أثر اختلاف عدد البدائل وموقع البديل الصحيح على ثبات الاختبار التحصيلي واتساقه الداخلي بين نظرية القياس الكلاسيكية والنموذج الإحصائي لبايزي

د. عبدالرحمن بن سالم بن محمد الشغيبي الشهري
جامعة الملك خالد /كلية التربية /أستاذ القياس والتقويم المساعد/ علم النفس
المملكة العربية السعودية

استلام البحث: ٢٠٢٣/٤/٧ قبول النشر: ٢٠٢٣/٨/٢ تاريخ النشر : ٢ /١/ ٢٠٢٤
https://doi.org/10.52839/0111-000-080-004

ملخص البحث

هدف البحث إلى التعرف على مدى تأثير اختلاف عدد بدائل السؤال الموضوعي وموقع البديل الصحيح على الثبات والاتساق الداخلي لاختبار في مقرر القياس والتقويم بين نظرية القياس الكلاسيكية والنموذج الإحصائي لبايزي لدى طلاب كلية التربية جامعة الملك خالد، ولتحقيق تلك الأهداف اختار الباحث عينة من طلاب كلية التربية حجمها (١٠٠) طالب يدرسون مقرر القياس والتقويم، وبناء اختبار تحصيلي موضوعي (أربع بدائل)، وتم صياغة الأسئلة لتناسب بديلين وثلاثة بدائل، تم استخدم الإحصاء البايزي Bayesian Statistics من خلال برنامج جافيز الإحصائي المعروف اختصارا JASP والذي يساعد في القيام بالاختبارات الإحصائية الكلاسيكية والبايزية معا، وتوصل الباحث إلى النتائج التالية: كلما زاد عدد البدائل كلما زاد ثبات الاختبار، كما أن الفروق بين قيمة معامل الثبات بإحصاء بايزي وقيمته بالإحصاء التقليدي، كانت فروقا بسيطة، وجود اتساق داخلي بين الأسئلة والدرجة الكلية للاختبار في حالة عدد البدائل أربعة، كما وجد اتفاق كبير بين نتائج إحصاء بايزي والإحصاء التقليدي، بينما على العكس في حالة البديلين فقد وجد اتساق داخلي ضعيف في حالة الإحصاء التقليدي ومتوسطا في حالة إحصاء بايزي، حيث كانت دلالات الارتباط وميلها جهة الفرض البديل في حالة إحصاء بايزي أكثر منها في حالة الإحصاء البديل، معامل الثبات يزداد في الإحصاء البايزي عنه في الإحصاء التقليدي، وأن موقع البديل الصحيح في الإحصاء البايزي له تأثير طفيف على قيمة معامل الثبات (٠,٦٩٣ مقابل ٠,٧٠٠)، بينما له تأثير أكثر وضوحا في الإحصاء التقليدي (٠,٧٢٨ في مقابل ٠,٧٥٦)، وجود تشابه في مستوى الدلالة بين النظرية التقليدية وإحصاء بايزي، وكانت جميع معاملات الارتباط في صالح الفرض البديل الذي يشير إلى وجود ارتباط موجب دال بين درجة السؤال، والدرجة الكلية للاختبار، فيما عدا سؤال واحد كانت لصالح الفرض الصفري الذي يشير إلى عدم وجود ارتباط موجب دال بين درجة السؤال والدرجة الكلية للاختبار، وجود تشابه في مستوى الدلالة بين النظرية التقليدية وإحصاء بايزي، وكانت جميع معاملات الارتباط في صالح الفرض البديل الذي يشير إلى وجود ارتباط موجب دال بين درجة السؤال، والدرجة الكلية للاختبار، فيما عدا (٣) أسئلة من بين الأسئلة العشرة، كانت لصالح الفرض الصفري الذي يشير إلى عدم وجود ارتباط موجب دال بين درجة السؤال والدرجة الكلية للاختبار.

الكلمات المفتاحية: عدد البدائل ، الاختبار التحصيلي ، نظرية القياس الكلاسيكية ، نموذج بايزي

The Effect of Varying the Number of Alternatives and the Location of the Correct Alternative on the Reliability of the Achievement Test and Its Internal Consistency between Classical Measurement Theory and Paizi's Statistical Model

Dr. Abdul Rahman bin Salem bin Mohammad Al-Shughaibi Al-Shehri
Assistant Professor of Measurement and Evaluation, College of Education- Psychology, King Khalid University,Saudi Arabia
alshehri92@gmail.com

Abstract

The research aims to determine the degree to which the variation in the number of alternatives to the objective question and the location of the correct alternative affects the stability and internal consistency of a test in the measurement and evaluation course between the Bayesian statistical model and the classical measurement theory among King Khalid University College of Education students. An objective accomplishment exam with four possibilities was created by (100) students who studied the measuring and assessment process. The questions were designed to accommodate options two and three. The JASP statistical software was utilized to apply Bayesian statistics, aiding in the execution of both conventional and Bayesian statistical tests. The researcher reached the following results: the greater the number of alternatives, the greater the stability of the test, and the differences between the value of the stability coefficient in Bayesian statistics and its value in traditional statistics were simple differences, and there was an internal consistency between the questions and the total score of the test in the case of the number of alternatives being four, and a large agreement was found. between the results of the Bayesian statistic and the traditional statistic, while on the contrary, in the case of the two alternatives, a weak internal consistency was found in the case of the traditional statistic and medium In the case of Bayesian statistics, where the correlation indicatives and their inclination towards the alternative hypothesis were more in the case of Bayesian statistics than in the case of the alternative statistics, the stability coefficient is higher in the Bayesian statistics than in the traditional statistics, and that the location of the correct alternative in the Bayesian statistics has a slight effect on the value of the stability coefficient (0.693 vs. 0.700), while it has a more clear effect in the traditional statistics (0.728 vs. 0.756), and there is a similarity in the level of significance between the traditional theory and Bayesian statistics, and all correlation coefficients were in favor of the alternative hypothesis, which indicates the presence of a significant positive correlation between the question score , and the total score of the test, with the exception of one question, was in favor of the null hypothesis, which indicates that there is no significant positive correlation between the score of the question and the total score of the test, and there is a similarity in the level of significance between the traditional theory and Bayesian statistics, and all correlation coefficients were in favor of the alternative hypothesis, which indicates There is no significant positive correlation between the score of the question and the total score of the test, except for three questions among the ten, which were in favor of the null hypothesis, which indicates that there is no significant positive correlation between the score of the question and the total score for testing.

Keywords: number of alternatives, achievement test, classical measurement theory, Paizi's model

مقدمة:

هناك اتجاه متزايد نحو استخدام الأسئلة الموضوعية في تقويم الطلاب في مختلف المراحل التعليمية، وخصوصا في الجامعات. وقد تزايد هذا الأمر في الوقت الحاضر بعد جائحة كورونا التي ضربت جميع بلدان العالم، وأدت إلى ضرورة حدوث التباعد بين الطلاب مع استمرار تعليمهم، وهذا يقتضي التعليم عن بعد، والتي تمثل الاختبارات الموضوعية جزءًا من تقييم الطلاب في هذا النوع من التعليم.

ولا تزال الاختبارات التحصيلية هي الأداة الأكثر انتشارا في عملية تقويم الطلاب، ولا يوجد حتى الآن بديل يجعلنا نستغني عنها، ولكن تبذل جهود لضمان إعدادها بالصورة التي تحقق تقييما حقيقيا للطلاب. وتكتسب عملية تقويم الطلاب أهمية كبيرة بقدر أهمية القرارات التي تبنى عليها، وبقدر خطورة القرارات الخاطئة التي يمكن أن تترتب على ذلك في المواقف والمجالات المتعددة، وبالتالي يجب أن توفر الاختبارات معلومات تتسم بالصدق والدقة، وهذا يمكن أن يتحقق من خلال التخطيط والإعداد الجيد لها (النصراوين، ٢٠١٩).

وتتميز الأسئلة الموضوعية عن أسئلة المقال، بوجود إجابات محددة من خلال عدد من الاختيارات، تبدأ من خلال إجابة من بين اختيارين (الصواب والخطأ)، وقد تمتد الاختيارات (البدائل) لثلاثة أو أربع أو خمس بدائل أو أكثر، وقد أوضح Budescu & Nevo,1985 ; Straton & Catts, 1980 بأن اختلاف عدد البدائل يؤثر على الخصائص السيكومترية (الثبات والصدق)، على مستوى الاختبار، أو أسئلته، فزيادة البدائل يزيد الثبات، وقد أوضح Eble & Frisbie 1986، أن اختلاف عدد البدائل في اختبارات الاختيار من متعدد يؤدي إلى تأثير على ثبات الاختبار من خلال التأثير على صعوبة السؤال ومعامل تمييزه. في حين توصل Grier, 1975 إلى أن قلة عدد البدائل يزيد من ثبات الاختبار. وفي خضم هذا التناقض أجريت العديد من الدراسات العربية منها، دراسة القحطاني (٢٠٢١)، والتي توصلت إلى أن عدد البدائل (٤) يزيد الثبات عنه في حالة عدد البدائل (٥)، مع العينة التي ليس لديه اكتئاب، بينما العكس صحيح مع العينة التي لديها اكتئاب. ودراسة حميدة (٢٠٢٠)، التي توصلت إلى أن عدد البدائل (٥) يزيد الثبات عنه في حالة عدد البدائل (٣). ودراسة المرواني وسليمان (٢٠١٩)، التي توصلت إلى أن اختلاف موقع البديل له تأثير على ثبات الاختبار الموضوعي. من هنا يتضح أن هناك عوامل كثيرة تلعب دورا في دقة نتائج الاختبارات التحصيلية، فبجانب طول الاختبار وتباين فقراته وزمن الاختبار التي تم تناولها من قبل الباحثين وأصبحت مسلمات في تأثيرها على الثبات، ظهرت عوامل أخرى مثل عدد البدائل وموقع البديل الصحيح. لذلك يحاول الباحث الحالي التحقق من مدى تأثير هذين المتغيرين على الثبات والاتساق الداخلي لاختبار تحصيلي في مقرر القياس والتقويم.

مشكلة البحث:

الموثوقية في نتائج الاختبارات التحصيلية –خصوصا ذات الأسئلة الموضوعية– مازالت في المحك وتحت الفحص والتحقق، وقد لاحظ الباحث الحالي أثناء عمله كعضو هيئة تدريس لمقررات القياس والتقويم، وجود تباين واضح في نتائج الاختبارات الموضوعية في القسم الواحد وحتى في الاختبارات الموحدة في حالة اختلاف المحاضرين، أي أن طريقة التدريس لها تأثير أيضا، وقد لاحظ الباحث وجود اختلافات بين الممتحنين في عدد البدائل لكل سؤال حيث تمتد بين بديلين إلى خمسة بدائل، كما أن هناك من يضع البديل الصحيح بين البديلين (أ أو ب) أو بين البديلين (ج أو د)، وهناك من يضعها بصورة عشوائية تمتد بين البدائل الأربعة. ومن المتوقع أن كل تلك المتغيرات لها تأثير على ثبات وصدق الاختبار التحصيلي ذي الأسئلة الموضوعية، ومن هنا يصيغ الباحث مشكلة البحث في الأسئلة التالية:

١.هل يختلف ثبات اختبار مقرر القياس والتقويم باختلاف عدد البدائل للسؤال الموضوعي بين نظرية القياس الكلاسيكية والنموذج الإحصائي لبايزي لدى طلاب كلية التربية جامعة الملك خالد؟

٢.هل يختلف الاتساق الداخلي لاختبار مقرر القياس والتقويم باختلاف عدد البدائل للسؤال الموضوعي بين نظرية القياس الكلاسيكية والنموذج الإحصائي لبايزي لدى طلاب كلية التربية جامعة الملك خالد؟

٣.هل يختلف ثبات اختبار مقرر القياس والتقويم باختلاف موقع البديل الصحيح للسؤال الموضوعي بين نظرية القياس الكلاسيكية والنموذج الإحصائي لبايزي لدى طلاب كلية التربية جامعة الملك خالد؟

٤.هل يختلف الاتساق الداخلي لاختبار مقرر القياس والتقويم باختلاف موقع البديل الصحيح للسؤال الموضوعي بين نظرية القياس الكلاسيكية والنموذج الإحصائي لبايزي لدى طلاب كلية التربية جامعة الملك خالد؟

أهداف البحث:

التعرف على مدى تأثير اختلاف عدد بدائل السؤال الموضوعي وموقع البديل الصحيح على الثبات والاتساق الداخلي لاختبار في مقرر القياس والتقويم بين نظرية القياس الكلاسيكية والأنموذج الإحصائي لبايزي لدى طلاب كلية التربية جامعة الملك خالد.

أهمية البحث:

الأهمية النظرية:

قد تفيد نتائج البحث في التعرف على أطر نظرية جديدة في مجال القياس والتقويم، وخصوصا في مجال إعداد أسئلة الاختبارات.

الأهمية التطبيقية:

قد تفيد نتائج البحث في تطوير عملية وضع الاختبارات الموضوعية بحيث تحقق أعلى درجة من الثبات والصدق وبالتالي الموثوقية فيما تفضي إليه من نتائج للطلاب

مصطلحات البحث:

بدائل السؤال:

اختيارات الإجابة التي توضح تحت كل رأس سؤال، وعلى المستجيب اختيار الصحيح منها، والمناسب لرأس السؤال

موقع البديل الصحيح:

هل يكون البديل الصحيح في بداية البدائل أم الوسط أم في نهايتها؟

ثبات الاختبار التحصيلي:

ثبات الاختبار يعني تقارب بين الدرجة الحقيقية والدرجة الفعلية لأداء الطالب على الاختبار، ويستدل عليه من خلال حصول الطلاب على نفس الدرجة تقريبا كلما أعيد تطبيق الاختبار عليهم (سكران، ٢٠١٣)

الاتساق الداخلي للاختبار:

مدى ارتباط السؤال بالدرجة الكلية للبعد الذي ينتمي إليه، وارتباط أبعاد الاختبار بالدرجة الكلية للاختبار

نظرية القياس الكلاسيكية:

هذه النظرية تستند على أنموذج الدرجة الحقيقية للفرد وينص على أن لكل فرد قدر ما من السلوك غير الملاحظ والذي لا يمكن أن يقاس بصورة مباشرة، فتفترض أن درجة الفرد الملاحظة في متغير ما هي مجموع درجاته على جميع مفردات المقياس الذي يقيس هذا المتغير وبالتالي قد يحصل فردان على نفس الدرجة رغم اختلاف المفردات التي أجاب عنها، كما أنه قد يتساوى الفرق بين درجتي فردين مرتفعين في القدرة أو في السمة التي يقيسها المقياس مع الفرق بين درجتي فردين منخفضين في القدرة رغم اختلاف صعوبة المفردات في الحالتين.

النموذج الإحصائي لبايزي:

الاستدلال البايزي وهي أحد طرائق الاستدلال الإحصائي. عند تطبيق الاستدلال البايزي، قد يكون للاحتمالات التي تنطوي عليها مبرهنة بايز (Bayesian interpretation of probability) مدلول مختلف عن المفهوم التكرار

حدود البحث:

الحدود الموضوعية:

أثر اختلاف عدد البدائل وموقع البديل الصحيح على ثبات الاختبار التحصيلي واتساقه الداخلي بين نظرية القياس الكلاسيكية والنموذج الإحصائي لبايزي

الحدود المكانية: كلية التربية جامعة الملك خالد.

الحدود الزمانية: الفصل الدراسي الأول من العام الدراسي ١٤٤٣/ ١٤٤٤هـ

الإطار النظري والدراسات السابقة:

التقويم التربوي والاختبارات التحصيلية:

يمكن استخدام التقييم التربوي لأغراض التشخيص وإضفاء الطابع الرسمي وكذلك تحديد مستوى الفرد، وهو كان ومازال أحد الطرق الرائدة لتحديد مستوى تحصيل الطلاب.

ويرتبط مفهوم التحصيل بالتغيرات في السلوكيات المعرفية التي يمكن تغييرها من خلال التدريس أو التعليم، وترتبط بشكل خاص بثلاثة مفاهيم أساسية: المعرفة والمهارات والقدرات. حيث يتطلب مفهوم المعرفة التذكر أو الفهم لتعلم المبادئ والحقائق. ويشير مفهوم المهارة إلى الحالات التي يمكن ملاحظتها وتتضمن الأداء. وأخيرًا، يعني مفهوم القدرة استخدام المعرفة والمهارات معًا، بينما تطويرها يتطلب مدة زمنية طويلة (Haladyna, 2004). وبالنظر إلى النظام التربوي والتعليمي، فمن الممكن تحديد تحصيل الطلاب للسلوكيات المحددة وفقا لعناصر المدخلات والعمليات والمخرجات. على سبيل المثال، تحديد مستوى استعداد الطلاب يخدم هدف تقييم مدخلات النظام التعليمي. ومن ناحية أخرى، فإن التساؤل عن مستوى تحقيق التعلم في نهاية الفصل الدراسي يخدم هدف تحديد عنصر المخرجات. نظرًا لاعتماد المنهج البنائي بشكل شائع اليوم، فقد أصبح مفهوم التقييم التكويني أكثر أهمية، والهدف الرئيس الآن هو مراقبة تعلم الطلاب. إن تحديد مستوى تحصيل الطلاب مفيدُ أيضًا في مراقبة تعلم الطلاب طوال العملية. ومن الواضح أن اختبارات التحصيل تستخدم بشكل شائع لتحديد مدى نجاح الطلاب.

(Sahin, Yildirim, Ozturk, 2023)

الاختبارات هي أدوات التقويم المستخدمة لتحديد مقدار أو درجة التعلم عدديًا في بيئة صممها المطورون (Haladyna, 2004). ويقسم كرونباخ (Cronbach, 1990) الاختبارات على قسمين اختبارات استجابة نموذجية واختبارات للأداء الأقصى.

تقيس اختبارات الاستجابة النموذجية التركيبات النفسية مثل الموقف، أو الإدراك أو الشخصية أو الدافع أو الاهتمام، في حين أن هذه الاختبارات لا تتضمن عناصر ذات إجابة صحيحة. في مثل هذه الاختبارات، يركز الأفراد بشكل عام على الإبلاغ الذاتي بدلاً من الإجابة على بنود الأسئلة

بشكل صحيح. ومن ناحية أخرى، تشمل اختبارات الأداء الأقصى اختبارات التحصيل واختبارات الذكاء واختبارات الكفاءة. في هذه الاختبارات، من المتوقع أن يظهر الأفراد أعلى مستوى من الأداء. تعتبر واحدة من أقصى اختبارات الأداء والإنجاز

يمكن تطوير الاختبارات من قبل المعلمين، أو يمكن أن تكون اختبارات قياسية. ومن الأهمية بمكان اتباع خطوات تطوير الاختبار التحصيلي للتأكد من أن الاختبارات التحصيلية تقيس نوع المخرج المقصود بما يتوافق مع الغرض منه والتأكد من أنها خالية من الأخطاء قدر الإمكان. في حين أن هناك العديد من الموارد التي تحدد عملية تطوير اختبارات التحصيل بالتفصيل (Sahin, Yildirim, Ozturk, 2023)

تطوير الاختبارات:

عملية تطوير الاختبار تتطلب منهجاً منظماً فيما يتعلق بصحة درجات الاختبار والقرارات المبنية عليه، مما يجعل عملية وضع خطة تطوير الاختبار التحصيلي ذات أولوية في العملية التعليمية

(Lane et al., 2016). وبعبارة أخرى، فإن إنشاء خطة لبناء الاختبار التحصيلي واتباع الخطوات المضمنة في الخطة يوفر لمطوري الاختبار معيارًا بالإضافة إلى دليل على الصلاحية.

تتضمن معايير الاختبارات التربوية والنفسية خمس خطوات يجب القيام بها، تم اتباعها أثناء تطوير الاختبار (Aera & et al., 2014). أول هذه الخطوات هو تحديد خصائص الاختبار. في هذه الخطوة، تتضمن المشكلات التي يجب اتخاذ قرار بشأنها هدف الاختبار، ومحتوى الاختبار، وتنسيق السؤال

(سؤال، مهمة أداء، وما إلى ذلك)، وكيفية تلقي الاستجابات، وكيفية تسجيل النتائج، ومدة الاختبار، وإجراءات إدارة الاختبار (الورقي، المعتمد على الكمبيوتر)، وأنظمة تطوير أسئلة الاختبار. يمكن أن يكون هدف الاختبار هو الاختيار والكشف عن الكفاءة والتصنيف وإجراء الفحص (Turgut, 1992). مجال الاختبار ودرجة الصعوبة، وشكل الأسئلة التي سيتم تضمينها في الاختبار، وطول الاختبار، وما إلى ذلك يمكن أن تتغير وفقًا لهدف الاختبار.

الخطوة الثانية لتطوير الاختبار هي كتابة الأسئلة ومراجعتها. في هذه المرحلة، من الضروري إنشاء مجموعة أسئلة ومراجعتها وإجراء اختبار تجريبي وإجراء التحليل المتعلق بالاختبار التجريبي وإنشاء نموذج تقييم إذا لزم الأمر بسبب تنسيق السؤال. ويعد إنشاء جدول المواصفات في هذه المرحلة ذا أهمية حيوية لصحة المحتوى. يمكن عدّ جدول المواصفات بمثابة نسخة أكثر تحديدًا لخطة الاختبار لأنه عبارة عن مسودة خطة حول ما يجب القيام به حتى يتم إنشاء أنموذج الاختبار. جدول المواصفات، يتضمن الموضوعات ونتائج التعلم والصفات التي سيكتسبها الطلاب والمستوى التصنيفي وعدد الأسئلة لكل نتيجة تعليمية. هناك قرار مهم آخر في الاختبار التجريبي وهو كيفية تحديد العينة. أول شيء يجب فعله هنا هو اختيار عينة ممثلة. في حين أنه من المناسب أن يكون هناك عينة تم اختيارها عشوائيًا تضم عددًا كافيًا من المشاركين حيث يشكل الطلاب بنية متجانسة، إلا أنه يمكن اعتماد طريقة أخذ العينات الطبقية إذا كانت

مجموعة الطلاب غير متجانسة أو إذا كان من الممكن اختيار طلاب من مدارس ذات مستويات نجاح مختلفة. هناك معايير مختلفة في الأدبيات لتحديد عدد الطلاب. وفقا لهذه المعايير المختلفة، يمكن أن يختلف عدد الطلاب بين

١٢٠ و٤٠٠ (Özçelik, 1992)، فمن الضروري أن يكون لديك مجموعة كبيرة من العينة تضم ١٠٠ إلى ٢٠٠ مشارك (Crocker & Algina, 2006) ومن المهم المضي قدمًا عندما تكون العينة أكبر من ٢٠٠ مشارك (Haladyna, 2004).

وفي نهاية الاختبار التجريبي، من الضروري التحقق من صدق وثبات الاختبار. في هذه المرحلة، من الممكن الحصول على إحصائيات الأسئلة والاختبارات. الإحصائيات التي سيتم حسابها في هذه المرحلة تشمل مؤشر تمييز السؤال، مؤشر صعوبة السؤال، تحليل التشتت، الثبات والصدق للاختبار

يعد دمج نماذج الاختبار وتقييمها هو الخطوة الثالثة في تطوير الاختبار. من المهم التأكد من أن العناصر موجودة في كل نموذج اختبار بما يتماشى مع القواعد المناسبة (عدم تضمين القرائن أو عدم تضمين أسئلة مماثلة في نفس نموذج الاختبار، وما إلى ذلك) بعد تلقي آراء الخبراء. وتتطلب الخطوة الرابعة تطوير الإجراءات والمواد اللازمة للإدارة والتسجيل. في هذه المرحلة، يتم إعداد تعليمات لإدارة الاختبار، ويتم تطوير بعض إجراءات الموثوقية لإدارة الاختبار وتسجيله. الخطوة الأخيرة هي مراجعة الاختبار. يتم في هذه المرحلة مراجعة الاختبار من خلال مراجعة تعليمات الاختبار حول كيفية إدارة الاختبار والإجابة على عناصر الاختبار بشكل دوري، وإجراء التغييرات اللازمة إذا كان هناك تغيير في محتوى الاختبار، بما يعكس التغييرات التي تطرأ على المنهج الدراسي على الاختبار، مع تغيير وقت إجراء الاختبار إذا لزم الأمر (Aera & et al., 2014).

توسيع نماذج وملحقات نظرية استجابة المفردة

هناك أنموذجان للاستخدام الواسع لدراسة التغييرات الفردية.

أحدهما يسمى أنموذج منحنى النمو الكامن (Bollen & Curran, 2004) والآخر هو النمذجة متعددة المستويات أو النمذجة الخطية الهرمية (Raudenbush & Bryk, 2002).

ومع ذلك، عندما يتم ملاحظة النتائج في أوقات زمنية متفاوتة بشكل فردي ومتباعدة بشكل غير منتظم، فإن الاستنتاجات من هذين النموذجين التقليديين لدراسة التغييرات الفردية قد تصبح مشكلة بسبب التعديل غير السهل للبناءات البارامترية في تلك النماذج (Geiser et al., 2013). علاوة على ذلك، في الاختبار المحوسب في التعليم، تكون الاستجابات عادةً ثنائية التفرع أو ترتيبية أو اسمية، في حين أن السمات الكامنة المطلوب استنتاجها غالبًا ما تكون مستمرة. هذه تجعل الاستدلال أكثر صعوبة بسبب فقدان المعلومات في التقدير

إجراء متغيرات الاستجابة الأساسية.

نحن نركز على امتداد إطار نموذج IRT الكلاسيكي لنمذجة البيانات ثنائية التفرع الطولية التي تم جمعها في أوقات زمنية غير منتظمة ومتغيرة. كما ذكرنا سابقًا، هناك أسلوبان مستخدمان على نطاق واسع ، أي الأسلوب متعدد الأبعاد والمتعدد المستويات، والمتوفر في الأدبيات الحالية لنماذج IRT لتحليل البيانات الطولية.

أولاً، بالنسبة للأسلوب متعدد الأبعاد، يتم استخدام نموذج IRT متعدد الأبعاد لتمثيل تغيير القدرة كقدرة أولية وواحد أو أكثر من المعدلات المحتملة في الاختبارات أحادية البعد أو متعددة الأبعاد Cho et al., (2013). **ومع ذلك، هذا يسمح باختلاف بسيط في الفقرات في مناسبات مختلفة وغالبًا ما يتطلب من الأفراد إجراء نفس الاختبارات. تمنعنا هذه العيوب من توسيع أساليب لتحليل سلسلة زمنية لبيانات الاختبار المحوسبة.**

ثانيًا ، بالنسبة للأسلوب متعدد المستويات، غالبًا ما يُفترض أن المستوى الأول يتبع نموذج IRT الكلاسيكي، بينما في المستوى الأعلى، هناك فكرتان شائعتان لنمذجة النمو. تتمثل إحدى الأفكار في افتراض نمو سمة كامنة لتكون دالة حدودية للوقت، مثل الانحدار الخطي أو متعدد الحدود لمتغير الوقت مع معاملات ثابتة أو عشوائية. هذه الفكرة هي التباين في نمذجة منحنى النمو الكامن في تحليل البيانات الطولية الثنائية / الفئوية (Hsieh et al., 2013**). فكرة أخرى هي استخدام نماذج سلسلة ماركوف، حيث يُفترض أن التغييرات في سمة كامنة بمرور الوقت تعتمد على قيمتها أو حالتها السابقة**
(Kim & Camilli, 2014)

ومع ذلك، هناك العديد من الحالات التي لا تكون فيها أي من هاتين الفكرتين مناسبة بما يكفي لوصف النمو (Bollen & Curran, 2004 **). أحد هذه الحالات هو الاختبار المحوسب، خاصةً عندما تكون الفواصل الزمنية بين الاختبارات متباعدة بشكل غير متساوٍ بين الأفراد وكذلك داخل الأفراد.**

لمواجهة التحديات في الاختبار المحوسب لنمذجة نمو الصفات الكامنة، طور Wang et al. **(٢٠١٣)، نموذجًا ديناميكيًا من خلال الجمع بين أفكار المعادلات البارامترية للوقت بالإضافة إلى نماذج سلسلة ماركوف لوصف النمو. قاموا بدمج نماذج IRT في فئة جديدة من نماذج تحليل البيانات الطولية الموجودة في التغايرات الفردية والمتباعدة بشكل غير منتظم خلال أوقات زمنية. على وجه الخصوص قد تكون المعلمة المتغيرة بشكل فردي والتي يمكن تسميتها "عامل النمو" في المستوى الثاني من نموذج IRT الديناميكي المقترح، مرتبطة بخصائص كل فرد، مثل الجنس، والدرجة ، وما إلى ذلك.**

نمذجة Bayesian وتوسيع نموذج IRT الديناميكي.

يعد تحديد المتغير في الطبقة العليا مشكلة طبيعية ومهمة في هذا النموذج. هناك العديد من طرق اختيار متغير بايزي متوفرة في الأدبيات الإحصائية، مثل معيار معلومات الانحراف (Spiegelhalter et (DIC)

(al., 2002، و Bayesian lasso، وعامل بايزي للتقارب. تمت مناقشة العديد من هذه المعايير مبدئيًا ضمن إطار نموذجي خطي، ويمكن تطبيقها على النماذج الهرمية كنموذج IRT الديناميكي. ومع ذلك، لا يزال من غير الواضح ما إذا كانت بعض خصائص اختيار النموذج المفضلة مثل اتساق اختيار النموذج ستظل قائمة في النماذج الهرمية.

ناقش Fernandez et al. (2001) and Liang et al. بعض خصائص اختيار النماذج المفضلة لعامل بايز عندما تم تحديد g-Prioris لمعاملات الانحدار ضمن إطار النموذج الخطي. على وجه الخصوص، افترض Liang et al. (2008) أنه في حالة حساب Inv-Gamma(1/2, n/2) مسبقًا على g ، فإن أسلوب عامل Bayes له العديد من الخصائص المفضلة، بما في ذلك اتساق اختيار النموذج اللاحق. Li and Clyde (2018) و Wu et al. (2016) ناقش خصائص اختيار النموذج مع مزيج من g priors ضمن إطار النموذج الخطي المعمم. هذا يحفزنا لاعتماد أسلوب عامل Bayes للاختيار المتغير في الطبقة المضافة تحت نموذج IRT الديناميكي.

ومع ذلك، نظراً للبناء المعقد لنموذج IRT الديناميكي والأبعاد العالية لمساحة المعلمة، لا توجد أشكال تحليلية متاحة لعوامل Bayes لنماذج IRT الديناميكية (بينما توجد أشكال تحليلية متاحة لنموذج الانحدار الخطي الذي تمت مناقشته في Liang et al. (2008) عندما تكون g ثابتة).

بالاعتماد على عمل Chen (٢٠٠٥)، قام Liu (2019) بتطوير أسلوب Monte Carlo لحساب عوامل Bayes وفقًا لنموذج IRT الديناميكي باستخدام ناتج MCMC واحد، وهو أسلوب ممكن إذا كان عدد التغايرات Covariates المشتركة معتدلاً إلى حد ما. على سبيل المثال، ٢٠ تغايرا ستؤدي إلى ٢٢٠ = ١٠٤٨٥٧٦ نموذجًا ممكنًا، حتى لو لم تكن هناك شروط تفاعل مأخوذة في الحسبان.

ومع ذلك، فإن افتراض وجود علاقة وظيفية معينة للنمو بشكل عام يمكن أن يكون مقيدًا ويصعب عادةً تبريره. بدلاً من ذلك، يمكن أن يكون الأنموذج اللامعلمي أكثر مرونة لوصف التغييرات في السمات الكامنة وتجنب أخطاء التحديد الخاطئ للأنموذج.

بناءً على النتائج الموضحة في Wang et al. (2013)، وجدنا أن مسار قدرة القراءة للموضوع غالبًا ما ينمو بسرعة أكبر في الفترة الأولية، ولكنه يتباطأ عندما يقترب من النضج. بشكل عام، تُظهر القدرة اتجاهًا متزايدًا ولكن غالبًا ما يكون لها منطقة مسطحة في النهاية. يتوافق اكتشاف الشكل مثل مسار النمو مع المعتقدات والخبرات السابقة من الممارسين.

في العديد من التطبيقات العلمية، يمكن اقتراح افتراضات حول شكل المسار، مثل الرتابة أو التحدب أو التقعر، قبل التحليل، للمساعدة في عملية النمذجة وتعزيز القابلية للتفسير (Rodrigues et al., 2018).

تلعب نماذج نظرية استجابة المفردة (IRT) دورًا مهمًا في الدراسات السيكومترية لتصميم الاختبارات وتحليلها. تأخذ نماذج IRT في الحسبان العلاقة بين صحة المفردات والقدرة الكامنة للفرد وصعوبة كل عنصر وعوامل محتملة أخرى مثل التخمين.

قام Liu (2019) بتطوير طرائق النمذجة Bayesian وتقنيات اختيار النماذج ضمن إطار نموذج IRT. لمقارنة أنموذج Bayesian ، ويعد عامل Bayes أداة مستخدمة على نطاق واسع ، والتي تتطلب حساب الاحتمالات الهامشية. بالنسبة للنماذج المعقدة مثل نماذج IRT ، فإن الاحتمالات الهامشية غير متوفرة تحليليًا. هناك مجموعة متنوعة من طرائق مونت كارلو لتقدير أو حساب الاحتمالات الهامشية، على الرغم من أن بعضها قد لا يكون مجديًا لنماذج IRT بسبب الأبعاد العالية لمساحة المعلمة. نقوم بمراجعة العديد من طرائق مونت كارلو المختلفة لحساب الاحتمال الهامشي في إطار نماذج IRT الكلاسيكية ، وقد قام بتطوير "أفضل" تنفيذ لهذه الأساليب لنماذج IRT ، وتطبيق هذه الطرق على مجموعة بيانات حقيقية للمقارنة بين نموذج IRT الكلاسيكي ذي المعلمة الواحدة ونموذج IRT ثنائي المعلمات.

مع زيادة توافر الاختبارات المحوسبة ، غالبًا ما يتم جمع الملاحظات في نقاط زمنية غير منتظمة ومتغيرة. نعتمد نموذج IRT الديناميكي القائم على نموذج IRT ذي المعلمة الواحدة لاستيعاب بنية البيانات هذه. تم بناء طبقة chical hierar على نموذج IRT الديناميكي لالتقاط العلاقة بين "عامل النمو" وخصائص الأفراد. نستخدم عامل بايز لإجراء اختيار متغير على المتغيرات المشتركة المرتبطة بالنمو، وتطوير نهج مونت كارلو لحساب عوامل بايز لجميع أزواج النماذج باستخدام سلسلة مونت كارلو (MCMC) لسلسلة sin gle Markov. (Liu, 2019)

وقد طور علماء النفس والتربية نظريات ونماذج قياس مختلفة لشرح السمة الكامنة وراء استجابة الأفراد لفقرة ما. أحد الأمثلة الأساسية والتي لا تزال معتمدة على نطاق واسع هي نظرية الاختبار الكلاسيكي (CTT) ، والتي تهدف إلى شرح النتيجة التي تم الحصول عليها من اختبار بناءً على درجات الصواب والخطأ. تعتمد CTT على افتراضات سهلة ولكنها ضعيفة. على الرغم من أن CTT قد خدمت المجال لسنوات عديدة، إلا أن معلمات المفردة واعتماد القدرة على المجموعة، والصعوبات في مقارنة الأفراد، ووجود عتبات أقل لـــتقديرات الصدق (Hambleton; Swaminathan & Rogers, 1991) قادت الباحثين للبحث عن نظريات قياس مختلفة. علاوة على ذلك، فإن المناقشات حول الدرجات الحقيقية والفعلية المستخدمة في اختبار CTT ليس لها نفس معنى درجات القدرة، والأدبيات المتعلقة بضرورة أن تكون درجات القدرة مستقلة عن الاختبار وفقرات الاختبار، بما يتعارض مع طبيعة اختبار CTT (Hambleton & Jones, 1993)

لقد تشكلت أسس نظرية جديدة بعد CTT، تعد نظرية استجابة المفردة (IRT) واحدة من أهم النظريات، والتي لها مكان في تاريخ القياس النفسي. IRT هي نظرية اختبار قوية تشرح السمات الكامنة للنماذج

الاحتمالية الكامنة وراء استجابة الفرد لفقرة ما من خلال افتراضات أقوى بكثير مقارنة بـــ CTT (Bobcock, 2009).

في الأدبيات الخاصة بـــ IRT، يُشار إلى الافتراض القائل بوجود ميزة كامنة واحدة في الغالب وراء استجابة المختبرين لمجموعة من الفقرات في الاختبار على أنها أحادية البعد. مما لا شك فيه أن هناك عوامل أخرى، مثل الوظيفة المعرفية والإثارة والتوتر، تؤثر أيضًا على أداء الاختبار للأفراد. لذلك، يعدّ العامل السائد المسؤول عن أداء الاختبار للأفراد هو السمة التي تم قياسها بواسطة الاختبار (Hambleton, 1989) تُعرف النماذج الرياضية التي تشرح أداء الأفراد في فقرات الاختبار باستخدام سمات كامنة متعددة بنماذج IRT في الأدبيات ذات الصلة.

الاستقلال المحلي هو افتراض أن احتمال استجابة الأفراد لفقرة اختبار بطريقة معينة مستقل عن احتمالية استجابتهم للمفردات الأخرى في نفس الاختبار، بالنظر إلى أن السمة الكامنة التي يقيسها الاختبار تظل ثابتة. على الرغم من أن الأدبيات تحتوي على حسابات رياضية محددة لتلبية هذا الافتراض، إلا أن الاستقلال المحلي يعتبر عادةً موازيًا لافتراض أحادية البعد. هذا لأنه إذا كان احتمال استجابة الأفراد لكل مفردة اختبار مستقلًا عن بعضهم البعض، فإن العامل الوحيد الذي يفسر احتمالية الاستجابة هو السمة الكامنة المقاسة باختبار عامل واحد (Hambleton, 1989)

الافتراض الأساسي الأخير لـــ IRT هو الرتابة، والذي يتشابك مع منحنى خاصية المفردة. على عكس CTT، في IRT، تكون قدرة الأفراد واحتمالية الاستجابة بشكل صحيح للمفردات في الاختبار منحنية الخطوط، ويتم عرض هذه العلاقة المنحنية باستخدام منحنى خاصية المفردة (ICC).

ذكر الباحثون الذين عرّفوا IRT على أنه الانحدار غير الخطي لأداء المفردة على السمة التي تم قياسها بواسطة الاختبار، أن IRT افترض أنه مع زيادة قدرة الفرد، زاد احتمال استجابته بشكل صحيح لفقرات الاختبار (Crocker & Algina, 1986)

نموذج بايزي كأحد نماذج نظرية الاستجابة للمفردات أحادية البعد

في السنوات الأخيرة، تركز الاهتمام في مجتمع التعليم على الحاجة إلى البحث القائم على الأدلة، ولا سيما السياسات التعليمية والتدخلات التي تعتمد على "البحث العلمي". وقد أدى التركيز على البحث العلمي في التعليم إلى زيادة في الدراسات المصممة لتوفير ضمانات قوية للادعاءات السببية. وقد أجريت هذه الدراسات بالاعتماد بشكل كامل تقريبًا على الإجراءات التحليلية المتجذرة في نموذج الإحصاء المتكرر (الكلاسيكي).

ومع ذلك، فقد تم إحراز تقدم كبير في العقد الماضي في مجال الاستدلال الإحصائي البايزي (Bayesian) ، ويرجع ذلك في الغالب إلى التطورات الحسابية والبرمجيات المتاحة بسهولة (Richardson , 1996

Gilks, Spieghallter &). مع هذه التطورات استخدمت تطبيقات مهمة لطرائق بايزي للمشكلات في العلوم الاجتماعية والسلوكية.

ومع ذلك ، يمكن العثور على تطبيقات قليلة لأساليب بايزي للمشكلات التعليمية ، باستثناء الأساليب الإحصائية البايزية لنماذج نظرية استجابة المفردة في التعليم. يقترح (Fox & Glas, 2003). أن الأساليب البايزية تقدم أنموذجًا تحليليًا بديلاً يمكن أن يدعم ويعزز البحث العلمي في العلوم التربوية.

لا يمكن المبالغة في أهمية فحص النمذجة الإحصائية في العلوم التربوية من منظور بايزي. لفترة طويلة جدًا، استندت الأساليب الإحصائية المطبقة على المشكلات التعليمية إلى اختبار الفرضيات الإحصائية المتكررة، التي طورها فيشر في الأصل (١٩٢٥/١٩٤١) ، ثم لاحقًا بواسطة نيمان وبيرسون (١٩٢٨). تم انتقاد هذه الأساليب لكونها غير متماسكة منطقيًا وأن نهج Neyman-Pearson لاختبار الفرضيات على وجه الخصوص قد يكون قد أحدث ضررًا كبيرًا للتقدم في العلوم الاجتماعية والسلوكية. Steiger, & .(Harlow, Mulaik, 1997)

في حالة النماذج الإحصائية، لا يعترف المنظور المتكرر (الكلاسيكي)، بأن النماذج نفسها مأخوذة من عالم أكبر من النماذج الممكنة، وليس أي منها صحيحًا بكل معنى الكلمة وبالتالي فإن الهدف من الاستدلال الإحصائي من منظور بايزي هو التأكد من الأنموذج الذي تفضله البيانات ((Kass & Raftery, 1995).

أخيرًا، على الرغم من أنه ليس نقدًا للإحصاءات المتكررة في حد ذاتها، فليس من غير المألوف قراءة الأوراق التي تحتوي على أوصاف على سبيل المثال : فترات الثقة ، والتي لم يتم تفسيرها بشكل صحيح من وجهة نظر متكررة، ولكن يتم تفسيرها بشكل صحيح من وجهة نظر بايزي. يشير هذا إلى أن منظور بايز يتوافق مع مفاهيم الفطرة السليمة لاختبار الفرضيات.

في مجال التربية وعلم النفس توجد معادلات رياضية تشرح سلوك واستجابات الأفراد من خلال ربط المتغيرات المستقلة والمتغيرات التابعة بناءً على مفهوم النموذج. في IRT يمكن تصنيف النماذج وفقًا لعدد المعلمات المضمنة والأبعاد ونظام التسجيل (McDonald, 1999) ، وكذلك اعتمادًا على نوع المعادلة الرياضية.

في القياس النفسي، تلعب نماذج نظرية الاستجابة للمفردات (IRT) دورًا حاسمًا في تصميم الاختبارات وتحليلها، وتربط نماذج النظرية الكلاسيكية لاستجابة المفردة بين صحة الإجابة (متغير عشوائي ثنائي التفرع) لكل مفردة بالقدرة الكامنة للفرد وصعوبة المفردة، والمعلمات الأخرى المحتملة بأنواع مختلفة من وظائف الارتباط ويشير نموذج Rasch المستخدم على نطاق واسع (Rasch, 1960) إلى نموذج IRT ذي المعلمة الواحدة مع الارتباط اللوجستي (نموذج PL1). يضيف نموذج IRT اللوجستي ذو المعلمتين (نموذج PL2) معلمة التمييز إلى نموذج PL1 للسماح بقياس مختلف لكل مفردة، وتشمل الإضافات والاختلافات الأخرى لنماذج IRT نموذج IRT ذي المعلمات الثلاثة، والذي يضيف معلمة للتخمين مقارنة

بنموذج IRT ذي المعلمتين (Harris, 1989) ؛ نماذج IRT متعددة المستويات (Fox, 2005) ، والتي تقدم طبقات إضافية لمعامل القدرة ؛ ونموذج IRT الديناميكي (Wang et al., 2013) ، القادر على التعامل مع الملاحظات غير المنتظمة والمتغيرة بمرور الوقت وتقدير المسار الكامن للقدرة خلال فترة زمنية.

نظرًا لتعقيد نماذج IRT، أصبحت طرق Bayesian أكثر شيوعًا للتحليل وتقدير المعلمات لنماذج IRT (Karabatsos, 2016). عامل بايز (Kass and Raftery, 1995) هو معيار مقارنة نموذج بايزي شائع الاستخدام، والذي يتطلب التوزيعات المسبقة المناسبة وحساب الاحتمالات الهامشية.

الدراسات السابقة:

أولاً: دراسات تناولت المقارنة بين النظرية التقليدية والنظرية الحديثة في معالجة البيانات

سعت دراسة بابان (2014)، إلى المقارنة في الخصائص القياسية لمقياس التقويم النمائي للذكاءات المتعددة (ميداس) (M.I.D.A.S)، (Multiple Intelligences Developmental Assessment Scales) وفقاً لنظرية القياس الحديثة والكلاسيكية، وقد اتبع الباحث خطوات علمية تتعلق بإجراءات تحليل المقياس وفق نظرية القياس الكلاسيكية(التقليدية) ، فقام بترجمة المقياس من اللغة الإنكليزية إلى اللغة العربية، ثم اجرى ترجمته من العربية الى الإنجليزية ، ثم تم ارسال المقياس الى خبراء في اللغة الإنجليزية للتأكد من الترجمة ، وبعد التأكد من اكتمال إجراءات صدق الترجمة للمقياس، استعان الباحث بـــ ١٨ مختصاً وخبيراً في القياس النفسي والعلوم النفسية لهذا الغرض، وتم الإبقاء على كل الفقرات. وللتحقُق من صلاحيتها منطقياً تم استخراج الصدق الظاهري للمقياس الذي تكون من ١١٩ فقرة تقيس (٨) أنواع من الذكاءات التي اقترحها هاوارد كاردنر في نظريته للذكاءات المتعددة، وتمت صياغة هذه الفقرات بأسلوب العبارات التقريرية تكونت من خمسة بدائل للإجابة ،وتم الاخذ بملاحظات الخبراء حيث تم تعديل بعض الفقرات، وتمت إعادة صياغة بعضها الآخر، وتم الاتفاق على إبقاء الفقرات جميعها من قبل المحكمين بنسبة ٨٣%، وهي نسبة مقبولة للاحتفاظ بالفقرات ،وبذلك تم التأكد من الصدق الظاهري للمقياس ، ولأجل إعداد الصيغة النهائية للمقياس ،تم اعداد تعليمات واضحة للطلبة لكيفية الاجابة، باستخدام ورقة الإجابة المنفصلة، وتم تطبيق المقياس على عينة مكونة من ١٠٠ طالب وطالبة تم اختيارهم بطريقة عشوائية من طلبة جامعة بغداد، وقد اتضح من هذا التطبيق أن تعليمات المقياس وفقراته كانت واضحة ومفهومة، وكان زمن الإجابة المستغرق على المقياس مناسب، ولغرض التحليل الإحصائي لفقرات المقياس واستخراج الخصائص القياسية لها، فقد تم تطبيق المقياس على عينة مكونة من ١٠٠٠ طالب وطالبة تم اختيارهم بالطريقة العشوائية العنقودية من جامعات (بغداد، المستنصرية، التكنولوجية، النهرين، العراقية). أما ما يتعلق بالخصائص القياسية للمقياس وفق نظرية القياس الكلاسيكية(التقليدية)، فقد تم التحقق من الصدق (الظاهري ،وصدق البناء بثلاثة مؤشرات هي: القوة التميزية للمفردات

،وارتباط المفردة ،بالدرجة الكلية، والصدق العاملي)، وتم حساب الثبات للمقياس بطريقة إعادة الاختبار، وكذلك باستخدام معادلة ألفا كرونباخ. أما التحليل الإحصائي وفق نظرية القياس الحديثة ، فقد تم من خلال اعتماد الباحث على أنموذج راش المتعدد الاستجابة كأحد نماذج نظرية السمات الكامنة في تحليل فقرات مقياس ميداس، وكما تم حسابه ببرنامج لغة الأوامر.ولتحقيق افتراضات الأنموذج، قام الباحث بمايلي: للتحقق من افتراض أحادية البعد، تم إخضاع المكونات الثمانية للتحليل العاملي بطريقة المكونات الأساسية، و تم الحصول على عامل واحد ذي معنى لكل مكون، واعتمد تفسير هذا العامل على الحدود الدنيا لـــ"جتمان" الذي يعد العامل دالاً إحصائياً عندما يكون الجذر الكامن الذي يمكن تفسيره يساوي أو يزيد عن ١، وقد تم اعتماد معادلة الخطأ المعياري (بورت- بانكس) على أنها نسبة تشبع فقرات المقياس بالعامل العام. وقد تم استبعاد فقرة من النوع الثاني وفقرة من النوع الرابع للذكاء ، لأن تشبع الفقرات كان أقل من (٠,١١٢، ٠,١٤٧) على التوالي عند مستوى دلالة ٠,٠١. فضلاًعن أن مطابقة الفقرات لنموذج راش المتعدد الإستجابة يعد دليلاً على أن الفقرات تقيس سمة أحادية البعد. استناداً إلى قيمة مربع كاي بمستوى دلالة ٠,٠٥، وتم حسابها من البرنامج ICL، وبناء على ذلك لم تم الاحتفاظ بكل الفقرات ولم يتم استبعاد أي منها. وفيما يتعلق باستقلالية القياس بما يحقق موضوعية القياس في نموذج راش المتعدد الإستجابة وعليه لم يستبعد أي فقرة. وبعد تحليل البيانات وفق نظرية القياس الكلاسيكية، والنظرية الحديثة في القياس (نظرية السمات الكامنة)، توصل الباحث إلى مجموعة من النتائج التالية: ١) إنّ كلا من النظريتين تتميزان بصدق البناء. ٢) ان هناك فروقاً ذات دلالة إحصائية عند مستوى ٠,٠٥.لصالح النظرية الحديثة في القياس عند المقارنة في قيمة معامل الثبات المحسوب وفق النظريتين. ٣) وعند المقارنة في قيمة الخطأ المعياري بين النظريتين تبين أن قيمة الخطأ المعياري للمقياس الذي بُني وفق نظرية القياس الكلاسيكي(التقليدي) اكبر من قيمة الخطأ المعياري للمقياس الذي حُلّل وفق النظرية الحديثة في القياس . ٤) عدد الفقرات لكلا الأسلوبين متساوٍ بصورته النهائية للمقياس.

دراسة Nishio, Akasaka, Sakamoto & Togashi, (2020) والتي هدفت إلى التحقق من صحة نموذج بايزي لنظرية استجابة المفردة (IRT). تم استخدام IRT لتقييم طريقة جديدة (الطرح الزمني ، TS) في دراسات المراقبة لأخصائي الأشعة، مقارنة بالطريقة التقليدية (التصوير المقطعي المحوسب).

المواد والطرائق: من الأوراق المنشورة سابقًا، حصلنا على مجموعتي بيانات لدراسات المراقبة السريرية لأخصائي الأشعة. استخدمت تلك الدراسات نموذجًا متعدد القراءة ومتعدد الحالات لتقييم قدرات الكشف لدى أخصائي الأشعة، وذلك في المقام الأول لتحديد ما إذا كان تحليل متلازمة توريت يمكن أن يعزز قابلية اكتشاف النقائل العظمية أو احتشاء الدماغ. طبقنا IRT على مجموعات بيانات هذه الدراسات باستخدام برنامج ستان. قبل تطبيق IRT، تم تسجيل ردود اختصاصي الأشعة كثنائيات لكل حالة

(١ = صحيح ، ٠ = غير صحيح). تم تقييم تأثير TS على قابلية الاكتشاف باستخدام نموذج IRT الخاص بناء وحساب الفاصل الزمني الموثوق به ٩٥٪ للتأثير.

النتائج: كان المتوسط والوسيط وفترة الثقة ٩٥٪ لتأثير TS هو ٠,٩١٣ و ٠,٨٨٥ و[٠,٢٤٣-١,٧٤٥] للكشف عن النقائل العظمية و ٢,٥٢٤ و ٢,٥٠ و[١,٨٢٧-٣,٣١٠] للكشف عن احتشاء الدماغ. لكلتا دراستي الكشف، لم تتضمن الفواصل الزمنية الموثوقة ٩٥٪ لتأثير متلازمة توريت صفرًا ، مما يشير إلى أن متلازمة توريت تحسن بشكل كبير من القدرة التشخيصية.

الخلاصة: كانت الأحكام المبنية على نتائج الدراسة الحالية متوافقة مع الدراستين السابقتين. أظهرت نتائج دراستنا أن النموذج الإحصائي البايزي لـــ IRT يمكن أن يحكم على فائدة طريقة جديدة.

ثانيا: دراسات تناولت أثر بدائل الاستجابة على الخصائص السيكومترية للاختبار

دراسة القحطاني (2021)، والتي هدفت إلى التعرف على تأثير مستوى الاكتئاب وعدد بدائل الاستجابة على الخصائص السيكومترية لمقياس الوحدة، وذلك لتحديد تأثير التفاعل بين مستويات للاكتئاب وعدد بدائل الاستجابة على الخصائص السيكومترية لمقياس الوحدة. تكونت العينة من (٩٠) طالبة من جامعة الملك سعود فرع الدرعية، من القسمين العلمي والأدبي، استبعد منهم ٤٢ طالبة. وكانت العينة النهائية (٤٨) طالباً تم توزيعهم على مجموعتين بالتساوي، استخدم مقياس بيك للاكتئاب، ومقياس الشعور بالوحدة لراسل ذو البدائل من (خمسة إلى أربعة)، ومقياس الوحدة لإبراهيم قشقوش. أظهرت النتائج أن معامل الثبات بالتجزئة النصفية في النسخة الرابعة أعلى من النسخة الخامسة في المجموعة غير المكتئبة عكس المجموعة الأخرى. صدقًا، تعدّ معاملات ارتباط بيرسون ذات دلالة إحصائية، والتي تختبر الصدق المتعلق بالمعيار بين نسختين من مقياس الوحدة لراسل، ومقياس الوحدة لإبراهيم قشقوش، مما يشير إلى أن أعلى معامل ارتباط تم تحقيقه في النسخة الخامسة من المجموعة بدون انخفاض ، بينما تم تحقيق أعلى معامل ارتباط بواسطة – الإصدار الرابع من المجموعة الأخرى.

دراسة حميدة (2020)، من أهداف هذه الدراسة التعرف على تأثير عدد بدائل الاستجابة على كلٍ من معاملات الثبات والصدق لأداة القياس، كما هدف إلى معرفة الاختلاف في بدائل الاستجابة لمقياس القلق الرقمي، والتعرف على الاختلاف في الخصائص السيكومترية لأداة القياس في ضوء تباين عدد بدائل الاستجابة تبعا لمتغيرات النوع، والمرحلة الدراسية، سارت الدراسة وفقا للمنهج الوصفي التحليلي، تم اختيار عينة (١٥٠) طالباً (٧٥ ذكور، ٧٥ إناث، تتروح أعمارهم بين ١٠- ٣٠ عاماً، ٥٠ من المرحلة الأساسية، ٥٠ من المرحلة الثانوية، ٥٠ من المرحلة الجامعية). تم عمل نموذجين من مقياس القلق الرقمي في ضوء عدد البدائل (ثلاثي وخماسي)، وبعد تحليل البيانات، تم التوصل إلى: وجود تباين في الخصائص السيكومترية لأداة القياس في تبعا لاختلاف عدد بدائل الاستجابة فب اتجاه البدائل الخمسة، ويوجد تباين تبعا لبدائل الاستجابة لمقياس القلق الرقمي لصالح البديل الخماسي، ويوجد اختلاف في

خصائص المقياس السيكومترية بناء على اختلاف عدد البدائل تبعاً لمتغير النوع لصالح الذكور، وتبعا لمتغير المرحلة الدراسية لصالح المرحلة الجامعية ولصالح المرحلة الثانوية لكل من المقياسين ثلاثي وخماسي البدائل.

دراسة النصراويين (2019)، سعت الدراسة إلى تعرف أثر عدد بدائل اختبار متعدد الاختيارات على دالة المعلومات والخطأ المعياري وفقا لنظرية السمات الكامنة النموذج الثلاثي، تم بناء اختبار تحصيلي من نوع الاختيار من متعدّد في مقرر الرياضيات لطلبة الصف العاشر الاساسي في المدارس الحكومية في عمّان، شمل الاختبار (٣٤) سؤالاً، وعمل (٣) نماذج منه(٥، ٤، ٣ بدائل). شملت العينة (١٥٣٠) طالبًا من طلبة الصف العاشر للعام الدراسي (٢٠١٧-٢٠١٨) في محافظة العاصمة عمان، وبعد تحليل البيانات بالبرامج الإحصائية المناسبة، أظهرت نتائج الدراسة عدم وجود فروق ذات دلالة إحصائية بين المتوسطات الحسابية للخطأ المعياري في تقدير دالة المعلومات يُعزى لمتغير عدد البدائل للمفردة .

دراسة عودة (2016) والتي كان هدفها تقدير معالم المفردة والقدرة بالاستناد إلى نظرية السمات الكامنة في ضوء طول الاختبار وعدد البدائل، ومن أجل ذلك تمّ بناء اختبار تحصيلي في الرياضيات للصف العاشر الأساسي مكون من ١٠٠ سؤال (اختيار من متعدد) بصورته الأساسية، واشتقاق تسعة اختبارات فرعية بأطوال مختلفة بدلالة عدد الفقرات، وبعدد بدائل مختلفة للمفردة ، وبعد التأكد من الخصائص السيكومترية للاختبار الأصلي والاختبارات الفرعية. جرى تطبيق الاختبارات على عينة ١٧٦٣ طالبًا وطالبة في الصف العاشر الأساسي، منهم ٨٦٧ ذكور، و٨٩٦ إناث. وباستخدام البرامج المناسبة، أشارت النتائج الى أن المتوسطات الحسابية لمعالم القدرة المقدرة وفق الطرق التي تستند إلى نظرية الاستجابة للمفردة كان معظمها قريبًا من الصفر. كما أشارت النتائج إلى عدم وجود اختلاف في تقدير معالم القدرة تعزى لمتغيرات طول الاختبار وعدد البدائل، وطريقة التقدير والتفاعلات بينها

دراسة أبوجراد (2015)، من بين أهداف الدراسة، التحقق من أثر اختلاف البدائل في أسئلة الاختبار الموضوعي في دقة تقدير صعوبة الفقرات وثبات الاختبار. تم بناء اختبار تحصيلي موضوعي (٢٥ سؤالاً)، في الرياضيات لعينة من طلبة الصف الثامن (٢٧٨٠) طالبا، تم عمل (٣) نماذج حسب عدد البدائل، تمّ تحليل استجابات الطلاب لجميع نماذج الاختبار الثلاثة وفق نموذج التقدير الجزئي، ومن أهم النتائج: وجود فروق دالة إحصائيا في الأوساط الحسابية للأخطاء المعيارية الخاصة بصعوبة الفقرات بين النموذج الثاني وكلّ من النموذجين الأول والثالث، لصالح الأنموذج الثاني (عدد البدائل الصحيحة=٢). كما بيّنت النتائج أنّ هناك فروقا دالة إحصائيا في الأوساط الحسابية الخاصة بقدرات الأفراد بين النموذج الثاني وكل من النموذجين الأول والثالث، لصالح النموذج الثاني، وكشفت النتائج أيضا عن عدم وجود فروق دالة إحصائيا بين معاملات الثبات للنماذج الثلاثة.

دراسة الرغيلات (2014)، كان الهدف الرئيس للدراسة هو معرفة أثر عدد البدائل على دقة تقديرات مؤشرات القدرة للأفراد ودالة المعلومات للاختبار، تم تطبيق صور الاختبار الثلاث والمتباينة في عدد البدائل. تكونت عينة الدراسة من ٣٥٠ طالبا من طلبة الصف الثاني الثانوي العلمي في الأردن، طبّق عليهم اختبار تحصيلي في الكيمياء، وكانت نماذج البدائل (٣، ٤، ٥). استخدمت الدراسة الحالية نظرية الاستجابة للمفردة في تحليل نتائج الاستجابة على نماذج الاختبار الثلاثة، والتي طرحت العديد من الحلول لمشاكل تتعلق في بناء الاختبارات وتطويرها، حيث تتميّز النظرية الحديثة بأن تقديرات خصائص المفردات متحررة من قدرات عينة الأفراد، وتقديرات قدرات الأفراد متحررة من خصائص المفردات. كما هدفت إلى تعرّف أثر عدد البدائل في اختبار الاختيار من متعدد على تقديرات مؤشرات القدرة للأفراد وفقا لنظرية القياس الحديثة ، ثم تحديد أي من صور الاختبار الثلاث أكثرها كفاءة، وذلك بعد التحقق من افتراضات النظرية الحديثة في القياس.

دراسة الشريفين ،وطعامنة (2009)، وكان أهم أهداف الدراسة التعرف على أثر عدد البدائل في الاختبار الموضوعي في تقديرات القدرة للأفراد، وتقديرات الصعوبة للمفردات، واقتران المعلومات للمفردات والاختبار، استخدم اختبار تحصيلي موضوعي في ممادة الرياضيات (٤٠ سؤالاً) اختيار من متعدد، تم تطبيقه على (٦٠٠) طالب وطالبة من طلبة الصف العاشر الأساسي، تم تحليل البيانات ومن أهم النتائج عدم وجود فروق ذات دلالة إحصائية (α=0.05) بين متوسطات الأخطاء المعيارية في تقديرات معالم الصعوبة للمفردات، وعدم وجود تأثير لاختلاف عدد البدائل على ثبات الاختبار. كما أظهرت وجود فروق ذات دلالة إحصائية (α=0.05) بين متوسطات الأخطاء المعيارية في تقديرات معالم القدرة للأفراد، حيث كانت التقديرات أقل دقّة لقدرات أفراد الصورة الثالثة للاختبار منها لقدرات أفراد الصورة الاولى له، وكانت التقديرات لقدرات أفراد الصورة الثانية أكثر دقة منها للصورة الثالثة، بينما لم تختلف دقة تقدير معالم القدرة للأفراد لكل من الصورتين الثانية والثالثة عند مستويات القدرة المنخفضة، بينما أعطت الصورة الثالثة للاختبار كمية أكبر من المعلومات عند مستويات القدرة المتوسطة والمرتفعة من الصورتين الأخريين. كما بينت النتائج وجود فروق دالة إحصائياً بين معاملات صدق المحك ولصالح الصورة الثانية للاختبار.

ثالثا: دراسات تناولت أثر موقع البديل الصحيح على الخصائص السيكومترية للاختبار

دراسة المرواني وسليمان (2019)، هدف الباحثان إلى الكشف عن أثر موقع البديل الصحيح في اختبار موضوعي فقراته من نوع الاختيار من متعدد (٤) بدائل على الصعوبة والتمييز للمفردات، وعلى تقديرات معاملات الثبات التجريبي ومعاملات ثبات الأفراد، وفق الأنموذج ثنائي المعلمة لنظرية الاستجابة للمفردة . وللوصول على نتائج البحث تم إعداد اختبار تحصيلي موضوعي لمادة العلوم (٣١) سؤال، أسئلته من نوع اختيار من متعدد وله (٤) بدائل، ومن ثلاثة نماذج لها نفس المحتوى والمتغير هو موقع البديل الصحيح،

حيث النموذج الأول يكون موقع البديل الصحيح على البديلين (أ، ب)، والنموذج الثاني يكون موقع البديل الصحيح على البديلين (ج، د)، والنموذج الثالث يتوزع البديل الصحيح بشكل عشوائي على جميع البدائل الأربعة، وتم تطبيق نماذج الاختبار على عينة حجمها (٥٠٠) طالبة من طالبات الصف الثالث متوسط في مدينة ينبع، وتم تحليل البيانات. ومن أهم النتائج عدم وجود فروق دالة إحصائيًا بين متوسطات دقة تقدير معلمة الصعوبة والتمييز للمفردات تعزى لموقع البديل الصحيح، عدم وجود فروق ذات دلالة إحصائية بين متوسطات تقديرات معاملات الثبات التجريبي تعزى لموقع البديل الصحيح. بينما توجد فروق ذات دلالة إحصائية بين متوسطات تقديرات معاملات ثبات الأفراد تعزى لموقع البديل الصحيح.

دراسة المساوي (2018)، سعى الباحث إلى تعرف أثر بنية اختبار اختيار من متعدد على معالم المفردة ودقة تقديرها وفق نظرية الاستجابة للمفردة، أعد الباحث اختبارا موضوعيا (٤٠ سؤال) من نوع الاختيار من متعدد (٤) بدائل في مادة الحاسب الالي، وتم إدخال أربع مخالفات على أسئلة الاختبار وهي: طول البديل الصحيح، موقع البديل الصحيح، جميع ما ذكر صحيح، كلاهما صحيحان، بحيث أصبح الاختبار يتكون من نموذجين الأول يتكون من فقرات سليمة البنية، والثاني يتكون من نفس محتوى فقرات النموذج الأول لكن أدخل عليها المخالفات الأربع السابقة، شملت العينة (٦٢٤) طالبا.

ومن أهم نتائج البحث ما يأتي: الفقرات المتحررة من المخالفات التالية: "موقع البديل الصحيح، جميع ما ذكر صحيح كلاهما صحيحان" أكثر دقة في تقدير معلمة الصعوبة من الفقرات المخالفة. كما دلت على وجود فروق دالة إحصائيا بين متوسط معلمة التمييز للمفردات التي تتضمن المخالفة، وأن دقة تقدير الفقرات المتحررة من المخالفات لاتختلف عن دقة تقدير الفقرات المخالفة لمعلمة التمييز .كما دلت النتائج على وجود فروق دالة إحصائيا بين المتوسط الحسابي لمعلمة تخمين المفردات المتحررة من المخالفات ،والمتوسط الحسابي لمعلمة تخمين الفقرات التي تتضمن المخالفات، و وكان الفرق لصالح الفقرات المتحررة من المخالفات.

رابعاً: دراسات تناولت أثر عدد البدائل وموقع البديل الصحيح على الخصائص السيكومترية للاختبار

سعت دراسة الشريفين (2012) إلي الكشف عن أثر قدرات الافراد وطريقة تقدير المعالم للفقرات علي قيم معالم الفقرة والخصائص السيكومترية للاختبار، في ضوء اختلاف حجم العينة ، حيث تم بناء اختبار تحصيلي موضوعي في مقرر الفيزياء يتكون من أربعة بدائل، وقد تكون الاختبار من (٣٣) فقرة ،وطبق على عينة عددها (١٠٠٠) طالب وطالبة من طلبة الصف الثاني الثانوي العلمي، وحللت النتائج بإستخدام برمجة البايلوق وفق النموذج الثلاثي المعلمة ،واظهرت النتائج وجود فروق ذات دلالة إحصائية عند (α= 0.05) في متوسطات الأخطاء المعيارية لتقديرات معالم الفقرات تعزى للتفاعل بين حجم العينة وطريقة التقدير ، في حين لم تظهر فروق ذات دلالة إحصائية تعزى لمتغير حجم العينة أو طريقة التقدير ،

واظهرت النتائج ايضاً إلي وجود فروق ذات دلالة إحصائية (α= 0.05) في متوسطات الأخطاء المعيارية لتقديرات القدرة للأفراد تعزى لمتغير حجم العينة، وللتفاعل بين طريقة التقدير وحجم العينة، وعدم وجود فروق ذات دلالة إحصائية تعزى لطريقة التقدير،وعدم وجودفروق بين معاملات الثبات المقدرة عند أحجام العينة المختلفة (١٠٠، ٥٠٠، ١٠٠٠) وأظهرت النتائج إلي زيادة الدقة في تقديرات الافراد ذوي القدرة المتوسطة بإستخدام طريقة الارجحية العظمى بغض النظر عن حجم العينة ، وأن دقة تقديرات معلمة القدرة تزيد عند الأفراد ذوي القدرة العالية ، والأفراد ذوي القدرة المنخفضة عند استخدام طريقة بييز التوقع.

أظهرت دراسة ابوفوده (2014)إلى توضيح أثر إعادة ترتيب بدائل صعوبة الاستجابة في فقرة اختبارالاختيار من متعدد ، وقد تم بناء اختبار تحصيلي في مادة الرياضيات لطلبة الصف الثالث المتوسط ، ويتألف الاختبار من (٢٠) فقرة ، وتم إعداد الاستجابة والتي تكونت من (٣٣) بديلاً في كراستين، وكانت تتباين هذه الاستجابات في ترتيب البدائل للاستجابة الصحيحة لفقرات نموذج الاختبار ، ثمّ بعد ذلك تم تطبيق فقرات الاختبار على عينة بلغت (٦٠٠) من الطلبة من الجنسين في محافظة جرش ، كان نصيب كلٍ من الذكور والاناث (٣٠٠) طالب و(٣٠٠) طالبة ،وبعد إجراءات التطبيق تم حساب معامل صعوبة الفقرات للاختبار ، وللفقرات المتناظرة ، والفروق بين مواقع الاستجابة الصحيحة حسب نموذجي كراستي الاستجابةوبينت النتائج انه لايوجدفروق بين معاملات صعوبة الفقرات للاختبار تعزى الى ترتيب البدائل للاستجابات ،وأشارت المحصلة النهائية للنتائج انه لايوجد نمط يتعلق بإعادة ترتيب بدائل الاستجابة وتأثيرهافي فقرات الاختبار الصعبة، وبينت الدراسة ان إعادة ترتيب البدائل في الاختبارات المتكافئة قديكون في غاية الصعوبة

دراسة بني عطاء والرباعي (2013)، أراد الباحثان اكتشاف أثر عدد بدائل المفردة وموقع المموه القوي في فقرات اختبار الاختيار من متعدد على الخصائص السيكومترية للاختبار ومعالم الفقرات، تم إعداد اختبار موضوعي مكون من (٤١) سؤال من نوع الاختيار من متعدد في الرياضيات لطلبة الصف العاشر، وقد اشتمل على أربعة نماذج حسب عدد البدائل وموقع المموه القوي، وباستخدام برنامج (Bilog- Mg 3)، وتم تحليل بيانات عينة حجمها ٢١١١ طالبا وطالبة لكل النماذج الأربعة للاختبار وفق النموذج اللوجستي الثلاثي المعلمة. وقد اظهرت النتائج لتحليل التباين الثنائي عدم وجود فروق ذات دلالة إحصائية بين متوسطات معالم الصعوبة للمفردات تعزى لموقع المموه القوي، وعدد البدائل للمفردة ، وايضاً عدم وجود فروق ذات دلالة إحصائية بين متوسطات معالم التخمين للمفردات تعزى لموقع المموه القوي، وعدد البدائل للمفردة ، في حين اظهرت النتائج عن وجود فروق ذات دلالة إحصائية بين متوسطات معالم التمييز–تعزى لموقع المموه القوي، ولصالح النموذج الثاني. (٥ بدائل موقع المموه القوي بعيد)، واظهرت النتائج كذلك عدم وجود فروق دالة إحصائيا بين متوسطات معلمة القدرة للأفراد

تعزى لموقع المموه القوي، وعدد البدائل للمفردة ، واظهرت النتائج ايضاً بتباين دالة المعلومات للاختبار مع تباين نماذج الاختبار، ووجدت هناك فروق دالة إحصائيا لقيم معاملات الثبات تعزى لصالح النموذج الثاني.

التعليق على الدراسات السابقة:

من حيث الهدف: توزعت أهداف الدراسات السابقة بين ثلاثة محاور أساسية، دراسات هدفت دراسة أثر اختلاف عدد البدائل على الخصائص السيكومترية مثل: دراسة القحطاني (٢٠٢١)، دراسة حميدة (٢٠٢٠)، دراسة النصراويين (٢٠١٩)، دراسة عودة (٢٠١٦)، دراسة أبوجراد (٢٠١٥)، دراسة بابان (٢٠١٤)، دراسة الرغيلات (٢٠١٤)، دراسة الشريفين وطعامنة(٢٠٠٩).

ودراسات هدفت إلى تعرف أثر موقع البديل الصحيح على الخصائص السيكومترية للأداة، مثل: دراسة المرواني وسليمان (٢٠١٩)، دراسة المساوي (٢٠١٨).

ودراسات هدفت إلى تعرف أثر اختلاف عدد البدائل وموقع البديل الصحيح على الخصائص السيكومترية للأداة، مثل: دراسة بني عطاء والرباعي (٢٠١٣)، ودراسة الشريفين(٢٠١٢).

ودراسات تناولت إحصاء بايزي مثل دراسة Nishio, Akasaka, Sakamoto & Togashi (2019)

من حيث العينة والأدوات: كانت الأداة الأكثر استخداما هي الاختبارات التحصيلية في الرياضيات والعلوم والقليل منها كان مقاييس شخصية. تراوحت العينات بين ٤٨ إلى ٢٧٨٠ بحسب النظرية التي اعتمدت عليها الدراسة ما بين الكلاسيكية والنظرية الحديثة، فالمقارنة التقليدية لا تشترط حجما كبيرا للعينة بعكس المقارنة حسب نظرية الاستجابة للمفردة التي تتطلب عينة لا تقل حجمها عن ١٥٠.

من حيث النتائج:

أيدت بعض النتائج تأثير عدد البدائل أو موقع البديل الصحيح على بعض خصائص الاختبارات مثل دراسة حميدة (2020)، ودراسة المرواني وسليمان (2019)، ودراسة الرغيلات (2014)، بينما لم تؤيد ذلك دراسات كل من النصراويين (2019)، وعودة (2016)، أبوجراد (2015)، بني عطاء والرباعي (2013).

فروض البحث:

من خلال نتائج الدراسات السابقة والإطار النظري، اشتق الباحث فروض البحث على النحو الآتي:

١.يختلف ثبات اختبار مقرر القياس والتقويم بين –باختلاف عدد البدائل للسؤال الموضوعي بين نظرية القياس الكلاسيكية والنموذج الإحصائي لبايزي لدى طلاب كلية التربية جامعة الملك خالد.

٢.يختلف الاتساق الداخلي لاختبار مقرر القياس والتقويم –في ضوء أنموذج بايزي– باختلاف عدد البدائل للسؤال الموضوعي بين نظرية القياس الكلاسيكية والنموذج الإحصائي لبايزي لدى طلاب كلية التربية جامعة الملك خالد.

٣.يختلف ثبات اختبار مقرر القياس والتقويم –في ضوء نموذج بايزي– باختلاف موقع البديل الصحيح للسؤال الموضوعي بين نظرية القياس الكلاسيكية والنموذج الإحصائي لبايزي لدى طلاب كلية التربية جامعة الملك خالد.

٤.يختلف الاتساق الداخلي لاختبار مقرر القياس والتقويم –في ضوء نموذج بايزي–باختلاف موقع البديل الصحيح للسؤال الموضوعي بين نظرية القياس الكلاسيكية والنموذج الإحصائي لبايزي لدى طلاب كلية التربية جامعة الملك خالد.

٥.يختلف ثبات اختبار مقرر القياس والتقويم –في ضوء نموذج بايزي– بتفاعل عدد البدائل وموقع البديل الصحيح للسؤال الموضوعي بين نظرية القياس الكلاسيكية والنموذج الإحصائي لبايزي لدى طلاب كلية التربية جامعة الملك خالد.

إجراءات البحث:

منهج البحث: تم إجراء البحث في ضوء خطوات المنهج الوصفي.

مجتمع البحث: طلاب قسم علم النفس كلية التربية جامعة الملك خالد.

عينة البحث: تكونت عينة البحث من (100) طالب من طلاب كلية التربية جامعة الملك خالد ممن يدرسون مقرر القياس والتقويم.

أداة البحث:

قام الباحث بإعداد اختبار في مقرر القياس والتقويم، وفق الخطوات التالية:

تحديد الهدف من الاختبار

يهدف الاختبار إلى قياس التحصيل الدراسي لطلاب كلية التربية جامعة الملك خالد في مقرر القياس والتقويم

تحليل المحتوى:قام الباحث بتحليل محتوى مقرر القياس والتقويم، وركز على أهداف التذكر والفهم والتطبيق، واعتمد على عدد الصفحات في تحديد الأهمية النسبية للموضوعات التي تم فيها الاختبار، وكان جدول الأهمية النسبية للموضوعات كما يأتي:

جدول (١) الأهمية النسبية لموضوعات اختبار القياس والتقويم في ضوء عدد الصفحات

الموضوع	عدد الصفحات	الأهمية النسبية	عدد الأسئلة من ٢٠
التقويم التربوي	٣٣	%٣١,١٣٢	٦ *تقريبا
القياس النفسي	٢٥	%٢٣,٥٨٤	٥ تقريبا
الفروق الفردية	٢٤	%٢٢,٦٤٢	٥ تقريبا
مجالات التقويم	٢٤	%٢٢,٦٤٢	٥ تقريبا
المجموع	١٠٦	%١٠٠	٢١

*تم الاكتفاء بخمسة أسئلة لموضوع التقويم التربوي

جدول (٢) جدول المواصفات للاختبار التحصيلي

الموضوع	الأهمية النسبية	الأهداف السلوكية			عدد الأسئلة
		التذكر	الفهم	التطبيق	
التقويم التربوي	%٣١,١٣٢	٢ %١٠ %٢٥	٢ %١٠ %٢٨,٥٧	١ %٥ %٢٠	٥ %٢٥
القياس النفسي	%٢٣,٥٨٤	١ %٥ %.١٢,٥	٢ %١٠ %٢٨,٥٧	٢ %١٠ %٤٠	٥ %٢٥
الفروق الفردية	%٢٢,٦٤٢	٣ %١٥ %٣٧,٥	١ %١٠ %١٤,٢٩	١ %١٠ %٢٠	٥ %٢٥
مجالات التقويم	%٢٢,٦٤٢	٢ %١٠ %٢٥	٢ %١٠ %٢٨,٥٧	%١٥ %٢٠	٥ %٢٥
المجموع	%١٠٠	٨	٧	٥	٢٠

وتم صياغة تعليمات الاختبار باعتباره اختبارا للأعمال الفصلية، وسوف تضاف درجته كجزء من الدرجة الكلية للطالب في المقرر نهاية الفصل الدراسي.

الصورة المبدئية للاختبار، يتكون من ٢٠ سؤالاً (٨ تذكر، ٧ فهم، ٥ تطبيق)، وتوجد اربعة اختيارات للإجابة عن كل سؤال، بحيث تكون هناك إجابة واحدة صحيحة.

الخصائص السيكومترية للاختبار:

من خلال بيانات التطبيق المبدئي للاختبار، قام الباحث بالتأكد من الخصائص السيكومترية للاختبار على النحو الآتي:

أولا: معاملات السهولة والصعوبة والتمييز

معاملات السهولة والصعوبة:

قام الباحث بحساب معاملات السهولة (عدد الإجابات الصحيحة/ عدد أفراد العينة)، ومعاملات الصعوبة (١-معامل السهولة)، ومعامل التباين (معامل السهولة*معامل الصعوبة)، معامل التمييز

(مجموع أعلى ٢٧% من العينة في الاختبار (وكانوا من الذين حصلوا على ٢٠/١٩ فأكثر) وبلغ عددهم ٢٠ طفل، مطروحا منه مجموع أقل ٢٧% من العينة في الاختبار (وكانوا من الذين حصلوا على ٢٠/٨ فأقل) وبلغ عددهم ١٧ طالبا. ويحسب درجاتهم في كل سؤال، ويتم طرح درجات المجموعة الدنيا من درجات المجموعة العليا، ويقسم الناتج على عدد إحدى المجموعتين أو متوسط المجموعتين المساوي (١٨) تقريبا. والنتائج موضحة في الجدول التالي:

جدول (٣) معاملات الصعوبة والسهولة والتباين والتمييز لأسئلة اختبار التحصيل في مقرر القياس والتقويم

رقم السؤال	معامل الصعوبة (ص)	معامل السهولة (س)	مجموع درجات العليا* (مج١)	مجموع درجات الدنيا (مج٢)	التمييز	
			ن=٢٠	ن=١٧	التباين س*ص	معامل التمييز (مج١- مج٢)/١٨
س١	٠,٤٠٤	٠,٥٩٦	٢٠	٥	٠,٢٤٩	٠,٧٥
س٢	٠,٤٠٤	٠,٥٩٦	٢٠	٢	٠,٢٤٩	٠,٩٠
س٣	٠,٢٩٨	٠,٧٠٢	١٩	٥	٠,٢١٥	٠,٧٠
س٤	٠,٢٢٨	٠,٧٧٢	٢٠	٨	٠,٢٢٦	٠,٦٠
س٥	٠,٢٦٣	٠,٧٣٧	٢٠	٥	٠,٢٤١	٠,٧٠
س٦	٠,٣٨٦	٠,٦١٤	٢٠	٢	٠,٢٤١	٠,٨٥
س٧	٠,١٧٥	٠,٨٢٥	١٩	٥	٠,٢٠٩	٠,٥٥
س٨	٠,٤٥٦	٠,٥٤٤	٢٠	٨	٠,١٧٦	٠,٨٥
س٩	٠,٤٢١	٠,٥٧٩	٢٠	٦	٠,١٩٤	٠,٧٠
س١٠	٠,١٥٨	٠,٨٤٢	٢٠	٣	٠,٢٣٧	٠,٥٥
س١١	٠,٢٤٦	٠,٧٥٤	٢٠	٩	٠,١٤٤	٠,٥٠
س١٢	٠,١٢٣	٠,٨٧٧	٢٠	٣	٠,٢٤٨	٠,٤٥
س١٣	٠,٣٥١	٠,٦٤٩	٢٠	٦	٠,٢٤٤	٠,٦٥
س١٤	٠,٣٥١	٠,٦٤٩	٢٠	٩	٠,١٣٣	٠,٧٥
س١٥	٠,٣٥١	٠,٦٤٩	١٩	٩	٠,١٨٥	٠,٦٠
س١٦	٠,٣٣٣	٠,٦٦٧	٢٠	١١	٠,١٠٨	٠,٦٥
س١٧	٠,٣٥١	٠,٦٤٩	٢٠	٧	٠,٢٢٨	٠,٨٥
س١٨	٠,٣٦٨	٠,٦٣٢	٢٠	٥	٠,٢٢٨	٠,٧٠
س١٩	٠,٣١٦	٠,٦٨٤	٢٠	٨	٠,٢٢٨	٠,٦٠
س٢٠	٠,٤٧٤	٠,٥٢٦	٢٠	٧	٠,٢٢٢	٠,٩٥

يتضح من الجدول رقم (٣)، أن غالبية الأسئلة حققت المستوى النموذجي من حيث السهولة والصعوبة (٠,٢٠-٠,٨٠)، فيما عدا الأسئلة أرقام (٧، ١٠، ١٢)، حيث كان معامل الصعوبة لكل منها أقل من ٠,٢٠، ومعاملات التمييز (جميعها أكبر من ٠,٣٩ أي معاملات تمييز عالية، كما أن معاملات التباين كانت جميعها أكبر من ٠,١٣ وأفضل سؤال هو ما كان معامل تباينه ٠,٢٥ أو قريب منه. ولذلك أبقى الباحث على الأسئلة جميعها لقدرتها التمييزية العالية.

الثبات:

للتأكد من ثبات الاختبار، قام الباحث باستخدام معادلة كيودر-ريتشاردسون، والتجزئة النصفية، وكانت النتائج كما يأتي:

جدول (٤) معاملات ثبات الاختبار

الأبعاد	معامل الثبات	
	التجزئة النصفية	كيودر-ريتشاردسون
التذكر	٠,٧٧٨	٠,٧٩١
الفهم	٠,٨١٨	٠,٨٣١
التطبيق	٠,٧٠٧	٠,٧١١
الاختبار ككل	٠,٨٢٦	٠,٨٤٣

يتضح من الجدول (٤)وجود معاملات ثبات مقبولة للاختبار.

صدق الاختبار:

صدق المحكمين:

تم عرض الاختبار على عشرة من المحكمين في مجال علم النفس والمناهج وطرائق التدريس، من أجل:

١.تحديد مدى قياس السؤال لما وضع له.

٢.وضوح صياغة السؤال.

٣.إضافة ما يرونه من أسئلة.

٤.وتم إجراء التعديلات التي اتفق عليها ٨٠% من المحكمين، وأصبح الاختبار مكون من (٢٠) سؤالا موزعة على (التذكر، الفهم، التطبيق).

الاتساق الداخلي للاختبار:

قام الباحث بحساب معامل الارتباط بين كل سؤال والدرجة الكلية للاختبار، وكانت النتائج كما يأتي:

جدول (٥) معاملات ارتباط أسئلة الاختبار التحصيلي بالدرجة الكلية للاختبار

م	معامل الارتباط بالدرجة الكلية	م	معامل الارتباط بالدرجة الكلية
١	**٠,٦٥٩	١١	**٠,٦٢٩
٢	*٠,٧٧٠	١٢	**٠,٥٥٠
٣	**٠,٥٥٣	١٣	**٠,٧٩٢
٤	**٠,٥٣٠	١٤	**٠,٥٤٥
٥	**٠,٦٢٥	١٥	**٠,٥٣٦
٦	**٠,٧٣٩	١٦	**٠,٥٦١
٧	**٠,٤٩٠	١٧	**٠,٥٦٧
٨	**٠,٧٢٥	١٨	**٠,٦٧٥
٩	**٠,٦٢٨	١٩	**٠,٦١٢
١٠	**٠,٤٩٨	٢٠	**٠,٨٠٥

****دال عند ٠,٠١**

يتضح من الجدول رقم (٥) ارتباط الأسئلة بالدرجة الكلية للاختبار وهذا يعني الاتساق الداخلي للاختبار مما سبق اطمأن الباحث لمناسبة الاختبار للتطبيق على العينة الحالية للبحث

وتكونت الصورة المبدئية للمقياس من (٢٠) فقرة اختيار من متعدد بعدد بدائل (٤) بدائل.

نتائج البحث:

نتائج التحقق من صحة الفرض الأول، والذي ينص على

" يختلف ثبات اختبار مقرر القياس والتقويم باختلاف عدد البدائل للسؤال الموضوعي بين نظرية القياس الكلاسيكية والنموذج الإحصائي لبايزي لدى طلاب كلية التربية جامعة الملك خالد". وللتحقق من صحة هذا الفرض استخدم الباحث الإحصاء البايزي Bayesian Statistics من خلال برنامج جافيز الإحصائي المعروف اختصارا JASP والذي يساعد في القيام بالاختبارات الإحصائية الكلاسيكية والبايزية معا، وكانت النتائج كما يلي:

جدول (٦) قيم معاملات ثبات الاختبار التحصيلي في القياس والتقويم في ضوء عدد البدائل

عدد البدائل	إحصاء بايزي	إحصاء تقليدي
	McDonald's ω أوميجا ماكدونالد	Kuder-Richardson20
بديلين	٠,٦١١	٠,٦٧٦
ثلاث بدائل	٠,٨٠٤	٠,٨٢٨
أربع بدائل	٠,٨١١	٠,٨٣٢

يتضح من الجدول رقم (٦) أنه كلما زاد عدد البدائل كلما زاد ثبات الاختبار، كما أن الفروق بين قيمة معامل الثبات بإحصاء بايزي وقيمته بالإحصاء التقليدي، كانت فروقا بسيطة،

نتائج التحقق من صحة الفرض الثاني، والذي ينص على:

يختلف الاتساق الداخلي لاختبار مقرر القياس والتقويم باختلاف عدد البدائل للسؤال الموضوعي بين نظرية القياس الكلاسيكية والانموذج الإحصائي لبايزي لدى طلاب كلية التربية جامعة الملك خالد وللتحقق من صحة هذا الفرض، استخدم الباحث معاملات الارتباط في ضوء النظرية اللكلاسيكسة والنموذج الإحصائي لبايزي، وكانت النتائج كما يأتي:

جدول (٧) معاملات الارتباط بين درجات الأسئلة والدرجة الكلية للاختبار في حالة اختلاف البدائل

الأسئلة	في حالة أربع بدائل				الأسئلة	في حالة بديلين			
	إحصاء تقليدي		إحصاء بايزي			إحصاء تقليدي		إحصاء بايزي	
	معامل الارتباط	الدلالة	معامل الارتباط	الدلالة للفرض		معامل الارتباط	الدلالة	معامل الارتباط	الدلالة
١	٠,٣١٩	٠,٠١	٢١,٤٣٥	قوي بديل	١	٠,١٣٥	غير دالة	٠,٣٠٣	متوسط صفري
٢	٠,٢٩٤	٠,٠١	٩,٦٣٦	متوسط بديل	٢	٠,٢٩١	غير دالة	٨,٥٥٩	متوسط بديل
٣	٠,٠٨٩	غير دالة	٠,١٨٢	متوسط صفري	٣	٠,٣٥٥	٠,٠١	٧٥,١٠٨	قوي جدا بديل
٤	–	غير	٠,١٥٩	متوس	٤	٠,٥٣	٠,٠	٩٨٠٨٤٩,٩	قوي جدا

الأسئلة	في حالة أربع بدائل				الأسئلة	في حالة بديلين			
	إحصاء تقليدي		إحصاء بايزي			إحصاء تقليدي		إحصاء بايزي	
	معامل الارتباط	الدلالة	معامل الارتباط	الدلالة للفرض		معامل الارتباط	الدلالة	معامل الارتباط	الدلالة
	٠,٠٧	دالة		ط صفري		٣	١	٠	بديل
٥	٠,٢٤٧	٠,٠٥	٢,٥٦٢	قولي بديل	٥	٠,١٩٧	غير دالة	٠,٨٢٧	قولي صفري
٦	٠,٢٩٦	٠,٠١	١٠,٣٥٧	قوي بديل	٦	٠,٢١٦	غير دالة	١,٢١٥	قولي بديل
٧	٠,٣٩٥	٠,٠١	٤٤٠,٠٥٩	أقصى بديل	٧	٠,٢٢٢	غير دالة	١,٣٨٨	قولي بديل
٨	-٠,٠٠٨	غير دالة	٠,١٢٥	متوسط صفري	٨	٠,٢٤٢	غير دالة	٢,٢١١	قولي بديل
٩	٠,٣٨١	٠,٠١	٢٣٢,١٠٩	أقصى بديل	٩	٠,٣٠٧	٠,٠٥	١٣,٧٣٨	قوي بديل
١٠	٠,١٨٨	غير دالة	٠,٧٠٧	قولي صفري	١٠	٠,٢٢٥	غير دالة	١,٤٨٤	قولي بديل
١١	٠,٢٥٥	٠,٠١	٣,١٨١	متوسط بديل	١١	٠,٣٤٩	٠,٠١	٦٠,٣٦٨	قوي جدا بديل
١٢	٠,٤٢١	٠,٠١	١٤٨٥,٢٩٩	أقصى بديل	١٢	٠,٢٢٠	غير دالة	١,٣٤٢	قولي بديل
١٣	٠,٥٠٠	٠,٠١	١٢٧٧٦١,٠٤٩	أقصى بديل	١٣	٠,٢٠٠	غير دالة	٠,٨٧٧	قولي صفري
١٤	٠,٤٩	٠,٠	٦٨٤٠٤,٦٥٠	أقصى	١٤	٠,٣٥	٠,٠	٨٠,٣٠٧	قوي جدا

الأسئلة	في حالة أربع بدائل				الأسئلة	في حالة بديلين			
	إحصاء تقليدي		إحصاء بايزي			إحصاء تقليدي		إحصاء بايزي	
	معامل الارتباط	الدلالة	معامل الارتباط	الدلالة للفرض		معامل الارتباط	الدلالة	معامل الارتباط	الدلالة
	٠	١		بديل		٦	١		بديل
١٥	٠,٣٨٧	٠,٠١	٣٠٨,٠٧٦	أقصى بديل	١٥	٠,٢٩٤	غير دالة	٩,٢٢٤	متوسط بديل
١٦	٠,٣١٢	٠,٠١	١٧,٣١٨	قوي بديل	١٦	٠,٢٨١	غير دالة	٦,٢٩٤	متوسط بديل
١٧	٠,٤٦٧	٠,٠١	١٦٥٩٧,٦٧٧	أقصى بديل	١٧	٠,٣٤٦	٠,٠١	٥٤,٨٧٠	قوي جدا بديل
١٨	٠,٤١٤	٠,٠١	١٠٢٢,٤١٨	أقصى بديل	١٨	٠,٣٤٤	٠,٠١	٤٩,٣٧٤	قوي جدا بديل
١٩	٠,٢٨٨	٠,٠١	٨,٠٧٤	متوسط بديل	١٩	٠,٢٦٩	غير دالة	٤,٤٩٠	متوسط بديل
٢٠	٠,٥٠٥	٠,٠١	١٦٨١٤٩,٨٠٥	أقصى بديل	٢٠	٠,٣٠٣	٠,٠٥	١٢,١٨٠	قوي بديل

يتضح من الجدول رقم (٧) وجود اتساق داخلي بين الأسئلة والدرجة الكلية للاختبار في حالة عدد البدائل أربعة، كما وجد اتفاق كبير بين نتائج إحصاء بايزي والإحصاء التقليدي، بينما على العكس في حالة البديلين فقد وجد اتساق داخلي ضعيف في حالة الإحصاء التقليدي ومتوسطا في حالة إحصاء بايزي، حيث كانت دلالات الارتباط وميلها جهة الفرض البديل في حالة إحصاء بايزي أكثر منها في حالة الإحصاء البديل

نتائج التحقق من صحة الفرض الثالث، والذي ينص على:

يختلف ثبات اختبار مقرر القياس والتقويم باختلاف موقع البديل الصحيح للسؤال الموضوعي بين نظرية القياس الكلاسيكية والنموذج الإحصائي لبايزي لدى طلاب كلية التربية جامعة الملك خالد.

وللتحقق من صحة هذا الفرض، استخدم الباحث معاملات الثبات في الإحصاء البايزي McDonald's ω أوميجا ماكدونالد، ومعامل الثبات في الإحصاء الكلاسيكي كيودر–ريتشاردسون Kuder-Richardson20، وكانت النتائج كما يأتي:

جدول (٨) قيم معاملات ثبات الاختبار التحصيلي في القياس والتقويم في ضوء موقع البديل الصحيح

موقع البديل	إحصاء بايزي	إحصاء تقليدي
	McDonald's ω أوميجا ماكدونالد	Kuder-Richardson20
أ أو ب	٠,٦٩٣	٠,٧٢٨
ج أو د	٠,٧٠٠	٠,٧٥٦

يتضح من الجدول رقم (٨) أن معامل الثبات يزداد في الإحصاء البايزي عنه في الإحصاء التقليدي، وأن موقع البديل الصحيح في الإحصاء البايزي له تأثير طفيف على قيمة معامل الثبات (٠,٦٩٣ مقابل ٠,٧٠٠)، بينما له تأثير أكثر وضوحا في الإحصاء التقليدي (٠,٧٢٨ في مقابل ٠,٧٥٦)

نتائج التحقق من صحة الفرض الرابع، والذي ينص على:

يختلف الاتساق الداخلي لاختبار مقرر القياس والتقويم باختلاف موقع البديل الصحيح للسؤال الموضوعي بين نظرية القياس الكلاسيكية والنموذج الإحصائي لبايزي لدى طلاب كلية التربية جامعة الملك خالد، وللتحقق من صحة هذا الفرض، استخدم الباحث معاملات الارتباط في ضوء النظرية الكلاسيكية والنموذج الإحصائي لبايزي، وكانت النتائج كما يلي:

جدول (٩) معاملات الارتباط بين درجات الأسئلة والدرجة الكلية للاختبار في حالة موقع البديل أ أو ب

الأسئلة	إحصاء تقليدي		إحصاء بايزي		الأسئلة	إحصاء تقليدي		إحصاء بايزي	
	معامل الارتباط	الدلالة	معامل الارتباط	الدلالة للفرض		معامل الارتباط	الدلالة	معامل الارتباط	الدلالة
١	٠,٣١٩	٠,٠١	٢١,٤٣٥	قوي بديل	١١	٠,٢٥٥	٠,٠١	٣,١٨١	متوسط بديل
٣	٠,٠٨٩	غير دالة	٠,١٨٢	متوسط صفري	١٣	٠,٥٠٠	٠,٠١	١٢٧٧٦١,٠٤٩	أقصى بديل
٥	٠,٢٤٧	٠,٠٥	٢,٥٦٢	قولي	١٥	٠,٣٨٧	٠,٠١	٣٠٨,٠٧٦	أقصى

			بديل					بديل
٧	٠,٣٩٥	٠,٠١	٤٤٠,٠٥٩	أقصى بديل	١٧	٠,٤٦٧	٠,٠١	١٦٥٩٧,٦٧٧	أقصى بديل
٩	٠,٣٨١	٠,٠١	٢٣٢,١٠٩	أقصى بديل	١٩	٠,٢٨٨	٠,٠١	٨,٠٧٤	متوسط بديل

يتضح من الجدول رقم (٩) وجود تشابه في مستوى الدلالة بين النظرية التقليدية وإحصاء بايزي، وكانت جميع معاملات الارتباط في صالح الفرض البديل الذي يشير إلى وجود ارتباط موجب دال بين درجة السؤال، والدرجة الكلية للاختبار، فيما عدا سؤالاً واحداً كانت لصالح الفرض الصفري الذي يشير إلى عدم وجود ارتباط موجب دال بين درجة السؤال والدرجة الكلية للاختبار.

جدول (١٠) معاملات الارتباط بين درجات الأسئلة والدرجة الكلية للاختبار في حالة موقع البديل ج أو د

الأسئلة	إحصاء تقليدي		إحصاء بايزي		الأسئلة	إحصاء تقليدي		إحصاء بايزي	
	معامل الارتباط	الدلالة	معامل الارتباط	الدلالة للفرض		معامل الارتباط	الدلالة	معامل الارتباط	الدلالة
٢	٠,٢٩٤	٠,٠١	٩,٦٣٦	متوسط بديل	١٢	٠,٤٢١	٠,٠١	١٤٨٥,٢٩٩	أقصى بديل
٤	-٠,٠٧	غير دالة	٠,١٥٩	متوسط صفري	١٤	٠,٤٩٠	٠,٠١	٦٨٤٠٤,٦٥٠	أقصى بديل
٦	٠,٢٩٦	٠,٠١	١٠,٣٥٧	قوي بديل	١٦	٠,٣١٢	٠,٠١	١٧,٣١٨	قوي بديل
٨	-٠,٠٠٨	غير دالة	٠,١٢٥	متوسط صفري	١٨	٠,٤١٤	٠,٠١	١٠٢٢,٤١٨	أقصى بديل
١٠	٠,١٨٨	غير دالة	٠,٧٠٧	قولي صفري	٢٠	٠,٥٠٥	٠,٠١	١٦٨١٤٩,٨٠٥	أقصى بديل

يتضح من الجدول رقم (١٠) وجود تشابه في مستوى الدلالة بين النظرية التقليدية وإحصاء بايزي، وكانت جميع معاملات الارتباط في صالح الفرض البديل الذي يشير إلى وجود ارتباط موجب دال بين درجة السؤال، والدرجة الكلية للاختبار، فيما عدا (٣) أسئلة من بين الأسئلة العشرة، كانت لصالح الفرض الصفري الذي يشير إلى عدم وجود ارتباط موجب دال بين درجة السؤال والدرجة الكلية للاختبار.

تفسير النتائج ومناقشتها:

توصل الباحث إلى عدد من النتائج بناء على اختبار صحة فروض البحث، حيث توصل إلى:

كلما زاد عدد البدائل كلما زاد ثبات الاختبار، كما أن الفروق بين قيمة معامل الثبات بإحصاء بايزي وقيمته بالإحصاء التقليدي، كانت فروقا بسيطة.

وجود اتساق داخلي بين الأسئلة والدرجة الكلية للاختبار في حالة عدد البدائل أربعة، كما وجد اتفاقاً كبيراً بين نتائج إحصاء بايزي والإحصاء التقليدي، بينما على العكس في حالة البديلين فقد وجد اتساقاً داخلياً ضعيفاً في حالة الإحصاء التقليدي ومتوسطا في حالة إحصاء بايزي

أن معامل الثبات يزداد في الإحصاء البايزي عنه في الإحصاء التقليدي، وأن موقع البديل الصحيح في الإحصاء البايزي له تأثير طفيف على قيمة معامل الثبات، بينما له تأثير أكثر وضوحا في الإحصاء التقليدي.

ومن خلال تلك النتائج، وبمراجعة الدراسات السابقة، لم يجد الباحث أي مقارنة بين نتائج النظرية الكلاسيكية والإحصاء البايزي من حيث مؤشرات الثبات والصدق، إلا دراسة بابان (2014)، والتي كان هدفها المقارنة بين النظريتين في خصائص اختبار للذكاء، والتي توصلت إلى وجود فروق لصالح النظرية الحديثة، ودراسة Nishio , Akasaka, Sakamoto & Togashi (2019)، ولم يكن الهدف هو مقارنة تأثير استخدام كلا من الآنموذجين على الثبات والصدق ولكن كان الهدف هو التحقق من دقة التقدير لكل منهما، والتي توصلت إلى أن الإحصاء البايزي أكثر ملاءمة. وهذا ما يميز هذا البحث، ولكن من حيث المتغيرات الرئيسة (عدد البدائل، وموقع البديل الصحيح)، فقد أشارت النتائج إلى وجود أثر لكل منها على الثبات والصدق، وهذا يتفق مع نتائج، دراسة القحطاني (2021)، دراسة حميدة (2020)، التي توصلتا إلى وجود أثر لعدد البدائل، حيث أشارت إلى أنه كلما زاد عدد البدائل زاد الثبات والصدق، ولكنها تختلف مع نتائج دراسة النصراويين (2019)، دراسة عودة (2016)، دراسة أبوجراد (2015)، دراسة الرغيلات (2014)، دراسة الشريفين (2009)، والتي توصلت إلى أن اختلاف عدد بدائل السؤال ليس له تأثير على الثبات والصدق للاختبار التحصيلي.

ومن حيث موقع البديل الصحيح فقد توصلت الدراسة الحالية إلى أنه ليس له أثر واضح على قيمة معامل الثبات في الإحصاء البايزي، ولكن له تأثير في حالة الإحصاء التقليدي، وبخصوص تأثيره على الاتساق الداخلي للاختبار فلم يكن له تأثير واضح سواء بالإحصاء البايزي أو الإحصاء التقليدي.

وتلك النتائج تتفق مع دراسة المرواني وسليمان (2019)، دراسة المساوي (2018)، ولكنها تختلف مع دراسة بني عطاء والرباعي (2013).

التوصيات

١.ضرورة زيادة عدد بدائل السؤال الموضوعي بحيث لا تقل عن اربع بدائل كحد أدني

٢.ضرورة التنوع في وضع البديل الصحيح بحيث لا يأخذ موضعا واحدا

٣.ضرورة التدريب على أساليب وطرائق الإحصاء البايزي لما لها من دقة التقدير واختبار صحة الفروض

Recommendations:

1.There is a need to increase the number of the alternatives of the objective questions, whereas the number should be four minimum .

2.The need to place the correct alternative in diverse places

3.The need to receive explicit training on Paizi's statistical method due to its accuracy in examining hypotheses.

المراجع العربية:

١.أبو جراد، حمدي يونس (٢٠١٥)، أثر اختلاف عدد البدائل الصحيحة في فقرات الصواب – خطأ المتعدد في دقة تقدير صعوبة الفقرات وقدرات الأفراد وثبات الاختبار، مجلة البحث العلمي في التربية، ١٦(٣)، ١-٢٠.

٢.أبو فودة، باسل خميس سالم (٢٠١٤)، أثر إعادة ترتيب بدائل الاستجابة في صعوبة فقرة الاختيار من متعدد، دراسات عربية في التربية وعلم النفس، ٥٣، ٢٦٣-٢٨٧.

٣.بابان، وليد خالد عبد الكريم (٢٠١٤)، الخصائص القياسية لمقياس ميداس للذكاءات المتعددة وفقا لنظريتي القياس التقليدية والسمات الكامنة، رسالة دكتوراه غير منشورة، كلية التربية، جامعة بغداد-العراق

٤.بني عطا، زايد صالح؛ الرباعي، إبراهيم محمد (٢٠١٣)، أثر عدد البدائل وتغيير موقع المموه القوي في فقرات اختبار الاختيار من متعدد على معالم الفقرات وقدرة الفرد ودالة المعلومات، المجلة الأردنية في العلوم التربوية، ٩(٣)، ٣١٩-٣٣٣.

٥.حميدة، إبراهيم عبد الرحيم إبراهيم (٢٠٢٠)، مدى اختلاف الخصائص السيكومترية لأداة القياس في ضوء تباين عدد بدائل الاستجابة : حالة مقياس القلق الرقمي، المجلة العربية لعلم النفس، ٥(٢)، ٧٨-٩٨.

٦.الزغيلات، أحمد عبد الحافظ عطا الل (٢٠١٤)، أثر عدد البدائل على دقة تقديرات مؤشرات الصعوبة والتمييز في اختيار من متعدد وفق النموذج ثنائي المعلم، عالم التربية، ٤٧(٢)، ٢٥٣-٢٧٩.

٧.سكران، السيد عبدالدائم .(٢٠١٣).مهارات استخدام حزم البرامج الإحصائية في البحوث العلمية ،أبها:مطابع السروات

٨.الشريفين، نضال كمال (٢٠١٢)، أثر طريقة تقدير معالم المفردة وقدرات الأفراد على قيم معالم المفردة ، والخصائص السيكومترية للاختبار، في ضوء تغير حجم العينة، المجلة التربوية، ٢٦(١٠٤)، ١٧٧-٢٣٨.

٩.الشريفين، نضال كمال؛ طعامنة، إيمان صالح صلاح (٢٠٠٩)، أثر عدد البدائل في إختبار الإختيار من متعدد في تقديرات القدرة للأفراد والخصائص السيكومترية للمفردات والإختبار وفق نموذج راش في نظرية الإستجابة للمفردة، المجلة الأردنية في العلوم التربوية، ٥(٤)، ٣٠٩-٣٣٥.

١٠.عودة، موسى عزت (٢٠١٦)، تقدير معالم المفردة والقدرة بالطرق التي تستند إلى نظرية الاستجابة للمفردة تبعا لمتغيري طول الاختبار وعدد البدائل، رسالة دكتوراة غير منشورة، الجامعة الأردنية، عمان

١١.القحطاني، نائلة معيض (٢٠٢١)، تأثير مستوى الاكتئاب وعدد بدائل الاستجابة على الخصائص السيكومترية لمقياس الشعور بالوحدة النفسية، مجلة العلوم التربوية، ٧(١)، ٢٩١-٣٢٠.

١٢.المرواني، أشواق ضيف الله سليم وسليمان، شاهرخالد (٢٠١٩)، أثر موقع البديل الصحيح في اختبار اختيار من متعدد على دقة تقدير معالم المفردة وفق النموذج ثنائي المعلمة لنظرية الاستجابة للمفردة، المجلة الدولية للدراسات التربوية والنفسية، ٦(١٩)، ٦٦-٨٨.

١٣.المساوي، محمد علي جابر (٢٠١٨)، أثر بنية اختبار اختيار من متعدد في الحاسب الآلي على معالم المفردة وفق نظرية استجابة المفردة، مجلة كلية التربية، ٣٤(٢)، ٤٠٧-٤٣٩.

١٤.النصراوين، معين (٢٠١٩)، دالة معلومات المفردة والاختبار والخطأ المعياري عند استخدام ثلاثة نماذج من اختبار الاختيار من متعدد في اطار نظرية الاستجابة للمفردة، المجلة الدولية للأبحاث التربوية، ٤٣(٣)، ١٥٨-١٨١.

المراجع الأجنبية:

1.AERA, APA, & NCME. (2014). Standarts for educational and psychological testing. Washington, DC: American Educational Research Association.

2.Bobcock, B.G.E.(2009). Estimating a Noncompensatory IRT Model Using a modified. Metropolis algorithm. Unpublished Doctoral Dissertation. The University of Minesota.

3.Bollen, K. A. & Curran, P. J. (2004). Autoregressive latent trajectory (ALT) models a synthesis of two traditions. Sociological Methods and Research, 32(3), 336–383.

4.Budescu, D. V. & Nevo, B. (1985).Optimal Number of Options: An Investigation of the Assumption of Proportionality. Journal of Educational Measurement, 22(3), 183-196.

5.Chen, M.-H. (2005). Computing marginal likelihoods from a single MCMC output. Statistica Neerlandica, 59(1), 16–29.

6.Cho, S.-J., Athay, M., & Preacher, K. J. (2013). Measuring change for a multidimensional test using a generalized explanatory longitudinal item response model. British Journal of Mathematical and Statistical Psychology, 66(2), 353–381.

7.Crocker, L., & Algina, J. (1986). Introduction to classical and modern test theory. Chicago: Holt Rinehart and Winston.

8.Crocker, L., & Algina, J. (2006). Introduction to classical and modern test theory. Ohio, Maison: Cengage Learning

9.Cronbach, L. J. (1990). Essentials of psychological testing (5. ed.). New York, NY: Harper & Row Publishers Inc.

10.Ebel, R.L. and Frisbie, D.A. (1991) Essentials of Educational Measurement. 5th Edition, Prentice-Hall, Englewood Cliffs.

11.Fernandez, C., Ley, E., and Steel, M. F. (2001). Benchmark priors for Bayesian model averaging. Journal of Econometrics, 100(2), 381–427.

12.Fox, j. p. & Glas, C.A.W. (2003). Bayesian modeling of measurement error in predictor variables using item response theory.Psychometrika, 68, 169–191.

13.Fox, j.-p. (2005). Multilevel IRT using dichotomous and polytomous response datd.British Journal of Mathematical and Statistical Psychology, 58(1), 145–172.

14.Geiser, C., Bishop, J., Lockhart, G., Shiffman, S., & Grenard, J. L. (2013). Analyzing latent state-trait and multiple-indicator latent growth curve models as multilevel structural equation models. Frontiers in Psychology, 4(975), 1–23.

15.Gilks W. R., Richardson S. & Spiegelhalter D. J. (1996), Markov chain Monte Carlo in Practices, Chapman and Hall, London.

16.Grier, J.B. (1975). The number of alternatives for optimum test reliability. Journal of Educational Measurement, 12 (2), 109-112.

17.Haladyna, T. M. (2004). Developing and validating multiple-choice test items (3. ed.). New Jersey, NJ: Lawrence Erlbaum Associates Publishers.

18.Hambleton, R. K. (1989). Principles and selected applications of item response theory. In R. L. Linn (Ed.), Educational measurement (pp. 147–200). Macmillan Publishing Co, Inc; American Council on Education
19.Hambleton, j. D. (1989). "A New Approach to the Economic Analysis of Nonstationary Tim Series and the Business Cycle". Educational Measurement: Issues and Practice, (28), 357-384.
20.Hambleton, R. K., & Jones, R. W. (1993). "Comparison of Classical Test Theory and Item Response Theory and their Applications to Test Development". Educational Measurement: Issues and Practice, 12(3), 38-47.
21.Hambleton, R.K., Swaminathan, H. and Rogers, H.J. (1991) Fundamentals of Item Response Theory. Sage, Newbury Park, CA.
22.Harlow, L.L., Mulaik, S., & Steiger, J. (Eds.) (1997). What if there were no significance tests? Mahwah, NJ: Erlbaum.
23.Harris, D. (1989). Comparison of 1-, 2-, and 3-parameter IRT models. Educational Measurement: Issues and Practice, 8(1), 35-41.
24.Hsieh, C.-A., von Eye, A., Maier, K., Hsieh, H.-J., & Chen, S.-H. (2013). A unified latent growth curve model. Structural Equation Modeling: A Multidisciplinary Journal, 20(4), 592–615.
25.Karabatsos, G. (2016). Bayesian nonparametric response models. In Handbook of Item Response Theory Review by: Robert L. McKinley. Journal of Educational Measurement, 24(2), 182-184.
26.Kass, R. E., & Raftery, A. E. (1995). Bayes factors. Journal of the American Statistical Association, 90, 773-795.
27.Kim, S. & Camilli, G. (2014). An item response theory approach to longitudinal analysis with application to summer setback in preschool language/literacy. Large-scale Assessments in Education, 2(1), 1.
28.Lane, S., Raymond, M. R., & Haladyna, T. M. (2016). *Handbook of test development* (2. ed.). New York, NY: Routledge.
29.Li, Y. and Clyde, M. A. (2018). Mixtures of g-priors in generalized linear models. Journal of the American Statistical Association, 113(524), 1828–1845.
30.Liang, F., Paulo, R., Molina, G., Clyde, M. A., and Berger, J. O. (2008). Mixtures of g priors for Bayesian variable selection. Journal of the American Statistical Association, 103(481), 410–423.
31.Liu,Yang.(2019).Bayesian Item Response Theory: Methods and Applications. Doctoral Dissertation. The University of Connecticut Graduate School.
32.McDonald, R. P. (1999). Test theory: A unified treatment. Mahwah, NJ: Lawrence Erlbaum.

33. Nishio, M; Akasaka, T.; Sakamoto, R. & Togashi, K. (2020). Bayesian Statistical Model of Item Response Theory in Observer Studies of Radiologists. Acad Radiol, 27(3):e45-e54.

34.Rasch, G. (1960). Probabilistic model for some intelligence and achievement tests. Copenhagen: Danish Institute for Educational Research

35.Raudenbush, S. W. and Bryk, A. S. (2002). Hierarchical linear models: Applications and data analysis methods, volume 1. Sage.

36.Rodrigues, A., Chaves, L. M., Silva, F. F., Garcia, I. P., Duarte, D. A. S., and Ventura, H. T. (2018). Isotonic regression analysis of Guzer´a cattle growth curves. Revista Ceres, 65(1), 24–27.

37.Sahin, M. G., Yildirim, Y., Ozturk, N. B. (2023). Examining the Achievement Test Development Process in the Educational Studies. Participatory Educational Research (PER) Vol.10(1), pp. 251-274.

38.Turgut, F. (1992). Eğitimde ölçme ve değerlendirme [Measurement and assessment in education] (8. ed.). Ankara: Saydam Publ.

39.Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 64(4), 583–639.

40.Özçelik, D. A. (1992). Ölçme ve değerlendirme [Measurement and assessment]. Ankara: ÖSYM Publ.

41.Wang, X., Berger, J. O., and Burdick, D. S. (2013). Bayesian analysis of dynamic item response models in educational testing, Annals of Applied Statistics, 7(1), 126-153.

42.Wu, H.-H., Ferreira, M. A., Gompper, M. E., & et al. (2016). Consistency of hyperg-prior-based Bayesian variable selection for generalized linear models. Brazilian Journal of Probability and Statistics, 30(4), 691–709.