

Computerized Adaptive Testing: Its Concept, Procedures, and the Importance of Its Implementation

الاختبارات التكيفية المحوسبة: مفهومها وإجراءاتها، وأهمية تطبيقها

alaakhtbaaraat attakeefya almahousba: mafhoumhaa wa'ijraa'athaa, wa'hmya
tatbeeqhaa

بهمية الراشدية¹، أ.د حسين الحروصي² ARRACHIDIA Bahia, ALKHEROSI Hussein

¹ جامعة السلطان قابوس (سلطنة عمان)، s42056@student.squ.edu.om

² جامعة السلطان قابوس (سلطنة عمان)، hussein5@squ.edu.om

تاريخ النشر: 2024/12/21

تاريخ القبول: 2024/12/6

تاريخ الاستلام: 2024/11/4

Abstract: The current article aims at exploring the concept of computerized adaptive tests (CAT) and its principles. It also examines the development procedures of CAT and its algorithms. Furthermore, the article highlights the main benefits of CAT for organizations, examinees and assessment systems such as measurement accuracy and efficiency and the reduction in test duration compared to traditional tests. Additionally, the article sheds light on the challenges facing CAT implementation, including high costs and the need for specialized expertise in its development, which necessitate solutions to expand its implementation. The article recommends adopting CAT by the assessment systems in the Ministry of Education and higher education institutions due to its advantages in saving time and improving measurement accuracy compared to traditional testing methods. CAT also enables teachers to identify students' strengths and weaknesses, facilitating the provision of appropriate academic support for each student.

Keywords: computerized; adaptive; tests; algorithms IRT

ملخص: يهدف المقال الحالي إلى التعرف على مفهوم الاختبارات التكيفية المحوسبة، والأسس التي تقوم عليها. كما يستعرض إجراءات بناءها والخوارزميات اللازمة لتطبيقها. كما يسلط المقال الضوء على أهم فوائد الاختبارات التكيفية المحوسبة للمؤسسات، وللمفحوصين ولأنظمة التقويم مثل دقة وكفاءة القياس، وتقليل مدة الاختبارات مقارنة بالاختبارات التقليدية. بالإضافة إلى ذلك فإن المقال يسلط الضوء كذلك على التحديات التي يواجهها تطبيق الاختبارات التكيفية المحوسبة مثل التكلفة العالية، والحاجة للخبرة التخصصية في بنائها، والتي تحتاج إلى إيجاد حلول للاستفادة القصوى، وإمكانية تطبيق مثل هذه الاختبارات. ويوصي المقال بتطبيق الاختبارات التكيفية المحوسبة في أنظمة التقويم في وزارة التربية والتعليم ومؤسسات التعليم العالي لما لها من مميزات في توفير الوقت ودقة القياس مقارنة بالاختبارات التقليدية، كما يمكن للمعلمين تحديد نقاط القوة والضعف لدى الطلبة مما يساعد على تقديم الدعم الأكاديمي المناسب لكل طالب.

الكلمات المفتاحية: اختبارات؛ تكيفية؛ محوسبة؛ خوارزميات؛ الاستجابة للمفردة.

1. المقدمة

تأثر مجال الاختبارات التربوية والنفسية بالتطور التكنولوجي، حيث أدى دمج التكنولوجيا في ممارسات التقويم إلى تطورات كبيرة في طريقة بناء الاختبارات وإدارتها وتصحيحها (Lim et al., 2024). فعلى سبيل المثال ظهر الاختبار المحوسب كأداة أكثر فعالية في تطبيق الاختبارات، مما وفر مرونة أكبر في تنسيق الأسئلة، والتصحيح الآلي، والحصول على النتائج بشكل أسرع (العمرى وعيادات، 2016). ومع هذا التقدم، ظل نظام التقويم المحوسب يستخدم اختبارات ثابتة لا يمكن أن تراعي الفروق الفردية بشكل كامل بين المفحوصين.

وساهم التقدم التكنولوجي كذلك في ظهور نظريات قياس متقدمة مثل نظرية الاستجابة للمفردة، والتي أدت إلى ظهور أساليب جديدة للاختبارات عوضاً عن الاختبارات الخطية سواء كانت اختبارات الورقة والقلم أو الاختبارات المحوسبة، ومنها الاختبارات التكيفية المحوسبة (الصبيحي وأحمد، 2011).

تمثل الاختبارات التكيفية المحوسبة تقدماً كبيراً في مجال التقويم، حيث أنها تقدم طريقة ديناميكية مفصلة للاختبار تهدف إلى تحسين الدقة والكفاءة في نتائج التقويم (نور الدين، 2019؛ Burr et al., 2023). تختلف الاختبارات التكيفية المحوسبة عن اختبارات الورقة والقلم أو الاختبارات الثابتة التي تقدم الأسئلة نفسها لكل مفحوص، فالاختبارات التكيفية المحوسبة تعمل على تكيف صعوبة ومحتوى الاختبار وفقاً لمستوى قدرة الفرد أثناء تنفيذ الاختبار، وتتيح الطبيعة التكيفية لهذه الاختبارات تقييماً أكثر دقة للقدرة باستخدام عدد أقل من الأسئلة مما يجعلها أكثر كفاءة من الاختبارات التقليدية (شواشرة والقرعان، 2021). ويشير أيبك وديميرتاسلي (Aybek & Demirtasli, 2017) أن منطق الاختبارات التكيفية يقوم على تجنب المفردات الصعبة جداً أو السهلة جداً للمفحوص، وإنما تختار المفردات التي تناسب قدرته، فالمفحوص يمكن أن يرتكب أخطاءً بلا مبالاة عندما يتعرض لاختبار سهل جداً بالنسبة لقدراته، أو أنه سيعتمد على التخمين عند الإجابة على الأسئلة التي لا يعرف إجاباتها. في هذه الحالة فإن الاختبارات الخطية التقليدية كاختبارات الورقة والقلم، ستعطي معلومات قليلة عن المفحوص، وبالتالي فإنه سيصبح من الصعب التمييز بين مجموعة أفراد أجابوا على جميع الأسئلة بشكل صحيح. بينما

يخضع المفحوص في الاختبار التكيفي المحسوب لمفردات تعطي معلومات أكثر عن مستوى قدرته، لذلك فإن الفروق بين المفحوصين ستحدد بشكل أوضح (فرحات، 2021). تستخدم الاختبارات التكيفية المحوسبة في الوقت الحالي على نطاق واسع في مجال التعليم، والاختبارات المهنية، والتقييمات النفسية، كما أنها أصبحت الطريقة المفضلة للاختبارات واسعة النطاق مثل اختبارات القبول الجامعي Graduate Record Examination، واختبارات رخصة التمريض NCLEX (الصبيحي وأحمد، 2011؛ نور الدين، 2021).

وعلى الرغم من الاهتمام العالمي الذي تحظى به الاختبارات التكيفية المحوسبة من قبل مؤسسات التقويم مثل الجمعية الدولية للاختبار التكيفي المحسوب International Association for Computerized Adaptive Testing (IACAT) (<http://www.iacat.org>)، التي تأسست بهدف تشجيع الجهود الأكاديمية من الاهتمام بتطوير وتطبيق الاختبارات التكيفية المحوسبة في مختلف المجالات وتطوير تكنولوجيا تطبيقها، وشركة أنظمة التقييم Assessment Systems التي توفر العديد من الخدمات لتطوير وتطبيق الاختبارات التكيفية المحوسبة مثل البرامج الالكترونية للمحاكاة، وتطوير بنوك الأسئلة وتحليل مفردات الاختبار (<http://www.assess.com>). وعلى الرغم من تطبيق الاختبارات التكيفية المحوسبة في عدة اختبارات دولية، إلا أن الاهتمام بهذا النوع من الاختبارات وتطبيقها - حسب علم الباحثين - لا زال محدودا في البيئة العربية عامة، وسلطنة عمان خاصة. كما أشارت دراسة العمري وعيادات (2016) أن تصورات أعضاء هيئة التدريس والطلبة حول الاختبارات المحوسبة في العملية التعليمية التعليمية بدرجة متوسطة، مما يشير للحاجة إلى التأكيد على أهمية وفوائد هذا النوع من الاختبارات. لذلك يسلط هذا المقال الضوء على أهمية الاختبارات التكيفية المحوسبة من خلال الإجابة عن الأسئلة التالية:

1. ما مفهوم الاختبارات التكيفية المحوسبة؟
 2. ما مراحل اختبارات التكيفية المحوسبة؟
 3. ما فوائد الاختبارات التكيفية المحوسبة؟
 4. ما التحديات التي تواجه الاختبارات التكيفية المحوسبة؟
 5. ما دور الذكاء الاصطناعي وتعلم الآلة في تطوير الاختبارات التكيفية المحوسبة.
- وتتمثل أهداف المقال في النقاط التالية:

1. التعريف بمفهوم الاختبارات التكيفية المحوسبة.
2. تحليل خطوات اعداد الاختبارات التكيفية المحوسبة.
3. عرض أهم فوائد وتحديات تطبيق الاختبارات التكيفية المحوسبة.
4. التعرف على دور الذكاء الاصطناعي في تطوير الاختبارات التكيفية المحوسبة. وتعتمد الورقة الحالية على المدخل الوصفي التحليلي للاختبارات التكيفية المحوسبة، ومراحل اعدادها، وأهم فوائدها، وتحدياتها.

2. مفهوم الاختبارات التكيفية المحوسبة

تعرف الاختبارات التكيفية المحوسبة بأنها أحد طرق التقويم التي تكيف مستوى صعوبة المفردات وفقاً لقدرات الطالب (Ebenbeck et al., 2024; Huda et al., 2024)، كما يعرفها قاسم (2021) بأنها "الاختبارات التي يقوم فيها الطلبة بالإجابة عن مفردة أو مجموعة مختلفة من المفردات، تطرح عليهم بناء على مستوى كل طالب، وبناء على إجاباتهم عن مفردة أو مجموعة من المفردات التي يختارها الحاسب الآلي لهم (ص. 147)". وتشير كلمة المحوسب إلى دمج علوم الحاسب الآلي مع القياس الذكي (Liu et al., 2024)، وبالمقارنة مع الاختبارات التقليدية؛ يوفر الاختبار التكيفي المحوسب فرصاً لتفصيل وتخصيص الاختبارات للأفراد باستخدام خوارزميات لتحديد مستوى صعوبة المفردات بشكل آلي بناء على أداء كل متعلم على مفردات سابقة مما يسهل التعرف على القدرة الحقيقية للمفحوص (Liu et al., 2024; Rice et al., 2022). فالطالب الذي يجيب إجابة صحيحة عن المفردة سينتقل إلى مفردة أصعب، بينما إذا أجاب إجابة خاطئة فإنه ستعرض عليه مفردة أقل صعوبة. أي أن هذا النوع من الاختبارات يُفصل لكل فرد على حدة؛ بحيث لا يتعرض المفحوص للمفردات الصعبة جداً أو السهلة جداً، وإنما تعرض عليه المفردات التي تتناسب مع مستوى قدرته فقط (الصبيحي والأحمدي، 2011). لذا فإن الاختبار التكيفي المحوسب يتيح قياساً أكثر دقة لمستوى قدرة كل طالب (Huda et al., 2024; Sorrel et al., 2020).

تساهم الاختبارات التكيفية المحوسبة في تحسين كفاءة ودقة نتائج التقييم (Liu et al., 2024; Wulandari et al., 2020)، حيث أنها تقدم المعلومات الأكثر أهمية في قياس قدرة المفحوصين. ويتطلب الاختبار التكيفي المحوسب عدداً أقل من البنود لتحقيق

الدقة في قياس مستوى قدرة المفحوص (Wulandari et al., 2020)، حيث أن الاختبار يستبعد المفردات الصعبة جداً أو السهلة جداً بالنسبة للمفحوص، وبذلك يتطلب عدداً أقل من المفردات للوصول إلى تقدير دقيق للقدرة، حيث تشير بعض الدراسات أن الاختبارات التكيفية المحوسبة يمكن أن تقلل عدد المفردات إلى 50 % مقارنة بالاختبارات الخطية التقليدية (Weiss & Şahin, 2024; Ayanwale & Ndlovu, 2024; Seo, 2017)، بينما وضع إيبينبيك وزملاؤه (Ebinbeck et al., 2024) أن طول الاختبار قد يقصر إلى 50-90%. كما أن هذا الاختبار بخلاف الاختبارات التقليدية يراعي الطلبة في الأطراف القصوى للمنحنى الطبيعي من ذوي القدرات العالية جداً أو المنخفضة جداً (Martin & Lazendic, 2018). لذلك فإنه يزيد من دافعية الطلبة، ويساهم في التقليل من قلق الاختبار لديهم، وقلة التركيز، والملل (Ebinbeck et al., 2024; Ayanwale & Ndlovu, 2024). (Ebinbeck & Gebhardt, 2022) خاصة لدى طلبة التربية الخاصة.

3. نظرية الاستجابة للمفردة

تستند أغلب الاختبارات التكيفية المحوسبة على نظرية الاستجابة للمفردة (IRT) (الصبيحي وأحمد، 2011؛ Ayanwale & Ndlovu, 2024; Thompson & Weiss, 2011)، وتهتم هذه النظرية بحساب احتمالية الإجابة الصحيحة لكل مفردة على حدة، حيث توضح العلاقة بين قدرة المفحوص واحتمال الاستجابة على الفقرة من خلال نماذج رياضية تصف احتمال الإجابة عن فقرة بدلالة قدرة المفحوص ومعالم الفقرة. وتتكون نماذج نظرية الاستجابة للمفردة على نماذج ثنائية الاستجابة (Dichotomous IRT Model)، تتضمن استجابتين للفقرة مثل أسئلة الاختيار من متعدد، ونماذج متعددة الاستجابة (Polytomous IRT Models) والتي تتطلب أكثر من استجابتين للسؤال مثل حل المشكلات الرياضية (Zanon et al., 2016).

وتفترض هذه النماذج أن السمة المقاسة هي خاصية من خصائص المفحوص حيث توجد علاقة منتظمة بين مستويات السمة المقاسة لدى مفحوصين مختلفين وبين احتمالات الاستجابة الصحيحة ل فقرات مختلفة (Ogunsakin & Shogbesan, 2018). وتتحدد خصائص الفقرات وفقاً لمنحنى خصائص الفقرة بدلالة معالم الصعوبة والتمييز

والتخمين حسب النموذج المعتمد كالاتي (الصبيحي والأحمدي، 2011؛ Aybek& Demirtasli, 2017؛ نور الدين، 2019) :

1. النموذج اللوجستي أحادي المعلم (نموذج راش) (One- Parameter Logistic Model):

هو النموذج الذي يكون بدلالة معلم صعوبة الفقرة فقط، وتشير صعوبة الفقرة إلى مدى صعوبة السؤال (ما هو مستوى القدرة الكامنة الذي يحتاجه الشخص ليكون لديه فرصة بنسبة 50% للإجابة الصحيحة على الفقرة؟) ويتم تقدير معلم الصعوبة لكل فقرات من فقرات الاختبار. ويفترض هذا النموذج أن جميع الفقرات متساوية في التمييز، وأن قيم التخمين لها تقترب من الصفر، والفقرات تتباين فيما بينها من خلال معلم الصعوبة. ويعد هذا النموذج أبسط نماذج الاستجابة للمفردة وأكثرها استخداما.

2. النموذج اللوجستي ثنائي المعلم (The Two- Parameter Logistic Model): يفترض هذا النموذج تباين الفقرات في مستويات الصعوبة والتمييز، بينما لا تتأثر بالتخمين، حيث أن قيم التخمين تقترب من الصفر.

3. النموذج اللوجستي ثلاثي المعلم (The Three- Parameter Logistic Model) وهو النموذج الذي يفترض أن فقرات الاختبار تتباين فيما بينها بدلالة ثلاثة معالم وهي: صعوبة الفقرة، وتمييزها بالإضافة إلى معلم التخمين الذي يحدد احتمال أن يجيب المفحوص ذو القدرة المنخفضة على فقرة ما إجابة صحيحة عن طريق التخمين، لذلك تتباين فقرات الاختبار في هذا النموذج من خلال مستوى صعوبتها، وتمييزها، وقيم التخمين عليها.

وتختلف نظرية الاستجابة للمفردة عن النظرية الكلاسيكية في الافتراضات التي تقوم عليها، من أجل الوصول إلى نتائج موثوقة. ومن أهم الافتراضات التي تقوم عليها هذه النظرية:

أ. افتراض أحادية البعد

تفترض نماذج نظرية الاستجابة للمفردة أحادية البعد وجود سمة كامنة واحدة تفسر أداء الفرد على الاختبار (الشقصي وآخرون، 2020). ويشير بيتشي وطالب (Bichi & Talib, 2018) أن أحادية البعد تعني أن المفردات تقيس

مجالاً واحداً فقط من القدرة أو المعرفة، ولا يشترط في أحادية البعد أن تكون المفردات مرتبطة إيجابياً مع بعضها، حيث أنه من الممكن أن ترتبط المفردات ارتباطاً سلبياً ومع ذلك تظل أحادية البعد.

ب. افتراض الاستقلال الموضوعي

ويعني أن إجابة المفحوص على فقرات الاختبار تكون مستقلة عن إجابته على أي فقرة أخرى، أي أن إجابة الفقرة لا تعطي تلميحات لإجابة فقرة أخرى، ويتطلب هذا الافتراض أنه باستثناء القدرة، فإنه لا توجد علاقة بين استجابات عناصر الاختبار غير العلاقة التي تحددها القدرة أو معلمات النموذج الأخرى (Bichi Talib, 2018). ويؤكد هذا الافتراض على أن قدرة المفحوص وخصائص الفقرة هي التي تؤثر على أدائه في الاختبار فقط (الشقصي وآخرون، 2020).

4. مراحل اعداد الاختبارات التكيفية المحوسبة

من الناحية البنائية، تمر عملية تطوير الاختبار التكيفي المحوسب بخمس مراحل أساسية تتضمن بناء بنك الأسئلة المعايير والذي يعبر عن محتوى الاختبار، أما المراحل الأربعة المتبقية فتتعلق جميعها بالقياس النفسي لا بالمحتوى ذاته وهي ترتبط بخوارزميات نظام الاختبار التكيفي المحوسب (Thompson & Weiss, 2011)، وتضاف إلى هذه المراحل مرحلة مهمة قبل البدء بالتصميم الفعلي للاختبار وهي مرحلة دراسة الجدوى من الاختبار التكيفي المحوسب. وفيما يلي تفصيل لهذه المراحل:

1.4 المرحلة الأولى: دراسات الجدوى والتخطيط وقابلية التخطيط

قبل اتخاذ قرار بناء وتطبيق اختبار تكيفي محوسب، يجب دراسة القرار وجدوى استخدام هذا الاختبار. نظراً للمزايا التي توفرها الاختبارات التكيفية المحوسبة، فإن بعض المختصين بالتقييم قد يتحمسون لتطبيق الفكرة دون الإلمام الكافي بها وبمتطلبات تطويرها. إن اتخاذ مثل هذا القرار مع ما يوفره من مزايا قد ينطوي على مخاطر محتملة لذلك كان من المهم أن يستند على دراسة للاعتبارات التجارية والعملية المتعلقة بتطبيق الاختبارات التكيفية. ويجب على متخذ القرار أن يجيب على أسئلة مهمة مثل Weiss (

& Shain, 2024; Thompson & Weiss, 2011)

- هل لدى المؤسسة الخبرة النفسية والتقنية الكافية لإدارة اختبار تكيفي محوسب؟ وإذا لم تكن لديها؛ هل ستكون قادرة ماليا على توفير تكاليف خبير خارجي؟
- هل تمتلك المؤسسة الميزانية الكافية لتطوير بنوك أسئلة وشراء البرامج اللازمة لإدارة الاختبار التكيفي المحوسب؟
- هل تمتلك المؤسسة منصة لإدارة الاختبار التكيفي المحوسب، أم تحتاج لتطوير منصة خاصة بها؟
- هل ستؤدي الاختبارات التكيفية المحوسبة إلى تقليل طول ومدة الاختبار المتوقع وتحسين دقة الدرجات؟

ويمكن الإجابة على معظم هذه الأسئلة عن طريق دراسات المحاكاة التي تتبع عدة طرق. وتعد دراسات المحاكاة خطوة أساسية في تطوير الاختبار التكيفي المحوسب الذين يكون قابلا للتطبيق ويلبي احتياجات المؤسسة. وتعد محاكاة مونت كارلو Monte Carlo هي النهج الأكثر شمولية، وغالبا ما تستخدم في المراحل الأولى من تطوير الاختبار التكيفي المحوسب، وذلك لأنها لا تتطلب بيانات فعلية، سواء عن الأسئلة أو المفحوصين رغم أن وجود بيانات فعلية - إذا توفرت - سيكون مفيدا، مما يجعلها مفيدة جدا في تقييم ما إذا كان الاختبار التكيفي المحوسب قابلا للتطبيق في المؤسسة قبل استثمار أي أموال في تصميمه وتطبيقه (Thompson & Weiss, 2011).

ويشير عمارة وكاظم (2020) أن طريقة مونت كارلو تمكن المهتمين بتطوير الاختبارات التكيفية المحوسبة من تقدير طول الاختبار ودقة تقديرات القدرة في الاختبار التكيفي، كما تساعد في تحديد أنسب الطرق لانتقاء مفردات الاختبار، ومحكات إنهاء الاختبار، بالإضافة إلى أن هذه الطريقة تساعد في تحديد الحجم المناسب لبنك الأسئلة، والتحكم في معدل التعرض للمفردات.

2.4 المرحلة الثانية: تطوير بنك الأسئلة

يتطلب تنفيذ الاختبارات التكيفية المحوسبة اعداد بنك أسئلة يحتوي على مجموعة كبيرة من المفردات معلومة البارامترات، مما يتيح استخدام صورا مختصرة من الاختبارات تضم عددا قليلا من المفردات في كل مرة تناسب مستوى الطلبة، وتحقق أهداف القياس (فرحات، 2021). ويشير ثومبسون وويس (Thompson & Weiss, 2011)، وقاسم

(2021) إلى أنه عند بناء بنك الأسئلة يجب التركيز على أن تتطابق دالة معلومات الاختبار مع الهدف من الاختبار، فإذا كان الاختبار سيوظف لتصنيف الطلبة بناء على درجة قطع (ناجح/راسب)، فإنه لا بد أن يتضمن على العديد من المفردات التي تقترب من هذه الدرجة، بينما إذا كان الهدف من الاختبار تصنيف الطلبة إلى عدة مستويات، فإنه من الضروري اعداد مفردات عالية، ومتوسطة، ومنخفضة الصعوبة بحيث تغطي جميع مستويات الطلبة. لذلك ينبغي التخطيط جيدا لبناء بنك المفردات. ويؤكد فرحات (2021) أنه يجب كتابة مجموعة كبيرة من المفردات الاختبارية التي تقيس مدى واسعا من القدرة، وهذا يتطلب تحديد مواصفات كل مفردة، والمحتوى الذي تقيسه، ونطاق المهارات التي تمثلها، وكلما كانت هذه المواصفات واضحة، أدى ذلك إلى بناء عدد كاف ومتنوع من المفردات التي تتباين في صعوبتها على مدى السمة المراد قياسها.

لذلك عند التخطيط لبناء بنك المفردات ينبغي اعداد جدول مواصفات الاختبار وخريطة الاختبار، حيث يوضح جدول مواصفات الاختبار وزن الاختبار ونوع الأسئلة التي سيتم استخدامها، بينما تمثل خريطة الاختبار المخطط التفصيلي الذي يتم فيه تمثيل مجال المحتوى بعدد الأسئلة والمحتوى المعرفي الذي سيتم اختباره. ويمكن استبدال جدول مواصفات الاختبار بجدول مواصفات بنك الاختبار ليحقق نفس الهدف. يقدم هذا الجدول المشابه لجدول مواصفات الاختبار، نظرة عامة على بنك أسئلة الاختبار التكيفي، بما في ذلك المجالات المعرفية، الأهداف، أنواع الأسئلة، والمستويات الإدراكية. ويساعد في توجيه مؤلفي الأسئلة ومطوري الاختبارات من خلال تقديم هيكل تمهيدي. ويعتمد مستوى التفاصيل في هذا الجدول على احتياجات المؤسسة، ويغطي أسئلة حول كيفية التصحيح، وزن المحتوى، ونوع الاختبار التكيفي المستخدم (مثل أحادي البعد أو متعدد الأبعاد) (Weiss & Şahin, 2024).

بعد اكتمال مرحلة التخطيط لبنك الأسئلة، تأتي مرحلة كتابة المفردات والتي تعد مرحلة مهمة تتطلب اهتماما دقيقا، حيث أن بنك الأسئلة هو الجانب الأكثر أهمية في الاختبارات التكيفية المحوسبة، ويمثل العامل الرئيسي في نجاح تنفيذ الاختبار. ويجب مراعاة ثلاثة جوانب في عملية كتابة المفردات من أجل انتاج مفردات عالية الجودة، وهي: التعليمات التي ستقدم للمفحوصين، وصياغة جذر السؤال، وخيارات الإجابات. بعد ذلك يتم مراجعة

المفردات من قبل مختصين للتأكد من الصياغة اللغوية، وتمثيل المفردات للمحتوى، وتغطيتها لمستويات معرفية مختلفة. ويجب الانتباه لكتابة عدد كبير من المفردات بناء على الهدف من الاختبار ومدة تطبيقه (قاسم، 2021؛ Weiss & Şahin, 2024).

3.4 المرحلة الثالثة: التطبيق القبلي وتحليل النتائج باستخدام نظرية الاستجابة للمفردة
عند الانتهاء من عملية بناء مفردات الاختبار، يجب أن يتم تجريب المفردات على عينة استطلاعية، لمطابقتها مع مستوى المفحوصين بناء على البارامترات الخاصة بالمفردة بناء على نظرية الاستجابة للمفردة (قاسم، 2021؛ Weiss & Şahin, 2024). ويتحدد حجم العينة الاستطلاعية على أساس نموذج نظرية الاستجابة للمفردة الذي سيتم توظيفه للمفردات (Weiss & Şahin, 2024; Thompson & Weiss, 2011). وبعد التطبيق يتم تقدير بارامترات المفردات الاختبارية عن طريق استخدام برنامج معايرة لنظرية الاستجابة للمفردة. ويقترح ويس وشاهين (Weiss & Şahin, 2024) البارامترات في الجدول (1) لضمان جودة المفردات في بنك الأسئلة لنموذج ثلاثي المعلم (3PL).
جدول (1). قيم معالم المفردات لاختبار مفردات بنك الأسئلة

المعلم	مفردة جيدة	مفردة مقبولة	مفردة ضعيفة
معلم التمييز (a)	أعلى من 1.5	بين 0.49 إلى 1.49	أقل من 0.49
معلم الصعوبة (b)	بين 1.99 إلى 1.99	بين 2 إلى 3	أقل من 3 - أعلى من 3
معلم التخمين (c)	بين 0 إلى 0.24	بين 0.25 إلى 0.39	أعلى من 0.4

المصدر: Weiss & Şahin, 2024, P.155

بينما يقترح قاسم (2021) أن معلم التمييز يفضل ألا يقل عن 0.8، ويفضل أن تنتوز قيم معلم الصعوبة بين 3- إلى 3، وينبغي ألا تزيد قيمة معلم التخمين عن 0.3. وبناء على البيانات الإحصائية المتعلقة بالمفردات يتم تحديد المفردات التي يجب حذفها أو مراجعتها أو إعادة اختبارها (Thompson & Weiss, 2011). كما يجب تحليل الأبعاد في هذه المرحلة، حيث تقترض نظرية الاستجابة للمفردة أن الاختبار يكون أحادي البعد، لذا لا بد من تحليل مفردات الاختبار عاملياً لتحقيق ذلك، ولا بد كذلك من التحقق من الاستقلال الموضوعي للمفردات (Aybek & Demirtasli, 2017). ومن المهم أيضاً التأكد من تدرج مفردات الاختبار وفق أحد نماذج نظرية الاستجابة للمفردة، كما أنه لا بد من التأكد أن مفردات البنك تنتوز وفق مستوى صعوبتها على متصل القدرة (الصبيحي والأحمدي، 2011).

المرحلة الرابعة: تحديد المواصفات النهائية لتصميم الاختبار التكيفي المحسوب

قبل البدء في تطبيق ونشر الاختبار التكيفي المحسوب، من المهم تحديد الخوارزميات التي يعمل بناءً عليها الاختبار. لذلك لابد من تحديد كيف سيتم انتقاء المفردات لكل مفحوص، وكيف سيتم تقدير قدرته، ومتى سيتم إيقاف الاختبار، بالإضافة إلى ذلك فإنه لابد من الأخذ بالحسبان معدل عرض المفردات. وسوف يتم استعراض محكات الاختبار التكيفي المحسوب في الجزء القادم.

1.3.4 محك بداية الاختبار

لتحديد بداية الاختبار للمفحوصين، هناك عدة طرق يمكن اللجوء إليها وتعتمد على توفر معلومات سابقة عن المفحوص من عدمها. ففي حالة وجود معلومات سابقة عن المفحوص تتعلق بالسمة التي يقيسها الاختبار التكيفي، فيمكن استخدام هذه المعلومات كقيمة تقديرية أولية لقدرة المفحوص، كاستخدام نتائج التقويم التكويني كمحكات لبداية الاختبار، كما يقترح ثومبسون وويس (Thompson & Weiss, 2011)، وقاسم (2021) أنه من الممكن انتقاء محك بداية الاختبار بناءً على عوامل خارجية مثل دافعية الطالب أو مستواه الدراسي السابق، أو الاعتماد على رأي المعلمين. أما في حالة عدم توفر معلومات سابقة عن المفحوصين، فيمكن تحديد بداية الاختبار لجميع الفحوصين بعرض مفردة متوسطة الصعوبة، يبلغ متوسط صعوبتها (صفر)، أو أن يتم الاختيار عشوائياً للمفردات التي تتراوح قيمتها مثلاً من (-0.5) إلى (0.5).

2.3.4 محك انتقاء المفردات

يحتاج المفحوصون في الاختبارات التكيفية المحوسبة إلى عدد قليل من المفردات التي تتناسب مع قدراتهم للحصول على تقييم دقيق، على عكس الاختبارات التقليدية التي تقدم مجموعة ثابتة من المفردات لجميع المفحوصين (Oladele & Ndlovu, 2021)، لذلك لا بد أن يغطي اختيار الأسئلة أكبر جزء من المجال الذي يتم قياسه. وقد أشار هان (Han, 2018) إلى أنه من المهم أن تركز عملية اختيار المفردات على ثلاث مكونات أساسية وهي اختيار المفردات، والتحكم في معدل عرض المفردات، وتحقيق التوازن في المحتوى. ويعتمد اختيار المفردات على مدى توفيرها للمعلومات المطلوبة مع ملاءمتها لمستوى المفحوص. وتستند قاعدة اختيار المفردات على اختيار المفردة

التالية التي سيتم تقديمها للمفحوص والتي توفر أكبر قدر من المعلومات بناء على التقدير الحالي لمستوى قدرة المفحوص (Weiss & Şahin, 2024).

وتوجد عدة طرق لاختيار المفردة تتباين في سهولة وتعقيد إجراءات تطبيقها، ومن هذه الطرق أسلوب أقصى المعلومات (Maximum Information Method)، والأسلوب البييزي (Bayesian Item Selection)، وأسلوب كولباك-ليبيلر (Kullback- Leibler)، وطريقة الاحتمالية العظمى الموزونة (Maximum Likelihood Weighted Information) (Aybek & Demirtasli, 2017). ومن أكثر الطرق شيوعاً لانتقاء المفردات هي أسلوب أقصى المعلومات، والأسلوب البييزي (قاسم، 2021)، وفيما يلي توضيح لها:

• طريقة المعلومات القصوى (Maximum Information Method)

وتعد هذه الطريقة أكثر الطرق استخداماً في قواعد انتقاء مفردات الاختبار التكيفي، وتعتمد على انتقاء المفردات التي تقدم المستويات القصوى لمعلومات المفردة عند المستوى الحالي لقدرة المفحوص (Cheng et al., 2017) أي أنها تنتقي أفضل المفردات ذات أكبر قيمة من المعلومات في ضوء مستوى القدرة الحالي للمفحوص، ويستخدم تقدير قدرة المفحوص في اختيار المفردة التالية. وتتميز هذه الطريقة بأنها تتيح مسارات متعددة ومختلفة للاختبارات، كما أنها تسهم في التقدم السريع نحو المدى الصحيح لصعوبة المفردة التي تقدم للمفحوص (قاسم، 2021).

• الطريقة البييزية (Bayesian Method)

تستخدم هذه الطريقة أسس ومبادئ نظرية بييز في انتقاء مفردة لم يسبق للمفحوص أن اختبر بها، بحيث يكون تباين التوزيع القبلي (Bayesian Posterior Variance) لتقدير مستوى قدرة المفحوص بعد استجابته على المفردة أقل ما يمكن، فتقوم هذه الطريقة بحساب التباين البعدي المتوقع لمستوى قدرة المفحوص لجميع المفردات التي لم تقدم له في حالة أجاب إجابة صحيحة أو خاطئة، ثم يتم حساب المتوسط للتباين البعدي المتوقع لكل مفردة في هذه الحالتين، والمفردة التي تعطي أقل تباين متوقع هي المفردة التي تقدم للمفحوص (قاسم، 2021). وهناك بعض القيود التي يمكن إضافتها لطرق انتقاء المفردة وذلك لضمان تمثيل المحتوى بصورة جيدة، وضمان عدم تكرار عرض مفردات محددة لجميع المفحوصين ومن هذه القيود (قاسم، 2021؛ Veldkamp & Matteucci, 2013):

- توازن محتوى المفردات، بحيث لا يتم انتقاء المفردات الملاءمة لمستوى الطالب بناء على خصائصها السيكومترية فقط حسب محك انتقاء المفردات، وإنما يراعى توازن المحتوى الذي تقيسه المفردات كذلك، كأن يتم تحديد نسب مئوية لعدد المفردات في كل مجال من مجالات المحتوى، أو أن يتم استخدام عناقيد الاختبارات، بحيث يمثل كل عنقود مجموعة من المفردات المرتبطة بمجال واحد، ويحتوي على عدد ثابت من المسارات لا تتداخل مع بعضها.
- التحكم في معدل عرض المفردات وهو جانب مهم اضافته مع طرق اختيار المفردة، ففي حالة البدء بقيمة ثابتة لجميع المفحوصين، فإنه من الممكن أن تكون المفردات الأولى للاختبار معروفة للطلبة، كما أنه إذا تم تطبيق قياسات متعددة على الاختبار نفسه، فإنه قد تظهر المفردات نفسها للمفحوص في حالة اتساق أدائه، بالإضافة إلى أنه في حالة اختيار المفردة بطريقة أقصى دالة معلومات، فإن ذلك سيؤدي إلى تكرار تطبيق بعض المفردات، بينما ستوجد مفردات أخرى لن يتم تطبيقها، مما يؤثر على سرية الاختبار، لذلك لابد من التحكم في معدل عرض المفردات بهدف التحقق من استخدام المفردات الموجودة في بنك الأسئلة بكفاءة مع التحقق من عدم تكرار عرضها كثيرا.

3.3.4 محك تقدير القدرة

تعتبر عملية تقدير قدرات المفحوصين عملية مهمة في الاختبارات التكيفية المحوسبة، فالتقدير غير الدقيق للسمة أو القدرة يؤثر في اختيار المفردات، وإنهاء الاختبار، والتقدير النهائي للسمة. وهناك طريقتان لتقدير قدرة المفحوص هما طريقة الأرجحية العظمى (Maximum Likelihood Estimation)، والطرق البييزية (Bayesian Methods) (الصبيحي والأحمدي، 2011؛ قاسم، 2021؛ Weiss & Şahin, 2024).

• طريقة الأرجحية العظمى (Maximum Likelihood Estimation)

تعتمد هذه الطريقة على نمط إجابة المفحوص لمجموعة من المفردات ذات الدرجات الثنائية التي إجابتها (صفر، 1) في تقدير قدرة المفحوص، وهي تمثل الحد الأقصى لدالة الاحتمال، ويمكن الحصول على الأخطاء المعيارية لتقديرات القدرة من خلال أعداد كافية من المفردات من خلال دالة المعلومات، وتشير هذه الدالة أنه إذا كان

مستوى قدرة المفحوص أعلى من متوسط صعوبة المفردة فإن التحيز يكون موجبا، أما إذا كان أقل فإن التحيز يكون سالبا، أما إذا كانت جميع المفردات تتناسب مستوى قدرة المفحوص فإن التحيز يقترب من الصفر. ولكن من عيوب هذه الطريقة أنها لا تستطيع تقدير قدرة المفحوص إذا أجاب إجابة صحيحة على جميع المفردات أو أجاب إجابات خاطئة على جميعها. ويمكن اللجوء في هذه الحالة للطرق البييزية.

• الطرق البييزية (Bayesian Methods)

يتم في هذه الطرق تقدير القدرة بناء على دالة القدرة، حيث أنها تعتمد على المعلومات السابقة عن توزيع قدرات المفحوصين، وتفترض هذه الطرق أن متوسط توزيع قدرة المجتمع يساوي (صفر)، بينما التباين يساوي (1). وتقوم هذه الطرق بضرب دالة الاحتمالية في توزيع قبلي مفترض، ينتج عنه توزيعا لاحقا ذا قمة واضحة وله قيمة قصوى، بغض النظر عن عدد المفردات التي لها نفس دالة الاستجابة. وهناك طريقتان شائعتان لتقدير قيمة القدرة باستخدام الطرق البييزية وهي طريقة التقدير البعدي الأقصى (Maximum A Posteriori MAP) وطريقة التقدير البعدي المتوقع (Expected A Posteriori EAP). تعمل طريقة التقدير البعدي الأقصى بطريقة مشابهة لطريقة الأرجحية العظمى من خلال العثور على القيمة العظمى في التوزيع البعدي الناتج عن دمج التوزيع القبلي المفترض مع دالة الاحتمالية، وعلى النقيض فإن طريقة التقدير البعدي المتوقع فتحسب التوقع أو المتوسط لدالة الاحتمالية المعدلة باعتبارها تقديرا للقدرة.

4.3.4 محك إنهاء الاختبار

تعتمد عملية التوقف في الاختبار التكيفي المحسوب على تقدير قدرة المفحوص التي تتم عند استجابته لمفردات الاختبار، وبناء على المفردات التي استجاب عليها يتم تقدير القدرة، ويمكن اعتماد محك التوقف في الاختبار التكيفي باعتماد طول ثابت للاختبار وذلك بإعطاء جميع المفحوصين نفس عدد المفردات، ولا يفضل استخدام هذا المحك لأن أهم ما يميز الاختبار التكيفي عن الاختبارات التقليدية هو ديناميكية ومرونة اختيار الفقرات وفقا لمستوى المفحوص (الشهري، 2023).

أما المعيار الأكثر شيوعاً هو الطول المتغير للاختبار ويعتمد إما على قاعدة الخطأ المعياري الأدنى، حيث تعتمد هذه الطريقة على تصميم الاختبار بحيث ينتهي عندما يصل إلى مستوى خطأ معياري معين أو مستوى دقة مكافئ، فعلى سبيل المثال ينتهي الاختبار عندما يصبح الخطأ المعياري 0.25 أو أقل (Weiss & Şahin, 2024). ويتميز هذا المعيار بتوفير درجات مكافئة دقيقة لجميع الطلبة، إذا افترضنا أن بنك الأسئلة تم بناءه بطريقة صحيحة، أما القاعدة الثانية فهي الإنهاء بأدنى معلومات حيث يتوقف الاختبار عندما لا يكون هناك مفردات متوفرة قادرة على توفير حد أدنى محدد مسبقاً من المعلومات لقدرة المختبر، وتتميز هذه القاعدة في أن المختبر لا يتعرض لمفردات غير ضرورية لا تحقق دقة قياس مرتفعة، بينما يتمثل عيب هذه الطريقة في أنها تقدم دقة قياس أقل من قاعدة الخطأ المعياري بسبب عدم توفيرها مفردات إضافية عندما تكون المعلومات المتاحة غير كافية (فرحات، 2021؛ Weiss & Şahin, 2024).

5. الدراسات السابقة

نشرت العديد من الدراسات التي طورت اختبارات تكيفية محوسبة، وفحصت دقة النتائج مع تطبيق خوارزميات مختلفة لبدء الاختبار وانهاؤه، وتقدير القدرة. فقد هدفت دراسة إيبينيك وجيبارت (Ebenbeck & Gebhardt, 2022) إلى تطوير اختبار تكيفي محوسب للطلبة ذوي الاحتياجات الخاصة، حيث هدف الباحثان إلى تطوير اختبارات تكيفية تعتمد على نظرية الاستجابة للمفردة لتقليل زمن الاختبارات مع تحسين دقتها. كما سعت الدراسة إلى البحث عن الخصائص المطلوبة لبنك الأسئلة للاختبار التكيفي المحوسب ليكون فعالاً في مجال التربية الخاصة، بالإضافة إلى فحص أثر قواعد التوقف المختلفة لإنهاء الاختبار على أداء الاختبار فيما يتعلق بتحقيق التوازن بين طول الاختبار ودقته. ولتحقيق الهدف من الدراسة، استخدم الباحثان بنوك أسئلة تم بناءها مسبقاً لمادتي الرياضيات والفهم القرائي، حيث تكون بنك الأسئلة الأصلي للرياضيات من 80 سؤالاً تم بناءها للطلبة الذين تتراوح أعمارهم بين 8-12 سنة، والذين يمتلكون مهارات الحساب الأساسية، بينما تكون بنك الأسئلة الأصلي للفهم القرائي من 219 سؤالاً، واعتمدت الدراسة على نموذج راش حيث تم معايرة الأسئلة بناءً على صعوبتها واحتمال أن يجيب الطالب عليها بشكل صحيح.

أجريت محاكاة مونت كارلو لفحص أداء قواعد التوقف المختلفة، حيث تم استخدام المحاكاة لتحديد عدد الأسئلة المناسبة لتحقيق مستوى مقبول من الدقة وفقاً لقيم الخطأ المعياري التي تم تحديدها. واختبرت المحاكاة ثلاث قيم للخطأ المعياري، وتم الوصول إلى مستويات دقة عالية مع القيمة الأقل (0.3)، ولكنها تطلبت مدة اختبار أطول وصلت إلى 84 سؤالاً، لذلك لتحقيق التوازن بين طول الاختبار ودقته، اختار الباحثان قيمة خطأ معياري (0.5) كقاعدة توقف. وبذلك تم تقليل متوسط طول اختبار الرياضيات إلى 15 سؤالاً بمتوسط مدة زمنية قدرها 3 دقائق، وتم تقليص طول اختبار الفهم القرائي إلى 18 سؤالاً، مع متوسط زمن إنهاء الاختبار 3 دقائق كذلك.

وأظهرت نتائج المحاكاة كفاءة الاختبار التكيفي في تقييم الطلبة ذوي الاحتياجات الخاصة، حيث يقدم اختبارات مخصصة وفقاً لأداء كل طالب، كما أنه يقلل من إجهاد الطلبة ويزيد من تفاعلهم بسبب تقليل مدة الاختبار.

هدفت دراسة هوانج وآخرون (Huang et al., 2022) إلى إعداد اختبار تكيفي محوسب لقياس الكفاءة في اللغة الإنجليزية لطلبة الجامعات في تايوان، وقد تم تحديد محتوى الاختبار ليشمل مهارتي القراءة والاستماع. ولتطوير بنود الاختبار تم وضع جدول مفصل لمواصفات الاختبار، كما تم تحديد واضعي مفردات الاختبارات والمراجعين للمفردات لاحقاً، وتم تنفيذ ورش عمل للتدريب على كتابة الأسئلة. وتكون بنك الاختبار من 567 مفردة تقيس مهارة القراءة، و172 مفردة تقيس مهارة الاستماع. تم تطبيق الاختبار على 21.330 طالباً من ست جامعات في تايوان، حيث أجاب كل طالب على 20 سؤالاً في كل مهارة. وتم استخدام نموذج راش لتحليل البيانات. وتم حذف الأسئلة الضعيفة ليتكون بنك الأسئلة في النهاية من 567 سؤالاً لمهارة القراءة، و172 سؤالاً لمهارة الاستماع. للبدء في الاختبار، تم تحديد قيمة ثابتة ليبدأ كل طالب بالسؤال ذو الصعوبة المتوسطة، وتم استخدام طريقة المعلومات القصوى (MFI) لانتقاء الأسئلة، بينما تم استخدام الطريقة البيزية (التقدير البعدي المتوقع EAP) لتقدير قدرة الطالب. ولإنهاء الاختبار، تم استخدام طريقة الطول الثابت، حيث يتوقف الاختبار عند اكمال الطالب 40 سؤالاً موزعة بالتساوي بين مهارتي القراءة والاستماع بغض النظر عن دقة التقدير.

أظهرت نتائج الدراسة أن الاختبار التكيفي المحوسب يقيس بشكل فعال الكفاءة اللغوية التي حددت في بنية الاختبار، حيث أظهرت النتائج أن الارتباط بين الكفاءة المقدرة والكفاءة الحقيقية كان مرتفعا حيث بلغ 0.93. وتمكن النظام التكيفي المحوسب من تقديم تقييما دقيقا للكفاءة في اللغة الإنجليزية مع تقليل مدة الاختبار.

وهدفت دراسة ستبانيك ومارتينكوفا (Štěpánek & Martinková, 2020) إلى تقييم فعالية وجدوى الاختبارات التكيفية المحوسبة من خلال استخدام محاكاة مونت كارلو والمحاكاة البعدية للبيانات، كما سعت الدراسة إلى الكشف عن كيفية تحسين الاختبار التكيفي المحوسب لطول الاختبار مع تحقيق تقديرات دقيقة لقدرات الطلبة. استخدمت الدراسة بيانات حقيقية من اختبار قبول في كلية الطب تكون من 100 سؤال في علم الأحياء موزعة على ثلاث مجالات، ولتقدير معلومات صعوبة الأسئلة وقدرتها التمييزية، تم استخدام نموذج نظرية الاستجابة للمفردة ثنائية المعالم. ينتهي الاختبار عند تحقيق مستوى محدد من الدقة (الخطأ المعياري أقل من أو يساوي 0.3).

في المحاكاة البعدية تم محاكاة أداء الطلبة الفعليين في نسخة تكيفية من الاختبار بناء على بياناتهم السابقة من الاختبار الثابت، وتم تقييم التغييرات في طول الاختبار ودقة التقديرات. بينما قامت محاكاة مونت كارلو بتوليد استجابات ممتحنين افتراضيين بقدرات تم توزيعها توزيعا طبيعيا.

أظهرت نتائج تحليل بيانات المحاكاة البعدية، ومحاكاة مونت كارلو أن الاختبار التكيفي المحوسب يمكن أن يقلل من عدد الأسئلة المطلوبة مع الحفاظ على تقديرات دقيقة للقدرة. فقد أظهرت التحليل البعدي أن متوسط عدد الأسئلة كان 40 سؤالا من أصل 100 سؤال، بينما أظهرت محاكاة مونت كارلو أن الاختبار التكيفي المحوسب يمكن أن يقلل من طول الاختبار بنسبة تصل إلى 75% (أي ما يعادل 25 سؤالا). كما أظهرت محاكاة مونت كارلو أن الاختبار التكيفي المحوسب يحقق دقة عالية في تقدير قدرات الطلبة حتى مع استخدام عدد أقل من الأسئلة، حيث بينت نتائج المحاكاة أن العلاقة بين القدرات الحقيقية المولدة والقدرات المقدرة في الاختبار التكيفي المحوسب كانت قوية جدا ($P=0.96$)، مما يؤكد على فعالية الاختبار التكيفي.

توضح الدراسات السابقة التي تم استعراضها أن الاختبار التكيفي المحوسب يتميز بالدقة في تقدير نتائج الطلبة، مع تقليل وقت الاختبار وعدد الأسئلة اللازمة للوصول إلى التقدير الحقيقي لقدرة الطالب.

6. فوائد الاختبارات التكيفية المحوسبة

توفر الاختبارات التكيفية المحوسبة عدة مميزات مقارنة بالاختبارات الخطية التقليدية، وتتوزع هذه الفوائد بين فوائد سيكومترية، وفوائد للمفحوص، وفوائد تنظيمية.

• دقة القياس

نظراً لأن الاختبار التكيفي يختار مجموعة من المفردات لكل مفحوص، ويتكيف مع مستوى السمات أثناء إجراء الاختبار، فإنه يوفر دقة قياس عالية لجميع المفحوصين بغض النظر عن مستوى السمات لديهم (Weiss & Sahin, 2024)، ويؤكد فرحات (2021) أن الاختبارات التكيفية عادة تكون أقصر من الاختبار التقليدي بنسبة قد تزيد على (50%)، ولا يؤثر ذلك على الخصائص السيكومترية. كما أظهرت دراسة إيبينبيك وجيبارت (Ebenbeck & Gebhardt, 2022)، وهوانج وآخرون (Huang et al, 2021) أن الاختبار التكيفي المحوسب يمكن أن يقلل الوقت اللازم للاختبار بنسبة تصل إلى 50% مع الحفاظ على دقة النتائج.

• تحسين كفاءة الاختبار

يجعل الاختبار التكيفي عملية الاختبار أكثر كفاءة من خلال استخدام وقت الاختبار والمفردات بشكل أكثر فعالية، حيث يتم تقديم أسئلة لكل مفحوص تكون لديهم فيها احتمالية متوقعة للإجابة الصحيحة تبلغ حوالي (0.5)، لأن الأسئلة يتم اختيارها من بنك أسئلة تم معايرته مسبقاً، بحيث تكون درجة صعوبتها أقرب إلى مستوى القدرة/ السمة المقدرة للمفحوص أثناء الاختبار، ومن بين تلك الأسئلة، يتم اختيار الأسئلة الأكثر تمييزاً عند ذلك المستوى من القدرة (Weiss & Şahin, 2024)

• كفاءة استخدام وقت الاختبار

نظراً لأن الاختبارات التكيفية المحوسبة تحتوي على عدد قليل من المفردات اللازمة لتحقيق الدقة المقبولة، فإنها توفر الوقت المستغرق في الاختبار بشكل كبير. كما أن المفحوص لا يتقيد بالوقت الذي يستغرقه زملاؤه للانتقال إلى القسم

التالي، فهو يستجيب وفقا لسرعته المناسبة لقدرته (فرحات، 2021). كل هذه العوامل تساهم في تقليل التعب والجهد الذي يبذله المفحوص أثناء الاختبار، حيث أشارت نتائج الدراسات إلى أن الاختبار التكيفي المحوسب يقلل من القلق لدى الطلبة مقارنة بالاختبارات التقليدية، حيث أنه يقدم أسئلة تتناسب مع مستوى الطالب مما يحفز للأداء بشكل أفضل (Huang et al., 2021)، كما أنه يقلل من الإجهاد الذي تسببه الاختبارات التقليدية مما يعطي الطلبة الدافع للاستمرار وخاصة للفئات من ذوي الاحتياجات الخاصة كما أشارت إليها نتائج دراسة إينبيك وجيبارت (Ebenbeck & Gebhardt, 2022).

• تقليل التخمين

يساعد الاختبار التكيفي الموحد في تقليل تأثير التخمين الذي يمثل مشكلة كبيرة في الاختبارات الموحدة (Akhtar & Kovacs, 2024)، خاصة أن معظم الاختبارات الموحدة تستخدم أسئلة بنمط الاختيار من متعدد. فبفضل النموذج اللوحيسي ثلاثي المعلومات في نظرية الاستجابة للمفردة، يتم حساب معلم التخمين لكل مفردة، وتؤخذ بعين الاعتبار عند تقدير مستوى القدرة أو السمة. كما أنه في الاختبارات التكيفية المحوسبة يتم عرض الأسئلة التي تتوافق مع مستوى القدرة لدى كل مفحوص، مما يقلل أو يلغي سلوك التخمين عند الطلبة ذوي القدرات المنخفضة. بالإضافة إلى ذلك فإن تطبيق الاختبارات التكيفية على الحاسوب يتيح استخدام أنواع مختلفة من الأسئلة لا تقتصر على نمط الاختيار من متعدد، وهذه الأسئلة لا تعطي فرصة للمفحوص للتخمين (Weiss & Şahin, 2024). كما أن الاختبارات التكيفية تتجنب الأسئلة غير ذات الصلة من خلال مراقبة القيم التقديرية لصعوبة الأسئلة وتمييزها، وبالتالي فإن المفحوص لن يتعرض للمفردات السهلة جدا أو الصعبة جدا والتي قد تثير لديه سلوكيات غير مرغوبة مثل التخمين والإهمال وأنماط الإجابة (فرحات، 2021).

• زيادة الأمان/ السرية

إن تطبيق خوارزميات انتقاء المفردات لكل مفحوص بناءً على استجاباته، يقلل من احتمالية تسريب الاختبارات مقارنة باختبارات الورقة والقلم، حيث يتم اختيار

المفردات من بنك أسئلة كبير ومشفر، مما يجعل من الصعب الوصول إلى الأسئلة، أو معرفة المفردات التي ستظهر لأي مفحوص آخر (Weiss & Şahin, 2024).

7. تحديات الاختبارات التكيفية المحوسبة

على الرغم من أن الاختبارات التكيفية المحوسبة تقدم العديد من الفوائد للمؤسسات وللمفحوصين، وتحسن من دقة وكفاءة الاختبارات، إلا أنه لا تزال هناك بعض التحديات التي يجب التعامل معها من قبل المؤسسات والمفحوصين والمتخصصين في القياس النفسي عند تطبيق الاختبارات التكيفية المحوسبة.

• التمويل والبنية التحتية

أول تحد تواجه المؤسسات والقائمين على الاختبارات هو التمويل اللازم للاستثمار في نظام حاسوبي يمكنه إدارة الاختبارات التكيفية المحوسبة من حيث تقديمها، وتصحيح الإجابات، واستخدام معلمات نظرية الاستجابة للمفردة. كما أن تطوير نظاما داخليا قد يكون مكلفا ويستغرق وقتاً طويلاً (Weiss & Şahin, 2024). ويمكن أن تكون البرمجيات التجارية الجاهزة مثل FastTest، أو المفتوحة المصدر مثل Concerto بدائل جيدة، لكنها قد تتطلب تكلفة إضافية أو خبرة تقنية عالية.

• الأمن السيبراني

يجب توفير نظام حاسوبي آمن لحماية بنك الأسئلة وبيانات المفحوصين، لأن اختراق النظام قد يعرض كل الجهود والاستثمارات في تطوير الأسئلة وتحليلها للخطر، مما يؤثر على سرية وخصوصية بيانات المفحوصين (Weiss & Şahin, 2024).

• الخبرة في نظرية الاستجابة للمفردة والاختبارات التكيفية المحوسبة

تحتاج المؤسسة إلى خبراء في نظرية الاستجابة للمفردة، وفي بناء الاختبارات التكيفية المحوسبة، لضمان تنفيذ البرنامج بكفاءة، كما أنه لا بد من دعم الخبراء بصورة مستمرة خاصة إذا كان يجب تحديث بنك الأسئلة باستمرار (Thompson & Weiss, 2011).

• توعية المفحوصين وأولياء الأمور

يجب توعية المفحوصين وأولياء الأمور حول كيفية عمل الاختبارات التكيفية المحوسبة، خاصة أنه قد يتعرض المفحوصين في الاختبار التكيفي المحسوب

ذي الطول المتغير إلى عدد مختلف من الأسئلة، مما قد يولد شعور بعدم العدالة (فرحات، 2021؛ Weiss & Şahin, 2024).

- عدم إمكانية مراجعة الإجابات

يواجه المفحوصون في الاختبارات التكيفية المحوسبة تحدي عدم القدرة على العودة للأسئلة مرة أخرى ومراجعة أو تغيير إجاباتهم، حيث يعتمد كل سؤال جديد على نمط الإجابات السابقة، مما يمنعهم من تغيير إجاباتهم بمجرد الانتقال للسؤال التالي (فرحات، 2021؛ الصبيحي والأحمدي، 2011).

- معدل التعرض للمفردات

قد تستخدم بعض الأسئلة بشكل متكرر أكثر من غيرها، مما يعرضها للتسريب أو التأثير على نزاهة الاختبار، ويتطلب ذلك استبدال المفردات القديمة بأخرى جديدة بشكل دوري (قاسم، 2021).

- تغطية مستويات السمة

يجب أن يحتوي بنك الأسئلة على عناصر تغطي مجموعة واسعة من مستويات القدرة أو السمة، حيث أن نقص المفردات في مستويات السمة العالية جدا أو المنخفضة جدا قد يؤدي إلى اختبارات أطول وأقل كفاءة (فرحات، 2021؛ Weiss & Şahin, 2024).

8. الذكاء الاصطناعي والتعلم الآلي والتطورات المستقبلية للاختبارات التكيفية المحوسبة

تتضمن الاتجاهات الحديثة في مجال الاختبارات التكيفية المحوسبة دمج الذكاء الاصطناعي، والتعلم الآلي والنماذج النفسية المتقدمة مما يعزز كفاءة عمليات الاختبارات التكيفية المحوسبة، حيث يساهم الذكاء الاصطناعي والتعلم الآلي في تطوير كيفية تصميم وتنفيذ الاختبارات التكيفية المحوسبة، حيث أنهما يوفران خوارزميات أكثر تطوراً لاختيار الأسئلة، والتوليد التلقائي للأسئلة. كما يمكن للذكاء الاصطناعي أن يقدم اختبارات أكثر تخصيصاً من خلال ليس فقط مستوى قدرة المفحوص، ولكن أيضاً من خلال سرعة الاستجابة والحالة النفسية للمفحوص (Mujtaba & Mahapatra, 2020).

ويرى وانج وزملاؤه (Wang et al., 2020) أنه يمكن للذكاء الاصطناعي نمذجة سلوك المفحوص بدقة، واتخاذ القرارات في اختيار الأسئلة، مما يعزز دقة وكفاءة الاختبار، وبذلك يضمن الذكاء الاصطناعي تكيف النظام مع احتياجات المفحوص الفردية مما يوفر بيئة تقييم متوازنة.

بالإضافة إلى ذلك فإنه يمكن أن يساهم الذكاء الاصطناعي بتعزيز تكامل الاختبارات التكيفية المحوسبة مع منصات التعلم والاختبارات عبر الانترنت، فيصبح الوصول للاختبارات أكثر سهولة مع تزايد توجه المؤسسات التعليمية نحو الحلول الرقمية.

1.8. الاعتبارات الأخلاقية لاستخدام الذكاء الاصطناعي

أصبح من المهم التركيز على الجوانب الأخلاقية، والحفاظ على الخصوصية مع تطور أنظمة الاختبارات التكيفية المحوسبة. فاستخدام الذكاء الاصطناعي والتعلم الآلي يثير تساؤلات حول أمان البيانات، وتحيز خوارزميات الاختبار، مثل التحيز عند اختيار الأسئلة مما يؤدي إلى تباين غير عادل في نتائج الاختبار. لذلك فإنه لا بد من إعطاء الأولوية لتصميم خوارزميات تراعي الجوانب الأخلاقية، وضمان اختبارها والتحقق من فعاليتها في العالم الحقيقي (Mujtaba & Mahapatra, 2020).

9. الخاتمة

تمثل الاختبارات التكيفية المحوسبة تقدما كبيرا في مجال القياس والتقييم، حيث توفر بديلا فعالا للاختبارات التقليدية. من خلال الاستفادة من مبادئ نظرية الاستجابة للمفردة، بالإضافة للخوارزميات التي تتبعها، تستطيع أنظمة الاختبارات التكيفية المحوسبة تقدير مستويات المفحوصين بدقة باستخدام عدد أقل من الأسئلة، مما يقلل من طول مدة الاختبار مع تحقيق دقة القياس. وقد أثبتت هذه الأنظمة فعاليتها في مجالات عدة، كالتقويم التربوي، والتقييم النفسي واختبارات الشهادات المهنية.

تمتلك الاختبارات التكيفية المحوسبة العديد من الميزات التي تشمل تحسين دقة وكفاءة الاختبارات، وتخصيص التقويم، فهي تعزز دافعية المفحوصين وتقلل من قلق الاختبارات لديهم، كما أنها توفر تغذية راجعة فورية وهو أمر مهم في التقويم التكويني، كما أن كفاءة هذا النوع من الاختبارات يجعلها مناسبة للاختبارات واسعة النطاق.

ومع ذلك تواجه أنظمة الاختبارات التكيفية المحوسبة بعض التحديات كالحاجة للتمويل، والاستثمار في البنية التحتية. كما يجب التركيز في هذه الاختبارات على الجوانب الأخلاقية خاصة مع دمج الذكاء الاصطناعي، مما يتطلب وضع التدابير اللازمة لحماية البيانات وضمان النزاهة في تصميم خوارزميات الاختبار. ويوصي هذا المقال بالبدء في استخدام الاختبارات التكيفية المحوسبة في أنظمة التقويم في وزارة التربية والتعليم، ومؤسسات التعليم العالي، لما لها من مميزات في توفير الوقت ودقة القياس مقارنة بالاختبارات التقليدية، كما يمكن للمعلمين تحديد نقاط القوة والضعف لدى الطلبة مما يساعد على تقديم الدعم الأكاديمي المناسب لكل طالب.

10. المراجع

- الشقصي، يعقوب زاهر، وأبو شندي، يوسف عبد القادر، والمحزري، راشد سيف. (2020). فعالية مؤشرات مطابقة الفرد في نماذج استجابة الفقرة عند اختلاف قوة الارتباط الموضوعي بين الفقرات. *مجلة الدراسات التربوية والنفسية- جامعة السلطان قابوس*، 14 (1)، 41- 53. DOI: <http://dx.doi.org/10.24200/jeps.vol14iss1pp41>
- الشهري، أحمد علي. (2023). أثر قواعد توقف الطول الثابت في إدارة الاختبارات التكيفية. *المجلة العلمية لكلية التربية بجامعة الوادي الجديد*، 15 (45)، 197- 216. <https://doi.org/10.21608/sjsw.2023.303772>
- شواشرة، لؤي، والقرعان، محمود. (2021). دقة قياس الاختبارات التكيفية المحوسبة في ظل ظروف اختبارية مختلفة. *المجلة الأردنية في العلوم التربوية*، 17 (2)، 285- 299.
- الصبيحي، علي أحمد، والأحمدي، شرف حامد. (2015). الاختبار التكيفي المحوسب Computerized Adaptive Teasing (CAT) “كتطبيق لتطوير أساليب التقييم والاختبارات الالكترونية. *مجلة كلية التربية- جامعة دمنهور*، 3 (1)، 83- 126.
- عمارة، إيهاب، كاظم، علي. (2020). دراسة جدوى تطوير صورة تكيفية محوسبة من اختبار المصفوفات المتتابعة الملونة لرافن للأطفال العمانيين باستخدام نظرية الاستجابة للمفردة. *المجلة النولية للأبحاث التربوية*، 44 (3)، 142- 181. <http://doi.org/10.36771/ijre.44.3.20-pp142-181>
- العمرى، محمد، وعيادات، يوسف. (2016). تصورات أعضاء هيئة التدريس والطلبة حول الاختبارات المحوسبة في العملية التعليمية، التعليمية في جامعة اليرموك. *المجلة الأردنية في العلوم التربوية*، 12 (4)، 469- 478.
- فرحات، رمضان السيد. (2021). الفرق بين الاختبارات المحوسبة الموائمة والخطية في تقدير القدرة وفقا لنظرية الاستجابة للمفردة لطلاب كلية التربية جامعة الأزهر. *مجلة التربية- كلية التربية بالقاهرة، جامعة الأزهر-*، 129 (3)، 1- 52.
- قاسم، مصطفى أحمد. (2021). الاختبارات التكيفية المحوسبة “النظرية والتطبيق”. دار الفكر العربي.
- نور الدين، أمين محمد صبري. (2019). فاعلية القياس التكيفي المحوسب في قياس تحصيل طلاب الجامعة. *المجلة السعودية للعلوم النفسية*، 2، 29- 47.

- Akhtar, H., & Kovacs, K. (2024). Measurement precision and user experience with adaptive versus non-adaptive psychometric tests. *Personality and Individual Differences*, 225, 112675. <https://doi.org/10.1016/j.paid.2024.112675>
- 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. <https://doi.org/10.33902/JPR.202425210>

- 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. DOI: 10.21890/ijres.327907
- Bichi, A. A., & Talib, R. (2018). Item response theory: An introduction to latent trait models to test and item development. *International Journal of Evaluation and Research in Education (IJERE)*, 7(2), 142–151. <https://doi.org/10.11591/ijere.v7i2.12900>
- Burr, S. A., Gale, T., Kisielewska, J., Millin, P., Pêgo, J. M., Pinter, G., Robinson, I. M., & Zahra, D. (2023). A narrative review of adaptive testing and its application to medical education. *MedEdPublish*, 13, 221. <https://doi.org/10.12688/mep.19844.1>
- Cheng, Y., Diao, Q., & Behrens, J. T. (2017). A simplified version of the maximum information per time unit method in computerized adaptive testing. *Behavior Research Methods*, 49(2), 502–512. <https://doi.org/10.3758/s13428-016-0712-6>
- Ebenbeck, N., & Gebhardt, M. (2022). Simulating computerized adaptive testing in special education based on inclusive progress monitoring data. *Frontiers in Education*, 7. <https://doi.org/10.3389/educ.2022.945733>
- Ebenbeck, N., Bastian, M., Mühlhng, A., & Gebhardt, M. (2024). Duration versus accuracy—What matters for computerised adaptive testing in schools? *Journal of Computer Assisted Learning*, jcal.13074. <https://doi.org/10.1111/jcal.13074>
- Han, K. (2018). Conducting simulation studies for computerized adaptive testing using SimuCat: An instructional peace. *Journal of Educational Evaluation for Health Professions*, 15(20). <https://doi.org/10.3352/jeehp.2018.15.7>
- Huang, H.-T. D., Hung, S.-T. A., Chao, H.-Y., Chen, J.-H., Lin, T.-P., & Shih, C.-L. (2022). Developing and validating a computerized adaptive testing system for measuring the english proficiency of taiwanese efl university students. *Language Assessment Quarterly*, 19(2), 162–188. <https://doi.org/10.1080/15434303.2021.1984490>
- Huda, A., Firdaus, F., Irfan, D., Hendriyani, Y., Almasri, A., & Lim, T., Gottipati, S., & Cheong, M. L. (2023). Educational Technologies and Assessment Practices: Evolution and Emerging Research Gaps. In *Reshaping Learning with Next Generation Educational Technologies*. IGI Global. DOI: 10.4018/979-8-3693 1310-7.
- 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. <https://doi.org/10.48550/arXiv.2404.00712>
- Martin, A., & Lazendic, G. (2018). Computer adaptive testing: Implications for students' achievement, engagement and subjective test experience. (2018). *Journal of Educational Psychology*, 110(1), 27- 45.
- Mujtaba, D. F., & Mahapatra, N. R. (2020). Artificial intelligence in computerized adaptive testing. *2020 International Conference on Computational Science and Computational Intelligence (CSCI)*, 649–654. <https://doi.org/10.1109/CSCI51800.2020.00116>

- Ogunsakin, I. & Shogbesan. Y. (2018). Item response theory (IRT): a modern statistical theory for solving measurement problems in 21st century. *International Journal of Scientific Research in Education*, 11(3B), 627- 635.
 - Rice, N., Pêgo, J. M., Collares, C. F., Kisielewska, J., & Gale, T. (2022). The development and implementation of a computer adaptive progress test across European countries. *Computers and Education: Artificial Intelligence*, 3, 100083. <https://doi.org/10.1016/j.caeai.2022.100083>
 - 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. <https://doi.org/10.3352/jeehp.2017.14.17>
 - Sorrel, M. A., Barrada, J. R., Torre, J. de la, & Abad, F. J. (2020). Adapting cognitive diagnosis computerized adaptive testing item selection rules to traditional item response theory. *PLOS ONE*, 15(1), e0227196. <https://doi.org/10.1371/journal.pone.0227196>
 - Štěpánek, L., & Martinková, P. (2020). Feasibility of computerized adaptive testing evaluated by Monte-Carlo and post-hoc simulations. *15th Conference on Computer Science and Information Systems (FedCSIS)*, 359–367. <https://doi.org/10.15439/2020F197>
 - Thompson, N. A., & Weiss, D. A. (2011). *A framework for the development of computerized adaptive tests*. <https://doi.org/10.7275/WQZT-9427>
 - Veldkamp, B. P., & Matteucci, M. (2013). Bayesian computerized adaptive testing. *Ensaio: Avaliação e Políticas Públicas Em Educação*, 21, 57–82. <https://doi.org/10.1590/S0104-40362013005000001>
 - Wang, N., Wang, D., & Zhang, Y. (2020). Design of an adaptive examination system based on artificial intelligence recognition model. *Mechanical Systems and Signal Processing*, 142, 106656. <https://doi.org/10.1016/j.ymssp.2020.106656>
 - Weiss, D. J., & Şahin, 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. <https://doi.org/10.13189/csit.2020.080302>
 - Zanon, C., Hutz, C. S., Yoo, H. (Henry), & Hambleton, R. K. (2016). An application of item response theory to psychological test development. *Psicologia: Reflexão e Crítica*, 29(1), 18. <https://doi.org/10.1186/s41155-016-0040-x>
- [Online Assessment Software, but Smarter | Assessment Systems](#)
[International Association for Computerized Adaptive Testing | Advancing the Science and Practice of Human Assessment](#)