جامعة سوهاج

كلية التربية

المجلة التربوية

الأداء التفاضلي لاختبار تحصيلي الكتروني في مقرر القياس والتقويم لدى طلبة الفرقة الثالثة بكلية التربية بالجامعة باستخدام نموذج التقدير الجزئي

إعداد

أ.د/محمد المري محمد اسماعيل خليل
أستاذ علم النفس التربوي المتفرغ
كلية التربية – جامعة الزقازيق

د/ ميمي السيد أحمد إسماعيل
كلية التربية – جامعة الزقازيق
جامعة التقنية والعلوم التطبيقية
الرستاق – عمان

تاريخ الاستلام : ٢٠ مارس ٢٠٢١م – تاريخ القبول : ١٠ أبريل ٢٠٢١م

DOI: 10.12816/EDUSOHAG.2021.

ملخص:

يهدف البحث إلى الكشف عن مدى مطابقة البيانات لافتراضات نموذج التقدير الجزئي في ضوء طريقتي تقدير الدرجات (التامة – والجزئية)، والتعرف على مدى اختلاف تقديرات قدرات الأفراد وأخطاؤها المعيارية باختلاف طريقتي تقدير الدرجات(التامة – والجزئية)، وكذلك التعرف على مدى اختلاف تقدير الأداء التفاضلي للاختبار التحصيلي الالكتروني باختلاف نوع الطالب(ذكور– اناث) وفق الطريقة التامة لتقدير الدرجات، بالإضافة إلى التعرف على أفضل نموذج للاختبار التحصيلي الالكتروني للحكم على كفاءة المعلومات وفق طريقتي تقدير الدجات (التامة والجزئية). تكونت عينة البحث من (١٠٧) طالباً وطالبة بالفرقة الثالثة بكلية التربية بجامعة التقنية والعلوم التطبيقية بالرستاق (خلال الفصل الدراسي الأول ٢٠٢٠/٢٠٢١م). ولتحقيق أهداف البحث قام الباحثان ببناء أداة البحث (اختبار تحصيلي الكتروني في مقرر القياس النفسي والتقويم التربوي من خلال نظام البلاك بورد) وتطبيقها بعد التحقق من افتراضات نظرية الاستجابة للمفردة. وباستخدام برنامج التحليل الاحصائيSTATA v.14 أظهرت نتائج البحث ما يلي: امتدت تقديرات صعوبة مفردات الاختبار التحصيلي الالكتروني بين (-٢.٥٤، ١.٦٨) باستخدام الطريقة التامة لتقدير الدرجات وفق نموذج التقدير الجزئي وجميعها ضمن المدى المقبول وهو بين (-٢.٩٥، ٢.٩٥)، كما امتدت تقديرات صعوبة الخطوة من(٠-١) بين (٠.٨٤-، ٧.٣٨)، والخطوة (١-٢) بين(-٧.٤٥، ٢.٦٣)، والخطوة(٢-٣) بين (-٣.٢٩، ٢.٥٧)، والخطوة (٣-٤) بين(-١٦.٥٣، ١,٠٨) باستخدام الطريقة الجزئية لتقدير الدرجات وفق نموذج التقدير الجزئي، وأن جميع قيم مربع كاي لمفردات الاختبار التحصيلي الالكتروني غير دالة إحصائيا مما يدل على مطابقة البيانات لافتراضات نموذج التقدير ا لجزئي باستخدام الطريقة التامة لتقدير الدرجات، كما أن جميع قيم مربع كاي لخطوات مفردات الاختبار التحصيلي الالكتروني غير دالة إحصائيا ما عدا الخطوة من(٣-٤) في المفردات(١، ٢، ٣، ٨، ٩)، وجاءت تقديرات القدرة باستخدام الطريقة الجزئية في تقدير الدرجات أعلى من الطريقة التامة في تقدير الدرجات، وجاء الخطأ المعياري لتقديرات القدرة باستخدام الطريقة التامة لتقدير الدرجات أعلى من الخطأ المعياري لتقديرات القدرة باستخدام الطريقة الجزئية لتقدير الدرجات، بالإضافة إلى أنه لا توجد فروق ذات دلالة إحصائية بين الذكور والاناث في تقدير الاداء التفاضلي لمفردات الاختبار

التحصيلي الالكتروني باستخدام طريقة مانتل هانزل ما عدا المفردة رقم (١) أظهرت أداءاً تفاضليا دالا احصائيا عند مستوى دلالة (٠.٠٥) لصالح الذكور، كما أنه لا توجد فروق ذات دلالة إحصائية بين الذكور والاناث في تقدير الأداء التفاضلي المنتظم لمفردات الاختبار التحصيلي الالكتروني باستخدام طريقة الانحدار اللوجستي ما عدا المفردة رقم (١) أظهرت أداءاً تفاضليا منتظما دالا إحصائيا عند مستوى دلالة (٠.٠١) ، ولا توجد فروق ذات دلالة إحصائية بين الذكور والاناث في تقدير الأداء التفاضلي غير المنتظم لمفردات الاختبار التحصيلي الالكتروني باستخدام طريقة الانحدار اللوجستي ما عدا المفردة (٣) أظهرت أداءاً تفاضليا دالا إحصائيا عند مستوى دلالة (٠.٠٥) ، وأن قيمة محك دالة المعلومات باستخدام نموذج (Akaike's) (AIC) يساوي(٤١٠٧.٥٨٩) باستخدام الطريقة التامة لتقدير الدرجات، (١٧٦٢.٣٤) باستخدام الطريقة الجزئية في تقدير الدرجات، مما يدل على أن أفضل نموذج للحكم على كفاءة المعلومات هو النموذج الذي يعتمد على الطريقة الجزئية في تقدير الدرجات.

الكلمات المفتاحية : الأداء التفاضلي– اختبار تحصيلي الكتروني– نموذج التقدير الجزئي – طلبة الجامعة.

---

*تم إجراء هذا البحث مناصفة بين الباحثين منذ بداية فكرة البحث إلى نهايته.

Differential items functioning for an electronic achievement test in measurement and evaluation for an education college third year students in a university using a partial credit model

Abstract:

This research aimed to identify the data to the assumptions of the partial estimation form in the light of grading estimation methods (full-partial), find out how the person's abilities estimation differs from its standard errors in the light of the differences of grading estimation method (full- partial) using partial credit model, find out how the Differential items functioning for the electronic achievement test differs according to the difference of the students' gender(males –females) according to the full method in grading estimation using partial credit model, in addition to find out the best electronic achievement test to judge the efficacy of the information. The sample consisted of (107) female and male students in education college in technology and applied science university in al Rustaq, in the academic year 2020/2021. The researchers made the research tool (an electronic achievement test) in the curriculum of measurement and evaluation through blackboard system, and they applied it after checking the item response theory assumptions. The results (using the statistical analysis program (STATA v.14) were: The electronic achievement test difficulty estimation for grading estimation were between (-2.54: 1.68) using the full method for grade estimation using the partial credit model, The step (0-1) difficulty estimation were between (0.84:7.38), step (1-2) were between (-7.45:2.63), step (2-3) were between (-3.29:2.57), step (3-4) were between (-16.53:1.08) using the partial method for estimating grade according to the partial estimation form, All Kai values of the electronic achievement test items were not statistically significant except for the step (3-4) in items (1, 2, 3, 8, and 9), The ability estimation using the partial method in grading estimation was higher than the full method in grading estimation, and the standard error of ability estimation using the full method in grading estimation was higher than the standard error of ability estimation using the partial method in grading estimation, There is no statistical differences between male students and female students in the Differential items functioning for the electronic achievement test items using Mantel Hanzel except for item number (1) that indicated a statistical differential performance at the point (0.05) for the male students, There are no statistical differences between male students and female students in the partial performance estimation of the electronic achievement test items using the logistic slope method except for item number (1) which indicated a statistical differential performance at the point of (0.01), as well as There are no statistical differences between male students and female students in the differential performance estimation of the electronic achievement test items using the logistic slope method except for item number

(3) which indicated a statistical differential performance at the point of (0.05), The value of the criterion of the information function using Akaike's model is (4107.58) using the full method in grading estimation, and it is (1762.34) using the partial method in grading estimation, which indicated the best model for judging information efficacy is the model depends on the partial method in grading estimation.

Key words: Differential items functioning - electronic achievement test- partial credit model – university students.

المقدمة:

يعد فحص الاداء التفاضلي للمفردة جزءا مهما في تحليل مفردات المقاييس. وقد أكد إصدار معايير الاختبارات التربوية والنفسية(AERA,APA& NCME,2014) على أن الأداء التفاضلي للمفردة هو دليل صدق قائم على البنية الداخلية. ويمثل وجود أداء تفاضلي للمفردة حضور أبعاد أخرى مزعجة تتطفل على السمة المصممة لقياسها، وهو شكل من أشكال التغير غير المرغوب في القياس، ويشير إلى اختلاف ما تقيسه المفردة من مجموعة إلى أخرى، وهذا قد يشكل مصدرا للتحيز في علامات الاختبار.

ويستخدم مفهوم الأداء التفاضلي للمفردة للتعبير عن الطرق والمعاني الاحصائية المستخدمة للكشف عن تحيز مفردات الاختبار(Ellis& Raju,2003). ويشير الأداء التفاضلي للمفردة DIF إلى الفرق في الاحتمال المشروط لإجابتها إجابة صحيحة بين مجموعتين(Penfield,2008). ويعني عدم وجود أداء تفاضلي للمفردة أن احتمالية الاستجابة الصحيحة متساوية للأفراد الذين ينتمون إلى مجموعتين مختلفتين ولكنهم متماثلون في القدرة.

ويمكن اعتبار أدوات القياس متحيزة إذا كانت تحابي مجموعة من المفحوصين من نفس مستوى القدرة على حساب مجموعات أخرى، وبالتالي ينتهك مبدأ العدالة بين المفحوصين. وفي هذا السياق فإن التحيز هو أحد مهددات صدق علامات أدوات القياس، فعندما تحابي إحدى المفردات مجموعة من المفحوصين، فإن ذلك يعني أن هذه المجموعة لها أفضلية دون المجموعات الاخرى في التوصل إلى الإجابة الصحيحة على المفردة (Pokropek & Kondratek,2012).

والأداء التفاضلي للمفردة من منظور نظرية الاستجابة للمفردة يكون بفحص العلاقة بين الاحتمالات المشروطة للإجابة الصحيحة على المفردة، وبين السمة الكامنة المراد قياسها، بحيث يتم مقارنة منحنيات خصائص المفردة في المجموعتين الفرعيتين لمجتمع إحصائي لكشف الأداء التفاضلي فيها والتي يجرى رسمها وفق النموذج اللوجستي المستخدم(Gybles,2004).

وهناك عدة طرق لتحديد الأداء التفاضلي للمفردات مثل الانحدار اللوجستي (Logistic regression) ، نظرية الاستجابة للمفردة (Item response theory)،

ونسبة الارجحية القصوى لمانتل هانزل (Mantel Hanzel) وتتميز هذه الطريقة عن الطرق الأخرى بأنها طريقة لامعلمية، ويمكن تنفيذها بسهولة لفحص الدالة المميزة للمفردة، وتعد من الطرق الأقوى للكشف عن الأداء التفاضلي (Haladyna,2004). وتعطي هذه الطريقة مقدار حجم الأثر واتجاهه بحسب معايير محددة، وتوفر معلومات كافية وذات معنى عن الأداء التفاضلي، ولا تحتاج إلى عينات كبيرة(Penfield,2008).

وقد تناولت العديد من الدراسات الاداء التفاضلي للمفردات في اختبارات ذات أهمية خاصة كاختبارات القبول أو بعض الاختبارات الدولية(Karakaya,2012 ; Pae,2004 ; Uiterwijk & Vallen,2005).

واهتمت دراسة ميدلتون ولايتوسيس Middleton and Laitusis(2007) بالكشف عن الاداء التفاضلي بين ذوي صعوبات التعلم الذين حصلوا على مساعدة (قراءة بصوت عال) والذين لم يحصلوا على مساعدة باستخدام النموذج اللوغاريتمي الخطي. وأظهرت الدراسة وجود ١٠ مفردات ذات أداء تفاضلي لصالح الذين حصلوا على مساعدة منها ٧ فقرات تتضمن أداء تفاضليا للمموهات.

وهدفت دراسة تسوسيس، وسيدريديس، والصاوي (Tsaousis, Sideridis and Al-Saawi 2018) بفحص الاداء التفاضلي كوسيلة لتحسين الخصائص السيكومترية لاختبار الكيمياء ضمن اختبار القبول التحصيلي المعياري. وقد أظهرت الدراسة أن هناك ٥ مفردات ذات أداء تفاضلي وقد تم اجراء بعض التعديلات على المفردات ذات الأداء التفاضلي من أجل تحسين الخصائص السيكومترية للاختبار.

وأثبتت العديد من الدراسات أن طريقة مانتل – هانزل من أكثر الطرق المستخدمة في الكشف عن الاداء التفاضلي للمفردات باستخدام نماذج الاستجابة للمفردة ثنائية الاستجابة من أهمها دراسة(أوعلا ومطارنة، ٢٠١٨)، ودراسة(Alquraan &Alkuwaiti, 2017)، ودراسة(Barnabas,2012)، ودراسة(Karakaya 2012,).

وتتضح أهمية نموذج التقدير الجزئي من خلال الاطلاع على الدراسات التي استخدمته في تحليل المفردات، حيث هدفت دراسة (Nakano & Primi,2014) إلى استخدام نموذج التقدير الجزئي في تحليل الاختبارات وذلك لأهمية النموذج في تقليل أخطاء القياس وبخاصة

بالنسبة للطلاب الذين يقترب مستواهم من درجة القطع، كما يتميز النموذج بالدقة في قياس القدرة من خلال تقليل درجة عدم مطابقة الطلاب الذين يقترب مستواهم من درجة القطع.

وقد أكدت دراسة كل من (Yau and Schwarz,2006; Zheng,2016) إلى أنه من الأفضل استخدام نموذج التقدير الجزئي في بناء الاختبارات التكيفية الالكترونية لأن الاعتماد على النماذج متعددة التدريج تقلل الفجوة في مستودع المفردات.

وفى ضوء ما أظهرته نتائج العديد من البحوث والدراسات الأجنبية –حول اختلاف تقديرات الاداء التفاضلي باختلاف الطريقة المستخدمة واختلاف الدراسات في تحديد الطريقة الأفضل للكشف عن الأداء التفاضلي لمفردات الصواب والخطأ المتعدد وندرة البحوث التي استخدمت نموذج التقدير الجزئي وفق طريقتي تقدير الدرجات (التامة والجزئية) يجعل هناك حاجة إلى إجراء هذا البحث، للمساعدة في حل هذه المشكلة، وذلك بحساب الأداء التفاضلي لاختبار تحصيلي الكتروني من نوع الصواب والخطأ المتعدد في مقرر القياس النفسي والتقويم التربوي لدى طلبة السنة الثالثة بكلية التربية بجامعة التقنية والعلوم التطبيقية باستخدام نموذج التقدير الجزئي.

مشكلة البحث:

إن دراسة دلالات الصدق والثبات لا يعطي مؤشراً كافياً لعدالة الاختبار ومراعاة البيئات الثقافية المختلفة، والكشف عن الاداء التفاضلي لمفردات الاختبار محل اهتمام المختصين في مجال الاختبارات والمقاييس النفسية لتحقيق مبدأ العدالة، وتناولت العديد من الدراسات الاداء التفاضلي، واهتمت بعض الدراسات بالكشف عن الاداء التفاضلي في تقييم الأداء التدريسي، بينما اهتمت دراسات أخرى بالمقارنة بين طرق الكشف عن الاداء التفاضلي في ضوء معايير محددة. وتكشف مراجعة الدراسات السابقة في هذا المجال عدم تناولها الاداء التفاضلي للمفردات الاختبارات التحصيلية الالكترونية باستخدام نموذج التقدير الجزئي. وانطلاقا مما سبق تمثلت مشكلة البحث في الكشف عن الاداء التفاضلي لمفردات اختبار تحصيلي الكتروني في مقرر القياس والتقويم لطلبة الفرقة الثالثة بكلية التربية باستخدام نموذج التقدير الجزئي، وذلك من خلال الاجابة عن الأسئلة الآتية:

– ما مدى مطابقة البيانات لافتراضات نموذج التقدير الجزئي في ضوء طريقتي تقدير الدرجات(التامة – والجزئية)؟

– هل تختلف تقديرات قدرات الأفراد وأخطاؤها المعيارية باختلاف طريقتي تقدير الدرجات(التامة – والجزئية) باستخدام نموذج التقدير الجزئي؟

– هل يختلف تقدير الأداء التفاضلي للاختبار التحصيلي الالكتروني باختلاف نوع الطالب وفق الطريقة التامة لتقدير الدرجات باستخدام نموذج التقدير الجزئي؟

– ما أفضل نموذج للحكم على كفاءة المعلومات في الاختبار التحصيلي الالكتروني وفق طريقتي تقدير الدرجات(التامة – والجزئية) باستخدام نموذج التقدير الجزئي؟

أهداف البحث:

يهدف البحث الحالي إلى تحقيق الأهداف التالية:

– التعرف على مدى مطابقة البيانات لافتراضات نموذج التقدير الجزئي في ضوء طريقتي تقدير الدرجات التامة والجزئية.

– التعرف على مدى اختلاف تقديرات قدرات الأفراد وأخطاؤها المعيارية باختلاف طريقة تقدير الدرجات التامة والجزئية باستخدام نموذج التقدير الجزئي.

– التعرف على مدى اختلاف تقدير الأداء التفاضلي للاختبار التحصيلي الالكتروني باختلاف نوع الطالب وفق الطريقة التامة لتقدير الدرجات باستخدام نموذج التقدير الجزئي.

– الكشف عن أفضل نموذج للحكم على كفاءة المعلومات في الاختبار التحصيلي الالكتروني وفق طريقتي تقدير الدرجات التامة والجزئية باستخدام نموذج التقدير الجزئي.

أهمية البحث:

يعد الموضوع الذي يتناوله البحث الحالي من الموضوعات المهمة في المجالات النفسية، وخصوصا مجال القياس والتقويم؛ حيث تبرز أهميته في معالجتها الاداء التفاضلي لمفردات اختبار تحصيلي الكتروني في مقرر القياس والتقويم لدى طلبة الفرقة الثالثة بكلية التربية باستخدام نموذج التقدير الجزئي، وبذلك تضيف موضوعا جديداً لم تتطرق إليه الدراسات العربية من قبل، أما الاهمية العملية تتمثل في تسليط الضوء على اهمية معرفة حجم الاداء التفاضلي للمفردة، وكيف يساعدنا على فهم الاداء التفاضلي للمفردات بشكل أفضل ، وإرشادات تساعد المهتمين على بناء الفقرات، لانتقاء أفضل الطرق في الكشف عن الاداء التفاضلي لمفردات التقييم، وذلك للتخلص من المفردات التي تظهر أداء تفاضلياً لمفردات التقييم، أو تعديلها لتحقيق العدالة بين الأفراد في الاستجابة على المفردات.

كما يعد موضوع الكشف عن الاداء التفاضلي للاختبارات التحصيلية الالكترونية باستخدام نموذج التقدير الجزئي هو الأول في سلطنة عمان، وهو من المواضيع النفسية الحديثة التي لم يتم التطرق لها من قبل وبناء عليه تتضح أهمية البحث في ما يلي:

- تقديم دراسة نفسية حديثة في مجال القياس والتقويم تقوم على الكشف عن الاداء التفاضلي لمفردات الاختبارات التحصيلية الالكترونية باستخدام نموذج التقدير الجزئي تساعد في حل المشكلات التي تهدد صدق وثبات الاختبارات والقرارات المترتبة عليها.
- يساعد في تقديم تعديلات على الاختبار التحصيلي الالكتروني بما يجعله خالياً من المفردات ذات الاداء التفاضلي.
- يزيد الباحثين كوسيلة للتحقق من عدالة الاختبارات والمقاييس النفسية.

مصطلحات البحث

الاداء التفاضلي للمفردة: (Differential items functioning)

يقصد به الطرق والمعالجات الاحصائية للكشف عن فروق الخصائص السيكومترية في أداء المفردة لدى مجموعتين مختلفتين، ويحدث حينما تظهر مجموعتان متكافئتان في السمة احتمالات مختلفة للإجابة الصحيحة عن نفس المفردة(Woods, 2009). ويعرفه (Abbott (2007بأنه احتمال اختلاف أداء المجموعات المختلفة من المختبرين من نفس مستوى السمة على المفردة. وتحدد اجرائيا من خلال اختلاف احتمالات الوصول للاجابة الصحيحة عن مفردات الاختبار التحصيلي الالكتروني لدى طلبة الجامعة. وهناك طريقتان لحساب الاداء التفاضلي تم استخدامها في البحث الحالي :

الطريقة الأولى: طريقة مانتل – هانزل

تعد من أكثر الطرق انتشارا واستخداما في الكشف عن الأداء التفاضلي المفردة وذلك لسهولة حساباتها واجراءاتها. فهي طريقة احصائية لفحص الاداء التفاضلي سواء لمفردات ثنائية الاستجابة أو لمفردات متعددة الاستجابة أو لفقرات متعددة الاستجابة .(Haladyna,2004)

الطريقة الثانية : طريقة الانحدار اللوجستي

هي احتمال الاجابة الصحيحة لفرد معين على مفردة في اختبار ما اعتمادا على المجموعة التي ينتمي اليها، وقدرته(مستوى السمة الكامنة لديه) وهي الدرجة الكلية التي يحصل عليها الفرد وتستخدم في مطابقة المستجيبين على السمة الكامنة(Finch ,Frensh &Immerkus,2016).

نموذج التقدير الجزئي: (Partial Credit Model)

أحد نماذج الاستجابة للمفردة ويستخدم معايرة وتدريج مفردات المقاييس والاختبارات أحادية البعد ومتعدة الاستجابة ويسمح باختلاف المفردات في معالم صعوبتها فقط ويفترض تساوي المفردات في تمييزها بين المفحوصين عند مستويات القدرة المختلفة(Abdelwahab,2019).

مفردات الصواب والخطأ المتعدد: (Multiple true and false questions)

هي شكل من أشكال مفردات الاختيار من متعدد وتتكون من متن واحد وعدة بدائل بعضها أو كلها صحيحة أو خطأ، ويطلب فيها من المستجيب الاستجابة لكل بديل من البدائل على حده بوضع علامة (√) أو علامة (×) (Brassil & Couch,2019)

- الإطار النظري البحث:

بدأ مصطلح الأداء التفاضلي للمفردة منذ بداية ثمانيات القرن العشرين للدلالة على الطرق والمعالجات الاحصائية التي تستخدم للكشف عن تحيز مفردات الاختبارات، والأداء التفاضلي للمفردة دالة مشتقة إحصائيا للتعبير عن الفرق في الاستجابة للمفردة بين مجموعتين عند مستوى قدرة واحد، وهناك من يستخدم مصطلح الأداء التفاضلي للمفردة كمرادف لمصطلح التحيز، ومثل هذا الاستخدام ليس له ما يبرره، فمصطلح الاداء التفاضلي للمفردة يستخدم بتحديد المفردات التي يكون احتمال الاجابة الصحيحة مختلفاً بين مجموعتين لهما مستوى القدرة، وعند تفسير اسباب ذلك يمكن وصف المفردات ذات الاداء التفاضلي بأنها مفردات متحيزة لمجموعة ما دون الأخرى لأسباب تتعلف بقدرة الأفراد فقط (Camili & Shepared,1994).

ويعرف Meade and Wright (2012) الأداء التفاضلي بأنه الخصائص السيكومترية لأداة القياس ومدى تماثل هذه الخصائص في المجموعات المختلفة من نفس مستوى السمة، ومدى تماثلها لنفس المجموعة وفقا لعامل الزمن كالدراسات الطولية ، ومدى تماثلها باختلاف صيغة الاجابة كالاختبارات الورقية والاختبار عبر الانترنت.

ويرى Ariffin, Idris and Ishak (2010) أن الأداء التفاضلي يصف المفردات التي تختلف في طريقة قياسها للبنية المستهدفة عندما يقدم الاختبار إلى مجموعات ذات خلفيات ديموغرافية مختلفة من نفس مستوى السمة.

ويذكر Uiterwijk and Vallen (2005) خطوات الكشف عن الأداء التفاضلي كالتالي: استخدام اجراءات إحصائية محددة للكشف عن المفردات ذات الأداء التفاضلي، والتعرف على العناصر المسئولة عن ظهور الأداء التفاضلي، وذلك من خلال استخدام طرق مختلفة للحصول على مصادر الأداء التفاضلي، واتخاذ قرار حول مصادر الأداء التفاضلي فيما إذا كانت ذات علاقة ببنية الاختبار والسمة المقاسة.

ولقد اقترح علماء القياس والتقويم عدة طرق وأساليب إحصائية للكشف عن الأداء التفاضلي للمفردات في جميع أشكال الاختبارات، وتهدف هذه الطرق والأساليب الاحصائية لتنقية الاختبارات من التحيز ولتطبيق مبدأ العدالة والمساواة بين المجموعات المختلفة وهي كالآتي :

الطرق القائمة على النظرية الكلاسيكية في القياس: تم تطوير هذه الطرق من خلال مفاهيم النظرية الكلاسيكية في القياس للكشف عن الاداء التفاضلي لمفردات الاختبار التحصيلي، ومن أهمها: طريقة تحليل التباين، وطريقة الصعوبة المحولة، وطريقة معامل التمييز(اللبدي، ٢٠٠٨).

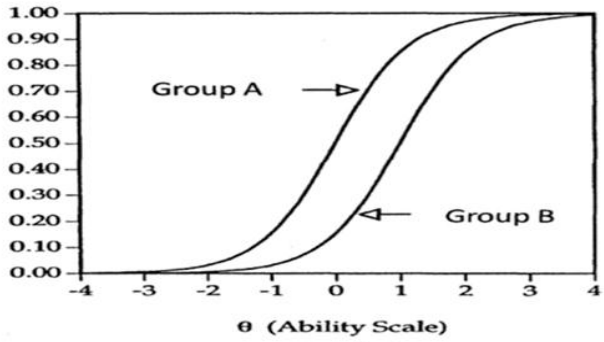
الطرق القائمة على نظرية الاستجابة للمفردة : توجد ثلاث طرق للكشف عن الأداء التفاضلي للاختبار التحصيلي وهي: طريقة منحنيات خصائص المفردة، وطريقة مقارنة المعالم المقدرة للمفردة، وطريقة منحنى نسبة الأرجحية(Chung & Huisu,2004).

الطرق القائمة على الاحصاء كاي تربيع: وتعمل هذه الطرق على مقارنة توزيع الاستجابات لمجموعتين من المختبرين، وتستخدم العلامة الكلية كمؤشر على القدرة التي تقيسها مفردات الاختبار التحصيلي. وتوجد ثلاث طرق للكشف عن الأداء التفاضلي لمفردات الاختبار

التحصيلي وهي: طريقة كاي تربيع وتعتمد هذه الطريقة على فحص احتمالية الاجابة الصحيحة على المفردة من قبل المستجيبين ذوي القدرة المماثلة في المجموعات المختلفة بشكل أفضل مما هو متوقع، فاذا اختلفت الاستجابة على المفردة فإنها تبدي أداء تفاضليا أما إذا لم تختلف فإن المفردة لا تبدي أداء تفاضليا، وطريقة الانحدار اللوجستي: قام بتطوير هذه الطريقة كل من Swaminathan &Rogers(1990) للاستخدام مع البنود الثنائية، وقد قام Zumbo(1999) بتوسيعها للاستخدام مع المفردات على شكل ليكرت، ونحاول عند تطبيق هذه الطريقة معرفة امكانية التنبؤ بالأداء على المفردة (الاجابة الصحيحة أو الخاطئة) من الدرجات الكلية والانتماء إلى المجموعة والتفاعل بينهما، وطريقة مانتل – هانزل طورت في اواخر الثمانينات من القرن الماضي من قبل هولاند وثابر (١٩٨٨)، تعد من أبرز الطرق في الكشف عن الأداء التفاضلي للمفردة وتعتمد على مقارنة الأداء لمجموعتين من الأفراد على كل مفردة حيث يتم تقسيم المجتمع إلى مجموعتين فرعيتين، تسمى الاولى بمجموعة الأغلبية أو المرجعية، وتسمى الثانية مجموعة الأقلية أو المستهدفة (In Yan Somg,2005).

وأشار Gybles (2004) أن هناك نوعين للأداء التفاضلي للمفردة هما:

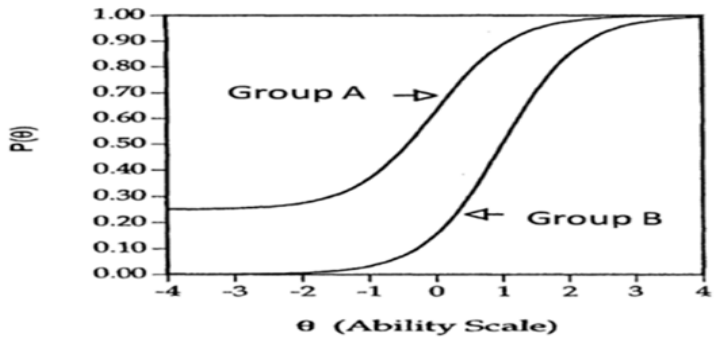
١) الاداء التفاضلي المنتظم(Uniform): يظهر عندما يكون منحنيا خاصية المفردة في المجموعتين المرجعية A، والمستهدفة B متوازيين، أي أن قيمة الفرق بين المجموعتين في احتمال الاجابة الصحيحة له نفس الاتجاه والمقدار عبر مستويات القدرة.



شكل (١) الأداء التفاضلي المنتظم للمفردة

٢) الأداء التفاضلي غير المنتظم (Nonuniform): يظهر عندما يكون منحنى خاصية المفردة في المجموعتين المرجعية A والمستهدفة B غير متوازيين. أي أن قيمة الفرق

بين المجموعتين في احتمال الاجابة الصحيحة للمفردة يختلف في الاتجاه أو المقدار عبر مستويات القدرة المختلفة.



شكل (٢) الأداء التفاضلي غير المنتظم للمفردة

<u>نموذج التقدير الجزئي:</u>

"يعتبر هذا النموذج تعميم لنموذج راش فقد قام العالم الاسترالي ماسترز (Masters,1982) بتطوير نموذج راش المتخصص في تحليل المفردات ثنائية الاستجابة ليشمل تحليل المفردات متعددة الاستجابة والمرتبة في فئتين أو أكثر كمفردات الاختبارات التحصيلية والتي يتطلب حلها إجراء العديد من الخطوات ومن ثم امكانية تصحيح الاجابة بشكل جزئي يتناسب مع عدد خطوات الحل وكذلك يمكن استخدامه في تحليل الاستجابة الناتجة عن مقاييس الاتجاهات ومقاييس الشخصية التي تعتمد على مقياس ليكرت" (علام، ٢٠٠٥، ص٦٩).

ويستخدم نموذج التقدير الجزئي (PCM) في حساب الأداء التفاضلي لمفردات الاختبار التي تتعامل مع المفردات متعددة التدريج والتي تتطلب الاجابة عنها المرور بعدة خطوات، بحيث تمثل كل خطوة نجاحا جزئيا في الاجابة عن المفردة وبالتالي يتم تقدير كل خطوة ، بحيث تمثل كل خطوة نجاحا جزئيا في الاجابة عن المفردة وبالتالي يتم تقدير كل خطوة تقديرا جزئيا في ضوء الدرجة الكلية للمفردة ، ومن ثم يهدف لدراسة النجاحات الجزئية للأفراد في الاجابة عن مفردات الاختبار، ولهذا يصلح استخدامه مع الاختبارات الموضوعية والاختبارات المقالية التي تتطلب الاجابة عن مفرداتها المرور بعدة خطوات متسلسلة ومتتابعة (Tutz & Berger ,2016).

ومن العرض السابق يتضح أن الأداء التفاضلي شرطاً أساسيا لاعتبار المفردة متحيزة ويدرس الخصائص السيكومترية للمفردة من حيث الكيفية التي تعمل بها المفردة في المجموعة بشكل مختلف عن المجموعة الاخرى، وتعتمد على الفروق في احتمالات الاجابة الصحيحة للمفردة في المجموعات المختلفة المتساوية في القدرة.

ويتبين من العرض السابق أن هناك العديد من الطرق للكشف عن الأداء التفاضلي للمفردات الاختبارات قائمة على الاحصاء كاي تربيع من أهمها طريقة الانحدار اللوجستي التي تساعد على امكانية التنبؤ بالأداء على المفردة من الدرجات الكلية والانتماء إلى المجموعة والتفاعل بينهما ، وطريقة مانتل هانزل التي تعد من أبرز الطرق في الكشف عن الأداء التفاضلي للمفردة وتعتمد على مقارنة الأداء لمجموعتين من الأفراد على كل مفردة.

البحوث والدراسات السابقة:

يتناول هذا الجزء عرضاً للدراسات السابقة ذات الصلة بموضوع الأداء التفاضلي للمفردات الاختبار التحصيلي وفق طريقتي تقدير الدرجات التامة والجزئية باستخدام نموذج التقدير الجزئي وفي ضوء متغير نوع الطالب لا توجد دراسات بحثت مباشرة بالموضوع نفسه، فهناك دراسات بحثت تقدير الأداء التفاضلي للمقاييس وفق نماذج الاستجابة للمفردة، ودراسات قارنت بين طرق تقدير الاداء التفاضلي باستخدام نماذج الاستجابة للمفردة الثنائية والمتعددة، ودراسة بحثت تقدير الفروق في تقدير الأداء التفاضلي للمقاييس النفسية في ضوء بعض المتغيرات باستخدام نماذج الاستجابة للمفردة ثناية ومتعددة الأبعاد، وقد تم عرض هذه الدراسات وفق الآتي :

قام جرادات(٢٠٠٣) بدراسة هدفت إلى مقارنة بين طريقتي الصعوبة المحولة ومانتل – هانزل في الكشف عن تحيز المفردات تبعا لمتغيري الجنس والحقل الأكاديمي، ومدى الاتفاق بين الطريقتين، وتم تطبيق اختبار كاليفورنيا لمهارات التفكير الناقد الذي تم تعريبه، وتكونت عينة الدراسة من(٩٩٦) طالباً، و(٧٨٩) طالبة، وقد تم استخدام معامل ارتباط فاي وكاي تربيع لاختبار درجة اتفاق الطريقتين والفروق بين نسب الاتفاق. وقد أظهرت نتائج الدراسة أن طريقة الصعوبة المحولة كانت أفضل من طريقة مانتل – هانزل في الكشف عن المفردات المتحيزة لمتغير الجنس، وكانت نسبة المفردات المتحيزة (٣٩%)، (١٨%) على

الترتيب، وأثبتت نتائج الدراسة أن هناك اتفاقاً بين الطريقتين في الكشف عن المفردات المتحيزة، وكانت نسبة الاتفاق (٥٩%) للجنس، و(٦٥%) للحقل الأكاديمي.

وهدفت دراسة Brad(2003) إلى استخدام أربع طرق لتقدير درجات مفردات الصواب والخطأ المتعدد ذات البدائل الخمسة في الجزء الخاص بالرياضيات في امتحان القبول بالجامعات الوطنية في تايوان لعينة تكونت من (٥٠٠٠) طالب، وتم حساب العلاقة الارتباطية بين الدرجات الخام المستمدة من كل طريقة من طرق تقدير الدرجات المقدرة باستخدام نموذج التقدير الجزئي، وتوصلت النتائج إلى وجود علاقة ارتباطية شبه تامة تراوحت بين(–0.97 0.99) بين الدرجات الخام وتقديرات القدرة باستخدام الطرق الأربعة لتقدير الدرجات وذلك على مستوى الاختبار المكون من (٥) مفردات، بينما تراوحت معاملات الارتباط بين (٠.٧٨– ٠.٩٨) على مستوى المفردة الواحدة في الحالات الأربعة، وأن الطريقة الرابعة في تقدير الدرجات كانت أكثر دقة من الطريقة الاولى والثانية والثالثة في تقدير قدرات الافراد، كما أن أفضل نموذج للحكم على كفاءة المعلومات كان وفق الطريقة الرابعة لتقدير الدرجات باستخدام نموذج التقدير الجزئي.

وأجري Pae(2004) بدراسة هدفت إلى الكشف عن الأداء التفاضلي في مفردات الجزء الانجليزي من اختبار القبول الكوري للعام (١٩٩٨) تبعاً لمتغير المجال الأكاديمي(علوم إنسانية، علوم) وذلك باستخدام نسبة الأرجحية وطريقة مانتل – هانزل باستخدام نموذج راش، حيث يضم الجزء الانجليزي في هذا الاختبار(٥٥) مفردة منها (١٧) مفردة تقيس مهارة الاستماع و(٣٨) مفردة تقيس مهارة استيعاب المقروء، وطبق الاختبار على (٨٣٩٨٣٧) طالباً ، وأظهرت نتائج الدراسة وجود (١٨) مفردة ذات أداء تفاضلي، فيها(٧) مفردات كانت أسهل للعلوم الانسانية.

وسعت (Innabi & Dodeen,2006) إلى تحليل مفردات اختبار الرياضيات التي تظهر أداءً تفاضليا وفقا لمتغير الجنس في الدراسة الدولية وتوجهات في الرياضيات والعلوم (Timss,1990) في الأردن، وتكونت عينة البحث من (٥٢٩٩) طالباً و(٢٨٢٩) طالبة من طلبة الصف الثامن الأساسي، وقد طبق الباحثان اختبار مكون من (٣٧) مفردة على الطلبة وتم حساب الأداء التفاضلي باستخدام طريقة مانتل –هانزل باستخدام نموذج التقدير الجزئي بعد التحقق من مدى مطابقة البيانات للنموذج من خلال حساب معالم صعوبة المفردات،

وأظهرت النتائج أن (٣٧) مفردة أظهرت أداءً تفاضليا بنسبة (٣٠%) من مفردات الاختبار، منها (١٧) مفردة لصالح الذكور، (٢٠) مفردة لصالح الاناث، وقد كانت أغلب المفردات التي أظهرت أداء تفاضليا لصالح الاناث تنتمي لموضوع الجبر والاحصاء، أما المفردات التي كانت سلبية باتجاه الاناث كانت موضوعاتها تتطلب المخاطرة في الحل، مثل موضوعات التقدير، والتوقع، والتقريب ، وقد كانت أغلب المفردات التي أظهرت أداءً تفاضليا لصالح الذكور تنتمي لموضوع القياس.

كما سعت دراسة Li–An We and Tsai (2010) إلى مقارنة أداء ثلاثة إجراءات(الانحدار اللوجستي (LogR)، واختبار نسبة الأرجحية (LRT)، واختلاف دالة المفردة والاختبار (DFTI)، للكشف عن دالة الأداء التفاضلي (DIF) تحت نموذج الاستجابة المتدرجة في دراسة محاكاة، وأظهرت النتائج أن اختبار نسبة الأرجحية الأقوى بين الثلاثة، كما أن اختلاف دالة المفردة والاختبار أقل قوة من اختبار نسبة الأرجحية، كما أظهرت النتائج أن إجراء الانحدار اللوجيستي لدالة الأداء التفاضلي كان حساسا فقط للمفردات ذات حجم دالة الأداء التفاضلي الكبير.

وهدفت دراسة (Karakaya (2012 إلى الكشف عن المفردات التي تظهر أداءً تفاضلياً تبعاً لمتغير الجنس في الاختبارات الفرعية في العلوم والتكنولوجيا والرياضيات والتي تم تطبيقها على الصفوف السادس والثامن ضمن اختبارات تحديد المستوى، حيث تكون الاختبار من خمسة مجالات فرعية (اللغة التركية، الرياضيات، والعلوم والتكنولوجيا، والعلوم الاجتماعية، واللغات)، ويهدف الاختبار إلى التعرف على مدى امتلاك الطلبة للمعارف والتحصيل العلمي في البرامج التعليمية في كل مستوى من الصفوف الأساسية، ثم طبق الاختبار على عينة مكونة من (٢٢٦٢٤) طالباً وطالبة، وتم التحقق من مطابقة النتائج للنموذج المستخدم، ثم استخدم الباحث طريقة مانتل – هانزل للكشف عن المفردات ذات الاداء التفاضلي باستخدام نماذج الاستجابة للمفردة ثنائية الاستجابة، وأكدت النتائج أنه يوجد (٣) مفردات ذات أداء تفاضلي لصالح الاناث، و(٢) لصالح الذكور في اختبار العلوم والتكنولوجيا، و(٣) مفردات لصالح الذكور، ومفردة واحدة لصالح الاناث في اختبار الرياضيات، بالإضافة إلى أن نماذج الاستجابة للمفردة ثنائية الاستجابة تعطي نتائج جيدة لدالة المعلومات لمفردات الاختبارات.

أما دراسة Ogbebor and Onuka (2013) هدفت إلى إستقصاء الأداء التفاضلي لمتغيري نوع المدرسة(الخاصة والعامة)، وموقع المدرسة (الحضرية والريفية) في مفردات اختبار مجلس الامتحانات الوطنية لعام ٢٠١٠، واستخدمت طريقة الانحدار اللوجستي لاستقصاء الأداء التفاضلي، وتوصلت النتائج إلى أن (١٠) مفردات أبدت أداء تفاضليا وفقا لنوع المدرسة، في حين أبدت (٨) مفردات أداء تفاضلياً وفقا لموقع المدرسة.

وتناول أحمد (٢٠١٧) دراسة هدفت إلى الكشف عن الأداء التفاضلي لمفردات اختبار تحصيلي مرجعي المحك في الرياضيات لدى تلاميذ المرحلة الإعدادية وفقاً لنظريتي القياس التقليدية والاستجابة للمفردة في ضوء بعض المتغيرات المتمثلة في: (حجم العينة، ونوع المدرسة، وترتيب المفردات) وذلك باستخدام خمس طرق للكشف عن الأداء التفاضلي وفقاً لنظرية القياس التقليدية CTT، وطريقتين للكشف عن الأداء التفاضلي للمفردة وفقاً لنظرية الاستجابة للمفردة IRT . وقام الباحث بإعداد اختبار تحصيلي في الرياضيات، وطبقه على عينة عشوائية كلية مكونة من (١٢٧٩) تلميذاً وتلميذة من تلاميذ الصف الثاني الإعدادي بإدارة ديرب نجم التعليمية، بمحافظة الشرقية، وباستخدام الأساليب والبرامج الإحصائية المناسبة لتحليل مفردات الاختبار، تم التوصل إلى النتائج التالية: لم تكشف طريقتي صعوبة المفردة المحولة TID، والتقنين المعياري STD عن أي مفردة ذات أداء تفاضلي بالنسبة لجميع أحجام العينات (المرجعية/ المستهدفة) المستخدمة في البحث سواء في حالة عدم/ مع تنقيح المفردة، وأن طريقة مانتل–هانسزل MH أكثر حساسية من طريقتي صعوبة المفردة المحولة TID، والتقنين المعياري STD في الكشف عن الأداء التفاضلي للمفردة من النوع المنتظم UDIF ولجميع متغيرات البحث (حجم العينة، نوع المدرسة، ترتيب المفردات) سواء في حالة عدم/ مع تنقيح المفردة وفقاً لنظرية القياس التقليدية ، بالإضافة إلى أنه لا يختلف الأداء التفاضلي للمفردة من النوع المنتظم باختلاف كل من: حجم العينة، ونوع المدرسة، وترتيب المفردات باستخدام جميع طرق الكشف وفقاً لنظريتي القياس التقليدية CTT والاستجابة للمفردة IRT .

وتناول Alquraan and Alkuwaiti (2017) دراسة بعنوان أثر التخصص للطالب كمصدر للأداء التفاضلي في مفردات تقييم فعالية وجودة التعليم العالي من وجهة نظر الطالب، وتكونت عينة الدراسة من (٣٦٤٥٩) طالباً وطالبة في كليات (التربية – الصحة – الهندسة) بجامعة الامام عبد الرحمن بن فيصل، وقد تم التحليل باستخدام النموذج أحادي المعلم المعدل(نموذج راش المعدل) وفق طريقة مانتل هانزل، وقد تم التحقق من الخصائص

السيكومترية للمفردات وفق النموذج المستخدم قبل استخدامه، وأظهرت النتائج وجود (٤) مفردات أبدت أداءً تفاضلياً حسب الكلية، وتبين أن محتوى المفردات يساعد على احتمالية أن تكون تلك المفردات متحيزة لكلية دون أخرى، كما أظهرت النتائج أن حذف المفردات ذات الأداء التفاضلي يسهم وبدلالة احصائية في تحسين صدق البناء للأداة، وأن نموذج راش المعدل من أكثر نماذج نظرية الاستجابة للمفردة فاعلية في تقديم معلومات جيدة حول تقدير بعض معالم المفردات.

وأجرى Shanmugam (2020) بدراسة هدفت إلى التعرف على الفروق بين الجنسين في تقدير الأداء التفاضلي لمفردات اختبار الرياضيات لدى الطلاب غير الناطقين اللغة الانجليزية باستخدام طريقة مانتل هانزل، تكونت عينة الدراسة من (٩٨٨) طالب، (١٣٨١) طالبة من طلبة الصف الثاني من ٣٤ مدرسة، (١٧) مدرسة تقع في جزيرة بينانغ، (١٢) مدرسة في بر بينانغ، (٥) مدارس في بيراك. واستخدم الباحث اختبار في مادة الرياضيات مكون من (٢٠) مفردة في موضوع الكسور والأرقام السالبة ويقيس هذا الاختبار مهارات التفكير العليا والدنيا لدى الطلاب. وقد تم استخدام برنامج WINSTEPS في ضوء نموذج راش. أظهرت النتائج أنه يوجد (٧) مفردات أظهرت أداءً تفاضليا منها (٢) أظهرت أداء تفاضليا معتدلا، (٥) أظهرت أداءً تفاضليا كبير، وأنه يوجد فروق في الاداء التفاضلي لصالح الذكور في المفردات التي تعتمد على مهارات التفكير العليا.

- تعليق على البحوث والدراسات السابقة:

يتضح من خلال استعراض البحوث والدراسات السابقة، أنها تباينت في أهدافها ومتغيراتها، فهدف بعضها إلى الكشف عن الأداء التفاضلي للمفردة تبعا لمتغير واحد مثل دراسة كل من (Alquraan & Karakaya,2012; shanmugam,2020 ;Alkuwaiti,2017)، وبعضها هدف إلى الكشف عن الاداء التفاضلي للمفردة تبعاً لمتغيرين مثل دراسة كل من (Ogbebor & Onuka,2013 ; جرادات، ٢٠٠٣)، في حين ركزت دراسات أخرى إلى استخدام طريقة واحدة للكشف عن الأداء التفاضلي للمفردة مثل دراسة كل من (Innabi & Ogbebor & Onuka,2013; shanmugam,2020 ;Alquraan & Alkuwaiti,2017; Dodeen,2006)، والبعض استخدم طريقتين أو أكثر مثل دراسة كل من (Pae,2004; ; 2010, Li–An We & Tsai جرادات، ٢٠٠٣)، كما

يلاحظ أن بعض الدراسات اهتمت بالكشف عن الاداء التفاضلي من خلال نماذج الاستجابة للمفردة ثنائية الاستجابة مثل دراسة كل من(Pae,2004 ; Shanmugam,2020 Alquraan &Alkuwaiti,2017; Karami,2011; Karakaya,2012;، والبعض استخدم نماذج الاستجابة للمفردة متعددة الاستجابات مثل نموذج التقدير الجزئي ونموذج الاستجابات المتدرجة مثل دراسة كل من(Li–An We & Ogbebor & Onuka,2013 ;2010 ,Tsai ;Innabi & Dodeen,2006 ;أحمد،٢٠١٧)، ويلاحظ أيضا ندرة الدراسات في البيئة العمانية التي تناولت الكشف عن الأداء التفاضلي من خلال نموذج التقدير الجزئي، ويأتي هذا البحث ليتميز عن سابقه في موضوعه وهدفه الذي يتمثل في الكشف عن الأداء التفاضلي لاختبار تحصيلي الكتروني باستخدام نموذج التقدير الجزئي وفق طريقتي تقدير الدرجات (التامة والجزئية).

فروض البحث:

بناءً على نتائج البحوث والدراسات السابقة، سعى البحث الحالي للتحقق من قبول أو رفض الفروض التالية:

– توجد مطابقة جيدة للبيانات لافتراضات نموذج التقدير الجزئي في ضوء طريقتي تقدير الدرجات(التامة– والجزئية).

– تختلف تقديرات قدرات الأفراد وأخطاؤها المعيارية باختلاف طريقتي تقدير الدرجات(التامة– والجزئية) باستخدام نموذج التقدير الجزئي.

– يختلف تقدير الأداء التفاضلي للاختبار التحصيلي الالكتروني باختلاف نوع الطالب(ذكو– اناث) وفق الطريقة التامة لتقدير الدرجات باستخدام نموذج التقدير الجزئي.

– أفضل نموذج للحكم على كفاءة المعلومات في الاختبار التحصيلي الالكتروني وفق طريقة تقدير الدرجات الجزئية باستخدام نموذج التقدير الجزئي.

إجراءات البحث:

– **منهج البحث:** تم استخدام المنهج الوصفي للكشف عن الاداء التفاضلي لمفردات الاختبار التحصيلي الالكتروني لدى طلبة الفرقة الثالثة بكلية التربية بالجامعة، وذلك لمناسبة هذا المنهج لطبيعة البحث وأهدافه، وباعتبار هذا المنهج لا يكتفي بجمع البيانات وتنظيمها، بل

يتعدى ذلك إلى تحليلها وتفسيرها للتوصل إلى مجموعة من النتائج التي تساعد على فهم الواقع ومن ثم العمل على تحسينه.

– **مجتمع وعينة البحث** : تألف مجتمع البحث من جميع طلاب وطالبات الفرقة الثالثة الدارسين لمقرر القياس النفسي والتقويم التربوي بجامعة التقنية والعلوم التطبيقية بالرستاق بسلطنة عمان للعام الدراسي ٢٠٢٠ / ٢٠٢١م والبالغ عددهم (١٥٧) طالب وطالبة، وتكونت عينة حساب ثبات وصدق الاختبار من (٥٠) طالب وطالبة من طلبة الفرقة الثالثة الدراسين لمقرر القياس النفسي والتقويم التربوي بكلية التربية بالرستاق، بينما تكونت عينة البحث الأساسية من(١٠٧) طالباً وطالبةً بالفرقة الثالثة بكلية التربية بالرستاق خلال العام الدراسي ٢٠٢٠/٢٠٢١م منهم (٦٣) طالباً و(٤٤) طالبةً، بمتوسط أعمار بلغ (١٨.٩) وانحراف معياري (١.٢٤)، ويوضح الجدول(١) توزيع أفراد العينة وفقا لمتغير نوع الطالب.

جدول (١)

توزيع أفراد العينة حسب متغير نوع الطالب

المتغير	المستوى	التكرار	النسبة
نوع الطالب	الذكور	٦٣	%٥٨.٨٨
	الاناث	٤٤	%٤١.١٢
	المجموع	١٠٧	%١٠٠

– **أداة البحث:**

اختبار تحصيلي الكتروني Electronic Achievement Test (إعداد الباحثان):

لإعداد هذا الاختبار قام الباحثان ببناء اختبار تحصيلي موضوعي وفق خطوات اعداد الاختبار التحصيلي الجيد للوحدة الأولى في مقرر القياس النفسي والتقويم التربوي لطلبة الفرقة الثالثة بجامعة التقنية والعلوم التطبيقية بالرستاق من خلال اتباع الخطوات الآتية:

١. تحديد الأهداف السلوكية التي تم تدريسها للطلاب في الوحدة الأولى للمقرر.
٢. تحديد الكفايات والأهداف السلوكية للوحدة الأولى لمقرر القياس النفسي والتقويم التربوي. حيث قام الباحثان بتحليل محتوى الوحدات، وتحديد الكفايات الأساسية التي يجب أن يتقنها الطالب ليعد متقنا للوحدة، ثم تنظيم هذه الكفايات وإعداد قائمة بها لعرضها على الخبراء في مجال علم النفس التربوي والقياس والتقويم وذلك لإبداء الرأي حول كفايات الوحدة من حيث:
   – دقة الصياغة اللغوية

- **قابليتها للقياس**
- **قابليتها للتعلم**
- **تقديم المقترحات حول إضافة أو تعديل أو حذف من الكفايات.**

وبعد الاطلاع على آراء الخبراء في مجال علم النفس والقياس والتقويم النفسي والتربوي، قام الباحثان بإجراء تعديلات على بعض الكفايات الفرعية، مع الابقاء على الكفايات الرئيسة الأربعة دون تعديل(ملحق ١)

٣. **تحليل الكفايات الأساسية إلى مكوناتها السلوكية: قام الباحثان بتحليل كفايات الوحدة التعليمية تحليل هرميا إجرائيا إلى مكوناتها من أهداف سلوكية في ضوء تقسيم مستويات بنية ناتج التعلم (ملحق ٢)**

٤. **إعداد المواصفات الاختبار: في هذه الخطوة قام الباحثان بإعداد مواصفات الاختبار ويقصد بها تلك القواعد التفصيلية الدقيقة التي ترشد معد الاختبار وتمكنه من صياغة مفردات متكافئة تقيس الهدف نفسه، ويتم بناء المفردات في ضوء تقسيم بنية ناتج التعلم فهو يركز على البنية للهدف والسؤال ولا يهتم بالعمليات العقلية، وقد قام الباحثان بإعداد تلك المواصفات في أربع خطوات رئيسة هي :**

- **الهدف**
- **عينة المثير**
- **عينة الاستجابة**
- **عينة المفردة**

٥. **تحديد طول الاختبار: اكتفى الباحثان بتحديد مفردة واحدة فقط لكل هدف، وذلك في ضوء عدد الأهداف المطلوب قياسها، والزمن المتاح لتطبيق الاختبار(٢٠) دقيقة وهي الفترة الزمنية التي استطاع جميع الطلاب من خلالها انهاء الاختبار، وبذلك أصبح طول الاختبار(١١) مفردة .**

٦. **بناء مفردات الاختبار: قام الباحثان بكتابة مفردات الاختبار استرشادا بالمواصفات التفصيلية التي تم تحديدها في الخطوة السابقة، وقد راعي الباحثان تنوع مفردات الاختبار وفقا لتنوع المستوى المعرفي للهدف في ضوء تقسيم مستويات بنية ناتج التعلم (ملحق ٣).**

٧. تقدير صدق المحتوى: قام الباحثان بحساب صدق المحتوى للاختبار بإعداد قائمة تتضمن الأهداف السلوكية وتصنيفاتها، مع المفردات التي تقيس كل هدف من أهداف الاختبار (ملحق ٣)، وتم عرضها على مجموعة من المحكمين المتخصصين في مجال علم النفس التربوي والقياس والتقويم النفسي والتربوي للحكم على الآتي :

- مناسبة أو عدم مناسبة الأهداف لمستويات بنية ناتج التعلم.
- مناسبة أو عدم مناسبة المفردات لقياس الأهداف السلوكية.
- دقة الصياغة اللغوية لكل من الأهداف السلوكية ومفردات الاختبار.

وقد تراوحت جميع قيم معاملات تطابق المفردات مع عناصر التحكيم بين (٠.٩٠– ١.٠٠)، مما يدل على وجود صدق محتوى مرتفع جدا للاختبار، وقد قام الباحثان بإجراء التعديلات التي أشار إليها السادة المحكمين، كما في الجدول (٢) الذي يوضح المفردات قبل التعديل وبعده.

جدول (٢)
التعديلات التي أشار إليها السادة المحكمين

رقم الهدف	قبل التعديل	بعد التعديل
١٠	يميز بين الاختبار والملاحظة والمقابلة	يميز بين الاختبار ووسائل التقويم الأخرى.
١١	يميز بين الاختبار والاستبيان	

وقد أشار المحكمين إلى ضرورة حذف المفردة رقم (١١) نظرا لضم الهدف العاشر مع الهدف الحادي عشر.

٨. حساب ثبات الاختبار: قام الباحثان بحساب ثبات الاختبار وفق طريقتي تقدير الدرجات التامة والجزئية باستخدام برنامج SPSS من خلال حساب معامل ألفا كرونباخ (في حال حذف درجة المفردة)، وأظهرت النتائج أن معامل ثبات الاختبار باستخدام الطريقة التامة في تقدير الدرجات يساوي (٠.٦١) ، ومعامل ثبات الاختبار باستخدام الطريقة الجزئية في تقدير الدرجات يساوي (٠.٧٢)، وهي معاملات ثبات جيده تؤكد على ثبات الاختبار. من الإجراءات السابقة تأكد للباحثين صلاحية الاختبار في صورته النهائية (المكون من: ١٠ مفردات) للتطبيق في البحث الحالي (ملحق ٤).

٩. إنشاء الاختبار على نظام البلاك بورد: قام الباحثان بناء وتصميم الاختبار من خلال نظام البلاك بورد، وتطبيقه الكترونيا على جميع الطلبة الدارسين لهذ المقرر من خلال نظام البلاك بورد، واشتمل هذا الاختبار على (١٠) مفردات من نوع الصواب والخطأ المتعدد، وتم تصحيح الدرجات الكترونيا من خلال نظام البلاك بورد بطريقتين الأولى: الطريقة التامة (١،٠)، والثانية: الطريقة الجزئية(١، ٠.٧٥، ٠.٥٠، ٠.٢٥، ٠) ثم حولت إلى (٤، ٣، ٢، ١، ٠) لتلائم طريقة الادخال في برنامج (STATA).

– الأساليب الإحصائية: تم استخدام برنامجي (SPSS v.26)، و(STATA v.14) في إجراءات ثبات وصدق اداة البحث والتحقق من قبول أو رفض فروضه؛ حيث تم استخدام برنامج (SPSS v.26) في حساب التحليل العاملي والمتوسطات الحسابية ومدى مطابقة البيانات لافتراضات نموذج التقدير الجزئي، بينما تم استخدام برنامج(STATA v.14) في حساب ثبات ألفا كرونباخ، ومدى مطابقة البيانات لافتراضات نموذج التقدير الجزئي، والتحقق من قبول أو رفض فروض البحث.

– التحقق من افتراضات نظرية الاستجابة للمفردة للاختبار التحصيلي : أحادية البعد، والاستقلال الموضعي، والمنحنى المميز للمفردة. قام الباحثان بالإجراءات الآتية:

١. التحقق من أحادية البعد: قام الباحثان بإجراء تحليل عاملي استكشافي بطريقة المكونات الأساسية قبل التدوير وكانت النتائج وفق الجدول التالي:

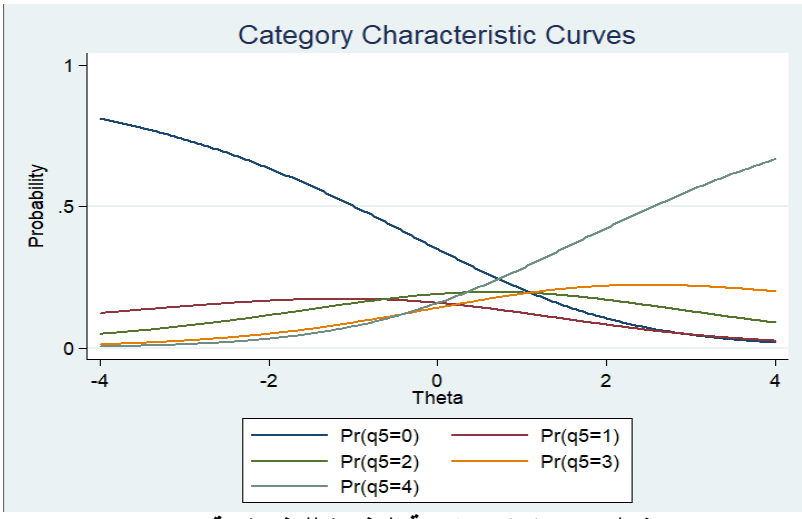
جدول (٣)
نتائج التحليل العاملي من وفق طريقتي تقدير الدرجات التامة والجزئية

الطريقة التامة			الطريقة الجزئية		
العوامل	الجذر الكامن	نسبة التباين المفسر	العوامل	الجذر الكامن	نسبة التباين المفسر
1	3.565	%35.65	١	3.667	%٣٦.٦٧
2	1.045	%١٠.٤٥	٢	1.075	%١٠.٧٥
3	0.984	%٩.٨٤	٣	0.915	%٩.١٥
4	0.866	%٨.٦٦	٤	0.883	%٨.٨٣
5	0.832	%٨.٣٢	٥	0.815	%٨.١٥
6	0.769	%٧.٦٩	٦	0.719	%٧.١٩
7	0.678	%٦.٧٨	٧	0.674	%٦.٧٤
8	0.496	%٤.٩٦	٨	0.613	%٦.١٣
9	0.485	%٤.٨٥	٩	0.451	%٤.٥١
10	0.280	%٢.٨٠	١٠	0.188	%١.٨٨

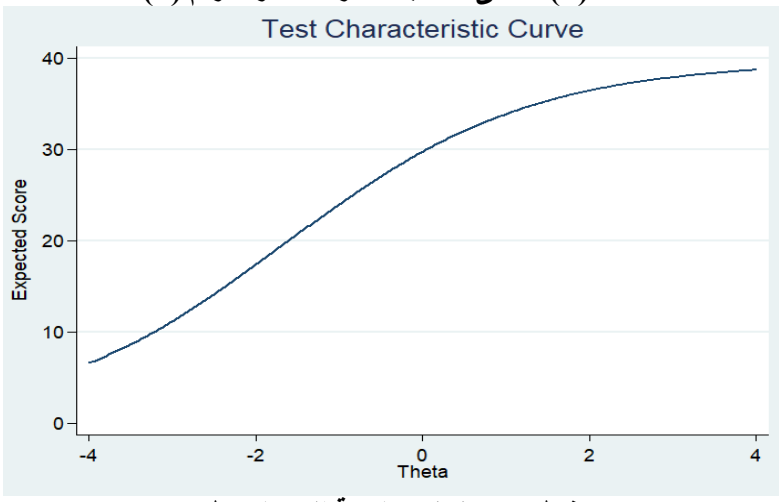
يتضح من الجدول السابق أنه يوجد عامل واحد سائد عند تحليل الاختبار على أنه يتكون من مجموعة من المفردات وفق الطريقة التامة والجزئية في تقدير الدرجات، حيث بلغ الجذر الكامن للعامل الأول والثاني وفق الطريقة التامة لتقدير الدرجات(3.565، ١.٠٤٥) على الترتيب، ونسبة التباين المفسر(٣٥.٦٥%، ١٠.٤٥%) على الترتيب، ويتبين أن العامل الأول أكثر من ثلاثة أمثال العامل الثاني، حيث بلغ قسمة الجذر الكامن للعامل الأول على الجذر الكامن للعامل الثاني(3.412) وهو يزيد عن القيمة (٢) التي تعد محكا لأحادية البعد (Lian & Idris, 2006)، وأن نسبة التباين المفسر للعامل الأول بلغ (٣٥.٦٥%) وتعتبر هذه النسبة مثالية للحكم على الاختبار بأنه أحادي البعد وهي أن يفسر العامل الأول أكثر من (٢٠%) من التباين الكلي، كما بلغ الجذر الكامن للعامل الأول والثاني وفق الطريقة الجزئية لتقدير الدرجات(3.667، ١.٠٧٥) على الترتيب، ونسبة التباين المفسر(٣٦.٦٧%، ١٠.٧٥%) على الترتيب، ويتبين أن العامل الأول أكثر من ثلاثة أمثال العامل الثاني، حيث بلغ قسمة الجذر الكامن للعامل الأول على الجذر الكامن للعامل الثاني(3.411) وهو يزيد عن القيمة (٢)، وأن نسبة التباين المفسر للعامل الأول بلغ (٣٦.٦٧%) وتعتبر هذه النسبة مثالية للحكم على الاختبار بأنه أحادي البعد وهي أن يفسر العامل الأول أكثر من (٢٠%) من التباين الكلي.

٢. **الاستقلال الموضعي: ويشير إلى عدم تأثر الاجابة عن أي مفردة من مفردات الاختبار بالإجابة عن أي مفردة أخرى تأثرا سالباً أو موجباً أي أن افتراض الاستقلال الموضعي يتحقق إذا كان احتمال الاجابة الصحيحة عن مفردة ما من مفردات المقياس لا يرتبط باحتمال الاجابة الصحيحة عن أي مفردة أخرى. وقد قام الباحثان بالتحقق من الاستقلال الموضعي عن طريق حساب معاملات الارتباط بين مفردات الاختبار، وقد أشارت النتائج إلى انخفاض معاملات الارتباط بين المفردات وبعضها البعض مما يدل على تحقق الاستقلال الموضعي وفق الطريقة التامة والجزئية في تقدير الدرجات، حيث جاءت جميع معاملات الارتباط أقل من(٠.٧) وهو المحك الذي إذا زادت عنه قيمة معامل الارتباط يتم رفض فرض الاستقلال الموضعي (Linacre,2018).**

٣. **منحنى خاصية المفردة: هو احتمال التوصل إلى الاستجابة الصحيحة على المفردة كدالة للسمة الكامنة أو القدرة المقاسة في ضوء الأداء على المفردات .وتوضح الرسوم البيانية لهذا المنحنى احتمال التوصل إلى الاستجابة الصحيحة على المفردة كدالة للسمة الكامنة أو القدرة المقاسة في ضوء الأداء على المفردات. وللتحقق من هذا الافتراض، تم استخراج منحنى خصائص المفردة باستخدام برنامج (STATA) لكل من مفردات الاختبار ويظهر الشكل (١) منحنى خاصية المفردة رقم (٥) كمثال من مفردات الاختبار في صورته النهائية باستخدام نموذج التقدير الجزئي ،ويظهر الشكل (٢) منحنى خاصية الاختبار ككل باستخدام نموذج التقدير الجزئي.**



شكل (٣) منحنى خاصية المفردة للمفردة رقم (٥)



شكل (٤) منحنى خاصية الاختبار ككل

يتضح من شكل (١)،(٢) أنه كلما زادت قدرة الفرد زادت درجته في المفردة ويلاحظ أن معامل الصعوبة يؤثر في احتمالية الدرجة وليس هناك تأثير لمعامل التمييز إلا أن التخمين له تأثير من الدرجة (صفر حتى أقل من ١٠)، أي أن نموذج التقدير الجزئي يساعد في تحقيق توازي منحنيات خصائص المفردة.

٤. التحرر من السرعة :

تم التحقق من هذا الفرض من خلال إخفاق بعض الأفراد في الاجابة على المفردات بسبب انخفاض قدراتهم وليس بتأثير عامل السرعة على إجابتهم أي أنهم يفشلون في الاجابة بسبب محدودية قدرتهم.

نتائج البحث وتفسيرها ومناقشتها :

نتائج التحقق من الفرض الأول: "توجد مطابقة جيدة للبيانات لافتراضات نموذج التقدير الجزئي في ضوء طريقتي تقدير الدرجات(التامة – الجزئية)". وللتحقق من هذا الفرض تم تحليل مفردات الاختبار التحصيلي الالكتروني باستخدام البرنامج الاحصائي ستاتا (STATA v.14) لدراسة مدى مطابقة بيانات الاختبار التحصيلي لافتراضات نموذج التقدير الجزئي في ضوء طريقة تقدير الدرجات(التامة والجزئية)، حيث تم تقدير معالم الصعوبة وأخطاؤها المعيارية، وقيم مربع كاي ودلالتها الاحصائية لكل مفردة وفق طريقتي تقدير الدرجات(التامة والجزئية) باستخدام نموذج التقدير الجزئي، كما قام الباحثان بحساب ثبات الاختبار وفق طريقتي تقدير الدرجات(التامة والجزئية) باستخدام برنامج (STATA v.14) من خلال حساب معامل ألفا كرونباخ باستخدام أداة (Multivariate analysis)، والجداول التالية توضح النتائج:

جدول (٤)
معاملات الصعوبة للاختبار التحصيلي باستخدام نموذج التقدير الجزئي وفق طريقتي تقدير الدرجات (التامة – والجزئية)

المفردات	الطريقة التامة				الطريقة الجزئية			
	معامل الصعوبة	الخطأ المعياري لمعامل الصعوبة	مربع كاي	الدلالة	معامل صعوبة الخطوات	الخطأ المعياري لصعوبة الخطوات	مربع كاي	الدلالة
Q1	١.٦٧٦	٠.٩٩٨	١.٦٨	٠.٠٩٣	(١-٠) ٠.٢٥٥-	١.٦٨٨	٠.١٥-	٠.٨٨٠
					(٢-١) ١.٧٢٦	٢.٠٦٩	٠.٨٣	٠.٤٠٤
					(٣-٢) ٠.٦٢٥	٢.٤٥٧	٠.٢٥	٠.٧٩٩
					(٤-٣) ١٠.٨٣٩-	٢.٤٦١	٤.٤١-	٠.٠٠٠**
Q2	٠.٨٥٩	٠.٧٧٧	١.١١	٠.٢٦٩	(١-٠) ٥.٢١٢	١.٩٩٦	١.٦١	٠.١٠٩
					(٢-١) ١.٥٢٩-	٢.١٤٩	٠.٧١-	٠.٤٧٧
					(٣-٢) ٠.٥٦١-	١.٨٩٥	٠.٣٠-	٠.٧٦٧
					(٤-٣) ٧.٣٧٣-	١.٧٦٦	٤.١٧-	٠.٠٠٠**
Q3	١.٢٦٤	٠.٨٧٥	١.٤٤	٠.١٤٩	(١-٠) ٠.٠٧٥	٢.٤٦١	٠.٠٣	٠.٩٧٦
					(٢-١) ٠.٥١٧-	٢.٥٨٧	٠.٢٠-	٠.٨٤٢
					(٣-٢) ٠.٧٦٩	٢.٧٩٨	٠.٢٨	٠.٧٨٣
					(٤-٣) ١٢.٤٩١-	٢.٨١٣	٤.٤٤-	٠.٠٠٠**
Q4	٠.٩٩٤-	٠.٨٠٦	١.٢٣-	٠.٢١٧	(١-٠) ٣.٧٠٣	١.٦١٨	١.٢٩	٠.٢٨٠
					(٢-١) ٢.٠٠٩-	١.٦٦٤	١.٢١-	٠.٤٠٤
					(٣-٢) ٠.٠٠٧-	١.٤٣٨	٠.٠١-	٠.٧٩٩
					(٤-٣) ٤.٥٣٧-	١.٣٤٩	١.٣٦-	٠.٤٠١
Q5	٢.٥٤٠-	١.٢٩٨	١.٩٦-	٠.٠٥٥	(١-٠) ٢.٨٨٤	١.١٩٧	١.٤١	٠.١١٦
					(٢-١) ٠.٦٦٨-	١.٢٤٩	٠.٥٣-	٠.٥٩٣
					(٣-٢) ١.٠٦٤	١.٢٧٤	٠.٨٣	٠.٤٠٤
					(٤-٣) ٠.٤١٤-	١.٢٧٩	٠.٣٢-	٠.٧٤٦
Q6	٠.٣٢٩-	٠.٦٩٦	٠.٤٧-	٠.٦٣٥	(١-٠) ٥.٢٤٦	١.٧٤٦	١.٩٣	٠.١٦٣
					(٢-١) ١.٨٤٨-	١.٧٨٥	١.٠٤-	٠.٣٠٠
					(٣-٢) ٢.٩٠٧	٢.٠١٩	١.٤٤	٠.١٥٠
					(٤-٣) ٨.٢١٦-	٢.١٢٤	١.٨٧-	٠.١٨٠
Q7	٠.٧٢٧-	٠.٧٥٠	٠.٩٧-	٠.٣٣٣	(١-٠) ٠.١٥٧	١.٢٥٢	٠.١٣	٠.٩٠٠
					(٢-١) ١.١١٩	١.٤٥٢	٠.٧٧	٠.٤٤١
					(٣-٢) ٠.٣٣٤-	١.٥٣٠	٠.٢٢-	٠.٨٢٧
					(٤-٣) ٤.٩١٣-	١.٤٠٧	١.٤٩-	٠.٠٧٦
Q8	١.١٢٨	٠.٨٣٩	١.٣٤	٠.١٧٩	(١-٠) ٠.٠٧٣	٢.٤٦٠	٠.٠٣	٠.٩٧٦
					(٢-١) ٠.٥٣٣	٢.٧٩٨	٠.١٩	٠.٨٤٩
					(٣-٢) ٢.١٤٥-	٢.٦٨٢	٠.٨٠-	٠.٤٢٤
					(٤-٣) ١٠.٥٨٥-	٢.٢٨٠	٤.٦٤-	٠.٠٠٠**
Q9	٠.٩٩٤	٠.٨٠٧	١.٢٣	٠.٢١٨	(١-٠) ٥.٥٦٥	٢.٩٦٤	١.٨٨	٠.٠٦٠
					(٢-١) ٥.٠١٢-	٣.٠٠١	١.٦٧-	٠.٠٩٥
					(٣-٢) ١.٥٠٨-	١.٨١٣	٠.٨٣-	٠.٤٠٥
					(٤-٣) ٧.٤١٥-	١.٦٤٦	٤.٥٠-	٠.٠٠٠**
Q10	١.٢٦٤	٠.٨٩٥	١.٤٤	٠.١٤٩	(١-٠) ١.٥٤٢-	٢.٤٥٧	٠.٦٣-	٠.٥٣٠
					(٢-١) ٠.٥٠٢-	٢.٣١٤	٠.٢٢-	٠.٨٢٨
					(٣-٢) ٢.٤١٠-	٢.٠٦٢	١.١٧-	٠.٢٤٢
					(٤-٣) ٨.١٧٠-	١.٧٥٢	١.٦٦-	٠.٠٩٣

** دال عند مستوى دلالة (٠.٠١)

يتضح من نتائج الجدول السابق ما يلي:

١. **تراوحت تقديرات صعوبة مفردات الاختبار بين(–٢.٥٤، ١.٦٨) وفق الطريقة التامة لتقدير الدرجات باستخدام نموذج التقدير الجزئي وجميعها ضمن المدى المقبول وهو بين (–٢.٩٥، ٢.٩٥).**

٢. **تراوحت تقديرات صعوبة الخطوة من(٠–١) بين(٠.٨٤، ٧.٣٨)، والخطوة (١–٢) بين(–٧.٤٥، ٢.٦٣)، والخطوة(٢–٣) بين(–٣.٢٩، ٢.٥٧)، والخطوة (٣–٤) بين(–١٦.٥٣، ١,٠٨) وفق الطريقة الجزئية لتقدير الدرجات باستخدام نموذج التقدير الجزئي.**

٣. **جميع قيم مربع كاي لمفردات الاختبار غير دالة إحصائيا وفق الطريقة التامة لتقدير الدرجات.**

٤. **جميع قيم مربع كاي لخطوات مفردات الاختبار غير دالة إحصائيا ما عدا الخطوة من(٣–٤) دالة احصائيا في المفردات(١، ٢، ٣، ٨، ٩) عند مستوى دلالة (٠.٠١) وفق الطريقة الجزئية في تقدير الدرجات باستخدام نموذج التقدير الجزئي.**

جدول (٥)

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

طريقة تقدير الدرجات	التامة	الجزئية
معامل ثبات الاختبار	٠.٦٠	٠.٧٦

يتبين من الجدول السابق أن معامل ثبات الاختبار باستخدام الطريقة التامة في تقدير الدرجات يساوي (٠.٦٠)، ومعامل ثبات الاختبار باستخدام الطريقة الجزئية في تقدير الدرجات يساوي (٠.٧٦)، وهي معاملات ثبات جيده تؤكد على صلاحية الاختبار ومفرداته في جميع الحالات، ومطابقة بيانات الاختبار التحصيلي لافتراضات نموذج التقدير الجزئي.

مناقشة وتفسير نتائج التحقق من الفرض الأول:

تُظهر نتائج التحقق من الفرض الأول وجود مطابقة جيدة للبيانات لافتراضات نموذج التقدير الجزئي في ضوء طريقتي تقدير الدرجات(التامة – الجزئية). ويعد هذا الاجراء ضروريا قبل الاعتماد على أي نموذج من نماذج الاستجابة للمفردة. وهذا يتفق مع ما نتائج دراسات كل من (أحمد،٢٠١٧Alquraan & ; alkuwaiti,2017; Karakaya,2012)، حيث أشاروا إلى وجود مطابقة جيدة للبيانات مع النموذج المستخدم من خلال حساب

الخصائص السيكومترية للمفردات، ودراسة (Innabi & Dodeen,2006) التي اعتمدت على نموذج التقدير الجزئي في تقدير معالم المفردات، وكذلك دراسة (Brad,2003) التي ركزت تقدير معالم المفردات(معلم الصعوبة) والأفراد (قدرات الأفراد) وفق الطرق الأربعة لتقدير الدرجات باستخدام نموذج التقدير الجزئي.

ويرى الباحثان أن هذه النتيجة ترجع إلى أن جميع تقديرات معالم صعوبة المفردات تقع ضمن المدى المقبول، وأن جميع قيم مربع كاي لمفردات الاختبار غير دالة إحصائيا وفق طريقتي تقدير الدرجات التامة والجزئية، وكانت قيمة معامل ثبات الاختبار مقبولة سيكومتريا وفق الطريقة التامة والجزئية لتقدير الدرجات، مما يدل على مطابقة البيانات لنموذج التقدير الجزئي.

نتائج التحقق من الفرض الثاني: ينص الفرض الثاني على أنه " تختلف تقديرات قدرات الأفراد وأخطاؤها المعيارية باختلاف طريقتي تقدير الدرجات(التامة – والجزئية) باستخدام نموذج التقدير الجزئي." وللتحقق من هذا الفرض تم حساب تقديرات قدرات الأفراد وفق طريقتي تقدير الدرجات التامة والجزئية باستخدام البرنامج الاحصائي ستاتا (STATA) من خلال أداة (Postestimation)، وتوضح النتائج أن تقديرات القدرة باستخدام الطريقة الجزئية في تقدير الدرجات أعلى من الطريقة التامة في تقدير الدرجات، وجاء الخطأ المعياري لتقديرات القدرة باستخدام الطريقة التامة لتقدير الدرجات أعلى من الخطأ المعياري لتقديرات القدرة باستخدام الطريقة الجزئية لتقدير الدرجات. ولتحديد الفروق في تقديرات قدرات الأفراد وفق طريقتي تقدير الدرجات التامة والجزئية تم استخدام برنامج التحليل الاحصائي SPSS وكانت النتائج كالتالي :

جدول (٦)
نتائج اختبار(ت) لدلالة الفروق بين متوسطات تقديرات قدرات الأفراد 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	0.744	0.018	١٠٦	٤.٨٥١	**٠.٠٠١
	الجزئية	0.755	0.021			
الخطأ المعياري لقدرات الأفراد	التامة	o.648	0.022	١٠٦	5.093	**0.000
	الجزئية	0.624	0.047			

** دال عند مستوى دلالة (٠.٠١)

يتضح من الجدول السابق أنه توجد فروق ذات دلالة احصائية بين متوسطات تقديرات قدرات الأفراد وفق الطريقة التامة والجزئية في تقدير الدرجات لصالح الطريقة الجزئية، كما أنه توجد فروق ذات دلالة إحصائية بين متوسطات الأخطاء المعيارية لقدرات الأفراد وفق الطريقة التامة والجزئية في تقدير الدرجات لصالح الطريقة التامة، مما يدل على تحقق الفرض الثاني للبحث.

وتبين النتائج أن الطريقة الجزئية في تقدير الدرجات أكثر دقة من الطريقة التامة في تقدير الدرجات في تقدير معالم قدرات الأفراد، وهذا يعني أنه كلما روعيت المعرفة الجزئية للأفراد بالمحتوى المراد قياسه كانت دقة تقديرات معالم قدرات الافراد أفضل، وتتفق هذه النتيجة مع نتائج دراسة (Brad,2003) التي أشارت إلى وجود علاقة ارتباطية شبه تامة بين الدرجات الخام وتقديرات قدرات الأفراد باستخدام الطرق الأربعة لتقدير الدرجات، وأن الطريقة الرابعة في تقدير الدرجات كانت أكثر دقة من الطريقة الاولى والثانية والثالثة في تقدير قدرات الأفراد.

وقد تعزي هذه النتيجة إلى أن الطريقة الجزئية في تقدير الدرجات تراعي المعرفة الجزئية بأعلى درجة ممكنة تتباين فيها الدرجات التي يستحقها الطالب بشكل أكبر، حيث تقدر درجة المفحوص على المفردة الواحدة بإحدى الدرجات (٤، ٣، ٢، ١، ٠)، وبالتالي فان هذه الطريقة يمكن من خلالها الكشف بشكل أكبر عن الفروق والاختلافات بين قدرات الطلاب، ومن ثم يمكن من خلالها تقدير قدرات الطلاب بدرجة عالية من الدقة، بينما تقل الدرجات

التي يستحقها المفحوص عن إجابته عن المفردة في الطريقة التامة الدرجات(١، ٠)، وبالتالي تقل القدرة في الكشف عن الفروق والاختلافات بين قدرات المفحوصين، أي أن زيادة عدد فئات تقدير الدرجات قد يكون السبب في زيادة دقة تقدير قدرات الأفراد.

نتائج التحقق من الفرض الثالث: **"توجد فروق دالة احصائيا في تقدير الأداء التفاضلي للاختبار التحصيلي الالكتروني باختلاف نوع الطالب(ذكور – اناث) وفق الطريقة التامة لتقدير الدرجات باستخدام نموذج التقدير الجزئي." وللتحقق من هذا الفرض تم تحليل مفردات الاختبار التحصيلي الالكتروني باستخدام البرنامج الاحصائي (STATA v.14) لتقدير الأداء التفاضلي للاختبار التحصيلي الالكتروني وفق الطريقة التامة لتقدير الدرجات باستخدام اختبار مانتل هانزل، واختبار الانحدار اللوجستي ، والجدول التالي يوضح النتائج:**

جدول (٧)

اختبار مانتل هانزل، والانحدار اللوجيستي للكشف عن الاداء التفاضلي وفق الطريقة التامة لتقدير الدرجات باستخدام نموذج التقدير الجزئي

مانتل هانزل				الانحدار اللوجيستي					
المفردة	كا^2	الدلالة	الأداء التفاضلي	المفردة	منتظم		غير منتظم		
					كا^2	الدلالة	الأداء التفاضلي	الدلالة	كا^2
1	5.42	*0.0199	**لصالح الذكور**	**1**	9.71	**0.0018	**يوجد**	0.4690	0.52	لا يوجد
2	0.33	0.5682	لا يوجد	**2**	0.44	0.5071	لا يوجد	0.1484	2.09	لا يوجد
3	0.65	0.4201	لا يوجد	**3**	1.26	0.2621	لا يوجد	*0.0373	4.34	**يوجد**
4	1.03	0.3096	لا يوجد	**4**	1.41	0.2347	لا يوجد	0.3931	0.73	لا يوجد
5	0.06	0.8133	لا يوجد	**5**	0.03	0.8672	لا يوجد	0.2739	1.20	لا يوجد
6	0.57	0.4483	لا يوجد	**6**	0.13	0.7196	لا يوجد	0.2303	1.44	لا يوجد
7	0.01	0.9689	لا يوجد	**7**	0.06	0.8134	لا يوجد	0.4427	0.59	لا يوجد
8	0.85	0.3576	لا يوجد	**8**	0.80	0.3699	لا يوجد	0.8553	0.03	لا يوجد
9	0.11	0.7394	لا يوجد	**9**	0.34	0.5604	لا يوجد	0.7576	0.10	لا يوجد
10	1.50	0.2205	لا يوجد	**10**	1.30	0.2535	لا يوجد	0.2632	1.25	لا يوجد

**** دال عند مستوى دلالة (٠.٠١)**

يتضح من الجدول السابق أنه لا توجد فروق ذات دلالة إحصائية بين متوسطات درجات الذكور والاناث في تقدير الاداء التفاضلي لمفردات الاختبار باستخدام طريقة مانتل هانزل ما عدا المفردة رقم(١) أظهرت أداءاً تفاضلياً دال إحصائيا عند مستوى دلالة (٠.٠٥) لصالح الذكور، كما أنه لا توجد فروق ذات دلالة إحصائية بين الذكور والاناث في تقدير الأداء التفاضلي المنتظم لمفردات الاختبار باستخدام طريقة الانحدار اللوجستي ما عدا

المفردة رقم (١) أظهرت أداءاً تفاضلياً منتظما دال إحصائيا عند مستوى دلالة (٠.٠١)، بالإضافة إلى أنه لا توجد فروق ذات دلالة إحصائية بين الذكور والاناث في تقدير الأداء التفاضلي غير المنتظم لمفردات الاختبار باستخدام طريقة الانحدار اللوجستي ما عدا المفردة (٣) أظهرت أداءاً تفاضلياً غير منتظم دال إحصائيا عند مستوى دلالة (٠.٠٥).

وتشير نتائج تحليل الأداء التفاضلي للمفردات إلى أن عدد المفردات ذات الأداء التفاضلي لا تزيد عن مفردة واحدة فقط وفق متغير نوع الطالب باستخدام طريقة مانتل هانزل، فالمفردة رقم(١) التي تتعلق بـ " القياس النفسي هو عملية ..... لصفات أو خصائص الأشياء أو الأشخاص في ضوء محكات محددة" فهي ذات أداء تفاضلي لصالح الذكور، مما يعني أن لدى الذكور مهارة أعلى من الإناث في حفظ وتذكر المعلومات، وباستخدام طريقة الانحدار اللوجستي توجد مفردة واحدة فقط ذات أداء تفاضلي منتظم وهي المفردة رقم (١)، ومفردة واحدة فقط ذات أداء تفاضلي غير منتظم وهي المفردة رقم (٣) التي تتعلق بـ " إذا اجتمع المعلم بمجموعة من الطلاب وألقى عليهم عددا من الأسئلة بانه يستخدم بذلك .....".

وقد ترجع هذه النتيجة إلى العلاقة بين الأداء التفاضلي للمفردة (DIF) والأداء التفاضلي للمموهات(DDF)، فالأداء التفاضلي للمموهات يعطي صورة أوضح عن أنماط اختيار المفحوصين لبدائل الاجابة، ويفيد في فهم الأداء التفاضلي للمفردات، فالأثر الكبير الناتج عن الأداء التفاضلي لأحد المموهات مقارنة مع الأثر الناتج عن بقية المموهات، وهذا يعني أن ذلك المموه كان جذاباً بإحدى المجموعتين دون الأخرى، مما يدل على أن سبب وجود الأداء التفاضلي للمفردة يتمثل في خصائص هذا المموه أو البديل الخاطئ أو في التفاعل بين خصائص هذا المموه وبين محتوى المفردة، ويتطلب الوضع في هذه الحالة إجراء تعديل للمموه وليس للمفردة أو الاجابات الصحيحة، وتساعد مراجعة المعلومات المتوافرة عن الاداء التفاضلي للمموهات في مراجعة المفردات ذات الأداء التفاضلي التي تنعكس بشكل ايجابي على الاختبار ككل لضمان درجة أعلى في عدالته.

ويفسر الباحثان هذه النتيجة إلى أن مستوى صعوبة جميع مفردات الاختبار متقاربة مع بعضها ما عدا المفردة رقم (١) التي تبدى صعبة إلى حد ما، والمفردة رقم (٣) التي

تحتاج إلى مهارات تفكير عليا والتي لا يستطيع الإجابة عليهما غير الطالب ذو القدرة العقلية المرتفعة والذي أتقن المحتوى التعليمي بشكل كامل.

وتتفق نتيجة هذا البحث مع دراسة (Shanmugam,2020) التي أشارت إلى أنه توجد فروق دالة إحصائيا في الأداء التفاضلي لصالح الذكور في المفردات التي تعتمد على مهارات التفكير العليا، ودراسة (Li–An We & Tsai ,2010) التي أشارت إلى أن إجراء الانحدار اللوجيستي لدالة الأداء التفاضلي كان حساسا فقط للمفردات ذات حجم دالة الأداء التفاضلي الكبير.

وتختلف هذه النتيجة مع دراسات كل من (Karakaya,2012) التي أكدت على وجود مفردات تبدي أداءاً تفاضلياً لكل من الذكور والاناث بعضها لصالح الذكور والأخرى لصالح الاناث، ودراسة (Innabi & Dodeen,2006) التي توصلت إلى أن (٣٧) مفردة تبدي أداءاً تفاضلياً بنسبة (٣٠%) من مفردات الاختبار، منها (١٧) مفردة لصالح الذكور، (٢٠) مفردة لصالح الاناث، وقد كانت أغلب المفردات التي أظهرت أداءاً تفاضليا لصالح الاناث، ودراسة جرادات(٢٠٠٣) التي اكدت على أن طريقة الصعوبة المحولة كانت أفضل من طريقة مانتل هانزل في الكشف عن المفردات المتحيزة لمتغير الجنس، وربما اختلاف النتائج يرجع إلى اختلاف البرنامج المستخدم في الكشف عن الاداء التفاضلي، أو المحتوى التدريسي للمقرر الدراسي، أو حجم العينة المستخدمة.

نتائج التحقق من الفرض الرابع: "أفضل نموذج للحكم على كفاءة المعلومات في الاختبار التحصيلي الالكتروني كان وفق طريقة تقدير الدرجات الجزئية باستخدام نموذج التقدير الجزئي." وللتحقق من هذا الفرض تم تحليل مفردات الاختبار التحصيلي الالكتروني باستخدام البرنامج الاحصائي ستاتا (STATA) من خلال أداة (Postestimation) لحساب محك المعلومات (AIC) وفق طريقتي تقدير الدرجات التامة والجزئية، والجدول التالي يوضح النتائج:

جدول (٨)
محك دالة المعلومات (AIC) وفق طريقتي تقدير الدرجات (التامة والجزئية) باستخدام نموذج التقدير الجزئي

النموذج	الطريقة التامة	الطريقة الجزئية
Akaike's	٤١٠٧.٥٨	١٧٦٢.٣٤

وأظهرت النتائج أن قيمة محك دالة المعلومات باستخدام نموذج(Akaike's) (AIC) يساوي (٤١٠٧.٥٨) وفق الطريقة التامة لتقدير الدرجات، ويساوي (١٧٦٢.٣٤) وفق الطريقة الجزئية في تقدير الدرجات، أي أن قيمة (AIC) للنموذج الثاني باستخدام الطريقة الجزئية لتقدير الدرجات أقل من قيمة (AIC) 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، وهذا يدل على قبول الفرض الرابع للبحث.

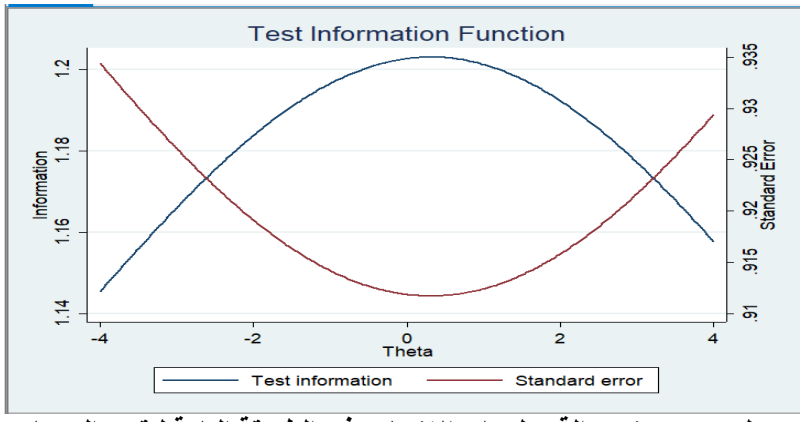
مناقشة وتفسير نتائج التحقق من الفرض الرابع:

تُظهر نتائج التحقق من الفرض الرابع أن أفضل نموذج للحكم على كفاءة المعلومات هو النموذج الذي يعتمد على الطريقة الجزئية في تقدير الدرجات وفق. وعادة عند اختيار معيار لاختيار النموذج، يكون قبول حقيقة أن النماذج تقارب الواقع فقط، وبالنظر إلى مجموعة البيانات، فإن الهدف هو تحديد أي من النماذج المرشحة تقارب البيانات بشكل أفضل، وهذا يضمن قدر الامكان تقليل فقدان المعلومات. ويعد محك المعلومات(AIC) معياراً للاختيار من بين النماذج الاحصائية المتداخلة، كما يعد بشكل أساسي مقياساً في تقدير الجودة لكل نموذج من النماذج الاحصائية المتاحة من حيث صلتها ببعضها البعض لمجموعة معينة من البيانات، مما يجعلها طريقة مثالية لاختيار النماذج، كما تساعد الباحثان على تقدير المعلومات المفقودة اذا تم استخدام نموذج معين لعرض العملية التي أنتجت البيانات.

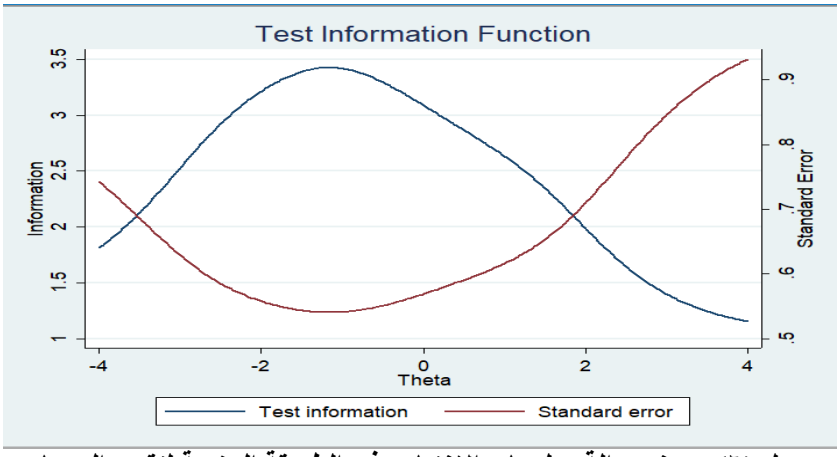
وتتفق هذه النتيجة مع نتائج دراسة (Brad,2003) التي كشفت عن وجود علاقة ارتباطية بين الدرجات الخام المستمدة من الطريقة الجزئية لتقدير الدرجات باستخدام نموذج التقدير الجزئي، وأن أفضل نموذج للحكم على كفاءة المعلومات كان وفق الطريقة الجزئية لتقدير الدرجات باستخدام نموذج التقدير الجزئي، كما تتفق مع بعض الأطر النظرية التي أثبتت أن الطريقة الجزئية أفضل من الطريقة التامة في تقدير الدرجات باستخدام النماذج المعلمية(Lee ,Wollak & Douglas,2009). بينما تختلف هذه النتيجة مع نتائج دراسة كل من (Alquraan & alkuwaiti,2017) التي أثبتت أن نموذج راش المعدل من أكثر نماذج نظرية الاستجابة للمفردة فاعلية في تقديم معلومات جيدة حول تقدير بعض معالم المفردات، ودراسة (Karakaya,2012) التي أظهرت أن نماذج الاستجابة للمفردة ثنائية

الاستجابة تعطي نتائج جيدة لدالة المعلومات لمفردات الاختبار، وقد يرجع هذا الاختلاف إلى اختلاف نموذج الاستجابة للمفردة المستخدم أو برنامج التحليل الاحصائي المستخدم في تحليل النتائج أو خصائص العينة التي طبق عليها الاختبار.

ويرى الباحثان أن هذه النتيجة ترجع إلى أن منحنى دالة معلومات الاختبار وفق الطريقة الجزئية في تقدير الدرجات أفضل في إعطاء المعلومات عن منحنى دالة معلومات الاختبار وفق الطريقة التامة في تقدير الدرجات باستخدام نموذج التقدير الجزئي كما يتضح من الجدولين (١) ،(٢) :



جدول (٥): يوضح دالة معلومات الاختبار وفق الطريقة التامة لتقدير الدرجات



جدول (٦): يوضح دالة معلومات الاختبار وفق الطريقة الجزئية لتقدير الدرجات

ويفسر الباحثان هذه النتيجة بأن الطريقة الجزئية في تقدير الدرجات تراعي المعرفة الجزئية بأعلى درجة ممكنة تتباين فيها الدرجات التي يستحقها الطالب بشكل أكبر، وبالتالي فان هذه الطريقة تعطي أكبر قدر ممكن من المعلومات الواضحة.

التوصيات والبحوث المقترحة:

ويمكن في ضوء ما توصل إليه البحث من نتائج اقترح بعض التوصيات مثل:

١. **استخدام نماذج الاستجابة للمفردة متعددة الاستجابات في تقدير معالم المفردات مما ينعكس ايجاباً على جودة التقديرات.**
٢. **ضرورة توجيه الباحثين لقضية التحقق من عدالة الاختبارات والمقاييس النفسية.**
٣. **التأكد من خلو مفردات الاختبارات التحصيلية الالكترونية من المفردات ذات الأداء التفاضلي.**
٤. **الاهتمام ببناء الاختبارات التحصيلية وفقا لخطوات سليمة ومناسبة، وإعداد مموهات ذات خصائص مناسبة.**

البحوث المقترحة:

١. **دراسة أثر المفردات ذات الاداء التفاضلي على مؤشرات صدق البناء الداخلي للاختبار التحصيلي الالكتروني باستخدام نموذج التقدير الجزئي.**
٢. **إجراء دراسات أخرى مماثلة باستخدام مقاييس التقدير المتدرجة والاختبارات بأشكالها المختلفة للوصول إلى نتائج أكثر قابلية للتعميم.**
٣. **إجراء دراسات عن الاداء التفاضلي لمفردات أداة تقييم من النوع المقالي، فلا توجد أي دراسة عربية تطرقت لهذ الموضوع.**

المراجع

أولاً: المراجع العربية:

أحمد، إبراهيم (٢٠١٧). الكشف عن الأداء التفاضلي لمفردات اختبار تحصيلي في الرياضيات لدى تلاميذ المرحلة الإعدادية وفقاً لنظريتي القياس التقليدية والاستجابة للمفردة . *رسالة دكتوراه غير منشورة* ، جامعة الزقازيق: مصر.

أوعلا، بوشري؛ والمطارنة، أحمد(٢٠١٨). الأداء التفاضلي لفقرات اختبار تحديد المستوى في مادة اللغة الانجليزية المطبق على طلبة جامعة مؤتة. *مجلة العلوم التربوية والنفسية*، جامعة البحرين، ١٩(٢)،٤٤٩–٢٧٥.

جرادات، عبد الله(٢٠٠٣). مقارنة بين طريقيتي مانتل هانزل وطريقة الصعوبة المحولة في الكشف عن تحيز المفردات. *رسالة ماجستير غير منشورة*، جامعة مؤته: الكرك.

علام، صلاح الدين(٢٠٠٥). *نماذج الاستجابة للمفردات الاختيارية أحادية البعد ومتعددة الأبعاد وتطبيقاتها في القياس النفسي والتربوي*، ط١ ،القاهرة: دار الفكر العربي.

اللبدي، نزار(٢٠٠٨). مقارنة أربعة طرائق للكشف عن دالة الفقرة التفاضلية(دراسة محاكاة). *رسالة دكتوراه غير منشورة*، الجامعة الأردنية: الأردن.

ثانياً: المراجع الأجنبية:

Abbott, M. L. (2007). A confirmatory approach to differential item functioning on an ESL reading assessment. *Language testing*, 24(1), 7-36.

Abdelwahab,M.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.

AlQuraan, M., & Kuwaiti, A. A. (2017). Differential Item Functioning in Students Rating of Teaching Effectiveness Surveys in Higher Education According to Academic Disciplines: Data from a Saudi University. *Journal of Educational and Psychological Studies* [JEPS], 11(4), 770-780.

Alquraan,M. & Alkuwaiti,A. (2017). Differential Item Functioning in Students Rating of Teaching Effectiveness Surveys in Higher Education According to Academic Disciplines: Data from a Saudi University. *Journal of Educational and Psyvhological Studies* – Sultan Qaboos University,11(4),770-780.

American Educational Research Association. (2014). Standards for educational and psychological testing. *American Educational Research Association American Psychological Association National Council on Measurement in Education*.

Ariffin, S. R., Idris, R., & Ishak, N. M. (2010). Differential item functioning in Malaysian generic skills instrument (MyGSI). *Jurnal Pendidikan Malaysia(Malaysian Journal of Education)*, 35(1), 1-10.

Barnabas,C.(2012).Analysis of Gender-Related Differential Item Functioning In Mathematics Multiple Choice Items Administered by West African Examination Council(WAEC),*Journal of Education and Practice*,3(8),71-78.

Brassil, C. E., & Couch, B. A. (2019). Multiple-true-false questions reveal more thoroughly the complexity of student thinking than multiple-choice questions: a Bayesian item response model comparison. *International Journal of STEM Education*, 6(1), 16.

Ellis, B. B., & Raju, N. S. (2003). Test and Item Bias: What They Are, What They Aren't, and How To Detect Them. *Assessment issue for teachers, counselors and administrator*. (ERIC Document Reproduction Service No.ED480042).

Finch, H., French, B. F., & Immekus, J. C. (2016). *Applied psychometrics using SPSS and AMOS*. IAP.

Haladyna, T. M. (2004). *Developing and validating multiple-choice test items*. Routledge. NJ: Lawrence Erlbaum Associates.

Innabi, H., & Dodeen, H. (2006). Content Analysis of Gender-related Differential Item Functioning TIMSS Items in Mathematics in Jordan. *School Science and Mathematics*, 106(8), 328-337.

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.

Karami, H. (2011). Detecting gender bias in a language proficiency test. *International Journal of Language Studies*, 5(2), 27-38.

Li-An Wu & Tsai,R.(2010). A Comparison of Three Polytomous Dif DIF Detection Methods. *Quiz Annual Digest of Statistics*,18,1-12.

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

Linacre, J. (2018). *Auserguidetowinsteps*, RASCHmodelcomputerprogram. available online: www.winsteps.com.

Meade, A. W., & Wright, N. A. (2012). Solving the measurement invariance anchor item problem in item response theory. *Journal of Applied Psychology*, 97(5), 1016.

Middleton, K., & Laitusis, C. C. (2007). Examining test items for differential distractor functioning among students with learning disabilities. *ETS Research Report Series*, 2007(2), i-34.

Nakano, T., & Primi, R. (2014). Rasch-master's partial credit model in the assessment of children's creativity in drawings. *The Spanish Journal of Psychology*, 17-31.

Ogbebor,U. & Onuka,A.(2013).Differential Item functioning method as an item bias indicator. *Educational Research*,4(4),367-373.

Pae, T. I. (2004). DIF for examinees with different academic backgrounds. *Language testing*, 21(1), 53-73.

Penfield, R. D.(2008). An odds ratio approach for assessing differential distractor functioning effects under the nominal response model. *Journal of Educational Measurement*, 45(3), 247-269.

Penfield, R. D., & Camilli, G. (2006). *5 Differential Item Functioning and Item Bias*. Handbook of statistics, 26, 125-167.

Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: a critical review of the literature and recommended remedies. *Journal of applied psychology*, 88(5), 879.

Pokropek, A., & Kondratek, B. (2012). Test equating. Definitions and examples of applications. *EDUKACJA*, 120(4), pp 52-71.

Schauberger, G., & Mair, P. (2020). A regularization approach for the detection of differential item functioning in generalized partial credit models. *Behavior research methods*, 52(1), 279-294.

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.

Shindi, Y. A., & Kazem, A. M. (2018). Sex Differential Item Functioning for Mathematics test in Cognitive Development Program in Sultanate of Oman by Mental-Haenszel and Item Characteristic Curve Methods. *Int. J. Learn. Man. Sys*, 6(2), 61-73.

Tsaousis, I., Sideridis, G., & Al-Saawi, F. (2018). Differential Distractor Functioning as a Method for Explaining DIF: The Case of a National Admissions Test in Saudi Arabia. *International Journal of Testing*, 18(1), 1-26.

Tutz, G., & Berger, M. (2016). Item-focussed trees for the identification of items in differential item functioning. *Psychometrika*, 81(3), 727-750.

Uiterwijk, H., & Vallen, T. (2005). Linguistic sources of item bias for second generation immigrants in Dutch tests. *Language Testing*, 22(2), 211-234.

Wang, W. C., & Su, Y. H. (2004). Effects of average signed area between two item characteristic curves and test purification procedures on the DIF

detection via the Mantel-Haenszel method. *Applied Measurement in Education*, 17(2), 113-144.

Welkenhuysen-Gybels, J. (2004). The performance of some observed and unobserved conditional invariance techniques for the detection of differential item functioning. *Quality and quantity*, 38(6), 681-702.

Woods, C. M. (2009). Empirical selection of anchors for tests of differential item functioning. *Applied Psychological Measurement*, 33(1), 42-57.

Wu, B. C. (2003). Scoring Multiple True False Items: A Comparison of Summed Scores and Response Pattern Scores at Item and Test Levels. (*ERIC Document Reproduction Service No. ED476 148).*

Yao, L., & Schwarz, R. D. (2006). A multidimensional partial credit model with associated item and test statistics: An application to mixed-format tests. *Applied psychological measurement*, 30(6), 469-492.

Zheng, Y. (2016). Online calibration of polytomous items under the generalized partial credit model. *Applied psychological measurement*, 40(6), 434-450.