

## دقة تقدير العلامات الحقيقية عند درجات مختلفة للارتباط الموضوعي بين فقرات الاختبار في توزيعات مختلفة للقدر

د. يوسف عبدالقادر أبوشندي

قسم علم النفس

كلية التربية - جامعة السلطان قابوس

yousefaaa@squ.edu.om

د. إيهاب محمد عماره

قسم علم النفس

كلية التربية - جامعة السلطان قابوس

ehab@squ.edu.om

د. راشد سيف المحرزي

قسم علم النفس

كلية التربية - جامعة السلطان قابوس

mehrzi@squ.edu.om

## دقة تقدير العلامات الحقيقية عند درجات مختلفة للارتباط الموضعي بين فقرات الاختبار في توزيعات مختلفة للقدرة

د. يوسف عبدالقادر أبوشندي

قسم علم النفس  
كلية التربية - جامعة السلطان قابوس

د. إيهاب محمد عماره

قسم علم النفس  
كلية التربية - جامعة السلطان قابوس

د. راشد سيف المحرزي

قسم علم النفس  
كلية التربية - جامعة السلطان قابوس

### الملخص

هدفت هذه الدراسة إلى تعرّف أثر الارتباط الموضعي بين الفقرات في دقة تقدير العلامات الحقيقية (True Scores) في توزيعات تكرارية مختلفة لقدرات المفحوصين، وذلك من خلال توليد استجابات اثنتي عشرة مجموعة عشوائياً بحيث احتوت كل مجموعة على ٢٠٠٠ مفحوصاً. نتجت مجموعات البيانات من خلال التحكم بمستويات عاملين: التوزيع التكراري لقدرات الأفراد (٣ توزيعات تكرارية: التوزيع الطبيعي، التوزيع المتلوي يميناً، التوزيع المتلوي يساراً) ودرجة الارتباط الموضعي للفقرات (٤ درجات: صفر، ٣، ٦، ٩، ٠) في كل مجموعة تم توليد استجابات لاختبار مكون من ٦٠ فقرة، ١٢ فقرة منها مرتبطة موضعياً (بنسبة ٢٠٪). كشفت نتائج الدراسة أن دقة تقدير العلامات الحقيقية نقصت بزيادة درجة الارتباط الموضعي بين الفقرات. وأظهر الارتباط الموضعي تضخيماً في تقدير العلامة الحقيقية بشكل عام، وخاصة عند القدرات الواقعة عند طرفي متصل القدرة. كما أظهرت النتائج اختلافاً في دقة تقدير العلامة الحقيقية باختلاف شكل توزيع قدرة المفحوصين.

الكلمات المفتاحية: العلامة الحقيقية، الارتباط الموضعي، توزيع القدرة، دقة تقدير العلامة الحقيقية.

## The Estimation Accuracy of True Scores at Different Degrees of Local Dependence Among Test Items for Different Ability Distributions

**Dr. Yousef A. Abu Shindi**

College of Education  
Sultan Qaboos University

**Dr. Rashid S. Alemhrizi**

College of Education  
Sultan Qaboos University

**Dr. Ehab M. Omara**

College of Education  
Sultan Qaboos University

### Abstract

The study aimed to investigate the effect of local dependence among items on the estimation of true scores under different frequency distribution for examinees' ability. The study used randomly twelve data sets; each with 2000 examinees. These data sets resulted from controlling two factors: ability distributions (3 distribution: normal, positively skewed, negatively skewed) and local dependence (4 levels: correlation of zero, 0.3, 0.6, 0.9). Each data sets manipulate the local dependence for 12 items out from 60 real items (20%). Results revealed that the estimation accuracy of true scores decreases as the local dependence increases. Generally, local dependence among items showed positive bias in the estimation of True Scores across most ability scales. Moreover, ability distribution influenced the effect of local dependence on the estimation accuracy of true scores.

**Keywords:** true score, local dependence, ability distribution, estimation accuracy of true score.

## دقة تقدير العلامات الحقيقية عند درجات مختلفة للارتباط الموضعي بين فقرات الاختبار في توزيعات مختلفة للقدرة

د. يوسف عبدالقادر أبوشندي

قسم علم النفس

كلية التربية - جامعة السلطان قابوس

د. إيهاب محمد عماره

قسم علم النفس

كلية التربية - جامعة السلطان قابوس

د. راشد سيف المحرزي

قسم علم النفس

كلية التربية - جامعة السلطان قابوس

### المقدمة

يشهد العصر الحالي تزايداً واضحاً في استخدام نظرية الاستجابة للفقرة، ولعل ما تتصف به هذه النظرية من تقديرات متميزة لخصائص الاختبارات ومعالم فقراتها ودقتها في تقدير قدرات الأفراد هو السبب الحقيقي خلف هذا النمو المتزايد في مجالات استخدامها. وتقوم هذه النظرية في الأساس على فكرة أن هناك علاقة ما بين أداء الفرد الملاحظ على الفقرة وبين السمة الكامنة (غير الملاحظة) وراء هذه الاستجابة. ولأن هذه النظرية تقوم على مجموعة متنوعة من النماذج الرياضية، فإن هذه النماذج كباقي النماذج الرياضية ترتكز على مجموعة من الافتراضات التي ينبغي التحقق منها أولاً قبل الاعتماد عليها واستخدامها في التطبيقات المختلفة. وتشترط نظرية الاستجابة للفقرة تحقق عدة افتراضات في مصفوفة البيانات المثلة لاستجابات عينة الأفراد على فقرات الاختبار، وهي أحادية البعد Unidimensionality والاستقلال الموضعي Local Independence والتحرر من السرعة Speededness وفرض ملاءمة منحنى خاصية الفقرة (Item Characteristic Curve) (Hambleton, 1989).

ويربط هامبلتون وسواميناثان (Hambleton & Swaminathan, 1985) بين افتراض أحادية البعد وافتراض الاستقلال الموضعي، فيرون أنه إذا تحقق شرط أحادية البعد في فقرات الاختبار فإنها تكون مستقلة موضعياً، وإذا لم تتحقق أحادية البعد فإن الفقرات تكون مرتبطة عند مستويات معينة من القدرة. ويتعارض مع هذا الطرح العديد من المتخصصين أمثال (Crocker & Algina, 1986)، فيرون أن الافتراضين السابقين غير متكافئين، وهو ما تم إثباته في دراسة (Warm, 1978) حيث أكد أن تحقق افتراض أحادية البعد في البيانات يضمن تحقق الاستقلال الموضعي للفقرات، بينما لا يضمن الاستقلال الموضعي بين فقرات

الاختبار وجود سمة أو قدرة واحدة وراء استجابة الأفراد على تلك الفقرات.

ولا يعتبر الاستقلال الموضوعي افتراضاً حديثاً لنظرية الاستجابة للفقرة ولكنه ضمن افتراضات النظرية الكلاسيكية في القياس (Lee, 2004). ويشير الاستقلال الموضوعي إلى أن استجابة الأفراد على أي فقرة من فقرات الاختبار لا تتأثر سلباً أو إيجاباً بالاستجابة على أي فقرة أخرى بالاختبار، بمعنى أن احتمال الفشل أو النجاح في إحدى الفقرات لا يؤثر ولا يتأثر بالفشل أو النجاح في أي فقرة أخرى (Linacre, 2009).

ويوضح هامبلتون وسواميناثان (1985) Hambleton & Swaminithan أنه ليس القصد بالاستقلال الموضوعي عدم وجود ارتباط بين استجابات الأفراد على فقرات الاختبار، إذ إنه قد ترتبط الفقرات أو بعض الفقرات فيما بينها عند المستويات المتباينة من القدرة، ولكن الاستقلال يكون عند كل مستوى من مستويات القدرة، بمعنى أن لا يكون هناك ارتباط بين الاستجابات عند تثبيت مستوى القدرة.

ولقد أظهر ين (1993) Yen بعض أسباب حدوث الارتباط الموضوعي بين الفقرات، بعضها مستقلة أو بعيدة عن محتوى الفقرات الاختبارية مثل تلقي مساعدات من المعلم، الشعور بالتعب في نهاية الامتحان فبعض الفقرات قد يبدو مستواها أصعب نتيجة الشعور بالتعب في نهاية موقف الامتحان، وكذلك السرعة قد تتسبب في الارتباط الموضوعي بين الفقرات وخاصة التي تقع في نهاية الاختبار ذلك أن الطلاب الذين لن يصلوا إلى سؤال معين في نهاية موقف الامتحان لن يصلوا إلى الذي يليه، وأطلق (Chen & Thissen, 1997) على هذا النوع الارتباط الموضوعي السطحي Surface Local Dependence.

كما أشار (Walter and Rose (2013); Yen, (1993) إلى بعض الأسباب التي قد تؤدي إلى الارتباط الموضوعي بين الفقرات ولها صلة بمحتوى هذه الفقرات كارتباط بعض الفقرات بنص معين أو قطعة ما أو مثير ما، أو معلومات تحملها بعض الفقرات تزيد من احتمال الإجابة الصحيحة عن غيرها، أو ترتيب الفقرات في الاختبار مثلما وضع (Tony (2002 أن وجود أسئلة الاختيار من متعدد في الاختبارات قبل الأسئلة المقالية التي تتطلب بناء الاستجابة يؤدي إلى استفادة الطلاب من المعلومات التي يحملها هذا النوع في الإجابة عنها.

ظهرت العديد من الطرق للكشف عن الارتباط الموضوعي بين فقرات الاختبار، ومنها حساب معامل الثبات بطريقة تحليل التباين لفقرات الاختبار عند كل مستوى من مستويات القدرة (زكري، ٢٠٠٩). ففي دراسة (Yen, 1984) تم تناول طريقة Q3 الذي تعتمد على العلاقة بين البواقي لزوج من الفقرات الاختبارية بعد ضبط مستويات القدرة، وفي ضوء هذا المؤشر

تعتبر الفقرات مرتبطة موضعياً كلما ابتعدت قيمة معامل الارتباط عند مستوى قدرة معين عن الصفر. كما أشار (Chen & Thissen (1997) إلى طريقة G2 والتي تعتمد على جداول التوافق والتي تستخدم للكشف عن الفروق بين التكرارات المتوقعة والملاحظة بين كل فقرتين، وقارنت هذه الدراسة كلاً من Pearson's G2، Q3،  $\chi^2$  في دقة الكشف عن الارتباط الموضوعي بين الفقرات، وقد أظهرت نتائجها قوة أقل لكل من الطريقتين G2، Pearson's  $\chi^2$  مقارنة بالطريقة Q3.

وتناولت دراسة (Iramaneerat, Myford & Yuldkowsky (2006) أسلوب Fisher's Z الذي يعتمد على تحويله لقيم الأخطاء الملاحظة إلى أخطاء معيارية، وتعد الفقرات مرتبطة موضعياً إذا وقعت قيمة Z الملاحظة خارج مدة معينة يتم تحديدها باستخدام قيمة الانحراف المعياري لهذا المؤشر. وقد ناقشت دراسة (Hashimoto, 2011) تحيز هذا الأسلوب من أساليب الكشف عن ارتباط الفقرات وخاصة عند زيادة عدد الفقرات المرتبطة.

وقد أجرى أبوشندي (٢٠١١) دراسته للمقارنة بين مؤشرات الكشف عن الاستقلال الموضوعي للفقرات الاختبارية، إذ قارن بين مؤشرات: Q3، Fisher's Z، G2، ووسط معاملات الارتباط البينية تحت تأثير قوة الارتباط بين الأبعاد (٠، ٥، ٠، ٩، ٠) ومستوى قدرة المفحوصين (منخفضة، متوسطة، مرتفعة) وبيّنت نتائجها عدم تأثر G2 بتغير قوة الارتباط بين أبعاد الاختبار، في حين تأثرت قيمة كل من Q3، Fisher's Z، وفيما يخص تأثير مستوى القدرة بينت النتائج أن نسبة أزواج الفقرات المنتهكة للاستقلال الموضوعي لم تتغير عند استخدام المؤشر G2، بينما كانت تتخفف بنقصان مستوى قدرة الأفراد عند استخدام Fisher's Z، Q3 وقوة الارتباط بين الأبعاد ٩، ٠.

ومن المتوقع أن يؤثر انتهاك افتراض الاستقلال الموضوعي للفقرات على النتائج المتعلقة بتقدير معالم الفقرات والمفحوصين، وذلك لأن دالة الأرجحية Likelihood Function هي المعادلة التي يتم من خلالها حساب أرجحية نمط استجابة لمفحوص، ولن تكون صحيحة عند انتهاك افتراض الاستقلال الموضوعي:

$$L(u_1, u_2, u_3, \dots, u_n / \theta) = \prod_{i=1}^n P_i u_i Q_i (1 - u_i)$$

حيث  $u_i$ : استجابة المفحوص على الفقرة، وتحمل قيمة الواحد الصحيح عند استجابة المفحوص على الفقرة استجابة صحيحة، والقيمة صفر عند الاستجابة الخاطئة للمفحوص على الفقرة، و  $L(u_1, u_2, u_3, \dots, u_n / q)$  تمثل الاحتمال المشروط لنمط استجابة المفحوص،

ومن خلال هذه المعادلة سيكون احتمال إجابة مفحوص إجابة صحيحة على أي فقرتين يساوي حاصل ضرب احتمال إجابة المفحوص على الفقرة الأولى إجابة صحيحة في احتمال إجابة المفحوص على الفقرة الثانية إجابة صحيحة، ويتم تحقق هذا الشرط في حالة الاستقلال الموضوعي للفقرات، وأما إذا لم يتحقق هذا الشرط فمن المتوقع أن تتأثر دقة تقدير كل من معالم الفقرات والمفحوصين (Embretson & Reise, 2000; Hambelton & Swaminathan, 1985).

وتؤكد بعض الدراسات كدراسة شن ووانج (Chen & Wang, 1997) ودراسة زانج (Zhang, 2010) أن انتهاك افتراض الاستقلال الموضوعي عند استخدام نظرية الاستجابة للفقرة يمكن أن يترتب عليه تبعات مؤثرة في دقة تقديرات معالم الفقرات والأفراد، ومن ثم في دقة التطبيقات المختلفة التي تعتمد عليها. إذ إنه قد يؤدي إلى انخفاض تقديرات الأخطاء المعيارية مما يترتب عليه تحيز تقديرات معالم الصعوبة وتضخم معالم التمييز للفقرات وكذلك الزيادة في دقة درجات المفحوصين وتضخم معاملات الثبات ومعلومات الاختبار فقد بين كاظم المشار اليه في الطراونة (2016) أن ثبات الاختبار يتحقق بتحقيق افتراض الاستقلال الموضوعي، وأنه يمكن زيادة تحقيق ثبات الاختبار من خلال حذف الفقرات التي لا تحقق استقلالية القياس، وهو ما قد يترتب عليه استدلالات غير صحيحة وغير دقيقة بناء على تلك الدرجات كالتصنيف غير الصحيح للأفراد.

وعلى الرغم من تزايد الدراسات المتعلقة بنظرية استجابة الفقرة في الآونة الأخيرة سواء منها ما يتعلق بتطوير المقاييس النفسية والتربوية بالاعتماد على هذه النظرية، أو التي اهتمت بدراسة أثر بعض العوامل في دقة تقديرات معالم الفقرات والمفحوصين. إلا أن المتصفح للدراسات التي اهتمت بدراسة تأثير عدم تحقق افتراض الاستقلال الموضوعي على تقدير المعالم يجد أنها لم تحظ بشكل كبير من الاهتمام ومن هذه الدراسات: دراسة (Yen, 1993) التي حاولت الوقوف على تأثير انتهاك افتراض الاستقلال الموضوعي في كل من دالة المعلومات للاختبار والخطأ المعياري للتقدير ومعالم التمييز للفقرات الاختبارية، بالاعتماد على بيانات حقيقية من اختبارين أولهما في الفهم القرائي والثاني في المفاهيم والتطبيقات الرياضية (فقرات من نوع الاختيار من متعدد ذات أربعة بدائل)، وباستخدام أحد أساليب الكشف عن انتهاك الاستقلال الموضوعي وهو Q3، وأشارت النتائج إلى تأثير انتهاك افتراض الاستقلال الموضوعي على دالة معلومات الاختبار التي ظهر عليها ارتفاع واضح في مقدار المعلومات إذا ما تم تصحيح ومعالجة الاختبار بافتراض الاستقلال الموضوعي لفقراته، كما ظهر الأمر نفسه

بالنسبة لمعاملات التمييز لل فقرات الاختبارية، في حين لم تظهر النتائج فروقاً ظاهرة بالنسبة للخطأ المعياري في التقدير.

بينما تناولت دراسة (Lynda 1999) أثر الارتباط الموضوعي بين فقرات الاختبار في طريقة التصحيح المعتمدة على نظرية الاستجابة للفقرة في الاختبارات التكيفية المحوسبة. واعتمدت على توليد البيانات لخمس اختبارات تكيفية محوسبة بحيث تضمن وجود مقادير مختلفة من الارتباط الموضوعي بين الفقرات لدراسة تأثير هذه المقادير على القيم المقدرة للقدرة والدرجات الحقيقية ودقة الاختبار بشكل عام. ولقد أظهرت النتائج تأثير المقادير المرتفع من الارتباط الموضوعي بشكل سلبي في الدرجة الكلية التي يحصل عليها الفرد كما أثرت في مستوى الدقة العامة للاختبار والمتمثلة في كمية المعلومات التي يوفرها الاختبار Test Information. وأشارت النتائج إلى أن المستويات المختلفة من الارتباط الموضوعي لم يظهر لها تأثير واضح في مستويات القدرة في المدة بين (-1، 0، 5) في منتصف متصل القدرة، بينما يزداد التباين الناتج عن اختلاف مستويات الارتباط الموضوعي كلما اتجهنا نحو طرف متصل القدرة. كما أوضحت قيم التحيز لتقديرات القدرة أنها تميل للانخفاض عن القيم الحقيقية في الاتجاه السالب للقدرة بينما تميل التقديرات لأن تكون مرتفعة عن القيم الحقيقية في الطرف الأيمن من متصل القدرة مع تباين واضح لقيم التحيز في الاتجاهين بين المستويات المختلفة من مقادير الارتباط الموضوعي لفقرات الاختبار.

وهدف دراسة (Zenisky, Hambleton and Sireci 2002) إلى التعرف على مدى الارتباط الموضوعي بين فقرات الاختبار من متعدد في بيانات من اختبار القبول بكلية الطب، كما هدفت إلى دراسة أثر الارتباط الموضوعي بين الفقرات على تقديرات الثبات وتقديرات قدرات الأفراد. وقد أظهرت النتائج وجود فقرات لم تحقق الاستقلال الموضوعي، وكان السبب وراء ذلك إما عامل الزمن أو وجود فقرات مرتبطة بنص أو سياق معين. وبالتحديد ظهر العدد الأكبر من الفقرات المرتبطة موضعياً في المقاطع المعتمدة على حل المشكلات وعلى الاستدلال اللفظي، في حين ظهرت أقل الأعداد من الفقرات في اختبار العلوم البيولوجية. كما كان هناك تأثير واضح للارتباط الموضوعي في تقديرات ثبات الاختبار إذ يلاحظ ارتفاع واضح في معاملات الثبات الكلاسيكية (ثبات ألفا) ومعاملات الثبات الهامشية المقدرة وفقاً لنظرية الاستجابة للفقرة عند التعامل مع الفقرات بافتراض أنها تحقق الاستقلال الموضوعي، كما كان هناك تأثير في مستوى تقديرات قدرات الأفراد، إذ يظهر بعض التباين في التقديرات كلما زاد عدد الفقرات التي لا تحقق الاستقلال الموضوعي.



وتفحصت دراسة Bastick (2002) أثر الارتباط الموضوعي بين فقرات الاختبار على تحصل الطلبة، وبدأت الدراسة بفرضية بحثية تتوقع وجود ارتباط موضوعي بين فقرات الاختبار عندما تستخدم الأسئلة الموضوعية في الاختبارات متبوعة بأسئلة إنشائية قصيرة وهو ما يؤدي في نهاية الأمر إلى تضخم درجات الطلاب على الجزء الإنشائي من الاختبار. وقد أكدت النتائج هذه الفرضية البحثية، فقد تم الكشف عن وجود الارتباط الموضوعي بين الفقرات في الحالة موضع الدراسة كما كان هناك ارتفاع ملحوظ في درجات الطلاب في الجزء الموضوعي من الاختبار، في حين لم يظهر هذا التأثير عندما جاءت أسئلة الإجابة القصيرة قبل الأسئلة الموضوعية عند تطبيق الاختبار على عينة الدراسة.

كما قام كل من Killer, Swaminathan & Sireci (2003) بدراسة تناولت تأثير انتهاك افتراض الاستقلال الموضوعي على كل من ثبات الاختبار ومعلومات الاختبار، واستخدم للكشف عن هذا الافتراض مؤشر Q3، واعتمدت على بيانات حقيقية من فقرات اختيار من متعدد تحقق شرط الاستقلال الموضوعي، ومجموعة من الفقرات التي يربط بينها نص معين أدى إلى الارتباط الموضوعي بين الفقرات، وهو ما أكدته نتائج المؤشر Q3. ولقد أظهرت النتائج تأثير ثبات ومعلومات الاختبار بانتهاك افتراض الاستقلال الموضوعي (أي بمعالجة الاختبار كما لو كان يحقق شرط الاستقلال)، فقد ظهر ارتفاع واضح في قيم كل منهما، وكان التأثير الأكبر في معامل ثبات كرونباخ ألفا مقارنة بمعامل الثبات الطرقي المستخدم في نظرية الاستجابة للفقرة.

وفي دراسة Chen and Wang (2007) التي استخدمت مؤشر Q3، في دراسة محاكاة تم توليد البيانات فيها باستخدام النموذج ثلاثي المعالم، أظهرت النتائج تأثير انتهاك شرط الاستقلال الموضوعي على تقديرات معالم التمييز للفقرات، وأشارت النتائج إلى أن التأثير يبدو موجباً بارتفاع قيم معالم التمييز إذا كان الارتباط الموضوعي بين الفقرات موجباً، بينما تظهر قيم معالم التمييز أقل من الواقع إذا ما كان الارتباط الموضوعي بين الفقرات سالباً، كما كانت معالم التخمين للفقرات الاختبارية من بين العوامل المؤثرة على تقديرات معالم التمييز في الحالات المختلفة التي تناولتها الدراسة.

اهتمت الدراسات السابقة المتعلقة بافتراض الاستقلال الموضوعي بالطرائق والمؤشرات المستخدمة في الكشف عن تحقق هذا الافتراض في البيانات الاختبارية سواء الفعلية منها أو المحاكاة، وبعضها تفحص أثر عدم تحقق الافتراض في تقديرات معالم الفقرات: صعوبتها وتمييزها وتخمينها، ومعلمة قدرة المفحوصين، وعلى ثبات الاختبار ودالة معلوماته.

أثر عدم تحقق الافتراض في البيانات في تقدير العلامة الحقيقية لم يحظ باهتمام الباحثين باستثناء دراسة (Lynda, 1999) وكان ذلك في الاختبارات التكيفية المحوسبة Computer Adaptive Testing CAT. ولم تتناول أي من الدراسات أثر انتهاك الاستقلال الموضوعي في تقدير العلامات الحقيقية عند اختلاف توزيع قدرة المفحوصين.

ويمكن القول إن تقدير الدرجات الحقيقية يتأثر بانتهاك افتراض الاستقلال الموضوعي في مرحلتين: مرحلة تقدير معالم الفقرات والمفحوصين ومرحلة حساب الدرجة الحقيقية من منحنيات خاصة الفقرات. لذلك جاءت هذه الدراسة لتفحص تأثير الارتباط الموضوعي بين الفقرات على دقة تقدير العلامة الحقيقية. وإلى أي درجة من الارتباط الموضوعي يمكن أن تتأثر تقديرات العلامات الحقيقية. فقد يؤدي وجود ارتباطات موضوعية بين الفقرات إلى تضخيم الدرجات الحقيقية ومن ثم تعطي تحيزاً موجباً (أي درجات حقيقية عالية)؛ وفي المقابل قد يتم الحصول على تقزيم الدرجات الحقيقية وبذلك تعطي تحيزاً سالباً (أي درجات حقيقية منخفضة). ويختلف مقدار التحيز واتجاهه باختلاف نسبة الفقرات المرتبطة موضعياً من فقرات الاختبار وباختلاف درجة الارتباط الموضوعي وباختلاف مستوى القدرات الحقيقية للطالب.

### مشكلة الدراسة

من المعالم الرئيسية التي يتم تقديرها في نماذج استجابة الفقرة العلامة الحقيقية True Score، إذ بين لورد (Lord, 1980) أن تقدير العلامة الحقيقية عنصراً يُعد جوهرياً لفهم عملية القياس، وأثر أخطاء القياس وخصائص العلامات المشاهدة كقياسات قابلة للخطأ. وبيّنت الحموري (2011) أن نظرية استجابة الفقرة تأخذ باعتبارها الأنماط الملاحظة للاستجابة على كل فقرة، وقارنت بين مفحوصين حصلوا على نفس المجموع من العلامات على الاختبار، ولكنهما اختلفا في درجتهما على السمة الكامنة، فقد أجاب المفحوص الأول عن فقرات ذات معاملي صعوبة وتمييز مرتفعين وأجاب الآخر على فقرات ذات معاملي صعوبة وتمييز منخفضين.

حاولت دراسات سابقة أن تفحص تأثير دقة تقدير العلامة الحقيقية بعدم تحقق شروط تطبيق نماذج نظرية استجابة الفقرة، حيث اهتمت بعضها بتأثير عدم تحقق افتراض أحادية البعد في البيانات عند استخدام نماذج تتطلب تحقق هذا الافتراض أبوشندي (مقبول للنشر) Wang, Harris & Rosses, 2002; De Champlain, 1995; Camilli, Wang & Fesq, 1992; Cook, Dorans, Eignor & Peterson, 1983. ومنها ركز على تأثير

التغير في طول الاختبار على تقدير العلامة الحقيقية (ذيات، ٢٠٠٧؛ Pommerich & Nicewander, 1998).

ونظراً لأهمية تحقق افتراض الاستقلال الموضوعي عند تطبيق نماذج نظرية استجابة الفقرة، تناولت العديد من الدراسات تأثير عدم تحققه في تقديرات معالم الفقرات. ومعلمة المفحوص (النعي، ٢٠٠٦؛ Yen, Zenisky, Hambleton & Sireci, 2002؛ 1993؛ Bastick, Chen & Wang, 2007؛ Killer, Swaminathan & Sireci, 2003). إذ يعد عدم تحقق افتراض الاستقلال الموضوعي بين الفقرات (الارتباط الموضوعي) من الانتهاكات التي تحد بشكل كبير من استخدام نماذج استجابة الفقرة، فتستخدم طرق تقدير معالم الفقرات والمفحوصين مثل طرق الاحتمالية القصوى (Maximum Likelihood) وطرق بيز (Bayesian Method) دالة احتمالية أنماط الاستجابة للمفحوصين. فإذا كان هناك ارتباط معنوي بين بعض الفقرات أو كلها فإن دالة احتمالية أنماط الاستجابة للمفحوصين تعطي قيمة غير دقيقة. كما أن حساب الدرجات الحقيقية يفترض توفر الاستقلال الموضوعي بين الفقرات، إذ يتم حساب الدرجات الحقيقية من خلال حساب المتوسط الحسابي المباشر لاحتمال الاستجابة الصحيحة على الفقرة (المستخرجة بأي نموذج لوجاريتمي) عبر جميع فقرات الاختبار.

$$T = \frac{1}{n} \sum_{i=1}^n p_i(\theta)$$

وتمثل  $T$  العلامة الحقيقية للطالب في اختبار يتكون من  $n$  من الفقرات ( $i = 1, \dots, n$ )، ويمثل  $p_i(q)$  احتمال استجابة الشخص ذي القدرة ( $q$ ) استجابة صحيحة على الفقرة  $q$ . وبناءً على ما تقدم، تسعى الدراسة الحالية لبحث أثر درجات مختلفة من الارتباط الموضوعي بين الفقرات في دقة تقدير العلامة الحقيقية عند حالات مختلفة من شكل توزيع قدرات المفحوصين.

### أسئلة الدراسة

حاولت الدراسة الإجابة عن التساؤلات الآتية:

١. ما أثر الارتباط الموضوعي بين فقرات الاختبار في دقة تقدير العلامات الحقيقية لدى المفحوصين الذين تتوزع قدراتهم توزيعاً طبيعياً؟
٢. ما أثر الارتباط الموضوعي بين فقرات الاختبار في دقة تقدير العلامات الحقيقية لدى المفحوصين الذين تتوزع قدراتهم توزيعاً ملتوياً إلى اليمين؟

٢. ما أثر الارتباط الموضوعي بين فقرات الاختبار في دقة تقدير العلامات الحقيقية لدى المفحوصين الذين تتوزع قدراتهم توزيعاً ملتوياً إلى اليسار؟

### أهمية الدراسة

تظهر أهمية هذه الدراسة فيما يأتي:

- أهمية تحقق افتراض الاستقلال الموضوعي عند استخدام جميع نماذج نظرية استجابة الفقرة ، فقد تناولت بعض الدراسات أثر انتهاك هذا الافتراض على تقدير معالم الفقرات كدراسة النعيمي (٢٠٠٦). وتضيف هذه الدراسة تفحص أنها إلى أي درجة من عدم تحقق افتراض الاستقلال الموضوعي يمكن أن تبقى تقديرات العلامات الحقيقية True Scores دقيقة.
- تناولت حالات مختلفة من درجة الارتباط الموضوعي بين الفقرات (٠,٩, ٠,٦, ٠,٣) ودرجة استقلال موضعي تام (٠, ٠) عند توزيعات مختلفة للقدرة والتي تمثلت في التوزيع الطبيعي لقدرات المفحوصين والتوزيع الملتوي لليمين والتوزيع الملتوي لليساار وهي توزيعات تقارب التوزيعات الاعتيادية لدرجات الطلبة الفعلية.
- تساعد الباحثين وطلبة الدراسة العليا المختصين في نظرية استجابة الفقرة في معرفة إمكانية تحمل تقديرات العلامات الحقيقية لعدم تحقق افتراض الاستقلال الموضوعي وبخاصة أنه افتراض يصعب تحقيقه كما أوضح أبوعلام (٢٠٠٥) في البيانات الاختبارية وبخاصة العقلية منها.

### مصطلحات الدراسة

**العلامة الحقيقية:** هي المتوسط الحسابي لاحتمالات اجابات المفحوص عند مستوى القدرة عن جميع فقرات الاختبار.

**الارتباط الموضوعي:** هو درجة الارتباط بين كل فقرتين عند تثبيت مستوى قدرة المفحوصين. وفي هذه الدراسة استخدمت ثلاث درجات للارتباط الموضوعي بين الفقرات (٠,٩, ٠,٦, ٠,٣) وحالة الاستقلال الموضوعي. وقد تم تثبيت نسبة الفقرات التي بينها ارتباط موضعي (٢٠٪) في هذه الدراسة.

**دقة تقدير العلامة الحقيقية:** في هذه الدراسة قدرت دقة تقدير العلامة الحقيقية من خلال: حساب التحيز في تقدير الدرجات الحقيقية لكل فرد من خلال حساب الفرق غير المطلق بين

الدرجات الحقيقية المقدرة والدرجات الحقيقية الأولية (المقدرة - الأولية)، وحساب الفرق المطلق للدرجات الحقيقية المقدرة من الدرجات الحقيقية الأولية لكل فرد.

**توزيع القدرة:** الخصائص المتعلقة بالتوزيع التكراري لمعلمة القدرة من حيث شكل التوزيع وفي هذه الدراسة تم تقصص ثلاثة توزيعات: الطبيعي، المتوي يمينا، المتوي يساراً.

### محددات الدراسة

- اقتصرت الدراسة على ثلاث حالات من الارتباط الموضوعي بين الفقرات، وحالة الاستقلال الموضوعي.
- اقتصرت الدراسة على نسبة ثابتة من الفقرات المرتبطة موضعياً (٢٠٪).
- تعتمد نتائج الدراسة على دقة تقدير البرامج الحاسوبية المستخدمة في تحليل البيانات.

### المنهجية والاجراءات

#### عينة الدراسة

تكونت عينة الدراسة من اثنتي عشرة مجموعة بيانات مولدة عشوائياً باستخدام برمجية Excel بحيث احتوت كل مجموعة على ٢٠٠٠ مضمون. نتجت مجموعات البيانات من خلال التحكم بمستويات متغيرين: التوزيع التكراري لقدرات الأفراد (٣ توزيعات تكرارية: التوزيع الطبيعي، التوزيع المتوي يمينا، التوزيع المتوي يساراً) ودرجة الارتباط الموضوعي للفقرات (٤ درجات: صفر، ٣، ٦، ٩، ٠).

### إجراءات توليد البيانات

في كل مجموعة بيانات، تم استخدام تقديرات معالم فقرات لاختبار مكون من ٦٠ فقرة مختلفة في معلمة الصعوبة ومعلمة التمييز، وتم الحصول على هذه المعالم من دراسة ابو شندي (٢٠٠٨)، الذي قام بتقسيم الفترة (-٢، ٢) لمعلمة الصعوبة إلى ٠٦ جزء بشكل متساوٍ وبفارق ٠,٠٦٧، بين المعلمة والتي تليها، وتقسيم الفترة (١٩، ٠، ٧٩، ١) لمعلمة التمييز إلى ٠٦ جزء بشكل متساوٍ وبفارق ٠,٠٢٧، بين المعلمة والتي تليها. ومن خلال التداخلات بين معلمتي الصعوبة والتمييز تم الحصول على ٣٦٠٠ فقرة مختلفة المعالم. حيث تم الاختيار وبشكل عشوائي لـ ٦٠ فقرة ولعشر مرات مختلفة Replication. ولكل مجموعة من المجموعات العشرة تم إدخال معلمتي الصعوبة والتمييز لفقرات هذه المجموعة كقيم

حقيقية إلى برمجية RESGEN لتوليد الاستجابات للمفحوصين لهذه الفقرات، ثم أدخلت هذه الاستجابات إلى برمجية BILOG-MG وقدرت معالمها. وبعد ذلك تم حساب دقة تقدير المعالم في كل حالة من الحالات العشر، واختيار الحالة التي تعطي أعلى دقة تقدير. وكان الهدف من هذه العملية هو الوصول إلى أفضل عينة عشوائية من الفقرات (60) يتم اختيارها من مجموعة الفقرات الكلية (3600)، والتي تكون فيها الفروق بين القيم الحقيقية والقيم المقدرة أقل ما يمكن، واعتماد هذه العينة لتمثل القيم الحقيقية لمعالم الفقرات بغرض توليد الحالات المختلفة من البيانات، للإجابة عن أسئلة الدراسة من خلالها.

من أجل الحصول على بيانات مولدة أقرب الى الواقع، تم اعتبار معالم فقرات الاختبار كمعالم حقيقية استخدمت في توليد استجابات 2000 فرداً بقدرات مختلفة تتبع توزيعاً تكرارياً معيناً على هذه الفقرات الستين باستخدام النموذج اللوغاريتمي ثنائي المعالم. في بيانات الاستقلال الموضوعي للفقرات (درجة ارتباط موضعي صفر)، تم الحصول على استجابات كل فرد على الفقرة الواحدة من خلال مقارنة احتمال الإجابة الصحيحة باستخدام النموذج اللوغاريتمي (وفقاً لمعالم الفقرة وقدرة الفرد) مع رقم عشوائي، بحيث تكون الاستجابة صحيحة (الاستجابة 1) إذا كان احتمال الإجابة الصحيحة أكبر من الرقم العشوائي أو يساويه، وتكون الاستجابة غير صحيحة (الاستجابة صفر) إذا كان احتمال الإجابة الصحيحة أصغر من الرقم العشوائي.

وفي حالة الدرجات المختلفة من الارتباط الموضوعي للفقرات، فقد تم استخدام طريقة (مرجع) في توليد الاستجابات على الفقرات. ويتم التحكم في درجة الارتباط الموضوعي من خلال عدد الفقرات التي بها ارتباط موضعي وبدرجة الارتباط الموضوعي بين كل فقرتين، وهي تتراوح بين صفر (استقلال موضعي تام) والواحد الصحيح (ارتباط موضعي تام). وفي هذه الدراسة تم تثبيت عدد الفقرات التي بها ارتباط موضعي بعدد 12 فقرة بنسبة 20% (12 من 60 فقرة). وقد تم استخدام ثلاث درجات للارتباط الموضوعي للفقرات وهي  $(a_{ij} = 0.3, 0.6, 0.9)$ . وتقوم طريقة توليد استجابة فقرة مرتبطة موضعياً  $(u_j)$  مع فقرة أخرى  $(u_i)$  من خلال مقارنة احتمال الاستجابة الصحيحة للفقرة  $(p_j)$  وفقاً للنموذج ثنائي المعالم مع درجة الارتباط الموضوعي. إذا كان  $(a_{ij} \leq p_j)$  فيتم توليد الاستجابة على الفقرة  $(u_j)$  بالطريقة المستخدمة في حالة الاستقلال الموضوعي المشروحة سابقاً. أما إذا كان  $a_{ij} > p_j$  فتأخذ الفقرة  $(u_j)$  نفس الاستجابة على الفقرة التي ترتبط بها  $(u_i)$ .

## الأساليب الإحصائية

بعد توليد بيانات المجموعات الاثنتي عشرة، تم إخضاع كل منها للإجراءات الإحصائية الآتية:

١. حساب احتمال الاجابة الصحيحة لكل فقرة باستخدام القيم الأولية لمعالم الصعوبة والتمييز والقيم الأولية لقدرات الأفراد باستخدام النموذج ثنائي المعالم. وبعد ذلك تم حساب متوسط احتمال الإجابة الصحيحة لجميع الفقرات الستين التي تمثل الدرجات الحقيقية الأولية للأفراد.

٢. معايرة البيانات باستخدام برنامج BILOG-MG لتقدير معالم الصعوبة والتمييز للفقرات الستين وتقدير قدرات ٢٠٠٠ فرد باستخدام النموذج ثنائي المعالم.

٣. حساب احتمال الإجابة الصحيحة لكل فقرة باستخدام القيم المقدره لمعالم الصعوبة والتمييز وقدرات الأفراد المستخرجة في الخطوة ٢ باستخدام النموذج ثنائي المعالم. بعد ذلك تم حساب متوسط احتمال الإجابة الصحيحة لجميع الفقرات الستين باستخدام Microsoft Excell والتي تمثل الدرجات الحقيقية المقدره للأفراد.

٤. تم حساب التحيز في تقدير الدرجات الحقيقية لكل فرد من خلال حساب الفرق غير المطلق بين الدرجات الحقيقية المقدره والدرجات الحقيقية الأولية (المقدره - الأولية).

٥. تم حساب الفرق المطلق للدرجات الحقيقية المقدره من الدرجات الحقيقية الأولية لكل فرد.

## نتائج الدراسة

### أولاً: نتائج السؤال الأول

نص السؤال الأول على: «ما أثر الارتباط الموضوعي بين فقرات الاختبار على دقة تقدير العلامات الحقيقية لدى المفحوصين الذين تتوزع قدراتهم توزيعاً طبيعياً؟» وللإجابة عن هذا السؤال، تم إيجاد الاحصائيات الوصفية لمطلق الفروق بين العلامات الحقيقية المقدره والمولده في مجموعات البيانات الأربع لدرجات الارتباط الموضوعي، عند توزيع قدرات المفحوصين بشكل طبيعي.

جدول ١  
الاحصائيات الوصفية لمطلق الفروق بين العلامات الحقيقية المقدرة والمولدة  
في التوزيعات الطبيعية لأربع درجات للارتباط الموضوعي بين الفقرات

الانحراف المعياري	الوسط الحسابي	العدد	درجة الارتباط الموضوعي
٠,٠٣٠٣٥	٠,٠٤٠٤٤	٢٠٠٠	٠,٠
٠,٠٣٦١٧	٠,٠٤٨١٤	٢٠٠٠	٠,٣
٠,٠٣٨٩١	٠,٠٥٢٤٢	٢٠٠٠	٠,٦
٠,٠٣٩٠٧	٠,٠٥٣٣٥	٢٠٠٠	٠,٩

يظهر من جدول ١ أن المتوسطات الحسابية للقيم المطلقة للفروق بين العلامات الحقيقية المقدرة والعلامات الحقيقية المولدة للمفحوصين ذوي قدرات متوزعة توزيعاً طبيعياً تختلف باختلاف درجة الارتباط الموضوعي بين الفقرات بل وتتزايد بزيادة الارتباط الموضوعي بين الفقرات. وهذا يشير إلى أن دقة تقدير العلامة الحقيقية تقل كلما زادت درجة الارتباط الموضوعي بين الفقرات المشككة للاختبار. كما يظهر نمط هذا التزايد أيضاً في الانحراف المعياري. ويشير ذلك إلى اختلاف أكبر للقيم المطلقة بين العلامات الحقيقية المقدرة والمولدة كلما زادت درجة الارتباط الموضوعي للفقرات. وبغرض تفحص دلالة الفروق الظاهرة في المتوسطات الحسابية بين درجات الارتباط الموضوعي بين الفقرات، استخدم تحليل التباين الأحادي والذي يعرض جدول ٢ نتائجه.

جدول ٢  
تحليل التباين الأحادي للقيم المطلقة للفروق بين العلامات الحقيقية المقدرة والمولدة في  
التوزيعات الطبيعية عند اختلاف درجة الارتباط الموضوعي بين الفقرات

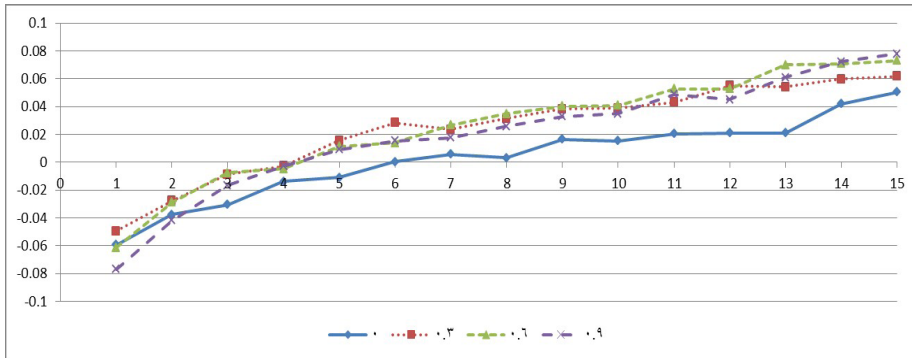
القيمة الاحتمالية	ف	متوسط المربعات	درجات الحرية	مجموع المربعات	مصدر التباين
*٠,٠٠٠	٥٢,٥٩٤	٠,٠٦٩	٣	٠,٢٠٨	بين المجموعات
		٠,٠٠١	٧٩٩٦	١٠,٥٣٥	داخل المجموعات
			٧٩٩٩	١٠,٧٤٣	الكلية

\*دالة احصائية عند مستوى دلالة ٠,٠١

توضح النتائج في جدول ٢ أن الاختلاف في القيم المطلقة للفروق بين العلامات الحقيقية المقدرة والمولدة كانت دالة إحصائياً عند مستوى دلالة ٠,٠١، ولتحديد مصادر هذه الفروق استخدمت طريقة شيفيه للمقارنات البعدية. وبينت نتائجها أن الفروق دالة إحصائياً بين حالة الاستقلال الموضوعي بين الفقرات والحالات المختلفة للارتباط الموضوعي، كما تبين وجود فروق



بين حالات الارتباط الموضوعي ٠,٢ وبقية الحالات وعدم وجود فروق دالة إحصائية بين حالتين الارتباط الموضوعي ٠,٦, ٠,٩ و٠,٠. وتشير هذه النتائج إلى أن زيادة الارتباط الموضوعي بين الفقرات عن مستوى معين لا يزيد من تأثيره على تقدير العلامات الحقيقية. وللتعرف على اتجاه تأثير درجة الارتباط الموضوعي للفقرات على دقة تقدير العلامة الحقيقية، تم حساب التحيز من خلال الفرق غير المطلق بين العلامات الحقيقية المقدر والمولدة عبر تدرج قدرات الطلبة. ولتليخيص قيم التحيز، تم حساب متوسط التحيز في ١٥ مجموعة من المفحوصين بعد ترتيبهم تصاعدياً حسب قدراتهم، بحيث تحتوي كل مجموعة على عدد متساوٍ من الطلبة. ويعرض شكل ١ متوسط التحيز في تقدير العلامات الحقيقية في الدرجات الأربع للارتباط الموضوعي؛ بحيث يحتوي المحور العمودي على مقدار التحيز في تقدير العلامات الحقيقية، ويحتوي المحور الأفقي على المجموعات الخمس عشرة لدرجات المفحوصين.



شكل ١  
التحيز في تقدير العلامات الحقيقية في التوزيع الطبيعي للدرجات  
الأربع للارتباط الموضوعي بين الفقرات

يظهر من شكل ١ أن التقديرات الأكثر دقة للعلامات الحقيقية كانت عندما تحقق افتراض الاستقلال الموضوعي بين الفقرات. بينما زاد نسبياً مقدار التحيز في تقدير العلامات الحقيقية في الدرجات الثلاث للارتباط الموضوعي بين الفقرات. وبشكل عام يعد مقدار التحيز في تقدير العلامات الحقيقية منخفضاً ومتقارباً في الدرجات الثلاث للارتباط الموضوعي بين الفقرات إذ انحصر مقدار التحيز بين -٠,٠٨ و ٠,٠٨. ويظهر الشكل أن مقدار التحيز كان قريباً من الصفر عند ذوي القدرات المتوسطة، بينما ظهر تحيزاً سلبياً في تقدير العلامات الحقيقية للطلبة ذوي القدرة المنخفضة وتحيزاً موجباً عند الطلبة ذوي القدرات المرتفعة.

## ثانياً : نتائج السؤال الثاني :

نص السؤال الثاني على «ما أثر الارتباط الموضوعي بين فقرات الاختبار على دقة تقدير العلامات الحقيقية لدى المفحوصين الذين تتوزع قدراتهم توزيعاً ملتوياً إلى اليمين؟» وللإجابة عن هذا السؤال، تم إيجاد الاحصائيات الوصفية لمطلق الفروق بين العلامات الحقيقية المقدره المولده في مجموعات البيانات الأربع لدرجات الارتباط الموضوعي.

## جدول ٣

الاحصائيات الوصفية لمطلق الفروق بين العلامات الحقيقية المقدره والمولده في التوزيعات الملتوية الى اليمين لأربع درجات للارتباط الموضوعي بين الفقرات

الانحراف المعياري	الوسط الحسابي	العدد	درجة الارتباط الموضوعي
٧٠,٠٣٢٩	٠,٠٤٥٨١	٢٠٠٠	٠,٠
٠,٠٣٤٨٢	٠,٠٤٥٧٧	٢٠٠٠	٠,٢
٠,٠٤٠٣٢	٠,٠٥٤٣٧	٢٠٠٠	٠,٦
٠,٠٤١٠٩	٠,٠٥٧٩٣	٢٠٠٠	٠,٩

يظهر من جدول ٣ أن المتوسطات الحسابية للقيم المطلقة للفروق بين العلامات الحقيقية المقدره والمولده تختلف باختلاف درجة الارتباط الموضوعي بين الفقرات عندما توزعت قدرات المفحوصين بشكل ملتوياً إلى اليمين، ولكن تظهر النتائج تقارب في دقة التقدير في حالتها الارتباط (٠,٣,٠,٠) واختلافهما عن حالات الارتباط الأخرى (٠,٦,٠,٩). وبغرض فحص دلالة الفروق الظاهرة في المتوسطات الحسابية بين درجات الارتباط الموضوعي بين الفقرات، استخدم تحليل التباين الأحادي والذي يعرض جدول ٤ نتائج.

## جدول ٤

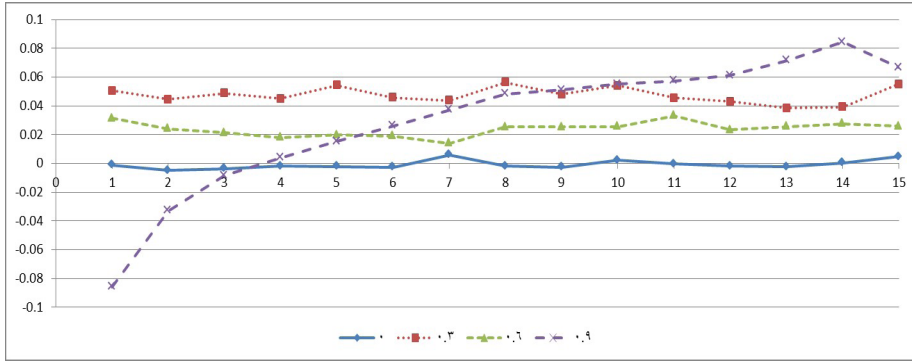
تحليل التباين الأحادي للقيم المطلقة للفروق بين العلامات الحقيقية المقدره والمولده في التوزيعات الملتوية إلى اليمين عند اختلاف درجة الارتباط الموضوعي بين الفقرات

مصدر التباين	مجموع المربعات	درجات الحرية	متوسط المربعات	ف	القيمة الاحتمالية
بين المجموعات	٠,٢٧٧	٣	٠,٠٧٦	٥٣,٢٢٨	٠,٠٠٠*
داخل المجموعات	١١,٢٥٧	٧٩٩٦	٠,٠٠١		
الكلي	١١,٥٨٤	٧٩٩٩			

\* دالة احصائياً عند مستوى دلالة ٠,٠١

توضح النتائج في جدول ٤ أن الفروق في القيم المطلقة للفروق بين العلامات الحقيقية المقدره والمولده كانت دالة إحصائياً عند مستوى دلالة ٠,٠١، وأظهرت نتائج مقارنات شيفية

البعديّة أن الفرق لم تكن دالة إحصائياً بين حالة الاستقلال الموضوعي وحالة الارتباط الموضوعي الضعيف بين الفقرات (٠, ٢)، ولكنها دالة في الحالات الأخرى. وللتعرف على مدى تأثر دقة تقدير العلامة الحقيقية بدرجة الارتباط الموضوعي للفقرات، تم حساب التحيز من خلال الفرق غير المطلق بين العلامات الحقيقية المقدرة والمولدة عبر تدرّج قدرات الطلبة. وتلخيص قيم التحيز تم حساب متوسط التحيز في ١٥ مجموعة طلابية بعد ترتيبهم تصاعدياً حسب قدراتهم بحيث تحتوي كل مجموعة على عدد متساوٍ من الطلبة. يعرض الشكل ٢ متوسط التحيز في تقدير العلامات الحقيقية في الدرجات الأربع للارتباط الموضوعي؛ بحيث يحتوي المحور العمودي على مقدار التحيز في تقدير العلامات الحقيقية، ويحتوي المحور الأفقي المجموعات الخمسة لدرجات المفحوصين.



شكل ٢

### التحيز في تقدير العلامات الحقيقية في التوزيع المتوي إلى اليمين للدرجات الأربع للارتباط الموضوعي

يظهر من شكل ٢ أن التحيز في تقدير العلامات الحقيقية بشكل عام كان منخفضاً في حالات الارتباط الموضوعي بين الفقرات، وانطبقت تقديرات العلامات الحقيقية مع المولدة في حالة الاستقلال الموضوعي. ويظهر عدم الاختلاف الواضح في دقة تقدير العلامة الحقيقية باختلاف مستوى قدرات المفحوصين في حالات الارتباط (٠, ٠, ٠, ٢, ٠, ٦)، والتباين الواضح في دقة التقديرات باختلاف مستوى قدرات المفحوصين في حالة الارتباط ٠, ٩، إذ يظهر ارتفاع تقدير العلامات الحقيقية عن القيم المولدة عبر جميع تدرّج قدرة الطلبة ما عدا عند ذوي القدرات المنخفضة والتي أظهرت تحيزاً سالباً في تقدير العلامات الحقيقية.

## ثالثاً: نتائج السؤال الثالث

نص السؤال الثالث على: "ما أثر الارتباط الموضوعي بين فقرات الاختبار في دقة تقدير العلامات الحقيقية لدى المفحوصين الذين تتوزع قدراتهم توزيعاً ملتوياً إلى اليسار؟" وللإجابة عن هذا السؤال، تم إيجاد الإحصائيات الوصفية لمطلق الفروق بين العلامات الحقيقية المقدره والعلامات الحقيقية الأولية.

## جدول ٥

الاحصائيات الوصفية لمطلق الفروق بين العلامات الحقيقية المقدره والمولدة في التوزيعات الملتوية الى اليسار لأربع درجات للارتباط الموضوعي بين الفقرات

الانحراف المعياري	الوسط الحسابي	العدد	درجة الارتباط الموضوعي
٠,٠٣٢٢٨	٠,٠٤٣١٧	٢٠٠٠	٠,٠
٠,٠٣٨٨٨	٠,٠٥١٤٣	٢٠٠٠	٠,٣
٠,٠٤٠١٦	٠,٠٥٢٣٦	٢٠٠٠	٠,٦
٠,٠٤٢٦٢	٠,٠٥٨٤٩	٢٠٠٠	٠,٩

يظهر من النتائج في جدول ٥ أن المتوسطات الحسابية للقيمة المطلقة للفروق بين القيم المقدره للعلامة الحقيقية والقيم الأولية ازدادت بزيادة قوة الارتباط الموضوعي بين الفقرات، وهذا يشير إلى أن دقة تقدير العلامة الحقيقية عندما توزعت قدرات المفحوصين بشكل ملتو الى اليسار تقل كلما زادت درجة الارتباط الموضوعي بين الفقرات المشكلة للاختبار. كما يظهر نمط هذا التزايد أيضاً في الانحراف المعياري. ويشير ذلك إلى اختلاف أكبر للقيم المطلقة بين العلامات الحقيقية المقدره والمولدة كلما زادت درجة الارتباط الموضوعي للفقرات. وبغرض تفحص دلالة الفروق الظاهرة في المتوسطات الحسابية بين درجات الارتباط الموضوعي بين الفقرات، استخدم تحليل التباين الأحادي والذي يعرض جدول ٦ نتائج.

## جدول ٦

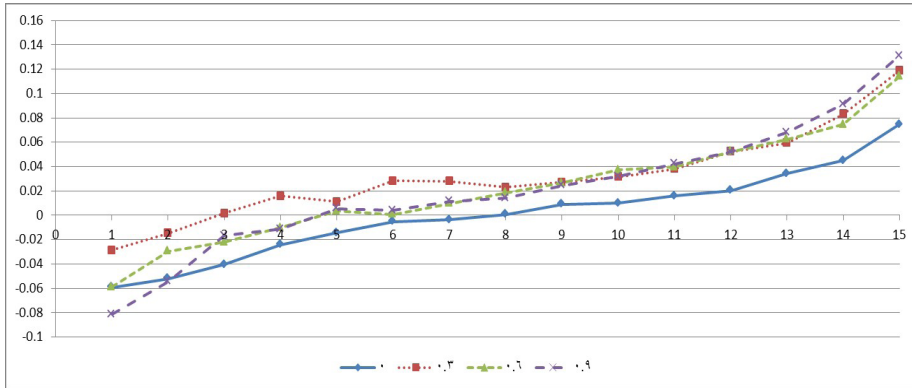
تحليل التباين الأحادي للقيم المطلقة للفروق بين العلامات الحقيقية المقدره والمولدة في التوزيعات الملتوية إلى اليسار عند اختلاف درجة الارتباط الموضوعي بين الفقرات

مصدر التباين	مجموع المربعات	درجات الحرية	متوسط المربعات	ف	القيمة الاحتمالية
بين المجموعات	٠,٢٣٨	٣	*٠,٠٧٩	٥٢,٩٧٠	*٠,٠٠٠
داخل المجموعات	١١,٩٦٠	٧٩٩٦	*٠,٠٠١		
الكلية	١٢,١٩٧	٧٩٩٩			

\*دالة احصائية عند مستوى دلالة ٠,٠١

توضح النتائج في جدول ٦ أن الاختلاف في القيم المطلقة للفروق بين العلامات الحقيقية المقدرة والمولدة كانت دالة إحصائياً عند مستوى دلالة ٠,٠١، وبينت نتائج مقارنات شيفيه البعدية أن الفروق دالة إحصائياً بين حالة الاستقلال الموضوعي بين الفقرات والحالات المختلفة للارتباط الموضوعي، ولم تكن هناك فروق دالة إحصائياً بين حالات الارتباط الموضوعي الثلاثة (٠,٢,٠,٦,٠,٩).

وللتعرف على مدى تأثر دقة تقدير العلامة الحقيقية بدرجة الارتباط الموضوعي للفقرات في حالة توزيع قدرات المفحوصين بشكل ملتبس إلى اليسار، تم حساب التحيز من خلال الفرق غير المطلق بين العلامات الحقيقية المقدرة والعلامات الحقيقية المولدة عبر تدرج قدرات الطلبة. ولتخيص قيم التحيز تم حساب متوسط التحيز في ١٥ مجموعة طلابية بحسب قدراتهم بحيث تحتوي كل مجموعة على عدد متساو من الطلبة. يعرض شكل ٣ متوسط التحيز في تقدير العلامات الحقيقية في الدرجات الأربع للارتباط الموضوعي؛ بحيث يحتوي المحور العمودي على مقدار التحيز في تقدير العلامات الحقيقية، ويحتوي المحور الأفقي المجموعات الخمس عشرة لقدرات المفحوصين.



شكل ٣

التحيز في تقدير العلامات الحقيقية في التوزيع الملتبس إلى اليسار للدرجات الأربع للارتباط الموضوعي

يظهر من شكل ٣ أن التحيز في تقديرات العلامة الحقيقية بشكل عام كان منخفضاً ومتقارباً في الدرجات الأربع للارتباط الموضوعي بين الفقرات، إذ انحصر مقدار التحيز بين -٠,٠٨ و ٠,١٣، وكان التحيز قريباً من الصفر عند ذوي القدرات المتوسطة. كما يظهر ارتفاع تقدير العلامات الحقيقية في حالة الاستقلال الموضوعي والحالات الثلاث للارتباط الموضوعي بشكل عام عن القيم المولدة عبر جميع مستويات القدرة باستثناء القدرات المنخفضة.

## مناقشة النتائج

بينت نتائج الدراسة انخفاض دقة تقدير العلامة الحقيقية بارتفاع الارتباط الموضوعي المقدر والمولدة؛ ففي حالة التوزيع الطبيعي لقدرات المفحوصين كانت الفروق دالة إحصائياً في العلامات الحقيقية باستثناء بين حالتي الارتباط ٠,٦ و ٠,٩، وفي حالة التوزيع المتوي إلى اليمين للقدرات كانت الفروق دالة إحصائياً باستثناء بين حالتي الاستقلال الموضوعي التام بين الفقرات وحالة الارتباط ٠,٢، وفي حالة التوزيع المتوي إلى اليسار للقدرات كانت الفروق دالة إحصائياً بين حالة الاستقلال الموضوعي التام بين الفقرات وبقية حالات الارتباط الموضوعي.

وأظهرت النتائج أن معدل تأثير انتهاك افتراض الاستقلال الموضوعي بين الفقرات على تقدير العلامات الحقيقية لا يتبع نمطاً واحداً، وإنما يختلف باختلاف مستوى قدرات المفحوصين. فتقدير العلامات الحقيقية للمفحوصين ذوي القدرات المتوسطة لا يتأثر بانتهاك افتراض الاستقلال الموضوعي، إذ تتشابه العلامات الحقيقية المقدره مع المولدة عند الدرجات الثلاث للارتباط الموضوعي في التوزيعات التكرارية الثلاث لقدرات المفحوصين. بينما نجد أن انتهاك افتراض الاستقلال الموضوعي بين الفقرات يظهر تحيزاً في تقدير العلامات الحقيقية للمفحوصين ذوي القدرات الطرفية. فبينما يظهر تحيزٌ موجبٌ في تقدير العلامات الحقيقية للمفحوصين من ذوي القدرات العالية؛ نجد أنه يظهر تحيزٌ سالباً في تقدير العلامات الحقيقية للمفحوصين من ذوي القدرات المنخفضة في التوزيعات التكرارية الثلاث لقدرات المفحوصين. وتشير هذه النتائج إلى الحاجة إلى الانتباه إلى توافر افتراض الاستقلال الموضوعي بين فقرات الاختبار عندما يكون التركيز منصباً أكثر على تقدير العلامات الحقيقية للمفحوصين مرتفعي التحصيل ومنخفضي التحصيل.

وهذه النتائج يمكن تفسيرها من خلال ما أشار إليه كل من هاملتون وسواميناثان (Hambleton & Swaminathan, 1985) بأن توزيع العلامات الحقيقية يكون في مدى واسع عندما تكون الفقرات صعبة (نقطة الانعكاس على منحنى خصائص الفقرة) وعالية التمييز (النقطة التي يكون عندها المنحنى أكثر انحداراً أو الميل أكبر ما يمكن) وعند وسط تدرج علامات القدرة. وتكون العلامات الحقيقية ضمن مدى ضيق إذا كانت علامات القدرة عالية أو متدنية والفقرات صعبة وعالية التمييز.

إن المتبع لنتائج هذه الدراسة يلاحظ جلياً توافقها مع ما يؤكده الأدب المتعلق بنظرية استجابة الفقرة من ضرورة تحقق افتراضات هذه النظرية في البيانات مما يزيد من الموثوقية

في استخدام نماذجها لتقدير المعالم سواء منها المتعلق ب فقرات الاختبار أو المفحوصين كالعلامة الحقيقية.

وجاءت نتائج هذه الدراسة مؤكدة على ضرورة تحقق افتراض الاستقلال الموضوعي في البيانات الأمر الذي يعتبر الافتراض المشترك بين جميع نماذج نظرية استجابة الفقرة (ثنائية الاستجابة Dichotomous ومتعددة الاستجابة Polytomous ومتصلة الاستجابة Continuous، أحادية المعلم وثنائية المعلم وثلاثية المعلم، وأحادية البعد Unidimensional ومتعددة الأبعاد Multidimensional).

ولغرض التوصل إلى تفسيرات لنتائج الدراسة تم حساب المتوسطات الحسابية والانحرافات المعيارية لمعالم الفقرات (الصعوبة والتمييز) ومعلمة المفحوصين (القدرة) عند حالات الارتباط الموضوعي المختلفة وحالات توزيع قدرات المفحوصين المختلفة أيضاً، لأن معالم الفقرات والمفحوصين هي المتغيرات التي تدخل في حساب احتمال استجابة الفرد استجابة صحيحة على الفقرة  $\pi(\theta)$  والتي بدورها تدخل في حساب العلامة الحقيقية، وبعدها حسب الفروق المطلقة بين القيم المقدرة لهذه المعالم والقيم المولدة لها، ومن ثم أجري فحص الدلالة الإحصائية لهذه الفروق. بينت النتائج في جميع الحالات أن الفروق في متوسطات معالم الفقرات لم تكن دالة إحصائياً، ولكنها كانت دالة لمعلمة القدرة.

مما يدل على أن سبب تأثير تقديرات العلامة الحقيقية عند الحالات المختلفة من الارتباط الموضوعي جاء من تأثير معلمة قدرة المفحوصين وهذا قد يعود إلى العلاقة الإيجابية التي وضحتها هامبلتون وسواميناثان (Hambleton & Swaminathan, 1985) بين توزيع العلامات الحقيقية وعلامات القدرة.

ويتفق هذا التفسير مع ما توصلت إليه دراسة (Zenisky, Hambleton & Sireci, 2002) بأن معلمة قدرة المفحوصين تقل دقة تقديرها بزيادة الارتباط الموضوعي بين الفقرات، واتفقت أيضاً النتائج مع نتائج دراسة (Lynda, 1999) التي توصلت إلى أن دقة التقدير تقل عند أطراف متصل القدرة وتزداد عند الوسط. ولكن نتائج هذه الدراسة تخالفت مع نتائج دراسات (Chen & Wang, 2007; Yen, 1993) من حيث تأثير الارتباط الموضوعي في دقة تقدير معلمة تمييز الفقرات.

## التوصيات

- بناء على نتائج البحث الحالي، فإنه يوصى بما يأتي:
- ضرورة التحقق من توافر افتراض الاستقلال الموضوعي في البيانات قبل تقدير العلامة الحقيقية والاعتماد عليها في إصدار أحكام على المفحوصين، وذلك بسبب تأثير دقة تقديرات العلامة الحقيقية بقوة العلاقة الارتباطية بين الفقرات. ويمكن معالجة الارتباط الموضوعي بين الفقرات قبل تطبيق نماذج الاستجابة للفقرة بدمج الفقرات المرتبطة موضعياً، واستخدام نماذج الاستجابات المتعددة.
  - إجراء دراسات مشابهه باستخدام نماذج أخرى لاستجابة الفقرة مثل نموذج راش أو النموذج اللوغاريتمي الثلاثي أو الرباعي.

## المراجع

- أبوشندي، يوسف (٢٠٠٨). تأثير تعدد الأبعاد للاختبار والعلاقة بينها على تقديرات معالم فقراته: دراسة محاكاة. (رسالة دكتوراه غير منشورة)، جامعة اليرموك، إربد، الأردن.
- أبوشندي، يوسف (٢٠١١). كواشف الاستقلال الموضوعي في نظرية الاستجابة للفقرة ومدى تأثيرها بقوة العلاقة الارتباطية بين الأبعاد ومستوى القدرة للمفحوصين. مجلة جامعة تشرين للبحوث والدراسات العلمية (سلسلة الآداب والعلوم الإنسانية). ٣٣(٢)، ٤٩-٦٢.
- أبوشندي، يوسف (مقبول للنشر). أثر استخدام معايير أحادية البعد لبيانات متعددة الأبعاد على التقديرات المختلفة لنظرية استجابة الفقرة: دراسة محاكاة. مجلة جامعة دمشق.
- أبوعلام، صلاح (٢٠٠٥). نماذج الاستجابة للمفردة الاختيارية أحادية البعد ومتعددة الأبعاد وتطبيقاتها في القياس النفسي والتربوي. القاهرة: دار الفكر العربي.
- الحموري، هند (٢٠١١). دراسة استكشافية لملاءمة نماذج نظرية الاستجابة للمفردة في بناء اختبار تحصيلي من إعداد المعلم. مجلة العلوم التربوية والنفسية. ١٢(٢)، ٤٧-٨١.
- ذيابات، لؤي (٢٠٠٧). أثر التغير في طول الاختبار على خصائص توزيع العلامات الحقيقية حسب النموذج الثلاثي المعلم. رسالة ماجستير غير منشورة، جامعة اليرموك، إربد، الأردن.
- زكري، علي بن محمد (٢٠٠٩). الخصائص السيكومترية لاختبار (أوتيس-لينون) للقدرة العقلية مقدره وفق القياس الكلاسيكي ونموذج راش لدى طلبة المرحلة المتوسطة بمحافظة صبيا التعليمية. رسالة دكتوراه غير منشورة، كلية التربية، جامعة أم القرى، المملكة العربية السعودية.



الطراونة، صبري (٢٠١٦). بناء اختبار محكي المرجع في الإحصاء التحليلي لطلبة الدراسات العليا في كليات العلوم التربوية في الجامعات الأردنية الحكومية باستخدام نظرية استجابة الفقرة. مجلة العلوم التربوية والنفسية. ١٧(١)، ٥١٥-٥٤٢.

النعمي، عزالدين (٢٠٠٦). أثر انتهاك افتراض الاستقلال الموضوعي على التقديرات المختلفة لنظرية استجابة الفقرة. رسالة دكتوراة غير منشورة، جامعة اليرموك، اربد، الأردن.

Bastick, T. (2002). *Demonstrating local item dependence for recognition and supply format tests*. Paper presented at the annual meeting of the Western Psychological Association, Irvine, CA, April 14-17, (ERIC No.ED465785)

Camilli, G; Wang, M & Fesq, J.(1992). *The effects of dimensionality on True score conversion tables for the law school admission test*. LSAC research report series (ERIC Document Reproduction Service No. ED 468754).

Chen, W. H., & Thissen, D. (1997). Local dependence index for item pairs using item response theory. *Journal of Educational and Behavioral Statistics*, 22, 265-289.

Chen, C. & Wang, W. (2007). Effect of ignoring item interaction on item parameter estimation and detection of interacting items. *Applied Psychological Measurement*, 31, 388-411.

Cook, L; Dorans, N; Eignor, D & Petersen, N. (1983). *An assessment of the relationship between the assumption of unidimensionality and the quality of IRT True – score Equating*. (ERIC Document Reproduction Service No. ED 235190).

Crocker, L. & Algina, J. (1986). *Introduction to classical and modern test theory*. New York: CBS College publishing.

De Champlain, A. (1995). *Assessing the effect of multidimensionality on IRT True score equating for subgroups of examinees*. (ERIC Document Reproduction Service No. ED 382643).

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

Hambleton, R. (1989). *Principles and selected applications of item response theory in R.L. Linn (Ed.), Educational measurement (3<sup>rd</sup> ed. PP. 200)* New York: American Council on Education and Macmillan.

Hambleton, R. & Swaminathan, H. (1985). *Item Response Theory: Principles and applications*. Boston. MA: Kluwer-Nijhoff.

Hashimoto, T. (2011). Local independence statistics applied to testlet items. *The National Center for University Entrance Examinations*, 2, 19-23.

- Iramaneerat, C., Myford, C. & Yuldkowsky, R. (2006). *Item dependence in an objective structured clinical examination*. Paper presented at the 25th International Objective Measurement Workshop, Berkeley, CA, April 2006.
- Killer, L., Swaminathan, H. & Sireci, S. (2003). Evaluating scoring procedures for context-dependent items sets. *Applied Measurement in Education*, 16(3), 207-222.
- Lee, y. (2004). Examining passage-related local item dependence (LID) and measurement construct using Q3 statistics in an EFL reading comprehension test. *Language Testing*, 21(1), 74-100
- Linacre, J. (2009). Local independence and residual covariance: a study of Olympic figure skating ratings. *Journal of Applied Measurement*, 10(2), 1-13.
- Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Lynda, M. R. (1999). *Impact of Local Item Dependence on Item Response Theory Scoring in CAT*. Law School Admission Council, Princeton, NJ.
- Pommerich, M. & Nicewander, W. (1998). *Estimating average True Score*. (ERIC Document Reproduction Service No. ED 425175).
- Tony, B. (2002). *Demonstrating Local Item Dependence for Recognition and Supply Format Tests*. Paper presented at the annual meeting Annual Meeting of the Western Psychological Association. (ED 465785), Irvine, CA.
- Walter, O. & Rose, M. (2013). Effect of item order on item calibration and item bank construction for computer adaptive tests. *Psychological Test and Assessment Modeling*, 55(1), 81-91.
- Wang, X. B., Harris, V., & Rosses, L. (2002). *Effects of multidimensionality on IRT item characteristics and true score estimates: implications for computerized test assembly*. Low school Admission council, Newton, PA.
- Warm, T. A. (1978). *A primer of item response theory*. Oklahoma City, OK: U.S Coast Guard Institute.
- Yen, W. M. (1984). Effects of local item dependence on the fit and equating performance of the three-parameter logistic model. *Applied Psychological Measurement*, 8, 125-145.
- Yen, W. M. (1993). Scaling performance assessments: Strategies for managing local item dependence. *Journal of Educational Measurement*, 30, 187-213.

Zenisky, A., Hambleton, R., & Sireci, S. (2002). *Effects of local item dependent on the validity of IRT item, test and ability statistics*. Retrieved from, [www.aamc.org/students/mcat/research/monograph5.df](http://www.aamc.org/students/mcat/research/monograph5.df).

Zhang, O. (2010). *Polytomous IRT or testlet model: An evaluation of scoring models in small testlet size situations*. Unpublished Master Thesis. Graduate School of The University of Florida.