بناء اختبار تكيُّفيّ محوسب باستخدام نظرية الاستجابة للفقرة في مادة الرياضيات للصف الرابع الأساسي بسلطنة عمان

Construction of a Computerized Adaptive Test Based on Item Response Theory in Mathematics for Fourth Grade Students in the Sultanate of Oman

إعداد

آمال بنت خلفان بن حميد الحسنية

Amal Khalfan Humaid Mohamed AL-Hasani

د. شريف عبد الرحمن السعودي

Dr. Sherif Abdelrahman Al-Saudi

جامعة الشرقية - سلطنة عمان

Doi: 10.21608/ejev.2025.447442

استلام البحث: ٩ / ٢ / ٢٠٢٥

قبول النشـــــر: ١٣ / ٤ / ٢٠٢٥

الحسنية، آمال بنت خلفان بن حميد والسعودي، شريف عبد الرحمن (٢٠٢٥ بناء اختبار تكيُّفيّ محوسب باستخدام نظرية الاستجابة للفقرة في مادة الرياضيات للصف الرابع الأساسي بسلطنة عمان . ***المجلة العربية للتربية النوعية***، المؤسسة العربية للتربية والعلوم والآداب، مصر، ٩(٣٨)، ٧٥ - ١٣٤.

https://ejev.journals.ekb.eg

بناء اختبار تكيُّفيّ محوسب باستخدام نظرية الاستجابة للفقرة في مادة الرياضيات للصف الرابع الأساسي بسلطنة عمان

المستخلص:

تهدف الدراسة إلى بناء اختبار تكيُّفيّ محوسب في مادة الرياضيات للصف الرابع الأساسي باستخدام نظرية الاستجابة للفقرة في سلطنة عمان خلال العام الدراسي (٢٠٢٤/٢٠٢٥م)، باستخدام المنهج الوصفي. تكوّنت عينة الدراسة النهائية لمعايرة بنك الفقرات من (٢٢٢٢) طالبًا وطالبة، من أربع محافظات تيسَّر العمل معها وهي: (مسقط، والداخلية، وجنوب الشرقية، وشمال الشرقية). وقد تبنّت الدراسة الحالية تصميم المجموعات غير المتكافئة ذات الفقرات المشتركة؛ فتكوّن الاختبار من (٩٦) فقرة واختيرت (١٢) لتكون فقرات رابطة مع مراعاة أن تكون متفاوتة في الصعوبة، وموزعة على كل المحتويات الدراسية مما يسهل المعادلة الأفقية، وتم تقسيم الفقرات المتبقية إلى أربعة نماذج، يحتوي كل نموذج على (٣٣) فقرة مع مراعاة تكافئ النماذج من حيث درجة الصعوبة وطبيعة المحتوى، وتكوّنت عينة التجريب للاختبار التكيُّفيّ المحوسب من (٥٨) طالبًا وطالبة من طلبة الصف الرابع بمحافظة شمال الشرقية؛ فتكوّن الاختبار في صورته النهائية بعد الحذف من (٩٤) فقرة، وتم اختيار العينة بالطريقة العشوائية العنقودية، وطُبق عليهم الاختبار التكيُّفيّ المحوسب وفق اشتراطات فنية محددة؛ فبدأ الاختبار بفقرات متوسطة الصعوبة عند مستوى قدره (٠.٦١)، وتم انتقاء الفقرات وفق طرق (فيشر) للمعلومات وطريقة الاحتمالية القصوى الموزونة تحديدًا، وقد انتهى الاختبار بمعيارين هما: تحقيق المستوى المطلوب للقدرة، والمستوى المنخفض للخطأ المعياري للقياس (SEM< 01>0). وأظهرت النتائج تقدير (متوسط) إلى (عالي الدقة) لقدرات الطلبة؛ إذ وفرت كمًّا متوسطًا نوعًا ما من المعلومات ومقدارًا قليل للخطأ. وفي ضوء نتائج الدراسة تم تقديم مجموعة من التوصيات أهمها: الاستفادة من إجراءات الدراسة الحالية وذلك بتبني الجهات التعليمية في سلطنة عمان الاختبارات التكيُّفيّة المحوسبة على نطاق واسع، في تقييم الطلبة في مختلف المواد الدراسية والمراحل عمرية، وأوصي باستخدام الاختبار التكيُّفيّ المُعدّ في هذه الدراسة من قِبَل سائر المعنيين في وزارة التربية والتعليم، وإجراء بحوث مستقبلية حول أثر هذه الاختبارات في تحسين التحصيل الدراسي، وتقليل القلق المرتبط بالاختبارات التقليدية.

الكلمات المفتاحية: اختبار تكيُّفي محوسب، نظرية الاستجابة للفقرة، الاختبارات التحصيلية، رياضيات.

ISSN: 2537-0448 eISSN: 2537-0456

Abstract:

The study aims to investigate the effectiveness of constructing a computerized adaptive test in mathematics for the fourth grade using item response theory in the Sultanate of Oman during the academic year 2024/2025 using the descriptive approach. The final study sample for calibrating the item bank consisted of (2222) male and female students from four available governorates of the Sultanate (Muscat, Ad Dakhiliyah, South Al Sharqiyah and North Al Sharqiyah). The current study adopted the design of unequal groups with common items, where the test consisted of (96) items.12 items were chosen as linking items , taking into account that they should be of varying difficulty and distributed in all academic contents, which facilitates horizontal equivalence. Remaining items were divided into four models, each model include (33) items, taking into account the equivalence of the models in terms of degree of difficulty and nature of the content. Initial experimental sample for the computerized adaptive test consisted of (58) male and female students from the fourth grade in North Al Sharqiyah Governorate, where the test was in its final form after deleting (94) items. The sample was selected using the cluster random method.Computrized adaptive test was applied to participants according to specific technical criteria. The test began with items of medium difficulty at an ability level of (0.61), and the items were selected according to Fisher's information methods, specifically the weighted maximum likelihood method was used. The test was completed using two criteria: achieving the required level of ability and a low standard error of measurement (SEM<0.01). Results showed a moderate to highly accurate estimation of the students' abilities, providing a moderate amount of information and a small amount of error. Future research is also needed to examine the impact of these

ISSN: 2537-0448 *eISSN: 2537-0456*

tests on improving academic achievement and reducing anxiety associated with traditional tests.

Keywords: Computerized adaptive testing, poverty response theory, Achievement Tests, Mathematics.

المقدمة:

يمثل التقويم أحد المكونات الأساسية للمنظومة التعليمية التي تتكون من مجموعة من العناصر المرتبطة بعلاقات تأثير وتأثر متبادلة، ويعتبر الخبراء التقويم نقطة البدء والاستمرار مع مراحل البناء، والتنفيذ، والمتابعة، والتطوير لتحقيق الغايات المنشودة. ويعرف التقويم بأنه عملية منهجية تتطلب جمع بيانات صادقة وموضوعية من مصادر مختلفة، باستخدام أدوات متنوعة في ضوء أهداف محددة؛ بغرض التوصل إلى تقديرات كمية وأدلة يُستند عليها في إصدار أحكام أو اتخاذ قرارات مناسبة تتعلق بالأفراد (بركات، ٢٠١٨).

وقد اهتم علماء القياس بابتكار طرق وأساليب تُوجِّه عمليات القياس والتقويم المعاصر، وركزوا جهودهم للتوصل إلى أعلى مستويات الموضوعية والدقة في القياس؛ بحيث تتحقق أدق علاقة بين أداة القياس والسمة المراد قياسها. ومن الأدوات المستخدمة في جمع البيانات الاختبارات، فالاختبار الجيد مقياس لا غنى عنه لمعرفة مدى نجاح المواقف التعليمية المختلفة (عليمات، ٢٠٢١). واهتم علماء القياس - أيضًا-بالتحقق من صدق وثبات الاختبارات والمقاييس النفسية والتربوية، إذ يبرز صدق علامات الاختبار عند اتخاذ قرارات مهمة على مستوى الفرد؛ مثل قبول فرد في برنامج أو استثناء آخر. (الشقصي وآخرون، ٢٠٢٠).

والاختبار الجيد هو الذي يتسم بدلالات الصدق والثبات الجيدة، اللذان يعتمدان على خصائص فقراته الاختبارية، لذلك كان لابد من الاهتمام بمعرفة الشروط الواجب وجودها في الفقرات الاختبارية وفقًا لمعايير الاختبار الجيد؛ إذ إنَّ جودة الاختبارات نابعة من حصيلة تخطيط سابق وسليم، ومهارة عالية من واضع فقرات الاختبار، وللتأكد من تحقيقها لمواصفات الفقرات الاختبارية الجيدة، ومعايرتها وفقا لنظريات القياس الحديثة (حماد، ٢٠١١). وعرفت هذه النظرية بمسميات أخرى مثل نظرية السمات الكامنة Latent Trait Theory، والنظرية المعاصرة في القياس Modern Test Theory، ونظرية الاستجابة للفقرة Item Response Theory. وأهم ما تتميز به افتراض أن تقدير معالم الأفراد مستقل عن عينة الفقرات، وأن تقدير معالم الفقرات مستقل عن عينة المفحوصين (البيايضة، ٢٠١١).

ومن الاتجاهات الحديثة في مجال التقويم التربوي ما يطلق عليه القياس التكيُّفيّ المحوسب computerized Adaptive Testing\CAT الذي بدأ مفهومه مع

ISSN: 2537-0448 eISSN: 2537-0456

محاولات بينيه عام (١٩٠٥)؛ باختيار عدد من الفقرات لتتناسب مع قدرة المفحوص أو استعداده أو المرحلة العمرية التي يمر بها، فعندما ينجح المفحوص في الإجابة عن مجموعة من الفقرات يتم تقديم فقرات أكثر صعوبة والعكس صحيح. ويعد تيرنبول Turnaball أول من استخدم لفظ (تكيفي)، وقد اقترح لورد عام (١٩٧٠) طريقة القياس التكيُّفيّ باستخدام نظرية الاستجابة للفقرة(IRT) (Chen, Wang Xin& Chang, 2017; Su, 2016; Van Der Linden& Glas, 2010; Wainer et al., 2000). وتعتبر نماذج نظرية الاستجابة للفقرة من النماذج السيكومترية المعاصرة التي ثبتت منفعتها في التغلب على كثير من مشكلات القياس النفسي والتربوي والقصور التي واجهت النموذج التقليدي أو ما تسمى بالنظرية الكلاسيكية Classical Test Theory، ويرى المختصون أن مبادئ النظرية الحديثة للقياس تختلف اختلافًا جوهريًا عن مبادئ النموذج التقليدي. فالنظرية المعاصرة في القياس تعد تطورًا حديثًا ومهمًا في القياس النفسي، إذ إنها تُمكن المختص من الإجابة عن أي فقرة في الاختبار (أبو شندي، ٢٠١١). وتفترض النظرية (IRT) إمكانية التنبؤ بأداء المفحوص أثناء الاستجابة لفقرات الاختبار من خلال تحديد خصائص المفحوصين، أو ما يعرف بسماتهم أو قدراتهم على افتراض وجود متصل للسمة، بحيث يمكن تقدير احتمال الإجابة الصحيحة للفرد عن فقرة ما إذا علم موقعه على متصل السمة. وأن العلاقة بين أداء الفرد على الفقرة وقدرته يمكن أن تحدد من خلال ما يسمى منحنى خصائص الفقرة Curve Characteristic Item (القضاة والشريفين، ٢٠٢٠).

مشكلة الدراسة وأسئلتها:

تُعتبر الاختبارات في العملية التعليمية الطريقة الأساسية لتقويم أداء الطلبة وتحصيلهم، فمن صفات الاختبار الجيد أن يمتلك خصائص سيكومترية جيدة من حيث الصدق والثبات والموضوعية. وفي الآونة الأخيرة حدث تطور كبير في علم القياس والتقويم؛ فقد تطور هذا العلم تطورًا واسعًا منذ ظهور النظرية الحديثة في القياس والمعروفة باسم نظرية الاستجابة للفقرة Item Response Theory: IRT، ومع هذا التطور تزايد الاهتمام في بناء وتطوير المقاييس والاختبارات العقلية والأدائية لقياس وتقييم قدرات الأفراد، وتحقيق معايير القياس العلمي الذي يضمن تقديرًا كميًا صادقًا وبدرجة مقبولة من الموضوعية والدقة للدرجات التي تم اعتمادها كتقدير كمي لهذه القدرات (الشهري، ٢٠٢٣). إذ توفر نظرية الاستجابة للفقرة إطارًا قويًا لتحليل عناصر الاختبار، بما يضمن قياس قدرات الطلاب بدقة، وتحسين تصميم الاختبارات المدرسية، وتستخدم هذه النظرية نماذج احتمالية لوصف التفاعل بين

ISSN: 2537-0448 eISSN: 2537-0456

المتقدمين للاختبار وفقراته، مع التركيز على خصائص الفقرات مثل الصعوبة والتمييز (Crocker & Algina, 2006).

كما أنه يمكن من خلال هذه النظرية تقييم جودة الفقرة بصورة فعالة في مواضيع مختلفة ومنها الرياضيات، مما يؤكد أنه يمكن تصميم الفقرات لتناسب قدرات الطلاب المحددة (Herosian et al., 2024; Arriza et al., 2023)، كما تساعد نظرية الاستجابة للفقرة تعزيزَ فهم خصائص فقرات الاختبار وأداء الطلاب، من خلال فهم كيفية استجابة الطلاب للاختبارات المدرسية بناءً على القدرات الكامنة وخصائص الفقرة (Ohiri, 2023).

وقد واكب ظهور نظرية الاستجابة للفقرة أساليب جديدة للاختبارات المعتمدة على الحاسب الآلي، التي احتلت مكان الورقة والقلم، وأدى ذلك لظهور مصطلح الاختبار التكيُّفيّ المحوسب Computerized Adaptive Testing الذي يتكيَّف مع قدرة الطالب؛ لأن المتقدمين للاختبار يجيبون عن مجموعات مختلفة من المفردات تطرح عليهم بناءً على مستوياتهم المعرفية، بحيث تؤثر إجاباتهم في مفردة ما على بقية المفردات التالية التي يختارها الحاسب لهم (حماد، ٢٠١١). وفي هذا السياق تسمح مرونة نظرية الاستجابة للفقرة بإجراء مثل هذا النوع من الاختبارات، فتُخصص الاختبارات بناءً على أداء الطالب الفردي مما يؤدي إلى تقييمات أكثر دقة (Crocker & Algina, 2006).

ونظرًا لعدم انتشار هذا النوع من الاختبارات، وأن هناك ندرة وخصوصًا في الدراسات العربية في سلطنة عمان -حسب اطلاع الباحثة-ومن هذا المنطلق لكون الباحثة تعمل في الميدان التربوي رأت ضرورة تبنّي أنماط تقويمية جديدة مع ظهور العديد من المستحدثات التكنولوجية في التعليم، وتجهيز البيئة المدرسية ببنية تحتية جيدة مثل الشبكات والأجهزة الحاسوبية في غرفة المصادر التي تساعد على تحسين تعلم الطلبة وإكسابهم المهارات المختلفة بدقة وسرعة وكفاءة؛ لذلك جاءت الدراسة لبناء اختبار تكيفي محوسب في الرياضيات للصف الرابع الأساسي بسلطنة عمان، حتى تكون مشجعة لبناء اختبارات تكيفية لجميع المراحل العمرية وفي مواضيع متعددة.

وتتمثل مشكلة الدراسة الحالية في السؤال الرئيس الآتي: **ما فاعلية بناء اختبار تكيُّفي محوسب في مادة الرياضيات للصف الرابع الأساسي باستخدام نظرية الاستجابة للفقرة؟**

ويتفرع عن تلك السؤال، الأسئلة الفرعية التالية:

١. ما دلالات الصدق البنائي والثبات لاختبار الرياضيات التكيُّفيّ المحوسب؟

ISSN: 2537-0448 eISSN: 2537-0456

٢. ما مدى تحقق بيانات اختبار الرياضيات التكيُّفيّ المحوسب لافتراضات نظرية الاستجابة للفقرة؟
٣. ما مدى مطابقة فقرات الاختبار التكيُّفيّ المحوسب للنموذج الثلاثي المعالم ثنائي التصحيح لمادة الرياضيات؟
٤. ما تقديرات معالم الصعوبة والتمييز والتخمين لفقرات الاختبار التكيُّفيّ المحوسب في مادة الرياضيات للصف الرابع الأساسي؟
٥. ما دالة المعلومات والخطأ المعياري لفقرات الاختبار التكيُّفيّ المحوسب لمادة الرياضيات للصف الرابع الأساسي؟
٦. ما فاعلية الاختبار التكيُّفيّ المحوسب لمادة الرياضيات للصف الرابع الأساسي في ضوء الاشتراطات الفنية؟

أهداف الدراسة

تهدف هذه الدراسة إلى ما يأتي:

١. تطوير الاختبار التكيُّفيّ المحوسب لمادة الرياضيات للصف الرابع الأساسي باستخدام نظرية الاستجابة للفقرة.
٢. التحقق من فاعلية الاختبار التكيُّفيّ المحوسب لمادة الرياضيات للصف الرابع الأساسي.

أهمية الدراسة

الأهمية النظرية

تكمن الأهمية النظرية في الآتي:

١. تعريف الباحثين بأهمية موضوع القياس التكيُّفيّ المحوسب.
٢. بناء اختبار تكيُّفي محوسب وفق نماذج نظرية الاستجابة للفقرة (IRT) يجعله أكثر دقة وموضوعية واستقلالية في القياس.
٣. بناء اختبار تكيُّفي محوسب لأي مادة دراسية يعدّ مساهمة نوعية تُضاف إلى مكتبة القياس النفسي والتربوي.
٤. تسليط الضوء على ضرورة الاهتمام ببناء اختبارات تكيُّفية محوسبة في مختلف المواد الدراسية وفق نظرية الاستجابة للفقرة.

الأهمية العملية

تكمن الأهمية العملية في:

١. إعداد وبناء اختبار تكيُّفي محوسب في مادة الرياضيات للصف الرابع الأساسي في سلطنة عمان واستخدامه في المؤسسات.

ISSN: 2537-0448 eISSN: 2537-0456

٢. تزويد الباحثين ببيانات عن واقع الاختبارات التكيُّفيّة المحوسبة والاستفادة من كيفية إجراء الدراسة والأداة المستخدمة فيها.
٣. توفير أداة قياس تتوافر فيها خصائص سيكومترية ملائمة ومقبولة للبيئة العمانية تمكّن الفاحص من استخدامها.
٤. مساعدة المعلم على إصدار أحكام موضوعية على مدى نجاح أساليب التدريس في تنظيم العملية التعليمية التعلمية من خلال تطبيق الاختبارات التكيُّفيّة المحوسبة المبنية وفق نظرية الاستجابة للفقرة.

مصطلحات الدراسة

تضمنت الدراسة العديد من المصطلحات، وهي:

الاختبار التكيُّفيّ المحوسب Computerized Adaptive Testing: "هو الاختبار الذي يُفصَل لكل مفحوص على حدة، وذلك بعرض الفقرات التي تتناسب مع مستوى قدرته، وبناءً على إجاباتهم عن مفردة ما أو مجموعة مفردات التي يختارها الحاسب الآلي لهم، مما يمكن من تقدير أدق للقدرة بأقل عدد من الفقرات" (عودة وعبيدات،٢٠١٣، ١٦٠٥).

وتعرفه الباحثة إجرائيًا: توفير تجمع فقرات لبناء اختبار تكيُّفي محوسب يُمكّن من قياس تحصيل طلبة الصف الرابع الأساسي بسلطنة عمان في الرياضيات باستخدام الحاسب الآلي وفق أحد النماذج اللوجستية المنبثقة من نظرية الاستجابة للفقرة بمستوى عالٍ من الدقة والموضوعية.

نظرية الاستجابة للفقرة Item Response Theory: "هي نموذج قياس يستخدم في الاختبار على نطاق واسع يفصل بين نمذجة معلمات العنصر ومعلمات الشخص، يستخدم التعبيرات اللوجستية لتأسيس العلاقة بين احتمال الحصول على إجابة صحيحة لعنصر ما والقدرة الكامنة للشخص ويتم تحديد دقة القياس من خلال كمية المعلومات في الاختبار" (Veldkamp,2005).

الإطار النظري والدراسات السابقة

المحور الأول: نظرية الاستجابة للفقرة

نظرية الاستجابة للفقرة Item Response Theory

تمثل هذه النظرية الاتجاه الحديث في القياس النفسي والتربوي، ويُشار إليها أحيانًا باسم نظرية السمات الكامنة Latent Traits Theory أو نظرية المنحنى المميز لخصائص الفقرة Item Characteristic Curve Theory. وقد ظهرت هذه النظرية لمعالجة أوجه القصور والتغلب على التحديات التي واجهت النظرية الكلاسيكية، والتي كانت النهج السائد في تطوير الاختبارات (علام، 2005).

ISSN: 2537-0448 eISSN: 2537-0456

تعاني النظرية الكلاسيكية من عدة أوجه قصور تمثلت في عدم استقلال قيم معالم فقرات الاختبار (الصعوبة والتمييز) عن خصائص عينة المفحوصين، كما أن تقدير درجات المفحوصين غير مستقل عن عينة فقرات الاختبار؛ أي أن تحديد موقع المفحوصين على متصل السمة يعتمد على عينة فقرات الاختبار الذي طبق عليهم. ومن جوانب القصور التي عجزت النظرية الكلاسيكية عن تقديم حلول لها الكشف عن التحيز في الفقرات والمعادلة بين علامات الاختبارات المختلفة، كما أن مفهوم الثبات في النظرية الكلاسيكية يُعرف بدلالة الاختبارات المتكافئة Parallel Forms والتي من الصعب الحصول عليها أو الاختبار وإعادة الاختبار، ولهذه الطريقة عيوب منها تأثر المفحوصين بعوامل التذكر، النسيان، الدافعية، والقلق، وقد يطور المفحوصين مهارات جديدة. وتفترض النظرية الكلاسكية أن الأخطاء المعيارية في القياس لها نفس التوزيع عند جميع مستويات القدرة وهذا يتنافى مع حقيقة أن تباين الخطأ يختلف باختلاف مستويات القدرة. إضافةً إلى ذلك فإن الاختبارات المبنيّة على أساس النظرية الكلاسيكية تكون فقراتها ملائمة للأفراد متوسطي القدرة أكثر من الأفراد ذوي القدرات المتدنية والمرتفعة (Hambleton & Swaminathan, 1985). ويرى ميرفي وديفيدشوفر (Murphy & Davidshofer, 1994) أن ظهور نظرية الاستجابة للفقرة ساعد في إيجاد الحلول للعديد من مشكلات القياس النفسي والتربوي من بناء بنوك الأسئلة وبناء الاختبارات التكيُّفيّة وتطويرها.

وتوضح نظرية الاستجابة للفقرة (IRT) العلاقة بين قدرة المفحوص واحتمال الاستجابة على الفقرة من خلال نماذج رياضية، وتصف نماذج (IRT) احتمال الإجابة عن فقرة بدلالة قدرة المفحوص ومعالم الفقرة، وتختلف هذه النماذج عن بعضها البعض من حيث كونها أحادية البعد أو متعددة الأبعاد، ومن حيث التدريج ظهرت النماذج متعددة التدريج Polytomous Models والنماذج ثنائية التدريج Dichotomous Models وأشهر النماذج أحادية البعد وثنائية التدريج كما جاءت في (Gruijter & Kamp, 2005; Hambleton & Swaminathan, 1985) وتؤكد كاظم (١٩٩٦) بأن هذه النماذج تسمح بإضافة أو حذف أو تعديل فقرات دون أن يتأثر الاختبار ككل. بينما يذكر علام (٢٠٠٥) أن نماذج هذه النظرية تفترض أن السمة المقاسة هي قدرة معينة أو خاصية من خصائص الفرد الذي يختبر بها بحيث إنه توجد علاقة منتظمة بين مستويات السمة المقاسة لدى أفراد مختلفيين وبين احتمالات الاستجابة الصحيحة لفقرات مختلفة، ويمكن رسم منحنى خصائص الفقرة بدلالة معلم واحد فقط وهو معلم صعوبة الفقرة، أو بدلالة معلمين هما صعوبة الفقرة ومعلم التمييز، أو بدلالة ثلاثة معالم هي الصعوبة والتمييز ومعلم التخمين وذلك بالاعتماد على مفاهيم نظرية الاستجابة للفقرة. وبناءً على هذا الاختلاف ظهرت

ISSN: 2537-0448 eISSN: 2537-0456

نماذج لوجستية متعددة ذات صور رياضية مختلفة. ويمكن توضيح هذه النماذج باستخدام الصور الرياضية لكل منها وعلى النحو التالي كما يذكرها هامبلتون وسوامينثان (Hambleton & Swaminathan, 1985):

١.النموذج اللوجستي الأحادي المعلم One- Parameter Logistic Model

يعدّ النموذج اللوجستي أحادي المعلمة من أبسط نماذج نظرية الاستجابة للفقرة، وهو النموذج الذي يكون بدلالة معلم صعوبة الفقرة فقط، ويفترض هذا النموذج أن جميع الفقرات متساوية في التمييز، وأن قيم التخمين تقترب من الصفر، وأن الفقرات تتباين فيما بينما من خلال معلم الصعوبة. ويُعتبر هذا النموذج من أكثر النماذج استخدامًا في بناء المقاييس وتحليل مفرداتها، ويُعزى ذلك إلى المزايا التي يتمتع بها. تعود جذور هذا النموذج إلى العالم الدنماركي جورج راش Rasch، حيث طوّر العالم بنجامين رايت Wright التطبيق العملي له. كان الهدف الأساسي لراش هو تحقيق الموضوعية في القياس، أي أن تكون الدرجة التي يحصل عليها الفرد تعكس مستوى قدرته الفعلية، بغض النظر عن طبيعة العينة التي أُجري عليها القياس. كما ينبغي أن تكون الدرجة متسقة مع مقياس التدريج الأصلي للفقرات المشمولة فيه، بحيث يحصل الفرد على الدرجة نفسها عند إجرائه لاختبارين يقيسان القدرة ذاتها أو السمة نفسها (علام، ٢٠٠٥). يعتمد نموذج راش الأحادي المعلمة على التقدير الثنائي للاستجابة، إذ يُفترض أن يجيب الفرد إجابة صحيحة عندما يحصل على الدرجة (١)، وأن يجيب إجابة خاطئة عندما يحصل على الدرجة (٠) (كاظم، ١٩٨٨ب).

٢.النموذج اللوجستي الثنائي المعلم Tow-Parameter Logistic Model

ظهر على يد عالم الإحصاء بيرنبوم Birnbaum عام ١٩٦٨، يفترض هذا النموذج وجود اختلاف بين الفقرات من حيث صعوبتها وتمييزها وعدم تأثر الإجابات بالتخمين، ويكون هذا النموذج بدلالة معلمين هما صعوبة الفقرة وتمييزها، وبذلك يفترض هذا النموذج أن قيم التخمين تقترب من الصفر وأن الفقرات تختلف فيما بينها من خلال معلمي الصعوبة والتمييز.

٣.النموذج اللوجستي الثلاثي المعلم Three-Parameter Logistic Model

يعتبر هذا النموذج الأكثر عمومية من النموذج الثنائي والأحادي المعلم؛ لأنه أضاف معلمًا ثالثًا وهو معلم التخمين وبذلك يكون هذا النموذج بدلالة ثلاث معالم، هي: الصعوبة والتمييز والتخمين الذي يحدد احتمال إجابة المفحوصين ممن يملكون قدرة متدنية للإجابة الصحيحة عن طريق التخمين، وبذلك فإن فقرات الاختبار تختلف عن بعضها البعض من خلال صعوبتها وتمييزها وقيم التخمين عليها، ويعتبر

ISSN: 2537-0448 eISSN: 2537-0456

من أكثر النماذج دقة في تقدير احتمال الاستجابة ومعلم القدرة، وهذا النموذج هو المستخدم في الدراسة الحالية.

افتراضات نظرية الاستجابة للفقرة

تعتمد نظرية الاستجابة للفقرة في أساسها على مجموعة من الافتراضات التي يجب تحققها لضمان موثوقية النتائج المستخلصة من البيانات. ومن أبرز هذه الفرضيات التي تم التحقق منها في هذه الدراسة هي، كما أشار إليها هامبلتون (Hambleton, 1987):

١. أحادية البعد Unidimensionality:

تفترض هذه الفرضية على وجود سمة واحدة تُفسر أداء الفرد أي المفحوص في الاختبار أو المقياس، بمعنى أن فقرات الاختبار متجانسة وتقيس قدرة واحدة، أي أن الدرجة على الاختبار تعكس السمة التي يقيسها الاختبار فقط، وقد أشار انستاسي ويوربينا (Anastasi & Urbina, 1997) إلى أن هذا الافتراض يكون محققًا إذا كان أداء الفرد يعتمد أساسًا على سمة رئيسية واحدة Predominant Trait، حتى وإن تأثرت نتائجه بعوامل ثانوية أخرى مثل الدافعية، القلق، السرعة وغيرها. أما بالنسبة للمؤشرات التي تدعم تحقق أحادية البعد، فقد ذكر هاتي (Hattie, 1985) أن من أبرزها:

- المؤشرات المستمدة من التحليل العاملي Factor Analysis وبخاصة تحليل المكونات الرئيسية:

وقد أشار ريكايس (Reckase, 1997) إلى مجموعة من المؤشرات المهمة للتحقق من هذا الافتراض، ومنها: نسبة التباين المفسَّر من قِبَل العامل الأول، والتي يجب أن تتجاوز20% لضمان وجود بُعد رئيسي واحد، ونسبة الفرق بين الجذر الكامن للعامل الأول والثاني، والتي يجب أن تكون أكبر من 2، معامل تشبع الفقرات بالعامل الأول، حيث يُفضّل أن يكون أعلى من ٠.٣٠ لضمان ارتباطها القوي بالبُعد الأساسي، وأيضًا التأكد من عدم وجود مشكلة التداخل الخطي المتعدد Multi-Collinearity بين المتغيرات عندما تكون قيمة محدد مصفوفة معامل الارتباط بين المتغيرات Determinant > 0.00001 وأيضًا كفاية حجم العينة من خلال إحصائي Kaiser-Meyer-Olkin وتكون أكبر من الحد الأدنى المقبول (0.5). كما يتم التحقق من مدى ملائمة مصفوفة معاملات الارتباط بين المتغيرات من خلال اختبار **Sphericity Bartlett's Test** (Field, 2009). وأيضًا حساب قيمة مربع كاي Chi-Square لتقييم مدى اعتدالية توزيع البيانات. ويُشترط أن تكون قيمة الاختبار دالة إحصائيًا عند مستوى دلالة أقل من ٠.٠٥ (نجيب والوفائي، ٢٠٠٦) لضمان صحة الفرضيات الأساسية للتحليل.

ISSN: 2537-0448 eISSN: 2537-0456

- نمط الاستجابة Indices Based On Answer Patterns: وذلك باستخدام أسلوب جتمان في الاستجابة على الفقرات، وذلك بترتيب الفقرات حسب صعوبتها، حيث يفترض جتمان أن السمة أحادية البعد إذا تحقق مفهوم التراكمية بحيث إجابة سؤال في مستوى ما تعتمد على قدرة المفحوص على إجابة سؤال في مستوى أدنى.
- مؤشرات تعتمد على الاتساق الداخلي: يعد معامل ارتباط الفقرة بالدرجة الكلية مؤشرًا على اتساق ما تقيسه هذه الفقرات وبالتالي مؤشرًا على أن الفقرات تقيس سمة واحدة ومن أمثلة ذلك معامل كرونباخ ألفا، أو معامل كودر-ريتشاردسون (٢٠) كحالة خاصة من كرونباخ α كذلك متوسط الارتباط الداخلية للفقرات Inter-Item Correlation.

٢. الاستقلال الموضعي Local Independence:

يشير هاميلتون وآخرون (Hambleton et al., 1991) إلى أن تحقيق افتراض الاستقلال الموضعي يعتمد على حساب الاحتمالية الخاصة بحصول فرد ما على نمط معين من الدرجات في الاختبار. فعلى سبيل المثال، إذا حصل فرد على درجات في أربع فقرات بالشكل (١, ١, ١, ٠)، فإن احتمال ذلك يُعبر عنه بالمعادلة: $P_1 \times P_2 \times P_3 \times Q_4$ حيث يُعرف هذا الأسلوب باسم التقدير بالأرجحية العظمى، والذي يفترض أن درجة الفرد تُحسب بناءً على ناتج ضرب احتمالات الإجابة الصحيحة في الفقرات التي أجاب عنها بشكل صحيح. في هذه الحالة: P_i يمثل احتمال الإجابة الصحيحة على الفقرة i، Q_i يمثل احتمال الإجابة الخاطئة على الفقرة i، ويُعبر عنه بالعلاقة $Q_i = 1 - P_i$ وبذلك، يعكس هذا الافتراض أن كل فقرة تقيس القدرة المستهدفة بشكل مستقل عن الفقرات الأخرى، ما لم يكن هناك عامل إضافي يربط بين الإجابات. ويمكن تقييم هذا الافتراض من خلال استخدام أسلوب التحليل العاملي.

٣. منحنى خصائص الفقرة Item Characteristic Curve-ICC:

يمثل منحنى خصائص الفقرة (ICC) أحد المفاهيم الأساسية في نظرية الاستجابة للفقرة، حيث يمثل العلاقة بين احتمال الإجابة الصحيحة ومستوى القدرة الكامنة للفرد (θ). يتخذ هذا المنحنى شكلًا مميزًا على هيئة الحرف (S) وفقًا لتطبيقات النظرية، مما يوضح كيف يؤدي ارتفاع قيمة القدرة الكامنة إلى زيادة احتمالية الإجابة الصحيحة. ويقترب المنحنى عند طرفه السفلي Lower Asymptote من الصفر، في حين يقترب عند طرفه العلوي Upper Asymptote من الواحد. أن المحور الأفقي يمثل قدرة المفحوص والمحور العمودي يمثل احتمالية الإجابة الصحيحة للمفحوص عن الفقرة، ويفترض أن المنحنى يزداد بشكل متزايد على طول متصل القدرة، مما يعني أن كلما زادت قدرة

ISSN: 2537-0448 eISSN: 2537-0456

المفحوص زاد احتمال اجابته الصحيحة عن الفقرة، وبالتالي يتوقع زيادة درجاته المتوقعة. (Crocker & Algina, 2006)

٤. التحرر من السرعة Speededness:

تفترض نماذج نظرية الاستجابة للفقرة أن عامل السرعة لا يؤثر في الإجابة عن فقرات الاختبار، أي أن المفحوصين الذين يفشلون في الإجابة عن فقرات الاختبار يرجع ذلك إلى محدودية قدرتهم وليس بسبب عدم تمكنهم، وهذا يعني أنه لو تم اختيار مفحوص بشكل عشوائي من مستوى قدرة معينة فإن أدائه على الفقرة لا يتأثر بالوقت المخصص أو سرعة إجابته على الفقرة. ونجد هذا العامل متضمن في أحادية البعد، حيث إن تأثير السرعة يدل على وجود قدرتين هما السرعة في الأداء والسمة المقاسة بواسطة محتوى الاختبار، كما يتم التحقق من هذا الافتراض من خلال عدة عوامل منها: مقارنة الأداء على الاختبار مع تحديد وقت الإجابة، وبدون تحديد وقت الإجابة (علام، ١٩٨٦).

أما فيما يتعلق **بخصائص الفقرات في ضوء نظرية الاستجابة للفقرة** والتي تعرف باسم المعالم، وفيما يلي عرض لهذه المعالم الثلاث (الصعوبة، والتمييز والتخمين):

الصعوبة (b): تعبّر عن سلوك الفقرة ضمن متصل القدرة، حيث يتم قياسها بوحدة مشتركة مع القدرة، وتُعرف بوحدة اللوجت Logit بمعنى آخر، تحدد الصعوبة موقع الفقرة على متصل القدرة، بحيث تكون الفقرات ذات الصعوبة المنخفضة أكثر ملائمة للأفراد ذوي القدرات المحدودة، في حين أن الفقرات الأكثر صعوبة تتطلب مستوى أعلى من القدرة لتحقيق إجابة صحيحة. وتمثل الصعوبة النقطة التي يكون عندها احتمال الإجابة الصحيحة ٥٠%، بشرط ثبات العوامل الأخرى أي عدم التخمين. وتمثل الصعوبة بيانيًا نقطة انقلاب المنحنى، حيث في الفقرات الصعبة تكون الإزاحة نحو اليمين، بينما الفقرات السهلة تكون الإزاحة نحو اليسار.

التمييز(a): يشير إلى مدى قدرة الفقرة على التفريق بين الأفراد ذوي المستويات المختلفة من القدرة، حيث يُظهر الفارق بين من يمتلكون قدرة عالية ومن لديهم قدرة منخفضة. رياضياً، يُعبَّر عن ذلك من خلال ميل المنحنى، فكلما زاد الميل، ارتفعت قدرة الفقرة على التمييز. وإذا كان الميل مساوية للصفر، فهذا يعني أن احتمال الإجابة الصحيحة لا يختلف بين الأفراد ذوي المستويات المتباينة من القدرة، مما يجعل الفقرة أقل قدرة على التمييز. وتُعرف الفقرات التي تمتلك ميلًا حادًا بأنها ذات تمييز مرتفع Steep Slope، حيث يؤدي التغير الطفيف في القدرة إلى تغير ملحوظ في احتمال الإجابة الصحيحة، بينما في حالة الميل المنخفض، يكون تأثير القدرة على احتمال الإجابة ضعيفًا.

ISSN: 2537-0448 eISSN: 2537-0456

التخمين (C): يمثل احتمال إجابة المفحوص بشكل صحيح على الفقرة نتيجة التخمين، وليس بناءً على مستوى قدرته الفعلية. ويُترجم ذلك بيانيًا كنقطة تقاطع المنحنى مع محور الصادات (الاحتمالية)، وأشار (Hattie, 1985) إلى أن قيمة معامل التخمين تتراوح بين صفر وواحد نظريًا، إلا أنه في الواقع العملي لوحظ أن يتخذ قيم أقل من (٠،٣).(Hambleton & Swaminathan, 1985) .

طرق المعايرة Calibration Methods

١. طريقة المعايرة المتزامنة Concurrent Calibration:

تعتمد هذه الطريقة على معايرة جميع بيانات النماذج الاختبارية الأربعة من مختلف العينات دفعة واحدة باستخدام برنامج الحاسوب، وتتميز هذه الطريقة بالكفاءة والفاعلية العالية، حيث تتيح الاستفادة القصوى من البيانات المتاحة وبالتالي نتائجها أكثر ثباتًا، وهذه الطريقة مناسبة أكثر مع النماذج أحادية البعد عندما يكون عدد الصفوف المراد ربطها قليل، وهي الطريقة المستخدمة في هذه الدراسة.

٢. طريقة المعايرة المنفصلة Separate Calibration:

في هذه الطريقة، تتم معايرة بيانات كل عينة أي كل صف بشكل مستقل عن الأخرى، مما يتطلب إجراء نوع من التحويل الخطي باستخدام الحاسوب عدة مرات. تُستخدم هذه الطريقة عادةً عندما تكون بيانات كل مجموعة منفصلة تمامًا عن الأخرى ولا يوجد تداخل كبير بينها. وعلى الرغم من أن هذه الطريقة قد تقلل من التأثيرات الناتجة عن تباين العينات، إلا أنها قد تكون أقل دقة مقارنةً بالمعايرة المتزامنة، خاصةً في الدراسات التي تعتمد على عينات مترابطة أو متداخلة. وتعدّ هذه الطريقة أكثر ملائمة للنماذج اللوجستية متعددة الأبعاد. (Jodoin, Keller & Swaminathan, 2003).

تقدير قيم معالم الفقرات والأفراد باستخدام نظرية الاستجابة للفقرة

١. طريقة بيز Bayesian method:

تُستخدم هذه الطريقة عندما تفشل طريقة الأرجحية العظمى (MLE) في العمل، أي عندما يجيب المفحوص عن جميع فقرات الاختبار إجابة صحيحة أو إجابة خاطئة. في هذه الحالة، تقدم الأرجحية العظمى القيمة الأولى للقدرة ٠ بـ (+∞)، والثانية بـ (-∞)، وهو أمر غير عملي. لا يوجد فرد لديه قدرة لانهائية في أي سمة، ولا يوجد فرد لديه سمة متكاملة بالكامل. سُميت هذه الطريقة بهذا الاسم لأنها تستخدم نظرية بيز في حساب الاحتمال المشروط، والتي تعبر عنها بالمعادلة الآتية: (Gruijter & Kamp, 2005)

حيث تعتبر القدرة (θ) في هذه الطريقة هي الحدث (B)، والاستجابات المشاهدة (u) هي الحدث (A).

ISSN: 2537-0448 eISSN: 2537-0456

٢. تقدير الأرجحية العظمى Maximum Likelihood Estimation-MLE:

تُعتبر هذه الطريقة من أكثر طرق التقدير شيوعاً، حيث يتم إيجاد المعالم من خلال إجراءات تعظيم الاحتمالية للمعالم المراد تقديرها. هذا الأسلوب شائع الاستخدام في الاختبارات التكوينية المحوسبة، ويعتمد على تقدير قيمة القدرة (θ) للمفحوص على نمط الاستجابة التي تأخذ (١,٠) ويرمز لنمط الاستجابة بالرمز u حيث إذا كانت الإجابة صحيحة u=1، وإذا كانت الإجابة خاطئة u=0. ومن المهم الإشارة إلى وجود نوعين من دالة الأرجحية العظمى: أحدهما دالة الأرجحية العظمى الشرطية (CML)، والآخر دالة الأرجحية العظمى الهامشية (MML). تم استخدام برنامج Bilog-mg لتحليل البيانات لتقدير معالم الفقرات والأفراد وفق طريقة الأرجحية العظمى الهامشية (MML). يشير (علام، ٢٠٠٥) إلى مميزات هذه الطريقة حيث يمكن استخدامها لتقدير معالم جميع النماذج أحادية البعد ومتعددة الأبعاد وتعتبر فعالة في كلا الحالتين سواء كانت فقرات الاختبار قليلة أم كثيرة، إمكانية الحصول على قيم تقديرية لمعالم المفحوصين الذين أجابوا إجابة صحيحة أو خاطئة على جميع الفقرات وتعطي تقديرات للعلامة الكلية وأيضًا القيم التقديرية للأخطاء المعيارية تمتاز بالدقة واستخدمت الدراسة الحالية هذه الطريقة.

المحور الثاني: الاختبارات التكيُّفيّة

مفهوم الاختبارات التكيُّفيّة

يعرّف أيضًا إمبرتسون وريز (2000) Embreton & Reise الاختبار التكيُّفيّ المحوسب على أنه أداة قياس تُطبَّق على المفحوص باستخدام الحاسب، وتُعد بمثابة بديل رقمي للاختبار التقليدي (الورقة والقلم)، ويعتمد هذا النوع من الاختبارات على بنك ضخم من الفقراتItem Pool، حيث يتطلب تحقيق دقة قياس عالية وأن يتضمن هذا البنك عددًا كافيًا من الفقرات التي تتمتع بدرجات صعوبة متفاوتة، موزّعة بشكل جيد على مقياس السمة المستهدفة. كما يتطلب وجود آليات محددة لاختيار الفقرات Item Selection Procedures، وطرق فعالة لتقدير القدرة Ability Estimation، بالإضافة إلى وضع قواعد واضحة لتحديد نقطة التوقف Stopping Rule.

أهمية الاختبارات التكيُّفيّة المحوسبة

تُساهم الاختبارات التكيُّفيّة المحوسبة في تحسين دقة وكفاءة التقييم من خلال تقديم معلومات أكثر تفصيلًا حول مستوى المتعلم، ما يجعلها أداة قوية لقياس القدرات بموضوعية عالية (Liu et al., 2024; Wulandari et al., 2020). كما تتيح إمكانية تقليل عدد الأسئلة المطلوبة للوصول إلى تقييم دقيق لمستوى الطالب، حيث تشير بعض الدراسات إلى أن هذه الاختبارات يمكن أن تخفض عدد المفردات بنسبة

ISSN: 2537-0448 eISSN: 2537-0456

تصل إلى ٥٠% مقارنة بالاختبارات الخطية التقليدية (Weiss & Şahin, 2024; Ayanwale & Ndlovu, 2024; Seo, 2017). ويُضاف إلى ذلك أن هذه الاختبارات تراعي الفروق الفردية، فتُصمم بحيث تناسب جميع المستويات، بما في ذلك الطلاب ذوي القدرات العالية أو المنخفضة، مما يجعلها أكثر تحفيزًا ويقلل من التوتر والقلق المصاحب للاختبارات التقليدية (Martin & Lazendic, 2018; Ebenbeck et al., 2024). وهذا بدوره يُسهم في تعزيز دافعية المتعلمين وتحقيق نتائج أكثر دقة وعدالة في التقييم الأكاديمي.

مراحل بناء الاختبارات التكيُّفيّة المحوسبة

المرحلة الأولى: دراسات الجدوى والتخطيط وقابلية التخطيط:

قبل اتخاذ القرار بشأن بناء وتطبيق اختبار تكيفي محوسب، من الضروري إجراء دراسة شاملة لتحديد جدوى هذا الاختبار وقابليته للتنفيذ. فعلى الرغم من المزايا العديدة التي توفرها هذه الاختبارات، قد يتردد بعض المختصين في تبنيها ما لم يكونوا على دراية كاملة بالمتطلبات اللازمة لتطويرها. لذا، يعتمد اتخاذ القرار على تحليل متوازن بين الفوائد المتوقعة والمخاطر المحتملة. ويُفضل في هذه المرحلة دراسة الجوانب التجارية والعملية المرتبطة بتطبيق الاختبارات التكيُّفيّة، مما يضمن اتخاذ قرار مستنير يستند إلى أسئلة محورية مثل تلك التي *طرحها* (Weiss & Shain, 2024; Thompson & Weiss, 2011)

- هل لدى المؤسسة الخبرة التقنية الكافية لإدارة اختبار تكيفي محوسب؟ وإذا لم يكن لديها؛ هل ستكون قادرة على توفير تكاليف خبير خارجي؟
- هل ستؤدي الاختبارات التكيُّفيّة المحوسبة إلى تقليل طول ومدة الاختبار وتحسين الدرجات؟
- هل تمتلك المؤسسة منصة لإدارة الاختبار التكيُّفيّ المحوسب، أم تحتاج لتطوير منصة خاصة بها؟

تُعد دراسات المحاكاة أداة أساسية للإجابة على العديد من التساؤلات المتعلقة بتطوير الاختبار التكيُّفيّ المحوسب، حيث تعتمد هذه الدراسات على عدة أساليب لمحاكاة أداء الاختبار، مما يسهم في تحديد مدى قابلية تطبيقه وملائمته لاحتياجات المؤسسة. ومن بين هذه الأساليب، تُعتبر محاكاة مونت كارلو Monte Carlo الأكثر شمولية، حيث تُستخدم عادة في المراحل الأولية من تطوير الاختبارات التكيُّفيّة المحوسبة. تتميز هذه الطريقة بعدم حاجتها إلى بيانات فعلية من أسئلة أو مفحوصين، ورغم ذلك، فإن توافر بيانات واقعية يعزز دقتها وفعاليتها، مما يجعل نتائجها أكثر فائدة عند تقييم مدى جاهزية الاختبار التكيُّفيّ المحوسب للتطبيق الفعلي

ISSN: 2537-0448 eISSN: 2537-0456

داخل المؤسسة، وبالتالي دعم اتخاذ القرار بشأن الاستثمار في تصميمه وتنفيذه (Thompson & Weiss, 2011).

المرحلة الثانية: تطوير بنك الأسئلة:

عند تصميم بنك الأسئلة للاختبارات التكيُّفيّة المحوسبة، من الضروري التركيز على تحقيق الاتساق بين المعلومات المقدمة وأهداف الاختبار. فإذا كان الغرض من الاختبار هو تصنيف أداء المختبرين وفقًا لمستويات محددة (مثل ناجح/راسب)، فإن الأسئلة يجب أن تتضمن مفردات متقاربة في مستوى الصعوبة، مما يسهم في تحقيق دقة التصنيف. أما إذا كان الهدف هو قياس مستوى أداء المختبر عبر عدة مستويات، فمن المهم تنويع مستويات الصعوبة، بحيث يشمل بنك الأسئلة مفردات ذات مستويات عالية ومتوسطة ومنخفضة من التعقيد. وتبرز أهمية هذا التنوع في ضمان دقة القياس وتغطية جميع المستويات المهارية للمختبرين. وعليه، فإن بناء بنك أسئلة متكامل يتطلب إعداد مجموعة واسعة من المفردات المصممة بعناية، بحيث تتيح تمييز الفروقات الفردية بدقة، مما يعزز موثوقية النتائج التي يقدمها الاختبار (قاسم، ٢٠٢١). ويؤكد فرحات (٢٠٢١) لضمان بناء بنك مفردات اختبارية شامل وفعال، يتطلب الأمر تخطيطًا دقيقًا يعتمد على تحديد خصائص كل مفردة، إلى جانب طبيعة المحتوى الذي تقيسه ومدى تغطيته للمهارات المستهدفة. فكلما كانت هذه الخصائص واضحة، زادت فعالية المفردات في تحقيق أهداف القياس. ولذلك، ينبغي أن يشتمل بنك المفردات على مجموعة واسعة ومتنوعة من المفردات التي تتفاوت في درجة صعوبتها، وفقًا لمجال القياس المستهدف. عند التخطيط لبناء بنك المفردات، يعد انشاء جدول لمواصفات الاختبار وخريطة اختبارية أمرًا ضروريًا. حيث يُحدد جدول المواصفات معايير وزن الاختبار، وأنواع الأسئلة المستخدمة فيه، بينما تمثل الخريطة الاختبارية مخططًا تفصيليًا يوضح عدد الأسئلة، وطبيعة المحتوى المعرفي. كما يمكن استبدال جدول مواصفات الاختبار بجدول مماثل خاص ببنك الأسئلة لتحقيق الهدف ذاته. يسهم جدول المواصفات في تقديم تصور شامل حول طبيعة الاختبار التكيُّفيّ، بما يشمل أهدافه، وأنواع الأسئلة المستخدمة، والمستويات الإدراكية التي يستهدفها. كما يساعد في ضبط مواقع الأسئلة ضمن الاختبار، مما يتيح تقديم هيكل منظم يضمن تغطية احتياجات المؤسسة التعليمية؛ بالإضافة إلى ذلك يتناول الجدول تفاصيل مثل آلية التصحيح، ووزن المحتوى، ونوع الاختبار التكيُّفيّ المستخدم، سواء كان أحادي الأبعاد أو متعدد الأبعاد (Weiss & sahin, 2024). ثم تأتي مرحلة كتابة المفردات ويجب مراعاة ثلاثة جوانب في عملية كتابة المفردات وهي: التعليمات التي ستقدم للمفحوصين، وصياغة جذر السؤال، وخيارات الإجابات. بعد ذلك يتم مراجعة المفردات من قبل

ISSN: 2537-0448 eISSN: 2537-0456

المختصين للتأكد من الصياغة اللغوية، وتمثيل المفردات للمحتوى، وتغطيتها لمستويات معرفية مختلفة. ويجب التركيز على الهدف من الاختبار ومدة تطبيقه (قاسم، ٢٠٢١؛ Weiss & Sahin, 2024).

المرحلة الثالثة: التطبيق القبلي وتحليل النتائج باستخدام نظرية الاستجابة للمفردة:

بعد الانتهاء من بناء مفردات الاختبار، من الضروري إجراء اختبار تجريبي لهذه المفردات على عينة استطلاعية، وذلك بهدف التحقق من مدى ملائمتها لمستوى الفئة المستهدفة. يعتمد هذا التحقق على تحليل المعايير الخاصة بكل مفردة وفقًا لنظرية الاستجابة للمفردة (قاسم، ٢٠٢١؛ Weiss & (Sahin, 2024.

يُحدد حجم العينة الاستطلاعية استنادًا إلى النموذج الإحصائي المستخدم في تحليل استجابات الأفراد للمفردات، حيث يتم توظيف نموذج نظرية الاستجابة للمفردة لتقدير خصائص هذه المفردات؛ وبعد استكمال عملية التحليل، يتم تقييم المعايير الإحصائية لمفردات الاختبار من خلال استخدام برامج مخصصة تستند إلى نظرية الاستجابة للمفردة، كما أوضح كل من ويس وشاهين(Weiss & Şahin, 2024).)، ويقترح قاسم (٢٠٢١) أن تكون قيمة معامل التمييز لا تقل عن ٠.٨، كما يُفضَّل أن تتراوح قيم معامل الصعوبة بين -٣ و٣، وألا تقل قيمة معامل التخمين عن ٠.٣. واستنادًا إلى البيانات الإحصائية المتعلقة بالمفردات، يجب حذف المفردات غير الملائمة أو مراجعتها وإعادة اختبارها (Weiss, 2011) Thompson & . علاوة على ذلك، يتطلب تحليل الأبعاد في هذه المرحلة ضمان أن يكون الاختبار أحادي البعد، وفقًا لنظرية الاستجابة المفردة، وكذلك لابد من التحقق من الاستقلال الموضعي للمفردات (Aybek & Demirtasli, 2017) . كما أنه من الضروري التأكد من أن مفردات الاختبار تتدرج وفق أحد نماذج نظرية الاستجابة المفردة، مع مراعاة تنوعها حسب مستوى القدرة (الصبيحي والأحمدي، ٢٠١١).

المرحلة الرابعة: تحديد المواصفات النهائية لتصميم الاختبار التكيُّفيّ المحوسب:

قبل البدء في تنفيذ ونشر الاختبار، من الضروري تحديد الخوارزميات التي سيعتمد عليها الاختبار. يتطلب ذلك وضع آلية دقيقة لاختيار المفردات المناسبة لكل مفحوص، وتحديد كيفية تقدير مستواه، واللحظة التي ينبغي فيها إنهاء الاختبار.

محكات الاختبار التكيُّفيّ المحوسب

بنك الفقراتItem Bank**:**

لتحقيق الهدف من الاختبار التكيُّفيّ المحوسب، ينبغي أن يحتوي الاختبار على عدد كبير من الفقرات ذات المعالم المقدّرة وفقًا لنماذج نظرية الاستجابة للمفردة **(IRT)**مما يضمن دقة القياس والمعايرة، بحيث تتناسب هذه الفقرات مع قدرات المفحوصين المختلفة.

ISSN: 2537-0448 eISSN: 2537-0456

يعد التنوع في مستوى صعوبة الفقرات وتوزيعها على طول مقياس السمة أمرًا ضروريًا لضمان التقدير الدقيق لقدرات المفحوصين، ويتم اختيار الفقرات من هذا البنك بناءً على توافقها مع نماذج نظرية الاستجابة للمفردة *(Boyd, 2003)*.

أساليب اختيار الفقرة Item Selection Procedures:

يُعد اختيار الفقرة من بنك الفقرات (item bank) أحد المراحل الأساسية في تصميم الاختبار، حيث يتم تقديمها للمفحوص وفقًا لاستراتيجيات محددة، تهدف إلى تحسين تقدير قدراته بشكل دقيق. وتختلف هذه الأساليب تبعًا لطبيعة الاختبار، ومن أبرزها: **الاختيار العشوائي Random Selection** حيث تختار الفقرات بطريقة غير موجهة، **الاختيار وفقًا لأقصى المعلومات Maximum Information Selection**حيث يتم اختيار الفقرة التي تقدم أقصى قدر من المعلومات حول قدرة المفحوص، **الاختيار البييزي Bayesian Item Selection** وهو نهج يعتمد على تقديرات احتمالية للقدرة، **أسلوب كولباك – ليبلر Kullback-Leibler** الذي يعتمد على قياس التباعد بين التوزيعات الاحتمالية لضمان دقة أعلى في انتقاء الفقرات.

طرق تقدير القدرة (Ability Estimstion):

فيما يخص البرامج المحوسبة للاختبارات التكيُّفيّة، أشار علام (٢٠٠٥) إلى أن هذه البرامج غالبًا ما تعتمد على أساليب الأرجحية العظمى Maximum Likelihood Methodsأو الأساليب المعتمدة على نظرية بييز Bayes Theorem. وتقوم الطريقة الأولى على تقدير القدرة من خلال استجابات المفحوص الأكثر احتمالاً، أما الطريقة الثانية فتفترض توزيعًا مبدئيًا للقدرة Prior Distributionثم تعمل على تعديل هذا التوزيع بناءً على استجابات المفحوص أثناء أداء الاختبار، لتقدير القدرة باستخدام التوزيع الاحتمالي اللاحق Posterior Distribution.

قاعدة التوقف أو إنهاء الاختبار Stopping Rule:

يشير لينكر Linacre (2000) إلى أن الوصول إلى مرحلة إنهاء الاختبار تعد من الخطوات الأكثر أهمية في الاختبارات التكيُّفيّة المحوسبة، لأنه عندما يكون الاختبار المحوسب قصيرًا جدًا فقد يؤدي إلى نتائج غير دقيقة في تقدير القدرة، أما في حال كان طويلًا جدًا، فإن ذلك يستغرق وقتًا أطول ويجهد المفحوص دون أن يضيف مستوى دقيقًا من النتائج.

طرق إنهاء الاختبار التكيُّفيّ

١.الطول الثابت للاختبار Fixed Length:

ISSN: 2537-0448 eISSN: 2537-0456

تعتمد هذه الطريقة على إعطاء كل مفحوص عددًا محددًا من الفقرات التي تتوافق مع قدرته. ومن خلال هذه الطريقة يمكن تحديد فيما إذا كنا نستطيع تقديم فقرة إضافية للمفحوص أم لا، إلا أنه لا يُفضل استخدامها كثيرًا نظرًا لأنها تفتقر إلى ما يميز الاختبارات التكيُّفيّة المحوسبة من مرونة وديناميكية في اختيار الفقرات خلال الاختبار (Boyd, 2003).

٢. الطول المتغير للاختبار Variable Length:

تعد هذه الطريقة من أكثر الأساليب شيوعًا في الاختبارات حيث تم استخدامها في الدراسة الحالية، حيث يستمر في تقديم الفقرات إلى أن يصبح الخطأ المعياري للتقدير أقل من قيمة معينة محددة مسبقًا (Boyd, 2003)، وبذلك يتم إنهاء الاختبار.

استراتيجيات الاختبارات التكيُّفيّة

١. **الاستراتيجية ثنائية المرحلة Two-Stage Strategy:**

تعرف بالتفرع الثابت Fixed Branch، حيث يتم تحديد مسار الاختبار وفقًا لنتيجة الفاحص في المرحلة الأولى.

٢. **الاستراتيجية الهرمية أو متعددة المراحل Multi-Stage Strategy:**

ويُطلق عليها أحيانًا التفرع المتغيرVariable Branch ، حيث يتم توجيه الفاحص إلى مسارات مختلفة بناءً على استجاباته.

٣. **الاستراتيجية المحوسبة Computerized Stage Strategy**:

وهي تعتمد على تقنيات الحوسبة في تقديم الفقرات وتحديد مسار الاختبار بصورة ديناميكية.

مميزات الاختبارات التكيُّفيّة

أشارت العديد من الدراسات إلى أن للاختبارات التكيُّفيّة مزايا متعددة تجعلها تتفوق على الاختبارات التقليدية. وقد تناول كل من كريست، لينكر، بويد (Crist, 1989; Linacre, 2000; Boyd, 2003) هذه المميزات كما يلي:

١. تتيح هذه الاختبارات كفاءة عالية من حيث الدقة، إذ يمكن من خلالها قياس القدرة بدقة ضمن نطاق واسع من مستويات المفحوصين.

٢. توفر مرونة كبيرة في إدارة الاختبارات، حيث تُمكّن من تطبيق نماذج متنوعة تتلائم مع مستوى قدرة كل مفحوص.

٣. لا تتطلب استخدام أوراق للإجابات أو للمراقبة، مما يقلل من الاعتماد على الوسائل الورقية.

٤. تعزز هذه الاختبارات سرية مضمونها، خاصة في الاختبارات الحساسة التي تتطلب الحد من احتمالية تسريب الأسئلة.

ISSN: 2537-0448 eISSN: 2537-0456

٥. تساهم هذه الاختبارات في تقليل فرص الغش أو التأثير على نتائج المفحوصين نتيجة تدخلات خارجية.
٦. تُعد هذه الاختبارات أسرع وأكثر كفاءة من غيرها، سواء من حيث التنفيذ أو التطوير أو تحليل النتائج.
٧. تُمكّن المفحوصين من الحصول على تقييم يتناسب مع قدراتهم الواقعية.
٨. تراعي الفروق الفردية، حيث لا يكون المفحوص مضطرًا لانتظار المفحوصين الآخرين للانتقال إلى الفقرة التالية.
٩. تصدر النتائج فورًا بعد انتهاء المفحوص من الاختبار، حيث يحصل على درجاته مباشرة عند الطلب.

الدراسات السابقة:

هدفت **دراسة خلف (٢٠٢٤)** إلى المقارنة في تقدير القدرة ودقتها لاختبار محبوك وفقا للاستراتيجيتين (ثنائية المرحلة والهرمية متعددة المراحل) وهي أحد استراتيجيات الاختبار التكيُّفيّ باستخدام نظرية القياس الحديثة، ولتحقيق هدف الدراسة اتبعت الباحثة خطوات علمية في إجراءات تحليل اختبار القدرة العامة من قبل إدارة التعليم والتدريب وتم نشره من المجلس الأسترالي للبحوث التعليمية للمرحلة الاعدادية المكون من ٦٠ فقرة، ولتأكد من صلاحية فقرات الاختبار للبيئة العراقية تم عرضها على الخبراء والمحكمين والبالغ عددهم ١٠ خبيرًا لمعرفة آراءهم حول صلاحية الفقرات والتعديلات التي أجريت عليها، واعتمدت الباحثة نسبة مئوية بلغت ١٠٠% وبذلك تم التأكد من صلاحية الفقرات، وقد طبق الاختبار على عينة عشوائية من طلبة المرحلة الإعدادية للمديريات العامة لتربية محافظة بغداد والبالغ عددهم ٤٠٠ طالب وطالبة، وتم إجراء التحليل الإحصائي باستخدام برنامج Bilog-mg3 وتبين أن جميع فقرات الاختبار مطابقة للنموذج الأحادي، ولكي تبدأ الباحثة بتصميم إستراتيجتين الاختبار المحبوك (ثنائية المرحلة والهرمية متعددة المراحل) قامت بتدريج فقرات الاختبار تنازليًا حسب معلم الصعوبة، وقد وزعت الاختبارات إلى خمسة مراحل بشكل هرمي حيث أن المرحلة الأولى تتضمن اختبار واحد، والمرحلة الثانية اختبارين وهكذا حتى المرحلة الخامسة خمسة اختبارات.

قام **جاسم (٢٠٢٣)** بدراسة هدفت إلى بناء اختبار تكيفي باستخدام استراتيجية الجزئية متعددة المراحل وفقًا لنظرية الاستجابة في البيئة العراقية، حيث تم تكييف اختبار الاستدلال ليمكل والذي تكون من ١٣٢ فقرة بصيغته النهائية، تضمنت المرحلة الأولى الاختبار الاستطلاعي صورة اختبارية مصغرة متوسطة الصعوبة، أما المرحلة الاختبارية الثانية فتضمنت ٢ صورة اختبارية مصغرة

ISSN: 2537-0448 eISSN: 2537-0456

(سهلة، صعبة) والمرحلة الاختبارية الثالثة تتضمن ٣ صور اختبارية مصغرة (سهلة، متوسطة الصعوبة، صعبة)، وقد تكونت عينة الدراسة من ١٠٠٠ طالب وطالبة من طلبة المرحلة الاعدادية، وعينة التدريج تكونت من ٢٠٠ طالب وطالبة، وقد أوضحت نتائج الدراسة إلى فاعلية الاختبار التكيُّفيّ باستخدام الاستراتيجية الجزئية متعددة المراحل وفقًا لنظرية الاستجابة للفقرة في تقدير القدرة الرياضية.

وقام **مرجان والبرصان (٢٠٢٣)** بدراسة هدفت إلى الكشف عن فاعلية القياس التكيُّفيّ متعدد المراحل باستخدام طريقة دلتا لتقدير الدرجات، وقد تم استخدام البيانات الناتجة عن تطبيق أوتيس لينون للقدرة العقلية العامة الصورة (J) على طلبة جامعة إب. وبلغ حجم العينة ١٦٠٠ طالبًا وطالبة، حيث تم تدريج البيانات باستخدام برنامج دلتا. وتكونت عينة القياس التكيُّفيّ من ١٣٠ مفحوصًا، استخدمت بيانات استجاباتهم في القياس التكيُّفيّ الذي تكون من أربع مراحل وتكونت كل مرحلة من خمس مفردات. ومن ثم تم حساب الدرجة لكل مفحوص في المراحل الأربع بشكل تراكمي وتم مقارنتها بدرجة المفحوص المقدرة من الاختبار الخطي بمفرداته الثمانيين. وأظهرت النتائج أن معامل الارتباط بين الدرجة المقدرة من الاختبار التكيُّفيّ والدرجة المقدرة من الاختبار الخطي بمفرداته الثمانين ٠.٨٧٨، ومتوسط الفروق في الدرجة المقدرة بمقدار ٥.٧٢٨ والخطأ المعياري في المتوسط ٠.٠٩٦، إضافة إلى أن الفرق بين الجذر التربيعي لمتوسط مربعات الخطأ في المتوسط كان ٠.٠٥٣، الأمر الذي يفيد بفاعلية القياس التكيُّفيّ المبنى باستخدام طريقة دلتا.

وهدفت **دراسة جرجس (٢٠٢١)** إلى معرفة أثر تصميم اختبار تكيفي إلكتروني بنائي برجع (تصحيحي، تفسيري) لطلاب الفرقة الثانية بكلية التربية - جامعة أسيوط وتم تحديد خطوات إنتاج الاختبار إلكتروني بنائي المرجع (تصحيحي، تفسيري)، وتم تطبيق تجربة البحث على مجموعتين تجريبيتين وهما المجموعة التجريبية الأولى طبق عليها الاختبار التكيُّفيّ الإلكتروني البنائي برجع تصحيحي، والمجموعة التجريبية الثانية طبق عليها الاختبار التكيُّفيّ الإلكتروني البنائي برجع تفسيري، وأظهرت النتائج وجود فروق دالة إحصائيًا عند مستوى (٠.٠١) بين متوسطي درجات المجموعتين التجريبيتين، كما يتضح أيضًا أن الاختبار التكيُّفيّ الإلكتروني البنائي برجع تفسيري له حجم أثر كبير على التقبل التكنولوجي.

التعقيب على الدراسات السابقة:

استنادا الى الدراسات السابقة نجد أن الميدان التربوي بسلطنة عمان بحاجة ماسة لبناء اختبار تكيفي محوسب في الرياضيات باستخدام نظرية الاستجابة للفقرة والتي يتم فيها تفصيل اختبارات خاصة لكل مفحوص اعتمادًا على قدرته عوضاً من تعرض المفحوص إلى جميع فقرات الاختبار، حيث أن كثرة الاختبارات الموجودة

ISSN: 2537-0448 eISSN: 2537-0456

والمستخدمة مع طلبة المدارس تطلب من المفحوص استخدام القلم والورقة من أجل الإجابة، وتفترض تعرض المفحوص إلى جميع فقرات الاختبار مما يؤدي إلى الشعور بالتعب والإرهاق بالإضافة إلى الاعتماد على النظرية الكلاسيكية في بناء هذه الاختبارات، ولذلك فإن هذه الدراسة تستخدم برمجية خاصة وهذه ميزة أخرى تضيفها الدراسة الحالية عن الدراسات السابقة من خلال التمكن من أداء الاختبار بشكل محوسب والذي ينتج عنه تقدير فعّال لقدرة المفحوص، وبالرجوع إلى الدراسات السابقة تجد الباحثة أن هناك ندرة في الدراسات العربية وخصوصًا في البيئة التعليمية العمانية التي اهتمت ببناء اختبارات تكيفية محوسبة في المواد الدراسية للطلبة سواء كان ذلك في الحلقة الأولى أو الحلقة الثانية؛ لذلك جاءت هذه الدراسة لبناء اختبار تكيفي محوسب في الرياضيات للصف الرابع الأساسي بسلطنة عمان لسد الفراغ الواضح في الاختبارات التكيُّنيّة المحوسبة ولتساعد المعلمين والمشرفين والمهتمين بقطاع التعليم على استخدامها.

منهجية الدراسة وإجراءاتها

منهج الدراسة

اعتمدت الدراسة الحالية على المنهج الوصفي نظرًا لملائمته لأهداف هذه الدراسة وطبيعتها.

مجتمع الدراسة

تمثل مجتمع الدراسة الحالية جميع طلبة الصف الرابع الأساسي بسلطنة عمان للعام الدراسي ٢٠٢٤/ ٢٠٢٥ البالغ عددهم ٣٩٩٤٢ إناثًا و٣٨٦٢٢ ذكورًا موزعين على ٣٠٣ مدرسة في إحدى عشرة محافظة (وزارة التربية والتعليم، ٢٠٢٥).

عينة الدراسة

في ضوء أهداف الدراسة وتعدد مراحلها، تم اختيار عينة عشوائية عنقودية ممثلة لمجتمع الدراسة حيث شملت على ثلاث عينات فرعية من طلبة الصف الرابع الأساسي في سلطنة عمان على النحو الآتي:

١. العينة الاستطلاعية: شملت ٣٥ طالب وطالبة من خارج عينة الدراسة للتجريب الأولي لبنك الفقرات والتحقق من الصدق والثبات (إيجاد الصعوبة والتمييز) لكل فقرة ومن ثم تم تقسيمها إلى أربعة نماذج متكافئة.

٢. عينة معايرة بنك الفقرات: شملت عينة المعايرة على ٢٢٢٢ طالب وطالبة من أربع محافظات السلطنة المتيسرة (مسقط، الداخلية، جنوب الشرقية وشمال الشرقية). واختيار مدرستين إلى ثلاث مدراس من هذه المحافظات بشكل عشوائي. ويوضح

ISSN: 2537-0448 eISSN: 2537-0456

الجدول (١) توزيع أفراد عينة الدراسة حسب المديريات التعليمية التابعة لكل محافظة.

الجدول ١ توزيع أفراد عينة الدراسة حسب المديريات التعليمية التابعة لكل محافظة لإحدى عشر مدرسة

المديرية	اسم المدرسة	العدد
مسقط	بساتين للتعليم الأساسي (١-٤)	٢٣٨
	المحامد للتعليم الأساسي (١-٤)	٢٧٣
	الفراسة للتعليم الأساسي (١-٤)	٣٦٣
الداخلية	الأجيال للتعليم الأساسي (١-٤)	٢٠٢
	إزكي للتعليم الأساسي (١-٤)	١٩٧
	المواهب للتعليم الأساسي (١-٤)	٢٠٣
جنوب الشرقية	الشراع للتعليم الأساسي (١-٤)	١٣٨
	العفية للتعليم الأساسي (١-٤)	٩٦
	البر للتعليم الأساسي (١-٤)	٢١١
شمال الشرقية	شموع الغد للتعليم الأساسي للصف الرابع	١٨٢
	الفضائل للتعليم الأساسي (١-٤)	١١٩
المجموع الكلي	١١	**٢٢٢٢**

٣. عينة التجريب للاختبار التكيُّفيّ المحوسب: تم جمع البيانات من ٥٨ طالب وطالبة من طلبة الصف الرابع الأساسي.

أدوات الدراسة:

لتحقيق أهداف الدراسة تم بناء اختبار تكيُّفي محوسب باستخدام نظرية الاستجابة للفقرة في مادة الرياضيات وفقًا للخطوات التالية:

أولاً: الاطلاع على الأدب النظري والدراسات السابقة العربية والأجنبية لاكتساب خلفية نظرية ومعرفية كاملة وشافية ودراستها حول بناء الاختبارات التكيُّفيّة المحوسبة باستخدام نظرية الاستجابة للفقرة في مادة الرياضيات.

ثانيًا: الاطلاع على عدد من نماذج سابقة للاختبارات التكيُّفيّة المحوسبة لمعرفة آلية تطبيق هذه الاختبارات ومساعدة الباحثة في صياغة وبناء الفقرات الخاصة بأداة الدراسة.

ثالثًا: اتباع الخطوات الثلاث العملية لبناء الاختبار التحصيلي وصياغة الفقرات في بنك الفقرات التي استعرضها بوتينج وآخرون (Boateng et al., 2018) والتي تشمل ما يلي:

١. بناء الفقرات من خلال تحديد الهدف من الاختبار، والتحقق من صدق المحتوى للمقياس عبر تحليل المحتوى وإعداد جدول المواصفات وكتابة الفقرات وفقًا للجدول.

ISSN: 2537-0448 eISSN: 2537-0456

وللتحقق من صدق المحتوى، قامت الباحثة بتحليل محتوى كتاب الرياضيات للصف الرابع الأساسي وذلك بعد الاطلاع على كتاب الطالب ودليل المعلم ووثيقة التقويم؛ حيث تناول الفصل الدراسي الأول الوحدات التالية (الأعداد ونظام الأعداد، الجمع والطرح(١)، الضرب والقسمة(١)، الكتلة، الوقت(١)، المساحة والمحيط(١)، الجداول والرسوم البيانية(١)، مخططات كارول وفن، نظام الأعداد وخصائص العدد، الجمع والطرح(٢)، الضرب والقسمة(٢). وقامت الباحثة بكتابة الأهداف التدريسية لهذه الوحدات وتعديل صياغة الأهداف المركبة وجعلها أهداف سلوكية قابلة للقياس والملاحظ.

٢. حساب الأوزان النسبية لمستويات الأهداف المعرفية.

في الخطوة الثانية، قامت الباحثة ببناء جدول المواصفات حيث تم الاعتماد على عدد الأهداف لحساب الوزن النسبي للوحدات (الموضوعات). كما تم حساب الأوزان النسبية لمستويات الأهداف المعرفية وفق تصنيف بلوم مع إجراء تعديل ليتوافق مع وثيقة التقويم الخاصة بالصف الرابع (وزارة التربية والتعليم، ٢٠٢٣) حيث تم تقليص الأهداف المعرفية إلى مستويين وهما: ٧٠% للمعرفة والفهم، و٣٠% لحل المشكلات .

٣. بناء فقرات الاختبار في صورته الأولية.

بعد الاطلاع على كتاب الطالب ودليل المعلم وبعد الرجوع للاختبارات الدولية تيمز، وإلى الدراسات السابقة التي تناولت موضوع الاختبارات التحصيلية في مادة الرياضيات للصف الرابع، كدراسة (أبو هادي وعالم، ٢٠٢٠؛ سلامة، ٢٠١٧؛ القضاة ٢٠١٠)، تم بناء الاختبار التحصيلي في مادة الرياضيات الذي يحتوي في صورته الأولية على ٩٦ فقرة من نوع الاختيار من متعدد نظرًا لسهولة تطبيقها وملائمتها للفئة العمرية وهم طلبة الصف الرابع، التي تتوافق مع وثيقة المجال الثاني لتقويم تعلم الطلبة للصفوف (١-٤) للعام الدراسي ٢٠٢٣-٢٠٢٤م، الصادرة من وزارة التربية والتعليم؛ ومن الأسباب الإضافية لاختيار هذا النمط من الفقرات هو قدرتها على قياس مدى واسع من القدرات والمستويات وفق تصنيف بلوم وتمتعها بدرجة عالية من الصدق والثبات، وخلوها من تأثير ذاتية المصحح (الحوسنية، ٢٠٢٣).

٤. تطوير الاختبار وبنك الفقرات من خلال التحقق من الصدق الظاهري، والتطبيق الأولي للفقرات على العينة الاستطلاعية، وحذف الفقرات غير مناسبة.

تم عرض فقرات الاختبار على مجموعة من المحكمين من المختصين والمشرفين التربويين والمعلمين الملحق (٣)، حيث تم الطلب منهم الحكم على هذه الفقرات من حيث وضوح الفقرات وصياغتها اللغوية ودقة الفقرة وخلوه من الأخطاء، وكذلك

ISSN: 2537-0448 eISSN: 2537-0456

تطابق الفقرة مع الهدف وتغطيتها للمحتوى وانتماء الأهداف المقاسة للمستوى المعرفي، كما هو موضح في الملحق (٤). كما أبدى المحكمين الاتفاق التام بنسبة ١٠٠% على وضوح الفقرات وانتمائها للوحدات التعليمية، في حين بلغت الفقرات التي تم تعديل صياغتها اللغوية عشر فقرات، كما أشار المحكمين إلى تغيير مستوى الهدف الذي تقيسه ثلاث فقرات معرفة وفهم وثلاث فقرات حل مشكلات. وثمان فقرات تم تعديل المموهات فيها. وفي ضوء ذلك تم تعديل بعض الفقرات وبناء النسخة النهائية للاختبار كما هو موضح في الملحق (٥) الذي تم تقسيمه إلى أربعة نماذج بمجمل (٣٣) فقرة لكل نسخة.

رابعًا: وبعد استكمال تسهيل مهمة الباحثة (الملحق٦)، تم تطبيق الفقرات على العينة الاستطلاعية لتحديد الخصائص السيكومترية المبدئية لبنك الفقرات. ثم تم استكمال الخطوات العملية المتبقية لإعداد الاختبار التحصيلي التكيُّفيّ المحوسب التي استعرضها ثرمبسون ووايس (٢٠١١) والمتمثلة فيما يلي:

٥. معايرة فقرات بنك الفقرات وفق النموذج ثلاثي المعالم ثنائي التصحيح. ويأتي اختيار هذا النموذج للأسباب التالية: عدد المعلمات/ البارامترات المراد تقديرها هي ثلاث المعالم، نوع الفقرات ومحتواها المتمثل في فقرات الاختيار من متعدد، ونمط الاستجابة أو آلية التصحيح الذي يعدّ ثنائي التصحيح. وجاء هذا الاختيار منسجمًا مع العديد من الأدبيات الإحصائية (De Ayala, 2009; Bond & Fox, 2015) والدراسات التطبيقية (Bulut & Kan, 2012; Garcia-Perez, 2014; Kalender, 2012; Thompson, 2016) وتم تطبيق هذا النموذج للتحقق من الافتراضات الإحصائية (أحادية البعد، والاستقلال الموضعي) نظرًا لعدم إمكانية تطبيق الاختبار المكون من (٩٦) فقرة على الطلبة بسبب الفئة العمرية وطبيعة المادة حيث بلغ عدد الوحدات في منهج الرياضيات للفصل الأول (١١) وحدة، فقد تم اختيار (١٢) كفقرة رابطة مع مراعاة أن تكون متنوعة ومتفاوتة من حيث الصعوبة والمحتوى أي تم اختيار من كل وحدة سؤال عدا الوحدة الدراسية الأكثر وزنًا تم اختيار سؤالين منها، مما يسهل المعادلة الأفقية (الكرامنة، ٢٠١٥). وتم تقسيم الفقرات المتبقية إلى أربعة نماذج، مع مراعاة تكافؤ النماذج من حيث درجة الصعوبة وطبيعة المحتوى كما هو موضح في جدول (٢). تم تطبيق النماذج الأربعة على العينة الثانية للدراسة (٢٢٢٢) طالب وطالبة؛ عند الدخول إلى الصف تم ترتيب الطلبة بحيث يعطى كل طالب نموذج مختلف عن الذي بجواره.

ISSN: 2537-0448 eISSN: 2537-0456

الجدول ٢ توزيع فقرات الاختبار على النماذج الأربعة

فقرات الاختبار	النموذج الأول	النموذج الثاني	النموذج الثالث	النموذج الرابع
الفقرات الرابطة ١٢ فقرة	• الوحدة الأولى: ٦* • الوحدة الثانية: ١٦ • الوحدة الثالثة: ٣٥* • الوحدة الرابعة: ٤٢* • الوحدة الخامسة: ٤٤*، ٤٨* • الوحدة السادسة: ٥٢ • الوحدة السابعة: ٦٠ • الوحدة الثامنة: ٦١ • الوحدة التاسعة: ٦٨ • الوحدة العاشرة: ٧٧، ٨٧			
الفقرات المتباينة ٢١ فقرة	٣، ٥*، ١٢، ١٨، ٢٠، *٢٦، ٣٢، ٣٣*، ٣٩، ٤٣، ٤٦، ٥٥، ٥٩*، ٦٢، ٧٢*، ٧٤*، ٧٨، ٧٩، ٨٠، ٩١، ٩٢*	٤، ٩، ١٠، ١٥، ١٩*، ٢١، ٣٧، ٣٨، ٤٠، ٤٧*، ٤٩*، ٥١، ٥٤، ٦٣*، ٦٦، ٦٧، ٧٦، ٨٢، ٨٣، ٨٨، ٩٠	٨، ١١، ١٤، ١٧*، ٢٢، ٢٣*، ٣١، ٣٤*، ٣٦، ٤٨*، ٥٣*، ٥٨، ٦٤، ٦٥، ٦٩، ٧٥، ٨١، ٨٤*، ٩٤*، ٩٥*، ٤١	١، ٧*، ١٣، ٢٤*، ٢٥*، ٢٧*، ٢٨، ٢٩، ٣٠، ٤٥، ٥٠، ٥٦، ٥٧، ٧٠، ٧١*، ٧٣*، ٨٥*، ٨٦*، ٨٩، ٩٣*، ٩٦*
العدد الإجمالي للفقرات	**٣٣**	**٣٣**	**٣٣**	**٣٣**

ملاحظة: تمثل فقرات حل المشكلات

خامسًا: وبعد إتمام إجراءات تصحيح الاختبار، تم استخدام البيانات للتحقق من الخصائص السيكومترية على عينة المعايرة باستخدام برنامج SPSS والتي شملت ما يلي:

- صدق الاختبار: التحقق من صدق الاختبار عن طريق صدق المحكمين كما تم استعراضه مسبقًا، والتحليل العاملي الاستكشافي باستخدام طريقة Principal axis factoring بطريقة التدوير المتعامد للنماذج الأربعة للاختبار.
- ثبات الاختبار: التحقق من ثبات الاتساق الداخلي لنماذج الاختبار الأربعة باستخدام ألفا كرونباخ نظرًا لمناسبة هذا المعامل لكافة أنواع الفقرات الاختبارية (عودة، ٢٠٠٥) بالإضافة إلى معامل ماكدونالدأوميجا.
- معايرة الفقرات واستخراج معاملات الصعوبة والتمييز والتخمين، وحساب دوال معلومات فقرات الاختبار ودالة معلومات الاختبار ككل مما يسهل الحصول على تدريج لفقرات الاختبار باستخراج قيم معامل الصعوبة والخطأ المعياري للتقدير لكل فقرة من فقرات النماذج بالإضافة للفقرات الرابطة.

ISSN: 2537-0448 eISSN: 2537-0456

سادسًا: تم تحديد الاشتراطات الفنية للاختبار التكيُّفيّ المحوسب النهائي.
تم تحديد الاشتراطات الفنية ونشر حيّ للاختبار التكيُّفيّ المحوسب (Publish Live CAT) من خلال منصة (FastTest v3 . 81. 3) كما قامت الباحثة بإعداد دليل للاختبار التكيُّفيّ المحوسب وفيه وصف موجز للاختبار، وطريقة التطبيق، وعدد فقرات الاختبار، وتعليمات للطلبة المختبرين، وتعليمات أخرى للمعلم المطبّق للاختبار.

نتائج الدراسة ومناقشتها

النتائج المتعلقة بالسؤال الأول:

نص السؤال الأول على: **ما دلالات الصدق والثبات باستخدام التحليل العاملي للنماذج الأربعة للاختبار التكيُّفيّ لمادة الرياضيات للصف الرابع؟**

للإجابة عن هذا السؤال، تم التحقق من الافتراضات الإحصائية ومنها التوزيع الطبيعي للبيانات ممثلة في الإحصاءات الوصفية (الوسط الحسابي، الانحراف المعياري، وقيم الالتواء والتفلطح) للفقرات في النماذج الأربعة للاختبار التحصيلي حيث جاء تقسيم فقرات الاختبار التحصيلي (ن= ٩٦ فقرة) إلى أربعة نماذج وذلك لتسهيل الاستجابة الصحيحة والدقيقة لفقرات الاختبار وتقليل الجهد والوقت على أفراد العينة، ومن ثم تم إجراء التحليل العاملي الاستكشافي. ويوضح الجدول (٣) هذه البيانات الوصفية.

الجدول ٣ البيانات الوصفية لفقرات النماذج الأربعة للاختبار التحصيلي

الفقرات	المتوسط الحسابي	الانحراف المعياري	الالتواء	التفلطح	المتوسط الحسابي	الانحراف المعياري	الالتواء	التفلطح
	النموذج الأول للاختبار				النموذج الثاني للاختبار			
الفقرة ١	٠.٧٩	٠.٣٨٠	-١.٧٢٢	٠.٩٧٠	٠.٦٨	٠.٤٦٦	-٠.٧٩١	-١.٣٨٠
الفقرة المشتركة ٢	٠.٤١	٠.٤٩١	٠.٣٨٥	-١.٨٥٣	٠.٤١	٠.٤٩١	٠.٣٨٥	-١.٨٥٣
الفقرة ٣	٠.٢٧	٠.٤٤٣	١.٠٥٧	-٠.٨٨٥	٠.٦٩	٠.٤٦٤	-٠.٨٠٨	-١.٣٥١
الفقرة ٤	٠.٦٠	٠.٤٨٩	-٠.٤٢٨	-١.٨٢٣	٠.٤٥	٠.٤٩٨	٠.١٩٨	-١.٩٦٨
الفقرة ٥	٠.٤٩	٠.٥٠٠	٠.٠٤٠	-١.٠٠٦	٠.٧٢	٠.٤٤٩	-٠.٩٨٥	-١.٠٣٤
الفقرة المشتركة ٦	٠.٥٦	٠.٤٩٦	-٠.٢٤٩	-١.٩٤٠	٠.٥٦	٠.٤٩٦	-٠.٢٤٩	-١.٩٤٠
الفقرة ٧	٠.٥٤	٠.٤٩٩	-٠.١٥٧	-١.٩٨٣	٠.٤٩	٠.٥٠٠	٠.٠٥٤	-٢.٠٠٤
الفقرة ٨	٠.٦٣	٠.٤٨٤	-٠.٥٣١	-١.٧٢٤	٠.٥٠	٠.٥٠٠	-٠.٠١٨	-٢.٠٠٧
الفقرة ٩	٠.٣٠	٠.٤٦١	٠.٨٥٠	-١.٢٨٢	٠.٥٨	٠.٤٩٣	-٠.٣٣٨	-١.٨٩٢
الفقرة ١٠	٠.٤٧	٠.٤٩٩	٠.١٣٥	-١.٩٨٩	٠.٤٠	٠.٤٩٠	٠.٤٠٦	-١.٨٤٢
الفقرة المشتركة ١١	٠.٣٤	٠.٤٧٥	٠.٦٥٦	-١.٥٧١	٠.٣٤	٠.٤٧٥	٠.٦٥٦	-١.٥٧١
الفقرة ١٢	٠.٤٨	٠.٥٠٠	٠.٠٩١	-١.٩٩٩	٠.٤٥	٠.٤٩٨	٠.٢٠٦	-١.٩٦٥

ISSN: 2537-0448 eISSN: 2537-0456

الفقرة المشتركة ١٣	٠.٣٨	٠.٤٨٦	٠.٤٨٢	-١.٧٧٠	٠.٣٨	٠.٤٨٦	٠.٤٨٢	-١.٧٧٠
الفقرة ١٤	٠.٧٧	٠.٤١٩	-١.٣٠٨	-٠.٢٩٠	٠.٦٦	٠.٤٧٥	-٠.٦٦١	-١.٥٦٩
الفقرة المشتركة ١٥	٠.٤٦	٠.٤٩٨	٠.١٦٣	-١.٩٧٥	٠.٤٦	٠.٤٩٨	٠.١٦٣	-١.٩٧٥
الفقرة المشتركة ١٦	٠.٥٣	٠.٤٩٩	-٠.١١٠	-١.٩٩٠	٠.٥٣	٠.٤٩٩	-٠.١١٠	-١.٩٩٠
الفقرة ١٧	٠.٦٩	٠.٤٦٥	-٠.٨٠٤	-١.٣٥٩	٠.٥٦	٠.٤٩٦	-٠.٢٥٧	-١.٩٤١
الفقرة المشتركة ١٨	٠.٤٧	٠.٤٩٩	٠.١٢٢	-١.٩٨٧	٠.٤٧	٠.٤٩٩	٠.١٢٢	-١.٩٨٧
الفقرة ١٩	٠.٦٩	٠.٤٦٥	-٠.٨٠٤	-١.٣٥٩	٠.٥١	٠.٥٠٠	-٠.٠٣٩	-٢.٠٠٦
الفقرة المشتركة ٢٠	٠.٤٩	٠.٥٠٠	٠.٠٥٨	-١.٩٩٨	٠.٤٩	٠.٥٠٠	٠.٠٥٨	-١.٩٩٨
الفقرة المشتركة ٢١	٠.٤٣	٠.٥٢٦	١.١٩٣	٢.٦٦٥	٠.٤٣	٠.٥٢٦	١.١٩٣	٢.٦٦٥
الفقرة المشتركة ٢٢	٠.٤١	٠.٤٩٢	٠.٣٦٤	-١.٨٦٩	٠.٤١	٠.٤٩٢	٠.٣٦٤	-١.٨٦٩
الفقرة ٢٣	٠.٣٤	٠.٤٧٥	٠.٦٧٢	-١.٥٥٤	٠.٥٠	٠.٥٠٠	٠.٠٠٠	-٢.٠٠٧
الفقرة المشتركة ٢٤	٠.٤٢	٠.٤٩٣	٠.٣٣٠	-١.٨٩٢	٠.٤٢	٠.٤٩٣	٠.٣٣٠	-١.٨٩٢
الفقرة ٢٥	٠.٥٩	٠.٤٩٣	-٠.٣٥٩	-١.٨٧٨	٠.٢٢	٠.٤١٦	١.٣٤٣	-٠.١٩٨
الفقرة ٢٦	٠.٣١	٠.٤٦٥	٠.٨٠٤	-١.٣٥٩	٠.٤٣	٠.٤٩٦	٠.٢٧٩	-١.٩٢٩
الفقرة ٢٧	٠.٣٧	٠.٤٨٣	٠.٥٣٩	-١.٧١٦	٠.٤٨	٠.٥٠٠	٠.٠٩٠	-١.٩٩٩
الفقرة ٢٨	٠.٣٣	٠.٤٧١	٠.٧٢٤	-١.٤٨٢	٠.٤٣	٠.٤٩٥	٠.٢٩٤	-١.٩٢١
الفقرة ٢٩	٠.٥٠	٠.٥٠٠	-٠.٠١١	-٢.٠٠٧	٠.٤٩	٠.٥٠٠	٠.٠٣٢	-٢.٠٠٦
الفقرة ٣٠	٠.٣٩	٠.٤٨٩	٠.٤٣٦	-١.٨١٧	٠.٢٠	٠.٣٩٩	١.٥١٥	٠.٢٩٧
الفقرة المشتركة ٣١	٠.٢٢	٠.٤١٥	١.٣٤٦	-٠.١٨٩	٠.٢٢	٠.٤١٥	١.٣٤٦	-٠.١٨٩
الفقرة ٣٢	٠.٤٢	٠.٤٩٤	٠.٣٢١	-١.٩٠٤	٠.٣٥	٠.٤٧٨	٠.٦٢٨	-١.٦١٢
الفقرة ٣٣	٠.٢٩	٠.٤٠٩	١.٤١١	-٠.٠١٠	٠.٣٨	٠.٤٨٥	٠.٥١٥	-١.٧٤١

ISSN: 2537-0448 eISSN: 2537-0456

الجدول ٤ البيانات الوصفية للنماذج الأربعة للاختبار التحصيلي

نموذج الاختبار	العدد	المتوسط الحسابي	الانحراف المعياري	الالتواء	التفلطح	ف	مستوى الدلالة
1	551	15.53	7.129	.439	-.765	١.١٣	٠.٣٤
2	559	15.10	6.558	.463	-.485		
3	594	15.86	7.232	.411	-.715		
4	518	15.57	7.192	.433	-.864		
المجموع الكلي	**2222**	**15.52**	**7.034**	**.442**	**-.706**		

أوضحت نتائج جدول (٣) تفاوت في متوسط الفقرات للنموذج الأول تراوحت بين (٠.٢٢-٠.٧٩) حيث حصلت الفقرة ١ على أعلى متوسط حسابي بمعدل ٠.٧٩، وحصلت الفقرة المشتركة ٣١ على أقل قيمة للمتوسط الحسابي بمعدل ٠.٢٢. كما بينت نتائج النموذج الثاني تراوحت بين (٠.٢٠-٠.٧٢) أن أكبر قيمة للمتوسط الحسابي للفقرة ٥ حيث بلغت ٠.٧٢، وبلغت أقل قيمة للمتوسط الحسابي ٠.٢٠ للفقرة ٣٠. وأوضحت نتائج النموذج الثالث متوسطات حسابية متنوعة تراوحت بين (٠.٢٢-٠.٧٩) حيث حصلت الفقرة ١٠ على أعلى متوسط حسابي بمعدل ٠.٧٩، وحصلت الفقرة ٣١ على أقل قيمة للمتوسط الحسابي بمعدل ٠.٢٢. وبيّنت نتائج النموذج الرابع بين (٠.٢٩-٠.٧٧) أن الفقرة ١٩ تمتعت بأعلى معدل للمتوسط الحسابي بلغت ٠.٧٧، والفقرة ٣٠ حصلت على أقل معدل للمتوسط الحسابي بلغت ٠.٢٩. وجاءت معاملات الالتواء ضمن المدى المقبول[+٣، -٣] ومعاملات التفلطح ضمن المجال المقبول [+٣، -٣] وفقًا لعدد من الأدبيات السابقة، مما أشار هذا إلى اعتدالية توزيع البيانات لفقرات كافة النماذج الأربعة (Hinkle et al., 2003). كما أوضحت نتائج الجدول (٤) إلى عدم وجود فروق جوهرية بين النماذج الاختبارية الأربعة في المتوسط الحسابي والانحراف المعياري وكذلك الالتواء والتفلطح ويشير هذا إلى وجود تكافؤ بين النماذج الاختبارية

بعد التحقق من توزيع البيانات، تم التحقق من الافتراضات الإحصائية للتحليل العاملي الاستكشافي لحساب الصدق العاملي للنماذج الأربعة للاختبار التحصيلي كما هو موضح في جدول(٥).

ISSN: 2537-0448 eISSN: 2537-0456

الجدول ٥ مؤشرات مناسبة العينة للتحليل العاملي الاستكشافي وفقًا لطريقة Principal axis factoring للنماذج الأربعة للاختبار التحصيل

الفقرات	النموذج الأول	النموذج الثاني	النموذج الثالث	النموذج الرابع
	التشبعات (Loading)			
١	٠.٣٣	٠.27	٠.38	٠.32
٢	٠.٣٢	٠.49	٠.48	٠.49
٣	٠.٣٧	٠.41	٠.32	٠.33
٤	٠.٣٧	٠.57	٠.50	٠.57
٥	٠.٥٥	٠.49	٠.55	٠.48
٦	٠.٥٢	٠.50	٠.54	٠.57
٧	٠.٢٩	٠.14	٠.37	٠.25
٨	٠.٤٥	٠.55	٠.61	٠.55
٩	٠.٤٧.	٠.33	٠.34	٠.45
١٠	٠.٤٦	٠.39	٠.31	٠.40
١١	٠.٥٠	٠.49	٠.46	٠.50
١٢	٠.٤٩	٠.42	٠.46	٠.38
١٣	٠.٤٣	٠.40	٠.33	٠.30
١٤	٠.٤٦	٠.44	٠.31	٠.47
١٥	٠.٤٧	٠.41	٠.32	٠.42
١٦	٠.٥٤	٠.41	٠.42	٠.59
١٧	٠.٣١	٠.35	٠.52	٠.36
١٨	٠.٥٣	٠.49	٠.26	٠.47
١٩	٠.٦٢	٠.48	٠.43	٠.48
٢٠	٠.٦١	٠.41	٠.45	٠.56
٢١	٠.٣٧	٠.35	٠.58	٠.46
٢٢	٠.٢٩	٠.55	٠.23	٠.51
٢٣	٠.٥٢	٠.50	٠.39	٠.43
٢٤	٠.٤٣	٠.24	٠.53	٠.50
٢٥	٠.٥١	٠.52	٠.58	٠.39
٢٦	٠.٥٤	-	٠.52	٠.32
٢٧	٠.٣٢	٠.47	٠.54	٠.40
٢٨	٠.٤٩	٠.19	٠.63	٠.52
٢٩	٠.٤٠	٠.42	٠.47	٠.53
٣٠	٠.٣٩	٠.47	٠.46	٠.48
٣١	٠.٤٩	٠.42	٠.45	٠.18
٣٢	٠.٤٩	٠.44	٠.49	٠.45

ISSN: 2537-0448 eISSN: 2537-0456

٣٣	٠.٤٦	٠.31	٠.42	٠.45
الجذر الكامن للعامل الأول	6.92	6.59	6.99	6.85
الجذر الكامن للعامل الثاني	٢.١٠	١.٧١	1.48	٨1.8
نسبة الجذر الكامن للعامل الأول إلى العامل الثاني	٣.٢٩	٣.٨٦	٤.٧٣	٣.٦٥
نسبة التباين المفسر في العامل الأول	%٢٠.٩٨	%١٩.٩٧	%٢١.٢٠	%٢٠.٧٧

تم التحقق من الافتراضات الإحصائية للتحليل العاملي الاستكشافي لحساب الصدق العاملي ومنها: تجنب الارتباطات العالية بين الفقرات Multicollinearity من خلال مراجعة مصفوفة الارتباطات بين الفقرات حيث لاحظت الباحثة خلو المصفوفة من أي معاملات تتجاوز ٠.٨٠ أو أقل عن ٠.٣٠ (Costello & Osborne, 2005). وكذلك تجنب الأحادية Singularity، فقد بلغ معامل التحديد Determinant في النموذج الأول ٠.٠٠١، وفي النموذج الثاني ٠.٠٠٥ وفي النموذج الثالث والربع ٠.٠٠٢ حيث تعتبر هذه القيم صغيرة جدًا وفي ذات الوقت لا تساوي صفر مما يعني أنها قيم مقبولة وقد تحقق شرط تجنب الأحادية (Beavers et al., 2013). وأمّا بخصوص قيم اختبار بارتلت فقد جاءت على نحوٍ دال إحصائيًا عند مستوى الدلالة ٠.٠٠١ في كافة النماذج الأربعة مما أكدّ أن مصفوفة الارتباطات ليست مشابهة لمصفوفة الوحدة (identity matrix; Pett et al., 2003). وجاءت قيم معاملات مايزر- ماير- أوكلن (KMO) ٠.٩٠٦، ٠.٨٨٥، ٠.٩٢٤، ٠.٩٠٠ للنماذج الأربعة للاختبار التحصيلي على التوالي. جاءت هذه القيم أعلى من ٠.٨٠ مما يشير إلى أن جميع القيم ممتازة ومناسبة وفقًا لعدد من الدراسات الإحصائية (Alhadabi et al., 2019; Pett et al., 2003). وأشارت هذه النتائج جميعها إلى تحقق افتراضات التحليل العاملي الإحصائية وصلاحية هذه البيانات للتحليل.

تم إجراء التحليل العاملي الاستكشافي وبالتحديد طريقة تحليل المحور الأساسي لاستجابات عينة الدراسة (ن=٢٢٢٢) على فقرات المقياس الاختبار التكيُّفيّ لمادة الرياضيات للصف الرابع الأساسي (٣٣ فقرة لكل نموذج) كما هو موضح في جدول (٦). ولتحديد عدد العوامل الضمنية الأمثل، تم الاعتماد على ثلاثة معايير وهي: معامل التشبع الأعلى من واحد لكايزر جوتمان (Guttman, 1954)، وشكل (Cattell, 1966) والتحليل المتزامن (Horn, 1965).

ISSN: 2537-0448 eISSN: 2537-0456

الجدول ٦ نتائج التحليل الاستكشافي وفقًا لطريقة Principal axis factoring للنماذج الأربعة للاختبار التحصيلي (ن= ٢٢٢٢) بطريقة التدوير المتعامد

الافتراض (المحك)	النموذج الأول	النموذج الثاني	النموذج الثالث	النموذج الرابع
معامل التحديد (Determinant)	٠.٠٠١	٠.٠٠٥	٠.٠٠٢	٠.٠٠٢
اختبار بارتلت (Bartlett's Test)	٠.٠٠١	٠.٠٠١	٠.٠٠١	٠.٠٠١
كاي سكوير (Chi-Square) /(df)	٣٧٥٤.٩١(٥٢٨)	٢٨٧٤.٨٨(٥٢٨)	٣٥٤٧.٦١(٥٢٨)	٣٢٨١.٦٧(٥٢٨)
معامل كايز-ماي-أوكلن (KMO)	٠.٩٠٦	٠.٨٨٥	٠.٩٢٤	٠.٩٠٠

أظهرت النتائج بناءً على معامل التشبع الأعلى من واحد لكايزر جوتمان Guttman وجود عامل ضمني واحد في النماذج الأربعة، وأشارت نتائج scree plot للنماذج الأربعة كما هو موضح في الشكل ٣ تركيبًا عامليًا مختلفًا يؤكد على وجود عامل ضمني واحد وفقًا لِ (Beavers et al., 2013)، حيث تعدّ أشكال scree plot في الدراسة الحالية من الأشكال الغامضة التي تظهر نقطة ذراع غير واضحة مما يستدعي إجراء التحليل الموازي Parallel analysis لحسم العدد الأدق للعوامل الضمنية في النماذج الأربعة، تم إجراء التحليل المتوازي وتوضح نتائج التحليل المتزامن لطريقة تحليل المكونات (PC) في شكل علامة(X)، وطريقة التحليل العاملي (FA) في شكل مثلثات. وقد تم الاقتصار في الدراسة الحالية على تفسير نتائج التحليل المتوازي للتحليل العاملي. وأشارت النتائج إلى وجود بعدًا ضمنيًا واحدًا في النماذج الأربعة، بلغت نسبة نسبة التباين المفسر في النموذج الأول ٢٠.٩٨%، و١٩.٩٧% من التباين في النموذج الثاني، و٢١.٢٠% في النموذج الثالث و٢٠.٧٧% من التباين في النموذج الرابع.

الجدول ٧ ثبات الاتساق الداخلي للنماذج الاختبارية الأربعة

الفقرات	النموذج الأول	النموذج الثاني	النموذج الثالث	النموذج الرابع
معامل ألفا كرونباخ للثبات (α)	٠.٨٨	٠.٨٥	٠.٨٨	٠.٨٨
معامل ماكدونالد أو ميجا للثبات (ω)	٠.٨٨	٠.٨٥	٠.٨٨	٠.٨٨

كما تم حساب ثبات الاتساق الداخلي باستخدام معاملي ألفا كرونباخ وماكدونالد أو ميجا كما هو موضح في الجدول ٧، حيث بلغ معامل الثبات بطريقة ألفا كرونباخ للنموذج الأول والثالث والرابع ٠.٨٨ والنموذج الثاني ٠.٨٥ كما أتت

ISSN: 2537-0448 eISSN: 2537-0456

قيم معاملات أوميجا مساوية لمعاملات ألفا كرونباخ وهذا يدل على تمتع الاختبار بثبات مرتفع.

مناقشة نتائج السؤال الأول:

بيّنت نتائج السؤال الأول تحقق جميع الافتراضات الإحصائية للتحليل العاملي الاستكشافي وكانت جميعها مؤشرات إيجابية، مما أكدّ على دقة النتائج والاستنتاجات. كما أظهرت نتائج التحليل العاملي الاستكشافي التركيبة العاملية أحادية البعد في كافة النماذج الأربعة للاختبار التحصيلي. اتفقت هذه النتائج مع دراسة خلف (٢٠٢٤) حققت أحادية البعد عند المقارنة لاختبار محبوك وفقا للإستراتيجيتين ثنائية المرحلة والهرمية متعددة المراحل باستخدام نظرية القياس الحديثة. وأظهرت نتائج هذا السؤال أيضًا معاملات الثبات بطريقتي ألفا كرونباخ وماكدونالد أو ميجا لكل من النماذج الأربعة للاختبار التحصيلي وكانت مرتفعة، حيث جاءت أعلى من ٠.٨٠ مشيرًا لوجود مستوى عالٍ من الاتساق الداخلي في الاختبار. اتفقت هذه النتائج مع دراسة مرجان والبرصان (٢٠٢٣).

النتائج المتعلقة بالسؤال الثاني:

نص السؤال الثاني على: **ما مدى تحقق الافتراضات الرئيسية لنظرية الاستجابة للمفردة لفقرات الاختبار التكيُّفيّ لمادة الرياضيات لطلبة الصف الرابع الأسي؟**

للإجابة عن هذا السؤال، تم التحقق من افتراضيين إحصائيين وهما: أحادية البعد والاستقلال الموضعي. للتحقق من أحادية البعد تم استخدام ثلاث طرق وهي التحليل العاملي الاستكشافي، والتحليل المتزامن كما هو موضح في نتائج السؤال الأول وتحليل راش للمكونات الأساسية في البواقي المعيارية Rasch Principal Component Analysis Of Standardizrd Residual Correlations. وتشير الأدبيات الإحصائية إلى وجود نقاط قطع متعددة للحكم على مقدار التباين غير المفسر. على سبيل المثال، يشير ليناكر (Linacre, 2002) أن قيمة التباين غير المفسر في القيمة المقارنة للعامل الأول (First Contrast) يجب أن تكون أقل من ٢.٠٠ للوصول إلى الاستنتاج الذي يؤكد أحادية البعد؛ بينما تقدم أدبيات أخرى قيمة أكثر تحررًا حيث تشير أن قيمة التباين غير المفسر في القيمة المقارنة للعامل الأول الأقل من ٣ مقبولة لتأكيد أحادية البعد (Linacre, 2024; Yang et al., 2023). ويوضح الجدول (٨) قيم التباين غير المفسر في النماذج الاختبارية الأربعة.

ISSN: 2537-0448 eISSN: 2537-0456

الجدول ٨ قيم التباين غير المفسر في تحليل راش للمكونات الأساسية في البواقي المعيارية للنماذج الأربعة

الفقرات	القيمة المقارنة الأولى	القيمة المقارنة الثانية	القيمة المقارنة الثالثة	القيمة المقارنة الرابعة	القيمة المقارنة الخامسة
النموذج الأول للاختبار	٢.٥٢	١.٧٥	١.٦٨	١.٤٢	١.٣٥
النموذج الثاني للاختبار	٢.١١	١.٦٨	١.٥٢	١.٣٨	١.٣٣
النموذج الثالث للاختبار	١.٨٣	١.٦٨	١.٤٩	١.٣٩	١.٣٦
النموذج الرابع للاختبار	٢.٢٩	١.٦٤	١.٦١	١.٤١	١.٣٥

أظهرت نتائج الجدول ٨ تحليل راش العاملي للبواقي المعيارية أن قيمة التباين في القيمة المقارنة الثانية والثالثة والرابعة والخامسة كانت أقل من ٢ في كافة النماذج الاختبارية الأربعة وقد أشارت هذه القيم إلى مقدار صغير جدًا للتباين غير المفسر مما تؤكد أحادية البعد في النماذج الأربعة وفقًا لقيمة القطع المتحفظة (Linacre, 2002). عدا القيمة المقارنة الأولى كانت أعلى من ٢ بقليل أي أقل من ٣ في النموذج الأول والثاني والرابع وما زالت قيمة مقبولة وفقًا لقيمة القطع المتحررة (Linacre, 2024; Yang et al., 2023). بينما القيمة المقارنة للعامل الأول في النموذج الثالث ١.٨٣ وهي أقل من ٢.

كما تم التحقق من الافتراض الإحصائي الثاني (الاستقلال الموضعي) من خلال مراجعة قيم معاملات الارتباطات بين البواقي. وتشير قيم المعاملات التي تزيد عن ٠.٣٠ فوق متوسط معاملات الارتباط بين البواقي على عدم استيفاء الاستقلال الموضعي (Christensen et al., 2017). وكان مدى معاملات الارتباط بين درجات البواقي للمفردات (٠.٢٩ إلى ٠.١٦-) للنموذج الأول، ومدى القيم (٠.١٧ إلى ٠.١٣-) للنموذج الثاني، ومدى القيم للنموذج الثالث (٠.١٦ إلى ٠.١٢-) ومدى القيم للنموذج الرابع (٠.٢٨ إلى ٠.١٤)؛ مما يشير إلى تحقق الاستقلال الموضعي.

مناقشة نتائج السؤال الثاني:

أظهرت نتائج هذا السؤال تحقق افتراضي نظرية الاستجابة للمفردة (أحادية البعد، والاستقلال الموضعي). فقد أوضحت نتائج راش للمكونات الأساسية في البواقي المعيارية تحقق أحادية البعد لأن التباين غير المفسر في القيمة المقارنة للعامل الأول كان صغير جدًا (أقل من ٢) في النموذج الثالث وفقًا لقيمة القطع المتحفظة، وفي النموذج الأول والثاني والرابع وفقًا لقيمة القطع المتحررة مما يدّعم ويعزز أحادية البعد في كافة النماذج الاختبارية الأربعة. اتفقت هذه الدراسة مع العديد

ISSN: 2537-0448 eISSN: 2537-0456

من الدراسات الإحصائية التي تؤكد أهمية التحقق من أحادية البعد قبل تقدير نماذج الاستجابة للمفردة (خلف، ٢٠٢٤)، واتبعت الدراسة الحالية إجراء إحصائي حديث إلى حدٍ ما عند التحقق من الافتراض الإحصائي الثاني وهو الاستقلال الموضعي وذلك من خلال مراجعة قيم معاملات الارتباطات بين البواقي. فقد جاءت قيم معاملات الارتباط بين البواقي في كافة النماذج الأربعة أقل من ٠.٢٠ فوق متوسط معاملات الارتباط بين البواقي مؤكدة تحقق الاستقلال الموضعي. وتماشيت هذه النتائج على عدد من الدراسات الإحصائية التي تبنت هذا الأسلوب (عمارة وكاظم، ٢٠٢٠) وتبني هذا السؤال على نتائج الدراسات السابقة التي اعتمدت على مربع كاي (x^2 LD) في التحقق من الاستقلال الموضعي (البرصان، ٢٠١٥؛ سلامة، ٢٠١٧).

النتائج المتعلقة بالسؤال الثالث:

نص السؤال الثالث على: **ما مدى مطابقة فقرات الاختبار التكيُّفيّ المحوسب للنموذج ثلاثي المعالم ثنائي التصحيح في مادة الرياضيات للصف الرابع الأساسي؟**

للإجابة عن هذا السؤال، تم مطابقة النموذج الثلاثي المعالم ثنائي التصحيح أحادي البعد لتقدير معاملات الصعوبة والتمييز والتخمين. ويوضح جدول (٩) مؤشرات مطابقة النموذج للاختبارات الأربعة.

الجدول ٩ قيم التباين غير المفسر في تحليل راش للمكونات الأساسية في البواقي المعيارية للنماذج الأربعة

مؤشرات المطابقة	النموذج الأول	النموذج الثاني	النموذج الثالث	النموذج الرابع
مؤشر $M_2(df)$	٤٨٨(٤٦٢)***	٧٢٤(٤٦٢)***	٦٢٧(٤٦٢)***	٩٢٢(٤٦٢)***
جذرمتوسط مربع الخطأ التقريبي (RMSEA)	٠.٠٤	٠.٠٣	٠.٠٣	٠.٠٤
مؤشر توكر لويس (TLI)	٠.٩٦	٠.٩٦	٠.٩٨	٠.٩٥
مؤشر المطابقة المقارن (CFI)	٠.٩٦	٠.٩٦	٠.٩٩	٠.٩٥
متوسط الجذر التربيعي للبواقي المعيارية (SRMSR)	٠.٠٥	٠.٠٥	٠.٠٤	٠.٠٥

أشارت نتائج مؤشر (M_2) لوجود مطابقة جيدة للنموذج لكونها دالة إحصائيًا (Immekus et al., 2019) كما أشارت نتائج الجذر التربيعي لمتوسط أخطاء التقدير (RMSEA وSRMSR) مطابقة جيدة للنموذج لأنها جاءت أقل من ٠.٠٥ لكافة النماذج الاختبارية الأربعة. وفقًا لعدد من الأدبيات الإحصائية (Schumacker & Lomax, 2016; Wang & Wang, 2012). وأوضحت نتائج مؤشرات توكر لويس (TLI)، والمطابقة المقارن(CFI) مطابقة جيدة للنموذج ثلاثي المعالم

ISSN: 2537-0448 eISSN: 2537-0456

ثنائي التصحيح نظرًا لأنها جاءت أعلى من ٠.٩٠ على نحوٍ مجمل. حيث أشارت هذه النتائج إلى مطابقة جيدة للنموذج الثلاثي المعالم ثنائي التصحيح أحادي البعد في جميع النماذج الاختبارية الأربعة لمادة الرياضيات.

مناقشة نتائج السؤال الثالث:

بينت نتائج السؤال الثالث وجود مطابقة جيدة للنموذج ثلاثي المعالم ثنائي التصحيح أحادي البعد وفقًا لمؤشر (M_2) فهو مؤشر حساس جدًا على عكس مربع كاي (X^2)، والجذر التربيعي لمتوسط الخطأ التقريبي (RMSEA)، وقيم القطع لمؤشر توكر لويس (TLI) ومؤشر المطابقة المقارن (CFI) وفقًا للمعايير التي استعرضتها الأدبيات الإحصائية (Schumacker & Lomax, 2016; Wang & Wang, 2012).

النتائج المتعلقة بالسؤال الرابع:

نص السؤال الرابع على: **" ما تقديرات معالم الصعوبة والتمييز والتخمين لفقرات الاختبار التكيُّفيّ المحوسب في مادة الرياضيات للصف الرابع الأساسي"؟**

للإجابة عن هذا السؤال، تم مطابقة النموذج ثلاثي المعالم ثنائي التصحيح أحادي البعد لتقدير معاملات الصعوبة، والتمييز، والتخمين والخطأ المعياري. ويوضح الجدول (١٠) معاملات النموذج ثلاثي المعالم ثنائي الاستجابة أحادي البعد في النموذجين الأول والثاني، بينما يوضح الجدول (١١) معاملات النموذج ثلاثي المعالم ثنائي الاستجابة أحادي البعد في النموذجين الثالث والرابع. حيث يتضمن كلا الجدولين قيم مؤشر *(S-X2)* وقيم الدلالة الخاصة بهذا المؤشر، ومعامل الصعوبة/ القدرة (bi)، ومعامل التمييز (ai)، ومعامل التخمين (ci) لتقييم مطابقة فقرات الاختبار.

وفقًا للعديد من الأدبيات الإحصائية (OrIando & Thissen, 2003; Toland, 2014) تدل قيم الدلالة الإحصائية *(p-value)* لمؤشر*(S-X2)* أعلى من ٠.٠١ على مطابقة جيدة للفقرات، وتعكس قيم الجذر التربيعي لمتوسط أخطاء التقدير (RMSEA) التي تكون أقل من ٠.٠٥ مطابقة جيدة للفقرات. وفيما يتعلق بمعامل الصعوبة/ القدرة (b_i) يشير بيكر (Baker, 2001) أن مدى القيم التي تتراوح بين +٢ و-٢ تعكس المدى المقبول لمعدل صعوبة الفقرات، أمّا المدى الذي يتراوح بين +٣ و -٣ محط تساؤل ونقد. فيما يتعلق بالتمييز(a_i) يقدم بيكر (Baker, 2001) تصنيفات متعددة على النحو التالي:

تدل القيم (a_i< ٠.٢٠) تمييز ضعيف جدًا، والقيم (٠.٢١< a_i <.٤٠) تمييز ضعيف، والقيم (.٤١< a_i <.٨٠) تمييز متوسط، والقيم (.٨١< a_i <١.٠٠) تمييز مرتفع

ISSN: 2537-0448 eISSN: 2537-0456

والقيم (a_i > ١.٠٠) تمييز عال جدًا. وإجمالاً تشير دي أيلا (De Ayala, 2009) أن القيم التي تتراوح بين ٠.٣٠ و٠.٥٠ تعكس مدى جيد لمعاملات التمييز.

ويقدم بيكر (2001) أن قيم معامل التخمين تتراوح بين صفر وواحد، كلما كانت القيم قريبة من صفر فإنها تعكس احتمالية منخفضة للتخمين. كما تشير الدراسات والأدبيات أن القيم التي تكون أقل من ٠.٣٥ تعكس مستوى مقبول للتخمين، بينما القيم التي تفوق عن ٠.٣٥ مدى غير مبرر للتخمين (Issayeva, 2022).

في النموذج الأول للاختبار، أشارت قيم الدلالة الإحصائية *(p-value)* لمؤشر *(S-X2)* مطابقة جيدة لكافة الفقرات ماعدا الفقرة(DI-18) لكونها أقل من ٠.٠١ (Toland, 2014). في المقابل أشارت (RMSEA) مطابقة جيدة لكافة الفقرات لكونها أقل من ٠.٠٥ إلا الفقرة (DI-18) كانت قيمة (RMSEA) تساوي ٠.٠٥ (Schumacker & Lomax, 2016). وتراوحت معاملات التمييز لفقرات النموذج الأول بين (٠.٦٦ - ١٢.١٤). تميزت عبارتين (CI-1, CI-7) بتمييز متوسط، وتميزت فقرة واحدة بتمييز مرتفع (CI-3)، وتميزت باقي الفقرات بتمييز عال جدًا. تراوحت معاملات الصعوبة لمفردات النموذج الأول بين (١.٥٩- إلى ١.٦٥) مشيرة لتمتع فقرات النموذج الأول بمدى صعوبة مقبول وفقًا لنقاط القطع الإحصائية (Baker, 2001; De Ayala, 2009; Toland, 2014). تمتعت (٣٢) فقرة بمستوى تخمين مقبول لأنها جاءت أقل من ٠.٣٥ تراوحت بين (٠ – ٠.٣٥) ماعدا الفقرة (DI-14) ٠.٣٦ ويعتبر مدى غير مبرر للتخمين وفقًا لـ (Issayeva, 2022).

في النموذج الثاني للاختبار، أظهرت النتائج، تمتع كافة الفقرات بمطابقة جيدة وفقًا لقيم الدلالة الإحصائية *(p-value)* لمؤشر *(S-X2)* حيث جاءت أعلى من ٠.٠١، ووفقًا لقيم RMSEA التي جاءت أقل من ٠.٠٥ (Baker, 2001; De Ayala, 2009; Toland, 2014). وتراوحت معاملات التمييز لمفردات النموذج الثاني بين (٠.٢٦ – ٤.٣٩) عدا الفقرة (DI-14=٢.٨٦-)، وتميز الفقرة (CI -7) بتمييز ضعيف، واتسمت ثلاثة فقرات بتمييز متوسط (CI-1, DI-5, DI-12) وباقي الفقرات كان لها تمييز عال جدًا. تراوحت معاملات الصعوبة لفقرات النموذج الثاني (١.٠١- إلى ١.٦٩) مشيرة لتمتع فقرات النموذج الثاني بمدى صعوبة مقبول وفقًا لنقاط القطع الإحصائية (Baker, 2001; De Ayala, 2009; Toland, 2014). عدا الفقرة (DI-14=٢.٥١-) وتم حذف هذه الفقرة وفقًا لنقطة القطع المتشددة (Baker, 2001). تمتعت غالبية الفقرات (٣٢ فقرة) بمستوى مقبول للتخمين لأنها جاءت أقل من ٠.٣٥ حيث تراوحت معاملات التخمين (٠ – ٠.٣٥) عدا فقرة واحد تميز بتخمين غير مقبول (DI-16= ٠.٤٢).

ISSN: 2537-0448 eISSN: 2537-0456

فيما يتعلق بالنموذج الثالث، أظهرت النتائج أن قيم الدلالة الإحصائية *(p-value)* لمؤشر *(S-X2)* مطابقة جيدة لكافة الفقرات حيث جاءت أعلى من ٠.٠١، في المقابل جاءت RMSEA أقل من ٠.٠٥ مما أشار بمطابقة جيدة لكافة الفقرات. وتباينت معاملات التمييز في المدى بين (٠.٩٤ – ٤.٢٧). تميزت كافة فقرات النموذج الثالث بتمييز عال جدًا حبث كانت أعلى من واحد ($a_i > ١.٠٠$). وتراوحت معاملات الصعوبة (0.97- إلى ١.٥١) مشيرة لتمتع الفقرات بمدى صعوبة مقبول وفقًا لنقاط القطع الإحصائية (Baker, 2001; De Ayala, 2009; Toland, 2014). تراوحت معاملات التخمين بمدى مقبول (٠ – ٠.٣٥) عدا فقرة واحدة مدى غير مقبول (DI-2= ٠.٤٤).

بالنسبة للنموذج الرابع للاختبار، بينت قيم الدلالة الإحصائية *(p-value)* لمؤشر *(S-X2)* مطابقة جيدة لكافة الفقرات حيث جاءت أعلى من ٠.٠١. أكدت هذا الاستنتاج قيم RMSEA التي جاءت أقل من ٠.٠٥ مما أشار بمطابقة جيدة لكافة الفقرات. تباينت معاملات التمييز في المدى بين (٠.٤٩ – ٣.٧٦). تميزت ثلاثة فقرات بتمييز متوسط (CI-3, DI-14, DI-19). وفقرة واحدة (DI-1) بتمييز مرتفع وباقي فقرات النموذج (٢٩ فقرة) لها تمييز عال جدًا. تمتعت فقرات النموذج الرابع بمدى صعوبة مقبول تراوحت بين (٠.٧٥- إلى ١.٣٩) ماعدا فقرة واحد (DI-19= ٣.٨٢) تعدى المدى المقبول للصعوبة. تراوحت معاملات التخمين (٠ – ٠.٣٥) مشيرة إلى تمتع الفقرات بمستوى تخمين مقبول لأنها جاءت أقل من ٠.٣٥ ماعدا ثلاثة فقرات تمتعت بمستوى تخمين غير مقبول (CI-2, DI-12, DI-20).

إجمالاً، أشارت النتائج مطابقة جيدة لكافة الفقرات في النماذج الاختبارية الأربعة وفقًا لقيم الدلالة الإحصائية لمؤشر *(S-X2)* ماعدا فقرة واحد فقط في النموذج الأول، كذلك أشار المؤشر RMSEA مطابقة جيدة في النماذج الأربعة مما يدعم قبول المطابقة الجيدة لكل الفقرات. كما أوضحت النتائج وجود تنوع في مستويات صعوبة الفقرات في النماذج الأربعة. كما أظهرت قدرة تمييزية متنوعة للفقرات حيث أظهر عدد قليل من الفقرات بتمييز ضعيف وفقرة واحدة في النموذج الثاني بالسالب وتم حذفها، بينما تراوحت القدرة التمييزية للغالبية العظمى من الفقرات بين متوسط والعالية جدًا. كما بينت النتائج مستوى مقبول للتخمين في معظم الفقرات. لتلخيص معاملات النموذج ثلاثي المعالم ثنائي الاستجابة أحادي البعد، ويوضح الشكل ٥ منحنى خصائص المفردات Item Characteristics Curve (ICC) في النماذج الأربعة للاختبار التحصيلي؛ لتلخيص معاملات النموذج ثلاثي المعامل ثنائي التصحيح أحادي البعد. كما يوضح الشكل ٦ مطابقة الفقرات infit and outfit في النماذج الأربعة للاختبار التحصيلي، ويلاحظ أن قيم إحصائيات

ISSN: 2537-0448 eISSN: 2537-0456

المطابقة (ZSTD) لإحصائي المطابقة الداخلية والخارجية لفقرات الاختبار كانت قريبة جدًا من الواحد الصحيح، وأن تكون قيم MNSQ لـــ outfit and infit في المدى بين (٠.٥ – ١.٥) وهذا الوضع المثالي كما يتوقعه النموذج (Linacre & Wright, 1993). كما أن القيم وقعت في المدى بين (٠.٧٠ - ١.١٠) في النماذج الأربعة مما يشير إجمالاً لمطابقة جيدة للفقرات.

الجدول (١٠) معاملات النموذج الثلاثي المعالم ثنائي الاستجابة أحادي البعد في النموذجين الأول والثاني

النموذج الأول						النموذج الثاني					
الفقرات	a_i (se)	b_i (se)	c_i (se)	*S-X2*	*p*	RMSEA	a_i (se)	b_i (se)	c_i (se)	S-X2	*p*	RMSEA
CI - 1	0.67 (0.39)	0.71 (1.07)	0.04 (0.32)	25.61	0.37	0.01	0.65 (0.84)	1.09 (2.00)	0.07 (0.60)	29.39	0.17	0.02
CI - 2	2.14 (0.42)	0.17 (0.15)	0.26 (0.06)	26.09	0.16	0.02	1.29 (0.1)	-0.15 (0.09)	0.00 (0.01)	19.31	0.44	0.01
CI - 3	1.37 (0.36)	0.88 (0.18)	0.15 (0.07)	23.49	0.43	0.01	1.52 (0.3)	1.20 (0.16)	0.12 (0.05)	23.25	0.45	٠0.0
CI - 4	2.17 (0.48)	0.75 (0.11)	0.12 (0.04)	19.61	0.55	0.00	4.43 (1.1)	0.76 (0.07)	0.15 (0.03)	31.88	0.02	0.04
CI - 5	1.55 (0.29)	0.31 (0.16)	0.13 (0.07)	20.02	0.52	0.00	1.74 (0.4)	٤0.6 ٠.١) (	0.15 (0.1)	26.20	0.16	0.02
CI - 6	2.39 (0.43)	0.38 (0.11)	0.24 (0.04)	18.24	0.57	0.00	1.89 (0.4)	0.29 (.19)	١0.2 (0.0)	29.82	0.05	0.03
CI - 7	0.68 (0.11)	0.32 (0.18)	0.00 (0.03)	26.32	0.34	0.01	0.26 (0.1)	0.61 (2.42)	0.02 (0.30)	34.31	0.08	0.03
CI - 8	1.78 (1.78)	0.51 (0.15)	0.18 (0.06)	16.27	0.75	0.00	2.11 (0.4)	0.40 (0.13)	0.15 0.06	20.84	0.35	0.01
CI - 9	1.73 (0.41)	0.93 (0.15)	0.26 (0.05)	27.36	0.29	0.02	1.63 (0.4)	1.22 (0.18)	0.27 (0.05)	10.34	0.99	0.00
CI - 10	1.79 (0.56)	0.99 (0.16)	0.24 (0.05)	22.25	0.51	0.00	4.39 (1.68	1.04 (0.10)	0.31 (0.03)	18.79	0.59	0.00
CI - 11	3.84 (0.87)	0.85 (0.09)	0.26 (0.03)	19.81	0.53	0.00	2.65 (0.7)	0.83 (0.12)	0.21 0.04	18.10	0.58	0.00
CI-12	3.33 (0.72)	1.17 (0.09)	0.14 (0.02)	20.69	0.60	0.00	2.00 (0.5)	1.54 (0.15)	0.08 (0.03)	16.43	0.79	0.00

ISSN: 2537-0448 eISSN: 2537-0456

0.00	0.75	15.38	0.00 (0.03)	-0.87 (0.15)	1.07 (0.1)	0.02	0.33	19.02	0.00) 0.04 (	-1.59) 0.19 (	1.24 (0.18)	DI-1
0.01	0.38	20.30	0.01 (0.06)	-0.80 (0.15)	1.26 (0.17	0.02	0.28	27.63	0.16 (0.03)	1.55 (0.17)	1.82 (0.53)	DI-2
0.03	0.14	27.00	0.22 (0.07)	0.74 (0.18)	1.76 (0.51	0.02	0.15	30.99	0.12 (0.35)	-0.23 (0.97)	0.98 (0.43)	DI-3
0.02	0.29	20.81	0.01 0.06	-1.01 0.18	1.18 0.16	0.00	0.65	20.84	0.31 (0.05)	0.88 (0.16)	1.92 (0.53)	DI-4
0.03	0.09	31.11	0.01 (0.25)	0.12 (0.76)	0.77 (0.27)	0.00	0.49	20.49	0.00 (0.05)	-0.12 (0.13)	1.58 (0.21)	DI-5
0.02	0.27	23.43	0.25 (0.05)	0.52 (0.14)	2.31 (0.56)	0.02	0.17	23.66	0.00 (0.01)	-0.46 (0.09)	1.63 (0.18)	DI-6
0.02	0.16	25.05	0.31 (0.06)	0.32 (0.16)	2.46 (0.654)	0.03	0.09	33.58	0.00 0.03	1.39 (0.27)	0.66 (0.13)	DI-7
0.00	0.93	12.44	0.23 (0.04)	0.99 (0.14)	2.01 (0.54)	0.03	0.09	31.58	0.31 (0.03)	0.86 (0.10)	3.59 (0.99)	DI-8
0.02	0.16	28.66	0.27 (0.08)	1.06 (0.22)	1.49 (0.58)	0.00	0.45	23.17	0.12 (0.09)	0.40 (0.23)	1.36 (0.33)	DI-9
0.02	0.29	17.47	0.15 (0.09)	-0.29 (0.17)	2.23 (0.4)	0.05	0.01	31.74	0.26 (0.16)	-0.67 (0.31)	2.03 (0.47)	DI-10
0.00	0.59	16.03	0.20 (0.09)	0.16 (0.20)	1.83 (0.43	0.00	0.58	16.25	0.08 (0.18)	-0.61 (0.34)	1.59 (0.34)	DI-11
0.03	0.08	34.06	0.18 (0.46)	0.71 (1.97)	0.59 (0.53)	0.01	0.39	18.92	0.17 (0.16)	-0.41 (0.33)	1.75 (0.43)	DI-12
0.00	0.76	14.41	0.13 (0.09)	0.27 (0.22)	1.72 (0.45)	0.02	0.29	26.09	0.13 (0.05)	1.05 (0.15)	1.50 (0.36)	DI-13
0.3	0.08	35.56	0.21 (0.02)	-2.51 (0.8	-2.86 (4.67	0.02	0.31	23.63	0.36 (0.05)	0.50 (0.15)	2.35 (0.60)	DI-14

ISSN: 2537-0448 eISSN: 2537-0456

				8)								
0.00	0.59	17.92	0.21 (0.05)	0.77 (0.13)	2.15 (0.50)	0.00	0.99	9.43	0.16 (0.03)	1.08 (0.09)	2.99 (0.72)	DI-15
0.00	0.89	16.06	0.42 (0.03)	1.69 (0.25)	1.98 (0.77)	0.00	0.56	19.46	0.17 (0.03)	0.85 (0.09)	2.91 (0.63)	DI-16
0.00	0.91	12.85	0.26 (0.04)	0.97 (0.13)	2.16 (0.57)	0.02	0.15	30.16	0.24 (0.03)	1.36 (0.12)	3.03 (0.84)	DI-17
0.00	0.52	19.96	0.25 (0.06)	0.63 (0.16)	1.93 (0.47)	0.05	0.00	39.62	0.34 (0.03)	0.73 (0.08)	3.00 (7.79)	DI-18
0.00	0.85	15.32	0.05 (0.04)	1.55 (0.17)	1.52 (0.43)	0.02	0.16	23.84	0.18 (0.02)	0.71 (0.07)	3.24 (2.30)	DI-19
0.03	0.03	37.31	0.07 (0.07)	0.91 (0.19)	1.15 (0.28)	0.04	0.04	28.63	0.21 (0.03)	0.68 (0.09)	3.68 (2.69)	DI-20
0.00	0.71	19.85	0.23 (0.06)	1.49 (0.24)	1.26 (0.43)	0.03	0.05	35.43	0.08 (0.03)	1.65 (0.19)	1.45 (0.42)	DI-21

ملاحظة: CI= الفقرات المشتركة (Common Items)، DI= الفقرات المختلفة (Different Items)، se= الخطأ المعياري (Standard error)، a_i= معامل التمييز، b_i= معامل الصعوبة/القدرة، c_i= معامل التخمين.

الجدول (١١) معاملات النموذج الثلاثي المعالم ثنائي الاستجابة أحادي البعد في النموذجين الثالث والرابع

النموذج الرابع						النموذج الثالث						
RMSEA	*p*	S-X2	c_i (se)	b_i (se)	a_i (se)	RMSEA	*p*	*S-X2*	c_i (se)	b_i (se)	a_i (se)	الفقرات
0.01	0.33	25.33	0.16 (0.12)	0.15 (0.33)	1.01 (0.43)	0.00	0.79	١٨.٢٢	0.21 (0.06)	0.93 (0.16)	1.58 (0.42)	CI -1
0.01	0.44	19.28	0.36 (0.05)	0.42 (0.12)	3.17 (0.80)	0.03	0.03	٣٧.٥٠	0.21 (0.07)	0.34 (0.16)	1.71 (0.34)	CI -2
0.00	0.53	22.91	0.00 (0.01)	1.05 (0.21)	0.69 (0.12)	0.00	0.49	٢٣.٤٦	0.16 (0.07)	1.42 (0.21)	1.21 (0.48)	CI -3
0.02	0.23	25.34	0.13 (0.04)	0.69 (0.10)	2.71 (0.57)	0.01	0.39	22.21	0.23 (0.03)	0.77 (0.08)	3.82 (0.81)	CI -4
0.00	0.89	13.61	0.17 (0.07)	0.54 (0.17)	1.66 (0.39)	0.00	0.59	19.85	0.16 (0.05)	0.51 (0.12)	2.32 (0.49)	CI -5
0.00	0.76	15.34	0.15 (0.07)	0.14 (0.15)	2.12 (0.43)	0.03	0.11	29.07	0.22 (0.05)	0.44 (0.12)	2.52 (0.56)	CI -6
0.03	0.09	33.53	0.35 (0.04)	1.39 (0.18)	2.06 (0.91)	0.00	0.68	20.35	0.11 (0.24)	0.27 (0.67)	1.01 (0.43)	CI -7
0.00	0.59	17.98	0.20 (0.05)	0.47 (0.12)	2.53 (0.54)	0.01	0.39	20.16	0.03 (0.07)	0.03 (0.14)	1.94 (0.32)	CI -8
0.00	0.97	11.35	0.19 (0.04)	0.90 (0.12)	2.09 (0.48)	0.00	0.80	17.15	0.13 (0.09)	0.69 (0.26)	1.18 (0.34)	CI -9
0.03	0.13	30.61	0.22 (0.04)	1.06 (0.13)	2.02 (0.57)	0.03	0.08	34.34	0.29 (0.05)	1.27 (0.16)	1.95 (0.79)	CI -10
0.00	0.97	9.92	0.22 (0.03)	0.84 (0.09)	3.76 (0.93)	0.00	0.48	22.64	0.22 (0.04)	0.76 (0.12)	2.09 (0.47)	CI -11

ISSN: 2537-0448 eISSN: 2537-0456

0.00	0.92	12.72	0.12 (0.02)	1.35 (0.11)	3.10 (0.78)	0.00	0.69	19.25	0.05 (0.02)	٥1.3 (0.12)	1.88 (٠.٤١)	CI-12
0.02	0.16	30.70	0.32 (0.17)	0.36 (0.62)	0.99 (0.41)	0.03	0.08	35.66	0.16 (0.13)	0.88 (0.37)	1.00 (0.38)	DI-1
0.03	0.12	30.13	0.14 (0.05)	0.77 (0.14)	1.83 (0.43)	0.00	0.64	21.01	0.44 (0.09)	0.49 (0.34)	1.35 (0.45)	DI-2
0.02	0.19	27.34	0.31 (0.05)	0.64 (0.14)	2.21 (0.56)	0.00	0.62	21.40	0.28 (0.03)	1.23 (0.11)	3.44 (1.19)	DI-3
0.00	0.57	17.25	0.18 (0.05)	0.37 (0.11)	3.00 (0.63)	0.03	0.07	33.59	0.21 (0.03)	1.04 (0.11)	2.48 (0.58)	DI-4
0.03	0.12	30.03	0.16 (0.23)	-0.12 (0.65)	1.01 (0.33)	0.03	0.09	26.65	0.31 (0.09)	-0.10 (0.19)	2.54 (0.61)	DI-5
0.03	0.06	31.64	0.28 (0.04)	0.61 (0.12)	2.58 (0.62)	0.01	0.31	26.97	0.30 (0.03)	1.49 (0.16)	2.33 (0.82)	DI-6
0.01	0.36	17.42	0.08 (0.25)	-0.75 (0.43)	1.64 (0.39)	0.00	0.49	22.53	0.28 (0.05)	0.82 (0.15)	2.14 (0.67)	DI-7
0.01	0.44	20.37	0.01 (0.07)	-0.20 (0.14)	1.62 (0.22)	0.02	0.20	20.39	0.33 (0.19)	-0.68 (0.40)	2.09 (0.58)	DI-8
0.02	0.04	36.93	0.00 (0.02)	-0.28 (0.11)	1.14 (0.15)	0.00	0.57	15.31	0.29 (0.04)	0.18 (0.09)	4.27 (0.90)	DI-9
0.00	0.66	17.77	0.06 (0.09)	0.31 (0.21)	1.42 (0.31)	0.00	0.51	23.22	0.35 (0.03)	1.51 (0.18)	2.09 (0.69)	DI-10
0.00	0.68	17.55	0.16 (0.12)	0.27 (0.32)	1.31 (0.36)	0.01	0.31	26.88	0.00 (0.01)	-0.27 (0.11)	0.94 (0.12)	DI-11
0.02	0.29	15.32	0.44 (0.08)	-0.24 (0.19)	3.47 (1.01)	0.00	0.56	18.42	0.17 (0.12)	-0.27 (0.24)	1.90 (0.41)	DI-12
0.00	0.72	17.86	0.35 (0.10)	0.33 (0.31)	1.50 (0.45)	0.00	0.69	16.28	0.01 (0.07)	0.03 (0.14)	1.75 (0.28)	DI-13
0.00	0.63	21.16	0.00 (0.01)	1.37 (0.25)	0.70 (0.12)	0.00	0.63	19.21	0.16 (0.06)	0.53 (0.14)	1.99 (0.44)	DI-14
0.00	0.87	14.09	0.24 (0.04)	0.70 (0.11)	2.76 (0.70)	0.00	0.92	11.92	0.25 (0.05)	0.30 0.11	2.77 (0.55)	DI-15
0.00	0.81	15.19	0.19 (0.08)	0.41 (0.19)	1.86 (0.48)	0.00	0.74	15.55	0.00 (0.02)	0.14 (0.08)	1.92 (0.21)	DI-16
0.03	0.08	30.46	0.18 (0.04)	0.87 (0.10)	3.67 (1.41)	0.02	0.15	27.75	0.33 (0.06)	0.44 (0.15)	2.55 (0.69)	DI-17
0.00	0.47	21.87	0.15 (0.03)	1.03 (0.10)	2.53 (0.57)	0.02	0.18	28.94	٤0.1 (0.04)	0.96 (0.12)	1.89 (0.44)	DI-18
0.03	0.14	29.24	0.00 (0.03)	3.82 (1.08)	0.49 (0.17)	0.03	0.05	34.36	0.19 (0.02)	1.03 (0.09)	3.04 (0.81)	DI-19
0.00	0.64	18.12	0.37 (0.07)	0.45 (0.19)	2.29 (0.72)	0.00	0.51	22.12	0.11 (0.05)	0.95 (0.12)	2.09 (0.59)	DI-20
0.01	0.39	22.20	0.33 (0.03)	0.78 (0.10)	3.68 (1.10)	0.02	0.16	27.46	0.00 (0.04)	-0.97 (0.14)	1.29 (0.16)	DI-21

ملاحظة: CI= الفقرات المشتركة (Common Items)، DI= الفقرات المختلفة (Different Items)، se= الخطأ المعياري (Standard error)، a_i= معامل التمييز، b_i= معامل الصعوبة/القدرة، c_i= معامل التخمين.

مناقشة نتائج السؤال الرابع:

أظهرت النتائج المتعلقة بهذا السؤال مطابقة النموذج الثلاثي المعالم ثنائي التصحيح لتقدير معاملات الصعوبة والتمييز والتخمين والخطأ المعياري لكل معامل. أشارت النتائج مطابقة جيدة لكافة الفقرات في النماذج الاختبارية الأربعة وفقًا لقيم الدلالة الإحصائية لمؤشر *(S-X2)* ماعدا فقرة واحد فقط في النموذج الأول، كذلك

ISSN: 2537-0448 eISSN: 2537-0456

أشار المؤشر RMSEA مطابقة جيدة في النماذج الأربعة مما يدعم قبول المطابقة الجيدة لكل الفقرات متوافقة مع الدراسات والأدبيات ذات العلاقة (Baker, 2001; De Ayala, 2009; Toland, 2014) تطابقت نتائج هذا الفقرة مع عدد من الدراسات التي أظهرت تمتع الفقرات بمطابقة جيدة عند استخدام النموذج ثلاثي المعالم ، وأوضحت النتائج وجود تنوع في مستويات صعوبة الفقرات في النماذج الأربعة. فقد تراوح مستوى صعوبة الفقرات في بعض النماذج بين (+٢ و -٢) متوافقًا مع بيكر (٢٠٠١)، وبعضها تراوحت بين (+٣ و -٣) متماشيًا مع تولاند (٢٠١٤). وتأتي هذه النتائج مع دراسة البيايضة (٢٠١١) التي أظهرت أن معاملات صعوبة الفقرات تراوحت بين (٢.١٢ و٣.٣٧-) عند معايرة مفردات اختبار تكيفي للقدرة الرياضية للصف السابع الأساسي.

وأظهرت أيضًا نتائج هذا السؤال قدرات تمييزية مختلفة ومتنوعة للفقرات حيث أظهر عدد قليل من الفقرات تمييز ضعيف وفي النموذج الثاني تم حذف الفقرة من الاختبار كونها أظهرت تمييز بالسالب، بينما تراوحت القدرة التمييزية للغالبية العظمى من الفقرات بين المتوسطة والعالية جدًا. كما بينت النتائج مستوى مقبول للتخمين في معظم الفقرات. وتوافقت هذه النتائج مع معايير الأدبيات الإحصائية (De Ayala, 2009; Issayeva, 2022; Toland, 2014).

النتائج المتعلقة بالسؤال الخامس:

نص السؤال الخامس على: **" ما دالة المعلومات والخطأ المعياري للاختبار التكيُّفيّ المحوسب لمادة الرياضيات للصف الرابع الأساسي"؟**

للإجابة عن هذا السؤال، تم تقدير أولاً دالة المعلومات لكافة الفقرات في النماذج الأربعة للاختبار التحصيلي ويلخص مستوى فاعلية المفردات ودقة قياسها للقدرة مع الأخذ في الاعتبار جميع الخصائص السيكومترية للمفردات وبالتحديد معاملات الصعوبة والتمييز والتخمين (De Ayala, 2009). أظهرت نتائج المفردات تنوع في مقدار المعلومات التي تقدمها كل فقرة حيث بلغ أقصى كم من المعلومات المقدمة عند مستويات مختلفة من القدرات. حيث قدمت ثلاثة فقرات (٣٠، ٣١، ٣٢) أكبر قدر من المعلومات في النموذج الأول. بينما قدمت فقرتين (٤، ١٠) أكبر كم من المعلومات في النموذج الثاني. وأظهر النموذج الثالث أن أكبر كم من المعلومات تم قياسه من سبعة فقرات (٤، ٥، ٦، ١٥، ٢١، ٢٧، ٣١) وأخيرً في النموذج الرابع أظهرت أغلب الفقرات أكبر قدر من المعلومات ماعدا تسعة فقرات (١،٣، ١٣، ١٧، ٢١، ٢٣، ٢٥، ٢٦، ٣١).

كما تم جمع كافة دوال معلومات المفردات لكل نموذج في دالة إجمالية لتقدير مدى دقة قياس كل نموذج للقدرة المقاسة، وتم التعبير عنها بدالة معلومات الاختبار

ISSN: 2537-0448 eISSN: 2537-0456

في الشكل (8) والذي تقدم مصدرين لدقة القياس وهما: حجم المعلومات ممثلاً بالخط المتصل، والخطأ المعياري ممثلاً بالخط المتقطع (Bond & Fox, 2015). ومن خلال الاطلاع على الأدبيات أشارت أن الشكل المثالي لدالة معلومات الاختبار يجب أن يظهر أكبر قدر من المعلومات أعلى من الخطأ المعياري في المدى المتوسط للقدرة الممثل في الخط الأفقي للشكل. وأيضًا يجب أن يظهر الشكل مستويات خطأ أعلى من كم المعلومات عند أطراف القدرة سواء على اليمين أو اليسار. هذا الوضع المثالي ينطبق على كافة دوال معلومات الاختبار في النماذج الأربعة (TIF). فقد أظهرت دوال معلومات الاختبار أكبر قدر من المعلومات عند مستوى قدرة يساوي واحد بمعدل ٥٠ في النموذج الأول، وبمعدل ٢٠ في النموذج الثاني والثالث، وفي النموذج الرابع بمعدل ٢٥. كما أظهرت الدوال في النماذج الأربعة مستويات خطأ منخفضة عند المستويات المتوسطة للقدرة.

مناقشة نتائج السؤال الخامس:

أظهرت نتائج منحنى معلومات الفقرات (IIF) تنوع في مقدار المعلومات التي تقدمها كل فقرة حيث بلغ أقصى كم من المعلومات المقدمة عند مستويات مختلفة من القدرات. كما أظهرت دوال معلومات الاختبار التحصيلي (TIF) لمادة الرياضيات لدى طلبة الصف الرابع قدرًا أكبر من المعلومات على نحو يفوق الخطأ المعياري في المدى المتوسط للقدرة في النماذج الأربعة، وربما يعود ذلك إلى معلمة التمييز للفقرات، حيث كانت قيم التمييز للفقرات عالية فكلما زاد معلم التمييز للفقرة زادت كمية المعلومات التي تسهم بها الفقرة وقد يعزى ذلك أيضًا إلى معلمة التخمين حيث كانت قيم التخمين مقبولة أي منخفضة نوعًا ما فكلما قلت معلمة التخمين زادت القيمة العظمى للمعلومات متوافقة مع عدد من الأدبيات الإحصائية (Bond & Fox, 2015; De Ayala, 2009) وأسفرت النتائج عن فعالية الاختبار التكيُّفيّ المحوسب عن فعالية الاختبار التكيُّفيّ المحوسب على الاختبار الورقي التقليدي (الخطي) إذ كان متوسط دالة المعلومات التي يوفرها ٧.٤٥ وأقل في قيمة الخطأ المعياري بمتوسط بلغ ٠.٣٩ مع تقدير للقدرة لا يختلف عن الورقي التقليدي بصورة دالة.

النتائج المتعلقة بالسؤال السادس:

نص السؤال السابع على: **" ما فاعلية الاختبار التكيُّفيّ المحوسب لمادة الرياضيات للصف الرابع الأساسي في ضوء الاشتراطات الفنية المتعلقة (ببدء الاختبار، اختيار الفقرات، تقدير قدرات الطلبة، ومعايير انهاء الاختبار من حيث تقدير القدرة، ومقدار الخطأ المعياري) "؟**

تم توحيد الاشتراطات الفنية المتعلقة ببدء وإنهاء الاختبار بما يتوافق مع المعايير التي استعرضتها الأدبيات الإحصائية (Chalmers, 2016). فيما يتعلق

ISSN: 2537-0448 eISSN: 2537-0456

ببدء الاختبار واختيار الفقرة الأول، تم استخدام طريقة الاحتمالية القصوى لتحديد الفقرة الأنسب لكل مستجيب من خلال جزيئة الكود (start_item= MI). أما فيما يتعلق بإنهاء الاختبار فقد اختارت الدراسة الحالية معياريين وهي: (١) صغر حجم الخطأ المعياري بحيث لا يكون أدنى من ٠.٢٥ والذي بدوره يعكس مقدارًا عاليًا من دقة القياس، و(٢) الحد الأعلى التي قد يُجيب عليها الطالب ٣٥ فقرة تقليلاً للجهد والتعب الذي قد يواجهه المستجيب. وتم تحديد زمن الاختبار وقدره ٤٥ دقيقة.

فيما يتعلق بطريقة اختيار الفقرات تشير الأدبيات الإحصائية إلى وجود أربعة طرق (Fixed, ML, EAP, MAP) لتقدير قدرة المستجيب أثناء إجراء الاختبار التكيُّفيّ. وتماشيًا مع الدراسات الإحصائية السابقة (Ince Araci & Tan, 2022; Chalmers, 2016) اقتصرت الدراسة الحالية على دراسة تأثير طريقتين فقط وهما: طريقة بيز باستخدام القيمة البعدية المتوقعة Bayesian Expected a Posteriori Estimation (EAP) وطريقة بيز باستخدام القيمة البعدية القصوى Bayesian Maximum a Posteriori Estimation (MAP) ؛ نظرًا لقدرتهما على التعامل الأمثل مع استجابات الطلبة التي توصف بـــ all-0r-none response patterns، والتي عادة ما تحصل عندما يحصل الطلبة على إجابات صحيحة لكافة الفقرات (all) أو إجابات خاطئة لكافة الفقرات (none).

تم التطبيق الفعلي للاختبار التكيُّفيّ المحوسب على العينة الثالثة للدراسة حيث بلغت ٥٨ طالب وطالبة. تم تحديد عدد من المواصفات للاختبار الحي فيما يتعلق ببدء الاختبار، وانتقاء الفقرات، وتقدير القدرة وإنهاء الاختبار. فيما يتعلق ببدء الاختبار، تم بدء الاختبار بفقرات متوسطة الصعوبة عند مستوى قدرة ٠.٦١. وبالنسبة لانتقاء الفقرات، تم انتقاء الفقرات وفق طرق فيشر للمعلومات وبالتحديد طريقة الاحتمالية القصوى الموزونة. وأيضًا تم تحديث قدرة الطالب وفق طرق فيشر للمعلومات وبالتحديد طريقة الاحتمالية القصوى الموزونة. تم تبني الطريقة التي يترتب عليها عدد غير موحد من الفقرات للطلبة (Variable-Iength) وبالتحديد معيارين وهما: تحقيق المستوى المطلوب للقدرة، ومستوى منخفض للخطأ المعياري للقياس (SEM< 0.01). ويوضح الجدول (١٢) قيم تقدير القدرة التي جاءت متفاوتة لكل الطلبة.

ISSN: 2537-0448 eISSN: 2537-0456

الجدول ١٢ تقدير القدرة للطلبة والخطأ المعياري وعدد الفقرات من خلال الاختبار التكيُّفيّ المحوسب

رقم الطالب	عدد الفقرات	عدد الإجابات الصحيحة	عدد الإجابات الخاطئة	تقدير القدرة	الخطأ المعياري
1	35	11	24	-0.877	0.299
2	10	4	6	0.632	0.238
3	10	8	2	1.081	0.224
4	10	6	4	0.867	0.171
5	10	5	5	0.779	0.166
6	21	16	5	1.403	0.236
7	23	13	10	0.078	0.246
8	10	6	4	0.684	0.184
9	10	6	4	0.684	0.184
10	10	6	4	0.684	0.184
11	10	5	5	0.593	0.246
12	10	5	5	0.608	0.183
13	10	8	2	1.121	0.232
14	21	11	10	-0.077	0.246
15	35	9	26	-1.259	0.361
١٦	10	5	5	0.765	0.169
١٧	35	12	23	-0.739	0.261
١٨	26	17	9	0.482	0.237
١٩	21	17	4	1.458	0.238
٢٠	5	4	1	0.825	0.186
21	10	4	6	0.632	0.238
22	10	5	5	0.712	0.181
23	١٠	7	3	0.741	0.167
24	10	8	2	1.054	0.218
25	30	16	14	-0.1	0.244
26	11	5	6	0.622	0.227
27	11	9	2	1.161	0.244
28	10	7	3	0.952	0.189
29	10	7	3	0.952	0.189
30	10	7	3	0.952	0.189
31	10	8	2	0.901	0.179
32	10	6	4	0.57	0.211

ISSN: 2537-0448 eISSN: 2537-0456

0.217	0.391	5	5	10	**33**
0.217	0.391	5	5	10	**34**
0.166	0.762	5	5	10	**35**
0.183	0.923	3	7	10	**36**
0.323	-0.982	22	13	35	**37**
0.297	1.99	3	29	33	**38**
0.218	-0.132	10	11	٢١	**39**
0.249	0.923	3	7	10	**40**
0.183	0.582	5	5	10	**41**
0.234	-1.181	28	7	35	**42**
0.322	0.217	12	15	27	**43**
0.246	0.924	3	7	10	**44**
0.183	0.632	6	4	10	**45**
0.238	0.641	3	7	10	**46**
0.156	0.309	5	5	10	**47**
0.242	-0.201	12	13	25	**48**
0.245	0.345	4	7	11	**49**
0.247	0.345	4	7	11	**50**
0.211	1.019	2	8	10	**51**
0.238	0.632	6	4	10	**52**
0.155	0.675	4	6	10	**53**
0.155	0.675	4	6	10	**54**
0.233	1.123	2	8	10	**55**
0.168	0.83	4	6	10	**56**
0.248	0.09	12	13	25	**٥٧**
0.246	0.075	10	14	24	**٥٨**

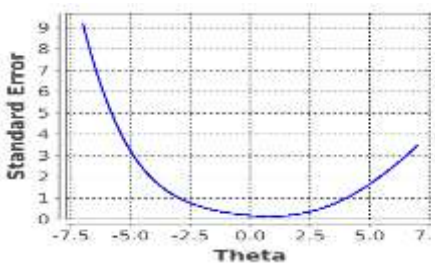
الجدول ١٣ الملخص الإحصائي للمتوسط الحسابي والانحراف المعياري للصعوبة والتمييز والتخمين

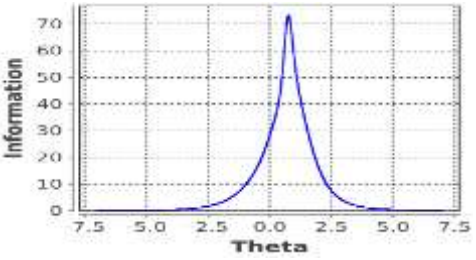
الإحصائية	المتوسط الحسابي	الانحراف المعياري
معلمة IRT a (التمييز)	٢.٠٩٧	١.٢٨٥
معلمة IRT b (الصعوبة)	٠.٥١٩	٠.٧٦٤
معلمة IRT c (التخمين)	٠.١٧٩	٠.١١٧

ويعرض الشكل (١) دالة معلومات الاختبار، والخطأ المعياري لتقدير القدرة بعد التطبيق الحي للاختبار التكيُّفيّ المحوسب. أظهر الشكل (١) وبالتحديد الجزيئة أ قيمة متوسطة نوعًا ما من المعلومات التي يقدمها الاختبار خاصة عند مستويات القدرة المتوسطة وذلك بسبب صغر حجم العينة وكذلك عدم اطلاع الطلبة على جميع

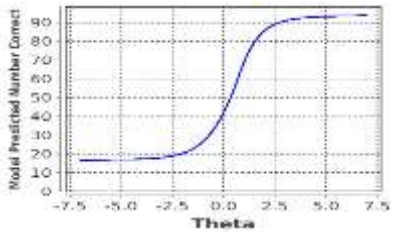
ISSN: 2537-0448 eISSN: 2537-0456

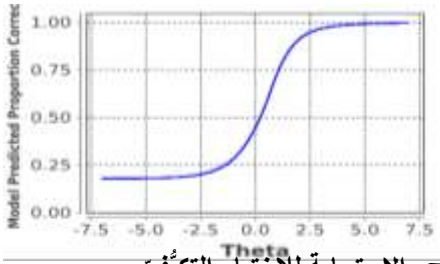
الفقرات بسبب المعايير التي تم وضعها للاختبار التكيُّفيّ الموحسب. في المقابل، وبالتحديد الجزئية ب قيمة منخفضة للخطأ المعياري لتقدير القدرة مما يشير إلى دقة عالية للاختبار. وأيضًا توضح كلاً من الجزئية ج ود **الاستجابة للاختبار التكيُّفيّ المحوسب الحي بشكل عددي ونسبي.**





أ. دالة معلومات للاختبار التكيُّفيّ المحوسب الحي **ب. الخطأ المعياري للاختبار التكيُّفيّ المحوسب الحي**





ج. الاستجابة للاختبار التكيُّفيّ المحوسب الحي (نسبيًا) **د. الاستجابة للاختبار التكيُّفيّ المحوسب الحي (عدديًا)**

الشكل ١

دالة المعلومات والخطأ المعياري لتقدير القدرة في الاختبار التكيُّفيّ المحوسب الحي

مناقشة نتائج السؤال السادس:

تم التطبيق الفعلي للاختبار التكيُّفيّ المحوسب FastTest v3.81.3 على العينة الثالثة للدراسة حيث بلغت ٥٨ طالب وطالبة، مع تحديد عدد من الاشتراطات الفنية للاختبار التكيُّفيّ المحوسب. فقد تم بدء الاختبار بفقرات متوسطة الصعوبة عند مستوى قدرة ٠.٦١ تماشيًا مع عدد من الدراسات السابقة. وتم انتقاء الفقرات وفق طرق فيشر للمعلومات وبالتحديد طريقة الاحتمالية القصوى الموزونة نظرًا لشيوع وسهولة استخدامها وتم تحديث قدرة الطالب وفق طرق فيشر للمعلومات، وبالتحديد

ISSN: 2537-0448 eISSN: 2537-0456

طريقة طريقة الاحتمالية القصوى الموزونة توافقًا مع الدراسات السابقة. وتم إنهاء الاختبار من خلال تبني الطريقة التي يترتب عليها عدد غير موحد من الفقرات للطلبة Variable-length وبالتحديد معيارين، وهما: تحقيق المستوى المطلوب للقدرة، ومستوى منخفض للخطأ المعياري للقياس (SEM< 0.01). وأظهرت النتائج تقدير متوسط إلى عال الدقة لقدرات الطلبة حيث وفرت كم متوسط نوعًا ما من المعلومات ومقدار قليل للخطأ.

التوصيات:

هدفت الدراسة الحالية إلى بناء اختبار تكيُّفي محوسب في مادة الرياضيات للصف الرابع الأساسي باستخدام نظرية الاستجابة للمفردة. في ضوء النتائج التي تم التوصل لها، يمكن الخروج بالتوصيات الآتية:

١. أوصي باستخدام الاختبار التكيُّفيّ المحوسب المُعدّ في هذه الدراسة من قِبَل كافة المعنيين في وزارة التربية والتعليم، بما في ذلك المعلمين، والمشرفين التربويين، ومتخذي القرار، وذلك لتطوير عمليات التقييم في مادة الرياضيات للصف الرابع الأساسي.

٢. تصميم منصة وطنية تتيح إنشاء وتنفيذ الاختبارات التكيُّفيّة بحيث تكون متاحة للمعلمين والطلاب وتدعم عمليات التحليل والإحصاء التربوي.

٣. تصميم برامج تدريبية للمعلمين حول بناء الاختبارات وفق نظرية الاستجابة للمفردة.

المقترحات:

- إجراء دراسات مماثلة للبحث الحالي على طلبة الجامعات والكليات في سلطنة عمان.
- إجراء المزيد من البحوث حول استراتيجيات جديدة لبناء الاختبارات التكيُّفيّة المحوسبة.
- إجراء بحوث مستقبلية حول أثر هذه الاختبارات على تحسين التحصيل الدراسي، وتقليل القلق المرتبط بالاختبارات التقليدية ودراسة مدى فاعليتها مقارنة بأنواع التقييم الأخرى.

ISSN: 2537-0448 eISSN: 2537-0456

المراجع

المراجع العربية

أبو شندي، يوسف (٢٠١١). كواشف الاستقلال الموضعي في نظرية الاستجابة للفقرة ومدى تأثرها بقوة العلاقة الارتباطية بين الأبعاد ومستوى القدرة للمفحوصين. *مجلة جامعة تشرين للبحوث والدراسات العلمية – سلسلة الآداب والعلوم الإنسانية*، ٣٣ (٣)، ٤٩ – ٦٢.

البرصان، إسماعيل سلامة سليمان وبخيت، صلاح الدين فرح عطا الله (٢٠١٥). فاعلية القياس التكيُّفيّ في الكشف عن الموهوبين باستخدام اختبار رافن للمصفوفات المتتابعة المعياري. *رسالة التربية وعلم النفس*، (٥٠)، ٧٥-٩٧.

بركات، زياد (٢٠١٨). القياس والتقويم النفسي والتربوي بين النظرية والتطبيق. فلسطين: عمادة البحث العلمي جامعة القدس المفتوحة.

جرجس، ماريان ميلاد منصور (٢٠٢١). تصميم اختبار تكيفي إلكتروني بنائي برجع (تصحيحي، تفسيري) بمقرر الحاسب الآلي لطلاب كلية التربية وأثره على تقبلهم التكنولوجي له. *مجلة جامعة جنوب الوادي الدولية للعلوم التربوية*، (٧)، ٣٣٣ – ٣٧٩.

حماد، شريف علي (٢٠١١). جودة فقرات الامتحانات النصفية لبرنامج التربية في جامعة القدس المفتوحة في ضوء معايير جودة الامتحان. ورقة علمية مقدمة إلى اليوم الدراسي "جودة الامتحانات الجامعية" الواقع والمأمول، غزة، ٢٨ أبريل.

الحوسنية، هدى (٢٠٢٣). فاعلية تطبيق هاتفي قائم على منحي التلعيب " Dr. Sciencex" في تنمية الدافعية للإنجاز والاتجاه نحو التعلم الذاتي واكتساب المفاهيم العلمية لدى طلبة الصف الرابع الأساسي بسلطنة عمان [رسالة دكتوارة غير منشورة، جامعة السلطان قابوس].

خلف، لقاء شامل (٢٠٢٤). المقارنة في تقدير القدرة ودقتها لاختبار محبوك وفقا للاستراتجيتين ثنائية المرحلة الهرمية متعددة المراحل باستخدام نظرية القياس الحديثة. *مجلة كلية التربية – جامعة واسط*، ٥٥ (١)، ٣٦٥ - ٣٨٦.

سلامة، حنان طه (٢٠١٧). *تقييم نتاجات التعلم المعرفية لدى طلبة الصفوف من السادس إلى التاسع الأساسي في الرياضيات باستخدام اختبار متعدد المستويات وفق نظرية الاستجابة للفقرة* [رسالة دكتوراة غير منشورة]. الجامعة الأردنية، عمان.

الشقصي، يعقوب بن زاهر بن سليمان وأبو شندي، يوسف عبد القادر علي والمحرزي، راشد بن سيف بن مصبح (٢٠٢٠). فعالية مؤشرات مطابقة الفرد

ISSN: 2537-0448 eISSN: 2537-0456

في نماذج استجابة الفقرة عند اختلاف قوة الارتباط الموضعي بين الفقرات. *مجلة دراسات التربوية والنفسية*، *١٤* (١)، ٤١ – ٥٣.

الشهري، أحمد علي (٢٠٢٣). أثر قواعد توقف الطول الثابت في إدارة الاختبارات التكيُّفيّة. *المجلة العلمية – كلية التربية – جامعة الوادي الجديد*، (٤٥)، ١٩٧-٢١٦.

الصبيحي، علي أحمد، والأحمدي، شرف حامد. (٢٠١٥). الاختبار التكيُّفيّ المحوسب Computerized Adaptive Teasing (CAT) كتطبيق لتطوير أساليب التقويم والاختبارات الالكترونية. *مجلة كلية التربية-جامعة دمنهور*، ٣ (١)، ٨٣-١٢٦.

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

علام، صلاح الدين محمود. (١٩٨٦). *تطورات معاصرة في القياس التربوي*. جامعة الكويت: إدارة التأليف والترجمة والنشر.

عليمات، مهدي محمد فهد (٢٠٢١). بناء اختبار تحصيلي وفق نظرية الاستجابة للفقرة " النموذج ثنائي المعلمة". *مجلة جامعة النجاح للأبحاث – العلوم الإنسانية*، ٣٥ (٧)، ١٠٧٩ – ١١٠٤.

عودة، أحمد سليمان وعبيدات، عمر سليمان (٢٠١٣). فاعلية الاختبار التكيُّفيّ المحوسب في تقدير القدرة العقلية باستخدام مصفوفات رافن. العلوم التربية، ٤٠ (٢)، ١٦٠٢-١٦٢١.

فرحات، رمضان السيد (٢٠٢١). الفروق بين الاختبارات المحوسبة الموائمة والخطية في تقدير القدرة وفقا لنظرية الاستجابة للمفردة لطلاب كلية التربية جامعة الأزهر. *مجلة التربية - كلية التربية – جامعة الأزهر*، (١٩٢)، ج٣، ٢ – ٥٢.

قاسم، مصطفى أحمد. (٢٠٢١). *الاختبارات التكيُّفيّة المحوسبة "النظرية والتطبيق"*. دار الفكر العربي.

القضاة، عبد الحميد محمد والشريفين، نضال كمال محمد (٢٠٢٠). أثر طول الاستبانة على دقة تقديرات القدرة والخصائص السيكومترية للفقرة والمقياس في ضوء نظرية الاستجابة للفقرة. *مجلة جامعة النجاح للأبحاث – العلوم الإنسانية*، ٣٤ (٦)، ٩٥٣-٩٨٢.

كاظم، أمينة (١٩٩٦). *دراسة نقدية حول القياس الموضوعي للسلوك في اتجاهات معاصرة في القياس والتقويم النفسي والتربوي*. القاهرة: مكتبة الأنجلو المصرية.

ISSN: 2537-0448 eISSN: 2537-0456

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

كاظم، علي مهدي وعمارة، إيهاب محمد (٢٠٢٠). دراسة جدوى تطوير صورة تكيفية محوسبة من اختبار المصفوفات المتتابعة الملونة لرافن للأطفال العمانيين باستخدام نظرية الاستجابة للمفردة. *المجلة الدولية للأبحاث التربوية، ٤٤* (٣)، ١٤٣ – ١٨١.

الكرامنة، محمد صالح محمد (٢٠١٥). *بناء اختبار تكيفي محوسب للذكاء المنطقي الرياضي للطلبة الموهوبين باستخدام نظرية الاستجابة للفقرة* [رسالة دكتوراه غير منشورة] . الجامعة الأردنية، عمان.

مرجان، رضا أمين غالب والبرصان، إسماعيل سلامة (٢٠٢٣). فاعلية القياس التكيُّفيّ متعدد المراحل باستخدام طريقة دلتا لتقدير الدرجات. *مجلة الأستاذ للعلوم الإنسانية والاجتماعية*، ٦٢ (٢)، ٤٠ – ٦٤.

مهدي، رشيد صالح وجاسم، خالد جمال (٢٠٢٣). بناء اختبار تكيفي باستخدام الاستراتيجية الجزئية متعددة المراحل وفقا لنظرية الاستجابة للفقرة. *مجلة الأستاذ للعلوم الإنسانية والاجتماعية*، ٦٢ (١)، ٢٥٢ – ٢٦٧.

نجيب، حسين علي والرفاعي، غالب عوض صالح. (٢٠٠٦). تحليل ونمذجة البيانات باستخدام الحاسوب: تطبيق شامل للحزمة SPSS. عمان (الأردن): الاهلية للنشر والتوزيع.

وزارة التربية والتعليم (2025). الكتاب الإحصائي السنوي للإحصاءات التعليمية لعام ٢٠٢٣/ ٢٠٢٤.

وزارة التربية والتعليم (٢٠٢٣). وثيقة تقويم تعلم التلاميذ في مواد المجال الثاني للصفوف (١-٤).

المراجع الأجنبية

Alhadabi, A., Aldhafri, S., Alkharusi, H., AL-Harthy, I., ALBarashdi, H., & Alrajhi, M. (2019). Psychometric assessment and cross-cultural adaptation of the Grit-S Scale among Omani and American universities'students. *European Journal of Educational Research, 8*(4), 1175-1191.

Anastasi, A., & Urbina, S. (1997). *Psychology testing*, (7th Ed.), New York: Prentice Hall.

ISSN: 2537-0448 *eISSN: 2537-0456*

Arriza, L., Retnawati, H., & Ayuni, R. T. (2024). Item analysis of high school specialization mathematics exam questions with item response theory approach. *Barekeng: Jurnal Ilmu Matematika dan Terapan, 18*(1), 0151- 0162.

Ayanwale, M. A., & Ndlovu, M. (2024). The feasibility of computerized adaptive testing of the national benchmark test: A simulation study. *Journal of Pedagogical Research, 8*(2), 95–112.

Aybek, E.C. & Demirtasli, R.N. (2017). Computerized adaptive test (CAT) applications and item response theory models for polytomous items. International Journal of Research in Education and Science (IJRES), 3(2), 475–487.

Baker, F. (2001). *The basics of item response theory* (2nd ed.). Clearinghouse on assessment and evaluation.

Beavers, A., Lounsbury, J., Richards, J., Huck, S., Skolits, G., & Esquivel, S. (2013). Practical Considerations for using exploratory factor analysis in educational research. *Practical Assessment, Research, and Evaluation 18*(1), 1-13.

Boateng, G., Neilands, T., Frongillo, E., Melgar-Quinonez, H., & Young, S. (2018). Best practices for developing and validating scales for health, social, and behavioral research: A primer. *Frontiers in Public Health, 6*(149), 1-18.

Bond,T., & Fox, C. (2015). Applying the rasch model: *Fundamental measurement in the* human sciences (3rd ed.) Erlbaum.

Boyd, A. (2003). S*trategies for controlling Testlet Exposure Ratesin Computerized Adaptive Testing System*, Unpublished Thesis, University of at Austin, Canterbury.

ISSN: 2537-0448 *eISSN: 2537-0456*

Bulut O., & Kan A. (2012). Application of computerized adaptive testing to entrance examination for graduate studies in Turkey. *Eurasian Journal of Educational Research, 49*, 61-80.

Chalmers, R (2016). Generating adaptive and non-adaptive test interfaces for multidimensional item response theory applications. *Journal of Statistical Software, 71*(5), 1-38

Christensen, K., Makransky, G., & Horton, M. (2017). Critical values for Yen's Q3: Identification of local dependence in the Rasch model using residual correlations. *Applied Psycchological Measurement, 41*(3), 178-194.

Costello, A., & Osborne, J. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment Research & Evaluation, 10*(7), 1-9.

Crist, S. (1989). Computerized Adaptive test: Eric Digest No.107, Eric database.

Crocker, L &Algina, J. (2006). Introduction to Classical and Modern Test Theory. Cengage Learning, Ohio, USA.

De Ayala, R. (2009). *The theory and practice of item response theory*. Guilford Publications. Eggen, T. (2001). *Overexposure and underexposure of items in computerized adaptive testing*. Measurement and Research Department Reports.

Ebenbeck, N., Bastian, M., Mühling, A., & Gebhardt, M. (2024). Duration versus accuracy—What matters for computerised adaptive testing in schools? *Journal of Computer Assisted Learning*, jcal.13074.

ISSN: 2537-0448 ***eISSN: 2537-0456***

Field, A. (2009). *Discovering statistics using SPSS (3rd ed.)*. SAGE Publications. Fliege, H., Becker, J., Walter, O., Rose, M., Bjorner, J., & Klapp, B. (2009). Evaluation of a computer- adaptive test for the assessment of depression (D-CAT) in clinical application. *International Journal of Methods in Psychiatric Research, 18*(1), 23-36.

Garcia-perez, M. (2014). Multiple- choice teste: polytomous IRT models misestimate item informatiom. *The spanish Journal of psychology, 17*(88), 1-18

Gruijter, D.N.M., & Kamp, L.J.T.H. (2005). S*tatistical test Theory for education and Psychology*.

Hambleton, R. K., (1987). Determining Optimal Test Lengths with A Fixed Total Testing Time. Educational and Psychological Measurement, 2(47), 338-339.

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of Item Response Theory*. London, SAGE Publications.

Hambleton, R., & Swaminathan, H. (1985). Item Response Theory: Multiple-Choice Item Writing Guidelines for Classroom Assessment. *Applied Measurement in Education*, 15(3), 309-334.

Hattie, J. (1985). *Methodology Review: Assessing Unidimensionality of Tests and Items.* Applied Psychological Measurement, 9, 139-164.

Herosian, M. Y., Sihombing, Y. R., & Pulungan, D. A. (2022). Item Response Theory Analysis on Student Statistical Literacy Tests. *PEDAGOGIK: Jurnal Pendidikan*, *9*(2), 203-215.

ISSN: 2537-0448 ***eISSN: 2537-0456***

Hinkle, D., Wiersma, W. & Jurs, S. (2003). *Applied statistics for the behavioral sciences* (5th ed.). Cengage Learning.

Immekus, J. C., Snyder, K. E., & Ralston, P. A. (2019). Multidimensional item response theory for factor structure assessment in educational psychology research. *Frontiers in Education, (4)* 45.

Ince Araci, F., & Tan, S. (2022). Multidimensional computerized adaptive testing simulations in R. *International Journal of Assessment Tools in Education, 9*(1), 118-137.

Issayeva, L. (2022). *The IRT item pseudo-guessing parameter.*

Jodoin, M.G., Keller, L.A., and Swaminathan, H. (2003). A comparison of Linear, Fixed Common Item and Concurrent Parameter Estimation Equating Procedures in Capturing Academic Growth.*Journal of Experimental Education*, 77(3), 229-25.

Kalender, I. (2012). Computerized adaptive testing for student selection to higher education. *Journal of Higher Education, 2*, 13-19.

Linacre, J. (2000).*Computer Adaptive Testing Methodology Whose Time has Com*. Seoul: Komesa Press.

Linacre, J. (2002). Optimizing rating scale category effectiveness. *Journal of Applied Measurement, 3*(1), 85-106.

Liu, Q., Zhuang, Y., Bi, H., Huang, Z., Huang, W., Li, J., Yu, J., Liu, Z., Hu, Z., Hong, Y., Pardos, Z. A., Ma, H., Zhu, M., Wang, S., & Chen, E. (2024). Survey of computerized adaptive testing: A machine learning perspective (arXiv:2404.00712). arXiv.

ISSN: 2537-0448 ***eISSN: 2537-0456***

Martin, A. J., & Lazendic, G. (2018). Computer-adaptive testing: Implications for students' achievement, motivation, engagement, and subjective test experience. *Journal of Educational Psychology*, *110*(1), 27-45.

Murphy, K and Davidshofer, C. (1994). *Psychological testing: Principles application*. (3ed). New Jersey: Prentice-Hall.

Ohiri, S. C.(٢٠٢٣). Application of Item Response Theory as a Modern Statistical Tool to Test Item Development and Analysis. *International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)*, *3*(1), 145 – 150.

Pett, M., Lackey, N., & Sullivan, J. (2003). *Making sense of factor analysis: The use of factor analysis for instrument development in health care researc*h.Sage Publications, Inc.

Reckase, M., D. (1997). The pest and future of multidimensional item response theory. *Applied Psychological Measurement*, 21(1), 25-36.

Schumacker, R. E., & Lomax, R. G. (2016). *A Beginner's Guide to Structural Equation Modeling* (4th ed.). Routledge.

Seo, D. G. (2017). Overview and current management of computerized adaptive testing in licensing/certification examinations. *Journal of Educational Evaluation for Health Professions, 14*, 17.

Su, Y. (2016). A comparison of constrained item selection methods in multidimensional computerized adaptive testing. *Applied psychological Measurement*, *40*(5), 346-360.

ISSN: 2537-0448 ***eISSN: 2537-0456***

Thompson, N. (2016). " *Dichotomous " Vs " Polytomous " in IRT* ?. https://assess.com/what-do-dichotomous-and-polytomous-mean-in-irt/

Thompson, N., & Weiss, D. (2011). A Framework for the development of computerized adaptive tests. *Practical Assessment, Research & Evaluation, 16*(1), 1-9.

Van Der Linden, W.J & Glas, C. A. (EDS) (2010) *Elements of Adaptive Testing*. New York, NY: Springer.

Veldkamp, B. P. (2005). Optimal test construction. In Encyclopedia of social measurement, 2 , 933-939

Wainer, H., Dorans, N.J., Eignor, D., Flaugher, R., Green, B. F., Mislevy, R. J., Steinberg L., & Thissen, D. (2000). *Computerized adaptive testing: A primer* (2nd ed). Mahwah, NJ: Lawrence Erlbaum.

Wang, J., & Wang, X. (2012). *Structural Equation Modeling: Applications* Using *Mplus.* Wiley.

Wang, K. (2017). *A fair comparison of the performance of computerized adaptive testing and multistage adaptive testing* (Doctoral dissertation, Michigan State University). ProQuest Dissertations Publishing.

Wang, X. (2013). *An investigation on computer-adaptive multistage testing panels for multidimensional assessment* (Doctoral dissertation). The University of North Carolina at Greensboro.

Weiss, D. (1983). *New horizonsin testing: Latent trait test theory and computerized adaptive testing*. MN: Assessment Systems Corporation.

ISSN: 2537-0448 ***eISSN: 2537-0456***

Weiss, D. J., & Sahin, A. (2024). *Computerized adaptive testing*: From concept to implementation. The Guilford Press.

Wulandari, F., Hadi, S., & Haryanto, H. (2020). Computer-based adaptive test development using fuzzy item response theory to estimate student ability. *Computer Science and Information Technology*, *8*(3), 66–73.

Yang, Y., Peng, Y., Li, W., Lu, S., Wang, C., & Zhong, J. (2023). Psychometric evaluation of the academic involution scale for college students in China: An application of Rasch analysis. *Frontiers in psychology*, *14*, 1135658.

ISSN: 2537-0448 *eISSN: 2537-0456*