٤٤٦٠٧٥	الحد النظري الأدنى لمؤشر ZQ3 الخاص بأزواج الفقرات
٠,١٠٠٧٤٠٩٣٨١٣٨	الحد النظري الأعلى لمؤشر ZQ3 الخاص بأزواج الفقرات

يلاحظ من الجدول (٧)، أنَّ قيمة المتوسط الحسابي لمؤشر ZQ3 الخاصة بأزواج الفقرات (٠,١٦٥٢٥٧٥) تقع بين الحد الأدنى والحد الأعلى النظري لمؤشر ZQ3 الخاصة بأزواج الفقرات، وهذا مؤشر على تحقق افتراض الاستقلال الموضوعي (Ho Kim, Cohen, & Lin, 2005).

كما تم حساب عدد أزواج الفقرات التي سقطت ضمن فترة الثقة المحققة لشرط الاستقلال الموضوعي، والجدول (٨) يبين مؤشرات الاستقلال الموضوعي وفقاً لنظرية استجابة الفقرة.

جدول (٨): مؤشرات الاستقلال الموضوعي وفقاً لنظرية استجابة الفقرة

حالة الاستقلال الموضوعي	التكرار	النسبة المئوية
التبعية الموضوعية	٤٨٤	٣٩,٥
الاستقلالية الموضوعية	٧٤١	٦٠,٥
الكل	١٢٢٥	١٠٠,٠

يلاحظ من الجدول (٨)، أن عدد أزواج الفقرات التي وقعت خارج المدى (٤٨٤) بنسبة (٣٩,٥٪)، بينما كان عدد أزواج الفقرات التي وقعت ضمن المدى (٧٤١) بنسبة (٦٠,٥٪)، وهذا يبين أن عدد أزواج الفقرات التي حققت الاستقلالية أعلى بضعفين تقريباً من عدد أزواج الفقرات التي حققت التبعية الموضوعية، وهذا مؤشر على تحقق افتراض الاستقلال الموضوعي (Ho Kim, Cohen, & Lin, 2005).

وأشار هامبلتون وسواميناثان (Hambelton & Swaminathan, 1985) إلى أنَّ افتراض الاستقلال الموضوعي يكافئ افتراض أحادية البعد، وهذا يعني أنه إذا تحقق افتراض أحادية البعد في الاختبار، فإنَّ الاختبار يحقق افتراض الاستقلال الموضوعي.

افتراض عامل السرعة Seededness: تأكد الباحثان من أنَّ عامل السرعة لم يلعب دوراً في الإجابة عن فقرات الاختبار، بمعنى أنَّ إحقاق الأفراد في الاختبار، يعود إلى انخفاض قدراتهم، وليس إلى تأثير عامل السرعة في الإجابة، وكان ذلك عن طريق إعطاء الطلبة الوقت الكافي للإجابة عن فقرات الاختبار، ولم يشترك أي طالب أثناء التطبيق من ضيق الوقت، علماً بأنَّ زمن الاختبار تمَّ تقديره، في ضوء المعلومات المتوافرة للباحث، بناءً على الوقت الذي استغرقته طلبة المجموعة الاستطلاعية، في الإجابة عن الاختبار.

قام الباحثان باستخراج الخصائص السيكومترية لفقرات الاختبار قبل حذف الفقرات غير المطابقة وفق النظرية التقليدية، كما قام الباحثان باستخراج الخصائص السيكومترية لفقرات الاختبار بعد التحقق من مطابقة الأفراد والفقرات للنموذج ثنائي المعلمة على النحو الآتي:

الخصائص السيكومترية لفقرات الاختبار وفق النظرية التقليدية: تم تحليل فقرات الاختبار وفق النظرية التقليدية باستخدام برنامج (SPSS)، حيث حُسبت معاملات الصعوبة، والقدرة التمييزية لفقرات الاختبار، ولحساب مستوى الصعوبة لكل فقرة من فقرات الاختبار؛ تم اعتماد النسبة المئوية للطلبة الذين أجابوا عن الفقرة إجابة صحيحة، أما القدرة التمييزية فقد حُسبت من خلال إيجاد معامل ارتباط الفقرة بالاختبار بعد حذف الفقرة من الاختبار (معامل الارتباط المصحح (Corrected Item-Total Correlation)). والجدول (٩) يظهر معاملات الصعوبة والقدرة التمييزية لكل فقرة من فقرات الاختبار قبل حذف الفقرات غير المطابقة للنظرية التقليدية.

جدول (٩): معاملات الصعوبة والتمييز لفقرات الاختبار بالطريقة التقليدية قبل حذف الفقرات غير المطابقة

رقم الفقرة	معامل الصعوبة	معامل التمييز	رقم الفقرة	معامل الصعوبة	معامل التمييز
١	٠,٧٩	٠,٤٧	٢٦	٠,٧٢	٠,٤٢
٢	٠,٤٧	٠,٤٠	٢٧	٠,٦٨	٠,٥٨
٣	٠,٤٨	٠,١٩	٢٨	٠,٤٦	٠,٠٥
٤	٠,٣٢	٠,١٢	٢٩	٠,٤٣	٠,٤٧
٥	٠,٥١	٠,١٩	٣٠	٠,٤٣	٠,٣٦
٦	٠,١٤	٠,٠٠	٣١	٠,٦٦	٠,٤٤
٧	٠,١٧	٠,٤٣	٣٢	٠,٦٦	٠,٣٦
٨	٠,٢٠	-٠,١٥	٣٣	٠,٧٠	٠,٤١

٩	٠,٧٠	٠,٢٣	٣٤	٠,٨٣	٠,٤١
١٠	٠,٤٢	٠,٢٨	٣٥	٠,٥٣	٠,٠٥
١١	٠,٦٣	٠,٠٣	٣٦	٠,٥٤	٠,٠٥
١٢	٠,٣٠	٠,٢٥	٣٧	٠,٨٠	٠,٥١
١٣	٠,٦٣	-٠,١٥	٣٨	٠,٧٢	٠,١٠
١٤	٠,٣٠	-٠,١٩	٣٩	٠,٣٦	٠,١١
١٥	٠,٧٢	٠,٣٤	٤٠	٠,٥٣	٠,٢٧
١٦	٠,٥٧	-٠,٠٣	٤١	٠,٥٩	٠,٤٢
١٧	٠,٥١	-٠,٠٥	٤٢	٠,٤٣	٠,٤٨
١٨	٠,٦٧	٠,٣٣	٤٣	٠,٥٦	٠,٥١
١٩	٠,٦٨	٠,٥٤	٤٤	٠,٦٨	٠,٥٥
٢٠	٠,٦٧	٠,١٧	٤٥	٠,٤٩	٠,٤٢
٢١	٠,٦١	٠,٣٧	٤٦	٠,٥٢	٠,٠٩
٢٢	٠,٧٣	٠,٣٥	٤٧	٠,٥١	٠,٥٠
٢٣	٠,٧٥	-٠,٠١	٤٨	٠,٣٩	٠,٤١
٢٤	٠,٦٢	٠,٢٣	٤٩	٠,٢٨	-٠,١٢
٢٥	٠,٧١	٠,٥٢	٥٠	٠,٦٣	٠,٢١

يلاحظ من الجدول (٩)، أنَّ معاملات صعوبة الفقرات، في نموذج الصورة النهائية للاختبار، قد تراوحت بين (٠,٨٣-٠,١٤) بوسط حسابي قدره (٠,٥٥)، وانحراف معياري (٠,١٧). ويذكر عودة (٢٠١٠) أن المدى المقبول لصعوبة الفقرة كلاسيكياً يتراوح بين (٠,٢٠-٠,٨٠)، وعليه فإن الفقرات التي يقل معامل صعوبتها عن (٠,٢٠) اعتبرت صعبة في هذا الاختبار وهي الفقرات (٧، ٦) لأن معامل صعوبتهما (٠,١٤، ٠,١٧) وعددها فقرتين في هذا الاختبار. بينما توجد فقرة واحدة متطرفة في سهولتها أي يزيد معامل صعوبتها عن (٠,٨٠) وهي الفقرة (٣٤) في هذا الاختبار لأن معامل صعوبتها (٠,٨٣)، ولذا فإنه وفقاً لمعيار الصعوبة يفضل إعادة النظر في هذه الفقرات أو استبعادها.

أما معاملات التمييز للفقرات، المحسوبة على أساس إيجاد معامل الارتباط المصحح بين نتائج المفحوصين في كل فقرة، ونتائجهم في الاختبار الكلي، فتراوحت بين (٠,١٩-٠,٥٨) بوسط حسابي قدره (٠,٢٦)، وانحراف معياري (٠,٢١). ويرى عودة (٢٠١٠) أن أي فقرة تتمتع بقدرة تمييزية سالبة تحذف ولا داعي للاحتفاظ بها، وأي فقرة ذات معامل تمييز من (صفر-٠,١٩) تعتبر ضعيفة التمييز وينصح بحذفها أيضاً، في حين أي فقرة ذات معامل تمييز بين (٠,٢٠-٠,٣٩) ذات تمييز مقبول وينصح بتحسينها، وكذلك أي فقرة ذات تمييز أعلى من (٠,٣٩) تعتبر فقرة ذات تمييز جيد ويمكن الاحتفاظ بها، وبناءً على ذلك فإن الفقرات: (٣، ٤، ٥، ٦، ٨، ١١، ١٣، ١٤، ١٦، ١٧، ٢٠، ٢٣، ٢٨، ٣٥، ٣٦، ٣٨، ٣٩، ٤٦، ٤٩) تتمتع بقدرة تمييزية أقل من (٠,٢٠) لذا يفضل إعادة النظر في مثل هذه الفقرات أو استبعادها، حيث تبلغ قيمة معامل التمييز لهذه الفقرات على الترتيب: (٠,١٩، ٠,١٢، ٠,١٩، ٠,٠٠، ٠,١٥، ٠,٠٣، ٠,١٥، ٠,٠٣، ٠,٠٥، ٠,٠٥، ٠,٠١، ٠,٠٥، ٠,٠٥، ٠,١٠، ٠,١١، ٠,٠٩، ٠,١٢) وعددها تسعة عشرة فقرة.

وبناءً على ما سبق؛ فقد تم استبعاد الفقرات وفق معيارَي الصعوبة والتمييز؛ مع مراعاة أن تتم عملية الحذف للفقرات وفقاً لمعامل التمييز الضعيف الخاص بكل فقرة بشكل متدرج؛ بمعنى أنه تم حذف الفقرة ذات معامل التمييز الأدنى ثم تم حساب معاملات التمييز لبقية الفقرات مرة أخرى، وهكذا إلى أن تم حذف كافة الفقرات ذات معامل التمييز الضعيف؛ مما ترتب عليه حذف الفقرات: (٣، ٤، ٥، ٦، ٨، ١٤، ١٦، ١٧، ٢٨، ٣٥، ٣٦، ٣٩، ٤٩) وعددها اثنتا عشرة فقرة، وحذف الفقرتين (٧، ٢٤) وفقاً لمعامل الصعوبة الخاص بكلاً منهما؛ أي أنه تم حذف ما نسبته ٢٨٪ من مجمل فقرات الاختبار. النتائج المتعلقة بالخصائص السيكومترية وفق النموذج ثنائي المعلمة:

فقد قام الباحثان بإجراء مطابقة لأفراد عينة الدراسة باستخدام برنامج (BILOG-MG 3)، بالاعتماد على استجابات أفراد الدراسة على نموذج الاختبار المُعد؛ إذ يُعد هذا الإجراء خطوة أساسية، وذلك لاختيار الأفراد المطابقين للنموذج اللوجستي ثنائي المعلمة، حيث أفرزت نتائج التحليل لأول مرة عند مستوى الدلالة ($\alpha=0,01$) مطابقة استجابات المفحوصين للنموذج ثنائي المعلمة، حيث كانت قيمة الاحتمالية لهم أكبر من (٠,٠١).

ومن ثم قام الباحثان بإعادة التحليل مرة أخرى لاختبار مدى مطابقة الفقرات للنموذج اللوجستي ثنائي المعلمة باستخدام برنامج (BILOG-MG 3). حيث قُدرت معلمة الصعوبة، ومعلمة التمييز لفقرات الاختبار، كما تم حساب مدى مطابقة الفقرة للنموذج ثنائي المعلمة عن طريق الإحصائي (χ^2) لحسن المطابقة، الذي يفيد في تحديد الفقرات ذات التقديرات الضعيفة لمعاملها، ومثل هذه الفقرات تحمل قيمة لهذا الإحصائي كبيرة نسبياً مقارنة مع الفقرات الأخرى، وبناءً عليه فإن مثل هذه الفقرات يفضل إعادة النظر فيها أو استبعادها من الاختبار. والجدول (١٠) يبين المؤشرات الإحصائية لكل فقرة وفقاً للنموذج ثنائي المعلمة.

جدول (١٠): معلمة الصعوبة والتمييز لفقرات الاختبار وفقاً لنظرية استجابة الفقرة قبل حذف الفقرات غير المطابقة

رقم الفقرة	التمييز	الخطأ المعياري للتمييز	الصعوبة	الخطأ المعياري للصعوبة	ك	درجة الحرية	مطابقة الفقرة
١	٢,٣٢٣	٠,٢٨٧	-٠,٩٩٦	٠,٠٦٧	٥,٣	٥	٠,٣٧٩٨
٢	١,١٥٥	٠,١٣٥	٠,٠٤	٠,٠٨٧	٧,١	٨	٠,٥٢٥١
٣	٠,٤٤٥	٠,٠٧٩	٠,١٥٤	٠,١٩٤	٧,٨	٩	٠,٥٥١٠
٤	٠,٢٦٧	٠,٠٦٥	٢,٨٥١	٠,٧٤١	٧٢,٧	٩	٠,٠٠٠٠
٥	٠,٨٥٢	٠,١٢٤	-٠,١٢٣	٠,١٠٦	٥,٣	٨	٠,٧٢٨٠
٦	٠,٢٤٨	٠,٠٦٨	٧,٠٧٤	٢,٠٠٦	٢٢,٣	٩	٠,٠٠٨٠
٧	١,٢٢٧	٠,١٢٥	١,٦٥٨	٠,١٨	٣٦	٨	٠,٠٠٠٠
٨	١,٢٣٩	٠,٢٣٦	٤,٠٨٢	٠,٥٣١	٤٥	٨	٠,٠٠٠٠
٩	١,٠٨٣	٠,١٦٧	-٠,٩٨٦	٠,١٢٦	٠,٨	٦	٠,٩٩١٢
١٠	١,٠٨٣	٠,١٣٥	٠,٢٧٩	٠,١٠٣	٨,٢	٨	٠,٤١٠٧
١١	٢,٢٢٨	٠,٢٤٤	-٠,٥١٦	٠,٠٥	٦,٢	٥	٠,٢٨٩٨
١٢	٠,٧٩٣	٠,١٠٦	١,١٧٦	٠,٢٠٤	٤٧,٣	٩	٠,٠٠٠٠
١٣	١,٢١٥	٠,١٦٣	-٠,٦٢٨	٠,٠٨٧	٥,٦	٧	٠,٥٩٢٣
١٤	٣,٤١٩	٠,٨١٠	-٤,١٠٥	٠,٨٧١	٩٩,٤	٨	٠,٠٠٠٠
١٥	١,٣١٧	٠,١٩٢	-٠,٩٩٩	٠,١١	٩	٦	٠,١٧٦٢
١٦	٠,١٢٢	٠,٠٣٢	-١,٩٧	٠,٨٤٢	٩٤,٩	٩	٠,٠٠٠٠
١٧	٠,١٣٢	٠,٠٣٥	-٠,٣٨	٠,٦٠٩	٦١,٦	٩	٠,٠٠٠٠
١٨	٠,٧٤	٠,١٢٣	-١,١٠٥	٠,١٨٧	٢١,٧	٨	٠,٠٠٠٥
١٩	٢,٣٠٣	٠,٢٨٣	-٠,٦٥	٠,٠٥	٣,٧	٤	٠,٤٤٢٢
٢٠	١,٤٩٩	٠,١٨٥	-٠,٧١٥	٠,٠٧٦	٤,٣	٦	٠,٦٣٧٢
٢١	٢,٣٨٩	٠,٢١٩	-٠,٤٥٢	٠,٠٤٨	٨,٤	٥	٠,١٣٤٦
٢٢	١,٧٨٨	٠,٢٠٣	-٠,٨٧١	٠,٠٧٤	٧,٧	٥	٠,١٧٥٩
٢٣	١,١٨١	٠,١٦٥	-١,٢٠٥	٠,١٤	٥,٨	٦	٠,٤٤٩١
٢٤	١,٣٩	٠,١٦٤	-٠,٥٣٣	٠,٠٧٥	١٤	٧	٠,٠٥٠٦
٢٥	١,٢٠٢	٠,١٧١	-٠,٩٨٥	٠,١١٧	٦,٨	٦	٠,٣٣٦٩
٢٦	١,٣٧٨	٠,١٨٢	-٠,٩٥٣	٠,١٠١	١٦,٤	٦	٠,٠١١٩
٢٧	٢,٠٧٥	٠,٢٣٦	-٠,٦٧٩	٠,٠٥٦	١,٩	٥	٠,٨٦٠٧
٢٨	٠,١٥٣	٠,٠٤	١,٠٤٦	٠,٥٨٥	١١١	٩	٠,٠٠٠٠
٢٩	٠,٩٢٨	٠,١١٩	٠,٣٠٧	٠,١١٥	١٣,٣	٩	٠,١٤٨٧
٣٠	١,٣٩٩	٠,١٥٩	٠,١٥٧	٠,٠٨٣	١٣,٣	٧	٠,٠٦٤٣
٣١	١,٨٣٤	٠,٢	-٠,٦٣٥	٠,٠٦١	٣,١	٥	٠,٦٨٣٦
٣٢	١,٩٠٨	٠,٢٠٦	-٠,٦٢٢	٠,٠٥٩	٧	٥	٠,٢٢١٢
٣٣	٢,٥٦١	٠,٢٨١	-٠,٧	٠,٠٤٨	٤,٨	٥	٠,٤٣٧٠
٣٤	١,٩٩٥	٠,٢٣	-١,٢١٩	٠,٠٩٢	١٤,١	٥	٠,٠١٥١
٣٥	٠,١٣٧	٠,٠٣٥	-٠,٧٠٢	٠,٦١٧	١٢٨	٩	٠,٠٠٠٠
٣٦	٠,١٥٧	٠,٠٤	-٠,٨٧٩	٠,٥٧١	١٤١	٩	٠,٠٠٠٠

٣٧	٢,٦٥٧	٠,٣٠٣	-٠,٩٧٨	٠,٠٥٨	٦	٥	٠,٣٠٩٦
٣٨	١,٤٤٩	٠,١٧٩	-٠,٩٤٣	٠,٠٩٤	٩,٢	٦	٠,١٦٥١
٣٩	٠,٢٢	٠,٠٥٦	٢,٤٧١	٠,٧١٣	١٠,٣	٩	٠,٠٠٠٠
٤٠	١,٣٦٢	٠,١٨٤	-٠,٢١٣	٠,٠٧٢	٦,١	٥	٠,٢٩٥٤
٤١	١,٥٢٨	٠,١٨٦	-٠,٤١	٠,٠٦٥	٠,٩	٦	٠,٩٩٠١
٤٢	١,٢٥٦	٠,١٤٥	٠,٢٠٩	٠,٠٨٨	٥,٣	٨	٠,٧٢٢١
٤٣	١,٢٣٧	٠,١٤٤	-٠,٣١٩	٠,٠٧٧	١,٥	٨	٠,٩٩٢١
٤٤	١,٤٣٤	٠,٢٠٧	-٠,٧٥٦	٠,٠٨٢	٩	٥	٠,١٠٩٩
٤٥	١,٤٧٢	٠,١٨٦	-٠,٠٥٥	٠,٠٧٢	٦,٥	٥	٠,٢٦٠٥
٤٦	١,٦٢٦	٠,١٨٥	-٠,١٨٩	٠,٠٦٤	١١,٥	٦	٠,٠٧٤٤
٤٧	١,٣٢٢	٠,١٦٨	-٠,١٣	٠,٠٧٥	٤,١	٦	٠,٦٦٥٤
٤٨	١,٤٩٥	٠,١٦٩	٠,٣٢٥	٠,٠٨٨	١٢,٨	٦	٠,٠٤٦٥
٤٩	٠,٢٢	٠,٩٢٨	٢,٤٧١	٠,٨٩٩	٧٤,٨	٩	٠,٠٠٠٠
٥٠	١,٠٧٨	٠,١٣٣	-٠,٦٧٥	٠,١	٧,٦	٨	٠,٤٧٦٢

يتضح من الجدول (١٠)، أنَّ الفقرات ذوات الأرقام: (٤، ٦، ٧، ٨، ١٢، ١٤، ١٦، ١٧، ١٨، ٢٨، ٣٥، ٣٦، ٣٩، ٤٩) لم تكن مطابقة للنموذج ثنائي المعلمة وفقاً لمؤشر المطابقة المعتمد على قيمة (χ^2) لحسن المطابقة عند مستوى دلالة إحصائية $(\alpha = ٠.٠١)$ ؛ أي أنه تم استبعاد هذه الفقرات وعددها أربعة عشر فقرة أي ما نسبته ٢٨٪.

النتائج المتعلقة بثبات الاختبار قبل الحذف وفقاً لنظرية استجابة الفقرة والنظرية التقليدية:

قام الباحثان باستخدام برنامج (SPSS) لتقدير معامل الثبات باستخدام معامل كودر-ريتشاردسون ٢٠ (KR-٢٠)، كما قام باستخدام برمجة (BILOG-MG 3) لتقدير الثبات الإمبريقي (التجريبي)، والجدول (١١) يبين قيم معاملات الثبات للاختبار قبل حذف الفقرات غير المطابقة وفقاً للمؤشرات التقليدية والنموذج ثنائي المعلمة.

جدول (١١): قيم معاملات الثبات للاختبار قبل حذف الفقرات غير المطابقة للنظرية التقليدية والنموذج ثنائي المعلمة

النظرية	الثبات	عدد الفقرات
التقليدية	٠,٨٦٨	٥٠
استجابة الفقرة	٠,٩٤٩	٥٠

يلاحظ من الجدول (١١)، أن قيمة معامل ثبات الاختبار وفق النظرية التقليدية قبل حذف الفقرات كانت (٠,٨٦٨)، بينما بلغت قيمة معامل الثبات الإمبريقي للاختبار قبل حذف الفقرات (٠,٩٤٩).

٦. نتائج الدراسة

للإجابة عن سؤال الدراسة الأول الذي نص على: "ما مدى التوافق في نتائج تصنيف الفقرات الخاصة باختبار مبحث العلوم للصف السادس الأساسي إلى (مطابقة أو غير مطابقة) وفقاً لـ: (نظرية استجابة الفقرة باستخدام النموذج الثنائي المعلمة، النظرية التقليدية)؟" وفي ضوء ما تقدم من كشفٍ للفقرات غير المطابقة وفقاً للنظرية التقليدية؛ فقد تم استبقاء (٣٦) فقرة كما في الجدول (١٢)، الذي يظهر معاملات الصعوبة والقدرة التمييزية لكل فقرة من فقرات الاختبار المستبقاة.

جدول (١٢): معاملات الصعوبة والتمييز كلاسيكياً بعد حذف الفقرات غير المطابقة

رقم الفقرة	معامل الصعوبة	معامل التمييز
١	٠,٧٩	٠,٤٤
٢	٠,٤٧	٠,٤١
٥	٠,٥١	٠,٢٩
٩	٠,٧٠	٠,٣٣
١٠	٠,٤٢	٠,٣٩
١١	٠,٦٣	٠,٥٤

١٢	٠,٣٠	٠,٢٦
١٣	٠,٦٣	٠,٣٨
١٥	٠,٧٢	٠,٣٦
١٨	٠,٦٧	٠,٢٦
١٩	٠,٦٨	٠,٥٣
٢٠	٠,٦٧	٠,٤٢
٢١	٠,٦١	٠,٥٧
٢٢	٠,٧٣	٠,٤٦
٢٣	٠,٧٥	٠,٣٥
٢٤	٠,٦٢	٠,٤٤
٢٥	٠,٧١	٠,٣٥
٢٦	٠,٧٢	٠,٤٠
٢٧	٠,٦٨	٠,٥٠
٢٩	٠,٤٣	٠,٣٥
٣٠	٠,٤٣	٠,٤٨
٣١	٠,٦٦	٠,٤٧
٣٢	٠,٦٦	٠,٥٠
٣٣	٠,٧٠	٠,٥٣
٣٧	٠,٨٠	٠,٤٧
٣٨	٠,٧٢	٠,٤١
٤٠	٠,٥٣	٠,٤٥
٤١	٠,٥٩	٠,٤٧
٤٢	٠,٤٣	٠,٤٤
٤٣	٠,٥٦	٠,٤٣
٤٤	٠,٦٨	٠,٤٠
٤٥	٠,٤٩	٠,٤٧
٤٦	٠,٥٢	٠,٥٠
٤٧	٠,٥١	٠,٤٥
٤٨	٠,٣٩	٠,٥٠
٥٠	٠,٦٣	٠,٣٧

وكذلك في ضوء ما تقدم من كشفٍ للفقرات غير المطابقة وفقاً لنظرية استجابة الفقرة؛ فقد تم استبقاء (٣٦) فقرة كما في الجدول (١٣)، الذي يظهر معلّمة الصعوبة والتمييز لكل فقرة من فقرات الاختبار المستبقاة.

جدول (١٣): معلمة الصعوبة والتمييز لفقرات الاختبار وفقاً لنظرية استجابة الفقرة بعد حذف الفقرات غير المطابقة

رقم الفقرة	التمييز	الخطأ المعياري للتمييز	الصعوبة	الخطأ المعياري للصعوبة	٢٤	درجة الحرية	مطابقة الفقرة
١	٢,٧٠٧	٠,٣٢٥	-٠,٩٢٢	٠,٠٥٧	٣,٧	٥	٠,٥٨٦٢
٢	١,٢٣٥	٠,١٤٢	٠,٠١٤	٠,٠٨١	١,٨	٨	٠,٩٨٦١
٣	٠,٤٧٥	٠,٠٨	٠,١٣٧	٠,١٨١	٩,٤	٩	٠,٣٩٧٦
٥	٠,٨٨٩	٠,١٢٣	-٠,١٣٢	٠,١٠١	٤,٣	٨	٠,٨٢٦٣
٩	١,٢١٦	٠,١٨٣	-٠,٩١٨	٠,١١١	١,٥	٦	٠,٩٦٢١
١٠	١,١٠٧	٠,١٣١	٠,٢٥٦	٠,٠٩٧	٤	٨	٠,٨٥٩٥
١١	٢,٥٢٨	٠,٢٦٨	-٠,٥٠٤	٠,٠٤٥	١,١	٤	٠,٨٨٧٧
١٣	١,٣٣٤	٠,١٧٢	-٠,٦٠٤	٠,٠٧٩	٨	٧	٠,٣٢٩٨
١٥	١,٤٦٩	٠,٢٠٥	-٠,٩٣٥	٠,٠٩٧	٤,٢	٦	٠,٦٥٥٩
١٩	٢,٦١	٠,٣١٨	-٠,٦٢٤	٠,٠٤٤	٣	٣	٠,٣٩١٩

٠,٤٣٨٨	٦	٥,٩	٠,٠٦٩	-٠,٦٨٦	٠,١٩١	١,٦٤١	٢٠
٠,٢٢٣٣	٤	٥,٧	٠,٠٤٣	-٠,٤٤٨	٠,٢٣٩	٢,٦٩٨	٢١
٠,٦٩٣٥	٥	٣	٠,٠٦٦	-٠,٨٢١	٠,٢٢٥	٢,٠٠٢	٢٢
٠,٣٥٦٧	٦	٦,٦	٠,١٢٥	-١,١٣٤	٠,١٦٩	١,٢٨٧	٢٣
٠,٣٨٩١	٦	٦,٣	٠,٠٧	-٠,٥٢٢	٠,١٧١	١,٤٩٣	٢٤
٠,٦٧٥٧	٦	٤	٠,١٠١	-٠,٩١٤	٠,١٨٩	١,٣٦٣	٢٥
٠,٩٥٢٦	٥	١,١	٠,٠٨٩	-٠,٨٩٢	٠,١٩٧	١,٥٤٢	٢٦
٠,٨٧٨٣	٥	١,٨	٠,٠٤٩	-٠,٦٤٨	٠,٢٦٧	٢,٣٦١	٢٧
٠,٤٧٧١	٩	٨,٦	٠,١٠٧	٠,٢٦٦	٠,١٢٧	١,٠٠٥	٢٩
٠,٠٤٤٣	٦	١٢,٩	٠,٠٧٧	٠,١٢	٠,١٦٧	١,٤٩٥	٣٠
٠,٨٦٩٦	٤	١,٣	٠,٠٥٤	-٠,٦٠٨	٠,٢٢٥	٢,٠٨٣	٣١
٠,٢٢٦١	٤	٥,٧	٠,٠٥٣	-٠,٥٩٩	٠,٢٢٧	٢,١٣٧	٣٢
٠,٩٥١٨	٥	١,١	٠,٠٤٢	-٠,٦٦٤	٠,٣٢٧	٣,٠٠٤	٣٣
٠,١٧٢٤	٤	٦,٤	٠,٠٧٥	-١,١٠٢	٠,٢٦٢	٢,٣٨٥	٣٤
٠,٤٦٩٤	٤	٣,٦	٠,٠٤٩	-٠,٩٠٥	٠,٣٥٨	٣,١٣٢	٣٧
٠,١٤١٨	٤	٦,٩	٠,٠٨٣	-٠,٨٨٣	٠,١٩٩	١,٦٢٩	٣٨
٠,٧٣١١	٤	٢	٠,٠٦٥	-٠,٢٢٧	٠,٢٠٧	١,٥٢٧	٤٠
٠,٨٧٢٤	٦	٢,٥	٠,٠٥٧	-٠,٤٠٥	٠,٢١٢	١,٧٤٥	٤١
٠,٩٩٢٥	٧	١,١	٠,٠٨١	٠,١٦	٠,١٦٣	١,٣٩١	٤٢
٠,٩٩٢	٧	١,٢	٠,٠٧	-٠,٣٢١	٠,١٦	١,٣٦٤	٤٣
٠,٠٣٩٤	٥	١١,٧	٠,٠٧٢	-٠,٧١٣	٠,٢٣٢	١,٦٢٥	٤٤
٠,٣٥٩١	٥	٥,٥	٠,٠٦٥	-٠,٠٨٦	٠,٢٠٩	١,٦٤٦	٤٥
٠,٠٨٣٧	٦	١١,٢	٠,٠٥٧	-٠,٢٠٩	٠,٢٠٧	١,٨٣٥	٤٦
٠,٧٩٤٢	٦	٣,١	٠,٠٦٩	-٠,١٤٩	٠,١٨٤	١,٤٤٥	٤٧
٠,١٧٢٩	٥	٧,٧	٠,٠٨٢	٠,٢٦٦	٠,١٨٩	١,٦٤٥	٤٨
٠,٧٢١١	٧	٤,٥	٠,٠٩١	-٠,٦٤٦	٠,١٤٣	١,١٨١	٥٠

قام الباحثان في ضوء نتائج الجدولين (١٤، ١٥)، بإجراء المقارنة بين نتائج المؤشرات التقليدية ونتائج مؤشرات النموذج ثنائي المعلمة مع بعضها البعض، حيث قام باستخراج عدد ونسبة الفقرات المستبعدة في كل نظرية من النظريتين، حيث يلاحظ منهما أن مؤشرات النموذج اللوجستي ثنائي المعلمة استبعدت الفقرات التي استبعدتها النظرية التقليدية عدا الفئتين (٣، ٣٤)، فقد قبلها النموذج في حين استبعدتها النظرية التقليدية، كما أن النموذج اللوجستي ثنائي المعلمة استبعد الفئتين (١٢، ١٨) في حين قبلتها النظرية التقليدية. والجدول (١٤) يوضح الفقرات التي تم استبعادها وفقاً لكل من: المؤشرات التقليدية والنموذج اللوجستي ثنائي المعلمة.

جدول (١٤): أرقام الفقرات المحذوفة وفقاً للمؤشرات التقليدية والنموذج اللوجستي ثنائي المعلمة

أرقام الفقرات المحذوفة وفقاً لنظرية	
التقليدية	استجابة الفقرة
٣	
٤	٤
٦	٦
٧	٧
٨	٨
	١٢
١٤	١٤
١٦	١٦

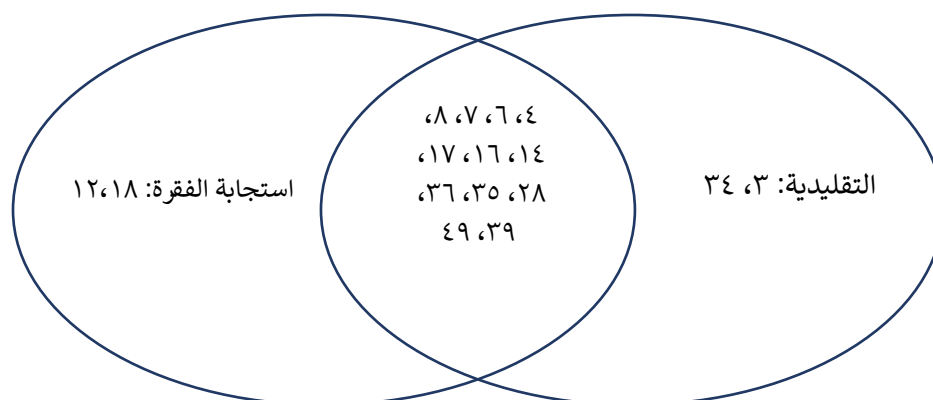
١٧	١٧
١٨	
٢٨	٢٨
	٣٤
٣٥	٣٥
٣٦	٣٦
٣٩	٣٩
٤٩	٤٩

يلاحظ من الجدول (١٤)، أنه يوجد اتفاق بين المؤشرات التقليدية والنموذج اللوجستي ثنائي المعلمة في عدد الفقرات التي تم استبعادها، حيث توجد (١٢) فقرة تم حذفها وفقاً للنظرية التقليدية ونظرية استجابة الفقرة مشكلة ما نسبته (٧٥٪) من مجمل الفقرات المحذوفة، في حين لا يوجد اتفاق بين المؤشرات التقليدية والنموذج اللوجستي ثنائي المعلمة في فقرتين تم استبعادهما وفقاً للمؤشرات التقليدية وقبلها النموذج الثنائي مشكلة ما نسبته (١٢,٥٪) من مجمل الفقرات المحذوفة، وكذلك لا يوجد اتفاق بين المؤشرات التقليدية والنموذج اللوجستي ثنائي المعلمة في فقرتين تم استبعادهما وفق النموذج الثنائي في حين قبلتها النظرية التقليدية مشكلة ما نسبته (١٢,٥٪) من مجمل الفقرات المحذوفة، والجدول (١٥) يوضح أعداد الفقرات التي تم حذفها والنسب المئوية الخاصة بها وفقاً لمؤشرات النظريتين.

جدول (١٥): مدى التوافق في الفقرات المحذوفة وفقاً للنظرية التقليدية ونظرية استجابة الفقرة

الفقرات المحذوفة وفقاً:	عدد الفقرات المحذوفة	نسبة الفقرات المحذوفة
للنظريتين	١٢	٧٥,٠
للكلاسيكية	٢	١٢,٥
لنظرية استجابة الفقرة	٢	١٢,٥
الكلية	١٦	١٠٠,٠

كما يمكن توضيح نسبة الاتفاق بين النظريتين في انتقاء الفقرات الواردة في الجدول (١٥) بأشكال فن، كما هو موضح في الشكل (٢).



الشكل (٢): نسبة الاتفاق بين المؤشرات التقليدية والنموذج اللوجستي ثنائي المعلمة في انتقاء الفقرات

وعليه فإن الفقرات التي تم حذفها وفقاً للمؤشرات التقليدية ومؤشرات النموذج ثنائي المعلمة صنفت على أنها فقرات غير مطابقة بينما الفقرات التي لم تستبعد فصنفت على أنها فقرات مطابقة للنموذج أو النظرية. والجدول (١٦) يبين عدد ونسبة الفقرات وفقاً لتصنيف الفقرات حسب حالة مطابقتها في كلاً من النظريتين.

جدول (١٦): تصنيف الفقرات حسب مطابقتها للنموذج اللوجستي ثنائي المعلمة والنظرية التقليدية

حالة الفقرة وفقاً:	عدد الفقرات المحذوفة	نسبة الفقرات المحذوفة
غير مطابقة في النظريتين	١٢	٢٤,٠
غير مطابقة كلاسيكياً ومطابقة في استجابة الفقرة	٢	٤,٠
غير مطابقة في استجابة الفقرة ومطابقة كلاسيكياً	٢	٤,٠
مطابقة في النظريتين	٣٤	٦٨,٠
الكل	٥٠	١٠٠,٠

يلاحظ من الجدول (١٦)، أن عدد الفقرات غير المطابقة في النظريتين (١٢) فقرة أي ما نسبته (٢٤٪) من مجمل فقرات الاختبار، بينما عدد الفقرات غير المطابقة للنظرية التقليدية ومطابقة في نظرية استجابة الفقرة (٢) فقرة أي ما نسبته (٤٪) من مجمل فقرات الاختبار، وكذلك فإن عدد الفقرات غير المطابقة في نظرية استجابة الفقرة ومطابقة كلاسيكياً (٢) ما نسبته (٤٪) من مجمل فقرات الاختبار، في حين كان عدد الفقرات المطابقة في النظريتين (٣٤) فقرة أي ما نسبته (٦٨٪) من مجمل فقرات الاختبار.

وكذلك قام الباحثان بإيجاد نسبة الاتفاق بين نتائج المؤشرات التقليدية ونتائج مؤشرات النموذج ثنائي المعلمة في تصنيف الفقرات إلى (مطابقة أو غير مطابقة)، وذلك على النحو الآتي:

$$\text{نسبة الاتفاق} = \frac{١٢ + ٣٤}{١٢ + ٢ + ٢ + ٣٤} \times ١٠٠\% = ٩٢\%$$

أي أن (٩٢٪) من فقرات الاختبار (٥٠ فقرة) اتفقت النظريتان في الحكم عليها، حيث اتفق النظريتان على بقاء (٣٤) فقرة من فقرات الاختبار، في حين اتفقتا على استبعاد (١٢) فقرة، أما النسبة المتبقية (٨٪)؛ فقد اختلفتا في الحكم عليها، فقرتين استبعدتهما المؤشرات التقليدية ولم يستبعدهما النموذج الثنائي المعلمة وهما (٣، ٣٤)، وفقرتين استبعدهما النموذج الثنائي ولم تستبعدهما المؤشرات التقليدية وهما (١٢، ١٨). وعليه يمكن القول إن النسبة هذه تعتبر نسبة عالية.

والجدول (١٧)، يوضح نتائج اختبار χ^2 لحسن المطابقة؛ للكشف عما إذا كان توزيع فقرات الاختبار الملاحظ وفقاً لحالة مطابقة الفقرة يختلف جوهرياً عن توزيعها المتوقع وفقاً لحالة مطابقة الفقرة.

جدول (١٧): نتائج اختبار χ^2 لحسن مطابقة توزيع الفقرات الملاحظ وفقاً لحالة مطابقة الفقرة

حالة مطابقة الفقرة وفقاً:					
الإحصائي	غير مطابقة في النظريتين	غير مطابقة كلاسيكياً	غير مطابقة في استجابة الفقرة	مطابقة في النظريتين	الكل
عدد الفقرات المحذوفة	١٢	٢	٢	٣٤	٥٠
نسبة الفقرات المحذوفة	٢٤	٤	٤	٦٨	١٠٠
التكرار المتوقع	١٢,٥	١٢,٥	١٢,٥	١٢,٥	٥٠
الباقى	-٠,٥	-١٠,٥	-١٠,٥	٢١,٥	
مربع الباقي	٠,٢٥	١١٠,٢٥	١١٠,٢٥	٤٦٢,٢٥	
χ^2 الفرعية	٠,٢	٨,٨٢	٨,٨٢	٣٦,٩٨	
الباقى المعياري	-٠,١٤١	-٢,٩٧٠	-٢,٩٧٠	٦,٠٨١	
χ^2				٥٤,٦٤	
درجة الحرية				٣	
الدلالة الإحصائية				٠,٠٠٠	

يتضح من الجدول (١٧)، أن تكرار الفقرات المطابقة في النظريتين؛ قد أسهم مساهمة إيجابية فاعلة في حساب χ^2 وبفارق جوهري عن التكرار المتوقع لها؛ بمعنى أن التكرار الملاحظ لها كان أكبر وبفارق جوهري عما هو متوقع له فيها، في

حين أن تكرر الفقرات (غير المطابقة في النظريتين، وغير المطابقة كلاسيكياً، وغير المطابقة في استجابة الفقرة)، قد أسهم مساهمة سلبية فاعلة في حساب χ^2 وبفارق جوهري عن التكرار المتوقع لها؛ بمعنى أن التكرار المشاهد لها كان أقل وبفارق جوهري عما هو متوقع له فيها.

للإجابة عن سؤال الدراسة الثاني الذي نص على: "هل يوجد اختلاف بين قيمة معامل ثبات الاختبار وفقاً للنظريتين (النظرية التقليدية، نظرية استجابة الفقرة للنموذج الثنائي المعلمة) عند مستوى الدلالة الإحصائية ($\alpha=0.05$)؟"؛ قام الباحثان باستخدام برنامج (SPSS) لتقدير معامل الثبات باستخدام معامل كودر-ريتشاردسون 20 (KR-20)، كما قام باستخدام برمجية (BILOG-MG 3) لتقدير الثبات الإمبريقي (التجريبي)، والجدول (١٨) يبين قيم معاملات الثبات للاختبار بعد حذف الفقرات غير المطابقة وفقاً للمؤشرات التقليدية والنموذج ثنائي المعلمة.

جدول (١٨): قيم معاملات الثبات للاختبار بعد حذف الفقرات غير المطابقة للنظرية التقليدية والنموذج ثنائي المعلمة

النظرية	الثبات	عدد الفقرات
التقليدية	٠,٩٠١	٣٦
استجابة الفقرة	٠,٩٥١	٣٦

يلاحظ من الجدول (١٨)، أن قيمة معامل ثبات الاختبار وفق النظرية التقليدية بعد حذف الفقرات غير المطابقة قد ارتفعت لتصبح (٠,٩٠١)، بينما بلغت قيمة معامل الثبات الإمبريقي للاختبار بعد حذف الفقرات غير المطابقة (٠,٩٥١)؛ وللكشف عن دلالة الفرق بين معاملي الثبات الكلاسيكي والإمبريقي؛ تم استخدام معادلة (F) Modified Feldt, (1969) للقياسات المترابطة في النظريتين، حسب العلاقة الرياضية الآتية:

$$F = \frac{1 - \alpha_1}{1 - \alpha_2}$$

حيث:

α_1 : التقدير العيني لمعامل ثبات النظرية التقليدية.

α_2 : التقدير العيني لمعامل ثبات نظرية استجابة الفقرة (الثبات الإمبريقي).

وبين الجدول (١٩) نتائج التحليل:

جدول (١٩): نتائج الإحصائي (F) لفحص مدى الاختلاف بين معاملي ثبات الاتساق الداخلي في نموذجي الاختبار

ثبات الاختبار	الإحصائي	وفقاً للنظرية التقليدية (٠,٩٠١٣١)
وفقاً لنظرية استجابة الفقرة (٠,٩٥٠٦)	قيمة F المحسوبة	١,٩٩٨
	معامل الارتباط	٠,٩٣١
	درجة الحرية	٤٢١٧,٣٨٧٤
	الدلالة الإحصائية	٠,٠٠٠٠

يتضح من الجدول (٢١)، وجود فرق دال إحصائياً عند مستوى ($\alpha=0.05$) بين معاملي ثبات الاتساق الداخلي لنموذج الاختبار، إذ بلغت قيمة F (١,٩٩٨).

للإجابة عن سؤال الدراسة الثالث الذي نص على: "هل يوجد اختلاف بين قيمة معامل الصدق المحكي للاختبار وفق النظريتين (النظرية التقليدية ونظرية استجابة الفقرة للنموذج الثنائي المعلمة) عند مستوى الدلالة الإحصائية ($\alpha=0.05$)؟"؛ تم حساب معامل الصدق المحكي للاختبار بعد حذف الفقرات غير المطابقة وفقاً لنظرية استجابة الفقرة والنظرية التقليدية، حيث بلغت قيمة معامل الصدق المحكي للاختبار بعد حذف الفقرات غير المطابقة وفقاً لنظرية استجابة الفقرة (٠,٩١)، بينما بلغت قيمة معامل الصدق المحكي للاختبار بعد حذف الفقرات غير المطابقة وفقاً للنظرية التقليدية (٠,٧٦)؛ وللكشف عن دلالة الفرق بين معاملي الصدق المحكي وفقاً للنظريتين التقليدية واستجابة الفقرة؛ تم

استخدام معادلة اختبار (t) لمعامل ارتباط للقياسات المترابطة في النظريتين، حسب العلاقة الرياضية الآتية:

$$t = \frac{(r_{xy} - r_{xz}) \times \sqrt{(n-3) \times (1 + r_{yz})}}{\sqrt{2 \times (1 - (r_{xy}^2 + r_{xz}^2 + r_{yz}^2) + (2 \times (r_{xy}r_{xz}r_{yz})))}}$$

حيث:

r_{xy} : معامل الارتباط بين العلامة الكلية على الاختبار وفقاً للنظرية التقليدية وبين المحك.

r_{xz} : معامل الارتباط بين العلامة الكلية على الاختبار وفقاً لنظرية استجابة الفقرة وبين المحك.

r_{yz} : معامل الارتباط بين العلامة الكلية على الاختبار وفقاً للنظرية التقليدية وبين العلامة الكلية على الاختبار وفقاً لنظرية استجابة الفقرة.

n: حجم العينة.

ويبين الجدول (٢٠) نتائج التحليل:

جدول (٢٠): نتائج الإحصائي (t) لفحص جوهرية الاختلاف بين معاملي الصدق المحكي للاختبار وفقاً للنظريتين

الاختبار	الارتباط مع المحك	IRT
		٠,٩١٠
	قيمة الإحصائي ت	٢,٠١٦
	الارتباط البيئي	٠,٩٩٤
	عدد أفراد العينة	٥٦٧
	درجة الحرية	٥٦٤
	الدلالة الإحصائية	٠,٠٤٤
CTT	٠,٧٦٠	

يتضح من الجدول (٢٠)، وجود فرق دال إحصائياً عند مستوى ($\alpha=0,05$) بين معاملي الصدق المحكي للاختبار وفقاً للنظريتين لصالح الصدق المحكي للاختبار وفقاً لنظرية استجابة الفقرة، إذ بلغت قيمة t (٢,٠١٦).

للإجابة عن سؤال الدراسة الرابع الذي نص على: "ما مدى التوافق في تدريج قدرة المفحوصين على الاختبار في مبحث العلوم للصف السادس الأساسي وفقاً للنظرية (التقليدية، استجابة الفقرة ذات النموذج الثنائي المعلمة)؟"؛ قام الباحثان بإيجاد القيمة التدريجية لأداء المفحوصين وفقاً للنظرية (التقليدية، استجابة الفقرة ذات النموذج الثنائي المعلمة)، ثم إيجاد العلاقة الارتباطية بينهما كمؤشر على مدى التوافق بينهما، وذلك كما في الجدول (٢١).

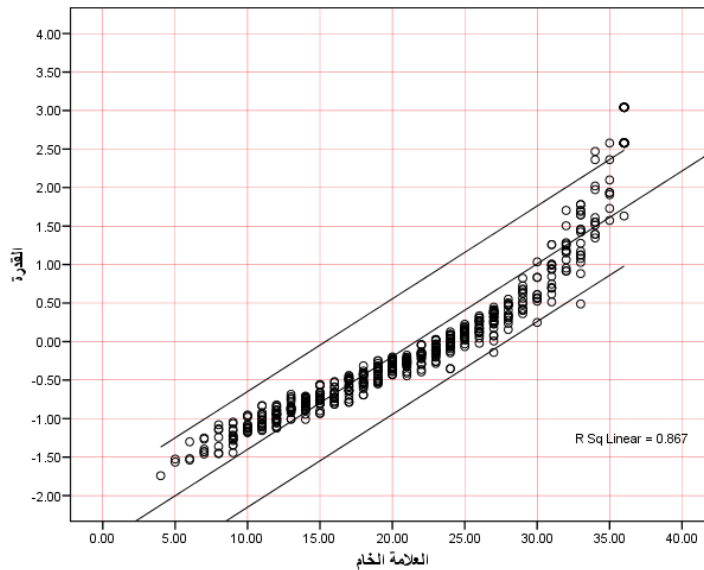
جدول (٢١): معاملات الارتباط بين القدرة من جهة وبين العلامة الخام والعلامة المحولة معيارياً من جهة أخرى

العلامة الارتباطية	القدرة
العلامة الخام المحولة معيارياً	٠,٨٥١
العلامة الخام	٠,٩٣١

يتضح من الجدول (٢١)، أن العلاقة الارتباطية بين القدرة من جهة وبين العلامة الخام من جهة أخرى؛ قد كانت طردية وذات دلالة إحصائية عند مستوى الدلالة ($\alpha=0,05$) مصنفة على أنها علاقة عالية جداً وفقاً لمعيار هنكل (Hinkle, Wiersma & Jurs, 1988).

كما يتضح من الجدول (٢١)، أن العلاقة الارتباطية بين القدرة من جهة وبين العلامة الخام المحولة معيارياً من جهة أخرى قد كانت طردية وذات دلالة إحصائية عند مستوى الدلالة ($\alpha=0,05$) مصنفة على أنها علاقة عالية.

كما تم تمثيل العلاقة الارتباطية بين القدرة وبين العلامة الخام تمثيلاً بيانياً، وذلك كما في الشكل (٣).



الشكل (٣): تمثيل بياني للعلاقة الارتباطية بين القدرة وبين العلامة الخام لتحصيل الطلبة على الاختبار وفقاً للنظرية التقليدية ونظرية استجابة الفقرة.

٧. مناقشة النتائج

يتناول هذا الجزء مناقشة نتائج الدراسة، والتوصيات التي ترتبت على تلك النتائج، حيث هدفت إلى توضيح مدى التوافق بين النموذج ثنائي المعلمة والنظرية التقليدية في بناء اختبار تحصيلي في مبحث العلوم العامة للصف السادس الأساسي.

أولاً: مناقشة النتائج المتعلقة بسؤال الدراسة الأول: "ما مدى التوافق في نتائج تصنيف الفقرات الخاصة باختبار مبحث العلوم للصف السادس الأساسي إلى (مطابقة أو غير مطابقة) وفقاً لـ: (نظرية استجابة الفقرة باستخدام النموذج الثنائي المعلمة، النظرية التقليدية)؟"

أظهرت النتائج الخاصة في تصنيف الفقرات للاختبار المعدّ في مبحث العلوم للصف السادس الأساسي إلى (مطابقة أو غير مطابقة) وفقاً لـ: (نظرية استجابة الفقرة باستخدام النموذج الثنائي المعلمة، النظرية التقليدية) وجود تشابه كبير بين النظريتين في تصنيف الفقرات، حيث بلغت نسبة الاتفاق بين نتائج المؤشرات التقليدية ونتائج مؤشرات النموذج ثنائي المعلمة في تصنيف الفقرات إلى (مطابقة أو غير مطابقة) (٩٢٪)، أي أن (٩٢٪) من فقرات الاختبار والبالغ عددها (٥٠ فقرة) اتفقت النظريتان في الحكم عليها، أما النسبة المتبقية (٨٪)؛ فقد اختلفت في الحكم عليها. كما كشفت النتائج أن عدد الفقرات غير المطابقة في النظريتين (١٢) فقرة أي ما نسبته (٢٤٪) من مجمل فقرات الاختبار ككل، بينما عدد الفقرات غير المطابقة للنظرية التقليدية ومطابقة في نظرية استجابة الفقرة (٢) فقرة أي ما نسبته (٤٪) من مجمل فقرات الاختبار ككل، وكذلك فإن عدد الفقرات غير المطابقة في نظرية استجابة الفقرة ومطابقة كلاسيكياً (٢) ما نسبته (٤٪) من مجمل فقرات الاختبار ككل، في حين كان عدد الفقرات المطابقة في النظريتين (٣٤) فقرة أي ما نسبته (٦٨٪) من مجمل فقرات الاختبار ككل. وهذا يؤكد وجود اتفاق بين المؤشرات التقليدية ونظرية الاستجابة الفقرة للنموذج الثنائي المعلمة.

وللكشف عما إذا كان توزع فقرات الاختبار الملاحظ وفقاً لحالة مطابقة الفقرة يختلف جوهرياً عن توزيعها المتوقع وفقاً لحالة مطابقة الفقرة، أظهرت نتائج اختبار χ^2 لحسن المطابقة؛ أن تكرار الفقرات المطابقة في النظريتين قد أسهم مساهمة إيجابية فاعلة في حساب χ^2 وبفارق جوهري عن التكرار المتوقع لها في حين أن تكرار الفقرات (غير المطابقة في النظريتين، و غير المطابقة كلاسيكياً)، قد أسهم مساهمة سلبية فاعلية في حساب χ^2 وبفارق جوهري عن التكرار المتوقع لها؛ بمعنى أن التكرار المشاهد لها كان أقل وبفارق جوهري عما هو متوقع له فيها. ويعزو الباحثان نتيجة عدم تأثر معالم الفقرات بخصائص العينة أو الأفراد الذين خضعوا للاختبار كون افتراضات هذه النظرية من السهل انتهاكها كما أشار علماء القياس من أمثال هامبلتون وجونز (Hambelton & Swaminathan, 1985)، علماً بأن النظرية التقليدية تعتمد في تحليل نتائج الاختبار على عينة الأفراد التي يجري عليها الاختبار، إذ تختلف صعوبة الفقرات باختلاف قدرات الطلبة الذين خضعوا

للاختبار. أما التمييز فيرتفع عندما يكون هناك تباين بين أفراد العينة الواحدة والعكس صحيح، في حين أن النظرية الحديثة تفترض أن تقدير معالم الفقرات مستقلاً عن خصائص عينة الأفراد المستخدمة في تقدير هذه المعالم. كما يعزو الباحثان هذه النتيجة ربما إلى التشابه الكبير بين المؤشرات الإحصائية للنظرية التقليدية ونظرية الاستجابة للفقرة للنموذج ثنائي المعلمة من الناحية النظرية رغم اختلافهما في آلية تقديرها، فالنظرية التقليدية تركز على الاختبار ككل نظرياً إلا أنها ضمناً تتعامل مع كل فقرة من فقرات الاختبار بشكل مستقل تماماً كما يحدث في نظرية الاستجابة للفقرة للنموذج ثنائي المعلمة. ومما تجدر الإشارة إليه أن هذه النتيجة اتفقت ونتائج دراسة كل من: (جمحاوي، ٢٠٠٠؛ Fan, 1998؛ Anil, 2008؛ Hernandez, 2009) من حيث المطابقة وعدد الفقرات، واختلفت مع نتائج دراسة كل من: (الحكماني، ٢٠٠٧؛ Raj, 2010) من حيث المطابقة وعدد الفقرات أيضاً.

ثانياً: مناقشة النتائج المتعلقة بسؤال الدراسة الثاني: "هل يوجد اختلاف بين قيمة معامل ثبات الاختبار وفقاً للنظريتين (النظرية التقليدية، ونظرية استجابة الفقرة للنموذج الثنائي المعلمة) عند مستوى الدلالة الإحصائية $\alpha = 0.05$ ؟"

أظهرت النتائج الخاصة بقيمة معامل ثبات الاختبار وفقاً للنظريتين (النظرية التقليدية، ونظرية استجابة الفقرة للنموذج الثنائي المعلمة) عند مستوى الدلالة الإحصائية $\alpha = 0.05$ ، أن قيمة معامل ثبات الاختبار وفق النظرية التقليدية بعد حذف الفقرات غير المطابقة قد ارتفعت لتصبح (٠,٩٠١)، بينما بلغت قيمة معامل الثبات الإمبريقي للاختبار بعد حذف الفقرات غير المطابقة (٠,٩٥١)؛ ولأن صورتَي الاختبار الناتجتين وفقاً للنظريتين (النظرية التقليدية، ونظرية استجابة الفقرة للنموذج الثنائي المعلمة) متساوية في عدد فقراتهما (أي في طولها)؛ فقد تم استخدام معادلة فلدت المعدلة Modified Feldet (F) للقياسات المترابطة (Feldet, 1969) وهو اختبار يستخدم للكشف عن دلالة الفرق بين معاملي ثبات اختبارين يقيسان المحتوى نفسه ومتساويين في عدد فقراتهما وقد طبقا على نفس العينة من المفحوصين. ومن نتائج الإحصائي (F) التي يعرضها الجدول (٢١)، تبين وجود فرق دال إحصائياً عند مستوى $\alpha = 0.05$ بين معاملي ثبات الاتساق الداخلي لنموذجي الاختبار، ولصالح نظرية الاستجابة للفقرة للنموذج الثنائي المعلمة إذ بلغت قيمة الدلالة الإحصائية (٠,٠٠٠) وعند مقارنة هذه القيمة مع قيمة $\alpha = 0.05$ نجد أن هذه القيمة أقل من قيمة $\alpha = 0.05$ ، وهذا مؤشر على رفض الفرضية الصفريّة وقبول الفرضية البديلة، وبذلك يمكن القول أن هناك فرقاً ذا دلالة إحصائية بين النظريتين (التقليدية، ونظرية الاستجابة للفقرة للنموذج الثنائي المعلمة) ولصالح نظرية استجابة الفقرة. ويعزو الباحثان هذه النتيجة ربما إلى خصائص أفراد العينة واعينة المفحوصين في تقدير معامل الثبات وفقاً للنظريتين (النظرية التقليدية، ونظرية استجابة الفقرة للنموذج الثنائي المعلمة)، فإذا أخذنا ثبات إحصائيات (الفقرة/ المفحوص) بعين الاعتبار فهي لا تتوفر وفق النظرية التقليدية لأن إحصائيات الفقرة تعتمد على عينة المفحوصين وخصائص المفحوص تعتمد على عينة الفقرات المستخدمة، وتتوفر وفق النظرية الحديثة لأن إحصائيات الفقرة مستقلة عن عينة المفحوصين وإحصائيات المفحوص مستقلة عن عينة الفقرات المستخدمة، إذا طابق النموذج المستخدم العينة (علام، ١٩٨٦؛ Fan, 1998؛ Hambelton & Swaminathan, 1985). وتتفق نتائج الدراسة الحالية مع دراسة فان (Fan, 1998) من حيث وجود فرق في معامل الثبات لصالح نظرية استجابة الفقرة على حساب النظرية التقليدية، إلا أنها اختلفت مع دراسة جمحاوي (٢٠٠٠) من حيث عدم وجود فرق جوهري بين معاملي الثبات للاختبار وفقاً للنظريتين (التقليدية واستجابة الفقرة).

ثالثاً: مناقشة النتائج المتعلقة بسؤال الدراسة الثالث: "هل يوجد اختلاف بين قيمة معامل الصدق المحكي للاختبار وفق النظريتين (النظرية التقليدية ونظرية استجابة الفقرة للنموذج الثنائي المعلمة) عند مستوى الدلالة الإحصائية $\alpha = 0.05$ ؟"

أظهرت النتائج الخاصة بقيمة معامل الصدق المحكي للاختبار وفق النظريتين (النظرية التقليدية ونظرية استجابة الفقرة للنموذج الثنائي المعلمة)، أن قيمة معامل الصدق المحكي للاختبار بعد حذف الفقرات غير المطابقة وفقاً لنظرية استجابة الفقرة بلغت (٠,٩١)، بينما بلغت قيمة معامل الصدق المحكي للاختبار بعد حذف الفقرات غير المطابقة وفقاً للنظرية التقليدية (٠,٧٦)؛ ولأن صورتَي الاختبار الناتجتين وفقاً للنظريتين (النظرية التقليدية، ونظرية استجابة الفقرة للنموذج الثنائي المعلمة) قد طبقتا على نفس العينة من المفحوصين؛ فقد تم استخدام معادلة اختبار (t) للقياسات المترابطة وهو اختبار يستخدم للكشف عن دلالة الفرق بين معاملي الارتباط (معاملي الصدق) لاختبارين يقيسان المحتوى نفسه كانا قد طبقا على نفس العينة من المفحوصين. ومن نتائج الإحصائي (t) التي يعرضها الجدول (٢٢) تبين وجود فرق دال إحصائياً عند مستوى $\alpha = 0.05$ بين معاملي الصدق المحكي للاختبارين وفق النظريتين (النظرية التقليدية ونظرية استجابة الفقرة للنموذج الثنائي المعلمة)، ولصالح نظرية الاستجابة للفقرة للنموذج الثنائي المعلمة إذ بلغت قيمة الدلالة

الإحصائية (٠,٠٤٤) وعند مقارنة هذه القيمة مع قيمة (٠,٠٥=α) نجد أن هذه القيمة أقل من قيمة (٠,٠٥=α)، وهذا مؤشر على وجود فرق ذي دلالة إحصائية بين النظريتين (التقليدية، ونظرية الاستجابة للفقرة للنموذج الثنائي المعلمة)؛ ولصالح نظرية استجابة الفقرة للنموذج الثنائي. ويعزو الباحثان هذه النتيجة ربما إلى الافتراضات الخاصة بالنظريتين، فقد أجمع علماء القياس من أمثال هامبلتون وجونز (Hambelton & Swaminathan, 1985) أن افتراضات النظرية التقليدية سهلة الانتهاك ويمكن لأي بيانات أن تحقق افتراضاتها، أما افتراضات النظرية الحديثة فهي قوية وليست من السهل لأي بيانات أن تحققها.

رابعاً: مناقشة النتائج المتعلقة بسؤال الدراسة الرابع: "ما مدى التوافق في تدرج قدرة المفحوصين على الاختبار في مبحث العلوم للصف السادس الأساسي وفقاً للنظرية (التقليدية، استجابة الفقرة ذات النموذج الثنائي المعلمة)؟"

أظهرت النتائج الخاصة بإيجاد العلاقة الارتباطية بين القدرة من جهة وبين العلامة الخام من جهة أخرى كمؤشر على مدى التوافق بين النظريتين (التقليدية، استجابة الفقرة ذات النموذج الثنائي المعلمة) أنها كانت طردية وذات دلالة إحصائية عند مستوى الدلالة (α=٠,٠٥) ومصنفة على أنها علاقة عالية جداً وفقاً لمعيار هنكل (Hinkle, Wiersma & Jurs, 1988). كما بينت النتائج أن العلاقة الارتباطية بين القدرة من جهة وبين العلامة الخام المحولة معيارياً من جهة أخرى كانت طردية وذات دلالة إحصائية عند مستوى الدلالة (α=٠,٠٥) ومصنفة على أنها علاقة عالية. ويعزو الباحثان هذه النتيجة ربما إلى أثر خصائص الأفراد على معالم الفقرات وأثر معالم الفقرات على خصائص الأفراد عند بناء الاختبار، فالاختبارات التي يتم بناؤها وفق أسس النظرية الحديثة والمفاهيم السيكمترية المرتبطة لا تختلف باختلاف خصائص أفراد العينة المستخدمة في المعالم بعكس النظرية التقليدية التي يكون فيها تقديرات القدرة مستقلة عن عينة الفقرات التي طبقت عليهم. كما أن الاختبارات المبنية على أساس النظرية التقليدية تكون فقراتها ملائمة للأفراد متوسطي القدرة أكثر من الأفراد ذوي القدرات المرتفعة والمتدنية؛ لذلك لا يمكن تقدير قدرات الأفراد ذوي القدرات المرتفعة والمتدنية (Croker & Algina, 1986)، في حين تفترض النظرية الحديثة أن مستوى قدرة الفرد على الإجابة يكون مستقلاً عن عينة الفقرات التي تطبق عليه بافتراض أن هناك عدداً كبيراً من فقرات الاختبار التي تقيس السمة، أي أن تقديرات الأفراد متحررة من خصائص الفقرات المستخدمة في تقدير القدرة. كما أنها توفر مؤشراً إحصائياً لتقدير درجة الدقة في قياس قدرة كل فرد، وهذا الإحصائي يختلف من فرد إلى آخر. وعليه فإن هذه النتيجة اتفقت مع دراسة كل من: (Pelton, 2002؛ جمحاوي، ٢٠٠٠؛ الحكامي، ٢٠٠٧؛ Raj, 2010)، واختلفت مع دراسة كورفيلي (Courville, 2004) من حيث وجود علاقة ارتباطية عالية جداً بين العلامة الخام لنتائج الطلبة على الاختبار وفقاً للنظرية التقليدية وبين نتائجهم وفقاً للقدرة في نظرية استجابة القدرة.

٨. التوصيات

وفي ضوء نتائج هذه الدراسة، يوصي الباحثان بما يلي:

١. إجراء المزيد من الدراسات المشابهة للدراسة الحالية في مواد دراسية مختلفة في ميادين التحصيل والقدرات العقلية المختلفة، تُحلل إحصائياً وفق النظرية الحديثة في القياس النفسي والتربوي.
٢. إجراء دراسة لسلوك النظريتين (التقليدية ونظرية الاستجابة للفقرة) للكشف عن مدى التوافق بين النظرية التقليدية ونظرية استجابة الفقرة باستخدام نماذج متعددة التدرج.
٣. إجراء دراسة لسلوك النظريتين (التقليدية ونظرية الاستجابة للفقرة) في ثبات وصدق إحصائيات الفقرة وضمن حجوم عينات مختلفة وفي مراحل عمرية مختلفة.
٤. اعتماد النظرية الحديثة في القياس النفسي— والتربوي عند تحليل نتائج الاختبارات التحصيلية المعدة من قبل المعلمين في المواد الدراسية المختلفة.

بيان تضارب المصالح

يقر جميع المؤلفين أنه ليس لديهم أي تضارب في المصالح.

- أبو غربية، إيمان. (٢٠٠٨). *القياس والتقويم التربوي* (ط.١). عمان، الأردن: دار البداية للنشر والتوزيع.
- التقي، احمد محمد. (٢٠٠٩). *النظرية الحديثة في القياس*. عمان: دار المسيرة للنشر والتوزيع.
- الجلبي، سوسن شاكر. (٢٠٠٥). *أساسيات بناء الاختبارات التربوية*. ط١، مؤسسة علاء الدين للطباعة.
- جمحاوي، إيناس محمود. (٢٠٠٠). *مقارنة خصائص الفقرات وفق النظرية التقليدية ونظرية استجابة الفقرة في مقياس للقدرة الرياضية* [رسالة ماجستير غير منشورة]. جامعة اليرموك، إربد، الأردن.
- الحكماني، رحاب بنت سعيد بن راشد. (٢٠٠٧). *مقارنة بين النظرية التقليدية للاختبار ونظرية الاستجابة للمفردة في تقديرات قدرات الأفراد ومدى استقرار مؤشرات المفردات الاختبارية* [رسالة ماجستير غير منشورة]. جامعة السلطان قابوس، مسقط، عُمان.
- الشرقاوي، أنور محمد وأمينه، كاظم والشيخ، سليمان الخضري. (١٩٩٧). *اتجاهات معاصرة في القياس والتقويم التربوي والنفسي*. القاهرة: الأنجلو المصرية.
- عدس، عبد الرحمن. (٢٠٠٢). *دليل المعلم في بناء الاختبارات التحصيلية*. عمان: دار الفكر للطباعة والنشر والتوزيع.
- علام، صلاح الدين محمود. (٢٠٠٧). *القياس والتقويم التربوي في العملية التدريسية* (ط.١). عمان، الأردن: دار المسيرة للنشر والتوزيع.
- علام، صلاح الدين. (١٩٩٠). *استخدام نموذج راش في بناء مقياس هدي المرجع للمعارف الأساسية في إعداد خطة البحوث النفسية والتربوية*. مجلة كلية التربية، جامعة الأزهر، (١٧)، ١٢٩-١٨٢.
- علام، صلاح الدين محمود. (١٩٨٧). *دراسة موازنة ناقدة لنماذج السمات الكامنة، والنماذج التقليدية في القياس النفسي والتربوي*. المجلة العربية للعلوم الإنسانية، ٧ (٢٧)، جامعة الكويت، ٤٤-١٨.
- علام، صلاح الدين. (١٩٨٦). *تطورات معاصرة في القياس النفسي والتربوي*. جامعة الكويت: إدارة التأليف والترجمة والنشر.
- علام، صلاح الدين. (١٩٨٥). *تحليل بيانات الاختبار العقلية باستخدام نموذج راش اللوغاريتمي الاحتمالي "دراسة تجريبية"*. المجلة العربية للعلوم الإنسانية، ٥ (١٧)، جامعة الكويت، ١٢٢-١٠٠.
- علام، صلاح الدين. (١٩٨٢). *استخدام نماذج السمات الكامنة في بناء بنوك الأسئلة*. صحيفة التربية، (٢)، ٤٤-٣٨.
- عمر، محمود أحمد وفخرو، حصة عبد الرحمن والسبيعي، تركي وتركي، آمنه عبد الله. (٢٠١٠). *القياس النفسي والتربوي*. ط ١، عمان، الأردن: دار المسيرة للنشر والتوزيع والطباعة.
- عودة، أحمد سليمان. (٢٠١٠). *القياس والتقويم في العملية التدريسية* (ط.٤). الأردن: دار الأمل للنشر والتوزيع.
- عودة، أحمد سليمان. (١٩٩٢). *مدى التوافق بين نموذج راش والمؤشرات التقليدية في اختيار فقرات مقياس اتجاه سباعي التدرج*. حولية كلية التربية ٧ (٨)، جامعة الإمارات، ١٧٩-١٥٣.
- عيد، غادة خالد. (٢٠٠٤). *الدرجة الحقيقية المقدرة باستخدام نظرية السمات الكامنة والنظرية التقليدية "دراسة سيكومترية"*. مجلة جامعة أم القرى للعلوم التربوية والاجتماعية، ٦ (٢) كلية التربية، جامعة الكويت، ٢٨٧-٢٢٨.
- كاظم، علي مهدي. (٢٠٠١). *القياس والتقويم في التعلم والتعليم* (ط.١). جامعة السلطان قابوس، مسقط: دار الكندي للنشر والتوزيع.
- مجيد، سوسن شاكر. (٢٠٠٧). *أسس بناء الاختبارات والمقاييس النفسية والتربوية* (ط.١). عمان، الأردن: ديونو للطباعة والنشر والتوزيع.
- النجار، نبيل جمعة. (٢٠١٠). *القياس والتقويم منظور تطبيقي مع تطبيقات برمجية SPSS*. عمان، الأردن: دار الحامد للنشر والتوزيع.

يوسف، عماد عبد المسيح. (١٩٩١). استخدام نموذج راش اللوغاريتمي- أحادي البارامتر في تحليل مفردات الاختبار المعرفية مرجعية المعيار ثنائية القطب. *مجلة البحث في التربية وعلم النفس*، جامعة المنيا، ٤(٤)، ص ٤٤٣-٤٧٥.

References

- Aiken, L. (1987). Testing with multiple-choice items. *Journal of Development in Education*, 20(4), 44-57.
- Allen, M. J., & Yen, W. M. (1979). *Introduction to Measurement Theory*. California: Cole Publishing Company.
- Anil, D. (2008). The prediction of item parameters based on classical test theory and latent trait theory. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi (H. U. Journal of Education)*, 34, 1-11.
- Brown, F. G. (1976). *Principles of education and psychological testing* (2nd ed.). New York: Holt, Rinehart and Winston.
- Courville, T. (2004). An empirical comparison of item response theory and classical test theory item/person statistics [Unpublished doctoral dissertation]. Texas A&M University, USA.
- Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. California: Cole Publishing Company.
- Fan, X. (1998). Item response theory and classical test theory: An empirical comparison of their item/person statistics. *Educational and Psychological Measurement*, 58(3), 357-381.
- Feldt, L. S. (1969). A test of the hypothesis that Cronbach's alpha or Kuder-Richardson coefficient twenty is the same for two tests. *Psychometrika*, 34(3), 363-373.
- Gay, L. (1980). *Educational evaluation and measurement*. Ohio: E. Merrill.
- Gronlund, N., & Linn, R. (1990). *Measurement and evaluation in teaching* (6th ed.). New York: Macmillan Publishing Company.
- Hambleton, R. K., & Jones, R. W. (1993). Comparison of classical test theory and item response theory and their applications to test development. *Educational Measurement: Issues and Practice*, 12(3), 38-47.
- Hambleton, R., & Swaminathan, H. (1985). *Item response theory: Principles and applications*. Boston: Kluwer-Nijhoff Publishing.
- Hattie, J. (1985). Methodology review: Assessing unidimensionality of tests and items. *Applied Psychological Measurement*, 9, 139-164.
- Hernandez, R. (2009). Comparison of the item discrimination and item difficulty of the Quick-Mental Aptitude Test using CTT and IRT methods. *The International Journal of Educational and Psychological Assessment*, 1(1), 12-18.
- Hinkle, D., Wiersma, W., & Jurs, S. (1988). *Applied statistics for the behavioral sciences*. Chicago: Rand McNally.
- Ho Kim, S., Cohen, A. S., & Lin, Y.-H. (2005). LDID: A computer program for local dependence indices for dichotomous items. National Taichung University, Taiwan, and The University of Georgia. Version 1.0. This software and manual may be reproduced for non-commercial use without the authors' permission.

-
- Hwang, D. Y. (2002). Classical test theory and item response theory: Analytical and empirical comparison. Paper presented at the Annual Meeting of the Southwest Educational Research Association, Austin, TX. (ERIC Document Reproduction Service No. ED 466779).
- 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). Johannes Gutenberg University Mainz, Germany.
- Lee, Y. (2004). Examining passage-related local item dependence (LID) measurement construct using Q3 statistics in an EFL reading comprehension test. *Language Testing*, 21(1), 74-100.
- Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. New York: Erlbaum Associates.
- Pelton, W. (2002). The accuracy of unidimensional measurement models in the presence of deviations from the underlying assumptions [Unpublished doctoral dissertation]. Brigham Young University, USA.
- Popham, W. (1980). *Criterion referenced measurement*. New Jersey: Prentice Hall, Inc.
- Raj-kamela. (2010). Investigating the invariance of person parameter estimates based on classical test and item response theories. *Department of Educational Foundations*, 2(2), 107-113.
- Warm, T. (1978). *A primer of item response theory*. Oklahoma City: U.S. Coast Guard Institute.
- Wright, B., Mead, R., & Bell, S. (1980). BICAL: Calibrating items with Rasch model. Chicago: University of Chicago, Research Memorandum No. 23.