أثر قواعد توقف الطول الثابت في إدارة الاختبارات التكيفية

أحمد على الشهري

طالب دكتوراه جامعة الملك سعود

ISSN 2735-5209

المستخلص:

هدفت الدراسة إلى تعرف أثر قواعد توقف الطول الثابت في ادارة الاختبارات التكيفية. بما يشمل الكشف عن تأثير اختلاف قواعد التوقف في الاختبار التكيفي المحوسب على دقة تقدير القدرة. وتحديد قاعدة التوقف في الاختبار التكيفي المحوسب الأكثر تأثيرًا على دقة تقدير القدرة للأفراد. وتضمنت عينة الدِّراسة من (١٢٠٠) طالب والتي تشكل نسبة مقدارها (٣١٪) تقريبًا من مجتمع الدِّراسة، وجرى اختيارها بالطريقة العشوائية العنقودية. وتضمنت أداة الدراسة اختبار تحصيلي في مقرر الفيزياء للصف الثالث الثانوي حيث تم إجراء تحليل محتوى كتاب مقرر فيزياء للصف الثالث ثانوي–(نظام المقررات الدراسية) من خلال تحديد الاهداف التعليمية للمقرر ومستويات تلك الاهداف طبقا لتصنيف بلوم. وأشارت النتائج إلى أن النموذج الثابت ٤٠ هو الأفضل بين النماذج الأربعة، حيث يوضح أداءً أفضل في مقاييس RMSE وMAE. ومع ذلك، فإن الاختيار بين النماذج الأربعة يعتمد على السياق والأهداف المحددة للنموذج. يمكن استخدام النموذج الثابت ٣٠ إذا كان الهدف هو تقليل التعقيد في النموذج، بينما يمكن استخدام النموذج الثابت ٤٠ إذا كان الهدف هو تحقيق أفضل أداء في التنبؤ بالقيم الحقيقية. يمكن استخدام النموذج الثابت ٥٠ أو النموذج الثابت ١٠٠ إذا كان الهدف هو تحقيق أداءً أفضل في مقاييس محددة مثل RMSE أو MAE.

الكلمات الدالة: الطول الثابت، إدارة الاختبارات التكيفية، الاختبار التكيفي المحوسب.

The effect of length Termination rules on the administration of Adaptive Testing

Abstract

The study aimed to know the effect of fixed length stop rules in administering adaptive tests. Including revealing the effect of different stopping rules in the computerized adaptive test on the accuracy of ability estimation. Determining the stopping rule in the computerized adaptive test has the most impact on the accuracy of the ability estimate for individuals. The study sample included (1200) students, which constitute approximately (31%) of the study population, and they were selected by the cluster random method. The study tool included an achievement test in the physics course for the third year of high school, where an analysis of the content of a physics course book for the third year of secondary school - (course system) was conducted by determining the educational goals of the course and the levels of those goals according to Bloom's classification. The results indicated that the fixed model 40 is the best among the four models, as it shows better performance in the RMSE and MAE scales. However, the choice between the four models depends on the specific context and objectives of the model. The static model 30 can be used if the goal is to reduce the complexity of the model, while the static model 40 can be used if the goal is to achieve the best performance in predicting true values. The Fixed Model 50 or the Fixed Model 100 may be used if the goal is to achieve better performance on specific measures such as RMSE or MAE.

Keywords: fixed length, adaptive test administration, Computerized adaptive testing

ISSN 2735-5209

مقدمة الدراسة:

في الآونة الأخيرة حدث تطور كبير في علم القياس والتقويم، فقد تطور هذا العلم تطورًا كبيرًا منذ ظهور النظرية الحديثة في القياس والمعروفة باسم نظرية الاستجابة للمفردة Item (Response Theory: IRT)، ومع هذا التطور تزايد الاهتمام في بناء وتطوير المقاييس والاختبارات العقلية والأدائية لقياس وتقييم قدرات الأفراد، وتحقيق معايير القياس العلمي الذي يضمن تقديرًا كميًا صادقًا وبدرجة مقبولة من الدقة والموضوعية للدرجات التي تم اعتمادها كتقدير كمي لهذه القدرات، وقد واكب هذا الاهتمام تطورًا في البرمجيات ذات الصلة بتطبيقات النظرية الحديثة في القياس، والذي أدى إلى ظهور الاختبارات المحوسبة التي توفر إمكانية التخزين والتطبيق للاختبارات .(Murphy& Davidshofer ,1994)

ويشير كايسر وآخرون (Cisar et al., 2010) إلى أن الاختبارات المحوسبة تختلف في طريقة تطبيقها، فقد تقدم جميع الفقرات لجميع المختبرين بنفس الترتيب، ويسمى ذلك الاختبار "الاختبار الخطي المحوسب"، أو تقدم للمفحوص الواحد الفقرات التي تتناسب مع مستواه فقط، ويطلق عليه "الاختبار التكيفي المحوسب" (Computer Adaptive Testing) CAT:وهو ما ستهتم به الدراسة الحالية، حيث يُعرض على المختبر بعض الفقرات لتحديد قدرته المبدئية، وبناءً على قدرته المبدئية من خلال أدائه على تلك الفقرات تُقدم له فقرات أخرى لاحقة من بنك الأسئلة بما يتناسب مع التقدير المستمر لمستواه، ويعتمد اختيار الفقرة اللاحقة على استجابة المختبر على الفقرات السابقة، وعلى خصائص تلك الفقرات.

ويذكر جاكوبس وبورين (Jacobusse & Buuren,2007) أن الخطوة الأهم في الاختبارات التكيفية هي تحديد قاعدة الإنهاء للاختبار، فقد ينتهي الاختبار التكيفي المحوسب عند تطبيق عدد محدد من الفقرات قاعدة الطول الثابت للاختبار (Fixed Length) بغض النظر عن درجة دقة القياس التي تم تحقيقها عند إنهاء الاختبار, ويمكن انهاؤه أيضا عند الوصول إلى درجة مقبولة من الدقة لكل مختبر قاعدة الطول المتغير(Variable Length)، ووفقًا لهذه القاعدة يأخذ كل مختبر مجموعة فقرات مختلفة عن المختبرين الآخرين وليس بنفس العدد، وينتهي تقديم الاختبار التكيفي عند الوصول لأدنى خطأ معياري (Minimum Standard Error) محدد مسبقًا, أو عند عدم وجود المزيد من الأسئلة المتاحة القادرة على توفير حد أدنى محدد مسبقًا من المعلومات التي تساعد على تقدير القدرة .

مشكلة الدراسة:

وتتمثل مشكلة الدراسة في التعرف على أفضل قواعد التوقف ذات الطول الثابت((١٠٠، ٥٠، ٤٠، ٣٠)؟ من حيث تأثيرها على إدارة الاختبار التكيفي، وذلك من خلال إعداد اختبار

ISSN 2735-5209

تحصيلي في مادة الفيزياء للصف الثالث ثانوي من نوع الاختيار من متعدد بواقع أربعة بدائل لكل مفردة، ويتمثل السؤال الرئيسي لهذه الدراسة فيما يلي:

- ما دقة استخدام قاعدة التوقف ذات الطول الثابت في ادارة الاختبار التكيفي (١٠٠، ٥٠، ٤٠، ٣٠)؟

أهداف الدراسة:

١- الكشف عن تأثير اختلاف قواعد التوقف في الاختبار التكيفي المحوسب على دقة تقدير القدرة.

٢- تحديد قاعدة التوقف في الاختبار التكيفي المحوسب الأكثر تأثيرًا على دقة تقدير القدرة للأفراد.

أهمية الدراسة:

أولا: من الناحية النظرية:

١- تكمن أهمية هذه الدراسة في تطبيق وسائل جديدة في القياس النفسي والتربوي وهي نظرية الاستجابة للمفردة والاختبارات التكيفية المحوسبة والتي قد تكون أكثر دقة وموضوعية في تقدير قدرة الطالب، وبالتالي تعد كأداة لتشخيص نقاط القوة والضعف عند كل فرد وبناء الخطط العلاجية المناسبة وتحديد نقاط القوة لتعزيزها.

٢- تُعد الدراسة الحالية استكمالًا لما توصلت إليه الدراسات السابقة، واستجابة لتوصياتها لتحديد أهم قواعد التوقف في الاختبار التكيفي الأكثر تأثيرًا على دقة تقدير القدرة ودالة معلومات الفقرة والاختبار، كما تتناول جوانب جديدة لم تتطرق لها الدراسات السابقة.

ثانيا: الأهمية التطبيقية:

١- تسلط الدراسة الحالية الضوء على مشكلة حقيقية في المجال التعليمي، وهي قصور بعض القائمين على الاختبارات في تحديد قواعد التوقف ذات الأثر الأكبر في دقة تقدير الفرد ودالة المعلومات للفقرة، وهذه الخاصية ذات أهمية كبيرة من حيث تقليص الوقت والجهد في تطبيق الاختبارات المختلفة وخاصة الاختبارات المكونة من عدد كبير من الفقرات.

٢- مساعدة مصممي الاختبارات التربوية في تصميم اختبارات تكيفية في ضوء خصائص الطلاب والاختبار المرغوب، الذي يتمتع بدرجة عالية من الدقة.

حدود الدراسة:

١. الحدود الموضوعية: تتحدد الدراسة الحالية بموضوعها: دراسة أثر قواعد التوقف ذات الطول الثابت على ادارة الاختبار التكيفي.

ISSN 2735-5209

٢. الحدود المكانية: تم إجراء الدراسة في المدارس الحكومية/ بنين، بمدينة الدمام التابعة للإدارة العامة للتعليم بالمنطقة الشرقية.
٣. الحدود الزمانية: تم تطبيق الدراسة في الفصل الدراسي الأول من العام الجامعي ١٤٤٤هـ/٢٠٢٣م.
٤. الحدود البشرية: عينة تم اختيارها عشوائيًا من طلاب الصف الثالث الثانوي.

الإطار النظري:

أولا: الاختبارات التكيفية:

يعرفها (الصبيحي، والأحمدي، ٢٠١١) بأنها الاختبارات التي يقوم فيها مؤدي الاختبار بالإجابة عن مجموعات مختلفة من الأسئلة تطرح عليهم بناء على مستوياتهم حيث تؤثر إجابتهم عن سؤال ما على نوعية الأسئلة التالية التي يختارها الحاسب لهم ففي حال أجاب الطالب عن السؤال إجابة صحيحة سيختار الحاسب سؤالًا أصعب قليلًا من السؤال السابق وأما إذا أجاب إجابة خاطئة فسيختار له الحاسب سؤالًا أقل صعوبة وهكذا ولهذا سمي ب "الاختبار التكيفي"

ويتكون الاختبار التكيفي من مكونين, وهما اختبار محوسب بالإضافة إلى مكونات متكيفة, والمقصود بالتكيف: هي القدرة على تفصيل مستوى صعوبة كل مفردة من مفردات الاختبار على أساس صحة إجابة الطالب عن المفردة السابقة (الصبيحي, والأحمدي, ٢٠١١).

التعريف الإجرائي للاختبار التكيفي: هو الاختبار التحصيلي في مادة الفيزياء للصف الثالث الثانوي الفصل الدراسي الأول الذي يفصل لكل طالب على حدة، وذلك بعرض الفقرات التي تتناسب مع مستوى قدرته، مما يمكن من تقدير أدق للقدرة بأقل عدد من الفقرات.

ثانيا: قواعد التوقف:

هي قواعد لإنهاء الاختبار التكيفي ويتم فيها تطبيق عدد محدد من الفقرات بدرجة مقبولة من الدقة لكل مفحوص، ووفقًا لهذه القواعد يأخذ كل مختبر مجموعة فقرات مختلفة من المختبرين الأخرين حتى يتحقق معيار الإنهاء المحدد سلفًا (عبيدات، وعودة ٢٠١٣).

قاعدة الطول الثابت للاختبار: Fixed Length

حيث يقدم لكل مفحوص نفس العدد من الأسئلة الاختبارية التي يحددها القائم على الاختبار، وهي تفتقد للاتساق في دقة تقديرات القدرة للمختبرين، حيث يتوقع اختلاف دقة تقدير المختبر بحسب موقعة على متصل القدرة، وتتطلب محاولات رفع الدقة باستخدام هذه الطريقة زيادة عدد الأسئلة وهو ما يعني عدم جدوى تكييف الاختبار وكذلك زيادة نسبة التعرض لكل سؤال في بنك الأسئلة وهو أيضًا الأمر الذي ينبغي الانتباه له عند تكييف الاختبارات, وتعد قاعدة إيقاف الطول الثابت هي الأكثر وضوحًا لأنها تنهي اختبار المختبر بعد اجتيازه لعدد محدد مسبقًا من الفقرات بغض النظر عن قدرة المختبر.(Leroux & Dodd,2014)

ISSN 2735-5209

التعريف الإجرائي لقواعد التوقف: هي القاعدة التي يتم من خلالها تحديد متى يتم إنهاء اختبار مقرر الفيزياء للصف الثالث الثانوي وتوقف عرض المفردات على أفراد عينة البحث.

الدراسات السابقة:

هدفت دراسة بابكوك وويس (Babcock & Weiss,2009) إلى فحص عددٍ من قواعد إنهاء الاختبار المحوسب (CAT) من خلال دراسة محاكاة باستخدام إطار نظرية الاستجابة للمفردة, أظهرت النتائج أن الاختبارات الأكثر طولًا كانت الأكثر دقة في تقدير سمات المختبرين, كما كانت نتيجة قاعدة إنهاء الخطأ المعياري ذات دقة عالية وأفضل من قاعدة الإنهاء ذات الطول الثابت من حيث تقدير القدرة في الاختبارات ذات المفردات قليلة العدد, خاصة اذا كانت مستوى الخطأ المعياري المستخدم منخفض بدرجة كافية, لكنها كانت حساسة لهيكل بنك الأسئلة.

وهدفت دراسة تشو وآخرين (Choi, Grady &Dodd ,2010) الى تقديم قاعدة إيقاف جديدة للاختبار التكيفي المحوسب. تستخدم قاعدة إيقاف تقليل الخطأ المعياري المتوقعة (PSER) التباين الخلفي التنبئي لتحديد الانخفاض في الخطأ القياسي الذي قد ينتج عن المزيد من المفردات. تمت مقارنة أداء قاعدة إيقاف الحد من الأخطاء المعيارية المتوقعة مع قاعدة إيقاف الحد الأدنى من الخطأ المعياري ونسخة معدلة من قاعدة إيقاف الحد الأدنى من المعلومات في سلسلة من الاختبارات التكيفية المحاكاة، المستمدة من عدد من مجموعات المفردات. تشير النتائج إلى أن قاعدة إيقاف الحد من الخطأ المعياري المتوقع تستخدم بكفاءة لمجموعات مفردات الاختبار التكيفي المحوسب، حيث يتم تطبيق عدد أقل من المفردات عندما تكون المكاسب التنبؤية في المعلومات صغيرة وتزيد من دقة القياس عندما تكون المعلومات وفيرة.

بينما أشارت دراسة ستافورد واخرون (Stafford et al.,2019) الى أحد الاعتبارات المهمة لأي برنامج اختبار تكيفي للحاسوب (CAT) هو المعيار المستخدم لإنهاء تطبيق المفردة – قاعدة التوقف، والتي تضمن أن يتم تقييم جميع الممتحنين وفقًا لنفس المعيار. على الرغم من وجود قواعد مختلفة للإيقاف، لم تتم مقارنة أي منها بموجب نموذج التقدير الجزئي المعمم. في دراسة المحاكاة، قام هؤلاء الباحثون بمقارنة أداء ثلاث قواعد إيقاف متغيرة الطول – الخطأ المعياري (SE)، والحد الأدنى من المعلومات (MI)، والتغيير في ثيتا (CT) – بشكل مستقل وبالاقتران مع متطلبات الحد الأدنى والأقصى لعدد من المفردات بالإضافة إلى قاعدة إيقاف ذات طول ثابت. تم فحص كل قاعدة توقف وفقًا لمعيارين إنهاء – أحدهما مطلب أكثر تساهلاً (SE = 0.35، MI = 0.56، CT = 0.05)، وواحد أكثر صرامة (SE = 0.30، MI = 0.42، CT = 0.02). تضمن تصميم المحاكاة أيضًا موازنة المحتوى وضوابط التعرض، وهي جوانب من الاحتبار التكيفى المحوسب CAT التي تم استبعادها في البحوث السابقة التي تقارن قواعد التوقف متغيرة الطول. أنتجت قاعدة إيقاف الحد الأدنى من المعلومات تقديرات ثيتا المتحيزة وتنوعت بشكل كبير في جودة

ISSN 2735-5209

القياس عبر توزيع ثيتا. كان لقاعدة إيقاف التغيير المطلق في ثيتا أداءً قويًا عند ربطها بمعيار أقل وطول اختبار أدنى. توفر قاعدة إيقاف الخطأ القياسية باستمرار أفضل توازن لدقة القياس والفعالية الاجرائية واستندت إلى أقل عدد من المفردات المطبقة اللازمة للحصول على تقديرات ثيتا دقيقة وموجزة، لا سيما عندما تم إقرانها مع قاعدة التوقف ذات الحد الأقصى لعدد المفردات.

منهج الدراسة وإجراءاتها:

منهج الدراسة: تعتمد الدراسة الحالية على استخدام المنهج الوصفي المقارن.

عينة الدِّراسة: تألَّفت عينة الدِّراسة من (١٢٠٠) طالب والتي تشكل نسبة مقدارها (٣١٪) تقريبًا من مجتمع الدِّراسة، وجرى اختيارها بالطريقة العشوائية العنقودية.

أداة الدراسة: قام الباحث ببناء اختبار تحصيلي في مقرر الفيزياء للصف الثالث الثانوي حيث تم إجراء تحليل محتوى كتاب مقرر فيزياء للصف الثالث ثانوي–(نظام المقررات الدراسية) من خلال تحديد الاهداف التعليمية للمقرر ومستويات تلك الاهداف طبقا لتصنيف بلوم للفصول التالية:
اساسيات الضوء– الانعكاس والمرايا– الانكسار والعدسات– تطبيقات العدسات– التداخل والحيود الكهرباء الساكنة– المجالات الكهربائية.

تدريج بنك الأسئلة: فيما يلي تحليل مفردات الاختبار بالنظرية التقليدية والثبات: حيث تم تحليل مفردات المستودع على مستوى نظرية القياس التقليدية ونظرية الاستجابة للمفردة.

تحليل المفردات بالنظرية الكلاسيكية:

تم تحليل المفردات باستخدام نظرية القياس التقليدية حيث تم استخراج معاملات الصعوبة والتمييز لكل مفردة من مفردات المستودع. فاتضح أن متوسط الصعوبة لمفردات المستودع بلغت (٠.٥١) وأن متوسط التمييز بلغ (٠.٤٣) وهى قيم مقبولة طبقا لنظرية القياس التقليدية حيث جاءت جميع قيم معاملات التمييز بالموجب وأعلى من(٠.٣)

ثبات الاختبار

الثبات بمعادلة كيودرريتشاردسون٢١:

تم التحقق من ثبات مستودع المفردات حيث تم استخدام الثبات التقليدى بمعادلة كيودرريتشاردسون لان المفردات ثنائية التدريج كما يتضح من الجدول (١)

جدول (١) تقدير الثبات للاختبار

عدد الاشخاص	عدد المفردات	متوسط الدرجات	معامل كيودرريتشاردسون
١٢٠٠	١٠٠	٥١.٢٧٥٨٣	٠.٩٦٠٨٣٩

ISSN 2735-5209

يتضح من الجدول ان الثبات تم تقديره على المستودع بطول (١٠٠) مفردة متوسها ٥١.٢٧ وبلغت قيمة ثبات الاتساق الداخلى بمعادلة كيودرريتشاردسون ٠.٩٦

تحليل مفردات الاختبار بالنظرية الحديثة:

– تحديد النموذج ثلاثى البارامتر للمفردات ثنائية التدريج

ركزت هذه الدراسة على CCT مع العناصر ثنائية التدريج. يُقال أن العنصر ثنائي التدريج عندما يكون له نتيجتان محتملتان فقط، على سبيل المثال، صحيح وغير صحيح. نماذج IRT ثنائية الأبعاد الثلاثة الأكثر استخدامًا هي النماذج اللوجيستية أحادية وثنائية وثلاثية المعلمات. كما توحي أسمائهم، فإنهم يختلفون في عدد المعلمات التي يدمجونها.

افتراض أحادية البعد:

للتحقق من أحادية البعد تم استخدام التحليل العاملى الاستكشافى من الدرجة الأولى باستخدام برنامج SPSSv.28 حيث تم اجراء التحليل العاملى الاستكشافى من الدرجة الاولى باستخدام طريقة المكونات الاساسية وتم استخراج المكونات العاملية والجذور الكامنة والتباين المفسر لتلك المكونات العاملية كما يتضح من الجدول (٢)

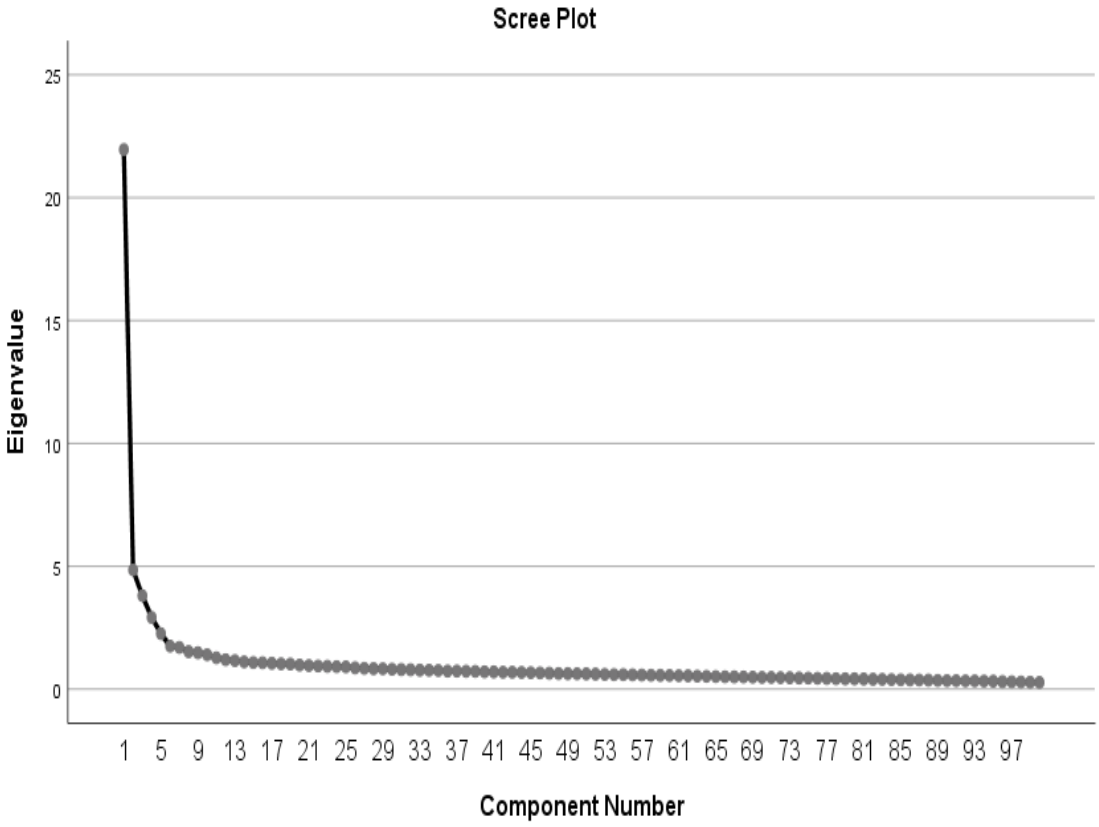
جدول (٢) المكونات العاملية وجذورها الكامنة والتباين المفسر

م	الجذر الكامن	التباين المفسر	التباين التراكمى
١	٢١.٩٥٧	٢١.٩٥٧	٢١.٩٥٧
٢	٤.٨٥١	٤.٨٥١	٢٦.٨٠٧
٣	٣.٨٠٣	٣.٨٠٣	٣٠.٦١١
٤	٢.٩١٩	٢.٩١٩	٣٣.٥٣٠
٥	٢.٢٦٢	٢.٢٦٢	٣٥.٧٩١
٦	١.٧٥٠	١.٧٥٠	٣٧.٥٤٢
٧	١.٦٩٨	١.٦٩٨	٣٩.٢٤٠
٨	١.٥٢٨	١.٥٢٨	٤٠.٧٦٨
٩	١.٤٨٣	١.٤٨٣	٤٢.٢٥٠
١٠	١.٤٠١	١.٤٠١	٤٣.٦٥٢
١١	١.٢٧٨	١.٢٧٨	٤٤.٩٣٠
١٢	١.١٩٥	١.١٩٥	٤٦.١٢٥
١٣	١.١٦٥	١.١٦٥	٤٧.٢٩٠
١٤	١.١١٢	١.١١٢	٤٨.٤٠٢

ISSN 2735-5209

م	الجذر الكامن	التباين المفسر	التباين التراكمى
١٥	١.٠٨٤	١.٠٨٤	٤٩.٤٨٦
١٦	١.٠٧٥	١.٠٧٥	٥٠.٥٦٠
١٧	١.٠٥١	١.٠٥١	٥١.٦١٢
١٨	١.٠٣٣	١.٠٣٣	٥٢.٦٤٤
١٩	١.٠٢١	١.٠٢١	٥٣.٦٦٦

يتضح من الجدول أن بنك الاسئلة يتشبع على (١٩) مكون جذورهم الكامنة أعلى من الواحد الصحيح وتفسر نسبة ٥٣.٦٦٪ كما يتضح ان المكون الاول جذره الكامن (٢١.٩٥٧) ونسبة التباين المفسر له (٢١.٩٥٧٪) بينما المكون الثانى جذره الكامن (٤.٨٥١) والتباين المفسر له (٤.٨٥١٪) ومن ثم فالعامل الاول يفسر أكثر من ٢٠٪ من التباين علاوة على أنه يفسر نسبة ٤.٥٢ % ضعف المكون الثانى. ومن ثم فإن المستودع في شكله النهائي يتوافر به أحادية البعد وما يؤكد ذلك قيا الباحث باستخراج Scree Plot كما يتضح من الشكل التالى:



شكل (١) يوضح scree plot

يتضح من الشكل (١) أن العامل الاول يشكل مساحة كبيرة من التباين بينما باقى العوامل متقاربة من بعضها البعض وهو ما يفيد بتوافر أحادية البعد.

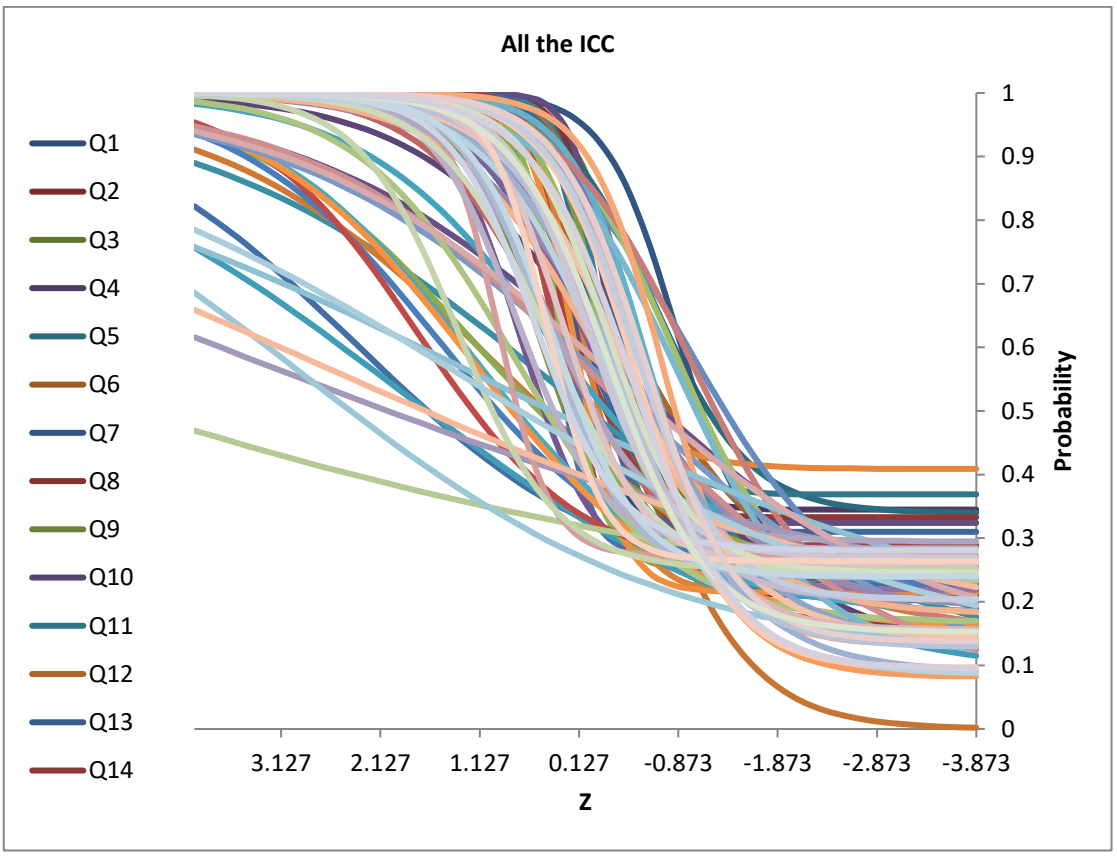
ISSN 2735-5209

التحقق من الاستقلال الموضعى:

أشارت ادبيات القياس أن توافر أحادية البعد دليل على توافر الاستقلال الموضعى ومن ثم فإنه يتوافر للمفردات المستودع الاستقلال الموضعى وما ما تحرى الباحث أن يقوم به أثناء بناء مفردات المستودع حيث تحرى الدقة في ان تقيس كل مفردة هدف محدد وان تمثل بدقة نواتج التعلم المستهدفة وتتلائم مع مستوى الاهداف.

منحنيات خصائص المفردات:

تم التحقق من توافر العلاقة الطردية بين تقدير القدرة واحتمال الاجابة الصحيحة عن المفردات على مستوى مفردات المستودع حيث تم استخراج منحنيات خصائص المفردات للمستودع لكل مفردة من مفردات طبقا للنموذج الثلاثى البارامتر كما يتضح من الشكل (٢)



شكل (٢) منحنيات خصائص المفردات

ISSN 2735-5209

يوضح الرسم البيانى السابق عدد (١٢) مفردة من مفردات المستودع تم عرضها في متن البحث كمثال توضيحى وباقى المنحنيات ستكون في ملاحق البحث حيث يتضح توافر العلاقة الاطرادية بين تقدير القدرة واحتمالية الاجابة الصحيحة عن المفردات ومن ثم يتوافر افتراض منحنيات خصائص المفردات فالعلاقات اطرادية هرمية.

إجراءات الدراسة:

قام الباحث بتطبيق الاختبار على عدد ١٢٠٠ طالب، من طلاب الصف الثالث الثانوي بالمدارس الحكومية بالدمَّام. ومن ثم استخدام برنامج (Bilog-MG3) لإيجاد معاملات الصعوبة والتمييز والتخمين وفق النموذج الثلاثي المعالم لنظرية الاستجابة للفقرة، وإيجاد دقة تقدير معالم الفقرة وفق النموذج الثلاثي المعلمة واختبار مربع كاي لحسن المطابقة للكشف عن مدى مطابقة الفقرات للنموذج اللوجيستي ثلاثي المعلمة. وبعد التحقق من جميع الشروط قام الباحث بتصميم الاختبار التكيفي وفقا لنوع قواعد الانهاء المستخدمة في الدراسة الحالية وتطبيقه على نفس عينة الدراسة ومن ثم المقارنة بينهم للتحقق من أفضل قواعد التوقف من حيث تأثيرها على دقة تقدير القدرة ودالة المعلومات.

الأساليب الاحصائية المستخدمة:

تم استخدام عدد من الاساليب الاحصائية:

– أساليب احصائية ضمن النظرية التقليدية في القياس مثل: معاملات الصعوبة والتمييز ومعامل ثبات الفا لكرونباخ

– أساليب إحصائية ضمن النظرية الحديثة تتضمن أساليب التحقق من افتراضات النظرية وتضمنت التحليل العاملي الاستكشافي من الدرجة الاولى، وبارامترات التمييز والصعوبة والتخمين

– تم استخدام اسلوب الارجحية القصوى واقصى دالة معلومات في ادارة الاختبارات التكيفية وانتقاء المفردات

– استخدام عدة مقاييس للكشف عن دقة الاختبار التكيفي طبقا لقواعد التوقف المحددة

نتائج الدراسة ومناقشتها:

- **ما دقة استخدام قاعدة الطول الثابت في ادارة الاختبار التكيفي (١٠٠، ٥٠، ٤٠، ٣٠)؟**

تعتمد المقاييس المناسبة للاختبارات التكيفية المحوسبة على الطبيعة والهدف من الاختبار. ومع ذلك، فإن بعض المقاييس المشتركة التي يمكن استخدامها لتقييم أداء النماذج في الاختبارات التكيفية المحوسبة هي:

ISSN 2735-5209

١. مقياس المتوسط المربع الجذري للخطأ (Root Mean Square Error – RMSE): يقيس مدي تفوق النموذج في التنبؤ بالقيم الحقيقية.

٢. مقياس الانحراف المطلق المتوسط (Mean Absolute Deviation – MAD): يقيس مدى اختلاف القيم المتوقعة عن القيم الحقيقية.

٣. مقياس معامل الانحياز (Bias coefficient): يقيس مدى انحياز النموذج عن القيم الحقيقية.

من ثم فإن استخدام مجموعة من المقاييس المختلفة يمكن أن يعطي صورة أوضح عن أداء النموذج في الاختبارات التكيفية المحوسبة. ويمكن تحديد المقاييس الأنسب للحالة الخاصة بك حسب الهدف والبيانات المتاحة والسياق العام للاختبار.

الطول الثابت ١٠٠

N = 1200

BIAS(Theta) = −0.05407

RMSE(Theta) = 0.38564

MAE(Theta) = 0.21970

جدول (٣) الطول الثابت ١٠٠

Theta Area	-٢,٥	-٢,٠	-١,٥	-١,٠	-٠,٥	٠,٠	٠,٥	١,٠	١,٥	٢,٠
عدد الحالات	١٤	١٥٠	٢٠٨	١٥٨	٢٣١	١٩٥	١٠١	٩٧	١٧	٢٩
CBias	-٠,٢٣٠	-٠,١٩٧	-٠,١٨٨	-٠,٠٠٣	٠,٠٠٩	٠,٠٠٢	٠,٠١٦	٠,٠٥٥	٠,٠٥٩	٠,٠٢٧
CMAE	٠,٧٢١	٠,٥٧٤	٠,٣٧٤	٠,١٢٦	٠,٠٧١	٠,٠٦٢	٠,٠٨٧	٠,١٩٤	٠,٢٢٨	٠,٣٢٢
CRMSE	٠,٧٧١	٠,٧١٩	٠,٥٧٩	٠,١٦٠	٠,٠٩٠	٠,٠٨٠	٠,١١٢	٠,٢٨٦	٠,٢٨٧	٠,٣٩١
طول الاختبار	١٠٠	١٠٠	١٠٠	١٠٠	١٠٠	١٠٠	١٠٠	١٠٠	١٠٠	١٠٠

تعد CBias وCMAE وCRMSE هي كلها مقاييس تستخدم عادة لتقييم دقة النماذج الإحصائية في عمليات التنبؤ. وإليك شرحًا موجزًا لكل منها:

١. CBias: هو اختصار ل "الانحياز الثابت"، ويقيس الفرق المتوسط بين القيم المتوقعة للنموذج والقيم الحقيقية عبر جميع قيم المتغير المتنبئ. تشير قيمة CBias السلبية إلى أن النموذج يميل إلى تقدير القيم الحقيقية بشكل غير كافٍ، في حين تشير القيمة الإيجابية إلى أنه يميل إلى تقديرها بشكل زائد. تشير قيمة CBias المساوية للصفر إلى أن النموذج غير محايد.

٢. CMAE: هو اختصار ل "الخطأ المطلق المتوسط الثابت"، ويقيس الفرق المطلق المتوسط بين القيم المتوقعة للنموذج والقيم الحقيقية عبر جميع قيم المتغير المتنبئ. تشير قيمة CMAE الأقل إلى أن النموذج يقدم توقعات أكثر دقة.

ISSN 2735-5209

٣. CRMSE: هو اختصار ل "الخطأ المربع المتوسط الجذري الثابت"، ويقيس الجذر التربيعي للفرق المربعية المتوسطة بين القيم المتوقعة للنموذج والقيم الحقيقية عبر جميع قيم المتغير المتنبئ. مثل CMAE، تشير قيمة CRMSE الأقل إلى أن النموذج يقدم توقعات أكثر دقة.

يجدر بالذكر أن لكل من هذه المقاييس مزاياها وعيوبها الخاصة، وعادة ما يكون من الأفضل تقييم النموذج باستخدام مقاييس متعددة للحصول على صورة أكثر اكتمالًا لأدائه. بالإضافة إلى ذلك، يمكن أن تختلف تفسير هذه المقاييس باختلاف السياق وأهداف النموذج المحدد.

لكل من هذه المقاييس مزايا وعيوبها، وفيما يلي بعض النقاط التي يجب مراعاتها عند استخدامها:

– CBias: يمكن استخدامها لتقييم مدى وجود تحيز في النموذج، وهذا يمكن أن يكون مفيدًا في بعض الحالات مثل توقعات الأسعار. ومع ذلك، فإن CBias ينظر إلى الانحياز الثابت فقط، ولا يأخذ في الاعتبار أي تغيير في الانحياز حسب قيم المتغيرات الأخرى، وهذا يعني أنه لا يمكن استخدامها لتقييم الانحياز المتغير.

– CMAE وCRMSE: تعطيان فكرة عن مدى دقة النموذج في التنبؤ بالقيم الحقيقية، ويعدان مناسبين لتقييم الأداء العام للنموذج. ومع ذلك، فإن القيم القياسية مثل CMAE وCRMSE يتأثران بالقيم الشاذة (الأوتلاينرز)، وقد يؤدي ذلك إلى تضخيم الأخطاء في بعض الحالات، مما يؤدي إلى تقييم غير دقيق لأداء النموذج.

– يجب مراعاة السياق: يجب أن يتم تقييم هذه المقاييس في السياق المناسب وحسب أهداف النموذج المحدد. فعلى سبيل المثال، يمكن أن يكون الهدف من النموذج تحقيق دقة عالية، ولكن يمكن أن يكون الهدف الآخر هو تقليل الخسائر، وهذا يعني أن الاختيار بين المقاييس يتوقف على الهدف المحدد.

– أهمية البيانات: يجب أن يتم اختبار النموذج على بيانات غير مستخدمة في التدريب لضمان عدم تحيز النموذج، كما يجب الانتباه إلى جودة البيانات وتوزيعها، حيث يمكن أن يؤثر ذلك على دقة النموذج وبالتالي على قدرة المقاييس على تقييم أدائه.

تتعلق هذه النتائج بنموذج تنبؤ يستخدم فيه متغير θ لتوقع قيمة متغير الاستجابة y من مدخلات أخرى. وتشير النتائج إلى أن النموذج يعمل بشكل جيد على البيانات التي تم استخدامها في تدريب النموذج.

– N = 1200: يشير إلى عدد الحالات التي تم استخدامها في تدريب النموذج.

– BIAS(Theta) = −0.02780: يشير إلى الانحياز الذي يتوقعه النموذج في القيمة المتوقعة للمتغير الاستجابة مقارنة بالقيمة الفعلية. هذا الرقم يشير إلى أن النموذج قد يكون قليلاً معرفيًا، حيث يميل إلى تقديم توقعات أقل من القيم الفعلية.

– RMSE(Theta) = 0.34300: يشير إلى مدى الانحراف المعياري للتوقعات التي يقدمها النموذج مقارنة بالقيم الفعلية. يتم حسابه بأخذ الفرق بين التوقعات والقيم الفعلية، ثم رفع هذه

ISSN 2735-5209

الفروقات إلى السلبيتين وجمعها وتقسيمها على عدد الحالات. هذا الرقم يشير إلى أن النموذج يوفر توقعات جيدة بشكل عام، حيث يبلغ متوسط الخطأ النسبي الجذري (RMSE) حوالي ٠.٠٣٤٣٠٠.

– MAE(Theta) = 0.20656: يشير إلى مدى الانحراف المطلق للتوقعات التي يقدمها النموذج مقارنة بالقيم الفعلية. يتم حسابه بأخذ الفرق بين التوقعات والقيم الفعلية وأخذ القيم المطلقة لهذه الفروقات وجمعها وتقسيمها على عدد الحالات. هذا الرقم يشير إلى أن النموذج يوفر توقعات جيدة بشكل عام، حيث يبلغ الانحراف المطلق المتوسط (MAE) حوالي ٠.٠٢٠٦٥٦.

– Theta Area: يشير إلى توزيع القيم التي يمكن أن تأخذها متغير θ.

– # of Cases: يشير إلى عدد الحالات التي تم استخدامها في كل منطقة من مناطق θ.

– CBias: يشير إلى الانحياز المركزي الذي يتوقعه النموذج في القيمة المتوقعة للمتغير الاستجابة مقارنة بالقيم الفعلية في كل منطقة من مناطق θ.

– CMAE: يشير إلى الانحراف المطلق المتوسط الذي يتوقعه النموذج في التوقعات المقدمة للمتغير الاستجابة في كل منطقة من مناطق θ.

– CRMSE: يشير إلى الانحراف المعياري الجذري الذي يتوقعه النموذج في التوقعات المقدمة للمتغير الاستجابة في كل منطقة من مناطق θ.

– Test Length: يشير إلى المدة التي تم اختبار النموذج خلالها على بيانات الاختبار. ويبدو أنه تم اختبار النموذج على جميع الحالات المتاحة في بيانات الاختبار.

يمثل الجدول Theta Area −٢.٥ −٢.٠ −١.٥ −١.٠ −٠.٥ ٠.٠ ٠.٥ ١.٠ ١.٥ ٢.٠ ٢.٥ عدد الحالات، والانحياز المركزي، والانحراف المطلق المتوسط، والانحراف المعياري الجذري، وطول الاختبار لنموذج التنبؤ.

وفقًا للجمعية الأمريكية لعلم النفس، فإن الانحياز المركزي هو قدر التحيز المتوقع في القيم المتوقعة للمتغير الاستجابة مقارنة بالقيم الفعلية في كل منطقة من مناطق θ. وفي هذا الجدول، يتم رصد الانحياز المركزي للنموذج في كل منطقة من مناطق θ. ويوضح هذا الجدول أن النموذج يسجل تحيزًا سلبيًا معتدلًا في المناطق −٢.٥ و−٢.٠ و−١.٥ و٠.٥ و١.٥ و٢.٠ و٢.٥، وتحيزًا موجبًا في المناطق −٠.٥ و١.٠.

بالنسبة للانحراف المطلق المتوسط، فإنه يقيس مدى الانحراف في التوقعات التي يقدمها النموذج مقارنة بالقيم الفعلية في كل منطقة من مناطق θ. ويوضح هذا الجدول أن النموذج يقدم توقعات دقيقة جدًا في المناطق −١.٠ و−٠.٥ و٠.٠، ويسجل انحرافًا معتدلًا في المناطق −٢.٥ و−٢.٠ و−١.٥ و٠.٥ و١.٥ و٢.٠ و٢.٥، ويسجل انحرافًا كبيرًا في المناطق ١.٠ و١.٥.

أما بالنسبة للانحراف المعياري الجذري، فإنه يقيس مدى الانحراف في التوقعات التي يقدمها النموذج مقارنة بالقيم الفعلية في كل منطقة من مناطق θ. ويوضح هذا الجدول أن النموذج يقدم توقعات دقيقة جدًا في المناطق −١.٠ و−٠.٥ و٠.٠، ويسجل انحرافًا معتدلًا في المناطق −٢.٥

ISSN 2735-5209

و-٢.٠ و-١.٥ و٠.٥ و١.٥ و٢.٠ و٢.٥، ويسجل انحرافًا كبيرًا في المناطق ١.٠ و١.٥ و٢.٠ و٢.٥.

وبالتالى يوضح هذا الجدول أن النموذج يقدم توقعات جيدة جدًا في بعض المناطق، ويسجل انحرافًا معتدلًا في المناطق الأخرى. ومع ذلك، يبدو أنه يوجد بعض التحيز في بعض المناطق، ويمكن أن يكون هذا الأمر مهمًا عند استخدام النموذج في توقعات المستقبل.

حيث يوضح الجدول مقاييس الأداء المختلفة مثل CBias وCMAE وCRMSE لنماذج مختلفة في اختبار تكيفي طول ثابت ١٠٠ مفردة، ويظهر أيضًا Theta Area المختلفة وعدد الحالات المستخدمة في الاختبار. وبالنسبة لمدى الدقة من خلال مقاييس الأداء المختلفة، فإن CBias هو قياس الانحياز، ويظهر من الجدول أن قيم CBias للنماذج تتراوح بين -٠.٢٣٠ و٠.٠٥٩، وهذا يشير إلى أن بعض النماذج قد تكون أكثر دقة من غيرها في تقدير القيم الحقيقية للمتغيرات.

أما بالنسبة لمقياس CMAE فهو قياس الانحراف المطلق المتوسط، ويظهر من الجدول أن قيم CMAE للنماذج تتراوح بين ٠.٠٦٢ و٠.٧٢١، وهذا يشير إلى أن بعض النماذج قد تكون أكثر دقة من غيرها في تقدير القيم الحقيقية للمتغيرات.

وأخيرًا، يبين المقياس CRMSE جذر المتوسط المربعي للخطأ، ويظهر من الجدول أن قيم CRMSE للنماذج تتراوح بين ٠.٠٨٠ و٠.٧٧١، وهذا يشير إلى أن بعض النماذج قد تكون أكثر دقة من غيرها في تقدير القيم الحقيقية للمتغيرات.

ويمكن القول أن النتائج تدل على أن بعض النماذج أكثر دقة من غيرها في تقدير القيم الحقيقية للمتغيرات، ويمكن استخدام هذه المعلومات لتحديد النماذج الأكثر دقة في المستقبل وتحسين أدائها.

الطول الثابت ٣٠

$$N = 1200$$

$$BIAS(Theta) = -0.04668$$

$$RMSE(Theta) = 0.38336$$

$$MAE(Theta) = 0.23366$$

جدول (٤) الطول الثابت ٣٠

Theta Area	-٢,٥	-٢,٠	-١,٥	-١,٠	-٠,٥	٠,٠	٠,٥	١,٠	١,٥	٢,٠
عدد الحالات	١٤	١٥٠	٢٠٨	١٥٨	٢٣١	١٩٥	١٠١	٩٧	١٧	٢٩
CBias	-٠,١٦١	-٠,٢١١	-٠,١٠٦	-٠,٠٧٧	٠,٠١٢	٠,٠١٤	٠,٠٣٤	٠,١٣٦	-٠,٠٢٥	-٠,٠٤٧
CMAE	٠,٦٣٧	٠,٥٣٨	٠,٣٢٦	٠,٢٠٧	٠,٠٩٦	٠,٠٨٧	٠,١١٦	٠,٢٦٠	٠,٢٩٦	٠,٣٥٧
CRMSE	٠,٧١٢	٠,٦٧٢	٠,٥٠٧	٠,٣٦٢	٠,١٢٤	٠,١٠٩	٠,١٥٦	٠,٣٥٤	٠,٣٧٣	٠,٤٢٢
طول الاختبار	٣٠	٣٠	٣٠	٣٠	٣٠	٣٠	٣٠	٣٠	٣٠	٣٠

ISSN 2735-5209

يوضح هذا الجدول مقاييس الأداء المختلفة مثل CBias وCMAE وCRMSE لنماذج مختلفة في اختبار تكيفي طول ثابت ٣٠ مفردة، ويظهر أيضًا Theta Area المختلفة وعدد الحالات المستخدمة في الاختبار.

بالنسبة لمدى الدقة من خلال مقاييس الأداء المختلفة، فإن CBias هو قياس الانحياز، ويظهر من الجدول أن قيم CBias للنماذج تتراوح بين −٠.٢١١ و٠.١٣٦، وهذا يشير إلى أن بعض النماذج قد تكون أكثر دقة من غيرها في تقدير القيم الحقيقية للمتغيرات.

أما بالنسبة لمقياس CMAE فهو قياس الانحراف المطلق المتوسط، ويظهر من الجدول أن قيم CMAE للنماذج تتراوح بين ٠.٠٨٧ و٠.٦٣٧، وهذا يشير إلى أن بعض النماذج قد تكون أكثر دقة من غيرها في تقدير القيم الحقيقية للمتغيرات.

وأخيرًا، يبين المقياس CRMSE جذر المتوسط المربعي للخطأ، ويظهر من الجدول أن قيم CRMSE للنماذج تتراوح بين ٠.١٠٩ و٠.٧١٢، وهذا يشير إلى أن بعض النماذج قد تكون أكثر دقة من غيرها في تقدير القيم الحقيقية للمتغيرات.

ومن ثم يمكن القول أن النتائج تدل على أن بعض النماذج أكثر دقة من غيرها في تقدير القيم الحقيقية للمتغيرات، ويمكن استخدام هذه المعلومات لتحديد النماذج الأكثر دقة في المستقبل وتحسين أدائها، ومن المهم تذكر أن استخدام طول الاختبار القصير (٣٠) يمكن أن يؤدي إلى قياسات أداء غير دقيقة للنماذج.

الطول الثابت ٤٠

N = 1200

BIAS(Theta) = −0.03866

RMSE(Theta) = 0.37267

MAE(Theta) = 0.22612

جدول (٥) الطول الثابت ٤٠

Theta Area	-٢,٥	-٢,٠	-١,٥	-١,٠	-٠,٥	٠,٠	٠,٥	١,٠	١,٥	٢,٠
عدد الحالات	١٤	١٥٠	٢٠٨	١٥٨	٢٣١	١٩٥	١٠١	٩٧	١٧	٢٩
CBias	٠,١١٦	-٠,٢٤٢	-٠,١١٠	-٠,٠٣٣	٠,٠٠٥	٠,٠٠٩	٠,٠٠٨	٠,٠٢٦	٠,٢٠٤	٠,٣١٦
CMAE	٠,٥٧٤	٠,٥٨٤	٠,٣١٥	٠,١٧٤	٠,٠٨٧	٠,٠٧٠	٠,١٢٠	٠,٢٠٥	٠,٣٨٠	٠,٤٠٤
CRMSE	٠,٦٤٩	٠,٧٢٢	٠,٤٦٩	٠,٢٧٦	٠,١١٩	٠,٠٨٩	٠,١٥٩	٠,٢٦٥	٠,٤٥٢	٠,٥١٧
طول الاختبار	٤٠	٤٠	٤٠	٤٠	٤٠	٤٠	٤٠	٤٠	٤٠	٤٠

يوضح هذا الجدول مقاييس الأداء المختلفة مثل CBias وCMAE وCRMSE لنماذج مختلفة في اختبار تكيفي طول ثابت ٤٠ مفردة، ويظهر أيضًا Theta Area المختلفة وعدد الحالات المستخدمة في الاختبار.

ISSN 2735-5209

بالنسبة لمدى الدقة من خلال مقاييس الأداء المختلفة، فإن CBias هو قياس الانحياز، ويظهر من الجدول أن قيم CBias للنماذج تتراوح بين –٠.٢٤٢ و٠.٣١٦، وهذا يشير إلى أن بعض النماذج قد تكون أكثر دقة من غيرها في تقدير القيم الحقيقية للمتغيرات.

أما بالنسبة لمقياس CMAE فهو قياس الانحراف المطلق المتوسط، ويظهر من الجدول أن قيم CMAE للنماذج تتراوح بين ٠.٠٧٠ و٠.٥٨٤، وهذا يشير إلى أن بعض النماذج قد تكون أكثر دقة من غيرها في تقدير القيم الحقيقية للمتغيرات.

وأخيرًا، يبين المقياس CRMSE جذر المتوسط المربعي للخطأ، ويظهر من الجدول أن قيم CRMSE للنماذج تتراوح بين ٠.٠٨٩ و٠.٧٢٢، وهذا يشير إلى أن بعض النماذج قد تكون أكثر دقة من غيرها في تقدير القيم الحقيقية للمتغيرات.

ومن ثم يمكن القول أن النتائج تدل على أن بعض النماذج أكثر دقة من غيرها في تقدير القيم الحقيقية للمتغيرات، ويمكن استخدام هذه المعلومات لتحديد النماذج الأكثر دقة في المستقبل وتحسين أدائها. ومن المهم تذكر أن استخدام طول الاختبار الأطول (٤٠) يمكن أن يؤدي إلى قياسات أداء أكثر دقة للنماذج.

الطول الثابت ٥٠

N = 1200

BIAS(Theta) = −0.04915

RMSE(Theta) = 0.38983

MAE(Theta) = 0.23079

جدول (٦) الطول الثابت ٥٠

Theta Area	-٢,٥	-٢,٠	-١,٥	-١,٠	-٠,٥	٠,٠	٠,٥	١,٠	١,٥	٢,٠
عدد الحالات	١٤	١٥٠	٢٠٨	١٥٨	٢٣١	١٩٥	١٠١	٩٧	١٧	٢٩
CBias	-٠,١٠٩	-٠,١٩٩	-٠,١٨٤	-٠,٠١٢	٠,٠٠٣	٠,٠٠٩	٠,٠٣٨	٠,٠٤٤	٠,١٢٥	٠,٠٧٠
CMAE	٠,٥٨٩	٠,٥٣٤	٠,٤٠٩	٠,١٥٥	٠,٠٧٩	٠,٠٦٧	٠,١٢١	٠,٢١٨	٠,٢٨٥	٠,٣٣٥
CRMSE	٠,٦٤٥	٠,٦٧٢	٠,٦٠٦	٠,٢٤٧	٠,١٠٣	٠,٠٨٣	٠,١٥٦	٠,٢٧٢	٠,٤٩٨	٠,٤١٣
طول الاختبار	٥٠	٥٠	٥٠	٥٠	٥٠	٥٠	٥٠	٥٠	٥٠	٥٠

يوضح هذا الجدول مقاييس الأداء المختلفة مثل CBias وCMAE وCRMSE لنماذج مختلفة في اختبار تكيفي طول ثابت ٥٠ مفردة، ويظهر أيضًا Theta Area المختلفة وعدد الحالات المستخدمة في الاختبار.

بالنسبة لمدى الدقة من خلال مقاييس الأداء المختلفة، فإن CBias هو قياس الانحياز، ويظهر من الجدول أن قيم CBias للنماذج تتراوح بين –٠.١٩٩ و٠.١٢٥، وهذا يشير إلى أن بعض النماذج قد تكون أكثر دقة من غيرها في تقدير القيم الحقيقية للمتغيرات.

ISSN 2735-5209

أما بالنسبة لمقياس CMAE فهو قياس الانحراف المطلق المتوسط، ويظهر من الجدول أن قيم CMAE للنماذج تتراوح بين ٠.٠٦٧ و٠.٥٨٩، وهذا يشير إلى أن بعض النماذج قد تكون أكثر دقة من غيرها في تقدير القيم الحقيقية للمتغيرات.

وأخيرًا، يبين المقياس CRMSE جذر المتوسط المربعي للخطأ، ويظهر من الجدول أن قيم CRMSE للنماذج تتراوح بين ٠.٠٨٣ و٠.٦٧٢، وهذا يشير إلى أن بعض النماذج قد تكون أكثر دقة من غيرها في تقدير القيم الحقيقية للمتغيرات.

ومن ثم يمكن القول أن النتائج تدل على أن بعض النماذج أكثر دقة من غيرها في تقدير القيم الحقيقية للمتغيرات، ويمكن استخدام هذه المعلومات لتحديد النماذج الأكثر دقة في المستقبل وتحسين أدائها. ومن المهم تذكر أن استخدام طول الاختبار الأطول (٥٠) يمكن أن يؤدي إلى قياسات أداء أكثر دقة للنماذج.

المقارنة بين النماذج الاربعة:

يمكن استخدام هذه النتائج لمقارنة الأداء بين النماذج الأربعة. وفيما يلي تحليل بسيط لكل مقياس:

١. BIAS: يشير إلى مدى انحياز النموذج عن القيم الحقيقية. يوضح النموذج الثابت ٣٠ انحيازًا قليلاً إلى الأسفل (-٠.٠٤٦٦٨)، في حين يوضح النموذج الثابت ٤٠ انحيازًا أقل (-٠.٠٣٨٦٦)، أيضًا إلى الأسفل، ويوضح النموذج الثابت ٥٠ انحيازًا قليلاً إلى الأسفل (-٠.٠٤٩١٥). ويوضح النموذج الثابت ١٠٠ انحيازًا أعلى قليلاً (-٠.٠٥٤٠٧) إلى الأسفل. وبالتالي، يمكن القول إن النموذج الثابت ٤٠ أفضل من النموذج الثابت ٣٠ والنموذج الثابت ٥٠ في هذا المقياس، ولكن النموذج الثابت ١٠٠ يوضح أداءً أسوأ من النموذج الثابت ٣٠ والنموذج الثابت ٤٠ والنموذج الثابت ٥٠ في هذا المقياس.

٢. RMSE: يشير إلى مدى تفوق النموذج في التنبؤ بالقيم الحقيقية. يوضح النموذج الثابت ٣٠ متوسط خطأ جذري مربعي (RMSE) يبلغ ٠.٣٨٣٣٦، في حين يوضح النموذج الثابت ٤٠ RMSE يبلغ ٠.٣٧٢٦٧، ويوضح النموذج الثابت ٥٠ RMSE يبلغ ٠.٣٨٩٨٣، ويوضح النموذج الثابت ١٠٠ RMSE يبلغ ٠.٣٨٥٦٤. وبالتالي، يمكن القول إن النموذج الثابت ٤٠ أفضل من النموذج الثابت ٣٠ والنموذج الثابت ٥٠ في هذا المقياس، ولكن النموذج الثابت ١٠٠ يوضح أداءً أسوأ من النموذج الثابت ٤٠ والنموذج الثابت ٣٠ في هذا المقياس.

٣. MAE: يشير إلى مدى تقارب القيم المتوقعة من القيم الحقيقية. يوضح النموذج الثابت ٣٠ متوسط خطأ مطلق (MAE) يبلغ ٠.٢٣٣٦٦، في حين يوضح النموذج الثابت ٤٠ MAE يبلغ ٠.٢٢٦١٢، ويوضح النموذج الثابت ٥٠ MAE يبلغ ٠.٢٣٠٧٩، ويوضح النموذج الثابت ١٠٠ MAE يبلغ ٠.٢١٩٧٠. وبالتالي، يمكن القول إن النموذج الثابت ٤٠ أفضل من النموذج الثابت

ISSN 2735-5209

٣٠ والنموذج الثابت ٥٠ في هذا المقياس، ولكن النموذج الثابت ١٠٠ يوضح أداءً أفضل من النموذج الثابت ٣٠ والنموذج الثابت ٤٠ والنموذج الثابت ٥٠ في هذا المقياس.

وبناء على ذلك يمكن القول إن النموذج الثابت ٤٠ هو الأفضل بين النماذج الأربعة، حيث يوضح أداءً أفضل في مقاييس RMSE وMAE. ومع ذلك، فإن الاختيار بين النماذج الأربعة يعتمد على السياق والأهداف المحددة للنموذج. يمكن استخدام النموذج الثابت ٣٠ إذا كان الهدف هو تقليل التعقيد في النموذج، بينما يمكن استخدام النموذج الثابت ٤٠ إذا كان الهدف هو تحقيق أفضل أداء في التنبؤ بالقيم الحقيقية. يمكن استخدام النموذج الثابت ٥٠ أو النموذج الثابت ١٠٠ إذا كان الهدف هو تحقيق أداءً أفضل في مقاييس محددة مثل RMSE أو MAE.

التوصيات:

١- إجراء دراسات أخرى لتقييم أداء النماذج باستخدام مقاييس الأداء المختلفة لتحديد أفضل النماذج في مجال الاختبار التكيفي المحوسب.

٢- يمكن دراسة تأثير عوامل أخرى على أداء النماذج في الاختبار التكيفي المحوسب، مثل تأثير البيانات المستخدمة وخوارزميات التكيف المختلفة.

٣- يمكن استخدام هذه النتائج لتحسين أداء النماذج الحالية في مجال الاختبار التكيفي المحوسب، وتحسين القدرة على التكيف مع بيانات جديدة ومتغيرة.

المراجع:

أولا: المراجع العربية:

الصبيحي، على بن أحمد، والأحمدي، شرف بنت حامد (٢٠١١). الاختبار التكيفي المحوسب" كتطبيق لتطوير أساليب التقويم والاختبارات الإلكترونية". *مجلة كلية التربية*. جامعة دمنهور، المجلد الثالث العدد(١)، ٨٤-١٢٦.

عبيدات، عمر سليمان، وعودة، أحمد سليمان (٢٠١٣). فاعلية الاختبار التكيفي المحوسب في تقدير القدرة العقلية باستخدام مصفوفات رافن. *مجلة دراسات العلوم التربوية*، ٤٠ (٢)، ١٦٠٢-١٦٢١.

ثانيا: المراجع الأجنبية:

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

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

ISSN 2735-5209

Jacobusse, G., & Buuren, S. V. (2007). Computerized adaptive testing for measuring development of young children. *Statistics in medicine*, *26*(13), 2629-2638.

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

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

Stafford, R. E., Runyon, C. R., Casabianca, J. M., & Dodd, B. G. (2019). Comparing computer adaptive testing stopping rules under the generalized partial-credit model. Behavior research methods, 51(3), 1305-1315.

ISSN 2735-5209