الفروق بين الاختبارات المحوسبة الموائمة والخطية في تقدير القدرة وفقاً لنظرية الاستجابة للمفردة لطلاب كلية التربية جامعة الأزهر

إعداد

د/ رمضان السيد فرحات

قسم علم النفس التعليمي والإحصاء التربوي بكلية التربية

بنين - جامعة الأزهر بالقاهرة

الفروق بين الاختبارات المحوسبة الموائمة والخطية في تقدير القدرة وفقًا لنظرية الاستجابة للمفردة لطلاب كلية التربية جامعة الأزهر

رمضان السيد فرحات
قسم علم النفس التعليمي والإحصاء التربوي بكلية التربية بنين، جامعة الأزهر بالقاهرة.
البريد الإلكتروني: Ramadanfrahat2254.el@azhar.edu.eg

مستخلص البحث:

هدف البحث إلى المقارنة بين الاختبارات الموائمة المحوسبة والاختبارات الخطية المحوسبة في تقدير القدرة بطريقتي الأرجحية القصوى وبييز وباستخدام قاعدتي إنهاء الاختبار بعدد محدد من المفردات وبأدنى خطأ معياري، وبلغ عدد المشاركين في البحث (100) طالبًا من طلاب الفرقة الثالثة بكلية التربية بنين جامعة الأزهر بالقاهرة، واستخدم الباحث بنكًا للأسئلة في بعض فصول مقرر علم النفس التعليمي (1) (سيكولوجية التعلم)، واستخدم الباحث الأسلوب الإحصائي اختبار "ت" لعينتين مرتبطتين للتحقق من فروض البحث، واعتمد الباحث على البرامج الإحصائية الآتية: SPSS 25، jMetrik 4.1.1، FastTest 3.75.39، وتوصل البحث إلى عدم وجود فروق ذات دلالة إحصائية بين متوسطي تقدير القدرة للاختبار الخطي المحوسب والاختبار الموائم المحوسب باستخدام طريقة الأرجحية القصوى أو بييز لتقدير القدرة وقاعدة إنهاء الاختبار بعدد محدد من المفردات أو بأدنى خطأ معياري، ووجود فروق ذات دلالة إحصائية بين متوسطي الخطأ المعياري لتقدير القدرة للاختبار الخطي المحوسب والاختبار الموائم المحوسب باستخدام طريقة الأرجحية القصوى أو بييز لتقدير القدرة وقاعدة إنهاء الاختبار بعدد محدد من المفردات أو بأدنى خطأ معياري لصالح الاختبار الموائم المحوسب.

الكلمات المفتاحية: الاختبارات الموائمة المحوسبة، الاختبارات الخطية المحوسبة، نظرية الاستجابة للمفردة.

The Differences between the Adaptive Computerized and Linear Testings in Terms of Estimating Ability According to the Item Response Theory among the Faculty of Education Students, Al-Azhar University

Ramadan El-Sayed Frahat
Department of Educational Psychology and Statistics, Faculty of Education for Boys in Cairo, Al-Azhar University
Email: Ramadanfrahat2254.el@azhar.edu.eg

ABSTRACT

This research aimed to compare the computerized adaptive and linear testings concerning estimating ability by the two methods of maximum liklihood and bayes and by using the two rules of terminating the test with a specified number of items and with minimum standard error. The number of the participants in the research were (100) students of the third year at the College of Education for Boys, Al-Azhar University in Cairo. The researcher used a question bank in some chapters of the educational psychology course (1) (Learning Psychology). The researcher used paired samples *t*- test to verify the research hypotheses. The researcher relied on the following statistical programs SPSS 25, jMetrik 4.1.1, FastTest 3.75.39. The results of the research revealed that there were no statistically significant differences between the two means of assessing ability of computerized linear testing and computerized adaptive testing concerning using the maximum likelihood method or bayes to estimate ability and the termination test rule with a specified number of items or minimum standard error. The results also revealed that there were statistically significant differences between the two means of standard error of assessing ability of computerized linear testing and computerized adaptive testing using the maximum likelihood method or bayes to estimate ability and the termination test rule with a specified number of items or minimum standard error in favor of the computerized adaptive testing.

Keywords: Computerized Adaptive Testing, Computerized Linear Testing, Item Response Theory.

مقدمة:

تسعى مؤسسات التعليم العالي إلى تسريع تحولها الرقمي، وتطوير منظومة القياس والتقويم بحيث تواكب التقدم الهائل في الحاسوب وصناعة البرمجيات ونظرية القياس الحديثة، وتتماشى مع التغيرات التي يشهدها العالم بسبب جائحة كورونا، مما يتطلب تطوير نظم الاختبارات وتقديمها بشكل أكثر كفاءة؛ لأنه يمكن استخدامها في الحكم على درجة تحقق نواتج التعلم المستهدفة وتحسين منظومة التعليم والتعلم وضمان جودتها.

وتعد الاختبارات من أكثر الأدوات استخدامًا في التعليم العالي، والهدف الرئيس منها هو قياس وتقييم معرفة الطالب في الموضوعات المختلفة، وتم استخدام طرق تقييم مختلفة في سياقات مختلفة منذ اعتُمِد التعليم كمؤسسة، مثل: العروض التقدمية الصفية، والمقالات، والمشروعات، والتدريب العملي، ومع ذلك فإن الاختبار الشفهي واختبار الورقة والقلم هما أكثر أدوات قياس الأداء شيوعًا، وكان ينظر إلى الحاسوب كأداة تعليمية على مدى العقود القليلة الماضية، وانتشر استخدامه بسرعة في جميع مستويات التعليم والتدريب، وزاد استخدام الاختبارات المحوسبة بشكل كبير للحصول على التراخيص والشهادات والقبول، والنوع الأكثر شيوعًا للاختبارات المحوسبة هو الاختبار الخطي Linear test حيث يقدم نفس عدد المفردات لكل مختبر في ترتيب محدد عبر الحاسوب (Triantafillou et al., 2008, 1320)*.

ويعتبر الاختبار الموائم المحوسب Computerized adaptive testing (CAT) حالة خاصة من الاختبارات المحوسبة، حيث يأخذ كل مختبر اختبارًا فريدًا مصممًا وفقًا لمستوى قدرته، ويعتمد على نظرية الاستجابة للمفردة، التي يمكن من خلالها تقدير ومقارنة مستويات قدرات المواد المختلفة من مجموعات مختلفة من المفردات، وتحديد المفردة الأكثر ملاءمة لمستوى قدرة معينة، وتوسع استخدامه بسرعة خلال الأعوام الأخيرة نتيجة التقدم الذي حدث في تكنولوجيا المعلومات والاتصالات والحاسوب، والنظريات السيكومترية، وأصبح يستخدم بشكل متزايد في التقييم التعليمي والبرامج الاختبارية التي تقدمها العديد من المؤسسات مثل: مركز خدمات الاختبارات التربوية بالولايات المتحدة الأمريكية، ومن أمثلة هذه الاختبارات: اختبار اللغة الإنجليزية كلغة أجنبية (TOEFL)، واختبار تقييم الخريجين (GRE)، واختبار القبول للدراسات العليا في مجال الإدارة (GMAT)، وبطارية الكفاءة المهنية للقوات المسلحة (ASVAB) (Runder, 2010,153؛ Jian-quan et al., 2007, 72؛ Hol et al., 2008, 12).

ويبدأ الاختبار الموائم المحوسب بتقديم مفردة متوسطة الصعوبة من أجل تقييم مستوى المختبر مبدئيًا، وتسجل إجابته على الفور أثناء الاختبار، فإذا أجاب بشكل صحيح، تقدر قدرته ثم تقدم له مفردة تتطابق مع هذه القدرة الأعلى، وإذا أجاب عن المفردة التالية بشكل صحيح مرة أخرى، فإنه يعيد تقدير القدرة وتقدم المفردة التالية التي تطابق تقدير القدرة الجديدة، ويحدث العكس إذا أجاب عن المفردة بشكل غير صحيح، ويعيد الحاسوب باستمرار تقييم قدرته حتى تصل دقة التقدير إلى مستوى مقبول إحصائيًا أو عند الوصول إلى الحد الأقصى لعدد مفردات الاختبار (Leroux & Dodd, 2016, 666؛ Triantafillou et al., 2008, 1320).

ويتمثل المبدأ الأساسي وراء الاختبار الموائم المحوسب في تجنب طرح الأسئلة التي تكون صعبة جدًا أو سهلة جدًا بالنسبة للطالب الذي يتم اختباره، لأن الفاحصين يتوقعون غالبًا أن

* يلتزم الباحث في توثيقه للمراجع بالنسخة السابعة لقواعد الجمعية الأمريكية لعلم النفس APA

الطلاب مرتفعو القدرة سيجيبون عن المفردات السهلة بشكل صحيح، والطلاب منخفضو القدرة سيتعثرون في الإجابة عن الأسئلة الصعبة، ويتم تعلم القليل نسبيًا من هذه الإجابات، بينما يتم تعلم الكثير من خلال الأسئلة التي تتحدى الطالب ولا تعجزه، فتحديد هذه الأسئلة بشكل صحيح ثم تقديمها هو هدف كل اختبار موائم محوسب (Davey, 2011,4).

ويعتمد الاختبار الموائم المحوسب على خوارزميات Algorithms معينة تحدد المفردة الأكثر ملاءمة لاستدعائها بسرعة من بنك الأسئلة وعرضها على شاشة الحاسوب للفرد المختبر، ويعتمد اختيار المفردة عادةً على مفهوم معلومات المفردة، والذي يسعى إلى تحديد فكرة أن بعض المفردات أكثر ملاءمة من غيرها لموقف معين ((Thompson & Weiss, 2011,6.

ويمكن تصنيف أنظمة الاختبارات الموائمة المحوسبة إلى أسلوبين، أسلوب المفردة الواحدة، حيث تطبق مفردة واحدة في كل مرة، ثم يتم تكرار دورة (الاختيار- التطبيق- الدرجة) عدة مرات، وأسلوب الحزم أو المجموعات، أو يطلق عليه الاختبار متعدد المراحل Multistage testing (MST)، حيث تطبق مفردات متعددة في كل مرة، ويبدأ بتطبيق مجموعة من المفردات متوسطة الصعوبة، ويطلق على هذه المرحلة الاختبار التوجيهي أو الاستطلاعي، ثم يتم توجيه المختبر بناءً على أدائه في اختبار التوجيه إلى أحد اختبارات المرحلة الثانية التي تختلف في الصعوبة، ويمكن أن يتكون من مرحلتين أو ثلاث مراحل، وتسمى المرحلة الأخيرة اختبار القياس (Davey, 2011,8-9 ؛ Magis et al., 2017,3).

ويعد اختيار الأسلوب المناسب لتقدير قدرة الفرد في الاختبار الموائم المحوسب مهمًا؛ لأن دقة كل من اختيار المفردة التي يتم تقديمها للفرد، وإنهاء الاختبار، والتقدير النهائي للقدرة المقاسة تتوقف على هذا الأسلوب، ويمكن تقدير القدرة بعدة أساليب منها أسلوب الأرجحية القصوى (ML) Maximum Liklihood، والأساليب البيزية Bayesian (الصبيحي والأحمدي، 2011، 106-107).

ويمكن تصنيف قواعد إنهاء الاختبارات الموائمة المحوسبة إلى قاعدة الطول الثابت للاختبار، حيث يتوقف الاختبار بعد تطبيق عدد محدد مسبقًا من المفردات بغض النظر عن قدرة المختبرين، وقاعدة الطول المتغير للاختبار مثل: قاعدة أدنى خطأ معياري، حيث يأخذ المختبرون عددًا مختلفًا من المفردات قبل استيفاء قاعدة إنهاء الاختبار (Stafford et al., 2019,1306).

وتعمل الاختبارات الموائمة المحوسبة على تناغم عملية الاختبار مع احتياجات الجيل الإلكتروني من المتعلمين حيث تجعلها مبتكرة ومرنة وفردية وفعالة وسريعة، وتساعد المؤسسات الأكاديمية في تقييم عدد كبير من الطلاب في وقت قصير، وتُقدم نتائج دقيقة وسريعة لهم، وتكون نتائج التقييم متاحة بسرعة كبيرة، ويمكن للطلاب تقييم أنفسهم في أي وقت وبالسرعة التي تناسبهم، وبالتالي يصبحون مستقلين، وتقلل الحاجة إلى الموارد البشرية، ولذلك ينبغي أن يكون المعلمون والطلاب على وعي بالتطورات الجديدة في مجال الاختبارات، وتنمية مهاراتهم وتدريبهم على كيفية استخدام الحاسوب (Rezaie & Golshan, 2015,133).

وتستخدم الاختبارات الموائمة المحوسبة في قياس القدرة والتحصيل وسمات الشخصية والاتجاهات، والغرض الرئيس منها هو إنتاج اختبار أكثر كفاءة ودقة من الاختبار الخطي، ويعطي أقل قيمة من الخطأ المعياري لتقدير القدرة (Magis et al., 2017,35).

ويتزايد انتشار الاختبارات الموائمة المحوسبة لأنها تتمتع بأمان عال، وتجرى عند الطلب، ولا تحتاج إلى أوراق إجابة، وتلائم قدرة المختبر، ومرنة في اختيار المفردات، ومستوى تقنينها أعلى من الاختبار الخطي، ومدة المراقبة فيها قصيرة، وتعد فيها التقارير بصورة فورية ودقيقة (Istiyono et al., 2020,92).

ويتضح مما سبق أهمية استخدام الاختبارات الموائمة المحوسبة لتعدد مميزاتها واستخدامها في العديد من المجالات، وإمكانية تطبيقها بأسلوب المفردة الواحدة، أو أسلوب الحزم متعددة المراحل، وتتضمن خطوتين أساسيتين هما اختيار السؤال وتقدير الدرجات؛ وتحدد الخطوة الأولى السؤال المناسب لمعرفة مستوى أداء الطالب، والذي يتم اختياره من بنك الأسئلة، وتستخدم الخطوة الثانية في تقدير أداء الطالب من خلال إجابته عن السؤال السابق، وبذلك تكون الأسئلة التالية أكثر ملاءمة له، ويمكن تقدير القدرة بطريقتين هما الأرجحية القصوى وطريقة بييز، كما يمكن إنهاء الاختبار وفقًا لعدد محدد من مفردات الاختبار أو لأدنى خطأ معياري، ولذلك يسعى البحث الحالي إلى إعداد بنك أسئلة في بعض فصول مقرر علم النفس التعليمي (1) (سيكولوجية التعلم) لطلاب الفرقة الثالثة بكلية التربية جامعة الأزهر، وتقدير ومقارنة قدرتهم في كل من الاختبار الموائم المحوسب والاختبار الخطي المحوسب.

مشكلة البحث:

يشهد ميدان التقويم التربوي تقدمًا ملموسًا وذلك مع التطور التكنولوجي الذي يمثل السمة الرئيسة للعصر الحالي، فمنذ منتصف القرن الماضي، ظهرت بدايات نظرية الاستجابة للمفردة لتواجه العديد من الصعوبات والتحديات التي لم تتمكن نظرية القياس الكلاسيكية من علاجها، ومن أهم التطبيقات التي ارتبطت بهذا التقدم بناء بنوك الأسئلة والاختبارات الموائمة المحوسبة والتي تعتمد بشكل أساسي على استخدام الحواسيب ليس فقط كوحدات تخزينية، ولكن لتطبيق وإدارة عملية التقييم (عمارة وكاظم، 2020، 146).

وتتعدد مصادر مشكلة البحث، ويمكن للباحث عرضها في العناصر الآتية:

1- توصيات البحوث والدراسات السابقة:

يوصي الصبيحي والأحمدي (2011، 121-122) بضرورة إنشاء مراكز متخصصة في القياس والتقويم في الجامعات، تتوفر بها البنية التحتية والبرامج الخاصة بتطوير وتحليل الاختبارات، وتتبني الأساليب المتقدمة في التقويم والاختبارات الالكترونية ممثلة في الاختبارات الموائمة المحوسبة، كما ينبغي على الباحثين إجراء دراسات للمقارنة بين نتائج الطلاب على الاختبارات التقليدية والاختبارات الموائمة المحوسبة في المواد المختلفة.

ويرى ((Rezaie and Golshan 2015,133 أن هناك حاجة إلى إجراء المزيد من البحوث التي تقارن بين النتائج التي تم الحصول عليها من الاختبار الموائم المحوسب والدرجات التي تم الحصول عليها من اختبار الورقة والقلم أو الاختبار الخطي المحوسب.

ويُوصي نور الدين (2019، 43-44) بضرورة تبني الجامعات للاختبارات الموائمة المحوسبة، والتوسع في تحويل الاختبارات التحصيلية الورقية والخطية المحوسبة إلى اختبارات موائمة محوسبة، والتحقق من مدى فاعلية الاختبارات الموائمة المحوسبة في تقدير قدرة الأفراد في السياق الثقافي للبيئة العربية، ومعرفة مدى اختلاف التقدير البييزي والأرجحية القصوى في تقدير قدرة الأفراد في الاختبارات التحصيلية.

2- تناقض الأطر النظرية ونتائج الدراسات والبحوث السابقة حول أسلوب تقدير القدرة وقاعدة إنهاء الاختبار المناسبين للاختبار الموائم المحوسب:

يشير كل من نور الدين (2019، 29)، Ozyurt and Ozyurt (2015,34)، Samsudin et al. (2019,548)، Triantafillou et al. (2008,1323) and إلى فاعلية الاختبارات الموائمة المحوسبة في تقدير القدرة باستخدام طريقة الأرجحية القصوى.

بينما اعتمدت بعض الدراسات والبحوث على استخدام طريقة بييز لتقدير القدرة، فيرى نور الدين (2006، 100) أن طريقة بييز تعطي خطأ معياريًا أقل وكفاءة أكبر من طريقة الأرجحية القصوى، كما أشار Jacobusse and Van Buuren (2007) إلى أن استخدام طريقة بييز يعطي تقدير القدرة أكثر دقة في الاختبار الموائم المحوسب.

وتشير بحوث كل من Stafford et al.(2019, 1306), and Leroux and (2014, 213-214), Dodd (2012, 1) Babcock and Weiss إلى أن قاعدة الطول المتغير التي تعتمد على محك الخطأ المعياري للقياس تعطي نتائج أكثر دقة، بينما قاعدة الطول الثابت لا تقيس السمات الكامنة للمختبرين بدقة متساوية، فتفتقد للاتساق في دقة تقديرات قدرتهم، أي في الخطأ المعياري لتقدير قدرتهم عبر مدى القدرات، وهذا يعد إشكالية، خاصة في الاختبارات عالية المخاطر مثل: اختبار الترخيص أو الشهادة، وتطبيق عدد كبير من المفردات لزيادة الدقة يمكن أن يقلل من جدوى استخدام الاختبارات الموائمة المحوسبة، ويمكن أن يؤدي إلى زيادة نسبة التعرض للمفردات، ويزيد من عبء المختبرين.

ويرى Choi et al. (2011,38) أن قاعدة الطول الثابت للاختبار تتميز ببساطتها، ولكن يتم فيها قياس المختبرين بدرجات متفاوتة من الدقة، وبالتالي يصبح خطأ القياس أكبر عند المستويات المتطرفة من القدرة المقاسة، كما أنها تحد من كفاءة الاختبار الموائم المحوسب عندما يتم تطبيق مفردات غير ضرورية تُسهم بقليل من المعلومات حول مستوى قدرة المختبر.

بينما توصلت نتائج بحوث عودة وعبيدات (2013، 1602) إلى أن قاعدة إنهاء الاختبار بعدد محدد من المفردات توفر تقديرات للقدرة أدق، ودالة معلومات أعلى من قاعدة أدنى خطأ معياري.

كما توصلت نتائج بحث Kalender and Berberoglu (2017, 588) إلى عدم وجود اختلافات ملاحظة بين قواعد إنهاء الاختبار الموائم المحوسب بأسلوبي الطول الثابت والخطأ المعياري، وإمكانية استخدامه للقبول في التعليم الجامعي في تركيا.

3- كثرة مشكلات الاختبارات الخطية أو التقليدية:

يشير علام (2005، 297- 298) إلى أنه لا يستطيع أفضل طلاب صف دراسي معين أن يظهروا المستوى الحقيقي لتحصيلهم أو قدرتهم في الاختبارات التقليدية؛ وذلك لأن معظم المفردات الاختبارية تكون غاية في السهولة بالنسبة لهم، كما أن أكثر الطلاب انخفاضًا في القدرة سوف يجدون أن معظم المفردات الاختبارية غاية في الصعوبة بالنسبة لهم، وبذلك يمكن فقط معرفة ما أخفقوا فيه، وليس ما يمكنهم أداؤه، ولذلك تُفصل الاختبارات الموائمة المحوسبة، واستناد هذه الاختبارات في تصميمها وبنائها إلى نماذج نظرية الاستجابة للمفردة يساعد في التغلب على مشكلات الاختبارات التقليدية وعدم فعاليتها في قياس قدرة الأفراد.

ويؤكد علام (2006، 733-734) على أنه لا ينبغي اختبار مجموعة من الأفراد المختلفين في قدرتهم بنفس الاختبار، ولكن الأجدى اختبارهم بمفردات مختلفة، ولذلك تعتمد الاختبارات الموائمة المحوسبة على فكرة أن الطريقة المثلى لقياس قدرة الفرد تتم من خلال المضاهاة بين صعوبة مفردة الاختبار، ومستوى القدرة التي يقيسها الاختبار.

ويرى Ozyurt and Ozyurt (2015, 28) أن طريقة التقييم والتقويم التقليدية تتناقض مع التعلم الحديث، الذي يركز على خصائص التعلم الفردي، وأدى هذا الاتجاه إلى تسريع تحويل الاختبارات التقليدية إلى اختبارات موائمة تكون أكثر ملاءمة للتعلم الحديث.

ويشير الكيال وآخران (2017، 224) إلى كثرة مشكلات الاختبارات التي يتم إعدادها وتقديمها بصورة تقليدية، والتي من الممكن التغلب عليها باستخدام مزايا الاختبارات الموائمة المحوسبة.

ويوضح Samsudian et al. (2019,547) أن معايير التقييم الحالية تتطلب مستوى عاليًا من الدقة والموضوعية ووقتًا قليلًا لتطبيق الاختبارات، مما يُقلل من كفاءة الاختبارات التقليدية التي سيطرت على مجال التقييم لسنوات طويلة جدًا، بينما يلبي الاختبار الموائم المحوسب معايير التقييم الحالية؛ لأنه يتميز بالموائمة مع قدرات المختبرين، ولذلك يعد بديلًا للاختبارات التقليدية.

4- تزايد الاهتمام في الآونة الأخيرة بالاختبارات الموائمة المحوسبة:

يرجع تزايد الاهتمام بالاختبارات الموائمة المحوسبة لكثرة مميزاتها بالمقارنة بالاختبارات التقليدية أو الخطية، ومن هذه المميزات كما يشير Cheng and Basu (2009, 2034) , and Cella et al. (2007, 135) اختيار المفردات الاختبارية بناءً على مستوى الطالب، وانخفاضًا كبيرًا في تكلفة إجراء الاختبارات، وتقليل زمن الاختبار، وإدخال البيانات وتقدير الدرجات بصورة فورية، وإجراء الاختبارات حسب الطلب، واستخدام الرسومات والألعاب التعليمية في تصميم المفردات لجذب الطلاب وزيادة دافعيتهم للتعلم.

وترى دعنا (2012، 155) أن العصر الحالي يشهد نموًا متزايدًا في تفريد التعليم والتعلم للإتقان، ولذلك أصبحت الحاجة متزايدة وملحة لقياس نتائج البرامج التعليمية بصورة صادقة وموضوعية وذات قيمة تنبؤية وتحقق محكات التفريد والإتقان، لتمكن القائمين على البرامج التعليمية من تقدير المستوى الفعلي لطلابهم، ولذلك تزايد تزايد الاهتمام بالاختبارات الموائمة

لأنها تلبي متطلبات التعليم المتطورة، وتواجه المشكلات السيكومترية لنظرية القياس الكلاسيكية.

ويشير Magis et al. (2017,1) إلى أن الاختبارات الخطية تعد الطريقة الأكثر شيوعًا لقياس معرفة ومهارات المختبرين في التقييمات التعليمية، ومع تقدم علوم وتكنولوجيا الحاسوب بشكل سريع في العقدين الماضيين، ازداد الطلب على الاختبارات المحوسبة بصفة عامة والاختبارات الموائمة المحوسبة بصفة خاصة بسبب دقتها وكفاءتها.

ويؤكد Su (2018,346) على مميزات الاختبار الموائم المحوسب واستخدامه على نطاق واسع في العديد من التقييمات التربوية والنفسية، لأنه يمكن من خلاله الحصول على تقدير فعال ودقيق

للقدرة باستخدام مفردات أقل من اختبار الورقة والقلم التقليدي.

5- ندرة الدراسات والبحوث العربية التي اهتمت بتطبيق الاختبارات الموائمة المحوسبة لتقدير قدرة الطلاب في أحد المقررات الدراسية:

لاحظ الباحث من خلال اطلاعه على بعض الدراسات والبحوث العربية أنها ركزت على تطبيق الاختبارات الموائمة المحوسبة في قياس وتقدير القدرات العقلية مثل: (الحسيني والعدروس، 2014؛ الكرامنة، 2015؛ عودة وعبيدات، 2013؛ نور الدين، 2006)، ما عدا بحثي (السعدني، 2019؛ نور الدين، 2019) حيث ركزا على تطبيق الاختبارات الموائمة المحوسبة في قياس وتقدير القدرة في أحد المقررات الدراسية.

وفي ضوء ما سبق يمكن صياغة مشكلة البحث الحالي في التساؤلات الآتية:

1- ما الفروق بين الاختبار الموائم المحوسب والاختبار الخطي المحوسب في متوسطي تقدير القدرة والخطأ المعياري لتقدير القدرة باستخدام طريقة الأرجحية القصوي وقاعدة إنهاء الاختبار بعدد محدد من المفردات؟

2- ما الفروق بين الاختبار الموائم المحوسب والاختبار الخطي المحوسب في متوسطي تقدير القدرة والخطأ المعياري لتقدير القدرة باستخدام طريقة الأرجحية القصوي وقاعدة إنهاء الاختبار بأدنى خطأ معياري؟

3- ما الفروق بين الاختبار الموائم المحوسب والاختبار الخطي المحوسب في متوسطي تقدير القدرة والخطأ المعياري لتقدير القدرة باستخدام طريقة بييز وقاعدة إنهاء الاختبار بعدد محدد من المفردات؟

4- ما الفروق بين الاختبار الموائم المحوسب والاختبار الخطي المحوسب في متوسطي تقدير القدرة والخطأ المعياري لتقدير القدرة باستخدام طريقة بييز وقاعدة إنهاء الاختبار بأدنى خطأ معياري؟

أهدف البحث:

يهدف البحث الحالي إلى المقارنة بين الاختبارات الموائمة المحوسبة والاختبارات الخطية المحوسبة في تقدير القدرة بطريقتي الأرجحية القصوى وبييز وباستخدام قاعدتي إنهاء الاختبار بعدد محدد من المفردات وبأدنى خطأ معياري.

أهمية البحث:

يمكن أن تتضح أهمية البحث الحالي في:

1-إثراء المكتبة العربية بإطار نظري عن الاختبارات الموائمة المحوسبة من حيث نشأتها، ومفهومها، والفرق بينها وبين الاختبارات الخطية، وخطوات إعدادها، ومميزاتها، ومشكلاتها.

2-إعداد بنك أسئلة في بعض فصول مقرر علم النفس التعليمي (1) لطلاب الفرقة الثالثة بكلية التربية جامعة الأزهر يمكن استخدامه في قياس التحصيل بأسلوبي الاختبار الخطي المحوسب والاختبار الموائم المحوسب.

3- توجيه انتباه القائمين على المؤسسات التعليمية بصفة عامة ومؤسسات التعليم العالي بصفة خاصة إلى إمكانية تطبيق اختبارات موائمة محوسبة تتميز بخصائصها السيكومترية المرتفعة، ودقة القياس، وتوفر الوقت والجهد والتكلفة، وتحد من الغش، ويمكن تطبيقها عن بعد.

التعريف الإجرائي لمصطلحات البحث:

1- الاختبار الموائم المحوسب ((Computerized Adaptive Testing CAT:

"الاختبار الذي يُفَصَّل لكل طالب على حدة بحيث يُقدِّم له الحاسوب المفردات التي تتناسب مع مستوى قدرته من خلال برنامج FastTest، بهدف الوصول إلى تقدير متوسط للقدرة يزداد أو لا يختلف اختلافًا دالًا إحصائيًا عن متوسط تقدير القدرة للاختبار الخطي المحوسب، ويعطي أقل قيمة من الخطأ المعياري لتقدير القدرة وفق طريقتي الأرجحية القصوى أو بييز وقاعدتي إنهاء الاختبار بعدد محدد من المفردات أو بأدنى خطأ معياري" (الباحث).

2- الاختبار الخطي المحوسب Linear test **Computerized**:

"الاختبار الذي يُقدِّم فيه الحاسوب نفس عدد المفردات لكل طالب بترتيب معين وزمن محدد بغض النظر عن قدرته من خلال برنامج FastTest " (الباحث).

3- بنك الأسئلة Item Banking:

"تجمع كبير نسبيًا من المفردات الاختبارية التي تقيس التحصيل الأكاديمي في بعض فصول مقرر علم النفس التعليمي (1) حيث تم تجريبها، وتقدير بارامترات الصعوبة والتمييز والتخمين لها، والتحقق من مطابقتها للنموذج اللوغاريتمي ثلاثي البارامتر، وتخزينها وتصنيفها في الحاسوب باستخدام برنامج FastTest " (الباحث).

3- **النموذج اللوغاريتمي ثلاثي البارامتر** (3PL): Three- Parameter Logistic Model

"أحد نماذج نظرية الاستجابة للمفردة Item Response Theory (IRT) أحادية البعد ثنائية الاستجابة، والذي يقوم بتقدير ثلاثة بارامترات للمفردة الاختبارية تتمثل في بارامتر صعوبة المفردة (b_i) وبارامتر تمييز المفردة (a_i)وبارامتر تخمين المفردة (c_i)، بجانب تقدير قدرة الأفراد المختبرين (θ) وفق طريقتي الأرجحية القصوى وبييز"(الباحث).

حدود البحث:

تتمثل حدود تعميم نتائج البحث الحالي في ضوء ما يأتي:

1- حدود زمانية: تتمثل في فترة تطبيق البحث خلال الفصل الدراسي الأول من العام الجامعي 2020/ 2021م.

2- حدود مكانية: تتمثل في كلية التربية بنين جامعة الأزهر بالقاهرة.

3- حدود موضوعية: تتمثل في بنك الأسئلة المستخدم في البحث الحالي، ونموذج نظرية الاستجابة للمفردة وهو النموذج ثلاثي البارامتر، وأسلوب المفردة الواحدة للاختبار الموائم المحوسب، حيث إن النسخة المجانية التي حصل عليها الباحث لبرنامج Test Fast من مؤسسة Assessment الأمريكية أتاحت له تطبيق هذا الأسلوب على عدد محدود من الطلاب، بينما لم يتمكن من تطبيق أسلوب الحزم المتعددة بسبب تكلفته المرتفعة، وطلاب الفرقة الثالثة بكلية التربية بنين جامعة الأزهر بالقاهرة.

الإطار النظري والدراسات والبحوث السابقة:

أولًا: نشأة الاختبارات الموائمة المحوسبة:

يرى Linacre (2000,4) أن Binet حقق تقدمًا هائلًا في الاختبارات الموائمة من خلال اختبارات الذكاء التي أعدها، نظرًا لاهتمامه بتشخيص المفحوص بصورة فردية، حيث استطاع تفصيل الاختبار له عن طريق ترتيب المفردات من حيث الصعوبة، ثم يبدأ الاختبار بمجموعة فرعية من المفردات وفق مستوى قدرته الذي تم تخمينه، فإذا نجح يتم إعطاؤه بشكل متدرج مجموعات فرعية أصعب حتى يخطئ في الإجابة بنسبة كبيرة، وإذا فشل يتم إعطاؤه بشكل متدرج مجموعات فرعية أسهل حتى يجيب عنها إجابة صحيحة بنسبة كبيرة.

ويشير Weiss (2004,71-72) إلى أن قياس الذكاء في اختبار Binet هو إجراء موائم تمامًا لأنه: (أ) يستخدم بنك لمفردات سبق معايرتها، والتي يتم اختيارها في كل مستوى عمري عندما يجيب عنها بشكل صحيح حوالي (50%) من المختبرين في هذا المستوى العمري، (ب) يُطبق بشكل فردي بواسطة أخصائي نفسي للتحقق من مستوى الصعوبة الأكثر ملاءمة لكل مفحوص، (ج) يمكن البدء في تطبيقه من المستوى المحتمل لقدرة المختبر والذي عادة ما يكون وفق عمره الزمني، وقد يكون مستوى البداية أقل أو أعلى إذا كانت هناك معلومات تحدده، (د) تُسجل فيه المفردات بصورة فورية بواسطة الأخصائي النفسي، (هـ) يعتمد اختيار المفردة التالية فيه على أداء المختبر في المفردات السابقة، فإذا أجاب عن بعضها أو معظمها

في مستوى عمري معين بشكل صحيح، فإنه يتم تقديم المفردات في المستوى العمري الأعلى التالي، والعكس صحيح، (و) يتم إنهاؤه عندما يُحدد المستوى القاعدي (المستوى العمري الذي يجيب فيه المختبر إجابة صحيحة عن جميع مفردات الاختبار) وسقف الاختبار (المستوى العمري الذي يجيب فيه المختبر إجابة خطأ عن جميع مفردات الاختبار) لكل مفحوص، ويقع المدى الفعال للقياس بين هذين المستويين، وتقدر النتيجة النهائية في ضوء المفردات التي أجاب عنها المختبر إجابة صحيحة، ويتم وزنها حسب المستوى العمري لحساب نسبة الذكاء.

ويرى Finkelman et al. (2010, 239), and Weiss (2004,72) أن اختبار Binet ظل الاختبار الموائم الأدائي الوحيد لأكثر من نصف قرن؛ لأن متطلبات الحرب العالمية أدت إلى سيطرة تقنية اختبارات الورقة والقلم التقليدية على الاختبارات النفسية والتعليمية خلال معظم القرن العشرين، ومع ذلك في خمسينيات القرن العشرين، بدأ الباحثون في الجيش الأمريكي استكشاف إمكانية تقديم اختبارات موائمة بدائية باستخدام كل من الورق والقلم وآلات الاختبار، ولكن لم تنجح هذه الطرق، وفي أواخر الستينيات من القرن العشرين، بدأ مكتب الولايات المتحدة الأمريكية للبحوث البحرية في دعم البحوث النظرية حول الاختبارات الموائمة ونظرية الاستجابة المفردة التي قام بها Lord، ودعم بعض البحوث التجريبية لتنفيذ التدريس بمساعدة الحاسوب، وقياس التحصيل الأكاديمي، مما أدى إلى بداية مجال الاختبارات الموائمة المحوسبة.

ويؤكد Finkelman (2004,1) أن الاهتمام بالاختبارات الموائمة المحوسبة زاد مع إصدار التشريعات القانونية في الولايات المتحدة الأمريكية التي تتطلب قياس دقيق لمستوى قدرة الطلاب وتقدير مستوى كفاءتهم بدقة في جميع المواد، مما يستلزم استخدام نظام التصنيف الإحصائي بدلًا من أساليب التقدير المعيارية لتقييم الطلاب.

ويشير علام (2005، 298-299) إلى أن الحاجة إلى الاختبارات الموائمة نشأت نتيجة بعض المشكلات التي تواجه الاختبارات الجماعية، حيث تكون معظم مفرداتها متوسطة الصعوبة، وعدد قليل من مفرداتها يكون غاية في السهولة أو الصعوبة، ويترتب على ذلك أن يجيب الفرد الأكثر كفاءة عن عدد كبير من المفردات التي تكون سهلة بالنسبة له قبل التوصل إلى قدر معقول من المعلومات فيما يتعلق بقدرته، وهذا يعد مضيعة للوقت والجهد، وربما يؤدي إلى تأثر عملية القياس بمتغيرات دخيلة، مثل: الملل، والأخطاء الناجمة عن الإهمال، أما الأفراد الأقل كفاءة، فإنهم يجيبون على المفردات السهلة باعتبار أنها تقيس قدرتهم، بينما لا تقدم المفردات الصعبة أية معلومات في هذا الشأن، ولكن ربما يؤدي إلى إحباطهم، أو إرباكهم، أو لجوئهم إلى التخمين العشوائي، مما يسهم في تأثر عملية القياس أيضًا بهذه المتغيرات الدخيلة، ولكي يقيس الاختبار جميع الأفراد المختبرين قياسًا فاعلًا، ينبغي أن يشتمل على مدى متسع من صعوبات المفردات وفقًا لمدى كفاءتهم.

ويرى الصبيحي والأحمدي (2011، 85) و Rezaie and Golshan (2015,128) أن تطور ونمو الاختبارات التربوية والنفسية وتطبيقاتها تأثر إيجابيًا بثورة تقنية الحاسوب، حيث ساعدت في إيجاد نظريات قياس متقدمة، وظهور أساليب جديدة للاختبارات المحوسبة والتي احتلت مكان اختبارات الورقة والقلم التقليدية، ومنها الاختبارات الموائمة المحوسبة.

ويشير الحسيني والعدروس (2014، 3) إلى بعض المشكلات التي يعاني منها نظام القياس والتقويم التقليدي والتي كانت سببًا في نشأة الاختبارات الموائمة المحوسبة ومنها أن الاختبارات

لا تراعي تباين القدرات العقلية للمفحوصين، وصعوبة استخراج تقرير تشخيصي شامل لتحديد نقاط القوة والضعف لديهم، بالإضافة إلى اتباع الطرق التقليدية في وضع مفردات الاختبار وما يصاحبها من محدودية التقويم، وتنظيم وتجهيز وإخراج الأوراق الامتحانية، وعمليات التصحيح وتقدير الدرجات وإدارة الامتحانات وما تتكلفه من ميزانيات كبيرة.

ويتضح مما سبق أنه تم توضيح المبادئ الأساسية للاختبارات الموائمة المحوسبة وكيفية تنفيذها بواسطة Binet في عام 1905م من خلال اختبار ستانفورد- بينيه للذكاء (IQ)، مع استبدال الأخصائي النفسي بالبرنامج الحاسوبي الذي يتم فيه تخزين مفردات الاختبار، ويحدد كيفية بدء الاختبار لكل مفحوص، ويختار المفردات التالية بناءً على استجاباته على المفردات السابقة، ويطبق قاعدة واحدة أو أكثر لإنهاء الاختبار، وزاد الاهتمام بهذه الاختبارات نتيجة المشكلات التي تواجه الاختبارات التقليدية الجماعية بالإضافة إلى ظهور نظريات حديثة في القياس، وتطور البرمجيات وتكنولوجيا المعلومات والاتصالات.

ثانيًا: مفهوم الاختبارات الموائمة المحوسبة والفرق بينها وبين الاختبارات الخطية:

يعد الاختبار الموائم المحوسب من التطبيقات الرئيسة لنماذج نظرية الاستجابة للمفردة، وهو أداة قياس يتم تطبيقها على المختبرين عبر الحاسوب، ويتمثل الهدف الرئيس منه في استخدام خاصية عدم التغير التي تتميز بها نظرية الاستجابة للمفردة لإنشاء خوارزمية يتلقى من خلالها كل مفحوص اختبارًا جيدًا لقياس السمة لديه، ويكون الاختبار مفصلًا أو مصممًا له بحيث لا تكون الأسئلة صعبة جدًا أو سهلة جدًا، وعادة ما يحتوي على مفردات أقل من اختبار الورقة والقلم التقليدي، كما أن جميع المختبرين لا يتلقون نفس مفردات الاختبار، ويعالج مشكلة عدم الفاعلية التي تتميز بها الاختبارات التقليدية من خلال تقديم اختبار لكل مفحوص يوائم مستوى قدرته وتكون احتمالية إجابته عن المفردات بشكل صحيح (0,5) وبالتالي فإن هذه المفردات توفر معلومات جيدة عن وضعه النسبي (Embretson & Reise,2000,263).

وتتضمن الأدبيات المختلفة تسميات متعددة للاختبارات الموائمة المحوسبة، مثل: الاختبارات المفصلة Tailored Tests، والاختبارات المتفاعلة Interactive Tests، والاختبارات المقترنة بالاستجابات Response- Contingent Tests، والاختبارات المبرمجة Programmed Tests، والاختبارات الآلية Automated Tests، والاختبارات التفريدية Individualized Tests، والاختبارات التفريعية Branched Tests، والاختبارات التسلسلية Sequential Tests (علام، 2005، 296؛ Domino & Domino, 2006, 473).

ومن خلال اطلاع الباحث على بعض الكتابات النظرية والدراسات والبحوث السابقة مثل: عودة وعبيدات (2013، 1605)، مسلم (2011، 28)، نور الدين (2006، 88-89)، نور الدين (2019، 32)، Davey and Pitoniak, (2006, 543) ، Mizumoto et al. (2019, 106)، Ramadan and Aleksandrovna (2018, 3) ، Veldkamp (2016, 212)، and Yasuda et al.(2021,1) لاحظ أنها تتفق على بعض العناصر في تعريف الاختبارات الموائمة المحوسبة، وبالتالي أمكن للباحث استخلاص مفهومها على النحو الآتي "الاختبارات التي تطبق من خلال الحاسوب، وتعتمد على نظرية الاستجابة للمفردة في تطبيقها، وتُفصل لكل مفحوص على حدة بحيث تقدم له المفردات التي تتناسب مع مستوى قدرته فلا يتعرض للمفردات السهلة أو

الصعبة جدًا، ولذلك يختلف عدد المفردات الاختبارية المطبقة وترتيبها وزمن الاختبار من مفحوص لآخر حسب قدرته، بهدف الوصول إلى تقدير أدق للقدرة بأقل عدد من المفردات".

ويعرف الاختبار الخطي بأنه تقييم ثابت الطول يقدم فيه نفس عدد المفردات لكل مختبر في ترتيب محدد، وتعتمد درجته على عدد المفردات التي أجاب عنها بشكل صحيح، ويتكون من مجموعة من المفردات الاختبارية السهلة والصعبة التي يتم اختيارها عشوائيًا من مستودع أكبر أو تطبق كل مفردات المستودع على المختبرين، بغض النظر عما إذا كانت المفردات سهلة جدًا أو صعبة جدًا بالنسبة لهم، فلا يهتم بقدرة كل مختبر(Triantafillou et al. , 2008, 1320).

ويطبق الاختبار الخطي بصورة ورقية غالبًا، ولذلك يسمى باختبار الورقة والقلم، ونظرًا لأنه من المحتمل أن يجيب المختبرون ذوي القدرة المرتفعة عن جميع المفردات السهلة بشكل صحيح، وذوي القدرة المنخفضة عن جميع المفردات الصعبة بشكل خطأ، فإن هذه المفردات السهلة والصعبة تسهم بقليل من المعلومات في قياس قدراتهم عند طرفي ميزان القياس المرتفع والمنخفض على التوالي، ولذلك تحتاج الاختبارات الخطية إلى عدد كبير من المفردات للحصول على قياس جيد، ويشبه الاختبار الخطي المحوسب اختبار الورقة والقلم باستثناء أنه يتم تطبيقه عبر الحاسوب، ولكنه يتمتع بالعديد من المميزات المرتبطة بالحوسبة مثل مرونة نظام الاختبار، وفاعلية تطبيقه، وتجميعه، وتقرير الدرجات، ولكنه يعاني من نفس القيود التي يعاني منها اختبار الورقة والقلم (Magis et al., 2017,1-2).

ويلخص محمد (2017، 12) و Jian-quan et al. (2007,72) الفرق بين الاختبارات الخطية والاختبارات الموائمة المحوسبة كما في جدول (1)، وقام الباحث بإضافة النقاط الأربعة الأخيرة في الجدول:

جدول (1)

المقارنة بين الاختبارات الخطية والاختبارات الموائمة المحوسبة

وجه المقارنة	الاختبارات الخطية	الاختبارات الموائمة المحوسبة
تركيب الاختبار	يطبق على جميع المختبرين اختبار واحد	يطبق على كل مختبر اختبار مختلف عن زملائه
صعوبة الاختبار	تستهدف المختبر متوسط القدرة، وتحدد صعوبته في ضوء متوسط إجابات المختبرين	تستهدف كل مختبر على حدة، وتحدد صعوبته في ضوء إجابة كل مختبر
طول الاختبار	اختبارات طويلة نسبيًا، وطول الاختبار واحد لكل المختبرين	اختبارات قصيرة نسبيًا، وطول الاختبار يختلف باختلاف المختبرين
بداية الاختبار	يبدأ جميع المختبرين في نفس اللحظة	يبدأ في أي وقت مناسب للمختبر
زمن الاختبار	زمن واحد لكل المختبرين، ولا	يختلف الزمن لكل مختبر،

وجه المقارنة	الاختبارات الخطية	الاختبارات الموائمة المحوسبة
	توجد مرونة في زمن الاختبار	وتوجد مرونة في زمن الاختبار
تنظيم الاختبار	يحتاج لوقت طويل	يحتاج لوقت قصير
فورية النتائج	نتائج غير فورية	نتائج فورية
فاعلية القياس	أقل فاعلية	أكثر فاعلية
التطبيق والتصحيح	سهولة التطبيق وتقدير الدرجات	صعوبة في التنفيذ وتقدير الدرجات
تطوير الاختبار وعرض مفرداته	جهد أقل في تطوير الاختبار، وسهولة في عرض مفرداته	جهد كبير في تطوير الاختبار، وصعوبة في التحكم في عرض مفرداته
تعيير المفردات	لا يتطلب تعيير المفردات الاختبارية	يتطلب مجموعة كبيرة من المفردات الاختبارية تم تعييرها

ثالثًا: خطوات إعداد الاختبارات الموائمة المحوسبة:

يشير علام (2005، 303-304) و Chang (2005, 218) إلى ثمان خطوات لإعداد الاختبارات الموائمة المحوسبة تتمثل في: بناء مجموعة كبيرة من المفردات الاختبارية أو مستودع مفردات، واختيار أحد نماذج نظرية الاستجابة للمفردة الاختبارية (IRM)، وتأسيس بنك أسئلة محوسب Computerized Item Bank يشتمل على مجموعة كبيرة من المفردات التي سبق تعييرها أو تدريجها Calibration باستخدام النموذج المستخدم، وتصميم نظام تزويدي Delivery System لاستدعاء المفردات الاختبارية بسرعة من بنك الأسئلة وعرضها على شاشة الحاسوب، وتحديد آليات انتقاء المفردة الأولية عند بدء العملية الاختبارية، وتحديد استراتيجية انتقاء المفردة التالية استنادًا إلى درجة الفرد في المفردة السابقة، واختيار أسلوب تقدير القدرة المناسب، وتحديد قاعدة إنهاء الاختبار.

الخطوة الأولى: بناء مجموعة كبيرة من المفردات الاختبارية أو مستودع مفردات:

ينبغي كتابة أو تجميع عدد كافٍ من مفردات اختبارية تقيس مدى متسعًا من القدرة المعينة، وهذا يتطلب البدء بتحديد مواصفات كل مفردة، ومحتواها، ونطاق المهارات التي تُمثلها، وكلما كانت هذه المواصفات واضحة، أمكن بناء عدد كافٍ ومتنوع من المفردات المتباينة في صعوبتها على مدى القدرة أو السمة المراد قياسها، كما ينبغي مراجعتها بدقة للتحقق من جودتها، وعدم تحيزها لجماعة معينة دون غيرها (علام، 2005، 304؛Erdem-Kara, 2019,46).

الخطوة الثانية: اختيار أحد نماذج نظرية الاستجابة للمفردة الاختبارية:

تتمثل الميزة الأساسية لنظرية الاستجابة للمفردة في وضع المفردات والأفراد على نفس الميزان بمتوسط= (صفر)، وانحراف معياري= (1) بحيث يمكن مقارنة الأفراد بالمفردات

والعكس، ويمكن تمييز نماذجها ثنائية الاستجابة إلى ثلاثة نماذج: نموذج Rasch أحادي البارامتر، ونموذج Lord ثنائي البارامتر، ونموذج Birnbaum ثلاثي البارامتر، وذلك بناءً على عدد بارامترات المفردات الشائعة والتي تتمثل في الصعوبة، والتمييز، والتخمين (احتمال توصل الأفراد من ذوي القدرة المنخفضة إلى الإجابة الصحيحة عن المفردة عن طريق التخمين)، وعلى الرغم من أن البارامترات الثلاثة مفيدة في حد ذاتها، إلا أنه طبقًا للهدف من الاختبارات الموائمة المحوسبة يتم دمجها في دالة معلومات المفردة Item Information Function (IIF)، والتي تصف مدى جودة أو دقة مفردة ما في كل مستوى للسمات التي يتم قياسها بواسطة اختبار معين (Boyd, 2003,9 ؛Weiss,2004,73).

ويعد النموذج اللوغاريتمي أحادي البارامتر One- Parameter Logistic Model (1PL) أبسط نماذج الاستجابة للمفردة أحادية البعد، ويفترض هذا النموذج أن جميع مفردات الاختبار تميز بين الأفراد بنفس القدر، إلا أنها تختلف في صعوبتها، حيث لا تتقاطع المنحنيات المميزة للمفردات المختلفة، ويعبر عنه بالمعادلة $P_i(\theta) = \frac{e^{D(\theta-bi)}}{1+e^{D(\theta-bi)}}$، بينما يعتمد النموذج اللوغاريتمي ثنائي البارامتر Two- Parameter Logistic Model (2PL) على بارامترين أساسيين للمفردة هما: صعوبة المفردة (b_i) وتمييز المفردة (a_i)، ويعبر عنه بالمعادلة $P_i(\theta) = \frac{e^{Dai(\theta-bi)}}{1+e^{Dai(\theta-bi)}}$، ويفترض النموذج اللوغاريتمي ثلاثي البارامتر Three- Parameter Logistic Model (3PL) أن المفردات تختلف في الصعوبة والتمييز والتخمين (C_i)، ويعبر عنه بالمعادلة $P_i(\theta) = c_i + (1- c_i)\frac{e^{Dai(\theta i-bi)}}{1+e^{Dai(\theta i-bi)}}$، حيث $P_i(\theta)$ تمثل احتمال استجابة المختبر الذي له قدرة (θ) استجابة صحيحة على المفردة (i)، و (e) الأساس اللوغاريتمي الطبيعي ويساوي (2,7183)، و(D) مقدار ثابت قيمته (1,7) ويشير إلى عامل القياس (سليمان، 2009، 79-81).

ويجب أن تتم معايرة جميع المفردات الاختبارية في مستودع المفردات وفقًا لنظرية الاستجابة للمفردة، والتي تقوم بوظيفتين أساسيتين هما: إنتاج درجات قابلة للمقارنة بين الطلاب الذين خضعوا لاختبارات مختلفة تمامًا، وتحديد المفردات الأكثر ملاءمة للمفحوصين، وهذا يعد عنصرًا مهمًا في الاختبار الموائم المحوسب، حيث تتحدد هذه المفردات تلقائيًا بناءً على المعلومات التي تم استخلاصها من المفردات السابقة، وتتطلب جميع نماذج النظرية توفير عينات كبيرة من بيانات الإجابة لكل مفردة قبل أن يتم تضمينها في المستودع، ولذلك فإن بيانات مستودع المفردات الأولية ينبغي تجميعها ومعايرتها قبل تطبيق برنامج الاختبار الموائم المحوسب (Davey, 2011,7؛ Tian & Dai, 2020,2).

وتعد نظرية القياس الكلاسيكية غير مناسبة للاختبارات الموائمة، فمؤشرات الصدق والثبات وصلاحية المفردة التقليدية تكون مناسبة عندما يجيب جميع المختبرين عن نفس المفردات الاختبارية، بينما عند معرفة بارامترات كل مفردة في الاختبار فإن نظرية الاستجابة للمفردة تتيح تقديرات قدرة المختبر على أساس استجابته على كل مفردة، واختيار المفردة المثلى للتطبيق من الاختبار على أساس ذلك التقدير، وتحديد الدرجة الرقمية التي تمثل قدرة المختبر

في نهاية الاختبار، ولذلك تتيح مميزات نظرية الاستجابة للمفردة تطوير الاختبارات الموائمة المحوسبة وتوفر طرقًا وتقنيات جديدة لمعادلة الاختبار ودراسة التحيز (الحسيني، 2019، 99؛ Aybek & Gulleroglu, 2021,129).

الخطوة الثالثة: تأسيس بنك أسئلة محوسب يشتمل على مجموعة كبيرة من المفردات التي سبق تعييرها باستخدام النموذج المستخدم:

يعد بنك الأسئلة مكانًا آمنًا توضع فيه مجموعات متنوعة من الأسئلة ذات مستويات مختلفة، ويسهل عن طريقه سحب أو إضافة مجموعة من الأسئلة المختلفة التي لها خصائص سيكومترية مميزة ومعلومة، وتخزن الأسئلة في ذاكرة حاسب آلي وفق برنامج سابق للتجهيز، ويتم

معايرتها عن طريق نماذج معينة وباستخدام برامج خاصة (مراد وسليمان، 2005، 402).

ويتطلب تنفيذ الاختبارات الموائمة المحوسبة تأسيس بنك أسئلة يحتوي على مجموعة كبيرة من المفردات معلومة البارامترات التي تعرف المتغير موضوع القياس إجرائيًا، مما يتيح الفرصة لاستخدام صورة مختصرة من الاختبارات تضم عددًا قليلًا من المفردات تناسب مستوى الطلاب، وتحقق أهداف القياس، مما يساعد على سرعة عملية التقويم (الحسيني، 2019، 76).

وينبغي أن يشتمل بنك الأسئلة على عدد كافٍ من المفردات المميزة تمييزًا مرتفعًا، وتنشتر قيم صعوبتها انتشارًا واسعًا على مدى هذه السمة، وربما يتطلب أن يشتمل البنك على (100) مفردة ثنائية الدرجة على الأقل، وذلك عندما يكون الهدف من الاختبار هو التوصل إلى قياسات دقيقة للأفراد على المدى الكلي للسمة، وإذا كان الهدف من الاختبار هو تصنيف الأفراد في مجموعتين إحداهما متمكنة والأخرى غير متمكنة مثل اختبارات الإتقان أو الكفاءة، فإن البنك ينبغي أن يشتمل على مفردات تقترب قيم صعوبتها من درجة القطع، وهذا يجعل منحنى الدالة المعلوماتية للاختبار مُدببًا، أي يتمركز تجاه درجة القطع (Sahin & Weiss, 2015, 1586؛ Embretson & Reise, 2000, 264-265).

الخطوة الرابعة: تصميم نظام تزويدي لاستدعاء المفردات الاختبارية بسرعة من بنك الأسئلة وعرضها على شاشة الحاسوب للفرد المختبر:

تعتمد الاختبارات الموائمة المحوسبة على مجموعة من الخوارزميات التي يكرر تطبيقها لاختيار المفردة التالية التي ينبغي استدعاؤها لتقدير القدرة، وتتمثل هذه الخوارزميات في: (أ) تقدير جميع المفردات التي لم يتم قياسها لتحديد أيها أفضل للقياس بعد ذلك بما يتوافق مع مستوى قدرة المختبر الحالية، (ب) تقديم المفردة التالية الملائمة لقدرته ثم يجيب عنها، (ج) حساب القدرة الجديدة المقيمة بناءً على الاستجابات على جميع المفردات المقاسة، (د) تكرار الخطوات السابقة حتى يتم الوصول إلى قاعدة إنهاء الاختبار (Cisar et al., 2010, 141؛ Chang, 2005,221).

الخطوة الخامسة: تحديد آليات انتقاء المفردة الأولية عند بدء العملية الاختبارية:

يعد الاختيار المعقول لبدء الاختبار هو البدء بتطبيق مفردة متوسطة الصعوبة، حيث تتراوح القيمة التقديرية لصعوبتها بين (-0,5، +0,5)، عندما يتوزع مجتمع المختبرين توزيعًا اعتداليًا، وفي حالة توفر بعض المعلومات المسبقة عن موقع المختبر على متصل القدرة المقاسة، فإنه يمكن استخدام هذه المعلومات في تحديد مستوى صعوبة المفردة الأولى، وقد يرغب بعض المختبرين البدء بمفردة سهلة لكي تحفزه على الاستمرار في الاختبار دون تأثره بعامل قلق الاختبار (Bjorner et al., 2007, 98-99؛Embretson & Reise, 2000, 266).

وتختار المفردة التالية بما يلائم قدرة كل مختبر، حيث يتم اختيار درجة صعوبتها بحيث لا تكون صعبة جدًا أو سهلة جدًا بالنسبة له، أي يكون احتمال إجابته عنها بشكل صحيح (50%)، ومع كل إجابة عن المفردات التالية، يصبح الحاسوب قادرًا بشكل متزايد على تقييم معرفة المختبر وتقدير قدرته، لتحديد المفردة الأفضل في التطبيق بعد ذلك، ولذلك يُعد الاختبار الموائم المحوسب نوعًا من الاختبارات التي تم تطويرها لزيادة كفاءة عملية تقدير معرفة المختبرين، وذلك عن طريق تعديل المفردات المقدمة إليهم بناءً على إجاباتهم السابقة أثناء الاختبار(Cisar et al. , 2010,140؛ Davey & Pitoniak, 2006,545).

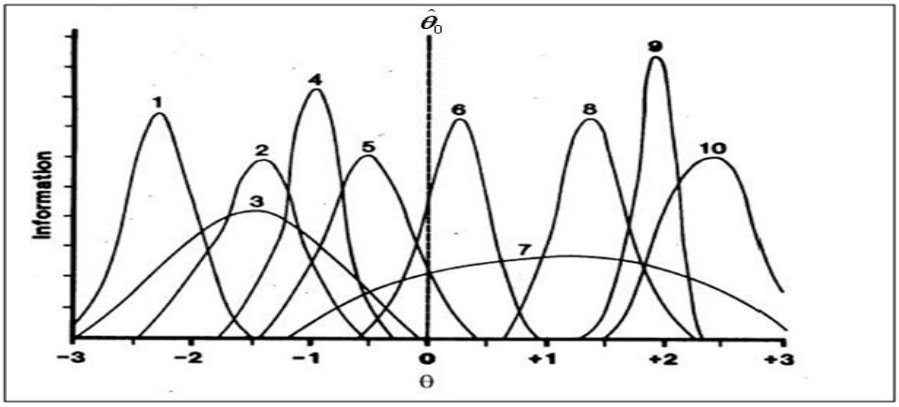
الخطوة السادسة: تحديد استراتيجية انتقاء المفردة التالية استنادًا إلى درجة الفرد في المفردة السابقة:

تعد استراتيجية المعلومات القصوى أكثر استراتيجيات اختيار المفردات شيوعًا، حيث تُقدر قدرة المختبر بناءً على استجاباته على المفردات التي تم قياسها لتحديد مقدار المعلومات التي توفرها المفردات المتبقية في بنك الأسئلة، ثم يتم تحديد المفردة التي توفر أكبر قدر من المعلومات لقياسها، وتكرر هذه العملية عادة بعد قياس كل مفردة، وتتوافق هذه الاستراتيجية مع تقدير القدرة بطريقة الأرجحية القصوى، ويمكن اختيار المفردات باستراتيجية تستند إلى نظرية بييز، حيث يتم اختيار المفردة التي تقلل من الانحراف المعياري المتوقع للمختبر، وتتوافق هذه الاستراتيجية مع تقدير القدرة بطريقة بييز (Embretson & Reise,2000, 267 ؛ Boyed, 2003,32-33).

وتُحدد المفردة التالية في المراحل الأولية من الاختبارات الموائمة المحوسبة من خلال قاعدة الخطوة، فإذا تمت الإجابة عن المفردة الأولى بشكل صحيح، يتم زيادة المقدار الأصلي السابق للقدرة بمقدار ما (على سبيل المثال 0,50)، وإذا تمت الإجابة عنها بشكل غير صحيح، يتم تقليل المقدار الأصلي السابق للقدرة بنفس المقدار، وبعد قياس كل مفردة وتسجيلها، تستخدم القدرة الجديدة لتحديد المفردة التالية، والتي يتم اختيارها من كل المفردات الموجودة في بنك الأسئلة التي لم يتم قياسها لهذا المختبر، من خلال تحديد المفردة التي توفر أكبر قدر من المعلومات للقدرة الحالية (Weiss,2004,75).

ويوضح الشكل (1) اختيار المفردة باستراتيجية المعلومات القصوى لعشر مفردات، حيث يشير موقع مركز دالة معلومات المفردة إلى صعوبتها، وارتفاع دالة معلومات المفردة إلى تمييزها، وعدم تماثل دالة معلومات المفردة إلى مقدار التخمين الزائف، وتعد المفردة الأولى أسهل مفردة؛ ولذلك فإن دالة معلومات المفردة لها توجد على الطرف الأيسر من القدرة (θ)، والمفردة العاشرة أصعب مفردة، والمفردة التاسعة أكثر المفردات تمييزًا؛ ولذلك فإن لديها أعلى دالة

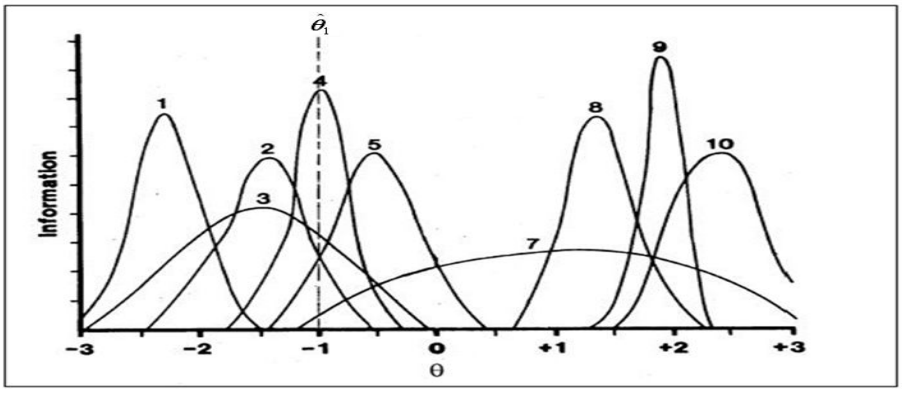
معلومات، والمفردة السابعة الأقل تمييزًا، ولا تحتوي أي من هذه المفردات على بارامتر تخمين زائف مرتفع لأن جميع دوال معلومات المفردات متماثلة بشكل معقول، وتُحسب قيم المعلومات لجميع المفردات عند متوسط مقياس القدرة أي عند (Θ=0)، حيث يتضح أن المفردة السادسة توفر أكبر قدر من المعلومات عندما تكون (Θ=0) ولذلك يتم قياسها وتسجيلها، وبافتراض الإجابة عنها بشكل غير صحيح، فإن القدرة الثانية تساوي (-1) على افتراض أن حجم الخطوة= (1)، ويتم اختيار وقياس المفردة الرابعة كما في الشكل (2)، وبافتراض الإجابة عنها بشكل صحيح، فإن القدرة الثالثة تساوي (-0,5)، ويتم اختيار وقياس المفردة الخامسة كما في الشكل (3)، ويستمر تقدير القدرة واختيار المفردة حتى يتم الوصول إلى قاعدة إنهاء الاختبار (Weiss,2011,15-17).



شكل (1)

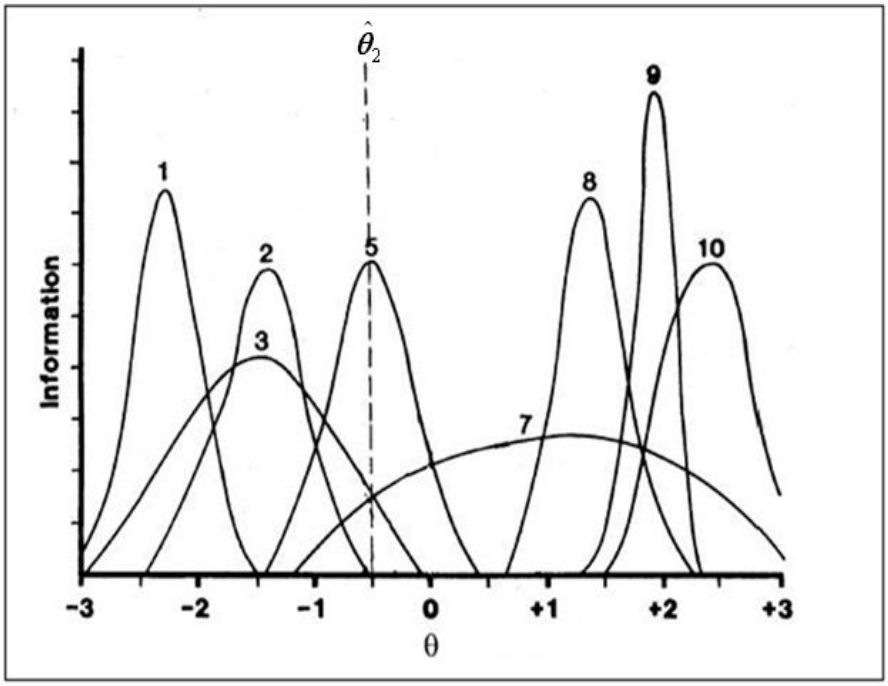
بنك أسئلة يحتوي على (10) مفردات لبدء الاختبار الموائم المحوسب

باستراتيجية المعلومات القصوى (Weiss,2011,16)



شكل (2)

تطبيق المفردة الأولى باستراتيجية المعلومات القصوى (Weiss,2011,16)



شكل (3)

تطبيق المفردة الثانية باستراتيجية المعلومات القصوى

(Weiss,2011,17)

وقام الباحث باستخدام استراتيجية المعلومات القصوى لأنها أكثر استراتيجيات اختيار المفردات شيوعًا، كما أن برنامج FastTest يعتمد عليها عند تطبيق الاختبارات الموائمة المحوسبة.

الخطوة السابعة: اختيار أسلوب تقدير القدرة المناسب:

يمكن تقدير القدرة بطريقة الأرجحية القصوى، وطريقة بييز التي تتضمن أسلوب التقدير البعدي المتوقع Expected A Posterior (EAP)، وأسلوب التقدير البعدي الأقصى Maximum A Posterior (MAP)، وتعتمد الأرجحية القصوى على نمط إجابة المختبر، ولا تسمح بتقدير قدرته إلا بعد أن يجيب عن مفردة واحدة على الأقل، ولذلك يمكن تطبيق قاعدة الخطوة لحساب الدرجات في بداية الاختبارات الموائمة المحوسبة، فإذا كان مستوى السمة الكامنة لدى المختبر (صفر) مثلًا، وأجاب عن المفردة الأولى إجابة صحيحة، فإن الخوارزمية ترفع تقدير مستوى سماته بحجم خطوة ثابت بمقدار (0,25) مثلًا، والعكس صحيح، وتستمر هذه العملية حتى يصبح تقدير القدرة بالأرجحية القصوى ممكنًا، ثم تستخدم بعد ذلك لحساب درجات المختبر عن الاستجابات التالية، بينما تفترض الأساليب البييزية توزيعًا قبليًا للقدرة، وتعدل هذا التوزيع بناءً على استجابة المختبر على المفردات الاختبارية (Reise,2000,266-267 & Embretson؛ Cella et al., 2007, 139).

وتقوم الأرجحية القصوى بتسجيل وتقدير مستويات قدرة المختبر من خلال قياس بارامترات الصعوبة والتمييز والتخمين لكل مفردة، وتأخذ في الاعتبار أيضًا إجابته عن المفردة بشكل صحيح، وتوفر العديد من الفروق بين المختبرين أكثر من مجرد تصحيح المفردات، نتيجة لدمج المعلومات المتعلقة بنمط استجاباتهم مع خصائص كل مفردة، فعلى سبيل المثال، يمكن أن يؤدي تصحيح الاختبار التقليدي الذي يتكون من عشرة مفردات إلى (11) درجة على الأكثر (من صفر إلى 10)، بينما يمكن أن تؤدي الأرجحية القصوى لنفس الاختبار إلى (1024) تقدير مختلف للقدرة، كما تقدم خطأ معياري للقياس لكل مختبر، اعتمادًا على كيفية إجابته عن مجموعة معينة من المفردات، والتي يمكن تحديدها من أي مجموعة فرعية من المفردات التي استجاب لها، بشرط أن يتم تقدير بارامترات هذه المفردات على نفس المقياس (Weiss,2004,73,75).

وتم اقتراح العديد من الأساليب للتغلب على عدم إمكانية استخدام الأرجحية القصوى في بداية الاختبارات الموائمة المحوسبة لتقدير القدرة، ومنها: (1) ثبيتها عند قيمة صغيرة أو كبيرة حتى يتم الحصول على تقديرات محددة، (2) تأجيل تقديرها حتى يتم الاستجابة لمجموعة كبيرة من المفردات، (3) استخدام طريقة بيز، (4) استخلاص التقديرات الأولية لها من المعلومات التجريبية المتوفرة عن المختبرين، ولا يوجد أسلوب مقبول بصفة عامة، فالأسلوبان الأول والثاني يتضمنان اختيارًا افتراضيًا لقيمة القدرة والمفردات على التوالي، ويعتمد الأسلوب الثالث على اختيار التوزيع القبلي الذي يتحكم تمامًا في اختيار المفردة الأولى عند عدم وجود بيانات الاستجابة، وإذا كان هذا التوزيع بعيدًا عن قدرة المختبر الحقيقية يؤدي إلى نتائج عكسية وينتج سلسلة أولية طويلة من الاستجابات الصحيحة أو غير الصحيحة، ولا توجد اعتراضات تقنية في الأسلوب الرابع، ولكن قد يرجع بعض المختبرين درجات الاختبار التي حصلوا عليها إلى المفردات السهلة أو الصعبة التي حصلوا عليها في البداية بسبب المعلومات السابقة، وتظهر مشكلة استنتاج تقدير القدرة الأولية فقط في الاختبارات القصيرة مثل الاختبارات المكونة من عشرة مفردات (Van der Linden & Pashley, 2010,9).

ويتضح للباحث أن تقدير القدرة في الاختبارات الموائمة المحوسبة يُجرى على ثلاث مراحل وهي: (1) عند البدء في إجراءات اختيار المفردات، (2) أثناء الاختبار ليتلاءم اختيار المفردات مع قدرة المختبر، (3) في نهاية الاختبار لتقدير قدرته النهائية.

الخطوة الثامنة: تحديد قاعدة إنهاء الاختبار:

تُقدر مستوى قدرة المختبر والخطأ المعياري لهذا التقدير عقب كل استجابة في الاختبارات الموائمة المحوسبة، ثم يختار الحاسوب المفردة التالية لقياسها، ولكن هذا لا يستمر إلى ما لانهاية، ولذلك تحتاج خوارزمية هذه الاختبارات إلى قاعدة لإنهاء الاختبار (Embretson & Reise,2000,267).

ويعتبر قرار إنهاء الاختبارات الموائمة المحوسبة هو العنصر الأكثر أهمية، فإذا كان الاختبار قصيرًا جدًا، فقد يكون تقدير القدرة غير دقيق، وإذا كان طويلًا جدًا، فإنه سيتم إهدار الوقت والموارد، والكشف عن المفردات دوان داعٍ، وإرهاق المختبر، وانخفاض مستوى أدائه، مما يؤدي إلى نتائج غير صادقة (Jian-quan et al. ,2007,77).

وتتميز الاختبارات الموائمة المحوسبة بأنه يمكن تغيير قاعدة إنهاء الاختبار طبقًا للهدف منه، حيث تستخدم بعضها للاختيار أو للتصنيف، مثل اختيار الطلاب الذين سيتم قبولهم في المدرسة أو الجامعة، أو تصنيفهم حسب درجة إتقانهم لبعض مجالات التحصيل، وبالتالي يعتمد إنهاؤها على مقارنة درجة المختبر بقيمة درجة القطع، وعندما لا تستخدم في التصنيف يتم تطبيق قاعدة إنهاء مختلفة، حيث يفضل في هذه الحالة قياس مدى وصول كل مختبر إلى مستوى الدقة المطلوب، كما هو محدد مسبقًا بالخطأ المعياري للقياس(≤ 0,25 لوجيت)، ويؤدي ذلك إلى قياسات متوازنة بحيث يحصل المختبرين على درجات دقيقة متكافئة (Weiss,2011,17-18 ؛ Weiss, 2004,75-76).

ويشير Jian-quan et al. (2007,77-78), and Linacre (2000,11-12) إلى أن الاختبار الموائم المحوسب يتوقف عندما: (أ) تُطبق جميع مفردات بنك الأسئلة على المختبر، ويحدث هذا مع بنوك الأسئلة الصغيرة، (ب) يتم الوصول إلى الحد الأقصى المحدد مسبقًا لطول الاختبار، (ج) تُقدر القدرة المقاسة بدقة كافية، حيث يصبح الخطأ المعياري للقياس منخفضًا (0,2 لوجيت) (د) يكون قياس القدرة بعيدًا بما فيه الكفاية عن محك النجاح والفشل، ويحدث ذلك عندما يكون تقدير القدرة بعيدًا عن المحك على الأقل بـ (2) خطأ معياري للتقدير، أو عندما لا يتبقى مفردات كافية في الاختبار لتغيير قرار النجاح والفشل للمختبر، (هـ) يُظهر المختبر سلوكًا خارج الاختبار، مثل اختيار إجابة غير ذات صلة أو اختيار نمط معين في الاستجابة، أو الاستجابة بسرعة كبيرة، أو ببطء شديد، ولا يمكن أن يتوقف قبل: (أ) تقديم الحد الأدنى من المفردات مثل (10) أو (20) مفردة على الأقل بغض النظر عن أداء المختبر، (ب) تغطية كل مجال في موضوع الاختبار، (ج) قياس المفردات الكافية التي تحقق صدق الاختبار.

ويرى Choi et al. (2011,38), and Wang et al. (2019, 750) أن محك الطول المتغير يعتمد على قاعدتين لإنهاء الاختبار، هما: (1) قاعدة الإنهاء بالخطأ المعياري وهي الأكثر استخدامًا حيث يتم إنهاء الاختبار عندما يصل الخطأ المعياري المصاحب لأحدث تقدير لقدرة المختبر إلى أقل من قيمة محددة سلفًا، ومن مميزات هذا القاعدة أنه عندما تكون دالة معلومات بنك الأسئلة مسطحة نسبيًا، فإنه عادة ما تكون دقة القياس مكافئة تقريبًا عبر متصل القدرة، ومن عيوبها أنها قد تحد من كفاءة الاختبار الموائم المحوسب، وتقلل من جدوى استخدامه، وتزيد من عبء المختبرين من خلال تطبيق عدد من المفردات غير الضروية التي لا تفي بمحك الخطأ المعياري المحدد مسبقًا، كما قد تحد من دقة القياس حيث توقف الاختبار بالرغم من أن المفردات الغنية بالمعلومات لا تزال متاحة للتطبيق، (2) قاعدة الإنهاء بأدنى معلومات حيث يتم وقف الاختبار عندما لا يكون هناك مزيد من المفردات المتاحة القادرة على توفير حد أدنى محدد مسبقًا من المعلومات للمختبر في أحدث تقدير للسمات، وتتمثل الميزة الرئيسة لهذه القاعدة في أنها لا تعطي للمختبر مفردات غير ضرورية أو لا تحقق دقة قياس مرتفعة، ويتمثل العيب الرئيس فيها في أنها تقدم دقة قياس أقل من قاعدة الخطأ المعياري بسبب منعها تطبيق مفردات إضافية عندما تكون المعلومات المتاحة قليلة نسبيًا.

رابعًا: مميزات الاختبارات الموائمة المحوسبة:

من خلال اطلاع الباحث على بعض الكتابات النظرية والدراسات والبحوث السابقة التي أشارت إلى مميزات الاختبارات الموائمة المحوسبة مثل: (علام، 2005، 300-302؛ Cisar et al., 2010, 141؛ Embretson & Reise,2000,268 ؛ Jian-quan et al.,2007, Stafford et al., 2019,1306؛ Tian & Dai, 2020,8؛ Weiss,2004,77-78) أمكن تلخيص تلك المميزات كما يأتي:

- زيادة دقة القياس: تعطي قياسات أكثر دقة رغم أنها تكون عادة أقصر من الاختبار التقليدي بنسبة (50%) أو أكثر بدون التأثير على الخصائص السيكومترية للاختبار، ويرجع ذلك إلى الانتقاء الأمثل للمفردات التي تضاهي صعوبتها مستوى قدرة المختبر.
- توفير الوقت بشكل كبير: لأنها تحتوي على عدد قليل من المفردات اللازمة لتحقيق الدقة المقبولة، مما يُقلل من التعب والإرهاق والجهد الذي يبذله المختبر، مما يؤثر بشكل كبير على نتائج اختباره، كما أنها أسرع في التطوير والتنفيذ والحصول على النتائج.
- يتم إجراؤها عند الطلب وبشكل فردي: حيث لا يضطر المختبر إلى انتظار انتهاء زملائه قبل الانتقال إلى القسم التالي، فيستجيب وفقًا لسرعته الذاتية المناسبة لقدرته، كما يمكن إعطاؤه وقتًا إضافيًا عندما يحتاج إليه، مما يقلل من مصدر قلق الاختبار.
- الأداء المنتج للمختبر: حيث تجعله نشطًا ومنتجًا، وتستثير دافعيته للأداء الجيد، وتزوده بخبرات أفضل، حيث يعتمد تقديم إحدى المفردات على استجابته للمفردة السابقة، مما يمثل تحديًا له دون التقليل من دافعيته.
- تعتمد على نظرية الاستجابة للمفردة عند تحديد مفردات الاختبار؛ لأنها توفر معلومات قصوى لتقدير كفاءة المختبر.
- تجنب قياس المفردات غير ذات الصلة: حيث يقوم الحاسوب بمراقبة القيم التقديرية لصعوبة المفردات وتمييزها بصورة مستمرة، لأن المفردات السهلة جدًا أو الصعبة جدًا تثير سلوكًا غير مرغوب فيه لدى المختبر، مثل: التخمين، والإهمال، وأنماط الاستجابة.
- التغذية المرتدة الفورية: فاستجابة المختبر لإحدى المفردات تحدد مستوى صعوبة المفردة التالية؛ ولذلك فإنه يكون على علم فوري بصحة أو خطأ استجابته على المفردات، وهذا لا يتوافر في الاختبارات التقليدية.
- تنوع صيغ وأنواع المفردات الاختبارية التي تشتمل عليها هذه الاختبارات، مثل: التمثيلات البيانية، والرسوم المتحركة، والوسائط المتعددة، وأقراص الفيديو، مما يضفي ديناميكية على العملية الاختبارية.

- توفر درجات دقيقة عبر مجموعة واسعة من القدرات، بينما تكون الاختبارات التقليدية عادة أكثر دقة للمختبرين ذوي القدرة المتوسطة، كما أنها تسجل الدرجات بصورة فورية وتحفظها في قاعدة بيانات.
- التغلب على مشكلة سرية الاختبارات: حيث تكون المفردات الاختبارية التي يتم تخزينها في الحاسوب أكثر أمانًا من حفظها في أدراج أو كبائن مغلقة، كما أن توافر عدد كبير من المفردات في بنك الأسئلة المختزن في الحاسوب يجعل ارتفاع درجة الطالب ارتفاعًا مصطنعًا نتيجة علمه السابق بعدد قليل من مفردات البنك أمرًا غير محتمل.
- التشابه الكبير بين درجاتها ودرجات الاختبارات التقليدية: على الرغم من أنه لا ينبغي توقع ارتباطات مرتفعة جدًا بينهما؛ لأن الاختبارات التقليدية لها قيود يتم التغلب عليها بواسطة الاختبارات الموائمة المحوسبة.
- تمكن المعلمين من معرفة طلابهم: من خلال درجات الاختبارات ومستويات قدرتهم، لأنها تُحدد خصائص الأسئلة التي يمكنهم الإجابة عنها، والأسئلة التي لا يستطيعون الإجابة عنها، كما أنها تحدد مستوى القدرة النهائية للطلاب، وبالتالي يمكن تزويدهم بالتعليم الفردي الملائم لهم.
- عدم احتياجها إلى أوراق إجابة ولا مدراء للاختبار، حيث يتم التخلص من اختلاف المراقبين على الاختبار كعامل في خطأ القياس.

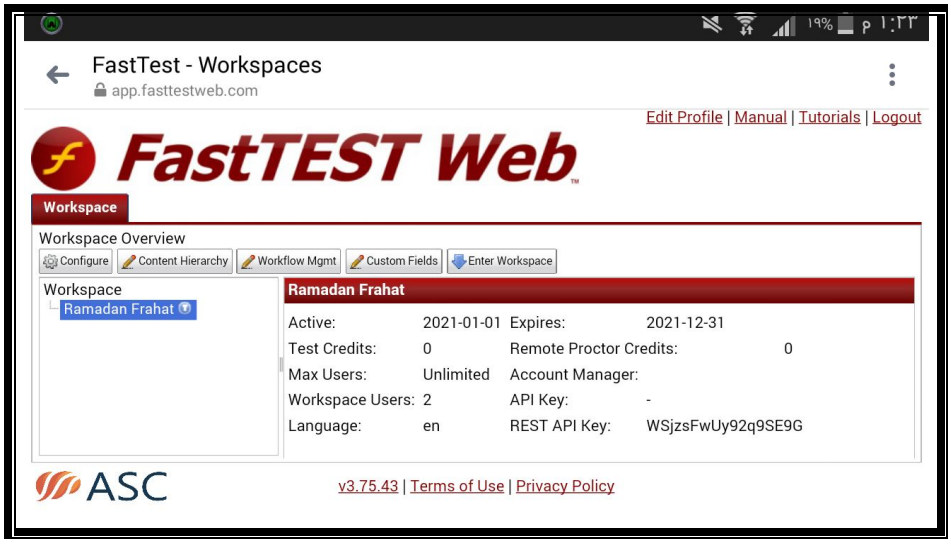
خامسًا: مشكلات الاختبارات الموائمة المحوسبة:

من خلال اطلاع الباحث على بعض الكتابات النظرية والدراسات والبحوث السابقة التي أشارت إلى بعض المشكلات المنهجية والعلمية، والعيوب والقيود والقضايا الفنية والإجرائية للاختبارات الموائمة المحوسبة مثل: (علام، 2005، 323؛ Cisar et al., 2010, 141؛ Magis et al., 2017,4؛ Runder, 1998) أمكن تلخيص تلك المشكلات كما يأتي:

- لا يوجد اتفاق حول أفضل استراتيجيات انتقاء المفردات، وهذه الاستراتيجيات المختلفة تؤدي أحيانًا إلى نتائج متباينة وربما غير دقيقة.
- تتطلب إجراءات تطبيقها إنشاء بنك أسئلة يشتمل على عدد كبير من المفردات تختزن في الحاسوب.
- تحتاج المشكلات المتعلقة بتأثير سياق عرض المفردات، وترتيبها، وتوازن محتواها إلى مزيد من الدراسات والبحوث التطويرية.
- لا تنطبق على جميع الموضوعات والمهارات، حيث تستند إلى نماذج نظرية الاستجابة للمفردة التي لا تنطبق على جميع أنواع المفردات والمهارات.
- تقتصر على بعض أنواع المفردات التي يمكن قياسها بواسطة الحاسوب، فقد يكون من الصعب تقديم المفردات التي تتطلب أعمال فنية تفصيلية ورسوم بيانية.

- تتطلب معايرة دقيقة للمفردات باستخدام أحد نماذج نظرية الاستجابة للمفردة، وهذا يتطلب خبرة فنية عالية وجهدًا كبيرًا وتمويلًا إضافيًا.
- تتطلب توفر عدد كاف من أجهزة الحاسوب في المؤسسة يتناسب مع عدد المختبرين، وأن يكون لديهم دراية بالحاسوب.
- يمكن أن تسبب مشكلات لبعض المختبرين لأن إجراءات القياس فيها مختلفة.
- عدم مساواة بين المختبرين نتيجة تلقيهم مجموعات مختلفة من الأسئلة.
- عدم السماح للمختبرين بتغيير الإجابات، كما أن المختبر الذكي قد يجيب إجابة خاطئة عن الأسئلة الأولية بشكل متعمد، وسيفترض البرنامج بعد ذلك أن لديه قدرة منخفضة فيختار له سلسلة من الأسئلة السهلة.

وتم التغلب على بعض المشكلات السابقة في الإصدار الحالي من برنامج الاختبارات الموائمة المحوسب FastTest حيث يتم إدخال بنك الأسئلة على موقع البرنامج دون الحاجة إلى إدخاله في كل أجهزة الحواسيب التي تستخدم في تطبيق الاختبار، كما يمكن تطبيق الاختبار على الهاتف المحمول لكل مختبر من خلال كود معين يرسله الباحث إليه، ويسمح للمختبر بتغيير إجابته عندما يكون إنهاء الاختبار الموائم المحوسب وفقًا لعدد محدد من المفردات، ويوضح شكل (4) حساب الباحث على موقع Fast Test.



شكل (4)

حساب الباحث على موقع FastTest

وقامت بعض الدراسات والبحوث السابقة بتطبيق الاختبارات الموائمة المحوسبة والتحقق من فاعليتها في العديد من المجالات، وسوف يركز الباحث على الدراسات والبحوث السابقة

التي تناولت تقدير القدرة التحصيلية للطلاب في أحد المقررات الدراسية من خلال الاختبارات الموائمة المحوسبة، فأجرى Triantafillou et al. (2008) بحثًا هدف إلى تصميم وتقييم اختبار موائم محوسب على الأجهزة المحمولة، وبلغ عدد المشاركين في البحث (12) طالبًا من طلاب الصف الثاني الثانوي، وتكونت أدوات البحث من بنك للأسئلة بإجمالي (80) سؤالًا في وحدة الكهرباء من مقرر الفيزياء، ومقياس الاتجاه نحو الاختبار الموائم المحوسب، واستخدم البحث النسبة المئوية لتحليل البيانات، وتم تطبيق كل من الاختبار الخطي والاختبار الموائم المحوسب على جميع المشاركين، وتم تقدير القدرة باستخدام طريقة الأرجحية القصوى كما تم تحديد عدد (20) مفردة كقاعدة لإنهاء الاختبارات أو الوصول إلى أدنى خطأ معياري (0,3)، وتوصل البحث إلى أن الاختبار الموائم المحوسب أكثر فاعلية من الاختبار الخطي في تقدير قدرة الطلاب.

وهدف بحث Ozyurt and Ozyurt, (2015) إلى تطوير نظام الاختبارات الموائمة المحوسبة في وحدة الاحتمالات من مادة الرياضيات لطلاب الصف الحادي عشر وتقييمها من حيث الثبات، وبلغ عدد المشاركين في البحث (84) طالبًا من طلاب الصف الحادي عشر بمدينة طرابزون بتركيا، واستخدم البحث خمسة اختبارات مختلفة لوحدة الاحتمالات تتعلق بالتباديل، والتوافيق، والتوزيع الثنائي، والاحتمال، واختبار على الوحدة بإجمالي (752) سؤالًا تم توزيعها كما يأتي (239) للتباديل، و(159) للتوافيق ، و(102) للتوزيع الثنائي، و(252) للاحتمال ، وتم دمج نظام الاختبار الموائم المحوسب في UZWEBMAT وهي بيئة تعليمية ديناميكية يمكنها تكييف نفسها مع أساليب التعلم وخصائص تعلم الطلاب، حيث تم دمج نظام الاختبارات الموائمة المحوسبة في هذه البيئة كنموذج للتقييم والقياس تحت مسمى UZWEBMAT-CAT ، وتم تقدير القدرة باستخدام طريقة الأرجحية القصوى كما تم تحديد عدد (15) مفردة كقاعدة لإنهاء الاختبارات ما عدا اختبار الوحدة تم تحديد (20) مفردة، وتوصل البحث إلى تمتع الاختبارات الموائمة المحوسبة بثبات مرتفع حيث بلغت قيمته للاختبارات الخمس (0,93، 0,93، 0,88، 0,93، 0,91).

وقام Martin and Lazendic (2018) ببحث هدف إلى دراسة أثر استخدام الاختبار الموائم المحوسب متعدد المراحل والاختبار الخطي المحوسب في الحساب على الدافعية والاندماج واستكشاف خبرة الطلاب الشخصية في الاختبار عبر الانترنت، وبلغ عدد المشاركين في البحث (12736) مشاركًا من المدارس الابتدائية (الصفين الثالث والخامس) والثانوية (الصفين السابع والتاسع) بإستراليا، تم تقسيمهم إلى مجموعتين حيث طبق الاختبار الخطي المحوسب على المجموعة الأولى وعددها (7152) مشاركًا بينما طبق الاختبار الموائم المحوسب متعدد المراحل على (5584) مشاركًا على المجموعة الثانية، بالتعاون مع هيئة المناهج والتقييم والتقارير الاسترالية التي دعت إلى مشاركة المدارس في جميع الولايات في تطبيق هذا الاختبار عبر الانترنت تحت إشراف ملاحظين خارجيين، واستخدم البحث اختبارًا في الحساب، ومقياس الدافعية، والاندماج (المشاركة)، وخبرة الاختبار الذاتي عبر الانترنت، واستخدم البحث البرنامجين الإحصائين R, M*plus* ، وتوصل البحث إلى أن الاختبار الموائم المحوسب أكثر دقة في قياس التحصيل.

وهدف بحث السعدني (2019) إلى تصميم اختبار إلكتروني بأنماطه (الموائم- الوسطي- الخطي)، وقياس أثره في تنمية التحصيل في مقرر المتاحف والمعارض التعليمية، وخفض مستوى القلق من الاختبارات الالكترونية (غير الطبيعي- المرضي)، وبلغ عدد المشاركين في

البحث (90) طالبًا من طلاب الفرقة الأولى شعبة تكنولوجيا التعليم بكلية التربية النوعية جامعة بورسعيد، واستخدم البحث اختبارًا تحصيليًا للمعارف والمهارات المتضمنة في الوحدات الخمسة الأولى من مقرر المتاحف والمعارض التعليمية، ومقياس القلق من الاختبارات الالكترونية، واستخدم البحث الأساليب الإحصائية الآتية للتحقق من فروض البحث: اختبار "ت" لعينتين مرتبطتين، وتحليل التباين الأحادي والثنائي، واختبار توكي، وتم تحديد قاعدة (50 %) من عدد الأسئلة كقاعدة لإنهاء الاختبار، وتوصل البحث إلى وجود فروق ذات دلالة إحصائية بين متوسطات درجات الطلاب في الاختبار التحصيلي لصالح الاختبار الموائم.

وأجرى نور الدين (2019) بحثًا هدف إلى التعرف على مدى فاعلية القياس الموائم المحوسب في قياس تحصيل طلاب الجامعة، وبلغ عدد المشاركين في البحث (25) طالبًا من طلاب الدبلوم العام بكلية التربية، جامعة الجوف بالمملكة العربية السعودية، واستخدم البحث اختبارًا تحصيليًا مكونًا من (48) سؤالًا من نمط الاختيار من متعدد في مقرر القياس والتقويم، بينما تكون الاختبار الموائم المحوسب من (20) سؤالًا من أسئلة الاختبار التحصيلي الورقي، واستخدم البحث البرنامج الإحصائي FastTest Professional للتحقق من فروض البحث، وتم تقدير القدرة باستخدام طريقة الأرجحية القصوى كما تم تحديد عدد (20) مفردة كقاعدة لإنهاء الاختبار، وتوصل البحث إلى فاعلية الاختبار الموائم المحوسب على الاختبار الورقي التقليدي أو الخطي حيث كان متوسط دالة المعلومات التي يوفرها (7,45)، ومتوسط الخطأ المعياري (0,39) ومعامل ثباته (0,85)، بينما بلغ متوسط دالة المعلومات للاختبار الورقي (5,47)، ومتوسط الخطأ المعياري (0,44) ومعامل ثباته (0,80).

وقام Samsudin et al. (2019) ببحث هدف إلى تقييم كفاءة الاختبار الموائم المحوسب في قياس أداء تلاميذ الصف الثامن على الاختبارات الدولية للعلوم (TIMSS)، وبلغ عدد المشاركين في البحث (68) تلميذًا تم اختيارهم من مدارس المنطقة الشمالية بماليزيا، وتم تقسيمهم إلى مجموعتين حيث تكونت المجموعة الأولى من (33) تلميذًا لا يزالون في مرحلة تعلم مادة العلوم للمرحلة الإعدادية والتي تعادل الصف الثامن حيث طبق النموذج (2)، بينما تكونت المجموعة الثانية من (35) طالبًا تعلموا مادة العلوم للمرحلة الإعدادية ويتعلمون الآن منهج العلوم للمرحلة الثانوية حيث طبق عليهم النموذج (4)، واستخدم البحث بنكًا للأسئلة في العلوم مكونًا من (122) مفردة تم اشتقاقها من أسئلة الاختبارات الدولية في العلوم والتي أصدرت ما بين عامي 2003- 2015م، وتم تقدير القدرة باستخدام طريقة الأرجحية القصوى كما تم تحديد عدد (20) مفردة كقاعدة لإنهاء الاختبار، واستخدم البحث البرنامج الإحصائي R من خلال منصة Concerto لإعداد وتطبيق الاختبار الموائم المحوسب، واختبار "ت"، وتوصل البحث إلى أن تطبيق (20) مفردة موضوعية في الاختبار الموائم المحوسب ذي الطول الثابت أنتج خطأ معياريًا أقل من أو يساوي (0,5) مما يشير إلى زيادة كفاءة التقييم بواسطة الاختبار الموائم المحوسب مع تطبيق أقل للمفردات، كما تبين وجود فروق ذات دلالة إحصائية في القدرة بين المجموعة الأولى والثانية لصالح المجموعة الثانية مما يدل على قدرة الاختبار الموائم المحوسب على التمييز بين مستويين من المعرفة.

وهدف بحث Istiyono et al. (2020) إلى تطوير اختبار موائم محوسب لقياس مهارات التفكير العليا وهي التحليل والتركيب والتقويم والابتكار في مادة الفيزياء، وقياس اتجاهات المعلمين والطلاب نحو تطبيق الاختبارات الموائمة المحوسبة، وبلغ عدد المشاركين في البحث (300) طالبًا من طلاب الصف العاشر بالمدرسة الثانوية العليا بمدينة يوجياكرتا بأندونسيا، و(10) من معلمي الفيزياء، وتكونت أدوات البحث من بنك للأسئلة يحتوي على (62) مفردة، واستبيان، وتم تحليل درجات الاختبار باستخدام نموذج التقدير الجزئي بينما تم تحليل الاستجابة على الاستبيانات وصفياً، وتوصل البحث إلى أن الاختبار الموائم المحوسب مناسباً لقياس مهارات التفكير العليا في الفيزياء كما أن (82,28%) من المعلمين والطلاب أشاروا إلى صلاحية تطبيق الاختبار الموائم المحوسب لقياس تلك المهارات.

وأجرى Hogenboom et al. (2021) بحثًا هدف إلى التحقق من صدق اختبار موائم محوسب في فهم مفاهيم البرمجة باستخدام نظرية الاستجابة للمفردة، وبلغ عدد المشاركين في البحث (93341) طفلًا من تلاميذ المدارس الابتدائية الهولندية تتراوح أعمارهم بين (4-13) عامًا، وتكونت أدوات البحث من اختبار مفاهيم البرمجة الذي تم تطبيقه عبر الانترنت، وتم استخدام تقديرات قدرة التلاميذ وصعوبة مفردات الاختبار للتحقق من صدق اختبار مفاهيم البرمجة الموائم المحوسب وفقًا لمنهج أو نظام التصميم المعرفي لـ Embretson، حيث يستلزم هذا التصميم استخدام المكونات المعرفية الكامنة وراء كل مفردة كمؤشرات لصعوبة المفردة بدلًا من تطبيق التحليل التقليدي بناءً على اختبارات الصدق الخارجية، وتوصلت نتائج البحث إلى أن اختبار مفاهيم البرمجة الموائم المحوسب يتمتع بخصائص سيكومترية جيدة.

وهدف بحث Yasuda et al.(2021) إلى فحص الطول الأمثل لاختبار موائم محوسب يقيس الفهم المفاهيمي لميكانيكا نيوتن وخاصة فيما يتعلق بمفهوم القوة في الفيزياء، وبلغ عدد المشاركين في البحث (86) طالبًا من الملتحقين بالدورات التمهيدية في الفيزياء بإحدى جامعات اليابان، وتكونت أدوات البحث من اختبار مفهوم القوة، وتم تقدير القدرة باستخدام طريقة بييز وتحديد قاعدة إنهاء الاختبار بعدد محدد من المفردات يتراوح بين (1-30) مفردة، وتم حساب التحيز والخطأ المعياري وجذر متوسط مربع خطأ لكوهين "d" للاختبار الموائم المحوسب لمقارنته بمؤشرات الاختبار التقليدي بالنسبة لحجم الصف، وتوصلت نتائج البحث إلى أنه يمكن تقليل طول الاختبار الموائم المحوسب إلى (15-19) مفردة إذا كان حجم الصف (40)، وتقليل زمن الاختبار إلى (50- 63%) مع انخفاض في الدقة والكفاءة بنسبة (5- 10%) فقط.

واتضح للباحث أن الدراسات والبحوث السابقة استخدمت الاختبار الموائم المحوسب في قياس وتقدير القدرة التحصيلية في بعض المقررات الدراسية مثل: الفيزياء، والرياضيات، والمتاحف والمعارض التعليمية، والقياس والتقويم، والعلوم، والبرمجة، وتراوحت أعداد المشاركين بين (12- 93341) مشاركًا، وتنوعت المراحل الدراسية التي اختير منها المشاركون، واعتمدت على طريقة واحدة لتقدير القدرة بالأرجحية القصوى أو طريقة بييز، وقاعدة واحدة لإنهاء الاختبار بعدد محدد من المفردات يتراوح بين (15-20) مفردة أو بأدنى خطأ معياري (0,3)، وتوصلت نتائجها إلى أن الاختبار الموائم المحوسب أكثر فاعلية في تقدير القدرة، ويتمتع بخصائص سيكومترية مرتفعة، كما تبين للباحث أنه لا توجد دراسات أو بحوث سابقة (في حدود ما اطلع عليه الباحث) حاولت بحث الفروق بين الاختبارات الموائمة المحوسبة والاختبارات الخطية المحوسبة في تقدير القدرة باختلاف طريقة تقدير القدرة (الأرجحية

القصوى- بييز)، واختلاف قاعدتي إنهاء الاختبار (بعدد محدد من المفردات- بأدنى خطأ معياري).

فروض البحث:

يسعى البحث الحالي إلى اختبار الفروض الآتية:

1- لا توجد فروق ذات دلالة إحصائية بين متوسطي تقدير القدرة والخطأ المعياري لتقدير القدرة (كل على حدة) للاختبار الخطي المحوسب والاختبار الموائم المحوسب باستخدام طريقة الأرجحية القصوى لتقدير القدرة وقاعدة إنهاء الاختبار بعدد محدد من المفردات.

2- لا توجد فروق ذات دلالة إحصائية بين متوسطي تقدير القدرة والخطأ المعياري لتقدير القدرة (كل على حدة) للاختبار الخطي المحوسب والاختبار الموائم المحوسب باستخدام طريقة الأرجحية القصوى لتقدير القدرة وقاعدة إنهاء الاختبار بأدنى خطأ معياري.

3- لا توجد فروق ذات دلالة إحصائية بين متوسطي تقدير القدرة والخطأ المعياري لتقدير القدرة (كل على حدة) للاختبار الخطي المحوسب والاختبار الموائم المحوسب باستخدام طريقة بييز لتقدير القدرة وقاعدة إنهاء الاختبار بعدد محدد من المفردات.

4- لا توجد فروق ذات دلالة إحصائية بين متوسطي تقدير القدرة والخطأ المعياري لتقدير القدرة (كل على حدة) للاختبار الخطي المحوسب والاختبار الموائم المحوسب باستخدام طريقة بييز لتقدير القدرة وقاعدة إنهاء الاختبار بأدنى خطأ معياري.

منهج البحث:

استخدم الباحث المنهج الوصفي المقارن لحساب الفروق في تقدير القدرة بين الاختبار الموائم المحوسب والاختبار الخطي المحوسب، ويتم تقدير القدرة من خلال حساب كل من متوسطات تقدير القدرة، والخطأ المعياري لتقدير القدرة.

المشاركون:

بلغ عدد المشاركين في البحث الحالي (1606) مشاركًا، وتم تقسيمهم على النحو الآتي:

1- المشاركون في حساب صدق وثبات المفردات الاختبارية وتقدير بارامتراتها:

قام الباحث بتطبيق المفردات الاختبارية على (1506) مشاركًا تم اختيارهم بطريقة عشوائية، من طلاب الفرقة الثالثة بكلية التربية بنين جامعة الأزهر بالقاهرة، شعب الدراسات الإسلامية واللغة العربية والتاريخ والجغرافيا والمكتبات وتكنولوجيا المعلومات والتاريخ الطبيعي والكيمياء والطبيعة متوسط أعمارهم (20,80) سنة بانحراف معياري (1,06)؛ لتقدير بارامترات

المفردات الاختبارية المتمثلة في معاملات الصعوبة والتمييز والتخمين لكل مفردة، حيث قام الباحث بتطبيق المفردات إلكترونيًا في قاعات الدراسة بالتعاون مع أستاذ المقرر*.

2- المشاركون في البحث الأساسي:

قام الباحث بتطبيق الاختبار الخطي المحوسب والاختبار الموائم المحوسب باستخدام برنامج FastTest على (100) مشاركًا تم اختيارهم بطريقة عشوائية، من طلاب الفرقة الثالثة شعبة الكيمياء والطبيعة بكلية التربية بنين جامعة الأزهر بالقاهرة متوسط أعمارهم (20,66) سنة بانحراف معياري (0,78).

أداة البحث:

استخدم الباحث بنكًا للأسئلة تم إعداده وفقًا للمراحل الآتية:

أولًا: مرحلة التخطيط: وتشمل هذه المرحلة الخطوات الآتية:

أ- تحديد المقرر الدراسي:

قام الباحث باختيار مقرر علم النفس التعليمي (1) المقرر على جميع شعب الفرقة الثالثة بكلية التربية جامعة الأزهر (عدا شعبة الخدمة الاجتماعية)، ثم اختار الباحث الفصول الأربعة الأولى (مدخل إلى علم النفس التربوي- التعلم- نظرية بافلوف- نظرية ثورنديك) حتى يتمكن الباحث من صياغة عدد من المفردات الاختبارية لا تقل عن (100) مفردة، ويكون لدى الباحث الوقت الملائم لتنفيذ إجراءات البحث قبل نهاية الفصل الدراسي.

ب- تحديد الهدف من بنك الأسئلة:

حدد الباحث الهدف من بنك الأسئلة وهو إعداد مجموعة من الأسئلة الاختبارية ذات خصائص سيكومترية معلومة لقياس التحصيل الأكاديمي في بعض فصول مقرر علم النفس التعليمي (1) في ضوء بعض مستويات بلوم المعرفية (التذكر – الفهم – التطبيق) لطلاب الفرقة الثالثة بكلية التربية جامعة الأزهر.

ج- تحليل محتوى المقرر الدراسي:

قام الباحث بتحليل محتوى الفصول التعليمية المختارة إلى مجموعة من الحقائق والمفاهيم والتعميمات والقوانين، كما في ملحق (1).

د- صياغة وتصنيف الأهداف التعليمية المطلوب تحقيقها:

قام الباحث بتحديد الأهداف التعليمية للفصول الأربعة، ثم صاغها في صورة أهداف إجرائية، وبلغ عددها (140) هدفًا، وصنفت هذه الأهداف إلى (51) هدفًا في مستوى التذكر، و(49) هدفًا في مستوى الفهم، و(40) هدفًا في مستوى التطبيق كما في ملحق (2)، واختار

* يتوجه الباحث بالشكر والتقدير إلى السادة أعضاء هيئة التدريس أ. م. د/ زين العابدين محمد علي وهبة، ود/ محمد عبد العزيز العلاف، ود/ ربيع شعبان حسن، بقسم علم النفس التعليمي والإحصاء التربوي بكلية التربية بنين جامعة الأزهر بالقاهرة على مساعدة الباحث في تطبيق المفردات الاختبارية على الطلاب المشاركين.

الباحث المستويات الثلاث لبلوم في ضوء نتائج تحليل المحتوى للفصول التعليمية المختارة، وتركيز معظم أسئلة الاختبارات والأنشطة والتدريبات المتضمنة في الكتاب المقرر لمادة علم النفس التعليمي (1) على المستويات الثلاثة.

هـ- إعداد جدول المواصفات للمفردات الاختبارية:

قام الباحث بإعداد جدول المواصفات لتحقيق التوازن بين المفردات الاختبارية والتأكد من أنها تقيس عينة ممثلة لمحتوى الفصول والأهداف التعليمية المراد قياس التحصيل فيها، ويوضح جدول (2) مواصفات المفردات الاختبارية في ضوء محكات عدد الساعات وعدد الأهداف ومستوياتها في كل فصل من الفصول التعليمية المختارة.

جدول (2)

جدول مواصفات المفردات الاختبارية في ضوء محكات عدد الساعات وعدد الأهداف ومستوياتها

الفصول التعليمية	عدد الساعات	أهداف التذكر	أهداف الفهم	أهداف التطبيق	نسبة الموضوع	عدد مفردات التذكر	عدد مفردات الفهم	عدد مفردات التطبيق	عدد مفردات الفصل
مدخل إلى علم النفس التربوي	2	7	6	5	%13,33	7	7	5	19
التعلم	4	14	13	13	%26,67	14	13	11	38
نظرية بافلوف	4	14	14	11	%26,27	14	13	11	38
نظرية ثورنديك	5	16	16	11	%33,33	17	16	12	45
المجموع	15	51	49	40	%100	52	49	39	140
مجموع الأهداف			140		المجموع %100		عدد المفردات الاختبارية 140 سؤال		
نسبة الأهداف		%36,43	%35	%28,57					

ويتضح من جدول (2) أن عدد المفردات الاختبارية يساوي (140) مفردة موزعة وفقًا لمستويات بلوم الثلاثة الدنيا على نحو يتوافق مع توزيع الأهداف السلوكية على هذه المستويات، وقد وزعت المفردات كما يلي: (52) مفردة تقيس مستوى التذكر، (49) مفردة تقيس مستوى الفهم، (39) مفردة تقيس مستوى التطبيق.

ويوضح جدول (3) توزيع المفردات الاختبارية في ضوء جدول المواصفات

جدول (3)

توزيع المفردات الاختبارية في ضوء جدول المواصفات

الفصول التعليمية	مفردات التذكر		مفردات الفهم		مفردات التطبيق		عدد المفردات
	العدد	أرقام المفردات	العدد	أرقام المفردات	العدد	أرقام المفردات	
مدخل إلى علم النفس التربوي	7	1، 2، 3، 4، 5، 6، 7	7	53، 54، 55، 56، 57، 58، 59	5	102، 103، 104، 105، 106	19
التعلم	14	8، 9، 10، 11، 12، 13، 14، 15، 16، 17، 18، 19، 20، 21	13	60، 61، 62، 63، 64، 65، 66، 67، 68، 69، 70، 71، 72	11	107، 108، 109، 110، 111، 112، 113، 114، 115، 116، 117	38
نظرية بافلوف	14	22، 23، 24، 25، 26، 27، 28، 29، 30، 31، 32، 33، 34، 35	13	73، 74، 75، 76، 77، 78، 79، 80، 81، 82، 83، 84، 85	11	118، 119، 120، 121، 122، 123، 124، 125، 126، 127، 128	38
نظرية ثورنديك	17	36، 37، 38، 39، 40، 41، 42، 43، 44، 45، 46، 47، 48، 49، 50، 51، 52	16	86، 87، 88، 89، 90، 91، 92، 93، 94، 95، 96، 97، 98، 99، 100، 101	12	129، 130، 131، 132، 133، 134، 135، 136، 137، 138، 139، 140	45
المجموع	52		49		39		140

ثانيًا: مرحلة بناء مجموعة من المفردات الاختبارية:

وتشمل هذه المرحلة الخطوات الآتية:

أ- تحديد أنواع المفردات الاختبارية:

صاغ الباحث المفردات بطريقة الاختيار من متعدد لأنها تتميز بالمرونة ويمكن تصحيحها بسهولة وموضوعية، وتقيس بكفاءة عالية نواتج التعليم لمختلف الأهداف التعليمية في كثير من المقررات الدراسية، كما أنها تقلل من التخمين، ويمكن أن تغطي مساحة تعليمية كبيرة من المحتوى.

ب- صياغة تعليمات الإجابة عن المفردات الاختبارية:

صاغ الباحث التعليمات بحيث توضح الهدف من التطبيق، وعدد المفردات، ونوعيتها، وعدد بدائل الإجابات، وكيفية الإجابة عن المفردات.

ج- صياغة المفردات الاختبارية في ضوء جدول المواصفات بحيث تقيس الأهداف التي تم تحديدها في المرحلة الأولى وإعداد مفتاح التصحيح:

قام الباحث بصياغة (140) مفردة من نوع الاختيار من متعدد في ضوء جدول المواصفات، ويكون لكل مفردة إجابة واحدة صحيحة من بين أربعة بدائل مستقلة بعضها عن بعض، بحيث تعطى الدرجة (واحد) للإجابة الصحيحة لكل مفردة، في حين تعطى الدرجة (صفر) للإجابة الخطأ.

د- تحكيم المفردات من خلال عرضها على بعض الأساتذة المحكمين[1] لإعطاء نوع من التغذية المرتدة حول تحليل محتوى الفصول التعليمية، وصحة الصياغة العلمية واللغوية للأهداف، وملاءمة توزيع الأهداف في ضوء تصنيف بلوم، وملاءمة صياغة المفردات الاختبارية لطريقة الاختيار من متعدد، وتمثيل المفردات الاختبارية لمستويات الأهداف التي وضعت لقياسها، وتغطية المفردات الاختبارية لموضوعات الفصول المختارة، واستقلالية المفردات الاختبارية عن بعضها البعض، وإضافة أو حذف أو تعديل المفردات الاختبارية، ويوضح ملحق (3) بنك الأسئلة بعد إجراء التعديلات التي اقترحها السادة المحكمون في صياغة بعض المفردات.

ثالثًا: مرحلة التجريب والتعيير للمفردات الاختبارية:

وتم فيها تطبيق المفردات الاختبارية على (1506) مشاركًا من طلاب الفرقة الثالثة بكلية التربية بنين جامعة الأزهر بالقاهرة، لحساب صدقها وثباتها، والتحقق من افتراضات نظرية الاستجابة للمفردة، وتقدير بارامتراتها المتمثلة في الصعوبة والتمييز والتخمين لكل مفردة، وهذا يتضح في الخطوات الآتية:

أ- التحقق من صدق المفردات الاختبارية:

قام الباحث بإجراء التحليل العاملى لمصفوفة الارتباط بطريقة المكونات الأساسية باستخدام البرنامج الإحصائي SPSS 25، وأخذ الباحث بمحك جيلفورد لمعرفة حد الدلالة الإحصائية للتشبعات وهو اعتبار التشبعات التي تصل إلى (0,30) أو أكثر تشبعات دالة، وتم حساب مدى كفاية حجم العينة لإجراء التحليل العاملي باستخدام اختبار Kaiser KMO Test) (Kaiser- Meyer- Olkin Measure of Sampling Adequacy –حيث تتراوح قيمة هذا الاختبار بين الصفر والواحد الصحيح، وبلغت قيمته في تحليل هذا المقياس (0,928) أي أكبر من قيمة الحد الأدنى الذي اشترطه Kaiser (0,50) وبالتالي فإنه يمكن الحكم بكفاية حجم العينة لإجراء التحليل العاملي، كما تم التأكد من ملاءمة المصفوفة للتحليل العاملي بحساب اختبار بارتليت Bartlett's Test حيث كان دال إحصائياً عند مستوى (0,01)، وتشبعت المفردات الاختبارية على

[1] **يتوجه الباحث بالشكر والتقدير إلى السادة الأساتذة أعضاء هيئة التدريس بقسم علم النفس التعليمي والإحصاء التربوي بكلية التربية بنين جامعة الأزهر بالقاهرة لتحكيمهم أداة البحث وهم أ.د/ حسين عبد العزيز الدريني، وأ.د/ سيف الدين يوسف عبدون، وأ.د/ أحمد مهدي مصطفى، وأ.د/ أحمد محمد شبيب، وأ.د/ جود السيد جودة شاهين، وأ.د/ محمد سعد محمد علي، وأ.م.د/ إبراهيم سيد أحمد عبد الواحد.**

العوامل المستخلصة ما عدا المفردات رقم (20، 33، 39، 40، 47، 53، 108، 120، 121، 135) فلم تتشبع على أي عامل من العوامل المستخلصة ولذلك تم حذفها من بنك الأسئلة.

ب- التحقق من ثبات المفردات الاختبارية:

قام الباحث بحساب ثبات المفردات الاختبارية بطريقتي:

1- كيودر- ريتشاردسون 20: حيث بلغ معامل الثبات (0,870) وهي نسبة مرتفعة تدل على ثبات الاختبار.

2- التجزئة النصفية: حيث بلغ معامل الارتباط بين نصفي الاختبار (0,793) وباستخدام معادلة التصحيح لسبيرمان وبراون بلغ معامل الثبات للاختبار ككل (0,884)، وهي نسبة مرتفعة تدل على أن الاختبار يتمتع بقدر مرتفع من الثبات.

ج- التحقق من افتراضات نظرية الاستجابة للمفردة:

1- التحقق من افتراض أحادية البعد:

تم إخضاع المفردات الاختبارية في بنك الأسئلة للتحليل العاملي بطريقة المكونات الأساسية باستخدام البرنامج الإحصائي SPSS 25؛ إذ تعد المكونات الأساسية الناتجة عن التحليل العاملي مؤشرًا لأحادية البعد عندما لا تقل نسبة التباين المفسر على العامل الأول بالمقارنة بنسبة التباين المفسر التراكمي للعوامل المستخلصة عن (20%) طبقًا لمحك Reckase.

ويوضح جدول (4) نسبة التباين المفسر للعوامل الثلاثة الأولى والعامل الأخير لبنك الأسئلة.

جدول (4)

قيم التباين المفسر للعوامل الثلاثة الأولى والعامل الأخير التي تشبعت بها المفردات الاختبارية

العامل	نسبة التباين المفسر	نسبة التباين المفسر التراكمي
1	13,478	13,478
2	8,187	21,665
3	4,346	26,011
32	0,723	58,540

يتضح من جدول (4) أن نسبة التباين المفسر للعامل الأول إلى التباين المفسر التراكمي لجميع للعوامل المستخلصة =(23,02%) أي أن بنك الأسئلة يحقق محك Reckase، وهذا يعد مؤشرًا على أنه يقيس عاملًا واحدًا.

2- التحقق من افتراض الاستقلال الموضعي:

ويقصد به أن احتمال الاجابة الصحيحة للفرد عن مفردة اختبارية يكون مستقلًا عن ناتج إجابته عن أي مفردة أخرى في الاختبار، عند ضبط كل من القيمة التقديرية لقدرته وصعوبة المفردة؛ ولذلك فإن الاستقلال الموضعي يشير إلى أن مفردات الاختبار تكون غير مرتبطة.

وتم التحقق من هذا الافتراض باستخدام البرنامج الإحصائي jMetrik 4.1.1 بحساب (Q3)وهو يشير إلى معاملات الارتباط للبواقي لزوج من المفردات بعد ضبط السمة المقدرة باستخدام البرنامج الإحصائي، ومن خلال فحص مصفوفة معاملات الارتباط للبواقي بين أزواج المفردات الاختبارية، اتضح أن جميع معاملات الارتباط للبواقي بين أزواج المفردات صغيرة وقد كانت قيمة متوسطات (Q3) لبنك الأسئلة (0,02) أي قريبة من الصفر، وهذا مؤشر على الاستقلال الموضعي.

3- التحقق من افتراض التحرر من السرعة:

قام الباحث بإعطاء الطلاب المشاركين الوقت الكافي للإجابة عن المفردات الاختبارية بحيث لا يؤثر عامل السرعة في أدائهم، بحيث يرجع عدم إجابتهم عن بعض المفردات الاختبارية إلى محدودية قدرتهم لا بسبب عدم كفاية الوقت للوصول إلى المفردة والإجابة عنها، وأنهى جميع الطلاب الإجابة عن المفردات ضمن الوقت المحدد للإجابة وهو ساعتان.

د- التحقق من مطابقة المفردات الاختبارية للنموذج ثلاثي البارامتر:

قام الباحث باختيار النموذج ثلاثي البارامتر لأنه حافظ على أكبر عدد من المفردات بالمقارنة بالنموذجين الأحادي والثنائي، وأعطى أكبر قيمة لدالة المعلومات، بالإضافة إلى أن مفردات الاختيار من متعدد قد تتأثر بالتخمين وهو ما أكدته نتائج التعيير حيث تراوحت قيمة بارامتر التخمين بين (0,06-0,5)، وأظهرت نتائج تحليل المفردات الاختبارية عدم مطابقة (28) مفردة أرقامها (18، 19، 27، 32، 38، 44، 57، 61، 62، 64، 66، 71، 75، 80، 85، 89، 91، 102، 104، 111، 113، 115، 118، 122، 123، 125، 126، 132) في بنك الأسئلة حيث كانت النسبة بين قيم كا^2 ودرجات الحرية أكبر من (5) وتم حذف هذه المفردات، وبذلك أصبح عدد المفردات في بنك الأسئلة بصورته النهائية (102) مفردة.

هـ- معايرة المفردات الاختبارية: تم تقدير بارامترات الصعوبة والتمييز والتخمين لكل مفردة اختبارية وفق النموذج ثلاثي البارامتر باستخدام البرنامج الإحصائي jMetrik 4.1.1 كما في جدول (5).

جدول (5)

قيم بارامترات الصعوبة والتمييز والتخمين للمفردات الاختبارية

المفردة	الصعوبة	التمييز	التخمين	المفردة	الصعوبة	التمييز	التخمين	المفردة	الصعوبة	التمييز	التخمين
1	2,75-	0,48	0,5	45	1,35-	1,06	0,2	92	0,1-	1,49	0,18
2	1,93-	0,56	0,15	46	1,32-	1,09	0,11	93	1,61	2,86	0,2
3	1,35-	0,88	0,15	48	0,6-	0,88	0,06	94	1,52	2,88	0,25
4	1,66-	0,73	0,11	49	1,09	3,88	0,26	95	3,18	0,82	0,17
5	1,1	2,21	0,43	50	1,22-	1,26	0,1	96	1,46	2,84	0,21
6	3,3	0,82	0,37	51	0,55-	0,77	0,07	97	1,56	2,86	0,18

المفردة	الصعوبة	التمييز	التخمين	المفردة	الصعوبة	التمييز	التخمين	المفردة	الصعوبة	التمييز	التخمين
7	1,34	2,8	0,3	52	0,59-	1,56	0,5	98	0,37-	1,17	0,5
8	3,97	0,72	0,18	54	2,07	1,97	0,2	99	3,43	0,82	0,21
9	1,26-	1,01	0,12	55	3,71	0,82	0,22	100	2,03	2,66	0,13
10	1,09-	1,23	0,17	56	1,24	2,88	0,33	101	3,98	1,58	0,32
11	3,68	0,82	0,38	58	3,96	0,82	0,21	103	1,49	2,9	0,18
12	1,39	2,73	0,28	59	3,2	0,82	0,15	105	0,98-	0,75	0,08
13	1-	1,13	0,26	60	0,41-	1,1	0,16	106	3,37	0,82	0,21
14	1,39	2,81	0,39	63	3,83	0,82	0,17	107	3,59	0,82	0,29
15	2,23-	1,14	0,5	65	3,65	3,4	0,29	109	1,43	2,83	0,28
16	1,5	2,85	0,44	67	0,44-	1,35	0,21	110	3,05	0,82	0,23
17	1,42-	0,95	0,15	68	3,74	0,82	0,29	112	1,4	2,91	0,15
21	3,67-	0,14	0,2	69	3,01	0,82	0,06	114	3,98	1,49	0,07
22	3,57-	0,07	0,23	70	3,98	1,72	0,07	116	1,24	2,91	0,3
23	3,34	0,82	0,31	72	1,37-	0,94	0,08	117	0,57-	1,52	0,5
24	0,56-	1,13	0,5	73	2,74	0,21	0,11	119	3,9	0,82	0,19
25	1,32-	1,08	0,1	74	1,27	2,91	0,1	124	0,65-	1,02	0,07
26	3,57-	0,82	0,27	76	1,15	2,88	0,25	127	0,6-	1,32	0,5
28	1,69	2,82	0,29	77	3,02	0,82	0,14	128	3,66	0,82	0,1
29	1,59	2	0,15	78	0,51-	1,59	0,09	129	3,33	0,82	0,2
30	0,95-	1,65	0,5	79	4,21	0,11	0,2	130	0,29-	0,9	0,22
31	3,69	0,82	0,2	81	2,52	0,18	0,13	131	2,66	0,21	0,1
34	1,37	0,22	0,13	82	3,34	0,82	0,15	133	3,05	0,82	0,15
35	3,97	1,37	0,29	83	1,31	2,89	0,35	134	3,98	1,58	0,38
36	3,84	0,82	0,19	84	2,83	0,09	0,21	136	3,97	0,82	0,29
37	1,16-	1,72	0,19	86	1,34-	1,08	0,17	137	3,13	0,82	0,08
41	2,15-	1,06	0,12	87	1,5	2,9	0,2	138	0,5	1,38	0,32
42	0,68-	1,21	0,5	88	1,28	2,91	0,23	139	1,47	2,72	0,25
43	0,95-	1,06	0,1	90	3,82	0,82	0,23	140	3,66	3,56	0,33

رابعًا: مرحلة التخزين والتصنيف:

قام الباحث بتخزين الأسئلة في الحاسوب باستخدام برنامج FastTest 3.75.39 وتصنيفها وفق نظام تصنيفي يتضمن: نوع السؤال، والهدف الذي يقيسه، ومحتواه، والصف المناسب له، وبارامترات كل سؤال، وتاريخ تجريبه وتعييره، وتاريخ استخدامه.

خامسًا: مرحلة الاستخدام:

قام الباحث بتطبيق الأسئلة الاختبارية على (100) طالبًا من الفرقة الثالثة شعبة الكيمياء والطبيعة بأسلوبي الاختبار الخطي المحوسب والاختبار الموائم المحوسب.

خطوات تنفيذ البحث:

لتنفيذ البحث اتبع الباحث الخطوات الآتية:

1- تطبيق المفردات الاختبارية على (1506) مشاركًا لحساب صدق وثبات المفردات الاختبارية، والتحقق من افتراضات نظرية الاستجابة للمفردة، وتحديد مدى مطابقة المفردات الاختبارية للنموذج ثلاثي البارامتر، وتقدير بارامترات الصعوبة والتمييز والتخمين لكل مفردة اختبارية.

2- تطبيق بنك الأسئلة في صورته النهائية على (100) مشاركًا، حيث قام الباحث بتقسيم الطلاب المشاركين إلى أربع مجموعات بواقع اختبارين أحدهما خطي محوسب عدد مفرداته (70) مفردة والآخر موائم محوسب لكل مجموعة على النحو الآتي:

المجموعة الأولى:

تم تطبيق الاختبار الخطي المحوسب والاختبار الموائم المحوسب على (25) مشاركًا، بحيث يتم تقدير القدرة في الاختبار الموائم المحوسب بطريقة الأرجحية القصوي مع تحديد قاعدة إنهاء الاختبار عند الوصول إلى عدد محدد من المفردات (25) مفردة.

المجموعة الثانية:

تم تطبيق الاختبار الخطي المحوسب والاختبار الموائم المحوسب على (25) مشاركًا، بحيث يتم تقدير القدرة في الاختبار الموائم المحوسب بطريقة الأرجحية القصوي مع تحديد قاعدة إنهاء الاختبار عند الوصول إلى خطأ معياري (0,25) أو أقل.

المجموعة الثالثة:

تم تطبيق الاختبار الخطي المحوسب والاختبار الموائم المحوسب على (25) مشاركًا، بحيث يتم تقدير القدرة في الاختبار الموائم المحوسب بطريقة بييز مع تحديد قاعدة إنهاء الاختبار عند الوصول إلى عدد محدد من المفردات (25) مفردة.

المجموعة الرابعة:

تم تطبيق الاختبار الخطي المحوسب والاختبار الموائم المحوسب على (25) مشاركًا، بحيث يتم تقدير القدرة في الاختبار الموائم المحوسب بطريقة بييز مع تحديد قاعدة إنهاء الاختبار عند الوصول إلى خطأ معياري (0,25) أو أقل.

3- قياس تقديرات القدرة والخطأ المعياري لتقديرات القدرة للاختبار الخطي المحوسب والاختبار الموائم المحوسب.

4- إجراء المعالجة الإحصائية اللازمة في ضوء فروض البحث، وعرض النتائج وتفسيرها في ضوء الإطار النظري ونتائج البحوث السابقة.

الأساليب الإحصائية المستخدمة:

اعتمد الباحث على الأسلوب الإحصائي اختبار "ت" لعينتين مرتبطتين للتحقق من فروض البحث.

نتائج البحث ومناقشتها وتفسيرها:

يعرض الباحث في هذا الجزء نتائج فروض البحث ومناقشة النتائج وتفسيرها في ضوء الإطار النظري والدراسات والبحوث السابقة على النحو الآتي:

نتيجة الفرض الأول:

ينص الفرض الأول على أنه "لا توجد فروق ذات دلالة إحصائية بين متوسطي تقدير القدرة والخطأ المعياري لتقدير القدرة (كل على حدة) للاختبار الخطي المحوسب والاختبار الموائم المحوسب باستخدام طريقة الأرجحية القصوى لتقدير القدرة وقاعدة إنهاء الاختبار بعدد محدد من المفردات".

ويوضح جدول (6) نتائج اختبار "ت" لعينتين مرتبطتين للتحقق من صحة هذا الفرض

جدول (6)

قيمة "ت" ودلالتها الإحصائية للفروق بين متوسطي تقدير القدرة والخطأ المعياري لتقدير القدرة للاختبار الخطي المحوسب والاختبار الموائم المحوسب باستخدام طريقة الأرجحية القصوى لتقدير القدرة وقاعدة إنهاء الاختبار بعدد محدد من المفردات (ن= 25)

الاختبار	القياس	العدد	المتوسط	الانحراف المعياري	قيمة "ت"	مستوى الدلالة
تقدير القدرة	الخطي المحوسب	25	0,46	0,74	0,66	0,51 غير دالة
	الموائم المحوسب	25	0,34	0,52		
الخطأ المعياري لتقدير القدرة	الخطي المحوسب	25	0,37	0,006	43,65	**0,01 دالة
	الموائم المحوسب	25	0,27	0,009		

ت (0,05)= 2,06 ت (0,01)= 2,80

يتضح من جدول (6) أن قيمة "ت" المحسوبة للفروق بين متوسطي تقدير القدرة للاختبار الخطي المحوسب والاختبار الموائم المحوسب تساوي (0,66)، وهي أقل من قيمة "ت" الجدولية، فتكون "ت" المحسوبة غير دالة إحصائيًا؛ وهذا يعني قبول الفرض الصفري؛ أي أنه "لا توجد فروق ذات دلالة إحصائية بين متوسطي تقدير القدرة للاختبار الخطي المحوسب والاختبار

الموائم المحوسب باستخدام طريقة الأرجحية القصوى لتقدير القدرة وقاعدة إنهاء الاختبار بعدد محدد من المفردات".

ويتضح من جدول (6) أيضًا أن قيمة "ت" المحسوبة للفروق بين متوسطي الخطأ المعياري لتقدير القدرة للاختبار الخطي المحوسب والاختبار الموائم المحوسب تساوي (43,65)، وهي أكبر من قيمة "ت" الجدولية، فتكون "ت" المحسوبة دالة إحصائيًا؛ وهذا يعني رفض الفرض الصفري؛ أي أنه " توجد فروق ذات دلالة إحصائية بين متوسطي الخطأ المعياري لتقدير القدرة للاختبار الخطي المحوسب والاختبار الموائم المحوسب باستخدام طريقة الأرجحية القصوى لتقدير القدرة وقاعدة إنهاء الاختبار بعدد محدد من المفردات لصالح الاختبار الموائم المحوسب".

وتتفق هذه النتيجة مع بحث نور الدين (2019) الذي توصل إلى عدم وجود فروق ذات دلالة إحصائية بين متوسطي تقدير القدرة التحصيلية للاختبار الخطي الورقي والاختبار الموائم المحوسب، باستخدام طريقة الأرجحية القصوى لتقدير القدرة وتحديد قاعدة إنهاء الاختبار بـ (20) مفردة، كما أن متوسط الخطأ المعياري لتقدير القدرة للاختبار الموائم المحوسب كان أقل بصورة دالة عند مستوى دلالة (0,05) عن متوسط الخطأ المعياري لتقدير القدرة للاختبار الخطي الورقي.

كما تتفق أيضًا مع بحث Samsudin et al. (2019) الذي توصل إلى كفاءة الاختبار الموائم المحوسب في قياس أداء طلاب الصف الثامن على الاختبارات الدولية في مادة العلوم (TIMSS) باستخدام طريقة الأرجحية القصوى لتقدير القدرة وتحديد قاعدة إنهاء الاختبار بـ (20) مفردة.

ويدعم هذه النتيجة ما أشار إليه Ozyurt and Ozyurt (2015) من أن الاختبارات الموائمة المحوسبة والتي تم إعدادها في وحدة الاحتمالات من مادة الرياضيات لطلاب الصف الحادي عشر تتمتع بثبات مرتفع باستخدام طريقة الأرجحية القصوى لتقدير القدرة وتحديد قاعدة إنهاء الاختبار بـ (15) أو (20) مفردة.

بينما تختلف هذه النتيجة مع ما وضحه كل من Leroux and Stafford et al.(2019, 1306), Babcock and Weiss (2012, 1), and Dodd (2014, 213-214) أن قاعدة الطول الثابت لا تقيس السمات الكامنة للمختبرين بدقة متساوية، فتفتقد للاتساق في دقة تقديرات القدرة لهم، أي في الخطأ المعياري لتقدير قدرتهم عبر مدى القدرات.

وتدل هذه النتيجة على أن استخدام طريقة الأرجحية القصوى لتقدير القدرة في الاختبار الموائم المحوسب وتحديد قاعدة إنهاء الاختبار بعدد محدد من المفردات يعطي نفس مستوى تقدير القدرة التي يعطيها الاختبار الخطي المحوسب، كما أنه أكثر دقة في تقدير القدرة لأنه يقدم أخطاء معيارية أقل من الاختبار الخطي المحوسب، حيث تعتمد طريقة الأرجحية القصوى على اختيار المفردات التي توفر أكبر قدر من المعلومات عن التقدير الحالي لقدرة الفرد أثناء عملية القياس، فتعتمد على دالة المعلومات عند اختيار المفردات، مما يعطي أقل خطأ للقياس عند مستوى القدرة المستهدفة، كما أنها تختار المفردات في المراحل الأولية للاختبار الموائم المحوسب من خلال قاعدة الخطوة، حيث تحدد مقدار الانتقال على متصل القدرة لتحديد قدرة المختبر

لإعطائه مفردة جديدة تتناسب مع القدرة الجديدة له حتى يتم الوصول إلى نمط الاستجابة الصحيحة والخاطئة.

واختار الباحث (25) مفردة كمحك لانهاء الاختبار أي (35%) من عدد مفردات الاختبار الخطي المحوسب، وهو عدد ملائم يمكن من خلاله الحصول على معلومات كافية تمكن الباحث من الحكم على مستوى المختبر بدقة، لأنه إذا كان عدد المفردات قصيرًا جدًا، فإن الاختبار الموائم المحوسب يعطي تقديرات غير دقيقة للقدرة، وإذا كان طويلًا جدًا، فإنه يُرهق المختبر ويؤثر بالسلب على أدائه، مما يعطي تقديرات غير دقيقة أيضًا للقدرة.

نتيجة الفرض الثاني:

ينص الفرض الثاني على أنه "لا توجد فروق ذات دلالة إحصائية بين متوسطي تقدير القدرة والخطأ المعياري لتقدير القدرة (كل على حدة) للاختبار الخطي المحوسب والاختبار الموائم المحوسب باستخدام طريقة الأرجحية القصوى لتقدير القدرة وقاعدة إنهاء الاختبار بأدنى خطأ معياري".

ويوضح جدول (7) نتائج اختبار "ت" لعينتين مرتبطتين للتحقق من صحة هذا الفرض

جدول (7)

قيمة "ت" ودلالتها الإحصائية للفروق بين متوسطي تقدير القدرة والخطأ المعياري لتقدير القدرة للاختبار الخطي المحوسب والاختبار الموائم المحوسب باستخدام طريقة الأرجحية القصوى لتقدير القدرة وقاعدة إنهاء الاختبار بأدنى خطأ معياري (ن= 25)

الاختبار	القياس	العدد	المتوسط	الانحراف المعياري	قيمة "ت"	مستوى الدلالة
تقدير القدرة	الخطي المحوسب	25	0,51	0,76	0,59	0,56 غير دالة
	الموائم المحوسب	25	0,63	1,02		
الخطأ المعياري لتقدير القدرة	الخطي المحوسب	25	0,46	0,007	7,85	**0,01 دالة
	الموائم المحوسب	25	0,37	0,056		

ت (0,05)= 2,06 ت (0,01)= 2,80

يتضح من جدول (7) أن قيمة "ت" المحسوبة للفروق بين متوسطي تقدير القدرة للاختبار الخطي المحوسب والاختبار الموائم المحوسب تساوي (0,59)، وهي أقل من قيمة "ت" الجدولية،

فتكون "ت" المحسوبة غير دالة إحصائيًا؛ وهذا يعني قبول الفرض الصفري؛ أي أنه "لا توجد فروق ذات دلالة إحصائية بين متوسطي تقديرات القدرة للاختبار الخطي المحوسب والاختبار الموائم المحوسب باستخدام طريقة الأرجحية القصوى لتقدير القدرة وقاعدة إنهاء الاختبار بأدنى خطأ معياري".

ويتضح من جدول (7) أيضًا أن قيمة "ت" المحسوبة للفروق بين متوسطي الخطأ المعياري لتقدير القدرة للاختبار الخطي المحوسب والاختبار الموائم المحوسب تساوي (7,85)، وهي أكبر من قيمة "ت" الجدولية، فتكون "ت" المحسوبة دالة إحصائيًا؛ وهذا يعني رفض الفرض الصفري؛ أي أنه "توجد فروق ذات دلالة إحصائية بين متوسطي الخطأ المعياري لتقدير القدرة للاختبار الخطي المحوسب والاختبار الموائم المحوسب باستخدام طريقة الأرجحية القصوى لتقدير القدرة وقاعدة إنهاء الاختبار بأدنى خطأ معياري لصالح الاختبار الموائم المحوسب".

وتتفق هذه النتيجة مع ما وضحه Embretson and Reise (2000,267-268) من أن قاعدة إنهاء الاختبار الأكثر استخدامًا هي الاستمرار في تطبيق المفردات إلى أن يصبح الخطأ المعياري للقياس أقل من قيمة محددة معينة، ويتطلب تنفيذ ذلك أن يحتوي بنك الأسئلة على مفردات تقدم معلومات تكفي لتحقيق هذا المحك، وبصفة عامة تحقق هذه القاعدة أفضل استخدام لخوارزمية الاختبارات الموائمة المحوسبة، ولذلك يفضل استخدامها.

كما تتفق مع ما أشار إليه Triantafillou et al. (2008) من أن الاختبار الموائم المحوسب في مادة الفيزياء أكثر فاعلية من الاختبار الخطي في تقدير قدرة الطلاب باستخدام طريقة الأرجحية القصوى وباستخدام (20) مفردة كقاعدة لإنهاء الاختبارات أو الوصول إلى أدنى خطأ معياري (0,3).

ويدعم هذه النتيجة ما أشار إليه كل من Leroux (2019, 1306) and Stafford et al. (2014, 213-214) Babcock and Weiss (2012, 1), and Dodd إلى أن قاعدة الطول المتغير التي تعتمد على محك الخطأ المعياري للقياس تعطي نتائج أكثر دقة.

وتوصل عودة وعبيدات (2013، 1614) أيضًا إلى وجود فروق ذات دلالة إحصائية بين متوسطي تقدير القدرة العقلية والخطأ المعياري لتقدير القدرة العقلية (كل على حدة) باستخدام مصفوفات رافن للاختبار الخطي المحوسب والاختبار الموائم المحوسب لصالح الاختبار الموائم المحوسب باستخدام طريقة الأرجحية القصوى لتقدير القدرة العقلية وقاعدة إنهاء الاختبار بأدنى خطأ معياري (0,25).

كما توصل الكرامنة (2015، 142) إلى انخفاض قيمة الخطأ المعياري لتقدير القدرة العقلية من خلال اختبارات الذكاء المنطقي الرياضي الموائمة المحوسبة باستخدام طريقة الأرجحية القصوى وقاعدة إنهاء الاختبار بأدنى خطأ معياري (0,20).

وتشير هذه النتيجة إلى أن استخدام طريقة الأرجحية القصوى لتقدير القدرة في الاختبار الموائم المحوسب وتحديد قاعدة إنهاء الاختبار بأدنى خطأ معياري يعطي نفس مستوى تقدير القدرة التي يعطيها الاختبار الخطي المحوسب، كما أنه أكثر دقة في تقدير القدرة لأنه يقدم أخطاء معيارية أقل من الاختبار الخطي المحوسب، حيث تعتمد طريقة الأرجحية القصوى في

تقدير القدرة على الاحتمال الأكبر لنمط استجابات المختبر، وتتجنب قياس المفردات غير الضرورية، كما أن تقدير بارامترات قدرة الأفراد فيها يتميز ببعض الخصائص مثل: التوزيع الاعتدالي، والاتساق، والفاعلية النسبية، والكفاية.

واعتمد الباحث على قاعدة إنهاء الاختبار عند الوصول إلى خطأ معياري (0,25) أو أقل، وهذا يتفق مع الأطر النظرية ومعظم الدراسات والبحوث السابقة التي اعتمدت على تحديد قيمة الخطأ المعياري بحيث تتراوح بين (0,20-0,30) أو أقل كمحك لإنهاء الاختبار الموائم المحوسب، حيث تؤدي هذه القاعدة إلى تحقيق قياسات متوازنة بحيث يحصل جميع المختبرين على درجات دقيقة متكافئة، أي يتم تقدير قدرتهم بنفس الدقة، وتتميز هذه القاعدة بتثبيت قيمة الدقة عند مستوى معين مما ييسر المقارنة والتحليلات الإحصائية، ويعطي ارتباطات مرتفعة بين تقديرات القدرة وتقديرات الاختبار الكلي.

نتيجة الفرض الثالث:

ينص الفرض الثالث على أنه "لا توجد فروق ذات دلالة إحصائية بين متوسطي تقدير القدرة والخطأ المعياري لتقدير القدرة (كل على حدة) للاختبار الخطي المحوسب والاختبار الموائم المحوسب باستخدام طريقة بييز لتقدير القدرة وقاعدة إنهاء الاختبار بعدد محدد من المفردات".

ويوضح جدول (8) نتائج اختبار "ت" لعينتين مرتبطتين للتحقق من صحة هذا الفرض.

جدول (8)

قيمة "ت" ودلالتها الإحصائية للفروق بين متوسطي تقدير القدرة والخطأ المعياري لتقدير القدرة للاختبار الخطي المحوسب والاختبار الموائم المحوسب باستخدام طريقة بييز لتقدير القدرة وقاعدة إنهاء الاختبار بعدد محدد من المفردات (ن= 25)

الاختبار	القياس	العدد	المتوسط	الانحراف المعياري	قيمة "ت"	مستوى الدلالة
تقدير القدرة	الخطي المحوسب	25	-0,14	0,59	0,95	0,35 غير دالة
	الموائم المحوسب	25	0,06	0,97		
الخطأ المعياري لتقدير القدرة	الخطي المحوسب	25	0,39	0,01	5,62	0,01** دالة
	الموائم المحوسب	25	0,28	0,09		
ت (0,05)= 2,06			ت (0,01)= 2,80			

يتضح من جدول (8) أن قيمة "ت" المحسوبة للفروق بين متوسطي تقدير القدرة للاختبار الخطي المحوسب والاختبار الموائم المحوسب تساوي (0,95)، وهي أقل من قيمة "ت"

الجدولية، فتكون "ت" المحسوبة غير دالة إحصائيًا؛ وهذا يعني قبول الفرض الصفري؛ أي أنه "لا توجد فروق ذات دلالة إحصائية بين متوسطي تقدير القدرة للاختبار الخطي المحوسب والاختبار الموائم المحوسب باستخدام طريقة بييز لتقدير القدرة وقاعدة إنهاء الاختبار بعدد محدد ممن المفردات".

ويتضح من جدول (8) أيضًا أن قيمة "ت" المحسوبة للفروق بين متوسطي الخطأ المعياري لتقدير القدرة للاختبار الخطي المحوسب والاختبار الموائم المحوسب تساوي (5,62)، وهي أكبر من قيمة "ت" الجدولية، فتكون "ت" المحسوبة دالة إحصائيًا؛ وهذا يعني رفض الفرض الصفري؛ أي أنه "توجد فروق ذات دلالة إحصائية بين متوسطي الخطأ المعياري لتقدير القدرة للاختبار الخطي المحوسب والاختبار الموائم المحوسب باستخدام طريقة بييز لتقدير القدرة وقاعدة إنهاء الاختبار بعدد محدد ممن المفردات".

ويدعم هذه النتيجة ما أشار إليه Jacobusse and Van Buuren (2007) من أن الاختبار الموائم المحوسب أكثر دقة في مراقبة نمو الأطفال من خلال استخدام طريقة بييز لتقدير القدرة وقاعدة إنهاء الاختبار بعدد محدد من المفردات.

ووضح نور الدين (2006، 100) أن طريقة بييز تعطي خطأ معياريًا أقل وكفاءة أكبر من طريقة الأرجحية القصوى، وأن قاعدة إنهاء الاختبار بـ (30) مفردة تعطي دقة في القياس.

كما أشار Kalender and Berberoglu (2017, 588) إلى عدم وجود اختلافات ملاحظة بين قواعد إنهاء الاختبار الموائم المحوسب بأسلوبي الطول الثابت والخطأ المعياري، وإمكانية استخدامه للقبول في التعليم الجامعي في تركيا.

وتختلف هذه النتيجة مع ما أشار إليه Choi et al. (2011,38) إلى أن محك الطول الثابت للاختبار يتم فيه قياس قدرة المختبرين بدرجات متفاوتة من الدقة، وبالتالي يصبح خطأ القياس أكبر عند المستويات المتطرفة من القدرة المقاسة.

وتدل هذه النتيجة على أن استخدام طريقة بييز لتقدير القدرة في الاختبار الموائم المحوسب وتحديد قاعدة إنهاء الاختبار بعدد محدد من المفردات يعطي نفس مستوى تقدير القدرة التي يعطيها الاختبار الخطي المحوسب، كما أنه أكثر دقة في تقدير القدرة لأنه يقدم أخطاء معيارية أقل من الاختبار الخطي المحوسب، حيث تفترض طريقة بييز توزيعًا قبليًا للقدرة وتعمل على تعديل هذا التوزيع استنادًا إلى استجابات المختبر، فيتم حساب التباين البعدي المتوقع لقدرة المختبر بعد إجابته عن المفردة، ويستمر اختيار المفردات حتى يصبح التباين البعدي صغيرًا، فيتم في هذه الطريقة تقدير القدرة والتباين المصاحب لها بعد كل مفردة تطبق على المختبر حتى يتم الوصول إلى قاعدة انهاء الاختبار.

واختار الباحث (25) مفردة كمحك لإنهاء الاختبار أي (35%) من عدد مفردات الاختبار الخطي المحوسب، وهذا يتفق مع الأطر النظرية ومعظم الدراسات والبحوث السابقة التي اعتمدت على عدد يتراوح بين (15-30) مفردة كمحك لإنهاء الاختبار الموائم المحوسب، مما يسهم في تقدير القدرة بشكل دقيق، لأنه يقلل الوقت والتعب والإرهاق والملل والجهد الذي يبذله المختبر أثناء إجابته عن المفردات الاختبارية.

نتيجة الفرض الرابع:

ينص الفرض الرابع على أنه "لا توجد فروق ذات دلالة إحصائية بين متوسطي تقدير القدرة والخطأ المعياري لتقدير القدرة (كل على حدة) للاختبار الخطي المحوسب والاختبار الموائم المحوسب باستخدام طريقة بييز لتقدير القدرة وقاعدة إنهاء الاختبار بأدنى خطأ معياري".

ويوضح جدول (9) نتائج اختبار "ت" لعينتين مرتبطتين للتحقق من صحة هذا الفرض

جدول (9)

قيمة "ت" ودلالتها الإحصائية للفروق بين متوسطي تقدير القدرة والخطأ المعياري لتقدير القدرة للاختبار الخطي المحوسب والاختبار الموائم المحوسب باستخدام طريقة بييز لتقدير القدرة وقاعدة إنهاء الاختبار بأدنى خطأ معياري (ن= 25)

الاختبار	القياس	العدد	المتوسط	الانحراف المعياري	قيمة "ت"	مستوى الدلالة
تقدير القدرة	الخطي المحوسب	25	0,03	1,12	0,10	0,91 غير دالة
	الموائم المحوسب	25	0,06	0,97		
الخطأ المعياري لتقدير القدرة	الخطي المحوسب	25	0,34	0,01	2,88	0,01** دالة
	الموائم المحوسب	25	0,28	0,10		
ت (0,05)= 2,06				ت (0,01)= 2,80		

يتضح من جدول (9) أن قيمة "ت" المحسوبة للفروق بين متوسطي تقدير القدرة للاختبار الخطي المحوسب والاختبار الموائم المحوسب تساوي (0,10)، وهي أقل من قيمة "ت" الجدولية، فتكون "ت" المحسوبة غير دالة إحصائيًا؛ وهذا يعني قبول الفرض الصفري؛ أي أنه "لا توجد فروق ذات دلالة إحصائية بين متوسطي تقدير القدرة للاختبار الخطي المحوسب والاختبار الموائم المحوسب باستخدام طريقة بييز لتقدير القدرة وقاعدة إنهاء الاختبار بأدنى خطأ معياري".

ويتضح من جدول (9) أيضًا أن قيمة "ت" المحسوبة للفروق بين متوسطي الخطأ المعياري لتقدير القدرة للاختبار الخطي المحوسب والاختبار الموائم المحوسب تساوي (2,88)، وهي أكبر من قيمة "ت" الجدولية، فتكون "ت" المحسوبة دالة إحصائيًا؛ وهذا يعني رفض الفرض الصفري؛ أي أنه "توجد فروق ذات دلالة إحصائية بين متوسطي الخطأ المعياري لتقدير القدرة للاختبار الخطي المحوسب والاختبار الموائم المحوسب باستخدام طريقة بييز لتقدير القدرة وقاعدة إنهاء الاختبار بأدنى خطأ معياري".

ويدعم هذه النتيجة ما توصل إليه عودة وعبيدات (2013، 1616) من وجود فروق ذات دلالة إحصائية بين متوسطي تقدير القدرة العقلية والخطأ المعياري لتقدير القدرة العقلية (كل على حدة) باستخدام مصفوفات رافن للاختبار الخطي المحوسب والاختبار الموائم المحوسب لصالح الاختبار الموائم المحوسب باستخدام طريقة بييز لتقدير القدرة العقلية وقاعدة إنهاء الاختبار بأدنى خطأ معياري (0,25).

وتدل هذه النتيجة على أن استخدام طريقة بييز لتقدير القدرة في الاختبار الموائم المحوسب وتحديد قاعدة إنهاء الاختبار بأدنى خطأ معياري يعطي نفس مستوى تقدير القدرة التي يعطيها

الاختبار الخطي المحوسب، كما أنه أكثر دقة في تقدير القدرة لأنه يقدم أخطاء معيارية أقل من الاختبار الخطي المحوسب، حيث تتطلب طريقة بييز افتراض احتمالات قبلية للبارامترات في ضوء اعتبارات نظرية أو إمبريقية، وهذا يعني أنه إذا توافرت معلومات عن توزيع القدرة لدى المختبرين، فإن طريقة بييز تساعد في الحصول على تقديرات جيدة للقدرة.

واعتمد الباحث على قاعدة إنهاء الاختبار عند الوصول إلى خطأ معياري (0,25) أو أقل، حيث يتوقف الاختبار عندما يصل إلى أدنى قيمة لمعلومات المفردة المصاحبة لتقدير القدرة، مما يعطي تقديرات دقيقة للقدرة عندما يشتمل بنك الأسئلة على مفردات تقدم معلومات تكفي لتحقيق هذه القاعدة.

توصيات البحث:

في ضوء ما أسفرت عنه نتائج البحث الحالي يقدم الباحث التوصيات الآتية:

1- ضرورة تطبيق أحدث نظم القياس والتقويم بما يتواكب مع التغيرات والمستجدات الأخيرة التي فرضت نفسها على المجتمع بالتعاون مع مراكز ووحدات القياس والتقويم بالجامعة.

2- إقامة ورش عمل لأعضاء هيئة التدريس بالجامعة لزيادة وعيهم وإدراكهم بطرق ووسائل القياس والتقويم الحديثة.

3- عقد دورات تدريبية لطلاب جامعة الأزهر لنشر الوعي بينهم بأحدث نظم الاختبارات ومميزاتها، وتنمية قدرتهم على التعامل مع نظم الاختبارات المحوسبة المختلفة.

4- تنظيم دورات إرشادية للعاملين بالجامعة لتوعيتهم بنظم القياس والتقويم الحديثة.

دراسات وبحوث مقترحة:

يمكن أن يثير البحث الحالي بعض المشكلات البحثية الآتية:

1- الفروق بين الاختبارات المحوسبة الموائمة متعددة المراحل والخطية في تقدير القدرة وفقًا لنظرية الاستجابة للمفردة لطلاب الجامعة.

2- أثر تفاعل أسلوبين للاختبارات المحوسبة الخطية والموائمة مع قلق الاختبار في تقدير القدرة وفقًا لنظرية الاستجابة لدى طلاب الجامعة.

3- الفروق بين الاختبارات المحوسبة الموائمة والخطية في تقدير القدرة باستخدام اختبارات أخرى مثل اختبارات الذكاء والتفكير.

4- أثر اختلاف استراتيجية انتقاء المفردات للاختبارات الموائمة المحوسبة في تقدير القدرة بطريقتي الأرجيحة القصوى وبييز.

5- فاعلية الاختبارات الموائمة المحوسبة في تقدير القدرة باستخدام برنامج R (دراسة محاكاة).

6- أثر تفاعل أسلوبين للاختبارات المحوسبة الخطية والموائمة مع الحكمة الاختبارية في تقدير القدرة وفقًا لنظرية الاستجابة لدى طلاب الجامعة.

قائمة المراجع

أولًا: المراجع العربية:

الحسيني، هشام. (2019، نوفمبر6-7). *آليات تفعيل دراسة تقويم الامتحانات العامة في الوطن العربي وتطويرها*. المؤتمر الحادي عشر لوزراء التربية والتعليم العرب بالبحرين "السياسات التعليمية ودورها في تحقيق الهدف الرابع من أهداف التنمية المستدامة 2030"، جامعة الدول العربية، المنظمة العربية للتربية والثقافة والعلوم، المنامة، مملكة البحرين.

الحسيني، هشام، والعدروس، مرفت. (2014، فبراير 6). *استخدام الاختبارات التواؤمية المحوسبة في قياس وتشخيص القدرات العقلية للمتعلمين*. ورقة عمل مقدمة إلى مؤتمر التعليم الإلكتروني بمؤسسات التعليم قبل الجامعي، المركز القومي للامتحانات والتقويم التربوي، القاهرة، مصر.

السعدني، محمد (2019). أنماط الاختبار الالكتروني (التكيفي، الوسطي، الخطي) وأثر تفاعلها مع مستوى القلق من الاختبار (غير الطبيعي- المرضي) على تنمية التحصيل وخفض القلق لدى طلاب تكنولوجيا التعليم. *مجلة تكنولوجيا التعليم، الجمعية المصرية لتكنولوجيا التعليم، 29*(8)، 11-94.

الصبيحي، علي، والأحمدي، شرف. (2011). الاختبار التكيفي المحوسب كتطبيق لتطوير أساليب التقويم والاختبارات الالكترونية. *مجلة الدراسات التربوية والإنسانية، كلية التربية، جامعة دمنهور، 3* (1)، 83-126.

الكرامنة، محمد. (2015). *بناء اختبار تكيفي محوسب للذكاء المنطقي الرياضي للطلبة الموهوبين باستخدام نظرية الاستجابة للفقرة* [رسالة دكتوراه، الجامعة الأردنية بالأردن]. قاعدة معلومات دار المنظومة، الرسائل الجامعية.

الكيال، مختار، ونور الدين، أمين، والظنحاني، ميسون. (2017). الخصائص السيكومترية لاختبار تواؤمي محوسب لتشخيص صعوبات تعلم الرياضيات. *مجلة الإرشاد النفسي، جامعة عين شمس، 50* (2)، 213-254.

دعنا، زينات. (2012). بناء اختبار محبوك محوسب في مقرر سيكولوجية الفروق الفردية وفق الاستراتيجية الهرمية (متعددة المستويات) في نظرية الاستجابة للفقرة في جامعة طيبة. *مجلة عالم التربية، المؤسسة العربية للاستشارات العلمية وتنمية الموارد البشرية، 37* (13)، 153-186.

سليمان، أمين (2009). *القياس والتقويم في العلوم الإنسانية أسسه وأدواته وتطبيقاته*. دار الكتاب الحديث.

علام، صلاح الدين. (2005). *نماذج الاستجابة للمفردة الاختبارية أحادية البعد ومتعددة الأبعاد وتطبيقاتها في القياس النفسي والتربوي*. دار الفكر العربي.

علام، صلاح الدين. (2006). *القياس والتقويم التربوي والنفسي أساسياته وتطبيقاته وتوجهاته المعاصرة*. دار الفكر العربي.

عمارة، إيهاب، وكاظم، علي. (2020). دراسة جدوى تطوير صورة تكيفية محوسبة من اختبار المصفوفات المتتابعة الملونة للأطفال العمانيين باستخدام نظرية الاستجابة للمفردة. *المجلة الدولية للأبحاث التربوية، جامعة الإمارات العربية المتحدة، 44* (3)، 142-181.

عودة، أحمد، وعبيدات، عمر. (2013). فاعلية الاختبار التكيفي المحوسب في تقدير القدرة العقلية باستخدام مصفوفات رافن. *مجلة دراسات العلوم التربوية، الجامعة الأردنية، 40* (2)، 1602-1621.

محمد، محمد. (2017). الاختبارات المحبوكة المحوسبة: تطوير لبنوك الأسئلة. *مجلة إبداعات تربوية، رابطة التربويين العرب*، (3)، 11-20.

مراد، صلاح، وسليمان، أمين. (2005). *الاختبارات والمقاييس في العلوم النفسية والتربوية خطوات إعدادها وخصائصها* (ط2). دار الكتاب الحديث.

مسلم، فؤاد. (2011). *أسلوب مقترح لاختيار الفقرات وتقدير القدرة في الاختبارات التكيفية المحوسبة* [رسالة دكتوراه، جامعة اليرموك بالأردن]. قاعدة معلومات دار المنظومة، الرسائل الجامعية.

نور الدين، أمين. (2006). أثر سلوك مراجعة وتعديل المفردات على تقدير القدرة وكفاية القياس في الاختبارات التواؤمية المحوسبة. *الجمعية المصرية للدراسات النفسية، 53*(16)، 75-132.

نور الدين، أمين. (2019). فاعلية القياس التكيفي المحوسب في قياس تحصيل طلاب الجامعة. *المجلة السعودية للعلوم النفسية، جامعة الملك سعود، الجمعية السعودية للعلوم التربوية والنفسية*، (64)، 29-47.

ثانيًا: المراجع العربية مترجمة باللغة الإنجليزية:

Al−Hossini, Hisham. (2019, November 6-7). *Mechanisms of activating study of evaluating public exams in the Arab world and developing it.* The Eleventh Conference of Arab Ministers of Education in Bahrain "Educational policies and their role in achieving the fourth goal of sustainable development goals 2030". League of Arab States, Arab Organization for Education, Culture and Science, Manama, Kingdom of Bahrain.

Al-Hossini, Hisham, & Al- Adroos, Mervet. (2014, February 6). *Using computerized adaptive testing to measure and diagnose learners' mental abilities.* Working paper presented to e-learning conference in pre-university education institutions, National Center for Examinations and Educational Evaluation, Cairo, Egypt.

Al-Saadany, Mohammed.(2019). Electronic test patterns (adaptive, middle, Linear) and their effect on interaction with test anxiety level (abnormal - pathological) on development of achievement and reduce anxiety among educational technology students. *Journal of Educational Technology, Egyptian Association for Educational Technology, 29* (8), 11-94.

Al-Subaihi, Ali, & Al-Ahmadi, Sharaf. (2011). Computerized adaptive testing as an application of developing evaluation methods and electronic tests. *Journal of Human and Educational Studies, Faculty of Education, Damanhour University, 3*(1), 83-126.

Al-Kramenh, Mohammed. (2015). *Constructing A computerized adaptive testing of logical mathematial intelligence for gifted students using item response theory* [Doctoral dissertation, Jordanian University, Jordan]. Dar Al-Mandumah Database, Dissertations.

El-Kayal, Mokhtar, Nour Al-Deen, Amin, & Al-Dhanhani, Maysoon. (2017). Psychometric properties of a computerized adaptive testing for diagnosing mathematics learning disabilities. *Journal of Psychological Counseling, Ain Shams University, 50* (2), 213-254.

Danaa, Zeenat. (2012). Building a computerized adaptive testing in a course of individual differences psychology according to the hierarchical strategy (multi-level) in item response theory at Taibah University. *Journal of Education World, The Arab Foundation for Scientific Consultation and Human Resource Development, 37* (13), 153-186.

Soliman, Amin. (2009). *Measurment and evaluation in human since (basics, tools, application)*. Dar Al-Kitab Al-Hadith.

Allam, Salah Al-Deen. (2005). *Unidimensional and multidimensional item response models and their applications in psychological and educational measurement.* Dar Al-Fikr Al-Arabi.

Allam, Salah Al-Deen. (2006). *Educational and psychological measurement and evaluation: its basics, applications and contemporary trends.* Dar Al-Fikr Al-Arabi.

Omara, Ehab, & Kazem, Ali.(2020). A Feasibility study for developing a computerized adaptive Form of colored sequential matrices test for Omani's children using the item response theory. *International Journal for Educational Researches, United Arab Emirates University, 44*(3), 142-181.

Odeh, Ahmed, & Obaidat, Omar.(2013). The Effectiveness of computerized adaptive testing in estimating mental ability using Raven's matrices. *Journal of Educational Sciences Studies, Jordanian University, 40* (2), 1602-1621.

Mohammed, Mohammed.(2017). Computrized adaptive testing: Develping item banking. *Journal of Educational Creations, Association of Arab's Educators*, (3), 11-20.

Morad, Salah, & Soliman, Amin.(2005). *Tests and measures in psychological and educational sciences steps of preparing it and their characteristics* (2 nd ed). Dar Al-Kitab Al-Hadith.

Mussallum, Fuad.(2011). *A Proposed style for selecting items and estimating ability in computer adaptive testing* [Doctoral dissertation, University of Yarmouk, Jordan]. Dar Al-Mandumah Database, Dissertations.

Nour Al-Deen, Amin.(2006). The Effect of behavior for revision and answer-changing item on estimating ability and sufficiency of measurment in computerized adaptive testing. *Egyptian Association for Psychological Studies, 53*(16), 75-132.

Nour Al-Deen, Amin.(2019). The Effectiveness of computerized adaptive measurment in measuring the achievement of university students. *Saudi Journal of Psychological Sciences, King Saud University, Saudi Association for Educational and Psychological Sciences,* (64), 29-47.

ثالثًا: المراجع الأجنبية:

Aybek, E., & Gulleroglu, H.(2021). Attitudes toward pirated content: A scale development study based on graded response model. *Eurasian Journal of Educational Research,91*, 127-144.

Babcock, B., & Weiss, D. (2012). Termination criteria in computerized adaptive tests: do variable-length CATs provide efficient and effective measurement?*Journal of Computerized Adaptive Testing, 1*(1), 1-18.

Bjorner, J., Chang, C., Thissen, D., & Reeve, B.(2007). Developing tailored instruments: item banking and computerized adaptive assessment. *Quality of Life Research, 16*, 95-108.

Boyd, A.(2003). *Strategies for controlling testlet exposure rates in computerized adaptive testing systems* [Doctoral dissertation, University of Texas at Austin, United States of America]. ProQuest Dissertations and Theses Global.

Cella, D., Gershon, R., Lai, J., & Choi, S.(2007). The future of outcomes measurment: item banking, tailored short-forms, and computerizd adaptive assessment. *Quality of Life Research, 16,* 133-141.

Chang, C .(2005). Developing tailored instruments: item banking and computerized adaptive assessment. *Evaluation and Program Planning Journal ,24*, 215-251.

Cheng, I., & Basu, A.(2009). Interactive graphis for computer adaptive testing. *Computer Graphics Forum, 28*(8), 2033–2045.

Choi, S., Grady, M., & Dodd, B.(2011). A new stopping rule for computerized adaptive testing. *Educational and Psychological Measurment, 71*(1), 37-53.

Cisar, S., Radosav, D., Markoski, B., Pinter, R., & Cisar, P.(2010). Computer adaptive testing of student knowledge. *Acta Polytechnical Hungarica, 7*(4), 139-152.

Davey, T.(2011). *A Guide to computer adaptive testing systems*. the Council of Chief State School Officers.

Davey, T., & Pitoniak, M.(2006). Designing computerized adaptive tests. In S. Dowing & T. Haladyna (Eds.), *Handbook of test development* (543-574). Lawrence Erlbaum Associates.

Domino, G., & Domino, M.(2006). *Psychological testing: An introduction* (2 nd ed) . Cambridge University Press.

Embretson, S., & Reise, S. (2000*). Item response theory for psychologists*. Lawrence Erlbaum Associates.

Erdem-Kara,B. (2019). Computer adaptive testing simulations in R. *International Journal of Assessment Tools in Education, 6*(5), 44-56.

Finkelman, M.(2004). *Statistical isssues in computerized adaptive testing* [Doctoral dissertation, University of Stanford, California, United States of America].ProQuest Dissertations and Theses Global.

Finkelman, M., Weiss, D., & Kim-Kang, G.(2010). Item selection and hypothesis testing for the adaptive measurment of change. *Applied Psychological Measurement, 34*(4), 238–254.

Hogenboom, S., Hermans, F., & Van der Mass, H.(2021). Computerized adaptive assessment of understanding of programming concepts in primary school children. *Computer Scince Education*, https://doi.org/10.1080/08993408.2021.191446.

Hol, A. Vorst, H., & Mellenberg, G.(2008). Computerized adaptive testing of personality traits. *Journal of Psychology, 216*(1), 12-21.

Istiyono, E., Brams, W., Setiawan, R., & Megawati, I.(2020). Developing of computerized adaptive testing to measure physics higher order thinking skills of senior high school students and its feasibility of use. *European Journal of Educational Research, 9*(1),91-101.

Jacobusse, G., & Van Buuren, S.(2007). Computerized adaptive testing for measuring development of young children. *Statistics in Medicine, 26*, 2629-2638.

Jian-quan, T., Dan-min, M., Xia, Z., & Jing-jing, G. (2007). An Introduction to the computerized adaptive testing. *Us- China Education Review, 4*(1), 72-81.

Kalender, I., & Berberoglu, G.(2017). Can computerized adaptive testing work in students' admission to higher education programs in Turkey*? Educational Sciences: Theory & Practice, 17*(2),573-596.

Leroux, A., & Dodd, B. (2014). A comparison of stopping rules for computerized adaptive screening measures using the rating scale model. *Journal of Applied Measurement, 15*, 213-226.

Leroux, A. & Dodd, B. (2016). A Comparison of exposure control procedures in CATs using the GPC model. *Journal of Experimental Education, 84*(4), 666-685.

Linacre, J.(2000). *Computer adaptive testing: a methodology whose time has com.* Komesa Press.

Magis, D., Yan, D., & Von Davier, A.(2017). *Computerized adaptive testing and multistage testing with R using packages catR and mstR*. Springer International Publishing.

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

Mizumoto, A., Sasao, Y., & Webb, S.(2019). Developing and evaluating a computerized adaptive testing version of the word part levels test. *Language Testing, 36*(1), 101-123.

Ozyurt, H., & Ozyurt, O. (2015). Ability level estimation of students on probability Unit Via computerized adaptive testing. *Eurasian Journal of Educational Research*, (58), 27-44.

Ozyurt, H., Ozyurt, O., Baki, A., & Guven B.(2012). An Application of individualized assessment in educational hypermedia: design of computerized adaptive testing system and its integration into UZWEBMAT. *Procedia- Social and Behavioral Science, 46*, 3191-3196.

Ramadan, A., & Aleksandrovna, E.(2018, November 27-29). *Computerized adaptive testing*. Conference paper: Information Technology in Education, Saratov National Research State University, Russia.

Rezaie, M., & Golshan, M.(2015). Computer adaptive test (CAT): Advantages and limitations. *International Journal of Educational Investigations, 2*(5), 128-137.

Runder, L.(1998). *An on- line, interactive, computer adaptive testing tutorial.* ERIC Clearinghouse on Assessment and Evaluation.

Runder, L.(2010). Implementing the graduate management admission test computerized adaptive test. In W. Van der Linden & C., Glas (Eds.), *Elements of adaptive testing* (151-166). Springer.

Sahin, A., & Weiss, D.(2015). Effects of calibration sample size and item bank size on ability estimation in computerized adaptive testing. *Educational Sciences: Theory & Practice, 15*(6), 1585-1595.

Samsudin, M., SomChut, T., & Ismail, M.(2019). Evaluating computerized adaptive testing efficiency in measurement student's performance in science timss. *Indonesian Journal of Science Education, 8*(4), 547-560.

Stafford, R., Runyon, C., Casabianca, J., & Dodd, B. (2019). Comparing computer adaptive testing stopping rules under the generalized partial-credit model. *Behavior Research Methods, 51*(3), 1305–1320.

Su, Y.(2018). A Comparison of constrained item selection method in multidimensional computerized adaptive testing. *Applied Psychological Measurment, 40*(5), 346-360.

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

Tian X., & Dai , B.(2020). Developing a computerized adaptive test to assess stress in Chinese college students. *Frontiers in Psychology, 11*(7), 1-9.

Triantafillou, E., Georgiadou, E., & Economides, A.(2008). The Design and evaluation of a Computerized adaptive test on mobile devices. *Computers & Education, 50*(4),1319-1330.

Van der Linden, W., & Pashley, P.(2010). Item selection and ability estimation in adaptive testing. In W. Van der Linden & C., Glas (Eds*.), Elements of adaptive testing* (3-30). Springer.

Veldkamp, B.(2016). On the issue of itme selection in computerized adaptive testing with response times. *Journal of Educational Measurement, 53*(2), 212-228.

Wang, C., Weiss, D., & Shang, Z.(2019). Variable-length stopping rules for multidimensional computerized adaptive testing. *Psychometrika, 84*(3), 749-771.

Weiss, D.(2004). Computerized adaptive testing for effective and efficient measurement in counseling and education. *Measurement and Evaluation in Counseling and Development, 37*(2), 70-84.

Weiss, D.(2011). Better data from better measurements using computerized adaptive testing. *Journal of Methods and Measurement in the Social Sciences, 2*(1), 1-27.

Yasuda, J., Mae, N., Hull, M., & Taniguchi, M.(2021). Optimizing the length of computerized adaptive testing for the force concept inventory. *The American Physical Society, Physical Review Physics Education Research, 17*, 1-15.