مجلة
كلية التربية
مجلة علمية دورية محكمة

رئيس مجلس إدارة المجلة

أ.د. محمد حماد هندى

نائب رئيس مجلس إدارة المجلة

أ.د. أسامة محمود قرنى

رئيس التحرير

أ.د. سليمان محمد سليمان

نائب رئيس التحرير

أ.د. سهام حنفى محمد

مديرى التحرير

أ.م.د. عزام عبدالنبى أحمد

أ.م.د. مروه عزت عبدالجواد

أ.م.د.عبدالمعز محمد إبراهيم

الهيئة الإستشارية الداخلية

أ.د. أمال عبدالسميع باظة

أ.د. نجوى يوسف جمال الدين

أ.د. ماهر إسماعيل صبرى

الهيئة الإستشارية الخارجية

أ.د. صالحة عبدالله عيسان

أ.د.على كاظم علي

أ.د.صالحة الزبير

Dr. Johnson Jessica

Dr. Valarie.

Dr. Sosa Eric

Dr . Mack C Shelley

Dr. Abdullah k abdullah

سكرتير تحرير(محرر)

م.م. أمل أحمد جابر

سكرتير تحرير (محرر صفحة)

م.م. ياسمين صلاح رشاد

سكرتير تحرير(محرر لغوي)

م. أحمد جمال عبدالجيد

الهيئة الإدارية

مسئول مالي أ . عمرو عبدالعاطى عبدالعزيز

مسئول إداري أ . كريمة عبد العال سليمان

استخدام بعض طرق نظرية الاستجابة للمفردة في الكشف عن الأداء التفاضلي لمتغير الجنس في مفردات اختبار القدرات العامة وأثرها على دقة تقدير القدرة والخصائص السيكومترية للاختبار

أ.د/ إسماعيل حسن فهيم الوليلي
أستاذ القياس والتقويم والإحصاء التربوي
بالمركز القومي للامتحانات والتقويم التربوي

د / سعيد بن عبد الله مبارك الدوسري
أستاذ القياس والتقويم المشارك
قسم علم النفس . كلية التربية . جامعة حائل

المستخلص

هدفت الدراسة للكشف عن المفردات ذات الأداء التفاضلي تبعاً لمتغير الجنس في مفردات اختبار القدرات العامة، وأثرها على دقة تقدير القدرات والخصائص السيكومترية للاختبار. تم استخدام بيانات أحد النماذج لاختبار القدرات العامة لعام ٢٠١٦ والمكون من ٩٦ مفردة، وبلغ عدد العينة ٧٤٠٨ طالباً وطالبة، منهم ٣٥٩٩ من الذكور و ٣٨٠٩ من الإناث باستخدام طريقتي مربع كاي لكاميللي Camilli Chi-Square، والفروق في المساحات تحت منحنيات خصائص المفردات .

أوضحت نتائج الدراسة التوافق بين الطريقتين في الكشف عن ٢٦ مفردة ذات أداء تفاضلي في اختبار القدرات العامة، وكان اغلبها لصالح الطلبة الذكور، وأن لها تأثير على دقة تقدير القدرات والخصائص السيكومترية للاختبار. كما أوضحت النتائج وجود فروق ذات دلالة إحصائية بين تقدير القدرات لكل من الذكور والإناث، وكذلك بين تقدير دوال معلومات الاختبار قبل وبعد حذف المفردات ذات الأداء التفاضلي، مما يدل على أن حذف هذه المفردات من الاختبار قد أدى إلى خفض قيم الأخطاء المعيارية لتقدير القدرات وتحقيق الدقة في تقدير دوال المعلومات المتعلقة بها، الأمر الذي يُمكن أن يُسهم في تحقيق الدقة والكفاءة في تقدير قدرات الطلبة.

الكلمات المفتاحية : الأداء التفاضلي للمفردة، منحنى خصائص المفردة، مربع كاي لكاميللي، نظرية الاستجابة للمفردة، اختبار القدرات العامة

Abstract :

The study aimed to discover the vocabulary with differential performance according to the gender variable in the general ability test vocabulary, and its effect on the accuracy of estimating the capabilities and the psychometric properties of the test. One model data was used to test the 2016 general abilities of 96 individuals, the sample number was 7,408 male and female students, 3599 were male and 3809 female using Camilli Chi-Square and the differences in the spaces under the curves of the vocabulary properties.

The results of the study showed the compatibility between the two methods in detecting 26 items with differential performance in the general abilities test, most of which were in favor of male students, and that they have an impact on the accuracy of estimating abilities and the psychometric properties of the test. The results also showed that there are statistically significant differences between the estimation of capabilities for both males and females, as well as between estimating the functions of test information before and after removing the items with differential performance, which indicates that the deletion of these items from the test has reduced the values of standard errors for estimating capabilities and achieving accuracy In estimating the functions of the information related to it, which can contribute to achieving accuracy and efficiency in assessing students' abilities.

مقدمة :

تمثل الاختبارات أحد أهم الوسائل المستخدمة فى عمليات التقويم التربوى , وتعتمد عليها المؤسسات التربوية بصورة أساسية للتعرف على مقدار التقدم فى تحصيل المتعلمين ومدى تحقيقهم للأهداف التعليمية . لذلك تعد نتائج الاختبارات أساساً مهماً يستند إليه التربويون فى صنع واتخاذ كثير من القرارات المصيرية لكل من تطبق عليه تلك الاختبارات , وبالتالى فإن دقة وسلامة القرارات التربوية تتوقف على مدى جودة وكفاءة الاختبارات المستخدمة فى عملية التقويم , وكذلك الأساليب المستخدمة فى تفسير وتحليل نتائجها .

وتعتمد الثقة في نتائج الاختبارات والاستخدام الفعال لها بدرجة كبيرة على مدى تحقق الخصائص السيكومترية في الاختبار، وبالتالي يركز مصمموا الاختبارات على تحقيق خصائص الصدق والثبات في الاختبارات للتأكيد على أنها مقاييس مقننة ويمكن الاستعانة بها واستخدامها بكفاءة في قياس ما وضعت لقياسه، والاستناد إلى نتائجها في صنع واتخاذ القرارات المصيرية بدرجة عالية من الثقة.

ونتيجة للاهتمام المستمر من قبل المتخصصين في القياس والتقويم بتطوير الاختبارات ورفع كفاءتها، اتجه الاهتمام في الآونة الأخيرة إلى خاصية سيكومترية مهمة تمثل أحد أهم خصائص المفردات الاختبارية وهي خاصية الأداء التفاضلي للمفردات Differential Item Function (DIF)، والتي لقيت اهتماماً واسعاً من قبل الخبراء والمختصين في القياس والتقويم لما لها من أهمية في ضمان تحرر الدرجة التي يحصل عليها المتعلم عن استجابته للمفردة الاختبارية من تأثير متغيرات أخرى غير القدرة التي يقيسها الاختبار كالعِرق، والثقافة، والجنسية، والجنس.

ويُعرف الأداء التفاضلي للمفردة بأنه مؤشر إحصائي يستخدم لتحديد المفردات التي لا تُؤدي وظيفتها بشكل منتظم وثابت مع المجموعات أو الفئات المختلفة التي تُطبق عليها. ففي بعض الأحوال أثناء تطبيق الاختبار على أكثر من مجموعة أو فئة تبدو بعض المفردات أكثر صعوبة لدى مجموعة أو فئة دون أُخرى، وهذا يتطلب تحليلاً منطقياً للسبب الذي أدى إلى ذلك وتحديد تلك المفردات، وفي ضوء ذلك تُوصف بأنها تظهر أداء تفاضلياً بين

المجموعات أو الفئات التي تُطبق عليها، أو أنها مُتحيزة لإحدى تلك المجموعات (Schumacker, 2010).

وتجدر الإشارة إلى أن هناك فرق بين مفهومي التحيز والأداء التفاضلي للمفردة الاختبارية، فالأداء التفاضلي للمفردة يهتم بالخصائص السيكومترية لها، من حيث الكيفية التي تعمل بها مع المجموعات المختلفة، بينما يهتم تحيز المفردة بالمعنى الاجتماعي لتطبيقها وما يترتب عليه من نتائج (Dorans & Holland, 1993).

ويُنظر للأداء التفاضلي كتحيز في حالة واحدة فقط هي عندما يكون الفرق في الأداء بين المجموعتين على المفردة غير مُبرر، ويُعزى إلى عوامل خارجية ليست ذات علاقة بالسمة أو القدرة المقاسة، ويُقال في هذه الحالة إن المفردة تقيس سمة أو سمات أخرى غير السمة الأساسية المقاسة (Jensen, 1980).

ويُمثل الأداء التفاضلي للمفردة وما يُمكن أن يؤول إليه من تحيز سواء كان للمفردة أو للاختبار كله مشكلة تُهدد صدق درجات الاختبار، ومبدأ العدالة بين المفحوصين. لذلك فقد برز كقضية هامة من قضايا القياس والتقويم، مما جعله الموضوع الرئيسي لاهتمامات الباحثين والخبراء في هذا المجال خلال العقدين الماضيين. ومن ثم وضعت شروط ومواصفات جديدة عند بناء وتصميم وتطوير الاختبارات والمقاييس النفسية والتربوية ونشرها بصفة عامة، فالاختبار يجب أن يتصف بالعدالة وأن يكون صادقاً بالنسبة لجميع الأفراد والمجموعات التي يطبق عليها، ويجب أن يُعطي فرصاً متساوية لجميع المفحوصين لإظهار ما لديهم من قدرات ومعرفة ومهارات مكتسبة تتعلق بغرض تطبيق الاختبار، بغض النظر عن جنسهم أو عرقهم أو ثقافتهم (Lee, 2004).

وفي ضوء ذلك فقد أولى المختصون في مجال القياس والتقويم قضية عدالة الاختبار بالغ الاهتمام والتقدير، نظراً لأهمية القرارات التي يتم اتخاذها استناداً إلى نتائج الاختبارات، سواء في اختيار الأفراد للوظائف المختلفة، أو تصنيفهم تبعاً لقدراتهم أو تحصيلهم، أو توجيه الطلبة مهنياً، أو تحديد الصفوف الدراسية المناسبة لهم. وقد أدى هذا الاهتمام إلى اعتبار خلو الاختبار أو مفرداته من الأداء التفاضلي من أهم الشروط التي ينبغي توافرها في الاختبار قبل

نشره ، لكي يُعتمد على نتائجه بدرجة أعلى من الثقة (نوافله، ٢٠١٧). فوجود مفردات متحيزة في الاختبار يؤدي إلى اختبار متحيز يهدد عدالة الاختبار وصدق نتائجه، الأمر الذي يؤدي إلى اختلاف معاني درجات الطلاب في الاختبار وكذلك التفسيرات المترتبة عليها، وبالتالي عدم صحة القرارات المتخذة في ضوئها.

مشكلة الدراسة :

على الرغم من استخدام الاختبارات في عمليات التقويم بصورة واسعة، إلا أنها تعاني من بعض المشكلات التي تفقدها دقتها وموضوعيتها، وفى سبيل التغلب على بعض تلك المشكلات لجأ علماء القياس إلى بعض الإجراءات والأساليب التي يُمكن باستخدامها تحقيق بعض متطلبات الدقة والموضوعية المنشودة (Sijtsma & Molenaar, 2002; Wainer & Thissen, 2001).

وقد تركزت جهود الباحثين في بناء الاختبارات وتطويرها على التدقيق في استخراج معالم المفردات الاختبارية المتمثلة في الصعوبة والتمييز والتخمين، وبالرغم من أهمية هذه المعالم إلا أنها ليست كافية للحكم على صلاحية مفردات الاختبار وكفاءتها للغرض الذي أعدت من أجله، وذلك لأن الاستجابة لمفردات الاختبار قد تتأثر بعوامل أخرى مختلفة عن القدرة المراد قياسها، كالجنس أو العرق أو بيئة السكن أو اللغة أو المستوى الاقتصادي والاجتماعي مما يُؤثر سلباً على نتيجة الاختبار، وبالتالي على القرارات المتخذة في ضوئها. وفي هذه الحالة تُوصف تلك المفردات بأنها تبدي أداء تفاضلياً نحو مجموعة أو فئة من المفحوصين دون أخرى، ويُعد ذلك مدخلاً لتحيز المفردات.

ويُعد موضوع الأداء التفاضلي لمفردات الاختبارات أحد أهم تطبيقات نظرية القياس الحديثة المعروفة باسم نظرية الاستجابة للمفردة Item Response Theory، حيث يتم تحديد الأداء التفاضلي للمفردة من خلال فحص العلاقة بين الاحتمالات المشروطة للإجابة الصحيحة على المفردة، وبين السمة الكامنة المراد قياسها (Cappaert, 2014).

وقد تعددت الدراسات التي تناولت موضوع الأداء التفاضلي للمفردة، حيث ركزت اغلبها على الكشف عن المفردات ذات الاداء التفاضلي في اختبارات القبول أو الاختبارات والمقاييس الدولية سواء باستخدام الطرق التي تعتمد على النظرية الكلاسيكية أو التي تعتمد

على نظرية الاستجابة للمفردة. كما ركزت معظم الدراسات على المقارنة بين الطرق المختلفة للكشف عن الأداء التفاضلي للمفردة، ونوع الأداء التفاضلي (منتظم-غير منتظم)، ومصدر الأداء التفاضلي، بينما لم ينل التأثير الناتج عن وجود مفردات ذات أداء تفاضلي الاهتمام البحثي الكافي نظرياً أو تطبيقياً، فلم تتناوله بالبحث والدراسة إلا عدد قليل من الدراسات. بالإضافة إلى انه لا توجد دراسة تناولت الكشف عن الأداء التفاضلي لمفردات اختبار القدرات العامة.

لذلك تتلخص مشكلة الدراسة الحالية في استخدام بعض طرق نظرية الاستجابة للمفردة للكشف عن الأداء التفاضلي لمتغير الجنس في مفردات اختبار القدرات العامة، ودراسة أثره على دقة تقدير قدرات الأفراد وتحقيق الخصائص السيكومترية للاختبار، وذلك من خلال الإجابة عن التساؤلات التالية :

١- ما نسبة المفردات ذات الأداء التفاضلي لمتغير الجنس في الصورة (١) من اختبار القدرات العامة ؟

٢- هل يؤثر وجود المفردات ذات الأداء التفاضلي لمتغير الجنس في الاختبار على دقة تقديرات القدرة (اللفظية-الكمية) ؟

٣- هل توجد فروق ذات دلالة إحصائية بين كل من تقديرات القدرة (الكمية – اللفظية) لكل من الذكور والإناث وما يتعلق بها من أخطاء معيارية قبل وبعد حذف المفردات ذات الأداء التفاضلي لمتغير الجنس ؟

٤- هل توجد فروق ذات دلالة إحصائية بين دقة تقدير دالة معلومات الاختبار قبل وبعد حذف المفردات ذات الأداء التفاضلى لمتغير الجنس ؟

٥- هل توجد فروق ذات دلالة إحصائية بين الخصائص السيكومترية للاختبار قبل وبعد حذف المفردات ذات الأداء التفاضلى لمتغير الجنس ؟

الهدف من الدراسة :

تهدفت الدراسة الحالية إلى الكشف عن المفردات ذات الأداء التفاضلي لمتغير الجنس في مفردات اختبار القدرة السعودي باستخدام طريقتين حديثتين تتبعان نظرية الاستجابة للمفردة هما طريقتي مربع كاي لكاميللي (Camilli Chi – Square) والفروق في المساحات تحت منحنيات خصائص المفردات ، ودراسة أثره على دقة تقديرات القدرة والخصائص السيكومترية للاختبار .

أهمية الدراسة :

تبدو أهمية الدراسة الحالية في النقاط التالية :

١- تكتسب الدراسة الحالية أهميتها من أهمية الاختبار المستخدم فيها ، فكونها تبحث في الأداء التفاضلي لمتغير الجنس في مفردات اختبار القدرات العامة ، والذي يمثل أحد الاختبارات الأساسية والمهمة في انتقاء الطلاب للجامعات والوظائف المختلفة على مستوى المملكة العربية السعودية ، فإن ذلك يمثل أهمية كبيرة ، حيث أن نتائجها سوف تفيد القائمين والمسئولين عن هذا الاختبار من خلال ما يمكن أن تقدمه لهم من معلوماتٍ حول خصائص المفردات الاختبارية التي يتضمنها هذا الاختبار، وتحديد المفردات التي تظهر أداء تفاضليا منها ، لتعديلها أو حذفها من الاختبار ، وأيضا لتجنب اشتماله على مثل هذه المفردات في المرات القادمة التي يعقد فيها الاختبار.

٢- تبدو أهمية الدراسة الحالية في إمكانية توجيه اهتمام معدي الاختبارات بشكل عام واختبار القدرات العامة بشكل خاص إلى ضرورة الانتباه لفحص الأداء التفاضلي لمفردات الاختبار كخاصية سيكومترية مهمة ، وتأثير ذلك على دقة نتائج هذه الاختبارات .

٣- كما تتبين أهمية الدراسة الحالية في كشفها عن المفردات التي تظهر أداء تفاضليا لمتغير الجنس في اختبار مهم ، وبالتالي يمكن أن تساعد نتائجها القائمين على هذا الاختبار بالأخذ بعين الاعتبار هذا الأداء التفاضلي الذي بينته الدراسة عند اتخاذ قرارات في ضوء نتائج هذا الاختبار.

٤- يمكن أن تساعد نتائج الدراسة الحالية المؤسسات والجهات التي تقوم ببناء وتطوير الاختبارات ، والمهتمين بموضوع الأداء التفاضلي للمفردة ، في وضع آلية لتقييم نتائجهم عند اشتمال اختباراتهم على مفردات ذات أداء تفاضلي على اختلاف نسبتها ومستواها في الاختبار ، في الحكم على مدى حجم أثر تلك المفردات على نتائج الاختبار .

فروض الدراسة :

١- لا توجد مفردات ذات أداء تفاضلى لمتغير الجنس من اختبار القدرات العامة.

٢- لا يؤثر وجود المفردات ذات الأداء التفاضلي لمتغير الجنس على دقة تقديرات القدرة (اللفظية-الكمية).

٣- لا توجد فروق ذات دلالة إحصائية بين كل من تقديرات القدرة (الكمية – اللفظية) لكل من الذكور والإناث وما يتعلق بها من أخطاء معيارية قبل وبعد حذف المفردات ذات الأداء التفاضلي لمتغير الجنس .

٤- لا توجد فروق ذات دلالة إحصائية بين دقة تقدير دالة معلومات الاختبار قبل وبعد حذف المفردات ذات الأداء التفاضلى لمتغير الجنس .

٥- لا توجد فروق ذات دلالة إحصائية بين الخصائص السيكومترية للاختبار قبل وبعد حذف المفردات ذات الأداء التفاضلى لمتغير الجنس .

حدود الدراسة :

اقتصرت الدراسة الحالية على ما يلي :

- نتائج تطبيق صورة واحدة من اختبار القدرات العامة المطبق عام ٢٠١٦ م .
- استخدام طريقتين لكشف الأداء التفاضلي للمفردات تتبعان نظرية الاستجابة للمفردة هما طريقتي مربع كاي لكاميللي (Camilli Chi – Square) والفروق في المساحات تحت منحنيات خصائص المفردات .
- دراسة الأداء التفاضلي لمفردات الاختبار في ضوء استخدام النموذج أحادي البارامتر (نموذج راش) .
- دراسة الأداء التفاضلي لمفردات الاختبار لمتغير واحد فقط هو الجنس .
- طلبة الصف الثالث الثانوي للعام الدراسي ٢٠١٥ /٢٠١٦(ذكور – إناث) الذين تقدموا لاختبار القدرات العامة عام ٢٠١٦ م .

منهج الدراسة :

اعتمدت الدراسة الحالية على المنهج الوصفي للكشف عن المفردات التي تُظهر أداءً تفاضلياً لمتغير الجنس، وأثره في دقة تقديرات القدرة والخصائص السيكومترية للاختبار.

عينة الدراسة :

تم الحصول على البيانات من المركز الوطني للقياس والتقويم للطلبة المتقدمين لاختبار القدرات العامة عام ٢٠١٦. واعتمدت الدراسة الحالية على البيانات الناتجة

من تطبيق نموذج واحد فقط من نماذج الاختبار، فبلغ عدد العينة ٧٤٠٨ فرداً، منهم ٣٥٩٩ من الذكور و ٣٨٠٩ من الاناث.

أداة الدراسة :

يُعد اختبار القدرات العامة أحد الاختبارات التي يُصممها المركز الوطني للقياس والتقويم في التعليم العالي في المملكة العربية السعودية لطلبة المرحلة الثانوية الراغبين في مواصلة الدراسة الجامعية، ومدته ساعتان ونصف.

يقيس الاختبار قدرتين أساسيتين هما: القدرة اللفظية والقدرة الكمية. تهدف أسئلة القدرة اللفظية للتعرف على قدرة الطالب لفهم تراكيب الجمل، وتحديد العلاقة بين زوجين من الكلمات، واستيعاب النصوص والقدرة على الاستنتاج، أما أسئلة القدرة الكمية فتركز على القياس والاستنتاج وحل المسائل الرياضية (المركز الوطني للقياس والتقويم، ٢٠١٣) .

يحتوي الاختبار على ٩٦ سؤالاً من نوع اختيار من متعدد، موزعة كالتالي: ٥٢ سؤالاً لقياس القدرة اللفظية و ٤٤ سؤالاً لقياس القدرة الكمية.

مصطلحات الدراسة :

الأداء التفاضلي للمفردة(DIF) Differential Item Function :

هو عبارة عن مؤشر إحصائي يمثل دلالة مشتقة إحصائيا للتعبير عن الفروق في الاستجابة على المفردة الاختبارية بين الذكور والإناث عند مستوى قدرة محدد (Kappaert, 2014) ، ويقدر في الدراسة الحالية بطريقتين هما مربع كاي لكاميلل – (Camilli Chi Square) والفروق في المساحات تحت منحنيات خصائص المفردات .

نسبة المفردات ذات الأداء التفاضلي :

هي عدد مفردات الاختبار التي تظهر أداء تفاضليا مقسوماً على عدد المفردات الكلية للاختبار.

نظرية الاستجابة للمفردة Item Response Theory :

هي نظرية حديثة في القياس النفسي والتربوي ، ويتم فيها تحديد العلاقة بين أداء المفحوص والسمة الكامنة موضوع القياس وفق دالة رياضية محددة .

منحنى خصائص المفردة Item Characteristic Curve (ICC) :

هو تمثيل بياني لانحدار احتمال استجابة الأفراد على المفردة استجابة صحيحة على القدرة التي تقيسها (علام ، ٢٠٠٥) .

الإطـــار النـــظري

نظرية الاستجابة للمفردة :

تعتمد هذه النظرية على فرضية أساسية مؤداها أن القيمة الاحتمالية لاستجابة فرد ما لمفردة اختبارية تكون دالة لكل من السمة أو القدرة التي يفترض أن الاختبار يقيسها لدى الفرد، وخصائص المفردة الاختبارية التي يحاول الإجابة عنها. أي أنها تفترض أن هناك دالة رياضية احتمالية تربط بين بارامترين أحدهما يتعلق بالفرد والآخر يتعلق بالمفردة التي يختبر بها، ويتركز الهدف الأساسي لهذه النظرية في التوصل إلى قيم تقديرية لكل من هذين البارامترين، واستخدامها في تقدير احتمال استجابة الفرد استجابة صحيحة عن كل مفردة من مفردات الاختبار، حيث يتم التقدير الإحصائي للعلاقة بين احتمال الاستجابة الصحيحة لمفردة اختبارية والسمة التي يفترض أن الاختبار يقيسها، والتحقق من صحتها أمبيريقياً (Hambleton & Swaminathan, 1985).

ويتطلب تحديد هذه العلاقة تحديد بعض المعلومات عن كل من الفرد والمفردة الاختبارية، فبالنسبة للفرد نحتاج إلى قيمة عددية واحدة هي بارامتر القدرة Ability Parameter، وبالنسبة للمفردة نحتاج إلى قيمة عددية واحدة أو أكثر تمثل البارامترات المطلوبة للمفردة والمتمثلة في الصعوبة والتمييز .

وقد تعددت النماذج المتعلقة بهذه النظرية وتباينت أشكال المنحنيات المميزة للمفردات تبعاً لاختلاف عدد البارامترات المتطلبة لكل نموذج (Hambleton, Swaminathan, & Rogers, 1991).

مزايا نظرية الاستجابة للمفردة :

يرى جورج راش صاحب نموذج راش أن القياس يقوم في جوهره على التفاعل بين السمات والخصائص المطلوب قياسها وبين أدوات القياس المستخدمة، وأنه يمكن الحكم على

القياس بالموضوعية إذا كانت نتائجه مستقلة عن الأداة المستخدمة في التوصل إلى تلك النتائج، وهذا هو ما تحققه نظرية الاستجابة للمفردة (كاظم، ١٩٩٦)، حيث أنها توفر متطلبات الموضوعية في القياس من خلال الجوانب الآتية:

• توحد القياس لكل من قدرات الأفراد وصعوبة المفردات: حيث يمكن التعبير عن كل من قدرات الأفراد وصعوبة المفردات بمقياس واحد، بحيث يمكن رد التقديرات الخاصة بالأفراد وكذلك التقديرات الخاصة بالمفردات إلى نقطة أصل واحدة (صفر تدريج مشترك)، أي أن أداة القياس تتدرج بوحدة قياس مطلقة ثابتة تتوافق مع تدرج مستويات القدرة التي تقيسها.

• تحرر صعوبة المفردات من توزيع أداء عينة الأفراد: حيث يمكن تقدير صعوبة كل مفردة من مفردات الاختبار بطريقة مستقلة عن عينة الأفراد الذين طبق عليهم الاختبار، أي أنه يمكن وصف المفردة الاختبارية بطريقة مستقلة عن مجموعة الأفراد الذين اختبروا بها، وهذا يعنى أن تدرج صعوبة أي مفردة بين باقي مفردات الاختبار يظل ثابتاً مهما اختلف الأفراد الذين يطبق عليهم الاختبار طالما أنهم مناسبون للأداء على مفرداته، وهذه الخاصية تسمح باختيار أي مفردة كنقطة أصل لتدريج صعوبة المفردات، حيث يتم تدريج صعوبات جميع المفردات الأخرى التي يشتمل عليها الاختبار بدءاً بهذه النقطة على التدريج.

• تحرر قدرات الأفراد من تأثير صعوبة المفردات: إن استخدام نظرية الاستجابة للمفردة في تحليل مجموعة من المفردات وتدريجها على تدريج واحد مشترك يتيح الفرصة لاستخدام هذه المفردات في تقدير قدرات الأفراد وفقاً لاستجاباتهم عليها، أي أنه يمكن وصف قدرة الفرد بطريقة مستقلة عن خصائص المفردات التي اختبر بها، وبالتالي فإن قدرة الفرد لا تختلف باختلاف مجموعة المفردات المستخدمة في تقديرها، أي أنه يمكن الموازنة بين قدرات الأفراد بطريقة مستقلة عن صعوبة المفردات المستخدمة في الموازنة، ويمكن استخدام القيمة العددية الممثلة لقدرة أي مجموعة حصل أفرادها على نفس الدرجة الكلية كنقطة أصل لميزان القدرة (كاظم، ١٩٩٦).

الأداء التفاضلي للمفردات الاختبارية :

يُعد الأداء التفاضلي لمفردات الاختبار أحد أهم تطبيقات نظرية الاستجابة للمفردة. وقد ميز هامبلتون وسوامنثان وروجرز (Hambleton, Swaminathan & Rogers, 1991) بين

الأداء التفاضلي للمفردة وتحيز المفردة، حيث رأوا أن المفردة التي تظهر أداء تفاضلياً هي تلك المفردة التي تكون دول استجابات المجموعات المختلفة عليها غير متماثلة، في حين يظهر التحيز في صورة خطأ منتظم يجعل أداء إحدى المجموعات على المفردة أفضل من أداء المجموعات الأُخرى .

ويُعرف الأداء التفاضلي للمفردة بأنه الفروق في احتمالات الإجابات الصحيحة للمفردة بين المجموعات المختلفة المتساوية في القدرة. ويمثل الأداء التفاضلي شرطاً ضرورياً لحدوث التحيز ولكنه غير كاف، فالمفردة قد تظهر أداء تفاضلياً ولا تكون متحيزة في حالة إذا كان الفرق بين المجموعات المختلفة أو الأفراد في الإجابة عنها يُعزى لفروق فعلية بينهم في القدرة المستهدفة من عملية القياس (Hambleton & Rogers, 1995).

ويمكن التمييز بين نوعين للأداء التفاضلي للمفردة هما: الأداء التفاضلي المنتظم Uniform DIF، والأداء التفاضلي غير المنتظم Non-Uniform DIF. فالأداء التفاضلي المنتظم للمفردة يحدث عندما يكون أداء إحدى المجموعتين أفضل من المجموعة الأُخرى عند جميع مستويات القدرة، أما إذا حدث هذا الأداء عند مستوى ما من مستويات القدرة ولم يحدث عند باقي المستويات فتعتبر المفردة ذات أداء تفاضلي غير منتظم (Zumbo, 1999).

وقد ظهرت العديد من الطرق للكشف عن المفردات ذات الأداء التفاضلي، منها ما يعتمد على النظرية الكلاسيكية كطريقة تحليل التباين وطريقة الصعوبة المحولة، ومنها ما يعتمد على نظرية الاستجابة للمفردة كطريقة فرق المساحة لمنحنى خصائص المفردة وطريقة نسبة الارجحية وطريقة مربع كاي (Magis & De Boeck, 2014).

إجراءات الدراسة :

أولا : التحقق من ملاءمة البيانات لنظرية الاستجابة للمفردة

أ – التحقق من توافر افتراض أحادية البعد Unidimensionality :

فيما يتعلق بالتحقق من مدى تحقيق البيانات الحالية لافتراضى أحادية البعد والاستقلال المركزى , فقد قام الباحثان بحساب قيم كل من "الاتساق الداخلى للمفردات , وقيمة معامل ألفا للمجموع الكلى للمفردات الاختبارية , وكذلك قيمة معامل ألفا فى حالة حذف كل مفردة على حدة , فتبين أن قيمة معامل ألفا للمفردات ككل تساوى (٠,٨٦) , وأن قيمته

فى حالة حذف كل مفردة على حدة قد تراوحت بين (٠,٨٥) و(٠,٨٦) , وهذه النتائج موضحة في ملحق (١) ، وتعد هذه القيم مؤشرًا على أن بيانات الدراسة الحالية تحقق افتراض أحادية البعد Unidimensionality [أندريش (2000) Andrich , علام (٢٠٠٥)]

كما تم إجراء التحليل العاملى الاستكشافى لبيانات اختبار القدرات (اللفظية – الكمية) كل على حدة ، باستخدام طريقة المكونات الأساسية وذلك لمعرفة العوامل المسئولة عن الأداء فى الاختبار، حيث تم استخراج قيم الجذر الكامن للعوامل المستخلصة ونسبة التباين المفسر لكل عامل بطريقة المكونات الأساسية Principal Component Analysis كما يبينها الجدول التالى:

جدول (١)

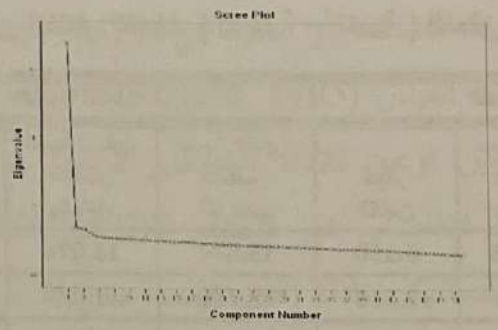
نتائج التحليل العاملى لاختبار القدرات (اللفظية – الكمية) للتحقق من أحادية البعد

العوامل	القدرة اللفظية			القدرة الكمية		
	الجذر الكامن	نسبة التباين المفسر بالعامل	نسبة التباين المفسر التراكمى	الجذر الكامن	نسبة التباين المفسر بالعامل	نسبة التباين المفسر التراكمى
١	6.799	13.074	13.074	6.190	14.068	14.068
٢	1.381	2.655	15.730	1.360	3.091	17.158
٣	1.327	2.553	18.282	1.136	2.581	19.739
٤	1.187	2.282	20.565	1.119	2.542	22.281
٥	1.100	2.115	22.680	1.094	2.487	24.768
٦	1.085	2.087	24.766	1.062	2.414	27.182
٧	1.073	2.063	26.829	1.043	2.370	29.552
٨	1.054	2.027	28.857	1.033	2.349	31.901
٩	1.047	2.013	30.869	1.011	2.297	34.197
١٠	1.038	1.995	32.865	1.004	2.283	36.480
١١	1.030	1.980	34.845			
١٢	1.012	1.946	36.791			

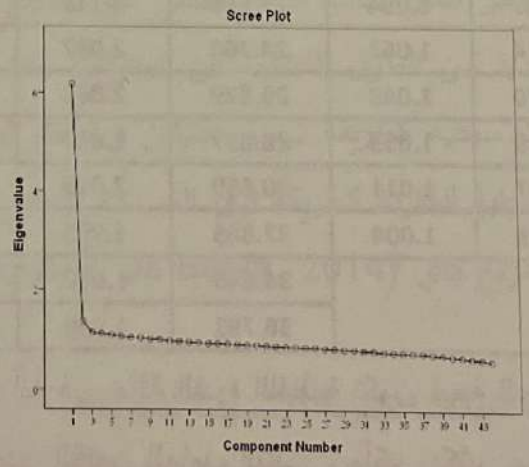
يتضح من الجدول السابق أنه فى حالة القدرة اللفظية تكون قيمة الجذر الكامن للعامل الأول أكثر من (٤) أضعاف الجذر الكامن للعامل الثانى ، وأكبر بكثير من العوامل الأخرى ، كما أن قيمة التباين المفسر بالعامل الأول كبيرة مقارنة بالعوامل الأخرى، وبالمثل فى حالة القدرة الكمية_تكون قيمة الجذر الكامن للعامل الأول أكثر من (٤) أضعاف الجذر الكامن للعامل الثانى، وأكبر بكثير من العوامل الأخرى ، أيضا قيمة التباين المفسر بالعامل الأول كبيرة

مقارنة بالعوامل الأخرى . وبالتالى يدل ذلك على سيطرة عامل واحد بشكل أكبر على استجابات الأفراد على مفردات الاختبار ، ممايدل على تحقق افتراض أحادية البعد ، حيث تشير الأدبيات إلى أن الأداة تعد أحادية البعد إذا كانت نسبة ما يفسره العامل الأول أكثر من (٢٠ %) ، وكذلك بالنسبة لبيانات الذكور وبيانات الإناث كل على حدة (Reckase as cited in Lee,2004) .

ويظهر ذلك من خلال الشكلين البيانيين التاليين :



شكل (١) قيم العوامل التى تشبعت عليها مفردات القدرة اللفظية



شكل (٢) قيم العوامل التى تشبعت عليها مفردات القدرة الكمية

وبذلك تكون النتائج الخاصة باختبار أحادية البعد لكل قدرة على حدة قد بينت وجود عامل واحد مسيطر على استجابات الأفراد على المفردات ، أي يتحقق في البيانات افتراض أحادية البعد.

ب – التحقق من توافر افتراض الاستقلال المركزي Local Independence:

قام الباحثان بالتحقق من هذا الافتراض من خلال حساب معاملات الارتباط الجزئية Partial correlations بين كل زوج من أزواج المفردات عند كل مستوى من مستويات القدرة المختلفة (لفظية – كمية) الموضح في ملحق (١) ، وبينت النتائج[١] تحقق افتراض الاستقلال المركزى في البيانات الحالية .

ج – افتراض التحرر من عامل السرعة فى الإجابة Speedness :

الاختبار الذي اعتمدت الدراسة الحالية على نتائجه هو اختبار قوة وليس فيه مجال للتنافس في سرعة الاستجابة . لذلك لم يعتمد في تطبيقه على السرعة في الإجابة ، بل أعطى للطلاب وقتا كافيا للتفكير لمحاولة الإجابة عن جميع الأسئلة ,

د – افتراض المنحنيات المميزة للمفردات Item characteristic curves :

إن استخدام برنامج "Winsteps" فى تحليل البيانات, يحقق الافتراض المتعلق بوجود منحنى مميز لكل مفردة من مفردات الاختبار , حيث أن مخرجات التحليل باستخدام هذا البرنامج تتضمن تحليلاً بيانياً للبيانات يشتمل على منحنى مميز لكل مفردة على حدة .

ثانيا : تحليل البيانات في ضوء نموذج راش باستخدام برنامج "Winsteps", حيث تمت الإجراءات التالية :

أ- حذف كل طالب لم يتمكن من الإجابة الصواب عن أى مفردة من مفردات الاختبار (أى حصل على الدرجة صفر) , حيث أن مستوى هذا الطالب أقل من مستوى الاختبار , كما تم حذف كل طالب أجاب صوابا عن جميع مفردات الاختبار (أى حصل على الدرجة النهائية للاختبار) , حيث أن مستوى هذا الطالب يكون أعلى من مستوى الاختبار .

ب- حذف كل مفردة لم يستطع جميع الطلاب الإجابة عنها , حيث تكون هذه المفردة أعلى من مستوى قدرات عينة الدراسة , كما تم حذف كل مفردة أجاب عنها جميع أفراد العينة إجابة صحيحة , حيث تكون هذه المفردة أقل من مستوى قدرات عينة الدراسة (كاظم, ١٩٩٦ : ٣٢٨ – ٣٢٩).

---

[١] ملحق (١) معاملات الارتباط الجزئية بين كل زوج من أزواج المفردات عند كل مستوى من مستويات القدرة المختلفة للتحقق من الاستقلال المركزي .

والجدول التالي يوضح مؤشرات الملاءمة التباعدية الناتجة عن مطابقة المفردات الاختبارية لكل من القدرتين (اللفظية – الكمية) للنموذج أحادى البارامتر :

جدول (٢)

نتائج مطابقة المفردات الاختبارية لكل من القدرتين(اللفظية – الكمية)للنموذج أحادى البارامتر

القدرة اللفظية		القدرة الكمية	
رقم المفردة	مؤشر الملائمة التباعدى OUT.ZSTD	رقم المفردة	مؤشر الملائمة التباعدى OUT.ZSTD
1	9.9	53	-0.01
2	9.56	54	9.9
3	4.55	55	8.11
4	1.97	56	9.32
5	9.9	57	9.9
6	3.85	58	-1.36
7	3.91	59	6.85
8	-2.65	60	9.9
9	9.9	61	9.9
10	-3.85	62	9.9
11	2.55	63	1.63
12	5.43	64	1.1
13	9.71	65	7.22
14	9.76	66	-0.54
15	-1.64	67	3.68
16	9.9	68	-9.9
17	9.9	69	-7.38
18	-9.9	70	4.61
19	1.44	71	4.93
20	7.95	72	5.07
21	-9.9	73	1.56
22	3.85	74	-5.75
23	9.9	75	1.39
24	-0.43	76	9.9
25	9.42	77	9.9
26	0.04	78	7.58
27	8.37	79	-0.92
28	-3.18	80	7.71
29	9.9	81	5.59
30	2.67	82	0.25
31	-9.9	83	-5.82
32	3.85	84	2.23
33	-9.9	85	-0.57
34	-8.28	86	3.59
35	-9.9	87	7.3
36	0.74	88	4.42
37	9.9	89	9.9
38	4.6	90	9.9
39	0.68	91	-9.9
40	2.56	92	6.03

القدرة اللفظية		القدرة الكمية	
41	5.38	93	4.97
42	-5.97	94	-9.9
43	-9.9	95	9.9
44	3.96	96	2.89
45	-5.28		
46	-9.9		
47	-7.1		
48	9.9		
49	4.66		
50	9.9		
51	6.3		
52	4.16		

يتضح من الجدول السابق وجود عدد من المفردات الاختبارية لكل من القدرتين(اللفظية – الكمية)غير ملائمة للنموذج أحادي البارامتر بلغ عددها(١٦) مفردة في حالة اختبار القدرة اللفظية ،و(١١) مفردة في حالة اختبار القدرة الكمية . وبهذا صار عدد مفردات الاختبار اللفظي (٣٦) مفردة ، في حين صار عدد مفردات الاختبار الكمي (٣٣) مفردة .

والجدول التالي يوضح نتائج مطابقة الأفراد (ذكور – إناث) في كل قدرة على حدة (اللفظية – الكمية) للنموذج أحادى البارامتر :

جدول (٣)

نتائج مطابقة الأفراد للنموذج أحادى البارامتر

م		عدد الأفراد المحذوفين			المجموع الكلى
	الجنس	القدرة اللفظية	القدرة الكمية	العدد الكلى الأفراد المحذوفين	
١	ذكور	٢٩١	٧٣	٣٦٤	٣٥٩٩
٢	إناث	١٧٣	٨٨	٢٦١	٣٨٠٩
المجموع الكلى		٤٦٤	١٦١	٦٢٥	٧٤٠٨

يتضح من الجدول السابق أن إجمالي عدد الأفراد غير المطابقين للنموذج أحادي البارامتر في الصورة الاختبارية (١) هو (٦٢٥) فردا ، حيث بلغ إجمالي المحذوفين في كلتا القدرتين (اللفظية – الكمية) (٣٦٤) فردا من الذكور ، و(٢٦١) فردا من الإناث .

ثالثا : إعادة التحليل بعد حذف الأفراد والمفردات غير الملائمة

قام الباحثان بإعادة التحليل بعد حذف الأفراد والمفردات غير الملائمة فى الخطوة السابقة , حيث تبين من مخرجات التحليل أن عدد المفردات الملائمة للنموذج هو (٦٩)

مفردة من إجمالي (٩٦) ، وأن عدد الأفراد الملائمين للنموذج هو (٦٧٨٣) فردا من الذكور والإناث من أصل (٧٤٠٨) المستجيبين على الصورة الاختبارية (١) ، منهم (٣٢٣٥) من الذكور ، و(٣٥٤٨) من الإناث .

نتائج الدراسة ومناقشتها :

فيما يلي عرضا للنتائج الخاصة بالإجابة عن أسئلة الدراسة واختبار صحة الفروض المنبثقة عنها .

قبل البدء في الإجابة عن أسئلة الدراسة وجد الباحثان أنه من المناسب التفحص النظري لاختلاف الخصائص السيكومترية لمفردات الاختبار بشقيه (اللفظي – الكمي) تبعا لمتغير الجنس ، من خلال تقدير متوسطات بارامترات مفردات كل من القدرتين اللفظية والكمية وانحرافها المعيارى حسب الجنس ، والجدول التالي يبين هذه الإحصاءات :

جدول (٤)

متوسطات بارامترات مفردات كل من القدرة اللفظية والقدرة الكمية وانحرافها المعيارى حسب الجنس (ذكور – إناث)

الجنس	متوسطات بارامترات مفردات القدرة اللفظية وانحرافها المعيارى						متوسطات بارامترات مفردات القدرة الكمية وانحرافها المعيارى					
	التمييز		الصعوبة		التخمين		التمييز		الصعوبة		التخمين	
	م	ع	م	ع	م	ع	م	ع	م	ع	م	ع
ذكور	1.87	0.30	1.56	0.69	0.27	0.04	1.04	0.11	2.79	0.85	0.19	0.13
إناث	1.01	0.14	1.22	0.25	0.15	0.11	1.51	0.32	1.93	0.53	0.41	0.08

يتضح من الجدول السابق أن هناك اختلافات واضحة في متوسطات بارامترات المفردات بين الذكور والإناث ، مما يدل على ضرورة تفحص الأداء التفاضلي لها تبعا لمتغير الجنس .

الإجابة عن السؤال الأول :

للإجابة عن هذا التساؤل استخدم الباحثان طريقتين للكشف عن المفردات ذات الأداء التفاضلى لمتغير الجنس في ضوء نظرية الاستجابة للمفردة هما :

أولا : نتائج طريقة مربع كاي لكاميللي لمقارنة بارامترات مفردات الاختبار

تعد المفردة ذات أداء تفاضلي إذا كان مجموع قيم مربع كاي عبر مستويات القدرة دالة إحصائيا ، حيث يتم الكشف في جدول الاحتمال ٢ × ٢ (المجموعة " مجموعتان × الاستجابة صحيحة أو خاطئة) لكل مستوى قدرة ، ثم تجمع قيم مربعات كاي عبر مستويات القدرة ، فتكون نسبة الإجابات الصحيحة عليها غير متساوية لمجموعتين لهما نفس الدرجات في الاختبار الذي يتضمن هذه المفردة .

والجدول التالي يوضح بارامترات مفردات القدرة اللفظية ومجموع مربعات كاي للفروق بين كل من الذكور والإناث

جدول (٥)

بارامترات مفردات القدرة اللفظية ومجموع مربعات كاي للفروق بين كل من الذكور والإناث

المفردات	ذكور			إناث			مجموع مربعات كاي للفروق
	التمييز	الصعوبة	التخمين	التمييز	الصعوبة	التخمين	
Item_001	1.36	-0.9	0	1.23	-1.1	0	2.810
Item_002	1.18	0.5	0	1.4	0.42	0	1.173
Item_003	0.7	-0.02	0.06	0.95	-0.12	0.04	1.916
Item_004	0.96	0.5	0.02	1.18	0.32	0	**8.009
Item_005	0.53	0.26	0.09	0.7	0.73	0.06	*4.610
Item_006	1.17	-1.44	0	1.03	-1.82	0	**6.452
Item_007	0.96	-1.14	0	0.95	-1.35	0	0.202
Item_008	1.11	0.39	0	1.06	0.48	0	2.100
Item_009	0.71	1.11	0.07	0.7	1.29	0.07	1.312
Item_010	1.18	0.54	0	1.11	0.78	0	1.118
Item_011	0.86	0.04	0.04	0.98	-0.16	0	2.031
Item_012	0.92	0.18	0.02	0.71	0.08	0.05	**7.239
Item_013	0.69	0.82	0.07	0.82	1.2	0.04	*4.091
Item_014	1.42	-0.91	0	1.15	-1.32	0	**6.905
Item_015	1.16	-0.28	0	1.03	-0.13	0.02	0.450
Item_016	0.61	0.6	0.1	0.84	0.86	0.04	**9.007
Item_017	1.4	-0.08	0	1.37	-0.62	0	*3.901
Item_018	1.41	-0.26	0	1.31	-0.37	0	3.614
Item_019	1.06	0.26	0	0.87	0.49	0.04	2.154
Item_020	1.19	0.03	0	1.34	-0.04	0	1.901
Item_021	1.62	-0.51	0	1.42	-0.53	0	2.663
Item_022	0.91	-0.59	0.01	0.89	-0.45	0.06	1.333
Item_023	0.77	-0.3	0.05	0.51	-0.16	0.17	*5.013
Item_024	0.95	0.44	0.02	1.1	0.51	0	0.905
Item_025	0.64	0.65	0.09	0.87	0.89	0.04	**7.601
Item_026	1.22	-0.64	0	0.96	-0.69	0	**6.199
Item_027	1.25	-0.65	0	1.24	-0.72	0	2.603
Item_028	1.04	0.55	0	1.19	0.57	0	3.212
Item_029	0.76	1.12	0.06	0.72	1.25	0.07	4.029
Item_030	0.85	0.65	0.04	1.01	0.87	0	1.701

Item_031	1.55	-0.59	0	1.22	-0.88	0	**8.119
Item_032	1.2	-0.07	0	1.1	0.02	0	2.318
Item_033	1.6	-0.51	0	1.25	-0.3	0	**9.901
Item_034	1.35	-0.53	0	1.2	-0.6	0	2.822
Item_035	1.29	0.79	0	1.45	0.64	0	1.072
Item_036	0.96	-0.17	0.01	1.05	0.04	0	**6.315
Item_037	0.54	1.12	0.1	0.54	1.48	0.1	*5.980
Item_038	1.21	-1.14	0	1.06	-1.28	0	2.127
Item_039	0.99	0.03	0.01	0.97	-0.18	0	3.028
Item_040	0.93	-0.99	0.02	0.96	-0.97	0.04	4.014
Item_041	0.77	1.17	0.05	0.94	1.05	0.02	4.610
Item_042	1.3	-0.37	0	1.2	-0.41	0	2.113
Item_043	1.44	-0.89	0	1.14	-1.25	0	*3.999
Item_044	1.27	-0.55	0	1.06	-0.48	0	1.127
Item_045	1.19	-0.92	0	1.09	-1.33	0	**6.221
Item_046	1.39	0.08	0	1.35	-0.35	0	*5.001
Item_047	1.28	-0.62	0	1.17	-0.87	0	1.200
Item_048	0.65	0.57	0.08	0.67	0.84	0.07	**7.133
Item_049	0.87	1.21	0.03	0.89	1.63	0.03	**9.724
Item_050	0.33	0.67	0.14	0.17	1.03	0.18	**9.198
Item_051	0.84	0.08	0.03	0.71	0.13	0.05	2.331
Item_052	0.86	0.7	0.04	0.93	0.89	0.02	3.103

يتضح من الجدول السابق وجود أداء تفاضلى لمتغير الجنس فى عدد (٢١) مفردة من مفردات القدرة اللفظية بنسبة (٤٠,٤ %) من إجمالي مفردات القدرة اللفظية البالغ عددها (٥٢) مفردة .

حيث أظهرت المفردات أرقام (٥ ، ٦ ، ١٣ ، ١٤ ، ١٦ ، ١٧ ، ٢٥ ، ٢٦ ، ٣١ ، ٣٧ ، ٤٣ ، ٤٥ ، ٤٦ ، ٤٨ ، ٤٩ ، ٥٠) أداء تفاضليا لصالح الذكور (المجموعة المرجعية) ، في حين أظهرت المفردات أرقام (٤ ، ١٢ ، ٢٣ ، ٣٣ ، ٣٦) أداء تفاضليا لصالح الإناث (المجموعة المستهدفة).

والجدول التالي يوضح بارامترات مفردات القدرة الكمية ومجموع مربعات كاي للفروق بين كل من الذكور والإناث :

جدول (٦)

بارامترات مفردات القدرة الكمية ومجموع مربعات كاي للفروق بين كل من الذكور والإناث

المفردات	ذكور			إناث			مجموع مربعات كاي للفروق
	التمييز	الصعوبة	التخمين	التمييز	الصعوبة	التخمين	
A53	1.11	-0.54	0	0.96	-0.81	0	*4.212
A54	0.6	-0.13	0.08	0.18	-0.07	0.17	1.408
A57	1.35	-0.59	0	1.29	-0.78	0	*5.021
A58	0.63	0.07	0.08	0.6	-0.19	0.08	3.718
A64	1.5	-0.58	0	1.57	-0.48	0	2.001
A66	1.01	-0.27	0	1	-0.34	0.01	**6.841
A70	1.27	0.07	0	1.18	-0.19	0	*3.431
A73	0.71	1.29	0.06	0.69	1.28	0.06	**9.107
A77	0.53	-0.07	0.07	0.39	-0.34	0.07	**7.900
A78	0.6	0.25	0.07	0.43	0.21	0.1	0.314
A85	0.89	0.35	0.04	1.03	0.34	0	0.154
A89	0.87	0.1	0.03	1.06	-0.2	0	**7.602
A91	0.73	0.33	0.04	0.83	0.59	0.02	*3.605
A95	1	-0.04	0.01	1.05	0.23	0	0.333
A55	1.14	-0.27	0	1.17	0	0	0.914
A56	1.48	-0.59	0	1.31	-0.3	0	**9.701
A61	1.3	-1.2	0	1.19	-1.29	0	**7.651
A63	1.32	-0.54	0	0.98	-0.68	0.08	0.099
A69	1.11	-0.58	0	1.21	-0.66	0	**9.613
A72	0.81	0.78	0.04	0.93	0.82	0.02	3.212
A75	0.89	0.01	0.04	1	-0.08	0.01	1.099
A79	1.22	-1.41	0	1.12	-1.86	0	**6.301
A80	0.9	0.22	0.03	1.03	0.72	0	0.896
A82	1.37	0.09	0	1.24	0.84	0	2.818
A83	1.7	-0.65	0	1.51	-0.61	0	1.511
A87	0.86	0.97	0.04	0.86	0.67	0.04	**7.822
A92	0.94	0.44	0.01	1.12	0.37	0	4.812
A93	0.75	-0.26	0.04	0.59	-0.41	0.07	*3.512
A96	0.84	0.16	0.02	0.85	0.41	0.02	1.173
A59	0.97	0.02	0.02	1.02	0.16	0.01	2.026
A60	1.1	0.89	0	1.18	0.56	0	0.018
A62	1	0.29	0.02	1.16	0.21	0	3.014

A65	1.03	0.17	0.01	0.96	0.2	0.02	1.711
A67	1.1	0.05	0	1.15	0.41	0	3.993
A68	0.83	0.31	0.06	0.74	0.36	0.07	1.119
A71	0.86	0.32	0.03	0.92	0.81	0.03	0.127
A74	0.76	0.89	0.06	0.77	0.82	0.06	2.311
A76	0.65	0.26	0.08	0.48	0.23	0.09	0.251
A81	1.63	-0.61	0	1.65	-0.53	0	0.221
A84	1.11	-0.23	0	1.35	-0.51	0	**8.105
A86	1.1	0.33	0	1.18	0.43	0	1.412
A88	1.82	-0.53	0	1.56	-0.84	0	1.118
A90	1.38	-0.14	0	1.43	-0.03	0	2.011
A94	0.94	0.6	0.02	0.94	0.52	0.02	*4.633

يتضح من الجدول السابق وجود أداء تفاضلى لمتغير الجنس فى عدد (16) مفردة من مفردات القدرة الكمية بنسبة (36.4%) من إجمالي مفردات القدرة الكمية البالغ عددها (٤٤) مفردة . حيث أظهرت المفردات أرقام (٥٣ ، ٥٧ ، ٦٦ ، ٧٧ ، ٩١ ، ٦١ ، ٦٩ ، ٧٩ ، ٩٣ ، ٨٤) أداء تفاضليا لصالح الذكور (المجموعة المرجعية) ، في حين أظهرت المفردات أرقام (٧٠ ، ٨٩ ، ٥٦ ، ٨٧ ، ٩٤) أداء تفاضليا لصالح الإناث (المجموعة المستهدفة) .

ثانيا : نتائج حساب الفروق في المساحات تحت منحنيات خصائص المفردات

لتقدير الأداء التفاضلي للمفردات بطريقة منحنى خصائص المفردة استخدم الباحثان برمجية IRT DIF ، وتقوم على حساب مصفوفة البواقي بين احتمالات الاستجابات الصحيحة على المفردة للمجموعة المرجعية والمجموعة المستهدفة عند قدرات محددة من خلال تقسيم فترة القدرة [3+ ,3-] إلى60 جزءا متساويا ، ومن ثم حساب الجذر التربيعي لمتوسط مربعات هذه البواقي Root Mean Square Differences (RMSD) . وقد اعتمد مكاليوم (Mc Callum, 2003) القيمة 0.05 كدرجة قطع للحكم على وجود أداء تفاضلي للمفردة .

والجدول التالي يوضح قيم الأرجحية و RMSD لكل مفردة من مفردات القدرتين (اللفظية – الكمية) :

جدول (٧)

قيم الأرجحية و RMSD لكل مفردة من مفردات القدرتين (اللفظية – الكمية)

أرقام المفردات	القدرة اللفظية		أرقام المفردات	القدرة الكمية	
	الأرجحية	RMSD		الأرجحية	RMSD
١	٠,٩٢	٠,٠١٨	١	١,٨٣	٠,٠٨١
٢	٠,٧٤	٠,٠٤٩	٢	١,٣٥	٠,٠٤٥
٣	٠,٩٨	٠,٠٢١	٣	٠,٩٦	٠,٠٢٤
٤	١,٢٥	٠,٠٣٥	٤	١,١٢	٠,٠٤٣
٥	١,١٩	٠,٠٨٦	٥	٠,٨٢	٠,٠٣٣
٦	١,٤٢	٠,١٤٠	٦	١,٢٨	٠,٠١٨
٧	١,٠٤	٠,٠١٩	٧	٠,٩٢	٠,٠٤٧
٨	١,٠٦	٠,٠٤٢	٨	٠,٩٩	٠,٠٣٧
٩	٠,٨٦	٠,٠٤٠	٩	١,٨٦	٠,٠٨٩
١٠	٠,٩٤	٠,٠٣٧	١٠	٠,٧١	٠,٠٧٧
١١	٠,٨١	٠,٠٤٥	١١	١,٧٦	٠,٠٣٨
١٢	٠,٩٧	٠,٠٨٣	١٢	٠,٩٤	٠,٠٢١
١٣	١,٣٣	٠,٠٦٨	١٣	١,٨٧	٠,٠١٢
١٤	٠,٩٢	٠,٠٦٧	١٤	٠,٩٧	٠,٠٥٢
١٥	١,٤٠	٠,٠٠٤	١٥	٠,٨٠	٠,٠٣٠
١٦	٠,٦٠	٠,٠٥٥	١٦	٢,١١	٠,٠٤٦
١٧	١,٣٥	٠,٠٨٠	١٧	١,٤٤	٠,٠٦٧
١٨	٠,٩٤	٠,٠٢٥	١٨	٠,٩٧	٠,٠١٨
١٩	٠,٨٥	٠,٠٥١	١٩	١,٠٩	٠,٠١٠
٢٠	٠,٥٥	٠,٠٣٧	٢٠	١,٣٠	٠,٠٤٦
٢١	١,٢٣	٠,٠٤٤	٢١	١,٦٢	٠,١١٢
٢٢	١,٠٠	٠,٠٢٣	٢٢	١,١١	٠,٠٣١
٢٣	١,٦٠	٠,٠٨٥	٢٣	٠,٨١	٠,٠٦٩
٢٤	١,١٢	٠,٠١٩	٢٤	٠,٩٥	٠,٠٦١
٢٥	٠,٧٦	٠,٠٩١	٢٥	٠,٨٨	٠,٠٤٠
٢٦	٠,٩٢	٠,٠٤٥	٢٦	٠,٩٢	٠,٠٠٦
٢٧	٠,٩٦	٠,٠٣٨	٢٧	١,٣٠	٠,١٤٠
٢٨	٠,٧٧	٠,٠١٥	٢٨	٠,٨١	٠,٠١٩
٢٩	١,١٨	٠,٠٣٤	٢٩	١,٢٩	٠,٠٤٢
٣٠	١,٠٦	٠,٠٥٧	٣٠	٠,٩٠	٠,٠٣٧
٣١	١,٣٩	٠,٠٩٧	٣١	١,١٣	٠,٠١٧
٣٢	٠,٩٩	٠,٠٤٦	٣٢	٠,٩٧	٠,٠٨٥
٣٣	١,٠٧	٠,٥٢٦	٣٣	٠,٨٥	٠,٠٤٣
٣٤	٠,٩٨	٠,٠٣٩	٣٤	٠,٥٥	٠,٠٢٨
٣٥	١,٥٠	٠,٠٣٣	٣٥	١,٢١	٠,٠٦٧
٣٦	٠,٨٤	٠,٠١٥	٣٦	٠,٩٥	٠,٠٤٤
٣٧	٠,٦٠	٠,٠٦٤	٣٧	١,٢٥	٠,٠٦٥
٣٨	١,٤٨	٠,٠٤٢	٣٨	١,١٧	٠,٠١٦
٣٩	٠,٩٧	٠,٠٤٠	٣٩	١,٣٢	٠,١٤٠
٤٠	١,٢٥	٠,٠٢٦	٤٠	١,٤٤	٠,٠١٩
٤١	٠,٨٧	٠,٠٢٥	٤١	١,٠٤	٠,٠٥٢
٤٢	١,١٩	٠,٠٠٥	٤٢	٠,٨٣	٠,٠٣٠

٤٣	١,٥٥	٠,١٦١	٤٣	١,٢٧	٠,٠٤٩
٤٤	٠,٧٨	٠,٠٢٦	٤٤	٠,٩٦	٠,٠٣٧
٤٥	١,٤٣	٠,٠٦٧			
٤٦	١,٣٥	٠,٠٨٠			
٤٧	٠,٩٢	٠,٠٤٣			
٤٨	١,٥٨	٠,٠٣٩			
٤٩	٠,٩٩	٠,٠٩١			
٥٠	٠,٦٨	٠,٠٦٩			
٥١	١,١٢	٠,٠٤١			
٥٢	١,٨٣	٠,٠٢٠			

أشارت قيم RMSD الموضحة في الجدول السابق إلى وجود أداء تفاضلي لمتغير الجنس فى عدد (19) مفردة من مفردات القدرة اللفظية بنسبة (36.5%) من إجمالي مفردات القدرة اللفظية البالغ عددها (٥٢) مفردة ، وكذلك وجود أداء تفاضلى لمتغير الجنس فى عدد (14) مفردة من مفردات القدرة الكمية بنسبة (31.8%) من إجمالي مفردات القدرة الكمية البالغ عددها (٤٤) مفردة .

وقد اتفقت نتائج كل من الطريقتين مربع كاي لكاميللي Camilli Chi – Square) (والفروق في المساحات تحت منحنيات خصائص المفردات في الحكم على وجود أداء تفاضلي لمتغير الجنس في (١٥) مفردة في حالة القدرة اللفظية تمثل نسبة (٢٨,٩ %) من إجمالي مفردات القدرة اللفظية ، وفي (١١) مفردة في حالة القدرة الكمية تمثل نسبة (٢٥ %) من إجمالي مفردات القدرة الكمية .

ولتحديد نوع الأداء التفاضلي : منتظم Uniform ، أو غير منتظم Non Uniform في المفردات التي أظهرت أداء تفاضليا تبعا لمتغير الجنس ، تم رسم منحنيات خصائص المفردات ذات الأداء التفاضلي في كل من القدرتين (اللفظية[٢] – الكمية[٣]) ، حيث أظهرت المفردات أرقام (5 ، 6 ، 13 ، 14 ، ١٦ ، ١٧ ، ١٩ ، ٢٣ ، ٣١ ، ٣٣ ، ٣٧ ، ٤٣ ، ٤٥ ، ٤٦ ، ٥٠) أداء تفاضليا منتظما في حالة القدرة اللفظية ، بينما أظهرت المفردات أرقام (١٢ ، ٢٥ ، ٣٠ ، ٤٩) أداء تفاضليا غير منتظم .

---

[٢] ملحق (٢) منحنيات خصائص المفردات التي أظهرت أداء تفاضليا تبعا لمتغير الجنس في القدرة اللفظية .

[٣] ملحق (٣) منحنيات خصائص المفردات التي أظهرت أداء تفاضليا تبعا لمتغير الجنس في القدرة الكمية .

أما في حالة القدرة الكمية فقد أظهرت المفردات أرقام (١٤ ، ١٧ ، ٢١ ، ٢٣ ، ٢٤ ، ٣٢ ، ٣٥ ، ٣٩ ، ٤١) أداء تفاضليا منتظما ، بينما أظهرت المفردات أرقام (١ ، ٩ ، ١٠ ، ٢٧ ، ٣٧) أداء تفاضليا غير منتظم .

الإجابة عن السؤال الثانى :

للإجابة عن هذا التساؤل قام الباحثان بحساب متوسطات الخطأ المعيارى لتقدير القدرة قبل وبعد حذف المفردات ذات الأداء التفاضلى لمتغير الجنس في حالتي تقدير كل من القدرة اللفظية والكمية ، وحساب قيمة "ت" للفروق بين متوسطي الأخطاء المعيارية لتقدير كل منهما ، وكذلك حساب حجم التأثير الناتج للمفردات ذات الأداء التفاضلي على قيم الأخطاء المعيارية لتقدير كل من القدرتين ، والجدول التالي يوضح هذه النتائج :

جدول (٨)

متوسطات القدرة اللفظية والكمية قبل وبعد حذف المفردات ذات الأداء التفاضلي وحجم تأثيرها على الأخطاء المعيارية لتقديرها والدلالة الإحصائية للفروق بين متوسطي الأخطاء المعيارية لتقديرات كل منهما

القدرة اللفظية							القدرة الكمية						
قبل الحذف		بعد الحذف		قيمة ت	الدلالة الاحصائية	حجم التأثير	قبل الحذف		بعد الحذف		قيمة ت	الدلالة الاحصائية	حجم التأثير
متوسط تقدير القدرة	متوسط الخطأ المعيارى	متوسط تقدير القدرة	متوسط الخطأ المعيارى				متوسط تقدير القدرة	متوسط الخطأ المعيارى	متوسط تقدير القدرة	متوسط الخطأ المعيارى			
-٠,٠١	٠,٣١٤	-٠,١٦	٠,٢٨٩	-٣٦١,٥١٤	٠,٠٠٠	٠,٩٩	-٠,٢٦	٠,٣٤٤	-٠,٣٢	٠,٢٩٨	-٢٤٠,٧٢٧	٠,٠٠٠	٠,٩٩

يتضح من الجدول السابق أن متوسط الخطأ المعيارى قبل حذف المفردات ذات الأداء التفاضلى لمتغير الجنس في حالتي تقدير القدرة اللفظية والقدرة الكمية أكبر منه بعد الحذف ، كما يتضح وجود دلالة إحصائية عند مستوى ٠,٠١ للفروق بين قيم متوسطي الخطأ المعياري لتقدير القدرة سواء اللفظية أو الكمية ، مما يدل على أن حذف المفردات ذات الأداء التفاضلي من الاختبار قد أدى إلى خفض قيم الأخطاء المعيارية لتقدير القدرة في الحالتين ، الأمر الذي يمكن أن يسهم في تحقيق الدقة والكفاءة في تقدير قدرات الطلاب.

كما يتضح أن بقاء المفردات ذات الأداء التفاضلي يؤثر بدرجة كبيرة على قيمة الخطأ المعياري لتقدير القدرة ، الأمر الذي يفقدها الدقة والكفاءة المطلوبة ، ويتضح ذلك من

خلال قيمتي حجم التأثير سواء في حالة تقدير القدرة اللفظية أو تقدير القدرة الكمية للطلاب ، حيث أن قيمتي حجم التأثير الناتجتين تدلان على حجم تأثير كبير للمفردات ذات الأداء التفاضلي على قيم الخطأ المعياري لتقدير كل من القدرتين .

الإجابة عن السؤال الثالث :

للإجابة عن هذا التساؤل قام الباحثان بحساب متوسطات تقدير كل من القدرة اللفظية والقدرة الكمية لكل من الذكور والإناث ومتوسطات الخطأ المعيارى المتعلقة بكل منها قبل وبعد حذف المفردات ذات الأداء التفاضلى لمتغير الجنس ، والجدول التالي يوضح تلك النتائج :

جدول (٩)

متوسطات القدرة اللفظية والكمية لكل من الذكور والإناث قبل وبعد حذف المفردات ذات الأداء التفاضلي والأخطاء المعيارية لتقديرات كل منهما والدلالة الإحصائية للفروق بينها

نوع الاختبار	قبل / بعد حذف المفردات ذات الأداء التفاضلى	متوسط تقديرات القدرة		قيمة ت		الدلالة الاحصائية		متوسط الخطا المعيارى لتقديرات القدرة		قيمة ت		الدلالة الاحصائية	
		ذكور	إناث	ذكور	إناث	ذكور	إناث	ذكور	إناث	ذكور	إناث	ذكور	إناث
القدرة اللفظية	قبل	-٠,١٧	-٠,١٤	٥١,٤٢	٤٧,٩٩	٠,٠٠٠	٠,٠٠٠	٠,٣٢	٠,٣١	-٢٣٤,٣٠	-٣٠٨,٣٩	٠,٠٠٠	٠,٠٠٠
	بعد	-٠,٣١	-٠,٠٣					٠,٤٨	٠,٤٧				
القدرة الكمية	قبل	-٠,٢٨	-٠,٢٥	٥٣,٨٩	٥٣,٥٥	٠,٠٠٠	٠,٠٠٠	٠,٣٥	٠,٣٤	-١٧١,٦٦	-١٨٧,٩٥	٠,٠٠٠	٠,٠٠٠
	بعد	-٠,٣٣	-٠,٢٩					٠,٥١	٠,٥٠				

يتضح من الجدول السابق وجود دلالة إحصائية عند مستوى ٠,٠١ للفروق بين كل من متوسطي تقديرات القدرة سواء اللفظية أو الكمية لكل من الذكور والإناث ، وكذلك بين متوسطي الخطأ المعياري لتقديرات القدرة قبل وبعد حذف المفردات ذات الأداء التفاضلي من الاختبار ، مما يدل على أن حذف المفردات ذات الأداء التفاضلي من الاختبار قد أدى إلى خفض قيم الأخطاء المعيارية لتقديرات القدرة سواء اللفظية أو الكمية ، الأمر الذي يمكن أن يسهم في تحقيق الدقة والكفاءة في تقدير قدرات الطلاب .

الإجابة عن السؤال الرابع : للإجابة عن هذا التساؤل قام الباحثان بحساب متوسطات قيم كل من دوال المعلومات لكل من الاختبارين الفرعيين والاختبار الكلي ومتوسطات الأخطاء المعيارية المتعلقة بكل منها وقيم "ت" لتحديد دلالة الفروق بينها قبل وبعد حذف المفردات ذات الأداء التفاضلى لمتغير الجنس ، والجدول التالي يوضح هذه النتائج :

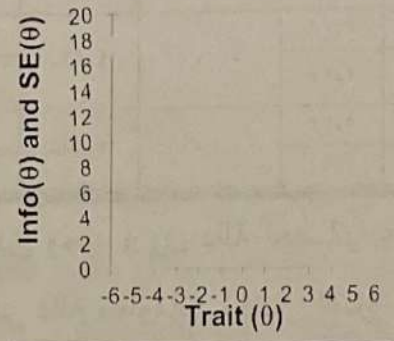
جدول (١٠)

متوسطات تقدير دالة معلومات الاختبار الكلي والاختبارات الفرعية قبل وبعد حذف المفردات ذات الأداء التفاضلي والأخطاء المعيارية لها وانحرافاتها المعيارية والدلالة الإحصائية للفروق بينها

نوع الاختبار	قبل / بعد حذف المفردات ذات الأداء التفاضلي	دالة المعلومات		قيمة ت	الدلالة الاحصائية	الخطأ المعيارى		قيمة ت	الدلالة الاحصائية
		م	ع			م	ع		
اختبار القدرة اللفظية	قبل	١,٩٥	٢,٤٨	١,٢٢-	٠,٠٠٧	٣,٤٧	٣,٩٣	٣,٠٣-	٠,٠٠٤
	بعد	٢,٠٥	٢,٨٤			٣,١٧	٣,٥٧		
اختبار القدرة الكمية	قبل	١,٦٩	٢,٥٤	٤,٤٦-	٠,٠٠٠	٣,٠٨	٤,٠٢	٢,٧١٢-	٠,٠٠٩
	بعد	١,٩١	٢,٧٦			٢,٩٥	٣,٤٠		
الاختبار الكلي	قبل	٣,٢٤	٢,٨١	٥,٦٢-	٠,٠٠٠	٥,٣٣	٤,٢٤	٤,٦٦-	٠,٠٠٠
	بعد	٣,٨٢	٢,٩٨			٤,٩٠	٣,٧٢		

يتضح من الجدول السابق وجود فروق دالة إحصائيا عند مستوى ٠,٠١ في الحالات الثلاث لقيم دوال المعلومات وهي دالة معلومات اختبار القدرة اللفظية ودالة معلومات اختبار القدرة الكمية ودالة معلومات الاختبار الكلي قبل وبعد حذف المفردات ذات الأداء التفاضلي لمتغير الجنس ، ويتضح أيضا وجود فروق دالة إحصائيا عند مستوى ٠,٠١ بين متوسطات قيم الأخطاء المعيارية لتقديرات دالة معلومات الاختبار في الحالات الثلاث ، سواء في الاختبارات الفرعية لتقدير القدرتين (اللفظية – الكمية) ، أو في الاختبار الكلي قبل وبعد حذف المفردات ذات الأداء التفاضلي لمتغير الجنس ، مما يدل على أن حذف المفردات ذات الأداء التفاضلي من الاختبار قد أدى إلى خفض قيم الأخطاء المعيارية لتقديرات دالة معلومات الاختبار ، سواء في الاختبارات الفرعية أو في الاختبار الكلي ، الأمر الذي يمكن أن يسهم في زيادة الدقة في تقدير قيمة دالة معلومات الاختبار ، مما يجعلها أكثر فعالية في إمدادنا بمعلومات أكثر عن قدرات الطلاب ومدى فعالية الاختبار في تقديرها ، بما يساعدنا في فهم قدرات الطلاب وتحديد جوانب القوة والضعف فيها بدرجة أكثر دقة وكفاءة ، وكذلك يمكن أن تساعدنا بدرجة أكبر في تفسير الضعف أو القوة في بعض جوانب القدرة والكشف عن أسباب الفروق بينها .

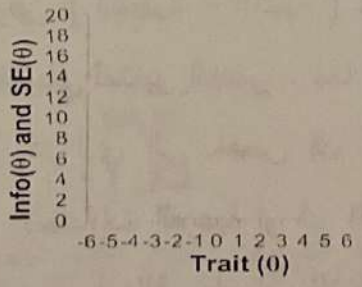
ثم قام الباحثان برسم دوال معلومات الاختبارات الفرعية المتعلقة بتقدير كل من القدرة اللفظية والقدرة الكمية ، وكذلك رسم دالة معلومات الاختبار الكلي بيانيا قبل وبعد حذف المفردات ذات الأداء التفاضلي لمتغير الجنس بغرض تقديم مزيد من التوضيح للنتائج ، وأيضا لمساعدتنا في تفسير النتائج بشكل أكثر دقة وموضوعية ، والشكلان البيانيان التاليان يمثلان دوال معلومات الاختبارين الفرعيين قبل حذف المفردات ذات الأداء التفاضلي لمتغير الجنس :



شكل (٣)

دالة معلومات اختبار القدرة اللفظية قبل حذف

المفردات ذات الأداء التفاضلي لمتغير الجنس



شكل (٤)

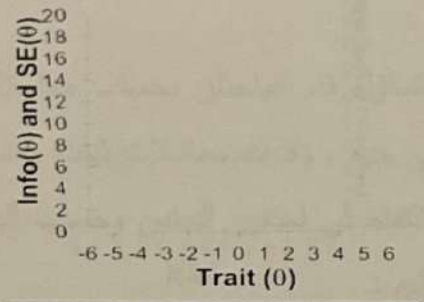
دالة معلومات اختبار القدرة الكمية قبل حذف

المفردات ذات الأداء التفاضلي لمتغير الجنس

يتضح من الشكلين السابقين أثر بقاء المفردات ذات الأداء التفاضلي لمتغير الجنس على دوال معلومات الاختبارين الفرعيين ، والذي يتمثل في عدم انسيابية المنحنى بشكل طبيعي ، كما أن حدود طرفيه عند مستوى متدني للمعلومات المتعلقة بقدرات الطلاب ،

وكذلك ضيق المساحات تحت المنحني ، مما يؤثر على كمية المعلومات التي يمكن أن تقدمها دالة معلومات كل من الاختبارين وكذلك على مستوى أقصى المعلومات لها .

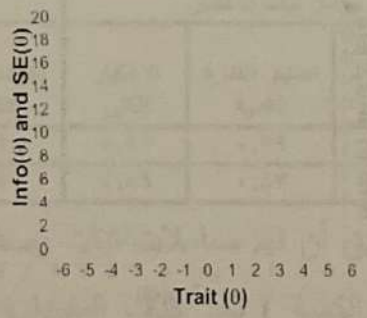
أما الشكلين التاليين فيمثلان دوال معلومات الاختبارين الفرعيين بعد حذف المفردات ذات الأداء التفاضلي لمتغير الجنس :



شكل (٥)

دالة معلومات اختبار القدرة اللفظية بعد حذف

المفردات ذات الأداء التفاضلي لمتغير الجنس



شكل (٦)

دالة معلومات اختبار القدرة الكمية بعد حذف

المفردات ذات الأداء التفاضلي لمتغير الجنس

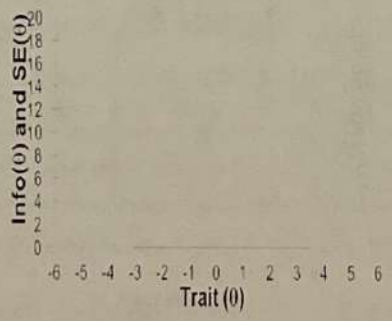
يتضح من الشكلين السابقين وجود اختلاف واضح عن الشكلين (٣) ، (٤) يتمثل في انسيابية المنحنى بصورة أفضل ، واتساع المساحات تحت المنحني ، وكذلك ابتعاد طرف المنحنى عن المحور الرأسي (محور المعلومات) ، مما يزيد من كمية المعلومات التي يمكن أن تقدمها دوال معلومات كل من الاختبارين وكذلك مستوى أقصى المعلومات لها .

أما الشكلين التاليين فيمثلان دالة معلومات الاختبار الكلي قبل وبعد حذف المفردات ذات الأداء التفاضلي لمتغير الجنس :

شكل (٧)

دالة معلومات الاختبار الكلي قبل حذف

المفردات ذات الأداء التفاضلي لمتغير الجنس



شكل (٨)

دالة معلومات الاختبار الكلي بعد حذف

المفردات ذات الأداء التفاضلي لمتغير الجنس

يتضح من الشكل (٧) أثر بقاء المفردات ذات الأداء التفاضلي لمتغير الجنس على دالة معلومات الاختبار الكلي ، والذي يتمثل في تحديد طرفيه عند مستوى متدني للمعلومات المتعلقة بقدرات الطلاب ، وكذلك ضيق المساحات تحت المنحني ، مما يؤثر على كمية المعلومات التي يمكن أن تقدمها دالة معلومات الاختبار الكلي وكذلك على مستوى أقصى المعلومات لها .

ويتضح من الشكل (٨) وجود اختلاف واضح عن الشكل(٧) لدالة معلومات الاختبار الكلي يتمثل في اتساع المساحات تحت المنحني ، وكذلك ابتعاد طرف المنحنى عن المحور الرأسي (محور المعلومات)، مما يزيد من كمية المعلومات التي يمكن أن تقدمها دالة معلومات الاختبار الكلي وكذلك مستوى أقصى المعلومات لها .

الإجابة عن السؤال الخامس :

للإجابة عن هذا التساؤل قام الباحثان بحساب معاملات ثبات وصدق الاختبارات الفرعية المتعلقة بكل قدرة على حدة ، وكذلك معاملات ثبات وصدق الاختبار الكلي قبل وبعد حذف المفردات ذات الأداء التفاضلى لمتغير الجنس وحساب الدلالة الإحصائية لكل منها ، والجدول التالي يبين هذه النتائج :

جدول (١١)

معاملات صدق وثبات الاختبار الكلي والاختبارات الفرعية قبل وبعد حذف المفردات ذات الأداء التفاضلي لمتغير الجنس

قبل / بعد حذف المفردات ذات الأداء التفاضلى	معامل ثبات الاختبار			معامل صدق الاختبار		
	اختبار القدرة اللفظية	اختبار القدرة الكمية	الاختبار الكلي	مفردات القدرة اللفظية	مفردات القدرة الكمية	الاختبار الكلي
قبل الحذف	٠,٧٦	٠,٧٩	٠,٨٠	٠,٧١	٠,٧٣	٠,٧٩
بعد الحذف	٠,٨١	٠,٨٣	٠,٨٦	٠,٧٤	٠,٧٧	٠,٨٢

يتضح من الجدول السابق أن قيم معاملات ثبات وصدق الاختبارات الفرعية المتعلقة بكل من القدرتين (اللفظية – الكمية) بعد حذف المفردات ذات الأداء التفاضلى لمتغير الجنس أكبر منها قبل حذف المفردات ذات الأداء التفاضلى لمتغير الجنس ، كما يتضح أيضا أن قيم معاملات ثبات وصدق الاختبار الكلي بعد حذف المفردات ذات الأداء التفاضلى لمتغير الجنس أكبر منها قبل حذف المفردات ذات الأداء التفاضلى لمتغير الجنس . وكذلك فقد تبين للباحثين باستخدام برنامج SPSS أن قيم معاملات ثبات وصدق الاختبارات سواء الفرعية أو الاختبار الكلي دالة إحصائيا عند مستوى ٠,٠٥ . وهذا يدل على أن بقاء المفردات ذات الأداء التفاضلى في الاختبار الكلي أو في الاختبارات الفرعية تؤدي إلى انخفاض

خصائصه السيكومترية ، الأمر الذي يشوه الاختبار ويفقده أهميته وقيمته القياسية والسيكومترية ، مما يؤدي إلى التشكيك في نتائجه وانخفاض الثقة فيها .

مناقشة النتائج والتعليق عليها :

بينت النتائج أن بعض مفردات اختبار القدرات السعودي تظهر أداء تفاضليا لمتغير الجنس ، وهو ما يمكن تفسيره في ضوء المناقشة التالية :

يظهر الأداء التفاضلي لبعض المفردات في كثير من الاختبارات لأسباب خارجة عن طبيعة الاختبار نفسه ، فالإجابة عن مفردات اختبار القدرات قد تتأثر بسمات أخرى بجانب السمة الرئيسة التي صمم الاختبار من أجل قياسها ، وهو ما يتفق مع ما أشار إليه كل من صلاح علام (٢٠٠٥ ، ٦٣) ، ولورد Lord ، حيث يروا أن افتراض أحادية البعد في الاختبارات المعرفية والعقلية لا يعد افتراضا صحيحا كليا ، فعلى الرغم من اعتماد التحليلات الخاصة بالكشف عن الأداء التفاضلي للمفردات على المزاوجة بين المجموعتين في السمة الكامنة المقاسة إلا أن المجموعتين قد تكونان مختلفتين في سمات أخرى لا تقيسها المفردة ، ومن هنا يمكن استنتاج أن اختلاف المجموعتين في السمات الأخرى المؤثرة في الإجابة عن المفردة قد يسهم في ظهور الأداء التفاضلي للمفردة .

ولذلك قد يكون الأداء التفاضلي ناتج عن الاختلافات بين الطلاب الذكور والإناث في طريقة الاستجابة للمفردة أو معنى المفردة لدى الفرد ، وارتباط محتوى المفردة بخبرات الفرد السابقة وسماته الشخصية . وهذا لا يتضح إلا بفحص المفردات التي ظهر لها أداء تفاضليا ، فقد أشار كل من Penfield , et al. , (2007, 453) , Ariffin , et al. (2010 , 2) إلى أن فحص محتوى المفردات التي ظهر أن لها أداء تفاضليا يعد من أفضل الطرق في تحديد أسباب الأداء التفاضلي للمفردة ، حيث أن ذلك يعد نوعا من التحليلات الكيفية المكملة للتحليلات الكمية .

وهنا لا بد من القول أن هناك عدة عوامل تتحكم في كون المفردات الاختبارية تظهر أداءً تفاضلياً أم لا ؟ وهذه العوامل يمكن إيجازها فيما يلي :

أ) كون الاختبار أحادي البعد أم لا ، بمعنى هل يقيس الاختبار سمة واحدة أم لا ؟

ب) الاختلاف في توزيع القدرة بين أفراد المجموعتين .

ج) الاختلاف في نوع المفردة .

د) التخمين.

ولا بد من تذكر عامل العملية العقلية التي تقيسها المفردة ، أهي استدلال ، أم عمليات روتينية ، أم تذكر أم عمليات أخرى . حيث أنها يمكن أن تُحدث أداءً تفاضلياً بغض النظر عن مستوى الطلاب (Leder, 1992). .

فعند تفسير الأداء التفاضلي في ضوء محتوى المفردة والسمات الثانوية المرتبطة بالإجابة الصحيحة ، يجب التفريق بين العوامل الثانوية الجوهرية التي تعد عاملا مساعدا في السمة المراد قياسها وتدخل في تحديد مستوى الفرد في السمة وبين العوامل الثانوية الخارجية ، والتي لا يهدف الاختبار قياسها ، ففي النوع الأول تكون الفروق في الأداء على المفردة بين المجموعات قد تعبر عن فروق حقيقية في السمة ، أما في النوع الثاني فإن الفروق في الأداء على المفردة يمكن اعتبارها أداء تفاضليا للمفردة ، وهو ما يتفق مع ما أشار إليه , Penfield et al. , (2007 , 453) حيث يؤكد على أن ظهور الأداء التفاضلي للمفردات يمكن أن يكون مؤكدا للصدق البنائي للاختبار ، وذلك في حالة كون العوامل الثانوية التي قد تكون مسؤولة عن الفروق في الأداء على المفردة ثابتة في كل المفردات التي أظهرت أداء تفاضليا ، وكونها جوهرية في تحديد مستوى السمة،أما إذا كانت هذه العوامل مختلفة من مفردة لأخرى وليست جوهرية في تحديد مستوى السمة ، فإن الأداء التفاضلي في هذه الحالة يكون مهددا لصدق الاختبار .

ولا يختلف أحد على علاقة كل من القدرة اللفظية والقدرة الكمية لدى الفرد بالعديد من القدرات العقلية الأخرى والتي قد تكون مشتركة في تحديد السمة أو القدرة لدى الفرد ، وليست هناك مشكلة في كون هذه السمات هي المسببة للأداء التفاضلي للمفردة ، ففي هذه الحالة تعكس المفردة فروقا حقيقية في السمة المستهدفة . ولكن يتضح من المناقشة السابقة أن إجابة بعض المفردات التي أظهرت أداء تفاضليا يمكن أن تتأثر بعوامل ثانوية ليست جوهرية في تحديد مستوى السمة ، وهو ما يتفق مع ما أكدته نتائج دراسات كل من (Gierl & Liu , 2011) , (Bolt , 2001 والتي أكدت على أن من أهم مسببات الأداء التفاضلي

للمفردات محتوى المفردة ، واختلاف معناه من مجموعة لأخرى ، وكذلك المتغيرات المعرفية المختلفة المطلوبة للإجابة عن المفردة ، وعدم تكافؤ هذه المتغيرات بين المجموعات .

وكذلك يتفق مع ما أكدت عليه دراسة (De Boeck , et al. , 2011 , 584) وأطلقت عليه الأداء التفاضلي الكامن ، حيث أرجعت الأداء التفاضلي لبعض المفردات إلى اختلاف المجموعتين في بعض الخصائص الأخرى المرتبطة بالسمة المستهدفة ، والتي تجعل أحد المجموعتين لها الأولوية في الإجابة الصحيحة للمفردة ، نظرا لأن التحليلات الإحصائية يتم فيها فقط تحييد مستوى السمة موضوع القياس في المجموعتين . وهو نفس ما أكدت عليه دراسة كل من (Price , 1999 , 7) , (Abbott , 2007 , 9) ، حيث اعتبرت أن وجود بعض المفردات التي تتأثر الإجابة عليها بعوامل ثانوية لا يهدف الاختبار إلى قياسها يسهم في زيادة احتمال ظهور الأداء التفاضلي لتلك المفردات ، وكذلك يتفق مع ما أشار إليه (Tay et al. , 2016, 23) من أن اعتماد التحليلات الخاصة بالكشف عن الأداء التفاضلي على المقارنة بين المجموعتين بعد تحييد مستوى السمة المقاسة قد يتأثر بعوامل ثانوية مرتبطة بالسمة ولم يتم تحييدها في المجموعتين .

كما بينت النتائج اختلاف نتائج طريقتي الكشف عن المفردات ذات الأداء التفاضلي المستخدمة في الدراسة الحالية (مربع كاي لكاميللي (Camilli Chi – Square) – الفروق في المساحات تحت منحنيات خصائص المفردات) ، لكنهما اتفقتا في الحكم على وجود أداء تفاضلي لمتغير الجنس في (١٥) مفردة في حالة القدرة اللفظية تمثل نسبة (٢٨,٩ %) من إجمالي مفردات القدرة اللفظية ، وفي (١١) مفردة في حالة القدرة الكمية تمثل نسبة (٢٥ %) من إجمالي مفردات القدرة الكمية .

وفي ضوء ما سبق فقد تأكد من النتائج الحالية تأثير المفردات التي لها أداء تفاضليا على الأداء التفاضلي للاختبار ككل ، حيث أن هذه المفردات تؤثر على تماثل البنية الكامنة المقاسة باستخدام الاختبار لدى كل من الذكور والإناث وهو ما يمكن تفسيره بأن زيادة عدد المفردات التي لها أداء تفاضليا في الاختبار ، وهو ما يتفق مع نتائج بعض الدراسات التي أشارت إلى أنه في حالة زيادة عدد المفردات التي لها أداء تفاضليا لا يكون للدرجات النهائية

في الاختبار ككل نفس المعنى في المجموعات المختلفة ويتسبب في تحيز تقدير السمة (le Bouthillier, et al. , 2015) , (Walker , 2011) , (Church , et al. ,2011) .

وفي الختام يجب أن نشير إلى أن فحص المنحنيات المميزة للمفردات التي أظهرت أداء تفاضليا على مستوى القدرتين اللفظية والكمية بشكل عام ، نلاحظ أن الأداء التفاضلي لمعظمها كان قليلا ، مما يؤكد أنه يرجع لعوامل خارجة عن طبيعة الاختبار وخصائصه السيكومترية ، مما يجعل تأثيره على الأداء التفاضلي للاختبار الكلي ضعيفا . ويمكن تفسير تلك النتيجة في ضوء قيمة β كمؤشر للأداء التفاضلي ، وهو ما يفسر تأثير الأداء التفاضلي للمفردات على الأداء التفاضلي للاختبار الكلي .

التوصيات والبحوث المقترحة

أولا : توصيات الدراسة

في ضوء ما أشارت إليه نتائج الدراسة الحالية ، يقدم الباحثان التوصيات التالية :

١. زيادة وعي القائمين على بناء وتطوير الاختبارات بخاصية الأداء التفاضلي للمفردات الاختبارية وأثرها الهادم لخاصية صدق الاختبارات .
٢. إعادة النظر في الاختبارات المستخدمة حاليا ودراسة مفرداتها للكشف عن الأداء التفاضلي فيها بغرض تحسين الاختبارات وزيادة فاعليتها في عمليات القياس والتقويم
٣. الانتباه إلى أهمية حذف أو تعديل المفردات ذات الأداء التفاضلي من الاختبارات في تحسين أداء الاختبارات وزيادة الثقة في نتائجها .
٤. إدراك خطورة الأداء التفاضلي للمفردات الاختبارية على دقة وكفاءة الاختبارات .
٥. مراعاة الآثار السلبية للمفردات ذات الأداء التفاضلي على الخصائص السيكومترية للاختبارات خصوصا الصدق .
٦. عقد دورات تدريبية للمعلمين والقائمين على بناء وتصميم الاختبارات ، لتدريبهم على كيفية الكشف عن المفردات ذات الأداء التفاضلي وكيفية التعامل معها .

ثانيا : البحوث المقترحة

في ضوء نتائج الدراسة الحالية تنبثق البحوث والدراسات التالية :

١- دراسة الأداء التفاضلي لمفردات اختبار القدرة السعودي لمتغيرات أخرى غير الجنس .

٢- تحليل محتوى المفردات التي تُظهر أداء تفاضلياً لمتغير الجنس أو متغيرات أخرى في اختبار القدرة السعودي من حيث المحتوى والعمليات العقلية.

٣- تشكيل لجنة من الخبراء لدراسة مفردات اختبار القدرة السعودي من حيث تحيز المفردات بناءً على نتائج دراسات الأداء التفاضلي للمفردات .

٤- إجراء دراسات مقارنة للأداء التفاضلي لمفردات اختبار القدرة السعودي باستخدام عدة طرق للكشف عن الأداء التفاضلي للمفردات .

المراجع

أولا : المراجع العربية

١- أمينة محمد كاظم (١٩٩٦) : نماذج السمات الكامنة , فى : أنور الشرقاوى , سليمان الخضرى , أمينة كاظم , نادية عبد السلام : اتجاهات معاصرة فى القياس والتقويم النفسى والتربوى , مكتبة الأنجلو المصرية , القاهرة , ص ص٢٧٩ – ٤٣٠ .

٢- إسماعيل حسن الوليلى (٢٠٠٢) : دراسة سيكومترية مقارنة لبعض نماذج الاستجابة للمفردة فى انتقاء مفردات الاختبارات مرجعية المحك, رسالة دكتوراة غير منشورة , كلية التربية – جامعة الأزهر , القاهرة .

٣- إسماعيل حسن الوليلى ، أيمن صابر حجازي (٢٠١٢) : فعالية تحليل نتائج الاختبار التحصيلي وفق النموذج ثلاثي البارامتر في التنبؤ بالتلاميذ ذوي صعوبات التعلم في مادة الرياضيات بالمرحلة الإعدادية ، مجلة كلية التربية ، جامعة بنها ، المجلد ٢٣ ، العدد ٩٢ ، (ج ٢) .

٤- إسماعيل سلامه البرصان (٢٠١٣) : الأداء التفاضلي لمتغير الجنس لفقرات الاختبار الوطني الأردني لضبط نوعية التعلم لمادة الرياضيات للصف العاشر، مجلة كلية التربية بالزقازيق ، ٢٨ ، (٧٩) ، ص ١ – ٣٥ .

٥- إياد حمادنة (٢٠٠٧) : الأداء التفاضلي لفقرات اختبارات تحديد الكفاءة اللغوية في اللغة الإنجليزية في الجامعات الأردنية : دراسة مقارنة ، رسالة دكتوراه غير منشورة ، جامعة اليرموك ، إربد ، الأردن

٦- إيهاب محمد نجيب (٢٠٠٨) : طرق الكشف عن تحيز مفردات الاختبار ، دراسة سيكومترية مقارنة ، رسالة ماجستير غير منشورة ، كلية التربية ، جامعة عين شمس .

٧- المركز الوطني للقياس والتقويم (٢٠١٣) : اختبار القدرات العامة ، الرياض ، المملكة العربية السعودية .

٨- جهاد العجلوني ، يوسف السوالمة (٢٠١٦) : فحص الأداء التفاضلي لمموهات اختبار الرياضيات في التقييم الوطني الأردني لاقتصاد المعرفة وفقا لجنس الطالب وموقع المدرسة ، مجلة جامعة النجاح لأبحاث العلوم الإنسانية ،٣٠ (١١) ، ص ٢٢٦٥ – ٢٢٨٦ .

٩- خالد الحياصات (٢٠١١) : مدى تحقق معايير الفاعلية في معادلة نموذجي اختبار مع بقاء الفقرات ذات الاداء التفاضلي للجنس وحذفها ، رسالة دكتوراه ، جامعة اليرموك ، إربد ، الأردن .

١٠- راتب أرحيل (٢٠١١) : أثر أداء الفقرات التفاضلي للجنس في دقة المعادلة العمودية لاختبا أروتيس لينون للقدرة العقلية ، رسالة دكتوراة غير منشورة ، جامعة اليرموك ، اربد ، الأردن .

١١- ربيع عبده رشوان (٢٠١٥) : الأداء التفاضلي لبنود "اختبار واطسون وجليسر للتفكير الناقد" باستخدام النماذج البارامترية واللابارامترية في نظرية الاستجابة للمفردة وعلاقته بالأداء التفاضلي للاختبار باستخدام التحليل العاملي التوكيدي متعدد المجموعات ، مجلة كلية التربية – جامعة الأزهر – ع ١٦٥ ، ج ٢ ، ص ٤٥٥ – ٥٢٥ .

١٢- رجاء محمود أبو علام ، عامر جبريل المرابحة (٢٠١٥) : أثر نسبة المفردات ذات الأداء التفاضلي وحجم العينة على قوة وفاعلية اختبار تحيز المفردة المتزامن : دراسة محاكاة ، المجلة العربية للعلوم الاجتماعية ، ع ,7 ج2 ، ص 139 – 165 .

١٣- صلاح الدين محمود علام (٢٠٠٢) : القياس والتقويم التربوى والنفسى (أساسياته وتطبيقاته وتوجهاته المعاصرة) , دار الفكر العربى , القاهرة .

١٤- صلاح الدين محمود علام (٢٠٠٥) : نماذج الاستجابة للمفردة الاختبارية أحادية البعد ومتعددة الأبعاد وتطبيقاتها فى القياس النفسى والتربوى , دار الفكر العربى , القاهرة .

١٥- علي صالح النوافلة (٢٠١٣) : أثر نسبة الفقرات ذات الأداء التفاضلي ومستواه في تقدير معالم فقرات الاختبار وقدرات الأفراد وفق نموذج الاستجابة للفقرة ثلاثي المعلم . رسالة دكتوراه ، جامعة اليرموك ، أربد، الأردن .

١٦- علي صالح زجا نوافله (٢٠١٧) : أثر الفقرات ذات الأداء التفاضلي المنتظم في تقديرات معالم فقرات الاختبار والافراد في ضوء بعض المتغيرات باستخدام بيانات مولدة وفقاً لنموذج استجابة الفقرة ثلاثي المعلمة ، العلوم التربوية ، ٤٤(٤) ، ١٨٧-٢٠٧ .

١٧- كوثر أبو حماد (٢٠٠٨) : تقصي فاعلية منحنى خصائص الفقرة في الكشف عن الأداء التفاضلي لمتغير الجنس في عينة مختارة من اختبار في الرياضيات ، رسالة ماجستير غير منشورة ، جامعة اليرموك ، إربد ، الأردن .

١٨- مايسة فاضل ابو مسلم (٢٠١٠) : الكشف عن تحيز بنود الاختبار وفقا لمدخل نظرية الاستجابة للمفردة (IRT) ومدخل مانتل هانزل : (H-M) دراسة مقارنة ، مجلة كلية التربية ، جامعة الإسكندرية ، مج 20 ، ع 2 ، ص 189 – 237 .

١٩- محمد إجبارة (٢٠١٥) : أثر نسبة الفقرات ذات الأداء التفاضلي على الخطأ من النوع الأول وقوة الاختبار في الكشف عنه باستخدام طريقتي الانحدار اللوجستي ونسبة الارجحية العظمى ، (رسالة دكتوارة) ، جامعة اليرموك ، اربد ، الأردن .

٢٠- معن محمد الهنائي (2017) : تقويم أسئلة اختبار الرياضيات للصف السابع في برنامج التنمية المعرفية بسلطنة عمان ، رسالة ماجستير غير منشورة ، جامعة السلطان قابوس ، سلطنة عمان

٢١- نبيل عبدالعزيز (٢٠٠٦) : فاعلية طريقة منحنى خصائص الفقرة في الكشف عن الأداء التفاضلي لمتغير الجنس في فقرات اختبار رياضيات متعدد التدريج وثنائي التدريج ، رسالة دكتوراه غير منشورة ، جامعة عمان العربية ، عمان ، الأردن .

٢٢- نزار اللبدي (٢٠٠٨) : مقارنة أربع طرق للكشف عن دالة الفقرة التفاضلية (دراسة محاكاة) ، رسالة دكتوراه غير منشورة ، الجامعة الأردنية ، عمان ، الأردن .

٢٣- وائل مبارك (٢٠١٠) : الأداء التفاضلي لفقرات اختبار العلوم في الدراسة الدولية بياز (٢٠٠٦) ، رسالة دكتوراه غير منشورة ، جامعة اليرموك ، إربد ، الأردن .

٢٤- يوسف عبدالقادر أبو شندي ، علي مهدي كاظم ٢٠١٨) : الأداء التفاضلي لفقرات اختبار الرياضيات في برنامج التنمية المعرفية بسلطنة عمان تبعا لمتغير الجنس باستخدام طريقتي مانتل هانزل ومنحنى خصائص الفقرة ، 2018) 73-61 ,2 .No ,6 .Sys .Man .Learn .J .Int) .

ثانيا : المراجع الأجنبية

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

26- Abdalaziz, N. (2010). A gender-related differential item functioning of mathematics test items. The international Journal of Educational and Psychological Assessment, 5(1), 101-116.

27- Andrich, D. (2002). Understanding resistance to the data-model relationship in Rasch's paradigm: A reflection for the next generation. Journal of Applied Measurement, 3, 325-359.

28- Andrich , David (2000): An Introduction to Rasch Models for Measurement, Office of Planning, Institutional Research Testing, Evaluation Service, The American University in Cairo, 24 – 30th October.

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

30- Camilli ،G. & Shepard ،L. A. (1994) : Methods for identifying biased test items.Thousand Oaks ،CA: sage .

31- Church, A. T., Alvarez, J. M., Mai, N. T. Q., French, B. F., Katigbak, M. S., & Ortiz, F. A. (2011). Are cross-cultural comparisons of personality profiles meaningful? Differential item and facet functioning in the Revised NEO Personality Inventory. Journal of Personality and Social Psychology, 101(5), 1068-1089.

32- Cohwn , A. & Others (1993) : Detection of Differential Item Functioning in the Graded Response Model , Applied Psychological Measurement , Vol. 17 , No. 4 , P. 335 – 350 .

33- De Boeck, P., Cho, S.-J., & Wilson, M. (2011). Explanatory secondary dimension modeling of latent differential item functioning. Applied Psychological Measurement, 35(8), 583-603.

34- Doolittle, A. E., & Cleary, T. A. (1987). Gender-based differential item performance in mathematics achievement items. Journal of Educational Measurement, 24(2), 157-166.

35- Dorans, N. J., & Holland, P. W. (1993). DIF detection and description: Mantel-Haenszel and standardization. In P. W. Holland & H. Wainer (Eds.), Differential item functioning (pp. 35–66). Hillsdale, NJ: Erlbaum

36- Fan, X. (1998). Item response theory and classical test theory: an empirical comparison of their item/person statistics. Journal of Educational Measurement, 58(3), 357-381..

37- Fidalgo, A. M., Mellenbergh, G. J., &Muñiz, J. (2000). Effects of amount of DIF, test length, and purification type on robustness and power of Mantel-Haenszel procedures. Methods of Psychological Research Online, 5(3), 43-53.

38- Gamer, M. & Engelhard, G. (1999) .Gender differences in performance on multiple -choice and constructed response of mathematics items. Applied Measurement in Education. 12(1),29-43.

39- Gierl M. J. ،Gotzman ،N. A. & Boughton ،K. A. (2004: (Performance Of SIBTEST When The Percentage Of Dif Items Is Large , Applied Measurement InEducation ،17(3),241-264.

40- Gierl, M. J., & Bolt, D. M. (2001). Illustrating the use of nonparametric regression to assess differential item and bundle functioning among multiple groups. International Journal of Testing,1(2&3), 249-270.

41- Hambleton ,R and Rogers, J.(1995) Item bias review Practical Assessment , Research and Evaluation .

42- Hambleton, K. Ronald & Swaminathan, H. & Rogers, H. Jane (1991) : Fundamentals of Item Response Theory, Sage Publications, Inc., California.

43- Hambleton, R. K., & Rogers, H. J. (1989). Detecting potentially biased test items: Comparison of IRT area and Mantel-Haenszel methods. Applied Measurement in Education, 2, 313-334 .

44- Hambleton, Ronald K. & Swaminathan, Hariharan (1985) : Item Response Theory Principles and Applications, Boston, MA : Kluwer Academic Publishers.

45- Harris, A. M., & Carlton, S. T. (1993). Patterns of gender differences on mathematics items on the scholastic aptitude test. Applied Measurement in Education, 6(2), 137-151.

46- 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.

47- Jensen, A. (1980). Bias in mental Testing . New York : A Division of Macmillan publishing Co . Inc.

48- Kamata, A., & Chu, k. (2005). Test equating in the presence of DIF items. Journal of Applied Measurement, 6(3), 342-354.

49- Kappaert, K. (2014). Dissecting the Impact of DIF/ DBF on Ability Estimation and Person Fit (Unpublished Doctoral Dissertation). University of Wisconsin Milwaukee.

50- Katherine, R., &Shuwan, C. (2001). An Examination of Item Context Effects, DIF, and Gender DIF. Applied Measurement in Education, 14(1), 73-90.

51- Kim, H., Plake, B. S., Wise, S. L., & Novak, C. D. (1990). A longitudinal study of sex-related bias in mathematics subtests of the California achievement test. Applied Measurement in Education, 3(3), 275-284.

52- Kim, S., Cohen, A., & Lin, Y. (2005). LDID: A computer program for Local dependence indices for dichotomous items. Version 1.0.

53- Lawrence, I. M. (1995). DIF data on free-response SAT I mathematical items. Princeton, NJ: ETS Report Number ETS-RR-95-22.

54- Lawrence, I. M., & Curley, W. E. (1989). Differential item functioning for males and females on SAT-verbal reading subscore items: Follow-up study. ETS-RR-89-22. Princeton, NJ: educational Testing Service.

55- Lawrence, I. M., Curley, W. E., & McHale, F. J. (1988). Differential item functioning for males and females on SAT-verbal reading subscore items. College Board Report No. 88-4. New York: College Board.

56- LeBouthillier, D., Thibodeau, M., Alberts, N., Hadjistavropoulos, H., & Asmundson, G. (2015). Do people with and without medical conditions respond similarly to the short health anxiety inventory? An assessment of differential item functioning using item response theory. Journal of Psychosomatic Research, 78(4), 384-390.

57- Leder, G. C. (1992). Mathematics and gender: Changing perspectives. In D. A. Grouws (Ed.), Handbook of research on mathematics teaching and learning: A project of the National Council of Teachers of Mathematics (pp. 597-622). New York, NY, England: Macmillan.

58- Lee, Y. (2004). Examining passage-related local item dependence (LID) and measurement construct using Q3 statistics in an EFL reading comprehension test. Language Testing, 21(1), 74-100.

59- Levine, M. V., &Drasgow, F. (1988). Optimal appropriateness measurement. Psychometrika, 53(2), 161-176.

60- Liu, Q. (2011). Item purification in differential using generalized linear mixed models (doctoral dissertation). Florida State University, USA .

61- Lord, F. M. (1980). Application of item response theory to practical testing problems. Hillsdale, NJ: Erlbaum.

62- Magis, D., & De Boeck, P. (2014). Type I error inflation in DIF identification with Mantel-Haenszel: An explanation and a solution. Educational and Psychological Measurement, 74(4), 713-728

63- Mc Callum, R. (2003). Handbook of Nonverbal Assessment. New York: Kluwer Academic/ Plenum Publishers.

64- Mendes-Barnett, S., & Ercikan, K. (2006). Examining sources of gender DIF in mathematics assessment using a confirmatory multidimensional model approach. Applied Measurement in Education, 19(4), 289-304.

65- Ning, W., & Lane, S. (1996). Detection of Gender- related Differential item functioning in mathematics performance assessment. Applied Measurement in Education.9, 171-211.

66- O'Neill, K. A., & McPeek, W. M. (1993). Item and test characteristics that are associated with differential item functioning. In W. Holland & H. Wainer (Eds.), Differential Item Functioning (pp. 255-276). Hillsdale, NJ: Erlbaum.

67- Pae, T. (2004). Gender effect on reading comprehension with Korean EFL learners. System, 32(2), 265-281.

68- Park, G. -P. (2008). Differential item functioning on an English listening test across gender. TESOL Quarterly, 42(1), 115-122.

69- Penfield, R. D., Giacobbi, P. R., & Myers, N. D. (2007). Using the cumulative common log-odds ratio to identity differential item functioning of rating scale items in the exercise and sport sciences. Research Quarterly for Exercise and Sport, 78(5), 451-464.

70- Plake, B. S., Hoover, H. D., & Loyd, B. H. (1980). An investigation of the lowa tests of basic skills for sex bias: A developmental look. Psychology in the Schools, 17(1), 47-52.

71- Price. L. (1999). Differential functioning of items and tests versus the Mantel-Haenszel technique for detecting differential item functioning in a translated test. Paper presented at the Annual Meeting of the American Alliance of Health, Psychical Education, Recreation, and Dance (Boston, MA, April 12-16).

72- Roever, C., (2005). That's not fair" Fairness, bias, and differential item functioning in language testing. Retrieved, February, 6, 2015, from: http // www.hawaii .edul /2ve / brounbag.doc.

73- Rojers, J. (2005). Encyclopedia of Statistics in Behavioral Science. john willy and sons LTD.

74- Schumacker, R. E. (2010). Test bias and item functioning. Retrieved from http://www.appliedmeasurementassociates.com/ama/assets/File/TEST_BIAS_AND_DIFFERENTIAL_ITEM_FUNCTIONING.pdf

75- Sijtsma, K., & Molenaar, I. W. (2002). Introduction to nonparametric item response theory. Thousand Oaks, CA: Sage.

76- Song, X., Cheng, L., & Klinger, D. (2015). DIF investigations across groups of gender and academic background in a large scale high-stakes language test. Papers in Language Testing and Assessment, 4(1), 97-124.

77- Tay, L., Huang, Q., & Vermunt, J. (2016). Item response theory with covariates (IRT-C): assessing item recovery and differential item functioning for the three-parameter logistic model. Educational and Psychological Measurement, 76(1), 22-42 .

78- Tsaousis, I., Sideridis, G., & Al-Saawi, F. (2017). Differential distractor functioning as a method for explaining DIF: The case of a national admissions test in Saudi Arabia. International Journal of Testing, 0, 1-26.

79- Wainer, H., & Thissen, D. (2001). True score theory: The traditional method. In D. Thissen & H. Wainer (Eds.), Test scoring (pp. 23-72). Mahwah, NJ: Erlbaum.

80- Walker, C. M. (2011). What's the DIF? Why differential item functioning analyses are an important part of instrument development and validation. Journal of Psychoeducational Assessment, 29(4), 364-376.

81- Wang, N., & Lane, S. (1996). Detection of gender-related differential item functioning in a mathematics performance assessment. Applied Measurement in Education, 9(2), 175-199.

82- William Gasparini, Gilles Vieille Marchiset. (2008). Le Sport Dans les Quartiers: Pratiques Sociales et Politiques Publiques. France: Presses Universitaires de France.

83- Zenisky, A. L., Hambleton, R. K., & Robin, F. (2003). Detection of differential item functioning in large scale state assessments: A study evaluating a two-stage approach. Educational and Psychological Measurement, 63(1), 51-64.

84- Zhang, Y., Matthews-Lopez, J., & Dorans, N. J. (2003). Using DIF dissection to assess effects of item deletion due to DIF on the performance of SAT I: Reasoning test sub-populations (Report). U.S.; New Jersey.

85- Zhen, L. & Zumbo, B. (2009). Impact of Differential Item Functioning on Subsequent statistical conclusions Based on observed Test Score Data. Psicologica, 30(2),343-370.

86- Zumbo, B. D. (1999). A Handbook on the theory and methods of differential item functioning (DIF): Logistic Regression modeling as a unitary framework for binary and likert-Type (ordinal) item scores. Ottawa, Directorate of Human Resources Research and Evaluation, Department of National Defense.

87- Zumbo, B. (2000). The Effect of DIF and Impact on Classical Test Statistics: Undetected DIF and Impact, and the Reliability and Interpretability of Scores from a language Proficiency Test. Presented at the Annual Conference of the National Counic on Measurement in Education (NCME), April, New Orleans, LA.S.

88- Zumbo, B. (2003). Does item-Level Dif manifest itself in scale- Level analyses? Implications for translating languge tests. Language Testing,20(2),136-147.