DOI: http://dx.doi.org/10.24200/jeps.vol14iss1pp41-53

فعالية مؤشرات مطابقة الفرد في نماذج استجابة الفقرة عند اختلاف قوة الارتباط الموضعي بين الفقرات

يعقوب زاهر الشقصي ويوسف عبدالقادرابوشندي* وراشد سيف المحرزي
جامعة السلطان قابوس، سلطنة عمان

استُلم بتاريخ: ٢٠١٩/٢/٣ قُبل بتاريخ: ٢٠١٩/٥/٣٠

ملخص: هدفت هذه الدراسة فحص فعالية مؤشرات مطابقة الفرد (رايت الموزون، ودرازجو، والمحرزي الموزون) في نماذج استجابة الفقرة عند اختلاف قوة الارتباط الموضعي بين الفقرات (٠.٠، ٠.٣، ٠.٦، ٠.٩) باستخدام بيانات مولدة، وقد تم توليد استجابات ٤٠ عينة تحتوي كل منها على ١٠٠٠٠فرد (بمجموع ٤٠٠٠٠٠ فرد) على اختبار مكوّن من ٦٠ فقرة تراوحت معالم تمييزها بين ٠,١٩ و ١.٧٩ ومعالم صعوبتها بين -٢ الى +٢ منها نسبة ٢٠% بينها ارتباط موضعي، وتوزعت قدرات المفحوصين بشكل طبيعي. تم التحقق من افتراضات نظرية استجابة الفقرة في البيانات باستخدام طريقتي: التحليل العاملي الاستكشافي وتحليل البواقي باستخدام برمجية NOHARM لافتراض أحادية البعد، وطريقة مؤشر Q_3 لافتراض الاستقلال الموضعي. أظهرت النتائج زيادة نسبة الأفراد ذوي الأنماط غير المطابقة بزيادة قوة الإرتباط الموضعي عند المؤشرات الثلاثة (رايت الموزون، درازجو، المحرزي الموزون)، وتوصلت أيضاً إلى أن نسبة عدد الأفراد غير المطابقين في مؤشر رايت الموزون أكبر من نسبة عدد الأفراد غير المطابقين في مؤشر درازجو وموؤشر المحرزي الموزون. كما وأظهرت المؤشرات الثلاثة نوعا من الثبات في خصائص توزيعها. وكذلك تشابه خصائص شكل توزيع مؤشري درازجو، والمحرزي. وأكبر معامل اتفاق توصلت له الدراسة كان بين مؤشري رايت الموزون ودرازجو.

كلمات مفتاحية: نظرية الاستجابة للفقرة، مؤشرات مطابقة الفرد، الارتباط الموضعي للفقرة.

Effectiveness of Person Fit Indices in Item Response Models with Different Degrees of Item Local Dependence

Yaqoub Z. Al Shaqsy,* Yousef A. Abu Shindi & Rashid S. Almehrizi
Sultan Qaboos University, Sultanate of Oman

Abstract: This study aimed to examine the effectiveness of person fit indices (Wright's weighted index, Drasgow index and Almehrizi's weighted index) in item response models with different degrees of item local dependence (0.0, 0.3, 0.6, and 0.9) using simulated item parameters. Item responses for 40 samples each with 10000 subjects (a total of 400000 subjects) were simulated on a test of 60 items. Item discrimination parameters ranged between 0.19 and 1.79 and item difficulty parameters ranged between -2 and +2. 20% of test items were manipulated to show local dependence for each level of local dependence degrees. Student ability was generated to follow a standard normal distribution. Assumptions of item response theory were examined in all data sets using exploratory factor analysis and residual analysis using NOHARM platform for unidimensionality and Q3 index for local independence. Results showed that there was an increase in the percentages of non-conforming persons when increasing the degree of items local dependence for the three person fit indices (Wright's weighted index, Drasgow index and Almehrizi's weighted index). Results showed also that the percentages of non-conforming persons were larger with Wright's weighted index than with Drasgow index and Almehrizi's weighted index. The distributional properties of the three indices showed relatively consistent in distributional properties. Drasgow index and Almehrizi's weighted index were very similar distributional properties. Also, there was a larger agreement index between Wright's weighted index and Drasgow index.

Keywords: Item response theory, person fit index, item local dependence.

*yousefaaa@squ.edu.om

اهتم علماء القياس بالتحقق من صدق وثبات الاختبارات والمقاييس النفسية والتربوية، وكان الهدف من وراء ذلك هو تحقيق أعلى درجة من الموضوعية والدقة في تلك الاختبارات، حيث تبرز أهمية صدق علامات الاختبار عند اتخاذ قرارات مهمة على مستوى الفرد، مثل قبول فرد في برنامج أو استثناء آخر.

ولأن أداء الاختبار وقياسه بدقة أمر بالغ الأهمية لصناع القرار، تعتبر إحصاءات مطابقة الفرد ضرورية للكشف عن الأنماط غير المطابقة للإختبار مع النموذج، حيث تعتمد القرارات التربوية بشكل كبير على القياس الدقيق لخصائص الفرد للتعرف على الأفراد المؤهلين ولمراقبة تقدمهم العلمي.

وتختلف المؤشرات الإحصائية لمطابقة الفرد من حيث معنى المؤشر والأساس الرياضي الذي بُني عليه، ألا أن جميعها تعمل على كشف أنماط إجابة الأفراد غير المطابقة، ويمكن أن يكشف مؤشر عن نمط إجابة غير مطابق، ومؤشر آخر لا يكشف عن هذا النمط. وقد تؤثر مجموعة من السمات على أداء الافراد في الاختبار مثل الغش، والقلق، والتخمين، والمرض، والتباطؤ في أداء الاختبار وغيرها، وهذه السمات قد تُظهر أنماطاً من الاستجابات غير المطابقة التي قد تؤدي إلى تقدير غير دقيق لقدرات الافراد ومعالم الفقرات (Yen, 1992).

تَفترض نظرية استجابة الفقرة وجود علاقة بين احتمال الإجابة الصحيحة للفقرة وقدرات الافراد على إجابة هذه الفقرة، ويمكن تفسير أداء الافراد على فقرات الاختبار في ضوء سمات معرفية أو سمات شخصية من خلال الاستجابة على مجموعة من الفقرات الاختبارية أو الشخصية يطلق عليها السمات الكامنة Latent Traits، والتي يستدل عليها من خلال تقدير أداء الافراد الملاحظ على فقرات الاختبار (Lord & Novick, 1968).

وتُعد نماذج نظرية استجابة الفقرة احادية البعد - بنوعيها ثنائية الاجابة Dichotomous response أو متعددة الإجابة polytomous - response دوال رياضية احتمالية، وتختلف الصيغة الرياضية لهذه النماذج باختلاف عدد معالم الفقرة المكونة لبنائها الرياضي، إذ تهدف هذه النماذج إلى تحديد العلاقة بين احتمالية إجابة الفرد على فقرة ما إجابة صحيحة وبين القدرة الكامنة التي تكمن وراء هذا الأداء، ومن هذه النماذج النموذج اللوجستي ثنائي المعلمة Two Parameter Logistic Model- 2PLM إذ يَفترض هذا النموذج أن الفقرات تختلف في معلمتي الصعوبة والتمييز، في حين تساوي معلمة التخمين من الصفر، وتعطى احتمالية الاجابة الصحيحة على الفقرة بالصيغة الرياضية الآتية (علام، ٢٠٠٥):

$$p_{ij} = \frac{e^{[a_i(\theta_j - b_i)]}}{1 + e^{[a_i(\theta_j - b_i)]}} \quad (1)$$

p_{ij}: احتمالية الإجابة للفرد j إجابة صحيحة للفقرة i.، b_i : معلمة الصعوبة للفقرة i، a_i: معلمة التمييز للفقرة i، θ_j: قدرة للفرد j. وقد استخدم هذا النموذج في هذه الدراسة لأن معلمة التخمين أقرب للبيانات الحقيقية منها للبيانات المولدة وذلك لتأثرها بعدد من السمات الشخصية لدى الأفراد يصعب التنبؤ بها من خلال البيانات المولدة، كما وتشير الدراسات إلى استقرار تقدير المعالم في النموذج ثنائي المعلم أكثر من ثلاثي المعلم. وللوصول إلى نتائج دقيقة يمكن الوثوق بها، فإن نماذج نظرية استجابة الفقرة تستند إلى افتراضات قوية يجب تحقيقها في البيانات قبل إجراء التقديرات للمعالم (Gruijter & Kamp, 2005):

افتراض أحادية البُعد Unidimensionaliy

تفترض نماذج نظرية استجابة الفقرة أحادية البُعد وجود قدرة أو سمة وحيدة كامنة تُفسر أداء الفرد على الاختبار، ويُقصد بافتراض أحادية البُعد كما يشير

إليه هامبلتون وسوامنثاون وروجرز (Hambleton, Swaminathan & Rogers, 1991) أن ثمة سمة معرفية واحدة تقف ورراء سلوك الفرد على الفقرة الإختبارية.

افتـراض التحـرر مـن السـرعة Lack of Speediness

حيث تفترض نماذج استجابة الفقرة، أن سرعة الفرد في الإستجابة على الفقرة الإختبارية لا تؤثر في نتيجته، مما يجعل إخفاق الافراد في الإجابة عن فقرات الاختبار يرجع إلى انخفاض قدراتهم وليس إلى سرعة استجابتهم.

افتراض الاستقلال الموضعي Local Independence

أشار هامبلتون وسوامينيثان (Hambleton& Swaminathan, 1985) إلى أن افتراض الاستقلال الموضعي يعني أن استجابة الافراد على فقرات الاختبار تكون مستقلة إحصائياً Statistically Independent عند مستوى قدرة معين، بمعنى ان إجابة الفقرة لا تعطي أية تلميحات أو معيقات لإجابة فقرة أخرى، وهذا الافتراض يوضح أن قدرة الفرد وخصائص الفقرة هي وحدها التي تؤثر في التحصيل أو الأداء على الاختبار. ويُعد افتراض الاستقلال الموضعي للفقرة من الافتراضات الرئيسة لكل من النظرية التقليدية في القياس (CTT) ونظرية استجابة الفقرة (IRT) (Lee, 2004).

وأوضح هامبلتون وسوامينا‌ثان (Hambleton & Swaminathan, 1985) أن الاستقلال الموضعي يمكن تحققه رغم ارتباط هذه الفقرات عند المستويات المتباينة من السمة أو القدرة المقيسة، إذ أنه يتطلب عدم ارتباط هذه الفقرات عند تثبيت مستوى هذه السمة أو القدرة المقيسة.

وتتوقع نماذج استجابة الفقرة أحادية البعد وجود متصل للسمة المراد قياسها، ويمكن تقدير احتمال إجابة الفرد على فقرة إجابة صحيحة إذا علم موقعه على هذا المتصل، ويكوَن هذا الاحتمال دالة رياضية متزايدة مطردة لمواقع الأفراد على متصل السمة، وبذلك يزداد احتمال استجابة الفرد بشكل صحيح على الفقرة بزيادة مقدار السمة لديه. وتمثل هذه الدالة الرياضية بيانياً بمنحنى تزايدي يُعرف بمنحنى خصائص الفقرة Item Characteristic Curve (ICC)، ففي حالة الفقرات ثنائية الاستجابة يمثل المنحنى العلاقة بين قدرات الأفراد (على محور السينات)، وإحتمال الاستجابة الصحيحة (على محور الصادات)، وفي حالة الفقرات متعددة التدريج فيمثل المنحنى انحدار احتمال الاستجابات لكل مستوى من التدريج (تدريج ليكرت مثلاً). أما في حالة الفقرات ثنائية البعد فيتحول منحنى خصائص الفقرة إلى سطح مميز للفقرة بحيث تقع كل سمة على محور، واحتمال الاستجابة الصحيحة على المحور الثالث (علام، ٢٠٠٥).

مؤشرات مطابقة الفرد

يُعـرّف مؤشـر مطابقـة الفـرد بأنـه مؤشـر إحصائي يحدد المدى أو البُعـد بـين البيانـات الفعلية المتمثلـة باسـتجابات الافـراد والقـيم المتوقعة من خـلال النمـوذج المسـتخدم مـن خلال مقارنة قيم هـذه المؤشـرات مـع قيمـة حرجـة لتحديـد أنمـاط الإجابـة المطابقـة وأنمـاط الإجابـة غيـر المطابقـة (Lopez & Montesinos, 2005; Meijer & Sijtsma, 2001).

وقد تناولت هذه الدراسة المؤشرات الإحصائية الآتية:

مؤشر رايت وماستر (Wright & Masters, 1982) اهتم رايت وماسترز (Wright & Masters, 1982) بدراسة المؤشرات المبنية على البواقي وقد طورا مؤشرين هما مؤشر المطابقة الخارجية (Outfit) ومؤشر المطابقة الداخلية (Infit). ويتصف مؤشر المطابقة الخارجية بتأثره في إجابة الفقرات السهلة جداً بصورة خاطئة من قبل

الأفراد ذوي القدرات العالية، وفي إجابة الفقرات الصعبة جداً بصورة صحيحة من قبل الافراد ذوي القدرات المتدنية (Smith, 1991).

لقد بدأت نشأة وتطور هذين المؤشرين من خلال مربع متوسط رايت Zij (Wright, 1977) الذي يعتبر واحد من مؤشرات مطابقة الفرد حيث تستخدم Z_{ij} كمؤشر للاستجابة غير المتوقعة للفرد j على الفقرة i (Smith, 1991).

$$Z_{ij} = \frac{x_{ij}-p_{ij}}{\sqrt{p_{ij}q_{ij}}} ; i = 1,2,\dots,n. \quad (2)$$

x_{ij}: درجة ملاحظة الفرد، p_{ij}: احتمال إجابة الفقرة i من قبل الفرد j إجابة صحيحة، q_{ij}: احتمال الإجابة الخاطئة وتساوي $(1\text{-}p_{ij})$. وقد استخدم رايت (Wright, 1977) الدرجة Z_{ij} لاشتقاق مؤشري مربع المطابقة الكلي: الموزون و غير الموزون.

مؤشر مربع المطابقة الكلي غير الموزون : UMS_j

$$UMS_j = \frac{1}{n-1}\sum_{i=1}^{n} Z_{ij}^2 = \frac{1}{n-1}\sum_{i=1}^{n}\frac{(x_{ij}-p_{ij})^2}{p_{ij}q_{ij}} \quad (3)$$

مؤشر مربع المطابقة الكلي الموزون :WMS_j

$$WMS_j = \frac{\sum_1^n (p_{ij}q_{ij})z_{ij}^2}{\sum_{i=1}^n p_{ij}q_{ij}} = \frac{\sum_1^n (x_{ij}-p_{ij})^2}{\sum_{i=1}^n p_{ij}q_{ij}} \quad (4)$$

وللتقليل من أثر الإجابات المتطرفة من قبل الافراد الذين تبعد قدرتهم عن مستوى صعوبة الفقرة، طورا مؤشر المطابقة الداخلية والذي يتصف بحساسيته في الكشف عن الإجابات غير المتوقعة للفقرات القريبة من مستوى قدرة الفرد (Smith, 1991). وفي الدراسة الحالية تم تناول مؤشر المطابقة الداخلية، ويعطى بالعلاقة بالمعادلة الآتية:

$$Zw = \left[\left(WMS_j{}^{1/3} - 1\right)\left(\frac{3}{\sigma}\right)\right] + \left[\left(\frac{\sigma}{3}\right)\right] \quad (5)$$

حيث $\sigma = \frac{\sqrt{\sum_{i=1}^n p_{ij}q_{ij}(p_{ij}-q_{ij})^2}}{\sum_{i=1}^n p_{ij}q_{ij}}$ ويتوقع ان يتوزع مؤشر Zw توزيعا طبيعياً (0, 1).

مؤشر الاحتمالية (Drasgow, Levine & William, 1985)

يعتمد هذا المؤشر على الأرجحية العظمى Maximum Likelihood، ويعتمد المؤشر Lz على المؤشر Lo الذي تم اقتراحه من قبل ليفين وروبن (Levine & Rubin, 1979)، ونظرا لوجود العديد من المشكلات في المؤشر الاحصائي Lo ، قام درازجو وليفين ووليام (Drasgow, Levine & William, 1985) بتطوير صورة معيارية للمؤشر Lz والذي يتوقع أن يتوزع توزيعاً طبيعياً بوسط حسابي صفر وانحراف معياري واحد. ويعطى بالعلاقة الآتية:

$$Lz = \frac{\sum_{i=1}^n (x_{ij}-p_{ij})In(p_{ij}/q_{ij})}{\sqrt{\sum_{i=1}^n p_{ij}q_{ij}[In(p_{ij}/q_{ij})]^2}} \quad (6)$$

مؤشر مربع البواقي للمحرزي *(Al-mahrazi, 2003)*

وهو معامل جديد لملائمة الفرد وكيفية تفسيره واستخدامه لكشف الافراد ذوي الاستجابات غير المطابقة، ويعتمد مؤشر المحرزي الموزون (Al-mahrazi, 2003) على البواقي، وهو مشابه لمؤشر متوسط رايت (Wright, 1979) لمطابقة الفرد، من حيث آنه يوظف منهج البواقي بين استجابة الفرد الملاحظة، واحتمال الإجابة الصحيحة على الفقرة. ويتوقع ان يتوزع المؤشر توزيعا طبيعيا (٠، ١)، ويعطى بالعلاقة الآتية:

$$Rz_j = \frac{\sum_{i=1}^n (x_{ij}-p_{ij})(p_{ij}-0.5)}{\sqrt{\sum_{i=1}^n p_{ij}q_{ij}(p_{ij}-0.5)^2}} \quad (7)$$

ويتشابه مؤشر رايت مع مؤشر المحرزي في ان كلاهما يستخدم البواقي المعيارية بين إجابة الفرد الملاحظة، واستجابته المتوقعة وفق نظرية الاستجابة للفقرة، إلا أن مؤشر المحرزي يختلف اجرائياً عن مؤشر رايت في (Almehrizi, 2010).

اهتمت الدراسات السابقة بتفحص دقة وفاعلية مؤشرات مطابقة الفرد، والمقارنة بينها، ومنها دراسة سنيجدرز (Snijders, 2001) التي تفحصت أثر طول الاختبار (١٠، ٣٠، ٥٠) فقرة على فاعلية الاحصائي L_z في الكشف عن أنماط الاستجابة غير المطابقة والتي تعود لعوامل: السرعة، الغش، وفقدان الدافعية. وقد أظهرت الدراسة ان قوة الاحصائي كانت عالية المعدل في الكشف عن أنماط عدم المطابقة والتي تعود لعامل الغش عند مستويات القدرة المتدنية، وكانت عالية أيضا في الكشف عن أنماط عدم المطابقة التي تعود لعامل فقدان الدافعية عند مستويات القدرة العالية، وتزداد قوة الاحصائي في الكشف بزيادة طول الاختبار. وأما قوة الاحصائي في الكشف عن أنماط الاستجابة غير المطابقة التي تعود لعامل السرعة، فقد كانت منخفضة جداً عبر جميع مستويات القدرة.

وأجرى لوبيز ومنتسينوس (Lopez & Montesinos, 2005) دراسة قارنا فيها بين خمسة مؤشرات مختلفة لمطابقة الافراد ضمن نموذج راش باستخدام بيانات مولدة تحت ظروف مختلفة في حجم العينة (١٠٠، ٢٥٠، ٥٠٠، ١٠٠٠) ومدى صعوبة الفقرات [١،١-] و [٢،٢-] وشكل توزيع القدرة (طبيعي، ومنتظم) وطول الاختبار (١٥، ٣٠)، ولتحقيق غرض الدراسة قام الباحثان بتوليد بيانات مطابقة للنموذج الثلاثي ثم عمل تقدير للبيانات على نموذج راش باستخدام برنامج ConQuest، أظهرت نتائج الدراسة أن الوسط والانحراف المعياري للمؤشرين رايت ودرازجو (Lz, Zw) لم يكونا كما هو متوقع لهما بسبب استخدام المعالم المقدرة بدلاً من القيم الحقيقية، على الرغم من اختلاف الأساس الرياضي لهما ألا انهما أظهرا نفس السلوك تقريباً في هذه الدراسة وتحت نفس الظروف المختلفة حيث انه كلما ضاق مدى صعوبة الفقرات وزاد طول الاختبار تحسنت خصائص التوزيع الطبيعي لهما، وأظهرت نتائج الدراسة أن المؤشر (ELC4) للين وتاتسوكا (linn&Tasouka, 1983) المبني على التباين المشترك وسطه وانحرافه المعياري قريبان من القيم المناظرة لهما في التوزيع الطبيعي، وأيضا كان معدل الكشف عن الأنماط غير المطابقة للمؤشرين (ELC2z, ELC4z) بين (٥%-١٠%)وهذه النسب أعلى منها للمؤشرات الأخرى (Lz, Zw, Zu).

وأجرى المحرزي (Almehrizi, 2010) دراسة قارنت من خلال بيانات مولدة الخصائص الإحصائية لمعاملات المحرزي الموزون وغير الموزون، ومعدلات قوته للكشف عن الأفراد ذوي الاستجابة غير المطابقة مع معاملات رايت الموزون وغير الموزون، لاثنتي عشر مجموعة من البيانات والتي نتجت عن اختبارين مختلفين في الطول (عدد الفقرات= ١٥، ٥٠)، وثلاث مستويات مختلفة من الصعوبة (متدنية، متوسطة، مرتفعة)، ومستويين مختلفين من التمييز (متدني، عالي). وقد أظهرت النتائج أن معاملات المحرزي الموزون وغير الموزون حصلت على خصائص إحصائية أفضل تحت ظروف اختبارية مختلفة بدرجة أعلى عن معاملات رايت.

ومن الملاحظ أن الدراسات التي تناولت مؤشرات مطابقة الفرد لم تتناول مقارنات فيما بين المؤشرات عند اختلاف قوة الارتباط الموضعي بين الفقرات، الأمر الذي تظهره هذه الدراسة لمدى تأثر المؤشر باختلاف قوة الارتباط الموضعي بين الفقرات.

مشكلة الدراسة

تتخذ المؤسسات قرارات تهم الفرد بناء على نتائج الأختبارات، لكن ليس من الضروري أن تكون هذه الدرجات صادقة، فقد تكون هناك استجابات غير مطابقة تعود لأسباب متعددة قد تؤدي الى تقدير غير دقيق لقدرة الافراد وكذلك لمعالم الفقرات، وهذا يؤثر على صدق وثبات الاختبار (Meijer & Van, 1999).

وتتعلق الفقرات المكونة للاختبارات بمحتوى معين في غالب الأحيان، وينتج عن

ذلك ارتباط موضعي بينها، وبالتالي لا يتحقق افتراض الاستقلال الموضعي للفقرات والذي هو حجر الزاوية لنماذج نظرية استجابة الفقرة، وتكمن مشكلة هذه الدراسة في درجة تأثير ارتباط الفقرات الموضعي على دقة تقديرات مؤشرات مطابقة الفرد، ونظراً لصعوبة تطوير اختبار يتحقق فيه افتراض الاستقلال الموضعي، فتظهر حاجتنا لمعرفة المؤشر الاحصائي المناسب لمطابقة الافراد وفاعليته في مختلف حالات الارتباط الموضعي، مما قد يساعد الباحثين ومطوري الاختبارات لاختيار المؤشر الاحصائي الأكثر فاعلية عند معرفة قوة الارتباط الموضعي بين فقرات الاختبار، وبالتالي يقودنا إلى نتائج أكثر دقة وضبط، وقدرة على التحكم والتصنيف، وهو الهدف الأسمى فيما يخص مطابقة الفرد. ولمعرفة المؤشر الأكثر ملائمة عند اختلاف مستوى الارتباط الموضعي (٠.٠، ٠.٣، ٠.٦، ٠.٩) بين الفقرات، جاءت فكرة هذه الدراسة والتي تحاول الإجابة عن الاسئلة الآتية:

١. **ما نسبة عدد الأفراد غير المطابقين التي يكشفها كل مؤشر من مؤشرات مطابقة الفرد عند كل حالة من حالات الارتباط الموضعي؟**
٢. **ما مستوى اتفاق مؤشرات مطابقة الفرد في الكشف عن أنماط استجابة الافراد غير المطابقة للنموذج عند حالات الارتباط الموضعي؟**
٣. **هل يؤثر الارتباط الموضعي بين الفقرات على خصائص التوزيع لمؤشرات مطابقة الفرد؟**

أهمية الدراسة

تعتمد دقة تقدير النماذج الرياضية لنظرية استجابة الفقرة للمعالم (الفقرات والمفحوصين والاحتمالية) على تحقق افتراضاتها في البيانات، وكذلك على مطابقة أنماط إستجابة المفحوصين لهذه النماذج، وبناء عليه اجتهد المختصون في اشتقاق الطرق والأساليب المناسبة للتحقق من الافتراضات، وطرق المطابقة، ومما لا يحتمل الشك أن طرق الكشف عن الأنماط غير المطابقة للنماذج تفترض تحقق الإستقلال الموضعي بين فقرات الاختبار، ولكن تحقق هذا الافتراض ليس بالأمر اليسير في البيانات الاختبارية، مما يثير قلق المشتغلون في هذا المجال حول تاثير عدم تحققه على دقة وفاعلية هذه الطرق، ويتساءلوا دائما حول حجم هذا التأثير.

جاءت هذه الدراسة والتي تُعد من الدراسات الحديثة، والتي من شأنها مساعدة الباحثين للتعرف على فاعلية مؤشرات مطابقة الفرد في نماذج استجابة الفقرة عند إختلاف قوة الارتباط الموضعي بين الفقرات، وتظهر أهميتها من خلال مقارنتها بين مؤشرات مطابقة الفرد (رايت الموزون، درازجو، المحرزي الموزون) عند وجود أزواج فقرات بينها ارتباط موضعي لمعرفة فاعلية المؤشرات عند قوة ارتباط معين، لاختيار المؤشر المناسب لكل اختبار عند معرفة قوة الارتباط بين فقراته، حيث لم يتم مقارنة هذه المؤشرات في دراسات سابقة على حد علم الباحثون في ظل ظروف أحادية البُعد واختلاف العلاقة الارتباطية (٠.٠، ٠.٣، ٠.٦، ٠.٩) بين الفقرات. وتُوفر هذه الدراسة معلومات تساعد علماء النفس والقياس وأصحاب القرار والتربويين لانتقاء المؤشرات لمطابقة الفرد الأكثر فاعلية عند وجود ظروف معينة للفقرات، من أجل تحقيق الصدق والعدالة في استخدام علامات الاختبار التي يتم بناء عليها اتخاذ قرارات تهم الفرد والمجتمع.

محددات الدراسة

١. تقتصر الدراسة على استخدام النموذج اللوجستي ثنائي المعلم 2-Parameters

(Logistic Model 2PLM) من نماذج نظرية الاستجابة للفقرة.

٢. استخدام مؤشر مربع البواقي (Al-mahrazi, 2003) ، ومؤشر البواقي المعيارية (Wright & Master, 1982)، ومؤشر الاحتمالية (Drasgow, Levine & William, 1985).

٣. تقتصر الدراسة على أربعة مستويات للارتباط الموضعي بين الابعاد (٠.٠، ٠.٣، ٠.٦، ٠.٩).

٤. نسبة واحدة لأزواج الفقرات المرتبطة موضعياً ١٢%.

٥. اعتمدت النتائج التي تم التوصل إليها على البرامج الحاسوبية المتوافرة في تحليل البيانات.

٦. اعتمدت الدراسة على القيم الحقيقية لمعالم الصعوبة والتمييز من دراسة أبو شندي (٢٠٠٨).

مصطلحات الدراسة

مطابقة الفرد Person Fit: يعرفه الشريفين (٢٠٠٣) بأنه درجة المطابقة بين النموذج الرياضي الذي يتم اختياره ونمط استجابة الفرد على الفقرات، ويمكن قياس درجة المطابقة بواسطة احصائي مطابقة الفرد. وبين ميجر (Meijer, 1996) أن الخصائص الفردية للمفحوصين من الأسباب التي تؤدي إلى حدوث أنماط غير مطابقة؛ كالغش، والتخمين، والقلق، والدافعية، والتباطؤ في أداء الأختبار، والتكاسل، وطرح ميجر مثالاً افتراضياً لاختبار مكون من ١٢ فقرة متدرجة في الصعوبة (من السهل إلى الصعب)، واشار إلى أنه من الممكن أن يقف سلوك الغش وراء عدم مطابقة نمط الاستجابة:١، ١، ٠، ١، ٠، ١، ٠، ٠، ٠، ١، ١، ١. وسلوك التباطؤ وراء عدم مطابقة نمط الاستجابة: ١، ١، ١، ١، ١، ١، ٠، ٠، ٠، ٠، ٠، ٠. وفسر ذلك بأن قدرة الفرد الأول المنخفضة لم تمنعه من الاستجابة الصحيحة على الفقرات الأولى السهلة، وسيلجأ للغش لبقية الفقرات، اما عن الفرد البطيء فهو حريص في أداءه على فقرات الاختبار، فلا ينتقل من فقرة إلى أخرى الا بعد التحقق من صحة اجابته، وهذا يؤدي الى انتهاء الوقت قبل أن يكمل الاختبار.

مؤشر مطابقة الفرد: مؤشر احصائي يحدد المدى أو البُعد بين البيانات الفعلية (استجابة الافراد) والقيم المتوقعة من خلال النموذج المستخدم، وتقارن قيم مؤشر المطابقة مع قيمة حرجة واحدة لتحديد أنماط الإجابة المطابقة وأنماط الإجابة غير المطابقة (Karabatsos, 2003).

الارتباط الموضعي بين زوج من الفقرات: معامل الارتباط بين البواقي للمفردتين عند نفس المستوى من القدرة وتعتمد القيمة المطلقة ٠.٠٥ للمؤشر Q_3 كحد فاصل بين ازواج الفقرات التي بينها انتهاك لافتراض الاستقلال الموضعي من غيرها من الفقرات (Hambleton & Swaminathan, 1985).

فعالية مؤشر مطابقة الفرد: مدى قرب أو بعد عدد نفس الافراد غير المطابقين للنموذج في حالة الارتباط الموضعي مع عدد الافراد غير المطابقين للنموذج في حالة الاستقلال الموضعي(Hatti, 1985).

المنهجية والاجراءات

عينة الدراسة

تكون أفراد الدراسة من ٤٠ عينة استجابات ثنائية Dichotomous مولدة عشوائياً باستخدام برمجية Excel بحيث احتوت كل عينة على استجابات ١٠٠٠٠ مفحوصاً (بما مجموعه ٤٠٠٠٠٠ فرد) على اختبار مكون من ٦٠ فقرة، منها ١٢ زوج ٢٠% بينها ارتباط موضعي. نتجت مجموعات البيانات من خلال التحكم بمستويات متغير درجة الارتباط الموضعي للفقرات (٤ درجات: صفر، ٠.٣، ٠.٦، ٠.٩)؛ واجري ذلك باستخراج قيم الدرجة الاحتمالية $Pi(\theta j)$ حيث تم توليد

استجابات الافراد على الفقرات في حالات الارتباط الموضعي المختلفة، بعدها تم تكرار الحالة الأولى عشر مرات باختلاف مجموعات (القدرات) مع ثبات معالم الفقرات (الصعوبة والتمييز)، وتم الحصول في هذه الحالة على عشر عينات تشترك فيما بينها في معالم الفقرات وتختلف في القدرات، وكذلك في الحالات الثلاث الأخرى.

التحقق من الارتباط الموضعي

للتأكد من دقة توليد الحالات الأربع المختلفة من قوة الارتباط الموضعي (٠.٠، ٠.٣، ٠.٦، ٠.٩)، تم استخدام مؤشر Q_3 لحساب الارتباط الموضعي بين زوج الفقرات ١٢ في كل عينة من العينات، حيث اقترحت ين (Yen, 1992) هذا المؤشر وهو معامل الارتباط للبواقي. ويبين جدول١ المتوسط الحسابي للارتباط بين بواقي الفقرات Q_3 للفقرات الاثنتا عشرة في العينات العشر في كل حالة من الحالات الأربع للمعالم المقدرة.

جدول ١

الوسط الحسابي للارتباط بين بواقي الفقرات Q_3 في الحالات الأربع من الارتباط الموضعي

قوة الارتباط الموضعي	الوسط الحسابي لمؤشر Q_3 للعينات العشر
٠.٠	٠.٠١
٠.٣	٠.٢٧
٠.٦	٠.٦٥
٠.٩	٠.٨٩

يلاحظ من جدول ١ التقارب الكبير بين قيم المتوسطات الحسابية للمؤشر Q_3 للمعالم المقدرة من قيمة قوة الارتباط الموضعي المقترحة، وفي ذلك اشارة واضحة على دقة توليد البيانات ضمن شرط اختلاف الارتباط الموضعي بين الفقرات.

التحقق من افتراض أحادية البُعد

تم التحقق من افتراض أحادية البُعد للنموذج المستخدم لما له من أثر على دقة التقديرات، وقد استخدمت الدراسة عدة طرق لذلك:

التحليل العاملي الاستكشافي بطريقة المكونات الرئيسية Principal Component Analysis: استخدم برنامج SPSS لحساب قيمة الجذور الكامنة للعينات، حيث استخرجت جميع العوامل التي لها جذر الكامن Eigenvalue أكبر من ١، وقورنت نسبة التباين المفسر Explained Variance للعامل الأول والثاني، وتوصلت النتائج إلى زيادة النسبة بين الجذر الكامن للعامل الأول الى الجذر الكامن للعامل الثاني عن القيمة ٢ كما هو موضح في جدول ٢ ،وذلك يحقق افتراض أحادية البعد في جميع العينات حسب رأي ريكيس (Reckase, 1985) كحد أدنى لتحقق افتراض أحادية البعد، وهذا يعني وجود عامل واحد طاغي يمكن ارجاع التباين بين الافراد إليه في كل عينة من العينات العشر، أي أن هناك سمة واحدة أو قدرة واحدة تقيسها فقرات الاختبار.

جدول٢

قيمة الجذر الكامن والتباين المفسر للعينة الاولى ونسبة التباين بين العامل الأول والعامل الثاني في الحالات الأربع.

قوة الارتباط الموضعي		العامل الاول	العامل الثاني	نسبة التباين
٠.٠	الجذر الكامن	٨.٣٤	١.٤٨	٥.٦٣
	التباين المفسر	١٣.٩١	٢.٤٧	
٠.٣	الجذر الكامن	٨.٢١	١.٨٦	٤.٤١
	التباين المفسر	١٣.٦٨	٣.١	
٠.٦	الجذر الكامن	٨.٧١	١.٩٧	٤.٤١
	التباين المفسر	١٤.٥١	٣.٢٩	
٠.٩	الجذر الكامن	٩.٠٥	١.٩٨	٤.٥٦
	التباين المفسر	١٥.٠٨	٣.٣٠	

كما وجاءت التمثيلات البيانية للجذور الكامنة Scree Plot لجميع العينات في الأربع حالات، لتؤكد على وجود انحدار واضح في التمثيل البياني لقيم الجذور الكامنة لجميع العينات وهذا يشير الى تحقق أحادية البعد فيها وفق ما أكده هايز (Hayes, 2012)

تحليل البواقي Residual Analysis: لتفحص تحقق افتراض احادية البعد بالطرائق التي تعتمد على نظرية استجابة الفقرة، تم إجراء تحليل البواقي Residual Analysis باستخدام برنامج NOHARM، حيث يتم فحص الموائمة بين عدد الابعاد التي يتم تحديدها والبيانات من خلال حساب مصفوفة البواقي Residual Matrix لإيجاد قيمة الجذر التربيعي لمتوسطات مربعات البواقي Root Mean Square of Residuals (RMSR)، ومؤشر تاناكا TANAKA index.

وقد بين ميلر (Miller, 1991) أنه إذا كانت قيمة مؤشر RMSR قريبة من الصفر فهذا دليل على تحقق أحادية البُعد للبيانات، وبالوقت نفسه إذا كانت قيمة مؤشر تاناكا Tanaka Index قريبة من الواحد يُعد دليلا آخر على أحادية البعد، ويلخص جدول ٣ الوسط الحسابي لقيم مؤشري: RMSR وتاناكا للبيانات في العينات.

جدول٣

الوسط الحسابي لقيم مؤشري:RMSR تاناكا للعينات العشر في الحالات الأربع من الارتباط الموضعي

قوة الارتباط الموضعي	RMSR	مؤشر تاناكا
٠.٠	٠.٠٠١٨	٠.٩٩٧٥
٠.٣	٠.٠٠٥٦	٠.٩٧٨٢
٠.٦	٠.٠١٢٦	٠.٩٠٥٠
٠.٩	٠.٠١٥٣	٠.٨٧١٧

يتضح من النتائج في جدول ٣ بأن قيمة الجذر التربيعي لمتوسط مربعات البواقي RMSR في جميع الحالات يقترب من الصفر، وبالوقت نفسه جاءت قيمة مؤشر تاناكا للمطابقة قريبة من الواحد، مما يبين تحقق افتراض أحادية البعد في البيانات، كما ويمكن ملاحظة زيادة قيمة مؤشر RMSR بزيادة قوة الارتباط الموضعي، ونقص قيمة مؤشر تاناكا، وهذا اشارة واضحة على العلاقة التي بينها البعض من المختصين بين افتراضي أحادية البعد والاستقلال الموضعي.

نتائج الدراسة

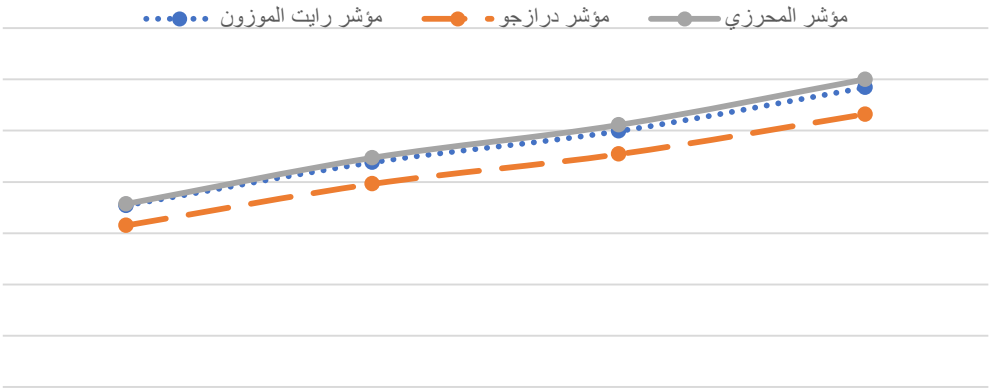
للإجابة عن السؤال البحثي الأول "ما نسبة عدد الأفراد غير المطابقين التي يكشفها كل مؤشر من مؤشرات مطابقة الفرد عند كل حالة من حالات الارتباط الموضعي؟" حسب الوسط الحسابي والنسبة لعدد الانماط غير المطابقة للعينات العشر في الحالات الاربع للارتباط الموضعي بين فقرات الاختبار لكل مؤشر من مؤشرات المطابقة ويظهر جدول ٤ ذلك.

من جدول٤ نلاحظ زيادة عدد الأفراد ذوي الانماط غير المطابقة بزيادة قوة الارتباط الموضعي عند المؤشرات الثلاثة (رايت الموزون- درازجو- المحرزي الموزون)، حيث تظهر النتائج أن فعالية مؤشرات الكشف عن الأنماط غير المطابقة تتأثر بالارتباط الموضعي بين الفقرات. وظهر ذلك من خلال مقارنة حالات الارتباط الموضعي (٠.٩، ٠.٦، ٠.٣) مع الحالة الأولى

جدول٤

الوسط الحسابي ونسبته عدد الانماط غير المطابقة التي كشفتها مؤشرات لحالات الارتباط الموضعي

قوة الارتباط الموضعي	مؤشر رايت الموزون		مؤشر درازجو		مؤشر المحرزي الموزون	
	الوسط الحسابي	النسبة	الوسط الحسابي	النسبة	الوسط الحسابي	النسبة
٠.٠	٣٥٤	٠.٠٣٥	٣١٥	٠.٠٣١	٣٥٧	٠.٠٣٥
٠.٣	٤٣٨	٠.٠٤٣	٣٩٦	٠.٠٣٩	٤٤٧	٠.٠٤٤
٠.٦	٤٩٩	٠.٠٤٩	٤٥٤	٠.٠٤٥	٥١١	٠.٠٥١
٠.٩	٥٨٤	٠.٠٥٨	٥٣٢	٠.٠٥٣	٦٠٠	٠.٠٦٠



شكل ١
التمثيل البياني لتغير نسبة الافراد غير المطابقين للاختبار للمؤشرات الثلاثة في جميع حالات الارتباط الأربع للبيانات الحقيقية

(حالة الارتباط الموضعي ٠.٠) كحالة مرجعية. ويوضح شكل ١ التشابه الكبير بين عدد الافراد غير المطابقين في المؤشرات الثلاثة عند اختلاف قوة الارتباط الموضعي، كما نلاحظ التغير في عدد الافراد غير المطابقين للمؤشرات بنفس النمط بين جميع حالات الارتباط الموضعي.

للإجابة عن السؤال البحثي الثاني "ما مستوى اتفاق مؤشرات مطابقة الفرد في الكشف عن أنماط استجابة الافراد غير المطابقة للنموذج عند حالات الارتباط الموضعي؟" حسب متوسط معامل اتفاق كابا بين كل مؤشرين لكل حالة من حالات الارتباط الموضعي للبيانات المقدرة.

يوضح جدول ٥ أن أكبر معامل اتفاق كان بين مؤشر رايت الموزون ومؤشر درازجو حيث بلغ في حالة الاستقلال الموضعي ٠.٩٣، وزاد الاتفاق لمعامل كابا في الحالات الثلاث للارتباط الموضعي. وتشير نتائج جدول ٥ ان معامل الاتفاق بين المؤشرات عند وجود حالات ارتباط في الفقرات اكبر من حالة الاستقلال الموضعي، كما أن جميع متوسطات معامل الاتفاق لكابا لجميع المؤشرات في كل حالات الارتباط الموضعي اكبر من ٠.٨٣.

للإجابة عن السؤال الثالث "هل يؤثر الارتباط الموضعي بين الفقرات على خصائص توزيع مؤشرات مطابقة الفرد ؟"

لمعرفة خصائص التوزيع للمؤشرات الثلاثة عند تغير مقدار الارتباط الموضعي في كل حالة، تم حساب المتوسطات الحسابية والانحرافات المعيارية والالتواء والتفلطح للمؤشرات عند كل حالة. ويوضح جدول ٦ المتوسطات الحسابية والانحرافات المعيارية والالتواء والتفلطح للمؤشرات عند جميع حالات الارتباط الموضعي أظهرت المؤشرات الثلاثة (رايت الموزون، درازجو، المحرزي الموزون) نوعا من الثبات في المتوسط والانحراف المعياري. وكذلك تشابه خصائص شكل توزيع مؤشري درازجو والمحرزي (التفلطح والالتواء).

جدول ٥
متوسط معامل الاتفاق لكابا بين كل مؤشرين في كل حالة من حالات الارتباط الموضعي للبيانات المقدرة

الحالة	رايت الموزون– درازجو	المحرزي الموزون–درازجو	المحرزي الموزون–رايت
٠.٠	٠.٩٣	٠.٨٣	٠.٨٤
٠.٣	٠.٩٤	٠.٨٥	٠.٨٥
٠.٦	٠.٩٥	٠.٨٥	٠.٨٥
٠.٩	٠.٩٥	٠.٨٥	٠.٨٥

جدول ٦

المتوسط الحسابي والانحراف المعياري والالتواء والتفلطح للمؤشرات عند حالات الارتباط الموضعي

المؤشر	الحالة	المتوسط	الانحراف	أعلى قيمة	أقل قيمة	التفلطح	الالتواء
رايت	٠.٠	٠.٢١	٠.٦٦	٣.٦٢	٠	٤.٤٢	٢.١٠
	٠.٣	٠.٣٠	٠.٥٤	٣.٦٦	٠	٤.٧٢	٢.١٤
	٠.٦	٠.٣١	٠.٥٦	٣.٧٦	٠	٤.٤٠	٢.١٠
	٠.٩	٠.٣٤	٠.٦١	٤.٣٥	٠	٤.٨٠	٢.١٥
درازجو	٠.٠	٠.٢٧	٠.٩٨	٢.٨٥	-٤.٥٢	٠.١٨	-٠.٤٤
	٠.٣	٠.٢٦	١.٠٣	٢.٩١	-٤.٧٤	٠.٣١	-٠.٥١
	٠.٦	٠.٢٥	٠.١١	٢.٩٤	-٥.١٢	٠.٤٠	-٠.٥٦
	٠.٩	٠.٢٥	١.١٢	٣.٠٣	-٥.٣٤	٠.٣٩	-٠.٦٠
المحرزي	٠.٠	٠.٢٧	١.٠٠	٣.٠٧	-٤.٤٢	٠.٠٤	-٠.٣١
	٠.٣	٠.٢٦	١.٠٥	٣.١٩	-٤.٥٢	٠.١٣	-٠.٣٨
	٠.٦	٠.٢٦	١.٠٨	٣.١٨	-٤.٩٦	٠.١٨	-٠.٤٢
	٠.٩	٠.٢٧	١.١٤	٣.٢٩	-٥.٠٤	٠.١٨	-٠.٤٦

مناقشة النتائج

تم التحقق من دقة توليد قوة الارتباط الموضعي للفقرات باستخدام مؤشر Q_3 (Yen, 1992) حيث توافقت النتائج مع حالات الارتباط الموضعي المفترضة في البيانات المقدرة، وكذلك تم التحقق من افتراض احادية السمة في البيانات باستخدام مؤشري: التحليل العاملي الاستكشافي وتحليل البواقي، وهذه اشارة واضحة على صحة توليد البيانات.

أظهرت النتائج زيادة نسبة الأفراد ذوي الانماط غير المطابقة بزيادة قوة الارتباط الموضعي عند المؤشرات الثلاثة (رايت الموزون- درازجو- المحرزي الموزون)، حيث تظهر النتائج أن فعالية مؤشرات الكشف عن الأنماط غير المطابقة تتأثر بالارتباط الموضعي بين الفقرات.

وأظهرت الدراسة أن نسبة عدد الافراد غير المطابقين في مؤشر رايت أكبر من نسبة عدد الافراد غير المطابقين في مؤشر درازجو، وهذه النتيجة اختلفت مع دراسة لي وأوليجنيك (Li & Olijnike, 1997) التي أظهرت أن مؤشر درازجو أظهر نسبة أعلى في الكشف عن الاستجابات غير المطابقة من مؤشر رايت الموزون، ويعود ذلك ربما لاختلاف الاساس الرياضي الذي بنيت عليه المؤشرات وشكل توزيع هذه المؤشرات، فيشترك مؤشري درازجو والمحرزي في انهما يتمتعان بخصائص التوزيع الطبيعي مع تحررهما من مستويات القدرة المختلفة (معدل كشف حالات عدم المطابقة لا يعزى لاختلاف مستويات قدرات الأفراد) وهذا من معايير المفاضلة بين المؤشرات.

واشارت النتائج بشكل مباشر إلى أن نسبة الافراد غير المطابقين تزداد بزيادة الارتباط الموضعي بين الفقرات للمؤشرات الثلاثة (رايت الموزون، ودرازجو، والمحرزي الموزون)، وهذا يوافق ما ذكر في دراسة الجراح (٢٠٠٩) أن انتهاك افتراضات نظرية استجابة الفقرة يعد من أسباب حدوث أنماط استجابة غير مطابقة.

وأشارت الإحصاءات الوصفية والالتواء والتفلطح لمؤشرات مطابقة الفرد الى أن قيم مؤشر المحرزي الموزون تتبع التوزيع الطبيعي الاعتدالي في حالة الارتباط الموضعي ٠.٠، وتتفق هذه النتيجة مع ما ورد في دراسة (Almehrizi, 2010) والتي أشار فيها إلى أنه إذا طابقت البيانات النموذج المستخدم فإن نتائج المؤشر تتبع التوزيع الطبيعي بمتوسط حسابي صفر وانحراف معياري واحد، أما في حالات الارتباط

الموضعي ٠.٣ و٠.٦ و٠.٩ فقد ابتعدت النتائج من التوزيع الطبيعي وحافظت على شكل التوزيع. كما أشارت الإحصاءات الوصفية لمؤشر درازجو في حالة الارتباط الموضعي ٠.٠ الى انها تتبع التوزيع الطبيعي المعياري وهذا يتفق مع ما ورد في دراسة (Drasgow, Levine, & William, 1985) التي أشارت الى أن المؤشر الاحصائي يتوزع توزيعا طبيعياً معيارياً. وأظهرت الإحصاءات الوصفية لقيم مؤشر رايت قربها من التوزيع الطبيعي عند حالة الارتباط ٠.٠، وهذه النتيجة تتفق مع القيم النظرية للمؤشر والتي أشار إليها رايت وستون المذكور في (Almehrizi, 2010) والتي أشارت الى أن قيم مؤشر رايت تتوزع توزيعا طبيعيا بمتوسط حسابي صفر وانحراف معياري واحد. كما وبينت النتائج أن أكثر المؤشرات استقراراً في خاصية التفلطح عبر حالات الارتباط الموضعي هو المحرزي، واكثرها استقراراً حسب خاصية الالتواء هو مؤشر رايت ووافقت هذه النتيجة دراسة الجراح (٢٠٠٩).

التوصيات والمقترحات

في ضوء النتائج التي تم التوصل إليها في هذه الدراسات يمكن التوصية بما يلي:

١. التحقق من افتراض الاستقلال الموضعي بعدة طرق بدلا من استخدام طريقة واحدة (Q_3، G^2، Fisher z، Pearsons X2، المقارنة بين تقديرين مختلفين للثبات) ثم المقارنة بين المؤشرات في ضوء اختلاف قوى الارتباط الموضعي.

٢. استخدام نموذج راش او النموذج اللوجستي الثلاثي لمعرفة فعالية المؤشر عند قوة ارتباط مختلفة مع كل نموذج.

٣. استخدام اختبار فقراته جميعها مرتبطة بقوة ارتباط موضعي (٠.٠، ٠.٣، ٠.٦، ٠.٩) ودراسة فعالية مؤشرات مطابقة الفرد على هذه الارتباطات الموضعية.

المراجع
References

أبو شندي، يوسف (٢٠٠٨). **تأثير تعدد الابعاد للاختبار والعلاقة بينها على تقديرات معالم فقراته: دراسة محاكاة** (رسالة دكتوراه غير منشوره). جامعة اليرموك، الأردن.

جراح، بندر (٢٠٠٩). **مقارنة مؤشرات مطابقة الشخص لنماذج استجابة الفقرة باستخدام بيانات فعلية** (رسالة دكتوراه غير منشورة). جامعة اليرموك، الأردن.

علام، صلاح (٢٠٠٥). **نماذج الاستجابة للفقرة الاختبارية أحادية البعد ومتعددة الابعاد وتطبيقاتها في القياس النفسي والتربوي**. القاهرة: دار الفكر العربي.

الشريفين، نضال (٢٠٠٣). بناء مقياس اتجاهات معلمي العلوم نحو العمل المخبري. **المجلة الأردنية في العلوم التربوية. ٢**(٣)، ١٦٩-١٨٧.

Al-mahrazi, R. (2003*). Investigating anew modification of the residual- based person fit index and its relationship with other indices in dichotomous item response theory* (Unpublished Doctoral Dissertation). University of Iowa, USA.

Almehrizi, R. (2010). Comparing among new residual-fit and wright's Indices for dichotomous Three-Parameter IRT model with standardized tests. *Journal of Educational and Psychological Studies. Sultan Qaboos University. 4*(2), 14-26

Drasgow, F., Levine, M. V., & Williams, E. A. (1985). Appropriateness measurement with polytomous item response models and standardized indices. *British Journal of Mathematical and statistical psychology, 38*, 67-86.

Gruijter, D., & Kamp, L. (2005). Statistical test theory for education and psychology. Retrieved December 30,

2017 from: www.leidenuniv.nl.

Hambleton, R. K. & Swaminathan, H. (1985). Item Response theory: principles and applications. Boston, MA: Kluwer- Nijhoff. Publishing.

Hambleton, R. K. and Swaminathan, H. & Rogers, H. J. (1991). Fundamental of item response theory. Newbury Park,CA:Sage.

Hatti J. (1985). Methodology review: Assessing unidimentionality of test and items. *Applied Psychological Measurement, 9,139*-164.

Hayes, H. (2012). A generalized partial credit FACETS model for investigating order effects in self-report personality data. (Unpublished Doctoral Dissertation) –School of Psychology in the Academic Faculty.

Karabatsos, G. (2003). Comparing the aberrant response detection performance of thirty-six Person-Fit Statistics. *Applied Measurement in Education, 16*(4), 277-298.

Lee, Y. (2004). Examining passage- related local item dependence (LID) and measurement construct using Q_3 statistics in an EFL reading comprehension test Language testing. *Applied Psychological Measurement, 21*(1), 74-100.

Levine, M. V. & Rubin D. B. (1979). Measuring the appropriateness of multiple - choice test scores. *Journal of Educational Statistics*, *4*, 269-290.

Li, M. F., & Olijnike, S. (1997). The power of rasch person-fit statistics in detecting unusual person patterns. *Applied Psychological Measurement, 21*, 215-231.

Linn, R & TatsukaK. (1983). Indications for detecting unusual patterns: Links between two general approaches and potential application. *Applied psychological measurement*, *7*, 81-96.

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

Lord, F. M. & Novick, M.R. (1968). Statistical theories of mental test scores. Reading, MA: Addison-Wesley

Meijer, R. (1996). Person-fit Research: An Introduction. *Applied Measurement in Education,* 9, 3-8.

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

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

Miller, T. R. (1991). Empirical estimation of standard errors of compensatory MI model parameters obtained from the NOHARM estimation program. (ACT Research Report No. onr91-2). Iowa City IA: ACT Inc.

Reckase, M. D. (1985). The difficulty of test items that measure more than one ability. *Applied Psychological Measurement, 9,* 401-412

Smith, R. M (1991). The distribution properties of Rasch Item fit statistics. *Education and Psychological Measurement*, 51,541-565.

Snijders, T. (2001). Asymptotic null distribution of person fit statistics with estimated person parameters. *Psychometrika, 66*, 331-345.

Wright, B. D. (1977). Solving measurement problems with the Rasch model. *Journal of Education Measurement, 114*, 96-115.

Wright, B. D., & Masters, G. N. (1982). *Rating scale analysis*. Chicago: MESA Press.

Wright, B. D., & Stone, M. H. (1979). *Best test design*. Chicago. MESA Press.

Yen, W. (1992). Scaling Performance Assessments: Strategies for Managing Local Item Dependence. *Journal of Educational Measurements, 30*(3), 187-213.