DOI: http://dx.doi.org/10.24200/jeps.vol13iss1pp27-45

أنماط الاستجابة الشاذة في اختبار اوتيس – لينون المقنن للبيئة الأردنية وأثرها في دقة تقديرات قدرة الفرد ودالة المعلومات

زايد صالح بني عطا*
جامعة اليرموك، الأردن

استُلم بتاريخ: ٢٠١٧/١٢/٢٥ قُبل بتاريخ: ٢٠١٨/٥/٣٠

ملخص: هدفت الدراسة إلى تقصي أنماط الاستجابة الشاذة في اختبار اوتيس – لينون المقنن للبيئة الأردنية وأثرها في دقة تقديرات معلمة القدرة للفرد ودالة معلومات الاختبار. ولتحقيق الهدف من الدراسة طُبّق اختبار اوتيس – لينون المستوى التمهيدي الثاني الصورة K المقنن للبيئة الأردنية على عينة مكونة من (٥٦٨) طالبا وطالبة من الصف الأول الأساسي في المدارس الحكومية التابعة لمحافظة عجلون للعام الدراسي ٢٠١٦/٢٠١٧، باستخدام مؤشر مطابقة الفرد Lz والنموذج اللوجستي ثلاثي المعلمة. وحُلّلت استجابات الطلبة عن فقرات الاختبار من أجل تقدير قدرة الأفراد ودالة معلومات الاختبار والكشف عن أنماط الاستجابة الشاذة. وأظهرت نتائج الدراسة وجود (٥٦) طالبا وطالبة كانت أنماط استجاباتهم شاذة بالاعتماد على المؤشر Lz، وبينت النتائج أن العوامل التي كانت مسؤولة عن ظهور هذه الأنماط الشاذة هي : التخمين، الغش، التكاسل والإبداع الاستثنائي .وأشارت النتائج كذلك إلى أنه عندما اُستُبعدت هذه الأنماط الشاذة من التحليل ازدادت دقة تقدير معلمة القدرة للفرد، وأن دالة معلومات الاختبار قد ازدادت بشكل ملحوظ عند مستويات القدرة المختلفة.

كلمات مفتاحية: أنماط الاستجابة الشاذة، النموذج اللوجستي ثلاثي المعلمة، دالة معلومات الاختبار، القدرة.

Aberrant Response Patterns on the Jordanian Version of Otis-Lennon Test and their Impact on the Accuracy of Estimating a Person's Ability and Information Function Test

Zaid S. Bani Ata*
Yarmouk University, Jordan

Abstract: This study aimed at investigating the aberrant response patterns and their impacts on the Jordanian version Otis- Lennon as well as the accuracy in the estimation of a person's' ability and information function test. To achieve this goal, the Jordanian version of the Ability Test primary II level form K was administrated to 568 first-grade male and female students of Ajloun district public schools during 2016/2017. The Lz person fit index and the three-parameter logistic model were used to analyze students' responses to test items to assess the person ability, information function test, and to detect aberrant response patterns. The results revealed that the response patterns of 56 students were aberrant based on Lz index; also the results showed that the factors responsible for the presence of this aberrant response were: guessing, cheating, laziness, and exponential creatively. The results indicated that when the aberrant patterns were excluded, both the accuracy person's estimating ability and the information function test had significantly increased at different ability levels.

Keywords: Aberrant response patterns, three-parameter logistic model, information function test, ability.

*zaid_baniata@yahoo.com

إن الهدف الرئيسي لعملية القياس النفسي والتربوي هو إعطاء تقديرات أو درجات للأفراد على السمات النفسية والتربوية المختلفة، ومن أجل إنجاح التطبيقات العملية للاختبارات والمقاييس النفسية في قياس تلك السمات، يجب أن تعكس التقديرات أو الدرجات الناتجة من استخدامها دقة عالية في وضع الأفراد على متصل السمة المستهدفة من عملية القياس (McDoland, 1999).

ولأن العلامة الناتجة من تطبيق الاختبار تعد الركيزة الأساسية التي يعتمد عليها في اتخاذ القرارات المتعلقة بالأفراد في المواقف العملية، فقد اهتم البحث السيكومتري بقضية صدق علامة الاختبار، من خلال النظر إلى صدق علامة الفرد وليس الاقتصار على الصدق من خلال المجموعة (Wright & Stone, 1999)؛ وذلك من أجل الوصول إلى أدق التفسيرات للأداء الحقيقي للمفحوص على الاختبار دون أخطاء تقود إلى قرارات مضللة (Johanson, Dulany & Bank, 2000). ونظرا لشيوع الاختبارات من نوع الاختيار من متعدد خصوصا اختبارات الذكاء، فإن استجابات المفحوصين على مثل هذا النوع من الاختبارات قد لا تعكس قدرة المفحوص الحقيقة بسبب عوامل كثيرة مثل الغش، والتخمين والتباطؤ (Belov & Armstrong, 2010; Meijer & Sijtsma, 2001)، حيث من المتوقع أن تتبع إجابة المفحوص نمطا معينا بحيث إن المفحوص من ذوي القدرة العالية يجيب إجابة صحيحة عن جميع الفقرات السهلة، والمفحوص من ذوي القدرة المتدنية يجيب إجابة خاطئة عن جميع الفقرات الصعبة، وعندما لا تكون إجابة المفحوص تتبع مثل هذا النمط فهذا يثير استفسارا عند تفسير علامته، ويعد نمط إجابته شاذا (Korir, 2014; Sotaridona,Pornal & Vallejo, 2003).

وتعد قضية الأنماط الشاذة للاستجابات (Aberrant Response Patterns) خصوصا عند استخدام الاختبارات من نوع الاختيار من متعدد من الخصائص التي اهتم بها البحث السيكومتري، لما لها انعكاس على صدق العلامة خاصة في المواقف التي تتطلب اتخاذ قرار بحق الفرد، مثل : التحصيل الدراسي، ومزاولة المهن، والقبول، والذكاء، والتنبؤ (Drasgow, Levine & Mclaughlin, 1987). ولقد وصف العديد من الباحثين (Birenbaum, 1986; Karabatsos, 2003; Meijier, 1996; Schmitt & Croker, 1984) أنواعا مختلفة من المفحوصين الذين تعد أنماط استجاباتهم الملاحظة لا تتفق مع نمط نموذجي باعتبارها أنماطا شاذة وتشمل: التخمين (Guessing) وهو الإجابة الصحيحة عن فقرة صعبة من المفحوص ذي القدرة المتدنية. الغش (Cheating)، وذلك بنقل الإجابة الصحيحة من المفحوص ذي القدرة المتدنية عن مفحوص من ذوي القدرة العالية. التباطؤ (Plodding)، وهي الحالة التي يكون فيها أداء المفحوص بطيئا أثناء تأدية الاختبار حتى يتأكد من أن إجابته صحيحة. التكاسل (Sleeping)، حيث يواجه المفحوص في هذه الحالة مشكلات عند بدء الاختبار وبعد التكيف مع الاختبار، لا يقوم بمراجعة الفقرات التي أجاب عنها في بداية الاختبار. القلق (Anxiety)، حيث يكون المفحوص في حالة اضطراب سببها الخوف من الفشل حيث يؤدي ذلك إلى تقليل التركيز أثناء تأدية الاختبار؛ مما قد يؤدي إلى إجابة الفقرات الصعبة وعدم إجابة الفقرات السهلة. وأخيرا الدافعية (Motivation)، فقد تؤدي الدافعية الزائدة، أو عدم الاهتمام بأداء الاختبار إلى أنماط شاذة.

وقد أشار اسوناس وبل وديفيد (Iasonas, Bill & Davied, 2000) بأن تلك الأنواع السابقة تعد بمثابة العوامل والأسباب التي تؤدي إلى حدوث الأنماط الشاذة أطلقوا عليها

الخصائص الشخصية للفرد (Personal Characteristics)، وأضافوا لذلك عامل الخصائص التربوية (Educational Characteristics)المتمثل بسوء فهم الفقرة الذي يشير إلى إجابة الفقرة بشكل خاطئ بناء على معلومات خاطئة مكتسبة، والإبداع الاستثنائي، وهو قيام المفحوص بتفسير الفقرات السهلة بطريقة إبداعية غير تقليدية؛ مما يجعله يجيب عن الفقرة إجابة خاطئة.

ونظرا لأهمية التعرف إلى أنماط الاستجابات الشاذة في الاختبارات والمقاييس النفسية والتربوية، فقد تم الدعوة خلال العقود الماضية إلى إيجاد النماذج والمؤشرات للحكم على مطابقة استجابة الفرد للنموذج السيكومتري المعتمد، وقد عرفت تلك المؤشرات باسم مؤشرات مطابقة الفرد (Person Fit Statistic)، وهي عبارة عن مؤشرات إحصائية لقياس مدى ابتعاد نمط الإجابة الملاحظ عن النمط المتوقع من النموذج السيكومتري (De Latorre & Deng, 2008; Lamprianeu, 2013; Karabatsos,2003; Meijer, Molenaar & Sijtsma, 1994).

وقد بلغ عدد تلك المؤشرات الإحصائية المعلمية واللامعلمية المستخدمة في الكشف عن مدى ابتعاد نمط الإجابة الملاحظ للمفحوص عن النمط المتوقع من النموذج السيكومتري (٣٦) مؤشرا، غالبيتها يتعامل مع الاستجابات ثنائية التدريج (Karabatsos, 2003). ومن تلك المؤشرات الإحصائية الشائعة الاستخدام في الكشف عن أنماط الاستجابة الشاذة عند استخدام النماذج المعلمية ثنائية التدريج، مؤشر Lz (Snijders, 2001).

ويعد مؤشر L_z لكل من دراسجو وليفين ووليمز (Drasgow, Levine & Williams, 1985) بمثابة نسخة مطورة عن المؤشر الإحصائي L_o الذي اقترحه الذي اقترحه ليفين وماكلوفلن (Levine & Rubin, 1979)؛ وذلك للتغلب على المشكلتين اللتين ظهرتا عند استخدام المؤشر L_o في الكشف عن أنماط الاستجابة الشاذة، والمتمثلة في أن القيم العددية للمؤشر L_o تعتمد على مستوى القدرة للفرد؛ إذ إنه من الممكن أن يختلف تصنيف استجابات الأفراد على اعتبار أنها مطابقة أو غير مطابقة للنموذج المستخدم حسب مستوى القدرة، وكذلك بأن هذا المؤشر لا يوجد له توزيع إحصائي محدد، حيث يعتمد التوزيع الإحصائي له على مستويات القدرة للأفراد، وبالتالي لا يمكن استخدام قيمة حرجة ثابتة عبر مستويات القدرة المختلفة للحكم على مدى مطابقة الاستجابة للنموذج، لذلك قاموا بمعايرة المؤشر L_o، وأصبح يعرف بالمؤشر L_z الذي يحسب وفق المعادلة رقم ١، ويتوقع أن يتوزع توزيعا طبيعيا بوسط حسابي مقداره صفر وانحراف معياري مقداره واحد

$$L_z = \frac{L_o - E(L_o)}{[Var(L_o)]^{0.5}} \quad \dots\dots\dots\dots\dots\dots 1$$

وقد أشار دراسجو وزملاؤه (Drasgow et al., 1987) من خلال دراستهما التي هدفت إلى المقارنة بين تسعة مؤشرات من ضمنها المؤشر L_z من حيث قوة المؤشر في الكشف عن الأنماط الشاذة، بأن المؤشر L_z كان أكثرها حساسية في تحديد أنماط الاستجابة الشاذة، وكذلك بين لي وليجينيك (Li & Olejnike, 1997) بأن مؤشر L_z قد كشف عن أعلى نسبة من الأنماط غير مطابقة في حالة البيانات أحادية البعد مقارنة مع أربعة مؤشرات معيارية (ECI42, ECI2z, ZU, ZW). ونظرا لفاعلية هذا المؤشر الإحصائي وحساسيته في الكشف عن أنماط الاستجابة الشاذة وشيوع استخدامه في الدراسات والأبحاث التي اهتمت بذلك (Seo & Weiss, 2013)، فقد تم استخدامه في هذه الدراسة.

ونظرا للاستخدام المتزايد للمؤشرات الإحصائية في تحديد أنماط الاستجابات الشاذة، فقد اتجهت غالبية الدراسات والأبحاث نحو الكشف عن الخصائص

الإحصائية لهذه المؤشرات الإحصائية ومقارنة هذه المؤشرات والعلاقة بينها للتحقق من مدى فاعليتها، فقد أجرى رايس وديو (Reise & Due, 1991) دراسة هدفت التعرف إلى أثر طول الاختبار (٧، ٢١، ٣٥، ٤٩)، ومستوى صعوبة الفقرات (- 3،3)، (- 2،2) و (- 1، 1) عند خمسة مستويات للتخمين (٠,٠، ٠,٠٥، ٠,١٠ ، ٠,١٥، ٠,٢) على قوة المؤشر الإحصائي L_z في تحديد الأنماط غير مطابقة للنموذج اللوجستي ثلاثي المعلمة بالاعتماد على المحاكاة. بينت نتائج الدراسة أنه بزيادة طول الاختبار تزيد نسبة الكشف عن الأنماط غير مطابقة، وكذلك باتساع مدى الصعوبة للفقرات، وبينت النتائج كذلك بأنه كلما قل التخمين زادت نسبة الكشف عن الأنماط غير مطابقة لذوي القدرة المتدنية، في حين لم تتأثر هذه النسبة عن الأنماط غير مطابقة لذوي القدرة العالية.

وفي دراسة نيرنغ (Nering, 1997) التي هدفت التعرف إلى خصائص توزيع المؤشرين (L_z, ζ_2) ضمن الاختبارات التكيفية، ومن أجل تحقيق الهدف من الدراسة تم توليد بيانات مطابقة للنموذج اللوجستي ثلاثي المعلمة تحت ثلاث شروط مختلفة من التوزيعات. أشارت نتائج الدراسة بأن توزيع خصائص المؤشرين لا تتبع خصائص التوزيع الطبيعي على غرار الاختبارات التقليدية، حيث تراوح الوسط الحسابي لقيم المؤشر L_z من (٠,٢٦٨) إلى (٠,٦٦٨)، في حين تراوح الوسط الحسابي لقيم المؤشر ζ_2 من (٠,٣٥٢) إلى (٠,٧٦٤) وكانت قيم الانحراف المعياري محصورة ما بين (٠,٥٢١) إلى (٠,٧٦٤) للمؤشر L_z ، بينما كانت للمؤشر ζ_2محصورة ما بين (٠.٥٦) الى (٠,٩١٧).

وقام لوبيز ومونتيسينوس (Lopez & Montesinos, 2005) بدراسة هدفت إلى مقارنة خمسة مؤشرات إحصائية (Lz, ECI42, ECI2z, Zw,Zu) بالاعتماد على المحاكاة من خلال توليد ثلاث مستويات لحجم العينة (١٠٠، ٢٥٠، ٥٠٠، ١٠٠٠)، ومستويين لمدى صعوبة الفقرات (- 1، 1) و (- 2، 2)، وشكل التوزيع للقدرة (طبيعي ، منتظم)، ومستويين لطول الاختبار (١٥، ٣٠). بينت نتائج الدراسة أن الوسط الحسابي والانحراف المعياري للمؤشرين (Lz, Zw) لم يكونا كما هو متوقع عندما تم استخدام معلمة القدرة المقدرة وليس الحقيقية، وعلى الرغم من اختلاف الأساس الرياضي لهما إلا أنهما اظهرا السلوك نفسه تحت جميع الظروف، حيث تحسنت خصائص التوزيع الطبيعي لهما عندما ضاق مدى الصعوبة للفقرات، وزاد طول الاختبار. وبينت النتائج كذلك بأن الوسط الحسابي والانحراف المعياري للمؤشر ECI42 كانا قريبين من خصائص التوزيع الطبيعي، وقد تراوحت نسبة الكشف عن الأنماط غير مطابقة للمؤشرين (ECI2z, ECI4z)) ما بين ٥% و ١٠%، وكانت أعلى من نسبة الكشف للمؤشرات.(Lz, Zw, Zu)

وأجري دي لا توري ودنغ (De La Toree & Deng, 2008)دراسة هدفت إلى اقتراح طريقة جديدة لتحسين فاعلية المؤشر الإحصائي L_z في الكشف عن أنماط الاستجابة غير المطابقة من خلال عمل تعديل للقدرة، وآخر في توزيع القدرة ضمن ثمان خطوات. أظهرت نتائج الدراسة أن طريقة تصحيح تقدير القدرة وتوزيعها المرجعي قد حسنت من فاعلية المؤشر الإحصائي في تحديد أنماط الاستجابة الشاذة، أو ما يسمى غير مطابق.

وهدفت دراسة حمادنة (٢٠١٥) إلى تقصي فاعلية المؤشر الإحصائي L_z^{*new} في الكشف عن الأنماط غير مطابقة للنموذج المستخدم، بدلالة الخطأ من النوع الأول واختلاف النموذج المستخدم (ثنائي المعلمة وثلاثي المعلمة) عند ١١ مستوى من مستويات

القدرة، وثلاثة مستويات مختلفة للدلالة الإحصائية بالاعتماد على المحاكاة. أظهرت نتائج الدراسة فاعلية المؤشر الإحصائي في تحديد أنماط الاستجابة غير مطابقة مهما اختلف النموذج، وأنه كان أكثر فاعلية في النموذج اللوجستي ثلاثي المعلمة مقارنة مع النموذج الثنائي، كما أشارت النتائج إلى أن الخطأ من النوع الأول كان قريبا من المستوى الاسمي في النموذج الثلاثي عبر مستويات القدرة المختلفة مهما اختلف مستوى الدلالة الإحصائية.

ومن جهة أخرى نجد قلة الدراسات التي اهتمت بدراسة انعكاس أنماط الاستجابات الشاذة على دقة القياس كونه يعد محورا رئيسا من محاور اهتمام عملية القياس النفسي والتربوي، حيث اهتمت بعض الدراسات بتناول أثر أنماط الاستجابات الشاذة على صدق الاختبارات وثباتها.

فقد أجرى شمت وكوريتنا وليتني (Schmitt, Cortina & Whiteny, 1993) دراسة هدفت إلى تقصي أثر الأنماط غير المطابقة على الصدق المرتبط بمحك باستخدام مؤشر(L_z). ولتحقيق الهدف من الدراسة تم تطبيق بطارية اختبارات في الميكانيكيا على (٢٠١) فردا. أظهرت نتائج الدراسة أن هناك تحسنا قليلا في معامل الارتباط بين المتنبئ والمحك عند حذف الأنماط غير المطابقة.

وفي السياق نفسه قام ميجر (Meijer, 1997) بدراسة هدفت إلى تقصي أثر وجود الأنماط غير المطابقة على صدق علامات الاختبار. ولتحقيق الهدف من الدراسة ولد بيانات وفق النموذج اللوجستي ثلاثي المعلمة ومؤشر المطابقة (L_z). كشفت نتائج الدراسة عن تأثر صدق علامات الاختبار بوجود الأنماط غير المطابقة عندما تزيد نسبتها على (١٥%) . وكذلك بين نيرنغ (Nering, 1998) بدراسته التي هدفت إلى محاولة فهم سلوك أنماط الاستجابة غير المطابقة على معالم الأفراد المطابقين، أظهرت النتائج بأن إدخال الاستجابات غير مطابقة في عملية المعايرة تؤثر في تقدير معالم الأفراد، وأن حذفها من عملية المعايرة يحسن من تقدير معالم الأفراد، ولهذا فقد اقترح عمل معايرتين، الأولى تكمن في تحديد الأنماط الشاذة، والمعايرة الثانية تتم بعد حذف تلك الأنماط.

وعلى الرغم من كثرة الدراسات التي اهتمت بمؤشرات مطابقة الفرد للكشف عن أنماط الاستجابات الشاذة وتطوير النماذج الرياضية التي تستند إليها، إلا انه يلاحظ ندرة الدراسات التي اهتمت بدراسة أثر مثل هذه الاستجابات في دقة تقدير معلمة القدرة خصوصا على المستوى المحلي والعربي والعوامل المسئولة عنها - في حدود علم الباحث - ، حيث أشار نيرنغ (Nering, 1998) أن وجود الأنماط غير مطابقة في عملية المعايرة قد يؤدي إلى تقديرات مضللة للقدرة. لذا، برزت الحاجة لمثل هذه الدراسة لتقصي العوامل المسئولة عن ظهور أنماط الاستجابة الشاذة وأثرها في دقة تقدير معلمة القدرة ودالة معلومات الاختبار والتي تعد صفة للاختبار ولأهميتها أيضا في مقارنة اختبارات تقيس السمة نفسها كما أشار كروكر والجينا (Croker & Algina, 1986)، أو مقارنة طريقتين مختلفتين لتصحيح الاختبار نفسه. وتجدر الإشارة كذلك بأن كمية المعلومات التي يعطيها الاختبار تتناسب عكسيا مع الخطأ المعياري لتقدير معلمة القدرة. واستخدمت الدراسة الحالية النموذج اللوجستي ثلاثي المعلمة في معايرة البيانات، ومؤشر L_z في تحديد أنماط الاستجابة الشاذة.

مشكلة الدراسة وأسئلتها

تعد دراسة أنماط الاستجابة الشاذة على الاختبارات والمقاييس النفسية والتربوية واحدة من القضايا السيكومترية التي يواجها الباحثون ضمن عملية القياس النفسي والتربوي خصوصا، في المواقف التي تتطلب

فيها اتخاذ قرارات بحق الأفراد، ويبدو هذا جليا بالاهتمام البحثي الذي تركز بإيجاد المؤشرات والنماذج الإحصائية للكشف عن مثل هذه الاستجابات، وكذلك تحديد العوامل والأسباب المحتملة التي أدت إلى ظهورها، لما لهذه الاستجابات من أثر في تقدير قدرة الأفراد عند استخدام نماذج نظرية الاستجابة للفقرة في معايرة الاختبارات والمقاييس، وجاءت هذه الدراسة لتكشف عن العوامل المحتملة وراء ظهور أنماط الاستجابات الشاذة، وأثرها في دقة تقدير معلمة القدرة للفرد ودالة معلومات الاختبار. لذا حاولت هذه الدراسة على وجه التحديد الإجابة عن السؤالين الآتين :-

١. ما نسبة أنماط الاستجابة الشاذة اعتمادا على المؤشر الإحصائي Lz، وما العوامل والأسباب التي أدت إلى ظهورها ؟
٢. هل تختلف دقة تقدير معلمة القدرة للفرد ودالة معلومات الاختبار عند كل مستوى من مستويات القدرة قبل وبعد حذف أنماط الاستجابة الشاذة ؟

أهمية الدراسة

تكمن الأهمية النظرية للدراسة الحالية في محاولتها الكشف عن الأسباب المحتملة التي تكمن وراء ظهور أنماط الاستجابة الشاذة في اختبار القدرة العقلية (اوتيس – لينون المقنن للبيئة الأردنية)، وأثر تلك الأنماط في دقة تقديرات معلمة الفرد ودالة معلومات الاختبار، الأمر الذي قد يتيح التنبؤ بالآثار السلبية لهذه الأنماط في دقة تقديرات القدرة ودالة معلومات الاختبار. وتكمن أهمية الدراسة كذلك في كونها تعد من الدراسات النادرة على المستوى المحلي والعربي التي اهتمت بالبحث عن العوامل التي تؤدي إلى ظهور أنماط الاستجابات الشاذة ودراسة انعكاساتها على تقديرات القدرة، ودالة المعلومات التي من المؤمل من نتائجها فتح الأبواب أمام الدراسات العربية في تناول أنماط الاستجابات الشاذة، ومؤشراتها، والتحقق من مدى أثرها في تقديرات القدرة، والبحث عن المؤشرات الإحصائية الأكثر فاعلية في الكشف عنها.

مصطلحات الدراسة

دقة التقدير: وهو مصطلح يشير إلى جودة التقدير وذلك بالاعتماد على قيمة الخطأ المعياري في التقدير لمعلمة قدرة الفرد الناتجة من معايرة البيانات باستخدام النموذج اللوجستي ثلاثي المعلمة.

نمط الاستجابة الشاذ: وهو النمط الملاحظ لاستجابات الفرد عن فقرات الاختبار، ولا يتوافق مع النمط المتوقع من النموذج اللوجستي ثلاثي المعلمة، ويتم الكشف عنه في هذه الدراسة باستخدام المؤشر الإحصائي L_z، ويعد نمط الاستجابة للفرد شاذا إذا كانت قيمة هذا المؤشر للفرد أقل من القيمة الحرجة وهي في هذه الدراسة (-١,١٤٨).

دالة معلومات الاختبار: وهي المجموع الحسابي لدوال معلومات الفقرات المكونة للاختبار، وتستخدم في تحديد الدقة التي تقدر بوساطتها قدرات المفحوصين من - ٤ إلى ٤

الطريقة والإجراءات

منهجية الدراسة

لتحقيق أهداف الدراسة الحالية في الكشف عن العوامل المسئولة عن ظهور أنماط الاستجابة الشاذة، وأثرها في دقة تقديرات معلمة القدرة للفرد ودالة معلومات الاختبار، استخدم المنهج الوصفي الذي يعد الأنسب لمثل هذه الدراسات.

مجتمع الدراسة وعينتها

تكون مجتمع الدراسة من جميع طلبة الصف الأول الأساسي المسجلين في المدارس الحكومية التابعة لمديرية التربية والتعليم

في محافظة عجلون للعام الدراسة ٢٠١٦/٢٠١٧ والبالغ عددهم (٢٥٤٥) طالبا وطالبة موزعين إلى (٧٠) مدرسة (إناث، ذكور، مختلطة) حسب السجلات الرسمية للمديرية. تكونت عينة الدراسة من (٥٦٨) طالبا وطالبا موزعين الى (١٨) مدرسة اختيرت بالطريقة العنقودية العشوائية حيث كانت وحدة الاختيار المدرسة.

أداة الدراسة

تم استخدام اختبار اوتيس – لينون المستوى التمهيدي الثاني الصورة K، حيث يطبق على طلاب الصف الأول الأساسي في النصف الأول من العام الدراسي، وقام بني عطا وحجازي (٢٠١١) بإيجاد دلالات الصدق والثبات لهذه الصورة المطورة للبيئة الأردنية، وتم الكشف عن دلالات صدق هذه الصورة من الاختبار بإيجاد كل من الصدق التمييزي وصدق البناء، والصدق المرتبط بمحك. أما دلالات الثبات فقد تم الكشف عنها من خلال طريقة الإعادة، وكانت قيمة معامل الثبات للاختبار(٠,٨٣)، والطريقة النصفية، وقد بلغت قيمة معامل الثبات بهذه الطريقة (٠,٨٧)، وأخيرا باستخدام معادلة كودر - ريتشاردسون ٢٠، حيث بلغت قيمة معامل الثبات المحسوب بهذه المعادلة (٠,٨٩) كمؤشر على الاتساق الداخلي. وكذلك تم اشتقاق معايير الأداء للاختبار، وتمثلت بنسبة الذكاء الانحرافية DIQ، والرتب المئينية، والتساعيات.

يتكون الاختبار من (٥٥) فقرة مصورة موزعة إلى قسمين ضمن كراس الاختبار، القسم الأول تكون من (٢٣) فقرة تتعلق بالتصنيف الهندسي، أما القسم الثني فتكون من (٣٢) فقرة مصورة موزعة كما يلي: (٦) فقرات مصورة تتعلق بإتباع التعليمات، (١٠) فقرات مصورة للاستدلال الكمي، وأخيرا (١٦) فقرة مصورة تتعلق بالمفردات المصورة والمعلومات العامة، وتصحح إجابة المفحوص بإعطائه العلامة واحد إذا أجاب إجابة صحيحة عن الفقرة والعلامة صفر إذا أجاب إجابة خاطئة عن الفقرة. ينفذ الاختبار على جلستين بينهما استراحة قصيرة، بحيث يكون زمن القسم الأول (١٨) دقيقة وزمن القسم الثاني (٢٠) دقيقة.

إجراءات التطبيق والتحليل

أولا: إجراءات التطبيق

تم تطبيق أداة الدراسة على عينة الدراسة، وذلك بالتنسيق مع معلمي ومعلمات الصف الأول الأساسي، حيث تم تقديم معلومات تفصيلية عن طريقة تطبيق الاختبار، وتحديد جدول زمني كل حسب شعبته للبدء بتطبيق الاختبار على الطلبة، حيث استغرقت عملية تطبيق الاختبار شهرا. وبعد الانتهاء من عملية تطبيق الاختبار على عينة الدراسة، قام الباحث بتصحيح استجابات الطلبة حسب مفتاح التصحيح للاختبار، وذلك بإعطاء العلامة واحد للإجابة الصحيحة، والعلامة صفر للإجابة الخاطئة، وبذلك يكون مدى العلامات على الاختبار من صفر إلى (٥٥). ومما تجدر الإشارة إليه أنه تم تغيب (٦) من الطلبة أثناء السير بتطبيق الاختبار بحيث أصبح عدد أفراد العينة (٥٦٢) طالبا وطالبة، حيث تم تجهيز ملف البيانات لاستجاباتهم لغايات التحليل.

ثانيا: إجراءات التحليل

من أجل الإجابة عن أسئلة الدراسة قام الباحث بتجهيز ملف استجابات الطلبة على صورة صفر وواحد، واستخدام برنامج SPSS، وبرنامج NOHARM، وبرنامج bilog-mg3، وبرنامج LIDD، وبرنامج Per fit من خلال برمجية R-Language لإجراء التحليلات الإحصائية اللازمة للإجابة عن أسئلة الدراسة. وقد مرت عملية التحليل بالمراحل التالية:

١- التحقق من افتراضات النموذج

تم التحقق من افتراضات النموذج اللوجستي ثلاثي المعلمة، وهو أحادية البعد من خلال

استخدام التحليل العاملي، وقد أفرزت نتائج التحليل العاملي (١٦) عاملا، الجذر الكامن (Eigen Value)لكل منها يزيد على واحد صحيح، وتفسر مجتمعة (٥٦,١٦٢%) من التباين. ويبين جدول ١ نسبة التباين التي يفسرها كل من العوامل الستة عشر وقيم الجذور الكامنة لها.

جدول ١

نسب التباين المفسر والجذر الكامن للعوامل الستة عشر لعينة الدراسة

العامل	الجذر الكامن	نسبة التباين المفسرة	نسبة التباين التراكمية
الأول	٩,٣٤٣	١٦,٩٨٨	١٦,٩٨٨
الثاني	٢,٥٢١	٤,٥٨٤	٢١,٥٧٢
الثالث	٢,٣٩٩	٤,٣٦٢	٢٥,٩٣٤
الرابع	١,٦٦٥	٣,٠٢٧	٢٨,٩٦٠
الخامس	١,٥٤٨	٢,٨١٥	٣١,٧٧٥
السادس	١,٤٧٨	٢,٦٨٧	٣٤,٤٦٢
السابع	١,٤٦٨	٢,٦٦٩	٣٧,١٣١
الثامن	١,٣٣٥	٢,٤٢٧	٣٩,٥٥٨
التاسع	١,٣١٠	٢,٣٨٢	٤١,٩٤٠
العاشر	١,٢١٧	٢,٢١٤	٤٤,١٥٣
الحادي عشر	١,١٩١	٢,١٦٦	٤٦,٣١٩
الثاني عشر	١,١٥٦	٢,١٠٢	٤٨,٤٢٢
الثالث عشر	١,١٠٦	٢,٠١١	٥٠,٤٣٢
الرابع عشر	١,٠٨٧	١,٩٧٦	٥٢,٤٠٨
الخامس عشر	١,٠٤٢	١,٨٩٥	٥٤,٣٠٤
السادس عشر	١,٠٢٢	١,٨٥٩	٥٦,١٦٢

تشير النتائج الواردة في الجدول ١ بأن ناتج قسمة الجذر الكامن للعامل الأول على العامل الثاني تزيد على ٢، مما يعني وجود سمة سائدة للاختبار؛ وهذا يعني تحقق افتراض أحادية البعد للاختبار (Hatti, 1985 ; Reckase, 1997).

كما تم استخدام برنامج NOHARM بغرض التحقق من أحادية البعد للبيانات من خلال إجراء تحليل البواقي Residual Analysis، حيث يتم فحص المواءمة بين عدد الأبعاد للاختبار والبيانات من خلال حساب مصفوفة البواقي Residual Matrix لإيجاد قيمة الجذر التربيعي لمتوسطات مربعات البواقي Root Mean Square of Residuals) (RMSR)، وقد بلغت قيمة الجذر التربيعي لمتوسطات مربعات البواقي (٠,٠١٣) وقيمة مؤشر تناكا بلغت (٠,٩٢٧)، وقد بين ميلر (Miller, 1991) أنه إذا كانت قيمة مؤشر RMSR قريبة من الصفر، وقيمة مؤشر تناكا (Tanaka Index) قريبة من الواحد، فيعد مثل هذا دليل على تحقق أحادية البعد. وبالنظر إلى قيمة RMSR نجد بأنها قريبة من الصفر، وكذلك اقتراب قيمة مؤشر تناكا من الواحد، مما يعني تحقق افتراض أحادية البعد في الاختبار.

ومن أجل التحقق من افتراض الاستقلال الموضعي، قام الباحث باستخدام برمجية LDID لكيم وكوهين ولين (Kim, Cohen & Lin, 2005) لإيجاد قيم المؤشر الإحصائي Q_3 لين (Yen, 1993) الذي يعد من المؤشرات المشهورة في الكشف عن الاستقلال الموضعي بين فقرات الاختبارات ثنائية التدريج، واعتمادا على درجة القطع المحددة لهذا المؤشر، وهي القيمة المطلقة للقيمة (٠,٢)(Yen, 1993). وقد بينت نتائج التحليل بأن قيم المؤشر الإحصائي Q_3 لأي زوج من أزواج الفقرات قد جاءت أقل من القيمة الحرجة، مما يعني تحقق افتراض الاستقلال الموضعي بين فقرات الاختبار، وفي الوقت نفسه كانت قيمة الوسط الحسابي لقيم المؤشر الإحصائي Q_3 تساوي (- ٠,٠٠٢) وهي أقل من القيمة المتوقعة للمؤشر الإحصائي ((-1/(k-1))، حيث يشير الرمز k الى عدد فقرات الاختبار، وقد أشارت ين (Yen, 1993) بأنه إذا كان الوسط الحسابي لقيم المؤشر الإحصائي Q_3 قريبة من القيمة المتوقعة له، فيعد مثل ذلك دليلا على أن افتراض الاستقلال الموضعي متحقق بين فقرات الاختبار.

٢- مطابقة البيانات للنموذج

للتحقق من مطابقة البيانات للنموذج اللوجستي ثلاثي المعلمة، فقد تم استخدام برنامج بايلوج (bilog-mg 3) لتحليل

البيانات، حيث أظهرت نتائج التحليل وبالاعتماد على المؤشر الإحصائي مربع كاي (χ^2) الذي يعتمده البرنامج للحكم على الفقرات غير مطابقة للنموذج بوجود فقرتين (فقرة ١١، فقرة ٢٣) غير متطابقتين مع النموذج. تم استبعاد هاتين الفقرتين بحيث أصبح الاختبار بصورته النهائية مكون من (٥٣) فقرة وعدد الاستجابات (٥٦٢) لتشكل العينة النهائية لغايات التحليل من أجل الإجابة عن أسئلة الدراسة.

٣- حساب قيم المؤشر الإحصائي L_z

من أجل حساب قيم المؤشر الإحصائي L_z لأنماط الاستجابة الشاذة وفق النموذج اللوجستي ثلاثي المعلمة، تم تشكيل ملف أوامر ضمن برمجية R (انظر ملحق ١) يتضمن الإجراءات التالية:

١. استدعاء رزمة foreign المسئولة عن استيراد وتصدير الملفات بصيغها المختلفة، وكذلك استدعاء رزمة per fit للتمكن من حساب المؤشر الإحصائي Lz .

٢. قراءة ملف البيانات الخاص بالدراسة بصيغة spss، بالإضافة إلى إصدار الأمر المتعلق بحساب قيم المؤشر الإحصائي L_z ، والخطأ المعياري له.

٣. إصدار أمر تفعيل العشوائية بقيم مبدئية ١٢٣ من أجل إيجاد درجة القطع للمؤشر L_z لأنماط الاستجابة التي تتم بعد إعادة التعيين resampling لألف مرة، حيث يسمى هذا الإجراء bootstrapping مع إجراء ألف مستنسخة لكل عملية إعادة.

٤. إصدار الأوامر لعرض معالم الفقرات وقدرة الأفراد وقيم المؤشر الإحصائي L_z والأخطاء المعيارية له.

٥. إصدار أمر إنشاء الرسم البياني يظهر الخصائص العينية للمؤشر L_z بالإضافة إلى درجة القطع وفترة الثقة لها.

المعالجة الإحصائية

تم استخدام برمجية (bilog-mg 3)، وبرمجية (spss)، وبرمجية (noharm)، وبرمجية (per fit)، وبرمجية (R-language) لإجراء التحليلات الإحصائية من أجل الإجابة عن أسئلة الدراسة ، وتمثلت هذه الإجراءات في ما يلي:

١. استخدام طريقة الأرجحية القصوى الهامشية (Marginal Maximum Liklilhod) لإيجاد تقديرات معلمة القدرة للأفراد والخطأ المعياري في التقدير وفق النموذج اللوجستي ثلاثي المعلمة

٢. إيجاد قيم المؤشر الإحصائي Lz باستخدام برمجية per fit من خلال برمجية R_ language واستخراج القيمة الحرجة له ليتم في ضوئها تحديد أنماط الاستجابة الشاذة.

٣. التحليل الكمي من خلال التحليل العنقودي والتحليل النوعي لأنماط استجابات الأفراد التي عدت استجاباتهم شاذة اعتمادا على قيم المؤشر الإحصائي L_z لهم

٤. إعادة تقدير معلمة القدرة للأفراد والخطأ المعياري في التقدير بعد حذف أنماط الاستجابة الشاذة باستخدام طريقة الأرجحية الهامشية القصوى (MML).

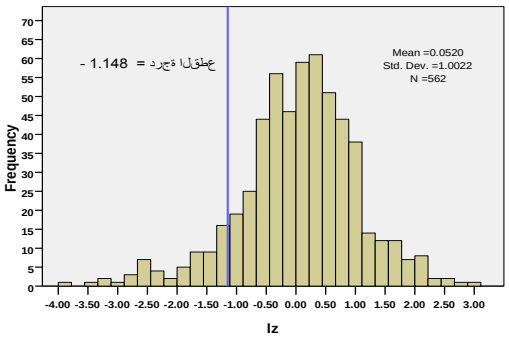
٥. إيجاد دالة معلومات الفقرة والاختبار قبل وبعد حذف أنماط الاستجابة الشاذة.

٦. استخدام اختبار t للبيانات المترابطة للوقوف على الفروق في دقة تقديرات معلمة القدرة للأفراد.

٧. حساب الكفاءة النسبية لدالة معلومات الاختبار قبل وبعد حذف أنماط الاستجابة الشاذة.

نتائج الدراسة ومناقشتها

للإجابة عن السؤال الأول فقد تم إيجاد قيم المؤشر الإحصائي L_z لأنماط الاستجابة وفق النموذج اللوجستي ثلاثي المعلمة، ودرجة القطع التي في ضوئها يتم الحكم على النمط الشاذ من خلال برمجية R، تمثلت هذه القيم بيانيا كما في الشكل ١



شكل ١: التمثيل البياني لقيم المؤشر L_z ودرجة القطع المحددة له

يتضح من شكل 1 بأن قيمة الوسط الحسابي لقيم المؤشر الإحصائي L_z قد بلغت (٠,٠٥٢) وقيمة الانحراف المعياري له بلغت (١,٠٠٢٢)، وجاءت هذه القيم قريبة من خصائص التوزيع الطبيعي ومتوافقة مع ما هو متوقع من التوزيع النظري للمؤشر L_z وهو التوزيع الطبيعي، كونه يعد مؤشرا معياريا يعتمد على دالة الأرجحية العظمى، ويتضح من الشكل كذلك بأن درجة القطع لهذا المؤشر قد بلغت (- ١,١٤٨) التي تعد بمثابة الحد الفاصل (الخط العمودي الازرق في الشكل ١) الذي في ضوئه يحدد نمط الاستجابة الشاذ للفرد إذا كانت قيمة المؤشر الإحصائي L_z للفرد أقل من هذه القيمة، وفي ضوء هذه القيمة فقد تبين وجود (٥٦) فردا كانت أنماط استجاباتهم شاذة شكلت ما نسبته (٩,٩٦%).

وبالنظر إلى قيمة الوسط الحسابي والانحراف المعياري يبدو واضحا اقترابه من خصائص التوزيع الطبيعي، واتفقت هذه النتيجة مع ما أشار إليه دراسجو وزملاؤه (Drasgow e al., 1985) عندما طوروا هذا المؤشر ليكون مؤشرا معياريا، مما يجعله مؤشرا فعالا في الكشف عن أنماط الاستجابة الشاذة وفق نماذج نظرية الاستجابة للفقرة ثنائية التدريج، حيث كشف عن نسبة عالية (٩.٩٦%)) من الأنماط الشاذة التي لا تتوافق مع النموذج اللوجستي ثلاثي المعلمة، ويعزو الباحث السبب في قوة المؤشر الإحصائي في الكشف عن أنماط الاستجابة الشاذة باعتباره مؤشرا يستند إلى دالة الأرجحية العظمى التي تعد من الطرق المشهورة في تقدير معالم الفقرات وقدرة الأفراد في سياق نظرية الاستجابة للفقرة.

ومن أجل معرفة الأسباب المحتملة وراء ظهور هذه النسبة من أنماط الاستجابة الشاذة في اختبار اوتيس – لينون للقدرة العقلية المستوى التمهيدي الصورة K، تم حصر استجابات الأفراد (٥٦) الذين اعتبرت أنماط استجاباتهم شاذة على هذا الاختبار كما جاءت من مخرجات نتائج حساب المؤشر Lz مرتبة حسب مستوى صعوبة الفقرة من الأسهل إلى الأصعب، وتم فصلها في ملف خاص من أجل تحليلها تحليلا نوعيا وكميا.

لقد استخدم التحليل العنقودي (cluster analysis) كأسلوب إحصائي في تجميع أنماط الاستجابة الشاذة من خلال طريقة التجميع K - Means إلى تجمعات متماثلة حسب القدرة، أفرزت نتائج التجميع عن أربعة تجمعات كانت متباعدة فيما بينها من حيث المسافة، في حين كانت المسافة أقل ما يمكن داخل التجمع الواحد، مما ساعدت هذه النتائج في التحليل النوعي لأنماط الاستجابة للوصول إلى الأسباب المحتملة وراء ظهورها.

وبالاستعانة بثلاثة خبراء من المختصين في القياس والتقويم التربوي تم دراسة استجابات الأفراد الذين اعتبرت أنماط استجاباتهم شاذة، حيث تم الإجماع على ترشيح عدة عوامل أدت إلي ظهورها، تلخصت في أربعة عوامل هي: التخمين (guessing)، التكاسل (sleeping)، الغش (cheating)، الإبداع الاستثنائي (exceptional creative). ومن الأمثلة على مثل تأثير هذه العوامل في إبراز الأنماط الشاذة لاستجابات الأفراد على اختبار اوتيس _لينون المقنن للبيئة الأردنية يتضح ذلك في جدول ٢.

يتضح من الأمثلة المعروضة في جدول ٢ تأثير سلوك التخمين العشوائي (guessing) في إبراز أنماط شاذة، حيث لوحظت الإجابة الصحيحة عن الفقرات الصعبة والإجابة الخاطئة عن الفقرات السهلة، وقد برز هذا النمط بوضوح في التجمع الرابع، وهم من الفئة ذوي القدرة المتدنية، حيث بلغت قيمة الوسط الحسابي لقدرتهم (- ٠,٩٢)، وهذا يتفق مع ما أشار إليه اسوناس وبل وديفيد (Iasonas et al., 2000) بأن التخمين يعد من العوامل الرئيسة في ظهور أنماط الاستجابة الشاذة عن فقرات الاختبار.

ويتضح كذلك عامل التكاسل (sleeping) في ظهور هذه الأنماط الشاذة، وذلك بعدم التكيف مع الاختبار في بدايته، مما يجعل الفرد لأن يجيب عن بعض الفقرات إجابة خاطئة دون مراجعتها، وهو يتفق مع ما أشار إليه ميجر (Meiger, 1996) إلى أن مثل هذا العامل قد يؤدي إلى بروز أنماط شاذة، وبرز مثل هذا العامل في التجمع الثاني الذي بلغ الوسط الحسابي لقدرتهم (٠,١٣١) الذين يعدون من ذوي القدرة المتوسطة، وربما يعزى السبب في ذلك إلى الفئة العمرية المستهدفة من الاختبار، وهم طلبة الصف الأول الأساسي، خاصة وأن مثل ذلك يعد بالنسبة لهم غير مألوف فهم معتادون على الاختبارات التحصيلية من إعداد المعلمين الذين يدرسونهم. ومما هو جدير بالذكر كذلك بأن مثل هذه الفئة العمرية ليس من السهل التزامهم بتعليمات الاختبار، وخصوصا فيما يتعلق بالغش من خلال النظر والحديث مع المجاورين أثناء الجلسة الاختيارية، وهذا أدى لأن تظهر أنماط شاذة في الاستجابات من خلال نقل إجابات بعض الفقرات عن الأفراد المجاورين، مثل هذا السلوك أكده غالبية المعلمين الذين يدرسون مثل هذه الفئة العمرية من خلال الالتقاء بهم أثناء تطبيق الاختبار، وقد برز مثل هذا النمط في التجمع الثالث.

جدول ٢

والعوامل المحتملة وراء ظهورها_أمثلة من الأنماط التي اعتبرت شاذة اعتمادا على درجة القطع لمؤشر

العامل	نمط الاستجابة على الفقرات (الاسهل ← الاصعب)	رقم الفرد
التخمين	١١٠٠٠٠١١١٠١٠١١٠٠١٠٠٠١٠٠١١٠٠١٠١٠١٠١١٠١٠١٠٠١١٠٠٠٠٠١٠١١	٣٦٩
	٠١١١٠٠٠٠١٠١٠١١٠٠١٠٠٠٠٠٠١١٠١٠٠٠١١٠٠٠١١٠١١١١٠٠١٠١٠١١١٠٠	١٠٤
	٠١١١٠٠٠٠٠١٠١١٠٠١٠٠٠٠٠٠١١٠١٠٠٠١١٠٠٠١١٠١١١١٠٠١٠١٠١١١٠١	٣٠٨
	١١٠٠٠١٠٠٠٠١٠١٠٠١١١١١٠٠٠٠٠٠٠٠٠٠٠٠١٠١١٠١٠١٠١١١١٠٠٠	٤٥٤
التكاسل	٠٠٠٠٠١٠٠١٠١١١١١١٠١٠١٠١١١١١١٠١١١١١١٠١١٠٠٠١١١٠١١١٠١٠١	١٢٥
	٠٠٠٠٠١١١١١٠١١١١١١١١١١١١٠١١١١١١١١١١١١٠١٠١٠١٠٠١١٠١١٠١	٤٣٤
	٠٠٠٠٠٠٠١١٠١١١١١١١١١١١١٠١١١١١١١٠١١١١١١٠٠١٠١١١١٠٠١٠١٠٠	١٣١
	٠٠٠٠١٠١٠١١١٠٠٠١١٠١٠٠١٠١١١٠١١٠٠١١١١١١١١١١١١٠٠١١١١١١٠١	٢٣٢
الغش	١٠١١٠١٠٠٠١٠١٠١١٠١١٠١١٠٠١٠١٠١٠٠١١١١١٠١٠١٠٠٠١١٠٠٠٠١٠١٠١	٥
	١١٠١٠٠١٠١١٠١١١١٠١٠١٠٠١١٠١١٠١٠١٠١١٠١٠١٠٠٠١١٠٠١٠٠١١٠١١١	٣٨٧
	١٠٠٠٠١٠٠١٠١٠٠١١١١٠٠٠١٠٠٠٠٠١٠١١١١١١١٠١٠١١٠١١١١١١٠٠٠	٣٥٠
	٠٠٠١٠١٠٠١٠١١١١٠٠٠٠١٠١١١١١١١٠١٠١٠٠١٠١٠١٠١١١١١١١١٠٠٠	١٩٥
الإبداع الاستثنائي	٠٠٠٠١١٠١١١١١١١١١٠١٠١١١١٠١٠١١١٠١١١١١١٠١١١١١٠١١٠١١١١١١١١	٢٩
	٠١٠١٠١١١٠٠١١١١١٠١١١١١١١١١١١١٠١١١١١١٠١١١١١٠١١١١١١١١٠١١	٣٧٣
	٠٠٠١١١٠١١١١١١١١١١١١١١١١١١١١٠١٠١١١١١٠١١١١١١١١١١١١١١١٠١١	٣٧٤
	٠٠١٠٠١٠٠١٠١١١١١١١١٠١٠١١١١١١١١١١١١١١١٠١١٠١٠١١١٠١١١١١٠١	١٢٥

ومن العوامل التي أدت كذلك إلى ظهور الأنماط الشاذة عامل متعلق بالخصائص التربوية للفرد، عامل الإبداع الاستثنائي، حيث يقوم الفرد بتفسير الفقرات السهلة بطريقة إبداعية مما يجعله لأن يجيب عنها إجابة خاطئة، وهذا بدوره يؤدي إلى ظهور أنماط شاذة، ويعد هذا العامل مرتبطا بالخصائص التربوية للفرد، وبرز أثر هذا العامل لدى فئة من أفراد القدرة العالية الذين تجمعت أنماط استجاباتهم الشاذة في التجمع الرابع، حيث بلغ الوسط الحسابي لقدرتهم (١,١٨٣).

إن تلك العوامل التي تم الإجماع عليها من الباحث والخبراء المختصين لدي تتبع أنماط الاستجابات الشاذة التي افرزها المؤشر L_z قد جاءت منسجمة مع نتائج الدراسات (Birenbaum, 1986 ; Iasonas et al., 2000; Karabatsos, 2003 ; Meijier,1996 ; Schmitt & Croker, 1984) التي اهتمت بالبحث عن العوامل والأسباب التي تحدث مثل تلك الأنماط، وهذه العوامل التي كانت الأبرز في حدوث الأنماط الشاذة في استجابات الأفراد على اختبار اوتيس _ لينون للقدرة العقلية المستوى التمهيدي الصورة K، لكونها ارتبطت بالخصائص الشخصية والتربوية لمثل هذه الفئة العمرية التي طبق عليها الاختبار،حيث لوحظ وجود نزعة لدى أفراد هذه الفئة العمرية أثناء تطبيق الاختبار لنقل الإجابة عن الأقران أو اللجوء إلى التخمين في حال عدم معرفة الإجابة الصحيحة، وكذلك لوحظ عملية التكاسل لدي بعض الأفراد أثناء بداية تطبيق الاختبار.

وللإجابة عن السؤال الثاني للدراسة فقد تم تقدير معالم الفقرات (الصعوبة، التمييز، التخمين) لفقرات الاختبار المكون من (٥٣) والأخطاء المعيارية لها بالإضافة إلى قدرة الأفراد، وفق النموذج اللوجستي ثلاثي المعلمة باستخدام طريقة الأرجحية القصوى الهامشية (MML)، بعد حذف أنماط الاستجابة الشاذة.

ولغاية الوقوف على الفروق في دقة تقديرات معلمة القدرة للأفراد قبل وبعد حذف أنماط الاستجابة الشاذة، فقد تم إيجاد الإحصاءات الوصفية لقيم الخطأ المعياري (standard error) لكل قدرة من قدرات الأفراد قبل وبعد حذف الأنماط الشاذة، والذي يعد أحد أهم المؤشرات التي يستند إليها للحكم على دقة التقدير عند استخدام أحد نماذج الاستجابة للفقرة، ويظهر جدول ٣ الإحصاءات الوصفية للخطأ المعياري في تقدير معلمة القدرة للفرد قبل وبعد حذف أنماط الاستجابات الشاذة.

جدول ٣
الإحصاءات الوصفية للخطأ المعياري في تقدير معلمة القدرة للفرد قبل وبعد حذف أنماط الاستجابات الشاذة

المجموعة	الوسط الحسابي	الانحراف المعياري	القيمة القصوى	القيمة الصغرى	المدى
قبل الحذف	٠,٣٤٠	٠,٢٤٦	١,٨٢٢	٠,٢٣٣	١,٥٨٩
بعد الحذف	٠,٢٩٢	٠,٢٢٨	١,٥٢٣	٠,١٦٨	١,٣٥٥

يتضح من النتائج الواردة في جدول ٣ بأن هناك تباينا ملحوظا في الأوساط الحسابية لدقة تقديرات معلمة القدرة للفرد (الخطأ المعياري) قبل وبعد حذف أنماط الاستجابة الشاذة، حيث ازدادت دقة تقدير معلمة القدرة للفرد بعد حذف أنماط الاستجابة الشاذة وذلك من خلال النظر إلى الوسط الحسابي الذي تناقص لتصبح قيمته (٠,٢٩٢) بعد الحذف بدلا من (٠,٣٤٠) قبل الحذف. وللوقوف على دلالة الفروق بين الأوساط الحسابية لدقة تقديرات معلمة القدرة (الخطأ المعياري) للفرد قبل وبعد حذف أنماط الاستجابات الشاذة، فقد تم استخدام اختبار t للبيانات المترابطة، ويظهر جدول ٤ نتائج اختبار t للبيانات المترابطة.

جدول ٤

الأوساط الحسابية والانحرافات المعيارية لدقة تقدير معلمة القدرة للفرد (الخطأ المعياري) وقيمة t والدلالة الإحصائية

المجموعة	العدد	الوسط الحسابي	الانحراف المعياري	درجات الحرية	قيمة t	الدلالة الإحصائية
قبل الحذف	٥٠٦	٠,٣٤٠	٠,٢٤٦	٥٠٥	١٨,٨٢٨	٠,٠٠٠*
بعد الحذف	٥٠٦	٠,٢٩٢	٠,٢٢٨			

* دال إحصائيا عند مستوى الدلالة ٠,٠٥ = α

يتضح من النتائج الواردة في جدول ٤ بوجود فروق ذات دلالة إحصائية بين الأوساط الحسابية في دقة تقدير معلمة القدرة (الخطأ المعياري)، وقد جاءت دقة التقدير لصالح حذف أنماط الاستجابة الشاذة، حيث ازدادت دقة التقدير من خلال تناقص قيم الخطأ المعياري لمعلمة القدرة للفرد المقدرة بعد حذف الأنماط الشاذة. ولمعرفة مقدار حجم تأثير أنماط الاستجابة الشاذة على دقة تقديرات معلمة القدرة للفرد، فقد تم إيجاد أحد مؤشرات الدلالة العملية وهو مربع ايتا ($\eta 2$) وفق ما أشار إليه كوهن (Cohen, 1988) لتقييم دلالة الفرق الإحصائي المحسوب في اختبار دلالة الفرق، حيث بلغت قيمة مربع إيتا (٤١,٢%) والتي تشير بأن حذف أنماط الاستجابة الشاذة من معايرة قدرة الأفراد قد فسر في تباين دقة التقدير لمعلمة القدرة (٤١,٢%) وهي نسبة عالية.

وبالنظر إلى النتائج الواردة في الجدول ٤ يتضح بأن هناك تحسنا ملحوظا في دقة تقديرات معلمة القدرة للفرد بعد حذف أنماط الاستجابة الشاذة وإعادة معايرة قدرة الأفراد وفق النموذج اللوجستي ثلاثي المعلمة باستخدام طريقة الأرجحية القصوى (MML) حيث قلت قيمة الوسط للخطأ المعياري لقدرة أفراد عينة الدراسة، وجاءت هذه النتيجة متوافقة مع نتيجة دراسة نيرنغ (Nering, 1998) الذي أشار إلى أن حذف الأنماط الشاذة تحسن من دقة تقدير قدرة الأفراد.

إن التحسن في دقة تقديرات معلمة القدرة للأفراد عند حذف أنماط الاستجابات الشاذة ربما يعزى لكون الأفراد الذين كانت استجاباتهم شاذة بناء على مؤشر Lz كان لديهم ميلا لتخمين الإجابة الصحيحة، وكما هو معلوم بأن التخمين العشوائي يعد أحد أخطاء القياس التي تؤثر في دقة تقدير معلمة القدرة للفرد عند استخدام أحد نماذج نظرية الاستجابة للفقرة، بالإضافة إلى سلوكي الغش والتكاسل اللذين اظهرا أنماطا شاذة انعكس تأثيرها على تقدير معالم الفقرات، وخصائصها، وقدرة الأفراد، مما جعل من حذف هذه الأنماط لأن يحسن من دقة تقدير معلمة القدرة. ومما يؤكد كذلك تحسن دقة تقدير معلمة القدرة للفرد بعد حذف أنماط الاستجابة الشاذة بأن الوسط الحسابي والانحراف المعياري لقدرة الأفراد قد اقترب من التوزيع الطبيعي، وجاء هذا متوافقا مع ما يفترضه النموذج اللوجستي ثلاثي المعلمة حيث بلغت قيمة الوسط الحسابي والانحراف المعياري لقدرة الأفراد على التوالي (٠,٠١٢، ١,٠١٨)، في حين كانت قيمة الوسط الحسابي والانحراف المعياري لقدرة الأفراد قبل الحذف (٠,٠٦٧، ٠,٩٦٨) على التوالي.

ومن الجدير بالذكر إنه لتقويم حسن المطابقة للبيانات قبل وبعد حذف أنماط الاستجابة الشاذة مع النموذج اللوجستي ثلاثي المعلمة، تم استخدام المؤشر الإحصائي G2 الذي يستخدم في تقويم التحسن النسبي في نسبة التباين التي تعزى لمقارنة نموذج مع نموذج آخر اعتمادا على

محك أكيك (AIC)، وهو يماثل R2 في مقارنة نماذج الانحدار (De Ayala, 2009)، حيث تم استخدامه في تقويم التحسن النسبي في درجة المطابقة قبل وبعد حذف أنماط الاستجابة الشاذة، وبلغت قيمته (٠,١٣٢)، مما يعني بأن حذف الأنماط الشاذة من عملية المعايرة قد حسن من درجة المطابقة مع النموذج المستخدم ما نسبته (١٣,٢%)، وبالتالي انعكست هذه النسبة من التحسن في دقة التقديرات لمعلمة القدرة. وتؤكد هذه النتيجة أهمية الاهتمام بأنماط الاستجابة الشاذة خصوصا في المواقف التي تستخدم فيها الاختبارات لغايات اتخاذ القرارات بحق الأفراد لما لها من آثار في دقة تقدير القدرة للفرد.

أما فيما يتعلق بدالة معلومات الاختبار فقد تم تقديرها عند كل مستوى من مستويات القدرة للأفراد وفق النموذج اللوجستي ثلاثي المعلمة قبل وبعد حذف أنماط الاستجابة الشاذة، وجدول ٥ يوضح قيم دالة معلومات الاختبار عند مستويات مختارة للقدرة من (- ٤) إلى (٤) بطول فترة (٠,٢).

يتضح من النتائج الواردة في جدول ٥ تحسن قيم دالة معلومات الاختبار عند كل مستوى من مستويات القدرة المختارة بعد حذف الأفراد الذين اعتبرت أنماط استجاباتهم عن فقرات الاختبار شاذة عند معايرة قدرة الأفراد وفق النموذج اللوجستي ثلاثي المعلمة، وللوقوف بشكل كلي على الفروق في دالة المعلومات التي يقدمها الاختبار عند كل مستوى من مستويات القدرة تم تمثيل دالة المعلومات للاختبار بيانيا قبل وبعد حذف أنماط الاستجابة الشاذة، والشكل ٢ يوضح هذه الدالة.

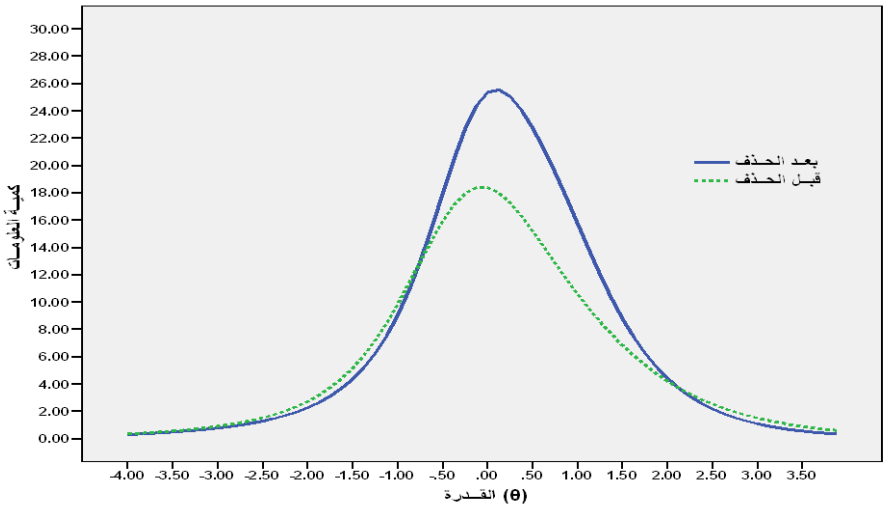
جدول ٥

قيم دالة معلومات الاختبار عند مستويات مختارة للقدرة (θ) قبل وبعد حذف أنماط الاستجابة الشاذة

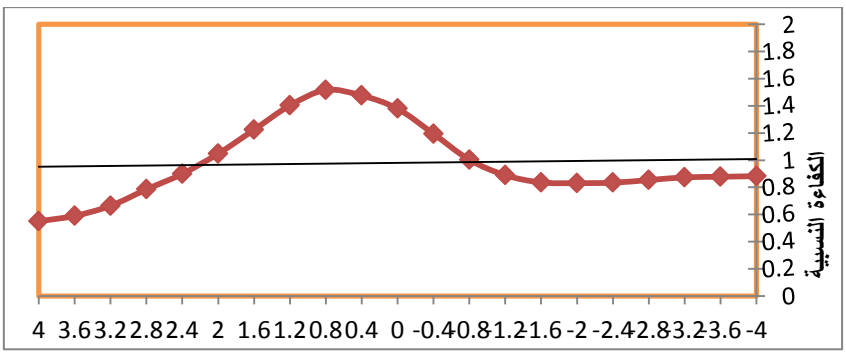
القدرة	دالة معلومات الاختبار		القدرة	دالة معلومات الاختبار	
(θ)	قبل الحذف	بعد الحذف	(θ)	قبل الحذف	بعد الحذف
– ٤,٠	٠,٣٤	٠,٣٠	٠,٢	١٧,٢٧	٢٥,٠٦
– ٣,٨	٠,٤٣	٠,٣٨	٠,٤	١٦,٣١	٢٤,٠٧
– ٣,٦	٠,٤٩	٠,٤٣	٠,٦	١٤,٠٤	٢١,٢٠
– ٣,٢	٠,٧٩	٠,٦٩	١,٠	١٠,٥٩	١٥,٧٦
– ٢,٨	١,١٦	٠,٩٩	١,٤	٧,٦٦	١٠,٣١
– ٢,٦	١,٣٣	١,١٢	١,٦	٦,٠٧	٧,٤٣
– ٢,٤	١,٧٥	١,٤٦	١,٨	٥,٣٩	٦,٢٦
– ٢,٢	٢,٣٤	١,٩٤	٢,٠	٤,٢١	٤,٤١
– ٢,٠	٢,٧٢	٢,٢٦	٢,٢	٣,٢٧	٣,٠٩
– ١.٨	٣.٧٣	٣.١٠	٢.٤	٢.٨٨	٢.٥٩
– ١,٦	٤,٣٩	٣,٦٧	٢,٦	٢,٢٢	١,٨٢
– ١,٤	٦,٠٩	٥,٢٠	٢,٨	١,٩٥	١,٥٣
– ١,٢	٨,٤٢	٧,٤٩	٣,٠	١,٥٠	١,٠٨
– ١.٠	٩.٨٢	٩.٠١	٣.٢	١.١٦	٠.٧٧
– ٠,٨	١٢,٩٢	١٢,٩٦	٣,٤	١,٠١	٠,٦٥
– ٠,٤	١٧,١١	٢٠,٤٣	٣,٨	٠,٦٩	٠,٣٩
٠,٠	١٨,٣٨	٢٥,٣٤			

يتضح من الشكل ٢ بأن دالة معلومات الاختبار كانت بشكل عام أكبر عندما تم حذف أنماط الاستجابة الشاذة عند جميع مستويات القدرة للأفراد باستثناء المستويات العالية والمنخفضة حيث كانت أعلى بفروق بسيطة عندما تم الاحتفاظ بأنماط الاستجابة الشاذة. ويتضح كذلك من الشكل ٢ بأن أكبر قدر من المعلومات كان يقدمها الاختبار بعد حذف أنماط الاستجابة الشاذة ضمن مستويات القدرة من (- ٠,٨) إلى (٢,٠)، حيث بلغت أعلى قيمة لدالة معلومات الاختبار والبالغة (٢٥,٣٤) عندما كان مستوى القدرة للفرد يساوي (٠,٠٠) في حين كانت أعلى قيمة لدالة معلومات الاختبار قبل حذف أنماط الاستجابة الشاذة تساوي (١٨,٣٨) عندما كان مستوى القدرة للفرد يساوي (٠,٠٠). وبالنظر إلى نتائج جدول ٥ والشكل ٢ يتضح بأن هناك فرقا ملاحظا وبشكل كبير بين قيم دالة معلومات الاختبار ضمن مستوى القدرة من (- ٠,٨) إلى (٢,٠) قبل وبعد حذف أنماط الاستجابة الشاذة، حيث كانت هذه القيم أعلى وبشكل ملحوظ عندما تم حذف أنماط الاستجابة الشاذة من التحليل.

وللوقوف بدرجة أكثر دقة في مقارنة كمية المعلومات التي يقدمها الاختبار قبل وبعد حذف أنماط الاستجابة الشاذة، فقد تم التمثيل البياني للكفاءة النسبية (Relative Efficiency) وذلك بقسمة المعلومات الكلية التي يقدمها الاختبار بعد حذف الأنماط الشاذة على المعلومات الكلية للاختبار قبل حذفها عند مستويات مختارة من القدرة تتراوح من (- ٤) إلى (٤) بطول فترة (٠,٤)، والشكل (٣) يوضح ذلك.



شكل ٢: دالة معلومات الاختبار قبل وبعد حذف أنماط الاستجابة الشاذة



شكل ٣: التمثيل البياني للكفاءة النسبية للاختبار بعد وقبل حذف أنماط الاستجابة الشاذة

يتضح من شكل ٣ فاعلية الاختبار بتقديم معلومات اكبر عند مستويات القدرة ضمن الفترة [- ٠,٨، ٢,٠] عندما تم استبعاد أنماط الاستجابة الشاذة من المعايرة، حيث الفاعلية النسبية أعلي من واحد، مما يعني بأن الاختبار أكثر دقة في تقدير قدرة الأفراد ضمن هذه الفترة، في حين عندما تم الاحتفاظ بها قدم الاختبار معلومات اكبر عند المستويات العالية والمنخفضة للقدرة. ويعزى ذلك إلى أن الفقرات من حيث الصعوبة (انظر ملحق ٢) كانت أسهل عندما تم الاحتفاظ بأنماط الاستجابة الشاذة حيث بلغ الوسط الحسابي لمعلمة الصعوبة (- ٠,٢٩١)؛ مما جعلها اقرب إلى موقع الفرد في المستويات المنخفضة للقدرة فأعطت معلومات أكثر عند هذه المستويات، وكانت سهلة للمستويات العالية، وعند حذف الأنماط الشاذة أعطت معلومات اقل في المستويات نفسها؛ وذلك لأن الوسط الحسابي لصعوبة الفقرات ازداد وبلغت قيمته (- ٠,٢٥٥)، فأصبحت معلمة الصعوبة أبعد عن قدرة الفرد؛ مما جعلها تقدم معلومات أقل لأن الأساس النظري لإيجاد دالة معلومات الفقرة يعتمد على خصائص الفقرة (التمييز، الصعوبة والتخمين) (Hulin, Drasgow & Parsons, 1983).

إن التحسن الملحوظ في قيم دالة معلومات الاختبار المقدرة عند كل مستوى من مستويات القدرة للأفراد بعد حذف أنماط الاستجابة الشاذة قد جاءت متوافقة مع أشار إليه غرين (Green, 1996) بأن حذف أنماط الاستجابة الشاذة يحسن من دلالات الثبات للاختبارلكون دالة معلومات الاختبار في سياق نظرية الاستجابة للفقرة هي المفهوم المناظر لثبات الاختبار، وهو ما أكده كل من سيجمت وكوريتينا وويتني (Schimitt, Cortina & Whitny, 1993) بأن وجود أنماط الاستجابة الشاذة يؤثر في قضايا ثبات الاختبارات على الرغم من أن ميجر (Meijer, 1997) أشار بأن صدق الاختبار المرتبط بمحك لا يتأثر بوجود أنماط الاستجابة الشاذة إلا أنه حذر من وجودها.

إن الأثر الواضح لأنماط الاستجابة الشاذة على قيم دالة معلومات الاختبار يبدو واضحا من نتائج الجدول (٥) والشكلين (٢،٣) وجاء هذا الأثر متوافقا مع نتائج الجدول (٤)، وذلك لارتباط دالة المعلومات بالخطأ المعياري لتقدير قدرة الفرد، حيث يعزى سبب تحسن قيم دالة معلومات الاختبار بعد حذف أنماط الاستجابة الشاذة إلى تناقص قيم الخطأ المعياري لكل مستوى من مستويات القدرة المقدرة للأفراد، حيث تتناسب قيمة دالة المعلومات عند أي مستوى من مستويات القدرة تناسبا عكسيا مع الجذر التربيعي للخطأ المعياري المقدر لذلك المستوى من القدرة، وبالتالي تناقص قيم الأخطاء المعيارية لمستويات القدرة أنعكس بشكل إيجابي على قيم دالة معلومات الاختبار.

وفي السياق نفسه، إن أحد التفسيرات التي تكمن كذلك وراء ظهور هذه النتيجة هو زيادة القدرة التمييزية لفقرات الاختبار (انظر ملحق 2) بعد حذف أنماط الاستجابة الشاذة من معايرة قدرة الأفراد، حيث بلغت قيمة الوسط الحسابي لمعلمة التمييز (١,٦٤٥) بعد حذف أنماط الاستجابة الشاذة، في حين كانت قيمة الوسط لمعلمة التمييز نفسها (١,٣١٩) قبل حذفها، فكلما زادت قيمة معلمة التمييز للفقرة زاد مقدار المعلومات التي تسهم فيه الفقرة حول معلمة الصعوبة (De Aayla, 2009; Baker, 2000). وبذلك، تقدم منحنيات

المعلومات للفقرات قدرا كبيرا من المعلومات في مدى ضيق لقيم القدرة، حيث القيم المنخفضة لمعلمة التمييز للفقرات تجعل منحنى دالة معلومات الاختبار أكثر انبساطا وانتشارا كما ظهر في الشكل ٢ عندما تم الاحتفاظ بأنماط الاستجابة الشاذة في عملية المعايرة.

الاستنتاجات والتوصيات

إن النتائج التي أظهرتها هذه الدراسة تبرز أهمية وضرورة الاهتمام بدراسة أنماط الاستجابة الشاذة في التطبيقات العملية للاختبارات والمقاييس وانعكاساتها على دقة تقدير قدرة الفرد، وتنبع هذه الأهمية من تعدد أغراض استخدام الاختبارات خصوصا عندما يتعلق الأمر باتخاذ قرارات متعلقة بالفرد نفسه. وفي ضوء النتائج التي توصلت إليها هذه الدراسة يوصي الباحث بما يلي:

١. إجراء دراسات مماثلة لمعرفة العوامل والأسباب التي قد تؤدي إلى ظهور الأنماط الشاذة لاستجابات المفحوصين على الاختبارات والمقاييس التربوية .

٢. إجراء دراسات مماثلة باستخدام مؤشرات إحصائية أخرى للكشف عن أنماط الاستجابة الشاذة في اختبارات اوتيس _لينون للقدرة العقلية.

المراجع
References

بني عطا، زايد صالح وحجازي، تغريد عبدالرحمن. (٢٠١١). تقنين اختبار اوتيس – لينون للقدرة العقلية "المستوى التمهيدي الثاني الصورة k " في البيئة الأردنية. **مجلة كلية التربية وعلم النفس، ٣٥**، ٢٦٧ – ٣٠٤.

حمادنه، اياد محمد.(٢٠١٥). تقصي فاعلية المؤشر (L_z^{*new}) في الكشف عن أنماط استجابة المفحوص غير المطابقة وفق نظرية استجابة الفقرة. **مجلة العلوم التربوية والنفسية ١٦** (٣)، ٥٦٥ – ٥٩٣.

Baker, F.B.(2000). *The basics of item response theory(2nd ed)*. College Park, MD: ERIC Clearing House on Assessment and Evaluation

Belov, D. I., & Armstrong, R. D. (2010). Automatic detection of answer copying via Kullback-Leibler divergence and K-Index. *Applied Psychological Measurement, 34*(6), 379–392.

Birenbaum, M. (1986). Effect of dissimulation motivation and anxiety on re-sponse pattern appropriateness measures. *Applied Psychological Measure-ment, 10*(2), 167-174.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences (2nd ed.).* Hillsdale, NJ: Erlbaum

Crocker, L., & Algina, J. (1986). *Introduction to Classical and Modern Test Theory*. N. Y.:Harcourt Jovanovich.

De Ayala, R.J. (2009). *The Theory and Practice of Item Response Theory*. NY: The Guildford Press

De laTorre, J. & Deng, W. (2008). Improving person-fit assessment by correcting the ability estimate and it reference distribution. *Journal of Educational Measurement, 45* (2), 159-177

Drasgow, F., Levine, M. V., & McLaughlin, M. E. (1987). Detecting inappropriate test scores with optimal and practical appropriateness indices. *Applied Psychological Measurement, 11* (1), 59-79.

Drasgow, F., Levine, M. V., & Williams, E. A. (1985). Appropriateness Measurement with Polychotomous Item Response Models and Standardized Indices. *British Journal of Mathematical and Statistical Psychology, 38* (1), 67-86.

Green, K. E. (1996). *The use of person fit statistics in mail surveys*. Paper presented to the annual meeting of the American Educational Research Association, New York, NY

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

Hulin, C. L.,Drasgow, F. & Parsons, C.K.(1983). *Item response theory*. Homewood Il: Dow Jones-Irwin.

Iasonas, C ., Bill, B . & David, W .(2000). The consistency of examinee misfit across tests on the same subject and across subject: the case of the KS2 mathematics and science National Curriculum tests in England. *Retrieved from http// www.man.edu.uk*

Johanson, S., Dulany, C.,& Banks, k.(2000). Measurement Error. For Good Measure. Retrieved from *http://www.eric.ed.gov/ ED445089*

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

Kim, S ., Cohen, A., & Lin, Y.(3005).LDID: *A computer program for local dependence indices for dichotomous items.* Version 1.0.

Korir, D. K.(2014). An assessment of the effects of item difficulty and examine abilty on the effect of LZ appropriateness index. *Journal of Education and Practice, 5*, 11 – 19.

Lamprianou, I .(2013). The tendency of individuals to responded to high-stakes tests in idiosyncratic way. *Journal of Applied Measurement, 14* (3), 299 – 317.

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

Lopez, A ., & Montesinos, H .(2005). Fittin Rasch model using appropriateness measure statistics. *The Spanish Journal of Psychology, 8* (1), 100 – 110.

Li, M. N. & Olejnik, S. (1997). The power of Rasch person-fit statistics in detecting unusual response patterns. *Applied Psychological Measurement, 21* (3), 215-231.

McDonald, R. P. (1999). *Test theory: A unified treatment*. Mahwah, NJ: Lawrence Erlbaum Associates.

Meijer, R. R. (1996). Person-fit research: an introduction. *Applied Measurement in Education, 9* (1), 3-8.

Meijer, R. R. (1997). Person fit and criterion-related validity: An extension of the Schmitt,Cortina, and Whitney study. *Applied Psychological Measurement, 21* (2), 99 - 113.

Meijer, R.R., Molenaar, I. W. & Sijtsma, K. (1994).Influence of test and person characteristics on nonparametric appropriateness measurement. *Applied Psychological Measurement, 18* (2), 111-120.

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

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

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.

Nering, M. L.(1998). *The influence of nonormal – fitting examine in estimating person parameter.* Paper presented at Annual Meeting of the American Educational Research Association conference. San Diego, CA.

Reckase, M.D.(1997). The past and future of multidimensional item response theory. *Applied Psychological Measurement,21* (1),25-36.

Reise, S. P., & Due, A. M. (1991). The influence of test characteristics on the detection of aberrant response

patterns. *Applied Psychological Measurement, 15* (3), 217-226

Schmitt, N., Cortina, J., & Whitney, D. (1993). Appropriateness fit and criterion-related validity. *Applied Psychological Measurement, 17* (2), 143-150

Schmitt, A. P. and L. Crocker (1984), *The relationship between test anxiety and person fit measures*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, April 23-2

Seo, D. G., & Weiss, D. J. (2013). Lz person-fit index to identify misfit students with achievement test data. *Educational and Psychological Measurement, 73* (6), 994- 1016.

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

Sotaridona, L ., Pornal, J . & Vallejo, A .(2003).Some application of Item Response Theory to testing. *The Philippine Statistician, 52* (1-4),81-92..

Wright, B .D., & Stone, M .H.(1999). *Measurement essentials*. Wilmington, Delaware.

Yen, W. M. (1993). Scaling performance assessments: Strategies for managing local item dependence. *Journal of Educational Measurement, 30* (3), 187-213.