المملكة العربية السعودية
وزارة التعليم
جامعة القصيم
كلية اللغات والعلوم الإنسانية
قسم علم النفس

تأثير نسبة المفردات ذات الأداء التفاضلي على الخصائص السيكومترية للاختبار وبنوده باستخدام النموذج اللوجستي ثنائي المعلمة

The Effect of Percentage of item with Differential Functioning on Psychometric properties for test and items According to the two-parameter logistic model

بحث مقدم لاستكمال متطلبات الحصول على درجة الماجستير في علم النفس التربوي مسار (القياس والتقويم)

إعداد الباحثة / رحاب بنت شعيب حمد الهميلي
الرقم الجامعي: 421200468
إشراف: محمد بن سليمان بن محمد الوطبان
أستاذ علم النفس في جامعة القصيم
1446هـ_2024م

ملخص الرسالة

هدفت الدراسة إلى الكشف عن تأثير نسبة المفردات ذات الأداء التفاضلي على الخصائص السيكومترية للاختبار وبنوده باستخدام النموذج ثنائي المعلمة لنظرية الاستجابة للمفردة. ولتحقيق هدف الدراسة تم استخدام المنهج التجريبي. وتمّ توليد البيانات باستخدام برنامج (Wingen v3) لتوليد نماذج اختبار ثنائي الاستجابة. وتكون الاختبار من 50 مفردة باستخدام الأداء التفاضلي المنتظم، وتمّ توليد استجابات (1000) فرد لكل مجموعة من مجموعات الدراسة. وتمّ تحليل البيانات باستخدام برمجية Bilog-Mg. توصّلت الدراسة إلى عدم وجود فروق دالة احصائيًا (α= 0.05) على دقة تقديرات القدرة ودقة تقدير معلمة التمييز تعزى إلى نسبة المفردات ذات الأداء التفاضلي (10%، 20%، 30%). وعدم وجود فروق دالة احصائياً (α= 0.05) على دقة تقدير معلمة الصعوبة تُعزى إلى نسبة المفردات ذات الأداء التفاضلي (10%، 20%)، ووجود فروق دالة احصائيًا (α= 0.05) على دقة تقدير معلمة الصعوبة تُعزى إلى نسبة المفردات ذات الأداء التفاضلي؛ لصالح المجموعة المستهدفة (30%)، وعدم وجود فروق دالة احصائيًا (α= 0.05) على متوسطات كمية معلومات الاختبار تُعزى إلى نسبة المفردات ذات الأداء التفاضلي (20%، 30%) وفقًا لمجموعات الدراسة. ووجود فروق دالة احصائيًا (α= 0.05) على متوسطات كمية معلومات الاختبار تُعزى إلى نسبة المفردات ذات الأداء التفاضلي؛ لصالح المجموعة المستهدفة (10%). ووجود فروق دالة احصائيًا (α= 0.05) على متوسطات كمية معلومات الاختبار وفق التفاعل بين مجموعات الدراسة وفئات القدرة عند مستويات القدرة المرتفعة؛ لصالح المجموعات المستهدفة بنسبة المفردات ذات الأداء التفاضلي (10%، 20%، 30%).

المقدمة:

ظهرت الاختبارات نتيجة التطورات الحاصلة على مدى قرنين من الزمان في مجالات متنوعة ومن أهمها علم النفس؛ بهدف التمييز بين الأفراد على أساس المهارات والمعارف والصفات والقدرات. لذا كان الدافع وراء التطورات في القياس النفسي هو التقدم المفاهيمي والإحصائي الذي أدى إلى تنظيم وتقدير الفروق الفردية بناءً على الاختبارات النفسية والتربوية (Wells & Faulkner-Bond, 2016).

وقد أسهم علم القياس النفسي والتربوي بتطورات أدت إلى إعادة النظر في النظريات والنماذج السيكومترية من خلال تطوير نماذج ونظريات معاصرة من أجل التغلب على كثير من المشكلات المتعلقة ببناء الاختبارات والمقاييس في الجوانب المختلفة للسلوك الإنساني. ومن أهم هذه التطورات المعاصرة نظرية الاستجابة للمفردة الاختبارية Item response theory (IRT) التي تجعل الباحث والممارس أكثر استبصارًا في بناءه وانتقاءه لأدوات القياس (علام، **2000**).

وبرزت نظرية الاستجابة للمفردة في قدرتها على تقدير معالم كل مفردة من مفردات الاختبار بشكل مستقل عن خصائص عينة المتقدمين للاختبار، إضافة إلى أنها تستخدم معادلات رياضية يُطلق عليها النماذج اللوجستية (Ogunsakini, Shogbesan,2018). وتهدف نظرية الاستجابة للمفردة الوصول إلى قيم تقديرية عددية، أحدها تتعلق بالفرد وتُسمَّى معلمة القدرة أو السمة المقاسة لدى الفرد (Ability Parameter)، وقيمة تقديرية عددية واحدة أو أكثر تتعلق بالمفردة الاختبارية وتُسمَّى بمعلمة أو معالم المفردة (Item Parameters)، وهذه النماذج احتمالية تصف التفاعل بين قدرة الفرد الممتحن (θ) مع مفردات الاختبار، بحيث تربط بين مَعْلَمَيْن أحدهما يتعلق بمعلمة القدرة للفرد، والآخر بخصائص المفردة الاختبارية (الناغي، **2011**).

وهناك ثلاثة نماذج لنظرية الاستجابة للمفردة بالنسبة للمفردات ثنائية الاستجابة، وتختلف هذه النماذج في عدد المعالم، وهي: معلمة الصعوبة (b)، ومعلمة التمييز (a)، ومعلمة التخمين (c) (Świst, 2015). ويُطلق على النموذج الأول، النموذج اللوجستي أحادي المعلمة The One-Parameter Logistic Model (1PL)، ويُسمَّى أحيانًا نموذج راش (Rasch Model)، والمعلمة الوحيدة التي يتم تقديرها باستخدام هذا النموذج هي معلمة الصعوبة، حيث تصف موقع المفردة على طول مقياس القدرة؛ بمعنى أن الممتحنين من ذوي القدرة المنخفضة يجيبون على المفردات السهلة، في حين أن الممتحنين من ذوي القدرة العالية يجيبون على المفردات الصعبة. أمّا النموذج الثاني فيُطلق عليه النموذج اللوجستي ثنائي المعلمة The Two-Parameter Logistic Model (1PL)، وهذا النموذج المكوّن من معلمتين له نفس الوظيفة التي تم تقديمها في النموذج الأول، إضافة إلى معلمة التمييز التي تميّز بين الممتحنين من ذوي المستويات المنخفضة والمرتفعة، وتعكس هذه الخاصية بشكل أساسي انحدار منحنى خصائص المفردة، فكلّما كان المنحنى أكثر حدة كلما كانت المفردة ذات تمييز أعلى (Baker & kim, 2017). أما النموذج الثالث فيُطلق عليه النموذج اللوجستي ثلاثي المعلمة The Three-Parameter Logistic Model (3PL)، ويرتكز هذا النموذج على ثلاث معالم هي: الصعوبة، والتمييز، والتخمين. وتشير معلمة التخمين إلى احتمالية أن يقوم الممتحن من ذوي القدرة المنخفضة بالإجابة على مفردات الاختبار بشكل صحيح (Teo, 2013).

واتجه اهتمام الباحثين ومطوّري الاختبارات في نظرية الاستجابة للمفردة الاختبارية على الخصائص السيكومترية للاختبار التي تشتمل على مفهوم الثبات، ووظّفت مفهوم الثبات الكلاسيكي بطريقتين مختلفتين، تبعًا للكيفية التي تقدر بها من خلال الثبات النظري والأمبريقي، فالثبات النظري(Theoretical reliability) يعتمد في تقديره

على معالم مفردات الاختبار، ولا يعتمد على تقديرات القدرة للأفراد، وبدلًا من ذلك يفترض أن قيم القدرة الحقيقية تتبع التوزيع الطبيعي، بوسط صفر وانحراف معياري واحد، والثبات التجريبي (Empirical reliability) هو تباين الدرجة الحقيقية مقسومًا على مجموع تباين الدرجة الحقيقية وتباين الخطأ (Du Toit, 2003).

كما اهتمت نظرية الاستجابة للمفردة، بمفهوم دالة المعلومات للاختبار وهو مجموع دوال معلومات المفردات في الاختبار، ودوال معلومات المفردات هي دوال رياضية لحساب مقدار المعلومات التي يحتوي عليها منحنى خصائص كل مفردة، بحيث يكون مقدار المعلومات مرتفعًا عندما يكون مستوى الصعوبة أقرب إلى مستوى القدرة، أو يكون مستوى التمييز عاليًا. كما يُستفاد منها في معرفة مدى جودة الاختبار في التمييز بين الأفراد على مستويات مختلفة من القدرة على السمة المقاسة، بالإضافة إلى تقدير معامل الثبات للاختبار، وأيضًا الخطأ المعياري في التقدير، وهو الجذر التربيعي لمقلوب دالة المعلومات (Secolsky, Denison, 2012).

وتعتبر نظرية الاستجابة للمفردة الاختبارية (IRT) إطارًا شائعًا وقيّمًا لنمذجة وتحليل البيانات في الاختبارات النفسية والتربوية، حيث تلعب تقنيات محاكاة الكمبيوتر أدوارًا رئيسية في الممارسات البحثية، وتتمثل إحدى الميزات المكتسبة من استخدام البيانات المولّدة في نظرية الاستجابة للمفردة الاختبارية في أنها تنشئ قيم معالم حقيقية ومعروفة، مما يتيح التقييم المباشر لتقديرات المعالم في ظل الظروف المختلفة للقياس. وغالبًا ما يكون هذا أمرًا بالغ الأهمية للتطبيقات بما في ذلك تقييم المقدرين، وتحليل ملاءمة النموذج والأداء التفاضلي للمفردات، وحساب الخطأ المعياري للقياس. ومن الناحية النسبيّة، يُعدُّ استخدام البيانات المولّدة أكثر كفاءة من استخدام البيانات التجريبية من حيث

الوقت والتكلفة، مما يجعلها أداة قيّمة للباحثين والممارسين؛ لتقييم برامج الاختبارات والتحقق من صحته (van der Linden, 2018).

وبالنظر إلى نتائج الاختبارات وأهميتها فإنّ التحقق من صحة الاختبار من خلال طرق مختلفة هو عملية ضرورية للتقييم العادل للممتحنين، وكذلك للتفسير الصحيح للاختبار (Pae & Park,2006). ومن المواضيع التي لاقت اهتمامًا دائمًا في البحوث النفسية والتربوية منذ أربعين عامًا، هو مصطلح التحيز الذي استبدل بعد ذلك بمفهوم الأداء التفاضلي، وتمّ اعتماد مصطلح الأداء التفاضلي للمفردة (Differential item functioning)، الذي قدّمه هولاند وثاير في عام **1988**، باعتباره تمثيلًا أكثر دقة لهدف الإجراءات الإحصائية من قبل المجتمع السيكومتري؛ لأنه كان قابلًا لتعريف مفاهيمي وتشغيلي واضح (Wells & Faulkner-Bond, 2016). ومصطلح التحيز أعم وأشمل من الأداء التفاضلي، ونظرًا لأن التقصي عن التحيز يُدّعم بالأدلة التجريبية المتعلقة بالأداء النسبي بين المجموعة المرجعية والمستهدفة على مفردات الاختبار، إلا أنّ مصطلح الأداء التفاضلي يُستخدم بشكل شائع بدلًا من التحيز لوصف الأدلة التجريبية التي تم الحصول عليها في تحقيقات التحيز، لذا فإنّ الأدلة التجريبية على الأداء التفاضلي ضرورية، ولكنها ليست كافية للحكم بوجود التحيز (Ellis, Raju, 2003).

وذُكر في دليل معايير الاختبارات النفسية والتربوية، الذي تقدّمه الجمعية الأمريكية للبحوث التربوية (AERA)، والجمعية الأمريكية لعلم النفس (APA) ، والمجلس الوطني للقياس في التعليم (NCME)، تعريف الأداء التفاضلي للمفردة، على أنه مؤشر إحصائي لمدى وجود مجموعات مختلفة من المتقدمين للاختبار، الذين هم من نفس مستوى القدرة، بأن لديهم احتمالات مختلفة للإجابة الصحيحة (American Educational Research Association et all, 2014). وفي نظرية الاستجابة للمفردة التي تم تأسيسها كإطار قياس قابل للتطبيق خلال السبعينيات،

فقد تمّ تعريف الأداء التفاضلي بشكل أكثر دقة في المصطلحات السيكومترية، على أنه اختلاف بين المجموعات في منحنيات خصائص المفردة (Lord,1977 كما ورد في Wells & Faulkner-Bond, 2016).
وأشار Ackerman (1992)، إلى أن هناك نوعين من الأداء التفاضلي هما: الأداء التفاضلي المنتظم Uniform-DIF، ويظهر عندما تختلف منحنيات خصائص المفردة للمجموعتين المرجعية والمستهدفة بحيث تكون متوازية. أمّا الأداء التفاضلي غير المنتظم Nonuniform-DIF، فيظهر عندما تكون منحنيات خصائص المفردة غير متوازية. ويمكن القول: بأن الأداء التفاضلي مرّ بثلاث مراحل من الأبحاث كما جاء في Zumbo (2007)، حيث تمركز الجيل الأول على تطوير المفاهيم، كما ركّز الجيل الثاني على الأساليب الإحصائية لتحديد الأداء التفاضلي للمفردة، في حين أنّ الجيل الثالث والحالي يهتم بتفسيرات الأداء التفاضلي للمفردة.
والجدير بالذكر أن العديد من الدراسات السابقة العربية والأجنبية اهتمت بدراسة تأثير وجود النسب المختلفة للأداء التفاضلي في الاختبارات وأثرها على بعض المتغيرات، واختلفت هذه النسب من دراسة إلى أخرى، إلا أن غالبية الدراسات أجمعت على دراسة تأثير ثلاث نسب (صغيرة – متوسطة – كبيرة). والنسب الصغيرة تراوحت بين **(5%- 10%)**، والنسب المتوسطة بين **(15%- 20%)**، والنسب الكبيرة بين **(30%- 40%)**، وظهر تأثير هذه النسب على المتغيرات المختلفة التي تم دراستها وهذا ما أدى بالباحثة إلى النظر في تأثير هذه النسب ذات الأداء التفاضلي على الخصائص السيكومترية للاختبار ومفرداته التي تعتبر من المعايير المهمة في بناء الاختبارات والمذكورة في دليل معايير الاختبارات النفسية والتربوية، حيث إنّ الخلل في تقدير الخصائص السيكومترية يؤدي إلى ضعف في تطبيق الاختبارات وبالتالي عدم دقة نتائج الاختبارات. بالإضافة إلى ندرة الدراسات العربية التي درست مفهوم ثبات الاختبار من منظور نظرية

الاستجابة للمفردة الاختبارية، فأغلب الدراسات التي درست تأثير الأداء التفاضلي اقتصرت على دراسة الثبات التجريبي وأهملت دراسة الثبات النظري. ففي حدود علم الباحثة لا توجد دراسة هدفت إلى بيان أثر نسبة المفردات ذات الأداء التفاضلي على الخصائص السيكومترية للاختبار ومفرداته وفق النموذج اللوجستي ثنائي المعلمة.

مشكلة الدراسة:

تُعدّ الاختبارات النفسية والتربوية من أهم الإسهامات في مجال العلوم المعرفية والسلوكية، حيث تتمتع الاختبارات المعدّة جيدًا بالتفسير الدقيق لنتائج الاختبارات، كما تُعتبر مصدرًا أساسيًّا ومهمًّا يُستند عليه في اتخاذ قرارات أفضل بشأن الأفراد والجماعات. ومن ناحية أخرى قد يسبب الاستخدام غير السليم للاختبارات ضررًا للمفحوصين، والأطراف الأخرى المتأثرة بالقرارات المستندة على الاختبارات. كما تستند الاختبارات في بناءها على عدة معايير؛ لتعزيز الممارسة الجيدة، وتوفير أساس لتقييم جودة تلك الممارسات، ومن أهمها الخصائص السيكومترية للاختبار ومفرداته التي تُعدُّ أساسًا لا يمكن الإخلال به في بناء الاختبارات (American Educational Research Association et all, 2014). كما تندرج هذه الخصائص تحت نظريات القياس وهي، النظرية الكلاسيكية في الاختبارات، ونظرية الاستجابة للمفردة، إلّا أن هناك اختلاف في تقدير الخصائص في كل من النظريتين، فالخصائص السيكومترية للاختبار وفقًا لنظرية الاستجابة للمفردة تشتمل على كلٍّ من الثبات النظري والتجريبي، وكذلك دالة المعلومات للمفردة والاختبار(Du Toit, 2003). كما تختلف خصائص المفردات باختلاف النموذج المستخدم لنظرية الاستجابة للمفردة. ويعتمد النموذج ثنائي المعلمة على تقدير معلمتي الصعوبة والتمييز(محاسنة، 2013)، غير أن هذه الخصائص السيكومترية قد تتأثر ببعض المتغيرات التي بدورها تؤثر على نتائج الاختبار ككل

مثل وجود الأداء التفاضلي للمفردة الذي يعتبر تهديدًا لصلاحية الاختبار، مما يجعل الأداء على بنود الاختبار لا تعكس الفائدة المرجوة (Finch & French, 2019).ويظهر الأداء التفاضلي للمفردة عندما تختلف الخصائص الإحصائية للمفردة وفقًا لمجموعات الدراسة بعد مطابقة السمة بوجود مجموعتان هما: المجموعة المستهدفة التي يظهر فيها الأداء التفاضلي، والمجموعة المرجعية التي تعتبر مجموعة الأقلية (Belzak & Bauer, 2020). ويعدُّ النموذج ثنائي المعلمة أفضل من النموذج ثلاثي المعلمة في دراسة تأثير نسب المفردات ذات الأداء التفاضلي؛ نظرًا لوجود بعض الاستقصاءات حول الأداء التفاضلي للمفردة فيما يتعلق بمعلمة التخمين، كما يجب أن يكون مختصو القياس حساسين لوجود معلمة التخمين لتأثيرها المحتمل على تقدير معالم النموذج، كما أنّ تجاهلها قد يؤدي إلى ضعف تقدير صعوبة المفردة ومعلمة التمييز، والنتائج الخاطئة لنوع الأداء التفاضلي المنتظم وغير المنتظم(Finch& French, 2019). كما أوضحت دراسة ضعضع **(2023)**، فعالية النموذج ثنائي المعلمة من بين النموذجين الأحادي والثلاثي المعلمة في إظهار النسبة الأعلى للمفردات ذات الأداء التفاضلي. بالإضافة إلى أن العديد من الدراسات السابقة التي درست تأثير نسب المفردات ذات الأداء التفاضلي باختلاف المتغير التابع كانت وفقًا للنموذج ثلاثي المعلمة، كما في دراسة النوافلة، حيث تناولت تأثير نسب المفردات ذات الأداء التفاضلي **(0%، 10%، 20%، 40%)** على معالم المفردات: الصعوبة، والتمييز، والتخمين وفق النموذج ثلاثي المعلمة، وتقديرات قدرات الأفراد، وكذلك الثبات التجريبي، ودالة المعلومات للاختبار. وترى الباحثة ضرورة استكمال ما قام به النوافلة مع اختلاف في نسب المفردات ذات الأداء التفاضلي **(0%، 10%، 20%، 30%)**، ودراسة الأثر على دقة تقدير قدرات الأفراد، ومعلمتي

الصعوبة والتمييز، والثبات النظري والتجريبي، وكذلك دالة المعلومات للاختبار باستخدام النموذج ثنائي المعلمة.

ويمكن تحديد مشكلة الدراسة في السؤال الرئيس التالي: ما تأثير نسبة المفردات ذات الأداء التفاضلي **(0%، 10%، 20%، 30%)** على الخصائص السيكومترية للاختبار ومفرداته وفق النموذج ثنائي المعلمة؟

ويتفرع منه الأسئلة التالية:

السؤال الأول: هل يوجد تأثير لنسب المفردات ذات الأداء التفاضلي **(0%، 10%، 20%، 30%)** على دقة تقديرات قدرات الأفراد باستخدام النموذج ثنائي المعلمة.

السؤال الثاني: هل يوجد تأثير لنسب المفردات ذات الأداء التفاضلي **(0%، 10%، 20%، 30%)** على دقة تقدير معلمتي الصعوبة والتمييز باستخدام النموذج ثنائي المعلمة.

السؤال الثالث: هل تختلف معاملات الثبات النظري والثبات التجريبي تبعًا لنسبة الفقرات ذات الأداء التفاضلي**(0%، 10%، 20%، 30%)** باستخدام النموذج ثنائي المعلمة.

السؤال الرابع: هل تختلف قيم دالة المعلومات للاختبار باستخدام النموذج ثنائي المعلمة تبعًا لنسبة الفقرات ذات الأداء التفاضلي **(0%، 10%، 20%، 30%).**

أهداف الدراسة:

- التحقق من مدى تأثير نسب المفردات ذات الأداء التفاضلي **(0%، 10%، 20%، 30%)** على دقة تقدير قدرات الأفراد باستخدام النموذج ثنائي المعلمة.
- التحقق من مدى تأثير نسب المفردات ذات الأداء التفاضلي **(0%، 10%، 20%، 30%)** على دقة تقديرات معالم المفردات للاختبار (الصعوبة والتمييز) باستخدام النموذج ثنائي المعلمة.

- التحقق من مدى تأثير نسب المفردات ذات الأداء التفاضلي **(0%، 10%، 20%، 30%)** على قيم معاملات الثبات النظري والثبات التجريبي باستخدام النموذج ثنائي المعلمة.
- التعرف على مدى تأثير نسب المفردات ذات الأداء التفاضلي **(0%، 10%، 20%، 30%)** على قيم دالة المعلومات للاختبار باستخدام النموذج ثنائي المعلمة.
- **أهمية الدراسة:**

الأهمية النظرية:

- التطرق إلى مفهوم الثبات في الاختبار الذي تمَّ توظيفه من خلال نظرية الاستجابة للمفردة، بالثبات النظري والثبات التجريبي.
- إلقاء الضوء على طريقة حديثة في مجال التحقق من الخصائص السيكومترية وفق النموذج ثنائي المعلمة.

الأهمية التطبيقية:

- مساعدة الباحثين ومطوّري الاختبارات في معرفة تأثير النسب المختلفة للمفردات ذات الأداء التفاضلي على معالم الفقرات، المتمثلة في الدراسة الحالية بمعلمتي الصعوبة والتمييز.
- توجيه اهتمام مُعدّي الاختبارات إلى ضرورة الانتباه لمدى تأثير المفردات ذات الأداء التفاضلي على دقة نتائج القياس وفقًا لدالة المعلومات في الاختبار، ومعاملات الثبات النظري والتجريبي للاختبار.
- لفت انتباه الباحثين عند اشتمال اختباراتهم على مفردات ذات أداء تفاضلي على اختلاف نسبتها، في إمكانية إبقاء الفقرات، أو استبعادها للحصول على نتائج أفضل.
- تحسين وتطوير الاختبارات من خلال الكشف عن تأثير مقدار نسب المفردات ذات الأداء التفاضلي في الاختبار.

التعريفات الاصطلاحية والإجرائية:

- **نسبة المفردات ذات الأداء التفاضلي Percentage of DIF :** أي عدد الأسئلة أو المفردات التي تُبدي دالة تفاضلية مقسوم على عدد بنود الاختبار (أبو علام والمرابحة،**2015**).

وفي الدراسة الحالية، هي عدد المفردات ذات الأداء التفاضلي التي تشكل النسب التالية **(10%-20%-30%).**

- **الخصائص السيكومترية للمفردات Items psychometric properties:** تتمثل في تقدير معلمة الصعوبة، ومعلمة التمييز، بالاعتماد على النموذج اللوجستي ثنائي المعلمة وهي كالتالي:

1. **معلمة الصعوبة Threshold Parameter**: هي النقطة التي تقع على متصل القدرة، وذلك عندما يكون احتمال الإجابة على المفردة إجابة صحيحة مساويًا للقيمة **(0.5)**، ويمكن أن تقع ضمن المدى النظري (-∞ إلى ∞)، وضمن المدى العملي **(-3 إلى3)**، وبالنظر إلى منحنى خصائص الفقرة الذي يأخذ شكل حرف s، فإنّ ميله يزداد بزيادة مستوى القدرة (محاسنة،**2013**).
2. **معلمة التمييز Slop Parameter** :خاصية تصف المفردة، وتأخذ قيمًا نظرية تتراوح من (-∞ إلى ∞)، وعمليًّا من **(-3 إلى 3)**، ويستدل عليها من خلال شدة انحدار أو ميل المماس لمنحنى خصائص الفقرة **(ICC)** عند نقاط انعطاف منحنى الخصائص للفقرة (الشريفين،**2018**).

- **الخصائص السيكومترية للاختبار Test psychometric properties:** تتمثل في تقدير كمية المعلومات التي يقدمها الاختبار من خلال دالة المعلومات للاختبار، والثبات النظري والثبات التجريبي بالاعتماد على النموذج اللوجستي ثنائي المعلمة.

1. **دالة المعلومات للاختبار Test information functions:** هي مجموع دوال معلومات المفردات في الاختبار، ودوال معلومات

المفردات عبارة عن دوال رياضية لحساب مقدار المعلومات التي يحتوي عليها منحنى خصائص كل مفردة Secolsky, (Denison,2012).

2. **الثبات النظري Theoretical reliability** : يعتمد في تقديره على معالم مفردات الاختبار، ولا يشتمل على تقديرات القدرة للأفراد، وبدلًا من ذلك يفترض أن قيم القدرة الحقيقية تتبع التوزيع الطبيعي بوسط صفر وانحراف معياري واحد (Du Toit, 2003).
3. **الثبات التجريبي Empirical reliability:** هو تباين الدرجة الحقيقية مقسومًا على مجموع تباين الدرجة الحقيقية وتباين الخطأ (Du Toit, 2003).

- **النموذج اللوجستي ثنائي المعلمة Two-parameter Logistic Model:** يعتمد هذا النموذج على الفرق بين القدرة التي يمتلكها الطالب s في السمة التي يراد تقديرها (القدرة الكامنة وراء إجابات الطلبة)، ودرجة صعوبة البند i التي يرغب الطالب الإجابة عنها التي يمثلها الرمز β مضروبًا في درجة التمييز التي يتصف فيها البند i التي يرمز لها بالرمز αi (محاسنة، **2013**).

حدود الدراسة:

- اقتصرت الدراسة الحالية على أربع نسب للمفردات التي تبدي أداءًا تفاضليًّا، وهي: **(0%، 10%، 20%، 30%)**.
- اعتمدت الدراسة على استخدام بيانات مولّدة باستخدام برنامج Wingen 3، ومعالجتها من خلال برمجية LDID، BILOG-MG، IRTPRO، SPSS، EXCEL.
- اقتصرت الدراسة على نموذج واحد من نماذج نظرية الاستجابة للمفردة، وهو النموذج ثنائي المعلمة.
- اقتصرت الدراسة على الأداء التفاضلي المنتظم.

فرضيات البحث:

- لا توجد فروق ذات دلالة إحصائية عند مستوى الدلالة **(0.05)** بين دقة تقديرات قدرات الأفراد تُعزى إلى نسبة المفردات ذات الأداء التفاضلي **(0%، 10%، 20%، 30%).**
- لا توجد فروق ذات دلالة إحصائية عند مستوى الدلالة **(0.05)** بين دقة تقديرات معالم الصعوبة والتمييز لمفردات الاختبار تُعزى إلى نسبة المفردات ذات الأداء التفاضلي **(0%، 10%، 20%، 30%).**
- لا توجد فروق ذات دلالة إحصائية عند مستوى الدلالة **(0.05)** بين قيم معاملات الثبات النظري والثبات التجريبي تُعزى إلى نسبة المفردات ذات الأداء التفاضلي **(0%، 10%، 20%، 30%).**
- لا توجد فروق ذات دلالة إحصائية عند مستوى الدلالة **(0.05)** بين قيم دالة معلومات الاختبار تُعزى إلى نسبة المفردات ذات الأداء التفاضلي **(0%، 10%، 20%، 30%).**

منهج الدراسة وإجراءاتها

الطريقة والإجراءات:

يتناول هذا الفصل منهجية الدراسة وكيفية توليد البيانات، والتحقق من افتراضات نظرية الاستجابة للمفردة، وكذلك الكشف عن المفردات ذات الأداء التفاضلي، والإجراءات التي تمّ اتباعها للإجابة عن أسئلة الدراسة.

منهجية الدراسة:

تم استخدام المنهج التجريبي للكشف عن تأثير نسب المفردات ذات الأداء التفاضلي **(0%، 10%، 20%، 30%)**، على كل من : دقة تقديرات قدرات الأفراد، ومعالم المفردات، والثبات النظري والتجريبي، ودالة المعلومات للاختبار وفق النموذج ثنائي المعلمة.

توليد البيانات:

تمّ توليد بيانات الدراسة باستخدام برنامج (Wingen 3) المصمم والمنتج من قبل هان (Han, 2007)، حيث يتمّ فيه توليد استجابات الأفراد على عدد محدد من بنود الاختبارات أُحادية البُعد أو متعددة الأبعاد، سواء كانت ثنائية التدريج أم متعددة التدريج، كما يمكن من خلال البرنامج توليد استجابات عدة مجموعات من الأفراد بأنواع مختلفة من التوزيعات، حيث يستخدم ثلاثة توزيعات لتوليد قدرة الأفراد، ومعالم الصعوبة والتمييز والتخمين وهي: التوزيع الطبيعي Normal (Distribution)، والتوزيع المنتظم (Uniform Distribution)، وتوزيع بيتا Beta) (Distribution، وتختلف معلمة التمييز في توزيع بيتا وتستخدم بدلًا من ذلك توزيع اللوغاريتم الطبيعي (Lognormal Distribution).

ويتم استخدام البيانات المولّدة بدلًا من البيانات الحقيقية كحلول محتملة لمشاكل القياس التطبيقية، وذلك نظرًا لأنها أقل تكلفة وأسهل وأسرع مقارنة بجمعها من أفراد حقيقيين، كما أنها توفر قيمًا حقيقية للمعالم؛ تساعد الباحث في مقارنتها مع القيم المقدرة لاختبار دقة التقدير، كما تسمح دراسات المحاكاة بتأكيد النتائج النظرية في الممارسة العملية. لذا فإن إظهار فعالية الإجراء قبل فترة من تطبيق النتائج المقاربة له أهمية عملية كبيرة (Davey et all., 1997).

وقد تم الاعتماد على الظروف التالية في توليد بيانات الدراسة:

1. الاعتماد على توليد (50) مفردة لكل مجموعة من مجموعات الدراسة.
2. الاعتماد على توليد نسب المفردات ذات الأداء التفاضلي (10%-20% - 30%) لكل مجموعة من المجموعات المستهدفة.
3. التركيز على المفردات ثنائية التدريج (0-1)، حيث إنّ الدرجة (0) تُعطى للاستجابة الخاطئة، والدرجة (1) للاستجابة الصحيحة.
4. تم الاعتماد على النموذج البارامتري ثنائي المعلمة.

وتم توليد البيانات وفق خطوتين رئيسيتين:

الخطوة الأولى: تمّ توليد بيانات المجموعة المرجعية من خلال توليد بيانات (1000) مفحوص، بتوزيع طبيعي (Normal) للقدرة بمتوسط **(0)**، وانحراف معياري **(1)**. وتمّ توليد معلمة الصعوبة باستخدام التوزيع الطبيعي (Normal)، بمتوسط **(0)**، وانحراف معياري **(1)**، كما تمّ توليد معلمة التمييز باستخدام التوزيع المنتظم (Uniform)، وكانت أدنى قيمة **(0.6)**، وأعلى قيمة هي **(1.6)**، ويوضح جدول **(1)** قيم معالم المفردات ثنائية التدريج من حيث الصعوبة والتمييز لدى المجموعة المرجعية تبعًا للنموذج البارامتري ثنائي المعلمة مرتبة ترتيبا تصاعديًّا من الأدنى إلى الأعلى وفقاً لمعلمة الصعوبة.

جدول 1 : معالم البنود (الصعوبة والتمييز) مرتبة ترتيبًا تصاعديًّا وفقًا لمعلمة الصعوبة

معالم المفردات			معالم المفردات		
المفردة	**التمييز**	**الصعوبة**	**المفردة**	**التمييز**	**الصعوبة**
1	1.569	-2.113	**26**	1.28	0.179
2	1.577	-2.006	**27**	0.826	0.283
3	0.948	-1.615	**28**	0.661	0.302
4	1.15	-1.388	**29**	1.53	0.367
5	0.709	-1.351	**30**	1.507	0.399
6	0.848	-1.333	**31**	1.465	0.401
7	0.817	-1.205	**32**	0.874	0.409
8	1.401	-1.197	**33**	1.137	0.418
9	1.347	-1.116	**34**	1.512	0.421
10	1.378	-1.108	**35**	0.969	0.421
11	0.886	-0.874	**36**	1.361	0.476
12	0.832	-0.796	**37**	1.144	0.534
13	1.578	-0.781	**38**	1.344	0.542
14	1.483	-0.656	**39**	1.073	0.628
15	0.706	-0.634	**40**	1.522	0.708
16	1.438	-0.5	**41**	1.561	0.843
17	1.019	-0.488	**42**	1.334	0.885
18	0.669	-0.486	**43**	0.901	0.999
19	1.093	-0.381	**44**	0.761	1.035

1.051	1.174	**45**	-0.157	1.463	**20**
1.231	1.121	**46**	-0.054	1.046	**21**
1.488	1.235	**47**	-0.009	1.487	**22**
1.801	1.442	**48**	0.056	0.973	**23**
1.926	1.214	**49**	0.12	1.501	**24**
2.465	0.906	**50**	0.134	1.481	**25**

الخطوة الثانية:

تمّ توليد بيانات المجموعات المستهدفة من خلال ترتيب معاملات الصعوبة تصاعديًّا من الأدنى إلى الأعلى في المجموعة المرجعية، واختيار المفردات لكل مجموعة من المجموعات المستهدفة لكي تبدي أداءً تفاضليًّا من خلال اختيار مفردات متباعدة عن بعضها بشكل منتظم. ومن ثمّ توليد المفردات ذات الأداء التفاضلي بإضافة مقدار ثابت **(1.5)** لمعلمة الصعوبة للمفردات ذات الأداء التفاضلي المنتظم للمجموعات المستهدفة. حيث تم اختيار خمس مفردات للمجموعة الأولى بنسبة **10%** من أصل **50** مفردة، وتمّ اختيار عشر مفردات للمجموعة الثانية بنسبة **20%** ، وكذلك اختيار خمس عشرة مفردة للمجموعة الثالثة بنسبة **30%**، ويظهر الجدول **(2)** المفردات التي تم اختيارها.

جدول 2: أرقام المفردات ذات الأداء التفاضلي وفقاً لنسبتها في الاختبار (10%-20%-30%) تبعًا للنموذج ثنائي المعلمة

النموذج الأول الذي يشتمل على مفردات ذات أداء تفاضلي بنسبة 10%	النموذج الثاني الذي يشتمل على مفردات ذات أداء تفاضلي بنسبة 20%	النموذج الثالث الذي يشتمل على مفردات ذات أداء تفاضلي بنسبة 30%
37-36-29-13-1	38-37-36-29-22-20-19-13-10-1	43-38-37-36-29-22-20-19-16-13-10-6-4-3-1

التحقق من افتراضات نظرية الاستجابة للمفردة:

أولًا: التحقق من افتراض أُحادية البُعد

تمّ التحقق من افتراض أُحادية البُعد، وذلك باستخدام التحليل العاملي الاستكشافي (Exploratory factor analysis)، باستخدام تحليل المكونات الأساسية (Principle component analysis) لاستجابات المفحوصين على مفردات الاختبار، وتمّ التحقق أولًا من مدى ملاءمة بيانات عينة الدراسة لإجراءات التحليل العاملي الاستكشافي باستخدام اختبار بارتلت (Bartlett's Test)، ومؤشر كايزر ماير أولكن [Keiser – Meyer Olken (KMO)]، وذلك كما هو مبين في جدول **(3)**.

جدول3 : اختبار بارتلت ومؤشر كايزر- ماير

	المجموعة المرجعية		المجموعة المستهدفة بنسبة 10%		المجموعة المستهدفة بنسبة 20%		المجموعة المستهدفة بنسبة 30%	
الاختبار المستخدم	**القيمة**	**الدلالة**	**القيمة**	**الدلالة**	**القيمة**	**الدلالة**	**القيمة**	**الدلالة**
Bartlett's Test	8904.689	.000	8738.093	.000	7962.482	.000	8161.565	.000
Keiser – Meyer Olken (KMO)	0.957		.956		. .950		.952	

يُلاحظ من الجدول السابق بأنّ اختبار بارتلت (Bartlett)، دال إحصائيًّا، كما أنّ مؤشر (KMO) بلغت قيمته الأقل (0.950) في المجموعة المستهدفة بنسبة 20%، وظهرت قيمته الأعلى **(0.957)** في المجموعة المرجعية، والحد الأدنى المقبول لملاءمة البيانات هو **(0.5)**، مما يدل ذلك على ملاءمة البيانات في جميع المجموعات لإجراءات التحليل العاملي الاستكشافي.

وبعد التحقق من ملاءمة البيانات، تمّ إجراء التحليل العاملي الاستكشافي بطريقة المكونات الرئيسية، ويوضح الجدول **(4)** المكوّن

الرئيسي الأول والثاني والثالث في كل مجموعة من مجموعات الدراسة.

جدول4 : تحليل المكونات الرئيسية، الأول والثاني والثالث عند المجموعة المرجعية والمجموعات المستهدفة

	المكونات	الجذر الكامن	نسبة التباين المفسر	التباين المفسر التراكمي
المجموعة المرجعية	1	9.883	19.766	**19.766**
	2	1.404	2.807	**22.573**
	3	1.216	2.43	**25.005**
المجموعة المستهدفة بنسبة 10%	1	9.736	19.471	**19.471**
	2	1.424	2.847	**22.318**
	3	1.17	2.35	**24.673**
المجموعة المستهدفة بنسبة 20%	1	9.025	18.050	**18.050**
	2	1.434	2.868	**20.918**
	3	1.21	2.43	**23.35**
المجموعة المستهدفة بنسبة 30%	1	9.195	18.389	**18.389**
	2	1.530	3.060	**21.449**
	3	1.20	2.409	**23.858**

أفرزت النتائج وجود عوامل كامنة تراوحت مابين **(11- 14)** عامل كامن في المجموعة المرجعية والمجموعات المستهدفة لكلّ نسبة من نسب المفردات ذات الأداء التفاضلي **(10%، 20%، 30%)**، وبلغت قيمة الجذر الكامن لكل واحد منها **(1)** وأكثر من الواحد الصحيح، كما أنّ المكوّن الرئيسي الأول في كل مجموعة من مجموعات الدراسة تجاوز جذره ستة أضعاف المكوّن الرئيسي الثاني، وهذا دليل على تحقق افتراض أُحادية البُعد في كل مجموعة من مجموعات الدراسة.

ثانيًا: التحقق من افتراض الاستقلال الموضعي

تمّ التحقق من افتراض الاستقلال الموضعي لأزواج مفردات الاختبار الارتباطية بين كل مفردتين باستخدام برنامج **[Local Dependence Indices for Dichotomous Items, (LDID)]** الذي وضعه كل من كيم وكوهين ولاين. ويوفر البرنامج عدة مؤشرات للتحقق من الاستقلال الموضعي وهي: مؤشر إحصائي مربع كاي بيرسون (Pearson chi-square statistic, X2)، وإحصائي مربع كاي لنسبة الاحتمال (likelihood-ratio chi-square statistic ,G2)، ومؤشر يّن للاعتماد الموضعي (Yen's index of local dependence Q3)، وإحصائي الفرق الارتباطي المحوّل من خلال فيشر (Fisher-transformed correlation difference statistic , ZD)، وتمّ كتابة البرنامج بلغة فورتران، كما يعمل البرنامج بشكل أساسي في وضع تفاعلي، حيث يجيب المستخدم على سلسلة من الأسئلة تخص اختيار نوع البرنامج الذي يعطي ملفات البيانات سواء كان برنامج (BILOG-Mg، MULTILOG، PARSCALE)، بالإضافة إلى عدد المفحوصين، وعدد مفردات الاختبار، وعدد فئات الاستجابة **(2)**، واختيار نوع النموذج المستخدم من نماذج نظرية الاستجابة للمفردة، واختيار أدنى قيمة **(0)**، وأخيرًا يتمّ إدخال فورمات بلغة فورتران فيما يخص ملف البيانات (Kim et all,

(2005. وقد تمّ استخدام مؤشر **(Q3)** للتحقق من الاستقلال الموضعي والمحوّل إلى قيم z الفشرية (ZQ3). وقد تمّ تزويد البرنامج بملف استجابات المفحوصين، حيث تمّ تغيير اسم الملف ليتناسب مع برنامج التشغيل (LDID.DAT)، وبمخرجات برنامج (BILOG-MG) الخاصة بقدرات الأفراد (LDID.SCO)، وبمعالم مفردات النموذج ثنائي المعلمة في نظرية الاستجابة للمفردة (LDID.PAR).

وتمّ تحديد القيمة الحرجة للفصل بين الأزواج المستقلة موضعيًّا، والأزواج المرتبطة موضعيًّا من خلال اختبار (ت) الخاص بمعاملات الارتباط وفقًا للمعادلة التالية $(t = (r \times \sqrt{df}\,)/\sqrt{1-r^2}\,)$ ، حيث إنّ **(df)** هي درجة الحرية التي تساوي عدد المفحوصين مطروحًا منها **2**، عند مستوى دلالة (α=0.05)، و (r) هو معامل الارتباط المحسوب من أزواج المفردات عند **(998)** درجة حرية في ضوء حجم العينة، وبلغت القيمة الحرجة لمؤشر z_{Q3} **(0.06185)** عند جميع المجموعات (Donnelly,2007).

وتمّ حساب أزواج المفردات التي حققت الاستقلال الموضعي والتبعية الموضعية عند جميع المجموعات، وجدول **(5)** يوضح ذلك.

جدول 5: الاستقلال الموضعي والتبعية الموضعية عند المجموعة المرجعية والمجموعات المستهدفة

	حالة التبعية الموضعية	التكرار	النسبة المئوية	الباقي المعياري
المجموعة المرجعية	ارتباط موضعي	133	10.86	-19.42
	استقلال موضعي	1092	89.14	19.42
	الكل	1225	100.00	
المجموعة المستهدفة بنسبة 10%	ارتباط موضعي	124	10.12	-19.66
	استقلال موضعي	1101	89.88	19.66
	الكل	1225	100.00	
المجموعة المستهدفة بنسبة 20%	ارتباط موضعي	139	11.35	-19.54
	استقلال موضعي	1086	88.65	19.54
	الكل	1225	100.00	

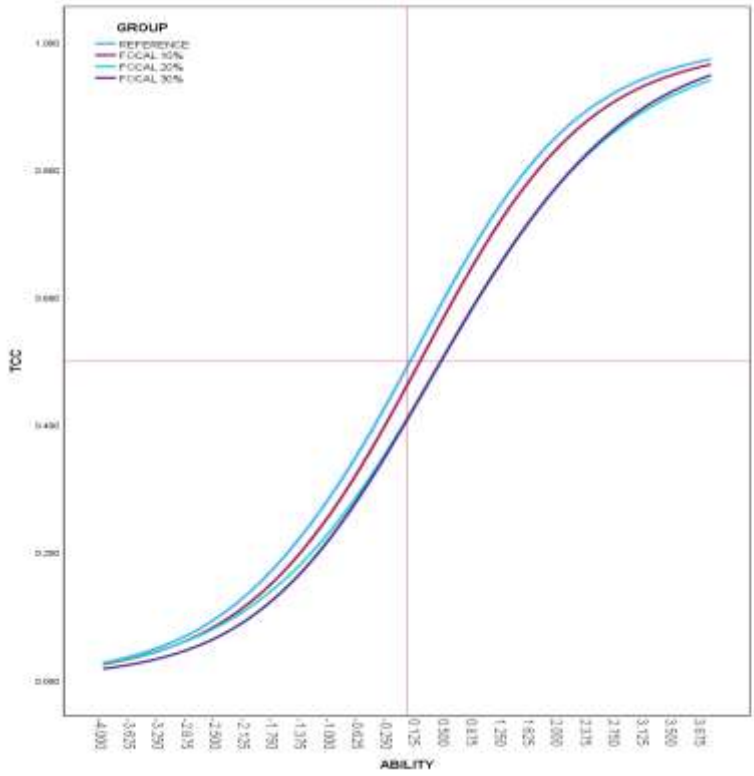
المجموعة المستهدفة بنسبة 30%	ارتباط موضعي	125	10.20	-19.74
	استقلال موضعي	1100	89.80	19.74
	الكل	1225	100.00	

يُلاحظ من جدول **(5)** أن الباقي المعياري في حالة التبعية الموضعية أصغر من **-2**، وفي حالة الاستقلال الموضعي أكبر من **2**، وهذا دليل على تحقق الاستقلال الموضعي (Hinkle, et all 2003). كما أن أزواج مفردات الاختبار التي حققت الاستقلال الموضعي وفق مؤشر ZQ3 جاءت أكبر من المعيار **(80%)**، وهذا دليل آخر على تحقق افتراض الاستقلال الموضعي.

ثالثًا: التحقق من افتراض منحنى خصائص البند (ICC)

للتحقق من افتراض منحنى خصائص المفردة وفقًا لنسبة البنود ذات الأداء التفاضلي في الاختبار.

فقد تمّ إنشاء رسم بياني لمنحنى خصائص الاختبار كما هو مبيّن في الشكل **(4)**.



شكل 1: منحنى خصائص الاختبار

يتضح من الشكل السابق أن مجموعات الدراسة المرجعية والمستهدفة تُحقق افتراض منحنى خصائص الاختبار؛ إذ لم يُلاحظ وجود تراجع يساوي أو يزيد عن مقدار **(0.03)** في قيمة احتمالية إجابة المفردة إجابة صحيحة في حال الانتقال من مستوى قدرة إلى مستوى قدرة لاحق عند مستويات قدرة منتقاة ابتداءً من مستوى قدرة **(-4)** إلى **(+4)** بزيادة مقدارها **0.125** بحسب اشتراط موكن (Molenaar et all, 2000) ، وجدول **(6)** يوضح ذلك.

جدول 6: رصد مقدار التراجع في احتمالية الإجابة على المفردة إجابة صحيحة

الرقم	نسب الأداء التفاضلي											
	المجموعة المرجعية. صفر%			%10			%20			%30		
	قيمتي الفرق بين مستويي القدرة			قيمتي الفرق بين مستويي القدرة			قيمتي الفرق بين مستويي القدرة			قيمتي الفرق بين مستويي القدرة		
	الصغرى	عدد الفروقات السالبة	العظمى	الصغرى	عدد الفروقات السالبة	العظمى	الصغرى	عدد الفروقات السالبة	العظمى	الصغرى	عدد الفروقات السالبة	العظمى
1	0.000	0	0.063	0.000	0	0.044	0.000	0	0.050	0.000	0	0.047
2	0.001	0	0.038	0.003	0	0.025	0.001	0	0.033	0.001	0	0.037
3	0.000	0	0.042	0.001	0	0.039	0.001	0	0.038	0.000	0	0.048
4	0.000	0	0.043	0.000	0	0.042	0.000	0	0.049	0.000	0	0.044
5	0.000	0	0.045	0.000	0	0.043	0.000	0	0.045	0.000	0	0.046
6	0.002	0	0.024	0.002	0	0.026	0.002	0	0.024	0.002	0	0.025
7	0.000	0	0.048	0.000	0	0.048	0.000	0	0.049	0.000	0	0.047
8	0.000	0	0.036	0.000	0	0.044	0.000	0	0.037	0.000	0	0.032
9	0.000	0	0.049	0.000	0	0.049	0.000	0	0.048	0.000	0	0.048
10	0.000	0	0.046	0.000	0	0.045	0.000	0	0.048	0.000	0	0.044
11	0.000	0	0.045	0.000	0	0.049	0.000	0	0.049	0.000	0	0.048
12	0.002	0	0.027	0.001	0	0.034	0.002	0	0.032	0.001	0	0.032
13	0.000	0	0.026	0.000	0	0.024	0.000	0	0.010	0.000	0	0.027
14	0.000	0	0.043	0.000	0	0.039	0.000	0	0.042	0.000	0	0.044
15	0.002	0	0.022	0.002	0	0.020	0.002	0	0.024	0.002	0	0.024
16	0.002	0	0.022	0.002	0	0.023	0.002	0	0.021	0.004	0	0.023
17	0.000	0	0.045	0.000	0	0.041	0.000	0	0.041	0.000	0	0.047
18	0.000	0	0.037	0.000	0	0.042	0.000	0	0.043	0.000	0	0.042
19	0.001	0	0.031	0.000	0	0.036	0.000	0	0.030	0.000	0	0.030
20	0.003	0	0.028	0.003	0	0.029	0.001	0	0.029	0.001	0	0.025
21	0.002	0	0.031	0.002	0	0.035	0.002	0	0.032	0.002	0	0.030
22	0.002	0	0.026	0.002	0	0.026	0.003	0	0.023	0.003	0	0.022
23	0.001	0	0.036	0.001	0	0.036	0.001	0	0.039	0.001	0	0.031
24	0.00	0	0.04	0.00	0	0.04	0.00	0	0.04	0.00	0	0.04

9		0	8		0	4		0	4		0	
0.045	0	0.000	0.056	0	0.000	0.043	0	0.001	0.043	0	0.001	25
0.026	0	0.003	0.032	0	0.001	0.030	0	0.002	0.027	0	0.002	26
0.052	0	0.000	0.045	0	0.000	0.046	0	0.000	0.039	0	0.001	27
0.032	0	0.001	0.029	0	0.001	0.033	0	0.001	0.037	0	0.001	28
0.044	0	0.000	0.052	0	0.000	0.043	0	0.000	0.047	0	0.000	29
0.042	0	0.000	0.030	0	0.001	0.039	0	0.000	0.034	0	0.000	30
0.042	0	0.000	0.048	0	0.000	0.040	0	0.001	0.037	0	0.001	31
0.029	0	0.001	0.027	0	0.002	0.027	0	0.002	0.031	0	0.001	32
0.033	0	0.001	0.034	0	0.001	0.038	0	0.000	0.039	0	0.000	33
0.053	0	0.000	0.048	0	0.000	0.052	0	0.000	0.047	0	0.000	34
0.037	0	0.000	0.040	0	0.000	0.041	0	0.000	0.038	0	0.000	35
0.042	0	0.000	0.041	0	0.000	0.038	0	0.000	0.041	0	0.000	36
0.046	0	0.000	0.045	0	0.000	0.040	0	0.001	0.041	0	0.000	37
0.045	0	0.000	0.038	0	0.000	0.043	0	0.000	0.045	0	0.000	38
0.032	0	0.001	0.027	0	0.001	0.029	0	0.001	0.027	0	0.002	39
0.052	0	0.000	0.042	0	0.000	0.048	0	0.000	0.041	0	0.000	40
0.027	0	0.001	0.027	0	0.001	0.028	0	0.001	0.031	0	0.000	41
0.024	0	0.002	0.020	0	0.002	0.025	0	0.002	0.025	0	0.002	42
0.024	0	0.002	0.021	0	0.004	0.018	0	0.005	0.017	0	0.005	43
0.039	0	0.000	0.036	0	0.000	0.034	0	0.000	0.032	0	0.001	44
0.035	0	0.001	0.033	0	0.001	0.035	0	0.001	0.030	0	0.002	45
0.025	0	0.003	0.022	0	0.004	0.025	0	0.004	0.026	0	0.003	46
0.039	0	0.000	0.042	0	0.000	0.047	0	0.000	0.046	0	0.000	47
0.025	0	0.001	0.022	0	0.002	0.026	0	0.001	0.026	0	0.001	48
0.020	0	0.004	0.020	0	0.004	0.021	0	0.004	0.020	0	0.005	49
0.036	0	0.001	0.037	0	0.001	0.043	0	0.001	0.039	0	0.001	50
1.291	0	0.102	1.280	0	0.133	1.389	0	0.135	1.382	0	0.151	الاختبار

يُلاحظ بأن جميع القيم صفرية **(0)** عند رصد مقدار التراجع في احتمالية الإجابة على المفردة إجابة صحيحة، لذا لا يوجد انتهاك لافتراض منحنى خصائص المفردة.

وللتحقق من افتراض منحنى خصائص المفردة لكل مفردة من مفردات الاختبار عند جميع المجموعات؛ فقد تمّ إنشاء رسوم بيانية لكل مفردة من مفردات الاختبار مُبينة في فصل الملاحق.

رابعًا: الكشف عن المفردات ذات الأداء التفاضلي

تمّ الكشف عن المفردات ذات الأداء التفاضلي في كل مجموعة من مجموعات الدراسة عن طريق برنامج (IRTPRO)، وهو تطبيق متقدم لمعايرة المفردات، وتسجيل النتائج باستخدام نظرية الاستجابة للمفردة (IRT)، ويتمتع البرنامج بمميزات عديدة، منها : الكشف عن المفردات ذات الأداء التفاضلي باستخدام اختبارات والد (Wald tests) المصممة وفقًا لاقتراح لورد (Lord, 1977, as cited in Lee & Suh, 2018) ، بالتضامن مع مصفوفات تباين – تغاير دقيقة لحساب الأخطاء المعيارية لتقديرات معالم المفردات، ويتمّ حسابها باستخدام خوارزمية (EM) التكميلية (SEM) (Cai, 2008). وتمّ الكشف عن المفردات ذات الأداء التفاضلي عند مستوى الدلالة (α= 0.01)، حيث تتمّ طريقة الكشف عن الأداء التفاضلي على مرحلتين لكل مجموعة من المجموعات المستهدفة مع المجموعة المرجعية، وتظهر المرحلة الأولى من خلال الجداول **(7، 8، 9)**.

جدول7 : المرحلة الأولى: المجموعة المرجعية مع المجموعة المستهدفة بنسبة مفردات ذات أداء تفاضلي 10%

عدد المفردات										
مجموعة 1	مجموعة 2	X^2 مجموع	درجات الحرية	قيمة الاحتمالية	X^2_a	درجات الحرية	قيمة الاحتمالية	$X^2_{c\|a}$	درجات الحرية	**قيمة الاحتمالية**
1	1	148.7	2	0.0001	6.9	1	0.0086	141.8	1	**0.0001**
3	3	13.1	2	0.0014	0.4	1	0.5238	12.7	1	**0.0004**
10	10	9.8	2	0.0074	0.1	1	0.8076	9.8	1	**0.0018**
11	11	8.7	2	0.0132	0.3	1	0.5926	8.4	1	**0.0038**
13	13	41.0	2	0.0001	0.0	1	0.9192	41.0	1	**0.0001**
29	29	247.7	2	0.0001	0.4	1	0.5527	247.3	1	**0.0001**
34	34	9.6	2	0.0083	1.3	1	0.2533	8.3	1	**0.0040**
36	36	126.4	2	0.0001	0.1	1	0.7070	126.3	1	**0.0001**
37	37	246.9	2	0.0001	0.0	1	0.8818	246.9	1	**0.0001**
46	46	8.3	2	0.0154	0.2	1	0.6334	8.1	1	**0.0044**

جدول8 : المرحلة الأولى: المجموعة المرجعية مع المجموعة المستهدفة بنسبة بنود ذات أداء تفاضلي 20%.

عدد المفردات										
مجموعة1	**مجموعة2**	X^2 مجموع	درجات الحرية	قيمة الاحتمالية	X^2_a	درجات الحرية	قيمة الاحتمالية	$X^2_{c\|a}$	درجات الحرية	**قيمة الاحتمالية**
1	1	416.3	2	0.0001	1.9	1	0.1645	414.3	1	**0.0001**
2	2	7.7	2	0.0210	0.3	1	0.6003	7.4	1	**0.0064**
3	3	20.9	2	0.0001	0.1	1	0.7971	20.8	1	**0.0001**
4	4	26.9	2	0.0001	1.9	1	0.1649	24.9	1	**0.0001**
5	5	35.4	2	0.0001	0.7	1	0.4143	34.7	1	**0.0001**
6	6	11.0	2	0.0042	0.4	1	0.5327	10.6	1	**0.0012**

0.0001	1	28.3	0.2979	1	1.1	0.0001	2	29.4	7	**7**
0.0037	1	8.4	0.5672	1	0.3	0.0127	2	8.7	8	**8**
0.0001	1	31.2	0.6535	1	0.2	0.0001	2	31.4	9	**9**
0.0001	1	166.5	0.3311	1	0.9	0.0001	2	167.5	10	**10**
0.0001	1	24.2	0.1812	1	1.8	0.0001	2	26.0	11	**11**
0.0001	1	43.5	0.6698	1	0.2	0.0001	2	43.7	13	**13**
0.0042	1	8.2	0.6040	1	0.3	0.0144	2	8.5	14	**14**
0.0037	1	8.4	0.4608	1	0.5	0.0111	2	9.0	15	**15**
0.0001	1	15.4	0.8513	1	0.0	0.0005	2	15.4	16	**16**
0.0002	1	14.1	0.9036	1	0.0	0.0009	2	14.1	17	**17**
0.0005	1	12.1	0.0372	1	4.3	0.0003	2	16.5	18	**18**
0.0001	1	83.5	0.7764	1	0.1	0.0001	2	83.6	19	**19**
0.0001	1	66.3	0.6252	1	0.2	0.0001	2	66.5	20	**20**
0.0001	1	18.1	0.3596	1	0.8	0.0001	2	18.9	21	**21**
0.0001	1	55.1	0.7004	1	0.1	0.0001	2	55.3	22	**22**
0.0002	1	13.8	0.2456	1	1.4	0.0005	2	15.2	23	**23**
0.0001	1	46.5	0.1876	1	1.7	0.0001	2	48.2	24	**24**
0.0001	1	29.5	0.0012	1	10.5	0.0001	2	40.0	25	**25**
0.0001	1	21.8	0.0577	1	3.6	0.0001	2	25.4	27	**27**
0.0001	1	14.9	0.1893	1	1.7	0.0002	2	16.6	28	**28**
0.0001	1	221.9	0.1606	1	2.0	0.0001	2	223.9	29	**29**
0.0001	1	15.2	0.6961	1	0.2	0.0005	2	15.4	30	**30**
0.0001	1	22.7	0.0025	1	9.1	0.0001	2	31.8	31	**31**
0.0001	1	17.8	0.5867	1	0.3	0.0001	2	18.1	33	**33**
0.0001	1	33.6	0.3117	1	1.0	0.0001	2	34.7	34	**34**
0.0046	1	8.0	0.347	1	0.9	0.0116	2	8.9	35	**35**

			3							
0.0001	1	122.1	0.7091	1	0.1	0.0001	2	122.2	36	**36**
0.0001	1	528.3	0.2376	1	1.4	0.0001	2	529.7	37	**37**
0.0001	1	151.6	0.6817	1	0.2	0.0001	2	151.7	38	**38**
0.0100	1	6.6	0.6266	1	0.2	0.0322	2	6.9	39	**39**
0.0002	1	14.3	0.2848	1	1.1	0.0004	2	15.4	42	**42**
0.0005	1	12.1	0.1156	1	2.5	0.0007	2	14.6	44	**44**
0.0001	1	27.9	0.1022	1	2.7	0.0001	2	30.6	45	**45**
0.0001	1	21.6	0.5956	1	0.3	0.0001	2	21.9	46	**46**
0.0001	1	26.3	0.9336	1	0.0	0.0001	2	26.3	47	**47**
0.0001	1	26.0	0.9350	1	0.0	0.0001	2	26.0	50	**50**

جدول9 : المرحلة الأولى: المجموعة المرجعية مع المجموعة المستهدفة بنسبة مفردات ذات أداء تفاضلي 30%

									عدد المفردات	
قيمة الاحتمالية	**درجات الحرية**	$X^2_{c\|a}$	**قيمة الاحتمالية**	**درجات الحرية**	X^2_a	**قيمة الاحتمالية**	**درجات الحرية**	X^2 مجموع	مجموعة 2	**مجموعة 1**
0.0001	1	99.1	0.0572	1	3.6	0.0001	2	102.7	1	**1**
0.0042	1	8.2	0.7461	1	0.1	0.0158	2	8.3	2	**2**
0.0001	1	105.2	0.2480	1	1.3	0.0001	2	106.5	3	**3**
0.0001	1	26.8	0.4871	1	0.5	0.0001	2	27.2	4	**4**
0.0001	1	28.6	0.5149	1	0.4	0.0001	2	29.0	5	**5**
0.0001	1	87.8	0.4706	1	0.5	0.0001	2	88.3	6	**6**
0.0001	1	30.5	0.8048	1	0.1	0.0001	2	30.6	7	**7**
0.0032	1	8.7	0.5441	1	0.4	0.0108	2	9.0	8	**8**
0.0001	1	26.3	0.9513	1	0.0	0.0001	2	26.3	9	**9**
0.0001	1	169.4	0.9726	1	0.0	0.0001	2	169.4	10	**10**
0.0001	1	28.0	0.3492	1	0.9	0.0001	2	28.9	11	**11**
0.0006	1	11.7	0.0777	1	3.1	0.0006	2	14.8	12	**12**
0.0001	1	26.3	0.5923	1	0.3	0.0001	2	26.6	13	**13**
0.0009	1	11.1	0.4104	1	0.7	0.0028	2	11.8	14	**14**
0.0009	1	10.9	0.5741	1	0.3	0.0036	2	11.2	15	**15**
0.0001	1	59.9	0.5612	1	0.3	0.0001	2	60.2	16	**16**
0.0001	1	24.9	0.4357	1	0.6	0.0001	2	25.5	17	**17**
0.0001	1	30.9	0.0855	1	3.0	0.0001	2	33.8	18	**18**
0.0001	1	89.1	0.9862	1	0.0	0.0001	2	89.1	19	**19**
0.0001	1	79.0	0.5152	1	0.4	0.0001	2	79.5	20	**20**
0.0001	1	20.1	0.9329	1	0.0	0.0001	2	20.1	21	**21**

0.0001	1	60.1	0.4091	1	0.7	0.0001	2	60.8	22	**22**
0.0001	1	15.9	0.3748	1	0.8	0.0002	2	16.7	23	**23**
0.0001	1	45.5	0.2578	1	1.3	0.0001	2	46.7	24	**24**
0.0001	1	21.6	0.3654	1	0.8	0.0001	2	22.4	25	**25**
0.0005	1	12.0	0.9655	1	0.0	0.0025	2	12.0	26	**26**
0.0001	1	16.6	0.0040	1	8.3	0.0001	2	24.9	27	**27**
0.0001	1	24.5	0.3814	1	0.8	0.0001	2	25.3	28	**28**
0.0001	1	137.8	0.7946	1	0.1	0.0001	2	137.8	29	**29**
0.0001	1	20.6	0.0530	1	3.7	0.0001	2	24.3	30	**30**
0.0001	1	28.1	0.1139	1	2.5	0.0001	2	30.7	31	**31**
0.0001	1	18.2	0.7560	1	0.1	0.0001	2	18.3	32	**32**
0.0001	1	15.7	0.2689	1	1.2	0.0002	2	16.9	33	**33**
0.0001	1	33.3	0.1370	1	2.2	0.0001	2	35.5	34	**34**
0.0001	1	21.0	0.9247	1	0.0	0.0001	2	21.0	35	**35**
0.0001	1	85.6	0.4760	1	0.5	0.0001	2	86.1	36	**36**
0.0001	1	155.8	0.1849	1	1.8	0.0001	2	157.6	37	**37**
0.0001	1	123.6	0.8099	1	0.1	0.0001	2	123.7	38	**38**
0.0001	1	15.3	0.0723	1	3.2	0.0001	2	18.6	39	**39**
0.0001	1	17.8	0.9493	1	0.0	0.0001	2	17.8	42	**42**
0.0001	1	78.5	0.0146	1	6.0	0.0001	2	84.5	43	**43**
0.0053	1	7.8	0.0416	1	4.2	0.0026	2	11.9	44	**44**
0.0001	1	33.9	0.1229	1	2.4	0.0001	2	36.3	45	**45**
0.0001	1	29.8	0.2753	1	1.2	0.0001	2	31.0	47	**47**
0.0088	1	6.9	0.7880	1	0.1	0.0310	2	6.9	48	**48**
0.0001	1	27.0	0.7633	1	0.1	0.0001	2	27.1	50	**50**

يتضح من خلال جداول المرحلة الأولى **(7،8،9)** وجود **(10)** مفردات ذات أداء تفاضلي في المجموعة المستهدفة الأولى، و**(42)** مفردة في المجموعة المستهدفة الثانية، و**(46)** مفردة في المجموعة المستهدفة الثالثة، لتأتي بعد ذلك المرحلة الثانية لإثبات صحة وجود أداء تفاضلي في تلك المفردات، من خلال إدخال المفردات التي تبدي أداء تفاضليا في مربع (Candidate items)، وبقية المفردات يتم إدخالها في (Anchor items). لتظهر المرحلة الثانية كما في الجداول **(10،11،12)**.

جدول10 : المرحلة الثانية: المجموعة المرجعية مع المجموعة المستهدفة بنسبة مفردات ذات أداء تفاضلي 10%.

عدد المفردات										
مجموعة 1	**مجموعة 2**	X^2 مجموع	درجات الحرية	قيمة الاحتمالية	X^2_a	درجات الحرية	قيمة الاحتمالية	$X^2_{c\|a}$	**قيمة الاحتمالية**
1	1	123.8	2	0.0001	6.5	1	0.0106	117.2	**0.0001**	1
13	13	47.3	2	0.0001	0.0	1	0.8853	47.2	**0.0001**	1
29	29	254.5	2	0.0001	0.4	1	0.5426	254.1	**0.0001**	1
36	36	138.4	2	0.0001	0.2	1	0.6730	138.2	**0.0001**	1
37	37	230.7	2	0.0001	0.0	1	0.8479	230.6	**0.0001**	1

جدول11 : المرحلة الثانية: المجموعة المرجعية مع المجموعة المستهدفة بنسبة مفردات ذات أداء تفاضلي 20%.

عدد المفردات										
مجموعة 1	**مجموعة 2**	X^2 مجموع	درجات الحرية	قيمة الاحتمالية	X^2_a	درجات الحرية	قيمة الاحتمالية	$X^2_{c\|a}$	**قيمة الاحتمالية**
1	1	349.5	2	0.0001	3.6	1	0.0584	345.9	**0.0001**	1
10	10	181.2	2	0.0001	0.0	1	0.8268	181.2	**0.0001**	1
13	13	50.7	2	0.0001	0.1	1	0.7657	50.6	**0.0001**	1
19	19	111.9	2	0.0001	0.1	1	0.7500	111.8	**0.0001**	1
20	20	93.1	2	0.0001	0.0	1	0.8999	93.1	**0.0001**	1
22	22	79.8	2	0.0001	0.9	1	0.3506	78.9	**0.0001**	1
24	24	9.1	2	0.0103	0.2	1	0.6577	9.0	**0.0028**	1
29	29	198.0	2	0.0001	1.1	1	0.2991	196.9	**0.0001**	1
34	34	7.0	2	0.0301	0.1	1	0.8178	6.9	**0.0084**	1
36	36	135.2	2	0.0001	0.0	1	0.9751	135.2	**0.0001**	1
37	37	460.8	2	0.0001	0.3	1	0.5894	460.5	**0.0001**	1
38	38	175.7	2	0.0001	1.0	1	0.3141	174.6	**0.0001**	1

جدول12 : المرحلة الثانية: المجموعة المرجعية مع المجموعة المستهدفة بنسبة مفردات ذات أداء تفاضلي 30%.

عدد المفردات										
مجموعة 1	**مجموعة 2**	X^2 مجموع	درجات الحرية	قيمة الاحتمالية	X^2_a	درجات الحرية	قيمة الاحتمالية	$X^2_{c\|a}$	درجات الحرية	**قيمة الاحتمالية**
1	1	97.7	2	0.0001	5.6	1	0.0184	92.2	1	**0.0001**
3	3	110.7	2	0.0001	0.4	1	0.5080	110.2	1	**0.0001**
4	4	40.6	2	0.0001	0.2	1	0.6887	40.4	1	**0.0001**
6	6	99.7	2	0.0001	0.1	1	0.8206	99.6	1	**0.0001**
10	10	155.2	2	0.0001	0.3	1	0.5764	154.9	1	**0.0001**
13	13	37.7	2	0.0001	0.1	1	0.7813	37.6	1	**0.0001**
16	16	74.0	2	0.0001	0.0	1	0.8956	74.0	1	**0.0001**
18	18	9.1	2	0.0106	1.3	1	0.2590	7.8	1	**0.0052**
19	19	104.7	2	0.0001	0.3	1	0.6123	104.5	1	**0.0001**
20	20	98.3	2	0.0001	1.3	1	0.2575	97.1	1	**0.0001**
22	22	79.4	2	0.0001	1.7	1	0.1962	77.7	1	**0.0001**
29	29	136.2	2	0.0001	0.7	1	0.4094	135.5	1	**0.0001**
36	36	93.1	2	0.0001	0.1	1	0.7688	93.0	1	**0.0001**
37	37	134.9	2	0.0001	0.5	1	0.4874	134.4	1	**0.0001**
38	38	120.2	2	0.0001	0.1	1	0.7775	120.1	1	**0.0001**
43	43	92.1	2	0.0001	4.1	1	0.0432	88.0	1	**0.0001**
45	45	7.9	2	0.0196	0.9	1	0.3398	7.0	1	**0.0084**

يُلاحظ من خلال جداول المرحلة الثانية، اختلاف عدد المفردات ذات الأداء التفاضلي عن المرحلة الأولى، حيث يُلاحظ بأن المفردات التي تبدي أداءً تفاضليًّا في المجموعة المستهدفة بنسبة **(10%)** هي **(5)** مفردات، وفي المجموعة المستهدفة بنسبة **(20%)** هي **(12)** مفردة، وفي المجموعة المستهدفة بنسبة **(30%)** هي **(17)** مفردة.

وأظهرت النتائج بأن المفردات التي صُممت على أن تبدي أداءً تفاضليًّا قد أظهرت وجود أداء تفاضلي، بالإضافة إلى تأثر بعض المفردات الأخرى التي لم تصمم على أن تبدي أداءً تفاضليًّا في المجموعتين المستهدفتين بنسبة **(20% - 30%)**، حيث كانت قيمة P-

value عند جميع المفردات تساوي أو أصغر من مستوى الدلالة (0.01 =α) مما يفيد بوجود أداء تفاضلي في تلك المفردات.

رابعًا: مطابقة الأفراد والمفحوصين للنموذج ثنائي المعلمة:

تمّ استخدام برنامج (Bilog-MG) للكشف عن مطابقة المفردات والأفراد للنموذج ثنائي المعلمة. وأظهرت النتائج عدم مطابقة **(6)** مفحوصين من أصل **(4000)** مفحوص في كل من المجموعات المرجعية والمستهدفة؛ نظرًا لتعذر حساب الخطأ المعياري في تقدير القدرة عند مستوى قدرة ما، لذا تمّ حذف الأفراد غير المطابقين في كل مجموعة بسبب عدم مطابقتهم للنموذج ثنائي المعلمة، ليصبح عدد المفحوصين بعد حذف الأفراد غير المطابقين في كل مجموعة (**950**) مفحوصًا. وجدول **(13)** يوضح عدد المفحوصين غير المطابقين للنموذج ثنائي المعلمة في كل مجموعة.

جدول 13: المفحوصين غير المطابقين في كل من المجموعة المرجعية والمجموعات المستهدفة.

المجموعة	التكرار	المفحوصين غير المطابقين
المجموعة المرجعية	2	62 ، 269
المجموعة المستهدفة بنسبة 10%	1	919
المجموعة المستهدفة بنسبة 20%	1	693
المجموعة المستهدفة بنسبة 30%	2	693 ، 930

وبعد حذف المفحوصين غير المطابقين، تمّ إعادة التحليل ومطابقة الفقرات في كل مجموعة من مجموعات الدراسة لتُظهر النتائج مطابقة جميع الفقرات للنموذج ثنائي المعلمة.

مناقشة النتائج.

عرض النتائج ومناقشتها:

السؤال الأول: هل يوجد تأثير لنسب المفردات ذات الأداء التفاضلي (0%، 10%- 20%- 30%) على دقة تقدير قدرات الأفراد وفق النموذج ثنائي المعلمة؟

للإجابة عن السؤال الأول تمّ استخراج الأوساط الحسابية، والأخطاء المعيارية لقدرات الأفراد من خلال مخرجات برنامج Bilog_MG، والجدول **(14)** يوضح ذلك.

جدول 14: إحصاءات القدرة وفق مجموعات الدراسة

المجموعة	الإحصائي	القدرة	الخطأ المعياري في تقدير القدرة
المجموعة المرجعية	العدد	995	**995**
	الوسط الحسابي	-0.0008	**0.3149**
	الانحراف المعياري	1.0861	**0.0535**
	القيمة الصغرى	-3.8649	**0.2799**
	القيمة العظمى	3.5634	**0.8528**
المجموعة المستهدفة 10%	العدد	995	**995**
	الوسط الحسابي	-0.0056	**0.3148**
	الانحراف المعياري	1.0774	**0.0597**
	القيمة الصغرى	-3.4983	**0.2781**
	القيمة العظمى	3.8206	**0.7798**
المجموعة المستهدفة 20%	العدد	995	**995**
	الوسط الحسابي	-0.0126	**0.3280**
	الانحراف المعياري	1.0817	**0.0644**
	القيمة الصغرى	-3.7276	**0.2876**
	القيمة العظمى	3.2000	**0.9056**
المجموعة المستهدفة 30%	العدد	995	**995**
	الوسط الحسابي	-0.0100	**0.3221**
	الانحراف المعياري	1.0799	**0.0653**
	القيمة الصغرى	-3.8452	**00.288**
	القيمة العظمى	3.2590	**1.0260**

كما تمّ استخدام معادلة اختبار (ت) للفرق بين كل وسطين حسابيين بالاعتماد على الخطأين المعياريين لهما $t = (x1 - x2)/\sqrt{se_1^2 + se_2^2}$ ، لعقد مقارنة بين فترتي ثقة للتأكد من مدى انتهاك خاصية اللاتغاير لدقة تقدير قدرات الأفراد بين المجموعة المرجعية والمجموعات المستهدفة بنسب مفردات ذات أداء تفاضلي (0%، 10%، 20%، 30%)، وذلك بسبب أن اختبار (ت) يُراعي خصوصية قدرة كل فرد، والخطأ المعياري في تقدير قدرته؛ لذا تمّ استخدامه بدلًا من اختبار تحليل التباين (ANOVA)، استنادًا إلى أن الخطأ المعياري في تقدير القدرة لا يتخذ شكل التوزيع

الطبيعي بل على العكس منه، كما أنه يتعامل مع الوسط الحسابي الكلي للخطأ المعياري في تقدير القدرة وفق كل نسبة من نسب الأداء التفاضلي (0%، 10%، 20%، 30%)، مما يترتب على ذلك ضياع حقيقة أثر نسب المفردات ذات الأداء التفاضلي على اختلاف نسبتها، والجدول **(15)** يوضح نتائج اختبار (ت):

جدول 14 :نتائج اختبار (ت) للتحقق من خاصية اللاتغاير في دقة تقدير القدرة

المجموعة	خاصية اللاتغاير	التكرار	النسبة المئوية	الباقي المعياري	القرار
المستهدفة بنسبة 10%	غير متحققة	52	5.23	-19.973	**متحقق**
	متحققة	943	94.77	19.973	**متحقق**
المستهدفة بنسبة 20%	غير متحققة	53	5.33	-19.929	**متحقق**
	متحققة	942	94.67	19.929	**متحقق**
المستهدفة بنسبة 30%	غير متحققة	59	5.93	-19.660	**متحقق**
	متحققة	936	94.07	19.660	**متحقق**

يتضح من خلال الجدول تحقق خاصية اللاتغاير عند جميع المجموعات المستهدفة بنسبة مفردات ذات أداء تفاضلي (10%، 20%، 30%) عند مقارنتها بالمجموعة المرجعية، ويُلاحظ بأن الباقي المعياري يساوي (19-)؛ أي أصغر من (2-) عند جميع المجموعات المستهدفة عندما كانت خاصية اللاتغاير غير متحققة، وعند تحقق خاصية اللاتغاير عند جميع المجموعات كان الباقي المعياري يساوي (19)؛ أي أكبر من (2+) عند جميع المجموعات المستهدفة، مما يفيد بتحقق خاصية اللاتغاير في دقة تقديرات قدرات الأفراد بين المجموعة المرجعية والمجموعات المستهدفة بنسب مفردات ذات أداء تفاضلي (10%، 20%، 30%). كما يُلاحظ من الجدول السابق زيادة عدد الأفراد المنتهكين لخاصية اللاتغاير بزيادة نسبة المفردات ذات الأداء التفاضلي مقارنة بالمجموعة المرجعية، مما يثبت ذلك عدم وجود فروق دالة إحصائيًّا عند مستوى دلالة (α= 0.05) بين دقة تقديرات

قدرات الأفراد تُعزى لنسب المفردات ذات الأداء التفاضلي **(10%، 20%، 30%).**

السؤال الثاني: هل يوجد تأثير لنسب المفردات ذات الأداء التفاضلي (10%- 20%- 30%) على معلمتي الصعوبة والتمييز لمفردات الاختبار وفق النموذج ثنائي المعلمة؟

للإجابة عن السؤال الثاني، فقد تمّ استخراج الأوساط الحسابية والأخطاء المعيارية لمعلمتي الصعوبة والتمييز من خلال مخرجات برنامج Bilog_MG ، وذلك كما هو مبين في الجدول **(16).**

جدول15 :إحصاءات معلمة الصعوبة والتمييز وفقًا لمجموعات الدراسة

المجموعة	الإحصائي	معلمة التمييز	الخطأ المعياري لها	معلمة الصعوبة	الخطأ المعياري لها
المرجعية	العدد	50	50	50	50
	الوسط الحسابي	1.131	0.095	0.004	0.096
	الانحراف المعياري	0.297	0.021	1.078	0.046
	القيمة الصغرى	0.527	0.066	-1.939	0.104
	القيمة العظمى	1.989	0.192	2.708	0.307
المستهدفة 10%	العدد	50	50	50	50
	الوسط الحسابي	1.136	0.097	0.167	0.108
	الانحراف المعياري	0.282	0.0212	1.199	0.109
	القيمة الصغرى	0.566	0.067	-2.09	0.142
	القيمة العظمى	1.679	0.158	4.49	0.809
المستهدفة 20%	العدد	50	50	50	50
	الوسط الحسابي	1.137	0.108	0.488	0.157
	الانحراف المعياري	0.322	0.0486	1.534	0.283
	القيمة الصغرى	0.613	0.069	-2.25	0.167
	القيمة العظمى	1.763	0.294	6.407	1.973

المستهدفة	العدد	50	50	50	50
30%	الوسط الحسابي	1.1520	0.1046	0.502	0.121
	الانحراف المعياري	0.3017	0.0354	1.276	0.111
	القيمة الصغرى	0.631	0.070	-1.952	0.118
	القيمة العظمى	1.648	0.281	4.158	0.721

كما تمّ استخدام معادلة اختبار (ت) للفرق بين وسطين حسابيين بالاعتماد على الخطأين المعياريين لهما، لعقد مقارنة بين فترتي ثقة؛ للتأكد من مدى انتهاك خاصية اللاتغاير لمعلمتي الصعوبة والتمييز بين المجموعة المرجعية والمجموعات المستهدفة، وذلك بسبب أن اختبار (ت) يُراعي خصوصية معلمتي الصعوبة والتمييز لكل مفردة من مفردات الاختبار، والخطأ المعياري في دقة تقدير معلمتي الصعوبة والتمييز وفق النموذج ثنائي المعلمة؛ لذا تمّ استخدام اختبار (ت) بدلًا من اختبار تحليل التباين (ANOVA)، استنادًا إلى أن الخطأ المعياري في تقدير معلمتي الصعوبة والتمييز لا يتخذ شكل التوزيع الطبيعي بل على العكس منه، كما أنه يتعامل مع الوسط الحسابي الكلي للخطأ المعياري في تقدير معلمتي الصعوبة والتمييز وفق كل نسبة من نسب المفردات ذات الأداء التفاضلي (0%، 10%، 20%، 30%)، مما يترتب على ذلك ضياع حقيقة أثر نسب المفردات ذات الأداء التفاضلي، والجدول **(17)** يوضح ذلك.

جدول 16: خاصية اللاتغاير في دقة تقدير معلمتي الصعوبة والتمييز.

المجموعة	معالم البنود	خاصية اللاتغاير	التكرار	النسبة المئوية	الباقي المعياري	القرار
المستهدفة 10%	التمييز	غير متحقق	3	6.00	-4.400	متحقق
		متحقق	47	94.00	4.400	متحقق
	الصعوبة	غير متحقق	8	16.00	-3.400	متحقق
		متحقق	42	84.00	3.400	متحقق
المستهدفة 20%	التمييز	غير متحقق	3	6.00	-4.400	متحقق
		متحقق	47	94.00	4.400	متحقق
	الصعوبة	غير متحقق	13	26.00	-2.400	متحقق
		متحقق	37	74.00	2.400	متحقق

المستهدفة 30%	التمييز	غير متحقق	3	6.00	-4.400	متحقق
		متحقق	47	94.00	4.400	متحقق
	الصعوبة	غير متحقق	17	34.00	-1.600	غير متحقق
		متحقق	33	66.00	1.600	غير متحقق

يُلاحظ من الجدول السابق تحقق خاصية اللاتغاير عند جميع المجموعات المستهدفة لمعلمة التمييز، كما يُلاحظ ثبات الباقي المعياري لمعلمة التمييز عند جميع المجموعات، حيث بلغ (4.400 -)؛ أي أصغر من (2-) عند انتهاك خاصية اللاتغاير، وبلغ (4.400) ؛ أي أكبر من (2+) عند تحقق خاصية اللاتغاير بسبب أن الباقي المعياري عندما يكون أكبر من (2-) عند انتهاك خاصية اللاتغاير، وأقل من (2+) عند تحقق خاصية اللاتغاير يؤدي ذلك إلى انتهاك خاصية اللاتغاير عند المجموعة المستهدفة بنسب مفردات ذات أداء تفاضلي (10%، 20%،30%) لمعلمتي الصعوبة والتمييز. ويرجع السبب في ذلك إلى عدم إجراء تغيير على معلمة التمييز عند جميع المجموعات المستهدفة بنسب مفردات ذات أداء تفاضلي (10%، 20%،30%)، مما يثبت ذلك عدم وجود فروق دالة إحصائيًّا عند مستوى دلالة (α= 0.05) بين دقة تقديرات معلمات التمييز تُعزى لنسب المفردات ذات الأداء التفاضلي (10%، 20%،30%). كما يظهر من خلال الجدول السابق تحقق خاصية اللاتغاير لمعلمة الصعوبة عند المجموعتين المستهدفة بنسب مفردات ذات أداء تفاضلي (10%، 20%)، كما تمّ انتهاك خاصية اللاتغاير عند المجموعة المستهدفة بنسبة مفردات ذات أداء تفاضلي (30%)، ويرجع سبب ذلك إلى أنه تمّ إدخال تغيير في قيمة معلمة الصعوبة مقدارها (1.5)، إلا أن المجموعتين المستهدفة بنسبة مفردات ذات أداء تفاضلي (10%-20%) لم تتأثر بمقدار التغيير، كما يُلاحظ بأن الباقي المعياري لمعلمة الصعوبة في المجموعات المستهدفة يقل بزيادة البنود ذات الأداء التفاضلي، مما يدل ذلك على تأثر معلمة الصعوبة في المجموعات المستهدفة بزيادة نسب المفردات ذات الأداء التفاضلي.

مما يثبت عدم وجود فروق دالة إحصائيًّا عند مستوى دلالة ($\alpha = 0.05$) بين دقة تقديرات معلمات الصعوبة تعزى لنسب المفردات ذات الأداء التفاضلي (10%، 20%). ووجود فروق دالة إحصائيًّا عند مستوى دلالة ($\alpha = 0.05$) بين دقة تقديرات معلمات الصعوبة تُعزى لنسب المفردات ذات الأداء التفاضلي (30%).

السؤال الثالث: هل يوجد تأثير لنسب المفردات ذات الأداء التفاضلي (10%- 20%- 30%) على معامل الثبات النظري والتجريبي وفق النموذج ثنائي المعلمة؟

للإجابة عن السؤال الثالث، فقد تمّ حساب الإحصائي (M) للعينات المستقلة؛ للكشف عن جوهرية الفروق بين قيم معاملات الثبات النظري والثبات التجريبي بين مجموعات الدراسة التي تُعزى إلى نسبة المفردات ذات الأداء التفاضلي (0%، 10%، 20%، 30%) عند درجة حرية مقدارها (3) وفقًا لتوزيع (χ^2)، وذلك كما هو مبين في الجدولين **(18)** و **(19)**.

الثبات النظري:

جدول 17: إحصائي (M) لدلالة الفروق بين الثبات النظري بين مجموعات الدراسة

مجموعة الأداء التفاضلي	طول الاختبار	حجم العينة	الثبات النظري	قيمة الإحصائي M	درجة الحرية	احتمالية الخطأ
المرجعية (صفر%)	50	995	0.9157	1.70	3	0.64
المستهدفة (10%)	50	995	0.9162			
المستهدفة (20%)	50	995	0.9099			
المستهدفة (30%)	50	995	0.9124			

يتضح من خلال الجدول السابق عدم وجود فروق دالة إحصائيًّا عند مستوى الدلالة **(α=0.05)** بين قيم معاملات الثبات النظري بين مجموعات الدراسة تُعزى لنسبة المفردات ذات الأداء التفاضلي (0%، 10%، 20%، 30%)، حيث بلغت قيمة الإحصائي (M) المحسوبة (1.70) وهذه القيمة أصغر من قيمة (χ^2) الحرجة عند درجة حرية (3)، والبالغة قيمتها (7.815).كما يُلاحظ من خلال الجدول أن قيم الثبات المقدرة كانت

عالية، وأعلى قيمة للثبات النظري كانت عند المجموعة المستهدفة بنسبة مفردات ذات أداء تفاضلي **(10%)** مقارنة بالمجموعة المرجعية والمجموعات المستهدفة بنسبة مفردات ذات أداء تفاضلي **(20%، 30%)**.

الثبات التجريبي:

جدول 18: إحصائي (M) لدلالة الفروق بين الثبات التجريبي بين مجموعات الدراسة.

مجموعة الأداء التفاضلي	طول الاختبار	حجم العينة	الثبات الأمبريقي	قيمة الإحصائي M	درجة الحرية	احتمالية الخطأ
المرجعية (صفر%)	50	995	0.9203	1.85	3	0.60
مستهدفة (10%)	50	995	0.9197			
مستهدفة (20%)	50	995	0.9140			
مستهدفة (30%)	50	995	0.9164			

يتضح من خلال الجدول السابق عدم وجود فروق دالة إحصائيًّا عند مستوى الدلالة **(α=0.05)** بين قيم معاملات الثبات التجريبي بين مجموعات الدراسة تُعزى لنسبة المفردات ذات الأداء التفاضلي (10%، 20%، 30%). حيث بلغت قيمة الإحصائي (M) المحسوبة (1.80) وهذه القيمة أصغر من قيمة (χ^2) الحرجة عند درجة حرية (3)، والبالغة قيمتها (7.815). كما يلاحظ من خلال الجدول أن قيم الثبات النظري كانت عالية عند جميع مجموعات الدراسة، وأعلى قيمة كانت لدى المجموعة المرجعية.

السؤال الرابع: هل يوجد تأثير لنسب المفردات ذات الأداء التفاضلي (0%، 10%، 20%، 30%) على قيم دالة معلومات الاختبار باستخدام النموذج ثنائي المعلمة؟

للإجابة عن سؤال الدراسة الرابع، فقد تمّ حساب الأوساط الحسابية والانحرافات المعيارية لكمية معلومات الاختبار وفقًا لمجموعات الدراسة بنسب مفردات ذات أداء تفاضلي (0%، 10%، 20%، 30%) ، وفئات القدرة وذلك كما هو مبين في الجدول **(20)**.

جدول 19: الأوساط الحسابية والانحرافات المعيارية لكمية معلومات الاختبار وفقًا لمجموعات الدراسة ومستويات القدرة.

المتغير التابع: دالة المعلومات						
فئة القدرة	الإحصائي	مجموعة الأداء التفاضلي				الكلي
		صفر%	%10	%20	%30	
بين [(2.49-)—(1.5-)]	العدد	86	81	91	89	347
	الوسط الحسابي	6.44	5.46	4.59	4.53	5.24
	الانحراف المعياري	1.50	1.56	1.24	1.39	1.62
بين [(1.49-)—(0.5-)]	العدد	233	235	223	218	909
	الوسط الحسابي	10.25	9.92	8.59	8.63	9.37
	الانحراف المعياري	0.91	1.24	1.01	1.11	1.31
بين [(0.49-)—(0.49)]	العدد	343	354	344	359	1400
	الوسط الحسابي	12.46	12.61	11.44	11.45	11.99
	الانحراف المعياري	0.32	0.35	0.58	0.59	0.73
بين [(1.49)—(0.50)]	العدد	256	252	269	251	1028
	الوسط الحسابي	11.17	11.55	11.42	11.47	11.40
	الانحراف المعياري	1.08	0.89	0.62	0.60	0.83
بين [(2.49)—(1.50)]	العدد	77	73	68	78	296
	الوسط الحسابي	6.54	7.41	8.20	8.38	7.62
	الانحراف المعياري	1.73	1.82	1.42	1.29	1.74
الكلي	العدد	995	995	995	995	
	الوسط الحسابي	10.63	10.74	9.95	9.98	
	الانحراف المعياري	2.24	2.39	2.31	2.32	

يلاحظ من خلال الجدول السابق وجود فروق ظاهرية بين الأوساط الحسابية لكمية معلومات الاختبار ناتجة عن اختلاف فئات القدرة ومجموعات الدراسة بنسب المفردات ذات الأداء التفاضلي **(0%،10%،20 %، 30%)** عند كل مجموعة من مجموعات الدراسة.

وللتحقق من جوهرية الفروق الظاهرية؛ فقد تمّ إجراء تحليل التباين ثنائي التفاعل وفقًا لمجموعات الدراسة وفئات القدرة، وذلك كما هو مُبيَّن في الجدول **(21).**

جدول 20: نتائج تحليل التباين ثنائي التفاعل لكمية معلومات الاختبار وفقًا لمجموعات الدراسة وفئات القدرة.

المتغير التابع: دالة المعلومات						
مصدر التباين	مجموع المربعات	درجة حرية	وسط مجموع المربعات	F	احتمالية الخطأ	حجم الأثر
مجموعة الأداء التفاضلي	181.25	3	60.42	*68.94	0.00	0.0496
فئات القدرة	16987.16	4	4246.79	*4845.86	0.00	0.8304
مجموعة الأداء التفاضلي×فئة القدرة	796.43	12	66.37	*75.73	0.00	0.1867
الخطأ	3470.45	3960	0.88			
الكلي	21842.11	3979				

* دال إحصائيًا (α=0.05)

يتضح من خلال الجدول السابق وجود فروق دالة إحصائيًّا عند مستوى الدلالة **(α=0.05)** بين الأوساط الحسابية لمجموعات الدراسة تُعزى لنسب المفردات ذات الأداء التفاضلي (0%، 10%، 20%، 30%)، ولفئات القدرة وللتفاعل بينهما. ولكون مجموعات الدراسة وفئات القدرة متعدد المستويات، فقد تمّ إجراء اختبار ليفين (Levene)؛ للكشف عن انتهاك تجانس التباين من عدمه بين الأوساط الحسابية لكمية معلومات الاختبار وفقًا لمجموعات الدراسة ولفئات القدرة، حيث بلغت قيمة (F) المحسوبة لاختبار ليفين (81.49) بدلالة إحصائية **(α=0.05)** عند درجتي حرية (19) للبسط، و(3960) للمقام؛ بما يفيد وجود انتهاك في تجانس التباين؛ مما أوجب استخدام أحد اختبارات المقارنات البعدية التي تراعي انتهاك تجانس التباين، حيث تمّ استخدام اختبار جيمس-هويل (Games-Howell) للمقارنات البعدية المتعددة؛ بهدف تحديد لصالح أيّ مجموعة من مجموعات الدراسة، وأيّ فئة من فئات القدرة قد كانت الفروق الجوهرية بين الأوساط الحسابية لكمية معلومات الاختبار، وذلك كما هو مبين في جدول **(22)** و**(23)**.

جدول 21: اختبار جيمس- هويل للمقارنات البعدية بين مجموعات الدراسة.

المتغير التَّابع: دالة المعلومات				
مجموعة الأداء التفاضلي		%20	%30	صفر%
{Games-Howell}	الوسط الحسابي	9.95	9.98	10.63
%30	9.98	0.03		
صفر%	10.63	*0.69	*0.65	
%10	10.74	*0.80	*0.77	*0.11

* دال إحصائيًا (α=0.05)

يتضح من خلال الجدول أنَّ الفروق الجوهرية بين الأوساط الحسابية لكمية معلومات الاختبار وفقًا لمجموعات الدراسة؛ لصالح المجموعة المرجعية مقارنة بالمجموعة المستهدفة بنسبة **(20%)**، ومن ثم المجموعة المستهدفة بنسبة **(30%)**، ثم يتضح أنَّ الفروق الجوهرية بين الأوساط الحسابية لكمية معلومات الاختبار وفقًا لمجموعات الدراسة لصالح المجموعة المستهدفة **(10%)** مقارنة بالمجموعة المرجعية.

جدول 22: اختبار جيمس- هويل للمقارنات البعدية بين فئات القدرة.

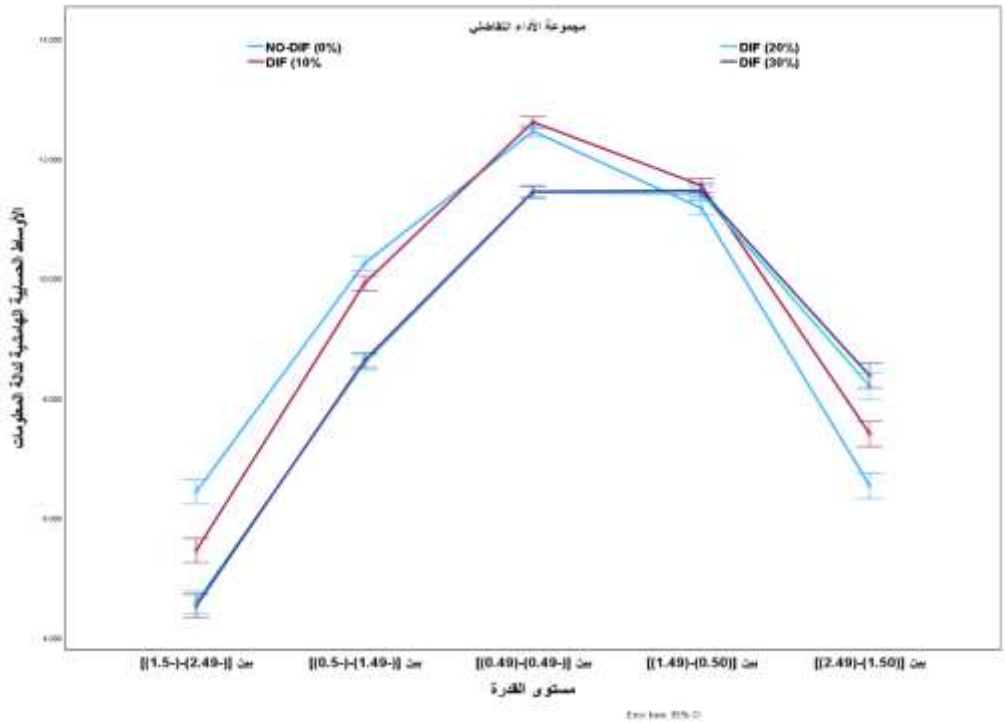
المتغير التَّابع: دالة المعلومات				
فئة القدرة		بين [(2.49-)—(1.5-)]	بين [(1.50)—(2.49)]	بين [(1.49-)—(0.5-)]	بين [(0.50)—(1.49)]
{Games-Howell}	الوسط الحسابي	5.24	7.62	9.37	11.40
بين [(1.50)—(2.49)]	7.62	*2.38			
بين [(1.49-)—(0.5-)]	9.37	*4.13	*1.75		
بين [(0.50)—(1.49)]	11.40	*6.17	*3.78	*2.03	
بين [(0.49-)—(0.49)]	11.99	*6.75	*4.37	*2.62	*0.59

* دال إحصائيًّا (α=0.05)

يتضح من خلال الجدول السابق أنَّ الفروق الجوهرية بين الأوساط الحسابية لكمية معلومات الاختبار وفقًا لفئات القدرة قد كانت على الترتيب: لصالح فئة القدرة بين [(0.49-)— (0.49)] مقارنة بفئة القدرة [(2.49-)— (1.5-)]، ومن ثمّ فئة القدرة [(1.50)— (2.49)]، ومن ثمّ فئة القدرة [(1.49-)— (0.5-)]، ومن ثمّ فئة القدرة [(0.50)— (1.49)]. كما أن الفروق الجوهرية بين الأوساط الحسابية لكمية معلومات الاختبار وفقًا لفئات القدرة قد كانت على الترتيب: لصالح فئة القدرة [(0.50)— (1.49)] مقارنة بفئة القدرة [(2.49-)— (1.5-)]، ومن ثمّ فئة القدرة [(1.50)

—(2.49)]، ومن ثمّ فئة القدرة [(-1.49)— (-0.5)]. كما أن الفروق الجوهرية بين الأوساط الحسابية لكمية معلومات الاختبار وفقًا لفئات القدرة قد كانت على الترتيب: لصالح فئة القدرة ما بين [(-1.49)— (-0.5)] مقارنة بفئة القدرة [(-2.49)— (-1.5)]، ومن ثمّ فئة القدرة [(1.50) —(2.49)]. ثمّ أن الفروق الجوهرية بين الأوساط الحسابية لكمية معلومات الاختبار وفقًا لفئات القدرة لصالح فئة القدرة [(1.50)— (2.49)] مقارنة بفئات القدرة [(-2.49)—(-1.5)].

ويتبين من جدول **(21)** وجود فروق دالة إحصائيًّا بين الأوساط الحسابية لمجموعات الدراسة وفئات القدرة تُعزى إلى التفاعل بينهما، ونظرًا لصعوبة تحديد مصادر الفروق الجوهرية بين الأوساط الحسابية وفق تفاعل متغيري المجموعات وفئات القدرة بالاعتماد على الرسم البياني في الشكل **(5)**، فقد تم استخدام اختبار بونفيروني للمقارنات البعدية الثنائية بين الأوساط الحسابية لمجموعات الدراسة



وفئات القدرة وفق التفاعل بينهما كما هو مبين في الجدول **(24).**

شكل 5: تفاعل متغيري المجموعات وفئات القدرة

جدول 23: اختبار بونفيروني للمقارنات البعدية الثنائية بين مجموعات الدراسة وفئات القدرة وفق التفاعل بينهما.

المتغير التَّابع: دالة المعلومات				
عندما تتراوح فئة القدرة بين [(-2.49)—(-1.50)]				
مجموعة الأداء التفاضلي	%30	%20	%10	
{Bonferroni}	الوسط الحسابي	4.53	4.59	5.46
%20	4.59	0.061		
%10	5.46	*0.927	*0.866	
صفر%	6.44	*1.908	*1.846	*0.98
عندما تتراوح فئة القدرة بين [(-1.49)—(-0.50)]				
مجموعة الأداء التفاضلي	%20	%30	%10	
{Bonferroni}	الوسط الحسابي	8.59	8.63	9.92
%30	8.63	0.044		
%10	9.92	*1.331	*1.288	
صفر%	10.25	*1.661	*1.618	*0.33
عندما تتراوح فئة القدرة بين [(-0.49)—(+0.49)]				
مجموعة الأداء التفاضلي	%20	%30	صفر%	
{Bonferroni}	الوسط الحسابي	11.44	11.45	12.46
%30	11.45	0.012		
صفر%	12.46	*1.02	*1.008	
%10	12.61	*1.176	*1.164	*0.156
عندما تتراوح فئة القدرة بين [(+0.50)—(+1.49)]				
مجموعة الأداء التفاضلي	صفر%	%20	%30	
{Bonferroni}	الوسط الحسابي	11.18	11.42	11.47
%20	11.42	*0.241		
%30	11.47	*0.295	0.054	
%10	11.55	*0.379	0.138	0.084
عندما تتراوح فئة القدرة بين [(+1.5)—(+2.49)]				
مجموعة الأداء التفاضلي	صفر%	%10	%20	
{Bonferroni}	الوسط الحسابي	6.54	7.41	8.20
%10	7.41	*0.868		
%20	8.20	*1.659	*0.791	
%30	8.38	*1.839	*0.97	0.179

* دال إحصائيًا (α=0.05)

يتضح من خلال الجدول السابق أنَّ الفروق الجوهرية بين الأوساط الحسابية لكمية معلومات الاختبار وفقًا لكل فئة من فئات القدرة والتفاعل مع مجموعات الدراسة قد كانت لصالح فئة القدرة بين

[(+0.50)—(+1.49)]، وفئة القدرة [(+1.5)—(+2.49)]، حيث يتضح بأنّ كمية المعلومات كانت أكبر عند المجموعات المستهدفة (10%، 20%، 30%) من المجموعة المرجعية. لذا طرأ تحسن في كمية المعلومات نتيجة إدخال نسبة مفردات ذات أداء تفاضلي (10%، 20%، 30%) عند فئات القدرة المرتفعة. ويتضح من خلال الجدول السابق وجود فروق دال احصائيًا (α=0.05) لكمية معلومات الاختبار عند فئات القدرة المرتفعة تُعزى لنسبة المفردات ذات الأداء التفاضلي (10%، 20%، 30%).

كما يلاحظ عدم وجود فروق دالة احصائيًا (α=0.05) لكمية معلومات الاختبار عند فئات القدرة المنخفضة تُعزى لنسبة المفردات ذات الأداء التفاضلي (10%، 20%، 30%).

مناقشة النتائج:

توصلت الدراسة إلى عدم وجود فروق ذات دلالة إحصائية بين دقة تقديرات قدرات الأفراد تُعزى إلى نسبة المفردات ذات الأداء التفاضلي (0%، 10%، 20%، 30%)، مما يُشير إلى أنّ دقة تقدير معلمة القدرة لا تختلف باختلاف نسبة المفردات ذات الأداء التفاضلي، وقد يُعزى ذلك إلى استقلالية القياس، حيث إنّ تقدير القدرات في نظرية الاستجابة للمفردة ثابت بتغير مفردات وظروف الاختبار وهو ما يُعرف بخاصية اللاتغاير (Invariance) والمتمثل بعدم تأثر تقدير قدرات الأفراد بمعالم مفردات الاختبار عندما أجري تغيير على معلمة صعوبة المفردة بمقدار (1.5) لكل مجموعة من المجموعات المستهدفة بنسب المفردات ذات الأداء التفاضلي (10%، 20%، 30%)، ويتفق ذلك مع دراسة الخطيب والعمري (2023). كما تتفق النتيجة مع ما توصل إليه النوافلة (2013)، بعدم تأثر تقدير قدرات الأفراد بنسب المفردات ذات الأداء التفاضلي **(10%، 20%)** عندما كان مقدار الأداء التفاضلي **(1.5)**، عند النموذج ثلاثي المعلمة. كما تتفق مع دراسة (Lee & Zhang,2016)، والتي

تشير إلى عدم وجود تأثير لنسب المفردات ذات الأداء التفاضلي (10%، 20%) على تقديرات القدرة عند النموذج ثنائي المعلمة. إلا أنها تختلف مع دراسة الوليلي و الدوسري (2019)، والتي توصّلت إلى نتيجة مفادها أن للأداء التفاضلي تأثير على دقة تقديرات القدرة، حيث أن حذف المفردات ذات الأداء التفاضلي أدى إلى انخفاض قيمة متوسط الخطأ المعياري في تقدير القدرة.

كما توصلت الدراسة إلى عدم وجود فروق دالة إحصائيًّا لمعلمة التمييز تُعزى إلى نسب المفردات ذات الأداء التفاضلي (10%، 20%، 30%)، ويرجع السبب في ذلك إلى عدم إجراء تغيير على مقدار معلمة التمييز عند جميع المجموعات المستهدفة. وتختلف الدراسة الحالية مع ما توصل إليه النوافلة بأن معلمة التمييز تنخفض بزيادة نسب المفردات ذات الأداء التفاضلي بمقدار أداء تفاضلي (1.5)، وقد يرجع السبب في ذلك إلى اعتماد الدراسة الحالية على الخطأ المعياري في دراسة أثر المفردات ذات الأداء التفاضلي على دقة تقدير معلمة التمييز، أما بالنسبة لدراسة النوافلة (2013)، فقد قارنت بين المتوسطات الحسابية لمعلمات التمييز عند كل مجموعة من مجموعات الدراسة، كما اختلفت الدراسة مع دراسة الشريفين (2018) والتي توصلت إلى وجود فروق دالة احصائيًا (α=0.05) بين الوسطين الحسابيين لمعلمة التمييز لصالح النموذج ثلاثي المعلمة عند نسبة مفردات ذات أداء تفاضلي (10%). وقد يرجع سبب الاختلاف إلى النموذج المستخدم في الدراسة، حيث تم استخدام النموذج ثلاثي المعلمة في كل من دراسة النوافلة والشريفين، بخلاف الدراسة الحالية والتي استندت النموذج ثنائي المعلمة.

كما توصلت الدراسة إلى عدم وجود فروق دالة إحصائيًّا (α=0.05) لمعلمة الصعوبة تُعزى إلى نسب المفردات ذات الأداء التفاضلي (10%، 20%)، حيث يوجد فروق دالة إحصائيًّا (α=0.05) لمعلمة

الصعوبة تُعزى إلى نسبة المفردات ذات الأداء التفاضلي (30%)، ويعود سبب ذلك إلى أنه تمّ إجراء تغيير في معلمة الصعوبة مقدارها **(1.5)** على المجموعات المستهدفة، إلا أنّ المجموعتين المستهدفة بنسب مفردات ذات أداء تفاضلي (10%،20%) لم تتأثر بمقدار التغيير، وتتفق الدراسة الحالية مع دراسة النوافلة (2013)، في أنه لم يطرأ تغيير على معلمة الصعوبة عندما كانت نسب المفردات ذات الأداء التفاضلي (10%،20%).

كما توصلت الدراسة إلى عدم وجود فروق ذات دلالة إحصائية (α=0.05) لمعاملات الثبات النظري والثبات التجريبي تُعزى لنسب المفردات ذات الأداء التفاضلي (10%، 20%،30%)، وأن جميع قيم معاملات الثبات النظري والتجريبي كانت عالية لدى مجموعات الدراسة. كما يُلاحظ أن أعلى قيمة للثبات النظري كانت عند المجموعة المستهدفة بنسبة مفردات ذات أداء تفاضلي **(10%)** مقارنة بالمجموعة المرجعية والمجموعات المستهدفة بنسبة مفردات ذات أداء تفاضلي **(20%،30%)**؛ ولأن الثبات يرتبط في حسابه على متوسط دالة المعلومات فقد يرجع السبب في ذلك إلى امتلاكها معلومات أكبر عند مستويات القدرة المرتفعة، بما يفيد تراجع مقدار خطأ القياس عند المجموعة المستهدفة بنسبة مفردات ذات أداء تفاضلي **(10%)**، نتيجة إقحام تلك المفردات ذات الأداء التفاضلي على بنية الاختبار.

كما أظهرت النتائج عدم وجود فروق ذات دلالة إحصائية (α=0.05) بين الأوساط الحسابية لكمية المعلومات وفقًا لمجموعات الدراسة تُعزى إلى نسبة المفردات ذات الأداء التفاضلي **(20%، 30%)**. ووجود فروق دالة احصائيًا (α=0.05) بين الأوساط الحسابية لكمية معلومات الاختبار وفقًا لمجموعات الدراسة لصالح المجموعة المستهدفة **(10%)** مقارنة بالمجموعة المرجعية.

ولكون متغيري مجموعة الدراسة وفئات القدرة ذا دلالة احصائية، فقد تمّ إجراء تحليل التباين ثنائي التفاعل على المتوسطات الحسابية لكمية المعلومات وفق متغيري مجموعة الدراسة وفئة القدرة، حيث أظهرت النتائج وجود فروق دالة احصائيًا (α=0.05) للمتوسطات الحسابية لكمية المعلومات وفقًا للتفاعل بين المتغيرين؛ لصالح المجموعات المستهدفة (10%،20%، 30%) وفئات القدرة المرتفعة. كما أظهرت النتائج وجود فروق دالة احصائيًا للمتوسطات الحسابية لكمية المعلومات وفقًا للتفاعل بين المتغيرين؛ لصالح المجموعة المستهدفة (10%) وفئات القدرة المتوسطة [(-0.49)—(+0.49)].وعدم وجود فروق دالة احصائيًا (α=0.05) للمتوسطات الحسابية لكمية المعلومات وفقًا للتفاعل بين المتغيرين؛ عند فئات القدرة المنخفضة [(-2.49)—(-1.50)]،[(-1.49)—(-0.50)].

ويستنتج من هذه الدراسة عدم وجود تأثير على دقة تقديرات قدرات الأفراد، ودقة تقدير معلمتي الصعوبة والتمييز ومتوسطات كمية معلومات الاختبار وفقًا لمجموعات الدراسة تُعزى إلى نسبة المفردات ذات الأداء التفاضلي (10%، 20%، 30%)، إلا أن التأثير كان على دقة تقدير معلمة الصعوبة عندما كانت نسبة المفردات ذات الأداء التفاضلي (30%)، ووجود تأثير على متوسطات كمية معلومات الاختبار تُعزى إلى التفاعل بين مجموعات الدراسة وفئات القدرة عندما كانت مستويات القدرة مرتفعة؛ لصالح المجموعات المستهدفة بنسبة مفردات ذات أداء تفاضلي (10%، 20%، 30%).

التوصيات:

اعتمادًا على الاستنتاجات التي تم التوصل إليها يمكن الخروج بجملة من التوصيات:

- تضمين نوع الأداء التفاضلي الغير منتظم ودراسة أثره على الخصائص السيكومترية للاختبار وبنوده باستخدام مختلف نماذج نظرية الاستجابة للمفردة.
- إجراء دراسة حول تأثير نسب المفردات ذات الأداء التفاضلي باستخدام النموذج أحادي المعلمة.
- دراسة تأثير نسب أخرى للأداء التفاضلي على الخصائص السيكومترية للاختبار وبنوده باستخدام نماذج نظرية الاستجابة للمفردة.
- اجراء أبحاث بنسب أعلى من (30%) للمفردات ذات الأداء التفاضلي مع الأداء التفاضلي الغير منتظم.

- **المراجع:**

أبو علام، رجاء محمود، والمرابحة، عامر جبريل. **(2015)**. أثر نسبة المفردات ذات الأداء التفاضلي وحجم العينة على قوة وفاعلية اختبار تحيز المفردة المتزامن: دراسة محاكاة. المجلة العربية للعلوم الاجتماعية، **(7)**2، **139-165.** **DOI:10.12816/0032962**

ضعضع، هبة عبداللطيف. **(2023)**. أثر اختلاف نموذج الاستجابة في الأداء التفاضلي للمفردة. *مجلة الأستاذ للعلوم* **(1PL,2PL,3PL)** للمفردة *الإنسانية والاجتماعية*، *62*(1)، 23-44. https://doi.org/10.36473/ujhss.v62i1.1964

علام، صلاح. **(2000)**. القياس والتقويم التربوي والنفسي أساسياته وتطبيقاته وتوجهاته المعاصرة. *دار الفكر العربي*.

محاسنة، إبراهيم محمد. **(2013)**. القياس النفسي في ظل النظرية التقليدية والنظرية الحديثة. دار جرير للنشر والتوزيع.

Ackerman, T. A. (1992). A didactic explanation of item bias, item impact, and item validity from a multidimensional perspective. Journal of Educational Measurement, 29, 67-91.
Alordia, C. O. (2021). A Comparison of three methods of detecting differential item functioning in dichotomous test items. ASSEREN journal of educational research and development,8,76-88. https://www.researchgate.net/publication/360317084
American Educational Research Association, American Educational Research Association, American Psychological Association, & National Council on (2014). Standards for Educational and Measurement in Education Psychological Testing: National Council on Measurement in Education. Washington DC: American Educational Research Association.

Apinyapibal, S., Lawthong, N., & Kanjanawasee, S. (2015). A comparative analysis of the efficacy of differential item functioning detection for dichotomously scored items among logistic regression, SIBTEST and Raschtree Methods. *Procedia - Social and Behavioral Sciences, 191*, 21–25. https://doi.org/10.1016/j.sbspro.2015.04.664

Baker, F. B., & Kim, S.-H. (2017). *The Basics of Item Response Theory Using R* (1st ed.). Springer Nature. https://doi.org/10.1007/978-3-319-54205-8

Belzak, W. C. M., & Bauer, D. J. (2020). Improving the assessment of measurement invariance: Using regularization to select anchor items and identify differential item functioning. **Psychological Methods, 25**(6), 673–690. https://doi.org/10.1037/met0000253

Berrío, Á. I., Gómez-Benito, J., & Arias-Patiño, E. M. (2020). Developments and Trends in Research on Methods of Detecting Differential Item Functioning. Educational Research Review, 31, 100340. https://doi.org/10.1016/j.edurev.2020.100340

Cai, L. (2008). SEM of another flavour: Two new applications of the supplemented EM algorithm. **British Journal of Mathematical and**

Statistical Psychology, 61(2), 309–329. https://doi.org/10.1348/000711007X249603

Carlson, J. E., & Davier, M. V. (2017). Item response theory. In R. E. Bennett & M. von Davier (Eds.), **Advancing human assessment: The methodological, psychological and policy contributions of ETS** (pp. 133–178). Springer Science + Business Media. https://doi.org/10.1007/978-3-319-58689-2_5

Carlson, J. E., & Davier, M. V. (2013). Item Response Theory. Research Report. ETS RR-13-28. ETS R&D Scientific and Policy Contributions Series. ETS SPC-13-05. *ETS Research Report Series*.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. **Psychological Bulletin, 56**(2), 81–105. https://doi.org/10.1037/h0046016

Chen, WH., Revicki, D. (2014). Differential Item Functioning (DIF). In: Michalos, A.C. (eds) Encyclopedia of Quality of Life and Well-Being Research. Springer, Dordrecht. https://doi.org/10.1007/978-94-007-0753-5_728

Cleary, T. A., & Hilton, T. L. (1968). An Investigation of Item Bias. *Educational and Psychological Measurement*, *28*(1), 61-75. https://doi.org/10.1177/001316446802800106

Davey, T., Nering, M. L. & Thompson, T.(1997). Realistic Simulation of item response data. *ACT research report series* *97*(4)1-36.

De Ayala, R. J. (2009). **The theory and practice of item response theory.** Guilford Press.

DeMars, C. E., Irwing, P., Hughes, D. J., & Booth, T. (2018). Classical Test Theory and Item Response Theory. In *The Wiley Handbook of Psychometric Testing* (pp. 49–73). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118489772.ch2

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). Lawrence Erlbaum Associates, Inc.

Donnelly, R. A. (2007). *The complete idiot's guide to statistics* (Second edition). Alpha

Downing, S. M., & Haladyna, T. M. (Eds.). (2006). *Handbook of test development.* Lawrence Erlbaum Associates Publishers.

Du Toit, M. (Ed.). (2003). IRT From SSI: BILOG-MG, MULTILOG, PARSCALE, TESTFACT. Scientific Software International.

Ellis, B. B. & Raju, N. S. (2003). Test and Item Bias: What They Are, What They Aren't, and How to Detect Them. EDUCATIONAL RESOURCES INFORMATION CENTER. https://files.eric.ed.gov/fulltext/ED480042.pdf

Finch, W.H., & French, B.F. (2019). Educational and Psychological Measurement (1st ed.). Routledge. https://doi.org/10.4324/9781315650951

Guttman, L. (1944). A basis for scaling qualitative data. **American Sociological Review, 9,(2),** 139–150. https://doi.org/10.2307/2086306

Hambleton, R. K., & Swaminathan, H. (1985). *Item response theory : principles and applications*. Klumer-Nijhoff.

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. Sage Publications.

Hambleton, R. K., & Zhao, Y. (2005). Item response theory (IRT) models for dichotomous data. *Encyclopedia of statistics in behavioral science*.

Han, K. T. (2007). WinGen: Windows software that generates item response theory parameters and item responses. **Applied Psychological Measurement, 31**(5), 457–459. https://doi.org/10.1177/0146621607299271

Hinkle, D. E., Wiersma, W., & Jurs, S. G. (2003). *Applied statistics for the behavioral sciences* (5th ed). Houghton Mifflin ; [Hi Marketing] (distributor). http://catalog.hathitrust.org/api/volumes/oclc/50716608.html

Holland, P. W., & Wainer, H. (Eds.). (1993). **Differential item functioning.** Lawrence Erlbaum Associates, Inc.

Hori, K., Fukuhara, H., & Yamada, T. (2020). Item response theory and its applications in educational measurement Part I: Item response theory and its implementation in R. *Wiley Interdisciplinary*

Reviews. Computational Statistics, 14(2), e1531-n/a. https://doi.org/10.1002/wics.1531

Ibrahim, A. w. (2017). Empirical Comparison of Three Methods for Detecting Differential Item Functioning in Dichotomous Test Items. Journal of Teaching and Teacher Education, 5(1), 1-17.http://dx.doi.org/10.12785/jtte/050101

Kim, S. H., Cohen, A. S., & Lin, Y. H. (2005). LDID: A Computer Program for Local Dependence Indices for Dichotomous Items.

Lee, Y. H., & Zhang, J. (2016). Effects of Differential Item Functioning on Examinees' Test Performance and Reliability of Test. International Journal of Testing,17(1), 23-54. https://doi.org/10.1080/15305058.2016.1224888

Lee, S., & Suh, Y. (2018). Lord's Wald Test for Detecting DIF in Multidimensional IRT Models: A Comparison of Two Estimation Approaches. **Journal of Educational Measurement**

Lim, H., Choe, E. M., & Han, K. T. (2022). A residual-based differential item functioning detection framework in item response theory. **Journal of Educational Measurement, 59**(1), 80-104. doi:10.1111/jedm.12313

Lord, F. M. (1977). PRACTICAL APPLICATIONS OF ITEM CHARACTERISTIC CURVE THEORY. *Journal of Educational Measurement, 14*(2), 117–138. https://doi.org/10.1111/j.1745-3984.1977.tb00032.x

Molenaar, I. W., & Sijtsma, K. (2000). MSP5 for Windows. A program for Mokken scale analysis for items. iecProGAMMA.

Ogunsakin, I. B., & Shogbesan, Y. O. (2018). Item response theory (IRT): a modern statistical theory for solving measurement problem in 21st century. *International Journal of Scientific Research in Education (IJSRE), 11*, 627-635.

Ohiri, S. C. , Christopher, M. O. & Benedict, I. C. (2024). Differential Item Functioning Detection Methods: An Overview. *International Journal of*

Research Publication and Reviews,5(2), 1555-1564. https://doi.org/10.55248/gengpi.5.0224.0505

Pae, T.-I., & Park, G.-P. (2006). Examining the relationship between differential item functioning and differential test functioning. Language Testing, 23(4), 475–496.https://doi.org/10.1191/0265532206lt338oa

Pashley, P. J. (1991). An alternative three-parameter logistic item response model. *ETS Research Report Series*, *1991*(1), i-35

PODSAKOFF, P. M., MACKENZIE, S. B., & PODSAKOFF, N. P. (2012). Sources of Method Bias in Social Science Research and Recommendations on How to Control It. *Annual Review of Psychology*, *63*(1), 539–569. https://doi.org/10.1146/annurev-psych-120710-100452

Reeve, B. B., Willis, G., Madans, J., Miller, K., & Maitland, A. (2011). Applying Item Response Theory for Questionnaire Evaluation. In *Question Evaluation Methods* (pp. 103–123). John Wiley & Sons, Inc. https://doi.org/10.1002/9781118037003.ch8

Secolsky, C., & Denison, D.B. (Eds.). (2012). Handbook on Measurement, Assessment, and Evaluation in Higher Education. Routledge.https://doi.org/10.4324/9780203142189

Selvi, H., & Özdemir Alici, D. (2018). Investigating the impact of missing data handling methods on the detection of differential item functioning. *International Journal of Assessment Tools in Education*, 1–14. https://doi.org/10.21449/ijate.330885

Sun, S., Fan, X., Matthes, J., Davis, C. S., & Potter, R. F. (2017). Item Response Theory. In *The International Encyclopedia of Communication Research Methods* (pp. 1–12). John Wiley & Sons, Inc. https://doi.org/10.1002/9781118901731.iecrm0129

Świst, K. (2015). Item analysis and evaluation using a four-parameter logistic model. *Edukacja*, *134*(3), 79-99.

Teo, T. (2013). Handbook of Quantitative Methods for Educational Research. Sense Publishers.

Van der Linden, W. J., (2005). Item response theory. In K. K. Leonard (Ed.), ***Encyclopedia of Social Measurement*** **(pp. 379-387) Elsevier Inc.**

Van der Linden, W. J. (2010). Item Response Theory. In P. Peterson, E. Baker, & B. McGaw (Eds.), **International Encyclopedia of Education, 3rd edition** (pp. 81-89). Elsevier

). Handbook of Item Response Theory VOLUME 8Van der Linden, W.J. (201 THREE Applications. https://doi.org/10.1201/9781315117430

Wells, C. S., & Faulkner-Bond, M. (Eds). (2016) Educational Measurement: From Foundation to Future. A Division of Guilford Publications.

Zumbo, B. D., (2007). Three Generations of DIF Analyses: Considering Where It Has Been, Where It Is Now, and Where It Is Going, Language Assessment Quarterly, 4:2, 223-233. https://doi.org/10.1080/15434300701375832