مجلة جامعة تشرين للبحوث والدراسات العلمية _ سلسلة الآداب والعلوم الإنسانية المجلد (37) العدد (6)2015
Tishreen University Journal for Research and Scientific Studies - Arts and Humanities Series Vol. (37) No. (6) 2015

الفروق في تقديرات معلم صعوبة الفقرة لبيانات محاكاة توائم النموذج التعويضي (MC1– PL) لنظرية استجابة الفقرة.

الدكتور إياد محمد حمادنه*
الدكتور وائل محمد الشرمان**

(تاريخ الإيداع 18 / 10 / 2015. قبل للنشر في 8 / 11 / 2015)

□ ملخّص □

هدفت الدراسة الحالية إلى الكشف عن الفروق في تقديرات معلم صعوبة الفقرة لبيانات محاكاة توائم النموذج التعويضي (MC1– PL)باختلاف عدد الأبعاد للاختبار وقوة الارتباط بين هذه الأبعاد واختلاف برمجية التقدير. واستخدمت في الدراسة بيانات محاكاة لعينة حجمها (1000)مفحوص تم توليدها باستخدام برمجية (RESGENT)، تقيس بعدين مرتبطين من الفقرات، تم من خلالهما خلق حالات مختلفة من تعددية البعد داخل الفقرات (ثلاثة أبعاد، بعدان، بعد واحد) بواقع (7)فقرات ثنائية التدريج لكل بعد، وعند ثلاثة مستويات مختلفة لقوة الارتباط بين الأبعاد (0.0, 0.50, 0.86) على الترتيب. وحللت البيانات باستخدام البرمجيات الإحصائية (NOHARM؛ Bilog– MG3).

أظهرت النتائج أن الفروق في تقديرات معلم صعوبة الفقرات المشكلة للاختبار متعدد الأبعاد داخل الفقرات لم تكن دالة إحصائيا عند مستوى الدلالة $(\alpha = 0.05)$ عند الحالات المختلفة لبعدية الأبعاد وباختلاف قوة الارتباط بين الأبعاد، وكذلك باختلاف برمجية التقدير (Bilog–MG3؛ NOHARM)، وكانت التقديرات بشكل عام مرتفعة ومتسقة. وتوصي الدراسة بضرورة الاعتماد على هاتين البرمجيتين الإحصائيتين في تحليل البيانات التي توائم النموذج، وخاصة عندما تتحقق الافتراضات التي اعتمدتها الدراسة الحالية في بيانات حقيقية.

الكلمات المفتاحية: نظرية الاستجابة للفقرة، أحادية البعد، النموذج التعويضي، النماذج متعددة الأبعاد.

* **أستاذ مشارك – كلية العلوم التربوية– جامعة آل البيت– المفرق – الأردن.**
** **أستاذ مساعد – كلية العلوم التربوية– جامعة آل البيت– المفرق – الأردن.**

مجلة جامعة تشرين للبحوث والدراسات العلمية _ سلسلة الآداب والعلوم الإنسانية المجلد (37) العدد (6)2015
Tishreen University Journal for Research and Scientific Studies - Arts and Humanities Series Vol. (37) No. (6) 2015

The Differences in Item Difficulty Parameter Estimations for Simulated Data having (MC1- PL) Model in Item Response Theory.

Dr. Iyad M. Hamadneh*
Dr. Wael M. Al-Shurman**

(Received 18 / 10 / 2015. Accepted 8 / 11 / 2015)

□ ABSTRACT □

This study aimed at finding out the differences in item difficulty parameter estimations for Simulated Data having (MC1- PL) Model in Item Response Theory (IRT) according to the differences in a test Dimensions (3D; 2D; 1D), correlations between these dimensions (0.0, 0.50, 0.86), and the statistical program used in analyzing the data (NOHARM ؛Bilog- MG3). The Monte Carlo simulations data having (MC1- PL) Model in IRT using (RESGENT) program; that fully filled in a 21-item was used to achieve the study aims. Data were analyzed using the statistical programs (NOHARM ؛Bilog- MG3).

The results revealed no statistically significant differences in item difficulty parameter estimations that construct the multidimensional test within items due to differences in a test, correlations between these dimensions, and the statistical program used in analyzing the data (NOHARM ؛Bilog- MG3), as well as the estimations were consistent and high. Finally, the study recommends using these statistical programs for a data having (MC1- PL) Model, especially when similar assumptions are satisfied in a real data.

Keywords: Item Response Theory, Unidimensionality, Compensatory Models, Multidimensional Models. .

* **Associate Professor, Faculty of Educational Sciences – Al al-Bayt University**
** **Associate Professor, Faculty of Educational Sciences – Al al-Bayt University**

مقدمة:

تعدّ النظرية الحديثة في القياس، أو ما يطلق عليها بنظرية استجابة الفقرة (Item Response Theory, IRT) إحدى أبرز التطورات الأساسية في مجال القياس النفسي والتربوي خلال العقد الماضي، والتي يُمكن من خلالها بناء وتطبيق المقاييس النفسية لتقييم خصائص الفقرات والمفحوصين بموجبها (Zhao, McMorris, Pruzek & Chen, 2002).

وتستند نظرية استجابة الفقرة (IRT) على عدة افتراضات، أهمها أحادية البعد (Unidimensionality)، وهو أكثر الافتراضات تعقيدا في نظرية استجابة الفقرة أحادية البعد، والذي يعني أن يقيس الاختبار سمة أو قدرة واحدة تفسر أداء المفحوصين عليها، وهذه السمة تعبّر عن وجود قدرة كامنة واحدة فقط، تفسر أداء المفحوصين على الفقرة، لذا يمكن تسميتها أحياناً بنظرية السمات الكامنة (Latent Trait Theory, LTT) (Hambleton, Swaminathan, & Rogers, 1991).

ونتج عن هذه النظرية مجموعة من النماذج تعرف باسم نماذج السمات الكامنة (Latent Trait Models)، تهدف إلى تحديد العلاقات بين أداء المفحوص في الاختبار والسمة التي تكمن وراء هذا الأداء وتفسّره. ونماذج نظرية استجابة الفقرة أحادية البعد، إما أن تكون لبيانات ثنائية التدريج (Unidimensional IRT models for dichotomous data)، أو أن تكون لبيانات متعددة التدريج(Unidimensional IRT models for polytomous data). وفيما يتعلق بنماذج نظرية استجابة الفقرة أحادية البعد للبيانات ثنائية التدريج فهي تلاءم البيانات الثنائية (Binary) التي يصنف الأداء عليها إلى ناجح أو راسب، أو يصنف إلى صح أو خطأ، ويوجد لها ميزتان، أولهما: مناسبتها لفقرات اختبار تحصيلي أو اختبار قدرات، وتناسب قوائم الشطب (Check lists) أو قوائم الشخصية والميول، بحيث تعطى الاستجابة إحدى قيمتين (0، 1) أي ثنائية التصحيح، وثانيهما: أنه يوجد بعض الحالات في الاستجابات متعددة التدريج يمكن تقليصها إلى ثنائية التدريج، مثلا في مقياس تقدير رباعي التدريج يمكن تقليصه إلى ثنائي التدريج؛ وذلك بدمج كل تدريجيين في واحد، هذا الوضع يناسب فقط الحالات التي يكون تدريجها زوجيا، وكذلك الحالات التي لا تؤدي إلى ضياع معلومات عن السمة الكامنة موضع القياس (كاظم والشرقاوي والشيخ وعبدالسلام، 1996).

ولنماذج نظرية الاستجابة للفقرة أحادية البعد ثنائية التدريج عدة أنواع، فمنها نماذج لوجستية أحادية أو ثنائية أو ثلاثية المعلم (صعوبة، تمييز، وتخمين)، إذ يوجد لكل منها معادلة خاصة تحسب فيها احتمالية الاستجابة الصحيحة على الفقرة. ففي النموذج اللوجستي أحادي المعلم أو ما يعرف بنموذج راش (One-Parameter Logistic Model, 1PL or Rasch Model) ، يتم حساب احتمالية الاستجابة الصحيحة لأداء فرد ما على فقرة معينة كما في المعادلة (1) الآتية:

$$P(X_{is}=1/\theta s, \beta i) = \frac{\exp(\theta_s-\beta_i)}{1+\exp(\theta_s-\beta_i)} \quad \dots (1)$$

حيث يمثل الرمز (θ_s) مستوى السمة الكامنة للمفحوص (s)، ويمثل الرمز (β_i) مستوى صعوبة الفقرة (i)، فيما تمثل $(\theta_s-\beta_i)$ الفرق بين مستوى القدرة للمفحوص (s)، ومعلم صعوبة الفقرة (i). أما في النموذج اللوجستي ثنائي المعلم (Two- Parameter Logistic, 2PL) فيتم إيجاد احتمالية الاستجابة الصحيحة لأداء فرد ما على فقرة معينة في هذا النموذج بإضافة معلم التمييز (Discrimination Parameter) للمعادلة السابقة، وفيما يتعلق بالنموذج اللوجستي ثلاثي المعلم (Three-Parameter Logistic Model, 3PL) يضاف في هذا النموذج معلم التخمين

(Guessing) أو ما يسمى بمعلم التقارب السفلي (Lower-asymptote) لمنحنى خصائص الفقرة (Hambleton & Swaminathan, 1985).

وبين ياو وشاوارتز (Yao & Shawartz, 2006) أن نماذج نظرية استجابة الفقرة أحادية البعد ومتعددة التدريج، تناسب الفقرات التي لا يمكن تصحيحها على أساس أنها ثنائية، ولها ميزتان، الأولى: أنها تعالج أنواع الفقرات بمرونة أكثر، إذ يمكن أن تستخدم نماذج متعددة التدريج مع أخرى ثنائية التدريج أي تجمع بين النوعين من النماذج. أما الميزة الثانية: فهي أنها تقدم معلومات أكثر من النماذج ثنائية التدريج المستخدمة للفقرات نفسها ، وبناء عليه، فإن زيادة المعلومات تؤدي إلى الزيادة في دقة تقدير قدرة المفحوص (θ).

ولنماذج نظرية الاستجابة للفقرة أحادية البعد متعددة التدريج عدة أنواع، هي: نموذج الاستجابة المتدرجة (Graded response model, GRM) الذي اقترحه سيمجيما (Samejima)، والنموذج المعدل لنموذج الاستجابة المتدرجة (Modified Graded Response Model, M-GRM) الذي طوره موراكي (Muraky) كتعديل لنموذج الاستجابة المتدرجة، بحيث يسهل استخدامه في تحليل الاستجابات على فقرات مقياس التقدير (Rating scale) كما في فقرات الاتجاهات، ونموذج التقدير الجزئي (Partial credit model, PCM) الذي طوره ماسترز (Masters, 1982) لتحليل فقرات الاختبار التي تحتاج الإجابة عنها إلى عدة خطوات ، ونموذج التقدير الجزئي المعمم (Generalized partial credit model, G-PCM) الذي طوره موراكي (Muraki) ليسمح باختلاف معلم الميل للفقرات ضمن المقياس الواحد (Lee & Terry, 2005).

وفي الواقع العملي، إن افتراض أحادية البعد لا يتحقق في جميع المواقف الاختبارية؛ فقد يكون في الاختبار مجموعة من الفقرات تقيس سمة معينة، ومجموعة أخرى من الفقرات تقيس سمة أخرى وهكذا، وبالتالي ينقسم الاختبار إلى مجموعة من الاختبارات الفرعية يتحقق في كل منها أحادية البعد، ويتم إيجاد معالم الفقرات ومعالم القدرة للمفحوصين لكل اختبار على حدة، لذلك لابد من التحقق من هذا الافتراض قبل استخدام النماذج أحادية البعد، وهنالك العديد من الطرق التي تستخدم للكشف عن البعدية لبيانات الاختبار، وأكثرها شيوعا تحليل المكونات الرئيسة، والتحليل العاملي الخطي، والتحليل العاملي غير الخطي (Ackerman, 1994; Walker & Beretras, 2000).

وفي حال عدم تحقق افتراض أحادية البعد لابد من اللجوء إلى نماذج نظرية الاستجابة للفقرة متعددة الأبعاد(Multidimensional Item Response Theory, MIRT). حيث عد ياو واشوارز (Yao & Schwarz, 2006) نظرية استجابة الفقرة متعددة الأبعاد بأنها تعميم لنظرية استجابة الفقرة أحادية البعد؛ إذ يمكن وصف خصائص المفحوص من خلال متجهة من المركبات يعبر عنها بالسمات. و تفترض هذه النماذج وجود بعدين أو أكثر يتم قياسهما في الاختبارات، ويزيد التعدد في الأبعاد من تطابق بيانات الفقرة عندما يظهر الاختلاف بين الأشخاص بشكل منتظم على الفقرات الصعبة أو السهلة (Meara , Robin & Sireci, 2000).

وتقسم النماذج متعدد الأبعاد (MIRT) إلى فئتين: الأولى، نماذج نظرية استجابة الفقرة متعددة الأبعاد الاستكشافية للبيانات ثنائية التدريج (Exploratory MIRT models for dichotomous data)، وتقدر معالم الفقرات والأشخاص على أكثر من بعد، للتحقق من تطابق النموذج مع البيانات، وهي تشبه التحليل العاملي الاستكشافي في تحديد عدد وطبيعة العوامل التي تكون أساسا في الأداء على الفقرة، ولا تشبه التحليل العاملي في أن بيانات مستويات الاستجابة على الفقرة تأخذ أحد نماذج نظرية الاستجابة للفقرة، بحيث تستخلص معلومات تامة من البيانات، وسمّي بالنموذج الاستكشافي؛ لأنه يكشف عن أفضل نموذج للمطابقة (McKinley and Way, 1992).

أما الفئة الثانية من النماذج متعددة الأبعاد (MIRT) فهي: نماذج نظرية استجابة الفقرة متعددة الأبعاد التوكيدية (Confirmatory MIRT) وتقدر المعالم لأبعاد معينة، وتتضمن أبنية تصاميم (Design structures) أو نماذج رياضية تربط الفقرات مع أبعاد محددة (علام، 2005). أو هي تقييم أي البيانات تتبع النموذج المفروض والمطور مسبقا على أساس المحتوى، ويعتبر هذا النوع من النماذج مهما عندما تكون الفروق بين الفقرات في المحتوى النظري، ومن الأمثلة على النماذج التوكيدية: نماذج للتعلم والتغيير (Models for learning and change)، ونماذج ببناءات مستوى سمة محدد (Models with specified trait level structures)، ونماذج لفئات متمايزة من الأفراد (Models for distinct classes of persons) (Embretson & Reise , 2000; Spencer , 2004).

ويعدّ نموذج راش متعدد الأبعاد (Multidimensional Rasch Model) من أشهر نماذج نظرية استجابة الفقرة متعددة الأبعاد الاستكشافية للبيانات ثنائية التدريج ، وأكثرها استخداما (Wang &Wang, 2003)، حيث يتم إيجاد احتمال الاستجابة الصحيحة لمفحوص ما على فقرة معينة ضمن نموذج راش كما في المعادلة (2)

$$P(X_{is}=1|\theta_s,\delta_i)=\frac{\exp\left(\sum_m \theta_{sm}+\delta_i\right)}{1+\exp\left(\sum_m \theta_{sm}+\delta_i\right)} \quad \ldots(2)$$

حيث: θ_{sm} السمة الكامنة للمفحوص s على البعد m ، δ_i مقطع السهولة (Easiness intercept) للفقرة i.

وعند مقارنة هذه المعادلة مع المعادلة (1) لنموذج راش البسيط أحادي البعد، يلاحظ أن مستوى السمة θs في المعادلة (2) يأخذ قيمة مركبة من عدة سمات بأوزان متساوية $\sum_m \theta_{sm}$ ، فقد بين ريكاس (Reckase as cited in Embretson & Reise, 2000) إلى أنه إذا أخذت الأبعاد أوزانا متساوية في كل فقرة، فإن نموذج راش أحادي البعد يطابق البيانات. وعلاوة على ذلك، لا يوجد مع الأوزان المتساوية نمط من الاستجابات على الفقرة يشير إلى تمايز موقع المفحوص على أبعاد مختلفة، لذلك فإن مستويات السمة المختلفة لا تقدر بشكل منفصل (Gruijter & kamp, 2005).

وأشار دراسغو وبارسونز (Drasgow & Parsons, 1983) إلى أن النماذج متعددة الأبعاد غير شائعة الاستخدام؛ وذلك لأن متطلبات تطبيقها كثيرة لتقدير معالم الفقرة والمفحوصين بشكل ملائم. وبالتالي فإن النماذج أحادية البعد هي الأكثر استخداما مما يؤدي إلى فقدان معلومات حول مستويات السمة للمفحوص.

وبين امبرتسون ورايس (Embretson & Reise, 2000) أن النماذج متعددة الأبعاد يمكن أن تصنف إلى نماذج تعويضية (Compensatory Models) ونماذج غير تعويضية (Non–Compensatory Models)، حيث أنه في النماذج التعويضية –قيد الدراسة الحالية– يستطيع المفحوصين الضعاف في سمة معينة أو أكثر تعويض هذا الضعف أو النقص من خلال قدراتهم الأخرى المرتفعة، بمعنى أكثر تحديد، هذه النماذج تسمح بالتفاعل بين السمات الكامنة أو القدرات متعددة الأبعاد، ولكن هذه الحالة لا يمكن تحقيقها في النماذج غير التعويضية. فيما ذكر كابار (Capar, 2000) أن المفحوص يعتمد في نماذج الاستجابة للفقرة متعددة الأبعاد (MIRT) على أكثر من قدرة للإجابة عن أسئلة الاختبار، وفي بعض الأحيان تتطلب أسئلة الاختبار مركبا لعدد من المهارات أو القدرات لكي يجيب المفحوص إجابة صحيحة.

ومن الجدير بالذكر، أن هناك العديد من البرمجيات الإحصائية التي تم تطويها لتقدير معالم الفقرات وقدرات المفحوصين، والمتعلقة بنظرية استجابة الفقرة أحادية البعد (IRT Software)، مثل: (BILOG MG; BIGSTEBS & WINSTEPS). وبالمقابل، فإن هناك برمجيات إحصائية أخرى تم تطويرها لتلاءم نماذج نظرية استجابة الفقرة

متعددة الأبعاد (MIRT Software)، مثل: (TESTFACT; NOHARM & MAXLOG) (الحجازي، 2010؛ Olivares, 2001; Wilson, 1999; Zimowski et al., 1996).

وعلى أية حال، فإن موضوع نماذج الاستجابة للفقرة متعددة الأبعاد (MIRT) وأثرها على تقدير معالم صعوبة الفقرات وقدرات المفحوصين من المواضيع التي يتناولها، حديثا، المختصون في مجال القياس والتقويم، وذلك من خلال العديد من الدراسات والبحوث:-

فقد أجرى داوادي (Dawadi, 1999) دراسة هدفت إلى استقصاء فاعلية نموذج الاستجابة للفقرة متعددة التدريج عند انتهاك افتراض أحادية البعد. واستخدم بيانات (1000) مفحوص يتمثل أداؤهم على اختبارات تقيس بعدين أو سمتين. وجرى تحليلها باستخدام التحليل العاملي بطريقة المكونات الرئيسة، وتم كذلك في الدراسة حساب التغيير في الأخطاء الناتجة عند انتهاك الافتراض على تقدير القدرة. أظهرت النتائج أنه عندما يقل الارتباط بين الأبعاد تزداد دقة تقدير القدرة للأفراد، وكذلك يزداد الارتباط مع البعد السائد.

وقام دنغ وأنسلي (Deng and Ansley, 2000) بدراسة هدفت إلى الكشف عن تعددية الأبعاد التعويضية وغير التعويضية باستخدام الإجراء الإحصائي (DIMTEST). واستخدم في الدراسة ثلاث عينات أحجامها (500، 1000، 2000)، وثلاثة اختبارات أطوالها (20، 40، 50) فقرة، وكان الارتباط بين القدرات مقدارها إما (0.3) أو (0.7) وباستخدام نموذجين هما، التعويضي وغير التعويضي. أظهرت النتائج ابتعاد البيانات أحادية البعد عن النموذج غير التعويضي عندما يكون حجم العينة كبيرا ، والارتباطات بين القدرات كانت منخفضة، كما بينت أن الإحصائي (DIMTEST) يعرّف البعدية بقوة أكثر عندما يزداد طول الاختبار وحجم العينة وعندما يكون الارتباط بين القدرات في حدها الأدنى.

وفي دراسة قام بها و وكر وبيرتراس (Walker & Beretras, 2000) هدفت إلى استخدام نماذج متعددة الأبعاد مقابل نماذج أحادية البعد في تقدير قدرة المفحوص في الرياضيات. طبقت الدراسة على طلبة الصفين الرابع والسابع وكان عدد المفحوصين (30000) تقريبا لكل صف، الذين تقدموا إلى اختبار يقيس القدرة الرياضية، بحيث تقدم طلبة الصف الرابع إلى (24) فقرة من نوع اختيار من متعدد و(16) فقرة من نوع الإجابة المصاغة، وطلبة الصف السابع إلى (30) فقرة اختيار من متعدد و(16) فقرة من نوع الإجابة المصاغة، حيث إن جميع الفقرات من نوع الإجابة المصاغة افترضت أن تكون متعددة الأبعاد؛ لأنها تحتاج إلى أكثر من قدرة رياضية. أظهرت النتائج أن استخدام نموذج أحادي البعد لبيانات متعددة الأبعاد يعتبر استنتاجا غير صحيح لتوزيع قدرات المفحوصين على الأبعاد الثانوية.

فيما أجرى غوس ووكر (Gosz & Walker, 2002) دراسة هدفت إلى المقارنة بين البرمجيتين TESTFACT & NOHARM في تحليل بيانات تمثل استجابة لاختبارات مؤلفة من فقرات متعددة الأبعاد. واستخدمت في الدراسة ستة اختبارات تكون كل منها من (40) فقرة ثنائية التدريج، ثلاثة اختبارات منها تألفت من (12) فقرة ثنائية البعد، و (28) فقرة أحادية البعد، بينما تألفت الثلاثة اختبارات الأخرى من (24) فقرة ثنائية البعد و(16) فقرة أحادية البعد. وطبقت أكثر من مرة على عدد كبير من المفحوصين. أظهرت النتائج أنه يفضل استخدام برمجية NOHARM عندما يكون الارتباط بين الأبعاد (θ_1 & θ_2) صغيرا وعندما يحتوي الاختبار أعدادا قليلة من الفقرات متعددة الأبعاد، بينما يجب استخدام برمجية TESTFACT عندما تكون الارتباطات بين الأبعاد (θ_1 & θ_2) كبيرة وعندما يحتوي الاختبار على عدد كبير من الفقرات متعددة الأبعاد.

وفي دراسة قام بها وانغ ووانغ (Wang &Wang, 2003) هدفت إلى معرفة أثر تقدير معالم أحادية البعد المشتق من بيانات ثنائية البعد أو ثلاثية البعد على تقديرات معالم نموذج راش أحادي المعلم متعدد الأبعاد. حيث عالجت الدراسة خمسة عوامل هي: بعد السمة (بعد واحد، اثنان، ثلاثة)، وحجم عينة المفحوصين (500، 1000، 2000)، ودرجة الارتباط بين أبعاد السمة الثلاثة (0 ، 0.3 ، 0.7)، والمهمة التي يقوم بها المفحوص وتتضمن مستويين (40، 80)، والمقدر (مقدر واحد، مقدرين، ثلاثة مقدرين). ومن أبرز النتائج التي توصلت إليها الدراسة أن عدم تحقق افتراض أحادية البعد يؤثر على تقدير المعالم.

وأجرى اسبنسر (Spencer, 2004) دراسة هدفت إلى مقارنة تقديرات معالم الفقرة لبيانات محاكاة توائم النموذج التعويضي متعدد الأبعاد والنموذج الأحادي المعلم. وأظهرت النتائج دقة النموذج التعويضي في تقدير معالم صعوبة الفقرات وقدرات المفحوصين للبيانات متعددة الأبعاد، وضرورة التأكد من أحادية البعد عند استخدام النموذج أحادي المعلم لنظرية استجابة الفقرة.

وأجرى فنش (Finch, 2006) دراسة هدفت إلى المقارنة بين استخدام طريقتين في التدوير هما، التدوير العمودي (Varimax)، والتدوير المائل (Promax) في الكشف عن بنية العوامل في فقرات ثنائية التدريج. أظهرت النتائج أن طريقتي التحليل العاملي تتضمن مصفوفات للتشبعات تستخدم لربط الفقرات بالعوامل، وأن كلا الطريقتين متكافئتان في الكشف عن البنية الأصلية الأساسية للفقرات، كما أظهرت النتائج أنه بالرغم من وجود ارتباط بين العوامل في التدوير المائل، إلا أن التحليل العاملي بالتدوير المائل يع د الأفضل في تحقيق التراكيب البسيطة (Simple structure)، كما أظهرت النتائج أن الطريقتين فعالت ان في تحديد أي الفقرات ترتبط مع أي العوامل مع الأفضلية لطريقة التدوير المائل.

وأجرى أبو شندي (2011) دراسة هدفت إلى تعرف أثر عدد أبعاد الاختبار قوة الارتباط بينها وتوزيع قدرة المفحوصين على تقدير العلامة الحقيقية، واستخدم بيانات تم توليدها لاستجابات (4000) مفحوص. وكشفت النتائج عن وجود اختلافات في تقدير العلامة الحقيقية تعزى لاختلاف عدد أبعاد الاختبار وتوزيع قدرة المفحوصين، وعدم تأثرها باختلاف قوة الارتباط بين الأبعاد (0.9 ;0.2). وأوصى الباحث بإجراء دراسات تتناول حالة تساوي عدد الفقرات التي تشكل كل بعد من أبعاد الاختبار، وحالة تعددية البعد داخل الفقرات.

وقامت وايبرغ (Wiberq, 2013) بدراسة لتقييم العواقب المحتملة من استخدام نماذج نظرية استجابة الفقرة التي تفترض أحادية البعد على بيانات اختبار قبول في الجامعة متعدّد الأبعاد ومكون من خمسة أبعاد، مرة على افتراض أنه يقيس بعد واحد، ومرة على افتراض أنه يقيس أكثر من بعد. أظهرت النتائج أن تقديرات معالم الفقرات متعددة الأبعاد كانت متقاربة من تقديرات معالم الفقرات أحادية البعد.

وبناء على ما تقدم، يتضح الاهتمام الكبير الذي توليه الدراسات في موضوع النماذج متعددة الأبعاد لنظرية الاستجابة للفقرة وأثرها في تقدير معالم الفقرات وقدرات المفحوصين، وما له من دور كبير في تطوير المقاييس والاختبارات في معظم المجالات. فقد تناولت الدراسات أنواعاً مختلفة من الاختبارات: منها ما هو ثنائي التدريج، ومنها ما هو متعدد التدريج، ومنها ما هو من إعداد الباحثين، ومنها ما تم توليد بياناته حاسوبياً، لتلاءم أحد افتراضات نماذج نظرية الاستجابة للفقرة. ومن جهة أخرى، ركزت بعض الدراسات على استخدام نماذج تطابق البيانات، وبعضها استخدم نماذج لا تطابق البيانات ودرس مدى تأثر كل منها على تقدير معالم الفقرات وقدرة الأفراد، والبعض الآخر تعرض إلى استخدام البرامج الحاسوبية وأثر كل منها كذلك على تقدير معالم الفقرات وقدرة الأفراد. و اتفقت غالبية نتائج الدراسات

حول وجوب استخدام النماذج متعددة الأبعاد لإيجاد معالم فقرات متعددة الأبعاد. أضف إلى ذلك، قلة الدراسات العربية –في حدود علم الباحثَين– التي تناولت تحديداً موضوع نموذج الاستجابة للفقرة متعدد الأبعاد أحادي المعلم . من هنا، تأتي هذه الدراسة في محاولة للكشف عن الفروق في تقديرات معلم صعوبة الفقرة لبيانات محاكاة توائم النموذج التعويضي (MC1– PL) باختلاف عدد الأبعاد (أحادي، ثنائي، ثلاثي) للاختبار وقوة الارتباط بين هذه الأبعاد (0.0، 0.5، 0.86) واختلاف برمجية التقدير (NOHARM; Bilog–MG3).

مشكلة الدراسة وأسئلتها:

إن الافتراض الشائع لنماذج نظرية استجابة الفقرة (IRT) هو قياس قدرة واحدة، أو بعد واحد بمجموعة من الفقرات، لكن هذا الافتراض في كثير من الأحيان لا يتحقق بشكل كامل؛ بسبب وجود العديد من العوامل التي تؤثر في الأداء على الاختبار، منها ما هو ذو طبيعة معرفية أو أنها ذات صلة ببعض الخصائص الشخصية وبعض الأنماط الخاصة بسلوك المفحوصين أثناء الاستجابة للاختبارات.

وفي كثير من الأحيان يستخدم الباحثون نماذج لنظرية الاستجابة للفقرة لا تطابق البيانات من حيث طبيعتها وعدد أبعادها، عند إيجادهم لمعالم الفقرات ومعالم القدرة للأفراد، دون الالتفات إلى افتراضات البرمجيات الحاسوبية التي يستخدمونها في التحليل، مما يؤدي إلى استخراج معالم مضللة وغير حقيقية للفقرات والأفراد، لا تعكس خصائص الفقرات والأفراد الحقيقية، وبالتالي الوصول إلى نتائج غير صحيحة خاصة عند استخدام نموذج يفترض أحادية البعد لبيانات متعددة الأبعاد (Luecht &Miller, 1992; Walker & Beretras, 2000). لذا أتت هذه الدراسة للكشف عن الفروق في تقديرات معلم صعوبة الفقرة لبيانات محاكاة توائم النموذج التعويضي (MC1– PL) باختلاف عدد الأبعاد للاختبار وقوة الارتباط بين هذه الأبعاد واختلاف برمجية التقدير. . وبالتحديد تسعى الدراسة الحالية للإجابة عن الأسئلة الآتية :

1 ـ هل تختلف تقديرات معلم صعوبة الفقرة لبيانات توائم النموذج التعويضي (MC1– PL) باختلاف عدد الأبعاد المكونة للاختيار وقوة العلاقة الإرتباطية بينها؟

2 ـ هل توجد فروق ذات دلالة إحصائية عند مستوى الدلالة ($\alpha = 0.05$) في متوسطات تقديرات معلم صعوبة الفقرة لبيانات توائم النموذج التعويضي (MC1– PL) تعزى لاختلاف البرمجية الإحصائية المستخدمة في التحليل؟

أهمية البحث وأهدافه:

تهدف الدراسة الحالية إلى الكشف عن الفروق في تقديرات معلم صعوبة الفقرة لبيانات محاكاة توائم النموذج التعويضي (MC1– PL) لنظرية استجابة الفقرة (IRT) باختلاف عدد الأبعاد (أحادي، ثنائي، ثلاثي) للاختبار وقوة الارتباط بين هذه الأبعاد (0.0، 0.5، 0.86) واختلاف برمجية التقدير (NOHARM; Bilog–MG3).

أهمية البحث:

تظهر أهمية الدراسة الحالية في الجوانب الآتية:

1. البحث في التحقق من التقدم الكبير الذي أحرزته نظرية الاستجابة للمفردة (IRT) ونماذج نظرية الاستجابة للفقرة متعددة الأبعاد (MIRT) والتي يمكن خلالها وصف خصائص المفحوص من خلال متجهة من المركبات يعبر عنها بالسمات، وتفترض هذه النماذج وجود بعدين أو أكثر يتم قياسهما في الاختبارات.

2. تساهم في تسليط الضوء على البعدية للبيانات الاختبارية، وقوة الارتباط بينها على تقديرات معلم صعوبة الفقرة لنماذج نظرية استجابة الفقرة في محاولة لتزويد علماء النفس والقياس وأصحاب القرار، بمنهجية علمية يمكن من خلالها التعرف على الفروق في تقديرات معلم صعوبة الفقرة لبيانات محاكاة توائم النموذج التعويضي (MC1– PL) لنظرية استجابة الفقرة.

4. تسليط الضوء على ضرورة توفير البرمجيات الإحصائية الحديثة (NOHARM; Bilog–MG3) للكشف عن الفروق في تقديرات معلم صعوبة الفقرة في الاختبارات متعددة الأبعاد داخل الفقرات.

مصطلحات البحث :

- **اختبار أحادي البعد (Unidimensional Test)**: اختبار يقيس قدرة (سمة) واحدة من خلال مجموعة الفقرات المشكلة (المكونة) لهذا الاختبار.
- **اختبار متعدد الأبعاد داخل الفقرات التعويضي (Within Items Multidimensional Test)** : اختبار يقيس سمتين أو أكثر، ويحتوي على مجموعة من الفقرات التي تقيس كل منها سمة أو أكثر. ويتطلب من المفحوص أكثر من قدرة حتى يستجيب على الفقرات بشكل صحيح.
- **معلم صعوبة الفقرة أحادية البعد :** مستوى القدرة الذي يقابل احتمال 0.50 للإجابة عن الفقرة إجابة صحيحة عندما يكون معامل التخمين مساوي صفر، وتمثل نقطة انعكاس على منحنى خصائص الفقرة. وتأخذ صعوبة الفقرة في نظرية استجابة الفقرة قيماً بين (−∞، +∞) ولكن بشكل عام، فإن القيم التي تأخذها تكون بين (3− و 3).
- **معلم صعوبة الفقرة متعددة الأبعاد :** هي النقطة على متصل القدرة التي يكون عندها سطح منحنى خصائص الفقرة أشد انحدارا، وأعلى نقطة تمييزية عند تلك النقطة.
- **النموذج التعويضي (MC1– PL)**: هو نموذج يستطيع المفحوصين الضعاف في سمة معينة أو أكثر تعويض ضعفهم من خلال قدراتهم الأخرى المرتفعة، بمعنى يسمح بالتفاعل بين السمات الكامنة أو القدرات متعددة الأبعاد.

محددات البحث:

1 – اقتصرت الدراسة على استخدام النماذج التعويضية، ولم تستخدم النماذج غير التعويضية.

2 – اقتصرت الدراسة على ثلاث حالات للبعدية (ثلاثي، ثنائي، أحادي)، عند ثلاثة مستويات لقوة الارتباط بين الأبعاد (0.00, 0.50, 0.86).

3 – اقتصرت الدراسة على اختبارات ثنائية التدريج (0, 1)، وموزعة بعدد متساو من الفقرات في ثلاثة أبعاد بواقع (7) فقرات لكل بعد.

4 – اقتصرت الدراسة على استخدام البرمجيات الإحصائية الحديثة (NOHARM ؛ Bilog– MG3) في تحليل البيانات.

منهجية البحث وإجراءاتها:

أسلوب توليد البيانات

أصل هذه الدراسة مبني على دراسة Monte Carlo التي تمثل الواقع الحقيقي لتقييم الرياضيات الذي يشمل بعدين مرتبطين هما: بعد العمليات الأساسية الضرورية، وبعد العمليات الحسابية. بعد العمليات الأساسية الضرورية تم تمثيله على محور الصادات (y-axis) والتي تمثل فقراته تقييم قدرة المفحوص على التعامل بشكل صحيح مع العمليات الحسابية م ثل الجمع والطرح والضرب والقسمة والأس، أما البعد الثاني فتم تمثيله على محور السينات (x-axis) والتي تمثل فقراته ترتيب المفحوص لتلك العمليات الحسابية بشكل صحيح، مع العلم أن الفقرات التي تم تمثيلها بالكامل على محور البعد (y-axis) لا تتطلب أي من العمليات الحسابية (x-axis)، وكذلك العكس تماماً. يعني هذا أن الفقرات التي تمثل كل بعد مستقلة بحد ذاتها. ولكن هنالك فقرات لا تمثل بالكامل أي من البعدين السابقين وتقع في المجال بين المحورين اللذان يمثلان البعدين.

ومن أجل تبسيط الدراسة، تم افتراض أن هذه الفقرات يمكن تمثيلها على محور بالمنتصف بين المحورين السابقين سمي هذا المحور بالمحور (z-axis). ولتحقيق هدف الدراسة تم افتراض أن معامل الارتباط بين البعدين (x-axis) و (y-axis) هو (0.50) وهو ما يعادل الزاوية 30 بين البعدين عند تمثيلهما على مستوى المحاور المائلة، وهذه الزاوية عبارة عن arc-cosine of 0.50. أي أن البيانات المتعلقة بالفقرة والمفحوص مثلت على مستوى مائل الزاوية بين بعدية هي 60 درجة، وبما أنه تم افتراض أن المحور (z-axis) يتوسط هاذين البعدين، أي يصنع زاوية مقدارها 30 درجة مع كل منهما ، الأمر الذي يوضح أن معامل الارتباط بين البعد (z-axis) وأي من البعدين الآخرين هو (0.86) الذي هو عبارة عن cosine of 30.

وبناء على ما سبق، إن معامل صعوبة الفقرات ومستوى قدرة المفحوص مثلت على الأبعاد الثلاثة، وإن أسئلة الدراسة تتطلب القيام بعملية إسقاط البيانات للمعالم من أحد الأبعاد على الأبعاد الأخرى (إسقاط الأبعاد على مستوى مشترك)، واستخدمت الخواص المثلثية لحساب تحويلات عملية الإسقاط من بعد على البعدين الآخرين وفقا للطريقة التي اعتمدها اسبنسر (Spencer, 2004) والتي بين من خلالها كيفية البناء المنطقي لمجموعة البيانات التي تناسب هذا النوع من الدراسات.

توليد البيانات

تم توليد البيانات في الدراسة الحالية باستخدام البرمجية الإحصائية (RESGENT) (Muraki, 2000)، حيث يمكن من خلال هذه البرمجية توليد استجابات ثنائية التدريج لاختبارات أحادية البعد (Unidimensionality)، واستجابات ثنائية التدريج لاختبارات متعددة الأبعاد (Multidimensionality)، وتتطلب هذه البرمجية تحديد المعالم الحقيقية للفقرات، وخصائص توزيع قدرة المفحوص، كما يمكن من خلالها توليد استجابات لاختبارات متعددة الأبعاد مع الأخذ بعين الاعتبار قوة ونوع العلاقة الارتباط بينها، والتي يتم حسابها اعتمادا على قيم القدرة الحقيقية المولّدة.

وتحتاج هذه البرمجية – كغيرها من برامج توليد البيانات- إلى تزويدها بالقيم الحقيقية لمعالم الفقرات (True Item Parameters) وخصائص توزيع معلمة القدرة للمفحوصين (θ)، حتى تقوم بتوليد الاستجابات (Responses) لأكثر من مجموعة من المفحوصين، وفق النماذج التعويضية المستخدمة في الدراسة، والتي تحتاج إلى معلم واحد ليشير إلى صعوبة كل فقرة من فقرات الاختبار؛ كون الدراسة الحالية اقتصرت على نموذج راش (أحادي

المعلم)، حيث تأخذ صعوبة الفقرة في نظرية استجابة الفقرة قيماً بين (− ∞، +∞) ولكن بشكل عام، فإن القيم التي تأخذها تكون بين (3− و 3) (Hambleton, Swaminathan & Rogers, 1991). أما معالم القدرة (θ) فقد افترض الباحثان أن قدرات المفحوصين تتوزع توزيعا طبيعيا بمتوسط حسابي يساوي صفر وانحراف معياري يساوي واحد صحيح.

وبناء على هذه البرمجية تم توليد 21 فقرة، موزعة بالتساوي على ثلاثة أبعاد (x, y, z) بواقع 7 فقرات لكل بعد، وروعي في بنائها أسلوب الإسقاط المنطقي الذي أشار إليه اسبنسر (Spencer, 2004)، وتم توليد استجابات 1000 مفحوص لـ 21 فقرة باستخدام القيم الحقيقية لمعالم الصعوبة نفسها، وكانت قوة العلاقة الإرتباطية بين الأبعاد (0.0، 0.50، 0.86).

التحقق من بعدية البيانات

استخدمت البرمجية الإحصائية (Bilog–MG3) (Zimowski et al., 1996) لتقدير معلم الصعوبة (β_i) وقدرات المفحوصين، وفق النموذج اللوجستي أحادي المعلم أو ما يعرف بنموذج راش (One–Parameter Logistic Model, 1PL or Rasch Model) المستخدم في الدراسة الحالية لنماذج نظرية استجابة الفقرة أحادية البعد (IRT).

أما فيما يتعلق بنماذج نظرية استجابة الفقرة متعددة الأبعاد (MIRT): تعدّ البرمجية الإحصائية (NOHARM) واحدة من البرمجيات الأكثر استخداما على نطاق واسع لتقدير معالم الفقرات الثنائية (binary) في الاختبارات متعددة الأبعاد (Multidimensionality) لنظرية استجابة الفقرة، وفي سياق الاختبارات التربوية يشار إليها في النموذج التراكمي الصاعد متعدد الأبعاد (Multidimensional Normal Ogive model, MNO) (Olivares, 2001)، وتستخدم لتفحص بعدية البيانات، وذلك بالتحقق من المواءمة (Harmonization) بين عدد أبعاد الاختبار والبيانات من خلال إجراء تحليل البواقي (Residual Analysis)، وحساب الجذر التربيعي لمتوسطات مربعات البواقي (Root Mean Square of Residuals, RMSR).

وللتحقق من بعدية البيانات، فقد بين ماكدونالد (McDonald, 1997) أن عدد أبعاد الاختبار– وفق هذه البرمجية (NOHARM)– توائم البيانات عندما تكون (RMSR > 0.003). ويبين الجدول رقم (1) قيم الجذر التربيعي لمتوسطات مربعات البواقي (RMSR) لتعدّدية البعد داخل (Within) الفقرات [(ثلاثي؛ 3D)، (ثنائي؛ 2D)، أحادي؛ 1D)] باختلاف قوة الارتباط بين الأبعاد (0.00، 0.50، 0.86).

جدول رقم (1)

قيم الجذر التربيعي لمتوسطات مربعات البواقي (RMSR) لتعددية البعد داخل الفقرات باختلاف قوة العلاقة الارتباطية بين الأبعاد .

قوة العلاقة / أبعاد الاختبار	0.00	0.50	0.86
3D	0.0017	0.0021	0.0024
2D	0.0013	0.0018	0.0019
1D	0.0016	0.0017	0.0018

يتبين من الجدول رقم (1)، أن جميع قيم الجذر التربيعي لمتوسطات مربعات البواقي (RMSR) كانت أقل من (0.003)، ووفقا لما ورد عن ماكدونالد (McDonald, 1997) فإن جميع الاختبارات تحقق افتراض أحادية البعد.

تقدير معالم الفقرات وقدرات المفحوصين

تم إجراء تقديرات لمعالم صعوبة فقرات الاختبار (21) فقرة، مرة على افتراض أنها تقيس ثلاث سمات كامنة (3D) (21 Items = z +y +x)، ومرة على اعتبار أنها تقيس سمتين كامنتين (2D) (14 =y +x)؛ (14 =z +y)؛ (14 =z +x)، ومرة على اعتبار أنها تقيس سمة كامنة واحدة (أحادية البعد؛ 1D) (7 =z ;7 =y ;7 =x). وتجدر الإشارة إلى أن عملية تقدير المعالم تمت في الحالات الثلاث السابقة، تارة باستخدام برمجية (Bilog–MG3) وتارة أخرى باستخدام برمجية (NOHARM)، ثم تم حساب معاملات الارتباط بين القيم الحقيقية والقيم المقدرة لمعالم صعوبة الفقرات باستخدام البرمجيتين عند اختلاف قوة الارتباط بين الأبعاد المعتمدة في الدراسة الحالية (0.00، 0.50، 0.86).

بعد تقدير معالم صعوبة الفقرات لأبعاد الاختبار، سواء كانت هذه الفقرات تمثل اختبار أحادي البعد، أو تمثل اختبار متعدد الأبعاد، فقد تم حساب معاملات ارتباط سبيرمان (Spearman Correlation) بين القيم الحقيقية والقيم المقدرة لمعالم صعوبة الفقرات عند اختلاف قوة الارتباط بين الأبعاد (0.00، 0.5، 0.86)، وفق البرمجيتين السابقتين.

وللكشف عن الفروق في تقديرات معلم صعوبة الفقرة لنموذج راش أحادي البعد ومتعدد الأبعاد وفق اختلاف البرمجية الإحصائية المستخدمة (Bilog–MG3؛ NOHARM) تم استخدام اختبار ت للعينات المرتبطة (T-test for Paired Samples).

المعالجة الإحصائية

للإجابة عن أسئلة الدراسة تم استخدام البرمجيتين الإحصائيتين (NOHARM; Bilog–MG3) لتقدير معاملات صعوبة الفقرة، واستخدمت معاملات ارتباط بيرسون، وتحليل التباين الأحادي ذو القياسات المتكررة (1-Way ANOVA for Repeated Measurement)، واختبار ت للعينات المرتبطة (T-test for Paired Samples) للكشف عن الفروق في التقديرات.

النتائج والمناقشة:

أولا: النتائج المتعلقة بالسؤال الأول "هل تختلف تقديرات معلم صعوبة الفقرة لبيانات توائم النموذج التعويضي (MC1- PL) باختلاف عدد الأبعاد المكونة للاختبار وقوة العلاقة الإرتباطية بينها؟"

للإجابة عن هذا السؤال، تم إجراء تقديرات لمعلم صعوبة الفقرة للبيانات التي توائم نموذج (MC1- PL) والتي تشكل اختبار ثنائي التدريج وثلاثي البعد، لثلاث حالات: الحالة الأولى، على اعتبار أن فقرات الاختبار تقيس ثلاث سمات كامنة (21 Items = z +y +x)، والحالة الثانية، على افتراض أنها تقيس سمتين كامنتين (14 =y +x)؛ (14 =z +y)؛ (14 =z +x)، والحالة الثالثة على افتراض أنها تقيس سمة كامنة واحدة (أحادية البعد) (7 =z ;7 =y ;7 =x). حيث تم حساب الإحصاءات الوصفية (Descriptive Statistics) للقيم

الحقيقية والقيم المقدرة (داخل الفقرات) لمعالم صعوبة الفقرة، وفقا لاختلاف عدد الأبعاد حسب الحالات الثلاث، ووفقا لاختلاف قوة العلاقة الإرتباطية بين الأبعاد (0.00، 0.50، 0.86)، كما هو مبين في الجدول رقم (2).

الجدول رقم (2)

الإحصاءات الوصفية للقيم الحقيقية والقيم المقدرة (داخل الفقرات) لمعلم صعوبة الفقرة لبيانات النموذج التعويضي (PL –MC1) باختلاف عدد الأبعاد (ثلاثي، ثنائي، أحادي) وقوة العلاقة الارتباطية بينها .

قوة العلاقة		0.00			0.50			0.86			معالم الصعوبة الحقيقية
أبعاد الاختبار / حالة الاختبار	الإحصائي	أحادي	ثنائي	ثلاثي	أحادي	ثنائي	ثلاثي	أحادي	ثنائي	ثلاثي	
3D	المتوسط الحسابي	−0.159	−0.165	−0.197	−0.165	−0.287	−0.280	−0.124	−0.261	−0.126	−0.194
	الانحراف المعياري	1.229	1.332	1.416	0.989	1.779	1.721	0.832	1.765	0.817	1.239
2D	المتوسط الحسابي	0.217	0.198		0.154	0.278		0.143	0.299		0.182
	الانحراف المعياري	1.419	1.364		1.152	2.138		0.998	2.217		1.277
1D	المتوسط الحسابي	−0.244									−0.254
	الانحراف المعياري	1.259									1.237

يتبين من الجدول رقم (2)، أن المتوسطات الحسابية لتقديرات معلم صعوبة فقرات الحالة الأولى (ثلاثي البعد) كانت قريبة من متوسطات القيم الحقيقية لمعلم الصعوبة، عندما كانت قوة الارتباط بين أبعاد الاختبار ثلاثي السمة (0.00)، كما يتبين أن هذه المتوسطات كانت أعلى ما يكون في الحالة الثالثة (أحادي البعد)، وأدنى ما يكون في الحالة الأولى (ثلاثي البعد). أما عندما كانت قوة الارتباط بين الأبعاد (0.50، 0.86) على الترتيب، فيلاحظ أن المتوسطات الحسابية لتقديرات معلم صعوبة الفقرة كانت تقترب من متوسط القيم الحقيقية عندما تم تقدير هذه المعالم على افتراض أن الاختبار أحادي البعد.

كما يتبين من الجدول رقم (2) نفسه، أن المتوسطات الحسابية لتقديرات معلم صعوبة فقرات الحالة الثانية (ثنائي البعد) كانت قريبة من متوسطات القيم الحقيقية لمعلم الصعوبة، عندما كانت قوة الارتباط بين الأبعاد (0.00). أما عندما كانت قوة الارتباط بين الأبعاد (0.50، 0.86) على الترتيب، فيلاحظ أن المتوسطات الحسابية لتقديرات معلم صعوبة الفقرة كانت تقترب من متوسط القيم الحقيقية عندما تم تقدير هذه المعالم على افتراض أن الاختبار أحادي البعد. أما فيما يتعلق بالانحراف المعياري، فيتبين من الجدول نفسه، أن الانحراف المعياري لتقديرات معلم صعوبة الفقرة كان أعلى من الانحراف المعياري للقيم الحقيقية لمعلم صعوبة الفقرة في الحالات الثلاث للاختبار عندما كانت قوة العلاقة الارتباطية بين الأبعاد (0.00). أما عندما كانت قوة الارتباط بين الأبعاد (0.50، 0.86) على الترتيب، فيلاحظ أن الانحراف المعياري لتقديرات معلم الصعوبة كان أعلى من الانحراف المعياري للقيم الحقيقية في الحالتين الثانية والثالثة.

وتتفق هذه النتائج مع النتائج التي توصلت إليها دراسة غوس ووكر (Gosz & Walker, 2002)حيث كانت الأوساط الحسابية والانحرافات المعيارية تزداد كتقدير أعلى (overestimate) لمعالم صعوبة الفقرات كلما زادت قوة الارتباط بين الأبعاد.

ولعدم تباين الفروق في تقديرات معلم صعوبة الفقرة للبيانات الموائمة للنموذج التعويضي (MC1– PL) فقد تم حساب معاملات ارتباط سبيرمان (spearman rank correlation coefficients) بين تقديرات معلم صعوبة فقرات الاختبار داخل الفقرات، لحالات الاختبار الثلاث (ثلاثي، ثنائي، أحادي) باختلاف قوة العلاقة الارتباطية (0.0، 0.50، 0.86) بين الأبعاد، كما هو مبين في الجدول رقم (3).

الجدول رقم (3)

معاملات ارتباط سبيرمان بين تقديرات معالم صعوبة الفقرة لحالات الاختبار الثلاث باختلاف قوة العلاقة الارتباطية بين الأبعاد.

قوة العلاقة	0.00			0.50			0.86		
أبعاد الاختبار / حالة الاختبار	أحادي ثنائي	أحادي ثلاثي	ثنائي ثلاثي	أحادي ثنائي	أحادي ثلاثي	ثنائي ثلاثي	أحادي ثنائي	أحادي ثلاثي	ثنائي ثلاثي
3D	0.973	0.965	0.969	0.981	0.976	0.979	0.981	0.981	0.987
2D	0.982			0.991			0.991		
1D									

يلاحظ من الجدول رقم (3)، أن قوة الارتباط بين تقديرات معالم صعوبة الفقرة (داخل الفقرات) لحالات الاختبار الثلاث تزداد بزيادة قوة العلاقة الارتباطية بين الأبعاد. كما أن أعلى قوة ارتباط بين التقديرات ظهرت في الحالة الثانية (ثنائي البعد) حيث كانت قريبة جدا من الواحد صحيح (1) بين التقديرات أحادية البعد والتقديرات ثنائية البعد عند قوة الارتباط (0.50، 0.86). أما أقل قوة ارتباط بين التقديرات المختلفة لمعلم صعوبة الفقرة فقد بلغت (0.973) في الحالة الأولى بين التقديرات الأحادية والثنائية عند قوة الارتباط بين الأبعاد (0.00).

وللكشف عن الفروق في تقديرات معلم صعوبة الفقرات المشكلة للاختبار متعدد الأبعاد داخل الفقرات عند الحالات المختلفة لبعدية الأبعاد وباختلاف قوة الارتباط بين الأبعاد، فقد تم استخدام اختبار تحليل التباين الأحادي ذو القياسات المتكررة (1-Way ANOVA for Repeated Measurement)، كماهو مبين في الجدول رقم(4).

الجدول رقم (4) نتائج اختبار تحليل التباين الأحادي ذو القياسات المتكررة لمعالم صعوبة الفقرات المشكلة للاختبار متعدد الأبعاد باختلاف قوة العلاقة الارتباطية بين الأبعاد

قوة العلاقة / حالة الاختبار	0.00	0.50	0.86
3D	0.0017	0.0021	0.0024
2D	0.0013	0.0018	0.0019
1D	0.0016	0.0017	0.0018

* دال عند مستوى الدلالة ($\alpha = 0.05$)

يتبين من الجدول رقم (4) أن الفروق في تقديرات معلم صعوبة الفقرات المشكلة للاختبار متعدد الأبعاد داخل الفقرات لم تكن دالة إحصائيا عند مستوى الدلالة ($\alpha = 0.05$) عند الحالات المختلفة لبعدية الأبعاد وباختلاف قوة الارتباط بين الأبعاد

ثانيا: النتائج المتعلقة بالسؤال الثاني " هل توجد فروق ذات دلالة إحصائية عند مستوى الدلالة ($\alpha = 0.05$) في متوسطات تقديرات معلم صعوبة الفقرة لبيانات توائم النموذج التعويضي (PL – MC1) تعزى لاختلاف البرمجية الإحصائية المستخدمة في التحليل؟ "

للإجابة عن هذا السؤال، تم تقدير معلم صعوبة الفقرة لبيانات الدراسة الحالية، في الحالات الثلاث التي تم افتراضها (ثلاثي، ثنائي، أحادي) البعد، عند ثلاثة مستويات (0.0، 0.50، 0.86) تمثل قوة العلاقة الارتباطية بين الأبعاد. حيث تم تقديرها بطريقتين: الطريقة الأولى باستخدام البرمجية الإحصائية (NOHARM)، أما الطريقة الثانية فكانت باستخدام البرمجية الإحصائية (Bilog–MG3)، وتم حساب معامل ارتباط سبيرمان بين القيم الحقيقية والقيم المقدرة لمعلم صعوبة الفقرة في الحالات الثلاث، وعند مستويات قوة العلاقة الارتباطية الثلاثة (0.0، 0.50، 0.86) بين الأبعاد كما هو مبين في الجدول رقم (5).

الجدول رقم (5)معاملات ارتباط سبيرمان بين تقديرات معالم صعوبة الفقرة وفق البرمجيتين الإحصائيتين (Bilog–MG3؛ NOHARM) وقيمها الحقيقية لحالات الاختبار الثلاث باختلاف قوة العلاقة الارتباطية بين الأبعاد.

قوة العلاقة		0.00			0.50			0.86		
أبعاد الاختبار / حالة الاختبار	البرمجية الإحصائية	أحادي	ثنائي	ثلاثي	أحادي	ثنائي	ثلاثي	أحادي	ثنائي	ثلاثي
3D	NOHARM	0.971	0.975	0.970	0.976	0.974	0.969	0.975	0.972	0.975
	Bilog–MG3			0.975			0.974			0.976
2D	NOHARM	0.977	0.978		0.977	0.977		0.978	0.979	
	Bilog–MG3		0.976		0.976			0.981		
1D	NOHARM	0.988								
	Bilog–MG3	0.988								

يتبين من الجدول رقم (5) مدى القدرة العالية لكلا البرمجيتين الإحصائيتين (Bilog–MG3؛ NOHARM) في إجراء تقديرات دقيقة لمعلم صعوبة الفقرة تقارب القيم الحقيقية لهذه المعالم في جميع الحالات الثلاث، وعند مستويات قوة العلاقة الارتباطية الثلاثة المختلفة (0.0، 0.50، 0.86) بين الأبعاد، حيث يلاحظ أن معاملات الارتباط بين القيم الحقيقية والقيم المقدرة لمعلم صعوبة الفقرة كانت مرتفعة، ومتقاربة إلى حد ما. وقد كان أعلى معامل ارتباط بين البرمجيتين في الحالة الثالثة (أحادي البعد)، حيث بلغ (0.9989) وذو قيمة متساوية تماما.

وللكشف عن دلالة الفروق بين تقديرات معلم صعوبة الفقرة في الحالات الثلاث (ثلاثي، ثنائي، أحادي) البعد، عند مستويات قوة العلاقة الارتباطية بين الأبعاد الثلاثة (0.0، 0.50، 0.86) باستخدام الطريقتين، بمعنى وفق

البرمجيتين الإحصائيتين (Bilog–MG3؛ NOHARM)، فقد تم استخدام الاختبار الإحصائي (ت) للعينات المترابطة (T– test for Paired Sample)، كما هو مبين في الجدول رقم (6).

الجدول رقم (6)

نتائج اختبار (ت) لتقديرات معلم صعوبة الفقرة وفق البرمجيتين الإحصائيتين (Bilog–MG3؛ NOHARM) لحالات الاختبار الثلاث باختلاف قوة العلاقة الارتباطية بين الأبعاد.

قوة العلاقة	0.00			0.50			0.86		
أبعاد الاختبار / حالة الاختبار	أحادي	ثنائي	ثلاثي	أحادي	ثنائي	ثلاثي	أحادي	ثنائي	ثلاثي
3D	0.248	0.057	0.901	0.495	0.411	0.337	0.186	0.368	0.269
2D	0.135	0.062		0.322	0.323		0.802	0.356	
1D	0.071								

* دال عند مستوى الدلالة ($\alpha = 0.05$)

يتبين من الجدول رقم (6)، عدم وجود فروق ذات دلالة إحصائية عند مستوى الدلالة ($\alpha = 0.05$) في متوسطات تقديرات معلم صعوبة الفقرة لبيانات الدراسة الحالية، المقدرة وفق البرمجيتين الإحصائيتين (Bilog–MG3؛ NOHARM)، لجميع حالات الاختبار الثلاث، وعند المستويات الثلاثة المختلفة لقوة العلاقة الإرتباطية بين أبعاد الاختبار. وربما يعود السبب في ذلك إلى أن النموذج المستخدم في توليد البيانات والنموذج المستخدم لتقدير معلم صعوبة الفقرة في الدراسة الحالية هي نماذج تعويضية (Compensatory Models) وتحتوي هذه النماذج على معلم واحد يمثل صعوبة الفقرة، وتتفق هذه النتيجة مع ما أشار إليه اكرمان (Ackerman, 1994) وأكده اسبنسر (Spencer, 2004) بأنه في حالة النماذج التعويضية تكون تقديرات معلم صعوبة الفقرة متقاربة ومرتفعة. وقد يعود السبب أيضا في الحصول على تقديرات مرتفعة ومتسقة لمعامل صعوبة الفقرة وفق البرمجيتين إلى ما أشار إليه كيرسكي وهسو (Kirsci & Hsu, 2001) بأن الأخذ بعين الاعتبار تعددية البعد يعطي تقديرات أكثر دقة عند تقدير المعالم للاختبارات متعددة باستخدام برمجيات حاسوبية متخصصة. واتفقت هذه الدراسة مع دراسة وايبرغ Wiberq, (2013) ودراسة (Luecht & Miller, 1992) التي توصلت إلى أن تقديرات معالم الفقرات متعددة الأبعاد مرة باستخدام برمجية Bilog–MG3 ومرة باستخدام برمجية NOHARM كانت متقاربة ومرتفعة.

الاستنتاجات والتوصيات:

في ضوء النتائج التي توصلت إليها هذه الدراسة، يوصي الباحثان بضرورة الاهتمام بالنماذج متعددة الأبعاد التعويضية (Compensatory Models) مع الأخذ بعين الاعتبار تساوي عدد الفقرات التي تشكل كل بعد، وأن تكون التعددية داخل الأبعاد، بغض النظر عن قوة الارتباط بين الأبعاد؛ لما لهذه الافتراضات من أثر كبير في دقة تقدير معلم صعوبة الفقرة وفق نظرية استجابة الفقرة، وكذلك يوصي الباحثان بضرورة توفير وتوظيف البرمجيات الإحصائية الحديثة في القياس لتحليل البيانات الموائمة للنماذج متعددة الأبعاد.

وبما أن الدراسة الحالية اقتصرت على المحددات السابقة فإن الباحثين يقترحان إجراء دراسات مشابهة باستخدام بيانات محاكاة تتناول النماذج متعددة الأبعاد غير التعويضية (Non–Compensatory Models) وتتناول حالة تعددية البعد بين الفقرات وفي حالة عدم تساوي عدد الفقرات التي تشكل كل بعد من أبعاد الاختبار، وتحليل البيانات باستخدام برمجيات أخرى تناسب نظرية استجابة الفقرة (IRT Software) مثل، BIGSTEPS أو TESTFACT أو ConQuest أو WINSTEPS.

المراجع:

أبو شندي، يوسف. مقاومة العلامة الحقيقية لاختلاف عدد أبعاد الاختبار وقوة الارتباط بينها واختلاف توزيع قدرة المفحوصين: دراسة محاكاة. *دراسات*، العلوم التربوية، 38(7) . (2011) ، 2295– 2313.

الحجازي، نايل *نماذج نظرية الاستجابة للفقرة متعددة التدريج متعددة الأبعاد وبرامجها الحاسوبية*، ط1. عمان: دار جليس الزمان للنشر. (2010).

علام، صلاح الدين. *نماذج الاستجابة للمفردات الاختبارية أحادية البعد ومتعدّدة الأبعاد وتطبيقاتها في القياس النفسي والتربوي*، ط.1. القاهرة: دار الفكر العربي. (2005).

كاظم، أمينة والشرقاوي، أنور والشيخ، سليمان وعبدالسلام، نادية *اتجاهات معاصرة في القياس والتقويم النفسي والتربوي*. القاهرة: مكتبة الأنجلو المصرية. (1996).

Ackerman, T. Using multidimensional item response theory to understand what items and tests are measuring. Applied Measurement in Education, 7, (1994). 255- 278.

Capar, N. (2000). Analyzing multidimensional response data structure represented by unidimensional IRT models increase the precision of scoring using out-of-scale information. Paper presented at the annual meting of the Florida educational research association (45th Tallahassee, FL, November 8-10, 2000).

Dawadi, B. (1999). Robustness of the polytomous IRT to violations of the unidimensionality assumption. Paper presented at the annual meeting of the American educational research association (Montreal, Canada, April 19-23, 1999).

Deng, H., & Ansley, T. (2000). Detecting compensatory and noncompensatory multidimensionality using DIMTEST. Paper presented at the annual meting of the national council on measurement in education (New Orleans, April 25-27, 2000).

Drasgow, F., & Parsons, C. (1983). Application of unidimensional item response theory models to multidimensional data. Applied Psychological Measurement, 7, 189-199.

Embretson, S., & Reise, S. (2000). Item response theory for psychologists. New Jersey: Lawrence Erlbaum Associates.

Finch, H. (2006). Comparison of the performance of varimax and promax rotations: factor structure recovery for dichotomous items. Journal of Educational Measurement, 43(1), 40-51.

Gosz, J., & Walker, C. (2002). An empirical comparison of multidimensional item response data using TESTFACT and NOHARM. Paper presented at the annual meeting of the national council for measurement in education (New Orleans, Louisiana, April 2 – 4, 2002).

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. (1985). Item Response Theory: Principles and Applications. Boston, MA: Kluwer Nij Publishing.

Hambleton, R., Swaminathan, H., & Rogers, J. (1991). Fundamentals of Item Response Theory. Newbury Perk California: Stage publications.

Lee, S., & Terry, R. (2005). MDIRT-FIT: SAS macros for fitting multidimensional item response. Paper presented at SUGI 31th conference in university of Oklahoma, Norman, OK.

Luecht, R., Miller, T. (1992). Unidimensional calibrations and interpretations of composite Abilities for multidimensional tests. Applied Psychological Measurement. 16, 279-294.

Masters, G. (1982). A Rasch model for partial credit scoring. Psychometrika, 47(2), 150-172.

McDonald, R. (1997). Normal ogive multidimensional model. In W.J. van der Linden and R. K. Hambleton (Eds.), Handbook of modern item response theory. New York: Springer.

Meara, K., Robin, F., & Sireci, S. (2000). Using multidimensional scaling to assess the dimensionality of dichotomous item data. Multivariate Behavioral Research, 35(2), 229-257.

Muraki, E. (2000). RESGENT: Item response generator. Princeton, NJ: Educational Testing Service.

Olivares, A. (2001). Multidimensional item response theory modeling of binary data: Large sample properties of NOHARM estimates. Journal of Educational and Behavioral Statistics, 26(1), 51- 71.

Spencer, G. (2004). The strength of multidimensional item response theory in exploring construct space that is multidimensional and correlated. Unpublished Doctoral Dissertation, Brigham Young University.

Walker, C., & Beretras, S. (2000). Using multidimensional versus unidimensional ability estimates to determine student proficiency in mathematics. Paper presented at the annual meeting of the American educational research association (New Orleans, LA, April 23-28, 2000).

Wang, S., & Wang, N. (2003). The effects of multidimensional polytomous response data on unidimensional Many-FACT Rasch model parameter estimates. Paper presented at the annual meeting of the American educational research association (Chicago, IL, April 21-24, 003).

Wiberq, M. (2013). Can a multidimensional test be evaluated with unidimensional item response theory?. Educational Research and Evaluation, 18(4), 307- 320.

Wilson, M. (1999). ACER conquests user manual. Melbourne, Australia: Quintec Group.

Yao , L., & Shawartz, R. (2006) . A multidimensional partial credit model with associated item and test statistics: an application to mixed-format tests [on line]. Available: http// apm.sagepub.com/cgi/content/refs/30/6/469

Zhao, J., McMorris, R., Pruzek, R., & Cohen, R. (2002). The robustness of the unidimensional 3-PL IRT model when applied to two-dimensional data in computerized adaptive tests. Paper presented at the annual meeting of the American educational research association (New Orleans, LA, 2002).

Zimowski, Michele F., Maraki, Eiji, Mislevy, Robert, J. & Back, D. (1996). BILOG-MG3: Multiple group IRT analysis and test maintenance for binary items [A Computer Program]. Chicago: Scientific Software.