المجلد 16 العدد 3 سبتمبر 2015

تقصي فاعلية الإحصائي (Lz^{*new}) في الكشف عن أنماط استجابة المفحوص غير المطابقة وفق نظرية استجابة المفردة

د. إياد محمد حمادنه
العلوم النفسية والتربية الخاصة
كلية العلوم التربوية – جامعة آل البيت
imh1232000@yahoo.com

تقصي فاعلية الإحصائي (Lz^{*new}) في الكشف عن أنماط استجابة المفحوص غير المطابقة وفق نظرية استجابة المفردة

د. إياد محمد حمادنه
العلوم النفسية والتربية الخاصة
كلية العلوم التربوية – جامعة آل البيت

الملخص

هدفت هذه الدراسة، إلى تقصي فاعلية الإحصائي (Lz^{*new}) في الكشف عن أنماط استجابة المفحوص غير المطابقة، بدلالة ضبط الخطأ من النوع الأول، في ضوء اختلاف النموذج اللوجستي (2PLM, 3PLM) لنظرية استجابة المفردة، عند (١١) مستوى مختلفا للقدرة، وثلاثة مستويات مختلفة للدلالة الإحصائية. لتحقيق أهداف الدراسة استخدمت بيانات ثنائية مولدة على اختبار مؤلف من (٤٠) مفردة. واستخدمت البرمجيات الإحصائية الحديثة (Wingen2 Bilog-MG3; R-Fit; SPSS) على الترتيب؛ لتحليل البيانات. أظهرت نتائج الدراسة فاعلية الإحصائي (Lz^{*new}) في الكشف عن أنماط الاستجابة غير المطابقة مهما اختلف نوع النموذج اللوجستي المستخدم، إلا أنه أكثر فاعلية في النموذج اللوجستي ثلاثي المعلم (3PLM) منه في النموذج الثنائي المعلم (2PLM). كما أشارت النتائج إلى أن الخطأ من النوع الأول في هذا النموذج (3PLM) يقترب أكثر من المستوى الاسمي، عبر جميع مستويات القدرة المختلفة، مهما اختلف مستوى الدلالة الإحصائية. وفي ضوء النتائج التي تم التوصل إليها قدم الباحث مجموعة من التوصيات والمقترحات

الكلمات المفتاحية: نظرية استجابة المفردة، الإحصائي (Lz^{*new})، أنماط الاستجابة غير المطابقة.

* تاريخ تسلم البحث: ٢٠١٣/١١/١٩م
* تاريخ قبوله للنشر: ٢٠١٤/٣/٣١م

The Effectiveness of (Lz^{*new}) Index in Detecting Person Aberrant Response Patterns According to IRT

Dr. Iyad M. Hamadneh
Faculty of Educational Sciences
Al al-Bayt University

Abstract

The present study aimed to investigate the effectiveness of Lz^{*new} index in detecting person aberrant response patterns, according to the Item Response Theory (IRT), by investigating the empirical type I error in the light of the different logistic models (2PLM and 3PLM) of IRT at 11 different ability levels, and three different levels of statistical significance. To achieve the objectives of the study, binary data generated from a 40-item test were used. Up-to-date statistical software packages (Wingen2, Bilog- MG3, R-Fit, and SPSS) were used chronologically to analyze the data. Findings of the study indicated the effectiveness of Lz^{*new} index in detecting person aberrant response patterns, no matter what the type of the logistic model used. However, it was more effective in the 3–Parameters Logistic Model (3PLM) than the 2–Parameters Logistic Model (2PLM). The results also indicated that type I error (3PLM) rates close to the nominal value for most ability level, no matter what the level of statistical significance was. In the light of the study findings, a number of conclusions and recommendations are suggested.

Keywords: item response theory, Lz^{*new} index, aberrant response patterns.

تقصي فاعلية الإحصائي (Lz^{*new}) في الكشف عن أنماط استجابة المفحوص غير المطابقة وفق نظرية استجابة المفردة

د. إياد محمد حمادنه
العلوم النفسية والتربية الخاصة
كلية العلوم التربوية – جامعة آل البيت

المقدمة

برزت نظرية استجابة المفردة (Item Response Theory, IRT)؛ لمعالجة جوانب القصور في النظرية الكلاسيكية، إذ أن أبرز مزايا نظرية استجابة المفردة (IRT) هي إن تقديرات الخصائص السيكومترية للفقرات مستقلة عن عينة المفحوصين التي استخدمت في تقدير هذه الخصائص (Person Free)، وإن تقديرات قدرات المفحوصين مستقلة عن عينة الفقرات التي تطبق عليهم (Item Free). علاوة على ذلك، توفر هذه النظرية مؤشراً إحصائياً لتقدير درجة الدقة في تقدير القدرة لكل مفحوص (Embretson & Reise, 2000).

وتستند نماذج نظرية استجابة المفردة إلى افتراضات قويّة يجب تحقيقها في البيانات؛ لكي تؤدي إلى نتائج يمكن الوثوق بها، وهذه الافتراضات هي: افتراض أحادية البعد (Unidimensionality)، وافتراض الاستقلال الموضعي (Local Independence)، وافتراض التحرر من السرعة (Speededness)، وافتراض منحنى خصائص المفردة (Item Characteristic Curve, ICC) (Gruijter & Kamp, 2005).

وهناك عدد من النماذج في نظرية استجابة المفردة (IRT Models) صُنّفت إلى نوعين، هما النماذج التي تستخدم المنحنى الطبيعي التراكمي، والنماذج اللوجستية (Logistic Models). وتُعد النماذج اللوجستية أوسع انتشارا في الوقت الحاضر، وهي ملائمة لفقرات ثنائية التدريج، وتختلف هذه النماذج فيما بينها باختلاف عدد معالم المفردة المقدرة (Hambleton, Swaminathan & Rogers, 1991)، وهذه النماذج هي:

أولاً: النموذج اللوجستي أحادي المعلم: (One-Parameter Logistic Model, 1PLM) ويُعرف بنموذج راش، وهو أكثر النماذج شهرة واستخداما وذلك لسهولة تطبيقه، حيث يفترض هذا النموذج أن جميع الفقرات تُميز بين الأفراد بنفس القدر، كما يفترض عدم تأثر الإجابات بعامل التخمين، ولكنها تتباين فقط في صعوبتها، والمعادلة الرياضية لهذا النموذج، هي:

$$P_i(\theta) = \frac{1}{1+\exp[-1.7(\theta - b_i)]} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (1)$$

حيث، $(P_i\,\theta)$: هي احتمال أن يجيب المفحوص ذو القدرة (θ) إجابة صحيحة على المفردة i.

D: عامل التدريج (Scaling Factor) وهو ثابت يساوي ١,٧.

ai: معامل تمييز المفردة i، bi: معامل صعوبة المفردة i، e: الأساس اللوغاريتمي الطبيعي =٢,٧١٨٣

ثانياً: النموذج اللوجستي ثنائي المعلم: (Two-Parameter Logistic Model, 2PLM)

يفترض هذا النموذج اختلاف الفقرات في صعوبتها وتمييزها، كما يفترض عدم تأثر الإجابات بعامل التخمين، والمعادلة الرياضية لهذا النموذج، هي:

$$P_i(\Theta) = \frac{1}{1 + e^{-Da_i(\theta - b_i)}} \quad \text{.......................................} (٢)$$

ثالثاً: النموذج اللوجستي ثلاثي المعلم: (Three-Parameter Logistic Model, 3PLM)

أضاف هذا النموذج معلماً ثالثا يمثل إجابة المفحوص إجابة صحيحة على المفردة عن طريق التخمين، والمعادلة الرياضية لهذا النموذج، هي:

$$P_i(\Theta) = C_i + (1 - C_i)\frac{e^{Da_i(\theta - b_i)}}{1 + e^{Da_i(\theta - b_i)}} \quad \text{..................................} (٣)$$

وهناك طرق متعددة لتقدير قدرات المفحوصين في نظرية استجابة المفردة، ومنها: طريقة تقدير الأرجحية العظمى (Maximum Likelihood Estimation, MLE)، وطريقة بيز (Bayesian Method, EAP) التي تُستخدم عادةً عندما لا يكون بالإمكان تطبيق طريقة الأرجحية العظمى، وتشير الدراسات السابقة إلى أن تقديرات القدرة وفق طريقة بيز (EAP) متوفرة لكل أنماط الاستجابة، بعكس طريقة (MLE) التي تكون مهمة في الاختبارات القصيرة فقط، كما أن طريقة بيز لها أقل تحيز وأقل خطأ معياري وتعطي تقديرات دقيقة للقدرة مقارنة بالطرق الأخرى لتقدير القدرة (Nering, 1995; Reeve, 2004; Reise, 1995).

ولكل مفردة من مفردات الاختبار هناك ما يسمى بدالة معلومات المفردة (Item Information Function, $Ii\theta$)، وقد أورد بيكر (Baker, 2001) صيغاً لمعادلة دالة معلومات المفردة وفق النماذج اللوجستية الثلاثة لنظرية استجابة المفردة (IRT) كما يلي:

١. معادلة دالة معلومات المفردة وفق النموذج اللوجستي ثلاثي المعلم (3PLM).

$$I_i(\theta) = a_i^2\left[\frac{Q_i(\theta)}{P_i(\theta)}\right]\left[\frac{P_i(\theta) - C_i}{1 - C_i}\right]^2 \quad \text{.................................} (٤)$$

٢. معادلة دالة معلومات المفردة وفق النموذج اللوجستي ثنائي المعلم (2PLM).

$$Ii(\theta) = ai^2 . Pi(\theta)\, Qi(\theta) \quad \text{.................................} (٥)$$

حيث إن معلم التخمين Ci = صفر.

٣. معادلة دالة معلومات المفردة وفق النموذج اللوجستي أحادي المعلم (1PLM)

$$I_i(\theta) = P_i(\theta)\, Q_i(\theta) \quad \ldots\ldots(٦)$$

وتعدّ دالة معلومات الاختبار من المفاهيم الأساسية في نظرية استجابة المفردة، وترتبط بالثبات ودقة القياس، وستساهم دالة معلومات الاختبار في الدراسة الحالية في حساب معامل الثبات الذي يُستخدم من أجل تصحيح تقدير القدرة من خلال العلاقات الآتية: (Darabi & Dodeen, 2009; Wainer, 2001)

$$\rho(\hat{\theta}) = \frac{1}{1 + \frac{1}{I(\hat{\theta})}} \quad \ldots\ldots(٧)$$

$$\hat{\theta}^{*} = \frac{\hat{\theta}}{\rho(\hat{\theta})} \quad \ldots\ldots(٨)$$

$$\hat{\theta}^{*} = \hat{\theta} \times [1 + \frac{1}{I(\hat{\theta})}]. \quad \ldots\ldots(٩)$$

$$\hat{\theta}^{*} = \hat{\theta} \times [1 + \mathrm{var}(\hat{\theta})]. \quad \ldots\ldots(١٠)$$

حيث، $\rho(\hat{\theta})$: هي معامل الثبات للقدرة المقدرة، $\hat{\theta}$: هي القدرة المقدرة، $\hat{\theta}^{*}$: هي تصحيح القدرة المقدرة،

$I(\hat{\theta})$: هي دالة معلومات الاختبار وفق القدرة المقدرة، $Var(\hat{\theta})$: هي $S^2(\hat{\theta})$ تباين القدرة المقدرة.

فيما أشار هامبلتون وزملاؤه (Hambleton et al., 1991) لدالة معلومات الاختبار بالعلاقة:

$$I(\Theta) = \sum_{i=1}^{N} I_i(\Theta) \quad \ldots\ldots(١١)$$

من جهة أخرى، فإنه يجب التأكد من صدق علامات الاختبار، حيث إن تقديرات القدرة للمفحوصين ذوي أنماط الاستجابة غير المطابقة (Aberrant Response Patterns, ARP) للنموذج المستخدم تكون غير دقيقة وغير صادقة، ولهذا فإن القرارات المبنية على هذه التقديرات حتما ستكون غير عادلة. ولقياس أداء المفحوصين على الاختبار بشكل دقيق، لا بد من الكشف عن مطابقة المفحوص (PF Person Fit) من خلال تحديد أنماط الاستجابة غير المطابقة (ARP)، ويعرّف مؤشر مطابقة الفرد بأنه مؤشر إحصائي يحدّد المدى او البعد بين

البيانات الفعلية المتمثلة باستجابات المفحوصين والقيم المتوقعة من خلال النموذج المستخدم (Lopez & Montesinos, 2005; Meijer & Sijtsma, 2001). ومن الجدير بالذكر، أنه إذا تمّ تصنيف نمط الاستجابة على أنها غير مطابقة، يُعتبر الاختبار في هذه الحالة اختباراً غير صادق، أو مقياساً غير ملائم للمفحوص (Chan & Schmitt, 1999; Deng, 2007).

وفي هذا السياق، تم اقتراح ثلاث طرق لبحث مطابقة المفحوص لنموذج نظرية استجابة المفردة، وهي: طريقة اختبارات مطابقة المفحوص (Person Response Tests, PFT)، وطريقة دالة استجابة المفحوص (Person Response Function, PRF) (Meijer & Van Krimpen, 1999; Trabin & Weiss, 1983)، وطريقة إحصائيات مطابقة المفحوص (Person Fit Statistics, PFS) والتي تُشير إلى درجة المطابقة بين النموذج اللوجستي الذي يتم اختياره ونمط علامة المفحوص على المفردة. ويمكن قياس درجة المطابقة بواسطة إحصائي مطابقة المفحوص الذي يعكس مدى إجابته عن أسئلة الاختبار وفقا لافتراضات النموذج، ويُعد نمط الاستجابة غير مطابق، إذا وجد أنه غير محتمل في ظل النموذج المستخدم حسب إحصائي مطابقة المفحوص المطبق (Karabatsos, 2003).

ويُمكن أن تُصنّف إحصائيات مطابقة المفحوص على أنها إحصائيات لا معلميه (Non-Parametric) أو معلميه (Parametric)، حيث إن إحصائي مطابقة المفحوص اللامعلمي لا يعتمد على معالم نظرية استجابة المفردة المقدّرة، وإنما يتم حسابه من مجموعة من البيانات لعدد من المفحوصين لهم إجابات على عدد من مفردات الاختبار، ومن الأمثلة على هذا النوع من الإحصائيات: الإحصائي C والإحصائي (U3) (Karabatsos, 2003). وأما إحصائي مطابقة المفحوص المعلمي فيقيس الفرق بين مجموعة بيانات الاختبار وأنماط الاستجابة المقدرة المشتقة من تقديرات معالم نموذج نظرية استجابة المفردة (IRT)، ومن الأمثلة على هذا النوع من الإحصائيات: الإحصائي L0، والإحصائي LZ، والإحصائي (Lz^{*new}) (Blais & Raiche, 2005; Torre & Deng, 2008). وهناك معياران للحكم على أفضلية إحصائي مطابقة المفحوص: يتعلق الأول بمدى تمتع الإحصائي بخصائص التوزيع الطبيعي (Birenbaum, 1985)، فيما يتعلق المعيار الثاني بقوة الإحصائي أو حساسيته في الكشف عن الأنماط غير المطابقة (Torre & Deng, 2008).

وتختلف إحصائيات مطابقة المفحوص من حيث معنى الإحصائي والأساس الرياضي الذي يُبنى عليه، ويُعد إحصائي المطابقة المعياري (Lz^{*new})، والمعروف بإحصائي لوغاريتم الأرجحية المعياري (Standardized Log-Likelihood Statistic, Lz^{*new}) الذي اقترحه

توري ودنج (Torre & Deng, 2008) أكثر إحصائي شيوعاً واستخداما وألفة؛ لكونه أسلوباً لتحسين مطابقة المفحوص من خلال تصحيح تقدير القدرة وتوزيعها المرجعي، حيث يتطلب هذا الأسلوب حساب قيمة الإحصائي Lz في ضوء قيم تصحيح تقدير القدرة $\widehat{\theta}^{*}$ ويتم التصحيح لعدم الثبات في الاختبار بالاعتماد على معامل الثبات الذي يرتبط بدالة معلومات الاختبار عند قيمة θ وفق المعادلة (٧). ويعطى التصحيح المقترح لقيمة θ المقدرة وفق المعادلة (٩). كما أن حساباته سهلة نسبياً، ويقيس بشكل مباشر المطابقة بين نموذج نظرية استجابة المفردة (IRT Model) ونمط استجابة المفحوص (ARP)، ويحسب من خلال المعادلات الآتية:

$$L_z = \frac{L_o - E(L_o)}{[Var(L_o)]^{\frac{1}{2}}} \quad \text{.....(١٢)}$$

$$E(L_o) = \sum_{j=1}^{n} \left(P_j(\theta) Ln P_j(\theta) + \left(1 - P_j(\theta)\right) Ln\left(1 - P_j(\theta)\right) \right) \quad \text{(١٣)}$$

$$Var(L_o) = \sum_{j=1}^{n} P_j(\theta)\left(1 - P_j(\theta)\right) \left[Ln \frac{P_j(\theta)}{1 - P_j(\theta)} \right]^2 \quad \text{.....(١٤)}$$

ومن أسباب حدوث الأنماط غير المطابقة للمفحوص (ARP)، الخصائص الديموغرافية له والتي تتعلق بالجنس، العرق، والتحيز الثقافي (Torre & Deng, 2008)، والغش (Cheating) والتخمين (Meijer,1996) (Guessing)، وقلق الاختبار (Birenbaum, Anxiety) (1985)، والتباطؤ في أداء الاختبار (Runder,1996) (Plodding Behavior)، وتعدّ هذه الأسباب بمجملها عوامل دخيلة على القدرة المقاسة.

وعموما، فقد حظي موضوع استجابة المفحوص غير المطابقة (ARP) بالعديد من الدراسات، حيث نال اهتماما متزايداً من المختصين في علم النفس والقياس والتربويون وأصحاب القرار.

فقد أجرى ريس (Reise, 1995) دراسة هدفت إلى الكشف عن أنماط المفحوص غير المطابقة باستخدام الإحصائي Lz، وباستخدام بيانات مولدة وفق النموذج اللوجستي ثنائي المعلم (2PLM)، عند استخدام القدرة الحقيقية (True θ) وفق خمسة مستويات للقدرة (θ= -٢، -١، ٠، ١، ٢) وثلاث طرق لتقدير القدرة (BIW, EAP & MLE). أظهرت النتائج أن نسبة الكشف عن الأنماط غير المطابقة كانت أعلى ما يمكن عند استخدام القدرة الحقيقية و كانت أقل ما يمكن عند استخدام طريقة الأرجحية العظمى MLE في تقدير القدرة.

فيما أجرى سنجدرز (Snijders, 2001) دراسة هدفت إلى تعرف أثر طول الاختبار

على فاعلية إحصائي مطابقة المفحوص Lz بدلالة الخطأ من النوع الأول، وكذلك بدلالة قوة الإحصائي في الكشف عن أنماط الاستجابة غير المطابقة (APR). واستخدمت بيانات ثنائية مولدة وفق النموذج اللوجستي ثلاثي المعلم (3PLM). واستخدمت كذلك القدرة المقدرة $\hat{\theta}^*$ لحساب الإحصائي Lz. أظهرت النتائج اقتراب الخطأ من النوع الأول من المستوى الاسمي عبر جميع مستويات القدرة عند مستويات الدلالة (α= 0.05, 0.10)، ولكن عند مستويات الدلالة (α= 0.01, 0.02)، كان الخطأ من النوع الأول يبتعد كثيراً عن المستوى الاسمي. كما أظهرت النتائج أن الخطأ من النوع الأول يكون أكثر اقتراباً من المستوى الاسمي كلما زاد طول الاختبار.

وقام دودين (Dodeen, 2003) بدراسة هدفت إلى استخدام إحصائي مطابقة المفحوص Lz لتحليل اختبارين، لقياس مستوى الطلبة في اللغة العربية واللغة الانجليزية، وفق النموذج اللوجستي ثلاثي المعلم (3PLM). تكونت عينة الدراسة من (٥٥٨) طالبا وطالبة تقدموا على الاختبارين. أظهرت النتائج عدم وجود ارتباط بين إحصائي مطابقة المفحوص Lz ومستوى القدرة. كما أظهرت النتائج أن توزيع الإحصائي Lz لكلا الاختبارين كان قريباً جداً من التوزيع الطبيعي المعياري، وأن نسبة عدم المطابقة للأفراد كانت قليلة جداً،

وقام لوبز ومونتيسينوس (Lopez & Montesinos, 2005) بدراسة هدفت إلى مقارنة خمسة مؤشرات مختلفة لمطابقة المفحوصين، وفق النموذج اللوجستي أحادي المعلم (1PLM) (نموذج راش)، باستخدام بيانات مولدة تحت ظروف مختلفة في حجم العينة، وصعوبة المفردات، وشكل توزيع القدرة (طبيعي، منتظم)، وطول الاختبار (٣٠، ١٥) مفردة. أظهرت النتائج أن الإحصائي الذي يتمتع بخصائص التوزيع الطبيعي ليس بالضرورة أن يكون أقوى مؤشر يكشف عن أعلى نسبة من أنماط استجابة المفحوص غير المطابقة (APR)، إلا أن إحصائي مطابقة المفحوص Lz كان أكثر حساسية في تحديد أنماط الاستجابة غير المطابقة.

وأجرى براون وفيلاريال (Brown & Villareal, 2007) دراسة هدفت إلى تعرف أثر كل من الجنس والعرق ومستوى الصف على مطابقة المفحوص وفق النموذج اللوجستي ثنائي المعلم (2PLM)، حيث تم تطبيق اختبار تكيفي محوسب (Computer Adaptive Test, CAT) في مادة الرياضيات على عينة تكونت من (١٦٠,٠٠٠) طالب وطالبة، واستخدم إحصائي مطابقة المفحوص Lz للكشف عن أنماط استجابة المفحوص غير المطابقة. أظهرت النتائج عدم وجود أثر للجنس على أنماط الاستجابة غير المطابقة، بينما كان هناك أثر للعرق على أنماط الاستجابة غير المطابقة حيث إن المفحوصين السود كان لهم أنماط استجابة غير

مطابقة أعلى من المفحوصين البيض. كما أظهرت النتائج أن أنماط الاستجابة غير المطابقة تزداد بزيادة مستوى الصف للمفحوص.

وقام دنج (Deng, 2007) بدراسة أثر طول الاختبار وعدد الأبعاد ومعامل الارتباط بين القدرات على فاعلية إحصائي مطابقة المفحوص Lz*new في الكشف عن أنماط استجابة المفحوص غير المطابقة، بدلالة الخطأ من النوع الأول، وعند خمسة مستويات من القدرة، وثلاثة مستويات مختلفة من الدلالة الإحصائية، باستخدام بيانات مولدة وفق النموذج اللوجستي ثلاثي المعلم (3PLM). حيث اقترح طريقة لتحسين مطابقة المفحوص اعتمادا على أسلوب تصحيح تقدير القدرة وتوزيعها المرجعي ($\widehat{\theta}^{*}$). أظهرت النتائج أن فاعلية إحصائي المطابقة Lz في الكشف عن الأنماط غير المطابقة كانت أكبر ما يمكن، ومقدار الفرق بين الخطأ من النوع الأول والمستوى الاسمي كان أقل ما يمكن كلما زاد كل من طول الاختبار، وعدد الأبعاد، ومعامل الارتباط بين القدرات، على الترتيب.

فيما قام توري ودنج (Torre & Deng, 2008) بدراسة هدفت إلى تعرف فاعلية الإحصائي Lz^{*new} في الكشف عن أنماط استجابة المفحوص غير المطابقة، بدلالة الخطأ من النوع الأول، عند أطوال مختلفة من اختبارات أحادية البعد، ومستويات مختلفة من القدرة، عند مستويات مختلفة من الدلالة، وباستخدام بيانات ثنائية مولدة وفق النموذج اللوجستي ثلاثي المعلم (3PLM). أظهرت النتائج أن هذه الطريقة قدمت أفضل النتائج من حيث التقارب بين الخطأ من النوع الأول المقدر والخطأ الاسمي مقارنة مع الطرق الأخرى التي استخدمت القدرة الحقيقية والقدرة المقدرة، وكذلك أظهرت النتائج أن هذه الطريقة هي الأكثر فاعلية في الكشف عن أنماط الاستجابة غير المطابقة وخاصة كلما زاد طول الاختبار.

وبناء على ما تقدم، يتضح الاهتمام الكبير الذي توليه الدراسات في موضوع أنماط استجابة المفحوص غير المطابقة (ARP)، وما له من دور كبير في تطوير المقاييس والاختبارات في معظم المجالات. فقد ركزت جميع الدراسات السابقة على إحصائي مطابقة المفحوص Lz من حيث دراسة خصائص توزيعه الطبيعي وقوته في الكشف عن أنماط الاستجابة غير المطابقة عند ظروف مختلفة، مثل: اختلاف كل من طول الاختبار، وطرق تقدير القدرة، وتوزيع القدرة، وخصائص المفردات، وعلى قوة هذا الإحصائي. كما استخدمت بعض الدراسات السابقة أسلوب القدرة الحقيقية لحساب إحصائي المطابقة Lz، وبعضها استخدم أسلوب القدرة المقدرة لحسابه، والبعض الآخر استخدم أسلوب تصحيح تقدير القدرة $\widehat{\theta}^{*}$، وفيما يتعلق بعينة الدراسة، فيلاحظ أن غالبية الدراسات استخدمت بيانات مولدة، وذلك لتوفير أكبر

حجم للعينة والتي تلزم لإجراء التحليلات وفق نظرية استجابة المفردة، من هذه البيانات ما تم توليده لتلاءم افتراضات النموذج اللوجستي ثلاثي المعلم (3PLM)، وأخرى تلاءم النموذج اللوجستي ثنائي المعلم (2PLM)، وأخرى تلاءم افتراضات النموذج اللوجستي أحادي المعلم (1PLM). ولم يجد الباحث أي دراسة تناولت أثر اختلاف النموذج اللوجستي المستخدم (2PLM, 3PLM) على فاعلية الإحصائي LZ في الكشف عن أنماط الاستجابة غير المطابقة. من هنا، تأتي هذه الدراسة في محاولة للكشف عن أنماط استجابة المفحوص غير المطابقة (ARP) باستخدام إحصائي لوغاريتم الأرجحية المعياري القائم على أسلوب تصحيح تقدير القدرة وتوزيعها المرجعي Lz^{*new}، في ضوء اختلاف النموذج اللوجستي، وباستخدام بيانات محاكاة.

مشكلة الدراسة

إن تقديرات القدرة للمفحوصين ذوي أنماط الاستجابة غير المطابقة للنموذج اللوجستي المستخدم، تكون غير دقيقة وغير صادقة، ولهذا فإن القرارات المبنية على هذه التقديرات حتما ستكون غير عادلة وغير ملائمة. ومن أجل تحقيق الصدق والعدالة في استخدام وتفسير علامات الاختبار، لا بد من قياس أداء المفحوصين على الاختبار بشكل دقيق، من هنا يأتي دور إحصائيات مطابقة المفحوص (PFS) في هذا المجال. ولقد اتفقت نتائج أحدث الدراسات السابقة على أن إحصائي مطابقة المفحوص Lz^{*new} هو الأكثر فاعلية ودقة في الكشف عن أنماط الاستجابة غير المطابقة مقارنة مع إحصائيات مطابقة المفحوص الأخرى.

ومن الجدير بالذكر، أن الدراسات السابقة تناولت إحصائي مطابقة المفحوص Lz^{*new} ، من حيث خصائص توزيعه الطبيعي وقوته في الكشف عن أنماط الاستجابة غير المطابقة باختلاف طول الاختبار وطرق تقدير القدرة وتوزيع القدرة وخصائص الفقرات، إلا أنها لم تولي اهتماما بأثر النموذج اللوجستي المستخدم. ومن هنا، فإن الدراسة الحالية تسعى للكشف عن فاعلية الإحصائي Lz^{*new} في الكشف عن أنماط الاستجابة غير المطابقة بدلالة الخطأ من النوع الأول في ضوء اختلاف النموذج اللوجستي المستخدم (2PLM, 3PLM) لنظرية استجابة المفردة، وباستخدام بيانات ثنائية مولدة لاختبار مكون من ٤٠ مفردة.

أسئلة الدراسة

تحاول الدراسة الحالية الإجابة عن الأسئلة الآتية:

١- ما فاعلية الإحصائي Lz^{*new} في الكشف عن أنماط استجابة المفحوص غير المطابقة،

بدلالة قدرته على ضبط الخطأ من النوع الأول، عند مستويات مختلفة من القدرة والدلالة الإحصائية (α= 0.01, 0.05, 0.10) في ضوء استخدام النموذج اللوجستي ثلاثي المعلم (3PLM) ؟

٢- ما فاعلية الإحصائي Lz^{*new} في الكشف عن أنماط استجابة المفحوص غير المطابقة، بدلالة قدرته على ضبط الخطأ من النوع الأول، عند مستويات مختلفة من القدرة والدلالة الإحصائية (α= 0.01, 0.05, 0.10) في ضوء استخدام النموذج اللوجستي ثنائي المعلم (2PLM) ؟

هدف الدراسة

تهدف الدراسة الحالية إلى الكشف عن أنماط استجابة المفحوص غير المطابقة (ARP) باستخدام إحصائي لوغاريتم الأرجحية المعياري القائم على أسلوب تصحيح تقدير القدرة وتوزيعها المرجعي Lz^{*new} في ضوء اختلاف النموذج اللوجستي (2PLM, 3PLM) لنظرية استجابة المفردة، بدلالة الخطأ من النوع الأول، عند مستويات مختلفة من القدرة والدلالة الإحصائية، باستخدام بيانات ثنائية مولدة لاختبار مكون من ٤٠ مفردة، عند (١١) مستوى مختلفا من القدرة تتراوح قيمها بين (٣- و ٣) بزيادة منتظمة مقدارها ٠,٦، وبتوزيع طبيعي، وعند ثلاثة مستويات مختلفة للدلالة الإحصائية (α= 0.01, 0.05, 0.10).

أهمية الدراسة

تتجلى أهمية الدراسة الحالية في الجوانب الآتية:

١. تبحث في التحقق من التقدم الكبير الذي أحرزته نظرية الاستجابة للمفردة (IRT) وتطبيقاتها في الكشف عن أنماط استجابة المفحوص غير المطابقة (ARP)، وما توفره من مؤشرات إحصائية للكشف عن هذه الأنماط.

٢. تساهم في تسليط الضوء على موضوع تقديرات القدرة للمفحوصين ذوي أنماط الاستجابة غير المطابقة للنموذج اللوجستي المستخدم، وما له من أثر على دقة وصدق القرارات المبنية على هذه التقديرات.

٣. يتوقع أن تزوّد هذه الدراسة علماء النفس والقياس وأصحاب القرار والتربويين، بمنهجية علمية يمكن من خلالها التعرف على آلية حساب إحصائي لوغاريتم الأرجحية المعياري القائم على أسلوب تصحيح القدرة وتوزيعها المرجعي Lz^{*new} في ضوء اختلاف النموذج اللوجستي لنظرية استجابة المفردة.

٤. تسليط الضوء على ضرورة توفير البرمجيات الإحصائية الحديثة (Wingen2؛ Bilog-MG3؛ R-Fit) للكشف عن أنماط استجابة المفحوص غير المطابقة وفق الإحصائي Lz^{*new}.

مصطلحات الدراسة

أحادية البعد (Unidimensionality): وجود قدرة (سمة) واحدة تفسر أداء المفحوص في الاختبار.

الاستقلال الموضعي (Local Independence) أن تكون علامات المفحوص مستقلة إحصائياً الواحدة عن الأخرى، عند أي نقطة على متصل السمة.

التحرر من السرعة (Non – Speededness): تفترض نماذج (IRT) أن إخفاق المفحوصين في الإجابة عن فقرات الاختبار يرجع إلى انخفاض قدرتهم، وليس إلى تأثير عامل السرعة على الإجابة.

معلم الصعوبة (bi): هو مقدار القدرة التي يكون عندها $P_i(q) = \frac{1 + C_i}{2}$ حيث (Ci) قيمة معلم التخمين، وتأخذ صعوبة المفردة في نظرية الاستجابة للمفردة قيماً بين $(-\infty, +\infty)$ ولكن بشكل عام، فإن القيم التي تأخذها تكون بين (٣- و ٣).

معلم التمييز (ai): هو مؤشر يربط بين التغير في القدرات والتغير في احتمال الإجابة الصحيحة. و(ai) هي ليست الميل، وإنما نسبة منه، ولكن كلما زاد الميل يزيد (ai).

معلم التخمين (ci): هو احتمال إجابة المفردة إجابة صحيحة من المفحوصين ذوي القدرة المتدنية.

الإحصائي Lz*new: هو إحصائي لوغاريتم الأرجحية المعياري Lz الذي تم حسابه اعتمادا على تصحيح تقدير القدرة الجديدة $(\hat{\theta}^*_{new})$ التي تنتمي لتوزيع معدل متوسطه الحسابي يساوي $\hat{\theta}^*$ وتباينه التقريبي يساوي $\frac{1}{1 + I(\hat{\theta}^*)}$.

الخطأ من النوع الأول: نسبة المفحوصين المطابقين للنموذج والذين يتم تصنيفهم في فئة غير المطابقين للنموذج اعتمادا على القيم الحرجة للإحصائي Lz حسب مستوى الدلالة α.

النماذج اللوجستية: نماذج تنبثق عن نظرية استجابة المفردة (IRT)، وتقوم على عدد من الافتراضات ينبغي التحقق منها. وفي الدراسة الحالية، تم تناول نموذجين هما (2PLM, 3PLM)، وتعطى بالمعادلتين ٢، ٣ على الترتيب.

محددات الدراسة

تتحدد الدراسة الحالية بما يلي:

١- اقتصرت عينة الدراسة على بيانات مولدة لـ (n = ١٠٠٠) مفحوص.

٢- اقتصرت الدراسة على استخدام الإحصائي Lz^{*new} في الكشف عن أنماط استجابة المفحوص غير المطابقة.

٣- اقتصرت أداة الدراسة على بيانات مولدة ثنائية التدريج (٠، ١) لاختبار واحد مكون من ٤٠ مفردة، وما يتوافر من برمجيات إحصائية حديثة (Wingen2؛ Bilog- MG3؛ R-Fit).

٤- استخدمت الدراسة مدى لقدرة المفحوصين يتراوح قيمها بين (٣- و ٣) بزيادة منتظمة مقدارها ٠,٦، وبتوزيع طبيعي، وعند ثلاثة مستويات مختلفة للدلالة الإحصائية (α= 0.01, 0.05, 0.10).

٥- اقتصرت الدراسة على نموذجين فقط لنظرية استجابة المفردة، هما: (2PLM, 3PLM).

منهجية الدراسة وإجراءاتها

تصميم الدراسة

لتعرّف فاعلية إحصائي لوغاريتم الأرجحية المعياري القائم على أسلوب تصحيح تقدير القدرة وتوزيعها المرجعي Lz^{*new} في الكشف عن أنماط استجابة المفحوص غير المطابقة (ARP)، بدلالة الخطأ من النوع الأول. فقد تم توليد ٤٠٠٠ استجابة لعينة من المفحوصين بلغت ١٠٠٠ مفحوص على اختبار مكون من ٤٠ مفردة وفق النموذجين (2PLM, 3PLM) لنظرية استجابة المفردة، عند ١١ مستوىً مختلفاً من القدرة تتراوح قيمها بين (٣- و ٣) بزيادة منتظمة مقدارها ٠,٦، وعند ثلاثة مستويات من الدلالة (α= 0.01, 0.05, 0.10)، وذلك باستخدام البرمجيات الإحصائية الملائمة لنظرية استجابة المفردة.

توليد البيانات

في ظل التقدم الكبير الذي أحرزته نظرية استجابة المفردة (IRT)، فقد ظهرت العديد من البرمجيات الإحصائية التي تخدم هذه النظرية وتطبيقاتها الرائدة مثل، برمجية Wingen2 (Harwell, 1996) وبرمجية (Bilog- MG3) (Zimowski, Mislevy & Back, 1996)، وبرمجية (R-Fit) (Lopez & Montesinos, 2005). حيث استخدمت برمجية Wingen2 في توليد استجابات ثنائية (١، ٠) على فقرات أحادية البعد (Unidimensionality) لعينة

من المفحوصين بلغت (١٠٠٠) مفحوص على اختبار مكون من (٤٠) مفردة، وقد تم توليد (٤٠٠٠) استجابة عند كل مستوى من مستويات القدرة، وفق النموذج اللوجستي ثلاثي المعلم (3PLM) والنموذج اللوجستي ثنائي المعلم (2PLM) لنظرية استجابة المفردة.

وسبب استخدام بيانات مولدة (Simulation) في الدراسة الحالية؛ لما توفره هذه البيانات من ظروف معيارية لا يمكن الحصول عليها في البيانات الحقيقية، مثل: توزيع مناسب لقدرات المفحوصين، و توزيعات مناسبة لمعالم المفردات، حيث اعتمدت الدراسة الحالية مستويات محددة للقدرة لتوليد بيانات عند كل مستوى من هذه المستويات، ومن الصعب تحديد هذه المستويات من القدرة وفق بيانات حقيقية، أي أن البيانات المولدة تساعد في السيطرة على الظروف التي يتم فيها اختبار فاعلية الإحصائي Lz^{*new} في الكشف عن أنماط استجابة المفحوص غير المطابقة بدلالة الخطأ من النوع الأول. علاوة على ذلك، إن البيانات المولدة توفر الوقت والجهد والتكلفة المادية، وخاصة في ظل الحاجة لعينات ذات أحجام كبيرة في نظرية استجابة المفردة.

خطوات حساب فاعلية الإحصائي (Lz^{*new}) في الكشف عن أنماط استجابة المفحوص غير المطابقة، بدلالة الخطأ من النوع الأول ، وفق النموذجين اللوجستي ثلاثي المعلم (3PLM) والنموذج اللوجستي ثنائي المعلم (2PLM) لنظرية استجابة المفردة، لعينة بلغت (١٠٠٠) مفحوص:

الخطوة الأولى: ١- تم استخدام البرمجية الإحصائية Wingen2 لتوليد معالم المفردات لاختبار مكون من (٤٠) مفردة وفقاً للنموذج اللوجستي ثلاثي المعلم (3PLM) وفقاً للشروط التالية:

أ. تتوزع معالم الصعوبة للمفردات توزيعاً طبيعياً (Normal) بمتوسط حسابي يساوي صفراً، وانحراف معياري يساوي ١.

ب. تتوزع معالم التمييز للمفردات توزيعاً منتظماً (Uniform) بحيث تتراوح القيم بين (٠,٥- ٢).

ت. تتوزع معالم التخمين توزيعاً طبيعياً بمتوسط حسابي يساوي ٠,١٤٨ وانحراف معياري يساوي ٠,٠٨.

ويبين الجدول (١) معالم المفردات المولدة وفق النموذج اللوجستي ثلاثي المعلم.

الجدول رقم (١)
معالم المفردات وفقاً للنموذج اللوجستي ثلاثي المعلم (3PLM)

Item	a	b	c	Item	a	b	c
١	٠,٦٥٠	-٠,٤٠١	٠,١٧٢	٢١	١,٢٣٨	-٠,٩٠٣	٠,٣٤٧
٢	١,٩٤٧	-١,٩٥٨	٠,٠٥٠	٢٢	١,٨٦٦	-٠,٨٨٤	٠,١١٧
٣	١,٤١٤	-٠,٢٩٢	٠,١١٤	٢٣	١,٧٦٧	٠,٦٨٢	٠,٢٠٣
٤	٠,٩٤٤	-٠,٥٥٨	٠,١٢٨	٢٤	٠,٦٦٥	٠,٤٤٩	٠,١١٨
٥	٠,٧٥٣	١,٢٤٨	٠,٠٨٦	٢٥	١,١٠٧	٠,١٦٩	٠,١٣٨
٦	١,٤٧٢	٠,٠٤٦	٠,٢٣٧	٢٦	١,٢٦٠	٠,٦٨٢	٠,١٦٢
٧	١,٤٧٥	-٠,٣١٨	٠,١٨٢	٢٧	١,٦٦٣	-٠,٥٨٨	٠,١٣٢
٨	٠,٥٥٦	٠,٥٣	٠,١٧٥	٢٨	١,٦٠٧	٠,٥٨٨	٠,٣٣٩
٩	١,٩١٧	٠,٢١٢	٠,٣٤١	٢٩	٠,٨٢٧	١,٠١٦	٠,٣٨٦
١٠	١,٢١٤	-٠,٦٢٢	٠,١٧٤	٣٠	١,٣٧٩	١,٩٥	٠,٣٢٤
١١	٠,٥٤٨	٠,٦٣٥	٠,٢١٤	٣١	٠,٨١٦	-٠,٢٩٢	٠,١٥٠
١٢	٠,٧٨٨	٠,٤٠٧	٠,٢٤٣	٣٢	٠,٧٨٩	٢,٣٧١	٠,٢٨٩
١٣	١,٩٢٨	-٠,٠٤٧	٠,٣٢٦	٣٣	٠,٦٥١	-١,٢٧٥	٠,٢٥٢
١٤	١,٤٩٤	-٠,٠٠٥	٠,١٩٦	٣٤	١,٢٦١	-٠,٩٤٩	٠,٣٨٥
١٥	٠,٧٢٥	٠,٨٨٢	٠,٢٥١	٣٥	١,٢٢٢	٠,٤٣٣	٠,٢٣٤
١٦	١,٩٧٩	-٠,٩٧٤	٠,٢٠٦	٣٦	٠,٥٤٧	-٢,٠٣٢	٠,١٤٩
١٧	٠,٩٠٨	-١,٥٥٧	٠,١٦٠	٣٧	١,٩٣٧	-٠,٣٠	٠,٢١٤
١٨	١,٤٩٠	٠,١٩٧	٠,١٥٧	٣٨	١,٧٣٢	٠,١٠٤	٠,١١٢
١٩	٠,٩٠٣	-٠,٢٦٣	٠,١٥٠	٣٩	١,٢٧٦	-٠,٠١٣	٠,١٨٧
٢٠	١,١٩٤	٠,٨٢٣	٠,٢٣٧	٤٠	١,٩١٨	٠,٢٥٢	٠,٣٠٦

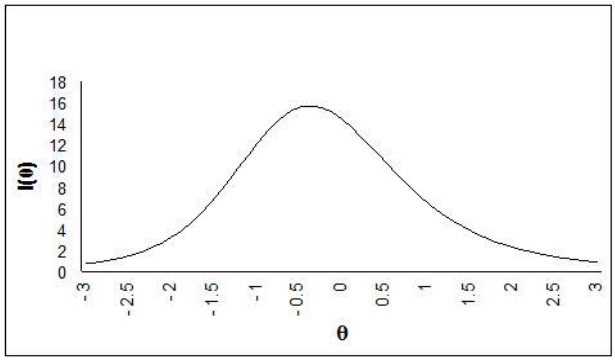
٢- كما استخدمت البرمجية الإحصائية نفسها Wingen2 لتوليد معالم المفردات لاختبار مكون من (٤٠) مفردة وفقاً للنموذج اللوجستي ثنائي المعلم (2PLM) وفقاً للشروط التالية:
أ. تتوزع معالم الصعوبة للفقرات توزيعاً طبيعياً (Normal). N(0, 1))، بمتوسط حسابي يساوي صفراً، وانحراف معياري يساوي ١.
ب. تتوزع معالم التمييز للفقرات توزيعاً منتظماً (Uniform) بحيث تتراوح القيم بين (٠,٥ - ٢).
ج. معلم التخمين يساوي صفراً.

ويبين الجدول (٢) معالم المفردات المولدة وفق النموذج اللوجستي ثنائي المعلم.

الجدول رقم (٢)
معالم المفردات وفقاً للنموذج اللوجستي ثنائي المعلم (2PLM)

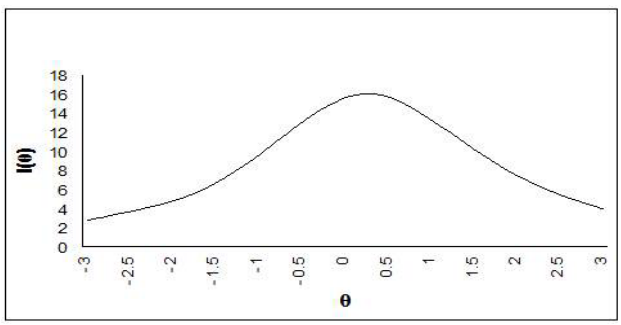
Item	a	b	Item	a	b
١	٠,٨٥٦	٠,٨٦١-	٢١	١,١٣١	٠,٠٥٦-
٢	٠,٤٨٣	٢,٠٤٥-	٢٢	٢,١١٥	٠,٧٠٢
٣	١,٦٥٧	١,٠٢٤	٢٣	١,٦٠٦	٠,١٥٣
٤	٢,٠١١	٠,٥٨٢-	٢٤	٠,٧٢٧	٠,٤٥٢-
٥	١,٣٨٤	١,٢١٥	٢٥	١,٣٢٧	١,٠٩٨
٦	٠,٥٦٩	٠,٠٢٣-	٢٦	٠,٦١١	١,٠٩٦-
٧	١,٣٧٦	٠,٨٦٤	٢٧	٠,٨٠١	٠,٧٤٢-
٨	٠,٥٦٠	٠,٢٦٩-	٢٨	١,٩٣٢	٠,٣١٨
٩	٢,٠٨٦	١,٤٣٢	٢٩	١,٢٠٢	٠,٠٨٨
١٠	٠,٥٩٤	٠,٥٢٦-	٣٠	١,٥١٥	٢,٧٢٣
١١	١,٣٥٥	٠,٧٧١	٣١	١,٩٩٧	٠,١٠٥-
١٢	٠,٨٣٨	٠,٨٥٠	٣٢	١,٤٩٢	٢,٩٨٦
١٣	٠,٨٧٤	٠,١١٠-	٣٣	١,٨٢٨	٠,٠٧٦-
١٤	٠,٥٠٣	٠,١٣٧-	٣٤	٠,٤٩٢	٠,٦١١-
١٥	١,١٦٨	٠,٠٢١	٣٥	٠,٧٤٠	٠,٥٢٨-
١٦	١,٩٥٣	٠,٣٢٢	٣٦	٠,٦٧٢	٠,٣٦٢-
١٧	١,٦٩٣	٠,٥٥٠	٣٧	٠,٦١٤	١,٠٠١
١٨	٢,٠١٤	٠,٧٢١-	٣٨	٠,٥٤٩	٠,٨٣٧
١٩	٢,١٣٥	٠,٢١٦	٣٩	١,٥٠٦	٠,٩٢٠-
٢٠	٠,٨٦٣	١,٤٤٨-	٤٠	٠,٩٤٣	٠,٤٩٦-

الخطوة الثانية: ١- تم حساب دالة معلومات الاختبار ($I(q)$) لـ ٤٠ مفردة، وفقاً للنموذج اللوجستي ثلاثي المعلم (PLM 3)، عند كل مستوى من مستويات القدرة التي تتراوح قيمها بين (-٣، ٣)، باستخدام المعادلات ٤، ١١ على الترتيب، كما هو موضح في الشكل (١).



الشكل (١): دالة معلومات الاختبار وفق النموذج اللوجستي ثلاثي المعلم (3 PLM)

٢- كما تم حساب دالة معلومات الاختبار ($I(q)$) لـ ٤٠ مفردة، وفقاً للنموذج اللوجستي ثنائي المعلم (2 PLM)، عند كل مستوى من مستويات القدرة التي تتراوح قيمها بين (-٣، ٣)، باستخدام المعادلات ٥، ١١ على الترتيب، كما هو موضح في الشكل (٢).



الشكل (٢): دالة معلومات الاختبار وفق النموذج اللوجستي ثنائي المعلم (2 PLM)

الخطوة الثالثة: وُلّدت معالم القدرة (θ) باستخدام برنامج Wingen2 لـ ١٠٠٠ مفحوص، حيث تتوزع القدرة توزيعاً طبيعياً بمتوسط حسابي يساوي صفراً وانحراف معياري يساوي ١. وفقاً للنموذجين (2PLM, 3PLM) على الترتيب.

الخطوة الرابعة: وُلدت استجابات ثنائية التدريج (٠، ١) باستخدام برمجية Wingen2 لـ ١٠٠٠ مفحوص على الـ ٤٠ مفردة، وفقاً للنموذج اللوجستي ثلاثي المعلم (3PLM)، في ضوء معالم المفردات التي تم توليدها في الجدول (١) ومعالم القدرة التي تمّ توليدها في الخطوة الثالثة. كما وُلّدت استجابات ثنائية التدريج (٠، ١) باستخدام البرمجية نفسها لـ ١٠٠٠

مفحوص على الـ ٤٠ مفردة، وفقاً للنموذج اللوجستي ثنائي المعلم (2PLM)، في ضوء معالم المفردات التي تم توليدها في الجدول (٢) ومعالم القدرة التي تمّ توليدها في الخطوة الثالثة.

الخطوة الخامسة: تم استخدام البرمجية الإحصائية Bilog-MG3 لتقدير قدرات المفحوصين $(\hat{\theta})$ باستخدام طريقة بيز (EAP) التي تفترض وجود توزيع طبيعي مسبق لقدرات المفحوصين، في كلا النموذجين.

الخطوة السادسة: تم استخدام البرمجية الإحصائية SPSS لحساب دالة معلومات الاختبار للقدرة المقدرة ($I(\hat{\theta})$) باستخدام المعادلات ٤، ١١ وفقا للنموذج اللوجستي ثلاثي المعلم (3PLM). وباستخدام المعادلات ٥، ١١ وفقاً للنموذج اللوجستي ثنائي المعلم (2PLM) ومن خلال البرمجية نفسها.

الخطوة السابعة: استخدمت البرمجية الإحصائية SPSS لتصحيح القدرة المقدرة ($\hat{\theta}^*$) باستخدام المعادلة (٩) في كلا النموذجين.

الخطوة الثامنة: تم حساب دالة معلومات الاختبار للقدرة المقدرة المصححة ($I(\hat{\theta}^*)$) باستخدام المعادلات (٤، ١١) وفقاً للنموذج اللوجستي ثلاثي المعلم (3PLM)، وباستخدام المعادلات (٥، ١١) وفقاً للنموذج اللوجستي ثنائي المعلم (2PLM)، ومن خلال البرمجية الإحصائية SPSS.

الخطوة التاسعة: تم حساب التباين التقريبي للقدرة المقدرة المصححة $(Var\ .(\hat{\theta}^*))$ باستخدام المعادلة (١٠) في كلا النموذجين.

الخطوة العاشرة: تم اعتماد (١١) مستوى من القدرة لـ ١٠٠٠ مفحوص تتراوح قيمها بين (-٣ ، +٣) بزيادة منتظمة مقدارها ٠,٦ لحساب الخطأ من النوع الأول عند كل مستوى من مستويات القدرة وعند ثلاثة مستويات مختلفة من الدلالة (α= 0.01, 0.05, 0.10)، في كلا النموذجين.

الخطوة الحادية عشرة: تم استخدام البرمجية الإحصائية Wingen2 لتوليد ٤٠٠٠ قدرة جديدة (θ^{new}) عند كل مستوى من مستويات القدرة التي تمّ اعتمادها في الخطوة السابقة، حيث إن هذه القدرات الجديدة تنتمي للتوزيع الجديد المعدّل ($\hat{\theta}^*$ ، $\frac{1}{1+I(\hat{\theta}^*)}$) N الذي متوسطه الحسابي يساوي $\hat{\theta}^*$ وتباينه التقريبي يساوي $\frac{1}{1+I(\hat{\theta}^*)}$ ، وفق كل نموذج.

الخطوة الثانية عشرة: وُلّدت استجابات ثنائية لقيم القدرة الجديدة التي تم توليدها في الخطوة السابقة، وفقاً للنموذج اللوجستي ثلاثي المعلم في ضوء معالم الفقرات التي تم توليدها في الجدول (١). وكذلك وُلّدت استجابات ثنائية لقيم القدرة الجديدة التي تم توليدها في

الخطوة السابقة، وفقاً للنموذج اللوجستي ثنائي المعلم (2PLM) في ضوء معالم الفقرات التي تم توليدها في الجدول (٢)، ومن خلال البرمجية نفسها.

الخطوة الثالثة عشرة: أعيدت الخطوات (٥، ٦، ٧) من أجل تقدير القدرات الجديدة التي تم توليدها $(\hat{\theta}_{new})$ وفقاً لطريقة بيز (EAP) وكذلك لحساب القدرة المقدرة المصححة الجديدة $(\hat{\theta}^*)$ لكل نمط استجابة تمّ توليده. لكلا النموذجين المستخدمين في الدراسة الحالية.

الخطوة الرابعة عشرة: تم حساب إحصائي مطابقة المفحوص (Lz^{*new}) باستخدام البرمجية الإحصائية R-Fit وفق المعادلة (١٢) لكل نمط استجابة تمّ توليده اعتماداً على القدرة المقدرة المصححة الجديدة. ولكل نموذج مستخدم. وتجدر الإشارة، إلى أن هذه البرمجية (R-Fit) تقوم بحساب إحصائيات مطابقة المفحوص على مستوى المفحوص والمفردة، للنماذج المختلفة بنوعيها اللوجستية والطبيعية، وتتطلب ثلاثة ملفات أساسية: ملفا يحوي على معالم المفردات، وملفا يحوي على معالم القدرة، وملفا يحوي على استجابات المفحوصين، وتستقبل هذه البرمجية مفردات ثنائية التدريج فقط، ويبلغ عدد المفردات التي تستقبلها ١٢٠ مفردة وعدد غير محدود من المفحوصين.

الخطوة الخامسة عشرة: استخدمت البرمجية الإحصائية SPSS لحساب الخطأ من النوع الأول عند كل مستوى من مستويات القدرة من خلال قيم الإحصائي L_z^{*new}، للنموذج اللوجستي ثلاثي المعلم (3PLM)، كما هو موضح في الجدول (٣) اعتماداً على القيمة الحرجة للإحصائي Lz حسب مستوى الدلالة. وكذلك استخدمت البرمجية الإحصائية نفسها لحساب الخطأ من النوع الأول عند كل مستوى من مستويات القدرة من خلال قيم الإحصائي L_z^{*new} للنموذج اللوجستي ثنائي المعلم (2PLM)، كما هو موضح في الجدول (٤) اعتماداً على القيمة الحرجة للإحصائي Lz حسب مستوى الدلالة. حيث تمّ اعتماد القيم الحرجة الآتية: القيمة الحرجة لـ Lz =٢,٣٣- عند المستوى (α = ٠,٠١)، والقيمة الحرجة لـ Lz =١,٦٥- عند المستوى (α = ٠,٠٥)، والقيمة الحرجة لـ Lz =١,٢٨- عند المستوى (α = ٠,١٠). وقد تمّ تصنيف المفحوص على أنه غير مطابق للنموذج إذا كانت قيمة L_z^{*new} المحسوبة أقل من القيمة الحرجة لـ LZ، كما تمّ تحديد الخطأ من النوع الأول في هذه الدراسة اعتماداً على الذيل السالب (Negative tail) للتوزيع، حيث إن القيم العالية الموجبة للإحصائي L_z تتضمن تطابقاً كبيراً مع النموذج أكثر مما هو متوقع (Torre & Deng, 2008).

نتائج الدراسة

أولاً: النتائج المتعلقة بالإجابة عن السؤال الأول

نص السؤال الأول على: "ما فاعلية الإحصائي Lz^{*new} في الكشف عن أنماط استجابة المفحوص غير المطابقة، بدلالة قدرته على ضبط الخطأ من النوع الأول، عند مستويات مختلفة من القدرة والدلالة الإحصائية (α= 0.01, 0.05, 0.10) في ضوء استخدام النموذج اللوجستي ثلاثي المعلم (3PLM)".

بعد تطبيق الخطوات السابقة، باستخدام البرمجيات الإحصائية المختلفة، المستخدمة في الدراسة الحالية، يُبيّن الجدول (٣) الخطأ من النوع الأول، عند (١١) مستوى مختلفا من القدرة (-٣، -٢,٤، -١,٨، -١,٢، -٠,٦، ٠، ٠,٦، ١,٢، ١,٨، ٢,٤، +٣) وثلاثة مستويات مختلفة للدلالة الإحصائية (α= 0.01, 0.05, 0.10)، في ضوء استخدام النموذج اللوجستي ثلاثي المعلم (3PLM)، لعينة من الفحوصين بلغت (١٠٠٠) مفحوص، على اختبار مكون من (٤٠) مفردة

الجدول رقم (٣)
الخطأ من النوع الأول ($\hat{\alpha}$) وفق النموذج اللوجستي ثلاثي المعلم (3PLM)
عند الاختلاف في مستوى القدرة ومستوى الدلالة (n= ١٠٠٠)*

n	(θ)	(α= 0.01)	(α= 0.05)	(α= 0.10)
١٠٠٠	-٣	٠,٠١٩	٠,٠٦٢	٠,١١٨
	-٢,٤	٠,٠١٧	٠,٠٦٠	٠,١١٦
	-١,٨	٠,٠١٥	٠,٠٥٨	٠,١١٦
	-١,٢	٠,٠١٣	٠,٠٥٧	٠,١١٤
	-٠,٦	٠,٠٠٧	٠,٠٤٦	٠,٠٩٦
	٠	٠,٠٠٩	٠,٠٥٢	٠,٠٩٨
	٠,٦	٠,٠٠٧	٠,٠٤٨	٠,٠٩٦
	١,٢	٠,٠٠٥	٠,٠٤٢	٠,٠٨٣
	١,٨	٠,٠٠٣	٠,٠٤١	٠,٠٨٠
	٢,٤	٠,٠٠١	٠,٠٣٩	٠,٠٧٧
	٣	٠,٠٠١	٠,٠٣٤	٠,٠٦١

*الأخطاء المعيارية المرتبطة بالقيم الاسمية ٠,١٠، ٠,٠٥، ٠,٠١ هي ٠,٠٠٤٣، ٠,٠٠٣٢، ٠,٠٠١٥ على الترتيب.

يتبين من الجدول (٣)، أن الخطأ من النوع الأول (المقدّر أو المحسوب، $\hat{\alpha}$) يقترب من المستوى الاسمي (α) عبر جميع مستويات القدرة المختلفة، وعند جميع مستويات الدلالة الثلاثة. كما نلاحظ، أن الخطأ من النوع الأول كان أكثر اقتراباً من المستوى الاسمي عند

مستويات القدرة المتوسطة (٠,٦ ، ٠ ، ٠,٦-=θ) عند مستويات الدلالة الإحصائية الثلاثة. فعلى سبيل المثال: عند مستوى القدرة ٠,٦-=θ ، كان الخطأ من النوع الأول يساوي (٠,٠٠٧، ٠,٠٤٦، ٠,٠٩٦) عند مستويات الدلالة الثلاثة (α= 0.01, 0.05, 0.10) على الترتيب. فيما نلاحظ أن الخطأ من النوع الأول، عند القدرات المتطرفة (Extreme)، يبتعد عن المستوى الاسمي، حيث كان أكبر من المستوى الاسمي (Overestimation) عند القدرات المتدنية (١,٨- ،٢,٤- ،٣-=θ) عند مستويات الدلالة الثلاثة. فعلى سبيل المثال: عند مستوى القدرة(٣-=θ)، كان الخطأ من النوع الأول يساوي (٠,٠١٩، ٠,٠٦٢، ٠,١١٨) عند مستويات الدلالة (α= 0.01, 0.05, 0.10) على الترتيب. وأمّا عند القدرات العالية (٢، ١,٨=θ ،٣،٤)، كان الخطأ من النوع الأول أقل من المستوى الاسمي عند جميع مستويات الدلالة الثلاثة. فعلى سبيل المثال: عند مستوى القدرة (١,٢=θ)، كان الخطأ من النوع الأول يساوي (٠,٠٠٥، ٠,٠٤٢، ٠,٠٨٣) عند مستويات الدلالة الثلاثة (α= 0.01, 0.05, 0.10) على الترتيب.

كما نلاحظ من الجدول رقم (٣)، أن الخطأ من النوع الأول كان أكثر اقتراباً من المستوى الاسمي عبر جميع مستويات القدرة وعند جميع مستويات الدلالة. فعلى سبيل المثال: عند مستوى الدلالة (α= 0.01)، كان الخطأ من النوع الأول ضمن خطأين معياريين من القيمة الاسمية ضمن مدى القدرة الذي يتراوح بين٠,٦- إلى ٠,٦.

ثانياً: النتائج المتعلقة بالإجابة عن السؤال الثاني

نص السؤال الثاني على: "ما فاعلية الإحصائي Lz^{*new} في الكشف عن أنماط استجابة المفحوص غير المطابقة، بدلالة قدرته على ضبط الخطأ من النوع الأول، عند مستويات مختلفة من القدرة والدلالة الإحصائية (α= 0.01, 0.05, 0.10) في ضوء استخدام النموذج اللوجستي ثنائي المعلم (2PLM) ؟"

بعد تطبيق الخطوات السابقة، باستخدام البرمجيات الإحصائية المختلفة، المستخدمة في الدراسة الحالية، وفي ضوء استخدام النموذج اللوجستي ثنائي المعلم (2PLM)، يُبيّن الجدول (٤) الخطأ من النوع الأول، عند (١١) مستوى مختلفا من القدرة (-٣، -٢,٤، -١,٨، -١,٢، -٠,٦، ٠، ٠,٦، ١,٢، ١,٨، ٢,٤، +٣) وثلاثة مستويات مختلفة للدلالة الإحصائية (α= 0.01, 0.05, 0.10) لعينة من الفحوصين بلغت (١٠٠٠) مفحوص، على اختبار مكون من (٤٠) مفردة.

الجدول رقم (٤)
الخطأ من النوع الأول ($\hat{\alpha}$) وفق النموذج اللوجستي ثنائي المعلم (2PLM)
عند الاختلاف في مستوى القدرة و مستوى الدلالة (n= ١٠٠٠)*

n	(θ)	(α= 0.01)	(α= 0.05)	(α= 0.10)
١٠٠٠	-٣	٠,٠٢٣	٠,٠٧٣	٠,١٢٦
	-٢,٤	٠,٠٢١	٠,٠٧١	٠,١٢٤
	-١,٨	٠,٠١٨	٠,٠٦٥	٠,١٢١
	-١,٢	٠,٠١٤	٠,٠٥٨	٠,١١٦
	-٠,٦	٠,٠٠٦	٠,٠٤٥	٠,٠٩٥
	٠	٠,٠٠٨	٠,٠٤٩	٠,٠٩٦
	٠,٦	٠,٠٠٦	٠,٠٤٥	٠,٠٩٥
	١,٢	٠,٠٠٣	٠,٠٣٩	٠,٠٧٧
	١,٨	٠,٠٠٢	٠,٠٣٨	٠,٠٧٤
	٢,٤	٠,٠٠٢	٠,٠٣٥	٠,٠٦١
	٣	٠,٠٠٠	٠,٠٣١	٠,٠٥٥

*الأخطاء المعيارية المرتبطة بالقيم الاسمية ٠,١٠، ٠,٠٥، ٠,٠١ هي ٠,٠٠٤٣، ٠,٠٠٣٢، ٠,٠٠١٥ على الترتيب.

يتبين من الجدول رقم (٤)، أن الخطأ من النوع الأول (المقدّر أو المحسوب، $\hat{\alpha}$) يقترب من المستوى الاسمي (a) عبر جميع مستويات القدرة عند جميع مستويات الدلالة. كما نلاحظ، أن الخطأ من النوع الأول كان أكثر اقتراباً من المستوى الاسمي عند مستويات القدرة المتوسطة (θ=-٠,٦، ٠ ، ٠,٦) وعند جميع مستويات الدلالة. فعلى سبيل المثال: عند مستوى القدرة θ=-٠,٦، كان الخطأ من النوع الأول يساوي (٠,٠٠٦، ٠,٠٤٥، ٠,٠٩٥) عند مستويات الدلالة (α= 0.01, 0.05, 0.10) على الترتيب. فيما تبين أن الخطأ من النوع الأول، عند القدرات المتطرفة (Extreme)، يبتعد عن المستوى الاسمي، حيث كان أكبر من المستوى الاسمي عند القدرات المتدنية (θ=-٣، -٢,٤، -١,٨) عند جميع مستويات الدلالة. فعلى سبيل المثال: عند مستوى القدرة (θ=-٣)، كان الخطأ من النوع الأول يساوي (٠,٠٢٣، ٠,٠٧٣، ٠,١٢٦) عند مستويات الدلالة (α= 0.01, 0.05, 0.10) على الترتيب. ونلاحظ أنه عند القدرات العالية (θ=١,٨، ٢، ٢,٤، ٣) كان الخطأ من النوع الأول أقل من المستوى الاسمي عند جميع مستويات الدلالة الثلاثة. فعلى سبيل المثال: عند مستوى القدرة (θ=١,٨)، كان الخطأ من النوع الأول يساوي (٠,٠٠٢، ٠,٠٣٨، ٠,٠٧٤) عند مستويات الدلالة (α= 0.01, 0.05, 0.10) على الترتيب.

كما نلاحظ من الجدول (٤)، أن الخطأ من النوع الأول كان أكثر اقتراباً من المستوى الاسمي عبر جميع مستويات القدرة المختلفة وعند جميع مستويات الدلالة الثلاثة. فعلى سبيل

المثال: عند مستوى الدلالة (α= 0.01)، كان الخطأ من النوع الأول ضمن خطأين معياريين من القيمة الاسمية ضمن مدى القدرة الذي يتراوح بين-٠,٦ إلى ٠,٦.

وبمقارنة الجدولين (٣ ،٤) معاً، نلاحظ أن الفرق بين الخطأ من النوع الأول والمستوى الاسمي وفق النموذج اللوجستي ثلاثي المعلم (3PLM) أقل من الفرق بين الخطأين وفق النموذج اللوجستي ثنائي المعلم (2PLM) عبر جميع مستويات القدرة وعند جميع مستويات الدلالة. فعلى سبيل المثال: عند مستوى القدرة (٣-=θ)، و مستوى الدلالة (α= 0.05)، كان الخطأ من النوع الأول يساوي (٠,٠٦٢) وفق النموذج اللوجستي ثلاثي المعلم (3PLM). وأمّا وفق النموذج اللوجستي ثنائي المعلم (2PLM)، فقد كانت قيمة الخطأ من النوع الأول تساوي (٠,٠٧٣).

ولمقارنة فاعلية النموذجين الثلاثي والثنائي المعلم (2PLM, 3PLM) في تحكّمهما بالخطأ من النوع الأول ($\hat{\alpha}$) ، فقد تمّ حساب الانحرافات المتوسطة المطلقة (Mean Absolute Deviations; MAD) للخطأ عن القيمة الاسمية لكل من النموذجين، بحسب اختلاف مستوى الدلالة، كما هو مبين في الجدول (٥) وفق المعادلة (١٥) الآتية (Torre & Deng, 2007)

$$MAD = \frac{\sum_{k=1}^{N} \left| \hat{\alpha}(\theta_k) - \alpha \right|}{N} \quad \text{.....(١٥)}$$

حيث، MAD : الانحرافات المتوسطة المطلقة، $\hat{\alpha}$: الخطأ من النوع الأول (المحسوب) عند مستوى القدرة q، q_k: مستوى القدرة (١١ مستوى)، a: المستوى الاسمي (٠,٠١، ٠,٠٥، ٠,١٠)، N : عدد مستويات القدرة (١١ مستوى).

الجدول رقم (٥)

الانحرافات المتوسطة المطلقة (MAD) للخطأ من النوع الأول ($\hat{\alpha}$) عن المستوى الاسمي (a) وفق النموذجين (3PLM, 2PLM)

n	α	MAD (3PLM)	MAD (2PLM)
١٠٠٠	٠,٠١	٠,٠٠٦	٠,٠٠٧
	٠,٠٥	٠,٠٠٨	٠,٠١٢
	٠,١٠	٠,٠١٤	٠,٠١٩

يتبين من الجدول رقم (٥)، أن الانحرافات المتوسطة المطلقة تزداد بزيادة مستوى الدلالة الإحصائية في كلا النموذجين الثلاثي والثنائي المعلم (2PLM, 3PLM). فعلى سبيل المثال:

وفق النموذج اللوجستي ثلاثي المعلم، كانت الانحرافات المتوسطة المطلقة تساوي (٠,٠٠٦، ٠,٠٠٨، ٠,٠١٤) عند مستويات الدلالة (α= 0.01, 0.05, 0.10) على الترتيب. و أمّا في النموذج اللوجستي ثنائي المعلم، كانت الانحرافات المتوسطة المطلقة تساوي (٠,٠٠٧، ٠,٠١٢، ٠,٠١٩) عند مستويات الدلالة (α= 0.01, 0.05, 0.10) على الترتيب.

كما يتبين من الجدول رقم (٥) نفسه، أن الانحرافات المتوسطة المطلقة وفق النموذج اللوجستي ثلاثي المعلم (3PLM) أقل من الانحرافات المتوسطة المطلقة وفق النموذج اللوجستي ثنائي المعلم (2PLM) وعبر جميع مستويات الدلالة الإحصائية. فعلى سبيل المثال: عند مستوى الدلالة (α= 0.05) ، كانت الانحرافات المتوسطة المطلقة تساوي ٠,٠٠٨ وفق النموذج اللوجستي ثلاثي المعلم، و تساوي ٠,٠١٢ وفق النموذج اللوجستي ثنائي المعلم.

وبناء على النتائج الموضحة في الجدول رقم (٥)، يمكن أن نستخلص أن الخطأ من النوع الأول أكثر اقتراباً (More Closeness) من المستوى الاسمي، وفق النموذج اللوجستي ثلاثي المعلم؛ لكونه أصغر انحرافات متوسطة مطلقة مقارنة بالنموذج اللوجستي ثنائي المعلم، عند جميع المستويات عند مستويات الدلالة (α= 0.10, 0.05, 0.01) على الترتيب.

مناقشة النتائج

أظهرت نتائج الدراسة، أن الخطأ من النوع الأول المحسوب أو المقدر ($\hat{\alpha}$) يكون أكثر اقتراباً (More Closeness) من المستوى الاسمي عند مستويات القدرة المتوسطة (θ=-0.6, 0, 0,6) مهما اختلف مستوى الدلالة، ومهما اختلف النموذج اللوجستي المستخدم (2PLM, 3PLM). ويبتعد أكثر عن المستوى الاسمي عند القيم المتطرفة لمستويات القدرة، حيث كان أكبر من المستوى الاسمي عند القدرات المتدنية (θ=-3, -2.4, -1.8)، و كان أقل من المستوى الاسمي عند القدرات العالية (θ=-3, -2.4, -1.8). إذ يُمكن أن تُفسّر هذه النتيجة اعتماداً على دالة معلومات الاختبار $I(\hat{\theta})$، ومعامل ثبات تقدير القدرة $\rho(\hat{\theta})$، حيث إن معامل ثبات تقدير القدرة يتناسب طردياً مع دالة معلومات الاختبار، حسب المعادلة (٧) في الدراسة الحالية، كما أن دالة معلومات الاختبار تتناسب عكسياً مع الخطأ المعياري في تقدير القدرة (Reeve, 2004). و بما أن دالة معلومات الاختبار تكون مرتفعة عند مستويات القدرة المتوسطة، لذلك يكون معامل الثبات في تقدير القدرة مرتفعاً عند هذه المستويات من القدرة، وبناءاً عليه فإن الخطأ المعياري في تقدير القدرة $(\hat{\theta})$ قليل جداً عند هذه المستويات من القدرة، وكما أن الخطأ في تصحيح تقدير القدرة $\hat{\theta}^*$ قليل، حيث إن القدرة

المصححة تتناسب عكسياً مع معامل ثبات تقدير القدرة، لذلك عند مستويات القدرة المتوسطة، كان الخطأ في تقدير هذه القدرات و تصحيحها قليلاً، وهذا ربّما يدل على أن التوزيع المعدل المصحح ($\hat{\theta}^*, \frac{1}{1+I(\hat{\theta}^*)}$) N عند هذه المستويات من القدرة كان يتصف بدقة عالية.

وعند توليد القدرات الجديدة ($4000\theta_{new}$) اعتماداً على التوزيع المعدل عند هذه المستويات، كان الخطأ في تقدير القدرات الجديدة و تصحيحها ($\hat{\theta}^*_{new}$) قليلاً، وكان معامل الثبات في تقدير القدرات الجديدة ($\hat{\theta}_{new}$) كبيراً. و هذا ربما جعل قيم الإحصائي L_z^{*new} التي تمّ حسابها اعتماداً على القدرات المصححة الجديدة تقترب بشكل كبير من القيم الحرجة المعيارية للإحصائي L_z التي تنتمي للتوزيع الطبيعي (0.1) N. و لهذا كان الخطأ من النوع الأول يقترب بشكل كبير من المستوى الاسمي عند هذه المستويات المتوسطة من القدرة، أي أن توزيع الإحصائي L_z كان أكثر اقتراباً من التوزيع الطبيعي المعياري عند هذه المستويات من القدرة. والعكس تماماً عند مستويات القدرة التي تبتعد عن المتوسط، سواء المستويات العالية للقدرة، أو المستويات المتدنية للقدرة.

وأظهرت نتائج الدراسة الحالية، أن الخطأ من النوع الأول المحسوب أو المقدر ($\hat{\alpha}$) يكون أكثر اقتراباً (More Closeness) من المستوى الاسمي وفق النموذج اللوجستي ثلاثي المعلم (3PLM) بأقل انحراف متوسط مطلق (MAD) مقارنة مع النموذج اللوجستي ثنائي المعلم (2PLM)، حيث إن الانحرافات المتوسطة المطلقة وفق النموذج اللوجستي ثلاثي المعلم أقل منها وفق النموذج اللوجستي ثنائي المعلم مهما اختلف مستوى الدلالة. و لكن هذه الانحرافات تقل كلما قلّ مستوى الدلالة. وهذا ربما يعود إلى أن البيانات التي تمّ توليدها وفق النموذج اللوجستي ثلاثي المعلم (3PLM) كانت تطابق النموذج بدرجة أفضل من البيانات التي تمّ توليدها وفق النموذج اللوجستي ثنائي المعلم (2PLM). و هذا ما أشار إليه توري ودنج (Torre & Deng, 2008) بأن الإحصائي L_z^{*new} يكون أداة قوية وفعّالة في مطابقة المفحوص عندما تكون البيانات مطابقة لنموذج نظرية استجابة المفردة (IRTM).

وتتفق نتائج الدراسة الحالية مع نتائج دراسة توري ودنج (Torre & Deng, 2008)، حيث كان الخطأ من النوع الأول يقترب كثيراً من المستوى الاسمي عبر جميع مستويات القدرة، وكان أكثر اقتراباً عبر مستويات القدرة المتوسطة. ولكن تباينت نتائج الدراسة الحالية مع نتائج دراسة سنجرز (Snijders, 2001) التي استخدمت القدرة المقدرة لحساب الإحصائي L_z، مع إجراء تصحيحات على تباين هذا الإحصائي، فقد كان الخطأ من النوع الأول وفق دراسة سنجرز يبتعد بشكل كبير عن المستوى الاسمي مقارنة مع نتائج الدراسة الحالية. وتتفق نتائج

الدراسة الحالية مع نتائج دراسة دودين (Dodeen, 2003) التي أظهرت أن توزيع الإحصائي L_z في النموذج اللوجستي ثلاثي المعلم كان قريباً جداً من التوزيع الطبيعي المعياري.

التوصيات والمقترحات

اعتماداً على نتائج الدراسة الحالية يوصي الباحث بما يلي:

١- استخدام إحصائي لوغاريتم الأرجحية المعياري القائم على أسلوب تصحيح تقدير القدرة وتوزيعها المرجعي Lz^{*new} في الكشف عن أنماط الاستجابة غير المطابقة، بدلالة ضبط الخطأ من النوع الأول، بغضّ النظر عن النموذج اللوجستي المستخدم لنظرية استجابة المفردة، في الاختبارات ثنائية التدريج، وأحادية البعد.

٢- إجراء المزيد من الدراسات حول فاعلية الإحصائي Lz^{*new} في الكشف عن أنماط الاستجابة غير المطابقة، بدلالة ضبط الخطأ من النوع الأول، وفق نظرية استجابة المفردة في ضوء اختلاف النموذج اللوجستي وتناول متغيرات أخرى لم تتطرق إليها هذه الدراسة مثل حجم العينة وطول الاختبار.

المراجع

Baker, F. (2001). *The basics of item response theory*. College park, MD, Wisconsin: University of Maryland.

Birenbaum, M. (1985). Comparing the effectiveness of several IRT based appropriateness measures in detecting unusual response patterns. *Educational and Psychological Measurement, 45,* 523-533.

Blais, J., & Raiche, G. (2005). *Characterization of the distribution of the Lz index of person fit according to the estimated proficiency level*. Paper presented at the International Meeting of the Psychometric Society,. Tilburg, Netherlands, Jul 2005, 1-28.

Brown, R., & Villareal, J. (2007). Correcting for person misfit in aggregated score reporting. *International Journal of Testing, 7*(1), 1-25.

Chan, D., & Schmitt, N. (1999). Correlates of person fit and effect of person fit on test validity. *Applied Psychological Measurement, 23(*1), 41-54.

Darabi, M., & Dodeen, H. (2009). Person – fit: relationship with personality tests in mathematics. *Research Paper in Education, 24*(1) 115-126.

Deng, W. (2007). *An innovative use of the standardized log-likelihood statistic to evaluate person-fit*. Dissertation submitted to the Graduate School-New Branswick, New Jersey.

Dodeen, H. (2003). The use of person fit statistics to analyze placement tests. *Educational Research Association,* Chicago, IL, 21-25.

Embretson, S., & Reise, S. (2000). *Item Response Theory for Psychologists*. New Jersey: Lawrence Erlbaum Associates.

Gruijter, D., & Kamp, L. (2005). *Statistical test theory for education and psychology*. Retrieved December 30, 2005 from: www.leidenuniv.nl /~ gruijterdnmde.

Hambleton, R., Swaminathan, H., & Rogers, J. (1991). *Fundamentals of item response theory*. Newbury Perk California: Stage publications.

Harwell, M. (1996). Monte Carlo studies in item response theory. *Applied Psychological Measurement, 20*(2), 112-128.

Karabatsos, G. (2003). Comparing the aberrant response detection performance of thirty – six person – fit statistics. *Applied Measurement in Education, 16*(4), 277-298.

Lopez, A., & Montesinos, H. (2005). Fitting rasch model using appropriateness measure statistics. *The Journal of Psychology, 8,* 11- 105.

Meijer, R. (1996). Person – fit research; an introduction. *Applied Measurement in Education, 9*(1), 3-8.

Meijer, R., & Sijtsma, K. (2001). methodology review: evaluating person fit. *Applied Psychological Measurement, 25*, 107-135.

Meijer, R., & Van Krimpen, S. (1999). The null distribution of person fit statistics for conventional and adaptive test. *Applied Psychological Measurement, 23,* 327-345.

Nering, M. (1998). The influence of non normal fitting examinee in estimating person parameter. *Applied Psychological Measurement, 27*, 135-147.

Reeve, B. (2004). *Application of item response theory (IRT)*. Modeling for building evaluating questionnaires measuring patient-reported outcomes. Retrieved from http://www.researchgate.net/puplication

Reise, S. (1995). Scoring method and the detection of person misfit in a personality assessment context. *Applied Psychological Measurement, 19*, 213-229.

Runder, L. (1996). The use of person– fit statistic with one high quality achievement test. *Applied Measurement in Education, 9*(1), 91-109.

Snijders, T. (2001). Asymptotic null distribution of person fit statistics with estimated person parameters. *Psychometrika, 66,* 331-342.

Torre, J., & Deng, W. (2008). Improving person fit assessment by correcting the ability estimate and its reference distribution. *Journal of Educational Measurement, 45*(2), 159-177.

Trabin, T., & Weiss, J. (1983). *The person response curve: fit of individuals to item response theory models*. In D.J. Weiss (Ed), new horizons in testing. New York: Academic Press.

Wainer, H. (2001). Augmented scores – "borrowing strength: to compute score based on small numbers of items. in d. thissen and wainer (Eds). *Test Scoring*, 343-3898. Mahwah, NJ: Erlbaum.

Zimowski, M., Mislevy, R., & Back, D. (1996). *Bilog-MG3: multiple- group IRT analysis & test maintenance for binary items*. Chicago: Scientific Software.