فاعليَّة استخدام أسلوب إعادة المعاينة في تقدير درجة القطع لإحصائي المطابقة (l_z^p) للفقرات متعددة التدريج في الكشف عن الاستجابات غير المطابقة في المقاييس غير المعرفية

علي الغزو* محمود القرعان**

تاريخ تسلم البحث 2017/7/9 تاريخ قبوله 2017/10/5

ملخص: هدفت الدّراسة الحالية إلى تقصي فاعليّة أسلوب إعادة المعاينة (Bootstrapping) في تقدير درجة القطع (Cutoff Score) لإحصائي مطابقة الفرد l_z^p كطريقة للكشف عن الاستجابات الشّاذة أو غير المطابقة في المقاييس غير المعرفيّة. ولتحقيق هدف الدّراسة، استخدم مقياس مكوّن من (62) فقرة بصورته النهائية صُمّم لتقييم جودة البرامج الأكاديميّة في جامعة اليرموك من وجهة نظر الطلبة. كما استخدم نموذج الاستجابة المتدرجة (Graded Response Model) كأحد نماذج نظرية الاستجابة للفقرة والمناسب للفقرات من نوع ليكرت الخماسي. تكوّن مجتمع الدّراسة من جميع طلبة البكالوريوس لمستوى السّنتين الدّراسيتين الثّالثة والرّابعة في جامعة اليرموك في الكليّات الإنسانيّة والعلميّة البالغ عددهم (15152) طالبًا وطالبةً. في حين تكوّنت عينة الدّراسة من (1000) طالبٍ وطالبة، تمّ اختيارهم بشكل عشوائي ثمّ عنقوديٍّ، كَوْن أنّ الشّعبة الدراسيّة هي وحدة الاختيار. أظهرت النتائج أن l_z^p ملتو نحو اليسار وبتفلطح موجب ولا يتوزع توزيعا طبيعيًا وأنّ استخدام درجة القطع المشتقّة من خلال أسلوب إعادة المعاينة هي الأنسب في تصنيف الاستجابات إلى مطابقة وغير مطابقة وذلك لكشفها عن عدد اكبر من الاستجابات غير المطابقة إذا ما قورنت بالقيمة الحرجة النظريّة لتوزيع (Z).

(الكلمات المفتاحية: إعادة المعاينة، إحصائي مطابقة الفرد l_z^p، نموذج الاستجابة المتدرجة، الاستجابات غير المطابقة).

مقدمة: تُعدُّ أساليب المسح القائمة على التقدير الذاتي (Self-Reported Surveys) من أشهر أدوات جمع البيانات وأكثرها استخدامًا في البحوث التربويّة وبحوث العلوم الإنسانيّة والاجتماعيّة. وتعتمد دقة النتائج التي يتم التوصل إليها بناءً على تلك الأدوات على جديّة المُستجيب في الإجابة عنها؛ لذا فإنّ الاستجابات غير العادية أو غير الشائعة (Unusual or Unexpected Responses) أو السلوكات الشّاذة (Aberrant Responses) في الإجابة عن فقرات المقاييس غير المعرفيّة كنقص الدافعيّة (Lack of Motivation)، والنمطيّة في الاستجابة (Stylistic Responding)،والمرغوبيّة الاجتماعيّة(SocialDesirability)، وسوء الفهم للفقرة (Misunderstanding of the item)، والتحيُّز في الإجابة (Response Bias) قد لاقت اهتمامًا كبيرًا من الباحثين مؤخرًا. وتُعرَّف الاستجابات الشّاذة أو غير المطابقة بأنّها نمط من الاستجابة عن فقرات أداة القياس التي لا تعكس القدرة الحقيقيّة (السمة المنوي قياسها) للمُستجيب (Barnes, 2016)؛ لذا فإنّ تقدير درجة الفرد على السمة بوجود استجابات غير مطابقة يؤدّي إلى تحديد خاطئ لموقعه على متصل السّمة (Tendeiro, 2017) .

The Effectiveness of Bootstrapping Procedure in Estimating l_z^p Person-Fit Statistic Cutoff- Score for Polytomous Items in Noncognitive Scales

Ali Algazo, Ministry of Education, Jordan.
Mahmoud Al-Quraan, *Collage of Education, Yarmouk University, Jordan.*

Abstract: This study aimed to assess the effectiveness of bootstrapping procedure in estimating l_z^p person-fit statistic cutoff score (denoted as l_z^p) to detect aberrant responses in noncognitive scales. To achieve the aim of this study, a sample of 1000 students was selected from the third and fourth year undergraduate students at Yarmouk University (N= 15152) using multi-stage sample procedures. A questionnaire consisted of 62 Likert Scale items designed to assess the quality of academic programs from students perspectives was used to collect the data. The datasets were analyzed utilizing graded response model of IRT. The results showed that l_z^p person-fit statistic distributions are not normally distributed and still have negative skewness and positive kurtosis. Moreover, the results showed that using cutoff scores calculated by bootstrapping procedures (l_z^p) to categorize responses to fit or misfit is preferred as l_z^p detected more aberrant responses when it is compared with Z critical value at the same level of statistical significance.

(**Keywords** Bootstrapping, Person-fit Statistic l_z^p, Graded Response Model (GRM), Aberrant Responses, Noncognitive Scale).

وقد استحوذ هذا الموضوع على اهتمام العديد من الباحثين في الآونة الأخيرة (Huang, Curran, Keeney, Poposki, & DeShon, 2012). وعليه فقد تمّ تطوير عدد من الطرق والإجراءات التي يمكن للباحثين استخدامها لغرض الكشف عن ذلك النوع من الاستجابات؛ ليُساعد الباحثين على فهم سلوكات الأفراد أثناء الاستجابة على فقرات المقاييس النفسيّة والتربويّة (Glas & Meijer,2003).

* وزارة التربية والتعليم، الأردن.
** قسم علم النفس الإرشادي والتربوي، كلية التربية، جامعة اليرموك.
© حقوق الطبع محفوظة لجامعة اليرموك، إربد، الأردن.

وتُسمَّى الطرق المُستخدمة في الكشف عن الاستجابات الشّاذة بإحصائيّات مطابقة الفرد (Person-fit Statistics). وتُستخدم هذه الإحصائيّات لتقييم سلوكات الأفراد غير المرغوبة في البيانات والتي تؤثر على نتائج عملية القياس (Jelínek, Květoň & Vobořil , 2013; Ferrando, 2012). وتُصنَّف إحصائيّات مطابقة الفرد في مجموعتين: المجموعة الأولى تمثّل المؤشرات التي تعتمد على المجموعة (Group-based indices) وتقوم بمقارنة الاستجابة الملاحظة للفرد مع الاستجابة المتوقعة عن الفقرة بناء على بيانات العينة (Karabtsos, 2003)، أما المجموعة الثانية فتمثل المؤشرات التي تعتمد على نظرية الاستجابة للفقرة (IRT-based indices) حيث تقوم على مقارنة نمط استجابة الفرد عن الفقرات مع نمط الاستجابة المتوقعة بناءً على أحد نماذج نظرية الاستجابة للفقرة (Meijer & Sijtsma, 2001).

ويُعدُّ الإحصائي(l_z) الذي طوّره دراسقو وزملاؤه (Drasgow, Levine.,&Williams, 1985) من أشهر الإحصائيّات المُستخدمة في الكشف عن الاستجابات الشّاذة أو غير المطابقة للأفراد التي تعتمد على نماذج نظرية الاستجابة للفقرة (Sinharay, 2016a). كما يُمكن استخدام(l_z) في حال الفقرات ثنائيّة التدريج (Dichotomous Data)أو الفقرات متعددة التدريج (Polytomous Data) (Tendeiro, 2017).

وقام ليفين وروبن (Levine & Rubin, 1979) باقتراح الإحصائي اقتران لوغارتيم الأرجحيّة العظمى(l_0) (log-likelihood function) للفقرات ثنائيّة التدريج الذي يُعطى بالمعادلة الآتية:

$$l_0 = \sum_{j=1}^{J} X_j \ln(\hat{\theta}) + (1 - X_j) \ln\left[1 - P_j(\hat{\theta})\right]$$

حيث: (X_j) تُمثل استجابة الفرد عن الفقرة j، (J) عدد فقرات الاختبار، ($P_j(\hat{\theta})$) احتماليّة استجابة الفرد ذي القدرة (θ) عن الفقرة (j) إجابة صحيحة. ولكون التوزيع الإحصائي لـ (l_0) غير معروف، فقد قام دراسقو وزملاؤه (Drasgow et al,1985) بتحويل (l_0) إلى صورة معياريّة مشابهة للعلامة المعياريّة الزائيّة (z-score) باستخدام المعادلات الآتية للفقرات ثنائيّة التدريج:

$$l_z = \frac{l - E(l_0)}{Var(l_0)^{\frac{1}{2}}}$$

حيث $E(l_0)$ و$Var(l_0)$ تأخذ الشكل الآتي:

$$E(l_0) = \sum_{j=1}^{j} \{P_j(\hat{\theta}) Ln P_j(\hat{\theta}) + [1 - P_j(\hat{\theta})] Ln[1 - P_j(\hat{\theta})]\}$$

$$Var(l_0) = \sum_{j=1}^{j} P_j(\hat{\theta})[1 - P_j(\hat{\theta})]\{\left\{Ln\frac{P_j(\hat{\theta})}{1 - P_j(\hat{\theta})}\right\}^2$$

أما في الفقرات متعددة التدريج، فإنّ (l_0^p) يبدأ بتعريف ($V = (V_1, V_2, \dots, V_j)$) كمتجه عشوائي (Random Vector) لبدائل الاستجابة (ككبيرة جدًا إلى قليلة جدًا في مقاييس ليكرت) على مجموعة من الفقرات عددها (J)، ويكون متجه الاستجابة على بدائل الفقرات الملاحظ هو ($v = (v_1, v_2, \dots, v_j)$) فتصبح المعادلات السابقة على النحو الآتي:

$$l_0^p = \sum_{j=1}^{J} \sum_{m=1}^{A} \delta_m(v_j) ln P_{jm}(\hat{\theta})$$

حيث (m) تُمثل بدائل الاستجابة عن الفقرة ($m = 1,2,\dots,A$)، و (P_{jm}) نسبة المستجبين عن البديل (m) للفقرة (j)، و ($\delta_m(v_j)$) تساوي (1) إذا كانت الاستجابة الملاحظة تساوي (m) و (0) إذا كانت غير ذلك، وعليه تكون الصورة المعياريّة لـ (l_0^p) للفقرات متعددة التدريج على النحو الآتي:

$$l_z^p = \frac{l_0^p - E(l_0^p)}{Var(l_0^p)^{\frac{1}{2}}}$$

وتكون ($E(l_0^p)$) و ($Var(l_0^p)$) تساوي:

$$E(l_0^p) = \sum_{j=1}^{J} \sum_{m=1}^{A} P_{jm}(\hat{\theta}) \ln P_{jm}(\hat{\theta})$$

$$Var(l_0^p) = \sum_{j=1}^{J} \left[\sum_{m=1}^{A} \sum_{k=1}^{A} P_{jm}(\hat{\theta}) P_{jk}(\hat{\theta}) \ln P_{jm}(\hat{\theta}) ln\left(\frac{P_{jm}(\hat{\theta})}{P_{jk}(\hat{\theta})}\right)\right]$$

وكون (l_z^p) هو صورة معياريّة لاحتماليّة نمط الاستجابة بناء على قيم محددة من مستوى القدرة ومعالم الفقرات، فإنّ زيادة قيمة (l_z^p) تدلُّ على مطابقة أفضل (Drasgow et al, 1985). وعليه، يتم تحديد الاستجابات غير المطابقة في الطرف الأيسر من التوزيع. وقد بيّنت عدد من الدراسات أنّ (l_z^p) لا يتوزع طبيعيًا عند استخدام القدرة المقدّرة ($\hat{\theta}$) بدلاً من القدرة الحقيقيّة (θ) كما في المعادلات السابقة (Nering, 1995; Schmitt, Chan, Sacco, McFarland, & Jennings, 1999; Meijer & Sijtsma, 2001). لذا فإنّ استخدام درجة القطع (-1.6449) عند مستوى الدلالة (α=0.05) لتحديد أنماط الاستجابة غير المطابقة تُعدُّ غير مناسبة. وبناءً على ما تقدّم، فقد قام سنجرز (Snijders, 2001) باشتقاق توزيع لـ (l_z) ليقترب من التوزيع الطبيعي وأسماه (l_z^*)، إلاّ أنّه خاص فقط بالفقرات ذات التدريج الثنائي، ولا يتناسب مع الفقرات ذات التدريج المتعدد (Sinharay, 2016b).

وتتّصف العديد من إحصائيّات الكشف عن الاستجابات غير المطابقة ومنها الإحصائي (l_z) بعدم وجود درجة قطع (Cutoff Score) لتصنيف الاستجابات إلى مطابقة أو غير مطابقة مشتقّة بصورة مناسبة (Barens, 2016). وقد وجد روب (Rupp, 2013) في مسح للدّراسات السابقة بأنّ اشتقاق درجة القطع بطريقة تجريبيّة وحسب التوزيع لـ (l_z) للبيانات تحت التحليل بالاعتماد على الرتبة المئينية (5) التي تقابل مستوى الدلالة (α=0.05) هي أشهر الطرق المُستخدمة في تقدير درجة القطع لإحصائيّات الكشف عن الاستجابات غير المطابقة. كما تمّ تقديم طريقة إعادة المعاينة (Bootstrapping or Resampling) لتقدير درجة القطع للإحصائي

(l_z) في الفقرات الثنائيّة والمتعددة التدريج (Tendeiro, Meijer, & Niessen, 2016; Tendeiro, 2016)، حيث تعتمد هذه الطريقة على سحب عينات كثيرة من العينة الأصليّة مع الإرجاع قد تصل إلى (1000) عينة. ويتم في كل عينة تقدير الإحصائي محور الاهتمام، وبعد ذلك أخذ الوسيط للإحصائي المُقدَّر من كافة العينات المسحوبة. وبالتالي يُمكن الحصول على تقدير غير متحيِّز للإحصائي محور الاهتمام (Efron, 1992). ويتم اتِّباع هذه الطريقة في تقدير الإحصائي (l_z) من خلال تقدير معالم الفقرات والقدرة للبيانات تحت الدِّراسة، ومن ثمَّ سحب عينات عشوائية بعدد (1000) مرة لتقدير الإحصائي (l_z)، ومن ثمَّ تحديد درجة القطع في كل مرَّة لتكون الرتبة المئينيّة (5) التي تمثل قيمة (l_z^p) التي يقل عنها (5%) من القيم. وتُعدُّ درجة القطع هي الوسط لتلك القيم (R Core Team, 2016).

وقد أظهرت الدراسات السابقة وجود بعض العوامل المؤثرة في فعاليّة الكشف عن الاستجابات غير المطابقة (الشّاذة) منها: عدد فقرات أداة القياس (Karabatsos, 2003; Reise & Due, 1991)، وعدد الاستجابات غير المطابقة (Meijer & Stijsma, 2001)، وطبيعة الإحصائي المُستخدم (Karabatsos, 2003)، ومستوى الدلالة (Meijer & Nering, 1997; Davis, 2009). وركَّز العديد من الباحثين في دراستهم للإحصائي (l_z) على شكل التوزيع، وإجراء التحويلات لجعله يتوزَّع توزيعًا طبيعيًّا (Standardization). وقد بيَّنت الدراسات الأولى (Drasgow et al., 1985 و Molenaar & Hoijtink, 1990) أنَّ الإحصائي (l_z) يقترب من التوزيع الطبيعي، ولكنه يبتعد عنه في حال استخدام القدرة المقدَّرة بدلاً من القدرة الحقيقيّة (Molenaar & Hoijtink, 1990, 1996; van Krimpen-Stroop & Meijer, 1999; Sinharay, 2016a,2016b; Tendeiro, 2017). وعليه، فإنَّ الحاجة إلى البحث عن درجة قطع للإحصائي (l_z) لا تفترض التوزيع الطبيعي له تصبح ذات قيمة عالية للباحثين في هذا المجال.

مشكلة الدِّراسة

توصلت العديد من الدراسات إلى أنَّ الإحصائي (l_z^p) لا يتوزع توزيعًا طبيعيًّا. وقد تمَّ مؤخرًا اقتراح طريقة إعادة المعاينة (Bootstrapping or Resampling)، لغرض تقريبه من التوزيع الطبيعي (Tendeiro, Meijer, & Niessen, 2016; Tendeiro, 2016)، غير أنَّه لم تتم مقارنة هذه الطريقة في تقدير درجة القطع مع طريقة القيمة الحرجة لتوزيع (Z)؛ الأمر الذي قد يُساعد في استخدام درجة القطع المُشتقة وفق هذا الأسلوب في تصنيف استجابات الأفراد إلى مطابقة (Fitting) وغير مطابقة (Misfitting) بدلاً من استخدام القيمة الحرجة النظرية لتوزيع (Z) البالغة (1.6449-)، وذلك بعد تقدير (l_z^p) لكل مُستجيب. ويُشار إلى أنَّه تمَّت إضافة هذه الإمكانيّة إلى بيئة R (R Core Team, 2016). وعليه، فإنَّ الدِّراسة الحالية جاءت لتقصي فاعليّة أسلوب إعادة المعاينة في تقدير درجة القطع للإحصائي (l_z^p) حيث بيَّنت غالبيّة الدراسات أنَّ الإحصائي (l_z^p) لا يتوزَّع طبيعيًّا، وغالبًا ما يكون ذا التواء سالب. كما تأتي هذه الدراسة استكمالاً لجهود الباحثين في هذا المجال، غير أنَّها تتمايز عنها في طبيعة المؤشر والنموذج الإحصائي اللذين تمَّ تبنيهما فيها، إضافة إلى تناولها بيانات واقعية. فضلاً عن تناولها لمجموعة من المتغيرات التي اتصفت بندرتها في مثل هذا النوع من البحوث، إذ إنَّ غالبية البحوث في هذا المجال كانت تركز على دراسة متغيرات من مثل حجم العينة، ومستوى الدلالة، وطول المقياس، ونوع النموذج الرياضي. وبناءً على ذلك، فقد حاولت الدِّراسة الإجابة عن السؤالين الآتيين:

1- ما خصائص التوزيعات العينيّة والأخطاء المعيارية للمؤشر الإحصائي l_z^p لأنماط الاستجابة لمقياس التقييم تبعًا لعينة الدِّراسة كاملة، ومستويي متغيرات (الجنس، الكليّة، المستوى الدراسي، التقدير الجامعي)؟

2- هل تختلف درجة القطع (Cutoff Scores) للإحصائي (l_z^p) المُشتقة بأسلوب إعادة المعاينة لألف مرة عن القيمة الحرجة النظرية لتوزيع (Z)؟

أهمية الدِّراسة

تبرز أهمية هذه الدِّراسة من **الناحية النظريّة** في أنَّ تحديد درجة القطع وفق الأسلوب الإحصائي المُستخدم فيها، يُعطي المؤشر الإحصائي لمطابقة الفرد دقَّة أكبر في الكشف عن أنماط الاستجابة غير المطابقة، إذ إن تصنيف الأفراد إلى مطابقين وهم في حقيقة الأمر غير مطابقين، ينطوي على آثار سلبيّة على دقة تقدير معالم الفقرات وقدرات الأفراد، خاصة وأنَّ هذا الأسلوب يتَّصف بندرة تناوله في الدراسات والبحوث التي أُجريت في هذا الحقل، **وتطبيقيًّا** تبرز أهمية هذه الدِّراسة في أنَّها تُقدِّم للباحثين والمهتمين في مثل هذا النوع من البحوث دليلاً يُمكنهم الاعتماد عليه في تحديد درجة القطع المناسبة بدلاً من الاعتماد على القيمة المرجعية وحدها، كون الدراسات السابقة قد بيَّنت أنَّ الإحصائي (l_z^p) لا يتوزَّع توزيعًا طبيعيًّا، وذلك عند دراستهم للإحصائيّات المتعلقة بمطابقة الفرد، كما أنَّها ستعود بفائدة على أصحاب القرار وكذلك المفحوصين؛ لما تُعطيه من نتائج دقيقة، خاصة فيما يتعلق بالمقاييس غير المعرفية .

مصطلحات الدراسة

إعادة المعاينة: أسلوب إحصائي يُستخدم لتحديد درجة القطع لإحصائي مطابقة الفرد في الفقرات الثنائية ومتعددة التدريج، حيث تعتمد هذه الطريقة على سحب عينات كثيرة من العينة الأصلية مع الإرجاع قد تصل إلى ألف عينة، ويتم في كل عينة تقدير الإحصائي. وبعد ذلك أخذ الوسيط المقدر من كافة العينات المسحوبة مما يُعطي توزيعا عينيًّا مُشابها للتوزيع في المجتمع.

الإحصائي R^2: أحد المؤشرات الإحصائية المبنية على طريقة تقدير الأرجحية العظمى، حيث يقارن بين نمط الاستجابة المتوقع من نموذج نظرية الاستجابة للفقرة ونمط استجابة الفرد على مقياس متعدد التدريج.

نموذج الاستجابة المتدرجة: أحد نماذج نظرية الاستجابة للفقرة، حيث يعتمد هذا النموذج على فقرات متعددة الاستجابة. ويعد هذا النموذج امتدادا للنموذج ثنائي المعلمة.

الاستجابات غير المطابقة: نمط من الاستجابة عن فقرات أداة القياس التي لا تعكس النمط المتوقع وفق نموذج نظرية الاستجابة للفقرة أو القدرة الحقيقيّة للمستجيب.

الطريقة

مجتمع الدِّراسة

تكوّن مجتمع الدِّراسة من جميع طلبة البكالوريوس للسنتين الدراسيتين الثالثة والرابعة في جامعة اليرموك في الكليات الإنسانيّة والعلميّة البالغ عددهم (15152) طالبًا وطالبةً، منهم (11533) طالبًا وطالبةً في الكليات الإنسانيّة و(3619) طالبًا وطالبةً في الكليات العلميّة. بالإضافة إلى أنّهم موزعون إلى (8277) طالبًا وطالبة في مستوى السّنة الدِّراسيّة الثالثة و(6875) طالبًا وطالبةً في مستوى السّنة الدِّراسيّة الرابعة. كما أنّهم موزعون وفقًا لجنس الطلبة إلى (5248) طالبًا و(9904) طالبةً خلال الفصل الدراسي الثاني من العام 2016/2015، بحسب إحصائيّات دائرة القبول والتسجيل في الجامعة.

عينة الدِّراسة

نظرًا لكون الأداة المُستخدمة في الدِّراسة تتعلّق بتقييم جودة البرامج الأكاديميّة في جامعة اليرموك من وجهة نظر الطلبة، فقد تمّ اختيار عينة منهم قضت وقتًا كافيًا في البرنامج الدِّراسي بحيث أصبح لديهم الخبرة والمعرفة لتقييمه. كما أصبح لديهم ألفة بأعضاء الهيئة التدريسية، والمساقات التي تُطرح في البرامج المختلفة، ونوع الخدمات التي تُقدِّمها الجامعة لهم؛ لذلك تمّ اختيار الطلبة من مستوى السنتين الدراسيتين الثالثة والرابعة في كلٍّ من الكليات الإنسانيّة والعلميّة. وقد تمّ تحديد الشعب والمساقات من مستوى السنة الثالثة والرابعة في كل من الكليّات العلميّة والإنسانيّة؛ أي أنّه تمّ تقسيم مجتمع الدِّراسة إلى مجموعتين وفقا للكليّة (إنسانيّة وعلميّة)، وبعد ذلك تمّ اختيار مجموعة من الشعب الدِّراسيّة عشوائيا في كلٍّ من الكليّات الإنسانيّة والعلميّة، وفي كل شعبة وقع عليها الاختيار، تمّ توزيع أداة الدراسة على كافة الطلبة المسجلين فيها (العينة العنقوديّة). ولأنّ خطأ المعاينة في العينات العنقودية يكون كبيرا، فقد تم معالجة هذا الخطأ من خلال زيادة حجم عينة الدِّراسة، بحيث تكون نسبة الطلبة والطالبات ممثلة لمجتمع الدِّراسة. وعليه، فقد بلغ حجم عينة الدِّراسة (1000) طالبٍ وطالبة، كما هو مبين في الجدول (1).

الجدول 1: توزيع أفراد عينة الدِّراسة وفقًا لمتغيرات (الجنس، الكليّة، المستوى الدراسي، التقدير الجامعي)

المتغير	مستويات المتغير	العدد	النسبة المئوية
جنس الطالب	ذكور	367	36.7
	إناث	633	633.3
الكليّة	إنسانيّة	557	55.7
	علمية	443	44.3
المستوى الدراسي	ثالثة	562	56.2
	رابعة	438	43.8
التقدير الجامعي	أقل من جيد	335	33.5
	جيد فأكثر	665	66.5

أداة الدِّراسة

لتحقيق أهداف الدِّراسة، فقد طوّر الباحثان مقياس تقييم جودة البرامج الأكاديميّة من وجهة نظر طلبة جامعة اليرموك بالرجوع إلى الأدب النظري ذي الصلة بموضوع الدِّراسة، وبناء على معايير الاعتماد في هيئة اعتماد مؤسسات التعليم العالي وضمان جودتها. تكوّن المقياس في المرحلة الأولى من (97) فقرة موزّعة على ثلاثة مجالات هي: بُعد الهيئة التدريسية، وتكوّن من (45) فقرة، وبُعد المنهاج والخطط الدراسيّة، وتكوّن من (25) فقرة، وبعد الخدمات الأكاديميّة المُساندة، وتكوّن من (27) فقرة. ثمّ تمّت مراجعة فقرات المقياس مرّة أخرى، حيث حُذفت (25) فقرة؛ لعدم مناسبتها للمجال. كما تمّ تعديل الصياغة اللغوية لاثنتي عشرة فقرة، وفصل الفقرة (83) من المجال الثالث إلى فقرتين. وعليه، فقد أصبح المقياس يتكون في صورته الأوليّة من 73 فقرة تتوزّع على ثلاثة مجالات هي: مجال الهيئة التدريسية وله (37) فقرة، ومجال المنهاج والخطط الدراسية وله (17) فقرة، ومجال الخدمات الأكاديمية المساندة وله (19) فقرة.

وللتحقق من الصدق الظاهري لمقياس التقييم، فقد عُرض بصورته الأولية على مجموعة مؤلفة من عشرة محكمين من أعضاء هيئة التدريس في مجالات القياس والتقويم، وعلم النفس التربوي، والتربية الخاصة، وأساليب تدريس الرياضيات، وأساليب تدريس اللغة الإنجليزية، وأساليب تدريس اللغة العربية، والإدارة التربوية في جامعة اليرموك، والجامعة الهاشمية، وجامعة جدارا، وقد تم الأخذ بملاحظات هؤلاء المحكمين بخصوص كل فقرة من فقرات الأداة، حيث حُذفت بعض الفقرات، وعُدِّلت الصياغة اللغوية للبعض الآخر. وعليه، فقد أصبحت الأداة تتكون من (72) فقرة.

وبهدف تقنين مقياس التقييم، فقد أُجري التحليل العاملي ذي التدوير المائل بطريقة (Promax) علمًا بأنّ فقرات المقياس تتوزّع على ثلاثة مجالات هي: مجال الهيئة التدريسيّة، ومجال المنهاج والخطط الدِّراسيّة، ومجال الخدمات الأكاديميّة المُساندة، وذلك

لعشرين عينة بحجم (500) طالبٍ وطالبةٍ من أصل (1000) طالبٍ وطالبةٍ تمَّ اختيارها من ملف بيانات التطبيق النهائي بطريقة عشوائيَّة مع الإرجاع باستخدام برنامج SPSS v23.0 وفق الشروط الآتية: عدم جواز اعتماد الفقرة ذات التَّشبُّع الأدنى من (0.40) ، وعدم جواز اعتماد العامل المتكوِّن من فقرتين فأدنى (Trivial Factor)، وعدم جواز تشبُّع الفقرة على عاملين بقيمتين تزيدان على (0.40) (Gorsuch,1983). وقد اتَّفقت نتائج التحليل العاملي لست عشرة عينة من أصل عشرين عينة على حذف عشر فقرات، لذا فقد حُذفت تلك الفقرات من الأداة كون التحليل العاملي يعتمد على التباين المشترك (Shared Variance)، فالأصل أن يقود إلى نتائج متشابهة (Gorsuch, 1997). والجدول (2) يبين أرقام الفقرات المُقترح حذفها من الأداة.

الجدول 2: نتائج التحليل العاملي لفقرات مقياس التقييم وفقا لخطوات التقنين

التحليل العاملي	عدد العوامل المفروزة	أرقام الفقرات المحذوفة	سبب الحذف
الأوّل	13		
الثاني	12	70	بعد زائف
الثالث	11	36، 37	بعد زائف
الرابع	11	31	متشبعة على بعدين بتشبع أكبر من 0.40
الخامس	11	26	متشبعة على بعدين بتشبع أكبر من 0.40
السادس	11	46	متشبعة على بعدين بتشبع أقل من 0.40
السابع	11	15	تشبعها أقل 0.40
الثامن	10	39	تشبعها أقل 0.40
التاسع	10	38	تشبعها أقل 0.40
العاشر	10	63	تشبعها أقل 0.40

وعليه، فقد أصبح المقياس يتكون في صورته النهائية من (62) فقرة. وللتحقق من الخصائص السيكومترية لفقرات مقياس التقييم، حُسبت معاملات ارتباط الفقرة المصحَّح بالمقياس وبالمجالات التي تتبع لها، وتراوحت قيم معاملات الارتباط المُصحَّح لفقرات مجال الهيئة التدريسيَّة بمجالها بين (0.35 - 0.65)، وبالمقياس فقد تراوحت بين (0.28 - 0.62)، في حين تراوحت قيم معاملات الارتباط المُصحَّح لفقرات مجال المنهاج والخطط الدِّراسيَّة بمجالها بين (0.54 - 0.67)، وبالمقياس فقد تراوحت بين (0.50 - 0.62)، وأخيرًا تراوحت قيم معاملات الارتباط المُصحَّح لفقرات مجال الخدمات الأكاديميَّة المُساندة بين (0.51 - 0.72)، وبالمقياس فقد تراوحت بين (0.45 - 0.62). وجميع معاملات الارتباط التي تم ذكرها تعدُّ مقبولة، وتدلُّ على ارتباط فقرات المجال بمقياس التقييم ارتباطًا جيِّدًا. كما تمَّ حساب معامل ارتباط الفقرة مع الأداة وتراوحت بين (0.90-0.81) وهي معاملات عالية تدل على قوة الارتباط بين الفقرات وأداة التقييم.

وللتحقق من ثبات مقياس التقييم، تمَّ تقدير ثبات الاتساق الداخلي لفقرات المقياس ومجالاته وفق معادلة كرونباخ ألفا، وقد بلغت قيمة ثبات الاتساق الداخلي للمقياس (0.962)، في حين بلغ (0.93) لمجال الهيئة التدريسية، و(0.92) لمجال المناهج والخطط الدراسية، و(0.93) لمجال الخدمات الأكاديمية المساندة.

النموذج المُستخدم في الدِّراسة والتحقق من افتراضاته

استخدم في الدِّراسة نموذج الاستجابة المتدرجة (Graded Response Model) كأحد نماذج نظرية الاستجابة للفقرة والمناسب للفقرات من نوع ليكرت الخماسي، وتمَّ التحقق من الافتراضات على النحو الآتي: للتحقق من أحادية البعد لفقرات مقياس التقييم، تمَّ إجراء التحليل العاملي باستخدام الرزمة الإحصائية للعلوم الاجتماعية (SPSS)، كما هو مبين في الجدول (3).

الجدول 3: نتائج التحليل العاملي لفقرات مقياس التقييم

المكون	الجذور الكامنة الاستهلالية			استخلاص مجاميع مربعات التشبعات		
	الكلي	التباين المفسر%	التباين المفسر التراكمي%	الكلي	التباين المفسر%	التباين المفسر التراكمي%
1	18.65	30.09	30.09	18.65	30.09	30.09
2	4.43	7.15	37.24	4.43	7.15	37.24
3	2.61	4.21	41.44	2.61	4.21	41.44
4	2.06	3.33	44.77	2.06	3.33	44.77
5	1.63	2.63	47.40	1.63	2.63	47.40
6	1.49	2.41	49.81	1.49	2.41	49.81
7	1.29	2.08	51.89	1.29	2.08	51.89
8	1.18	1.91	53.80	1.18	1.91	53.80
9	1.08	1.74	55.54	1.08	1.74	55.54
10	1.04	1.69	57.22	1.04	1.69	57.22
11	0.97	1.56	58.79			

يلاحظ من الجدول 3 أنّ التحليل العاملي لفقرات مقياس التقييم قد حقّق أحادية البعد بأربعة مؤشرات على النحو الآتي: ناتج قسمة الجذر الكامن للعامل الأوّل على الجذر الكامن للعامل الثاني أكبر من 2، وناتج قسمة حاصل طرح الجذر الكامن الثاني من الجذر الكامن الأوّل على حاصل طرح الجذر الكامن الثالث من الجذر الكامن الثاني ذي قيمة عالية، وقيمة التباين المفسّر التراكمي تزيد على 54.0%، وأخيرًا التباين المفسّر للعامل الأوّل أكبر من 20.0% (Hattie, 1985).

وللتحقق من افتراض الاستقلال الموضعي لفقرات مقياس التقييم، تمّ حساب قيمة χ^2 للاستقلال الموضعي المعياريّة (Standardized LD χ^2) لكل زوج من أزواج فقرات المقياس التي يبلغ عددها 210 وذلك عن طريق ضرب (62) فقرة بـ (61) ثمّ القسمة على 2 باستخدام برنامج IRTPro v3.1.21505.4001، ثمّ رصد التكرارات والنسب المئوية لكلا حالتي الاستقلال الموضعي شريطة أن تزيد قيمة الاستقلال الموضعي المعياري على 10، مما يشير إلى تبعية موضعية، والعكس صحيح إذا كانت قيمته 10 فأقل فهي تشير إلى استقلال موضعي، كما هو مبين في الجدول (4).

الجدول 4: التكرارات والنسب المئوية للاستقلال الموضعي لفقرات مقياس التقييم

حالة الموضعية	التكرار	النسبة المئوية
معتمدة	124	6.56
مستقلة	1767	93.44
الكلي	1891	100.00

يلاحظ من الجدول 4 أنّ الاستقلال الموضعي متحقق في 1767 زوج من أصل 1891 زوج لفقرات المقياس مُشكّلًا ما نسبته 93.44%.

الإجراءات

لتحقيق أهداف الدِّراسة والإجابة عن أسئلتها، طُوِّر مقياس تقييم جودة البرامج الأكاديمية من وجهة نظر طلبة مستوى السنتين الدراسيتين الثالثة والرابعة، بعد الحصول على كتاب تسهيل المهمة من رئاسة الجامعة. بعد ذلك تمّ التنسيق مع مدرسي المساقات التي وقع عليها الاختيار، لتحديد موعد تطبيق أداة الدراسة. كما تمّ الاتفاق مع مدرسي المساقات على مغادرة القاعة قبل البدء بعملية التطبيق، حتى لا تتأثر استجابات الطلبة على المقياس بشخص المدرس، خاصة وأنّ أحد المجالات المراد تقييمها هو عضو هيئة التدريس. وبعد ذلك تمّ شرح هدف الدراسة للطلبة، كما تمّ الطلب منهم الإجابة عن فقرات مقياس التقييم كما يرونها معبرة عن وجهة نظرهم بكل صدق وموضوعية، مع التركيز على أنّ الأداة تحتوي على فقرات مرتبطة ببرنامجهم الأكاديمي ولا تقتصر على المساق الحالي أو مُدَرِّسِهِ، وإنّما يشمل كافة المساقات والمدرسين في البرنامج. وبعد ذلك تمّ جمع البيانات من أجل معالجتها إحصائيا باستخدام البرمجيات المناسبة.

ومن أجل التحقق من افتراضات نموذج الاستجابة المتدرجة (GRM)، استخدمت برمجية (R-Perfit) لتقدير معالم القدرة والفقرات وفقا لنموذج الاستجابة المتدرجة، ثم استخدم خيار (Bootstrapping) لتحديد درجة القطع للإحصائي (l_z^p) من خلال إعادة المعاينة (Resampling) لألف مرة، حيث اتُبعت هذه الإجراءات للعينة كاملة، وكذلك لكل مستوى من مستويات المتغيرات (الجنس، والكليّة، والمستوى الدِّراسي، والتقدير الجامعي). كما حُسبت درجة القطع للإحصائي (l_z^p) بأسلوب إعادة المعاينة لألف مرة اعتمادًا على الرتبة المئينيّة (5) التي تقابل (α=.05) باعتبار أنّ الفرضية البديلة متجهة نحو اليسار.

النتائج ومناقشتها

للإجابة عن سؤال الدِّراسة الأوّل، حُسبت الإحصاءات الوصفية للتوزيعات العينية للمؤشر الإحصائي l_z^p لأنماط الاستجابة وفق نموذج الاستجابة المتدرجة لمقياس تقييم جودة البرامج الأكاديميّة في جامعة اليرموك من وجهة نظر الطلبة، والتحقق من مطابقته للتوزيع الطبيعي باستخدام اختبار (Kolmogorov-Smirnov) تبعًا لعينة الدِّراسة الكليّة ومستويي كلٍّ من متغيرات (الجنس، الكليّة، المستوى الدِّراسي، التقدير الجامعي)، كما هو مبين في الجدول (5).

الجدول 5: الإحصاءات الوصفية للتوزيعات العينية للمؤشر الإحصائي l_z^p لأنماط الاستجابة لمقياس التقييم، ونتائج اختبار (Kolmogorov-Smirnov) تبعًا لعينة الدِّراسة الكليّة ومستويي (الجنس، الكليّة، المستوى الدِّراسي، التقدير الجامعي)

المتغير		الوسط الحسابي	الإنحراف المعياري	الخطأ المعياري للوسط	فترة الثقة لوسط المؤشر		الإلتواء	التفلطح	قيمة المؤشر:		K-S
					%2.5	%97.5			الصغرى	العظمى	
العينة الكليّة		0.107	2.490	0.08	-0.047	0.261	-0.82	1.03	-11.12	5.39	*0.065
الجنس	طالب	0.098	2.432	0.13	-0.151	0.346	-0.75	0.80	-9.70	5.83	*0.062
	طالبة	0.105	2.454	0.10	-0.086	0.296	-0.83	1.11	-10.86	4.89	*0.072
الكليّة	إنسانية	0.125	2.475	0.10	-0.080	0.331	-0.96	1.50	-11.15	5.12	*0.075
	علمية	0.092	2.518	0.12	-0.142	0.327	-0.67	0.57	-10.13	5.44	*0.058

المتغير		الوسط الحسابي	الإنحراف المعياري	الخطأ المعياري للوسط	فترة الثقة لوسط المؤشر		الإلتواء	التفلطح	قيمة المؤشر:		K-S
					%2.5	%97.5			الصغرى	العظمى	
السنة الدراسية	الثالثة	0.134	2.515	0.11	-0.074	0.342	-0.79	1.00	-10.25	5.23	*0.067
	الرابعة	0.114	2.422	0.12	-0.113	0.341	-0.84	1.07	-10.78	5.35	*0.070
المعدل التراكمي	مقبول	0.110	2.474	0.14	-0.154	0.375	-0.71	0.63	-8.83	5.88	*0.077
	جيد فأعلى	0.119	2.436	0.09	-0.067	0.304	-0.83	1.22	-11.97	4.93	*0.061

* دال إحصائيا عند مستوى الدلالة α = 0.05

يلاحظ من الجدول 5 أنّ فترات الثقة للوسط الحسابي لقيم المؤشر الإحصائي I_2^{IV} لأنماط الاستجابة تبعًا لعينة الدِّراسة الكليّة ومستويي كلٍّ من (الجنس والكليّة والمستوى الدِّراسي والتقدير الجامعي) قد كانت متقاطعة، بما يفيد عدم وجود فرق جوهري بين الأوساط الحسابية لقيم المؤشر الإحصائي I_2^{IV} لأنماط الاستجابة، حيث تراوحت أوساطه الحسابية تبعًا لها من (0.092) لطلبة الكليات العلميّة وحتى (0.134) لطلبة مستوى السنة الدِّراسيّة الثالثة. وكون المعادلات المستخدمة في حساب الاحصائي تُشير الى اخذ حجم العينة بعين الاعتبار(Drasgow et al., 1985)، مما قد يؤدي الى عدم تأثُّر المؤشر الإحصائي I_2^{IV} لأنماط الاستجابة بشكل مباشر بحجم عينة الدِّراسة الكليّة أو بحجم عينة كل من مستويي (الجنس والكليّة والمستوى الدِّراسي والتقدير الجامعي) أثناء تقدير قيم المؤشر بالاعتماد على قدرات الطلبة، ومعالم عتبات الصعوبة لفقرات مقياس التقييم، في حين أنّه يتأثّر بشكل غير مباشر بحجم عينة الدِّراسة الكليّة أو بحجم عينة كلٍّ من مستويي (الجنس والكليّة والمستوى الدِّراسي والتقدير الجامعي) تبعًا لاختلاف كيفيّة توزُّع قدرات الطلبة فيما إذا كانت متوزِّعة بشكل منتظم ضمن مدى القدرة الخاص بكلٍّ منها؛ الأمر الذي يُقلِّص من قيمة الخطأ المعياريّ للمؤشر تبعًا لأيٍّ من عينة الدِّراسة الكليّة أو مستويي (الجنس والكليّة والمستوى الدِّراسي والتقدير الجامعي)، أم أنّها متوزِّعة بشكل غير منتظم ضمن مدى القدرة الخاص بكلٍّ منها. فإذا كانت قدرات الطلبة متمركزة على أحد طرفي التوزيع الطبيعي المُفترض لمعلمة القدرة أو عليهما معًا، فإنّ هذا يعود بأثر سلبي على دقة تقدير المؤشر الإحصائي، وبالتّالي يزيد في قيمة الخطأ المعياريّ له تبعًا لأيٍّ من عينة الدِّراسة الكليّة أو مستويي (الجنس والكليّة والمستوى الدِّراسي والتقدير الجامعي).

إضافة إلى ذلك، فإنّ المؤشر يتأثر بشكل مباشر باحتماليّة إجابة كل فقرة عند كل تدريج من تدريجاتها الخمسة لكل مستوى قدرة من مستويات قدرات الطلبة، فإذا كانت الاحتماليّات متوزِّعة بشكل منتظم ضمن مدى تدريجات الفقرة لكل مستوى قدرة من مستويات قدرات الطلبة، فإنّه يقلِّص مقدار تأثُّره بها؛ ممّا يُعني زيادة في موثوقيّة وفاعليّة تقدير المؤشر الإحصائي. أمّا إذا كانت متوزِّعة بشكل غير منتظم ضمن مدى تدريجات الفقرة لكلٍّ مستوى قدرة من مستويات قدرات الطلبة، فإنّه يتضخم مقدار تأثُّره بها؛ ممّا يُعني انخفاض في موثوقيّة وفاعليّة تقدير المؤشر الإحصائي (Drasgow *et al*, 1985).

أما قيم الإنحرافات المعياريّة للمؤشر الإحصائي I_2^{IV} لأنماط الاستجابة تبعًا لعينة الدِّراسة الكليّة ومستويي كلٍّ من (الجنس، الكليّة، المستوى الدِّراسي، التقدير الجامعي)، فقد تراوحت بين (2.422) لطلبة مستوى السنة الدِّراسية الرابعة - ممّا يعكس أنماط استجابات متقاربة قد تفتقر للموضوعيّة – و(2.518) لطلبة الكليات العلميّة - ممّا يعكس أنماط استجابات متباعدة قد تفتقر للموضوعيّة.

فيما يلاحظ من الجدول 5 أنّ قيم الأخطاء المعياريّة للأوساط الحسابية للمؤشر الإحصائي I_2^{IV} لأنماط الاستجابة تبعًا لعينة الدِّراسة الكليّة ومستويي كلٍّ من (الجنس والكليّة والمستوى الدِّراسي والتقدير الجامعي) قد تراوحت بين (0.09) لطلبة التقدير الجامعي جيد فأعلى، و(0.14) لطلبة التقدير الجامعي المقبول - علمًا بأنّ قيمته على مستوى العينة الكليّة قد بلغت (0.08). وهذا يعكس صورة عن وجه فاعليّة متقدِّم للمؤشر الإحصائي لدى طلبة التقدير الجامعي جيِّد فأعلى، مقارنةً بغيرهم من الطلبة، بدلالة تراص قيم المؤشر الإحصائي لديهم. وبالتّالي يعكس موثوقيّة أعلى في قيمة الوسط الحسابي للمؤشر الإحصائي. في حين أنّ قيمة الخطأ المعياريّ للوسط الحسابي للمؤشر الإحصائي I_2^{IV} لدى طلبة التقدير الجامعي المقبول يعكس صورة عن تراجع فاعليّة المؤشر الإحصائي لديهم مقارنة بغيرهم من الطلبة - علمًا بأنّ قيمته على مستوى العينة الكليّة قد بلغت (0.08)، وذلك بدلالة عدم تراص قيم المؤشر الإحصائي لديهم، ما يعكس موثوقيّة أقل في قيمة الوسط الحسابي للمؤشر الإحصائي.

كما يلاحظ من الجدول 5 أنّ قيم التواء المؤشر الإحصائي I_2^{IV} لأنماط الاستجابة تبعًا لعينة الدِّراسة الكليّة ومستويي كلٍّ من (الجنس والكليّة والمستوى الدِّراسي والتقدير الجامعي) قد تراوحت بين (-0.96) لطلبة الكليات الإنسانية و(-0.67) لطلبة الكليات العلميّة - ما يعكس صورة عن اعتداليّة ضعيفة لقيم المؤشر الإحصائي وبُعده عن التوزيع الطبيعي، وذلك لتمركُز معظم أنماط الاستجابة المطابقة في الجانب الأيمن من التوزيع العيني للمؤشر الإحصائي - و(-0.67) لطلبة الكليّات العلميّة - ما يعكس صورة عن اعتداليّة أقوى لقيم المؤشر الإحصائي وقربه من التوزيع الطبيعي،

وذلك لتمركُز معظم أنماط الاستجابة المطابقة في الجانب الأيمن من التوزيع العيني للمؤشر الإحصائي. وهذا يتّفق مع دراسة فان وميجر (Van & Meijer, 1999).

ويشير الجدول 5 أنّ قيم تفلطح المؤشر الإحصائي l_z^p لأنماط الاستجابة تبعًا لعينة الدِّراسة الكليّة ومستويي كلٍّ من (الجنس والكليّة والمستوى الدِّراسي والتقدير الجامعي) قد تراوحت بين (0.57) لطلبة الكليات العلميّة، و(1.50) لطلبة الكليات الإنسانية. أما القيم الصغرى للمؤشر الإحصائي l_z^p لأنماط الاستجابة تبعًا لعينة الدِّراسة الكليّة ومستويي كلٍّ من (الجنس، الكليّة، المستوى الدِّراسي، التقدير الجامعي) فقد تراوحت بين (11.97-) لطلبة التقدير الجامعي جيد فأعلى و(8.83-) لطلبة التقدير الجامعي المقبول، ما يعكس صورة عن اعتدالية أقوى لقيم المؤشر الإحصائي وقربه من التوزيع الطبيعي، وذلك لعدم تمركُز معظم أنماط الاستجابة المطابقة في المنطقة الوسطى من التوزيع العيني للمؤشر الإحصائي، و(1.50) لطلبة الكليّات الإنسانيّة، ما يعكس صورة عن اعتدالية ضعيفة لقيم المؤشر الإحصائي وبعده عن التوزيع الطبيعي؛ وذلك لتمركُز معظم أنماط الاستجابة المطابقة في المنطقة الوسطى من التوزيع العيني للمؤشر الإحصائي. وهذا يتّفق مع دراسة فان وميجر (Van & Meijer,1999).

كذلك يلاحظ من الجدول 5 أنّ القيم العظمى للمؤشر الإحصائي l_z^p لأنماط الاستجابة تبعًا لعينة الدِّراسة الكليّة ومستويي كلٍّ من (الجنس والكليّة والمستوى الدِّراسي والتقدير الجامعي) قد تراوحت بين (4.89) للطالبات و(5.88) لطلبة التقدير الجامعي المقبول، حيث تعكس القيمة العظمى للطالبات صورة عن ميلهن لعدم المطابقة أكثر في ضوء انخفاض قيمة القيمة العظمى للمؤشر الإحصائي لأنماط الاستجابة مقارنة بطلبة عينة الدِّراسة الكليّة، والطُلّاب، وطلبة مستويي (الكليّة، المستوى الدِّراسي، التقدير الجامعي) على الرغم من تطرفهن في مطابقتهن في ضوء قيمة القيمة العظمى للمؤشر الإحصائي لأنماط الاستجابة، إذ إنّ القيمة العظمى للمؤشر الإحصائي لأنماط الاستجابة البالغة (4.89) تتخطى القيمة الزائيّة المُناظرة لمستوى الدلالة α=0.05، ما يعكس صورة عن فاعليّة أكثر للمؤشر الإحصائي لأنماط الاستجابة، في حين أن قيمة القيمة العظمى لطلبة التقدير الجامعي المقبول تعكس صورة عن ميلهم للمطابقة أكثر مقارنة بطلبة عينة الدِّراسة الكليّة، وطلبة التقدير الجامعي جيِّد فأعلى، وطلبة مستويي (الجنس، الكليّة، المستوى الدِّراسي) على الرغم من تطرفهم في مطابقتهم في ضوء قيمة القيمة العظمى للمؤشر الإحصائي لأنماط الاستجابة، إذ إنّ القيمة العظمى للمؤشر الإحصائي لأنماط الاستجابة البالغة (5.88) تتخطى القيمة الزائيّة المُناظرة لمستوى الدلالة α=0.05، ما يعكس صورة عن فاعليّة أكثر للمؤشر الإحصائي لأنماط الاستجابة. وأخيرًا، يلاحظ من الجدول 5 أنّ قيم المؤشر الإحصائي l_z^p لأنماط الاستجابة تبعًا لعينة الدِّراسة الكليّة ومستويي كلٍّ من (الجنس، الكليّة، المستوى الدِّراسي، التقدير الجامعي) قد اختلفت عن التوزيع الطبيعي بدلالة إحصائية عند مستوى الدلالة (α=0.05) وفقًا لاختبار (Kolmogorov-Smirnov). حيث إنّ جميع القيم كانت تزيد على (0.05).

وبناء على الجدول 5، فإنّ التوزيع العيني للإحصائي l_z^p ملتو نحو اليسار وبتفلطح موجب ولا يتوزع توزيعا طبيعيًا. وتتفق هذه النتيجة مع غالبية الدراسات التي بينت أنّ l_z^p يميل للالتواء السالب والتفلطح الموجب (de la Torre & Deng, 2008; Reise, 1995; van Krimpen-Stoop & Meijer, 1999)، وقد يعود ذلك إلى أنّ معالم القدرات والفقرات المستخدمة في حسابه هي المقدرة من خلال نموذج الاستجابة المتدرجة وليس المعالم الحقيقة (Nering, 1995; Schmitt, Chan, Sacco, McFarland, & Jennings, 1999). ولذا فإنّ الاعتماد على درجة القطع التي تعتمد على القيمة الحرجة النظرية لتوزيع (Z) قد يؤدي إلى قرار خاطئ في تصنيف الاستجابات إلى مطابقة وغير مطابقة. كما أنّ الاستجابات الشاذة أو غير المطابقة في المقاييس النفسية غير المعرفية قد تكون ناتجة عن الاستجابات العشوائية للمستجيب أو عدم الجدية في الاستجابة عن الفقرات في المقياس(Karabtosos, 2003)، وهذا يجعل من استجابات الأفراد قيمًا متطرفة يسهل الكشف عنها، وبالتالي قد تؤدي إلى الالتواء السالب لتوزيع l_z^p.

كما حُسبت الإحصاءات الوصفية للتوزيعات العينية للخطأ المعياري للمؤشر الإحصائي l_z^p لأنماط الاستجابة وفق نموذج الاستجابة المتدرجة لمقياس تقييم جودة البرامج الأكاديميّة في جامعة اليرموك من وجهة نظر الطلبة، تبعًا لعينة الدِّراسة الكليّة ومستويي كلٍّ من (الجنس، الكليّة، المستوى الدِّراسي، التقدير الجامعي)، كما هو مبين في الجدول 6.

الجدول 6: الإحصاءات الوصفية للتوزيعات العينية للخطأ المعياري للمؤشر الإحصائي I_2^{ν} لأنماط الاستجابة لمقياس التقييم تبعًا لعينة الدّراسة الكليّة ومستويي كلٍّ من (الجنس، الكليّة، المستوى الدّراسي، التقدير الجامعي)

المتغيّر		الوسط الحسابي	الانحراف المعياريّ	الخطأ المعياريّ للوسط	فترة الثقة للخطأ المعياريّ		الالتواء	التفلطح	قيمة الخطأ المعياريّ	
					2.5	97.5			الصغرى	العظمى
العينة الكليّة		0.907	0.3064	0.0097	0.888	0.926	0.37	0.08	0.081	2.107
الجنس	طالب	0.949	0.3187	0.0166	0.916	0.981	0.36	0.11	0.122	2.167
	طالبة	0.930	0.3204	0.0127	0.906	0.955	0.80	1.46	0.150	2.403
الكليّة	إنسانيّة	0.980	0.3061	0.0130	0.955	1.005	0.63	0.77	0.152	2.325
	علميّة	1.169	0.3848	0.0183	1.133	1.205	0.86	0.97	0.462	2.763
المستوى الدّراسي	السّنة الثّالثة	0.956	0.3247	0.0137	0.929	0.983	0.54	0.58	0.227	2.217
	السّنة الرّابعة	1.040	0.3317	0.0158	1.009	1.071	0.47	-0.26	0.395	2.043
التقدير الجامعي	مقبول	1.119	0.4678	0.0256	1.069	1.169	1.72	4.73	0.341	3.633
	جيّد فأعلى	0.971	0.3328	0.0129	0.946	0.996	0.82	1.73	0.302	2.656

يلاحظ من الجدول 6 أنّ فترات الثقة للوسط الحسابي لقيم الخطأ المعياري للمؤشر الإحصائي I_2^{ν} لأنماط الاستجابة تبعًا لمستويي كلٍّ من (الجنس، الكليّة، المستوى الدّراسي، التقدير الجامعي) قد كانت غير متقاطعة، بما يفيد وجود فرق جوهري بين الأوساط الحسابية لقيم الخطأ المعياري للمؤشر الإحصائي I_2^{ν} لأنماط الاستجابة، لصالح قيمة الخطأ المعياري الأصغر (طلبة الكليات الإنسانية، طلبة المستوى الدّراسي الثالثة، طلبة التقدير الجامعي جيد فأعلى)، ما يُعني أنّ فاعليّة وموثوقيّة تقدير المؤشر الإحصائي لديهم قد كانت أفضل ممّا هي عليه لدى نظرائهم في المستوى الآخر للمتغيّر المعني، ومرجعيّة هذه الفاعليّة الموثوق بها لديهم مردُّها إلى وجود نمطيّة سائدة في أنماط الاستجابة لدى طلبة الكليّات الإنسانيّة، وطلبة مستوى السّنة الدّراسيّة الثّالثة، وطلبة التقدير الجامعي جيّد فأعلى، حدّت من تضخم الخطأ المعياريّ للمؤشر الإحصائي لديهم مقارنةً بطلبة الكليّات العلميّة، وطلبة مستوى السّنة الدّراسيّة الرّابعة، وطلبة التقدير الجامعي المقبول الذين تتوافر لديهم أنماط استجابة متباينة تخلو من النمطية، في حين تقاطعت فترتا الثقة للوسط الحسابي لقيم الخطأ المعياري للمؤشر الإحصائي I_2^{ν} لأنماط الاستجابة تبعًا لمستويي متغير الجنس، ممّا يُعني أنّ فاعليّة وموثوقيّة تقدير المؤشر الإحصائي لديهم قد كانت متكافئة.

كما يلاحظ من الجدول 6 أنّ قيم الأوساط الحسابية لقيم الخطأ المعياري للمؤشر الإحصائي I_2^{ν} لأنماط الاستجابة قد تراوحت تبعًا لها بين (0.930) للطالبات، و(1.169) لطلبة الكليات العلميّة -علمًا بأنّ قيمته على مستوى العينة الكليّة قد بلغت (0.907) - ما يعكس فاعليّة وموثوقيّة أكبر للمؤشر الإحصائي لديهم مقارنةً بما هو عليه لدى طلبة مستويي كلٍّ من (الجنس، الكليّة، المستوى الدّراسي، التقدير الجامعي).

وكذلك يتبين من الجدول 6 أنّ قيم الانحرافات المعياريّة لقيم الخطأ المعياري للمؤشر الإحصائي I_2^{ν} لأنماط الاستجابة قد تراوحت تبعًا لها بين (0.3061) لطلبة الكليات الإنسانية - ما يعكس المؤشر الإحصائي لأنماط استجابات متقاربة قد تفتقر للموضوعيّة - و(0.4678) لطلبة التقدير الجامعي المقبول - ما يعكس المؤشر الإحصائي لأنماط استجابات متباعدة قد تفتقر للموضوعيّة.

كما يبين الجدول 6 أنّ قيم الأخطاء المعياريّة للأوساط الحسابية لقيم الخطأ المعياري للمؤشر الإحصائي I_2^{ν} لأنماط الاستجابة قد تراوحت تبعًا لها بين (0.0127) للطالبات - ما يعكس صورة عن وجه فاعليّة متقدم للمؤشر الإحصائي لديهن مقارنةً بغيرهم من الطلبة بدلالة تراص قيم الخطأ المعياريّة للمؤشر الإحصائي لديهن ما يعكس موثوقيّة أعلى في قيمة الوسط الحسابي للخطأ المعياريّ للمؤشر الإحصائي- و(0.0256) لطلبة التقدير الجامعي المقبول - علمًا بأنّ قيمته على مستوى العينة الكليّة قد بلغت (0.0097)، ما يعكس صورة عن فاعليّة متراجعة للمؤشر الإحصائي لديهم مقارنة بغيرهم من الطلبة بدلالة عدم تراص قيم الخطأ المعياريّ للمؤشر الإحصائي لديهم وممّا يعكس موثوقيّة أقل في قيمة الوسط الحسابي للخطأ المعياريّ للمؤشر الإحصائي.

أما قيم الإلتواء لقيم الخطأ المعياري للمؤشر الإحصائي I_2^{ν} لأنماط الاستجابة، فقد تراوحت تبعًا لها بين (0.36) للطلاب - ما يعكس صورة عن اعتداليّة أقوى لقيم الخطأ المعياريّ للمؤشر

الإحصائي وقربه من التوزيع الطبيعي لتمركُز كمٍّ ما من قيم الخطأ المعياريّ للمؤشر الإحصائي لأنماط الاستجابة المطابقة في الجانب الأيمن من التوزيع العيني للمؤشر الإحصائي - و(1.72) لطلبة التقدير الجامعي المقبول - ما يعكس صورة عن اعتداليّة ضعيفة لقيم الخطأ المعياريّ للمؤشر الإحصائي وبُعده عن التوزُّع الطبيعي لتمرُكز معظم قيم الخطأ المعياريّ للمؤشر الإحصائي لأنماط الاستجابة المطابقة في الجانب الأيمن من التوزيع العيني للمؤشر الإحصائي.

فيما يلاحظ من الجدول 6 أنّ قيم التفلطح لقيم الخطأ المعياري للمؤشر الإحصائي l_z^p لأنماط الاستجابة قد تراوحت تبعًا لها بين (0.26-) لطلبة المستوى الدِّراسي الرابعة - ما يعكس صورة عن اعتداليّة أقوى لقيم الخطأ المعياريّ للمؤشر الإحصائي وقربه من التوزيع الطبيعي؛ لعدم تمركُز معظم قيم الخطأ المعياريّ للمؤشر الإحصائي لأنماط الاستجابة المطابقة في المنطقة الوسطى من التوزيع العيني للخطأ المعياريّ للمؤشر الإحصائي - و(1.73) لطلبة التقدير الجامعي جيد فأعلى -علمًا بأنّ قيمته على مستوى العينة الكليّة قد بلغت (0.08)، ما يعكس صورة عن اعتداليّة ضعيفة لقيم الخطأ المعياريّ للمؤشر الإحصائي وبُعده عن التوزيع الطبيعي لتمركُز معظم قيم الخطأ المعياريّ للمؤشر الإحصائي لأنماط الاستجابة المطابقة في المنطقة الوسطى من التوزيع العيني للخطأ المعياريّ للمؤشر الإحصائي. في حين يوضح الجدول 6 أنّ القيم الصغرى لقيم الخطأ المعياري للمؤشر الإحصائي l_z^p لأنماط الاستجابة قد تراوحت تبعًا لها بين (0.122) للطلاب و(0.462) لطلبة الكليات العلميّة -علمًا بأنّ أصغر قيمة له على مستوى العينة الكليّة قد بلغت (0.081)، حيث تعكس القيمة الصغرى لقيمة الخطأ المعياريّ للمؤشر الإحصائي لدى الطُلاب صورة عن ميلهم للمطابقة أكثر في ضوء انخفاض قيمة القيمة الصغرى للخطأ المعياريّ للمؤشر الإحصائي لأنماط الاستجابة مقارنة بطلبة عينة الدِّراسة الكليّة، والطّالبات، وطلبة مستويي (الكليّة، المستوى الدِّراسي، التقدير الجامعي) بما يتزامن مع مطابقتهم في ضوء قيمة القيمة الصغرى للخطأ المعياريّ للمؤشر الإحصائي لأنماط الاستجابة، إذ إنّ القيمة الصغرى للخطأ المعياريّ للمؤشر الإحصائي لأنماط الاستجابة البالغة قيمتها (0.122) بعد تحويلها إلى قيمة زائيّة لا تتخطى القيمة الزائيّة المُناظرة لمستوى الدلالة $\alpha = 0.05$، ما يعكس صورة عن فاعليّة أكثر للمؤشر الإحصائي لأنماط الاستجابة، في حين أنّ القيمة الصغرى لقيمة الخطأ المعياريّ للمؤشر الإحصائي لدى طلبة الكليّات العلميّة، ما يعكس صورة عن ميلهم لعدم المطابقة أكثر في ضوء ارتفاع قيمة القيمة الصغرى للخطأ المعياريّ للمؤشر الإحصائي لأنماط الاستجابة مقارنة بطلبة عينة الدِّراسة الكليّة وطلبة الكليّات الإنسانيّة وطلبة مستويي (الجنس، المستوى الدِّراسي، التقدير الجامعي) بما يتزامن مع مطابقتهم في ضوء قيمة القيمة الصغرى للخطأ المعياريّ للمؤشر الإحصائي لأنماط الاستجابة؛ إذ إنّ القيمة الصغرى للخطأ المعياريّ للمؤشر الإحصائي لأنماط الاستجابة البالغة قيمتها (0.462) بعد تحويلها إلى قيمة زائيّة لا تتخطى القيمة الزائيّة المُناظرة لمستوى الدلالة α=0.05، ما يعكس صورة عن فاعليّة أكثر للمؤشر الإحصائي لأنماط الاستجابة.

وأخيرًا، يلاحظ من الجدول 6 أنّ القيم العظمى لقيم الخطأ المعياري للمؤشر الإحصائي l_z^p لأنماط الاستجابة قد تراوحت تبعًا لها بين (2.043) لطلبة المستوى الدِّراسي الرابعة و(3.633) لطلبة التقدير الجامعي المقبول. حيث تعكس القيمة العظمى لقيمة الخطأ المعياريّ للمؤشر الإحصائي صورة عن ميلهم للمطابقة أكثر في ضوء انخفاض قيمة القيمة العظمى للخطأ المعياريّ للمؤشر الإحصائي لأنماط الاستجابة مقارنة بطلبة عينة الدِّراسة الكليّة، وطلبة السّنة الدِّراسيّة الثّالثة، وطلبة مستويي (الجنس، الكليّة، التقدير الجامعي)، على الرغم من تطرفهم في عدم مطابقتهم في ضوء قيمة القيمة العظمى للخطأ المعياريّ للمؤشر الإحصائي لأنماط الاستجابة؛ إذ إنّ القيمة العظمى للخطأ المعياريّ للمؤشر الإحصائي لأنماط الاستجابة البالغة قيمتها (2.043) بعد تحويلها إلى قيمة زائيّة تتخطى القيمة الزائيّة المُناظرة لمستوى الدلالة α=0.05، ما يعكس صورة عن فاعليّة أكثر للمؤشر الإحصائي لأنماط الاستجابة، في حين أنّ القيمة العظمى لقيمة الخطأ المعياريّ للمؤشر الإحصائي لدى طلبة مستوى السّنة الدِّراسيّة الرّابعة صورة عن ميلهم لعدم المطابقة أكثر في ضوء ارتفاع قيمة القيمة العظمى للخطأ المعياريّ للمؤشر الإحصائي لأنماط الاستجابة مقارنة بطلبة عينة الدِّراسة الكليّة، وطلبة التقدير الجامعي جيِّد فأعلى، وطلبة مستويي (الجنس، الكليّة، المستوى الدِّراسي)، على الرغم من تطرفهم في عدم مطابقتهم في ضوء قيمة القيمة العظمى للخطأ المعياريّ للمؤشر الإحصائي لأنماط الاستجابة، إذ إنّ القيمة العظمى للخطأ المعياريّ للمؤشر الإحصائي لأنماط الاستجابة البالغة (3.633) بعد تحويلها إلى قيمة زائيّة تتخطى القيمة الزائيّة المُناظرة لمستوى الدلالة α=0.05، ما يعكس صورة عن عدم فاعليّة أكثر للمؤشر الإحصائي.

ومن المتوقع أن يكون هنالك هامش من الخطأ في تقدير الإحصائي l_z^p، وقد يكون ذلك ناتجاً عن الأسلوب المستخدم في تقدير معالم الفقرات والقدرة. ففي الدِّراسة الحالية، اعتمد أسلوب بيز في التقدير، الذي يفترض المعرفة المسبقة للقدرات الحقيقية للأفراد التي تتوزع طبيعيًّا (Zimowski, Muraki, Mislevy, & Bock, 2003). كما أنّ وجود نسبة ملاحظة من الاستجابات غير المطابقة التي كانت تقريبا تساوي (21%) من العينة سيسهم في وجود هامش من الخطأ في حساب المؤشر l_z^p. كما أنّ الجدول 6 يشير إلى تقارب الوسط الحسابي للأخطاء المعياريّة للمؤشر l_z^p ، الذي قد يكون ناتجًا عن أحجام العينات المرتفع نسبيا حسب مستويات المتغيرات المستقلة، التي كان أقلها (335) للطلبة ذوي التقدير الجامعي أقل من جيد (مقبول)، وكذلك عدد فقرات الأداة التي بلغت 62 فقرة (Reise & Due, 1991).

وللإجابة عن السؤال الثاني، فقد حُسبت درجة القطع للإحصائي l_z^p من خلال استخدام الأسلوب نفسه وهو إعادة المعاينة لألف مرة، وفي كل مرة تمّ حساب قيمة l_z^p التي يقع أقل منها 5% من القيم، ومن ثم حساب الوسط الحسابي لتلك القيم لتشكل درجة القطع (Cutoff Score) التي يرمز لها بالرمز (l_z^b)، وكذلك حُسبت فترة الثقة (95%)، والجدول رقم (7) يبين ذلك.

الجدول 7: درجة القطع للإحصائي l_z^p من خلال استخدام أسلوب إعادة المعاينة لألف مرة

العينة	مستويات المتغير	N	l_z^b	الخطا المعياري لـ l_z^b	فترة الثقة لـ l_z^b	
					2.5%	97.5%
	العينة كاملة	1000	-1.5851	0.0881	-1.7685	-1.4401
الجنس	ذكور	367	-1.6112	0.0701	-1.7275	-1.4589
	إناث	633	-1.6278	0.0967	-1.8164	-1.4115
الكليّة	إنسانية	557	-1.4414	0.0818	-1.6291	-1.3293
	علمية	443	-1.6347	0.0954	-1.846	-1.4989
المستوى الدّراسي	ثالثة	562	-1.448	0.0875	-1.633	-1.3421
	رابعة	438	-1.5756	0.0731	-1.6863	-1.4151
التقدير الجامعي	مقبول	335	-1.5458	0.0741	-1.708	-1.4042
	جيد فاعلى	665	-1.4985	0.0691	-1.5824	-1.3085

يتبين من الجدول السابق أنّ القيمة الحرجة النظرية لتوزيع (Z) هي (1.6449-) تقع ضمن فترة الثقة عندما حُسبت نقطة القطع بالاعتماد على بيانات العينة كاملة، في حين أنّه عندما حُسبت نقطة القطع بالاعتماد على بيانات كل مستوى من مستويات المتغيرات الديمغرافية، فقد وُجد تعارض، حيث وقعت القيمة الحرجة النظرية لتوزيع (Z) في فترة الثقة لبيانات الذكور والإناث والكليات العلميّة وطلبة السنة الرابعة والطلبة الذين معدلهم التراكمي مقبول. في حين وقعت القيمة الحرجة خارج فترة الثقة لقيمة القطع المشتقة من خلال إعادة المعاينة لألف مرة في بيانات طلبة الكليات الإنسانية، وطلبة السنة الثالثة، وبيانات الطلبة الذين تقديرهم الجامعي جيد فأعلى.

ويمكن اعتبار قيم القطع المشتقة من خلال إعادة المعاينة متقاربة (وبشكل عام أقل من) القيمة الحرجة النظرية للتوزيع (Z). وقد يرجع ذلك إلى أنّ أسلوب إعادة المعاينة في تقدير درجة القطع قد أسهم في تقريب توزيع l_z^p إلى التوزيع الطبيعي، ولكن لم يصل إليه. ولكون الخطورة عند تصنيف الاستجابات إلى مطابقة وغير مطابقة تكون في تصنيف الاستجابة إلى مطابقة، عندما تكون في حقيقية الأمر غير مطابقة، لأنّ ذلك يؤدي إلى التحيز في تقدير معالم القدرة للأفراد والفقرات، ويؤدي كذلك إلى عدم التصنيف الصحيح للاستجابات إلى مطابقة وغير مطابقة (Hendrawan, Glas., & Meijer, 2005; Meijer & Nering, 1997)، لذا فإنّه من الأفضل الاعتماد على درجة القطع المشتقة من خلال إعادة المعاينة، لأنّها تستطيع الكشف عن عدد أكبر من الاستجابات غير المطابقة.

التوصيات

1- استخدام نقطة القطع للإحصائي l_z^b المشتقة من خلال أسلوب إعادة المعاينة وليس القيمة الحرجة النظرية لتوزيع (Z).

2- إجراء دراسات مشابهة يتم من خلالها استخدام طرق أخرى في تقدير القدرات ومعالم الفقرات غير طريقة بيز وباستخدام أطوال مختلفة للاستبانة أو المقياس.

3- مقارنة فعالية الإحصائي l_z^p المعتمد على درجة القطع المُشتقة من إعادة التعيين مع طرق أخرى للكشف على الاستجابات غير الجدية أو غير الشائعة.

المراجع

Barnes, T. (2016). *Detecting insufficient effort responding: An item response theory approach* (Master thesis, Wright State University, Dayton, USA).

Davis, J. (2009). *A comparative study of item-level fit indices in item response theory* (Doctoral dissertation, University of Minnesota, Minneapolis, USA).

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

Drasgow, F; Levine, M. & Williams, E. (1985). Appropriateness measurement with polychotomous item response models and standardized indices. *British Journal of Mathematical and Statistical Psychology,* 38, 67–86.

Efron, B. (1992). Bootstrap methods: Another look at the jackknife. In *Breakthroughs in Statistics* (pp. 569-593). New York: Springer

Ferrando, P. (2012). Assessing inconsistent responding in E and N measures: An application of person-fit analysis in personality. *Personality and Individual Differences*, *52*(6), 718-722.

Reise, S. (1995). Scoring method and the detection of person misfit in a personality assessment context. *Applied Psychological Measurement*, *19*(3), 213-229.

Reise, S. & Due, A. (1991). The influence of test characteristics on the detection of aberrant response patterns. *Applied Psychological Measurement, 15,* 217–226.

Rupp, A. (2013). A systematic review of the methodology for person fit research in item response theory: Lessons about generalizability of inferences from the design of simulation studies. *Psychological Test and Assessment Modeling*, 55(1), 3-38.

Schmitt, N; Chan, D; Sacco, J; McFarland, L. & Jennings, D. (1999). Correlates of person fit and effect of person fit on test validity. *Applied Psychological Measurement*, *23*(1), 41-53.

Sinharay, S. (2016a). Assessment of person fit using resampling-based approaches. *Journal of Educational Measurement*, *53*(1), 63-85.

Sinharay, S. (2016b). Asymptotically correct standardization of person-fit statistics beyond dichotomous items. *Psychometrika*, *81*(4), 992-1013.

Snijders, T. (2001). Asymptotic null distribution of person fit statistics with estimated person parameter. *Psychometrika*, *66*(3), 331-342.

Tendeiro, J. (2016). *PerFit: Person Fit. R package version 1.4.1*, URL https://CRANR-project.org/package=PerFit.

Tendeiro, J. (2017). The lz (p)* person-fit statistic in an unfolding model context. *Applied Psychological Measurement*, *41*(1), 44-59.

Tendeiro, J; Meijer, R. & Niessen, A. (2016). *PerFit: An R package for person-fit analysis in IRT*. *Journal of Statistical Software*, *74*(5). DOI: 10.18637/jss.v074.i05.

Tendeiro, J.& Meijer, R. (2014). Detection of invalid test scores: The usefulness of simple nonparametric statistics. *Journal of Educational Measurement*, *51*(3), 239–259. doi: 10.1111/jedm.12046.

van Krimpen-Stoop, E. & Meijer, R. (1999). The null distribution of person-fit statistics for conventional and adaptive tests. *Applied Psychological Measurement*, *23*(4), 327-345.

Ziegler, M; MacCann, C. & Roberts, R. (2011). Faking: Knowns, unknowns, and points of contention. *New perspectives on faking in personality assessment*, 3-16.

Zimowski, M; Muraki, E; Mislevy, R; & Bock, R. (2003). *BILOG-MG 3: Item analysis and test scoring with binary logistic models*. Chicago: Scientific Software.

Glas, C. & Meijer, R. (2003). A bayesian approach to person fit analysis in item response theory models. *Applied Psychological Measurement*, 27, 217–233.

Gorsuch, R. (1983). Factor analysis. 2nd. *Hillsdale, NJ: LEA*.

Gorsuch, R. (1997). New procedure for extension analysis in exploratory factor analysis. *Educational and Psychological Measurement*, *57*(5), 725-740.

Hattie, J. (1985). Methodology review: Assessing unidimensionality of tests and ltenls. *Applied Psychological Measurement*, *9*(2), 139-164.

Hendrawan, I; Glas, C. & Meijer, R. (2005). The effect of person misfit on classification decisions. *Applied Psychological Measurement*, *29*(1), 26-44.

Huang, J; Curran, P; Keeney, J; Poposki, E. & DeShon, R. (2012). Detecting and deterring insufficient effort responding to surveys. *Journal of Business and Psychology*, *27*(1), 99-114.

Jelínek, M; Květoň, P. & Vobořil, D. (2013). Assessment of response pattern aberrancy in eysenck personality inventory. *Feedback*, *58*(14), 36-47.

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

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

Meijer, R. & Nering, M. (1997). Ability estimation for non-fitting response vectors. *Applied Psychological Measurement, 21*, 321–336.

Meijer, R. & Sijtsma, K. (2001). Methodology Review: Evaluating person fit. *Applied Psychological Measurement*, *25*(2), 107-135.

Molenaar, I. & Hoijtink, H. (1996). Person-fit and the rasch model, with an application to knowledge of logical quantors. *Applied Measurement in Education*, *9*(1), 27-45.

Molenaar, I. & Hoijtink, H. (1990). The many null distributions of person fit indices. *Psychometrika, 55*, 75–106.

Nering, M. (1995). The distribution of person fit using true and estimated person parameters. *Applied Psychological Measurement*, *19*(2), 121-129.

Nering, M. (1997). The distribution of indexes of person fit within the computerized adaptive testing environment. *Applied Psychological Measurement*, *21*(2), 115-127.

R Core Team (2016). *R: A language and environment for statistical computing*. R foundation for statistical computing, Vienna, Austria. URL https://www.R-project.org/.