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 items With Differential Functioning On Psychometric properties Of Test And Its Items Using the Two-Parameter Logistic Model

رسالة مقدمة لاستكمال متطلبات الحصول على درجة الماجستير في علم النفس التربوي مسار (القياس والتقويم)

إعداد: رحاب بنت شعيب حمد الهميلي

الرقم الجامعي: 421200468
إشراف: أ.د/ محمد بن سليمان بن محمد الوطبان
أستاذ علم النفس التربوي بكلية اللغات والعلوم الإنسانية - جامعة القصيم
1446ه_2025م

شكر وتقدير

لله شكري، وعظيم امتناني، على جزيل نعمائه وكثير فضله لما يسّر لي من إتمام هذا البحث وإكماله، فله الحمد أولا وآخرًا وظاهرًا وباطنًا. واعترافًا منّي لذوي الفضل بفضلهم، لا يسعني في هذا المقام إلا أن أتقدم بجزيل الشكر، وعظيم العرفان إلى كل من تكرّم عليّ فأولاني نصيبًا من رعايته، وأعانني على تحمل مشاق البحث، وما أكثرهم، ولكن لابد أن أخص بالشكر العميق والثناء الجزيل بعضًا منهم وفي مقدمتهم الأستاذ الدكتور/ محمد بن سليمان الوطبان المشرف على هذه الرسالة والذي كان له بالغ الأثر في إخراج هذه الرسالة إلى حيز الوجود، فجزاه الله عني خير الجزاء. كما يطيب لي في هذا المقام أن أتقدم بجزيل الشكر والعرفان إلى الأستاذة الدكتورة / سهام رمضان عواد والتي أشرفت علي أثناء اعدادي لخطة البحث، والأستاذ الدكتور/ ربيع عبده أحمد رشوان على ما بذله من نصح وإرشاد أثناء كتابة الرسالة. كما وأتقدم بأسمى آيات الشكر والعرفان لأعضاء لجنة المناقشة؛ لتفضلهم بقبول مناقشة هذه الرسالة، وعلى ما بذلاه من جهد وعناء في قراءتها وتدقيقها وتصويبها من الأخطاء؛ من أجل الارتقاء بهذا العمل وخروجه بأفضل صورة وهما: الأستاذة الدكتورة / لمياء سليمان إبراهيم الفنيخ، والدكتورة نسرين محمد سعيد زارع، فجزاهما الله عني خير الجزاء. وأحقُّ من تُزجى إليه قوافل الشكر، أبي قدوتي وسندي والذي مهد لي طريق العلم والمعرفة وغرس في نفسي الجد والمثابرة حفظه الله ورعاه وأمد في عمره وألبسه ثوب الصحة والعافية. وإلى جنتي أُمي فلها عليّ من فضل دعائها ودعمها ما لا يُحصى، فكان لها الفضل بعد الله فيما وصلتُ إليه. كما أشكر زوجي ورفيق دربي فقد آزرني ووقف بجانبي وساعدني في إكمال مشواري العلمي فجزاه الله عني خير الجزاء. كما أشكر اخواني وأخواتي وبنات أختي وصديقاتي وأقاربي على ما بذلوه من صادق الدعاء، وجميل الثناء، وتهيئة الأجواء. والشكر موصول لكل من أرشدني على فهم البرامج الإحصائية وأخص بالشكر فني تحليل البيانات الأستاذ/ غسان عكور فجزاه الله عني خير الجزاء. وفي الختام اسأل الله العلي العظيم أن يجعل عملي خالصًا لوجهه الكريم، وأن ينفع به، فقد حرصت على إخراجه في أكمل صورة وأبهى حلة، فإن كان كذلك فبفضل من الله وحده، وما توفيقي إلا بالله عليه توكلت وإليه أُنيب.

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. وهدفت الدراسة إلى التحقق من تأثير نسب المفردات ذات الأداء التفاضلي (10%، 20%، 30%) على معلمتي الصعوبة والتمييز والثبات النظري والتجريبي ودالة المعلومات للاختبار. 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 بواقع (4) مجموعات. وتمّ الكشف عن المفردات ذات الأداء التفاضلي باستخدام برنامج IRTPRO من خلال طريقة (Wald Test). 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%).

ووجود فروق دالة احصائيًا (α=0.05) بين الأوساط الحسابية لكمية معلومات الاختبار وفقًا لكل فئة من فئات القدرة والتفاعل مع مجموعات الدراسة عند فئتي القدرة المرتفعة، لصالح نسبة المفردات ذات الأداء التفاضلي (10%، 20%، 30%).كما يوجد فروق دالة احصائيًا (α=0.05) بين الأوساط الحسابية لكمية معلومات الاختبار عند فئات القدرة المتوسطة تُعزى لنسبة المفردات ذات الأداء التفاضلي (10%) مقارنة بالمجموعة المرجعية (0%). وعدم وجود فروق دالة احصائيًا (α=0.05) بين الأوساط الحسابية لكمية معلومات الاختبار عند فئتي القدرة المنخفضة تُعزى لنسبة المفردات ذات الأداء التفاضلي (10%، 20%، 30%).

الكلمات المفتاحية: نسب المفردات ذات الأداء التفاضلي، الخصائص السيكومترية، معلمة الصعوبة، معلمة التمييز، الثبات النظري، الثبات التجريبي، دالة المعلومات للاختبار، النموذج اللوجستي ثنائي المعلمة.

الفصل الأول
المدخل إلى الدراسة

- **المقدمة.**
- **مشكلة الدراسة.**
- **أهداف الدراسة.**
- **أهمية الدراسة.**
- **مصطلحات الدراسة.**
- **حدود الدراسة.**

المقدمة:

ظهرت الاختبارات والمقاييس 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 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 Ogunsakin & (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 أربعة نماذج لنظرية الاستجابة للمفردة بالنسبة للمفردات ثنائية الاستجابة، وتختلف هذه النماذج في عدد المعالم، وهي: معلمة الصعوبة (b)، ومعلمة التمييز (a)، ومعلمة التخمين (c)، ومعلمة الإهمال (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 (2PL)، 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 التي تتجاوز قدرته بشكل صحيح (Teo, 2013). بالإضافة الى المعالم السابقة تأتي معلمة رابعة في النموذج اللوجستي رباعي المعلمة The Four-Parameter Logistic (4PL) Model وهي معلمة الإهمال على افتراض أن الممتحنين ذوي الكفاءة العالية من الممكن أن يرتكبوا أخطاء بسبب الإهمال (Świst, 2015).

واتجه اهتمام الباحثين ومطوّري الاختبارات في نظرية الاستجابة للمفردة الاختبارية على الخصائص السيكومترية للاختبار التي تشتمل على مفهوم الثبات والذي يدل على "مدى خلو درجات أداة القياس أو التقدير من الأخطاء العشوائية أو غير المنتظمة" (تيغزة، 2009، ص 640). ومن المؤشرات التي يستدل منها على معامل الثبات في نظرية الاستجابة للمفردة الاختبارية دالة المعلومات للاختبار والخطأ المعياري في القياس (الشريفين،2018). ووظّفت نظرية الاستجابة للمفردة الاختبارية 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).

وفي نظرية الاستجابة للمفردة الاختبارية تلعب تقنيات محاكاة الكمبيوتر أدوارًا رئيسية في الممارسات البحثية، وتتمثل إحدى الميزات المكتسبة في استخدام البيانات المولّدة 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 المحكمين، 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 بمفهومه الاجتماعي ولكن استخدم بدلًا منه مصطلح 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 يُدعّم 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 al., 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).

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 مثل: (دراسة النوافلة، 2013؛ ودراسة الخطيب، 2014؛ ودراسة اجبارة، 2015؛ ودراسة أبو علام والمرابحة، 2015؛ ودراسة Lee and Zhang, 2016؛ ودراسة الحربي والمدني، 2021؛ ودراسة الخطيب والعمري، 2023) 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%- 70%)، 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؛ حيث 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 (2013)، 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 (2013)، 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 (10%، 20%، 30%) على الخصائص السيكومترية للاختبار ومفرداته وفق النموذج ثنائي المعلمة؟

ويتفرع منه الأسئلة التالية:

السؤال الأول: هل يوجد تأثير لاختلاف نسب المفردات ذات الأداء التفاضلي (10%، 20%، 30%) على دقة تقديرات القدرات باستخدام النموذج ثنائي المعلمة.

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

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

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

أهداف الدراسة:

- التحقق من مدى تأثير نسب المفردات ذات الأداء التفاضلي (10%، 20%، 30%) على دقة تقدير القدرات باستخدام النموذج ثنائي المعلمة.
- التحقق من مدى تأثير نسب المفردات ذات الأداء التفاضلي (10%، 20%، 30%) على دقة تقديرات معالم المفردات للاختبار (الصعوبة والتمييز) باستخدام النموذج ثنائي المعلمة.
- التحقق من مدى تأثير نسب المفردات ذات الأداء التفاضلي (10%، 20%، 30%) على قيم معاملات الثبات النظري والثبات التجريبي باستخدام النموذج ثنائي المعلمة.
- التعرف على مدى تأثير نسب المفردات ذات الأداء التفاضلي (10%، 20%، 30%) على قيم دالة المعلومات للاختبار باستخدام النموذج ثنائي المعلمة.

أهمية الدراسة:

الأهمية النظرية:

- مساعدة الباحثين ومطوّري الاختبارات في معرفة تأثير النسب المختلفة للمفردات ذات الأداء التفاضلي على معالم الفقرات، المتمثلة في الدراسة الحالية بمعلمتي الصعوبة والتمييز.
- 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.
- لفت انتباه الباحثين عند اشتمال اختباراتهم على مفردات ذات أداء تفاضلي على اختلاف نسبتها، في إمكانية إبقاء الفقرات، أو استبعادها للحصول على نتائج أفضل.

الأهمية التطبيقية:

- 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.

التعريفات الاصطلاحية والإجرائية:

- **الأداء التفاضلي للمفردة Differential Item Functioning:** مؤشر إحصائي يُستخدم لتحديد مدى اختلاف أداء مجموعات مختلفة من المتقدمين للاختبار على بند معين، رغم تماثلهم في مستوى القدرة أو السمة المقاسة (American Educational Research Association et al., 2014, p. 218).
- **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، ص. 142). 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، ودالة المعلومات للمفردة، بالاعتماد على النموذج اللوجستي ثنائي المعلمة".(الشريفين،2018، ص615) وهي كالتالي:
  - **معلمة الصعوبة Difficulty 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، ص ص186-187).
  - **معلمة التمييز Slope 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 نقطة انعطاف منحنى الخصائص للفقرة" (الشريفين،2018، ص615).

- **دالة المعلومات للمفردة Item Information Function**: "تعتمد على ميل دالة استجابة البند والتباين الشرطي عند كل مستوى من مستويات القدرة (θ) ".(Hambleton & Swaminathan,1985, p. 105))
- **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"(الشريفين،2018، ص.615).
- **دالة المعلومات للاختبارTest Information Function:** "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, P. 157)
- **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 P. 33)
- **الثبات التجريبي Empirical Reliability:** "هو تباين الدرجة الحقيقية مقسومًا على مجموع تباين الدرجة الحقيقية وتباين الخطأ". (Du Toit, 2003, P. 34)
- **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". (التقي، 2009، ص ص21-22)

حدود الدراسة:

- اقتصرت الدراسة الحالية على ثلاثة نسب للمفردات التي تبدي أداءًا تفاضليًّا، وهي: (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.

الفصل الثاني

أدبيات الدراسة

أولًا: نظرية الاستجابة للمفردة

- المقدمة
- افتراضات النظرية
- نماذج النظرية
- خصائص نماذج النظرية
- الخصائص السيكومترية في نظرية الاستجابة للمفردة

ثانيًا: الأداء التفاضلي

- مقدمة
- الأداء التفاضلي وعلاقته بمفهوم التحيز
- أنواع الأداء التفاضلي للبنود
- الأداء التفاضلي والبيانات المولّدة
- الكشف عن الأداء التفاضلي

ثالثًا: الدراسات السابقة

- الدراسات السابقة
- التعقيب على الدراسات السابقة
- فروض الدراسة

■ **أولاً: نظرية الاستجابة للمفردة Item Response Theory (IRT):**

ظهرت نظرية الاستجابة للمفردة نتيجة التطورات في القياسات النفسية، التي عُرفت لأول مرة باسم نظرية السمات الكامنة Latent Trait Theory، التي يمكن استخدامها في تحليل كل من البيانات ثنائية الاستجابة ومتعددة الاستجابة (Wells & Faulkner-bond, 2016). وعلى الرغم من أن نظرية الاستجابة للمفردة (IRT) لها جذورها في عدد قليل من الأبحاث المنشورة في أربعينيات القرن العشرين، إلا أن الفكرة الأصيلة لنموذج الاستجابة يرجع إلى طريقة القياس المطلق التي طورها ثورستون في منتصف عشرينيات القرن العشرين، وتعتمد طريقة ثورستون على افتراض وجود مقياس كامن يقوم على الاستجابات التي يمكن من خلالها رسم خريطة لكل من المفردات والأفراد (Van der linden, 2005).

كما ساهم جتمان Guttman (1944) بالجذور التاريخية لنظرية الاستجابة للمفردة عندما أطلق خط تتبع المفردة الذي يُعرف حاليًا بمنحنى خصائص المفردة. وفي عام 1952 كتب فريد لورد Fred Lord أطروحة نُشرت كدراسة نفسية وضعت الأسس النظرية لنظرية الاستجابة للمفردة (IRT)، من خلال ما ناقشه من منحنيات خصائص المفردة، ومنحنى خصائص الاختبار، وكذلك افتراض الاستقلال الموضعي، وثبات المعالم التي لا تعتمد في توزيعها على قدرات المفحوصين في الاختبار (Carlson & Von davier, 2017).

كما يرجع الفضل إلى خدمات الاختبارات التربوية التي تأسست في عام 1947، حيث كان باحثيها من طليعة المساهمين في تطوير النظرية بدءًا من العمل المبكر الذي قام به تاكرTucker ولورد Lord في الأربعينيات والخمسينيات من القرن العشرين، ووصولًا إلى القرن الحادي والعشرين عندما واصل الباحثين مساهمتهم في تطوير نظرية الاستجابة للمفردة من خلال نماذج النظرية، أمثال هابرمان Haberman، وسينهاراي Sinharay، وفان ريجن van Rijn، وفون ديفير von Davier ، وريجمين Rijmen (Carlson & Von davier, 2013).

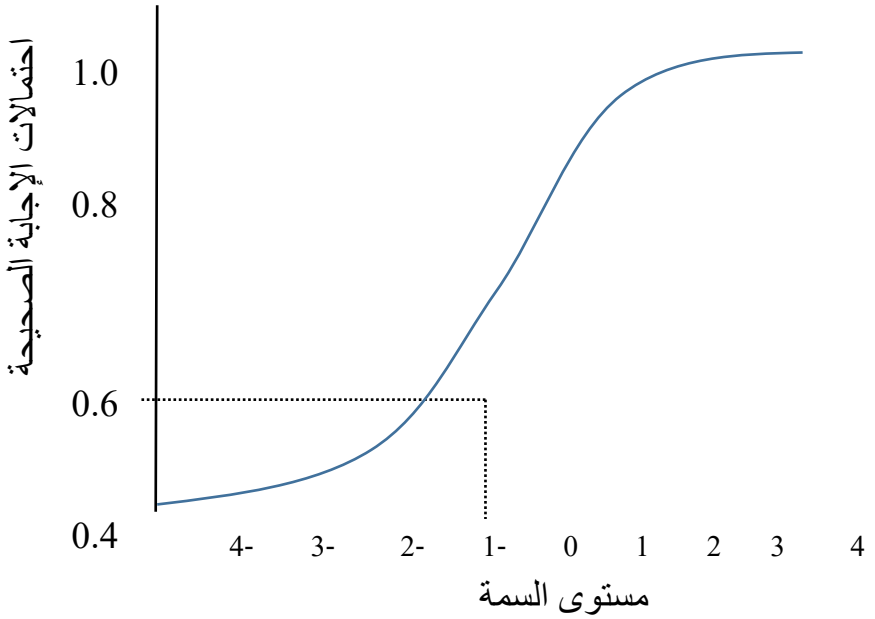
وتفترض نظرية الاستجابة للمفردة أنه يمكن التنبؤ بأداء الممتحنين في الاختبار من خلال خصائصهم التي يُشار إليها بالقدرات أو السمات، ونظرًا لأن هذه السمات لا يمكن قياسها مباشرة، فإن نموذج الاستجابة للمفردة يقيس هذه السمات من خلال ملاحظة استجابة الممتحنين على مفردات الاختبار التي تقيس تلك السمة. وتقدير الدرجات للمفحوصين على هذه السمات تسمى بدرجات القدرة، وتستخدم هذه الدرجات للتنبؤ بأداء المفحوصين على مفردات الاختبار (Lord & Novick, 1968 as cited Hambleton & Swaminathan, 1985).

وتقوم نظرية الاستجابة للمفردة على افتراض وجود سمة أو قدرة كامنة، يُرمز لها بالرمز θ،كما تفترض نماذج النظرية وجود علاقة محددة بين القدرة θ، والاستجابة على مفردات الاختبار(Demars, et al., 2018). وتعتبر نظرية الاستجابة للمفردة خيارًا عمليًّا في القياس النفسي والتربوي نظرًا لتحسن طرق التقدير والخوارزميات، وكذلك التحسن الكبير في أداء الكمبيوتر(Hori et al., 2020).

وهناك مجموعة من الافتراضات التي تستند عليها نظرية الاستجابة للمفردة، للوصول إلى نتائج بيانات دقيقة وموثوقة، وذلك من أجل الاستخدام الصحيح لنماذج النظرية في تحليل بيانات الاختبارات النفسية والتربوية. (حبيش ولكحل وبراهيمي، 2022). **وأهم هذه الافتراضات هي:**

- **أُحادية البُعد Unidimensionality:** بمعنى أن مفردات الاختبار تقيس سمة كامنة واحدة مستمرة، ولا يمنع أن تكون المفردات لها أبعاد ثانوية فرعية، إلا أنه يفترض أنّ بُعدًا واحدًا مهيمنًا يُفسر البنية الأساسية. ويمكن تقييم افتراض أُحادية البُعد للمقياس من خلال إجراء تحليل عاملي على مستوى المفردة الذي تمَّ تصميمه لتقييم بنية العامل التي تكمن وراء التباين المشترك الملحوظ بين استجابات العناصر، وفي حالة وجود أبعاد متعددة يمكن للباحث أن يستخدم نماذج نظرية الاستجابة للمفردة متعددة الأبعاد (Reeve, et al., 2011).
- **الاستقلال الموضعي Local Independence:** ويسمى أيضًا الاستقلال المشروط حيث تكون الاستجابات مشروطة فقط بالسمات الكامنة (Sun et al., 2017). وينص هذا الافتراض على أن استجابات الممتحن لمفردات الاختبار مستقلة إحصائيًّا؛ بمعنى أن أداء الممتحن في مفردة معينة لا يؤثر على إجاباته في مفردات أخرى في الاختبار، بحيث أن محتوى المفردة لا يوفر أدلة على إجابات المفردات الأخرى في الاختبار، سواء كان للأفضل أو الأسوأ. بالإضافة إلى أن ترتيب عرض مفردات الاختبار يجب أن لا يؤثر على أداء الاختبار (Hambleton & Swaminathan,1985). وهذه الخاصية هي أساس دالة الاحتمالية المستخدمة لتقدير مواقع المفحوصين المتقدمين للاختبار على المقياس (Carlson & Von davier, 2013).
- **التحرر من السرعة في الأداء Non-Speedness:** الافتراض الضمني لجميع نماذج نظرية الاستجابة للمفردة شائعة الاستخدام هو أن الاختبارات التي تناسب النماذج لا يتم إجراؤها في ظل ظروف سريعة؛ أي أنّ الممتحنين الذي يفشلون في الإجابة على مفردات الاختبار يفعلون ذلك بسبب محدودية قدراتهم، وليس لأنهم فشلوا في الوصول إلى الإجابة على مفردات الاختبار. ونادرًا ما يتم ذكر هذا الافتراض لأنه ضمني في افتراض أُحادية البُعد، فعندما تؤثر السرعة على أداء الاختبار، فإنّ ثمة سمتين تؤثران على ذلك هما: سرعة الأداء، والسمة التي يتم قياسها من خلال محتوى الاختبار. ويمكن تقييم مدى سرعة الاختبار بشكل تقريبي من خلال حساب عدد الممتحنين الذين فشلوا في إكمال الإجابة على مفردات الاختبار(Hambleton & Swaminathan,1985).

منحنى خصائص المفردة Item Characteristics Curve (ICC): دالة رياضية تصف العلاقة بين احتمالية الإجابة الصحيحة على بند معين ومستوى القدرة الذي يقيسه الاختبار، وتأخذ شكل منحنى الترجيح اللوغاريتمي (اللوجستي) Logistic Curve الذي يسمى المنحنى المميز للمفردة (ICC). ويوضح شكل المنحنى كيفية تغيُّر مستوى السمة في احتمالات الإجابة على مفردة ما، حيث يعبّر المنحنى المميز للمفردة عن احتمالية الإجابة الصحيحة كدالة على مستوى السمة أو القدرة المقاسة(Hambleton & Swaminathan,1985)، ويمكن تمثيل المنحنى المميز للمفردة كما في الشكل (1).



شكل 1: مثال على منحنى خصائص المفردة.

يتضح من الشكل أن المنحنى يزداد طرديًا على طول متصل القدرة؛ بمعنى أنه كلما زادت قدرة الفرد زادت احتمالية أن يجيب على المفردة إجابة صحيحة.

ويتضح الفرق بين كل نموذج من نماذج نظرية الاستجابة للمفردة من خلال الشكل الرياضي لمنحنى خصائص المفردة (ICC). وإذا كانت درجات المادة والقدرة لمجموعة من الممتحنين معروفة، فمن الممكن الكشف عن شكل منحنى خاصية المفردة من خلال النظر في توزيع درجات المادة عند مستويات ثابتة من القدرة (Hambleton & Swaminathan,1985).

- **نماذج نظرية الاستجابة للمفردة:**

إنّ الميزة الرئيسية لنماذج نظرية الاستجابة للمفردة هي خاصية الثبات لمعالم المفردات والقدرة، وثبات المعلمة يعني أن قيم معالم المفردات التي تتمثل في (الصعوبة، والتمييز، والتخمين)، ومعلمة القدرة لا تعتمد على عينة المفحوصين أو عينة المفردات التي بُني عليها الاختبار؛ بمعنى أنه عند إعطاء الاختبار لمجموعة مختلفة من المفحوصين فإن خصائص المفردات تكون ثابتة، وكذلك بالنسبة إلى تقديرات قدرات المفحوصين عند تغير مفردات الاختبار Wells & Faulkner-bond, (2016).

وتتعدد نماذج نظرية الاستجابة للمفردة سواء كانت (أُحادية البُعد، متعددة الأبعاد) أو (ثنائية الاستجابة، متعددة الاستجابة) (Demars, et al,. 2018). وهنا سوف يأتي ذكر النماذج أُحادية البُعد ثنائية الاستجابة التي تتناول هذه الدراسة واحدًا منها وهو: النموذج ثنائي المعلمة.

- **النموذج أحادي المعلمة (One-Parameter Logistic Model (1-PL:** يعرف باسم نموذج راش، وهو نموذج يدمج بين معلمة الصعوبة للمفردة ويرمز لها بالرمز b ، وكذلك معلمة قدرة الفرد θ التي تشير إلى قدرة الممتحن على الاستجابة لمفردات الاختبار بشكل صحيح. ويفترض النموذج تثبيت معامل التمييز عند قيمة (1.0) لجميع البنود؛ ويمكن فقط لمعامل الصعوبة أن يأخذ قيمًا مختلفة (baker & kim, 2017). ففي المعادلة التالية فإنّ خيار الرقم (1) يعني الإجابة عن المفردة بشكل صحيح، وخيار الرد (0) يعني الإجابة عن المفردة بشكل خاطئ، وتمثل معلمة الصعوبة للمفردة

مستوى السمة المطلوب الذي يكون لدى الفرد فيه فرصة بنسبة 50٪ للإجابة عن مفردة ما بشكل صحيح، لذا كلما ارتفعت قيمة المعلمة b، كلما كانت الإجابة عن المفردات أكثر صعوبة بالنسبة للممتحنين.

ويتمثل النموذج في المعادلة الرياضية التالية:

$$p(X_{is} = 1|\theta_s, b_i) = \frac{\exp(\theta_s - b_i),}{1 + \exp(\theta_s - b_i)}, i = 1, 2, \dots, n$$

X_{is} : وتعني استجابة المفحوصين للمفردة مع خيارات الاستجابة (1،0).
θ_s : وتعني قدرة المفحوص على السمة الكامنة التي يقيسها الاختبار.
b_i : وتعني مستوى صعوبة المفردة.
$(X_{is} = 1|\theta_s, b_i)$: يُشير إلى احتمال قيام فرد واحد بمستوى السمات θs بالاستجابة بشكل صحيح في الاتجاه المتوافق مع السمات (Sun et al., 2017).

- **ثانيًا: النموذج ثنائي المعلمة (Two-Parameter Logistic Model (2-PL:**

يتضمن هذا النموذج معلمة الصعوبة والتمييز، وتتناسب معلمة التمييز مع ميل منحنى خاصية المفردة عند مستوى الصعوبة على طول سلسلة السمات. وكلما زاد حجم معلمة التمييز زادت قوة المفردة في الفصل بين القدرة المرتفعة والقدرة المنخفضة. ومن الناحية النظرية، يمكن أن تتراوح معلمة التمييز من اللانهاية السالبة إلى اللانهاية الموجبة. والقيمة السالبة لمعلمة التمييز تعني انخفاض احتمالية الاستجابة بشكل صحيح عن مفردة ما مع زيادة القدرة، مما يدل على وجود خطأ أو خلل في المفردة (Hambleton et al., 1991). ومن الناحية العملية تتراوح قيم معلمة التمييز الجيدة من 0.8 إلى 2.5 (De Ayala, 2009).

ويتم تمثيل النموذج رياضيًا بالمعادلة التالية:

$$p(X_{is} = 1|\theta_s, b_i, \alpha_i) = \frac{\exp[\alpha_i(\theta_s - b_i)]}{1 + \exp[\alpha_i(\theta_s - b_i)]}, i = 1, 2, \dots, n$$

X_{is} : وتعني استجابة المفحوصين للمفردة مع خيارات الاستجابة (1،0).
θ_s : وتعني قدرة المفحوص على السمة الكامنة التي يقيسها الاختبار.
a_i : قوة التمييز للمفردة.
b_i : وتعني السمات الكامنة لدى المفحوصين (Sun et al., 2017).

- **النموذج ثلاثي المعلمة (Three-Parameter Logistic Model (3-PL:**

قام بيرنبوم Birnbaum في عام (1968) بتعديل النموذج اللوجستي ذو المعلمتين ليشمل معلمة التخمين التي تمثل الحد الأدنى لاحتمالية الإجابة الصحيحة عن المفردة نتيجة التخمين، خاصةً بين الأفراد ذوي القدرة المنخفضة. إلا أنه تم فقدان بعض الخصائص الرياضية للدالة اللوجستية، ومع ذلك أصبح النموذج الناتج يعرف بالنموذج اللوجستي ثلاثي المعلمة (Pashley, 1991)

ومعلمة التخمين c لا تختلف كدالة لمستوى القدرة. وبالتالي فإن المفحوصين من ذوي القدرة المنخفضة يكون لديهم احتمالية الحصول على الإجابة الصحيحة عن طريق التخمين. وتحتوي معلمة التخمين على نطاق نظري $c \leq 1.0$ ، ولكن في القيم العملية القيمة الأعلى من 0.35 لا تعتبر مقبولة(baker & kim, 2017).

ويتمثل النموذج في المعادلة الرياضية التالية:

$$p(X_{is} = 1|\theta s, \theta_s, b_i, \alpha_i, c_i) = c_i + (1 - c_i)\frac{\exp[\alpha i(\theta_s - b_i)]}{1 + \exp[\alpha_i(\theta_s - b_i)]}, i = 1, 2, \dots, n$$

X_{is} : وتعني استجابة المفحوصين للمفردة مع خيارات الاستجابة (0،1).
θ_s : وتعني السمات الكامنة لدى المفحوصين.
a_i : قوة التمييز للمفردة.
b_i : مستوى الصعوبة للمفردة.
c_i : عامل التخمين العشوائي للمفردة (Sun et al., 2017).

- **النموذج رباعي المعلمة (4-PL) Four-Parameter Logistic Model :**

هو امتداد للنماذج السابقة حيث يضيف معلمة جديدة لتحليل استجابات المفحوصين وهي معلمة الإهمال أو الخطأ؛ وتمثل الحد الأعلى لاحتمال الإجابة الصحيحة، مما يأخذ في الاعتبار أن المفحوصين من ذوي القدرة المرتفعة قد يخطئون بسبب الإهمال.

$$p(u_{ij}|\theta_i; a_j, b_j, \alpha_j, c_i, d_i) = (1 - d_i)c_j + d_i\,(c_j + 1 - c_j)\frac{e^{1.7\,a_j(\theta_i - b_j)}}{1 + e^{1.7\,a_j(\theta_i - b_j)}}$$

$$= c_j + d_i(1 - c_j)\frac{e^{1.7\,a_j(\theta_i - b_j)}}{1 + e^{1.7\,a_j(\theta_i - b_j)}}$$

α_{ij}: وتعني استجابة المفحوص للمفردة مع خيارات الاستجابة (0،1).
θ_i : وتعني السمات الكامنة لدى المفحوصين.
α_j: معلمة التمييز.
b_j: معلمة الصعوبة
c_j: معلمة التخمين.
d_i: معلمة الإهمال(Świst, 2015).

- **خصائص نماذج نظرية الاستجابة للمفردة:**
  - يفترض نموذج الاستجابة للمفردة أن أداء المفحوص في الاختبار يمكن التنبؤ به من خلال واحدة أو أكثر من الخصائص المعروفة بالسمات.
  - يحدد نموذج الاستجابة للمفردة العلاقة بين أداء المفردة الذي يمكن ملاحظته والسمات أو القدرات التي يفترض أنها تشكل أساس الأداء في الاختبار.
  - يعد نموذج الاستجابة للمفردة الذي يظهر ملاءمة جيدة للبيانات أداة فعالة لتقدير درجات المفحوصين على السمات الكامنة المستهدفة.
  - لا تقاس السمات الكامنة بشكل مباشر، بل يتم تقديرها أو استنتاجها من خلال تحليل أنماط الاستجابة الظاهرة على مجموعة من مفردات الاختبار (Hambleton & Swaminathan, 1985).

- **الخصائص السيكومترية في نظرية الاستجابة للمفردة:**

تشمل الخصائص السيكومترية على كل من معلمة القدرة، ومعالم المفردات التي تُمثل خصائص المفردة وهي: الصعوبة، والتمييز، والتخمين، ومعاملات الثبات النظري والتجريبي، ودالة المعلومات للمفردة والاختبار.

- **معلمة القدرة Ability Parameter:** تمثل قدرة الشخص أو مهارته أو سمة من سماته الكامنة ويتم قياسها من خلال استجاباته على مفردات الاختبار، ويرمز لها بالرمز ثيتا θ (Van der Linden, 2010) .
- **معالم المفردات:**
  - **معلمة الصعوبة Difficulty Parameter :** يُشار إليها بالرمز b، وتحدد المستوى المطلوب من السمة الكامنة (θ)ليكون احتمال الإجابة الصحيحة على المفردة 50%، بحيث كلما ارتفع مستوى صعوبة المفردة، زادت القدرة أو السمة الكامنة المطلوبة للإجابة عن هذه المفردة إجابة صحيحة (Hambleton & Zhao, 2005).
  - **معلمة التمييز Slope parameter** :خاصية تصف المفردة، ويشار إليها بالرمز α وتمثل قدرة المفردة على التمييز بين الأفراد من ذوي السمات المختلفة، وتأخذ قيمًا نظرية تتراوح من (-∞ إلى ∞) وعمليًا من (-3 إلى 3)، ويستدل

عليها من خلال شدة انحدار أو ميل المماس لمنحنى خصائص المفردة (ICC) عند نقاط انعطاف منحنى الخصائص للفقرة (الشريفين، 2018).

- **معلمة التخمين (Pseudo-chance (guessing parameter:** تمثل الحد الأدنى لاحتمالية الإجابة الصحيحة على المفردة بواسطة المفحوصين ذوي القدرة المنخفضة، ويُشار إليها بالرمز c وتعرف بمعلمة الفرصة الزائفة، وهي تمثل احتمالية قيام المفحوصين من ذوي القدرات المنخفضة بالإجابة عن المفردة إجابة صحيحة Secolsky & Denison, (2012).

- **الخصائص السيكومترية للاختبار:**

- **الثبات النظري Theoretical Reliability :** تنطبق قيمة الثبات النظري على درجات نظرية الاستجابة للمفردة المقدرة باستخدام طريقة الاحتمالية القصوى في برنامج Bilog-Mg، ويعتمد في تقديره على معالم المفردة فقط التي تم تمريرها من المرحلة الثانية، ولا يعتمد على درجات القدرة المحسوبة في المرحلة الثالثة، وبدلًا من ذلك يفترض أن درجات القدرة الحقيقية يتم توزيعها طبيعيًّا بمتوسط صفر وتباين واحد في مجتمع المفحوصين.
- **الثبات التجريبي Empirical Reliability :** تختلف صيغ تقدير الخطأ وتباين الدرجات الحقيقية لحساب الثبات التجريبي اعتمادًا على كيفية تقدير درجات قدرة المفحوصين في العينة، **ويتم ذلك من خلال طريقتين:**

الطريقة الأولى: من خلال درجات الاحتمالية القصوى، حيث إنّ تباين الخطأ المقدر هو مقلوب متوسط أعلى كمية للمعلومات التي تم تقييمها وفقًا لتقديرات قدرة جميع المفحوصين في العينة. وتباين النتيجة هو مجرد تباين للحد الأقصى لدرجات الاحتمالية للعينة، وبالتالي يمكن تقدير تباين النتيجة الحقيقية عن طريق طرح تباين الخطأ من تباين النتيجة، ويتم بعد ذلك إعطاء الثبات التجريبي في كل عينة من خلال قيمة تباين النتيجة الحقيقية مقسومة على تباين النتيجة.

الطريقة الثانية: من خلال درجات بايز، حيث إنّ تقدير تباين الخطأ هو متوسط تباينات قدرة التوزيع الخلفي لجميع المفحوصين في العينة نظرًا لأن درجات القدرة عبارة عن تقديرات متراجعة في تقدير بايز، ويتم تقدير تباين النتيجة الحقيقية مباشرة من خلال تباين درجات التوزيع الخلفي في العينة. وبالتالي فإن الثبات التجريبي هو تباين النتيجة الحقيقية مقسومًا على مجموع تباين النتيجة الحقيقية وتباين الخطأ(Du Toit, 2003) .

- **دالة المعلومات للمفردة والاختبار:**

تمثل دالة معلومات الاختبار تمثل مدى دقة الاختبار في قياس القدرة الكامنة. ودالة معلومات الاختبار هي ببساطة مجموع دوال معلومات المفردات في الاختبار، حيث يكون منحنى معلومات الاختبار في أعلى مستوياته عند النقطة التي تتركز فيها القدرة التمييزية لمفردات الاختبار. وارتفاع المنحنى عند درجة معينة يشير إلى مقدار المعلومات التي يوفرها الاختبار حول المفحوصين الذين تقع قدراتهم عند تلك الدرجة، ويقابل ذلك انخفاض في خطأ القياس. لذا فإن النقاط التي يكون فيها المنحنى مرتفعًا تمثل درجات يُحتمل أن تكون قريبة من القدرة الحقيقية للمفحوص، بينما تمثل النقاط المنخفضة درجات مصحوبة بخطأ قياس أكبر. ودالة معلومات المفردة تتمثل في مدى قدرة المفردة على التمييز بين المفحوصين من ذوي القدرات المتماثلة على مستويات مختلفة من القدرة. ويفترض أن يعكس شكل منحنى معلومات الاختبار الغرض الذي تم تصميم الاختبار من أجله (Downing & Haladyna, 2006). ويعتقد الكثير من الباحثين أن التركيز يكون على نقطة القدرة الأعلى وهذا ليس صحيح دائما؛ فمثلا: اختبار تشخيصي يجب أن يركز معلوماته عند القدرة المنخفضة، بينما اختبار تحصيلي معياري يُفضل أن يتركز عند المستوى المتوسط أو الأعلى.

واقترح بيرنباوم Birnbaum إجراءً من أربع خطوات لبناء الاختبار، استنادًا إلى منحنيات معلومات المفردات:

أولًا: تحديد الشكل المطلوب لدالة معلومات الاختبار، الذي يسميه لورد منحنى المعلومات المستهدف.
ثانيًا: تحديد المفردات التي تحتوي على منحنيات معلومات المفردة التي تملأ الفراغات الموجودة أسفل منحنى المعلومات المستهدف.
ثالثًا: الاستمرار في إضافة منحنيات معلومات المفردة، ومراقبة دالة معلومات الاختبار الفرعي الناتجة من المفردات المحددة بالفعل.
رابعًا: الاستمرار في إضافة (وربما طرح) المفردات حتى يتم تقريب منحنى المعلومات المستهدف بشكل مُرضٍ Lord, (1977).

ويستخدم هذا الإجراء العلاقة بين دالة معلومات المفردة والاختبار، ويضمن اتباع هذا الإجراء أن المفردات المختارة لنموذج الاختبار كافية لإنجاز المهمة التي تم تصميم الاختبار من أجلها. بالإضافة إلى أن الاختبار يُظهر دقة أكبر في الدرجات وثبات أعلى من الاختبار الذي تم إنشاؤه عن طريق اختيار المفردات عشوائيًا(Downing & Haladyna, 2006). ومن الناحية العملية، يكون الهدف عادةً هو دفع دالة معلومات الاختبار إلى ما لا يقل عن 10، وهو ما يتوافق تقريبًا مع تقدير الثبات الكلاسيكي الذي يبلغ حوالي 0.90 أحيانًا. لذا بدلًا من أن تكون دالة معلومات الاختبار نتيجة لاختيار المفردات، فإنه يتم تحديدها على أنها الهدف، ومن ثم يمكن اختيار المفردات لإنتاج دالة معلومات الاختبار موضع الاهتمام Hambleton & (Zhao, 2005).

- **ثانيًا: الأداء التفاضلي للمفردة: Differential Item Functioning (DIF)**

في الستينيات، تزايد الاهتمام بضمان التقييمات النفسية والتربوية العادلة، فأصبح تحليل الأداء التفاضلي للمفردة هو النهج المفضل للكشف عن التحيز المحتمل للمفردة(McAllister, 1993 as cited in Finch & French, 2019) ؛ لذا توفر هذه التحليلات أدلة للصلاحية من خلال فحص المدى الذي تقيس به المفردات ما تهدف إلى قياسه بطريقة عادلة عبر المجموعات، وتعتبر عنصرًا أساسيًا في عملية تطوير الاختبار. كما أن غياب الأداء التفاضلي للمفردة يساعد في ضمان إمكانية استخدام الاختبار بشكل أخلاقي مع جميع المفحوصين المحتملين، وأن الدرجات على الاختبار ذات مغزى، وقابلة للمقارنة بين أعضاء المجموعات المختلفة (Finch & French, 2019).

يُعرف الأداء التفاضلي للمفردة بأنه حالة تشير إلى وجود تحيّز إحصائي في مفردة اختبارية، يظهر عندما تختلف احتمالية الإجابة على المفردة بين مجموعتين من المفحوصين يمتلكان نفس مستوى السمة الكامنة، مثل القدرة أو المهارة، ولكن تختلف خصائصهم الديموغرافية كالجنس أو الخلفية الثقافية أو اللغة. ويمثل هذا النوع من التحيّز تهديدًا مباشرًا لصلاحية الاختبارات النفسية والتربوية؛ لأنه قد يُشير إلى أن المفردة لا تقيس فقط السمة المستهدفة، وإنما تتأثر بعوامل خارجية غير مرغوب فيها (Dorans & Holland,1993).كما يعرفه دليل معايير الاختبارات النفسية والتربوية بأنه مؤشر إحصائي لمدى وجود مجموعات مختلفة من المتقدمين للاختبار الذين هم من نفس مستوى القدرة، ولديهم معدلات مختلفة لاختيار الإجابة الصحيحة (American Educational Research Association et all, 2014). ويعرفه شين وريفيكي (Chen and Revicki (2014, بأنه الحالة التي يكون فيها الأفراد من مجموعات مختلفة، على سبيل المثال: (العمر، الجنس، العرق، التعليم، الثقافة)، على نفس المستوى من السمة الكامنة، ولديهم احتمالات مختلفة في إعطاء استجابة معينة على مفردة معينة.

وللحفاظ على صلاحية الاختبار؛ يجب ألاّ يميز الاختبار أداء المفحوصين على أساس المجموعة الديمغرافية التي ينتمون إليها. كما أن الأداء التفاضلي لمفردات الاختبار قد يهدد صحة وسلامة الاختبار من خلال إنتاج درجات متحيزة؛ لذا من الضروري تحديد المفردات التي تبدي أداءًا تفاضليًّا لمراجعتها وتصحيحها أو حذفها وفقًا لذلك (Lim et al., 2022).

الأداء التفاضلي وعلاقته بمفهوم التحيز:

تم استخدام مصطلح التحيز في وقت مبكر جدًّا لمعنيين مختلفين تمامًا، فالأول اجتماعي يشير إلى الظلم أو عدم المساواة، والآخر إحصائي يتعلق باختلاف الأداء بين المجموعات. مما أدى إلى التباس في استخدام مصطلح التحيز الذي كان يزعم بأنه السبب في تباين الدرجات، لذلك تم تقديم بعض الاقتراحات لاستخدام مصطلح غير التحيز للملاحظة الإحصائية، ومصطلح آخر لوصف الحكم وتقييم التحيز بالمعنى الاجتماعي. لذا تمّ استخدام مصطلح الأداء التفاضلي للمفردة مشيرًا إلى الملاحظة البسيطة التي مفادها أن المفردة تعرض خصائص إحصائية مختلفة في إعدادات مجموعة مختلفة بعد التحكم في الاختلافات في قدرات المجموعات (Holland & Wainer, 1993). ووجود التحيز في المفردات يدل على وجود عملية منهجية بخلاف السمة المقاسة، التي تؤدي إلى اختلافات في كيفية استجابة الأفراد لبنود الاختبار. ووصف بودساكوف وآخرون الطبيعة المعقدة لانحياز القياس، مؤكدين على عمل كامبل وفيسك، اللذين وصفا التحيز بأنه ناتج عن تباين غير ذي صلة في الأسلوب يمكن أن يجعل نتائج الاختبار غير صالحة في تفسيرها. والتحيز يتضمن تفسيرًا لسبب الاختلاف، بينما الأداء التفاضلي يقتصر على النتيجة الإحصائية فقط (Campbell & Fiske ,1959 as cited Podsakof et al.,2012).

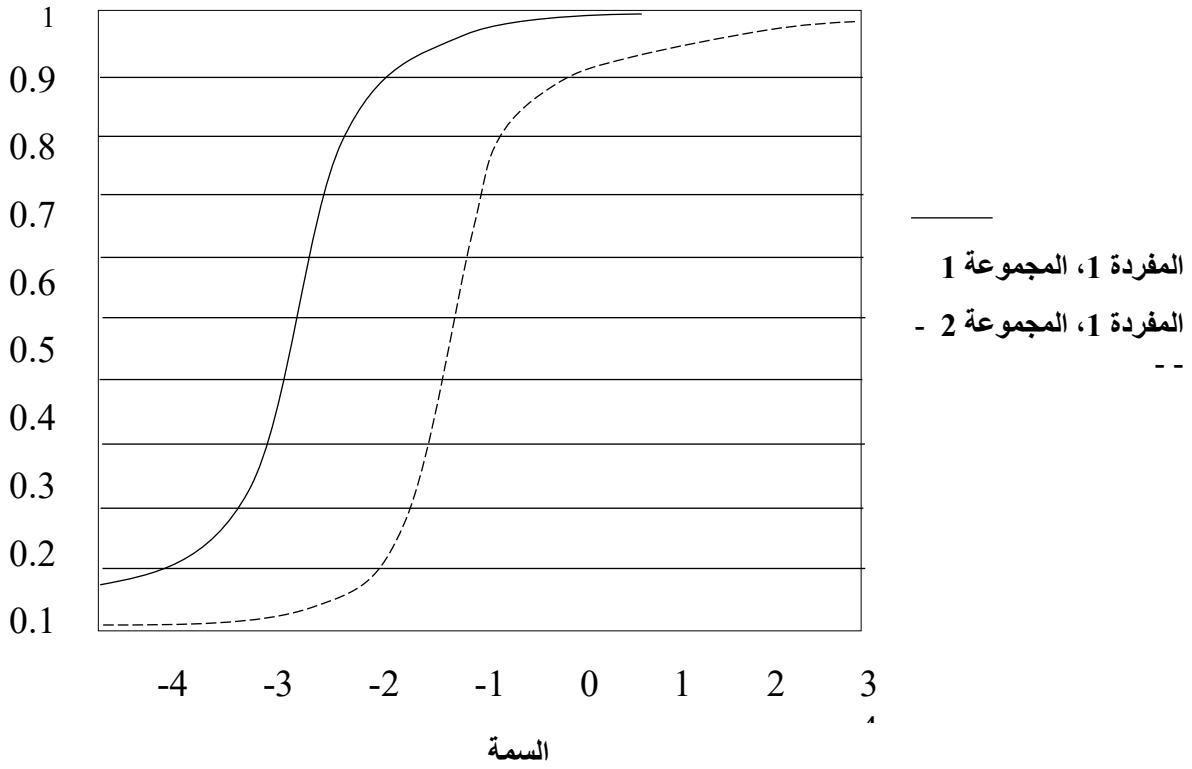
والنقطة الأساسية التي يجب ذكرها حول التحيز في أنه يمكن أن يؤثر على الخصائص السيكومترية للمفردات، مما يجعل الأداء على هذه المفردات غير صالح، كما أنه لا يعكس الفائدة المرجوة(Finch & French, 2019) . لتحليل التحيز، لا يتم

الاكتفاء بالتحليل الإحصائي مثل اختبار وجود الأداء التفاضلي للمفردات، بل يجب النظر أيضًا في سياق المفردة والغرض من الاختبار وتجربة المفحوصين، وكذلك استخدام الأدلة النوعية والمحتوى لتحديد ما إذا كان مصدر الاختلاف ناتجًا عن متغيرات غير ذات صلة بالقدرة المستهدفة (Holland & Wainer, 1993).

أنواع الأداء التفاضلي للمفردات:

يوجد نوعان من الأداء التفاضلي للمفردة هما:

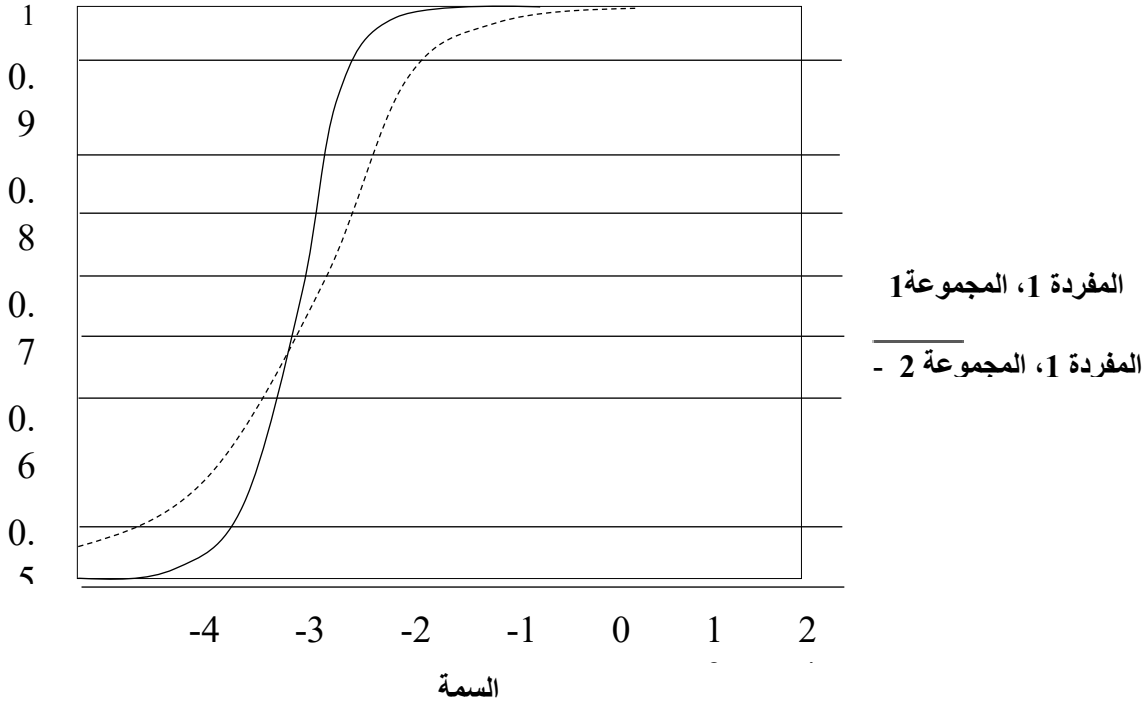
الأداء التفاضلي المنتظم Uniform DIF: يُشير إلى الحالة التي يختلف فيها احتمال الاستجابة الصحيحة لمفردة ما بين مجموعتين متطابقتين في السمة التي يقيسها الاختبار، ويكون الاختلاف ثابتًا عبر جميع مستويات السمة؛ بمعنى آخر يحدث الأداء التفاضلي المنتظم عندما يكون لدى الأفراد من مجموعات مختلفة (على سبيل المثال، الذكور والإناث) اللذين لديهم نفس المستوى من السمة الكامنة التي يتم قياسها (كالتحصيل في القراءة) احتمالات مختلفة للاستجابة بشكل صحيح على مفردة ما(Finch & French, 2019).. ويمكن تصور عمل الأداء التفاضلي للمفردة باستخدام المنحنى المميز للمفردة (ICC) الذي يظهر في الشكل (2).



شكل 2: الأداء التفاضلي المنتظم.

الأداء التفاضلي غير المنتظم Non-Uniform DIF:

يُشير إلى اختلاف احتمالية الاستجابة الصحيحة للمفردة بين المجموعات بعد مطابقة الأفراد عند بعض مستويات السمة التي يقيسها الاختبار، ويدل ذلك على تقاطع المنحنى المميز للمفردة(Finch & French, 2019). الذي يظهر في الشكل (3).



شكل 3: الأداء التفاضلي الغير منتظم.

الكشف عن الأداء التفاضلي:

خلال أكثر من خمسين عامًا سعت الطرق المبكرة إلى تحديد المفردات المتحيزة عن طريق تحليل التباين، ولكن سرعان ما حلت محلها أساليب أخرى أكثر دقة وقوة (Cardall & Coffman, 1964 as cited in Cleary & Hilton, 1968) . لذا تمّ اقتراح طرق مختلفة للكشف عن الأداء التفاضلي للمفردة، فكانت البداية في الطرق التي تعتمد على الكشف عن الأداء التفاضلي المنتظم التي تعتمد على المفردات ثنائية الاستجابة والاختبارات أُحادية البُعد. ولكن جاءت الحاجة إلى التعديل على التقنيات الموجودة وتطوير تقنيات جديدة تأخذ في الاعتبار التنسيقات والظروف المختلفة للاختبارات مثل الأداء التفاضلي غير المنتظم والمفردات متعددة الأبعاد.

وأدى اتساع نطاق الكشف عن الأداء التفاضلي إلى قيام بعض المؤلفين بتصنيفها وفقًا لخصائصها الرئيسية، فصُنّفت في مجموعات إلى ثلاث فئات: الأولى قائمة على تحليل التباين والنظرية الكلاسيكية للاختبار، والثانية قائمة على نظرية الاستجابة للمفردة الاختبارية (IRT)، والثالثة قائمة على الطرق المعتمدة على جداول الاحتمالات والتي تراعي عدة أبعاد تصنيفية، من بينها نوع المفردة: مفردات ثنائية الاستجابة (خياران فقط) مقابل مفردات متعددة الاستجابة (أكثر من خيارين). أسلوب المطابقة: هل تعتمد الطريقة على الدرجة المشاهدة كمقياس للقدرة، أم على السمة الكامنة نوع النموذج التحليلي: هل هو معلمي مثل النماذج اللوجستية، أم لا معلمي مثل مانتل هانزل (Berrío et al., 2020).

وأبرز طرق الكشف عن الأداء التفاضلي التي صُنّفت تحت النظرية الكلاسيكية في القياس، مثل : الانحدار اللوجستي (logistic regression)، وطرق الارتباط (correlation methods)، والتحليل الخطي اللوغاريتمي (log-linear analysis)، وتحليل التباين (analysis of variance)، ومربع كاي شونمان (scheuneman’s chi-square)، وطريقة مانتل هانزل (mantel-Haenszel method)، وطريقة صعوبة المفردات المحولة أو ما تُعرف بمخطط دلتا (transformed item difficulty or delta plot). بالإضافة إلى الطرق التي تندرج تحت نظرية الاستجابة للمفردة، مثل:

إحصائي والد (Wald statistics)، ونسبة الاحتمالية (likelihood-ratio)، ومربع كاي لورد (Lord's chi-square) (Ohiri et all ,2024).

وفي الدراسات التجريبية عادةً ما يتم تقييم الخصائص السيكومترية للاختبارات، ولكن هذا التقييم غير كاف لدعم الجودة السيكومترية، بل يجب أن يسمح أي اختبار يتمّ استخدامه بإجراء تقييم مناسب للبنية التي يتم قياسها التي يمكن تحليلها من خلال دراسة الثبات، وبشكل أكثر تحديدًا عن طريق الكشف عن الأداء التفاضلي للمفردات (Berrío et al., 2020).

الأداء التفاضلي وبيانات المحاكاة:

نظرًا لأن التطورات في البحوث حول الأداء التفاضلي للمفردات مرتبطة بتحديد الظروف التي تعمل في ظلها طرق الكشف عن الأداء التفاضلي، فقد أصبح استخدام بيانات المحاكاة شائعًا في هذا المجال لأنه يسمح للباحثين بتقييم ما إذا كان لبعض المتغيرات تأثيرًا على الخصائص الإحصائية لطرق عمل المفردات ذات الأداء التفاضلي (Berrío et al., 2020).كما تلعب تقنيات محاكاة الكمبيوتر أدوارًا رئيسية في البحوث وممارسة نظرية الاستجابة للمفردة (IRT)، وهي إطار شائع وقيّم لنمذجة وتحليل البيانات في الاختبارات النفسية والتربوية. وتتمثل إحدى الميزات المكتسبة من استخدام البيانات المولّدة في الكمبيوتر في نظرية الاستجابة للمفردة في أنها تنشئ قيم معلمات حقيقية ومعروفة، مما يتيح التقييم المباشر لتقديرات المعلمات في ظل الظروف المختلفة للقياس. وغالبًا ما يكون هذا أمرًا بالغ الأهمية للتطبيقات بما في ذلك تقييم المحكمين، وتحليل ملاءمة النموذج، والأداء التفاضلي للمفردات، وحساب الخطأ المعياري الشرطي للقياس. ومن الناحية النسبية، يُعدّ استخدام البيانات المولّدة أكثر كفاءة من استخدام البيانات التجريبية من حيث الوقت والتكلفة، مما يجعلها أداة قيّمة للباحثين والممارسين لتقييم برامج الاختبارات والتحقق من صحتها (Van der linden, 2018).

الدراسات السابقة:

هناك العديد من الدراسات السابقة التي تطرقت لموضوع الأداء التفاضلي، وتناولته من عدة جوانب ومنها الدراسات العربية والأجنبية، وسوف تستعرض هذه الدراسة عددًا من الدراسات التي تمّ الاستفادة منها، وبيان الفجوة العلمية التي تعالجها الدراسة الحالية.

هدفت دراسة النوافلة (2013) إلى بيان أثر نسبة الفقرات ذات الأداء التفاضلي، ومستوى الأداء التفاضلي للفقرة في تقدير معالم فقرات الاختبار وقدرات الأفراد وفق نموذج الاستجابة للفقرة ثلاثي المعلمة، باستخدام بيانات مولدة، وكانت نسبة الفقرات ذات الأداء التفاضلي في الاختبار هي (10%-20%-40%)، وكان مستويا الأداء التفاضلي للفقرات هما: (مستوى متدنٍ بمقدار 0.5، ومستوى عالٍ بمقدار 1.5). ونتج عن ذلك (7) مجموعات للدراسة. وقد تم استخدام برنامج (Wingen) لتوليد استجابات (510) أفراد في كل مجموعة، وقد تكونت العينة الكلية من (3570) مفحوصًا. وقد أشارت النتائج إلى أن تقدير معالم الفقرات، وكمية المعلومات التي يقدمها الاختبار، جميعها تتأثر سلبًا تبعًا لنسبة الفقرات ذات الأداء التفاضلي، ومستوى الأداء التفاضلي للفقرة، وذلك عندما كانت نسبة الفقرات ذات الأداء التفاضلي في الاختبار (40%)، ومستوى الأداء التفاضلي (1.5)، ولم يكن هناك أي تأثير على المتغيرات السابقة عندما كانت نسبة الفقرات ذات الأداء التفاضلي في الاختبار (10%-20%)، ومستوى الأداء التفاضلي (0.5).

وهدفت دراسة أبو صهيون (2014) إلى الكشف عن مدى فاعلية نموذج راش الهرمي في تقدير معالم الفقرة وقدرات الأفراد بوجود الأداء التفاضلي لنسب متفاوتة من الفقرات وبدرجات أداء مختلفة باستخدام بيانات مولدة. ولتحقيق هدف الدراسة تمّ استخدام المنهج الوصفي من خلال توليد فقرات ذات استجابة ثنائية (0،1) لعينات تحاكي المجتمع الأصلي بطريقة مونت كارلو (Monte Carlo) باستخدام برنامج (WinGen v.3)، وبلغت العينة (12000) فرد بواقع (1200) فرد لكل ظرف من ظروف الدراسة العشرة الناتجة عن تفاعل متغيري الدراسة [نسبة المفردات ذات الأداء التفاضلي (10%،25%،50%)، ومستوى الأداء التفاضلي (0.5، 1.0، 1.5)]، بوجود (60) فقرة في الاختبار لكل ظرف بحثي. وأظهرت نتائج الدراسة أن أدق التقديرات لمعلمة الصعوبة المقدرة عند خلو فقرات الاختبار من الأداء التفاضلي، وكذلك عندما تكون نسبة الفقرات ذات الأداء التفاضلي (10%) لجميع مستويات الأداء التفاضلي المعتمدة في هذه الدراسة، وعندما تكون نسبة الفقرات ذات الأداء التفاضلي (25%) عند مستوى الأداء التفاضلي (1.5)، وكذلك عندما تكون نسبة الفقرات ذات

الأداء التفاضلي (50%) ومستوى الأداء التفاضلي (0.5). في حين كانت أقل التقديرات دقة لمعلمة الصعوبة المقدرة عندما تكون نسبة المفردات ذات الأداء التفاضلي (25%) عند مستويين من مستويات الأداء التفاضلي (0.5، 1.0).

وفي ذات السياق هدفت دراسة الخطيب (2014) إلى الكشف عن الأثر التراكمي للأداء التفاضلي للفقرات بشقيه منتظم وغير منتظم على الأداء التفاضلي للاختبار، بالإضافة إلى الكشف عن أثر وجود فقرات ذات أداء تفاضلي في الاختبار على دالة معلوماته، عن طريق بيانات مولدة للمجموعتين المرجعية والمستهدفة. واستخدم البرنامج الإحصائي WinGen v3؛ لتوليد معالم القدرة للمفحوصين البالغ عددهم (520) مفحوصًا، ومعلمتي الصعوبة والتمييز للفقرات البالغ عددها (80) فقرة ثنائية الاستجابة، من توزيعات منتظمة تتراوح بين (-3 ،3) لمعلمة القدرة، وبين (-1.5،1.5) لمعلمة الصعوبة، وبين (0.8 ، 1.2) لمعلمة التمييز، ثمّ أحدث تغييرًا مقصودًا في معالم الفقرات المولّدة لتبدي أداءً تفاضليًّا بنسب (25%- 50%-70%). ولكل موقف تمّ توليد الاستجابات لأفراد المجموعتين، وتمّ استخدام برنامج Bilog-MG3 لتقدير معالم الأفراد والفقرات في كل مجموعة بصورة منفصلة. كما تمّ بحث أثر هذه الفقرات على دالة معلومات الاختبار عن طريق حساب الكفاءة النسبية، وقد توصلت الدراسة إلى وجود أثر تراكمي للفقرات التي تبدي أداءً تفاضليًّا على الأداء التفاضلي للاختبار في المواقف البحثية التي توزعت فيها فقرات الأداء التفاضلي بين المجموعتين المرجعية والمستهدفة، سواء كان في حالة الأداء المنتظم أو غير المنتظم، كما ظهر أثر بسيط للأداء التفاضلي للفقرات على دالة معلومات الاختبار عند النسب العالية.

وهدفت دراسة اجباره (2015) إلى بحث أثر نسبة الفقرات ذات الأداء التفاضلي على الخطأ من النوع الأول وقوة الاختبار في الكشف عنه باستخدام طريقتي الانحدار اللوجستي ونسبة الأرجحية العظمى. ولتحقيق هدف الدراسة تمّ استخدام أسلوب المحاكاة (توليد البيانات) وفق نظرية الاستجابة للفقرة للنموذج أحادي المعلمة، حيث تمّ توليد فقرات الاختبار بطول (40) فقرة، و (1500) مفحوص لكل طريقة، وتمّ تصميم الفقرات ذات الأداء التفاضلي وفقًا للنسب المستخدمة في البحث (10%-20%-40%). وكشفت نتائج الدراسة عن وجود فرق دال إحصائيًّا بين الأوساط الحسابية لمعدل الخطأ من النوع الأول للأداء التفاضلي يُعزى لنسبة الفقرات ذات الأداء التفاضلي؛ ولصالح النسبة الأقل، كما أظهرت النتائج وجود فرق دال إحصائيًّا بين الأوساط الحسابية لمعدل الخطأ من النوع الأول للأداء التفاضلي يُعزى لطريقة الكشف عن الأداء التفاضلي؛ ولصالح طريقة (نسبة الأرجحية)، كما أظهرت النتائج عدم وجود فروق دالة إحصائيًّا بين الأوساط الحسابية لمعدل الخطأ من النوع الأول للأداء التفاضلي تُعزى لتفاعل المتغيرين.

وأجرى أبو علام والمرابحة (2015) دراسة هدفت إلى تقييم فاعلية قوة الإحصائي لاختبار تحيز المفردة في الكشف عن دالة المفردة التفاضلية من خلال التعرف على مدى تأثير نسبة المفردات التي تُبدي أداءً تفاضليًّا وحجم العينة، والتفاعل بينهما باستخدام أسلوب المحاكاة في توليد البيانات وفقًا للنموذج اللوغاريتمي ثلاثي المعالم (3PLM). ولتحقيق هدف الدراسة تمّ توليد بيانات الاستجابة على المفردات وفقًا للنموذج اللوجستي ثلاثي المعلمة باستخدام برمجية Wingen-3، وتمّ تحديد خصائص المفردات الاختبارية بحيث تكون متماثلة لكل من المجوعتين المرجعية والمستهدفة، إذ جاء توزيع معلمة قدرة المفحوصين، ومعلمة صعوبة المفردة وتمييزها في كلا المجموعتين توزيعًا طبيعيًّا بمتوسط حسابي صفر وانحراف معياري واحد، في حين تمّ تثبيت معلمة التخمين لكلا المجموعتين عند القيمة (0.20). أما عينات المفحوصين وعدد مفردات الاختبار فقد صممت لتشكل (9) اختبارات لتحقيق أغراض المقارنة المتضمنة في الدراسة، إذ تمّ اختيار ثلاث نسب (10%-20%-30%) لمفردات تُبدي دالة تفاضلية لاختبار مكوّن من (50) مفردة، وتمّ اختيار ثلاث عينات أحجامها (400-800-1200) مفحوصًا، لدراسة قوة الاختبار الإحصائي، وتوصلت النتائج إلى أن قوة الاختبار الإحصائي لتحيز المفردة المتزامن تزداد بزيادة نسبة المفردات ذات الأداء التفاضلي في الاختبار.

كما هدفت دراسة أبينيابيبل وآخرون (2015) .Apinyapibal et al إلى مقارنة الكشف عن كفاءة الأداء التفاضلي باستخدام ثلاث طرق: الانحدار اللوجستي المعتمد على النظرية الكلاسيكية، واختبار تحيز المفردة المتزامن Simultaneous Item Bias Test (SIBTEST) المعتمد على نظرية الاستجابة للمفردة، وأشجار راش Raschtree المعتمد على نموذج التجزئة المتكرر. تمّ توليد البيانات باستخدام نسبتين من الأداء التفاضلي (10%،20%)، وثلاث أحجام من العينات (250،500،1000)، وطولين للاختبار (40 فقرة، و60 فقرة). توصلت النتائج إلى أن طريقة الانحدار اللوجستي، وطريقة اختبار تحيز المفردة المتزامن فعّالتان في اكتشاف كل من الأداء التفاضلي المنتظم وغير المنتظم.

وفي دراسة لي وزانج Lee and Zhang (2016) تمّ إجراء دراسة محاكاة لفحص تأثير نسب المفردات ذات الأداء التفاضلي على نتائج القياس مثل: تقديرات القدرة لنظرية الاستجابة للمفردة، والثبات التجريبي للاختبار، والخطأ المعياري لتقدير معلمة القدرة. تمّ اختيار ثلاث نسب للمفردات ذات الأداء التفاضلي، تتمثل في (5%،10%،20%)؛ لتوضيح إلى أيّ مدى يمكن أن تتأثر نتائج القياس مع زيادة النسبة للمفردات ذات الأداء التفاضلي، وقد توصلت النتائج إلى أن لـلأداء التفاضلي تأثير محدود على الثبات التجريبي، لكن تأثيره على الخطأ المعياري لتقدير معلمة القدرة كان واضحًا لقيم قدرة معينة.

وأجرى الشريفين (2018) دراسة هدفت إلى الكشف عن أثر نوع الأداء التفاضلي للفقرات (منتظم، غير منتظم) على الخصائص السيكومترية للفقرات، والاختبار وفق النموذج اللوجستي ثلاثي المعلمة، والنموذج اللوجستي الثلاثي اللامعلمي لنظرية الاستجابة للفقرة، واقتصرت الدراسة على نسبة واحدة للفقرات التي تُبدي أداءً تفاضليًّا (10%). ولتحقيق هدف الدراسة تمّ استخدام برنامج (WINGEN) لتوليد نموذجي اختبار من نوع الاختيار من متعدد، تكوّن كل منهما من (50) فقرة، تضمن الأول (10) فقرات ذات أداء تفاضلي منتظم، والثاني (10) فقرات ذات أداء تفاضلي غير منتظم، وتمّ توليد استجابات (1000) فرد، وقد تمّ تحليل البيانات وفق النموذج الثلاثي المعلمة باستخدام برمجية (BILOG-Mg)، ووفق النموذج الثلاثي اللامعلمي باستخدام برنامج (Test Graf). أشارت النتائج إلى وجود فرق دال إحصائيًّا بين الوسطين الحسابيين لكل من معالم التمييز، ومعالم التخمين لفقرات الاختبار لصالح النموذج المعلمي، كما أشارت النتائج أيضا إلى وجود فرق دال إحصائيًّا بين الوسطين الحسابيين لمعلمة الصعوبة لصالح النموذج اللامعلمي، ووجود فروق دالة إحصائيًّا بين الأوساط الحسابية لكل من معلمتي الصعوبة والتخمين لفقرات الاختبار تُعزى لتفاعل النموذج المستخدم، ونوع الأداء التفاضلي. كما أشارت النتائج إلى وجود فرق دال إحصائيًّا بين الوسطين الحسابيين لكمية معلومات الاختبار لصالح النموذج اللامعلمي، حيث كانت كمية المعلومات التي يقدمها الاختبار أكبر عند جميع مستويات القدرة وفق النموذج اللامعلمي، سواء كان الأداء التفاضلي منتظمًا أم غير منتظم، وبينت النتائج أن قيم معاملات الثبات المقدرة وفق النموذج المعلمي كانت الأعلى، سواء كان الأداء التفاضلي للفقرات منتظمًا أم غير منتظم.

وهدفت دراسة الحاج (2018)، إلى الكشف عن الأداء التفاضلي لفقرات ومموهات اختبار محكي المرجع في العلوم للصف السادس الابتدائي، ومصمم وفق نظرية الاستجابة للفقرة. ولتحقيق هدف الدراسة صممت الباحثة اختبارًا يتكون من (75) فقرة من نوع الاختيار من متعدد. طُبّق الاختبار على عينة من (520) طالبًا وطالبة من طلاب الصف السادس الابتدائي، في مدينة جدة، منهم (418) طالبًا من المدارس الحكومية، و(102) طالبًا من المدارس الأهلية. أظهرت نتائج البرنامج الإحصائي (BILOG-MG3) مطابقة (61) فقرة من فقرات الاختبار للنموذج ثلاثي المعلمة، وظهر الأداء التفاضلي في (16) فقرة منها، كما أظهر اختبار مربع كاي للكشف عن الأداء التفاضلي لمموهات الفقرات، وجود أداء تفاضلي لمموهات (6) فقرات منها.

وفي دراسة سيلفي واليجي Selvi and Alici (2018) والتي هدفت إلى التحقيق في تأثير مختلف طرق معالجة البيانات المفقودة على الكشف عن البنود ذات الأداء التفاضلي باستخدام طريقة مانتل هانزل والطرق المقننة القائمة على النظرية التقليدية واختبار نسبة الأرجحية القصوى القائمة على نظرية الإستجابة للمفردة. وبناء على البيانات التي تم الحصول عليها من 1046 مرشحاً منهم 523 اناث و 523 ذكور دخلوا في امتحان الطلاب الأجانب (FNSE)، في جامعة مرسين (MEU)وأجابوا على الاختبار الفرعي للمهارات الأساسية، حيث يتكون اختبار المهارات الأساسية من 80 بند متعدد الخيارات وجميع البنود هي بنود ثنائية الإستجابة (0،1). أظهرت النتائج أن عدد البنود التي تم وضع علامة عليها على أنها ذات أداء تفاضلي قد تغيرت مع أساليب معالجة البيانات المفقودة المستخدمة.

كما هدفت دراسة الوليلي والدوسري (2019) إلى الكشف عن المفردات ذات الأداء التفاضلي تبعًا لمتغير الجنس في مفردات اختبار القدرات العامة وأثرها على دقة تقدير القدرات والخصائص السيكومترية للاختبار. تم استخدام بيانات أحد النماذج لاختبار القدرات العامة والمكوّن من 96 مفردة، وبلغ عدد العينة 7408 طالبًا وطالبة، منهم 3599 من الذكور و3809 من الإناث باستخدام طريقتين تتبعان نظرية الاستجابة للمفردة هما: مربع كاي لكاميلي Camilli Chi-Square، والفروق في المساحات تحت منحنيات خصائص المفردات. وقد توصلت النتائج إلى اتفاق كل من الطريقتين في الحكم على وجود (26) مفردة ذات أداء تفاضلي وكان أغلبها لصالح الطلبة الذكور. كما أن قيم معاملات ثبات وصدق الاختبار بعد حذف المفردات ذات الأداء التفاضلي لمتغير الجنس أكبر منها قبل حذف المفردات ذات الأداء التفاضلي لمتغير الجنس. وباستخدام برنامج

(SPSS) تبيّن أن قيم معاملات ثبات وصدق الاختبار دالة إحصائيًا عند مستوى 0.05، وهذا يدل على أن بقاء المفردات ذات الأداء التفاضلي في الاختبار تؤدي إلى انخفاض خصائصه السيكومترية.

وسعت دراسة رشوان (2021) إلى التعرف على الأداء التفاضلي لبنود اختبار واطسن وجليسر "النسخة القصيرة" للتفكير الناقد في ضوء التخصص الأكاديمي باستخدام النموذج البارامتري ثنائي المعلمة ونموذج Ramsay للانحدار اللابارامتري الممهد، والكشف عن تأثير الأداء التفاضلي للبنود على الأداء التفاضلي للاختبار باستخدام التحليل العاملي التوكيدي متعدد المجموعات، وتمّ تطبيق الاختبار على 396 طالبًا من طلاب التخصصات العلمية والأدبية في جامعة القصيم، وتوصلت الدراسة إلى أن العديد من بنود الاختبار لها أداء تفاضلي لصالح التخصصات العلمية، والبعض الآخر له أداء تفاضلي لصالح التخصصات الأدبية. كما توصلت النتائج إلى تأثير الأداء التفاضلي للبنود على الأداء التفاضلي للاختبار ككل.

وهدفت دراسة الورديا (2021) Alordia إلى المقارنة بين ثلاث طرق للكشف عن البنود ذات الأداء التفاضلي، (طريقة مانتل هانزل ومربع كاي ونموذج راش). تم اختيار عينة الدراسة بالطريقة العشوائية الطبقية باختيار 1900 طالب من أصل 65961 طالباً من المرحلة الثانوية في المدارس العامة في نيجيريا. أظهرت النتائج أن طريقة نظرية الاستجابة للبند تعد أعلى في قيمة معامل الثبات، ولكن طريقة مربع كاي كانت إيجابية أيضاً. ووصت الدراسة بأهمية استفادة الخبراء من طرق الكشف عن الأداء التفاضلي بناءً على نظرية استجابة البند ولكن عندما لا تكون متاحة بسهولة، يمكن استخدام طريقة مربع كاي.

كما هدفت دراسة الحربي والمدني (2021) إلى القيام بالكشف عن أثر نسبة المفردات ذات الأداء التفاضلي على قوة وفاعلية اختبار منحنى خصائص المفردة باستخدام أسلوب المحاكاة. ولتحقيق هدف الدراسة تمّ توليد بيانات لثلاثة اختبارات مكوّن كل منها من 100 مفردة، تحوي على الترتيب نسب تفاضل (10%-20%-30%) طُبقت عبر برنامج المحاكاة على 3000 فرد، تتوزع قدراتهم توزيعًا طبيعيًا بمتوسط مقداره (صفر)، وانحرافه المعياري (1). وأشارت النتائج إلى فاعلية اختبار منحنى خصائص المفردة في الكشف عن نسبة المفردات ذات الأداء التفاضلي، كما أظهرت النتائج أن قوة الاختبار بلغت قيمتها (619.0) للاختبار ذو الأداء التفاضلي 10%، وبلغت قيمة قوة الاختبار ذو الأداء التفاضلي 20% (703.0)، أما قوة الاختبار للاختبار ذو نسبة الأداء التفاضلي 30% فقد بلغت قيمتها (786.0)، وجميعها أتت بمقدار قوة أعلى من المتوسط حسب تصنيف اختبار كوهين، مما يعني فعالية اختبار منحنى خصائص الفقرة في استخراج المفردات ذات الأداء التفاضلي.

وجاءت دراسة علي وعيّال (2021) بهدف معرفة تأثير اختلاف مستوى قدرة الأفراد في الأداء التفاضلي لاختبار القدرة العقلية غير اللفظي لماستر براين والمكون من (60) فقرة موزعة بواقع عشر فقرات وعلى ست مكونات وهي: التماثل أو التناظر، والمصفوفة، واستبعاد الشكل الغريب، والتشابه، والرموز – الكود، والسلسلة أو السلاسل. و قد اقتصر البحث على طلبة المرحلة الإعدادية لمدينة بغداد وللدراسة الصباحية. ولتحقيق أهداف البحث جرى تطبيق الاختبار على عينة مكونة من (600) طالب وطالبة تمّ اختيارهم بالطريقة العشوائية الطبقية، وتمّ التحقق من افتراضات نظرية الاستجابة للمفردة (أُحادية البُعد، الاستقلال الموضعي، المنحنى المميز للفقرة، الاختبار ليس اختبار سرعة). أظهرت النتائج ملاءمة (55) مفردة من مفردات اختبار ماستر براين للنموذج ثنائي المعلمة، كما أظهرت النتائج أنّ 10.90% من المفردات أداءً تفاضليًّا وفقًا للنموذج البارامتري ثنائي المعلمة لجميع أفراد العينة، و 18.18% للأفراد منخفضي القدرة، و 16.36% للأفراد مرتفعي القدرة، وتوصلت النتائج أيضا إلى عدم وجود فروق ذات دلالة إحصائية في نسبة المفردات ذات الأداء التفاضلي وفق نظرية الاستجابة للمفردة والنموذج ثنائي المعلمة تبعًا لمستوى القدرة.

وهدفت دراسة ضعضع (2023) إلى دراسة أثر اختيار نموذج الاستجابة للمفردة الثنائية (أحادي المعلمة وثنائي المعلمة وثلاثي المعلمة) في الكشف عن المفردات ذات الأداء التفاضلي، والمقارنة في نسبة المفردات ذات الأداء التفاضلي، ومعرفة دلالة الفروق في قيمة معامل الأداء التفاضلي ومتوسطه. ولتحقيق هدف الدراسة اعتمدت الباحثة نتائج اختبار قيد الماجستير المقام في المعهد العالي للغات في جامعة حلب المكون من (60) مفردة من النوع الموضوعي ثنائي الاستجابة (0 ،1) مطبقة على (1000) متقدم ومتقدمة من الكليات النظرية والتطبيقية. وبعد التأكد من تحقق افتراضات نظرية الاستجابة للمفردة الاختبارية أحادية البعد وملائمة المفردات للنماذج ثنائية الاستجابة باستخدام برنامج Jmetrik، أظهرت النتائج أن طريقة المنحنى المميز للمفردة للكشف عن الأداء التفاضلي كانت الأعلى في النموذج ثنائي المعلمة بنسبة قدرها (68.5%)، ومن ثم النموذج أحادي المعلمة بنسبة قدرها (22.22%)، ومن ثم النموذج ثلاثي المعلمة بنسبة قدرها (1.9%).

وهدفت دراسة الخطيب والعمري (2023)، إلى تقصي قيم دالة معلوماتية الاختبار المقدرة، وفحص التوافق في مستويات دالة معلوماتية الاختبار المقدر باستخدام النموذجين (النموذج ثلاثي المعلمة ونموذج موكن اللامعلمي) في ضوء نوع ونسبة الأداء التفاضلي (10%، 20%، 30%). ولتحقيق هدف الدراسة تم استخدام المنهج التجريبي باستخدام بيانات مولدة وفق برمجية (WinGen) على عينة مكونة من (2000) مفحوص موزعين بالتساوي على المجموعة المرجعية والمجموعة المستهدفة بحيث تتبع قدرتهم التوزيع الطبيعي بوسط حسابي (0) وانحراف معياري (1)، وتكونت أداة الدراسة من اختبار افتراضي بطول (60) فقرة. أظهرت نتائج الدراسة أن النموذج اللامعلمي أكثر كفاءة عند مستويات القدرة المتوسطة، في حين أنه كان أكثر كفاءة عند الأطراف في النموذج المعلمي، حيث القدرة المتطرفة، وهذا يعكس دقة أعلى للنموذج عند المستويات الدنيا والعليا للقدرة، كما أظهرت النتائج أن دالة معلوماتية الاختبار تقدر بشكل مختلف عبر مستويات القدرة باختلاف النموذج المستخدم ولصالح النموذج اللامعلمي.

التعليق على الدراسات السابقة:

- في ضوء ما تقدم عرضه في الدراسات السابقة نلاحظ أنها تنوعت في أهدافها، فقد هدف بعضها إلى دراسة تأثير نسب المفردات ذات الأداء التفاضلي على المتغيرات التابعة وفق نظرية الاستجابة للمفردة، والبعض هدف إلى دراسة فعالية طرق الكشف عن المفردات ذات الأداء التفاضلي في الاختبارات المختلفة، وكذلك هدف بعضها إلى دراسة مدى دقة بعض الاختبارات في تقدير قدرات الأفراد معالم الفقرات.
- تنوعت عينات الدراسة، حيث تكونت بعض العينات من طلاب وطالبات الجامعة، والبعض الآخر طلاب وطالبات المرحلة الابتدائية والمتوسطة والثانوية، في حين أن غالبية الدراسات اعتمدت على توليد البيانات ودراسة أثر المتغير المستقل على المتغير التابع من خلال برامج المحاكاة.
- غالبية الدراسات السابقة تمت وفق نماذج نظرية الاستجابة للمفردة الاختبارية المعلمية، وهي النموذج البارمتري أحادي المعلمة، وثنائي المعلمة، وثلاثي المعلمة، وكان النموذج البارامتري ثلاثي المعلمة هو الأكثر دراسة من بين النماذج المستخدمة.
- تفاوتت أحجام العينات في الدراسات، حيث تراوحت أحجام عينات المفحوصين التي تناولتها الدراسات مابين (250-3809) مفحوص لكل مجموعة من مجموعات الدراسة، كما أن بعض الدراسات درست تأثير ثلاث أحجام مختلفة للعينات.
- يلاحظ اختلاف نسب المفردات ذات الأداء التفاضلي التي تناولتها الدراسات، حيث تناولت نسبًا صغيرة ومتوسطة وكبيرة، وقد تناولت بعض الدراسات تأثير نسبة واحدة، ونسبتين، وثلاث نسب لمعرفة تأثيرها على المتغيرات التابعة، وأوضحت الدراسات أنه كلما زادت نسبة المفردات ذات الأداء التفاضلي أثّر ذلك سلبًا على الخصائص السيكومترية في الاختبار، وأن بعض اختبارات الكشف عن الأداء التفاضلي تزداد فاعليتها بزيادة نسب المفردات ذات الأداء التفاضلي، وبعضها تقل فاعليتها بزيادة المفردات ذات الأداء التفاضلي.
- تفاوتت الدراسات التي اعتمدت على توليد البيانات في طول الاختبارات التي تناولتها، فقد بلغ أطوال الاختبارات في الدراسات التي تم استعراضها في الدراسة الحالية (20-40-50-60) مفردة في كل اختبار، ويُلاحظ أن غالبية الدراسات التي درست تأثير طول واحد اختارت 50 مفردة لكل اختبار؛ لذا تمّ اختياره في الدراسة الحالية.
- جميع الدراسات السابقة التي تم استعراضها درست تأثير نوع الأداء التفاضلي، أو نسبة المفردات ذات الأداء التفاضلي على الثبات وفق نظرية الاستجابة للمفردة، ولم تدرس إلا الثبات التجريبي.
- أظهرت نتائج الدراسات المتعلقة بالأداء التفاضلي للمفردة والخصائص السيكومترية للاختبار أنّ الثبات التجريبي لا يتأثر بوجود الأداء التفاضلي للمفردة، وأنّ الثبات وفق النظرية الكلاسيكية ينخفض بوجود الأداء التفاضلي للمفردة، إضافة إلى أنه لم يظهر تأثير للأداء التفاضلي للمفردة على دالة المعلومات للاختبار إلا عند النسب الكبيرة للمفردات ذات الأداء التفاضلي.
- تم توليد البيانات في الدراسات المبنية على نظرية الاستجابة للمفردة الاختبارية من خلال برنامج Wingen، وتم تحليل البيانات باستخدام برنامج Bilog-MG وبرنامج Jmetrik.

- **فرضيات الدراسة:**

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

الفصل الثالث

منهج الدراسة وإجراءاتها

- **الطريقة والإجراءات.**
- **منهجية الدراسة.**
- **توليد البيانات.**
- **التحقق من افتراضات نظرية الاستجابة للمفردة.**

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.

منهج الدراسة:

تم استخدام المنهج التجريبي للكشف عن تأثير اختلاف 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.

جدول1: 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

0.401	1.465	**31**	-1.333	0.848	**6**
0.409	0.874	**32**	-1.205	0.817	**7**
0.418	1.137	**33**	-1.197	1.401	**8**
0.421	1.512	**34**	-1.116	1.347	**9**
0.421	0.969	**35**	-1.108	1.378	**10**
0.476	1.361	**36**	-0.874	0.886	**11**
0.534	1.144	**37**	-0.796	0.832	**12**
0.542	1.344	**38**	-0.781	1.578	**13**
0.628	1.073	**39**	-0.656	1.483	**14**
0.708	1.522	**40**	-0.634	0.706	**15**
0.843	1.561	**41**	-0.5	1.438	**16**
0.885	1.334	**42**	-0.488	1.019	**17**
0.999	0.901	**43**	-0.486	0.669	**18**
1.035	0.761	**44**	-0.381	1.093	**19**
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 توليد بيانات (1000) مفحوص. وترتيب 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).

جدول 3: 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 إحصائيًا، 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.

جدول 4: 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	2.432	**25.005**
المجموعة المستهدفة بنسبة 10%	1	9.736	19.471	**19.471**
	2	1.424	2.847	**22.318**
	3	1.178	2.355	**24.673**
المجموعة المستهدفة بنسبة 20%	1	9.025	18.050	**18.050**
	2	1.434	2.868	**20.918**
	3	1.219	2.438	**23.355**
المجموعة المستهدفة بنسبة 30%	1	9.195	18.389	**18.389**
	2	1.530	3.060	**21.449**
	3	1.205	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 ($t = \left(r \times \sqrt{df}\right) / \sqrt{1 - r^2}$) ، 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 ZQ3 (0.06185) عند جميع المجموعات (Donnelly,2007).

حيث تَمَّ اختبار تحقُّق افتراض الاستقلال الموضعي عن طريق: رصد الأزواج الارتباطية بين كُلِّ فقرتين من فقرات الاختبار التي تُحقِّق التبعية الموضعية، ورصد الفقرات التي تُحقِّق الاستقلال الموضعي وِفْقًا للنموذج البارامتري ثنائي المعلمة المُقدَّر (2PL) باستخدام الأرجحية العُظمى (MLE) قبل مُطابقة الأفراد والفقرات، كما هو مُبيَّن في جدول 5.

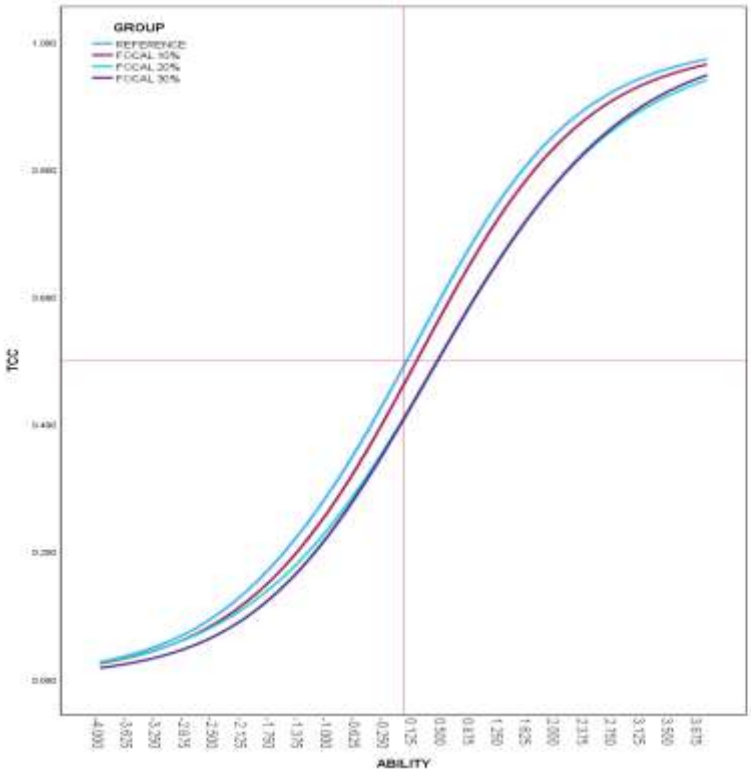
جدول 5: نتائج اختبار الباقي المعياري للمقارنات البعدية اللامعلمية لتكرارات الأزواج الارتباطية لفقرات الاختبار قبل مُطابقة الأفراد والفقرات

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	

يلاحظ من جدول 5 أنَّ جميع القيم الموجبة لاختبار الباقي المعياري للمقارنات البعدية اللامعلمية كانت أكبر من المعيار ذي القيمة (+1.96)؛ بما يُفيد أنَّ التكرارات المرصودة/الملاحظة للأزواج الارتباطية لفقرات الاختبار تُسْهِم مُساهَمةً موجبة في حساب مؤشر الاستقلال الموضعي (Q_3)؛ كونها تختلف جوهريًّا بدلالةٍ إحصائيةٍ (α=0.05) عن التكرارات المتوقعة لها، مِمَّا يقود إلى تحقُّق افتراض الاستقلال الموضعي لفقرات الاختبار وفق النموذج ثنائي المعلمة المُقدَّر (2PL) بطريقة الأرجحية العظمى (MLE) قبل المُطابقة.

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 من خلال برنامج SPSS كما هو مبيّن في الشكل (4).



شكل 4: منحنى خصائص الاختبار

يتضح من الشكل السابق أن مجموعات الدراسة المرجعية والمستهدفة تُحقق افتراض منحنى خصائص الاختبار؛ إذ لم يُلاحظ تناقص يزيد عن مقدار (0.02) 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 & Sijtsma, 2000) ، وجدول (6) يوضح ذلك.

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

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

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 اختبار والد (Wald Test) المصممة وفقًا لاقتراح لورد (Lord, 1977). و اختبار إحصائي يُستخدم لفحص ما إذا كانت معلمة معينة تختلف دلاليًا بين مجموعتين ويحسب بقسمة مربع الفرق بين تقدير المعلمة للمجموعتين مقسوم على المقام وهو التباين المقدر للفارق بين التقديرين (يُحسب من مصفوفة التباين–التغاير).ثم تُقارن إحصائية والد بتوزيع كاي تربيع (χ^2) بدرجة حرية = 1 أو أكثر حسب عدد المعلمات المقارنة. بالتضامن مع مصفوفات تباين – تغاير دقيقة لحساب الأخطاء المعيارية لتقديرات معالم المفردات، ويتمّ حسابها باستخدام خوارزمية التوقع والتعظيم (EM) التكميلية (SEM) وهي التي اقترحها كاي في عام 2008، وهذه الخوارزمية تستخدم لحل المشكلات التي تتضمن متغيرات كامنة (latent variables) غير قابلة للملاحظة بشكل مباشر، مما يجعل المشكلة معقدة. وتعتمد خوارزمية (EM) على عملية تكرارية لتقدير هذه المتغيرات الكامنة والوصول إلى تقدير أمثل للمعلمات المطلوبة (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).

جدول 7: 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**
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**

جدول 8: المرحلة الأولى: المجموعة المرجعية مع المجموعة المستهدفة بنسبة بنود ذات أداء تفاضلي 20%.

										عدد المفردات
مجموعة 1	**مجموعة 2**	X^2 مجموع	درجات الحرية	قيمة الاحتمالية	X^2_a	درجات الحرية	قيمة الاحتمالية	$X^2_{c\|a}$	درجات الحرية	**قيمة الاحتمالية**
1	1	416.3	2	0.0001	1.9	1	0.1645	414.3	1	**0.0001**
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**
0.0001	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**
0.0001	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**
0.0001	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**
0.0002	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**

جدول 9: 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**
0.0001	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**
0.0001	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**
0.0001	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 (8) مفردات ذات أداء تفاضلي في المجموعة المستهدفة الأولى، و(36) مفردة في المجموعة المستهدفة الثانية، و(43) 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).

جدول 10: 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	$X^2_{c\|a}$	درجات الحرية	**قيمة الاحتمالية**
1	1	123.8	2	0.0001	6.5	1	0.0106	117.2	1	**0.0001**
13	13	47.3	2	0.0001	0.0	1	0.8853	47.2	1	**0.0001**
29	29	254.5	2	0.0001	0.4	1	0.5426	254.1	1	**0.0001**
36	36	138.4	2	0.0001	0.2	1	0.6730	138.2	1	**0.0001**
37	37	230.7	2	0.0001	0.0	1	0.8479	230.6	1	**0.0001**

جدول 11: 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	$X^2_{c\|a}$	درجات الحرية	**قيمة الاحتمالية**
1	1	349.5	2	0.0001	3.6	1	0.0584	345.9	1	**0.0001**
10	10	181.2	2	0.0001	0.0	1	0.8268	181.2	1	**0.0001**
13	13	50.7	2	0.0001	0.1	1	0.7657	50.6	1	**0.0001**

0.0001	1	111.8	0.7500	1	0.1	0.0001	2	111.9	19	**19**
0.0001	1	93.1	0.8999	1	0.0	0.0001	2	93.1	20	**20**
0.0001	1	78.9	0.3506	1	0.9	0.0001	2	79.8	22	**22**
0.0001	1	196.9	0.2991	1	1.1	0.0001	2	198.0	29	**29**
0.0001	1	135.2	0.9751	1	0.0	0.0001	2	135.2	36	**36**
0.0001	1	460.5	0.5894	1	0.3	0.0001	2	460.8	37	**37**
0.0001	1	174.6	0.3141	1	1.0	0.0001	2	175.7	38	**38**

جدول 12: المرحلة الثانية: 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	92.2	0.0184	1	5.6	0.0001	2	97.7	1	**1**
0.0001	1	110.2	0.5080	1	0.4	0.0001	2	110.7	3	**3**
0.0001	1	40.4	0.6887	1	0.2	0.0001	2	40.6	4	**4**
0.0001	1	99.6	0.8206	1	0.1	0.0001	2	99.7	6	**6**
0.0001	1	154.9	0.5764	1	0.3	0.0001	2	155.2	10	**10**
0.0001	1	37.6	0.7813	1	0.1	0.0001	2	37.7	13	**13**
0.0001	1	74.0	0.8956	1	0.0	0.0001	2	74.0	16	**16**
0.0001	1	104.5	0.6123	1	0.3	0.0001	2	104.7	19	**19**
0.0001	1	97.1	0.2575	1	1.3	0.0001	2	98.3	20	**20**
0.0001	1	77.7	0.1962	1	1.7	0.0001	2	79.4	22	**22**
0.0001	1	135.5	0.4094	1	0.7	0.0001	2	136.2	29	**29**
0.0001	1	93.0	0.7688	1	0.1	0.0001	2	93.1	36	**36**
0.0001	1	134.4	0.4874	1	0.5	0.0001	2	134.9	37	**37**
0.0001	1	120.1	0.7775	1	0.1	0.0001	2	120.2	38	**38**
0.0001	1	88.0	0.0432	1	4.1	0.0001	2	92.1	43	**43**

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 (10) مفردات، وفي المجموعة المستهدفة بنسبة (30%) هي (15) مفردة.

مطابقة المفحوصين والمفردات للنموذج ثنائي المعلمة:

تمّ استخدام برنامج (Bilog-MG) للكشف عن مطابقة المفحوصين والمفردات 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 المفحوصين غير المطابقين في كل مجموعة بسبب عدم مطابقتهم للنموذج ثنائي المعلمة، ليصبح عدد المفحوصين بعد الحذف في كل مجموعة (995) 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	269 ، 62
المجموعة المستهدفة بنسبة 10%	1	919
المجموعة المستهدفة بنسبة 20%	1	693
المجموعة المستهدفة بنسبة 30%	2	930 ، 693

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

الفصل الرابع

- **الإجابة عن السؤال الأول.**
- **الإجابة عن السؤال الثاني.**
- **الإجابة عن السؤال الثالث.**
- **الإجابة عن السؤال الرابع.**
- **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	**0.2880**
	القيمة العظمى	3.2590	**1.0260**

تم استخدام معادلة اختبار (ت) $t = (x1 - x2)/\sqrt{se_1^2 + se_2^2}$ (وردات والحسن، 2019). لمقارنة دقة تقدير القدرات بين المجموعة المرجعية والمجموعات المستهدفة التي تختلف في نسب المفردات ذات الأداء التفاضلي (10%، 20%، 30%)، وذلك بهدف التحقق من مدى تحقق خاصية اللاتغاير لهذه التقديرات. وقد تم تفضيل اختبار (ت) على تحليل التباين (ANOVA) لعدة اعتبارات منهجية، أهمها أن الخطأ المعياري المرتبط بدقة تقدير القدرات وفق النموذج ثنائي المعلمة لا يتبع بالضرورة التوزيع الطبيعي، مما يجعل استخدام ANOVA أقل دقة في هذا السياق.

كما أن اختبار (ت) يُراعي الخطأ المعياري لكل قدرة على حدة، مما يسمح بتحليل الفروق بين فترتي الثقة لمتوسطات القدرات، وهو ما يوفر تقديرًا أكثر دقة لحجم الفروق الواقعة بين المجموعات. ويُعد هذا الإجراء ضروريًا للتحقق من مدى ثبات دقة تقدير القدرات، وضمان أن التغير في نسب الأداء التفاضلي لا يؤدي إلى انتهاك خاصية اللاتغاير، والتي تُعد من الشروط الجوهرية لصحة استخدام نماذج نظرية الاستجابة للمفردة، خاصة عند مقارنة قدرات الأفراد أو تفسير نتائج الاختبار، والجدول (15) يوضح نتائج اختبار (ت):

جدول 15: 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 إحصائيًا 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).

جدول16: 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 30%	العدد	50	50	50	50
	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 17: 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 (α= 0.05) بين دقة تقديرات معلمات الصعوبة تعزى لنسب المفردات ذات الأداء التفاضلي (10%، 20%). ووجود فروق دالة إحصائيًّا عند مستوى دلالة (α= 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 استخراج معاملات الثبات النظري والتجريبي من خلال مخرجات برنامج Bilog_MG ، وذلك كما هو مبين في الجدول (18).

جدول18: معاملات الثبات النظري والتجريبي في مجموعات الدراسة.

مجموعة الأداء التفاضلي	الثبات النظري	الثبات الأمبريقي
المرجعية (صفر%)	**0.9157**	**0.9203**
مستهدفة (10%)	**0.9162**	**0.9197**
مستهدفة (20%)	**0.9099**	**0.9140**
مستهدفة (30%)	**0.9124**	**0.9164**

وتمّ حساب الإحصائي (M) للعينات المستقلة $M = \left[\sum_{j=1}^{k}\left(\frac{(l_j-1)(9S_j-11)^2}{18l_j(S_j-1)}\right)\right] - \left[\frac{\left[\sum_{j=1}^{k}\left(\frac{(l_j-1)(9S_j-11)^2}{18l_j(S_j-1)}\right)(1-\hat{\alpha}_j)^{-\left(\frac{1}{3}\right)}\right]^2}{\sum_{j=1}^{k}\left(\frac{(l_j-1)(9S_j-11)^2}{18l_j(S_j-1)}\right)(1-\hat{\alpha}_j)^{-\left(\frac{2}{3}\right)}}\right]$ (القضاة، 2018). لتحليل الفروق بين معاملات الثبات النظري والثبات التجريبي الناتجة عن اختلاف نسب المفردات ذات الأداء التفاضلي في مجموعات الدراسة.

ويُعد الإحصائي (M) من الأساليب المناسبة لمقارنة عدة مجموعات مستقلة، ويُستخدم للكشف عن وجود فروق جوهرية في المؤشرات السيكومترية، خاصة عندما لا تتبع البيانات التوزيع الطبيعي أو عندما تكون مستمدة من نماذج احتمالية مثل نموذج الاستجابة للمفردة.

ويُفسّر هذا الإحصائي في ضوء توزيع كاي تربيع (χ^2) بدرجة حرية تساوي عدد المجموعات ناقص واحد ($df = k - 1$)، ويُستخدم كبديل لتحليل التباين في الحالات التي تتطلب تقديرات أكثر ملاءمة لخصائص بيانات نظرية الاستجابة للمفردة. ويُعد مناسبًا لمقارنة المؤشرات المستخرجة من بيانات تجريبية، مثل معاملات الثبات المقدرة باعتماد طريقتي الاحتمالية القصوى أو بايز، مما يعزز دقة التفسير حول مدى تأثير نسب الأداء التفاضلي على الخصائص السيكومترية للاختبار، وذلك كما هو مبين في الجدولين (19) و (20).

الثبات النظري:

جدول 19: 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 إحصائيًا (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 20: 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 إحصائيًا (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 (21).

جدول 21: 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 [(0.50)—(1.49)]	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 [(1.50)—(2.49)]	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 (α=0.05) 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 (22).

جدول 22: 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 جوهرية (α=0.05) بين الأوساط الحسابية لمجموعات الدراسة تُعزى لمجموعات الدراسة بنسب المفردات ذات الأداء التفاضلي (0%، 10%، 20%، 30%)، ولفئات القدرة وللتفاعل بينهما. ولكون مجموعات الدراسة وفئات القدرة والتفاعل بينهما 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 والتفاعل بينهما، 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 (23) و(24).

جدول 23: اختبار جيمس- هويل للمقارنات البعدية بين مجموعات الدراسة.

المتغير التّابع: 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 إحصائيًّا (α=0.05)

يتضح من خلال الجدول وجود فروق جوهرية (α=0.05) بين الأوساط الحسابية لكمية معلومات الاختبار وفقًا لمجموعات الدراسة؛ لصالح المجموعة المرجعية مقارنة بالمجموعة المستهدفة بنسبة (20%)، ومن ثم المجموعة المستهدفة بنسبة (30%)، كما يتضح وجود فروق جوهرية (α=0.05) بين الأوساط الحسابية لكمية معلومات الاختبار وفقًا لمجموعات الدراسة لصالح المجموعة المستهدفة (10%) مقارنة بالمجموعة المرجعية.

جدول 24: اختبار جيمس- هويل للمقارنات البعدية بين فئات القدرة.

المتغير التّابع: 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)

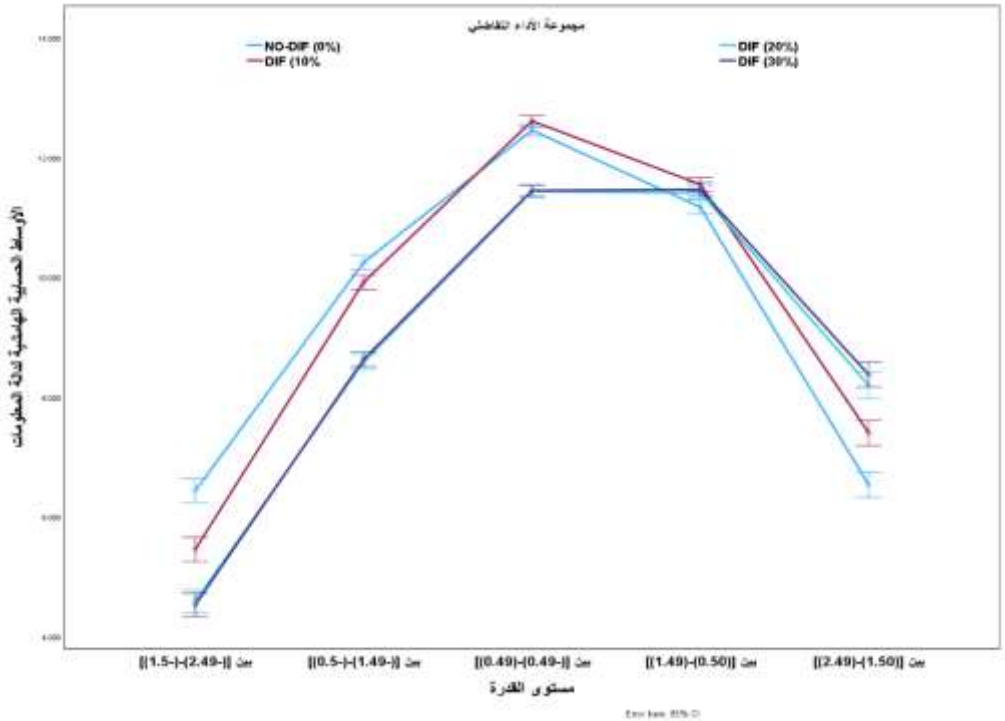
يتضح من خلال الجدول السابق وجود فروق جوهرية (α=0.05) بين الأوساط الحسابية لكمية معلومات الاختبار وفقًا لفئات القدرة قد كانت على الترتيب: لصالح فئة القدرة بين [(0.49)-)— (0.49)] مقارنة بفئة القدرة [(2.49)-)— (1.5-)]، ومن ثمّ فئة القدرة [(1.50)— (2.49)]، ومن ثمّ فئة القدرة [(1.49)-)— (0.5-)]، ومن ثمّ فئة القدرة [(0.50)— (1.49)]. كما

يتضح وجود فروق جوهرية (α=0.05) 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 يتضح وجود فروق جوهرية (α=0.05) 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)]. كما يتضح وجود فروق جوهرية (α=0.05) 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 (22) وجود فروق جوهرية (α=0.05) بين الأوساط الحسابية لمجموعات الدراسة وفئات القدرة تُعزى إلى التفاعل بينهما، ونظرًا لصعوبة تحديد مصادر الفروق الجوهرية بين الأوساط الحسابية وفق تفاعل متغيري مجموعات الدراسة 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 (25)

شكل 5: تفاعل متغيري المجموعات وفئات القدرة



جدول 25: اختبار بونفيروني للمقارنات البعدية الثنائية بين مجموعات الدراسة وفئات القدرة وفق التفاعل بينهما.

المتغير التّابع: 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

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

* دال إحصائيًّا (α=0.05)

يتضح من خلال الجدول السابق وجود فروق جوهرية (α=0.05) 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.

كما يوجد فروق دالة احصائيًا (α=0.05) بين الأوساط الحسابية لكمية معلومات الاختبار عند فئات القدرة المتوسطة بين [(-0.49)—(+0.49)] تُعزى لنسبة المفردات ذات الأداء التفاضلي (10%) مقارنة بالمجموعة المرجعية (0%).

كما يلاحظ عدم وجود فروق دالة احصائيًا (α=0.05) بين الأوساط الحسابية لكمية معلومات الاختبار عند فئات القدرة المنخفضة [(-2.49)—(-1.50)] وفئة القدرة (-1.49)—(-0.50)] 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 تقديرات القدرة في نظرية الاستجابة للمفردة ثابت بتغير مفردات وظروف الاختبار وهو ما يُعرف بخاصية اللاتغير (Invariance) (التقي، 2009). 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%)، وتتفق هذه النتيجة مع نتيجة (النوافلة، 2013) بعدم تأثر تقدير قدرات الأفراد بنسب المفردات ذات الأداء التفاضلي (10%، 20%، 40%) عندما كان مقدار الأداء التفاضلي (1.5)، عند النموذج ثلاثي المعلمة. كما تتفق جزئيًا مع دراسة (Lee & Zhang,2016)، في عدم تأثر تقديرات قدرات

الأفراد بنسب المفردات ذات الأداء التفاضلي (5%، 10%) وتختلف مع نسبة (20%) والتي تشير إلى وجود تأثير لنسبة المفردات ذات الأداء التفاضلي على تقديرات القدرة عند النموذج ثنائي المعلمة. كما أنها تختلف مع دراسة الوليلي والدوسري (2019)، والتي توصّلت إلى نتيجة مفادها أن للأداء التفاضلي تأثير على دقة تقديرات القدرة عندما كانت نسبة المفردات ذات الأداء التفاضلي (27%) حيث أن حذف المفردات ذات الأداء التفاضلي أدى إلى انخفاض قيمة متوسط الخطأ المعياري في تقدير القدرة. وقد يرجع سبب الاختلاف إلى أن دراسة الوليلي والدوسري (2019) استخدمت اختبار (ت) للفرق بين متوسطي الأخطاء المعيارية لقدرات الأفراد قبل وبعد حذف المفردات ذات الأداء التفاضلي، والتي اعتمدت على المتوسط الكلي للخطأ المعياري عند تقدير قدرة كل فرد، وليس على قيمة اختبار (ت) القائم على الخطأ المعياري في تقدير قدرة كل فرد؛ مما أدى إلى ضياع حقيقة أثر نسبة المفردات ذات الأداء التفاضلي. كما قد يرجع السبب إلى أن دراسة الوليلي والدوسري (2019) استخدمت النموذج أحادي البارامتر (نموذج راش) بخلاف الدراسة الحالية والتي استخدمت النموذج ثنائي المعلمة. وهذا مايتفق مع دراسة ضعضع (2023) والتي أكدت على أن لنموذج نظرية الاستجابة للمفردة أثر في اختلاف نتائج الكشف عن الأداء التفاضلي، وكان أفضل نموذج في الكشف عن الأداء التفاضلي هو النموذج ثنائي المعلمة.

كما توصلت الدراسة إلى عدم وجود فروق دالة إحصائيًا 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؛ بسبب اقتصار الدراسة على الأداء التفاضلي المنتظم وعدم تطرقها للأداء التفاضلي الغير منتظم الذي يرتبط بمعلمة التمييز. وتتفق الدراسة مع ما توصل إليه النوافلة (2013) بأن معلمة التمييز لا تتأثر بنسب المفردات ذات الأداء التفاضلي (10%، 20%، 40%).

كما توصلت الدراسة إلى عدم وجود فروق دالة إحصائيًا (α=0.05) لمعلمة الصعوبة تُعزى إلى نسب المفردات ذات الأداء التفاضلي (10%، 20%)، بينما توجد فروق دالة إحصائيًا (α=0.05) لمعلمة الصعوبة تُعزى إلى نسبة المفردات ذات الأداء التفاضلي (30%)، ويعود السبب في ذلك إلى أنه تمّ إجراء تغيير في معلمة الصعوبة مقدارها (1.5) على المجموعات المستهدفة، إلا أنّ المجموعتين المستهدفتين بنسب مفردات ذات أداء تفاضلي (10%،20%) لم تتأثر بحجم الأداء التفاضلي، وتتفق الدراسة الحالية مع دراسة النوافلة (2013) في أنه لم يطرأ تغيير على معلمة الصعوبة عندما كانت نسب المفردات ذات الأداء التفاضلي (10%،20%) بينما طرأ التغيير عندما كانت نسبة المفردات ذات الأداء التفاضلي (40%).

وقد يرجع السبب في ذلك إلى أن المفردات ذات الأداء التفاضلي عند النسب المنخفضة والمتوسطة (10%،20%) لم تؤثر على قوة معاملات الارتباط بين البنود لمعلمة الصعوبة مما أدى إلى عدم ظهورها على أنها فقرات تبدي أداء تفاضلي، أما مايتعلق بالنسبة العالية (30%) فقد تأثرت معلمة الصعوبة بوفرة تلك المفردات التي تبدي أداء تفاضلي مما أدى إلى تشوه في قوة العلاقات الارتباطية بين الفقرات التي تبدي أداء التفاضلي.

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 النظري 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 يتضح وجود فروق جوهرية (α=0.05) 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)]، لصالح نسبة المفردات ذات الأداء التفاضلي (10%، 20%، 30%)، 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. وتتفق هذه النتيجة مع ماتوصل إليه الخطيب (2014) بأن كمية معلومات الاختبار تتأثر عند مستويات القدرة العالية وعند جميع المواقف البحثية (25%، 50%، 75%). كما يوجد فروق دالة احصائيًا (α=0.05) بين الأوساط الحسابية لكمية معلومات الاختبار عند فئات القدرة المتوسطة بين [(-0.49)—(+0.49)] تُعزى لنسبة المفردات ذات الأداء التفاضلي (10%) مقارنة بالمجموعة المرجعية (0%).

كما يلاحظ عدم وجود فروق دالة احصائيًا (α=0.05) بين الأوساط الحسابية لكمية معلومات الاختبار عند فئات القدرة [(-2.49)—(-1.50)]، وفئة القدرة (-1.49)—(-0.50)] تُعزى لنسبة المفردات ذات الأداء التفاضلي (10%، 20%، 30%).
وتتفق نتائج كمية المعلومات للاختبار مع نتيجة دراسة الخطيب والعمري (2023) بأن دالة معلوماتية الاختبار تقدر بشكل مختلف عبر مستويات القدرة. حيث كانت نتيجة كمية المعلومات في هذه الدراسة أكبر عند مستويات القدرة المرتفعة، وكان هناك تراجع في كمية المعلومات عند مستويات القدرة المتوسطة والمنخفضة.
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 نسب المفردات ذات الأداء التفاضلي (10%، 20%، 30%)، إلا أن التأثير كان على دقة تقدير معلمة الصعوبة عندما كانت نسب المفردات ذات الأداء التفاضلي (30%)، ووجود تأثير على المتوسطات الحسابية لكمية معلومات الاختبار تُعزى إلى التفاعل بين مجموعات الدراسة وفئات القدرة عندما كانت مستويات القدرة مرتفعة؛ لصالح المجموعات المستهدفة بنسب مفردات ذات أداء تفاضلي (10%، 20%، 30%).

التوصيات والمقترحات:

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

- التأكيد على الباحثين والعاملين في مجال الاختبارات النفسية والتربوية، على الكشف عن الأداء التفاضلي، وخلو الاختبار والمقاييس من الأداء التفاضلي؛ بسبب تأثيره على الخصائص السيكومترية للاختبار وبنوده؛ مما ينتج عنه تشويه في نتائج الاختبار.
- اجراء دراسة حول طريقة والد تست (Wald test) في الكشف عن الأداء التفاضلي، وعقد مقارنات بينها وبين الطرق الأخرى في الكشف عن الأداء التفاضلي.
- اجراء المزيد من الدراسات حول تأثير الأداء التفاضلي على الخصائص السيكومترية للاختبار وبنوده مع التركيز على أهمية اجراء تكرار في عملية توليد البيانات (Replication) لنتائج أكثر موثوقية.
- اجراء دراسات مقارنة بين الأداء التفاضلي المنتظم والأداء التفاضلي الغير منتظم وأثرهما على الخصائص السيكومترية للاختبار وبنوده.

- **المراجع:**

أبو صهيون، محمد عبدالجبار عبدالرحمن. (2014). *فاعلية نموذج راش الهرمي في تقدير صعوبة الفقرة وقدرة الفرد في مواقف تختلف في نسبة الفقرات ذات الأداء التفاضلي ودرجته* (رقم المنشور 741813)[أطروحة دكتوراة، جامعة اليرموك-الأردن]. قاعدة بيانات دار المنظومة، الرسائل الجامعية.

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*،2 (7)، 139-165. https://doi.org/10.12816/0032962

التقي، أحمد محمد. (2009). *النظرية الحديثة في القياس*. دار المسيرة.

تيغزة، أمحمد. (2009). البنية المنطقية لمعامل ألفا كرونباخ، ومدى دقته في تقدير الثبات في ضوء افتراضات نماذج القياس .*مجلة العلوم التربوية والدراسات الاسلامية*، جامعة الملك سعود، 21(3)، 637 – 688.

اجبارة، محمد عدنان محمد. (2015). *أثر نسبة الفقرات ذات الأداء التفاضلي على الخطأ من النوع الأول وقوة الاختبار في الكشف عنه باستخدام طريقتي الانحدار اللوجستي ونسبة الأرجحية العظمى* (رقم المنشور759470)[أطروحة دكتوراة، جامعة اليرموك]. قاعدة بيانات دار المنظومة، الرسائل الجامعية.

الحاج محمود، فريال محمود محمد. (2018). الأداء التفاضلي لفقرات ومموهات اختبار محكي المرجع في العلوم للصف السادس الابتدائي مصمم وفق نظرية استجابة الفقرة. *مجلة جامعة القدس المفتوحة للبحوث الإنسانية والاجتماعية*،44(2)،123-135.028-044-000-0507/10.33977/DOI

حبيش، بشير؛ لكحل، مصطفى وبراهيمي، سعاد. (2022). نظرية الاستجابة للمفردة الاختبارية (افتراضاتها ونماذجها). *مجلة العلوم الإنسانية والاجتماعية*، 8(1)، 198-211.

الحربي، آلاء نواف؛ المدني، فاطمة رمزي أحمد. (2021). الكشف عن أثر نسبة المفردات ذات الأداء التفاضلي على قوة وفاعلية اختبار منحنى خصائص المفردة. *مجلة العلوم التربوية والنفسية*، 5(14)،46-62 DOI:10.26389/AJSRP.H241120

الخطيب، حسين محمود أحمد. (2014). *علاقة الأداء التفاضلي للفقرات بالأداء التفاضلي للاختبار ودالة معلوماته* (رقم المنشور741826) [أطروحة دكتوراة، جامعة اليرموك]. قاعدة بيانات دار المنظومة، الرسائل الجامعية.

الخطيب، إسراء محمد؛ العمري، حسّان غازي.(2023). أثر نوع ونسبة الفقرات ذات الأداء التفاضلي على معلمة القدرة للأفراد وفق النموذج ثلاثي المعلمة ونموذج موكن اللامعلمي لنظرية الاستجابة للفقرة. *المجلة الأردنية للعلوم التطبيقية - سلسلة العلوم الإنسانية*، 8(1)، 1-17. https://doi.org/10.35192/jjoas-h.v38i1.651

رشوان، ربيع عبده أحمد. (2021). الأداء التفاضلي للبنود وتأثيره على الأداء التفاضلي للاختبار باستخدام نماذج نظرية الاستجابة للمفردة والتحليل العاملي التوكيدي متعدد المجموعات. *مجلة العلوم التربوية والدراسات الإنسانية*، 6(15)،44-93.002-015-000-2152/10.55074/DOI:

الشريفين، نضال كمال. (2018). أثر نوع الأداء التفاضلي للفقرات على الخصائص السيكومترية للفقرات والاختبار وفق النماذج المعلمية والنماذج اللامعلمية لنظرية الاستجابة للفقرة. *دراسات: العلوم التربوية*،4(45)،632-605. https://search.emarefa.net/detail/BIM- 909188

ضعضع، هبة عبداللطيف. (2023). أثر اختلاف نموذج الاستجابة للمفردة (1PL,2PL,3PL) في الأداء التفاضلي للمفردة. *مجلة الأستاذ للعلوم 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.*

علي، سعدي أحمد؛ عيّال، ياسين حميد.(2021). تأثير اختلاف مستوى القدرة في الأداء التفاضلي وفقًا للأنموذج ثنائي المعلم لنظرية الاستجابة للفقرة. *مجلة الأداب* 138(1)، 137-162.

القضاة، عبدالحميد. (2018). *أثر طول الاستبانة على دقة تقدير معالم القدرة والخصائص السيكومترية للفقرة والمقياس في ضوء نظرية الاستجابة للفقرة*. [أطروحة دكتوراة غير منشورة]. جامعة اليرموك.

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

الناغي، هبه إبراهيم. (2011). أثر عدد مفردات الاختبار على تقدير قدرات الأفراد ودالة المعلومات لاختبار تحصيلي مرجعي المحك في الرياضيات باستخدام نماذج نظرية الاستجابة للمفردة. *مجلة كلية التربية*، *10*(10)،676-721. 10.21608/JFTP.2011.40282

النوافلة، علي صالح. (2013). *أثر الفقرات ذات الأداء التفاضلي المنتظم في تقديرات معالم فقرات الاختبار والأفراد في ضوء بعض المتغيرات باستخدام بيانات مولدة وفقًا لنموذج استجابة الفقرة ثلاثي المعلمة*. [أطروحة دكتوراة- جامعة اليرموك].

وردات، عبدالله محمد؛ الحسن، أحمد ضياء الدين. (2023). *الدور التربوي وعلاقته بالدور التعليمي لجمعيات ومراكز تحفيظ القرآن الكريم في المملكة الأردنية الهاشمية من وجهة نظر أولياء الأمور والعاملين فيها* (دراسة غير منشورة). [جامعة اليرموك].

الوليلي، إسماعيل حسن فهيم، والدوسري، سعيد بن عبدالله مبارك. (2019). 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. مجلة كلية التربية، 16(89)،156-201.
DOI: 10.21608/jfe.2019.129797

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 Measurement in Education (2014). *Standards for Educational and 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

Kim, S. H., Cohen, A. S., & Lin, Y. H. (2005). LDID: A computer program for local dependence indices for dichotomous items. *Applied Psychological Measurement, 29*(5), 364–365. https://doi.org/10.1177/0146621605276933

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 Educatio*n. 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

Van der Linden, W.J. (2018). *Handbook of Item Response Theory VOLUME Wells THREE Applications*. https://doi.org/10.1201/9781315117430

Wells, C. S., & Faulkner-Bond, M. (Eds.). (2016). *Educational measurement: From foundations to future*. Guilford Press.

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