أثر طول الاختبار وعدد بدائل الإجابة وحجم العينة على دقة تقدير قدرات الافراد ومعلمات المفردة وفقا لنموذج ثلاثي المعلمة

د . محمد بن حسن يحيى الزبيدي

أستاذ القياس والتقويم المشارك بقسم علم النفس - جامعة الطائف

m.alzubidy@tu.edu.sa

ملخص:

يهدف البحث الحالي الى فحص تأثير الاختلافات في طول الاختبار، عدد خيارات الاستجابة وحجم العينة على دقة تقدير القدرات الافراد ومعلمات للمفردة وفقا لنموذج لوجستي ثلاثي المعلمات باستخدام الأساليب التجريبية .تم اختبار البيانات التي تهدف إلى مقارنة معلمات الفقرة بدقة تقديرات قدرات الفرد مقابل سبعة وعشرين حالة اختبار بناء على نموذج لوغاريتمي ثلاثي المعلمات.

وعليه؛ فقد تم توليد عينات بحجم 500)، 1000، (2500فرد، واختبارات بأطوال 30)، 40، (50مفردة، وبعدد بدائل استجابة 3)، 4، (5بديل .وتمت معالجة البيانات باستخدام البرامج الإحصائية R, Bilog-Mg3, LDID, SPSS, Excel وتوصل الباحث إلى النتائج التالية :لم تتأثر دقة تقدير القدرة ودقة تقدير معلمتي التمييز والصعوبة بطول الاختبار، بينما تأثرت دقة تقدير معلمة التخمين بطول الاختبار .وكانت الفروق في دقة تقدير القدرة لصالح عدد البدائل الأكبر عند مقارنة أي منهما. كما أشارت النتائج إلى أن دقة تقدير معلمة التمييز لا تتأثر بعدد بدائل فقرات الاختبار بينما دقة تقدير كل من معلمتي الصعوبة والتخمين تتأثر بعدد بدائل فقرات الاختبار، وأن دقة تقدير القدرة كانت أكبر عندما كان حجم العينة أقل(500) فرد، ودقة تقدير معلمة التمييز كانت أكبر عندما كان حجم العينة أكبر .(2500)

الكلمات المفتاحية :طول الاختبار، عدد بدائل الاستجابة، حجم العينة، معالم الفقرات، قدرات الأفراد، النموذج اللوجستي ثلاثي البارامتر.

The effect of test length, number of answer alternatives, and sample size on the accuracy of estimating individuals' abilities and item parameters according to a three-parameter model

Preparation
Dr . Muhammad bin Hassan Yahya Al-Zubaidi
Associate Professor of Measurement and Evaluation
Department of Psychology, Taif University
m.alzubidy@tu.edu.sa

ABSTRACT

The current study aimed to investigate the effect of varying test length, number of response alternatives, and sample size on the accuracy of estimating individuals' abilities and item parameters according to the three-parameter logistic model using the experimental approach. Data aimed at comparing the accuracy of estimating individuals' abilities and item parameters for twenty-seven test conditions were generated based on the three-parameter logit model, and each test condition was defined by a combination of three factors: test length, number of response alternatives, and sample size.

Accordingly; Sample sizes of (500, 1000, 2500) individuals, tests of lengths of (30, 40, 50) and number of response alternatives (3, 4, 5) were generated. The data was processed using the statistical programs: R, Biolog-Mg3, LDID, SPSS, and Excel, and the most important findings are: The accuracy of both ability estimation, discrimination and difficulty parameters were not affected by the length of the test, while the accuracy of estimating the estimation parameter was affected by the length of the test. The differences in accuracy of ability estimation were in favor of the larger number of alternatives when comparing either. The results also indicated that the accuracy of the discrimination estimation parameter is not affected by the number of alternatives to the test items, while the accuracy of both difficulty and guessing estimation parameters is affected by the number of alternatives to the test items, and that the accuracy of ability estimation was greater when the sample size was less (500) individuals, and the accuracy of the discrimination ability estimation parameter was larger when the sample size was larger (2500).

Keywords: Test length, Number of response alternatives, Sample size, Item parameters, Individual abilities, Three-parameter logistic model.

مقدمة :

تعد نظرية القياس التقليدية أو ما تسمى بالنظرية الكلاسيكية للاختبارات Classical Test Theory (CTT) من أقدم النظريات في مجال القياس والتقويم والتي استخدمت لفترة طويلة في بناء الاختبارات النفسية والتربوية وتطويرها وتحليلها وتفسير البيانات المستخرجة منها .(Gregory, 2014)

كما تعتبر الاختبارات الأداة الأبرز من أدوات تقويم أداء الطالب لكونها تعمل على إعطاء قياسات تناسب كثيرًا من مخرجات الطالب التعليمية .وهي الأداة الأكثر استخداماً في تقدير مدى امتلاك الفرد لسمة أو قدرة معينة من خلال استجابته لمجموعة من المثيرات التي تمثل تلك السمة أو القدرة.

إن المتتبع لمجال القياس والتقويم يجد أن نظرية القياس التقليدية أو ما تسمى بنظرية الاختبار (CTT) على الرغم من أنها استخدمت بشكل كبير في تصميم وبناء الاختبارات النفسية والتربوية إلا أنها واجهتها العديد من العيوب والانتقادات، فقد اعترتها جوانب قصور كبيرة في تحقيق الموضوعية، وعدم وجود وحدة ثابتة للقياس، بالإضافة إلى القياس في أكثر من بعد .كذلك اعتماد الظاهرة السلوكية في قياسها وتقديرها على الأداة المستخدمة في القياس والعينة التي تم تطبيق تلك الأداة عليها)علام، .(2005

كل تلك الانتقادات وغيرها أدت بعلماء القياس للقيام بمحاولات بحثية متتالية لتطويرها ومواجهة أوجه القصور المتعلقة بها .حيث تم إجراء العديد من الدراسات والبحوث من أجل التغلب على جوانب القصور ومن ثم الوصول إلى قياس موضوعي يشبه القياس الفيزيائي الذي لا تتأثر نتائج القياس فيه باختلاف الأداة المستخدمة ولا العناصر التي استخدمت هذه الأداة التي تتدرج بوحدة قياس مطلقة وثابتة .ولقد أثمرت تلك الجهود في العقود الأخيرة بالتوصل إلى نظرية حديثة تم إطلاق عدة مسميات لها منها :نظرية السمات الكامنة، أو نظرية الاستجابة للمفردة Item Response Theory (IRT) التي تتميز بأن تقدير معالم الفقرات :الصعوبة،

التمييز والتخمين مستقل عن قدرة الأفراد التي استخدمت في تقدير هذه المعالم، كما أن تقدير قدرات الأفراد يكون مستقلاً عن عينة الفقرات المستخدمة في عملية التقدير .(Bond & Fox, 2015)

انبثقت عن هذه النظرية (IRT) نماذج رياضية لوغاريتمية تقوم على افتراضات معينة، تهدف إلى تحديد العلاقة بين أداء الفرد في الاختبار، وبين القدرات التي تكمن خلف هذا الأداء وتفسره .ومن أكثر هذه النماذج شيوعاً وأشهرها :النموذج أحادي البارامتر أو ما يسمى بنموذج راش Rasch Model، والنموذج ثنائي البارامتر والذي يسمى نموذج لورد Lord Model، والنموذج ثلاثي البارامتر لبيرنبوم Birnbaum Model) علام، .(2005

يعتبر النموذج اللوجستي ثلاثي البارامتر هو الحالة العامة بين النماذج الثلاثة السابقة، حيث يتم وصف المنحنى المميز للمفردة وفق هذا النموذج من خلال بارامترات ثلاثة هي بارامتر صعوبة الفقرة، بارامتر التمييز وبارامتر التخمين .كما يتميز هذا النموذج الثلاثي عن النموذجين الآخرين بمراعاة عامل التخمين الذي يتوقع أن يحصل في اختبارات الاختيار من متعدد مما قد يؤثر على دقة تقدير قدرات الافراد في مثل هذه الأنواع من الاختبارات)آن أنستازي، سوزانا أوربينا، .(2015

يتبين من بعض الدراسات وأدبيات القياس تأثر معالم الفقرات وقدرات الأفراد أيضاً باختلاف طول الاختبار وحجم العينة وكذلك عدد البدائل .وقد ساعد توافر البرامج الحاسوبية في إيجاد إحصائيات نماذج نظرية الاستجابة للفقرة ومنها النموذج ثلاثي البارامتر .لذا فإن هذه الدراسة تعد محاولة للتعرف على أثر اختلاف طول الاختبار وعدد بدائل الاستجابة وحجم العينة في دقة تقدير قدرات الأفراد ومعالم الفقرات وفق النموذج اللوجستي ثلاثي البارامتر، حيث أن هذا الميدان مازال بحاجة إلى المزيد من الدراسات والبحوث.

مشكلة الدراسة وتساؤلاها:

تعتبر نظرية الاستجابة للفقرة تطورًا هائلاً وتقدماً كبيرًا في مجال القياس النفسي والتربوي كونها تراعي عدد من المتغيرات أغفلتها النظرية التقليدية، مما أدى إلى قياس أكثر دقة وموضوعية في تقدير قدرات الأفراد .فبينما تتنوع وتزداد عيوب النظرية التقليدية في القياس، فإن مزايا نظرية الاستجابة للفقرة بنماذجها المختلفة تتعدد وتظهر في استقلال خصائص الفقرات عن الأفراد، واستقلال تقدير قدرات الأفراد عن الفقرات، كما أن تفسير درجاتهم يتم في ضوء الفقرات وليس في ضوء الجماعة المرجعية كما الحال في النظرية التقليدية .وبالنسبة لتقدير الخطأ المعياري فإنه يتم تقديره لكل مختبر على حده، مع القدرة على المقارنة بين أداء الأفراد الذين تم تطبيق مقاييس مختلفة عليهم لنفس السمة)علام،.(2005

تعتبر اختبارات الاختيار من متعدد أداةً شائعة لقياس القدرة أو التحصيل إلا أن هذا النوع لا يخلو من بعض المآخذ، ففقرات الاختبار تتكون من مجموعة صغيرة من بدائل الاستجابة لكل فقرة، وإذا لم يتعرف الفرد على الإجابة الصحيحة فمن الممكن أن يلجأ إلى تخمين الإجابة الصحيحة .والتخمين الصحيح في هذه الحالة يمكن أن يؤدي إلى زيادة الدرجة النهائية للفرد في ذلك الاختبار عن الدرجة المتوقعة منه، كما أنه يتعارض مع استخلاص القدرة الحقيقية للفرد من الاستجابة الملاحظة في الاختبار، خاصة أنه يصعب معرفة هل كانت تلك الاستجابة الصحيحة عن معرفة وإلمام بالمحتوى الذي تم التعرض له أم كانت بأثر التخمين.

لقد قام علماء القياس بالبحث والتطوير في نظرية الاستجابة للفقرة باعتبارها إحدى الطرق للوصول إلى الدرجة الحقيقية لقدرات الأفراد حتى مع مشكلة التخمين التي قد تواجههم في اختبارات الاختيار من متعدد، ذلك لأنها نظرية اختبار تعمل على محاولة تفسير أداء الفرد في فقرة اختبارية)استجابة (باعتبارها وظيفة لخاصية)غير ملحوظة (من خصائص الفرد)قدرة (إضافة إلى مؤشرات فقرة الاختبار نفسه، كما يمكن التنبؤ بأداء الفرد في فقرة اختبارية معينة عن طريق

كمية القدرة التي يتمتع بها هذا الفرد (Hambleton, Swaminathan & Rogers, 1991).

إن التطبيقات العملية لنظرية الاستجابة للفقرة لا يمكن الاستفادة منها في حال عدم توافر افتراضاتها والتي منها أحادية البعد، والمنحنى المميز للفقرة .وقد تحدث عدم ملائمة البيانات مع النموذج أحادي البارامتر لأن المنحنى المميز للفقرة يختلف في بارامتر التمييز أو التخمين، كما قد تحدث عدم الملائمة مع النموذج ثنائي البارامتر لأن المنحنى المميز للفقرة يختلف في بارامتر التخمين (Linden & Hambleton, 1997).وعليه؛ فإن عدم ملائمة البيانات ذات المستوى المنخفض من القدرة للنموذجين أحادي وثنائي البارامتر يُعدّ منطقياً، ذلك لأن الأفراد مع هذا المستوى من القدرة يلجؤون إلى التخمين في بدائل الاختيار من متعدد مما يؤدي إلى زيادة بارامتر التخمين عن الصفر الأمر الذي يتعارض مع افتراضات النموذجين، وبالتالي فإن النموذج ثلاثي البارامتر يعطي نتائج أفضل ودقة في القياس لاحتوائه بارامتر التخمين.

على الصعيد ذاته، بالرغم من وفرة ما كُتب عن نظرية الاستجابة للفقرة في الأدب العربي أو الأجنبي واستخداماتها إلا أن المجال يحتاج إلى المزيد من البيانات التجريبية ونتائجها وفق شروط ومعايير مختلفة .لذا وبناءً على ما تم عرضه من طرح نظري فإن اهتمام هذه الدراسة ينصبّ على معرفة تأثير طول الاختبار، وعدد بدائل الاستجابة، وكذلك حجم العينة في تقدير معالم الأفراد والفقرات باستخدام النموذج ثلاثي البارامتر.

وتحديدًا فإن هذه الدراسة تسعى للإجابة عن السؤال الرئيس التالي" :**ما أثر اختلاف طول الاختبار وعدد بدائل الاستجابة وحجم العينة في دقة تقدير قدرات الأفراد ومعالم الفقرات وفق النموذج اللوجستي ثلاثي البارامتر؟"**

ويتفرع عن هذا السؤال مجموعة من الأسئلة الفرعية تتمثل بالآتي:

١. ما أثر طول الاختبار في تقدير معلمة القدرة للأفراد وفق النموذج اللوجستي ثلاثي البارامتر؟

٢. ما أثر طول الاختبار في تقدير معالم الفقرات وفق النموذج اللوجستي ثلاثي البارامتر؟
٣. ما أثر عدد بدائل الاستجابة في تقدير معلمة القدرة للأفراد وفق النموذج اللوجستي ثلاثي البارامتر؟
٤. ما أثر عدد بدائل الاستجابة في تقدير معالم الفقرات وفق النموذج اللوجستي ثلاثي البارامتر؟
٥. ما أثر حجم العينة في تقدير معلمة القدرة للأفراد وفق النموذج اللوجستي ثلاثي البارامتر؟
٦. ما أثر حجم العينة في تقدير معالم الفقرات وفق النموذج اللوجستي ثلاثي البارامتر؟

أهداف الدراسة:

تسعى الدراسة الحالية إلى تحقيق الأهداف التالية:

١. التحقق من افتراضات نظرية الاستجابة للفقرة على استجابات أفراد عينة الدراسة على نماذج الاختبار المختلفة.
٢. التعرف على تقديرات قدرات الأفراد في اختبارات الاختيار من متعدد المستخدمة في الدراسة في ضوء النموذج اللوجستي ثلاثي البارامتر.
٣. التعرف على مدى دقة النموذج ثلاثي البارامتر في تقدير قدرات الأفراد ومعالم الفقرات من خلال ظروف معينة افتراضية تتضمن أطوال متفاوتة لاختبار وأحجام عينات مختلفة.
٤. الكشف عن تأثير طول الاختبار، حجم العينة، وعدد بدائل اختبار الاختيار من متعدد في دقة النموذج ثلاثي البارامتر في تقدير قدرات الأفراد ومعالم الفقرات.

الإطار النظري والدراسات والبحوث السابقة

تمهيد:

لقد اهتم علماء القياس والعلوم الإنسانية وكذلك علماء العلوم الطبيعية بعملية القياس على حد سواء، بيد أن اهتمام علماء القياس النفسي أكثر من غيرهم بمنطق القياس والأساليب الكمية بسبب تعقيد الظواهر النفسية وتعدد متغيراتها . حيث تهتم عملية القياس النفسي بقياس السمات الكامنة بطرق غير مباشرة من خلال سلوك المفحوصين الملاحظ على أدوات القياس المختلفة والتي لها علاقة مباشرة بتلك السمات موضوع القياس .لذا يعتبر القياس من القضايا المهمة والرئيسية في مجال العلوم الإنسانية باستخدامه الأساليب الكمية للتعبير الموضوعي عن الظواهر النفسية المختلفة.

إن المتأمل في أدوات القياس المستخدمة في العملية التربوية أو التعليمية ليجد أن الاختبار من أكثرها شيوعاً واستخداماً وبخاصة على مستوى المدارس ومؤسسات التعليم المختلفة، فالاختبار يعد إجراء منظم يقارن أداء الفرد بمستوى محدد سابقاً . حيث يعتبر أداة القياس التي يتم إعدادها وفق مجموعة من الخطوات والإجراءات المحددة لمعرفة ما يمتلكه الفرد من درجة لقدرة معينة أو سمة من خلال تفاعله واستجابته على مجموعة من الأسئلة والمفردات)المثيرات (التي تعبر عن السمة المراد قياسها) .علام، .(2010

يعتبر قياس السمة النفسية قياساً غير مباشر، حيث لا يتم معرفة ما يمتلكه الفرد من تلك السمة مباشرة بل من خلال أدائه في مواقف لها علاقة بتلك السمة، أي من خلال سلوكه الظاهري القابل للملاحظة والقياس .كما أن عملية القياس تلك غير تامة، أي لا يتم قياس جميع السلوكيات المكونة للمجال السلوكي للسمة وإنما بقياس عينة فقط من هذا المجال .كذلك تتصف عملية قياس السمات النفسية بأنها نسبية، لأنه لا يمكن تفسير درجة الفرد على هذا الاختبار النفسي دون أن يتم مقارنة أداء الفرد بالمجموعة التي ينتمي إليها، أو بمعيار أو محك معين .كل

ذلك يجعل من احتمال وقوع الخطأ في القياس النفسي وارداً بشكل أكبر من وقوعه في القياس الطبيعي الفيزيائي، مما أدى إلى زيادة الاهتمام من قبل المختصين في القياس النفسي والتربوي)عودة، (2014، الأمر الذي أدى إلى ظهور نظريات تمكّن افتراضاتها إيجاد طرق لتقدير دقة القياس نتج عنها مجموعة من النماذج لتفسير نتائج تلك الأدوات، والتي على ضوئها يتم اتخاذ القرار المناسب .ومن تلك القرارات : تصنيف الطلبة، وترفيعهم لمرحلة أعلى وغير ذلك .وتعد الاختبارات التحصيلية إحدى تلك الأدوات، فهي توفر معلومات يمكن استخدامها في قياس مستوى تحصيل الطلبة، وتشخيص نواحي الضعف لديهم، وتصنيفهم والتمييز بينهم)علام، .(1986

يعرّف الاختبار التحصيلي بأنه" :عينة من الأسئلة أو المثيرات التي تمثل السمة أو القدرة المقاسة، مصاغة بطرقة منظمة لتحديد مستوى الطالب فيما يمتلكه من مهارات ومعلومات في المادة الدراسية التي تعلمها سابقاً وبشكل رسمي من خلال استجاباته) "عودة، .(2010

يعتبر الاختبار الموضوعي أحد أنواع الاختبارات التحصيلية، وهو حديث نسبياً بالمقارنة مع الاختبارات المقالية، وأطلق عليه موضوعياً لما يمتاز به من دقة وموثوقية، وعدم تأثره بذاتية المصحح)علام، .(2000كما يعد من أكثرها استخداماً وشيوعاً في العملية التعليمية وبخاصة اختبارات الاختيار من متعدد، حيث تقيس الفقرة الجيدة فيه الأهداف البسيطة والمركبة كذلك .ويتميز هذا النوع من الاختبارات بمجموعة من الميزات منها:

- كفاءته وبخاصة عند مراعاة البناء الجيد لفقراته.
- تعدد استخداماته في الدراسات، والمراحل التعليمية والدراسية المختلفة.
- يسمح بمعاينة أكبر قدر من المجال السلوكي للمقياس)ظاظا، (2012

اختبار الاختيار من متعدد:

يعدّ اختبار الاختيار من متعدد أحد أنواع الاختبارات الموضوعية التي يكثر المعلمون من استخدامها في اختباراتهم بسبب مرونة استخدامها في قياس المستويات

المعرفية المختلفة عن غيرها من أنواع الفقرات الاختبارية، وسهولة تصحيحها، وكذلك إمكانية استخدامها في مجالات دراسية مختلفة)علام، .(2009

ويعرّف بأنه عبارة عن سؤال يتكون من جزأين :الأول ويسمى رأس السؤال)الجذر(، وهو عبارة تقديمية للمشكلة المعروضة في جملة ناقصة أو أكثر .والجزء الثاني يمثل البدائل التي يتم اختيار الإجابة الصحيحة منها، حيث تكون الإجابة الصحيحة هي أحد تلك البدائل وما عداها تعتبر مشتتات أو مموهات تشبه الإجابة الصحيحة ولكنها خاطئة وتشتت المفحوص.

أوضح علام (2009)أن سؤال الاختيار من متعدد يشمل عبارة تقديمية تليها مجموعة من الإجابات المقترحة، وعلى المفحوص اختيار الإجابة الصحيحة من بين تلك الإجابات، تسمى الإجابات المقترحة)بدائل(، وتكون وظيفة العبارة التقديمية عرض السؤال المراد إجابته أو المهمة الواجب تأديتها أو المشكلة التي تحتاج إلى حل .أما الخيارات أو البدائل فإنها تحتوي إجابة واحدة صحيحة ويطلق على ما سواها مشتتات وظيفتها تقديم إجابات أو حلول تبدو صحيحة لأصحاب القدرة المتدنية، ولكنها ليست كذلك لأصحاب القدرة العالية.

إن الهدف الرئيسي من وضع البدائل هو إبعاد وتشتيت التفكير لدى المختبرين الذين لا يعرفون الإجابة الصحيحة، ومعرفة مدى حاجتهم للمعلومات الموصلة للإجابة الصحيحة، وتحديد جوانب الضعف لديهم عند توجههم نحو بديل خاطئ، ولذلك يفترض أن تكون تلك البدائل أو المشتتات قريبة من الإجابة الصحيحة (McMillan, 2004).

يؤكد كل من ألين، و يَن (1979)على ضرورة أن تكون بدائل الإجابة متجانسة حتى لا يلجأ المفحوص إلى استبعاد بعض الإجابات للتوصل للإجابة الصحيحة، وبالتالي يجب أن يكون طول البديل الصحيح أو الإجابة الصحيحة مساو لبقية البدائل أو المشتتات، ذلك لأن اختلاف أطوالهم يعطي المفحوص مؤشراً للإجابة الصحيحة .مع التأكيد أن تكون كل البدائل متناسبة مع مقدار المعلومة في جذر

السؤال .كما يجب مراعاة عمر المفحوص وخبرته التعليمية، فما يناسب فئة عمرية قد لا يناسب أخرى.

تتأثر الخصائص السيكومترية للاختبار، ومعالم فقراته بمجموعة من العوامل منها:

١. أسلوب الصياغة :يحتاج متن الفقرة أو جذر السؤال إلى صياغة دقيقة، بحيث يكون السؤال واضحاً ودقيقاً ومحددًا، ذا مضمون واحد عند جميع المفحوصين .ذلك لأن الفقرات ذات المتن غير المحدد وغير الواضح تؤدي إلى اختلاف فهم المطلوب لدى المفحوصين، وبالتالي زيادة احتمالية الاستجابة الغير صحيحة عليها (Gronland & Linn, 1990).

٢. أسلوب اختيار البدائل :عند اختيار بدائل الاستجابة أو الموهات يجب أن يؤخذ في الاعتبار أن تكون متجانسة في محتواها، لأن تجانسها يؤدي إلى زيادة جذب انتباه المفحوصين وبالتالي زيادة معامل صعوبة الفقرة .أما عدم تجانسها فإن ذلك يؤدي إلى انخفاض معامل صعوبتها وبالتالي انخفاض قدرتها التمييزية، الجدير ذكره هنا أن الفقرات الصعبة جدًا أو السهلة جدًا يكون أيضاً تمييزها ضعيف (Murphy & Hofer, 1994).

٣. شكل البدائل :تختلف بدائل الاختيار من اختبار لآخر، حيث ذكر (Gronlund, 1997) أن هناك فقرات اختبار تتطلب وجود)بديلاً (إجابة واحدة صحيحة يتم اختيارها من بين البدائل المعطاة، وهذا الشكل هو الأبسط والأكثر استخداماً في الاختبارات خصوصاً التحصيلية .وهناك فقرات تتطلب اختيار أفضل إجابة، أي أنها تحتمل حلولاً وإجابات مختلفة ويجب على المفحوص اختيار الأفضل .وهناك بعض البدائل المركبة بحيث يجمع البديل بين بديلين معاً (Crehan & Haladyan, 1993).كما أن هناك بعض الفقرات تتطلب اختيار أكثر من بديل، حيث يكون للسؤال أكثر من إجابة صحيحة، إلا أن هذا النوع لا يفضل استخدامه في الاختبارات التحصيلية.

٤. عدد البدائل :في فقرات الاختيار من متعدد تختلف عدد البدائل من اختبار لآخر، لكن يجب أن يكون عددها كافياً حتى تقل محاولات التخمين للوصول إلى الإجابة الصحيحة .ومع ذلك؛ يفترض ألا يزيد عدد البدائل عن خمسة بشرط أن تكون مناسبة ودقيقة، لأن العبرة ليست بكثرتها وإنما بنوعيتها .كما يفضل مع طلبة الصفوف المبكرة تقليل عدد تلك البدائل)علّام، .(2009

نظرية الاستجابة للفقرة:

تعتبر نتائج القياس النفسي أقل دقة من نتائج القياس في العلوم الطبيعية نظرًا لتنوع مصادر الأخطاء التي ترافق عملية القياس النفسي، لذلك اهتم علماء القياس النفسي بمنطق القياس والأساليب الكمية أكثر من غيرهم بسبب تعقد الظواهر النفسية وتعدد المتغيرات فيها، فقاموا بابتكار وتطوير لأساليب بناء أدواتهم وتحليل نتائجها حتى يصلون بها إلى أعلى درجة من الدقة والموضوعية، فظهرت ما تسمى بالنظرية الكلاسيكية للاختبار Classical Test Theory (CTT) التي قدمت نموذجاً لتقدير قدرة الفرد يعبر عنها بالدرجة الحقيقية الخالية من أخطاء القياس، لذلك أطلق عليها البعض نظرية الدرجة الحقيقية True Score Theory، وقد حاولت هذه النظرية الوصول إلى أقصى ما يمكن من الدقة في عملية القياس النفسي من خلال معامل الثبات الذي يمثل نسبة التباين الحقيقي إلى التباين الكلي)عودة، 2004).

يذكر علّام (2005)أن هذه النظرية تهتم بشكل رئيسي بأخطاء القياس، فعندما نريد قياس سمة معينة فإنه ينبغي تقدير مصادر الخطأ التي تؤثر في تباين درجات الاختبار بهدف التحقق من ثبات تلك الدرجات.

توصف افتراضات النظرية التقليدية للاختبار بالضعيفة، حيث أنه من السهولة تحقيقها في بيانات الاختبار الحقيقية لأن معظم تركيزها ينصب على خصائص سيكومترية بسيطة كمعاملات الصعوبة والتمييز، ومن تلك الافتراضات ما يلي:

١. الدرجة الملاحظة للفرد تعبر عن الدرجة الحقيقة (T) مضافاً إليها خطأ القياس (E)، بحيث .X= T+E

٢. الدرجة الحقيقية للفرد هي عبارة عن متوسط مجموع الدرجات الملاحظة التي حصل عليها في عدة مواقف اختبارية لنفس الاختبار أو صورًا مكافئة له.
٣. لا يوجد ارتباط بين الدرجات الحقيقية وأخطاء القياس لعدد من الأفراد على نفس الاختبار.
٤. لا يوجد ارتباط بين أخطاء القياس في اختبارات مختلفة لنفس الفرد (Allen & Yen, 1979).
٥. اعتدالية توزيع الدرجات على متصل القدرة التي يقيسها الاختبار، حيث عدد الأفراد عند مستوى معين من القدرة يناظر العدد المتوقع من دالة الكثافة الاعتدالية الاحتمالية.
٦. تساوي تباين أخطاء القياس لجميع الأفراد الذين يطبق عليهم الاختبار (علاّم، 2005)

لقد كان للنقد الموجه للنظرية التقليدية في القياس الأثر الكبير في البحث عن طرق وأساليب حديثة، فعلى الرغم من كثرة استخدامها من قبل المختصين في مجال التربية وعلم النفس إلا أنها تعاني من بعض المشكلات .فعلى سبيل المثال : تفترض أن الاختبار يقيس متغيرًا أحادي البعد، وهذا قد لا يكون صحيحاً دائماً، كما أنها تفترض أن الدرجات التي يحصل عليها الفرد على فقرات اختبار يمكن جمعها كما لو كانت تمثل تدريجاً خطياً، والفقرات المتعلقة بالمتغير المراد قياسه تحمل نفس المعنى لدى جميع الأفراد، كما أنها تفترض أن درجة الاختبار التي تمثل السمة المقاسة دالة خطية مطردة، وأيضاً فإن تكوين الاختبار يتغير بمرور الزمن (علاّم، 2002).

تكمن المشكلة الأساسية التي عانت منها النظرية التقليدية هي أن جميع الخصائص السيكومترية تعتمد بشكل رئيسي على خصائص الأفراد المفحوصين، فمعاملات الصعوبة والتمييز تتغير بتغير قدرات هؤلاء الأفراد الذين طبق عليهم الاختبار، حيث سيرتفع معامل صعوبة الفقرات إذا كانت قدرة أفراد العينة مرتفع عن

متوسط قدرة المجتمع الذي تنتمي له تلك العينة .كذلك معاملات التمييز ستميل إلى الارتفاع عندما يتم تقديرها اعتماداً على عينة غير متجانسة في القدرة أكثر مما لو تم تقديرها على عينة متجانسة في القدرة، ويتأثر أيضاً متوسط ومدى مستوى القدرة بقيم معالم الفقرات بحيث أن الأفراد الذين يتعرضون لفقرات ذات معاملات صعوبة مرتفعة سيظهرون بقدرة منخفضة، والأفراد الذين يتعرضون لفقرات معاملات صعوبتها منخفضة سيظهرون بقدرة مرتفعة (Hambleton, Swaminthan & Rogers, 1991).

كما أن المقارنة بين المفحوصين حسب السمة المقاسة عن طريق فقرات اختبار معين تكون محدودة في الحالات التي يجيبون فيها على نفس الاختبار، أو اختبار مكافئ، حيث معظم الاختبارات التحصيلية هي مناسبة لأصحاب القدرة المتوسطة . وعليه؛ فإن الاختبار لا يزودنا بتقديرات جيدة لقدرات الأفراد ذوي القدرات المرتفعة أو المنخفضة .إضافة إلى أن هذه النظرية تفترض ثبات الخطأ المعياري لمجتمع الأفراد الذين يطبق عليهم الاختبار، مع أن لدى بعضهم أداءً أكثر اتساق من غيرهم لاختلاف مستوى القدرة لديهم .(Hambleton & Swaminathan, 1985)

كما عانت النظرية التقليدية أيضاً بأنه ليس هناك تدريج مشترك بين الأفراد وفقرات الاختبار، حيث أنه بمجرد ملاحظة المجموع الكلي لدرجات كل فقرة وفرد نجد أن أكبر الدرجات تقابل الأكثر قدرة من الأفراد، وأن أقلها تقابل الفقرات ذوات معامل الصعوبة الأعلى، وبالتالي يصعب تفسير درجات الأفراد بالنسبة لدرجات الفقرات .حتى أن هذه النظرية لا توفر أي قاعدة أو أسلوب للتنبؤ بأداء الفرد على أي فقرة من فقرات الاختبار .مما سبق تتجلى لنا الحاجة الماسة إلى نظرية جديدة في القياس تلبي بعض الاحتياجات الموضوعية مما يلي:

١. عدم اعتماد خصائص معاملات الثبات وكذلك معاملات الصعوبة والتمييز للمفردات على العينة التي حصلنا منها على خصائص المقياس.
٢. ألا تكون الدرجات التي تصف الأفراد معتمدة على الاختبار.

٣. أن تكون نموذجاً يعبر عن مستوى الفقرة أكثر من تعبيره عن الاختبار .

٤. ألا تحتاج لوجود نماذج مكافئة من الاختبار لتحقيق ثباته.

٥. تزويد المختصين بقياس واضح وصريح لقدرات الأفراد (Hambleton, Swaminthan & Rogers, 1991).

كل تلك المطالب، مع التقدم العلمي نتج عنها ظهور نظرية سميت بنظرية الاستجابة للفقرة Item Response Theory (IRT) في محاولة لمواجهة جوانب الضعف والقصور التي واجهتها النظرية التقليدية للاختبار)الطريري، (1997، وقد أطلق عليها عدة مسميات منها :نظرية السمات الكامنة Latent Trait Theory (LTT)، ونظرية منحنى خصائص الفقرة Item Characteristic Curve Theory.

لقد جاءت نظرية الاستجابة للفقرة امتداداً للنظرية التقليدية للقياس ومكملة لمبادئها ومفاهيمها، حيث تغلبت على الكثير من مشكلات القياس التي لم تستطع مواجهتها والتغلب عليها)علام، .(2005

تهدف نظرية الاستجابة للفقرة ونماذجها الرياضية المرتبطة بها إلى تقدير جميع معالم الفقرات والأفراد، وكلما كانت هناك ملائمة بين مجموعة البيانات والنموذج المستخدم كلما أدى إلى الحصول على تقدير دقيق لتلك المعالم، بحيث يتم وضع الفقرات والأفراد على مقياس للقدرة من خلال عمليات التقدير طالما توجد علاقة تقارب ممكنة بين الاحتمالات المتوقعة للأفراد وأدائهم عند كل مستوى من مستويات القدرة.(Hambleton & Swaminathan, 1985)

تتفق نظرية الاستجابة للفقرة مع النظرية التقليدية في وجود متصل للسمة المقاسة، وهناك إمكانية تقدير احتمال إجابة الفرد على فقرة معينة إجابة صحيحة إذا عُرف موقع الفرد على متصل السمة، وأن مقدار احتمال إجابته إجابةً صحيحة يكون دالة متزايدة مطردة مع مواقع الأفراد على نفس المتصل، أي أن احتمال إجابة الفرد بشكل صحيح يتناسب طردياً ومقدار السمة التي يمتلكها)علام، .(1987

تميز نظرية الاستجابة للفقرة بين كل من الدرجة الحقيقية والدرجة المقدرة للفرد حيث أن قدرة الفرد ثابتة لا تتأثر بخصائص العينة بعكس النظرية التقليدية)عودة، (2010، وهي تتطلب إجراء العديد من العمليات الرياضية المطولة والمعقدة، الأمر الذي جعلها قبل انتشار برامج الحاسب الآلي غير عملية .(Baker, 2001)

إن أهم ما يميز نظرية القياس الحديثة عن التقليدية هو اهتمامها بنمط استجابة الفرد على فقرات الاختبار، بينما تركز التقليدية على درجة الاختبار الخام ككل، وسماحها بتقييم قياس الخطأ عند أي مستوى من مستويات القدرة، بعكس التقليدية التي تفترض تساوي الخطأ للجميع عند أي مستوى لدرجات الاختبار .(Kline, 2005)

افتراضات نظرية الاستجابة للفقرة:

على عكس النظرية التقليدية ذات النماذج الضعيفة التي يمكن تحقيق افتراضاتها بسهولة، وبواسطة بيانات الاختبار، فإن نظرية الاستجابة للفقرة تنطلق من افتراضات قوية ليس من السهل تحقيقها في البيانات التي ستطبق عليها نماذج النظرية لكي تؤدي بالتالي إلى نتائج يمكن الوثوق بها عند اختبار ملائمة البيانات للنموذج (Hambleton & Jones, 1993)، وتتلخص هذه الافتراضات فيما يلي:

١ -الافتراض الأول، أحادية البعد :Unidimensionality

إن نماذج نظرية السمات الكامنة تفترض وجود سمة وحيدة تفسر الاستجابة على فقرات الاختبار، حيث يفترض مطور الاختبار أن استجابة الأفراد على فقرات الاختبار تُعزى إلى سمة كامنة واحدة تفسر تلك الاستجابات، ومع ذلك، يصعب تحقيق هذا الافتراض بصورة مطلقة بسبب تداخل العديد من العوامل الشخصية أو المعرفية التي قد تؤثر على استجابة الفرد كالدافعية، وقلق الاختبار وغيرها .لذا فإنه حتى يتحقق افتراض أحادية البعد يتطلب ذلك وجود عامل سائد واحد يؤثر في الأداء على الاختبار، وهو ما يسمى بالقدرة التي يقيسها ذلك الاختبار، وحتى يتم تفسير نتائجه بدقه وصدق لابد من التأكد بأنه يقيس سمة واحدة، أي أنه كلما زادت احتمالية تحقيق هذا الافتراض كلما زادت دقة وصدق تفسير النتائج .(Yang, 2007)

تجدر الإشارة بأنه عند استخدام نماذج نظرية الاستجابة للفقرة بهدف تدريج القدرة يجب فحص هذا الافتراض عند كل نقطة زمنية باعتبار أن التدريب والتعليم تؤثران على أحادية البعد لمجموعة من فقرات الاختبار، وأن الاختبار أحادي البعد لمجتمع ما قد لا يكون كذلك لآخر .ويشير التقي (2009)إلى أن النماذج التي تفترض وجود أكثر من قدرة لتفسير أداء الأفراد على اختبار معين تسمى بالنماذج متعددة الأبعاد .

عندما تكون المفردات متجانسة فيما بينها، وتقيس نفس الصفة، فإنه للإجابة على أي فقرة من فقرات الاختبار، يتطلب نفس العمليات السلوكية والإجراءات، وفي الغالب يتم استخدام التحليل العاملي للوصول إلى العامل المؤثر في أداء الأفراد على الاختبار .وإذا كانت تلك الفقرات تقيس أكثر من صفة فإنه بالإمكان تجميع الفقرات المتجانسة فيما بينها بعد إجراء التحليل العاملي ثم يتم استخدام أحد النماذج مع كل مجموعة متجانسة لتدريجها)أبو جراد، .(2012

وللدلالة على أحادية البعد فإنه توجد مجموعة من المؤشرات التي يمكن الاعتماد عليها للتأكد من هذا الافتراض منها ما يعتمد على المكونات الرئيسية كمؤشرات التحليل العاملي، ومنها ما يعتمد على الثبات كمعامل كرونباخ، وغيره، وبعضها يعتمد على نماذج السمات الكامنة، فإن وجدت مجموعة من الفقرات طابقت في استجاباتها نموذجاً معيناً فإنها تمثل أحادية البعد .(Hattie, 1985)

٢ -الاستقلال المحلي أو الموضعي :Local Independence

يقصد بهذا الافتراض أن إجابة الفرد على فقرة معينة يجب ألا تتأثر بإجابته على أخرى يتضمنها نفس الاختبار، أي أن القدرة التي يمتلكها ذلك الفرد هي العامل الوحيد المؤثر في استجابته .ينتج عن هذا الافتراض أحادية البعد حيث يعتبر افتراضاً مكافئاً له، غير أنهما ليسا مفهوماً واحداً، فقد يكون الاختبار ثنائي البعد إذا وجدت سمتان كامنتان بحيث تكون الفقرات مستقلة بالنسبة للأفراد المتجانسين في كل سمة منهما، كما أن أبعاد اختبار معين يساوي عدد السمات الكامنة المطلوبة لتحقيق

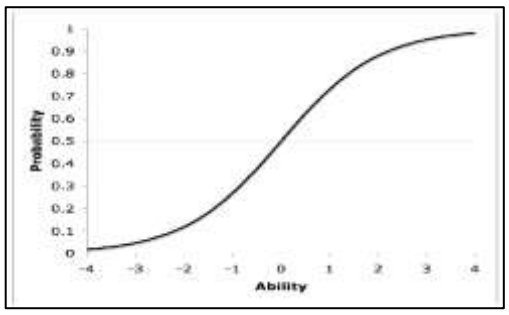
الاستقلال الموضعي .(Croker & Algina, 1986)كما أنه لا يعني بالعموم أن الاستجابة على أي فقرتين لا ترتبطان، فهو يتضمن أن الاستجابة على أي فقرتين لا ترتبطان ضمن المجموعات الجزئية المتجانسة من الأفراد في نفس المستوى من القدرة (Hulin, Drasgow & Parsons, 1983).

وللدلالة والكشف عن الاستقلال الموضعي بين الفقرات التي تتطلب الاستجابة الثنائية توجد بعض المؤشرات التي يمكن استخدامها في هذا الصدد كمؤشر (Q3) الذي يعبر عن معامل الارتباط بين البواقي لفقرتين، فإن وجد على أنه أكبر من الصفر فهذا مؤشر على عدم وجود استقلال موضعي بينهما، كما يمكن استخدام مؤشر (G2) للكشف عن الفروق بين ما هو متوقع وما هو ملاحظ بين فقرتين، ومؤشر (Fisher Z) كذلك، من خلال تحويل الأخطاء الملاحظة إلى معيارية، حيث القيم التي تزيد عن قيمة انحرافين معياريين يمكن اعتبارها غير مستقلة موضعياً)حمادنة، 2011).

٣ -منحنى خصائص الفقرة :Item Characteristic Curve (ICC)

يمثل هذا المنحنى العلاقة الرياضية التي تربط احتمالية إجابة الفرد إجابة صحيحة على إحدى فقرات الاختبار وبين القدرة التي يقيسها الاختبار، وهي دالة انحدار غير خطية، فبمعرفة درجات الأفراد لاختبار معين عند كل مستوى قدرة محدد يمكن رسم منحنى خصائص أي فقرة من فقرات ذلك الاختبار والذي يمثل خط الانحدار المار بمتوسط التوزيعات الشرطية لكل مستوى من مستويات القدرة (Hambleton & Swaminathan, 1985)

شكل (1)منحنى خصائص الفقرة ICC



بالتأمل في الشكل السابق يمثل المحور الأفقي متصل القدرة المقاسة للفقرة (θ)، والمحور الرأسي احتمالية الإجابة الصحيحة على تلك الفقرة P(θ)، وبالتالي تزيد احتمالية إجابة الفرد على فقرة إجابة صحيحة بازدياد قدرته.

إن احتمالية استجابة الفرد استجابة صحيحة على فقرة معينة يعتمد على شكل منحنى خصائص تلك الفقرة فقط وليس على عدد الأفراد الذين يقعون عند نفس مستوى القدرة، وهو ما يشار إليه بثبات منحنيات خصائص المفردات لمجتمع من المفحوصين تم معايرة الفقرات لهم، وهذا من أهم خصائص نماذج الاستجابة للفقرة (Baker, 2001).

٤ -التحرر من عامل السرعة :Speededness

يعتبر هذا الافتراض افتراضًا ضمنياً في افتراض أحادية البعد، حيث أن الاختبارات لا تطبق تحت شرط السرعة في الأداء، أي أن الأفراد الذين يفشلون في الإجابة على فقرات الاختبار فإن سبب ذلك هو محدودية القدرة وليس بأن الوقت ضيق أو غير كاف للوصول إلى الإجابة الصحيحة .وعندما تؤثر السرعة على الأداء فإنه يكون هناك عاملان يؤثران على الأداء هما سرعة الأداء، والسمة المقاسة من خلال الاختبار (Hambleton & Swaminathan, 1985).

نماذج نظرية الاستجابة للفقرة:

توجد العديد من النماذج الرياضية التي تفسر العلاقة بين احتمالية الإجابة الصحيحة على الفقرة والقدرة المقاسة، يتألف كل نموذج منها من معادلة رياضية خاصة به تربط بين أداء الفرد على الفقرات وقدرته الحقيقية.

تتميز تلك النماذج بأنها تفترض إمكانية وجود نطاق كبير جدًا من الفقرات لقياس نفس السمة، ولكن بشرط أن يكون تقدير القدرة للفرد مستقلاً عن عينة فقرات الاختبار المأخوذة من النطاق الكلي للفقرات .كذلك وجود مجتمع كبير جدًا من الأفراد الذين يتم من خلالهم توصيف كل فقرة من فقرات الاختبار شريطة أن تكون معالمها مستقلة عن عينة الأفراد التي تم استخدامها عند تقدير تلك

المعالم .كما أنها تفترض إمكانية الحصول على مؤشر إحصائي يبين مدى دقة تقدير القدرة لكل فرد بواسطة فقرات الاختبار (Hambleton & Swaminathan, 1985).

تهدف نماذج الاستجابة للفقرة Item Response Model إلى تحديد العلاقة بين أداء الفرد على فقرات الاختبار، وهذا الأداء يمكن ملاحظته بصورة مباشرة، وقدرته التي تفسر هذا الأداء على اعتبار أنها نماذج احتمالية وأن العلاقة التي تحددها تخضع لنظرية الاحتمالات.

تشترط النظرية تطابق البيانات التي تم الحصول عليها مع أحد هذه النماذج، ويتم اختيار النموذج المناسب بناءً على طبيعة الفقرات، عددها، وحجم العينة التي يتم استخدامها في تقدير المعالم .لكن المحدد الأهم لاختيار النموذج هي طبيعة البيانات المستخدمة، والتي تكون إما ثنائية الاستجابة 0)، (1أو متعددة الاستجابة . ومن أشهر وأهم تلك النماذج الرياضية ثنائية الاستجابة لنظرية الاستجابة للفقرة ما يلي:

١ -النموذج اللوغاريتمي أحادي البارامتر The One-Parameter Logistic Model (1PLM):

يعتبر هذا النموذج من النماذج البسيطة، وهو الأشهر استخداماً في نظرية الاستجابة للفقرة، ويسمى أيضاً بنموذج راش Rasch Model نسبة إلى عالم الرياضيات الدنماركي جورج راش .ويعد أحد الحلول المتاحة عند تقدير البارامترات في حالة العينات الصغيرة، ذلك لأن النماذج المتضمنة لبارامترات أقل عددًا تتطلب بيانات أقل دقة لتقدير تلك البارامترات.

يعتمد هذا النموذج على افتراض أن معلمة الصعوبة هي فقط التي تميز أداء الفرد، كما يفترض أن كل الفقرات للاختبار متساوية التمييز، وأن خط المقاربة السفلي للمنحنى المميز للقرة يساوي الصفر، بمعنى أنه لا يفترض احتمالية إجابة الأفراد ذوي القدرة المتدنية استجابة صحيحة على الفقرة، وبذلك ينفي وجود عامل التخمين.

يمكن التعبير عن النموذج بالصيغة الرياضية الاحتمالية التالية:

$$P_i(\theta) = \frac{e^{D(\theta - bi)}}{1 + e^{D(\theta - bi)}}$$

حيث أن:

= $P_i(\theta)$ احتمال أن يجيب فرد تم اختياره عشوائياً بمستوى قدره (θ) إجابة صحيحة عن الفقرة i.

$=(\theta)$ مستوى قدرة الفرد.

= b_i بارامتر صعوبة الفقرة i .

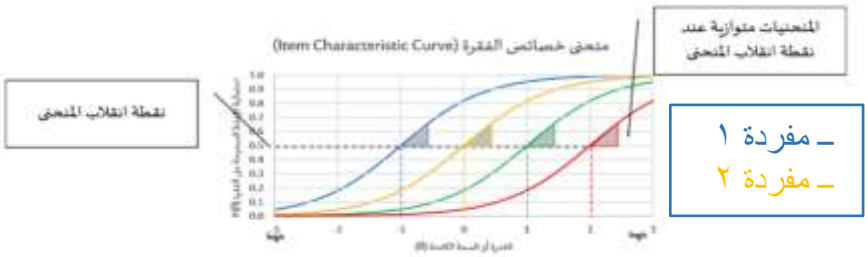
= D عامل التدريج الذي يساعد في الوصول إلى أقرب لوغاريتم ممكن لعملية التراكم الطبيعي، وهو ثابت لكل الفقرات ويساوي تقريباً 1.7

= i رقم الفقرة، =n عدد الفقرات للاختبار

= e الأساس اللوغاريتمي الطبيعي، وهو ثابت يساوي تقريباً 2.7183

والشكل (2) التالي يوضح المنحنى المميز للفقرة في النموذج اللوغاريتمي أحادي البارامتر :

شكل (2)



يتضح من الشكل السابق أن المنحنيات الخاصة بهذا النموذج جميعها متوازية ولا تتقاطع أبداً، ذلك لأن لها نفس التمييز، وتختلف في معامل الصعوبة حيث أن الفقرة)٤ (يجيب عنها الفرد الذي قدرته ٢ لوجيت، بينما الفقرة)٣ (يجيب عنها الفرد الذي قدرته ١ لوجيت إي أن الفقرة)٤ (أصعب من الفقرة)٣ (وكذلك مع باقي

الفقرات. كما نلاحظ أن كل المنحنيات تبدأ من احتمال إجابة =صفر، أي أن معامل التخمين ثابت للفقرات .وبما أن المنحنيات مختلفة فقط في معامل الصعوبة فإننا نستخدم النموذج أحادي البارامتر) نموذج راش .(Wells, 2004) (

٢ -النموذج اللوغاريتمي ثنائي البارامتر The Tow-Parameter Logistic Model: 2PLM

في هذا النموذج الذي طوّره لورد يأخذ فيه المنحنى المميز للفقرة شكل التوزيع اللوغاريتمي ذا البارامترين :التمييز، والصعوبة .أي بإضافة بارامتر التمييز إلى نموذج راش، بسبب صعوبة إيجاد عدة فقرات تميز بدرجة واحدة بين مستويات القدرة التي يقيسها الاختبار، وفي نفس الوقت لا يوجد مجال للتخمين، حيث لا تتأثر استجابات الأفراد على فقرات الاختبار بعامل التخمين)علام، .(2005

تتمثل الصيغة الرياضية لهذا النموذج كالتالي:

$$\text{Pi}\ (\theta) = \frac{e^{Dai\,(\theta - bi)}}{1 + e^{Dai\,(\theta - bi)}}$$

حيث أن:

= $P_i\ (\theta)$ احتمال أن يجيب فرد تم اختياره عشوائياً بمستوى قدره (θ) إجابة صحيحة عن الفقرة .i

=(θ) مستوى قدرة الفرد.

= a_i بارامتر تمييز الفقرة . i

= b_i بارامتر صعوبة الفقرة . i

= D عامل التدريج الذي يساعد في الوصول إلى أقرب لوغاريتم ممكن لعملية التراكم الطبيعي، وهو ثابت لكل الفقرات ويساوي تقريباً 1.7

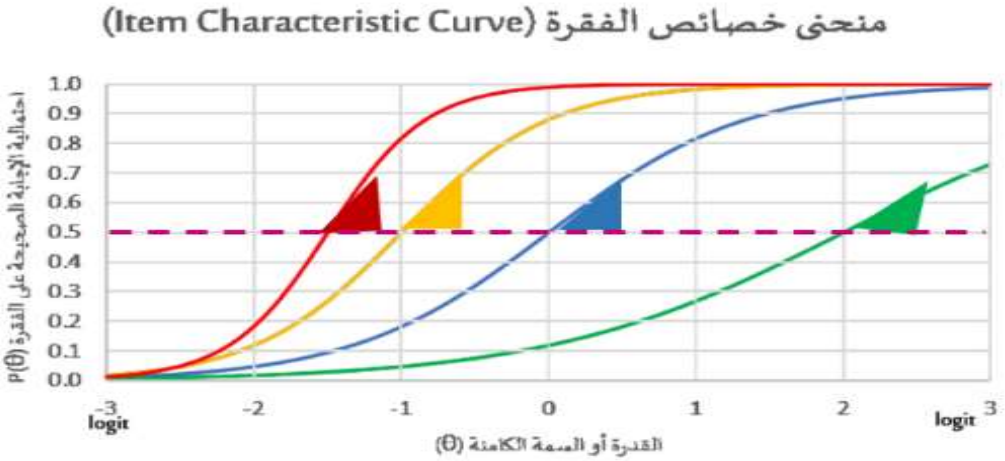
= iرقم الفقرة، =nعدد الفقرات للاختبار

= eالأساس اللوغاريتمي الطبيعي، وهو ثابت يساوي تقريباً 2.7183

والشكل (3)التالي يوضح المنحنى المميز للفقرة في النموذج اللوغاريتمي ثنائي البارامتر :

شكل (3)

يتضح من الشكل السابق أن المنحنيات الأربعة تختلف في الجزء الذي تلتقي فيه بالمحور الأفقي، أي أن فقراتها تختلف في الصعوبة، كذلك عند نقطة الانقلاب



يختلف ميل الزاوية وبذلك فإن الفقرات تختلف في بارامتر التمييز، فكلما زاد الميل كلما كانت الفقرة مميزة أكثر من غيرها) علام، .(2005

٣ -النموذج اللوغاريتمي ثلاثي البارامتر The Three-Parameter Logistic Model: 3PLM

يعتبر هذا النموذج امتداد للنموذج ثنائي البارامتر، حيث أضاف بيرنيوم بارامترًا ثالثاً ليأخذ المنحنى المميز للفقرة فيه شكل التوزيع اللوغاريتمي لبارامترات الصعوبة، التمييز، والتخمين .وتتمثل الصيغة الرياضية لهذا النموذج كالتالي:

$$Pi\ (\theta) = c_i + (1- c_i)\ \frac{e^{Dai\ (\theta - bi)}}{1+ e^{Dai\ (\theta - bi)}}$$

حيث أن:

$P_i\ (\theta)$ = احتمال أن يجيب فرد تم اختياره عشوائياً بمستوى قدره (θ) إجابة صحيحة عن الفقرة i.

=(θ) مستوى قدرة الفرد.

a_i= بارامتر تمييز الفقرة . i

b_i = بارامتر صعوبة الفقرة . i

c_i = بارامتر تخمين الفقرة . أويمثل الخط التقاربي الأسفل للمنحنى المميز للفقرة.

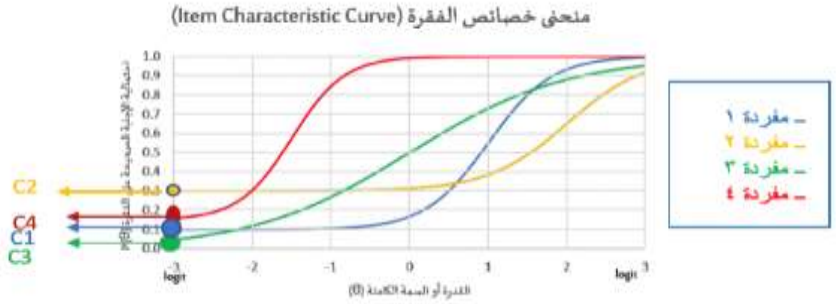
D = عامل التدريج الذي يساعد في الوصول إلى أقرب لوغاريتم ممكن لعملية التراكم الطبيعي، وهو ثابت لكل الفقرات ويساوي تقريباً 1.7

i = أرقم الفقرة، n=عدد الفقرات للاختبار

e = الأساس اللوغاريتمي الطبيعي، وهو ثابت يساوي تقريباً 2.7183

والشكل (4)التالي يوضح المنحنى المميز للفقرة في النموذج اللوغاريتمي ثلاثي البارامتر :

شكل (4)



من الشكل السابق يتضح أن المنحنيات المميزة للقرات الأربعة تختلف عن بعضها في الصعوبة، كذلك في درجة الميل مما يعني اختلافها في مستوى التمييز لكل منهم، وتختلف أيضاً في الخط التقاربي الأسفل الذي ينشأ بسبب تأثير التخمين في الاستجابة .

الجدير ذكره أنه ينبغي على الباحث أو مصمم الاختبار أن يقوم بتحديد النموذج الذي سيتعامل معه مسبقاً حتى يمكنه التحقق من ملائمة البيانات للنموذج الذي تم تحديده وذلك من خلال المنحنيات المميزة للفقرات والتي ستساعده في التأكد من

مدى تحقق افتراضات النموذج المستخدم (Hambleton & Swaminathan, 1985).

ثانيًا :الدراسات السابقة

لقد اهتم علماء القياس والباحثون بدراسة أثر طول الاختبار، وعدد فقراته، وعدد البدائل في اختبارات الاختيار من متعدد، وأحجام العينات على خصائص الاختبار وخصائص فقراته .وأظهرت دراساتهم سواء الأجنبية أو العربية عدم الاتفاق على عدد معين للفقرات أو البدائل أو حجم العينة .لذا فإنه في هذا الجزء من الفصل الحالي يستعرض الباحث مجموعة الدراسات والبحوث التي لها ارتباط بمتغيرات هذه الدراسة، حيث وجد أن هناك أربعة أنواع من الدراسات المرتبطة بهذه الدراسة:

- دراسات متعلقة بطول الاختبار، أو عدد فقرات، وأثرها على دقة تقدير قدرات الأفراد ومعالم الفقرات.
- دراسات متعلقة بعدد بدائل اختبارات الاختيار من متعدد وأثرها على دقة تقدير قدرات الأفراد ومعالم الفقرات.
- دراسات متعلقة بحجم العينة المناسب وأثرها على دقة تقدير قدرات الأفراد ومعالم الفقرات.
- دراسات متعلقة بنماذج نظرية الاستجابة للفقرة المستخدم في تقدير قدرات الأفراد ومعالم الفقرات.

وسيتم استعراض نتائج بعض تلك الدراسات بناء على تاريخها من الأقدم إلى الأحدث.

دراسة (Costin, 1970) والتي أُجريت للمقارنة بين معاملات صعوبة الفقرات وتمييزها لاختبارات الاختيار من متعدد عند اختلاف عدد بدائل الاستجابة، حيث تم إعداد صورتي اختبار مؤلف من 110فقرة لكل صورة في دروس الإدراك 25)فقرة(، الدافعية 30)فقرة(، التعلم 30)فقرة(، والذكاء 25)فقرة .(عدد بدائل الاستجابة في الصورة الأولى أربعة بدائل، وفي الصورة الثانية تم حذف أحد البدائل عشوائياً

لتكون الصورة بثلاثة بدائل لكل فقرة .تم تطبيق الاختبار على (207)من طلاب جامعة إلينوى، وتبين من النتائج أن متوسطات معاملات التمييز والصعوبة لفقرات الأربعة بدائل أقل منها في فقرات البدائل الثلاثة.

وفي دراسة قام بها كلًا من (Reckase & Mark, 1978) هدفت إلى مقارنة دقة تقدير معلمة القدرة للأفراد ومعالم الفقرات في نموذجي راش الأحادي، واللوجستي الثلاثي مستخدمين أسلوب المحاكاة في توليد البيانات، حيث أشارت نتائج دراستهما أن النموذج ثلاثي البارامتر طابق بيانات الاختبار بشكل أفضل من النموذج الأحادي لراش، كذلك بالنسبة لدقة تقدير معالم الفقرات .كما أن هناك ارتباطاً عالياً بين تقديرات القدرة وفق النموذجين لمعظم البيانات، وبينت الدراسة أيضاً أن النموذج ثلاثي البارامتر يحتاج حجم عينة أكبر لمعايرة الفقرات، وأنه يفضل استخدام نموذج راش في حال العينات ذات الأحجام الصغيرة.

وقد أجرى كل من (Straton & Catts, 1980) دراسة بهدف المقارنة بين خصائص اختبار الاختيار من متعدد بتثبيت عدد الفقرات، واختلاف عدد البدائل . حيث تم بناء اختبار مؤلف من (60)فقرة اختيار من متعدد تم اختيارها من اختبارات الاقتصاد للمستوى الثاني .بعد ذلك تم تشكيل (4)اختبارات منه :الأول يحوي (60) فقرة لكل منها بديلين، والثاني (40)فقرة بثلاثة بدائل لكل فقرة حيث تم حذف البديل الرابع عشوائيًا، ويحوي الثالث (40)فقرة أيضاً وبثلاث بدائل بعد حذف البديل الأقل تمييزًا، أما الاختبار الرابع فلم يتم إجراء أي تعديل عليه .وقد تم تطبيق الصور الأربع عشوائيًا على عينة مؤلفة من (260)وأشارت النتائج إلى أن عدد بدائل الفقرات يتناسب عكسيًا مع معامل الصعوبة، وطرديًا مع معامل التمييز .كما أن الثبات والخطأ المعياري في القياس للفقرات ذات الثلاثة بدائل أفضل منها في غيرها.

أما دراسة (Hambleton & cook, 1980) والتي تهدف إلى التعرف على أثر حجم العينة وطول الاختبار في تقدير القدرة وفقاً للنموذج اللوجستي ثلاثي البارامتر، فقد قامت على اختيار أحجام مختلفة العدد لعينات 50)، 200، (1000فرد، وأيضًا

مستويات مختلفة لعدد الفقرات 10)، 20، (80فقرة .حيث أوضحت النتائج أن أكبر قيمة للخطأ المعياري بلغت (2.19)عند نموذج الاختبار ذو 10)فقرات، 50 فرد .(وتراوحت قيمته بين 0.88وَ 1.5للاختبار ذو 20)فقرة (عند حجم 200فرد، وتقل قيمته عند نموذج 80)فقرة، 1000فرد .(وبشكل عام فإن دقة التقدير تتناسب طردياً مع كل من حجم العينة وعدد فقرات الاختبار .

وفي الدراسة التي تهدف إلى معرفة العدد الأنسب من بدائل الاستجابة لفقرات اختبارات الاختيار من متعدد والتي قام بها (Budescu & Nevo, 1985) حيث استخدما ثلاثة اختبارات :للرياضيات، الاستيعاب القرائي، واختبار للمفردات .لكل منهم أربعة صور تختلف في عدد البدائل لكل فقراتها .الصورة الأولى بديلين، والثانية ثلاثة بدائل، والثالثة أربعة بدائل، والرابعة خمسة بدائل، وكانت العينة (1018) طالباً .تم حساب معاملات الثبات للصور باستخدام معادلة Kudder-Richardson-20 وجاءت معاملات الثبات لهم على التوالي0.797) :، 0.834، 0.840، (850، وبلغت المتوسطات الحسابية16.79) :، 15.52، 14.27، (14.27 على التوالي .وجاءت في نتائج الدراسة أن معاملات الثبات تزداد بزيادة عدد بدائل الاستجابة، كما أوصت الدراسة باستخدام صورة الاختبار ذات الثلاثة بدائل لقرب معامل ثباته من معاملات الصورتين ذات الأربعة والخمسة بدائل، كما أن متوسط الدرجات أعلى من متوسطي الصورتين الأخريين.

وقدّم (Owen & Forman, 1987) دراسة هدفت إلى المقارنة بين خصائص فقرات اختبارين :الأول له ثلاثة بدائل، والثاني له خمسة بدائل، حيث تم تطبيق اختبارين على عينة من (114)طالب وعدد فقراتهما (100)فقرة بفاصل زمني (10)أيام .بعد ذلك تم تقليص عدد بدائل فقرات الاختبار ذا الخمسة بدائل إلى ثلاثة بدائل باستخدام الحذف التجريبي للبدائل وذلك بحذف البديل الأقل فاعلية، ثم تكوين صورتين كاختبارين، الأول اشتمل على (100)فقرة، النصف الأول منها كانت ذات ثلاثة بدائل، والنصف الآخر ذات خمسة بدائل .وتم تطبيق الصورتين عشوائياً على

العينة وطلب منهم كتابة الزمن المستغرق في الإجابة عن الجزء الأول، والجزء الثاني كذلك .وأظهرت النتائج إلى عدم وجود فروق ذات دلالة إحصائية عند مستوى الثقة (0.05)لمعاملات الصعوبة والتمييز للفقرات ذات الخمسة والثلاثة بدائل، كما كان هناك فروق ذات دلالة إحصائية في الزمن المستغرق للإجابة عن كل بعد .ثم تم حساب معامل الصدق التلازمي لكل اختبار مع نتائج الطلبة على الاختبار البعدي، حيث كان (0.75)للاختبار ذا الثلاثة بدائل، و (0.73)للاختبار ذا الخمسة بدائل.

وأجرى كل من (Barnes & Wise, 1991) دراسة بهدف التعرف على أثر حجم العينة وطول الاختبار في تقدير قدرات الأفراد وفقاً للنموذج الأحادي، حيث كانت أحجام العينات المستخدمة 50)، 100، (200فردًا، لاختبارين الأول (25)فقرة، والثاني (50)فقرة .وقد أظهرت نتائج تلك الدراسة أن هناك تأثير لعدد فقرات الاختبار على قدرات الأفراد، بينما لا يوجد تأثير لحجم العينة على قدرتهم .كما أن عدد فقرات الاختبار)طول الاختبار (لم يؤثر على صعوبة الفقرات، وإنما كان هناك تغيرًا بسيطاً مع تغير حجم العينة بسبب أن الفرق بين أحجام العينات لم يكن كبيرًا.

وفي دراسة قام بها الشبلي (1995)بهدف معرفة أثر عدد البدائل في اختبار الاختيار من متعدد على التجانس الداخلي وخصائص فقراته السيكومترية .تم تطبيق اختبار مكون من (45)فقرة، وبثلاثة نماذج :بدائل الاستجابة لفقرات الأول خمسة بدائل، وأربعة بدائل بعد حذف أحدها عشوائياً للنموذج الثاني، وثلاث بدائل بعد حذف أحد بدائل النموذج الثاني لتكوين النموذج الثالث .ثم تم تطبيق كل نموذج على (150) طالب وبإجمالي (450)للعينة .وأظهرت النتائج إلى وجود فروق ذات دلالة إحصائية بين متوسطات الصعوبة للفقرات ذات الثلاثة بدائل والفقرات ذات الخمسة بدائل لصالح الفقرات ذات الثلاثة بدائل .كما أظهرت النتائج عدم وجود فروق ذات دلالة إحصائية بين صعوبة الفقرات ذات الأربعة بدائل مع كلٍ من صعوبة فقرات الثلاثة بدائل أو الخمسة بدائل .كذلك أظهرت النتائج عدم وجود فروق ذات دلالة إحصائية بين أي من متوسطات معاملات التمييز لفقرات النماذج الثلاثة.

كما أجرى (Lio, Sinica & Yu, 1999) دراسة بهدف فحص دقة تقديرات معلم القدرة باستخدام أسلوب المحاكاة في ضوء النموذج اللوجستي ثلاثي البارامتر، حيث تم توليد فقرات اختبارية ثنائية الاستجابة لتصحيح ثلاثة نماذج لاختبارات مكونة من 20)، 40، (60فقرة لكل منهم، وعدد (500)فرد للعينة، وتم استخدام طريقة الأرجحية العظمى (MLE) لتقدير معلمة القدرة، وأشارت نتائج تلك الدراسة أن الخطأ المعياري لتقدير القدرة كان مرتفعاً عند كل من المستويات المرتفعة والمنخفضة للقدرة.

وفي دراسة أجراها (Fitzpzatric & Ann, 2001) هدفت إلى التعرف على أثر حجم العينة وطول الاختبار في الثبات للاختبار ومعادلة الاختبار باستخدام نموذج التقدير الجزئي، حيث تم استخدام ثلاث عينات 200)، 500، (1000فردًا، وخمسة مستويات لطول الاختبار 2)، 4، 8، 12، (20فقرة، وتم توليد البيانات باستخدام أسلوب المحاكاة، كما تم استخدام معيار جذر معدل مربعات الفروق لتقييم دقة تقديرات المعالم .وأشارت نتائج الدراسة أن دقة تقدير المعالم تزيد بزيادة عدد فقرات الاختبار)طول الاختبار.(

دراسة الدرابيع (2001)هدفت إلى التعرف على أثر حجم العينة وعدد فقرات الاختبار في تقدير قدرات الأفراد ومعالم الفقرات وفقاً للنموذج الأحادي، حيث كانت العينة 50)، 100، (500فرد، لاختبارات بأطوال مختلفة 25)، 50، (300فقرة . وأوضحت النتائج وجود فروق دالة لتفاعل حجم العينة وعدد الفقرات في دقة تقدير قدرات الأفراد تعزى لتفاعل حجم العينة وعدد الفقرات.

أما دراسة (Huang & Lin & Shin, 2001) والتي ركزت في أحد جوانبها على فحص أثر حجم العينة على دقة تقدير معالم الفقرات، حيث تم استخدام النموذج اللوجستي ثلاثي البارامتر لمعايرة (360)فقرة ثنائية التدريج باستخدام طريقة المحاكاة بافتراض أن أول (120)فقرة تشكل الجزء المشترك من الاختبار بينما باقي الفقرات شكلت أربعة اختبارات يتضمن كل منها (60)فقرة .وتم اختيار مدى حجم

العينة ما بين (250)، و (1500)فرد، بافتراض أن تقديرات الأفراد تتوزع طبيعياً، وبذلك تم تشكيل (14)حالة حسب حجم العينات وتوزيع القدرة، ولكل حالة من الحالات تم معايرة الاختبار المكون من (360)فقرة باستخدام برامج -Bilog, Bilog MG, PIC وتم استخدام معيار جذر معدل مربعات الفروق RMSD لمقارنة دقة تقديرات المعالم، وتوصلت الدراسة إلى أن أخطاء التقدير لمعلم الصعوبة ومعلم التمييز تكون أكبر عندما يقل حجم عينة المفحوصين .

كما قامت التعمري (2003)بدراسة هدفت إلى تطبيق النموذج اللوجستي ثلاثي البارامتر لتقدير قدرة الأفراد ومعالم الفقرات لاختبار الاختيار من متعدد باختلاف عدد بدائل الاستجابة .تم التطبيق على عينة مؤلفة من (1200)طالب، وباستخدام برنامج BILOG في تقدير قدرات الأفراد ومعالم الفقرات في ضوء النظرية الحديثة اتضح عدم وجود فروق ذات دلالة إحصائية في قدرة الأفراد على اختبار الاختيار من متعدد، وأشارت النتائج إلى وجود فروق ذات دلالة إحصائية في معلمة الصعوبة باختلاف عدد البدائل لصالح الاختبار ذو الثلاثة بدائل، كما أشارت النتائج إلى وجود فروق ذات دلالة إحصائية في تمييز الفقرات على اختبار الاختيار من متعدد باختلاف عدد البدائل لصالح الاختبار ذو الخمسة بدائل، وأشارت أيضاً إلى وجود فروق ذات دلالة إحصائية في تخمين الفقرة على اختبار الاختيار من متعدد باختلاف عدد البدائل لصالح الاختبار ذو الثلاثة بدائل.

في دراسة أجراها عبابنة (2004)هدفت إلى تفحص أثر حجم العينة وطريقة انتقائها وعدد الفقرات وطريقة انتقائها على دقة تقدير معالم الفقرة والقدرة وفق النموذج اللوجستي ثلاثي البارامتر، حيث تم بناء اختبار مؤلف من أربعة اختبارات فرعية وبعدد (71)فقرة بحجم عينة بلغ (1000)فرد، وتم استخدام برمجية Bilog 3.11 لتقدير قدرات الأفراد ومعالم الفقرات والأخطاء المعيارية في التقدير ومطابقة البيانات للنموذج اللوجستي ثلاثي البارامتر، وقد أظهرت النتائج بأن هناك علاقة طردية بين حجم العينة والدقة في تقدير معالم الفقرات، وأن تقديرات القدرة للأفراد

تتسم بالاستقرار عندما يتم استخدام عينات معايرة بأحجام كبيرة، وتتأثر الدقة في تقدير القدرة بوضع مطابقة البيانات للنموذج، فيكون من غير المؤكد الحصول على دقة أكبر عند زيادة أحجام العينات المستخدمة في المعايرة عن مستوى معين .ودلت النتائج على ازدياد الدقة في تقدير معلم التمييز بزيادة تباين قدرة المفحوصين، وتزيد الدقة في تقدير معلم الصعوبة والقدرة عندما يكون مدى القدرة للمفحوصين متوافقاً مع مدى صعوبة الفقرات، وتزداد كذلك الدقة في تقدير معلم التخمين عند استخدام عينة من ذوي القدرة المتدنية في معايرة الفقرات، كما أظهرت النتائج أيضاً زيادة الدقة في تقدير معلم القدرة بزيادة عدد فقرات الاختبار أو نسبتها إلى الاختبار الكلي.

وفي دراسة تهدف إلى الكشف عن مناسبة اختبارات الاختيار من متعدد ذات الثلاثة بدائل عن غيرها من عدد البدائل، فقد أجرى (Rodriguez, 2005) دراسة بتحليل مجموعة من البحوث والدراسات أجريت خلال (80)سنة للتعرف على أثر البدائل ودورها على خصائص الاختبار السيكومترية، حيث قام بتجميع نتائج دراسات من عام 1920م حتى 1999م بواقع (27)دراسة تشتمل على (56)اختبار .أظهرت نتائج الدراسة أن الاختبار ذو الثلاثة بدائل كان الأفضل في اختبارات الاختيار من متعدد في معظم الدراسات، وأن التغير في عدد البدائل من خمسة إلى أربعة يؤدي إلى انخفاض كل من :معامل صعوبة الفقرة بمتوسط (0.02)، ومعامل التمييز بمتوسط (0.04)، ومعامل الثبات بمتوسط .(0.035)وفي حالة التغيير لعدد البدائل من خمسة إلى ثلاثة باستخدام طريقة حذف البديل الأقل جاذبية فإن ذلك أدى إلى خفض معامل الصعوبة بمتوسط (0.07)، ولم يؤثر ذلك على معاملات التمييز والثبات، وعند حذف البدائل عشوائيًا للوصول إلى ثلاثة بدائل انخفض معامل الثبات بمتوسط .(0.06)أما في حالة التغير في عدد البدائل من الأربعة إلى الثلاثة فإن ذلك أدى إلى خفض معامل الصعوبة بمتوسط (0.04)وزيادة معاملي التمييز والثبات بمتوسط (0.03)و (0.02)على الترتيب .كما أنه في حالة تغيير عدد البدائل إلى

بديلين اثنين سواءً من الخمسة أو الأربعة، أصبحت فقرات الاختبار أكثر سهولة وقلت أيضاً معاملات التمييز فيها.

أجرى (Glass, 2005) دراسة تهدف إلى التعرف على أثر طول اختبار وعدد أفراد العينة في تقدير قدرات الأفراد باستخدام طريقة بييز، حيث تم توليد بيانات ثنائية التدريج لاختبارات بمستويين لعدد الفقرات 200)، (440فقرة، ولعدد 500)، 1000، (2000فرد .وقد أظهرت نتائج الدراسة أن هناك علاقة بين طول الاختبار ودقة التقدير، حيث تزيد دقة تقدير قدرات الأفراد ومعالم الفقرات بزيادة حجم العينة، وتقل في ذات الوقت الأخطاء المعيارية.

دراسة (Shizuka & al, 2006) والتي تهدف إلى معرفة مدى تأثير تقليل عدد البدائل في اختبار الاختيار من متعدد على خصائصه السيكومترية، حيث تم إعداد نموذجي اختبار، الأول مؤلف من (38)فقرة بأربعة بدائل لكل منها، وتم تطبيقه على (1000)طالب، والثاني مؤلف من (38)فقرة أيضاً على عينة بلغت (192) طالب ولكنه مجزأ إلى قسمين :الأول (10)فقرات بأربعة بدائل، والثاني (28)فقرة بثلاثة بدائل بعد حذف البديل الأقل جاذبية، وكان عدد الفقرات المشتركة بين الاختبارين (10)فقرات لكل منها أربعة بدائل .وقد أظهرت النتائج أنه لا توجد فروق ذات دلالة إحصائية تعزى لمعاملات الصعوبة والتمييز بين الاختبارين، وأن الاختبار ذو البدائل الثلاثة يؤدي الغرض منه مثله مثل الاختبار ذو الأربعة بدائل، حيث أن زيادة عدد البدائل يزيد من فرصة عامل التخمين لقلة الوقت المقدر.

وفي دراسة (Pommerich, 2007) والتي تم إجراؤها بهدف دراسة استقرار تقدير معالم الفقرات، وذلك باختبار البرامج التي تستخدم الكمبيوتر والاختبار الورقي من خلال اختبارات مختلفة في الأطوال، حيث أظهرت نتائج تلك الدراسة أن هناك ارتباط وعلاقة طردية بين عدد الفقرات واستقرار معالمها، أي أنه كلما زاد طول الاختبار بزيادة عدد فقراته كلما زاد الاستقرار في معالم فقرات ذلك الاختبار.

وأجرى عيلبوني (2007)دراسة تهدف إلى معرفة أثر عدد البدائل وجاذبيتها في فقرات اختبارات الاختيار من متعدد على التوافق مع النموذج اللوجستي ثلاثي البارامتر .حيث تم بناء اختبار تحصيلي من ثلاثة أشكال اختبارية لكل منها (50) فقرة، الأول بخمسة بدائل للفقرة، والثاني بثلاثة بدائل للفقرة بعد حذف بديلين عشوائياً من الأول، والثالث بثلاثة بدائل بعد حذف البديلين الأقل تمييزًا من الشكل الأول، وتم تطبيق الاختبار على (1656)طالب .وأظهرت النتائج توافق الفقرات مع النموذج ثلاثي البارامتر من خلال التحقق من افتراضات نظرية الاستجابة للفقرة، وتطابق الفقرات للأشكال الثلاثة للاختبار مع النموذج الثلاثي البارامتر بعد حذف (3)فقرات من الشكل الأول للاختبار، وبدون حذف أي فقرة لبقية الأشكال .كما أظهرت النتائج أيضاً من خلال معاملات الصعوبة والتمييز والتخمين ومحنى معلومات الاختبار ومعاملات الثبات إلى أن الشكل الأول ذو الخمسة بدائل هو الأفضل مقارنة بالشكلين الآخرين بغض النظر عن طريقة الحذف، وأنه لا يوجد أثر لتمييز البدائل في فقرات الاختيار من متعدد حتى بعد حذف البديلين الأقل تمييزاً من الشكل الثالث أو حذفهم عشوائيًا في الثاني.

أما دراسة (Fitzpatric, 2009) والتي هدفت إلى التعرف على أثر عدد فقرات الاختبار في تقدير معالم الفقرات وفقاً للنموذج الأحادي، حيث كانت عدد فقرات الاختبار 5)، 10، (15فقرة .وأظهرت نتائج الدراسة أنه حتى يكون هناك زيادة استقرار في تقدير معالم الفقرات لابد من استخدام اختبارات تحتوي أكثر من (15) فقرة.

في دراسة أجراها (De la Torre & Hong, 2010) بهدف معرفة أثر حجم العينة على دقة تقدير معالم الفقرة والقدرة في اختبارات مطورة حسب نماذج نظرية الاستجابة للفقرة، حيث تم توليد بيانات لعينات مكونة من 500)، (1000فرد، وأطوال اختبارات 10)، (20فقرة، وأظهرت النتائج أن كلاً من حجم العينة وعدد فقرات الاختبار يؤثران في تقدير معالم الفقرات عند عينة (1000)فرد و (20)فقرة،

كما أشارت النتائج إلى عدم تأثر دقة تقدير معلمة القدرة بحجم العينة وتأثرها بطول الاختبار.

دراسة (Fu, 2010) التي تهدف إلى التعرف على مدى دقة تقدير معلمة القدرة وصعوبة الفقرة باستخدام (5)نماذج من نماذج نظرية الاستجابة للفقرة، حيث كانت النماذج متباينة في مستوى التخمين، وأحجام العينة، وطول الاختبار . وتم توليد استجابات بلغت (50)مجموعة من البيانات المولدة باستخدام ظروف مختلفة .أشارت النتائج إلى أن هناك تبايناً في دقة تقدير معالم الفقرة والأفراد حسب مستوى التخمين الموجود في الاختبار، وحجم العينة، وطول الاختبار .كما أظهرت النتائج أن دقة تقدير معلمة القدرة، ومعالم الفقرات تعتمدان على معيار الدقة في كل واحد من نماذج الاستجابة للفقرة.

وفي دراسة الثوابية (2010)التي تهدف إلى استقصاء أثر حجم العينة في تقدير معالم الفقرات باستخدام نظرية الاستجابة للفقرة، حيث تناولت العينة أحجام مختلفة تراوحت بين 200)، (11292فرد، طبّق عليهم اختبار تحصيلي مؤلف من (80)فقرة، وتم استخدام برنامج BILOG-MG لتقدير المعالم .أظهرت النتائج أن معلمة صعوبة الفقرة تزيد بزيادة حجم العينة، حيث بلغ متوسطها (0.31)عند حجم (200)فرد، وازداد عند حجم (11292)، كما أشارت النتائج بأن الخطأ المعياري في تقدير معلمة الصعوبة يتناقص بزيادة حجم العينة، حيث بلغ (0.32)عند حجم عينة (200)فرد، و (0.07)عند حجم (11292)فرد.

أما في الدراسة التي أجراها الرباعي (2012) بهدف بيان أثر عدد البدائل، وتغير موقع المشتت القوي على خصائص الاختبار السيكومترية ومعالم الفقرات والقدرة وفق نظرية الاستجابة للفقرة، فقد تم بناء اختبار تحصيلي من (54)فقرة، وطبّق على (2123)فرد .حيث اشتمل الاختبار أربعة نماذج :يتكون الأول من خمسة بدائل، وموقع المشتت القوي قريب من الإجابة الصحيحة، ويتكون الثاني من خمسة بدائل أيضاً لكن موقع المشتت بعيد عن الإجابة الصحيحة، وفي الثالث والرابع تم

حذف البديلين الأضعف، ليتكون النموذج الثالث من ثلاثة بدائل يقع المشتت القوي قريباً من الإجابة الصحيحة، والرابع من ثلاثة بدائل، ويقع المشتت القوي بعيداً من الإجابة الصحيحة .وقد تم استخدام النموذج اللوجستي الثلاثي البارامتر، وبرنامجي SPSS، BILOG-MG3 للتحليل، وأظهرت النتائج عدم وجود اختلاف في تقدير معلمتي الصعوبة والتخمين للفقرة باختلاف عدد البدائل وموقع المشتت الأقوى والتفاعل فيما بينهما، كما أظهرت وجود اختلاف في تقدير معلمة التمييز يعزى لموقع المشتت القوي وتفاعله مع عدد البدائل.

وفي دراسة الشريفين (2012)التي تهدف للكشف عن أثر طريقة تقدير معالم الفقرات وقدرات الأفراد على خصائص الاختبار السيكومترية، ومعالم الفقرات في ضوء تغير حجم العينة باستخدام طريقتي بييز والأرجحية العظمى للتقدير .فقد تم بناء اختبار اختيار من متعدد مؤلف من (33)فقرة بأربعة بدائل لكل منها، وكانت العينة (1000)طالب، وتم التحليل باستخدام برمجية BILOG-MG وفق النموذج اللوجستي ثلاثي البارامتر .وأظهرت النتائج وجود فروق ذات دلالة إحصائية عند مستوى 0.05بين متوسطات الأخطاء المعيارية لتقديرات معالم الفقرات تعزى للتفاعل بين طريقة التقدير وحجم العينة، ولا توجد فروق ذات دلالة إحصائية تعزى لمتغير حجم العينة وطريقة التقدير .كما أظهرت النتائج وجود فروق ذات دلالة إحصائية بين متوسطات الأخطاء المعيارية لتقديرات القدرة للأفراد تعزى لحجم العينة، وللتفاعل بين طريقة التقدير وحجم العينة، كما لم تظهر فروق ذات دلالة إحصائية تعزى لطريقة التقدير، وأيضاً عدم وجود فروق دالة إحصائيًا بين معاملات الثبات المقدرة عند أحجام العينة المختلفة100) ، 500، (1000فرد.

دراسة الزبون (2013)التي أجراها بهدف الكشف عن أثر حجم العينة على تقدير دالة المعلومات للاختبار والخطأ المعياري في تقديرها باستخدام نظرية الاستجابة للفقرة، حيث استخدم بيانات استجابات لـ (7500)تم اختيارهم عشوائياً موزعين على مجموعات خمس مختلفة الأحجام 500)، 1000، 1500، 2500،

(3000فرد، لاختبار مكون من (40)فقرة من نوع الاختيار من متعدد .ولتحليل البيانات تم استخدام برنامج BILOG-MG3 وفق النموذج اللوجستي ثلاثي البارامتر . وتبين من خلال النتائج أن مقدار تقديرات دالة المعلومات يتناسب طرديًا مع حجم العينة، ويتناسب الخطأ المعياري في تقدير دالة المعلومات عكسيًا مع حجم العينة.

وفي دراسة أجراها بني عطا (2014)بهدف التعرف على أثر حجم العينة وطول الاختبار في تقدير القدرة للأفراد ومعالم الفقرات في ضوء النموذج اللوجستي ثلاثي البارامتر، وباستخدام بيانات مولدة ثنائية الاستجابة لعينات مختلفة الأحجام 100)، 250، 500، 1000، 2000، (4000وبعدد فقرات 10)، 25، 50، 75، 100، (300، حيث أشارت النتائج أن دقة تقدير قدرات الأفراد ومعالم الفقرات تزيد كلما زاد طول الاختبار عن (50)فقرة، وزاد حجم العينة المستخدم عن (2000)فرد.

دراسة (Chen et al., 2014) التي تهدف إلى تقييم نتائج تحليل نموذج راش باستخدام عينات ذات أحجام صغيرة، حيث تم الاعتماد على اختبار مؤلف من (10)فقرات لعينات ذات أحجام 30)، 50، 100، (250فرد .تم تحليل البيانات باستخدام برنامج Mplus، وأظهرت النتائج أنه الأخطاء المعيارية في تقدير معلمتي الصعوبة للفقرات والقدرة للأفراد لأحجام 30)، (50تكون أكبر من قيمها عند استخدام عينات 100)، (250فرد.

دراسة (Jiang et al., 2016) تهدف للمقارنة بين دقة تقدير معالم الفقرات في ضوء نموذج الاستجابة المتدرجة متعددة الأبعاد عند اختلاف حجم العينة، باستخدام الأرجحية العظمى الهامشية لتقدير المعالم .حيث تم توليد بيانات ذات أحجام 500)، 1000، 1500، (2500فرد، لاختبارات مكونة من 30)، 90، (240 فقرة .وبعد تحليل النتائج عن طريق برمجية flexMIRT أظهرت النتائج إلى أن حجم عينة (500)فرد يقدم تقديرات دقيقة لمعالم الفقرات عندما يتم استخدام الاختبارات ذات الأطوال 30)، (90فقرة، أما عند استخدام الاختبار المكون من (240)

فقرة فمن الضروري استخدام عينة لا تقل عن (1000)فرد .كما أن زيادة حجم العينة عن (1000)فرد لا يزيد من دقة تقدير المعالم.

دراسة (Sahin & Anil, 2017) بهدف معرفة أثر حجم العينة وعدد فقرات الاختبار على دقة تقدير معالم نظرية الاستجابة للفقرة، باستخدام طريقة الأرجحية العظمى الهامشية للتقدير معالم الفقرة عن طريق برمجية Xcalibre 4.1 لتقدير المعالم .حيث تم بناء ثلاث اختبارات ذات أطوال 10)، 20، (30فقرة تم تطبيقها على عينات ذات الأحجام 150)، 250، 350، 500، 750، 1000، 2000، 3000، (5000فرد .وأظهرت النتائج أنه يمكن استخدام حجم عينة لا يقل عن (150)مع الاختبارات المكونة من 10)، 20، (30فقرة لتقدير معلمة الصعوبة بدقة وفق نموذج راش، وأن أوساط الأخطاء المعيارية في تقدير معلمة الصعوبة تقل كلما زاد حجم العينة المستخدم .

دراسة (Finch & French, 2019) التي تهدف لمقارنة طرق تقدير معالم الفقرات وفق نماذج نظرية الاستجابة للفقرة وذلك باختلاف عدد أفراد العينة، حيث تم توليد بيانات لعينات ذات أحجام 25)، 50، 100، 250، 500، (1000فرد، لاختبارات بعدد 10)، 20، 30، 40، (50فقرة، وباستخدام برمجية R في تحليل البيانات، أشارت النتائج أن استخدام طريقة المزاوجة أعطت تقديرات أدق لمعلمة الصعوبة مقارنة بطريقتي الأرجحية العظمى وبييز، وأن الأخطاء المعيارية في تقدير معلمة الصعوبة تقل بزيادة حجم العينة، كما أن زيادة عدد فقرات الاختبار ليس بالضرورة أن يؤدي إلى دقة أفضل في تقدير معلمة الصعوبة.

وفي دراسة الياسين (2021)بعنوان أثر حجم العينة وطول الاختبار على دقة التقدير للخصائص السيكومترية للفقرات والاختبار باستخدام طريقة المزاوجة للتقدير، والتي تهدف إلى الكشف عن دقة تقدير معلمة صعوبة الفقرة لنموذج راش في ضوء تغير حجم العينة وعدد فقرات الاختبار .حيث تم توليد بيانات ثنائية الاستجابة لأربعة نماذج من الاختبارات المختلفة 20)، 30، 40، (70فقرة، وخمسة أحجام للعينة

50)، 100، 250، 500، (1000فرد، وتم استخدام برمجية R لتقدير معلمة الصعوبة للفقرات باستخدام طريقة المزاوجة، وتقدير قدرة الأفراد باستخدام طريقة الأرجحية الموزونة .وأظهرت النتائج وجود فروق ذات دلالة إحصائية عند مستوى 0.05 بين متوسطات الأخطاء المعيارية لتقديرات معلمة الصعوبة تعزى لحجم العينة لصالح حجم (1000)فرد، كما أظهرت النتائج أخطاء معيارية قليلة في تقدير معلمة الصعوبة في جميع اختبارات الدراسة المطبقة على حجوم العينات تراوحت بين 0.002و .0.05وأشارت النتائج إلى وجود فروق ذات دلالة إحصائية عند مستوى 0.05بين متوسطات الأخطاء المعيارية لتقديرات معلمة القدرة تعزى لحجم العينة وطول الاختبار والتفاعل بينهما .

وفي دراسة الزهراني (2022)التي تهدف إلى استقصاء أثر عدد فقرات اختبار الاختيار من متعدد على دقة تقدير معالم الفقرات وقدرات الأفراد وفق النموذج اللوجستي ثلاثي البارامتر، حيث تم تطبيق نموذجي اختبار مؤلفين من 30)، (60 فقرة، وقد تم استخدام برمجية BILOG-MG3 لإجراء التحليل وتقدير معالم الفقرات وقدرات الأفراد .أظهرت النتائج أن متوسطات معلمتي الصعوبة والتمييز لاختبار (60)فقرة أعلى من متوسطات معلمتي الصعوبة والتمييز لاختبار (30) فقرة، كما أظهرت عدم وجود فروق ذات دلالة إحصائية بين متوسطي معاملات التخمين ومتوسط الخطأ المعياري في تقديرها تبعاً لنموذجي الاختبار .وأشارت نتائج الدراسة إلى وجود فروق ذات دلالة إحصائية بين متوسطي قدرة الأفراد ومتوسطي الأخطاء المعيارية في تقدير القدرة، حيث كان متوسط الاختبار (60)فقرة أعلى من متوسط الاختبار (30)فقرة، وهذا يدل على أن عدد الفقرات (60)أثرت بشكل أكبر على تقدير القدرة.

التعليق على الدراسات والبحوث السابقة:

- الدراسات السابقة التي أتيح للباحث الاطلاع عليها تشترك مع الدراسة الحالية في الاهتمام بدقة تقديرات معالم الفقرات وقدرات الأفراد في نماذج الاستجابة

للفقرة.

- العديد من الدراسات السابقة كانت عبارة عن دراسات محاكاة تم توليد بياناتها باستخدام البرامج الحاسوبية مثل(1978) Reckase & Mark :، Huang & ،Fitzpzatric & Ann (2001) ،Lio, Sinica & Yu (1999) Lin & Shin (2001)، الزبون (2013)، بني عطا (2014)، Jiang et al (2016)، Sahin & Anil (2017)، Finch & French (2019)، الياسين (2021).
- المتغيرات التجريبية التي استخدمت في الدراسات السابقة ركزت في الغالب على متغير واحد إلى متغيرين من حجم العينة وطول الاختبار وعدد البدائل.
- تنوعت النماذج المستخدمة لمقارنة دقة تقدير المعالم للفقرات والأفراد .حيث تم استخدام النموذج ثلاثي البارامتر كما في دراسة Reckase & Mark (1978)، Hambleton & cook (1980)، Lio, Sinica & Yu (1999)، التعمري (2003)، عبابنة (2004)، عيلبوني (2007)، الزبون (2013)، بني عطا (2014)، الزهراني .(2022)
- تم استخدام برامج حاسوبية متنوعة مثل WinGen لتوليد البيانات، BILOG، Xcalibre، برمجية R لتقدير معالم الفقرات وقدرات الأفراد.
- تمثلت أغلب مقاييس التقدير في الدراسات السابقة بمعامل الارتباط، مؤشر الكفاءة، ومتوسطات الانحراف، التحيز وحجم التأثير.
- مما سبق أيضاً فإن الدراسة الحالية تعد استكمالاً لما أوصت به الدراسات السابقة من إجراء بحوث مستقبلية تتضمن بيانات محاكاة لمعرفة أثر حجم العينة، وطول الاختبار وعدد بدائل الاستجابة في دقة تقدير معالم الفقرات وقدرة الأفراد باستخدام النموذج اللوجستي ثلاثي البارامتر .وتتميز بأن لها ثلاثة مستويات العينات 500)، 1000، (2500، وثلاثة مستويات من عدد فقرات الاختبار 30)، 40، (50، وأيضًا ثلاثة مستويات من عدد بدائل الاستجابة 2)، 3، 5).

منهج وإجراءات الدراسة

منهج الدراسة:

استخدم الباحث في الدراسة الحالية المنهج التجريبي للإجابة على تساؤلاتها والمتعلقة بدراسة أثر اختلاف كل من طول الاختبار، وعدد بدائل الاستجابة، وحجم العينة في دقة تقدير قدرات الأفراد ومعالم الفقرات .

يرجع سبب استخدام المنهج التجريبي دون غيره كونه المنهج المناسب لدراسة أثر المتغيرات التجريبية التي تم التحكم بها وضبطها من خلال تصميم المحاكاة على المتغير التابع، كما أنه يوفر فهماً عن العلاقة بين المتغيرات المستقلة والمتغيرات التابعة وهي علاقة سببية موجهة.

إن استخدام البيانات القائمة على المحاكاة بطرق المونت كارلو (MCM)تحقق أقصى درجات الضبط التجريبي، حيث يتم اختيار البيانات بطريقة عشوائية تماماً من خلال العينات المولدة بواسطة برنامج . WinGen (Han, 2007)

في الدراسة الحالية، تم ضبط جميع المتغيرات التي يمكن أن تؤثر على النتائج، مثل خصائص الأفراد الممثلة بقدرات الأفراد، حيث تم تحديد مدى معين بمتوسط وانحراف معياري محددين، وخصائص الفقرات أيضاً، حيث كانت جميع الفقرات أحادية البعد ثنائية الاستجابة 0)، (1ولها توزيعات محددة في مدى معين بمتوسط وانحراف معياري محدد لكل بارامتر من بارامترات الصعوبة، التمييز، والتخمين.

وعليه؛ فإن المتغيرات التجريبية الممثلة في الدراسة الحالية بطول الاختبار 30)، 40، (50مفردة، وعدد بدائل الاستجابة 3)، 4، (5بديل لكل فقرة، وحجم العينة 500)، 1000، (2500فرد، والنموذج اللوجستي ثلاثي البارامتر هي فقط دون غيرها التي ستؤثر في المتغيرات التابعة المتمثلة في تقدير قدرات الأفراد ومعالم الفقرات وبمعزل عن أي متغيرات أخرى.

مجتمع الدراسة:

يعتبر مجتمع الدراسة هو جميع الأفراد الذين يملكون مستوى القدرة المحدد في

التصميم، وجميع الاختبارات أحادية البعد ثنائية الاستجابة 0)، (1والتي لها نفس توزيع البارامترات المحددة في تصميم الدراسة.

عينة الدراسة:

عينة عشوائية مكونة من الأفراد بحجم 500)، 1000، (2500فرد، وعينات عشوائية من الاختبار بطول 30)، 40، (50فقرة، لعدد بدائل 3)، 4، (5بديل استجابة يتم اختيارها من البيانات المولدة بواسطة برنامج WinGen وفقاً لخطوات البرنامج والبيانات التي تحقق تصميم الدراسة .

ظروف الاختبار:

تم توليد بيانات الدراسة الحالية التي تهدف إلى المقارنة بين دقة تقديرات قدرات الأفراد ومعالم الفقرات لسبعة وعشرين ظرفاً اختبارياً استنادًا إلى النموذج اللوغاريتمي ثلاثي البارامتر، وكل ظرف اختباري منها تم تعريفه بتركيبة مكونة من ثلاثة عوامل هي :طول الاختبار، عدد بدائل الاستجابة، وحجم العينة.

وعليه؛ فقد تم توليد عينات بحجم 500)، 1000، (2500فرد، واختبارات بأطوال 30)، 40، (50مفردة، وبعدد بدائل استجابة 3)، 4، (5وذلك من خلال تصميم تجريبي يحاكي الظروف التجريبية الموضحة في الجدول (1)التالي:

جدول (1)

ملخص الظروف التجريبية

الظرف	طول الاختبار	عدد البدائل	حجم العينة
1	30	3	500
2	30	3	1000
3	30	3	2500
4	30	4	500
5	30	4	1000
6	30	4	2500
7	30	5	500
8	30	5	1000
9	30	5	2500

الظرف	طول الاختبار	عدد البدائل	حجم العينة
10	40	3	500
11	40	3	1000
12	40	3	2500
13	40	4	500
الظرف	طول الاختبار	عدد البدائل	حجم العينة
14	40	4	1000
15	40	4	2500
16	40	5	500
17	40	5	1000
18	40	5	2500
19	50	3	500
20	50	3	1000
21	50	3	2500
22	50	4	500
23	50	4	1000
24	50	4	2500
25	50	5	500
26	50	5	1000
27	50	5	2500
حجم العينة الكلي			650

يوضح الجدول (1)السابق جميع الظروف التجريبية للاختبار، حيث تم توليد استجابات افتراضية تحاكي احتمالية استجابة الافراد الافتراضيين على الاختبارات المولدة، وقد نتج عن ذلك 27 مصفوفة استجابات بمجموع فقرات بلغ 1080فقرة ومجموع عينات بلغ 36000فردًا.

تم استخدام اختبارات بأطوال 30)، 40، (50فقرة، بحيث كانت معايير الأسئلة للفقرات الـ (30)هي نفسها معايير الـ (40)و الـ (50)فقرة، بالإضافة لاستخدام بدائل استجابة لكل فقرة من فقرات نماذج الاختبار بأطوال 3)، 4، (5بديل استجابة، لثلاث عينات من الاختبار 500)، 1000، (2500فرد وذلك بعد مراجعة الدراسات السابقة ذات الصلة .وقد تم توليد الافراد بقدرة توزيعها طبيعي بمتوسط

حسابي (0)وانحراف معياري (1)لكل ظرف من الظروف، وتم توليد معلمة التمييز وفقا للتوزيع المنتظم بقيمة ابتدائية (0.5)وقيمة عليا (1.5)ومن خلال التوزيع الطبيعي اللوغاريتمي، وتم توليد معلمة الصعوبة بتوزيع مشابه لتوزيع القدرة حيث خضعت للتوزيع الطبيعي بمتوسط حسابي (0)وانحراف معياري (1)، بينما تم توليد معلمة التخمين وفقا للظروف التجريبية حيث تم توليد وفقاً للتوزيع المنتظم بقيمة ابتدائية وقيم عظمى بلغت 0.33)، 0.25، (0.2عندما كان عدد البدائل 3)، 4، (5 على الترتيب.

المعالجة الإحصائية للبيانات:

للإجابة على تساؤلات الدراسة الحالية فقد قام الباحث بتحديد الأساليب والبرامج الإحصائية في ضوء مشكلة الدراسة وأهدافها متمثلة في الآتي:

- حساب بعض الإحصاءات الوصفية كالمتوسطات الحسابية، والانحرافات المعيارية، والتحليل العاملي، تحليل التباين، واختبار شيفيه للمقارنات البعدية عن طريق برنامج .spss
- تحليل استجابات الأفراد لتقدير قدراتهم ومعالم الفقرات وفق نظرية الاستجابة للفقرة عن طريق برامج .(R, Bilog-Mg3, LDID)

نتائج الدراسة وتفسيرها

يتضمن هذا الجزء عرضاً لنتائج الدراسة ومناقشتها وتفسيرها وفقاً لأسئلة الدراسة الستة والتي تهدف بشكل رئيسي إلى معرفة أثر طول الاختبار 30)، 40، (50 وعدد بدائل الاستجابة 3)، 4، (5، وحجم العينة 500)، 1000، (2500على تقديرات قدرات الأفراد ومعالم الفقرات.

وفيما يلي عرض النتائج التي تم التوصل لها:

التحقق من افتراضات نظرية الاستجابة للفقرة والمطابقة:

تم التحقق من افتراضات نظرية الاستجابة وهي:

أولاً :التحقق من افتراض أحادية البعد. حيث تم إجراء تحليل عاملي استكشافي للاستجابات الافتراضية وحساب نسبة الجذر الكامن الأول إلى الجذر الكامن الثاني والتي يجب أن تتجاوز القيمة .(2)

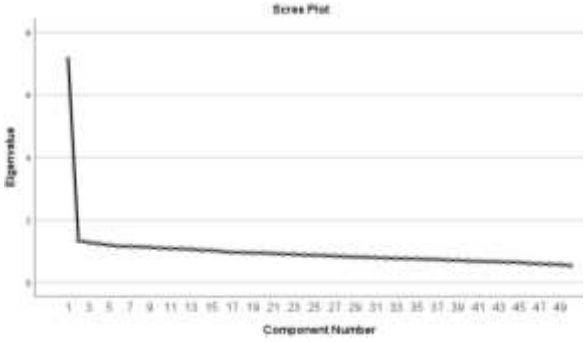
الجدول (2)

التحليل العاملي الاستكشافي للتحقق من أحادية البعد

العامل	قيمة الجذر الكامن	الجذر الكامن الأول / الجذر الكامن الثاني	المعيار
الأول	7.134	5.351	2<
الثاني	1.333		

كما تم تمثيل الجذور الكامنة بيانياً باستخدام (scree-plot) للتحقق من أحادية البعد نظريه من خلال الانعطاف في منحنى (scree-plot)كما هو مبين في الشكل .(7)

شكل (7)التمثيل البياني للجذور الكامنة للتحقق من أحادية البعد



ثانياً :التحقق من افتراض الاستقلال الموضعي. حيث تم استخدام برمجية (Bilog-MG3)وبرمجية (R)لتحليل استجابات الأفراد الافتراضية ومن ثم استخدام برمجية (LDID)المتخصصة في الاستقلال الموضعي لحساب معاملات الارتباط بين البواقي وبعض المؤشرات الإحصائية الدالة على الاستقلال الموضعي مثل (X^2 ,G^2 ,Z_Q, Z_{Q3})وقد تبين أن البيانات تحقق هذا الافتراض حيث كانت نسبة

الأزواج المستقلة تشكل 10أضعاف الأزواج غير المستقلة على أقل تقدير ويبين الشكل التالي جزء من المخرجات البر مجية لأحد الظروف التجريبية.

الشكل (8)

مؤشرات الاستقلال الموضعي من مخرجات برمجية LDID

ITEM	ITEM	O(COR)	E(COR)	D(COR)	
1	3	.13236	.13019	.00217	
ITEM	ITEM	X^2	G^2	Q_3	Z_D
1	3	41.54898	38.56534	.03542	.06767
				Z_Q_3	
P-VALUE		.00000	.00000	.03544	.94605
ITEM	ITEM	MISSING			
1	4	0			
ROW	COL	O(R,C)	E(R,C)		
1	1	157.0000	149.1783		
1	2	107.0000	62.6604		
2	1	362.0000	409.3793		
2	2	374.0000	378.7731		
ITEM	ITEM	O(COR)	E(COR)	D(COR)	
1	4	.09074	.15204	-.06130	
ITEM	ITEM	X^2	G^2	Q_3	Z_D
1	4	37.31199	31.99090	-.00792	-1.96764
				Z_Q_3	
P-VALUE		.00000	.00000	-.00792	.04911
ITEM	ITEM	MISSING			

ثالثاً :مطابقة الأفراد للنموذج. حيث تم تحليل الاستجابات باستخدام برمجية BilogMG3ومن خلال مؤشر الاحتمال الهامشي تبين أن جميع الافراد مطابقين للنموذج وبقي عددهم كما هو عند مرحلة التوليد.

رابعاً :مطابقة الفقرات للنموذج .حيث تم الاعتماد على مؤشر مربع كاي لمطابقة الفقرات للنموذج من مخرجات برمجية BilogMG3وتبين أن (81) فقرة غير مطابقة للنموذج من مجموع الفقرات البالغ عددها (1080)حيث تم استبعاد تلك الفقرات وإعادة التحليل على الفقرات المطابقة وعددها .(999)

خامساً :تم إعادة التحليل على الاستجابات على الفقرات المطابقة للإجابة عن أسئلة الدراسة.

وللإجابة على تساؤل الدراسة الرئيسي **"ما أثر اختلاف طول الاختبار وعدد بدائل الاستجابة وحجم العينة في دقة تقدير قدرات الأفراد ومعالم الفقرات وفق النموذج اللوجستي ثلاثي البارامتر؟"** فقد تم الإجابة عن كل التساؤلات الفرعية التالية:

النتائج المتعلقة بالسؤال الاول الذي ينص على "هل يوجد أثر لطول الاختبار على تقدير معلمة القدرة للأفراد وفق النموذج اللوجستي ثلاثي البارامتر؟"

للإجابة عن السؤال فقد تم تحليل استجابات الأفراد الافتراضيين على الاختبارات المولدة التي تحاكي فقراتها ظروف الدراسة، حيث تم تقدير قدرة كل من أفراد العينة وفق النموذج ثلاثي المعلمة واستخراج دقة تقدير القدرة ممثلة بالخطأ المعياري لدقة التقدير .كما تم استخراج الوسط الحسابي لخطأ التقدير المعياري (se) لتقدير القدرة في ضوء محاكاة الظروف التجريبية وتحديداً طول الاختبار بمستوياته الثلاثة30) ، 40، (50فقرة كما هو مبين في الجدول .(3)

الجدول (3)

دقة تقدير القدرة باختلاف طول الاختبار

طول الاختبار	العدد	الوسط الحسابي	الانحراف المعياري
30	12000	.4522	.05312
40	12000	.4101	.06864
50	12000	.3694	.08719
الكلي	36000	.4106	.07868

من الجدول (3)يتضح وجود فروق ملاحظة في الأوساط الحسابية للخطأ في دقة تقدير القدرة باختلاف طول الاختبار .حيث أن الوسط الحسابي للخطأ المعياري في دقة تقدير القدرة بلغ 0.4522)، 0.4101، (0.3694عندما كان طول الاختبار 30)، 40، (50على الترتيب أي أن دقة تقدير القدرة كانت أكبر عندما كان الاختبار اطول (50)، حيث يعتبر مؤشر خطأ التقدير المعياري مؤشر عكسي لدقة التقدير .

وللكشف عن دلالة تلك الفروق تم إجراء تحليل تباين كما هو مبين في الجدول (4).

الجدول (4)تحليل التباين في الخطأ المعياري لتقدير القدرة باختلاف طول الاختبار

مصدر التباين	مجموع المربعات	درجة الحرية	متوسط المربعات	Fقيمة	الدلالة
بين المجموعات	.007	2	.004	.527	.590
داخل المجموعات	6.675	996	.007		
الكلي	6.682	998			

من نتائج تحليل التباين يلاحظ عدم وجود فروق دالة إحصائياً في دقة تقدير القدرة تعزى إلى طول الاختبار .أي أن دقة تقدير القدرة لا تتأثر بطول الاختبار.

النتائج المتعلقة بالسؤال الثاني الذي ينص على "هل يوجد أثر لطول الاختبار على تقدير معالم الفقرات وفق النموذج اللوجستي ثلاثي البارامتر؟"

وللإجابة عن السؤال تم تحليل استجابات الأفراد الافتراضيين على الاختبارات المولدة التي تحاكي فقراتها ظروف الدراسة حيث تم تقدير معالم الفقرات وفق النموذج ثلاثي المعلمة ودقة تقدير كل منها ممثلة بالخطأ المعياري لدقة التقدير . كما تم استخراج الوسط الحسابي لخطأ التقدير المعياري (se)لتقدير كل من معالم فقرات الاختبار الافتراضي الذي يحاكي الظروف التجريبية وتحديداً طول الاختبار وله ثلاثة مستويات 30)، 40، (50كما هو مبين في الجدول.(5)

الجدول (5)دقة تقدير معالم فقرات الاختبار باختلاف طول الاختبار

طول الاختبار		الخطأ المعياري لدقة تقدير المعلمة		
		التمييز	الصعوبة	التخمين
30	الوسط الحسابي	.1889	.2538	.0670
	الانحراف المعياري	.07912	.11734	.01854
	العدد	228	228	228
40	الوسط الحسابي	.1869	.2322	.0635

طول الاختبار		الخطأ المعياري لدقة تقدير المعلمة		
		التمييز	الصعوبة	التخمين
	الانحراف المعياري	.07774	.10338	.02046
	العدد	340	340	340
50	الوسط الحسابي	.1825	.2352	.0619
	الانحراف المعياري	.08634	.11246	.02211
	العدد	431	431	431
الكلي	الوسط الحسابي	.1855	.2384	.0636
	الانحراف المعياري	.08183	.11083	.02085
	العدد	999	999	999

من الجدول (5)يلاحظ وجود فروق ملاحظة في الأوساط الحسابية للخطأ في دقة تقدير معالم فقرات الاختبار باختلاف طول الاختبار .فيلاحظ أن الوسط الحسابي للخطأ المعياري في دقة تقدير معلمة التمييز بلغ 0.1889)، 0.1869، (0.1825 عندما كان طول الاختبار 30)، 40، (50على الترتيب. أي أن دقة تقدير معلمة التمييز كانت أكبر عندما كان الاختبار أطول (50)، إذ يعتبر مؤشر خطأ التقدير المعياري مؤشر عكسي لدقة التقدير .كما يلاحظ أن الوسط الحسابي للخطأ المعياري في دقة تقدير معلمة الصعوبة بلغ 0.2538)، 0.2322، (0.2352عندما كان طول الاختبار 30)، 40، (50على الترتيب .ويلاحظ أن الوسط الحسابي للخطأ المعياري في دقة تقدير معلمة التخمين بلغ 0.0670)، 0.0635، (0.0619 عندما كان طول الاختبار 30)، 40، (50على الترتيب أي ان تقدير معلمة التخمين كان أكبر عندما كان الاختبار أطول (50)، حيث يعتبر مؤشر خطأ التقدير المعياري مؤشر عكسي لدقة التقدير .

وللكشف عن دلالة تلك الفروق تم اجراء تحليل تباين كما هو مبين في الجدول (6).

الجدول (6)تحليل التباين في الخطأ المعياري لتقدير كل من معالم الفقرة باختلاف طول الاختبار

المعلمة	مصدر التباين	مجموع المربعات	درجة الحرية	متوسط المربعات	Fقيمة	الدلالة
التمييز	بين المجموعات	.007	2	.004	.527	.590
	داخل المجموعات	6.675	996	.007		
	الكلي	6.682	998			
الصعوبة	بين المجموعات	.072	2	.036	2.926	.054
	داخل المجموعات	12.187	996	.012		
	الكلي	12.259	998			
التخمين	بين المجموعات	.004	2	.002	4.541	.011
	داخل المجموعات	.430	996	.000		
	الكلي	.434	998			

من نتائج تحليل التباين يلاحظ عدم وجود فروق دالة إحصائياً في دقة تقدير كل من معلمتي التمييز والصعوبة تعزى إلى طول الاختبار .أي أن دقة تقدير معلمتي التمييز والصعوبة لا تتأثر بطول الاختبار .بينما يلاحظ أن الفروق في دقة تقدير معلمة التخمين والتي تعزى لطول الاختبار ذات دلالة إحصائية .أي أن دقة تقدير معلمة التخمين تتأثر بطول الاختبار. ولمعرفة لصالح من كانت الفروق تم إجراء مقارنات بعدية بطريقة شيفيه كما هو مبين في الجدول .(7)

الجدول (7)المقارنات البعدية للفروق في دقة تقدير معلمة التخمين العائدة لطول الاختبار بطريقة شيفيه

طول الاختبار	30	40	50
30	-	.00351	.00512*
40	.143	-	.00161
50	.011	.564	-
*القيم فوق القطر تشير للفروق والقيم تحت القطر تشير لدلالة الفروق. :			

من نتائج المقارنات البعدية يلاحظ أن الفروق في دقة التقدير كانت لصالح الاختبار الأطول (50) فقرة بمقابل الاختبار الأقصر (30) فقرة.

النتائج المتعلقة بالسؤال الثالث الذي ينص على "هل يوجد أثر لعدد بدائل الاستجابة لفقرات الاختبار على تقدير معلمة القدرة للأفراد وفق النموذج اللوجستي ثلاثي البارامتر؟"

وللإجابة عن السؤال فقد تم تحليل استجابات الأفراد الافتراضيين على الاختبارات المولدة التي تحاكي فقراتها ظروف الدراسة حيث تم تقدير قدرة كل من أفراد العينة وفق النموذج ثلاثي المعلمة واستخراج دقة تقدير القدرة ممثلة بالخطأ المعياري لدقة التقدير .كما تم استخراج الوسط الحسابي لخطأ التقدير المعياري (se)لتقدير القدرة في ضوء محاكاة الظروف التجريبية وتحديداً عدد بدائل الاستجابة لفقرات الاختبار بثلاثة مستويات3) ، 4، (5كما هو مبين في الجدول .(8)

الجدول (8) دقة تقدير القدرة باختلاف عدد بدائل الاستجابة لفقرات الاختبار

عدد بدائل الاستجابة لفقرات الاختبار	العدد	الوسط الحسابي	الانحراف المعياري	
3	12000		.4179	.07655
4	12000		.4133	.07561
5	12000		.4004	.08263
الكلي	36000		.4106	.07868

من الجدول (8)يتبين وجود فروق ملاحظة في الأوساط الحسابية للخطأ في دقة تقدير القدرة باختلاف عدد بدائل الاستجابة لفقرات الاختبار .حيث أن الوسط الحسابي للخطأ المعياري في دقة تقدير القدرة بلغ 0.4179)، 0.4133، (0.4004 عندما كان عدد بدائل الاستجابة لفقرات الاختبار 3)، 4، (5على الترتيب أي أن دقة تقدير القدرة كانت أكبر عندما كان عدد البدائل أكبر (5)، إذ يعتبر مؤشر خطأ التقدير المعياري مؤشر عكسي لدقة التقدير .وللكشف عن دلالة تلك الفروق تم إجراء تحليل تباين كما هو مبين في الجدول .(9)

الجدول (9)تحليل التباين في الخطأ المعياري لتقدير القدرة باختلاف عدد بدائل الاستجابة لفقرات الاختبار

مصدر التباين	مجموع المربعات	درجة الحرية	متوسط المربعات	Fقيمة	الدلالة
بين المجموعات	1.981	2	.990	161.417	.000
داخل المجموعات	220.846	35997	.006		
الكلي	222.827	35999			

من نتائج تحليل التباين يتبين وجود فروق في دقة تقدير القدرة تعزى الى عدد البدائل .وهي ذات دلالة إحصائية عند مستوى الدلالة .(α<0.05)ولمعرفة لصالح من كانت الفروق تم اجراء مقارنات بعدية بطريقة شيفيه كما هو مبين في الجدول (10).

الجدول (10)المقارنات البعدية للفروق في دقة تقدير القدرة العائدة لعدد بدائل فقرات الاختبار بطريقة شيفيه

عدد البدائل	3	4	5
3	-	.00462*	.01753*
4	.000	-	.01291*
5	.000	.000	-
:*القيم فوق القطر تشير للفروق والقيم تحت القطر تشير لدلالة الفروق.			

من نتائج المقارنات البعدية يتبين أن الفروق في دقة تقدير القدرة كانت لصالح عدد البدائل الأكبر عند مقارنة أي منهما.

النتائج المتعلقة بالسؤال الرابع الذي ينص على "هل يوجد أثر لعدد بدائل الاستجابة لفقرات الاختبار على تقدير معالم الفقرات وفق النموذج اللوجستي ثلاثي البارامتر؟"

وللإجابة عن هذا السؤال تم تحليل استجابات الأفراد الافتراضيين على الاختبارات المولدة التي تحاكي فقراتها ظروف الدراسة حيث تم تقدير معالم الفقرات وفق النموذج ثلاثي المعلمة ودقة تقدير كل منها ممثلة بالخطأ المعياري لدقة التقدير.

كما تم استخراج الوسط الحسابي لخطأ التقدير المعياري (se)لتقدير كل من معالم فقرات الاختبار الافتراضي الذي يحاكي الظروف التجريبية وتحديداً عدد البدائل للاستجابة على فقرات الاختبار ولها ثلاث مستويات 3)، 4، (5كما هو مبين في الجدول .(11)

الجدول (11)دقة تقدير معالم فقرات الاختبار باختلاف عدد بدائل فقرات الاختبار

عدد البدائل		الخطأ المعياري لدقة تقدير المعلمة		
		التمييز	الصعوبة	التخمين
3	الوسط الحسابي	.1884	.2563	.0659
	الانحراف المعياري	.08245	.12452	.02119
	العدد	324	324	324
4	الوسط الحسابي	.1826	.2351	.0641
	الانحراف المعياري	.08094	.10393	.01980
	العدد	339	339	339
5	الوسط الحسابي	.1856	.2247	.0607
	الانحراف المعياري	.08225	.10117	.02129
	العدد	336	336	336
الكلي	الوسط الحسابي	.1855	.2384	.0636
	الانحراف المعياري	.08183	.11083	.02085
	العدد	999	999	999

من الجدول (11)يلاحظ وجود فروق ملاحظة في الأوساط الحسابية للخطأ في دقة تقدير معالم فقرات الاختبار باختلاف عدد بدائل فقرات الاختبار .حيث بلغ الوسط الحسابي للخطأ المعياري في دقة تقدير معلمة التمييز 0.1884)، 0.1826، (0.1856عندما كان عدد بدائل فقرات الاختبار 3)، 4، (5على الترتيب .كما بلغ الوسط الحسابي للخطأ المعياري في دقة تقدير معلمة الصعوبة 0.2563)، 0.2351، (0.2247عندما كان عدد بدائل فقرات الاختبار 3)، 4، (5على الترتيب .ويلاحظ

أن الوسط الحسابي للخطأ المعياري في دقة تقدير معلمة التخمين بلغ 0.0659)، 0.0641، (0.0607عندما كان عدد بدائل فقرات الاختبار 3)، 4، (5على الترتيب أي أن تقدير معلمة التخمين كان أكبر عندما كان عدد البدائل أكبر (5)، إذ يعتبر مؤشر خطأ التقدير المعياري مؤشر عكسي لدقة التقدير .وللكشف عن دلالة تلك الفروق تم إجراء تحليل تباين كما هو مبين في الجدول .(12)

الجدول (12)

تحليل التباين في الخطأ المعياري لتقدير كل من معالم الفقرة باختلاف عدد بدائل فقرات الاختبار

المعلمة	مصدر التباين	مجموع المربعات	درجة الحرية	متوسط المربعات	قيمةF	الدلالة
التمييز	بين المجموعات	.006	2	.003	.416	.660
	داخل المجموعات	6.676	996	.007		
	الكلي	6.682	998			
الصعوبة	بين المجموعات	.171	2	.085	7.031	.001
	داخل المجموعات	12.088	996	.012		
	الكلي	12.259	998			
التخمين	بين المجموعات	.005	2	.002	5.373	.005
	داخل المجموعات	.429	996	.000		
	الكلي	.434	998			

من نتائج تحليل التباين يلاحظ عدم وجود فروق دالة إحصائياً في دقة تقدير معلمة التمييز تعزى إلى عدد بدائل فقرات الاختبار .أي دقة تقدير معلمة التمييز لا تتأثر بعدد بدائل فقرات الاختبار .بينما يلاحظ أن الفروق في دقة تقدير معلمتي الصعوبة والتخمين والتي تعزى لعدد بدائل فقرات الاختبار ذات دلالة إحصائية .أي أن دقة تقدير كل من معلمتي الصعوبة والتخمين تتأثر بعدد بدائل فقرات الاختبار. ولمعرفة لصالح من كانت الفروق تم إجراء مقارنات بعدية بطريقة شيفيه كما هو مبين في الجدولين (13)، .(14)

الجدول (13)المقارنات البعدية للفروق في دقة تقدير معلمة الصعوبة العائدة لعدد بدائل فقرات الاختبار بطريقة شيفيه

عدد البدائل	3	4	5
3	-	.02119*	.03161*
4	.047	-	.01042
5	.001	0.470	-
*القيم فوق القطر تشير للفروق والقيم تحت القطر تشير لدلالة الفروق:			

من نتائج المقارنات البعدية يلاحظ أن الفروق في دقة تقدير معلمة الصعوبة كانت لصالح كل من عددي البدائل 4)، (5بمقابل عدد بدائل .(3)

الجدول (14)المقارنات البعدية للفروق في دقة تقدير معلمة التخمين العائدة لعدد بدائل فقرات الاختبار بطريقة شيفيه

عدد البدائل	3	4	5
3	-	.00182	.00521*
4	.530	-	.00339
5	.006	.106	-
*القيم فوق القطر تشير للفروق والقيم تحت القطر تشير لدلالة الفروق. :			

من نتائج المقارنات البعدية يلاحظ أن الفروق في دقة تقدير معلمة التخمين كانت لصالح عدد بدائل (5)بمقابل عدد بدائل .(3)

النتائج المتعلقة بالسؤال الخامس الذي ينص على "هل يوجد أثر لحجم العينة على تقدير معلمة القدرة للأفراد وفق النموذج اللوجستي ثلاثي البارامتر؟"

وللإجابة عن السؤال تم تحليل استجابات الأفراد الافتراضيين على الاختبارات المولدة التي تحاكي فقراتها ظروف الدراسة حيث تم تقدير قدرة كل من أفراد العينة وفق النموذج ثلاثي المعلمة واستخراج دقة تقدير القدرة ممثلة بالخطأ المعياري لدقة التقدير. كما تم استخراج الوسط الحسابي لخطأ التقدير المعياري (se)لتقدير القدرة في ضوء محاكاة الظروف التجريبية وتحديداً حجم العينة بمستوياته الثلاثة 500)، 1000، (2500 فرد كما هو مبين في الجدول .(15)

الجدول (15) دقة تقدير القدرة باختلاف حجم العينة

حجم العينة	العدد	الوسط الحسابي	الانحراف المعياري	
500	4500		.3908	.08619
1000	9000		.4115	.07952
2500	22500		.4141	.07615
الكلي	36000		.4106	.07868

من الجدول (15)يلاحظ وجود فروق ملاحظة في الأوساط الحسابية للخطأ في دقة تقدير القدرة باختلاف حجم العينة .حيث بلغ الوسط الحسابي للخطأ المعياري في دقة تقدير القدرة 0.3908)، 0.4115، (0.4141عندما كان حجم العينة 500)، 1000، (2500على الترتيب. أي أن دقة تقدير القدرة كانت أكبر عندما كان حجم العينة أقل(500) فرد، إذ يعتبر مؤشر خطأ التقدير المعياري مؤشر عكسي لدقة التقدير .وللكشف عن دلالة تلك الفروق تم اجراء تحليل تباين كما هو مبين في الجدول .(16)

الجدول (16) تحليل التباين في الخطأ المعياري لتقدير القدرة باختلاف حجم العينة

مصدر التباين	مجموع المربعات	درجة الحرية	متوسط المربعات	Fقيمة	الدلالة
بين المجموعات	2.054	2	1.027	167.473	.000
داخل المجموعات	220.772	35997	.006		
الكلي	222.827	35999			

من نتائج تحليل التباين يتبين وجود فروق في دقة تقدير القدرة تعزى إلى حجم العينة، وهي ذات دلالة إحصائية عند مستوى الدلالة .(α<0.05)ولمعرفة لصالح من كانت الفروق تم إجراء مقارنات بعدية بطريقة شيفيه كما هو مبين في الجدول (17).

الجدول (17)المقارنات البعدية للفروق في دقة تقدير القدرة العائدة لحجم العينة بطريقة شيفيه

حجم العينة	3	4	5
500	-	-.02075*	-.02334*
1000	.000	-	-.00259*
2500	.000	.000	-
*القيم فوق القطر تشير للفروق والقيم تحت القطر تشير لدلالة الفروق.			

من نتائج المقارنات البعدية يتبين أن الفروق في دقة تقدير القدرة كانت لصالح حجم العينة الاصغر عند مقارنة أي منهما.

النتائج المتعلقة بالسؤال السادس الذي ينص على "هل يوجد أثر لحجم العينة على تقدير معالم الفقرات وفق النموذج اللوجستي ثلاثي البارامتر؟"

وللإجابة عن السؤال تم تحليل استجابات الأفراد الافتراضيين على الاختبارات المولدة التي تحاكي فقراتها ظروف الدراسة حيث تم تقدير معالم الفقرات وفق النموذج ثلاثي المعلمة ودقة تقدير كل منها ممثلة بالخطأ المعياري لدقة التقدير . كما تم استخراج الوسط الحسابي لخطأ التقدير المعياري (se)لتقدير كل من معالم فقرات الاختبار الافتراضي الذي يحاكي الظروف التجريبية وتحديداً حجم العينة وله ثلاثة مستويات 500)، 1000، (2500كما هو مبين في الجدول .(18)

الجدول (18)دقة تقدير معالم فقرات الاختبار باختلاف حجم العينة

حجم العينة		الخطأ المعياري لدقة تقدير المعلمة		
		التمييز	الصعوبة	التخمين
500	الوسط الحسابي	.2400	.2634	.0693
	الانحراف المعياري	.08792	.10475	.01829
	العدد	352	352	352
1000	الوسط الحسابي	.1790	.2444	.0644

حجم العينة		الخطأ المعياري لدقة تقدير المعلمة		
		التمييز	الصعوبة	التخمين
	الانحراف المعياري	.06469	.11257	.02011
	العدد	337	337	337
2500	الوسط الحسابي	.1306	.2037	.0562
	الانحراف المعياري	.04295	.10711	.02220
	العدد	310	310	310
الكلي	الوسط الحسابي	.1855	.2384	.0636
	الانحراف المعياري	.08183	.11083	.02085
	العدد	999	999	999

من الجدول (18)يلاحظ وجود فروق في الأوساط الحسابية للخطأ في دقة تقدير معالم فقرات الاختبار باختلاف حجم العينة .حيث بلغ الوسط الحسابي للخطأ المعياري في دقة تقدير معلمة التمييز 0.2400)، 0.1790، (0.1306عندما كان حجم العينة 500)، 1000، (2500على الترتيب. أي أن دقة تقدير معلمة التمييز كانت أكبر عندما كان حجم العينة أكبر (2500)، ويعتبر مؤشر خطأ التقدير المعياري مؤشر عكسي لدقة التقدير .كما يلاحظ أن الوسط الحسابي للخطأ المعياري في دقة تقدير معلمة الصعوبة بلغ 0.2634)، 0.2444، (0.2037عندما كان حجم العينة 500)، 1000، (2500على الترتيب .وأن الوسط الحسابي للخطأ المعياري في دقة تقدير معلمة التخمين بلغ 0.0693)، 0.0644، (0.0562عندما كان حجم العينة 500)، 1000، (2500على الترتيب أي أن تقدير معلمة التخمين كان أكبر عندما كان حجم العينة أكبر (2500) فرد، حيث يعتبر مؤشر خطأ التقدير المعياري مؤشر عكسي لدقة التقدير .وللكشف عن دلالة تلك الفروق تم إجراء تحليل تباين كما هو مبين في الجدول .(19)

الجدول (19) تحليل التباين في الخطأ المعياري لتقدير كل من معالم الفقرة باختلاف حجم العينة

المعلمة	مصدر التباين	مجموع المربعات	درجة الحرية	متوسط المربعات	Fقيمة	الدلالة
التمييز	بين المجموعات	1.993	2	.996	211.658	.000
	داخل المجموعات	4.689	996	.005		
	الكلي	6.682	998			
الصعوبة	بين المجموعات	.605	2	.302	25.852	.000
	داخل المجموعات	11.654	996	.012		
	الكلي	12.259	998			
التخمين	بين المجموعات	.028	2	.014	34.900	.000
	داخل المجموعات	.406	996	.000		
	الكلي	.434	998			

من نتائج تحليل التباين يلاحظ وجود فروق دالة إحصائياً في دقة تقدير كل من معالم الفقرة تعزى إلى حجم العينة .أي أن دقة تقدير معالم فقرات الاختبار تتأثر بحجم العينة .ولمعرفة لصالح من كانت الفروق تم اجراء مقارنات بعدية بطريقة شيفيه كما هو مبين في الجداول .(20)

الجدول (20) المقارنات البعدية للفروق في دقة تقدير معلمة التمييز العائدة لحجم العينة بطريقة شيفيه

حجم العينة	500	1000	2500
500	-	.06102*	.10936*
1000	.000	-	.04834*
2500	.000	.000	-
*القيم فوق القطر تشير للفروق والقيم تحت القطر تشير لدلالة الفروق.			

من نتائج المقارنات البعدية يلاحظ أن الفروق في دقة تقدير معلمة التمييز كانت لصالح حجم العينة الأكبر بمقابل حجم العينة الأقل ولكل زوج منها.

الجدول (21)المقارنات البعدية للفروق في دقة تقدير معلمة الصعوبة العائدة لحجم العينة بطريقة شيفيه

حجم العينة	500	1000	2500
500	-	.01900	.05968*
1000	.071	-	.04068*
2500	.000	.000	-
*القيم فوق القطر تشير للفروق والقيم تحت القطر تشير لدلالة الفروق. :			

من نتائج المقارنات البعدية يلاحظ أن الفروق في دقة تقدير معلمة الصعوبة كانت لصالح حجم العينة (2500)بمقابل كل من الحجمين الاقل 500)، .(1000

الجدول (22) المقارنات البعدية للفروق في دقة تقدير معلمة التخمين العائدة لحجم العينة بطريقة شيفيه

حجم العينة	500	1000	2500
500	-	.00485*	.01305*
1000	.007	-	.00820*
2500	.000	.000	-
*القيم فوق القطر تشير للفروق والقيم تحت القطر تشير لدلالة الفروق. :			

من نتائج المقارنات البعدية يلاحظ أن الفروق في دقة تقدير معلمة التخمين كانت لصالح حجم العينة الأكبر بمقابل حجم العينة الأقل ولكل زوج منها.

ملخص نتائج الدراسة والتوصيات والبحوث المقترحة

- **تفسير النتائج المتعلقة بالسؤال الأول الذي ينص على "هل يوجد أثر لطول الاختبار على تقدير معلمة القدرة للأفراد وفق النموذج اللوجستي ثلاثي البارامتر؟"**

توصلت نتائج هذا السؤال لعدم وجود فروق في دقة تقدير القدرة تعزى إلى طول الاختبار. ويمكن تفسير هذه النتيجة من خلال آلية تقدير القدرة حيث أنها تعتمد على الاحتمال الهامشي المشروط والذي يقارن ما بين العلامة المقدرة باستخدام النموذج واحتمالية الاستجابة على الفقرات بغض النظر عن الدرجة الكلية على

الاختبار. وحسب الإشارات السابقة مثل إشارة لورد فان أحجام العينات المستخدمة في هذه الدراسة كافية لتقدير القدرة بدقة عالية وهذا بدروه أذاب الفروق في دقة التقدير باختلاف أطوال الاختبار .وتتفق نتيجة هذا السؤال مع نتيجة دراسة (Hambleton @ Cook, 1980)، ودراسة عبابنة (2004)، بينما أكدت الدراسة التي أجراها بني عطا (2014)إلى أن دقة تقدير قدرات الأفراد تزداد بزيادة عدد فقرات الاختبار.

- **تفسير النتائج المتعلقة بالسؤال الثاني الذي ينص على "هل يوجد أثر لطول الاختبار على تقدير معالم الفقرات وفق النموذج اللوجستي ثلاثي البارامتر؟"**

توصلت نتائج هذا السؤال لوجود أثر لطول الاختبار على دقة تقدير معلمة التخمين فقط ولصالح الاختبار الأطول .ويمكن تفسير ذلك من خلال مستويات القدرة المتدنية التي تحاكي معلمة التخمين وأن زيادة عدد فقرات الاختبار يزيد من دقة تقدير معلمة التخمين حيث أنها تعتبر معلومات إضافية قادرة على تحديد التقدير الأمثل لمعلمة التخمين. على العكس من معلمتي الصعوبة والتمييز، فمعلمة الصعوبة تحتاج إلى فقرات ذات صعوبة متباينة .ويمكن تأييد ذلك من خلال دالة معلومات الاختبار والتي تعتبر الوجه العكسي لخطأ التقدير حيث يلاحظ من معادلات دالة معلومات الاختبار أن معلمة التخمين تنفرد بعلاقة مختلفة عن معلمتي الصعوبة والتمييز وتنفرد مواقع مختلفة في دالة معلومات الاختبار.

وتتفق نتيجة هذا السؤال مع نتيجة دراسة(Hambleton @ Cook, 1980) ، وبني عطا .(2014).

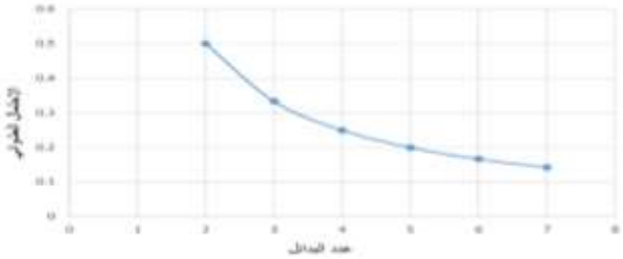
- **تفسير النتائج المتعلقة بالسؤال الثالث الذي ينص على "هل يوجد أثر لعدد بدائل الاستجابة لفقرات الاختبار على تقدير معلمة القدرة للأفراد وفق النموذج اللوجستي ثلاثي البارامتر؟"**

توصلت نتائج هذا السؤال لوجود أثر لعدد البدائل على دقة تقدير القدرة ولصالح الاختبار ذو البدائل الأكثر عدداً .ويمكن تفسير ذلك من خلال قدرة البدائل على تحديد قدرة الفرد فعدد البدائل في علاقة عكسية مع الاستجابات الصحيحة العشوائية. ومن هنا فإن زيادة عدد البدائل يقلل من عدد الاستجابات الصحيحة التي تتم بشكل عشوائي، مما يمكن النموذج من تحديد قدرة الفرد بشكل أدق .والشكل (9) يوضح تلك العلاقة.

الشكل (9)

العلاقة بين عدد البدائل

والاحتمال العشوائي للاستجابة الصحيحة



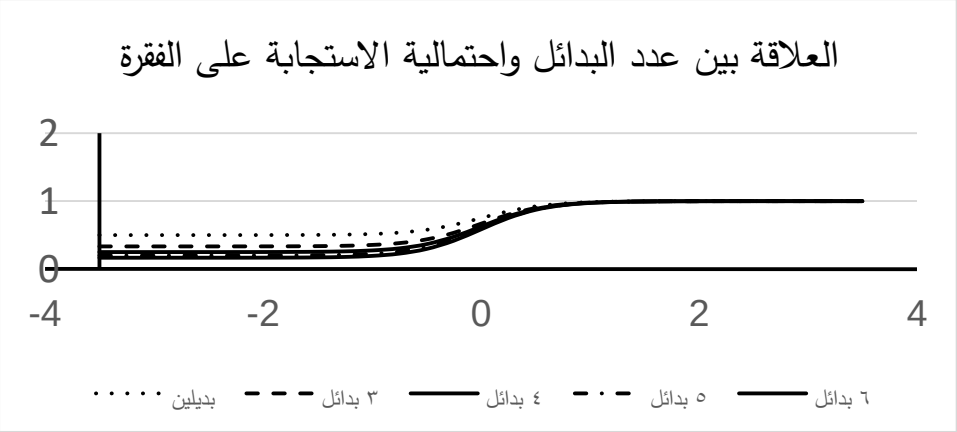
يلاحظ من خلال الشكل (9) السابق أن مستوى الاحتمال العشوائي ينخفض بزيادة عدد البدائل وبذلك يمكن قدير قدرة الفرد بشكل أدق.

- **تفسير النتائج المتعلقة بالسؤال الرابع الذي ينص على "هل يوجد أثر لعدد بدائل الاستجابة لفقرات الاختبار على تقدير معالم الفقرات وفق النموذج اللوجستي ثلاثي البارامتر؟"**

توصلت الدراسة إلى أن عدد البدائل يؤثر على دقة تقدير معلمتي الصعوبة والتخمين فقط وبعلاقة طردية، فزيادة عدد البدائل يزيد من دقة تقدير كل من معلمتي الصعوبة والتخمين ويمكن تفسير ذلك من خلال زيادة قدرة النموذج على التنبؤ بزيادة عدد البدائل وهذا ما يشار له أحياناً بمساحة النموذج الاحتمالية، ويوضح الشكل (10)ذلك.

الشكل (10)

العلاقة بين عدد البدائل واحتمالية الاستجابة على الفقرة



يلاحظ من الشكل أن عدد البدائل يلعب دوراً مهما في تخمين الفقرة مما يؤثر في المدى الاحتمالي للنموذج، ففي الفقرة ذات البديلين يتبين أن المدى الاحتمالي يبدأ من 0.5وحتى 1أي بمدى 0.5وهي قيمة منخفضة وبالتالي فإن النموذج الثلاثي يكون قادرًا على التنبؤ بنسبة أقل، وهذا ينعكس على قدرته على التقدير .بينما في الفقرة ذات الخمسة بدائل يلاحظ أن المدى الاحتمالي يبدأ من 0.2وحتى 1، أي بمدى 0.8 وهي قيمة مرتفعة، وبالتالي يصبح النموذج الثلاثي قادرًا على التنبؤ بنسبة عالية، وهذا ما يفسر الدقة الأعلى في التقدير.

وتتفق نتيجة هذا السؤال مع نتيجة دراسة (Rodriguez, 2005) التي أظهرت أن الاختبار ذو الثلاثة بدائل كان الأفضل في اختبار الاختيار من متعدد.

- **تفسير النتائج المتعلقة بالسؤال الخامس الذي ينص على "هل يوجد أثر لحجم العينة على تقدير معلمة القدرة للأفراد وفق النموذج اللوجستي ثلاثي البارامتر؟"**

توصلت الدراسة لوجود أثر لحجم العينة على دقة تقدير القدرة ولصالح حجم العينة الأقل وهذا ما يمكن تفسيره من خلال الدراسات السابقة التي كانت تشير لأحجام العينة الأنسب في ضوء طول الاختبار، وبالتحديد إشارة لورد التي كانت

تؤكد على أن حجم العينة (1000) مناسب عندما يكون الاختبار بطول (50) فقرة، وفي هذه الدراسة كان متوسط طول الاختبار أقل من (40) فقرة مما يشير إلى أن حجم العينة يجب أن يكون أقل من .(1000) أي أن الحجم الأنسب هو (500)، وهي نفس النتيجة التي خلصت لها الدراسة حيث كانت دقة التقدير أفضل عند حجم العينة (500) بمقابل كل من الحجمين (1000) و .(2500)

والوجه الآخر لما تم ذكره هو اعتماد التقدير على أنماط الاستجابة، فبوجود عدد من الفقرات فإننا بحاجة إلى عدد من الأفراد لتغطية أنماط الاستجابة المحتملة ليتم التقدير من خلال النموذج بفعالية .إلا أن زيادة عدد الأفراد ينتج أنماطاً مكرره أو متضاربة تعيق عملية التقدير بالشكل الأمثل .فزيادة حجم العينة مع ثبات عدد الفقرات يقلل من فاعلية النموذج في تقدير القدرة بمقابل زيادة فعاليته في تقدير معالم الفقرات .وتتفق نتيجة هذا السؤال مع نتيجة دراسة (Hambleton & cook, 1980) التي أكدت أن دقة التقدير تتناسب طردياً مع كل من حجم العينة وعدد فقرات الاختبار .

- **تفسير النتائج المتعلقة بالسؤال السادس الذي ينص على "هل يوجد أثر لحجم العينة على تقدير معالم الفقرات وفق النموذج اللوجستي ثلاثي البارامتر؟"**

توصلت الدراسة لوجود أثر لحجم العينة على دقة تقدير كل من معالم الفقرات ولصالح حجم العينة الأكبر، وهذا ما يمكن تفسيره من خلال اعتماد التقدير على أنماط الاستجابة. فبوجود عدد من الفقرات فإننا بحاجة إلى عدد من الأفراد لتغطية أنماط الاستجابة المحتملة ليتم التقدير من خلال النموذج بفعالية، وأن زيادة عدد الأفراد ينتج أنماطاً مكرره تؤكد على تقدير معالم الفقرات بشكل أكبر .فزيادة عدد الأفراد يولّد استجابات تتبع التوزيع الطبيعي حول معالم الفقرات، أي أن زيادة عدد الأفراد يعزز من تحقيق التوزيع الطبيعي للاستجابات .وزيادة حجم العينة مع ثبات

عدد الفقرات يزيد من فاعلية النموذج في تقدير معالم الفقرات بمقابل انخفاض فعاليته في تقدير القدرة.

التوصيات:

في ضوء النتائج التي تم التوصل لها فإن الباحث يوصي بما يلي:

أولاً :توصيات عملية

- أن تكون الاختبار بطول 30فقرة على الأقل عندما يكون الهدف تقدير القدرة.
- أن تكون الاختبارات طويلة 50)فقرة (عندما يكون الهدف تقدير معالم وفق النموذج ثلاثي المعلمة وبالتحديد التخمين.
- زيادة عدد البدائل إلى أقصى حد ممكن للحصول على أكبر قدر ممكن من الدقة في تقدير القدرة وتقدير معالم الفقرة.
- اختيار حجم عينة مناسب يتوافق مع طول الاختبار عندما يكون الهدف هو تقدير القدرة .وأن يكون أكبر ما يمكن عندما يكون الهدف تقدير معالم الفقرات.

ثانيا :توصيات نظرية

- يوصي الباحث بإعادة إجراء الدراسة على نماذج نظرية الاستجابة للفقرة متعددة الاستجابة.
- يوصي الباحث بإعادة إجراء الدراسة باستخدام النموذجين الأحادي والثنائي ومقارنة نتائجهما بالنموذج الثلاثي.

قائمة المراجع

أولا :المراجع العربية

أبو جراد، حمدي .(2008)استخدام نموذج)راش (في تطوير اختبار كاتل الثالث للذكاء الصورة)أ .(مجلة الجامعة الإسلامية .شؤون البحث العلمي والدراسات العليا بالجامعة الإسلامية .غزة، .583-555 .(2)16

إسماعيل، محمد عبد الرحمن .(2015)أثر عدد بدائل الاستجابة في مقياس ليكرت على الخصائص السيكومترية للمقياس وقياس الاتجاهات .مجلة الإدارة العامة -معهد الإدارة العامة(4)55 ,، 875 - 833

بني عطا، زايد صالح إبراهيم .(2014)تقصي دقة تقدير النموذج اللوجستي ثلاثي المعلمة لمعالم الفقرة وقدرة الأفراد في ضوء تغير طول الاختبار وحجم العينة :دراسة محاكاة .مجلة جامعة الشارقة للعلوم الإنسانية والاجتماعية، (2)11، .37 - 1

بيكر، ف.ب (2010)أسس نظرية الاستجابة للمفردة) .عبدالرحمن الطريري، السيد أبو هاشم، سوسن شلبي، مترجم .(الرياض :جامعة الملك سعود للنشر العلمي.

حسن، ياسر عبد الله حفني .(2019)أثر اختلاف طريقة المعادلة وطرق تقدير الدرجات وقواعد صياغة الفقرات على دقة تقدير معالم الفقرات وقدرات الأفراد في ضوء القياس الكلاسيكي والنموذج اللوجستي ثلاثي البارامتر . مجلة كلية التربية -جامعة أسيوط –كلية التربية، (7)35، – 352 434

الحواري، أروى .(2015)أثر طول الاختبار وشكل تقدير القدرة في تقدير قدرات الأفراد وفق نموذج راش في نظرية استجابة الفقرة .مجلة جامعة النجاح للأبحاث، العلوم الإنسانية، .1488-1463 .(8)29

خريبه، إيناس محمد صفوت مصطفى .(2020)أثر تدريب مفردات الاختبار التحصيلي على خصائص مفرداته في ضوء النموذج اللوجستي ثلاثي المعلم وعلى الاداء الأكاديمي لدى طلبة الدبلوم العام في التربية .مجلة دراسات تربوية ونفسية -جامعة الزقازيق -كلية التربية (107)، 146 – 87

الدرابيع، ماهر .(2001)فعالية النموذج اللوغاريتمي ذي المعلمة الواحدة)نموذج راش (في دقة تقدير قدرة الفرد ومعامل صعوبة الفقرة باختلاف حجم العينة وطول الاختبار .مجلة دراسات العلوم الإنسانية، .208-197 .(6)11

زكري، علي .(2008)الخصائص السيكومترية لاختبار)أوتيس-لينون (للقدرة العقلية مقدرة وفق القياس الكلاسيكي ونموذج راش لدى طلبة المرحلة المتوسطة بمحافظة صبيا التعليمية، رسالة دكتوراة غير منشورة، كلية التربية، جامعة أم القرى.

الزهراني، أحمد رزق الله .(2022)أثر عدد فقرات الاختيار من متعدد وترتيبها على دقة تقدير معالم الفقرات وقدرات الأفراد وفقاً لنظرية الاستجابة للمفردة . رسالة دكتوراة غير منشورة .جامعة أم القرى.

الزهراني، فهد بن شرف محمد .(2016)الخصائص السيكومترية لاختبارات تحصيلي من نوع الاختيار من متعدد والصواب والخطأ باستخدام طريقة التقدير :MLEدراسة مقارنة باستخدام النموذج اللوجستي ثلاثي البارامتر .مجلة الثقافة والتنمية -جمعية الثقافة من أجل التنمية , 17(106)، 122 - 88

سليمان، شاهر خالد .(2017)أثر موقع البديل الصحيح في اختبار اختيار من متعدد على تقديرات معالم الفقرات والقدرة وفق النموذج اللوجستي ثلاثي المعلمة .مجلة دراسات عربية في التربية وعلم النفس -رابطة التربويين العرب، (90)، 120 - 98

الشريفين، نضال كمال .(2017)خصائص توزيع قدرات الأفراد ومعالم فقرات الاختبار وفق نماذج نظرية الاستجابة للفقرة المعلمية واللا معلمية :دراسة مقارنة .المجلة التربوية -جامعة الكويت -مجلس النشر العلمي، (124)31، 265 - 320

الشريفين، نضال كمال محمد .(2019)أثر نموذج نظرية الاستجابة للفقرة ذات الاستجابة المتعددة التدريج على دقة تقدير القدرات للأفراد والخصائص السيكومترية للفقرات والاختبار .المجلة التربوية -جامعة الكويت -مجلس النشر العلمي، (130)33، 295 - 241

الشهري، بندر بن سالم .(2016)فاعلية استخدام كل من النموذج اللوجستي الثلاثي المعلم والنظرية الكلاسيكية في تقدير معالم الفقرات :دراسة مقارنة .المجلة العلمية لكلية التربية -جامعة الوادي الجديد -كلية التربية (23)، 262 - 224

الصيخان، رحاب بنت صالح .(2021)مقارنة بين النظرية الكلاسيكية والنموذج ثلاثي المعلم في انتقاء فقرات اختبار تحصيلي للغة الإنجليزية .المجلة الدولية للدراسات التربوية والنفسية -مركز رفاد للدراسات والأبحاث، 10(1)، 156 – 135

الطريري، عبد الرحمن سليمان .(1997)القياس النفسي والتربوي :نظرياته، أسسه وتطبيقاته .الرياض :مكتبة الرشد.

طنطاوي، شيرين فاروق محمد .(2022)تأثيرات طول الاختبار وطريقة التصحيح ونموذج تحليل المفردة على تقديرات القدرة ومعايير التحسين للاختبارات ذات المفردات المختلطة .مجلة الدراسات النفسية المعاصرة -جامعة بني سويف -كلية الآداب، (1)4، 96 – 38

عاشور، أحمد تيسير صالح .(2018)أثر نسب الفقد في بيانات الجذع المشترك ضمن طريقة التعويض في صورة الاختبار للمستويين المنخفض والمرتفع في دقة

تقديرات القدرة ومعالم الفقرات .مجلة جامعة القدس المفتوحة للأبحاث والدراسات التربوية والنفسية، (27)10، .174 - 145

عبد الجبار، عادل .والشافعي، محمد .(2005)أثر حجم عينة التحليل على مؤشرات الملائمة الإحصائية وتقديرات الصعوبة المتضمنة ببرنامجي)مايكروسكال (و)بايلوج (للمفردات باستخدام نموذج)راش" (دراسة محاكاة ."رسالة التربية وعلم النفس، (25)، .64-1

عبد الوهاب، محمد محمود .(2010)استخدام نماذج الاستجابة للمفردة الاختبارية في تدريج مفردات بعض الاختبارات المعرفية، رسالة دكتوراة غير منشورة، جامعة المنيا، مصر.

عبيدات، عمر سليمان علي .(2022)مقارنة طريقتي ألفا وأوميغا في دقة تقدير الثبات في ضوء عدد الأبعاد وطول الاختبار وحجم العينة استنادا لبيانات مولدة .المجلة التربوية -جامعة الكويت -مجلس النشر العلمي، (142)36، 75 - 95

العطيان، إخلاص .(2018)تقديرات المعالم للفقرات ودقتها وفقاً للنموذج اللوجستي ثلاثي المعلمة أثناء معالجة القيم المتطرفة في اختبارات تيمس الدولية TIMSS, 2011.مجلة اتحاد الجامعات العربية للتربية وعلم النفس - جامعه دمشق -كليه التربية، (2)16، .248 – 201

علام، صلاح الدين .(1987)دراسة موازنة ناقدة لنماذج السمات الكامنة والنماذج الكلاسيكية في القياس النفسي والتربوي .المجلة العربية للعلوم النفسية، 7(27). 18-43.

علام، صلاح الدين (2005)نماذج الاستجابة للمفردة الاختبارية أحادية البعد ومتعددة الأبعاد وتطبيقاتها في القياس النفسي والتربوي .القاهرة :دار الفكر العربي.

علام، صلاح الدين .(2011)القياس والتقويم التربوي والنفسي أساسياته وتطبيقاته وتوجهاته المعاصرة .القاهرة :دار الفكر العربي.

عليمات، مهدي محمد فهد .(2022)بناء اختبار محكي المرجع في الرياضيات وفق النموذج اللوجستي ثلاثي المعلمة .مجلة جامعة القدس المفتوحة للأبحاث والدراسات التربوية والنفسية، (38)13، .62 - 46

عيسى، جابر محمد عبد الله .(2014)تأثير عدد بدائل الاستجابة لبنود الاختبار التحصيلي اختيار من متعدد على خصائصه السيكومترية .مجلة التربية - جامعة الأزهر -كلية التربية (157)، 170 – 121

عيلبوني، سمير فؤاد .(2007)أثر عدد البدائل وتمييز الموهات في فقرات الاختيار من متعدد على التوافق مع نموذج ثلاث المعلم .رسالة دكتوراه غير منشورة . جامعة اليرموك.

القضاة، عبد الحميد محمد .(2019)أثر طول الاستبانة وطريقة تطبيقها على دقة تقديرات معالم القدرة والفقرات .رسالة دكتوراه :جامعة اليرموك

القضاة، عبد الحميد محمد .(2020)أثر طول الاستبانة على دقة تقديرات القدرة والخصائص السيكومترية للفقرة والمقياس في ضوء نظرية الاستجابة للفقرة .مجلة جامعة النجاح للأبحاث -العلوم الإنسانية، (6)34، - 953 982.

القضاة، محمد عواد سالم .(2023)تأثير مستوي طول الاختبار في دقة تقدير معلمتي الصعوبة والتمييز باستخدام النموذج ثنائي المعلمة .مجلة كلية التربية -جامعة أسيوط -كلية التربية، (2)39، .243 – 226

مجيد، سوسن شاكر .(2013)أسس بناء الاختبارات والمقاييس النفسية والتربوية . عمان :مركز ديبونو لتعليم التفكير

محمد، شحتة .(1999)تقويم بناء الاختبارات المرجعة إلى محك/إلى المعيار في ضوء نظرية الاستجابة للمفردة .IRTرسالة دكتوراة غير منشورة، القاهرة . جامعة عين شمس.

مناصرة، سوسن عاطف .(2014)خصائص توزيع قدرات الأفراد ومعالم الفقرات للاختبار الوطني لضبط نوعية التعليم للصف الرابع الأساسي وفق نماذج نظرية الاستجابة للفقرة المعلمية واللامعلمية .رسالة دكتوراة غير منشورة .جامعة اليرموك.

نور الدين، أمين .(2003)فاعلية استخدام الاختبار الموائم باستخدام الحاسب في تقدير قدرة الأفراد وتحديد الخصائص السيكومترية للمقياس .رسالة دكتوراة غير منشورة، القاهرة .جامعة عين شمس.

الياسين، محمد .(2022)دقة تقدير معلمة صعوبة الفقرة للنموذج اللوجستي أحادي المعلمة باستخدام طريقة المزاوجة في ضوء تغير حجم العينة وطول الاختبار .المجلة الأردنية للعلوم التطبيقية -سلسلة العلوم الانسانية، 31(2)، 14. – 1

ثانيا :المراجع الأجنبية

Agresti, A., & Finnlay, B. (2009). Statistical Methods for the Social Science (4th Ed). Upper Saddle River, NJ: Person- Prentice Hall.

Al-mehrej, H. N. (1990). The Sensitivity of Item Response Theory Methodology to Differential Instructional Coverages.Unpublished doctoral dissertation, Kansas: Kansas University.

ALKursheh, T. O., Al-zboon, H. S., & AlNasraween, M. E. S. (2022). The Effect of Item Form on Estimating Person's Ability, Item Parameters, and Information Function According to Item Response Theory (IRT). International Journal of Instruction, 15(3), 1111-1130.

Al-zboon, H. S., Alnasraween, M. E. S., & Alkursheh, T. O. (2021). The Effect of the Percentage of Missing Data on Estimating the Standard Error of the Items' Parameters and the Test Information Function According to the Three-Parameter Logistic Model in the Item Response Theory. Ilkogretim Online, 20(1).

Arslan, Y. K., Alkan, A., & Elhan, A. H. (2023). Comparison of the response time-based effort-moderated IRT model and three-parameter logistic model according to computerized adaptive test performances: a simulation study. Communications in Statistics-Simulation and Computation, 1-14.

Baker, F. B.; and Kim, S. (2004). Item Response Theory: Parameter Simulation Techniques. New York: Marcel Dekker, Inc.

Barnes, L.& Wise (1991). The Utility of AModified One-Parameter IRT Model with Small Samples, Journal of Applied Measurement in Education, 4(2),143-153

Baneke, J. J.; and Meijer, R. R. (2004). Analyzing Psychopathology Items: A Case for Nonparametric Item Response Theory Modeling. Psychological Methods, 9(3), 354-366.

Bond TG, Fox CM. (2007). Applying the Rasch Model -Fundamental Measurement in the Human Sciences. Mahwah, NY: Lawrence Erlbaum Associates.

Carvalho, D. J., Costa, M. E. L., & Koide, S. (2022). Assessment of Diffuse Pollution Loads in Peri-Urban Rivers—Analysis of the Accuracy of Estimation Based on Monthly Monitoring Data. Water, 14(15), 2354.

Chen, H. (2012). The Moderating Effects of Item order Arranged by Difficulty on the Relationship between Test Anxiety and Test Performance. Creative Education, 3 (3), 328-333.

Crocker, L. & Algina, J. (1986). Introduction to Classical and Modern Test Theory. New York: Holt, Rinehart and Winston.

Dmitriev, Y. G., & Koshkin, G. M. (2019). INFLUENCE OF USING ADDITIONAL INFORMATION ON THE ACCURACY OF ESTIMATION OF THE RELIABILITY FUNCTION CHARACTERISTICS AT A FINITE SIZE OF OBSERVATIONS. Russian Physics Journal, 62(4), 649-658.

Ebel, R. L. & Frisbie, D. A. (1991). Essentials of Educational Measurement (5th ed.). New Jersey, NJ: Prentice-Hall, Englewood Cliffs.

Ekal, M., & Ventura, R. (2020). On the accuracy of inertial parameter estimation of a free-flying robot while grasping an object. Journal of Intelligent & Robotic Systems, 98(1), 153-163.

Fischer, G. H. (1997). Unidimensional Linear Logistic Rasch Models. In Wim J. van der Linden and Ronald K. Hambleton (Ed.), Handbook of Modern Item Response Theory, (225-245), New York: Springer.

Fitzpatric, A. R. (2009). The Impact of Anchor Test Configuration on Student Proficiency Rates. Educational Measurement: Issues and Practice. 27(4). 34-40 Win 2008

Fitzpatrick,& Ann. R. (2001). The Effects of Test Length and Sample Size on the Reliability and Equating of Tests Composed of Constructed - Response Item. Applied Measurement in Education, 14 (1): 412-425.

Fox, J. P.; and Glas, C. A.. W. (2003). Bayesian Modeling of Measurement Error in Predictor Variables Using Item Response Theory. Pychometrika, 68, 169-191.

Hadwin, A. F.; Winne, Ph. H.; Nesbit, J. C. (2005). Roles for Software Technologies in Advancing Research and Theory in Educational Psychology. British Journal of Educational Psychology, 75(1), 1-24.

Hambleton, R. K.; and Swaminathan, H. (1985). Item Response Theory: Principles and Applications. The Netherlands: Kluwer Nijhoff Publishing.

Hambleton, R. K.; Swaminathan, H.; and Rogers, H. J. (1991). Fundamentals of Item Response Theory. California, Newbury Park: Sage Publications, Inc.

Hambleton, R. (1994). Item Response Theory: A broad Psychometric Frame Work for Measurement Advances. Psicothema, 6 (1), 535-556.

Han, K. T. (2007). WinGen2: Windows Software that Generates IRT Parameters and Item Responses. [computer program], Amherst, MA: Center for Educational Assessment, University of Massachusetts Amherst, Published on line at: http://www.umass.edu/remp/software/wingen: 3/8/2007

Han, K. T.; and Hambleton, R. K. (2007). User's Manual: WinGen Amherst, MA: University of Massachusetts, School of Education. Published on line at: http://www.umass.edu/remp/software/wingen: 3/8/2007.

Harwell, M.; Stone, C. A.; Hsu, T.; Kirisci, L. (1996). Monte Carlo Studies in Item Response Theory. Applied Psychological Measurement, 20(2), 101-125.

Koch, I., Hazeltine, E., Petersen, G., & Weissman, D. H. (2023). Response-repetition costs in task switching do not index a simple response-switch bias: Evidence from manipulating the number of response alternatives. Attention, Perception, & Psychophysics, 1-11.

Lee, Y. (2002). Applications of Isotonic Regression in Item Response Theory. Unpublished doctoral dissertation, Madison: Wisconsin University.

Lee, Y.; Wollack, J. A.; Douglas, J. (2004). On the Use of Nonparametric ICC Estimation Techniques For Checking Parametric Model Fit. Paper presented at the annual conference of the American Educational Research Association, San Diego, CA.

Mohammadloo, J., & Izanlou, B. (2020). Comparing Psychometric Properties of Lower Asymptote Three-Parameter Logistic model (3PL), Upper asymptote 3PLu model and both asymptotes (4PL) based on University Entrance Exam Data. Journal of Modern Psychological Researches, 15(57), 50-66.

Mousavi, A., & Cui, Y. (2020). The effect of person misfit on item arameter estimation and classification accuracy: a imulation study. Education Sciences, 10(11), 324.

Muraki, E. (1992). RESGEN: Item Response Generator. [computer program]. Princeton, NJ: Educational Testing Service.

Muraki, E., and Bock, R. D. (2003). PARSCALE 4: IRT Item Analysis and Test Scoring for Rating-Scale Data. [computer program]. Chicago, IL: Scientific Software.

Shi, D., Siceloff, E. R., Castellanos, R. E., Bridges, R. M., Jiang, Z., Flory, K., & Benson, K. (2021). Revisiting the effect of varying the number of response alternatives in clinical assessment: Evidence from measuring ADHD symptoms. Assessment, 28(5), 1287-1300.

Simms, L. J., Zelazny, K., Williams, T. F., & Bernstein, L. (2019). Does the number of response options matter? Psychometric perspectives using personality questionnaire data. Psychological assessment, 31(4), 557.

Thomas, R. L. (1997). Amonte Carlo Investigation of Parameter Estimation Efficacy Using Modified Fixed "C" Three arameter log(3pl) Item Response Theory Models with Small Sample Sizes. Unpublished doctoral dissertation, Michigan: Wayne State University.

Truskosky, D. M. (1999). An Empirical Examination of Classical Test Theory and Item Response Theory Parameters: Implications for Research and Practice in Small- and Large- Sample Assessments,

Unpublished doctoral dissertation, Illinois: Southern Illinois University at Carbondale.