

مقارنة ثلاثة طرق للكشف عن انتهاك افتراض الاستقلال الموضوعي في

الاختبارات الوطنية لضبط نوعية التعليم في الأردن

د. عز الدين عبدالله عواد النعيمي

جامعة الإمام محمد بن سعود الإسلامية

كلية العلوم الاجتماعية

قسم علم النفس

بحث منشور في مجلة كلية الاداب/جامعة جنوب الوادي/جمهورية مصر

العربية/العدد (42)/ لعام 2014

مقارنة ثلاثة طرق للكشف عن انتهاك افتراض الاستقلال الموضوعي في الاختبارات الوطنية

لضبط نوعية التعليم في الأردن

المخلص

هدفت الدراسة المقارنة بين ثلاث طرق للكشف عن انتهاك افتراض الاستقلال الموضوعي، وهذه الطرق، هي: (Z, G2 Fisher, Q3)، ولتحقيق هذا الهدف تم استخدام بيانات الاختبار الوطني لضبط نوعية التعليم في الأردن للعام الدراسي 2007/2008، حيث طبق هذا الاختبار على الصفوف الثامن والعاشر الأساسيين. وأظهرت النتائج أنه بالنسبة للصف الثامن، اتفقت المؤشرات الثلاثة (Z, G2 Fisher, Q3) في الكشف عن انتهاك افتراض الاستقلال الموضوعي، حيث أشارت النتائج أن جميع المباحث قد انتهكت هذا الافتراض بشكل كبير، حيث كانت نسب أزواج الفقرات التي بينها انتهاك تتجاوز 90%. وبالنسبة للصف العاشر الأساسي، ولوحظ انسجام المؤشرات الثلاثة في الكشف عن أعلى المباحث انتهاكاً لافتراض الاستقلال الموضوعي، وكانت في مبحث اللغة العربية، و انسجام المؤشرين Q3، z Fisher في الكشف عن أقل المباحث انتهاكاً لافتراض الاستقلال الموضوعي، وكانت في مبحث اللغة الإنجليزية.

Abstract

The purpose of the study was to compare three methods (G2, Fisher Z, and Q₃) to detect the violation of local independence assumption. To achieve this purpose, the researcher used the national test data for educational quality control in the academic year 2007/2008 as the test was administrated to eighth and tenth grade students. The results showed that for the eighth grade students, the three methods were consistent in detecting the violation of local independence assumption, as results showed high degrees of violation for such assumption in all study subjects, the items pairs that reported violation exceeded 90%. As for the tenth grade students, there was an obvious consistency between the three methods in detecting violation of local independence assumption, as they all methods reported similar results for the more violated school subject " Arabic" and consistency between Fisher Z, and Q₃ in detecting the least violated school subject " English".

مقدمة

تستخدم نظرية استجابة الفقرة (IRT) Item Response Theory بشكل واسع من قبل المتخصصين في القياس النفسي والتربوي (Xie,2001). ونماذج استجابة الفقرة أحادية البعد نماذج رياضية تعرّف العلاقة بين أداء المفحوص الملاحظ على الاختبار وبين السمة أو القدرة الغير الملاحظة للمفحوص والتي تكون مسؤولة عن أداء المفحوص على الاختبار، وترتكز نماذج نظرية استجابة الفقرة IRT Modals، مثل كثير من النماذج الرياضية على افتراضات عدة يجب تحققها في البيانات التي تخضع للتحليل باستخدام هذه النماذج، وهي أحادية البعد للسمة التي يقيسها الاختبار Unidimensionality، والاستقلال الموضوعي لل فقرات Local Item Independence (LII)، والعلاقة الوتيرية بين القدرة والأداء على الفقرة Item Characteristic Curve (ICC)، وعدم السرعة Non-Speededness (Hambleton, 1989).

وباستخدام نظرية استجابة الفقرة (IRT) في تحليل بيانات الاختبارات النفسية والتربوية، تم إيجاد حلول لمعظم المشكلات التي تعاني منها النظرية التقليدية في القياس Classical Test Theory (CTT)، ومن أهم هذه الحلول الحصول على تقديرات لمعالم الفقرة متحررة من قدرات المفحوصين، وكذلك الحصول على تقديرات لمعالم القدرة متحررة من خصائص الفقرات (Hambleton & Swaminathan, 1985).

ويعتبر افتراض الاستقلال الموضوعي للفقرة من الافتراضات الرئيسة لكل من النظرية التقليدية في القياس (CTT) ونظرية استجابة الفقرة (IRT) (Lee, 2004)، حيث أشار هامبلتون وسوامينيثن (Hambleton & Swaminathan, 1985) إلى أن افتراض الاستقلال الموضوعي يعني أن استجابات المفحوصين على فقرات الاختبار تكون مستقلة إحصائياً عند مستوى قدرة معينة، أي أن استجابة المفحوص لفقرة ما يجب أن لا تؤثر سلباً أو إيجاباً على استجابته لفقرة أخرى؛

بمعنى أن إجابة الفقرة لا تعطي أية تلميحات أو معيقات لإجابة فقرة أخرى ، وهذا الافتراض يوضح أن قدرة المفحوص وخصائص الفقرة هي وحدها التي تؤثر في استجابته على فقرات الاختبار. ويتحقق افتراض الاستقلال الموضوعي للفقرة (LII) إذا كانت مجموعة الفقرات أحادية البعد، أي إذا كانت تقيس قدرة أو سمة واحدة، فإذا لم يتحقق افتراض أحادية البعد فإن درجات فقرات الاختبار ستكون مرتبطة عند مستوى قدرة معين لأن فقرات الاختبار تقيس سمة ثانية أو أكثر (Hambleton & Swaminathan,1985).

وأشار عدد من الباحثين (Chen & Thisson,1997;Tuerlinckx & Deboeck,2001;Yen,1984) إلى أن من أهم الأسباب التي تقف وراء ظهور انتهاك افتراض الاستقلال الموضوعي بين فقرات الاختبار أي ظهور اعتماد موضعي (LD) Local Dependence هو تعددية الأبعاد لفقرات الاختبار، كما ذكرت ين (Yen,1993) أسباباً أخرى لظهور الاعتماد الموضوعي (LD) بين فقرات الاختبار، كان من أهمها المساعدة الخارجية والسرعة والتبعية لنص معين وغيرها من الأسباب التي تؤدي إلى ظهور الانتهاك لافتراض الاستقلال الموضوعي.

وبينت الكثير من الدراسات (Allen & Sudweeks,2001;Killer,Swaminathan & Sireci, 2003;Reese,1999;Thompson & Pommerrich;1996;Zenisky,Hambelton & Sireci,2002) أن انتهاك افتراض الاستقلال الموضوعي يؤدي إلى تقديرات غير دقيقة لمعالم الفقرات و قدرات المفحوصين ودالة المعلومات للفقرة والاختبار، بالإضافة إلى تأثيره على معاملات الثبات والخطأ المعياري في التقدير.

وقد ظهرت العديد من المؤشرات التي تستخدم للكشف عن انتهاك افتراض الاستقلال الموضوعي، ومن أكثر هذه المؤشرات استخداماً المؤشر Q_3 ، وهو عبارة عن معامل الارتباط للبوقي لزوج من الفقرات بعد ضبط السمة المقدرة، وقد استخدم في عدد من الدراسات للكشف عن انتهاك

الاستقلال الموضوعي (Chen & Thissen,1997; Reese,1999; Zenisky, Hambleton & Sireci,2002). والمؤشر الإحصائي G^2 ، وهو يبين الفروق بين ما هو متوقع وبين ما هو ملاحظ لتكرار الاستجابات لزوج من الفقرات، وقد استخدم هذا المؤشر في عدد من الدراسات (Chen & Thissen,1997; Thompson & Pommerich,1996)، والمؤشر فيشر (Z) Fisher Z الذي استخدمه شن (Shen,1997)، وهو تحويل للإحصائي Q_3 ، ويتم ذلك بتحويل الأخطاء الملاحظة إلى أخطاء معيارية.

وهناك عدد قليل من الدراسات التي تناولت المقارنة بين المؤشرات المستخدمة في الكشف عن انتهاك الاستقلال الموضوعي، وذلك لبيان مدى فاعليتها في الكشف عن الفقرات التي ينتهك فيها افتراض الاستقلال الموضوعي، و كانت ابرز النتائج التي توصلت إليها هذه الدراسات أن مؤشر فيشر (Z) أكثر كفاءة وحساسية وثباتا في الكشف عن انتهاك أو عدم انتهاك افتراض الاستقلال الموضوعي عند مقارنته مع المؤشر Q_3 (Shen,1997). كما أظهرت النتائج في دراسة تومبسون وبومرج (Thompson & Pommerich,1996) أن مؤشر G^2 حساس جدا لحجم العينة، حيث إن توزيع المؤشر G^2 قريب جدا من توزيع مربع كاي وبدرجة حرية واحدة، ومن المعروف أن توزيع مربع كاي حساس جدا لحجم العينة (Hambleton & Swaminathan,1985).

ولابد من التأكيد هنا على أهمية افتراض الاستقلال الموضوعي، حيث إن انتهاك هذا الافتراض سيؤثر على النتائج المتعلقة بتقدير معالم الفقرات والمفحوصين، وذلك لأن المعادلة المحورية التي تقوم عليها نماذج نظرية استجابة الفقرة أحادية البعد هي المعادلة التي يتم من خلالها حساب أرجحية نمط استجابة المفحوص، وانتهاك الاستقلال الموضوعي سيؤدي إلى التأثير على المعالم المقدره والمعادلة هي :

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

حيث u : استجابة المفحوص على الفقرة، وتحمل قيمة الواحد الصحيح عند استجابة المفحوص على الفقرة استجابة صحيحة ، والقيمة صفر عند الاستجابة الخاطئة للمفحوص على الفقرة، و $L(u_1, u_2, u_3, \dots, u_n / \theta)$ 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 الموضعي للفقرات، أما إذا لم يتحقق هذا الشرط فان هنالك خطورة تقع على دقة تقدير كلٍ من معالم الفقرات والمفحوصين . (Hambelton & Swaminalthan,1985; Embretson & Reiese,2000)

الاختبار الوطني لضبط نوعية التعليم في الأردن

يعد مستوى أداء الطلبة على اختبارات التحصيل معياراً أساسياً للحكم على فعالية النظم التربوية القائمة ووسيلة تساعد في التخطيط الفعال لمستقبل المسيرة التربوية في أي بلد، إذ يمكن في ضوء نتائج ذلك الأداء رسم سياسة تربوية شاملة تهدف إلى تحديد مدخلات العملية التربوية الفعالة، ومن ثم تحسينها، وكذلك مراقبة وتطوير عملياتها وتقويم مخرجاتها. ومن هنا جاء اهتمام القائمين على العملية التعليمية في الأردن على إعداد اختبار وطني لقياس مدى امتلاك طلبة الفئة المستهدفة المهارات الأساسية التراكمية لعدد من المباحث الدراسية، حيث تقوم عدة لجان مختارة بعناية بعملية إعداد هذا الاختبار، ويتم انتقاء فقرات الاختبار النهائية بالاعتماد على التحليلات الإحصائية للفقرات، ويتم الحكم على جودة الفقرات بالاعتماد على بعض الإحصائيات الخاصة بالفقرة كمعاملات الصعوبة والتمييز، والتي يتم تحليلها وفق النماذج الكلاسيكية

والحديثه. وللوقوف على دقة الاختيار لهذه الفقرات لابد من القائمين على مثل هذه الاختبارات أن يكونوا واعين لمخاطر عدم تحقق الافتراضات التي تقوم عليها هذه النماذج، ومن أهم هذه الافتراضات افتراض الاستقلال الموضوعي للفقرات، والذي إذا ما تم انتهاكه أدى إلى تقديرات غير دقيقة للإحصائيات الخاصة بالاختبار ومعالم الفقرات، ويعد قياس مستوى أداء الطلبة على الاختبار الوطني لضبط نوعية التعليم معياراً يمكن من خلاله تحديد مدى تحقيق الطلبة للنتائج المتوقع منهم تحقيقها، والكشف عن فاعلية العملية التعليمية التعلمية، وتحديد نقاط القوة لتعزيزها، ونقاط الضعف لمعالجتها.

المؤشرات المستخدمة للكشف عن الاستقلال الموضوعي

• مؤشر Q_3

اقترحت ين (Yen,1984) المؤشر الإحصائي Q_3 كمؤشر للكشف عن الارتباط الموضوعي بين فقرات الاختبار. ومؤشر Q_3 هو معامل الارتباط للبواقي لزوج من الفقرات بعد ضبط السمة المقدره . ولحساب المؤشر Q_3 ينبغي تقدير القدرة لكل مفحوص وذلك لاستخدام هذه القدرة لتقدير احتمال الإجابة الصحيحة للمفحوص الذي يمتلك هذه القدرة لكل الفقرات، ويتم حساب الباقي ويرمز له بالرمز (d_{ja}) ، وذلك بأخذ الفرق بين الأداء الملاحظ والأداء المتوقع للمفحوص على الفقرة، ويمكن التعبير عن ذلك رياضياً كما يلي :

$$d_{ja} = U_{ja} - P_j(\theta a) \dots \dots \dots (2)$$

حيث (U_{ja}) تمثل العلامة الملاحظة للمفحوص (a) على الفقرة (j) وهي إما واحد عند الاستجابة الصحيحة أو صفر عند الاستجابة الخاطئة، و $P_j(\theta a)$ تمثل احتمال إجابة الفقرة (j) إجابة صحيحة من قبل المفحوص (a) وهي تقع بين الصفر والواحد الصحيح. وبالتالي لو تم

أخذ الفقرتين (i ، j) ، فان Q_3 يكون عبارة عن معامل الارتباط بين البواقي لهاتين الفقرتين لجميع المفحوصين ،ويمكن التعبير عنه كما يلي:

$$Q_{3ji} = r(d_j, d_i) \dots \dots \dots (3)$$

وقد تم استخدام هذا المؤشر الإحصائي وأثبت فاعليته في الكشف عن الارتباط الموضوعي

للفقرات في عدة دراسات منها (Chen & Thission,1997;Lee,2004;Yen,1993).

• مؤشر G^2 : صمم مؤشر G^2 للكشف عن الفروق بين ما هو متوقع وما هو ملاحظ لزوج

من الفقرات ،وبذلك يمكن عمل جدول توافقي Contingency Table ،يتم من خلاله تبيان

التكرارات الملاحظة والمتوقعة لأي زوج من الفقرات كما يأتي :

		item j	
		0	1
Item i	0	O_{11}	O_{12}
	1	O_{21}	O_{22}

		Item j	
		0	1
Item i	0	E_{11}	E_{12}
	1	E_{21}	E_{22}

ويمكن حساب المؤشر الإحصائي كما يأتي :

$$G^2 = -2 \sum_{i=1}^2 \sum_{j=1}^2 O_{ij} \ln \left(\frac{E_{ij}}{O_{ij}} \right) \dots \dots \dots (4)$$

حيث يمكن معرفة التكرارات المتوقعة (E_{ij}) من خلال برمجة Bilog ، ويؤكد شن وثيسن (Chen & Thissen,1997) بأن توزيع G^2 يتوزع تقريباً حسب توزيع مربع كاي بدرجة حرية واحدة، وقد استخدم هذا المؤشر أيضاً من قبل ثومبسون وبومرش (Thompson & Pommerich,1996) .

• فيشر (ز) Fisher Z

قام الباحث شن (Shen,1997) باعتماد هذا الأسلوب في الكشف عن انتهاك افتراض الاستقلال الموضوعي، حيث حملت دراسته عنوان " Quantifying Item Dependency by Fisher's Z"، وتم اعتماد نموذج راش كأساس في هذه الدراسة، وبالتالي فإن جميع التحليلات الخاصة بالفقرات والمفحوصين قد تم تحليلها باستخدام هذا النموذج، ويمكن تلخيص إجراءات هذا الأسلوب بالخطوات الثلاث الآتية:

1- الحصول على الأخطاء المعيارية standardized residuals لكل مشاهدة من قبل المفحوص (n) على الفقرة (i)، وذلك كما في المعادلة الآتية :

$$d_{ni} = \frac{\text{observed}_{ni} - \text{Expectd}_n}{SE_n} \dots\dots\dots(5)$$

2- الحصول على معاملات الارتباط للأخطاء المعيارية بين أي زوج من الفقرات (d_{ni}) عند جميع المفحوصين .

3- القيام بحساب فيشر (ز) Fisher's Z لمعاملات الارتباط المحولة التي تم حسابها في الخطوة السابقة، وذلك كما في المعادلة الآتية :

$$Z_{ij} = 1/2 \log \left[\frac{1+r_{ij}}{1-r_{ij}} \right] \dots\dots\dots(6)$$

يرتكز المعيار الذي يتم اعتماده لوجود الارتباط الموضوعي على قيم فيشر (ز) التي يتم حسابها لل فقرات المستقلة، والمنطق من وراء ذلك هو أنّ توزيع قيم فيشر (ز) لل فقرات المستقلة ينبغي أن تتوزع بوسط مقداره صفر، وبالتالي فإن أي قيمة من قيم فيشر (ز) لل فقرات كانت أعلى بمقدار انحرافين معياريين لقيم فيشر (ز) يمكن اعتبارها فقرات غير مستقلة موضعياً .

مشكلة الدراسة وأسئلتها :

يعد افتراض الاستقلال الموضوعي من الافتراضات الأساسية التي تقوم عليها نظرية استجابة الفقرة، وذلك لان ظهور أي انتهاك لافتراض الاستقلال الموضوعي سيؤثر على التقديرات المختلفة والمتعلقة بمعالم الفقرات أو الاختبار أو معالم المفحوصين ، ومن هنا لابد من استخدام مؤشرات فعالة في الكشف عن الفقرات التي بينها انتهاك لافتراض الاستقلال الموضوعي، حتى نتقاضي أي تأثير على تقديرات نماذج استجابة الفقرة.

وهدفت هذه الدراسة إلى المقارنة بين ثلاثة مؤشرات هي : Q_3 والذي اقترحه ين (Yen,1984)، والمؤشر Fisher Z والذي اقترحه شن (Shen,1997)، والمؤشر G^2 والذي اقترحه كل من بايشوب وفينبرغ وهولاند (Bishop,Fienerg & Holland,1975;as Cited in Chen &Thissen,1997)، وهي مؤشرات تُستخدم للكشف عن الاستقلال الموضوعي لل فقرات وذلك من خلال بيانات الاختبار الوطني لضبط نوعية التعليم المستخدم في وزارة التربية والتعليم الأردنية، وبشكل أكثر تحديد حاولت الدراسة الإجابة عن السؤال الرئيس الآتي:

هل تختلف نسبة الفقرات التي كشفت المؤشرات ($Q_3, Fisher Z, G^2$) عن وجود

ارتباط موضوعي بينها ؟ وهل تختلف هذه النسبة باختلاف المباحث الدراسية؟

الطريقة الإجراءات

يحتوي هذا الجزء وصفاً للبرامج المستخدمة في الدراسة، ووصفاً لعينة الدراسة، والطريقة التي تمّ بها استخدام البيانات لحساب المؤشرات الثلاثة: $Q3$ و Fisher Z و G^2 التي تكشف عن أزواج الفقرات التي بينها انتهاك لافتراض الاستقلال الموضعي، وذلك للإجابة عن أسئلة الدراسة حول درجة انتهاك مباحث الاختبار الوطني لضبط نوعية التعليم في الأردن لافتراض الاستقلال الموضعي لكل من الصفين الثامن والعاشر الأساسيين.

البرامج المستخدمة في الدراسة

- BILOG : وهو برنامج يستخدم لتحليل بيانات تحتوي على فقرات ثنائية Binary Items للاختبارات ذات التصحيح الثنائي ويستخدم لتقدير كلٍّ من معالم الفقرات (الصعوبة، والتمييز، والتخمين) ، ومعالم المفحوصين، حيث تتضمن عملية التحليل هذه ثلاث مراحل، وهي :
1. المرحلة الأولى (Phase 1) وفي هذه المرحلة يتم تقدير إحصائيات الفقرات حسب النظرية التقليدية في القياس مثل معامل صعوبة الفقرة وهو نسبة الاستجابات الصحيحة على الفقرة، ومعامل تمييز الفقرة وهو معامل ارتباط بوينت بيسيريال بين علامة الفقرة والعلامة الكلية على الاختبار.
 2. المرحلة الثانية (Phase 2) وفي هذه المرحلة يتم تقدير معالم الفقرة ، كما يتم تحديد الفقرات الملائمة للنموذج .
 3. المرحلة الثالثة (Phase 3) وفي هذه المرحلة يتم تقدير علامات القدرة و معالم الفقرة حسب النموذج المستخدم في التحليل.

وقد استخدم في هذه الدراسة النموذج اللوجستي الثلاثي المعلم، والذي يتم من خلاله تقدير معالم الفقرة الثلاثة: الصعوبة، والتمييز، والتخمين، وتقدير معلمة القدرة للمفحوصين.

- برنامج مؤشرات الاعتماد الموضوعي لل فقرات الثنائية:

A Computer program for Local Dependence Indices For Dichotomous Items (LDID) وهو برامج مكتوب بلغة فورتران 90، يتم من خلاله حساب كل من المؤشرات Fisher Z, G² ، Q₃. ويحتاج تشغيل هذا البرنامج إلى ثلاثة ملفات هي: ملف قدرات المفحوصين المقدر، و ملف البيانات الخام المولدة، وملف معالم الفقرات المقدر، والتي يتم استلالها من مخرجات التحليل لبرنامج BILOG ، وبالاعتماد على برنامج LDID تم اعتماد القيمة 0.05 كحد فاصل بين أزواج الفقرات التي بينها انتهاك لافتراض الاستقلال الموضوعي من غيرها من الفقرات عندما تم التعامل مع المؤشر Q₃ ؛ حيث إذا كانت القيمة المطلقة لأزواج الفقرات عند استخدام المؤشر Q₃ أعلى من 0.05 اعتبرت هذه الأزواج منتهكة لافتراض الاستقلال الموضوعي، وإذا كانت القيمة المطلقة للمؤشر Q₃ اقل من 0.05 كانت أزواج الفقرات مستقلة موضعياً، وتعني القيمة الموجبة للمؤشر Q₃ أن إجابة فقرة ما تؤثر إيجاباً على إجابة فقرة أخرى بينما القيمة السالبة تعني أن إجابة فقرة ما تؤثر سلباً على إجابة فقرة أخرى، وكذلك نفس الأمر عند استخدام المؤشر Fisher Z، واعتبرت العلامة الزائفة 2 كحد فاصل بين أزواج الفقرات التي بينها انتهاك لافتراض الاستقلال الموضوعي من غيرها عندما تم التعامل مع المؤشر Fisher Z ؛ حيث إذا كانت القيمة المطلقة لأزواج الفقرات عند استخدام هذا المؤشر أعلى من العلامة الزائفة 2 كانت الفقرات منتهكة لافتراض الاستقلال الموضوعي، وإذا كانت اقل كانت أزواج الفقرات مستقلة موضعياً، والقيمة 10 كحد فاصل بين الفقرات التي بينها انتهاك لافتراض الاستقلال الموضوعي من غيرها عندما تم التعامل مع المؤشر G²؛ حيث انه إذا كانت القيمة أعلى من 10 لأزواج الفقرات عند

استخدام هذا المؤشر كانت الفقرات منتهكة لافتراض الاستقلال الموضوعي والعكس صحيح
(Kem, Cohen & lin, 2005).

عينة الدراسة

تم تطبيق الاختبار الوطني لضبط نوعية التعليم للعام الدراسي 2008/2007 على جميع طلبة الصف الثامن الأساسي في الأردن، وتم تطبيقه على عينة عشوائية لطلبة الصف العاشر الأساسي، والجدول (1) الآتي يوضح عينة الدراسة التي تم تطبيق هذه الدراسة عليهم، حيث اعتمد الباحث على نفس عينة طلبة الصف العاشر الأساسي، وأختار عينة عشوائية من 10000 طالب وطالبة للصف الثامن الأساسي.

الجدول 1

عينة الدراسة لطلبة الصف الثامن والعاشر الأساسي حسب مباحث الاختبار الوطني لضبط نوعية التعليم

المباحث					
الصف	اللغة العربية	اللغة الإنجليزية	الرياضيات	العلوم	التربية الاجتماعية والوطنية
الثامن	10000	10000	10000	10000	10000
العاشر	2982	2982	3055	3055	3119

النتائج

للإجابة عن سؤال الدراسة : هل تختلف نسبة أزواج الفقرات التي كشفت
المؤشرات (Q3,Fisher Z,G2) عن وجود انتهاك لافتراض الاستقلال ؟ وهل تختلف هذه
النسبة باختلاف المباحث الدراسية؟

تمّ حساب نسبة أزواج الفقرات التي بينها انتهاك لافتراض الاستقلال الموضوعي من خلال قسمة
عدد أزواج الفقرات و التي بينها انتهاك لافتراض الاستقلال الموضوعي على عدد أزواج الفقرات
الكلية والعلاقة الآتية توضح كيفية حساب نسبة أزواج الفقرات :

$$\text{نسبة أزواج الفقرات} = \text{عدد أزواج الفقرات} / 435 \dots\dots\dots (7)$$

وعدد أزواج الفقرات في كل حالة من الحالات المستخدمة قد بلغ 435، وذلك لان عدد الفقرات
في كل حالة من الحالات بلغ 30 فقرة، ومن خواص التوافق الرياضية فان عدد أزواج الفقرات
لاختبار عدد فقراته 30 فقرة يساوي 435 وهذه قيمة ثابتة في كل الحالات.

و يبين الجدول (2) نسبة أزواج الفقرات التي يوجد بينها انتهاك لافتراض الاستقلال
الموضوعي ولجميع المباحث الدراسية لكل من الصف الثامن والعاشر الأساسيين ، وذلك لكل
مؤشر من المؤشرات (Q3,Fisher Z,G²) .

الجدول 2

نسبة أزواج الفقرات للمؤشرات Q_3 و Fisher Z و G^2 والتي بينها انتهاك لافتراض الاستقلال الموضوعي لمباحث الاختبار الوطني لكل من الصف الثامن والعاشر الأساسيين

مباحث الاختبار الوطني						المؤشر	الصف
التربية الاجتماعية والوطنية	التربية الإسلامية	العلوم	الرياضيات	اللغة الإنجليزية	اللغة العربية		
0.906	1	0.986	1	0.998	0.986	Q_3	الثامن
0.931	0.886	0.917	0.910	0.906	0.940	Fisher Z	
1	1	1	1	1	1	G^2	
0.198	0.214	0.189	0.193	0.129	0.938	Q_3	العاشر
0.628	0.614	0.402	0.638	0.230	0.638	Fisher Z	
0.931	0.857	0.283	0.628	0.805	1	G^2	

بالنسبة للصف الثامن الأساسي، يتبين من الجدول (2) أن جميع المباحث كانت فقراتها منتهكة لافتراض الاستقلال الموضوعي، حيث كانت نسبة الفقرات التي بينها اعتماد موضوعي في هذه المباحث تتجاوز 90%، وكانت تتراوح هذه النسب من 0.886 إلى 1، مما يعني أن هذه الافتراض قد تم انتهاكه بصورة كبيرة، وقد اتفقت جميع المؤشرات في الكشف عن انتهاك هذا الافتراض، و باستخدام المؤشر Q_3 كانت أقل نسبة أزواج الفقرات التي يوجد بينها انتهاك لافتراض الاستقلال الموضوعي هي في مبحث التربية الوطنية والاجتماعية، وكانت هذه النسبة تساوي 0.906، وباستخدام المؤشر Fisher Z هي في مبحث التربية الإسلامية، وكانت النسبة تساوي 0.886، وباستخدام المؤشر G^2 كانت جميع النسب متساوية حيث بلغت 100%، في حين أن نسبة أزواج الفقرات التي بينها انتهاك لافتراض الاستقلال الموضوعي كانت أكبر في مبحثي الرياضيات و التربية الإسلامية باستخدام المؤشر Q_3 ، وكانت النسبة تساوي 100%،

وباستخدام المؤشر Z Fisher هي في مبحث اللغة العربية، وكانت النسبة تساوي 0.940،
وباستخدام المؤشر G^2 كانت جميع النسب متساوية حيث بلغت 100%.

وبالنسبة للصف العاشر الأساسي، يتبين من الجدول (2) أن معظم المباحث كانت نسبة الفقرات التي بينها انتهاك لافتراض الاستقلال الموضوعي أقل مما هو عليه في الصف الثامن الأساسي، حيث تراوحت نسبة الفقرات التي بينها انتهاك لافتراض الاستقلال الموضوعي من 0.129 إلى 0.938، ونلاحظ من خلال الجدول 2، أنه وباستخدام المؤشرين Q_3 و Z Fisher كانت أقل نسبة أزواج للفقرات التي يوجد بينها انتهاك لافتراض الاستقلال الموضوعي هي في مبحث اللغة الإنجليزية ، وكانت النسبة تساوي 0.129 و 0.230 على التوالي، وباستخدام المؤشر G^2 كانت في مبحث العلوم، وكانت النسبة تساوي 0.283، في حين أن نسبة أزواج الفقرات التي بينها انتهاك لافتراض الاستقلال الموضوعي كانت أكبر في مبحث اللغة العربية باستخدام المؤشر Q_3 ، وكانت النسبة تساوي 0.938 وباستخدام المؤشر Z Fisher هي في مبحثي اللغة العربية والرياضيات، وكانت النسبة تساوي 0.638، وباستخدام المؤشر G^2 هي في مبحث اللغة العربية ، وكانت النسبة تساوي 100%.

وتم إيجاد الوسيط لكل مؤشر من المؤشرات الثلاث (Q_3, Z, G^2) ولكل مبحث من مباحث الاختبار الوطني في الصفين الثامن والعاشر الأساسيين ، والجدول (3) يوضح هذه الوسيطات للمباحث المختلفة ولكل مؤشر من المؤشرات الثلاثة.

الجدول 3

الوسيط لكل مؤشر من المؤشرات Q_3 و Fisher Z و G^2 للمباحث المختلفة في كل من الصف

الثامن والعاشر الأساسيين

مباحث الاختبار الوطني						المؤشر	الصف
التربية الاجتماعية والوطنية	التربية الإسلامية	العلوم	الرياضيات	اللغة الإنجليزية	اللغة العربية		
0.168	0.290	0.209	0.297	0.308	0.243	Q_3	الثامن
12.098	10.163	13.974	9.024	12.832	14.154	Fisher Z	
5826	4012	3670	1156	1467	3513	G^2	
0.028-	0.029-	0.021-	0.025-	0.013-	0.261	Q_3	العاشر
2.611	2.462	1.537	2.461	1.128	3.547	Fisher Z	
29	24	5	13	5	665	G^2	

يتبين من الجدول (3) أنه بالنسبة إلى الصف الثامن الأساسي كانت الوسيطات للمؤشر Q_3 أعلى من 0.05 وهذا يؤكد بأن جميع المباحث قد انتهت فيها افتراض الاستقلال الموضوعي، وكانت أكثر المباحث انتهاكاً لهذا الافتراض مبحث اللغة الإنجليزية، وكان الوسيط يساوي 0.308 ، في حين كان أقلها مبحث التربية الاجتماعية والوطنية، وكان الوسيط يساوي 0.168، وفيما يخص المؤشر Fisher Z، كانت الوسيطات جميعها أعلى من 2، مما يعني أن جميع المباحث قد انتهت هذا الافتراض، حيث كانت أعلاها انتهاكاً في مبحث اللغة العربية، وكان الوسيط يساوي 14.154، وأقلها انتهاكاً في مبحث الرياضيات، وكان الوسيط يساوي 9.024، وفيما يخص المؤشر G^2 ، كانت الوسيطات جميعها أعلى من 10، مما يؤكد انتهاك هذه المباحث لافتراض الاستقلال الموضوعي، حيث كانت أعلاها انتهاكاً في مبحث التربية الاجتماعية والوطنية، وكان الوسيط يساوي 5826، وأقلها انتهاكاً في مبحث الرياضيات، وكان الوسيط يساوي 1156.

ويتبين من الجدول (3)، أنه بالنسبة للصف العاشر الأساسي، كانت الوسيطات للمؤشر Q3 أقل من 0.05، فيما عدا مبحث اللغة العربية، وهذا يؤكد بأن جميع المباحث قد استجابت لافتراض الاستقلال الموضوعي، وكانت أكثر المباحث استجابةً لهذا الافتراض مبحث اللغة الإنجليزية، وكان الوسيط يساوي -0.013، في حين أن مبحث اللغة العربية قد انتهك هذا الافتراض، حيث كان الوسيط يساوي 0.261، وفيما يخص المؤشر Fisher Z، كانت الوسيطات جميعها أعلى من 2، فيما عدا مبحثي اللغة الإنجليزية والعلوم، حيث كانت الوسيطات لهذين المبحثين تساوي 1.128 و 1.537 على التوالي، ويلاحظ أن الانتهاك في باقي المباحث كان طفيفاً، حيث كانت الوسيطات قريبة جداً من القيمة 2، وفيما يخص المؤشر G2، كانت الوسيطات جميعها أعلى من 10، فيما عدا مبحثي اللغة الإنجليزية والعلوم، حيث كان الوسيط يساوي 5 لكل منهما، ويلاحظ أن أكثر المباحث انتهاكاً لافتراض الاستقلال الموضوعي مبحث اللغة العربية، حيث كان الوسيط يساوي 665.

مناقشة النتائج والتوصيات

يتناول هذا الجزء مناقشة النتائج التي توصلت إليها الدراسة والتي هدفت إلى مقارنة ثلاث طرق Q3 و Fisher Z و G² في الكشف عن الاستقلال الموضوعي لفقرات الاختبار الوطني لضبط نوعية التعليم في الأردن في الصفين الثامن والعاشر الأساسيين والتوصيات المنبثقة عنها.

أولاً: مناقشة نتائج الدراسة

أظهرت النتائج أنه بالنسبة للصف الثامن، اتفقت جميع المؤشرات (Q3, Fisher Z, G2) في الكشف عن انتهاك افتراض الاستقلال الموضوعي، وبينت النتائج أن جميع المباحث قد انتهكت هذا الافتراض بشكل كبير، حيث كانت نسب أزواج الفقرات التي بينها انتهاك تتجاوز 90%، وتبين أن المؤشرين Q3, Fisher Z انسجما من حيث قدرتهما في الكشف عن انتهاك هذا

الافتراض لجميع المباحث، مع ملاحظة أن نسبة أزواج الفقرات التي بينها انتهاك لافتراض الاستقلال الموضعي باستخدام المؤشر Fisher Z كانت أقل مقارنةً مع المؤشر Q3، واتفقت نتائج الدراسة مع دراسة شن (Shen,1997) في اتفاق كلا المؤشرين في الكشف عن انتهاك افتراض الاستقلال الموضعي مع أفضلية للمؤشر Fisher Z، والسبب في ذلك يعود إلى وجود البواقي المعيارية باستخدام مؤشر Fisher Z، وأيضاً وجود معيار تطبيقي لهذا المؤشر كونه يحول البواقي الملاحظة إلى بواقي معيارية تتبع التوزيع الطبيعي، وكما أشار شن (Shen,1997) أن مؤشر Fisher z أكثر صدقا وثباتا وأكثر حساسية للاعتماد الموضعي عند مقارنته مع مؤشر Q3. وأظهرت نتائج الدراسة أن نسبة أزواج الفقرات التي بينها انتهاك لافتراض الاستقلال الموضعي باستخدام المؤشر G^2 لم تتغير باختلاف المباحث، حيث كانت النسبة تساوي 100% لجميع المباحث، حيث من الممكن عزو عدم التغير في نسب أزواج الفقرات التي بينها انتهاك لافتراض الاستقلال الموضعي إلى تأثير المؤشر G^2 بحجم العينة، وذلك لأن توزيع المؤشر G^2 شبيه بتوزيع مربع كاي و بدرجة حرية واحدة، ومن المعروف أن توزيع مربع كاي يتأثر بحجم العينة كما أشار هامبلتون وسوامينيثن (Hambleton & Swaminathan,1985)، وهناك دراسات مثل دراسة تومسون وبميرج (Thompson & Pommerich,1996) التي بينت أن مؤشر G^2 يتأثر بحجم العينة، حيث إن حجم العينة الضخم سوف يزيد من قيمة G^2 لأي زوج من أزواج الفقرات، وبالتالي فإن ذلك يزيد من نسبة أزواج الفقرات التي بينها انتهاك لافتراض الاستقلال الموضعي .

وأظهرت النتائج، أنه بالنسبة للصف العاشر الأساسي، كانت نسب أزواج الفقرات التي بينها انتهاك لافتراض الاستقلال الموضعي أقل مما هو عليه للصف الثامن الأساسي، ولوحظ انسجام المؤشرات الثلاثة في الكشف عن أعلى المباحث انتهاكاً لافتراض الاستقلال الموضعي، وكانت

في مبحث اللغة العربية، وكانت هذه النسب تساوي 0.986 و 0.940 و 1 لكل من Q3، Fisher z، G2 على التوالي. وبالرجوع إلى الوسيطات، أشارت النتائج أن المؤشرات الثلاثة قد انسجمت في الكشف عن أعلى المباحث انتهاكاً لافتراض الاستقلال الموضوعي، وكانت في مبحث اللغة العربية، وهذه الوسيطات تساوي 0.261 و 3.547 و 665 لكل من Q3، Fisher z، G2 على التوالي. ولوحظ انسجام المؤشرين Q3، Fisher z في الكشف عن أقل المباحث انتهاكاً لافتراض الاستقلال الموضوعي، وكانت في مبحث اللغة الإنجليزية، حيث كانت نسبة أزواج الفقرات التي بينها انتهاك لافتراض الاستقلال الموضوعي باستخدام Q3 و Fisher z تساوي 0.129 و 0.230 على التوالي. وبالرجوع إلى الوسيطات نلاحظ اتفاق المؤشرات الثلاثة في الكشف عن أقل المباحث انتهاكاً لافتراض الاستقلال الموضوعي، وكانت في مبحث اللغة الإنجليزية، حيث كانت الوسيطات للمؤشرات Q3 و Fisher z و G2 تساوي -0.013 و 1.128 و 5 على التوالي، وجميعها تؤكد أن هذا الافتراض لم يتم انتهاكه في مبحث اللغة الإنجليزية، وذلك حسب المعايير المعتمدة لكل مؤشر.

وفيما يخص المباحث الدراسية ودرجة حساسيتها لافتراض الاستقلال الموضوعي، أشارت النتائج المتعلقة بالصف الثامن الأساسي، أن أعلى المباحث انتهاكاً لافتراض الاستقلال الموضوعي كانت في مبحث الرياضيات والتربية الإسلامية باستخدام المؤشر Q3، حيث كانت نسبة أزواج الفقرات التي بينها انتهاك لافتراض الاستقلال الموضوعي تساوي 1، وفي مبحث اللغة العربية باستخدام المؤشر Fisher Z، حيث كانت نسبة أزواج الفقرات التي بينها انتهاك لافتراض الاستقلال الموضوعي تساوي 0.940، في حين لوحظ أن جميع المباحث كانت نسبة الفقرات المنتهكة لهذا الافتراض تساوي 100% باستخدام المؤشر G2. وبالرجوع إلى الوسيطات، أشارت النتائج إلى أن أكثر المباحث انتهاكاً لهذا الافتراض كانت في مبحث اللغة الإنجليزية باستخدام المؤشر Q3،

حيث كان الوسيط يساوي 0.308، ومبحث اللغة العربية باستخدام المؤشر **Fisher Z**، حيث كان الوسيط يساوي 14.154، ومبحث التربية الوطنية الاجتماعية باستخدام المؤشر G^2 ، حيث كان الوسيط يساوي 5.826. وفي المقابل نجد أن أقل المباحث انتهاكاً لافتراض الاستقلال الموضوعي كانت في مبحث التربية الوطنية والاجتماعية باستخدام المؤشر Q_3 ، حيث كانت نسبة أزواج الفقرات التي بينها انتهاك لافتراض الاستقلال الموضوعي تساوي 0.906، ومبحث التربية الإسلامية باستخدام المؤشر **Fisher z**، حيث كانت نسبة أزواج الفقرات التي بينها انتهاك لافتراض الاستقلال الموضوعي تساوي 0.886، في حين لاحظنا أن المؤشر G^2 كشف أن جميع المباحث قد انتهكت افتراض الاستقلال الموضوعي وبنسبة وصلت إلى 100%. وبالرجوع إلى الوسيطات، أشارت النتائج إلى أن أقل المباحث انتهاكاً لافتراض الاستقلال الموضوعي باستخدام المؤشر Q_3 مبحث التربية الاجتماعية والوطنية، وكان الوسيط يساوي 0.168، ومبحث الرياضيات باستخدام كل من المؤشرين **Fisher Z** و G^2 ، حيث كان الوسيط يساوي 9.024 و 1156 على التوالي.

وأشارت النتائج أنه بالنسبة للصف العاشر الأساسي، كانت أكثر المباحث انتهاكاً لافتراض الاستقلال الموضوعي باستخدام المؤشر Q_3 مبحث اللغة العربية، حيث كانت نسبة الفقرات التي بينها انتهاك لافتراض الاستقلال الموضوعي تساوي 0.938، ومبختي اللغة العربية والرياضيات باستخدام المؤشر **Fisher Z**، وكانت النسبة تساوي 0.638، ومبحث اللغة العربية باستخدام المؤشر G^2 ، وكانت النسبة تساوي 100%. وبالرجوع إلى الوسيطات، أشارت النتائج إلى أن المؤشرات الثلاثة قد اتفقت في الكشف عن أكثر المباحث انتهاكاً لافتراض الاستقلال الموضوعي، وكانت في مبحث اللغة العربية، حيث كانت الوسيطات تساوي 0.261 و 3.547 و 665 لكل من Q_3 و **Fisher z** و G^2 على التوالي. في حين كانت أقل المباحث انتهاكاً لمبحث اللغة

الإنجليزية باستخدام المؤشرين Q3 و Fisher z ، وكانت نسبة الفقرات التي بينها انتهاك لافتراض الاستقلال الموضوعي تساوي 0.129 و 0.230 على التوالي، في حين كانت أقل المباحث انتهاكاً باستخدام المؤشر G2 مبحث العلوم، وكانت النسبة تساوي 0.285. وبالرجوع إلى الوسيطات، أشارت النتائج إلى أن أقل المباحث انتهاكاً لافتراض الاستقلال الموضوعي مبحث اللغة الإنجليزية باستخدام المؤشرات الثلاثة، حيث كان الوسيط يساوي -0.013 و 1.128 و 5 لكل من Q3 و Fisher z و G2 على التوالي.

ثانياً: التوصيات

1. ينبغي على القائمين على الاختبار الوطني لضبط نوعية التعليم في الأردن أن يكونوا واعين لمخاطر عدم التحقق من الافتراضات القائمة عليها كل من النظرية الكلاسيكية في القياس ونظرية استجابة الفقرة، والتي يتم من خلالها تقدير معالم الفقرات والاختبار والمفحوصين، وخاصة افتراض الاستقلال الموضوعي للفقرات.
2. هناك العديد من الطرق التي يمكن استخدامها من أجل التخفيف من مقدار الارتباط الموضوعي في فقرات الاختبار، وكذلك التقليل من الأثر الذي يمكن أن ينجم عن هذا الارتباط على الإحصائيات المتعلقة بفقرات الاختبار. ويقول كل من الن و سدويكس (Allen & Sudweeks, 2001) بأن الحل المباشر هو كتابة فقرات مستقلة إحصائياً عن بعضها البعض. ومن الطرق المتبعة لتحليل العلامات لاختبار يحتوي على فقرات مرتبطة هو استخدام ما يسمى testlet scoring. وهذا الأجراء يتطلب تصحيح الفقرات داخل كل مجموعة من الفقرات (تشكل عنقود) باعتبارها فقرة واحدة. واستجابة المفحوص لهذه الفقرات تأخذ علامة واحدة تتراوح ما بين (صفر إلى ن) ،حيث (ن) عدد الفقرات داخل العنقود.

3. ذكر كيلر وسواميناثن وسيريسي (Killer, Swaminathan & Sireci, 2003) بأن من الأساليب المتبعة لتصحيح الفقرات المرتبطة موضعياً هو استخدام نموذج استجابة العنقود testlet response model الذي يحتوي على معلمة للتفاعل ما بين المفحوص والعنقود، ويستخدم هذا النموذج في كل من النموذج اللوجستي ثنائي وثلاثي المعلمة، وفي إتباع هذا الإجراء يتم تصحيح الفقرات المرتبطة موضعياً باستخدام التصحيح الثنائي، مع الأخذ بعين الاعتبار الارتباطات الموضعية للفقرات داخل العنقود.

- Allen, S. & Sudweeks, R. (2001). *Identifying and managing local item dependence in context-dependent item sets*. (ERIC Document Reproduction Service No. ED 453236).
- Chen, W.H. & Thissen, D. (1997). Local dependence indexes for item pairs using item response theory. *Journal of Educational and Behavioral Statistics*, 22, 265-289.
- 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* (3rd 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.
- Killer, L., Swaminathan, H. & Sireci, S. (2003). Evaluating scoring Procedures for context– dependent item sets. *Applied Measurement in Education*, 16(3), 207-222.
- Kim, S.; Cohen, A. & Lin, Y. (2005). LDID: A Computer Program for Local Dependence Indices for Dichotomous Items. Version 1.0.
- 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.

Reese,L.(1999).*A classical Test Theory perspective on LSAT local Item Dependence.*

Report No.LSAC-R-96-01.Law School Admission council, Newtown,

PA(Document Reproduction Service No.ED 469172).

Shen,L.(1997).Quantifying Item Dependency by Fisher's Z.paper presented At the

annual meeting of the American Educational Research Associat-Ion, Chicago,

IL, March24-28,1997(ERIC Document Reproduction Service No.ED 410241).

Thompson,T & pommerich,M,(1996).*Examining the sources and effects of local item*

*dependence.*paper presented at the Annual meeting of the American Educational

Research Association. NewYork,NY, April8-12,1996(ERIC Document

Reproduction Service No.ED 400311).

Tuerlinckx,F. & De Boeck,P.(2001).Nom-modeled item interactions lead to distorted

discrimination parameters:A case study.*Methods of Psychological Research on*

line,6(2).

Yen, W. (1984). Effects of local item dependence on the fit and equating performance

of the three- parameter logistic model. *Applied psychological Measurement*, 8,

125-145.

Xie.y.(2001). Dimensionality, Dependence, or Both?An Application of the Item

Bundle Model to Multidimensional Data. Policy,Organization, Measurement

& EvaluationGraduate School of Education:University of California at erkeley.

Yen,W.(1993).Scaling Performance Assessments:Strategies for Managing Local item

Dependence.*Journal of Educational Measurements*,30(3),pp187-213.

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

Zenisky, A., Hambleton, R., & Sireci, S. (2002). Identification and evaluation of local item dependencies in the medical college admissions test. *Journal of ducational Measurement, 39*, 1-16.