دقة تقدير قدرة الافراد وفق نموذجي متعدد التدريج (GPCM; GRM) في ضوء نسبة القيم المفقودة: دراسة محاكاة

د. هيثم فؤاد درويش[i]

تاريخ الاستلام
2024/6/15
تاريخ القبول
2024/7/16

الملخص:

هدفت الدراسة الحالية إلى التحقق من دقة تقدير القدرة للأفراد وفق نموذجي متعدد التدريج (GPCM; GRM) في ضوء نسبة القيم المفقودة، ولتحقيق ذلك تم اتباع المنهج التجريبي في إجراء الدراسة، حيث تم استخدام بيانات افتراضية ل (1000) مفحوص و(40) فقرة خماسية التدريج ذات قيم مفقودة بنسب مختلفة (0%، 5%، 10%، 20%) واستخدام نموذجي (GPCM; GRM) متعددة التدريج لتحليل البيانات. وقد بينت النتائج وجود فروق في دقة تقدير القدرة لصالح النموذج المتدرج (GRM) مقابل النموذج المعمم (GPCM) عند نسب فقد أقل أو تساوي (10%)، في حين لم تكن هناك فروق في دقة تقدير القدرة بين النموذجين عند نسب الفقد (20%)، كما أوصت الدراسة باعتماد النموذج المتدرج عند عدم وجود نسب منخفضة من البيانات المفقودة أقل أو تساوي (10%) عندما يكون الهدف تقدير قدرة الأفراد.

الكلمات المفتاحية: دقة التقدير، القدرة، النموذج متدرج الاستجابة، النموذج الجزئي المعمم، نظرية الاستجابة للفقرة.

[i] الأونروا/الأردن

Accuracy of Persons Ability Estimation According to Graded Response Models (GPCM; GRM) in Light of the Percentage of Missing Values: A Simulation Study

Abstract:

The study aimed to verify the accuracy of individuals' ability estimation according to the Graded response models (GPCM; GRM), in light of the percentage of missing values. To achieve this, the experimental approach was followed in conducting the study. Hypothetical data were used for (1000) subjects and (40) five-point scale items with missing values in different percentages (5%, 10%, 20%) and they used Graded response models (GPCM; GRM) to analyze the data. The results showed that there were differences in the accuracy of power estimation in favor of the graded model (GRM) versus the generalized partial graded model (GPCM) at loss rates missing values less than or equal to (10%). At the same time, there were no differences in the accuracy of power estimation between the two models at loss rates (20%). The study also recommended adopting the graded response model when there are no low percentages of missing data less than or equal to (10%) when the goal is to estimate the ability of individuals.

Keywords: Estimation Accuracy, Ability, Graded Response Model, Generalized Partial Model, Item Response Theory.

المقدمة:

تعدّ الاختبارات والمقاييس ذات أهمية قصوى في اتخاذ القرارات على مستوى الطلبة أو المؤسسة التعليمية ككل؛ لذا فإن عملية تحليل النتائج وتفسيرها تعد ركيزة أساسية للحكم على جودة وفاعلية هذه الاختبارات والمقاييس في قياس السمة الكامنة التي أعدت لأجلها، التي تستند في بنائها إلى مبادئ وافتراضات متعددة.

وحتى يكون تفسير درجات الاختبار صحيحًا لا بد أن تتميز هذه الاختبارات بالدقة وصدق القياس (Kubiszyn & Borich, 2024) إذ إن هناك ما يؤثر على خصائص القياس لوجود مشاكل مختلفة، ومن هذه المشاكل التي قد تواجه الباحث أثناء تطبيق هذه الاختبارات هي ظاهرة القيم المفقودة (missing value)، أو الاستجابات المفقودة المتمثلة بترك بعض فقرات الاختبار أو المقياس دون إجابة، فالحالة المثالية هي الحصول على استجابات كاملة على كافة الفقرات من قبل كافة أفراد عينة الدراسة، وهي الحالة التي لا تظهر في أغلب الأحيان، الأمر الذي يقود إلى تقديرات متحيزة وأقل كفاءة (Little & Rubin, 2019).

وقد أشار موهان (Mohan, 2023) إلى أن القيم المفقودة تظهر بسبب عدم الوصول إلى بعض الفقرات، أو عدم الإجابة عنها بسبب عدم الاهتمام من قبل المستجيب، أو عدم قدرته على الإجابة عنها. كما أشار ميكنايت وآخرون (Mcknight et al., 2007) إلى ثلاث فئات لأسباب فقد القيم، وهي الأسباب التي تعود إلى المستجيبين أنفسهم، أو إلى تصميم الاختبار، أو إلى التفاعل ما بين المستجيبين وتصميم الدراسة. وهكذا تتسبب مشكلة فقدان القيم في إثارة العديد من المشاكل لعل أهمها وفقًا لما ذكره كوبيسزين وبوريتش (Kubiszyn & Borich, 2024) احتمال تقليل البيانات المفقودة لقوة نتائج الاختبار الإحصائي (Test Power)؛ نتيجة للخلل الحاصل في حجم العينة، والمشكلة الثانية تكمن في تحيز تقدير معالم القياس، وخاصة عند تقدير الصدق والثبات، كما يضيف الصرايرة (2018، ص485) "إن مشكلة فقدان استجابات الأفراد على عدد من فقرات الاختبار له تأثير متوقع على دقة تقدير معالم الفقرات ومعلمة القدرة".

من هنا جاءت النظرية الحديثة في القياس أو التي تسمى نظرية الاستجابة للفقرة (IRT) (IRT: Item Response Theory) كمكمل وبديل في بعض الاحيان للنظرية التقليدية، حيث قدمت النظرية الحديثة حلول لكثير من المشكلات التي واجهت الباحثين في النظرية التقليدية مثل مشكلة اللاتغاير وبنوك الأسئلة وغيرها من القضايا المهمة كظاهرة القيم المفقودة (Little & Rubin, 2019).

تستند نظرية الاستجابة للفقرة إلى افتراضات قوية تعطي الضوء الأخضر لاستخدامها في التقدير والتدريج والتقييم، وذلك من خلال نماذج مختلفة حيث ظهرت نماذج مختلفة للنظرية الحديثة ابتداء بالنماذج ثنائية التدريج أحادي المعلمة 1PLM: One Parameter Logistic Model، ثنائي المعلمة 2PLM: Tow Parameter Logistic Model، ثلاثي المعلمة 3PLM: Three Parameter Logistic Model ورباعي المعلمة 4PLM: Four Parameter Logistic Model ومن خلال هذه النماذج يتم تقدير المعالم بدقة عالية بالاعتماد على أنماط الاستجابة على فقرات المقياس. ولتفي النظرية الحديثة بمتطلبات القياس الشامل ظهرت نماذج متعددة مثل: نموذج الاستجابة المتدرجة GRM: Graded Response Model، نموذج الاستجابة المتدرجة المعدل MGRM: ModifiedGraded Response Model؛ نموذج التقدير الجزئي المعمم GPCM: General Partial

Credit Model؛ نموذج التقدير الجزئي Santoso) PCM: Partial Credit Model et al., 2022).

بتعدد النماذج ظهر تساؤل مهم وهو أي النماذج يجب استخدامه وبخاصة عندما تتشابه الشروط وتتساوى المعالم المقدرة، فمثلا يتم استخدام النموذجين المتدرج GRM والجزئي المعمم GPCM لنفس الغرض حيث يستخدمان لتقدير نفس المعالم، وهو ما كان محط البحث والدراسة من قبل الكثيرين (Kumar et al., 2021).

وظهور هذه النماذج بالأساس كان للحصول على تقديرات لمعالم الفقرات وقدرات الأفراد لاتخاذ جملة من القرارات المتعلقة بالطلبة أو المؤسسات التعليمية ككل؛ لذا فإن عملية تحليل النتائج وتفسيرها تعد ركيزة أساسية للحكم على جودة وفاعلية هذه الاختبارات والمقاييس في قياس السمة الكامنة التي أعدت لأجلها، والتي تستند في بنائها إلى مبادئ وافتراضات متعددة (Mohan, 2023). وحتى يكون تفسير درجات الاختبار صحيحًا لا بد أن تتميز هذه الاختبارات بالدقة وصدق القياس (Kubiszyn & Borich, 2024).

إلا أن هناك جملة من المشاكل التي قد تحد من دقة تقدير تلك المعالم، ومن هذه المشاكل ظاهرة القيم المفقودة (missing value)، أو الاستجابات المفقودة المتمثلة بترك بعض فقرات الاختبار أو المقياس دون إجابة، فالحالة المثالية هي الحصول على استجابات كاملة على كافة الفقرات من قبل كافة أفراد عينة الدراسة، وهي الحالة التي لا تظهر في أغلب الأحيان، الأمر الذي يقود إلى تقديرات متحيزة وأقل كفاءة (Little & Rubin, 2019). ولمعالجة مشكلة القيم المفقودة ظهرت العديد من الطرق لمعالجة تلك القيم، سواء التي فقدت بشكل عشوائي تام أو التي فقدت بشكل غير عشوائي، كالحذف أو التعويض الصريح أو الضمني.

نماذج الاستجابة للفقرة أحادية البعد متعددة التدريج:

هناك العديد من النماذج أحادية البعد متعددة التدريج المشتقة من نظرية الاستجابة للفقرة ومن هذه النماذج:

أولًا: نموذج الاستجابة المتدرجة (GRM)؛ يعدّ هذا النموذج تعميمًا للنموذج ثنائي المعلمة والذي يستخدم فقرات ذات فئات استجابة متعددة وغير متساوية للفقرات المكونة للمقياس؛ إذ أن هذا النموذج لا يشترط تساوي عدد فئات التدريج ولا ينتج عنها أي تعقيد في تقدير وتفسير معالم الفقرات (Xiong et al., 2020).

يستخدم النموذج المعادلة الرياضية الآتية في تحليل بيانات الدراسة:

$$p_{ix}(\theta) = \frac{\exp[\alpha_i(\theta - ß_{ij})]}{1+\exp[\alpha_i(\theta - ß_{ij})]} \quad \text{.....1}$$

حيث: P_{ix}: احتمالية حصول المفحوص من ذوي القدرة (θ) على علامة أكبر أو تساوي (x) على الفقرة (i)؛ α_i: معلمة تمييز الفقرة (i)، $ß_{ij}$: الحد الفئوي (حدود فئة التدريج المتعددة) للفقرة (i) بين العلامة الفئوية (j-1) وهي تشير إلى مستوى السمة الكامنة اللازمة للاستجابة فوق العتبة (j) باحتمالية تساوي (0.50).

ثانيًا: نموذج الاستجابة المتدرجة المعدل (MGRM)؛ يتم في هذا النموذج تقسيم معالم الفقرات بين عتبات الفئات ($ß_{ij}$) إلى جزأين، هما: معلمة الصعوبة (الموقع) (b_i) لكل فقرة،

ومعلمة عتبة الفئة (c_j) للمقياس كاملًا، وهو بذلك يعد تعديًلا للنموذج المتدرج ويستخدم في حالة تساوي عدد فئات الاستجابة لفقرات المقياس، ويسهل استخدامه في مقاييس الاتجاهات بشكل عام، إذ أن ($\theta_{ij}=b_i-c_j$) باستخدام المعادلة الرياضية الآتية (Kehinde et al., 2022):

$$p_{ix}=\frac{\exp\left[\alpha_i\left(\theta-\left(b_i-c_j\right)\right)\right]}{1+\exp\left[\alpha_i\left(\theta-\left(b_i-c_j\right)\right)\right]} \quad \text{.....2}$$

حيث: P_{ix}: احتمالية حصول المفحوص من ذوي القدرة (θ) على علامة أكبر أو تساوي (x) على الفقرة (i)؛ α_i: معلمة تمييز الفقرة (i)، b_i: معلم الموقع (الصعوبة)، c_j: معالم عتبة الفئة للفقرة.

ثالثًا: نموذج التقدير الجزئي المعمم (GPCM)؛ يختلف هذا النموذج عن النماذج السابقة في توفر خاصية اختلاف معلم الميل للفقرات في المقياس الواحد، وهو يتمثل من خلال المعادلة الرياضية الآتية:

$$P_{IX}(\theta)=\frac{exp\sum_{j=}^{x}\alpha_i(\theta-\delta_{ij})}{\sum_{r=0}^{M}\left[exp\sum_{j=0}^{x}\alpha_i(\theta-\delta_{ij})\right]} \quad \text{.....3}$$

حيث: α_i معلم الميل (تمييز الفقرة)؛ δ_{ij} معلم تقاطع الفئة (صعوبة الخطوة) (Falani et al., 2020).

رابعًا: نموذج التقدير الجزئي (PCM)؛ يعد هذا النموذج تطويرًا لنموذج راش ثنائي التدريج وهو يتيح إمكانية تقدير معالم الفقرات متعددة التدريج، أي التي يمكن أن تأخذ أكثر من قيميتين او فئتين في عملية التصحيح، كأن تكون ثلاثية أو رباعية أو غير ذلك، وبصورة يمكن أن تعكس القدرة التي تقيسها تلك الفقرة، وبالتالي فإن احتمالية حصول المفحوص من ذوي القدرة (θ) على علامة أكبر أو تساوي (x) على الفقرة (i)؛ يتم من خلال المعادلة الرياضية الآتية (Santoso et al., 2022):

$$P_{IX}(\theta)=\frac{exp\sum_{k-0=}^{x}(\theta-bik)}{\sum_{h=0}^{m1}exp\sum_{k=0}^{h}(\theta-bik)} \quad \text{.....4}$$

يعتمد استخدام نماذج التقدير الجزئي على عدد من الافتراضات المتعلقة بالبيانات، لذلك يجب التحقق منها قبل استخدام هذه النتائج، ليتم الحصول على نتائج موثوقة، وهذه الافتراضات هي (Kirmayr et al., 2021):

1. أحادية البعد unidimensionality: ينص هذا الافتراض على ضرورة أن تكون الفقرات المكونة للمقياس تقيس سمة واحدة فقط عند المفحوصين، أي أنها ذات بعد أو قدرة واحدة تفسر أداء الأفراد على المقياس؛ إذ أنه في حال كان هناك أكثر من سمة تفسر أداء المفحوصين على المقياس أو الفقرة تكون النماذج في الوقت نفسه فتسمى النماذج متعددة الأبعاد وليست متعددة التدريج كما يعتقد البعض.

وللتحقق من هذا الافتراض يتم تحليل استجابات المفحوصين أو البيانات المجمعة من خلال استخدام أسلوب التحليل العاملي، ومن ثم ملاحظة قيم الجذور الكامنة ونسب التباين المفسر، والتي تشترط أن يكون العامل الأول يفسر أكثر من (20%) من التباين أو تكون نتيجة قسمة قيمة الجذر الكامن الأول على الثاني أكبر من (2).

2. الاستقلال الموضعي Local independence: يقصد بهذا الافتراض تحرر استجابات المفحوصين من الفقرات وتحرر معالم الفقرات من قدرات المفحوصين (Item Free; Person Free)، بمعنى إن احتمال استجابة المفحوص على فقرة معينة لا يؤثر على احتمال استجابته على فقرة أخرى، وأن تكون تقديرات قدرات الأفراد مستقلة عن معلم صعوبة الفقرة الفئوية.

3. منحنى خصائص الفقرة Item characteristics curve (ICC): تم استخدام المصطلح العام "القدرة" ضمن نظرية استجابة الفقرة للإشارة إلى السمات الكامنة، ولأنه من الصعب توفر مثل هذا المقياس الذي تعبر تدرجاته عن كمية ما يمتلك الفرد من السمة، فإن هذه العملية تتم بافتراض أنه مهما كانت القدرة فإنه يمكن قياسها باستخدام مقياس متوسطه (صفر) ووحدات قياس تساوي كل منها (واحد) وبمدى يتراوح من (-∞) إلى (+∞) فعندئذ يمكن ترجمة القدرة المقيسة بقيم تدرجات المقياس.

4. التحرر من السرعة (speediness): تفترض نظرية استجابة الفقرة أن عامل السرعة (الوقت) لا يؤدي دورًا في الإجابة عن فقرات الاختبار، بمعنى أن فشل الأفراد في الإجابة عن فقرة معينة إجابة صحيحة يرجع إلى تدني قدراتهم وعدم امتلاكهم للسمة وليس إلى تأثير عامل السرعة في الإجابة عن فقرات الاختبار، أي أن قدرة المفحوص أقل من صعوبة الفقرة وهناك طريقة عديدة للتحقق من هذا الافتراض مثل إحصاء المفحوصين الذين لم يكملوا الاختبار.

5. تساوي تمييز الفقرات: يقصد بهذا الافتراض أن تتشكل لكل فقرة مجموعة من منحنيات خصائص فئات الاستجابة لتمثل تعدد الاستجابات المرتبط بتدرجات الاستجابة على الفقرة.

تقدير معالم الفقرات والقدرة وفق نماذج نظرية الاستجابة للفقرة المعلمية:

يمكن تقدير معالم الفقرة في النماذج المعلمية من خلال عدة طرق أهمها (Hambloteton & Swaminathan, 1985):

1. طريقة الأرجحية العظمى المشتركة: يتم في هذه الطريقة تقدير معالم الفقرات وقدرات الأفراد في آن واحد، بالاعتماد على نمذجة احتمالات الاستجابات باستخدام تقديرات مبدئية لمستويات قدرة الأفراد.

2. طريقة الأرجحية العظمى المشروطة: تعمل هذه الطريقة على فصل المعالم الإحصائية للمفحوصين أثناء عملية التدريج؛ إذ تطبق على النموذج اللوجستي أحادي المعلمة حيث يكون اقتران الاحتمالية مشروطًا بعدد من الإجابات الصحية للأفراد على فقرات الاختبار، وهو الأمر الذي يتيح الحصول على تقديرات الأرجحية العظمى الشرطية لمعلمة الصعوبة بغض النظر عن معلمة القدرة.

3. طريقة الأرجحية العظمى الهامشية: تطبق هذه الطريقة على النماذج اللوجستية أحادية البعد ومتعددة الأبعاد بغض النظر عن عدد فقرات الاختبار، حيث يتم في هذه الطريقة إيجاد قيمة (θ) أكبر ما يمكن من خلال معادلة رياضية.

وكما يشير لورد (Lord, 2012) فإن تقدير القدرة يتم من خلال أسلوبين هما:
1. تقديرات الأرجحية العظمى المشتركة: يستخدم هذا الأسلوب عندما تكون معالم الفقرات معلومة بطريقة الأرجحية العظمى الهامشية؛ إذ يتم تقدير معلمة القدرة للأفراد من خلال تعظيم دالة الأرجحية.
2. أسلوب بييز (Bayesian Estimation): يتميز هذا الأسلوب باستخدامه للمعلومات الأولية المتوفرة من خبرات سابقة حول العينة، ويتم استخدامه في حال لم يتم تطبيق طريقة الأرجحية العظمى.

الدراسات السابقة:
فيما يلي عرض للدراسات العربية والأجنبية ذات الصلة بموضوع الدراسة، التي روعي في عرضها أهميتها وتسلسلها الزمني من الأقدم إلى الأحدث تبعًا لسنة النشر.
هدفت دراسة العبدالله (2021) إلى معرفة أثر اختلاف حجم العينة على تدريج مفردات بنك الأسئلة باستخدام نموذج التقدير الجزئي المعمم (GPCM)، بتدريج مفردات اختبار تحصيلي محكي المرجع من تصميم الباحث، وتم تدريج مفردات بنك الأسئلة باستخدام ثلاث عينات مختلفة الحجم (1000؛ 800؛ 400) وفي كل مرة كان يعاد معادلة صورتين لاختبارين (سهل وصعب) للحصول على تقديرات صعوبة المفردات، وفي النهاية تمت المقارنة بين نتائج التعديلات الثلاثة الناتجة عن الحجوم المختلفة، وقد بينت النتائج أن أفضل تدريج للمفردات كان يتوافق مع حجم العينة (1000) حيث كانت نسبة المفردات غير الملائمة (1.5%) وذلك حين انخفضت العينة إلى (800) وهي أقل من النسبة الحاصلة، حين انخفض حجم العينة من (800) إلى (400) حيث ارتفعت نسبة المفردات غير الملائمة إلى (5.8%) من العدد الكلي للمفردات؛ مما يشير إلى ارتفاع جودة التدريج بازدياد حجم العينة، وقد أوصت الدراسة على أن لا تقل حجم العينة المستخدمة عن (1000) وذلك حين استخدام نماذج استجابة للمفردة متعددة التدريج.
وكذلك دراسة القرني وأبو هاشم (2021) التي هدفت التعرف على أثر طريقة معالجة القيم المفقودة على دقة معادلة الاختبارات في نظرية الاستجابة للفقرة، وتكونت عينة الدراسة من 2372 تلميذا من الصفوف العليا بالمرحلة الابتدائية، وكذلك من 3276 من طلاب المرحلة المتوسطة بمدينة جدة. ولتحقيق هدف الدراسة تم استخدام اختبار أوتيس لينون للقدرات بصورتيه (الابتدائي الثنائي والمتوسط) وتتكون كل صورة من 80 فقرة اختبارية تم تطبيقها على عينة استطلاعية بحجم 208 طلبا، وتم اختيار 10 فقرات من كل صورة بواقع 20 فقرة تمثل الجذع المشترك. ومن ثم طبقت الأدوات على العينة وتم الحصول على بيانات كانت نسبة القيم المفقودة 5.5% للمرحلة الابتدائية و(5.6%) للمرحلة المتوسطة. وتمت معالجة البيانات للقيم المفقودة من خلال ثلاث طرق للمعالجة وهي خوارزمية تعظيم التوقعات (EM)، وحساب قيم تعويضية متعددة (MI)، وحساب المعالجة بطريقة دالة الاستجابة للفقرة (IRFI). وتم استخدام برنامج (X-calibre) في التقدير وكذلك للتحقق من مطابقة البيانات للنموذج. وللكشف عن دقة المعادلة العمودية في نظرية الاستجابة للفقرة باختلاف طريقة المعالجة للقيم المفقودة وطريقة التقدير تم استخدام محك الخطأ المعياري. وقد أظهرت النتائج أن أفضل طريقة معالجة للقيم المفقود كانت طريقة المعالجة بدالة الاستجابة للفقرة (IRFI) وذلك عندما نستخدم الطريقة الافتراضية للتقدير في البرنامج وهي طريقة (EAP).

كما هدفت دراسة فلاني وآخرين (Falani et al., 2020) إلى مقارنة دقة تقدير القدرة في النماذج المتدرجة (GRM) ونموذج (GPCM) ل (1625) طالبا من طلبة المدارس الإعدادية على اختبار للرياضيات، وقد بينت النتائج أن متوسط الخطأ التربيعي (RMSE) كان أقل عند تحليل بيانات الطلبة باتباع نموذج الاستجابة المتدرجة (GRM) مقارنة بنموذج التقدير الجزئي العام (GPCM)، وبهذا توصلت الدراسة إلى أن استخدام نموذج الاستجابة المتدرجة أكثر دقة في تقدير قدرات المفحوصين بشكل عام.

وهدفت دراسة الشريفين (2019) إلى الكشف عن أثر نموذج نظرية الاستجابة للفقرة ذات الاستجابة المتعددة التدريج المستخدم في التحليل وهي: نموذج الاستجابة الاسمي لبوك، ونموذج التقدير الجزئي العام لموراكي، ونموذج الاستجابة المتدرج لساميجما على دقة تقدير القدرات للأفراد، والخصائص السيكومترية للفقرات، والاختبار. ولتحقيق هدف الدراسة تم استخدام برنامج (WinGen 3) في توليد بيانات خاصة باختبار مكون من 50 فقرة وفق نموذج التقدير الجزئي العام، وكل فقرة مكونة من خمسة مستويات، يتم الانتقال بينها عبر أربع عتبات فارقة. أشارت نتائج الدراسة إلى أن نموذج التقدير الجزئي العام هو أكثر دقة في تقدير القدرات من نموذجي الاستجابة: الاسمي والمتدرج. وأشارت النتائج إلى أن نموذج الاستجابة الاسمي هو نموذج أكثر دقة في تقدير معالم التمييز للفقرات مقارنة بالنموذجين الآخرين، كما بينت النتائج أن نموذج الاستجابة المتدرج هو أكثر دقة في تقدير معالم الصعوبة للعتبات الفارقة من النموذجين الآخرين. وأنه قدم أقصى كمية من المعلومات التي يقدمها الاختبار، وأكثر دقة في تحليل البيانات مقارنة بالنموذجين الآخرين، كما كشفت النتائج عن عدم وجود فروق عند مستوي الدلالة (α=0.05) بين معاملات ثبات كرونباخ ألفا تبعا لنموذج الاستجابة المستخدم.

كما هدفت دراسة عبد الوهاب (2019) إلى معرفة أثر طرق تقدير الدرجات على مفردات الصواب والخطأ المتعدد، ونوع نموذج الاستجابة للمفردة المستخدمة (التقدير الجزئي والاستجابات المتدرجة) والتفاعل بينهما على دقة تقدير بارامترات الأفراد والمفردات، وفي سبيل ذلك تم بناء اختبار تحصيلي في مقرر علم النفس التربوي لطلاب كلية التربية بجامعة أم القرى (411) طالبًا، بحيث يتكون من (35) مفردة من نوع الصواب والخطأ المتعدد، وتم تقدير الأخطاء المعيارية في تقدير قدرات الطلاب على هذا الاختبار وبارامترات الصعوبة وفق أربع طرق مختلفة لتقدير الدرجات وباستخدام نموذجي التقدير الجزئي والاستجابات المتدرجة، كما تم تقدير الأخطاء المعيارية في تقدير بارامترات تمييز المفردات باستخدام نموذج الاستجابات المتدرجة، وقد أسفرت النتائج عن وجود أثر لكل من طريقة تقدير الدرجات ونوع النموذج المستخدم والتفاعل بينهما على دقة تقدير بارامترات صعوبة المفردات وبارامترات القدرة للطلاب، كما وجد أثر لطريقة تقدير الدرجات على دقة تقدير بارامترات تمييز المفردات وفق نموذج الاستجابات المتدرجة، وقد أوصى البحث بأفضلية الاعتماد على الطريقة التي تراعي المعرفة الجزئية بأكبر قدر ممكن في تقدير الدرجات على المفردات، وكذلك أفضلية استخدام نموذج الاستجابات المتدرجة في تدريج هذه المفردات.

كما هدفت دراسة أباديو وبستري (Abadyo & Bastari, 2015) إلى تحديد النموذج الأكثر دقة في تقدير قدرات الأفراد باختلاف طول الاختبار وحجم العينة، ولتحقيق ذلك تم توليد بيانات افتراضية مبنية على التصميم العاملي للتأثيرات الثابتة (2*3*3*3*5) مما أنتج (270) مجموعة من البيانات، وبالاعتماد على قيمة مربع متوسط الجذر (RAMSE) للقدرة، أظهرت النتائج أن تحليل البيانات باستخدام نموذج (GPCM) يوفر تقديرًا دقيقًا أعلى من نموذج

(GRM)، بالإضافة إلى أن معلومات الاختبار الناتجة من نموذج (GPCM) أعلى بثلاث مرات من المعلومات الناتجة من نموذج (GRM).
وهدفت دراسة هيبة (2013) إلى التحقق من تأثير طريقة معالجة البيانات المفقودة على الخصائص السيكومترية للمقاييس ذات الاستجابات المتعددة، وقد اعتمد الباحث على الطريقة الحديثة في معالجة البيانات المفقودة وهي التعويض المتعدد للبيانات المفقودة (Multiple Imputation)، وطريقة تعظيم التوقعات (Expectation Maximization)، وطريقة الانحدار (Regression Method)، وقد تكونت عينة الدراسة من 231 طالباً من طلاب الفرقة الثالثة والرابعة بكلية التربية جامعة عين شمس، واعتمد الباحث في توليد بيانات متكافئة في دراسة المحاكاة على برنامج (IRT-Lab) بأحجام عينات مختلفة 50، 100، 200 وبتوزيع بيانات مفقودة بنسب 10%، 20%، 40% على جميع الملفات المولدة البالغ عددها 90 عينة، وتوصلت نتائج البحث إلى أنه لا يوجد تأثير للطرق الثلاث على قيم ثبات ألفا، ولكن ظهرت أفضلية طفيفة جدا لطريقة تعظيم التوقعات عند نسبة فقد للبيانات 40%، وعند التحقق من صدق البنية، أظهرت النتائج عدم اختلاف الطرق الثلاث في تأثيرها على صدق البنية عند نسبة فقد 10% في جميع أحجام العينات، لكن في حالتي نسب الفقد 20%، 40% ظهرت أفضلية واضحة لطريقة تحليل الانحدار في عدم تأثيرها على صدق البنية.

التعقيب على الدراسات السابقة:
من خلال الاطلاع على الدراسات السابقة يتضح لنا ان حجم العينة ذو تأثير دال إحصائيًا في دقة تقدير معلمة القدرة للأفراد عند استخدام النماذج متعددة التدريج؛ إذ بينت النتائج أن هناك تباينا في دقة تقدير القدرة باختلاف النموذج المستخدم، وهو الأمر الذي يؤكد على أن أي فقد في البيانات من شأنه أن يقود إلى تأثير في دقة تقدير القدرة للأفراد، فعدم مطابقة الأفراد للنموذج المستخدم سيؤدي إلى تدني حجم العينة وهو أمر مشابه نوعًا ما لوجود بيانات مفقودة. وعلى ذلك فإن الدراسة الحالية تميزت عن باقي الدراسات في أنها عملت على تقصي دقة تقدير قدرة الأفراد وفق نموذجي متعدد التدريج (GPCM; GRM) في ضوء نسبة القيم المفقودة المختلفة باستخدام بينات افتراضية.

مشكلة الدراسة وسؤالها:
تعد الاختبارات والمقاييس إحدى أهم الوسائل المستخدمة لتقييم الأداء وقياس المهارات، والمعرفة، والقدرات والكفاءة، أو الاستعداد في موضوع معين، وهو الأمر الذي يستوجب أن تكون على درجة عالية من الدقة والموضوعية والثبات، وفي ضوء أن هذه الأدوات تختلف في تدرجاتها فقد ظهرت العديد من النماذج المعلمية متعددة التدريج، التي يمكن من خلالها تقدير قدرات الأفراد ومعالم الفقرات، إلا أن هذه الأدوات قد تتعرض لجملة من المشاكل بعد تطبيقها على المفحوصين كعدم مطابقة البيانات للنموذج المفترض أو وجود فقد في البيانات لسبب أو لآخر؛ إذ إنّ نظرية الاستجابة للفقرة أفرزت عددا من النماذج متعددة التدريج التي تواكب جملة من الافتراضات كأحادية البعد والاستقلال الموضعي وتواكب خصائص الفقرات مثل تساوي تمييز الفقرات، حيث ظهرت بعض النماذج التي تتشابه في متطلباتها، مثل تساوي التمييز وانعدام التخمين كنموذجي الاستجابة المتدرجة (GRM) والاستجابة المتدرجة المعمم (GPCM)، لكنها تختلف في مبدأ عملها مما دعا كثيرا من الباحثين للمقارنة بينها في ضوء ظروف تجريبية مختلفة، فقد خلصت العديد من الدراسات كدراسة العبدالله (2021) إلى أن

حجم العينة المناسب لاستخدام النماذج متعددة التدريج يجب أن لا يقل عن (1000) مفحوص لضمان مطابقة المفحوصين للنموذج المقترح، كما أن وجود نسب مختلفة من البيانات المفقودة قد يستدعي في بعض الأحيان استخدام طريقة الحذف، مما يخل في شرط حجم العينة، ولذلك تم اللجوء إلى طريقة التعويض عن القيم المفقودة كبديل لطريقة الحذف، إلا أنه مع تعويض القيم المفقودة فإن نسبة قد تخل في مدى فاعلية النموذج في تقدير المعالم وبخاصة معلمة القدرة للأفراد (زريقات، 2024).

من هنا تبرز مشكلة الدراسة في التحقق من دقة تقدير القدرة للأفراد وفق نموذجي متعدد التدريج (GPCM; GRM) في ضوء نسبة القيم المفقودة، وتحديد النموذج الأكثر منعة في حال كانت هناك قيم مفقودة.

سؤال الدراسة:

هل تختلف دقة تقدير قدرة الأفراد باختلاف نسبة القيم المفقودة (0%؛ 5%؛ 10؛ 20%) وفق نموذجي متعدد التدريج (GRM, GPCM)؟.

أهمية الدراسة:

تظهر أهمية هذه الدراسة أولًا: في الجانب النظري من حيث إنها تقارن بين نموذجين من نماذج الاستجابة المتدرجة وهما نموذج الاستجابة المتدرجة (GRM)؛ ونموذج التقدير الجزئي المعمم (GPCM) في ضوء نسب الفقد في البيانات (0%؛ 5%، 10%، 20%)، كما أنها تزود الأدب المتعلق بنظرية الاستجابة للفقرة بمؤشرات حول صلاحية النماذج اللوجستية متعددة التدريج ومدى الوثوق بها في ضوء نسب الفقد من حيث مطابقة الأفراد للنموذج المستخدم، وثانيهما: الجانب العملي حيث إن هذه الدراسة تزود الباحثين ومطبقي الاختبارات بمؤشرات لاختيار النموذج اللوجستي متعدد التدريج الأكثر ملاءمة لتحليل بياناتهم عند وجود نسب مختلفة من القيم المفقودة.

أهداف الدراسة:

تهدف هذه الدراسة إلى التحقق من دقة تقدير قدرة الأفراد باختلاف نسبة القيم المفقودة (0%؛ 5%، 10%، 20%)، وفق النماذج متعددة التدريج (GPCM, GRM).

المصطلحات المفاهيمية والإجرائية:

دقة التقدير (Accuracy of Estimation): تعبير يشير إلى جودة التقدير. التي تميزها الأرجحية العظمى في أن التقدير قريب من القيمة الحقيقية، حيث يمكن الوصول. إلى ذلك باختيار مقدر غير متحيز (unbiased estimator)، الذي يتصف تباينه بأنه أقل تباين. من أي مقدر آخر غير متحيز، وذلك باستخدام الجذر التربيعي لمتوسط مربعات الأخطاء (RMSE) ((Baker, 2001).

القيم المفقودة (Missing-value): قيم ناتجة عن عدم استجابة المفحوص. عن بعض فقرات الاختبار، وترك هذه الفقرات دون إجابة، بصرف النظر عن سبب ذلك (Graham, 2009).

نظرية الاستجابة للفقرة (IRT): هي إحدى النظريات في مجال القياس والتقويم التي تفترض. أن هناك سمات كامنة معينة يشترك. فيها مجموعة من الأفراد، ولكنهم يختلفون في مقدارها، وبالرغم من أن هذه السمات غير ملاحظة، إلا أنه يمكن الاستدلال عليها من السلوك. الملاحظ للفرد المتمثل في استجاباته على فقرات الاختبار، حيث تعتمد هذه النظرية. على نماذج رياضية تستخدم في عمليات. التقدير لكل من المعالم التي تصف الفقرة، والقدرات الكامنة التي تصف الأفراد (علام، 2005).
نموذج الاستجابة المتدرجة: أحد نماذج نظرية الاستجابة للفقرة متعددة التدريج والذي يستخدم فيه تدريج ليكرت (الثلاثي؛ الرباعي؛ الخماسي .....) (Abadyo & Bastari, 2015).

نماذج التدريج: هي تلك النماذج التي تمثل علاقة غير خطية بين مستوى قدرة الفرد واحتمال استجابته في مستوى معين من مستويات التدريج وهو ما يشار اليه بالعتبة Threshold وهي التي تمثل مستوى السمة اللازمة لتخطي استجابة الفرد للعتبة الفارقة باحتمال قدره 0.5، وعدد مستويات تدريج المقياس تساوي عدد العتبات مضافا إليها واحد (Kehinde et al., 2022).

محددات الدراسة:
تتحدد نتائج هذه الدراسة بالمحددات التالية:
-تقييم فاعلية النماذج اللوجستية متعددة التدريج فقط.
-محاكاة عينة افتراضية بطول (1000) مفحوص.
-محاكاة اختبار افتراضي بطول (40) فقرة.

الطريقة والإجراءات
منهجية الدراسة: تم اتباع المنهج التجريبي في إجراء البحث، حيث تم استخدام بيانات افتراضية تحاكي بيانات خماسية التدريج ذات قيم مفقودة بنسب مختلفة (0%؛ 5%، 10%، 20%) واستخدام نموذجي (GPCM; GRM) متعددة التدريج لتحليل البيانات.

عينة الدراسة:
تكونت عينة الدراسة من (1000) مفحوص افتراضي تم توليدهم باستخدام برمجية (WinGen) بحيث تتبع قدراتهم التوزيع الطبيعي بمتوسط حسابي (0) وانحراف معياري (1).

أداة الدراسة:
تكونت أداة الدراسة من اختبار افتراضي يتكون من (40) فقرة خماسية التدريج، تتبع معلمة الصعوبة التوزيع الطبيعي بوسط حسابي (0) وانحراف معياري (1)، وتتبع معلمة التمييز التوزيع المنتظم من (0.5–2).

إجراءات الدراسة:
للإجابة عن سؤال الدراسة تم اتباع الإجراءات التالية:

1. توليد عينة افتراضية من المستجيبين حجمها (1000) فرد تتبع قدرتهم التوزيع الطبيعي بمتوسط حسابي (0) وانحراف معياري (1)، باستخدام برمجية (WinGen).
2. توليد اختبار افتراضي بطول (40) فقرة خماسية التدريج بحيث تتبع معلمة الصعوبة التوزيع الطبيعي المعياري (0، 1) باستخدام برمجية (WinGen) وتتبع معلمة التمييز التوزيع المنتظم من (0.5–2).
3. توليد استجابات للأفراد الافتراضيين في العينة المولدة في الخطوة (1) على الاختبار الافتراضي المولدة معالم فقراته في الخطوة (2)، لتنتج مصفوفة الاستجابات الافتراضية، باستخدام برمجية (WinGen).
4. إجراء فقد عشوائي بنسب مختلفة (0%، 5%، 10%، 20%) من مصفوفة الاستجابات الافتراضية الأصلية.
5. إجراء تعويض للنسب المفقودة بطريقة الوسط الحسابي في كل من مصفوفات الاستجابة المتضمنة قيم مفقودة. لتتكون 4 مصفوفات للاستجابات المكتملة سميت حسب نسب الفقد (0%، 5%، 10%، 20%).
6. التحقق من افتراض أحادية البعد من خلال التحليل العاملي الاستكشافي باستخدام برمجية (SPSS).
7. التحقق من افتراض الاستقلال الموضعي من خلال البواقي باستخدام برمجية R.
8. التحقق من افتراض منحنى خصائص الفقرة باستخدام برمجية R.
9. تحليل كل من مصفوفات استجابات الأفراد وفقا لكل من النماذج متعددة التدريج (GPCM; GRM).
10. تقدير قدرة الأفراد واستخراج مؤشر دقة التقدير المتمثل في الخطأ المعياري (SE) في هذه الدراسة.
11. استخدام برمجية SPSS للتحقق من أثر نسبة القيم المفقودة على دقة تقدير الأفراد في ضوء النموذجين متعددي التدريج الجزئي المعمم (GPCM) والمتدرج الاستجابة (GRM).

التحقق من افتراضات نظرية الاستجابة للفقرة كالاتي:

أولًا: التحقق من افتراض أحادية البعد (Unidimensional)

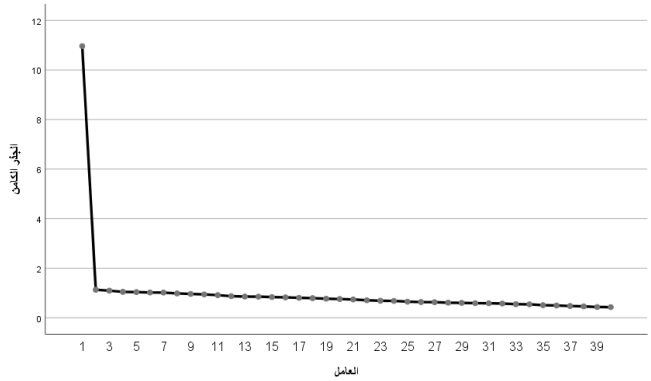
تم التحقق من هذا الافتراض من خلال إجراء اختبار التحليل العاملي الاستكشافي بالاعتماد على تحليل المكونات الرئيسية (Principal component analysis)، لإيجاد قيم الجذور الكامنة (Eigen value)، ونسب التباين المفسر لكل من العاملين الأول والثاني لمصفوفة الاستجابات الاصلية كما يأتي:

جدول 1: التحليل العاملي الاستكشافي للتحقق من أحادية البعد.

العامل	الجذر الكامن	الجذر الكامن الأول / الجذر الكامن الثاني	نسبة التباين المفسر
العامل الأول	10.962	9.692	27.404
العامل الثاني	1.131		2.828

تؤكد البيانات الواردة في جدول 1 الناتجة من عملية التحليل العاملي الاستكشافي، أن الاستجابات تحقق افتراض أحادية البعد، حيث إن نسبة الجذر الكامن للعامل الأول إلى نسبة الجذر الكامن للعامل الثاني قد بلغت (9.692)، وهي أكبر من (2) كما أن نسبة التباين المفسر من خلال العامل الأول أكبر من (20%)، وهي مؤشرات تؤكد على تحقق افتراض أحادية البعد (Hatti, 1985)، وهذا ما يمكن ملاحظته أيضًا من خلال التمثيل البياني للجذور الكامنة (Scree Plot) الآتي:

شكل 1: التمثيل البياني للجذور الكامنة (Scree Plot)



يظهر لنا من شكل 1 والخاص بالتمثيل البياني للجذور الكامنة (Scree Plot) للمقياس الذي تم توليده، أن هناك وجود انعطاف حاد عند وجود عاملين وهو الأمر الذي يدل على وجود سمة واحدة فقط سائدة في المقياس.

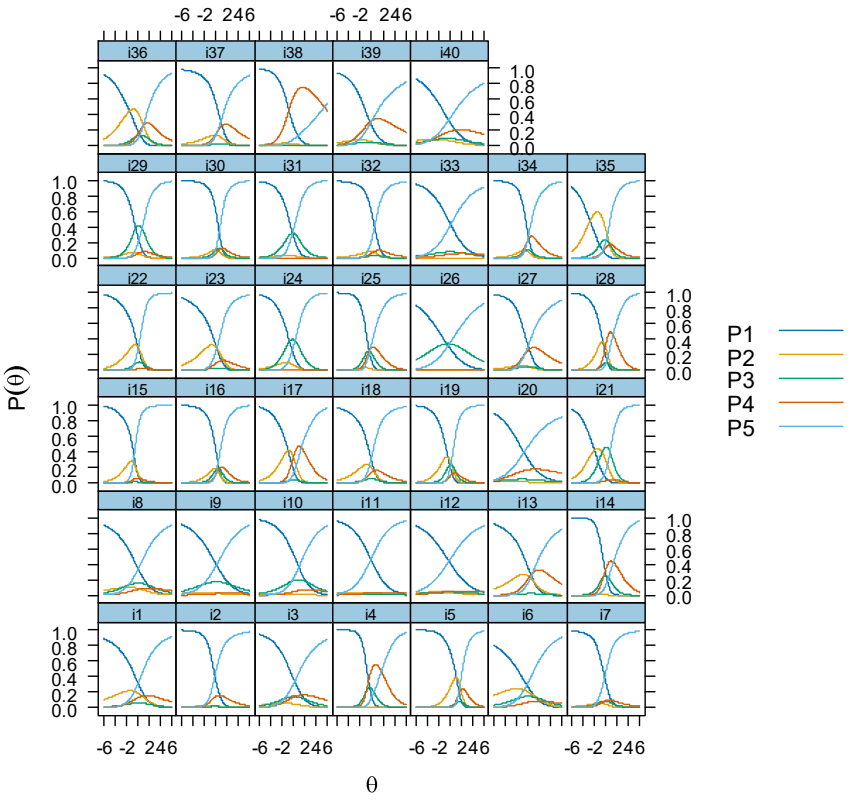
ثانيًا: التحقق من افتراض الاستقلال الموضعي للمقياس

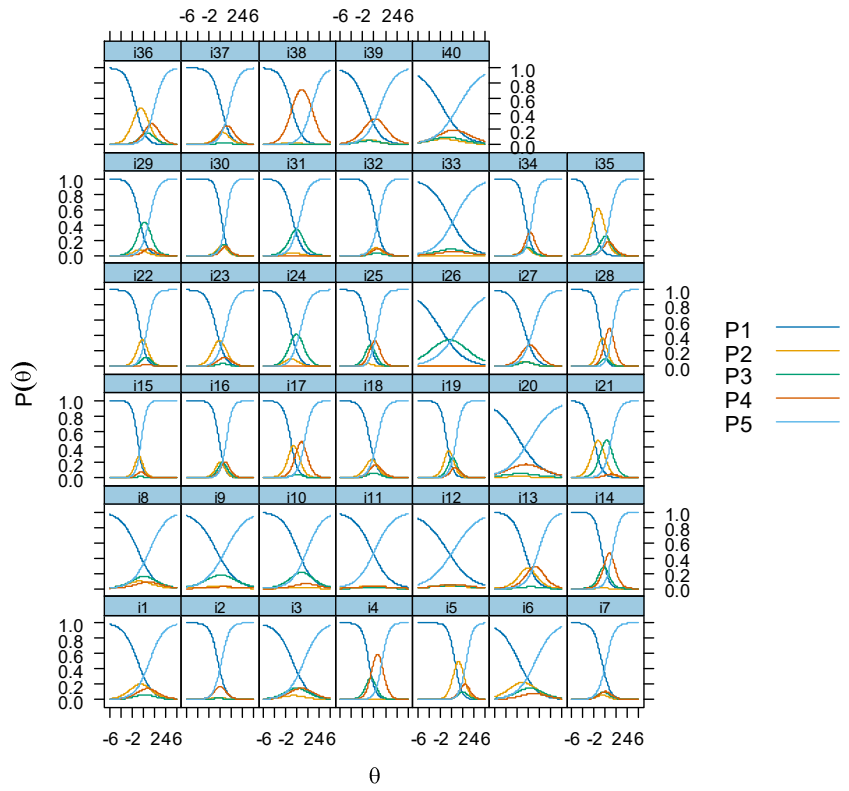
تم التحقق من هذا الافتراض بالاعتماد على القيم المعيارية للبواقي (Z_{Q3})، حيث تم الكشف عن أزاوج الفقرات المرتبطة من خلال قيم معاملات ارتباط البواقي لها، وتبين أن عدد الأزواج الكلي يساوي (780) زوجا، منها (53، 66) زوجا غير مستقل عند التحليل وفقا للنماذج متعددة التدريج (GPCM، GRM) على الترتيب بنسبة مئوية بلغت (7%، 8%) من مجمل الأزواج الكلية، وهذا يبين أن عدد أزاوج الفقرات التي حققت افتراض الاستقلال الموضعي أعلى من أربعة أمثال عدد الأزواج التي حققت التبعية الموضعية. وهذا مؤشر على تحقق افتراض الاستقلال الموضعي (Kim et al., 2005).

ثالثًا: التحقق من افتراض منحنى خصائص الفقرة

تم التحقق من افتراض منحنى خصائص الفقرة من خلال رسم منحنى خصائص الفقرة باستخدام برمجية R عند تحليل البيانات وفقا للنماذج متعددة التدريج (GPCM، GRM) كما هو مبين في شكل 2.

شكل 2: منحنى خصائص الفقرة وفقا للنماذج متعددة التدريج.

النموذج	منحنى خصائص الفقرة
GPCM	



من شكل 1 يتبين أن الافتراض متحقق حيث يشير منحنى خصائص الفقرة الى علاقة مضطرده غير خطية بين القدرة واحتمالية الاستجابة على كل من منحنيات الفقرة الفرعية.

نتائج الدراسة ومناقشتها:

النتائج المتعلقة بسؤال الدراسة: **"هل تختلف دقة تقدير قدرة الأفراد باختلاف نسبة القيم المفقودة (5%، 10%، 20%) وفق نموذجي متعدد التدريج (GPCM، GRM)؟"**

للإجابة عن هذا السؤال تم تحليل مصفوفة الاستجابات الافتراضية باستخدام برمجية R وفقا للنماذج متعددة التدريج الجزئي المعمم (GPCM) والمتدرج الاستجابة (GRM) وبعد فحص مطابقة الفقرات للنموذج تبين أن جميع الفقرات مطابقة لكل من النموذجين المتدرج والجزئي المعمم، وتم تقدير قدرة الأفراد لكل من مصفوفات الاستجابة الأربع ذات نسب الفقد المختلفة، وتم استخراج الخطأ المعياري في تقدير القدرة واستخراج الوسط الحسابي للخطأ المعياري حسب نسبة الفقد ولكل من النماذج المتعددة كما هو في جدول 2.

جدول 2: الأوساط الحسابية والانحرافات المعيارية لدقة تقدير القدرة وفقا للنماذج متعددة التدريج باختلاف نسبة الفقد.

		النموذج				الكلي	
		GPCM		GRM		الكلي	
		الوسط الحسابي	الانحراف المعياري	الوسط الحسابي	الانحراف المعياري	الوسط الحسابي	الانحراف المعياري
نسبة الفقد	0%	.2506	.05048	.2444	.03303	.2475	.04276
	5%	.2606	.04263	.2556	.02757	.2581	.03598
	10%	.2706	.03530	.2675	.02309	.2690	.02986
	20%	.2868	.02491	.2874	.01753	.2871	.02153
	الكلي	.2672	.04166	.2637	.03045	.2654	.03653

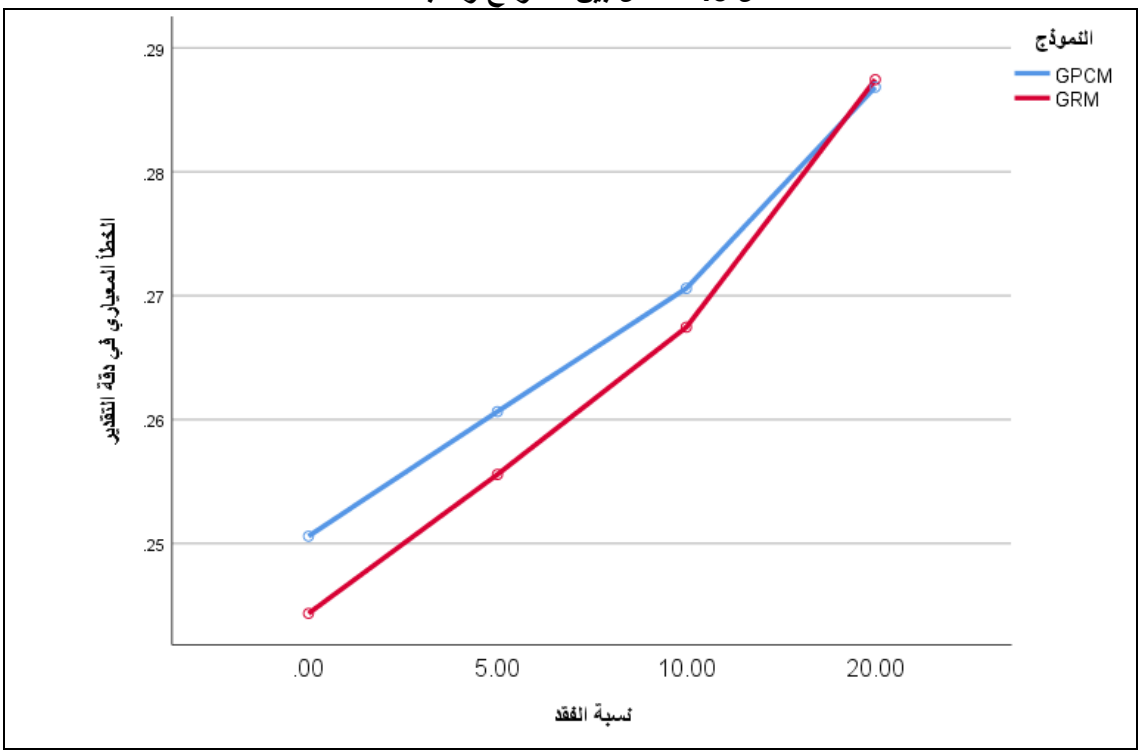
يلاحظ من جدول 2 وجود اختلاف في الأوساط الحسابية لدقة تقدير معلمة القدرة باختلاف النموذج المستخدم حيث يلاحظ أن الوسط الحسابي للخطأ المعياري في تقدير القدرة بلغ (0.2637) وفقا للنموذج المتدرج (GRM) و(0.2672) وفقا للنموذج الجزئي المعمم (GPCM). كما يلاحظ اختلاف الوسط الحسابي للخطأ المعياري في تقدير القدرة باختلاف نسبة القيم المفقودة. ولفحص دلالة الاختلافات باختلاف نسبة الفقد والنموذج المستخدم تم إجراء تحليل تباين ثنائي كما هو مبين في جدول 3.

جدول 3: تحليل التباين في دقة تقدير القدرة العائد للنموذج ونسبة الفقد.

مصدر التباين	مجموع المربعات	درجة الحرية	متوسط المربعات	قيمةF	القيمة الاحتمالية
النموذج	.024	1	.024	21.460	.000
نسبة الفقد	1.720	3	.573	514.083	.000
التفاعل	.013	3	.004	4.025	.007
الخطأ	8.915	7992	.001		
الكلي	10.673	7999			

من نتائج تحليل التباين في جدول 3 يتبين وجود أثر دال إحصائيا للتفاعل بين نسبة الفقد والنموذج على دقة تقدير معالم القدرة حيث لم تتجاوز القيمة الاحتمالية مستوى الدلالة الإحصائية (0.05). كما يتضح ذلك من خلال التمثيل البياني في شكل 3.

شكل 3: التفاعل بين النموذج ونسبة الفقد



يلاحظ أن الخطأ المعياري يختلف بين النموذجين عند عدم وجود فقد في البيانات وقد بدأت الفروق بالتلاشي مع زيادة نسبة الفقد ولفحص دلالة الفروق تم اجراء تحليل لفحص التأثيرات البسيطة كما هو مبين في جدول 4.

جدول 4: تحليل التباين في دقة تقدير القدرة العائد للنموذج في ضوء نسبة الفقد.

نسبة الفقد	مصدر التباين	مجموع المربعات	درجة الحرية	متوسط المربعات	قيمةF	القيمة الاحتمالية
%0	النموذج	.019	1	.019	10.679	.001
	الخطأ	3.636	1998	.002		
	الكلي	3.655	1999			
%5	النموذج	.013	1	.013	9.990	.002
	الخطأ	2.575	1998	.001		
	الكلي	2.587	1999			
%10	النموذج	.005	1	.005	5.528	.019
	الخطأ	1.778	1998	.001		
	الكلي	1.783	1999			
%20	النموذج	.000	1	.000	.396	.529
	الخطأ	.927	1998	.000		
	الكلي	.927	1999			

يلاحظ من جدول 4 أن الفروق في دقة تقدير القدرة العائدة للنموذج تكون دالة إحصائيا عند نسب فقد منخفضة أقل من 20% وتكون الفروق أكثر وضوحا عندما تكون نسب الفقد أقل. وتكون الفروق في الخطأ المعياري لصالح النموذج الجزئي المعمم (GPCM) مقابل النموذج المتدرج (GRM) وهذا يعني أن النموذج المتدرج أكثر دقة في تقدير قدرة الأفراد من النموذج الجزئي المعمم حيث إن الخطأ المعياري مؤشر عكسي للدقة. ويمكن تفسير ذلك من خلال ما تحدثه القيم المفقودة من خلل في نمط الاستجابة؛ مما يقلل من دقة التقدير حيث إن القيم التعويضية بدل القيم المفقودة تعتمد على مجموعة استجابات الأفراد على الفقرة وليس على نمط استجابة الفرد على الفقرات (المقياس) وهو ما كان واضحا في التمثيل البياني لدقة التقدير لكل من النموذجين بزيادة نسب الفقد. ويمكن تفسير انخفاض الفروق بين النموذجين في دقة التقدير بزيادة نسبة الفقد من خلال اعتماد كل من النموذجين على نمط الاستجابة وتسلسلها أو تدرجها بشكل مختلف فالنموذج المتدرج يعتمد على تدرج استجابة الفرد بتدرج صعوبة الفقرات وبزيادة نسب الفقد حصل انتهاك لتسلسل وتدرج الاستجابة؛ مما خفض من جودة مطابقة النموذج للبيانات، الذي بدوره قلل من دقة تقدير القدرة، لكن النموذج الجزئي المعمم تأثر بشكل أكبر حيث إن أي اختلاف جزئي في تسلسل النمط يعد انتهاكا بالنسبة لهذا النموذج ما أحدث عددا أكبر من الانتهاكات بالنسبة للنموذج الجزئي المعمم وخفض بدرجة أكبر من مطابقة النموذج للبيانات ودرجة أقل من دقة تقدير القدرة.

وقد جاءت هذه النتيجة متفقة مع نتائج دراسة عبد الوهاب (2019) التي بينت وجود أثر لكل من طريقة تقدير الدرجات ونوع النموذج المستخدم (GRM; GPCM) والتفاعل بينهما في دقة تقدير بارامترات صعوبة المفردات وبارامترات القدرة للطلاب.

بينما اختلفت نتيجة هذه الدراسة مع نتائج دراسة الشريفين (2019) التي أشارت إلى أن نموذج التقدير الجزئي العام هو أكثر دقة في تقدير القدرات من نموذجي الاستجابة: الاسمي والمتدرج. ونتائج دراسة أباديو وبستري (Abadyo & Bastari, 2015) التي أظهرت أن تحليل البيانات باستخدام نموذج (GPCM) يوفر تقديرًا دقيقًا أعلى من نموذج (GRM).

التوصيات:

وفي ضوء النتائج التي تم التوصل لها فان الباحث يوصي بما يلي:

- اعتماد النموذج المتدرج عند عدم وجود نسب منخفضة من البيانات المفقودة أقل أو تساوي (10%) عندما يكون الهدف تقدير قدرة الأفراد.
- إجراء دراسة للمقارنة بين النماذج متعددة الاستجابة في ضوء حجم العينة وطول الاختبار.
- إجراء مقارنة بين النماذج متعددة الاستجابة في تقدير معالم الفقرات.

المصادر والمراجع:

زريقات، هيثم (2024). أثر طرق معالجة القيم المفقودة باختلاف نسبها على دقة التقدير لمعالم الفقرات والأفراد تبعًا للنماذج المعلمية واللامعلمية. أطروحة دكتوراه غير منشورة، كلية العلوم التربوية، جامعة اليرموك، الأردن.

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

عبد الوهاب، محمد (2019). أثر طريقة تقدير الدرجات ونموذج الاستجابة للمفردة في دقة تقدير بارامترات الأفراد والمفردات لاختبار صواب وخطأ متعدد. المجلة الدولية للأبحاث الدولية، 43(2)، 59-85.

العبدالله، زياد (2021). أثر حجم العينة على تقدير مفردات بنك الأسئلة. مجلة جامعة بابل للعلوم الإنسانية، 29(10)، 94-114.

علام، صلاح الدين (2005). نماذج الاستجابة للفقرة الاختبارية أحادية البعد ومتعددة الأبعاد وتطبيقاتها في القياس النفسي والتربوي. دار الفكر العربي للنشر والتوزيع.
القرني، أحمد وأبو هاشم، السيد (2021). أثر طريقة معالجة القيم المفقودة على دقة معادلة الاختبارات في نظرية الاستجابة للفقرة. دراسات-العلوم التربوية، 29(2)، 455-491.

هيبة، محمد أحمد (2013). تأثير طريقة معالجة البيانات المفقودة على الخصائص السيكومترية للمقاييس ذات الاستجابات المتعددة (دراسة إمبيريقية ومحاكاة). مجلة جامعة عين شمس للقياس والتقويم، 3(5)، 1-57.

Refrences:

Abadyo, A., & Bastari, B. (2015). Estimation of ability and. item parameters in mathematics testing by using the. combination of 3PLM/GRM and MCM/GPCM scoring model. REiD (Research and Evaluation in Education), 1(1), 55-72.

Baker, F. b. (2001). The basics of item Response theory. ERIC. Clearinghouse on Assessment and Evaluation.

Falani, I., Akbar, M., & Naga, D. S. (2020). The Precision of Students'. Ability Estimation on. Combinations of Item Response Theory Models. International Journal of Instruction, 13(4), 545-558.

Graham, J. w. (2009). Missing data analysis: making it. work in the real world. Annual review of psychology, 60(1), 549-576.

Hambleton, R. K., Swaminathan, H., Hambleton, R. K., & Swaminathan, H. (1985). Estimation of. Ability. Item Response Theory: Principles and Applications, 75-99.

Hattie, J. (1985). Methodology review: assessing unidimensionality of. tests and Itenls. Applied psychological measurement, 9(2), 139-164.

Kehinde, O. J., Dai, S., & French, B. (2022, September). Item parameter estimations for multidimensional. graded response model under complex structures. In Frontiers in Education (Vol. 7, p. 947581). Frontiers.

Kim, S., Cohen, A., & Lin, Y. (2005). LDID: A computer program for local dependence. indices for dichotomous items. Applied Psychological Measurement.

Kirmayr, M., Quilodrán, C., Valente, B., Loezar, C., Garegnani, L., & Franco, J. V. A. (2021). The GRADE approach, Part 1: how to assess the. certainty of the evidence. Medwave, 21(02), 9-38.

Kubiszyn, T., & Borich, G. d. (2024). Educational testing and measurement. John Wiley & Sons.

Kubiszyn, T., & Borich, G. d. (2024). Educational testing and. measurement. John Wiley & Sons.

Kumar, Y., Gupta, A., & Tounsi, A. (2021). Size-dependent vibration. Response of porous graded nanostructure with FEM and nonlocal. continuum model. Advances in nano research, 11(1), 1-17.

Little, R. j., & Rubin, D. B. (2019). Statistical analysis with missing data (Vol. 793). John Wiley & Sons.

Little, R. j., & Rubin, D. B. (2019). Statistical analysis with. Missing data (Vol. 793). John Wiley & Sons.

Lord, F. m. (2012). Applications of item response theory to practical. testing problems. Routledge.

Mcknight, P., Mcknight, K., Sidani, S., & Figueredo. A. (2007). Missing data: A gentle introduction. New York: Guilford press.

Mohan, R. (2023). Measurement, evaluation and assessment in education. PHI Learning Pvt. Ltd.
Mohan, R. (2023). Measurement, evaluation and assessment in education. PHI Learning Pvt. Ltd.
Santoso, A., Pardede, T., Djidu, H., Apino, E., Rafi, I., Rosyada, M. N., & Abd Hamid, H. S. (2022). The effect of scoring correction and. model fit on the estimation of ability parameter and person fit on polytomous item response theory. REID (Research and Evaluation in Education), 8(2), 140-151.

Xiong, J., Ding, S., Luo, F., & Luo, Z. (2020). Online calibration of. polytomous items under the graded response model. Frontiers in psychology, 10, 1-12.