الأداء التفاضلي لاختبار تحصيلي وفق طريقتي تقدير الدرجات (التامة، والجزئية) باستخدام نموذج التقدير الجزئي لدى عينة من طلبة جامعة إربد الأهلية

د. رياض أحمد الدرابسه[i]

تاريخ الاستلام 2024/1/24
تاريخ القبول 2024/5/14

الملخص:

هدفت الدراسة إلى الكشف عن الأداء التفاضلي لاختبار تحصيلي في مادة القياس والتقويم باستخدام نموذج التقدير الجزئي لدى عينة من طلبة جامعة إربد الأهلية، وتكونت عينة الدراسة من (120) طالبًا وطالبة. وقد أظهرت نتائج الدراسة وجود مطابقة جيدة للبيانات لافتراضات نموذج التقدير الجزئي في ضوء طريقتي تقدير الدرجات (التامة–الجزئية)؛ وأن الطريقة الجزئية في تقدير الدرجات أكثر دقة من الطريقة التامة في تقدير الدرجات في تقدير معالم قدرات الطلبة، وأشارت النتائج إلى تفوق الذكور على الإناث في الفقرات (1، 9) باستخدام طريقة مانتل-هانزل، وتفوقت الإناث على الذكور في الفقرة (1) باستخدام طريقة الانحدار اللوجستي للأداء التفاضلي المنتظم، بينما تفوق الذكور على الإناث في الفقرة (20) في نفس الطريقة. بينما تفوق الذكور على الإناث في الفقرات (5، 12، 29) باستخدام طريقة الانحدار اللوجستي للأداء التفاضلي غير المنتظم.

الكلمات المفتاحية: الأداء التفاضلي، اختبار تحصيلي، طريقة تقدير الدرجات (التامة، والجزئية)، نموذج التقدير الجزئي، طلبة جامعة إربد الأهلية.

[i] جامعة إربد الأهلية

Item Differential Functioning of an Achievement Test According to the Full and Partial Grades Estimation Methods Using the Partial Credit Model Among Irbid Private University Students

Abstract

The current study aimed to reveal the item differential performance of an achievement test measurement and evaluation using the partial estimation model among Irbid private university students. The study sample consisted of (120) students. The results showed that the data the assumptions of the partial credit model and the two methods of estimating grades (full and partial). The partial method of estimating grades was more accurate in estimating the parameters of students' abilities. The results also indicated that males outperformed females in item (1, 9) using the Mantel-Haenszel method, and females outperformed males in item (1) using the logistic regression method for uniform differential performance, while males outperformed females in item (20) using the same method. While males outperformed females in item (5, 12, 29) using the logistic regression method for irregular differential performance.

Keywords: Item Differential Functioning, Achievement Test, Full and Partial Grades Estimation Methods, Partial Credit Model, Irbid National University Students.

المقدمة

تتضمن عملية التعليم والتعلم في المؤسسات التعليمية دائمًا إجراء التقييمات بشكل مستمر ودوري باعتباره أمرًا مهمًا للغاية يجب القيام به، فالتقييم التعليمي أوسع نطاقًا من الاختبارات، فهو يركز على نقاط الضعف والقوة لدى الأفراد ويحددها؛ والاختبار هو إجراء قياس يقوم به المدرس لمعارف ومهارات الطلبة لمعرفة مدى تحقيق الأهداف باستخدام أدوات معينة. وتُعد الاختبارات التعليمية إحدى أهم أدوات القياس التعليمي والتربوي، الهادفة إلى تزويد المدرسين ببيانات لاتخاذ قرارات مختلفة تتعلق بانتقاء الأفراد وتصنيفهم، وبوجه عام فإن درجة دقة القرارات المستندة إلى درجات الطلبة على الاختبارات تعتمد على عدد من العوامل، من بينها جودة الاختبار ودقة وشمولية المعلومات التي يعطيها وصحة طريقة تحليل البيانات وتفسيرها.

وتقع على المتخصصين في الاختبارات التعليمية مسؤولية معالجة هذه المخاوف من خلال تحسين جودة التقييمات التعليمية، وجعلها عادلة ومحايدة قدر الإمكان؛ لتقليل ومحاولة القضاء على فجوة درجات الاختبار التي تظهر باستمرار بين الطلبة مرتفعي ومنخفضي التحصيل، في حين أننا لا نستطيع القضاء بشكل كامل على الفوارق في مجموعات الأفراد التي تقدمت للاختبار والتي هي نتيجة لفروق عدم ضبط المساواة بينهم، بل يمكننا الاستفادة من الأدوات الإحصائية والقياسية النفسية لتحديد فقرات الاختبار التي تسهم في المشكلة وإزالتها. ولحسن الحظ، توجد مثل هذه الأدوات، وأحدها هو تحليل الأداء التفاضلي للفقرة (DIF) (Lucey and Saguil, 2020).

يُعد الأداء التفاضلي للفقرة خاصية إحصائية توضح مدى قدرة الفقرة على قياس القدرات المختلفة لأعضاء مجموعات فرعية منفصلة، تتم مقارنة متوسط درجات الفقرات للمجموعة الفرعية التي لها نفس الدرجة الإجمالية في الاختبار؛ لتحديد ما إذا كان يتم قياس الفقرة بنفس الطريقة بشكل أساسي لجميع المجموعات الفرعية، ويتطلب وجود الأداء التفاضلي المراجعة والحكم، ولا يدل بالضرورة إلى وجود تحيز (Holland & Wainer, 2012). ويوفر تحليل الأداء التفاضلي إشارة إلى السلوك غير المتوقع للفقرات في الاختبار، ولا يعرض الأداء التفاضلي للفقرة إذا كان لدى الأفراد من مجموعات مختلفة احتمالية مختلفة لإعطاء استجابة معينة، فإن الأداء التفاضلي يعرض إذا كان لدى الأفراد من مجموعات مختلفة لديهم نفس القدرة الحقيقية احتمالية مختلفة لاستجابة معينة (Wells & Faulkner-Bond, 2016).

ومن الإجراءات الشائعة لتقييم الأداء التفاضلي: طريقة مانتل – هانزل Mantel-Haenszel، وطريقة الانحدار اللوجستي logistic regression، والأساليب المستندة إلى نظرية استجابة للفقرة IRT (Wells & Faulkner-Bond, 2016). حيث تُعدُّ طريقة مانتل–هانزل Mantel-Haenszel الإجراء الأكثر استخدامًا للكشف عن الأداء التفاضلي، وهي عبارة عن نهج قائم على إحصائي مربع كاي chi-squared الذي يدرس الاختلافات بين المجموعة المرجعية والمجموعة المستهدفة لجميع فقرات الاختبار؛ ويتم تقسيم متصل القدرة التي تحددها مجموع درجات الاختبار إلى فترات زمنية (k) والتي تعمل بعد ذلك كأساس لمطابقة أعضاء كلا المجموعتين؛ ويستخدم جدول (2 x 2) في كل فترة زمنية (k) لمقارنة المجموعتين على كل فقرة. وتتوافق صفوف جدول الاحتمالات مع المجموعة (المرجعية أو المستهدفة) بينما تتوافق الأعمدة مع الإجابات الصحيحة أو غير الصحيحة (Wainer & Braun, 2013).

وتُعدُّ نظرية الاستجابة للفقرة (IRT) هي طريقة أخرى مستخدمة على نطاق واسع لتقييم الأداء التفاضلي، حيث تسمح هذه النظرية بإجراء فحص نقدي للاستجابات لفقرات معينة من اختبار أو قياس ما، ولأن الأداء التفاضلي يقوم بفحص احتمالية الاستجابة بشكل صحيح أو تأييد فقرة مشروطة بالسمات الكامنة؛ ولأن نظرية الاستجابة للفقرة تفحص العلاقة بين الاستجابات والسمات الكامنة، فهناك ثلاث مزايا رئيسة لاستخدام نظرية الاستجابة للفقرة في كشف الأداء التفاضلي هي: 1. لا يتم الخلط بين تقديرات معلمات نظرية الاستجابة للفقرة وخصائص العينة بالمقارنة مع نظرية الاختبار الكلاسيكية. 2. يمكن التعبير عن الخصائص الإحصائية للفقرات بدقة أكبر مما يزيد من دقة تفسير الأداء التفاضلي بين مجموعتين. 3. يمكن التعبير عن هذه الخصائص الإحصائية للفقرات بيانيًا؛ مما يُحسن إمكانية التفسير وفهم كيفية عمل الفقرات بشكل مختلف بين المجموعات (Linden, 2017).

كما يُعد اختبار نسبة الاحتمالية طريقة أخرى تعتمد على نظرية الاستجابة للفقرة لتقييم الأداء التفاضلي، إذ يتضمن هذا الإجراء مقارنة نسبة نموذجين، في ظل النموذج الأول تكون معلمات الفقرة مقيدة بأن تكون متساوية أو ثابتة بين المجموعتين المرجعية والمستهدفة، بينما في ظل النموذج الثاني تكون معلمات الفقرة قابلة للتغيير؛ إذ تعمل الفقرات المقيدة لتكون متساوية كفقرات أساسية لهذا الإجراء بينما يسمح للفقرات المشتبه بها في الأداء التفاضلي باختلاف بحرية، وباستخدام الفقرات الأساسية والسماح لمعلمات الفقرات المتبقية بالاختلاف يمكن تقنين فقرات متعددة في وقت واحد للأداء التفاضلي (Reise & Revicki, 2015).

ويُعدُّ الانحدار اللوجستي طريقة أخرى لتقييم الأداء التفاضلي، إذ تتضمن أساليب الانحدار إجراء تحليل منفصل لكل فقرة، فالمتغيرات المستقلة المدرجة في التحليل هي المجموعة، ومتغير مطابقة القدرة عادة ما يكون النتيجة الإجمالية، فالمتغير التابع للفحص هو احتمال أو احتمالية الحصول على إجابة صحيحة أو تأييد فقرة ما؛ ونظرًا لأن نتيجة الفحص يتم التعبير عنها من حيث الاحتمالات فإن تقدير أقصى احتمال هو الإجراء المناسب، ويمكن التعبير عن هذه المجموعة من المتغيرات بمعادلة الانحدار الآتية:

$$Y = \beta_0 + \beta_1 M + \beta_2 G + \beta_3 MG$$

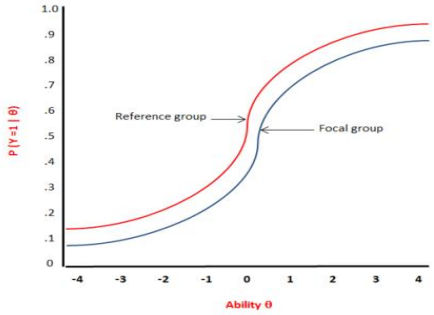
حيث (β0) يتوافق مع التقاطع أو احتمال الاستجابة عندما تكون (M & G) متساوية لـ (0) مع (βs) المتبقية، حيث يشير الحرف (M) إلى المتغير المستقل (متغير المطابقة)، وهو المستخدم لربط الأفراد من حيث القدرة؛ بينما يشير (G) إلى المجموعة (متغير التجميع)، ويشير (MG) إلى التفاعل بين المتغيرين المذكورين (متغير التفاعل) (Lane, Raymond & Haladyna, 2016).

يتم تحديد الأداء التفاضلي من خلال تقييم إحصائية مربع كاي chi-squared التي تم الحصول عليها بدرجتين من الحرية، ويتم اختبار أهمية تقدير المعلمة من نتائج الانحدار اللوجستي، ويتم الإشارة إلى الأداء التفاضلي إذا كان الأفراد المتطابقون في القدرة لديهم احتمالات مختلفة تمامًا للاستجابة لفقرة ما وبالتالي اختلاف منحنيات الانحدار اللوجستي، وعلى العكس من ذلك إذا كانت المنحنيات لكلا المجموعتين هي نفسها، فإن الفقرة غير متحيزة وبالتالي لا يوجد أداء تفاضلي (Holland & Wainer, 2012). ويشير الأداء التفاضلي إلى الاختلاف في أداء الفقرات عبر المجموعات، فمن المهم ملاحظة أنه عند فحص الفقرات

الخاصة بالأداء التفاضلي يجب أن تكون المجموعات متطابقة على السمة المقاسة، وإلا فقد يؤدي ذلك إلى اكتشاف غير دقيق للأداء التفاضلي، في كثير من الأحيان الديموغرافية تتطابق مع السمة الكامنة أو بشكل عام السمة التي يتم قياسها بواسطة الفقرات أو الاختبار (Matsumoto & Vijver, 2011).

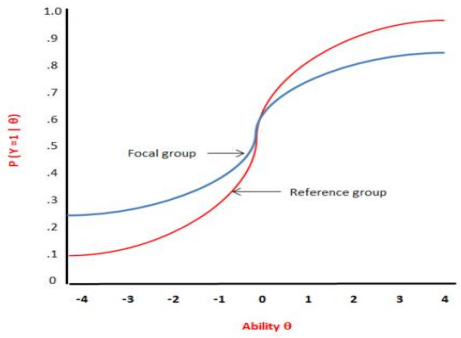
يُعدُّ الأداء التفاضلي المنتظم أبسط أنواع الأداء التفاضلي، حيث يكون حجم التبعية المشروطة ثابتًا نسبيًا عبر سلسلة السمات الكامنة (θ)، وتمنح الفقرة محل الاهتمام مجموعة واحدة باستمرار ميزة عبر جميع مستويات القدرة؛ في إطار نظرية الاستجابة للفقرة يمكن للسمات الكامنة إثبات ذلك عندما تكون منحنيات خصائص الفقرة (ICC) متساوية في التمييز، ولكن تظهر اختلافات في معلمات الصعوبة أي (ar = af and br < bf) كما هو موضوح بالشكل (1).

الشكل (1): منحنيات خصائص الفقرة للأداء التفاضلي المنتظم



ويقدم الأداء التفاضلي غير المنتظم حالة مثيرة للاهتمام، بدلًا من إعطاء ميزة متسقة للمجموعة المرجعية عبر مستويات القدرة المختلفة، تتحرك التبعية المشروطة وتغير اتجاهها في مواقع مختلفة على متصل السمة (θ)، وقد تمنح إحدى فقرات المجموعة المرجعية ميزة بسيطة في الطرف الأدنى من السلسلة بينما تمنح ميزة كبيرة في الطرف الأعلى، على العكس من الأداء التفاضلي المنتظم يمكن أن تختلف الفقرة في نفس الوقت في التمييز بين المجموعتين، بينما تختلف أيضًا في الصعوبة؛ على سبيل المثال (ar ≠ af and br < bf)، والأكثر تعقيدًا هو الأداء التفاضلي المتقاطع (الأداء التفاضلي غير المنتظم) كما هو في الشكل (2).

الشكل (2): منحنيات خصائص الفقرة للأداء التفاضلي غير المنتظم



يحدث هذا عندما تعطي فقرة ما ميزة لمجموعة مرجعية في أحد طرفي سلسلة السمات الكامنة (θ)، بينما يفضل المجموعة المرجعية في الطرف الآخر (Wells, 2021). ومن ناحية أخرى، إذا كان من الممكن أن يعزى الأداء التفاضلي لمجموعتين إلى اختلاف حقيقي في مستويات قدراتهما، فإن هذا يسمى التأثير وليس التحيز (Matsumoto & Vijver, 2011).
إن الحكم المنطقي له أهمية حاسمة عند تقييم الأداء التفاضلي للفقرات، وقد يتم الحصول على نتائج مختلفة عند الكشف عن الأداء التفاضلي اعتمادًا على إجراء إحصائي معين، فبعض الإجراءات أكثر دقة وبعضها الآخر أقل دقة؛ فيتطلب إجراء طريقة مانتل -هانزل من الباحث بناء مستويات القدرة بناءً على إجمالي درجات الاختبار، بينما تضع نظرية الاستجابة للفقرة الأفراد بشكل أكثر فعالية على طول متصل السمة الكامنة، وبالتالي يشير أحد الإجراءات إلى الأداء التفاضلي لفقرة معينة بينما لا يشير الآخر إلى ذلك (Lucey & Saguil, 2020).
يُعد تحليل الأداء التفاضلي للفقرة أسلوبًا مقبولاً من الناحية النفسية أكثر من مجرد استبعاد الفقرات التي تظهر فروقًا إجمالية كبيرة بين مجموعتين. بدلاً من ذلك، يقارن الأداء التفاضلي للفقرة الأداء في فقرة اختبار بين مجموعة مرجعية (أغلبية) ومجموعة محورية (مستهدفة)، مع التحكم في الأداء في الاختبار ككل (تقدير للفرق الحقيقي في القدرة ذات الصلة بالبناء)، تم استخدام تحليل الأداء التفاضلي للفقرة بنجاح للكشف عن إشكالية الفقرات في التقييمات واسعة النطاق مثل الاختبارات التعليمية، وقد تمت الدعوة لاستخدامه في تطوير واختبار التقييمات عالية المخاطر (Martinková et al., 2017). وإن تحليل الأداء التفاضلي للفقرة ليس سوى أحد الأساليب التي يجب على صناعة الاختبار استخدامها لتقليل الاختلافات الديموغرافية في التقييمات التعليمية عالية المخاطر (Koljatic, Silva & Sireci, 2021).

نموذج التقدير الجزئي

يعدُّ نموذج التقدير الجزئي من أكثر النماذج شيوعًا في بناء البيانات التي تتناسب مع نموذج القياس، ويحدد نموذج التقدير الجزئي أن كل فقرة لها هيكل أو تصنيف خاص بها. وهو مشتق من اختبارات الاختيار من متعدد حيث يتم منح الإجابات غير الصحيحة، رصيدًا جزئيًا للإجابة الصحيحة، ويختلف مقدار الصحة الجزئية عبر الفقرات (Kolen & Brennan, 2013).
ويسمح هذا النموذج بتحديد معلمات إضافية وإدخالها في التقدير؛ لأن المزيد من المعلمات يعني ملاءمة أفضل لبيانات النموذج، إذ إن معلمات نموذج التقدير الجزئي يمكن تفسيرها على أنها معلمات صعوبة للخطوات المعنية (Fischer & Molenaar, 2013).
ويشير فيشر ومولينار (Fischer & Molenaar, 2013) إلى نموذج التقدير الجزئي بأنه حالة خاصة من نموذج قياس راش، حيث يمتاز بعدم التحيز في عملية القياس، والقدرة على تسليط الضوء على تلك الفقرات التي تحتاج إلى تحسين، وتوفير معلومات عن التطابق بين البيانات والنموذج، والتغذية الراجعة لفئات الاستجابة؛ وهو وسيلة لتطوير التقييمات القائمة على الأداء، وذلك بسبب قدرة النموذج على استيعاب مجموعة متنوعة من إجراءات التصحيح.
ويعدُّ نموذج راش (Rasch, 1960) أحد أشهر نماذج نظرية الاستجابة للفقرة الذي يناسب الفقرات الثنائية فقط، يتم تمييز الفقرات من خلال صعوبتها التي يتم تعريفها على أنها السمة الكامنة للفرد الذي لديه احتمالية الاستجابة نفسها لكل من الإجابتين المقترحتين عن الفقرة (Wati, Mahtari, Hartini & Amelia, 2019).

تعدُّ نماذج نظرية الاستجابة للفقرة (IRT) هذه مناسبة بشكل خاص لإجراء قياسات على أساس السمات الكامنة للأفراد، ولدراسة تأثير المتغيرات المشتركة المرتبطة بها على هذه المتغيرات الكامنة. تتكون هذه النماذج من نماذج مختلطة: تعدُّ السمة الكامنة متغيرًا عشوائيًا، وتعتبر صعوبات الفقرة والمتغيرات المشتركة المحتملة متغيرات ثابتة. يتم بعد ذلك تقدير المعلمات بشكل كلاسيكي باستخدام هامش الأرجحية القصوى (Marginal MaximumLikelihood)، ويتزايد الاهتمام بتنفيذ نماذج نظرية الاستجابة للفقرة لإرشاد تصميم الأدوات والتحقق من صحتها عبر السياقات الاجتماعية والتعليمية (French & Vo, 2020).

يمكن استخدام نموذج التقدير الجزئي حتى إذا اختلف عدد فئات الاستجابة حسب الفقرات، ويتميز المجيبون بسماتهم الكامنة (أي قيمة المستجيبون للمتغير الكامن محل الاهتمام)، وتتميز الفقرات بالصعوبات المرتبطة بكل فئة من فئات الاستجابة الخاصة بها. ويُعرف نموذج التقدير الجزئي بأنه أحد نماذج نظرية الاستجابة للفقرة، ويستخدم ضمن معايرة وتدريج فقرات المقاييس والاختبارات أحادية البعد ومتعددة الاستجابة، ويسمح باختلاف الفقرات في معالم صعوبتها فقط، ويفترض تساوي الفقرات في تمييزها بين المفحوصين عند مستويات القدرة المختلفة (Abdelwahab, 2019).

يقدم نموذج القياس الجيد أكبر قدر ممكن من المعلومات النوعية، ودقة تفسير الأداة وفقًا لوظيفتها، ويوفر قياسًا خطيًا، ويمكنه التغلب على البيانات المفقودة، والعثور على بيانات غير صحيحة (غير ملائمة) أو غير شائعة (القيم المتطرفة)؛ ويمكن لنموذج التقدير الجزئي الخاص بـ Rasch تحديد جودة الاختبار بشكل أكثر دقة من النماذج الأخرى، حيث تتم الإشارة إلى قيمة الصدق في نموج راش من خلال مؤشرات الانفصال (فصل الأفراد) وفصل الفقرات، وإحصائيات ملخص المخرجات للحصول على معلومات حول صدق الفرد والفقرة ومعامل كرونباخ ألفا، ينتج التحليل باستخدام نموذج راش تحليلاً إحصائيًا للملاءمة (إحصاءات الملاءمة) الذي يوفر معلومات توضح البيانات التي تم الحصول عليها بشكل جيد أن الأفراد الذين لديهم قدرة عالية يقدمون أنماطًا من الإجابات عن الفقرات وفقًا لمستوى الصعوبة الخاص بهم (Wahyuningsih, 2020).

الدراسات السابقة

تتناول هذه الجزئية من الدراسة عرضًا للدراسات السابقة ذات الصلة بموضوع الأداء التفاضلي للفقرات وفق طريقتي تقدير الدرجات التامة والجزئية باستخدام نموذج التقدير الجزئي في ضوء متغير جنس الطالب، وسوف يتم عرض هذه الدراسات وفق الآتي:

أجرى خليل وإسماعيل (2021) دراسة هدفت لمعرفة الأداء التفاضلي لاختبار تحصيلي إلكتروني لدى عينة من طلبة الجامعة بالرستاق باستخدام نموذج التقدير الجزئي، وتكونت عينة الدراسة من (107) طالبًا وطالبة. أظهرت نتائج الدراسة أن جميع تقديرات صعوبة فقرات الاختبار التحصيلي الالكتروني باستخدام الطريقة التامة لتقدير الدرجات وفق نموذج التقدير الجزئي جاءت ضمن المدى المقبول، وأن جميع قيم مربع كاي لفقرات الاختبار غير دالة إحصائيًا؛ مما يدل على مطابقة البيانات لافتراضات نموذج التقدير الجزئي باستخدام الطريقة التامة لتقدير الدرجات. كما بينت النتائج عدم وجود فروق ذات دلالة إحصائية بين الجنسين في تقدير الأداء التفاضلي لفقرات الاختبار باستخدام طريقة مانتل–هانزل، وعدم وجود فروق ذات

دلالة إحصائية بين الجنسين في تقدير الأداء التفاضلي المنتظم وغير المنتظم لفقرات الاختبار باستخدام طريقة الانحدار اللوجستي.

وهدفت دراسة أبو مسرة والسوالمة (2021) إلى استخدام طريقة أشجار راش للكشف عن الأداء التفاضلي لفقرات الاختبار الوطني لضبط نوعية التعليم في الرياضيات للصف الرابع الأساسي في الأردن، وتكونت عينة الدراسة من (1500) طالبٍ وطالبة. بينت النتائج أن هناك أداءً تفاضليًا وفقًا لمتغير الجنس فقط؛ إذ تم الكشف عن (8) فقرات ذات أداء تفاضلي، أي ما نسبته (32%) من الفقرات. وكان الأداء التفاضلي لصالح الذكور في (5) فقرات، ولصالح الإناث في (3) فقرات. وأظهرت النتائج حساسية طريقة أشجار راش لحجم العينة؛ إذ زادت فرص وجود أداء تفاضلي للفقرات بازدياد حجم العينة.

وقام شانموغام (Shanmugam, 2020) بدراسة هدفت إلى التعرف على الفروق بين الجنسين في تقدير الأداء التفاضلي لفقرات الاختبار باستخدام طريقة مانتل – هانزل في جزيرة بينالغ، وتكونت عينة الدراسة من (1381) طالبًا وطالبة. أظهرت نتائج الدراسة أنه يوجد فروق دالة إحصائيًا في الأداء التفاضلي لصالح الذكور في الفقرات التي تعتمد على مهارات التفكير العليا، وأظهرت فقرتين أداءً تفاضليا معتدلاً، وأظهرت (5) فقرات أظهرت أداءً تفاضليا كبيرا.

وأجرى الشريفين (2018) دراسة هدفت للكشف عن أثر نوع الأداء التفاضلي للفقرات (منتظم، غير منتظم) على الخصائص السيكومترية للفقرات، وتكونت عينة الدراسة من (1000) فرد. أشارت النتائج إلى وجود فرق دال إحصائيًا بين الوسطين لكل من معالم التمييز، ومعالم التخمين لفقرات الاختبار يُعزى لنموذج نظرية الاستجابة للفقرة المستخدم لصالح النموذج المعلمي مقارناً بالنموذج اللامعلمي، ووجود فرق دال إحصائيًا بين الوسطين الحسابيين لمعالم الصعوبة يُعزى للنموذج المستخدم لصالح النموذج اللامعلمي مقارنًا بالنموذج المعلمي، ووجود فروق دالة إحصائيًا بين الأوساط الحسابية لكل من الصعوبة والتخمين لفقرات الاختبار تُعزى للتفاعل بين النموذج المستخدم ونوع الأداء التفاضلي؛ كما أشارت النتائج إلى وجود فرق دال إحصائيًا بين الوسطين الحسابيين لكمية معلومات الاختبار يُعزى للنموذج لصالح النموذج اللامعلمي مقارنة بالنموذج المعلمي، حيث كانت كمية المعلومات التي يقدمها الاختبار أكبر عند جميع مستويات القدرة وفق النموذج اللامعلمي سواء كان الأداء التفاضلي منتظمًا أم غير منتظم، باستثناء أن النموذج المعلمي الثلاثي قدم معلومات أكبر من النموذج اللامعلمي عند مستويات القدرة المرتفعة؛ وأن قيم معاملات الثبات المقدر وفق النموذجين كانت عالية، وأن قيم معاملات الثبات المقدرة وفق النموذج المعلمي كانت الأعلى سواء أكان الأداء التفاضلي للفقرات منتظما أم غير منتظم.

وهدفت دراسة أحمد (2017) للكشف عن الأداء التفاضلي لفقرات اختبار تحصيلي مرجعي المحك وفقًا لنظريتي القياس التقليدية والاستجابة للفقرة في ضوء بعض المتغيرات، وتكونت عينة الدراسة من (1279) طالبًا وطالبة بمحافظة الشرقية. أظهرت نتائج الدراسة أن طريقتي صعوبة الفقرة المحولة والتقنين المعياري لم تكشف عن أي فقرة ذات أداء تفاضلي بالنسبة لجميع أحجام العينات (المرجعية/المستهدفة)، وأن طريقة مانتل -هانزل أكثر حساسية من الطريقتين السابقتين في الكشف عن الأداء التفاضلي المنتظم للفقرة لجميع متغيرات البحث. وأشارت النتائج لعدم اختلاف في الأداء التفاضلي المنتظم للفقرة باختلاف حجم العينة، ونوع

المدرسة، وترتيب الفقرات باستخدام جميع طرق الكشف وفقًا لنظريتي القياس التقليدية والاستجابة للفقرة.

وقام كاراكايا (Karakaya, 2012) بدراسة هدفت للكشف عن الفقرات التي تظهر أداءً تفاضليًا في ضوء متغير الجنس لاختبارات العلوم والتكنولوجيا والرياضيات، وتكونت عينة الدراسة من (22624) طالبًا وطالبة. أظهرت نتائج الدراسة أنه يوجد (3) فقرات ذات أداء تفاضلي لصالح الإناث، و(2) لصالح الذكور في اختبار العلوم والتكنولوجيا؛ و(3) فقرات لصالح الذكور، وفقرة واحدة لصالح الإناث في اختبار الرياضيات، وأن نماذج الاستجابة للفقرة ثنائية الاستجابة تعطي نتائج جيدة لدالة المعلومات لفقرات الاختبارات.

مشكلة الدراسة وأسئلتها

أصبح استخدام التقييمات التعليمية ذات القرارات الهامة أمرًا صعبًا، حيث تجمعت الانتقادات الموجهة للاختبار الموحد حول قضية عدم المساواة في الاختبار، أي أن التقييمات التعليمية المستخدمة حول العالم تميز بشكل منهجي ضد بعض المجموعات من الطلبة. وتقوم الإحصائيات بدور حيوي، ولكن لا ينبغي بأي حال من الأحوال أن تكون الأساس الوحيد للقرارات والاستنتاجات التي يتم التوصل إليها. كما تُعد أنماط الاستجابة مشكلة مهمة في القياس النفسي وبيانات المسح، والمشكلة في أنماط الاستجابة هي أنها يمكن أن تؤثر في صحة درجات المقياس؛ لأن تقديرات السمة الموضوعية قد تكون متحيزة إذا تم تجاهل أسلوب الاستجابة في الاختبارات متعددة البدائل. النماذج التي تراعي أنماط الاستجابة بشكل واضح قادرة على تقليل التحيز.

يُعدُّ الأداء التفاضلي من أكثر الأمور التي تهدد صدق البناء والثبات للاختبار، بالإضافة إلى تأثيره في دقة تقدير معالم الفقرات، ودقة تقديرات القدرة للأفراد؛ مما يؤدي إلى عدم وضوح النتائج، ويؤدي بالتالي إلى اتخاذ قرارات خاطئة تبعًا لمعلومات غير دقيقة. ويعتمد تفسير درجات الاختبار وعمل المقارنات العادلة بين المجموعات أو الأفراد على توفر تقييمًا لثبات القياس بين المتقدمين للاختبار عبر المجموعات المختلفة والذي يتم من خلال دراسة الأداء التفاضلي لفقراته لضمان الفائدة المرجوة من العدالة في الدرجات. وقد تعمل هذه المتغيرات كمصادر محتملة للأداء التفاضلي في فقرات الاختبار. يُوصى غالبًا بفحص جميع فقرات الاختبار الخاصة بالأداء التفاضلي في وقت واحد، حيث سيوفر هذا معلومات حول جميع الفقرات، ويسلط الضوء على إشكالية الفقرات بالإضافة إلى تلك التي تعمل بشكل مماثل لكل من المجموعة المرجعية والمجموعة المستهدفة؛ وانطلاقًا مما سبق تمثلت مشكلة الدراسة في الكشف الأداء التفاضلي لاختبار تحصيلي وفق طريقتي تقدير الدرجات (التامة، والجزئية) باستخدام نموذج التقدير الجزئي لدى عينة من طلبة جامعة إربد الأهلية، وذلك من خلال الإجابة عن الأسئلة الآتية:

1. هل تختلف تقديرات قدرات الأفراد وأخطاؤها المعيارية باختلاف طريقة تقدير الدرجات؟
2. هل يختلف تقدير الأداء التفاضلي للاختبار التحصيلي لدى عينة الدراسة باختلاف الجنس وفق الطريقة التامة لتقدير الدرجات باستخدام نموذج التقدير الجزئي؟
3. ما النموذج الأفضل لتقييم كفاءة المعلومات في الاختبار التحصيلي وفق طريقتي تقدير الدرجات (التامة، والجزئية) باستخدام نموذج التقدير الجزئي؟

أهداف الدراسة

تهدف الدراسة الحالية إلى تحقيق الأهداف الآتية:

- التعرف إلى مدى اختلاف تقديرات قدرات الأفراد وأخطاؤها المعيارية باختلاف طريقة تقدير الدرجات التامة والجزئية باستخدام نموذج التقدير الجزئي.
- التعرف إلى مدى اختلاف تقدير الأداء التفاضلي للاختبار التحصيلي باختلاف جنس الطالب وفق الطريقة التامة لتقدير الدرجات باستخدام نموذج التقدير الجزئي.
- الكشف عن أفضل نموذج للحكم على كفاءة المعلومات في الاختبار التحصيلي وفق طريقتي تقدير الدرجات التامة والجزئية باستخدام نموذج التقدير الجزئي.

أهمية الدراسة

تبرز الأهمية النظرية للدراسة من خلال النظرة للأداء التفاضلي باعتباره أحد الموضوعات المهمة في مجال القياس النفسي بشكل خاص، والقياس والتقويم بشكل عام؛ حيث تبرز أهميتها في معالجتها للأداء التفاضلي لفقرات اختبار تحصيلي والتقييم باستخدام نموذج التقدير الجزئي لدى عينة من طلبة الجامعات. وهو من المواضيع النفسية الحديثة التي لم يتم الطرق لها من قبل في الدراسات السابقة (بحسب علم الباحث)، كما تأتي أهمية الدراسة من خلال تقديم دراسة نفسية حديثة في مجال القياس والتقويم تقوم على الكشف عن الأداء التفاضلي لفقرات الاختبارات التحصيلية باستخدام نموذج التقدير الجزئي تساعد في حل المشكلات التي تهدد صدق وثبات الاختبارات والقرارات المترتبة عليها، ويساعد في تقديم تعديلات على الاختبار التحصيلي بما يجعله خاليًا من الفقرات ذات الأداء التفاضلي.

وتأتي الأهمية العلمية في معرفة وفهم الأداء التفاضلي للفقرة بشكل أفضل، ومساعدة المهتمين في الاختبارات والقياس والتقويم وإرشادهم لكيفية بناء الفقرات؛ وانتقاء أفضل الطرق في الكشف عن الأداء التفاضلي لفقرات التقييم، وذلك للتخلص من الفقرات التي تُظهر أداءً تفاضليًا لفقرات التقييم أو تعديلها لتحقيق العدالة بين الأفراد في الاستجابة على الفقرات. كما يسمح الأداء التفاضلي للفقرة للباحثين بتحديد ما إذا كانت اختلافات الدرجات بين المجموعات في فقرة اختبار معين تنبع من اختلافات حقيقية في القدرة للسمة التي يتم قياسها. كما توفرالدراسة الحالية مساهمة العاملين في مجال الاختبارات في بناء اختبارات موضوعية وعادلة في العملية التعليمية، وتكتسب أهميتها من خلال استخدام طريقة تقدير الدرجات (التامة، الجزئية) التي توفر أدلة علمية بشأن الأداء التفاضلي لفقرات الاختبار؛ كما أنها ستفتح الباب أمام الباحثين للتعرف إلى هذه الطريقة ولاستخدامها في تحليلات الأداء التفاضلي المستقبلية مقارنة بالطرق الأخرى.

مصطلحات الدراسة

الأداء التفاضلي للفقرة: دالة مشتقة إحصائيًا للتعبير عن الفروق في الاستجابة على الفقرة بين المجموعات الفرعية للمفحوصين ممن لهم مستوى القدرة نفسه (أبو مسرة والسوالمة، 2021). ويُعرَّف إجرائيًا بأنه اختلاف الاحتمالات للوصول للإجابة الصحيحة عن فقرات اختبار تحصيلي في القياس والتقويم لدى طلبة مساق القياس والتقويم.

نموذج التقدير الجزئي: هو نموذج إحصائي للاستقلال العشوائي في نموذج راش، وهو نموذج للحلول التدريجية للفقرات ذات العلامات المتعددة، ويتم تفسير معلمات الفقرات على أنها

صعوبات في الخطوات (Verhelst & Verstralen, 2008). ويُعرَّف إجرائيًا بأنه أحد نماذج نظرية الاستجابة للفقرة الذي يعتبر معامل الصعوبة هو المؤشر الوحيد المؤثر في أداء المفحوصين.

طريقة تقدير الدرجات: هي الطريقة التي تُستخدم لتحديد الدرجة التي يستحقها المستجيب على إجابته عن الفقرة، ويُطلق عليها بطرق التصحيح (أبو جراد، 2015). وتُعرَّف إجرائيًا بأنها طريقة تقدير درجات الفقرات من خلال الطريقة التامة والجزئية.

طريقة التقدير التامة: هي أحد طرق تقدير الدرجات والتي تقوم على الكشف عن دالة الأداء ومعرفة منحنيات خصائص الفقرة (Chung & Huisu, 2004). وتُعرَّف إجرائيًا بأنها أحدى طرق تقدير الدرجات بحسب نموذج التقدير الجزئي.

طريقة التقدير الجزئية: هي إحدى طرق تقدير الدرجات والتي تقوم على رصد مستوى معرفة الطالب بكل اختيار مقدم في فقرات الاختبار (Fehintola & Akingbade, 2019). وتُعرَّف إجرائيًا بأنها أحد طرق تقدير الدرجات بحسب نموذج التقدير الجزئي.

حدود الدراسة

يتحدد نطاق تطبيق الدراسة الحالية على ما يأتي:

حدود زمانية: اقتصرت الدراسة الحالية على طلبة الفصل الدراسي الأول من العام الدراسي 2024/2023م، حيث طبقت الدراسة خلال شهر كانون الأول 2023.

حدود مكانية: اقتصر إجراء الدراسة الحالية على جامعة إربد الأهلية في مدينة إربد شمال الأردن.

حدود بشرية: اقتصرت الدراسة الحالية على طلبة المرحلة الجامعية الأولى (البكالوريوس) في جامعة إربد الأهلية، ولم تشمل طلبة الدراسات العليا.

حدود موضوعية: تتحدد نتائج الدراسة بأداة الدراسة التي استُخدمت فيها، وما تتمتع به من دلالات صدق وثبات، لذلك تتحدد إمكانية تعميم نتائج الدراسة على المجتمعات المماثلة لمجتمع الدراسة فقط، وبمدى تمثيل العينة لمجتمعها.

الطريقة والإجراءات

منهج الدراسة

اتبعت الدراسة الحالية المنهج الوصفي، إذ أنها بحثت الأداء التفاضلي لاختبار تحصيلي في مادة القياس والتقويم باستخدام نموذج التقدير الجزئي لدى عينة من طلبة جامعة إربد الأهلية.

مجتمع الدراسة

تَكوَّن مجتمع الدراسة من جميع طلبة المرحلة الجامعية الأولى (البكالوريوس) في كلية العلوم التربوية بجامعة إربد الأهلية، والذي يبلغ عددهم (834) طالباً وطالبة، وهم على مقاعد الدراسة في الفصل الدراسي الأول من العام الدراسي 2024/2023م.

عينة الدراسة

تكونت عينة الدراسة من (120) طالباً وطالبة من طلبة جامعة إربد الأهلية، فقد اختيرت شعب من مساق القياس والتقويم بالطريقة القصدية لتمثيل عينة الدراسة، لكون هذا المساق إجبارياً

لجميع طلبة كلية العلوم التربوية، والمسجلين به من كلا الجنسين ومختلف المستويات والتخصصات الأكاديمية. وجدول (1) يبين التكرارات والنسب المئوية للخصائص الديمغرافية لأفراد عينة الدراسة.

جدول (1): التكرارات والنسب المئوية حسب متغيرات الدراسة

	الفئات	التكرار	النسبة
الجنس	ذكر	48	%40
	أنثى	72	%60
	المجموع	120	%100
التخصص	الإرشاد النفسي والتربوي	39	%32.5
	معلم الصف	47	%39.2
	التربية المهنية	21	%17.5
	التربية الخاصة	13	%10.8
	المجموع	120	%100

أدوات الدراسة

أولاً: اختبار تحصيلي من إعداد الباحث.

لإعداد هذا الاختبار قام الباحث ببناء اختبار تحصيلي موضوعي وفق خطوات إعداد الاختبار التحصيلي الجيد لمادة القياس والتقويم لطلبة كلية العلوم التربوية من خلال اتباع الخطوات الآتية:

1. قام الباحث بتحديد الأهداف السلوكية لمحتوى المادة التي تُدرَّس للطلبة في المادة الدراسية.
2. قام الباحث بتحليل محتوى المادة لتحديد الكفايات الأساسية والأهداف السلوكية لمادة القياس والتقويم التي يجب على الطالب إتقانها، ثم تنظيم هذه الكفايات وإعداد قائمة بها لعرضها على (7) محكمين متخصصين في مجال القياس والتقويم في جامعتي إربد الأهلية واليرموك وذلك لإبداء الرأي حول كفايات المادة من حيث (دقة الصياغة اللغوية، وقابليتها للقياس، وقابليتها للتعلم، وتقديم المقترحات حول إضافة أو تعديل أو حذف أيٍّ من الكفايات). وبعد الاطلاع على آراء المحكمين في مجال القياس والتقويم النفسي والتربوي، قام الباحث بإجراء تعديلات على بعض الكفايات الفرعية، وقد أجمع المحكمين على الفقرات بنسبة (81.1%).
3. قام الباحث بتحليل الكفايات الأساسية للمادة التعليمية تحليلًا هرميًا إجرائيًا إلى مكوناتها من أهداف سلوكية في ضوء تقسيم مستويات بنية ناتج التعلم.
4. قام الباحث بوضع مواصفات الاختبار من خلال قواعد تفصيلية دقيقة تُرشد مُعِد الاختبار وتُمَكنه من صياغة فقرات مكافئة تقيس نفس الهدف. يتم بناء الفقرات في ضوء بنية نتائج التعلم. يركز على بنية الهدف والسؤال ولا يهتم بالعمليات العقلية. ويجب على الباحث إعداد هذه المواصفات في أربع مراحل رئيسة هي: الهدف، وعينة المثير، وعينة الاستجابة، وعينة الفقرة.
5. تحديد مدة الاختبار: قام الباحث فقط بتحديد فقرة واحدة لكل هدف واتخذ قراره بناءً على عدد الأهداف المراد قياسها. وكان الوقت المخصص للاختبار (15) دقيقة، وهو مقدار الوقت المخصص لجميع الطلاب للاختبار.

6. بناء فقرات الاختبار: قام الباحث بتصميم فقرات الاختبار بناءً على المواصفات التفصيلية التي تم تحديدها في الخطوة السابقة. وراعى الباحث تنوع فقرات الاختبار كدالة لتمييز المستوى المعرفي للهدف من حيث تقسيمه إلى مستويات بُنية نواتج التعلم.
7. تقييم صدق المحتوى: لتقييم صدق محتوى الاختبار أعد الباحث قائمة بالأهداف السلوكية وتصنيفاتها مع الفقرات التي تقيس كل هدف من أهداف الاختبار، وعرضها على مجموعة من المحكمين المتخصصين في القياس والتقويم النفسي التربوي لتقييم الآتي (كفاية أو عدم كفاية الأهداف لمستويات بنية نتائج التعلم، مدى كفاية أو عدم كفاية الفقرات لقياس الأهداف السلوكية، مدى كفاية الصياغة اللغوية للأهداف السلوكية وفقرات الاختبار). وقد تراوحت جميع قيم معاملات تطابق الفقرات مع عناصر التحكيم بين (0.90-1.00)؛ مما يدل على وجود صدق محتوى مرتفع للاختبار، وقد قام الباحث بإجراء التعديلات التي أشار إليها المحكمين.
8. قام الباحث بحساب ثبات الاختبار وفق طريقتي تقدير الدرجات التامة والجزئية باستخدام برنامج التحليل الإحصائي SPSS من خلال حساب معامل ألفا كرونباخ، حيث أظهرت نتائج التحليل أن معامل ثبات الاختبار باستخدام الطريقة التامة في تقدير الدرجات يساوي (0.73)، ومعامل ثبات الاختبار باستخدام الطريقة الجزئية في تقدير الدرجات تساوي (0.79)، وهي معاملات ثبات جيدة تؤكد على ثبات الاختبار.
9. قام الباحث ببناء وإنشاء الاختبار وتطبيقه على جميع الطلبة الدراسين لمادة القياس والتقويم، حيث اشتمل المقياس على (30) فقرة من نوع الصواب والخطأ المتعدد، وتم تصحيح الدرجات من خلال الطريقة التامة (0.1) والطريقة الجزئية (1.0، 0.75، 0.50، 0.25، 0) ثم حولت إلى (4، 3، 2، 1، 0) لتناسب طريقة الإدخال في برنامج STATA (خليل وإسماعيل، 2021).

الأساليب الإحصائية

تم استخدام برنامجي SPSS وSTATA في إجراءات ثبات وصدق أداة الدراسة، حيث تم استخدام برنامج التحليل الإحصائي SPSS في حساب التحليل العاملي والمتوسطات الحسابية ومدى مطابقة البيانات لافتراضات نموذج التقدير الجزئي، بينما استخدم برنامج STATA في حساب ثبات ألفا كرونباخ ومدى مطابقة البيانات لافتراضات نموذج التقدير الجزئي، والتحقق من قبول أو رفض أسئلة الدراسة.

قام الباحث بالتحقق من افتراضات نظرية الاستجابة للفقرة للاختبار التحصيلي: أُحادية البُعد، والاستقلال الموضعي، والمنحنى المميز للفقرة من خلال الإجراءات الآتية:

1-التحقق من أحادية البعد: قام الباحث بإجراء تحليل عاملي استكشافي بطريقة المكونات الأساسية قبل التدوير، وكانت النتائج وفق الجدول الآتي:

جدول (2): نتائج التحليل العاملي وفق طريقتي تقدير الدرجات التامة والجزئية

الطريقة التامة						الطريقة الجزئية					
العوامل	الجذر الكامن	نسبة التباين المفسر	العوامل	الجذر الكامن	نسبة التباين المفسر	العوامل	الجذر الكامن	نسبة التباين المفسر	العوامل	الجذر الكامن	نسبة التباين المفسر
1	3.795	%37.95	16	0.739	%7.39	1	3.928	%39.28	16	0.573	%5.73
2	1.029	%10.29	17	0.718	%7.18	2	1.098	%10.98	17	0.549	%5.49
3	0.997	%9.97	18	0.697	%6.97	3	0.982	%9.82	18	0.518	%5.18
4	0.970	%9.70	19	0.677	%6.77	4	0.957	%9.57	19	0.485	%4.85
5	0.958	%9.58	20	0.653	%6.53	5	0.921	%9.21	20	0.453	%4.53
6	0.931	%9.31	21	0.638	%6.38	6	0.893	%8.93	21	0.400	%4.00
7	0.917	%9.17	22	0.615	%6.18	7	0.867	%8.67	22	0.377	%3.77
8	0.898	%8.89	23	0.599	%5.99	8	0.833	%8.33	23	0.358	%3.58
9	0.873	%8.73	24	0.573	%5.73	9	0.801	%8.01	24	0.339	%3.39
10	0.859	%8.59	25	0.555	%5.55	10	0.778	%7.78	25	0.317	%3.17
11	0.835	%8.35	26	0.534	%5.34	11	0.743	%7.43	26	0.301	%3.01
12	0.816	%8.16	27	0.518	%5.18	12	0.713	%7.13	27	0.297	%2.97
13	0.792	%7.92	28	0.493	%4.93	13	0.682	%6.82	28	0.273	%2.73
14	0.775	%7.75	29	0.479	%4.79	14	0.655	%6.55	29	0.264	%2.64
15	0.751	%7.51	30	0.451	%4.51	15	0.629	%6.29	30	0.248	%2.48

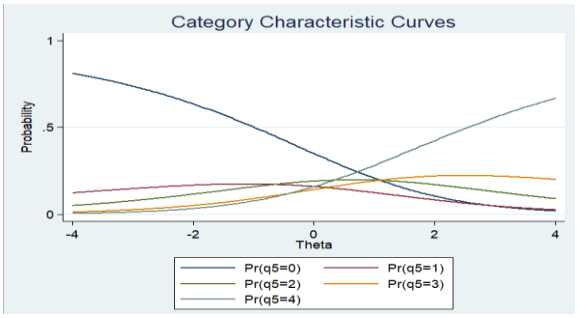
يتبين من جدول (2) أنه عند تحليل الاختبار المكون من مجموعة من الفقرات وفق الطريقة التامة والجزئية وجود عامل واحد سائد، إذ إن الجذر الكامن للعاملين الأول والثاني وفق الطريقة التامة لتقدير الدرجات بلغ (3.795؛ 1.029) على التوالي؛ وبلغت نسبة التباين المفسر (37.95% و10.29%) على التوالي. وتبين أن العامل الأول أكبر بثلاث مرات من الثاني، حيث إن قسمة الجذر الكامن للعامل الأول على جذر العامل الثاني يعطي (3.688)، وهي أكبر من قيمة (2) التي تعتبر محكا لأحادية البُعد (Lian & Idris, 2006).

وبلغت نسبة التباين المفسر للعامل الأول (37.95%). وهذه النسبة تعدُّ مقبولة لتقييم الاختبار على أنه أحادي البعد، أي أن العامل الأول يفسر أكثر من (20%) من إجمالي التباين. وكان الجذر الكامن للعاملين الأول والثاني وفق الطريقة الجزئية لتقدير الدرجات (3.928؛ 1.098) على التوالي، وكانت نسبة التباين المفسر (39.28%، 10.98%) على التوالي، ويتبين أن العامل الأول أكبر بثلاث مرات من العامل الثاني، وبلغت قسمة الجذر الكامن الأول على الجذر الكامن للعامل الثاني (3.577) وهي أكبر من القيمة (2)، وأن نسبة التباين المفسرة للعامل الأول (39.28%). وتعدُّ هذه النسبة مقبولة للحكم على الاختبار بأنه أحادي البعد، أي أن العامل الأول يفسر أكثر من (20%) من إجمالي التباين.

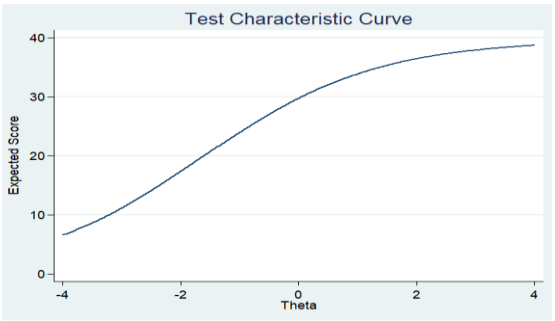
2-الاستقلال الموضعي: هذا يعني أن الاستجابة لفقرة اختبار ما ليس لها تأثير سلبي أو إيجابي على الاستجابة لفقرة أخرى؛ ويعني ذلك أن افتراض الاستقلال الموضعي يتحقق إذا كان احتمال الإجابة عن فقرة ما بشكل صحيح لا يرتبط باحتمال الإجابة على فقرة أخرى بشكل صحيح. وقام الباحث بالتحقق من استقلالية الموضعي من خلال حساب معاملات الارتباط بين فقرات الاختبار. وأظهرت النتائج انخفاض معاملات الارتباط بين الفقرات بالنسبة لبعضها بعضًا، مبينة أن الاستقلال الموضعي قد تحققت وفق طريقة تقدير الدرجات التامة والجزئية، حيث كانت جميع معاملات الارتباط أقل من (0.7) وهو المحك الذي إذا زادت عنه قيمة معامل الارتباط يتم رفض فرضية الاستقلال الموضعي.

3-منحنى خاصية الفقرة: هذا يعني احتمال الحصول على استجابة صحيحة على الفقرة كدالة للسمة الكامنة أو القدرة المقاسة في ضوء خصائص الفقرات. وللتحقق من هذه الفرضية تم استخدام برنامج STATA لاستخراج منحنى خاصية الفقرة لكل فقرة من فقرات الاختبار، ويظهر الشكل (3) منحنى خاصية الفقرة رقم (5) كمثال من فقرات الاختبار في صورته النهائية باستخدام نموذج التقدير الجزئي، ويظهر الشكل (4) منحنى خاصية الاختبار ككل باستخدام نموذج التقدير الجزئي.

شكل (3) منحنى خاصية الفقرة للفقرة رقم (5)



شكل (4) منحنى خاصية الاختبار ككل



يتضح من شكل (3) و(4) بأن قدرة الفرد تتزايد بشكل مباشر مع زيادة درجته على الفقرة، ويلاحظ أن معامل الصعوبة يؤثر في احتمالية الدرجة وليس هناك تأثير لمعامل التمييز إلا أن التخمين له تأثير من الدرجة (صفر حتى أقل من 10) أي أن نموذج التقدير الجزئي يساعد في تحقيق توازي منحنيات خصائص الفقرة.

التحقق من افتراضات النظرية

ما مدى التطابق للبيانات مع افتراضات نموذج التقدير الجزئي في ضوء طريقتي تقدير الدرجات (التامة، والجزئية)؟

للإجابة عن هذا الفرض تم تحليل فقرات الاختبار التحصيلي باستخدام البرنامج الإحصائي STATA v.17 لدراسة مدى مطابقة بيانات الاختبار التحصيلي لافتراضات نموذج التقدير الجزئي في ضوء طريقة تقدير الدرجات (التامة والجزئية)، حيث تم تقدير معالم تقدير الدرجات (التامة والجزئية) باستخدام نموذج التقدير الجزئي، كما قام الباحث بحساب ثبات الاختبار وفق طريقتي تقدير الدرجات (التامة والجزئية) باستخدام برنامج STATA من خلال حساب معامل كرونباخ ألفا باستخدام أداء (Multivariate analysis) وملحق (1) يوضح النتائج.

يتضح من ملحق (1) أن صعوبة فقرات الاختبار تراوحت ما بين (-2.13 – 2.14) وفق الطريقة التامة لتقدير الدرجات باستخدام نموذج التقدير الجزئي وجميعها ضمن المدى المقبول وهو بين (-2.95 – 2.95). كما تراوحت تقديرات صعوبة الخطوة من (0-1) بين (-1.77 – 5.01)، والخطوة (1-2) بين (-5.14 – 1.96)، والخطوة (2-3) بين (-2.93–2.80)، والخطوة (3-4) بين (-11.25 – 1.94) وفق الطريقة الجزئية لتقدير الدرجات باستخدام نموذج التقدير الجزئي.

كما أظهرت النتائج أن جميع قيم مربع كاي لفقرات الاختبار غير دالة إحصائيًا وفق الطريقة الجزئية لتقدير الدرجات. وأن جميع قيم مربع كاي لخطوات فقرات الاختبار غير دالة إحصائيًا ما عدا الخطوة من (3-4) دالة إحصائيًا في الفقرات (1، 3، 8، 11، 14، 16، 18، 24، 25، 28، 30) عند مستوى دلالة (0.01) وفق الطريقة الجزئية في تقدير الدرجات باستخدام نموذج التقدير الجزئي.

جدول (3): معامل ثبات الاختبار وفق طريقتي تقدير الدرجات (التامة والجزئية) باستخدام نموذج التقدير الجزئي

طريقة تقدير الدرجات	التامة	الجزئية
معامل ثبات الاختبار	0.64	0.79

يتضح من جدول (3) أن معامل ثبات الاختبار باستخدام الطريقة التامة في تقدير الدرجات يساوي (0.64)، ومعامل ثبات الاختبار باستخدام الطريقة الجزئية في تقدير الدرجات يساوي (0.79)، وهي معاملات ثبات جيدة تؤكد على صلاحية الاختبار وفقراته في جميع الحالات، ومطابَقة بيانات الاختبار التحصيلي لافتراضات نموذج التقدير الجزئي.
ويعزو الباحث هذه النتيجة إلى أن جميع تقديرات خصائص صعوبة الفقرات تقع ضمن المقبول، وأن جميع قيم مربع كاي لفقرات الاختبار غير دالة إحصائيا وفقًا لطريقة تقدير الدرجات الكاملة والجزئية، وأن قيم عامل استقرار الاختبار مقبولة نفسيا وفقا لطريقة تقدير الدرجات التامة والجزئية، وأن البيانات تتطابق مع نموذج التقدير الجزئي. يظهر أن هناك تطابق جيد مع البيانات مع افتراضات نموذج التقدير الجزئي في ضوء حقيقة أن النموذج لديه تطابق جيد مع البيانات مع افتراضات نموذج التقدير الجزئي. وتتفق هذه النتيجة مع نتيجة دراسة كل من خليل وإسماعيل (2021)، وأحمد (2017) وكاراكايا (Karakaya, 2012) الذين أشاروا إلى وجود مطابقة جيدة للبيانات مع النموذج المستخدم من خلال حساب الخصائص السيكومترية للفقرات.

نتائج الدراسة

السؤال الأول: هل تختلف تقديرات قدرات الأفراد وأخطاؤها المعيارية باختلاف طريقة تقدير الدرجات (التامة، والجزئية)؟
للإجابة عن هذا السؤال تم حساب تقديرات قدرات الطلبة وفق طريقتي تقدير الدرجات التامة والجزئية باستخدام البرنامج الإحصائي STATA من خلال أداء Postestimation، وتوضح النتائج أن تقدير القدرة باستخدام الطريقة الجزئية في تقدير الدرجات أعلى من الطريقة التامة في تقدير الدرجات، وجاء الخطأ المعياري لتقديرات القدرة باستخدام الطريقة الجزئية لتقدير الدرجات. ولتحديد الفروق في تقديرات قدرات الطلبة وفق طريقتي تقدير الدرجات التامة والجزئية تم استخدام برنامج التحليل الإحصائي SPSS وكانت النتائج كالآتي:

جدول (4): نتائج اختبار (ت) لدلالة الفروق بين متوسطات تقديرات قدرات الطلبة وفق الطريقة التامة والطريقة الجزئية في الاختبار التحصيلي باستخدام نموذج التقدير الجزئي

المتغير	طريقة التقدير	المتوسط الحسابي	الانحراف المعياري	درجات الحرية	قيمة (ت)	مستوى الدلالة
تقديرات القدرة	التامة	0.761	0.021	119	5.919	*0.001
	الجزئية	0.784	0.023			
الخطا المعياري لقدرات الطلبة	التامة	0.697	0.024	119	6.183	*0.000
	الجزئية	0.671	0.039			

* دال عند مستوى دلالة (0.01)

يتضح من جدول (4) وجود فروق ذات دلالة إحصائية بين متوسطات تقديرات قدرات الطلبة وفق الطريقة التامة والجزئية في تقدير الدرجات لصالح الطريقة الجزئية، كما أنه توجد فروق ذات دلالة إحصائية بين متوسطات الأخطاء المعيارية لقدرات الطلبة وفق الطريقة التامة والجزئية في تقدير الدرجات لصالح الطريقة التامة.

ويعزو الباحث هذه النتيجة إلى أن الطريقة الجزئية كانت أكثر دقة من الطريقة التامة في تقدير معايير قدرة الطلبة، وأرجع الباحث هذه النتيجة إلى المعرفة الجزئية المتعارف عليها أكثر لدى الطلبة، حيث تزيد من دقة تقديرات معلمات قدرة الطالب في ضوء ما يتم قياسه، وتأخذ الطريقة الجزئية لتقدير الدرجات الحد الأقصى من المعرفة الجزئية، حيث تكون الدرجة التي يستحقها الطالب أكثر تنوعا؛ لأن درجة المفحوص في الفقرة الواحدة تكون على أساس الدرجات (4، 3، 2، 1، 0)، ولذلك فإن هذه الطريقة تمكن من اكتشاف الفروق بين قدرات الطلبة بشكل أفضل، حيث يمكن من خلالها تقدير قدرات الطلبة بدقة عالية مع تقييم أداء الطالب، ويستحق المفحوص أقل من إجابته على البند المنفرد بالطريقة التامة (1، 0)، وبالتالي لديهم قدرة أقل على اكتشاف اختلافات القدرة وفروق القدرة بين المفحوصين، أي زيادة عدد تقدير الدرجات مسؤول عن زيادة دقة تقييم قدرات الطلبة. وتتفق هذه النتيجة مع نتيجة دراسة خليل وإسماعيل (2021) التي أشارت إلى أن الطريقة الجزئية في تقدير الدرجات أكثر دقة في تقدير معالم قدرات الطلبة.

السؤال الثاني: هل يختلف تقدير الأداء التفاضلي للاختبار التحصيلي لدى عينة الدراسة باختلاف الجنس وفق الطريقة التامة لتقدير الدرجات باستخدام نموذج التقدير الجزئي؟

للإجابة عن هذا السؤال تم تحليل فقرات الاختبار التحصيلي باستخدام البرنامج الإحصائي لتقدير الأداء التفاضلي للاختبار التحصيلي وفق الطريقة التامة لتقدير الدرجات باستخدام اختبار مانتل – هانزل، واختبار الانحدار اللوجستي، وجدول (5) يوضح ذلك.

جدول (5): اختبار مانتل-هانزل والانحدار اللوجستي للكشف عن الأداء التفاضلي وفق الطريقة التامة لتقدير الدرجات باستخدام نموذج التقدير الجزئي

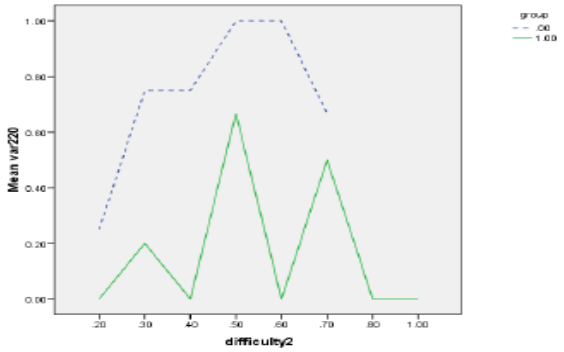
مانتل هانزل				الانحدار اللوجستي							
				منتظم			غير منتظم				
الفقرة	كا 2	الدلالة	الأداء التفاضلي المنتظم	الفقرة	كا 2	الدلالة	الأداء التفاضلي	كا 2	الدلالة	الأداء التفاضلي
1	6.79	*0.0168	الذكور	1	8.16	**0.0009	الإناث	0.73	0.3258	لا يوجد
2	1.28	0.5349	لا يوجد	2	1.24	0.4378	لا يوجد	1.08	0.1349	لا يوجد
3	0.43	0.6370	لا يوجد	3	0.88	0.8316	لا يوجد	0.87	0.4375	لا يوجد
4	1.28	0.3168	لا يوجد	4	1.37	0.9867	لا يوجد	1.99	0.9173	لا يوجد
5	0.75	0.1372	لا يوجد	5	0.58	0.7315	لا يوجد	5.91	**0.0001	الذكور
6	0.98	0.3519	لا يوجد	6	1.01	0.2227	لا يوجد	0.72	0.3733	لا يوجد
7	1.19	0.8288	لا يوجد	7	0.99	0.8346	لا يوجد	0.76	0.4359	لا يوجد
8	0.38	0.2725	لا يوجد	8	1.22	0.7349	لا يوجد	1.39	0.7349	لا يوجد
9	5.64	**0.0001	الذكور	9	0.41	0.6635	لا يوجد	0.29	0.4346	لا يوجد
10	1.73	0.9834	لا يوجد	10	1.51	0.5428	لا يوجد	1.14	0.8355	لا يوجد
11	0.91	0.4358	لا يوجد	11	0.44	0.3195	لا يوجد	0.37	0.1532	لا يوجد
12	0.82	0.5438	لا يوجد	12	1.12	0.2358	لا يوجد	5.34	*0.0012	الذكور
13	0.73	0.8315	لا يوجد	13	0.26	0.3524	لا يوجد	0.62	0.4197	لا يوجد
14	1.83	0.4382	لا يوجد	14	0.33	0.5372	لا يوجد	1.49	0.4322	لا يوجد
15	0.38	0.3468	لا يوجد	15	1.09	0.4365	لا يوجد	0.26	0.1257	لا يوجد
16	0.88	0.8953	لا يوجد	16	0.68	0.7349	لا يوجد	0.66	0.5376	لا يوجد
17	0.16	0.9273	لا يوجد	17	1.22	0.9834	لا يوجد	1.77	0.4359	لا يوجد
18	1.20	0.7329	لا يوجد	18	0.43	0.4395	لا يوجد	0.95	0.1375	لا يوجد
19	0.97	0.5379	لا يوجد	19	0.49	0.5233	لا يوجد	0.43	0.5349	لا يوجد
20	1.01	0.7321	لا يوجد	20	6.72	*0.0135	الذكور	1.35	0.1386	لا يوجد
21	0.53	0.4389	لا يوجد	21	1.37	03268	لا يوجد	0.71	0.9538	لا يوجد
22	0.64	0.8351	لا يوجد	22	0.90	0.5340	لا يوجد	1.91	0.8351	لا يوجد
23	7.21	*0.0123	الإناث	23	1.10	0.8319	لا يوجد	0.80	0.2346	لا يوجد
24	1.29	0.5379	لا يوجد	24	0.45	0.1312	لا يوجد	0.61	0.3265	لا يوجد
25	0.85	0.1385	لا يوجد	25	1.66	0.3211	لا يوجد	0.36	0.8349	لا يوجد
27	1.15	0.3265	لا يوجد	27	0.27	0.4365	لا يوجد	0.20	0.3193	لا يوجد
28	0.73	0.7256	لا يوجد	28	0.77	0.6321	لا يوجد	0.89	0.4359	لا يوجد
29	0.89	0.4387	لا يوجد	29	0.56	0.7485	لا يوجد	8.53	*0.0003	الذكور
30	1.62	0.5372	لا يوجد	30	1.47	0.6438	لا يوجد	1.13	0.9531	لا يوجد

** دال عند مستوى دلالة (0.01)، * دال عند مستوى دلالة (0.05)

يتضح من جدول (5) عدم وجود فروق ذات دلالة إحصائية بين متوسطات درجات الذكور والإناث في تقدير الأداء التفاضلي لفقرات الاختبار باستخدام طريقة مانتل–هانزل لجميع الفقرات ما عدا الفقرات ذات الأرقام (1، 9، 23) التي أظهرت أداءً تفاضليًا دالًا إحصائيًا؛ كما أنه لا توجد فروق ذات دلالة إحصائية بين الذكور والإناث في تقدير الأداء التفاضلي المنتظم لفقرات الاختبار باستخدام طريقة الانحدار اللوجستي ما عدا الفقرات ذات الأرقام (1، 20) التي أظهرت أداءً تفاضليًا منتظمًا دالًا إحصائيًا، بالإضافة إلى أنه لا توجد فروق ذات دلالة إحصائية بين الذكور والإناث في تقدير الأداء التفاضلي غير المنتظم لفقرات الاختبار باستخدام طريقة الانحدار اللوجستي ما عدا الفقرات ذات الأرقام (5، 12، 29) التي أظهرت أداءً تفاضليًا غير منتظم دالًا إحصائيًا.

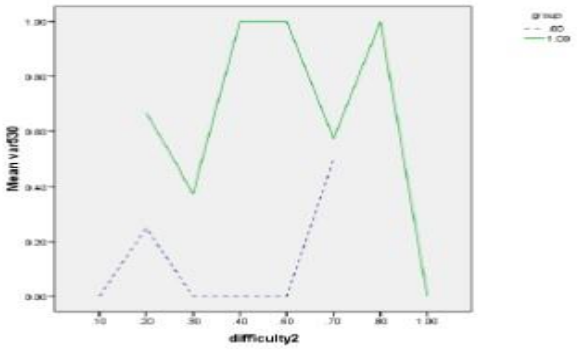
ويعزو الباحث هذه النتيجة إلى تفوق الذكور على الإناث في الفقرات (1، 9) باستخدام طريقة مانتل–هانزل، وهذه الطريقة تسمح بحساب معامل صعوبة الفقرة لكل مجموعة، ومن ثم حساب مقدار الأداء التفاضلي للفقرة، بحيث تُظهر الفقرة أداءً تفاضليًا إذا كانت الدلالة الاحتمالية لقيمة إحصائي مانتل-هانزل أقل من (α = 0.05).
ويُظهر الشكل رقم (5) تفوق الإناث على الذكور في الفقرة (9).

الشكل رقم (5): تفوق الإناث على الذكور في الفقرة (9)



وتفوقت الإناث على الذكور في الفقرة (1) باستخدام طريقة الانحدار اللوجستي للأداء التفاضلي المنتظم، بينما تفوق الذكور على الإناث في الفقرة (20) في نفس الطريقة. بينما تفوق الذكور على الإناث في الفقرات (5، 12، 29) باستخدام طريقة الانحدار اللوجستي للأداء التفاضلي غير المنتظم. حيث يعزو الباحث هذه النتيجة إلى الاختلاف في قدرات الطلبة والظروف المحيطة بالمفحوص سواء كانت شخصية أو اجتماعية أو بيئية وغيرها. حيث تعتمد هذه الطريقة في الكشف عن الأداء التفاضلي للفقرة على تكرار احتمال الاستجابة الصحيحة، والعلامة الكلية، وتستخدم هذه الطريقة جنباً إلى جنب مع طريقة مانتل هانزل، وتعتمد في حسابها على العلامة الكلية، وتقدير التفاعل بين العلامة الكلية وعضوية المجموعة، واختبار الدلالة الإحصائية لمتغيري عضوية المجموعة والتفاعل وتعتمد هذه المعادلة في حسابها على متغيرات: المجموعة، ومستوى القدرة والتفاعل بين المجموعة مع القدرة. ويظهر الشكل رقم (6) تفوق الذكور على الإناث في الفقرة (20).

الشكل رقم (6): تفوق الذكور على الإناث في الفقرة (20).



وتتفق هذه النتيجة مع نتيجة دراسة خليل وإسماعيل (2021) التي أشارت إلى تفوق الذكور على الإناث في طريقة مانتل – هانزل.

السؤال الثالث: ما هو النموذج الأفضل لتقييم كفاءة المعلومات في الاختبار التحصيلي وفق طريقتي تقدير الدرجات (التامة، والجزئية) باستخدام نموذج التقدير الجزئي؟
للإجابة عن هذا السؤال تم تحليل فقرات الاختبار التحصيلي باستخدام البرنامج الإحصائي STATA من خلال أداء Postestimation لحساب محك المعلومات (AIC) وفق طريقتي تقدير الدرجات التامة والجزئية، وجدول (6) يوضح ذلك.

جدول (6): محك دالة المعلومات (AIC) وفقت طريقتي تقدير الدرجات (التامة، والجزئية) باستخدام نموذج التقدير الجزئي

النموذج	الطريقة التامة	الطريقة الجزئية
Akaike's	5259.26	2107.38

يظهر جدول (6) أن قيمة محك دالة المعلومات وفق الطريقة التامة لتقدير الدرجات تساوي (5259.26)، بينما تساوي (2107.384) وفق الطريقة الجزئية في تقدير الدرجات، أي أن قيمة المحك للنموذج الثاني باستخدام الطريقة الجزئية لتقدير الدرجات أقل من قيمة المحك للنموذج الأول باستخدام الطريقة التامة لتقدير الدرجات؛ مما يدل على أن أفضل نموذج للحكم على كفاءة المعلومات هو النموذج الذي يعتمد على الطريقة الجزئية في تقدير الدرجات لأن قيمة محك المعلومات له أقل من قيمة محك المعلومات للنموذج الأول.
ويعزو الباحث هذه النتيجة إلى أن النموذج الأفضل للحكم على كفاءة المعلومات هو النموذج الذي يعتمد على طريقة درجة التقدير الجزئي، إذ يرى الباحث أنه عند اختيار معايير اختيار النموذج يجب قبول حقيقة أن النموذج قريب من الواقع، وبالنظر إلى مجموعة البيانات فإن الهدف هو تحديد أي من النماذج المرشحة تقارب البيانات بشكل أفضل؛ مما يضمن أقل قدر ممكن من فقدان المعلومات، ويعد محك دالة المعلومات (AIC) معيارًا للاختيار بين النماذج الإحصائية المتداخلة، حيث يتم تقييم كل منها على أساس علاقتها المتبادلة مع مجموعة بيانات محددة، وهو أيضًا مقياس لجودة النموذج الإحصائي القابل للاستخدام، مما يجعله طريقة مثالية لاختيار النموذج. ويمكن أن يساعد أيضًا الباحثين في تقدير المعلومات المفقودة إذا تم استخدام نموذج محدد لتمثيل العملية التي أنتجت البيانات. وتتفق هذه النتيجة مع نتيجة دراسة خليل وإسماعيل (2021) التي أشارت إلى أن أفضل النماذج للحكم على كفاءة المعلومات هو النموذج الذي يعتمد على الطريقة الجزئية في تقدير الدرجات؛ وتختلف نتيجة هذه الدراسة مع دراسة كاراكايا (Karakaya, 2012) التي أظهرت أن نماذج الاستجابة للفقرة ثنائية الاستجابة تعطي نتائج جيدة لدالة المعلومات لفقرات الاختبار، وقد يرجع هذا الاختلاف إلى اختلاف نموذج الاستجابة للفقرة المستخدم أو برنامج التحليل الإحصائي المستخدم في تحليل النتائج أو خصائص العينة التي طبق عليها الاختبار.
ويرى الباحث أن هذه النتيجة ترجع إلى أن منحنى دالة معلومات الاختبار المعتمد على طريقة تقدير الدرجات الجزئية أفضل في توفير المعلومات من منحنى دالة معلومات الاختبار المعتمد على طريقة تقدير الدرجة التامة. كما أوضح الباحث النتيجة إلى أن الطرق الجزئية لتقدير

الدرجات تأخذ في الاعتبار المعرفة الجزئية بأعلى درجة ممكنة، حيث تختلف الدرجة التي يستحقها الطلاب بشكل أكبر، وبالتالي تعطي أكبر قدر ممكن من المعلومات التي لا لبس فيها.

التوصيات

في ضوء هذه النتائج يوصي الباحث بما يأتي:

- ضرورة توعية الباحثين ومصممي الاختبارات للتحقق من عدالة الاختبارات والمقاييس النفسية.
- على الباحثين ومصممي الاختبارات الاهتمام ببناء الاختبارات والمقاييس وفق خطوات سليمة ومناسبة، وإعداد فقرات ذات خصائص أداء مناسبة.
- إجراء وعمل دراسات تتعلق بالأداء التفاضلي لفقرات أنواع أخرى من الاختبارات باستخدام نموذج التقدير الجزئي وفق طريقتي تقدير الدرجات (التامة، والجزئية).

قائمة المراجع

المراجع العربية

أبو جراد، حمدي يونس (2015). أثر اختلاف عدد البدائل الصحيحة في فقرات الصواب-خطأ المتعدد في دقة تقدير صعوبة الفقرات وقدرات الأفراد وثبات الاختبار. مجلة البحث العلمي في التربية، 16 (3)، 1 – 20.

أبو مسرة، آيات والسوالمة (2021). استخدام طريقة أشجار راش للكشف عن الأداء التفاضلي لفقرات الامتحان الوطني في الرياضيات للصف الرابع الأساسي في الأردن. المجلة الأردنية في العلوم التربوية، 17 (1)، 119-134.

أحمد، إبراهيم (2017). الكشف عن الأداء التفاضلي لفقرات اختبار تحصيلي في الرياضيات لدى تلاميذ المرحلة الإعدادية وفقًا لنظريتي القياس التقليدية والاستجابة للفقرة. رسالة دكتوراه غير منشورة، جامعة الزقازيق، مصر.

خليل، محمد وإسماعيل، ميمي (2021). الأداء التفاضلي لاختبار تحصيلي إلكتروني في مقرر القياس والتقويم لدى طلبة الفرقة الثالثة بكلية التربية بالجامعة باستخدام نموذج التقدير الجزئي. المجلة التربوية لكلية التربية بسوهاج، 2 (87)، 661-702.

الشريفين، نضال (2018). أثر نوع الأداء التفاضلي للفقرات على الخصائص السيكومترية للفقرات والاختبار وفق النماذج المعلمية والنماذج اللامعلمية لنظرية الاستجابة للفقرة. دراسات – العلوم التربوية، 45 (4)، 605-632.

هاشم، كمال الدين والخليفة، حسن (2011). التقويم التربوي: مفهومه – أساليبه – مجالاته. الرياض: مكتبة الرشد.

المراجع الأجنبية

Abdelwahab, M. (2019). The effect of method of estimating scores and types of item response model on accuracy of estimating persons and items parameters for multiple true and false test. International Journal for Research in Education,43 (2),59-85.

Chung, W. &Huisu, Y. (2004). Effects of averge signed area between two item characteristic curves and purification procedures on the DIF detection via the mantel-haenszel method. Applied Measurement in Education, 17(2), 113-144.

Fehintola, J. & Akingbade, O. (2019). Comparison of psychometric properties of multiple-choice test using confidence and number right scoring among senior secondary school students in ibadan metropolis. Al-Hikmah Journal of Education, 6 (1), 12-25.

Fischer, G. & Molenaar, I. (2013). Rasch models: Foundations, recent developments, and applications. New York: Springer-Verlag.

French, B. & Vo, T. (2020). Differential item functioning of a truancy assessment. Jounral Psychoeducational Assess. 38 (5), 642–648.

Holland, P. & Wainer, H. (2012). Differential item functioning. New York: Routledge.

Karakaya, I. (2012). An investigation of item bias in science and technology subtests and mathematic subtests in level determination exam (LDE). Educational Sciences: Theory and Practice, 12(1), 222-229.

Kolen, M. & Brennan, R. (2013). Test equating, scaling, and linking: Methods and practices. Stuttgart: Springer Nature Singapore.

Koljatic, M., Silva, M.& Sireci, S. (2021). College admission tests and social responsibility. Educational Measurement: Issues and Practice, 40 (4), 22-27.

Lane, S., Raymond, M. & Haladyna, T. (2016). Handbook of test development. New York: Routledge.

Lian, L. & Idris, N. (2006). Assessing algebraic solving ability of form four students. International Electronic Journal of Mathematics Education, 1(1), 55-76.

Linden, W. (2017). Handbook of item response theory. New York: CRC Press.

Lucey, C. & Saguil, A. (2020). The consequences of structural racism on MCAT scores and medical school admissions: The past is prologue. Academic Medicine, 95(3), 351–356.

Martinková, P., Drabinová, A., Liaw, Y., Sanders, E., McFarland, J. & Price, R. (2017). Checking equity: Why differential item functioning analysis should be a routine part of developing conceptual assessments. CBE—Life Sciences Education, 16 (2), rm2.

Matsumoto, D. & Vijver, F. (2011). Cross-Cultural Research Methods in Psychology. New York: Cambridge University Press.

Maydeu-Olivares, A., Cai, L. & Hernández, A. (2011). Comparing the fit of item response theory and factor analysis models. Struct. Equation Model. A Multidisciplinary Journal, 18 (3), 333–356.
Rasch, G. (1960). Probabilistic models for some intelligence and attainment Tests (G. Leunbach, Trans.). Copenhagen: The Danish Institute for Educational Research.
Reise, S. & Revicki, D. (2015). Handbook of item response theory modeling: Applications to typical performance assessment. New York: Routledge.
Shanmugam, S. (2020). Gender-related differential item functioning of mathematics computation items among non-native speakers of english. The Mathematics Enthusiast, 17(1), 108-140.
Verhelst, N. & Verstralen, H. (2008). Some considerations on the partial credit model. Psicologica, 29, 229-254.
Wahyuningsih, S. (2020). Using the rasch's partial credit model to analyze the quality of an essay math test. Advances in Social Science, Education and Humanities Research, 550, 257-265.
Wainer, H. & Braun, H. (2013). Test validity. New York: Routledge.
Wati, M., Mahtari, S., Hartini, S. & Amelia, H. (2019). A rasch model analysis on junior high school students' scientific reasoning ability. International Journal of Interactive Mobile Technologies (iJIM), 13(07), 141–149.
Wells, C. (2021). Assessing measurement invariance for applied research. New York: Cambridge University Press.
Wells, C. & Faulkner-Bond, M. (2016). Educational measurement: From foundations to future. New York: The Guilford Press.