التوافق بين أداء مؤشرات مطابقة الفرد اللامعلمية G_P ومؤشر U_{3P} باختلاف عدد فئات التدريج والمراحل العمرية

هديل محمد بني حمد(*)
أ.د زايد بني عطا(**)

تاريخ الإرسال
2023/8/27
تاريخ القبول
2023/9/23

الملخّص

هدفت هذه الدراسة إلى البحث في أداء المؤشرين U3p , Gp ومدى التوافق بينهما وفق نموذج الاستجابة المتدرجة في ضوء المتغيرات فئات التدريج والمراحل العمرية . ولتحقيق هدف الدراسة تم استخدام مقياس مكون من (22) فقرة ،لقياس قلق القراءة في اللغة الانجليزية ،وقد تم التحقق من دلالات صدق وثبات المقياس، إضافة إلى التحقق من مطابقة المقياس لافتراضات نموذج الاستجابة المتدرجة الذي تم تبنيه في الدراسة الحالية، تكون مجتمع الدراسة من جميع طلاب وطالبات الصف السابع والأول ثانوي في لواء الكورة –إربد والبالغ عددهم (5523) للعام الدراسي 2021–2022، في حين تكونت عينة الدراسة من (1537) طالبا وطالبة، تم اختيارهم بشكل طبقي ثم عنقودي ، كون أنّ الشعبة الدراسية هي وحدة الاختيار. أظهرت نتائج الدراسة عدم وجدو اختلاف في التوافق بين مؤشري المطابقة U3p , Gp باختلاف عدد فئات التدريج، وكذلك عدم وجود اختلاف في التوافق بين مؤشري المطابقة باختلاف الصف .

الكلمات المفتاحية : مطابقة الفرد ، المؤشرات اللامعلمية Gp, U3p ، نموذج الاستجابة المتدرجة (GRM).

(*) جامعة اليرموك
(**) جامعة اليرموك

Compatibility Between the Performance of Nonparametric Person-Fit Statistics Indicators G_P and U_{3P} According to the Graded Response Model and Age Stages

Abstract

This study aimed to investigate the performance of two indicators of nonparametric person-fit, namely Gp and U3p, and to determine their compatibility within the graded response model across two variables: gradation levels and age stages. The study used a scale to measure reading anxiety in English, and the validity, reliability, and conformity of the scale's indicators to the assumptions of the progressive response model were verified. The results showed no difference in the scale's compatibility between Gp and U3p based on the number of scoring categories, nor was there a difference in compatibility between the two indicators across different age groups.

Keywords: Prson-fit Statistics (PFSs), graded response model (GRM), nonparametric indicators Gp, U3p

خلفية الدراسة

جميع المقاييس التربوية والنفسية تتطلب أدوات لجمع البيانات وتبويبها في إطار قابل للقياس، فلا بد من أن تتمتع بدرجة عالية من الصدق والدقة حتى يكون لديها القدرة على تحديد خصائص المستجيبين باختلاف الهدف من المقياس أو الاختبار الذي تم إعداده لقياس و تقويم ظاهرة ما. ومع ذلك، لا ينبغي الاكتفاء بالتحقق من صدق هذه الأدوات من حيث ملاءمتها وقدرتها على تحقيق الاهداف المرجوة منها، بل يجب التأكد من صحة البيانات التي تم جمعها من خلال تلك الأدوات.

أدى التطور المستمر في علم القياس والتقويم إلى التحقق من صدق الاستجابة على أدوات القياس، والتحقق من صدق وثبات الاختبارات والمقاييس النفسية والتربوية، حيث تظهر أهمية صدق الاستجابة على المقاييس عند اتخاذ قرارات مهمة على مستوى الفرد، مثل قبول فرد في برنامج أو استثناء فرد من القبول في هذا البرنامج (علام، 1986). نتيجة لذلك، ظهرت بعض الاتجاهات الحديثة في مجال القياس والتقويم التربوي، وابرز هذه الاتجاهات نظرية استجابة الفقرة Item Response Theory (IRT) أو ما يطلق عليها نظرية السمات الكامنة Latent Trait Theory، إذ تفترض أنه يمكن التنبؤ بأداء الافراد، أو يمكن تفسير أدائهم في اختبار نفسي أو معرفي معين في ضوء خاصية مميزة لهذا الاداء تُسمى القدرات أو السمات Traits ،ويصعب ملاحظة هذه السمة بشكل مباشر. (Hambleton el al., 1991). وتقوم (IRT) كغيرها من النظريات على مجموعة من الافتراضات، هي: احادية البعد(Unidimensionality)، والاستقلال الموضعي (Local Independence)، ومنحنى خصائص الفقرة (Characteristic Curve Item)، وعامل السرعة(Seededness)(Boeck & Wilson, 2004). كما انبثق عن هذه النظرية مجموعة من النماذج التي تم استخدامها في بناء المقاييس والاختبارات التي يمكن من خلالها الحصول على مؤشرات إحصائية للفقرة دون الاعتماد على خصائص المستجيبين ،وتقديرات للمستجيبين من دون الاعتماد على صعوبة فقرات المقياس (Crocker & Algina, 1986).

وقد صنف هامبلتون وسوامنثان (Hambleton & Swaminathan, 1985) نماذج نظرية استجابة الفقرة IRT للفقرة أحادية البعد بحسب استجابة المفحوص إلى ثلاثة نماذج :

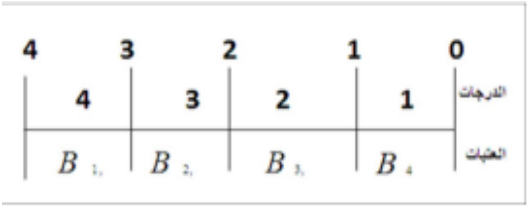
أولا: النماذج ثنائية الاستجابة (Dichotomous): نماذج تستخدم مع الفقرات التي يجيب عنها المفحوص بشكل ثنائي (صح أو خطأ) أو الاختيار بين (نعم أو لا)، فيحصل المفحوص على درجة اما 0 أو 1 .

ثانيا : النماذج متعددة الاستجابة (Polytomous): نماذج تستخدم في تدريج الفقرات ذات الاستجابات المتعددة، وغالبا ما تستخدم هذه النماذج في مقاييس الاتجاهات والشخصية التي يزيد فيها عدد البدائل عن بديلين (Cavanagh & Romanoki, 2006). وبما أنّ أداة الدراسة هي مقياس متعدد الاستجابة بفئات تدريج مختلفة (ثلاثي ، رباعي ، خماسي) وقد تم استخدام نموذج الاستجابة المتدرجة (Graded Response Model) من النماذج متعددة الاستجابة Polytomous ، وبالتالي سيتم توضيح هذا النموذج في هذه الدراسة الحالية .

نموذج الاستجابة المتدرجة GRM (Graded Response Model): اقترحت هذا النموذج ساميجما (Samejima, 1969) ، ليكون هناك إطار نظري للتعامل مع درجات الفقرة المتدرجة m_g..... , 0, 1, 2 ، في نظرية استجابة الفقرة IRT وهو يستند على النموذج ثنائي المعلمة (Reise & Yu, 1990). ويعد نموذج الاستجابة المتدرجة امتدادا لطريقة ثيرستون (Thurston Method)، في تحليل الفقرات ذات الاستجابات المتدرجة في الاختبارات التربوية، (Tang & Eignor, 1997). وكذلك يعد نموذج الاستجابة المتدرجة (Graded Response Model) تطبيقا لنموذج راش وتعميما لنموذج بيرنيوم الثنائي الاستجابة ، حيث يتم من خلاله قياس الشخصية والاتجاهات والنواحي المعرفية ، وإعطاء درجات جزئية على الإجابات الصحيحة (Rauch et al., 2008) .

شكل 1

عدد العتبات الفارقة لاستجابات الفقرة ذات خمسة اقسام



ويتم تفسير قيمة معلمة العتبة الفارقة (B_{ij}) على أنها تمثل مستوى السمة اللازم ، حتى تتخطى الاستجابة الفارقة (j) باحتمال مقداره (0.5) ويتم في هذا النموذج معالجة الفقرة من خلال سلسلة من الاجزاء الثنائية ، فاذا كان عدد بدائل الاستجابة خمسة بدائل ، والدرجات الخاصة بهذه الخيارات هي : 0 , 1, 2 , 3 , 4 فإن معالجة الفقرة في هذه الحالة يكون من خلال سلسلة من الاجزاء الثنائية التالية : (0 مقابل 1, 2, 3, 4) ، (0, 1 مقابل 2, 3, 4) ، (0,1 ,2 مقابل 3, 4) ، (0 ,1 ,2 ,3 مقابل 4) (Nering & Ostini, 2011; Samejima, 1997) . ويتم في نموذج الاستجابة المتدرجة (GRM) الحصول على منحنيات الاستجابة للاجزاء الثنائية من خلال خطوتين ، تكون الخطوة الأولى بحساب احتمال اجابة الفرد عن كل جزء من الأجزاء الثنائية للحصول على المنحنيات المميزة من خلال النموذج الثنائي حسب معادلة 1 الآتية :

$$P_{ix}^{*} = \frac{\exp\left(a_{i(\theta-\beta_{ij})}\right)}{1+\exp\left(a_{i\,(\theta-\beta_{ij})}\right)} \dots\dots\dots\dots\dots\dots\dots\dots (1)$$

تتضمن الخطوة الثانية حساب مقدار الاحتمال الفعلي لبدائل الاستجابة الخمسة ، وذلك عن طريق طرح احتمال اجابة الفرد لكل جزء من الأجزاء الثنائية من احتمال اجابة الفرد للجزء الذي يليه ، حسب معادلة 2 الآتية:

$$P_{ix}(\theta) = P_{ix}^{*}(\theta) - P_{ix+1}^{*}(\theta) \dots\dots\dots\dots\dots\dots\dots\dots (2)$$

ثالثا: نموذج الاستجابات المتصلة (Continuous Responses Model) : هذا النموذج هو حالة خاصة من نموذج الاستجابة المتدرجة، تم اقتراح هذا النموذج من قبل ساميجما (Samejima, 1969) ، حيث يتم وضع استجابات المفحوص عن الفقرة على مقياس مستمر، ويعد هذا النموذج هاما ونافعا للباحثين في علم النفس الاجتماعي، وكذلك للمهتمين بدراسة الاتجاهات (Hambelteon & Swaminathan, 1985) .

من خلال افتراضات نظرية استجابة الفقرة (IRT) والنماذج التي تقدمها، يمكن الاستفادة من تطبيق العديد من المؤشرات الإحصائية المختلفة لمطابقة استجابة الفرد، تُسمى(PFSs)

Person Fit Statistic بهدف تحديد ومعرفة نمط استجابة الفرد لفقرات المقياس (response pattern) (Reise & Flannery, 1996).

مطابقة الفرد Person-fit: يشير مفهوم نمط الاستجابة غير مطابق للفرد Person Fit وفق نظرية الاستجابة للفقرة(IRT) على أنه نمط استجابة ملاحظ ينحرف عن نمط الاستجابة المتوقع ، استنادًا إلى قدرة الفرد وخصائص نموذج الاستجابة للفقرة المستخدم (Meijer & Sijtsma, 2001; Reise,1990).

لقد حظيت مؤشرات مطابقة الفرد PFSs باهتمام كبير من قبل علماء النفس التربوي والقياس النفسي لعلاقتها بتحسين جودة الأداء في الاختبارات والمقاييس التربوية، حيث تساهم في التأكد من صدق نتائج القياس، وقابلية النتائج للتفسير، وفي تحديد الاستجابات غير المطابقة لنمط ما، والتي حتماً ستؤثر على صدق نتائج القياس.(Reise & Flannery, 1996).حيث تؤدي الخصائص الديموغرافية مثل الجنس والعرق والتحيز الثقافي لمجموعة من الافراد دور في توليد انماط استجابة غير مطابقة ،ولوجود العديد من الاسباب التي تؤدي إلى حدوث انماط غير مطابقة ، ولصعوبة تحديد السبب الذي أدى إلى حدوث أنماط غير مطابقة لتداخل عدة عوامل تؤثر في استجابة الفرد، ظهرت مؤشرات مطابقة الفرد لتحديد مدى مطابقة الأنماط غير مطابقة مع النموذج المستخدم. (Meijer, 1996). ويفيد تحليل مطابقة الفرد في الكشف عن أنماط الاستجابة غير المطابقة، ويمكن استخدام العديد من مؤشرات المطابقة لذلك. ويمكن الحكم على مؤشر المطابقة بالاعتماد على معيارين. الأول مدى امتلاك المؤشر الإحصائي لخصائص التوزيع الطبيعي، حيث يمكن مقارنة قيم الإحصائي مع درجة قطع واحدة لجميع قيم القدرة ويمكن أن يستدل على هذا المعيار من خلال حساب خصائص التوزيع (الوسط الحسابي، الانحراف المعياري، الالتواء والتفلطح) ومقارنتها بخصائص التوزيع الطبيعي أو من خلال حساب معامل الارتباط بين احصائي مطابقة الفرد ومعلمة القدرة (Birenbaum, 1985). أما المعيار الثاني فهو قوة المؤشر الإحصائي في الكشف عن الأنماط غير المطابقة، وهذا المعيار هو دالة لعوامل متعددة تتضمن طول الاختبار، ومستوى القدرة، ونوع عدم المطابقة . ويستدل على قوة الإحصائي من خلال النسبة التي يكشفها من الأنماط غير المطابقة أو من خلال معاملات الارتباط بين مؤشرات احصائيات مطابقة الفرد والتي تشير إلى مدى اشتراكها في تصنيف نفس الافراد كمطابقين أو غير مطابقين للنموذج (Deng & Terre, 2008). ويتم الاتفاق بين الإحصائيات في تحديد الأنماط غير المطابقة من خلال معاملات

الارتباط بين الإحصائيات ، فكلما كانت معاملات الارتباط أعلى كانت نسبة الاتفاق أعلى (Meijer & Sijtsma, 1995). ومن المهم معرفة المؤشرات الإحصائية التي يتفوق بعضها على الآخر، أي المؤشرات التي تكون أكثر قوة في اكتشاف الاستجابة غير المطابقة ، ويمكن ذلك من خلال استخدام بيانات محاكاة أو بيانات حقيقية (Karabatsos, 2003).

مؤشرات مطابقة الفرد Person-fit Indicators: صنف كل من ميجر وسيجسما (Meijer & Sijtsma, 1999) مؤشرات مطابقة الفرد إلى ثلاث فئات:

أولاً: احصائيات مطابقة الفرد المعتمدة على المجموعة group-based indices : احصائيات مطابقة الفرد التي تستند إلى خصائص المجموعة نمط درجة الفقرة فرديا مع أنماط درجة الفقرة الخاص بالأفراد الآخرين في العينة . (Meijer & Sijtsma, 1999).

ثانياً: مؤشرات مطابقة الفرد المعلمية Parametric person-fit indices : تستند إلى نموذج نظرية استجابة الفقرة المعلمي . نماذج نظرية الاستجابة هي افتراضات حول ما يحدث عندما يستجيب الفرد على فقرة أو بعبارة أخرى يشرح احتمال الحصول على درجة فقرة معينة حسب خصائص الفرد (القدرة) وخصائص الفقرات (مثل صعوبة الفقرة) ..

ثالثاً: مؤشرات مطابقة الفرد اللامعلمية Nonparametric person-fit indices : تستند في بحث مطابقة الفرد اللامعلمية لا يتم تحديد احتمال الاستجابة الصحيحة للفقرة بشكل معلمي كما هو الحال في البحث المعلمي لمطابقة الفرد . النماذج اللامعلمية عموما أقل تقييدا من النماذج المعلمية. الافتراضات (الاستقلال الموضعي ، وتيرية دالة استجابة الفقرة و أحادية البعد للمتغير الكامن) إلى حد كبير القياس الرتبي لا يزال ممكنا.وبالتالي، فإن استخدام المؤشرات اللامعلمية قد تطابق النموذج أكثر من استخدام المؤشرات المعلمية. (Meijer, 1994) .

وقد ركزت الدراسة الحالية على استخدام مؤشرات لامعلمية لمطابقة الفرد للفقرات متعددة الاستجابة ودراسة نمط استجاباتهم باستخدام نموذج GRM، للبحث في مدى التوافق والارتباط بين هذه المؤشرات GP و U3P في الكشف عن عدم مطابقة الافراد ، وهناك عدة أسباب لاستخدام الإحصاءات اللامعلمية ،تتمثل في كونها أكثر مرونة في افتراضاتها، كما تتميز الإحصاءات اللامعلمية بأنها غالبا ما تكون أقل تقييداً فيما يتعلق بالبيانات مقارنة بالإحصاءات المعلمية ،ومع ذلك يقتصر قياسها على المستوى الرتبي (Junker & Sijtsma, 2001).

حيث تتناول هذه الدراسة تحليل مطابقة الفرد في سياق IRT اللامعلمية والتي لا تفترض شكل محدد لدالة استجابة الفقرة IRF، وبالرغم من أنّ نموذج (GRM) من نماذج نظرية استجابة الفقرة المعلمية PIRT؛ إلا أنه يمكن ان يستخدم مع الإحصاءات اللامعلمية كونها الحالة العامة للإحصاءات المعلمية (Sijtsma, 1998). كما يتمثل السبب الرئيسي لتحليل البيانات استنادا إلى GRM(من نماذج PIRT) هو ان آلية GRM تعتبر شكل خاص من نموذج التجانس الوتيري اللامعلمي (MHM Monotone Homogeneity Model) ،حيث أن البيانات التي تتكيف مع GRM تتكيف ايضا مع MHM (Sijtsma el al., 2008). و يتم تعريف MHM الذي تم تطويره للفقرات المتعددة الاستجابة على أنه GRM اللامعلمي ،وقد تم تطويره لقياس الاستجابات الترتيبة مثل مقاييس ليكرت Likert – scales.(Hemker el al., 1996) .

المؤشر الإحصائي Gpoly (Guttman errors) :يشير المؤشر الإحصائي GP إلى أنه عندما تتوافق الفقرات مع مقياس جتمانGuttman scale ،يتم ترتيبها بشكل هرمي بحيث تؤدي الإجابة عن فقرة واحدة إلى الإجابة عن الفقرات التي تليها على نمط البنية الكامنة التي يتم قياسها ،عدد أخطاء جتمان الخاصة بالفقرات متعددة الاستجابة polytomous items سيتم تحديدها باستخدام المؤشر الإحصائي GP(Molenaar, 1991)). يتم إعطاء صيغة المعادلة للمؤشر الإحصائي GP من خلال معادلة 3 الآتية :

$$G^p = \sum_{l<k}^{JM} y_k(1-y_l) \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (3)$$

حيث :

J : هو عدد الفقرات ،والتي يُفترض هنا أن تحتوي جميعها على نفس عدد فئات الاستجابة(M+1)، وبالتالي نفس عدد خطوات الفقرة (M) لكل فقرة .

إذا كان M=1 ، فإن المؤشر الإحصائي GP يتم تحديده للفقرات ثنائية الاستجابة Dichotomous.

y_l : يمثل جميع فقرات المتجه Y التي تسبق y_k .

حيث أنه كلما تم تسجيل الكثير من أخطاء جتمان للمستجيب ،زادت قيمة المؤشر الإحصائي GP ، مما يُشير إلى عدم تطابق الفرد (المستجيب) بشكل أكبر .كما ثبت أن المؤشر الإحصائي GP

يكشف الاستجابة غير المطابقة بشكل جيد نسبيًا مقارنة بإحصاءات مطابقة الفرد اللامعلمية الأخرى عندما يعكس المستجيب سلوك استجابة غير مبالية(غير مطابقة) ، اي الاستجابة العشوائية أو فشل المستجيبون في ملاحظة الفقرات المصاغة في الاتجاه المعاكس (Emons, 2008).

المؤشر الإحصائي U_{3P} : قام فان دير فلاير (Van der Flier, 1980) بتعميم مؤشر مطابقة الفرد U3 الذي يستخدم مع الفقرات ثنائية الاستجابة ،للتطبيق على فقرات متعددة الاستجابة polytomous ، مما أدى إلى إيجاد المؤشر الإحصائي U_{3P} . حيث تم تقصي فعالية هذا المؤشر عبر الدراسات التي أشارت إلى تأثير فاعلية هذا المؤشر بشكل كبير من بين مؤشرات احصائيات مطابقة الفرد اللامعلمية والمعلمية الأخرى (Emons, 2008; Karabatsos, 2003; Tendeiro & Meijer, 2014; Turner,2018).

يمكن تعريف المؤشر الإحصائي اللامعلمي U_{3P} في بضع خطوات .أولاً ، يتم حساب المجموع الموزون /المرجح weighted sum لسجل احتمالات log–odds صعوبات خطوة الفقرة للخطوات التي مر بها المستجيب ، (y) W ، حسب معادلة 4 الآتية (Emons, 2008):

حيث:

Y: عبارة عن استجابة تمت ملاحظتها لفقرات (J) مع فئات استجابة (C+1)

yk: هي درجة خطوة الفقرة لخطوة الفقرة k (yk is the item step score for item–step k)

(مع أخذ القيمة 1 إذا تم تجاوز الخطوة k أو 0 إذا لم يتم تجاوز الخطوة k)

πk: هي صعوبة خطوة الفقرة لخطوة الفقرة k .

ثم يتم تحديد اللوجت لنسبة الترجيح

بعد ذلك يتم تحديد W(y) مما ينتج عنه مؤشر مطابقة الفرد U_{3P}كما في معادلة 5 الآتية : (Emons, 2008)

$$U3^{p} = \frac{\max(W|X_{+}) - W(\mathbf{y})}{\max(W|X_{+}) - \min(W|X_{+})} \quad \text{... ... ...(5)}$$

حيث :

max(W|X+) :احتمال تجاوز الحد الاقصى لعدد خطوات الفقرة التي تم اجتيازها نظرًا لإجمالي درجة الفردn.

min(W|X+): احتمال تجاوز الحد الادنى لعدد خطوات الفقرة التي تم اجتيازها نظرًا لإجمالي درجة الفرد n.

تتراوح قيم U3 من 0 إلى 1، حيث يشير(0) إلى أن نمط الدرجة الملاحظة للفقرة هو متجه جتمان Guttman vector و يشير (1) إلى أن النمط الملاحظ هو معكوس متجه جتمان reversed Guttman vector. كذلك تشير القيم المتزايدة إلى أن الأنماط بعيدة عن أنماط جتمان المثالية (Meijer el al., 1996).

مشكلة الدراسة

تعد دراسة مطابقة نمط استجابة الفرد واحدة من اهم المشكلات التي يواجهها الباحثون ضمن إطار نظرية استجابة الفقرة (IRT)، حيث لا يزال الغموض قائما حول أسباب عدم المطابقة، وكيفية اكتشافها والتحكم والتنبؤ بها . انطلاقا من ذلك ، ظهرت الحاجة إلى معرفة نتائج استجابة الأفراد على أدوات جمع البيانات التي يتم تطبيقها لأغراض متعددة وذلك بهدف اصدار الاحكام بناءً عليها .تناولت هذه الدراسة مؤشرين من المؤشرات الإحصائية اللامعلمية (NIRT)، وهي المؤشر الإحصائي (G_P)، و(U_{3P}) لبحث في أداء هذه المؤشرات ومدى التوافق بينها في الكشف عن الأنماط غير المطابقة، مما يسهل على الباحثين ومستخدمي نماذج استجابة الفقرة معرفة المؤشر الافضل في الكشف عن نسبة الأنماط غير المطابقة، وهي من المؤشرات اللامعلمية التي لم تستخدم بشكل كبير في الدراسات وقد طغى استخدام الإحصائيات المعلمية بالرغم من محدودية قدرتها على إيجاد معالم مهمة مثل نسبة الإجابة الصحيحة، والصعوبة، والقدرة التمييزية للفقرات وذلك بسبب امتلاكها افتراضات متشددة يصعب تحقيقها في حالات كثيرة، تبلورت مشكلة هذه الدراسة بسبب عدم امتلاك النماذج المعلمية مؤشرات مطابقة تتصل بتحديد نمط الإجابة للمستجيب كما هو الحال في الإحصاءات اللامعلمية، مما يؤدي إلى رفض الكثير من الفقرات التي لا تتطابق مع الشكل اللوجستي، وهذا يعني فقدان الكثير من الفقرات التي تفيد في دقة ترتيب المستجيبين، والتي تمتلك دوال استجابة غير متناقضة مما يؤثر بشكل واضح في النتائج (Sijtsma & Molenaar,

2002). ويتم استخدام نموذج الاستجابة المتدرجة (GRM) بشكل كبير في تصميم الاختبارات والمقاييس التربوية، وهو من نماذج نظرية استجابة الفقرة يستخدم مع الإحصاءات المعلمية، حيث تم استخدام (GRM) مع المؤشرات الإحصائية اللامعلمية باعتبار أنه تم تطويره للفقرات متعددة الاستجابة اللامعلمية، حيث لعبت نظرية استجابة الفقرة اللامعلمية دورًا مهمًا في تطوير نظرية استجابة الفقرة، وهي تمثل الحالة العامة لنظرية استجابة الفقرة التي تشكل النظرية المعلمية حالة خاصة منها، وكذلك تقوم بأداء جيد مقارنة بإحصاءات مطابقة الفرد الأخرى، الأمر الذي يساهم في إمكانية الكشف عن عدم مطابقة الفرد لنماذج نظرية استجابة الفقرة، وتزويد الباحثين بمعلومات قد تفيد في استخدام مؤشرات مطابقة لامعلمية أخرى لديها القدرة في الكشف عن الاستجابة غير المطابقة لدى المستجيبين، مما يؤدي إلى امكانية التعامل مع البيانات للحصول على نتائج أكثر دقة، وهو الهدف الاسمى في مجال مطابقة الفرد.

تسعى الدراسة الحالية إلى الإجابة عن السؤال التالي:

– ما مدى التوافق بين المؤشرين (G_P) و($\boldsymbol{U_{3P}}$) في الكشف عن الافراد غير المطابقين من خلال معاملات الارتباط بين الإحصائيات باختلاف عدد فئات التدريج (ثلاثي، رباعي، خماسي) واختلاف المراحل العمرية المتمثلة (الصف السابع، الصف الأول ثانوي) ؟

هدف الدراسة

ثبت أن الاستجابة غير المطابقة تؤثر تأثيرا سلبيا على نتائج البيانات، والتي يمكن أن تؤدي إلى عدم دقة القرارات التي يتم اتخاذها استنادا إلى تلك النتائج ،لذلك قامت هذه الدراسة بأخذ مؤشرين من مؤشرات مطابقة الفرد اللامعلمية ،هما مؤشر(GP)، و($\boldsymbol{U_{3P}}$) بهدف البحث في مدى التوافق بينهما في الكشف عن الاستجابات غير المطابقة، من خلال نظرية استجابة الفردIRT باستخدام مقياس قلق القراءة باللغة الانجليزية، يتكون من فقرات متعددة الاستجابة (Polytomous) حيث تقدم هذه الدراسة طريقة لاستخدام نموذج (GRM) في البحث في مدى التوافق بين المؤشرين في الكشف عن الاستجابات غير المطابقة.

أهمية الدراسة

الأهمية النظرية

تكمن أهمية هذه الدراسة من خلال سعيها في البحث عن أداء اثنين من مؤشرات مطابقة الفرد اللامعلمية Per-Fit ومدى التوافق بينهما باستخدام بيانات حقيقية لمقياس قلق القراءة باللغة الانجليزية، مكون من مجموعة من الفقرات متعددة الاستجابة باستخدام نموذج الاستجابة المتدرجة (GRM)، وفئات متعددة للتدريج، ولمراحل عمرية مختلفة.

الأهمية العملية

تكمن أهمية هذه الدراسة من خلال سعيها في البحث حول استخدام نموذج الاستجابة المتدرجة(GRM) بفئات تدريج مختلفة وتوفير قرارات حول اختلاف فئات التدريج وأثره على أداء المؤشرات المستخدمة في الدراسة للاستفادة منه في توفير أدوات قياس تربوية ونفسية بفئات تدريج ملائمة وتناسب الفئات العمرية المختلفة.

مصطلحات الدراسة

اشتملت الدراسة على العديد من المصطلحات والمفاهيم التربوية، والتي يمكن تعريفها على النحو الاتي:

اداء المؤشر (المطابقة) : نسبة الافراد المطابقين للمؤشرات الإحصائية المستخدمة.

مؤشرات مطابقة الفرد اللامعلمية: هي مؤشرات لا تستند إلى افتراضات، ولا تعتمد على معالم نظرية استجابة الفقرة المقدر وانما يتم حسابه من مجموعة من البيانات لعدد من الافراد.

نموذج الاستجابة المتدرجة (GRM):هو أحد نماذج استجابة الفقرة متعددة الاستجابة حيث يعتمد على فقرات متعددة الأستجابة ويعد هذا النموذج تعميماً لنموذج بيرنبوم ثنائي المعلمة (2PLM).

حدود الدراسة

تم إجراء الدراسة ضمن الحدود الآتية:

1. اقتصرت على البحث في مدى التوافق بين مؤشرين فقط من المؤشرات الإحصائية لمطابقة الفرد وهي (G_P)، و(U3).
2. اقتصرت فقط على فئات التدريج (الثلاثي، الرباعي، الخماسي) للفقرات.

3. اقتصر تطبيق المقياس على مرحلتين عمريتين والمتمثلة بصفوف دراسية (الصف السابع ، الصف الأول ثانوي).

الدراسات السابقة:

تناولت الباحثة في هذا الفصل عرضًا للدراسات السابقة ذات الصلة المباشرة بموضوع اداء المؤشرات الإحصائية لمطابقة الفرد، حيث تم استعراضها من الاقدم إلى الأحدث على النحو التالي:

قام رودنر (Rudner, 1983) بدراسة فاعلية المؤشرات r_i و rb_i و NCl_i و C_i و Wl_i و $U3_j$ و $W3_i$ و $L3_i$ و Ul_i باستخدام بيانات مولدة وتم تحديد فقرات الاختبار، تم محاكاة المستجيبين واستجاباتهم لفقرات الاختبار، تم استخدام النموذج اللوجستي ثلاثي المعلمة لـ بيرنبأوم لربط خصائص الفقرة باستجابات الافراد على الفقرات ،حيث تم إنشاء مجموعتين مستقلتين من البيانات، تم تصميم مجموعة البيانات الأولى، التي يشار إليها بالاختبار الأول ،لمحاكاة أداء المستجيبين في اختبار تجاري عالي الجودة واستند إلى معلمات لورد (1968) Lord، تكون من 80 فقرة من اختبار القدرات اللفظية، تم تصميم مجموعة البيانات الثانية، التي يشار إليها بالاختبار الثاني، لمحاكاة استجابات الفقرات من اختبار مطور للفصل الدراسي واستند إلى معلمات (1977) Bejar, Weiss, and Kingsbury's ، تكون من 45 فقرة من اختبار الأحياء العامة ، كان متوسط تمييز الفقرة والصعوبة والتخمين لمعلمات الفقرة من خلال نموذج بيرنبأوم على التوالي 1.07, 58. ،و، 16., على التوالي للاختبار الأول، و 1.09، 08. ، 25. على التوالي للاختبار الثاني ،بعد ذلك تم تحويل احتمال الاستجابة الصحيحة إلى استجابة يمكن ملاحظتها صحيحة أو خاطئة من خلال مقارنتها بعدد عشوائي موزع بشكل موحد بين 0 و 1 ، تعتبر الإجابة صحيحة اذا كان الاحتمال المرتبط بها اكبر من الرقم العشوائي ($u_{ij} = 1$) ،وغير صحيحة اذا كان أقل ($u_{ij} = 0$)، كانت الدرجات الاجمالية لأنماط الاستجابة غير المطابقة منخفضة ، تم حساب الارتباطات بين المؤشرات التسعة، اظهرت النتائج أنه يوجد ترابط كبير بين قيم المؤشرات r_i و rb_i و NCl_i و C_i و Wl_i ، بينما تميل قيم المؤشرات و$U3_j$ و $W3_i$ و $L3_i$ و Ul_i إلى ان تكون أقل ارتباطا.

قام ميجر وتنديرو (Meijer & Tenderio, 2014) بدراسة بهدف المقارنة بين ثلاثة مؤشرات إحصائية لامعلمية (C^* , H^T , U3)؛ لتقييم جودة نتائج الاختبارات الفردية في ضوء متغيري (الجنس ، والعرق)، حيث اهتم الباحثان بالتعرف إلى أنماط درجة الفقرة غير المطابقة،

ودراسة العلاقة بين درجات مطابقة الفرد في المجموعات الفرعية للجنس والعرق، واستخدم الباحثان البيانات التجريبية لاختبار تربوي (اختبار قبول الطلبة في كلية الحقوق)، وقد اظهرت نتائج الدراسة عدم وجود فروق في نمط درجة الفقرة بين الذكور والإناث المتقدمين للاختبار، كما اظهرت نتائج الدراسة وجود فروق دالة احصائيًا في وسط المؤشر الإحصائي (H^T) داخل المجموعات الفرعية لمتغير العرق .

اجرى سينجل افسر (Şengül AvŞar, 2019) دراسة للمقارنة بين مؤشرات مطابقة الفرد (, U3p Gp, GNp وlzp) للفقرات متعددة الاستجابة في ظروف اختبار مختلفة، وكان الهدف من دراسته فحص فاعلية مؤشرات مطابقة الفرد المعلمية واللامعلمية في مجموعات البيانات التي تتكون من فقرات متعددة الاستجابة ، تمت محاكاة البيانات باستخدام نموذج الاستجابة المتدرجة (GRM)، تضمنت ظروف الاختبار مستويات مختلفة من حجم العينة، وأشكال مختلفة لتوزيع قدرة الفرد ،واربعة معدلات خطأ من النوع الأول، ومستويات مختلفة من طول الاختبار ومستوين من عدم المطابقة (منخفض وعالي). ظهر ان 20% من المشاركين انخفض مستوى عدم المطابقة سلوك الاستجابة غير المطابق في نصف الفقرات، وبالنسبة للمستوى العالي من عدم المطابقة، 30% من المستجيبين أظهر سلوك استجابة غير مطابق في جميع الفقرات أظهرت النتائج انه كلما زاد حجم العينة، زاد معدل الكشف عن العديد من ظروف الاختبار، وزادت معدلات الكشف مع زيادة مستويات عدم المطابقة. و أظهر المؤشر اللامعلمي Gp أداء أفضل لاكتشاف عدم المطابقة ،كذلك وجد ان مؤشر U3p و GNp اللامعلمية مؤشرات نتائجها قريبة جدا من بعض عند معدلات الخطأ من النوع الأول.

التعقيب على الدراسات السابقة

من خلال الاطلاع على الدراسات السابقة فقد لاحظت الباحثة وجود العديد من الدراسات التي تناولت موضوع المقارنة بين أداء المؤشرات الإحصائية لمطابقة الفرد وكان بعضها باستخدام مؤشرات معلمية (Param)، ومؤشرات لامعلمية (NParam)، كما في دراسة(1983, Rudner)، ودراسة آفزر (AVŞAR, 2019). وهناك دراسات تناولت المقارنة بين المؤشرات الإحصائية لمطابقة الفرد باستخدام بيانات محاكاة كما في دراسة رودنر(1983ز Rudner)، ودراسة تيندرو ومايجر (Tendeiro & Meije, 2014)، ودراسة آفزر(AVŞAR, 2019). كما

تبين هناك اختلاف بين الدراسات التي قارنت بين المؤشرات الإحصائية لمطابقة الفرد من حيث متغيرات الدراسة ؛ ففي دراسة رودنر(Rudner , 1983) ودراسة آفزر(AVŞAR, 2019) كان المتغير طول الاختبار ،بينما دراسة ميجر وتنديرو (Meijer & Tenderio, 2014) كانت المتغيرات هي متغير الجنس والعرق .

الطريقة والإجراءات

تناول هذا الفصل وصفًا لمنهجية البحث المستخدمة في هذه الدراسة، ومجتمع وأفراد الدراسة، بالإضافة إلى أدوات الدراسة المستخدمة، وطرق التأكد من مدى صدقها وثباتها، واجراءاتها، والمعالجة الإحصائية.

منهج الدراسة

اتبعت الدراسة الحالية المنهج الوصفي التحليلي وذلك للمقارنة بين مؤشرين احصائيين من مؤشرات مطابقة الفرد (PFSs)، والتعرف على مدى التوافق بين المؤشرين في الكشف عن أنماط الاستجابة غير المطابقة للفرد، باستخدام نموذج الاستجابة التدريجية (GRM) من نماذج نظرية استجابة الفقرة (IRT)، من خلال مقياس قلق القراءة باللغة الانجليزية، بفئات تدريج مختلفة (ثلاثي، رباعي، خماسي) في ضوء اختلاف الفئات العمرية (الصف السابع، والأول ثانوي) .حيث تم أخذ عينة استطلاعية ،للبحث في صلاحية هذا المقياس والتحقق من افتراضات نموذج الاستجابة المتعددة قبل تطبيقه على عينة الدراسة للتمكن من تطبيقه و التعديل على فئات تدريج المقياس بصور مختلفة.

مجتمع الدراسة

تكون مجتمع الدراسة من جميع طلاب وطالبات الصفين السابع والأول ثانوي في الفصل الدراسي الثاني من العام 2022/2021م في المدارس التابعة للواء الكورة في محافظة إربد والبالغ عددهم(5523) لعام 2022/2021. إضافة إلى أنهم موزعين وفقا لجنس الطلبة إلى (1362) طالباً من مرحلة الأول ثانوي و(1291) طالبة من مرحلة الأول ثانوي و (1383) طالبا من

الصف السابع و (1487) طالبة من الصف السابع بحسب إحصاءات دائرة التخطيط في مديرية التربية والتعليم في لواء الكورة .

عينة الدراسة

تم اختيار عينة عشوائية عنقودية من طلبة المدارس في لواء الكورة التابع لمحافظة إربد، من الصفين السابع والأول ثانوي من العام الدراسي 2021 /2022، حيث كانت وحدة الاختيار هي الصف الدراسي ، وقد بلغ حجم عينة الدراسة (1537) طالبًا وطالبةً، تم تطبيق المقياس عليها حيث كان حجم العينة لفئة الصف السابع (779) طالبًا وطالبةً، كما بلغ حجم العينة لفئة الصف الأول ثانوي (758)، وتم توزيع الأداة على المستجيبين بشكل عشوائي من خلال ترتيب الأداة بشكل تسلسلي من حيث فئات التدريج للفقرة (ثلاثي، رباعي، خماسي)، وعند توزيع الأداة على عينة الدراسة أُعطي الطالب الأول المقياس بصورة التدريج الثلاثي والثاني المقياس بصورة التدريج الرباعي ،والثالث بصورة التدريج الخماسي ، وتم تكرار ذلك على العينة كاملة بهدف تحقيق العشوائية.

أداة الدراسة

تم استخدام مقياس قلق القراءة باللغة الانجليزية والذي أعدَّه سايتو وزملاؤه (Saito et al.,1999)، والذي تمت ترجمته لاحقا من قبل عطروز(2018) وتم تقنينه للبيئة الأردنية للكشف عن قلق القراءة لدى الطلبة في اللغة الانجليزية، بحيث تضمن المقياس (20) فقرة، يتم الإجابة عليها وفق تدريج ليكرت الخماسي، الذي يبدأ من (1) "لا أوافق بشدة " وينتهي ب (5) "أوافق بشدة". باستثناء الفقرات السلبية (13، ،14، 18، 12)، اذ يتم عكس التدريج عند ادخال البيانات، ومع الاطلاع على مقاييس القراءة الأخرى تم إضافة فقرتين من مقياس هيو (Hui-Ju, 2011) ، ليكون مجموع فقرات المقياس بصورته النهائية (22) فقرة عطروز(2018).

ولتحقيق أغراض الدراسة قامت الباحثة في الدراسة الحالية بالتحقق من الخصائص السيكومترية للمقياس على عينة استطلاعية، للتمكن من تطبيقه والتعديل على فئات تدريج المقياس بصور مختلفة لتصبح بالتدريج الثلاثي، والرباعي، إضافة إلى الصورة الاصلية للمقياس وهي التدريج الخماسي.

أولاً: صدق المحتوى

للتحقق من صدق محتوى المقياس تم عرضه على (10) محكمين من الأساتذة المتخصصين في مناهج اللغة العربية وأساليب تدريسها، والقياس والتقويم وعلم النفس التربوي، لتعديل ما يرونه مناسبًا من الفقرات لتكون أكثر وضوحاً وقابلية للقياس. وقد أجمع المحكمون على صلاحية المقياس وفقراته لتحقيق أغراض الدراسة من دون حذف أي فقرة منها و من دون إجراء أي تعديلات.

ثانيا :صدق بناء المقياس.

تم التحقق من صدق البناء العاملي لمقياس قلق القراءة باللغة الانجليزية في الدراسة الحالية، من خلال تطبيقه على عينة استطلاعية مكونة من (n=432) طالبًا وطالبة من طلبة الفئتين العمريتين (السابع، والحادي عشر) في مدارس لواء الكورة التابع لمحافظة إربد بواقع (199) مستجيب من الصف السابع و(233) مستجيب من الصف الحادي عشر بتدريجه الخماسي المعتمد في الصورة الأصلية للمقياس والذي يعتبر التدريج الأكثر حساسية والأكثر قدرة على تقديم معلومات أكثر من خلال فئاته الخمس. وقد تم استخدام التحليل العاملي التوكيدي (CFA) للتحقق من البنية العاملية للمقياس باستخدام برمجية (AMOS)، أظهرت قيم مؤشرات مطابقة البيانات للنموذج التوكيدي للمقياس ان النموذج مطابق للبيانات حيث بلغت قيمة مؤشر نسبة مربع كأي إلى درجة الحرية (χ^2/df) (1.023) وهي قيمة أقل من (2) حسب (Ullman, 2001). وبلغت قيمة جذر متوسط مربعات الخطأ(RMSEA) (0.007) وهي قيمة أقل من (0.08) وبلغت قيمة مؤشر المطابقة المقارنة (CFI) (0.99) وهي قيمة اكبر من (0.95) حسب (Schreiber el al., 2006). وان حجم العينة مناسب من خلال مؤشر هويلتر الذي بلغت قيمته (524) وهي قيمة أكبر من (75) حسب (Garson, 2009). مما سبق يتضح ان المقياس يتمتع ببنية عاملية صادقة ومقبولة لغايات هذه الدراسة.

ثبات المقياس.

للتحقق من ثبات المقياس تم حساب معامل كرونباخ الفا لاستجابات الطلبة في العينة الاستطلاعية وقد بلغت قيمته (0.839) وهي قيمة مقبولة لغايات إجراء هذه الدراسة .

مطابقة البيانات للنموذج

للتحقق من مدى مطابقة فقرات المقياس لنموذج الاستجابة المتدرج تم اتباع الإجراءات التالية:

أولا: تم التحقق من افتراضات أحادية البعد (Unidimensionality) والاستقلال الموضعي(Local Independence) :

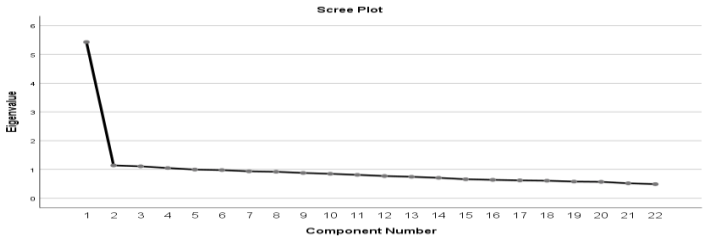
للتحقق من افتراض أحادية البعد قامت الباحثة باستخدام التحليل العاملي الاستكشافي حيث أظهرت نتائج التحليل الجذور الكامنة ، حيث أشارت إلى أربعة عوامل كانت قيمة الجذر الكامن لها اكبر من 1 . كما هو مبين في جدول1.

جدول 1: التحليل العاملي الاستكشافي للتحقق من أحادية البعد.

العامل	الجذر الكامن	الجذر الكامن الأول الجذر الكامن الثاني
1	5.428	4.757
2	1.141	
3	1.109	
4	1.050	

يتضح من جدول 1 أن نتائج التحليل العاملي الاستكشافي تشير إلى تحقق أحادية البعد ، لأنّ ناتج قسمة الجذر الكامن للعامل الأول على الجذر الكامن للعامل الثاني أكبر من 2 ، كما أن ناتج قسمة حاصل طرح الجذر الكامن الثاني من الجذر الكامن الأول على حاصل طرح الجذر الكامن الثالث من الجذر الكامن الثاني ذو قيمة عالية (Hoyle&Rattray, 2004). وللتاكيد على ذلك تم تمثيل الجذور الكامنة بيانيا (Scree Plot) كما هو مبين في شكل 2.

شكل 2: التمثيل البياني للجذور الكامنة.



يتبين من شكل 2 أن أحادية البعد متحققة حيث يلاحظ أن الانعطاف في التمثيل البياني يقابل الجذر الكامن الثاني أي بعد الجذر الكامن الأول.

وللتحقق من افتراض الاستقلال الموضعي قامت الباحثة بتحليل استجابات الطلبة في العينة الاستطلاعية باستخدام برمجية (IRTPRO) وتم استخراج مؤشر مربع كأي للاستقلال الموضعي، أشارت النتائج إلى تحقق الاستقلال الموضعي، حيث إن نسبة الأزواج المستقلة إلى الأزواج غير المستقلة كبيرة وتتجاوز الأربعة اضعاف، حيث لوحظ وجود زوجين معتمدين (غير مستقلين) فقط من أصل 231 زوج، لتشكل ما نسبته 99% وهي نسبة عالية.

منحنى خصائص الفقرة (ICC: Item Charactrastic Curve):

تم التحقق من هذا الافتراض باستخدام برمجية IRTPRO حيث تم تحليل استجابات الطلبة على نموذج الاستجابة المتدرجة وتم تمثيل منحنى خصائص كل من فقرات المقياس في الملحق.

التحقق من اطراديه فقرات المقياس (Monotonicity):

للتحقق من افتراض اطرادية (Monotonicity) منحنيات خصائص فقرات مقياس قلق القراءة باللغة الانجليزية؛ فقد تم تقدير معالم فقرات المقياس باستخدام برمجية IRTPRO، حيث كانت قيم العتبات تتزايد وهذا مؤشر على اطراديه الفقرات عندما يكون التمييز موجب، حيث يلاحظ على منحنى خصائص الفقرة أن العتبة الثانية أكبر من العتبة الأولى والعتبة الثالثة أكبر من العتبة الثانية والعتبة الرابعة أكبر من العتبة الثالثة، كما يلاحظ بأن تلك المنحنيات تتقاطع بتسلسل هرمي وهو ما يشار له بالاطرادية.

ثانياً: التحقق من مطابقة الفقرات للنموذج (Item Fit).

تم التحقق من مطابقة الفقرات لنموذج الاستجابة المتدرجة من خلال مؤشر (χ^2) المستخدم في برمجية IRTPRO ،و بينت النتائج من خلال مؤشر (χ^2) ان جميع الفقرات مطابقة لنموذج الاستجابة المتدرج حيث كانت القيمة الاحتمالية لهذا الاختبار أكبر من (0.05) ولجميع فقرات المقياس.

ثالثاً: التحقق من مطابقة النموذج للبيانات.

تم حساب قيمة مؤشر جذر متوسط الأخطاء المعياري (RMSEA) (0.04) وتؤكد قيمته لمطابقة النموذج للبيانات وهو ما تؤكده الدلالة الإحصائية لاختبار M2 البالغة (238.04).

المعالجات الإحصائية

تمت المعالجة الإحصائية للبيانات في هذه الدراسة باستخدام برمجية R وبرمجية SPSS وبرمجية AMOS وبرمجية IRTPRO حيث تم:

للعينة الاستطلاعية:

- استخدام برمجية AMOS للتحقق من البناء العاملي للمقياس من خلال التحليل العالمي التوكيدي.
- استخدام برمجية SPSS للتحقق من أحادية البعد من خلال التحليل العاملي الاستكشافي.
- استخدام برمجية IRTPRO لتحليل البيانات وفق نظرية الاستجابة للفقرة بهدف التحقق من أحادية البعد وتقدير معالم الفقرات والافراد والتحقق من المطابقة.

للإجابة عن سؤال الدراسة

استخدام برمجية SPSS للتحقق من التوافق في المطابقة بين المؤشرين من خلال اختبار مربع كأي للتوافق ومعامل التوافق.

- التحقق من اختلاف معاملات توافق المطابقة عبر الظروف من خلال اختبار مربع كأي باستخدام برمجية R.

نتائج الدراسة ومناقشتها

– للإجابة عن سؤال الدراسة ، تم تقدير القيم الحرجة (القيم الحدية) لمؤشري Gp وU3p من خلال برمجية R لاستجابات الطلبة على المقياس ثلاثي ورباعي وخماسي التدريج ولكل من الصفين كما هو مبين في الجدول2:

جدول2:القيم الحديه لمؤشرات المطابقة حسب عدد فئات التدريج والمرحلة العمرية (الصف).

			عدد فئات التدريج		
المؤشر			3	4	5
Gp	الصف	السابع	189.00	300.00	545.15
		الأول ثانوي	177.00	310.00	537.15
U3p	الصف	السابع	.54	.45	.44
		الأول ثانوي	.48	.54	.54

يلاحظ أنّ القيم الحدية لمؤشر Gp تتزايد بازدياد عدد فئات التدريج بشكل واضح ،حيث كانت أصغر قيمة لدرجة القطع للتدريج الثلاثي وبلغت قيمتها (189.00) في ضوء الصف السابع ،وأكبر قيمة للتدريج الخماسي وبلغت قيمتها (545.15) في ضوء الصف السابع ، بينما كانت قيمه متقاربة في للصفين السابع والأول ثانوي في عبر فئات التدريج ، حيث كانت أصغر قيمة لدرجة القطع للتدريج الثلاثي وبلغت قيمتها (177.00) في ضوء الصف الأول ثانوي ،وأكبر قيمة للتدريج الخماسي وبلغت قيمتها (537.15) في عبر الصف الأول ثانوي ، بينما كانت قيمه متقاربة في للصفين السابع والأول ثانوي في عبر فئات التدريج .أما فيما يتعلق بالقيم الحدية لمؤشر U3p فقد كانت القيم متفاوتة في عبر عدد فئات التدريج والمراحل العمرية (الصف).

– تم تحديد التوافق بين المؤشرين في عملية الكشف عن مطابقة الافراد لنموذج الاستجابة المتدرجة من خلال معامل التوافق (Contingency coefficient) واختبار مربع كأي لكل من عدد فئات التدريج والمبينة في جدول 3.

جدول3:اختبار مربع كأي لفحص التوافق بين مؤشري Gp و U3p في الكشف عن الافراد غير المطابقين في ضوء عدد فئات التدريج

عدد فئات التدريج			حسب مؤشر U3p				اختبار مربع كأي				التوافق	
	مؤشر		غير مطابق	مطابق	المجموع		قيمة الاختبار	درجة الحرية	القيمة الاحتمالية		معامل التوافق	القيمة الاحتمالية
ثلاثي		غير	38	9	47							
	Gp	مطابق	10	463	473		316.327	1	0.000		0.615	0.000
		المجموع	48	472	520							
رباعي		غير	42	10	52							
	Gp	مطابق	10	448	458		314.438	1	0.000		0.618	0.000
		المجموع	52	458	510							
خماسي		غير	50	16	66							
	Gp	مطابق	5	436	441		330.555	1	0.000		0.628	0.000
		المجموع	55	452	507							

من نتائج اختبار مربع كأي يتبين وجود توافق مرتفع ودال احصائيا بين المؤشرين في الكشف عن الافراد غير المطابقين ولكل من عدد فئات التدريج، وهو ما يمكن ملاحظته من تركز الأفراد في خلايا التوافق في جدول التوافق حيث يلاحظ أن (38) من الأفراد كانوا غير مطابقين بالنسبة لأداء المؤشرين Gp و U3p و(463) من الأفراد كانوا مطابقين بالنسبة لأداء المؤشرين Gp و U3p وعدد الأفراد غير المطابقين بالنسبة لأداء مؤشر Gp (9) ولمؤشر U3p (10)بالنسبة للتدريج الثلاثي وكذلك الأمر بالنسبة للتدريج الرباعي والخماسي .كما يلاحظ اختلاف قيم معاملات التوافق حسب عدد فئات التدريج حيث كانت للتدريج الثلاثي (0.615) اما التدريج الرباعي فكانت قيمة معامل التوافق (0.618) في حين كانت للتدريج الخماسي (0.628). ولفحص دلالة الاختلاف في معاملات التوافق تم استخدام اختبار مربع كأي والمبينة نتائجه في جدول 4.

جدول 4:

فحص الفروق في معاملات ارتباط مؤشري Gp و U3p باختلاف عدد فئات التدريج.

قيمة اختبار مربع كأي	درجة الحرية	الدلالة
0.125	2	0.939

من نتائج اختبار مربع كأي لفحص الاختلاف في معاملات الارتباط يتبين عدم وجود اختلاف في التوافق بين مؤشري مطابقة الفرد باختلاف عدد فئات التدريج أي أن درجة التوافق بين المؤشرين عالية وهناك تقارب بين المؤشرين في الكشف عن الاستجابات غير المطابقة باختلاف عدد فئات التدريج .

-تم تحديد التوافق بين المؤشرين في عملية الكشف عن مطابقة الأفراد لنموذج الاستجابة المتدرجة من خلال معامل التوافق (Contingency coefficient) واختبار مربع كأي لكل من الصفين السابع والأول ثانوي والمبينة في جدول5.

جدول 5:اختبار مربع كأي لفحص التوافق بين مؤشري Gp و U3p في الكشف عن الأفراد غير المطابقين في ضوء المرحلة العمرية (الصف).

الصف			حسب مؤشر U3p				اختبار مربع كأي			التوافق	
	مؤشر		غير	مطابِ	المجمو		قيمة	درجة	القيمة	معامل	القيمة
السابع		غير	71	17	88						
	Gp	مطابق	12	681	693		512.437	1	0.000	0.62	0.000
		المجموع	83	698	781						
الأول		غير	59	18	77						
ثانوي	Gp	مطابق	13	666	679		447.958	1	0.000	0.61	0.000
		المجموع	72	684	756						

من نتائج اختبار مربع كأي يتبين وجود توافق مرتفع ودال احصائيا بين المؤشرين في الكشف عن الأفراد غير المطابقين في كل من الصفين السابع والأول ثانوي، وهو ما يمكن ملاحظته من تركز الأفراد في خلايا التوافق في جدول التوافق حيث يلاحظ أن (71) من الأفراد كانوا غير مطابقين بالنسبة لأداء المؤشرين Gp و U3p و(681) من الأفراد كانوا مطابقين بالنسبة لأداء المؤشرين Gp و U3p وعدد الأفراد غير المطابقين بالنسبة لأداء

مؤشر Gp (17) ولمؤشر U3p (12)بالنسبة للصف السابع وكذلك الأمر بالنسبة للصف الأول ثانوي. كما يلاحظ اختلاف قيم معاملات التوافق حسب المرحلة العمرية (الصف) حيث كان معامل التوافق للصف السابع (0.629) بينما للصف الأول ثانوي كانت قيمته (0.610). ولفحص دلالة الاختلاف في معاملات التوافق تم استخدام اختبار مربع كأي والمبينة نتائجه في جدول 6.

جدول 6: فحص الفروق في معاملات ارتباط مؤشري Gp و U3p باختلاف الصف.

قيمة اختبار مربع كأي	درجة الحرية	الدلالة
0.364	1	0.546

من نتائج اختبار مربع كأي لفحص الاختلاف في معاملات الارتباط يتبين عدم وجود اختلاف في التوافق بين مؤشري مطابقة الفرد باختلاف المرحلة العمرية (الصف) أي أن درجة التوافق بين المؤشرين عالية وهناك تقارب بين المؤشرين في الكشف عن الاستجابات غير المطابقة باختلاف المرحلة العمرية.

توصلت النتائج المتعلقة بهذا السؤال إلى أن المؤشرين Gp و U3p متوافقين بدرجة مرتفعة ولكل من عدد فئات التدريج. وان درجة التوافق في المؤشرين لا تختلف باختلاف عدد فئات التدريج. ويمكن تفسير ذلك من خلال اعتماد كل من المؤشرين على نمط الاستجابة والتراكمية هذا من جهة ومن جهة أخرى فان المؤشرين يتأثران بعدد فئات التدريج بشكل متواز وهو ما حافظ على نفس المستوى من التوافق. كما توصلت النتائج المتعلقة بهذا السؤال إلى ان المؤشرين Gp و U3p متوافقين بدرجة مرتفعة ولكل من الفئات العمرية. وان درجة التوافق في المؤشرين لا تختلف باختلاف الفئة العمرية. ويمكن تفسير ذلك من خلال اعتماد كل من المؤشرين على نمط الاستجابة بدلا من عينة الأفراد فالفرد ينظر له كنمط استجابة لا كفئة عمرية.

وتتفق نتيجة هذا السؤال مع نتائج دراسة روندر(Rudner, 1983) من جهة في التوافق بين مجموعة من المؤشرات وجد أن بينها ترابط مرتفع وتوافق كبير في الكشف عن أنماط الاستجابة غير المطابقة ومن جهة أخرى تختلف نتيجة هذا السؤال مع نتائج دراسة روندر لمجموعة أخرى من المؤشرات التي وجد أنها أقل ترابطا والتوافق بينها كان غير عال في الكشف عن أنماط الاستجابة غير المطابقة ومن ضمنها المؤشر اللامعلمي U3p ، كما تتفق هذه النتيجة مع الدراسة التي أجراها

سينجل افسر (Şengül AvŞar, 2019) حيث وجد أن المؤشرات اللامعلمية U3p و GNp كانت نتائجها قريبة جدا من بعضها أي أنّ التوافق بين أدائها مرتفع .

التوصيات:

- الاعتماد على نتائج عدم مطابقة الأفراد بالتوافق بين مؤشري Gpو U3p.
- إجراء مقارنة بين المؤشرين في مدى التوافق مع المؤشرات المعلمية مثل مؤشر Zh.
- مقارنة أداء المؤشرات Gp و U3p والتوافق بينها في ضوء حجم العينة وعدد فقرات المقياس.
- استخدام مؤشر واحد من Gp و U3p يغني عن الاخر بسبب درجة التوافق المرتفع بينهم

المراجع العربية :

علام، صلاح الدين. (1986). تطورات معاصرة في القياس النفسي والتربوي. الكويت: كلية الآداب.

عطروز، ناديه. (2018). أثر برنامج تدريسي مستند إلى استراتيجيات القراءة ما وراء المعرفية على قلق القراءة ودافعيتها لدى طالبات الصف الثامن في اللغة الإنجليزية . *المجلة الأردنية في العلوم التربوية*، مجلد 15 ،عدد 3 ،2019، 341.

المراجع الاجنبية :

Birenbaum, M. (1985). Comparing the Effectiveness of Several Irt Based Appropriateness Measures in Detecting Unusual Response Patterns. *Educational and Psychological Measurement, 45(3), 523–534*

Boeck, P. and Wilson, M. (2004). Descriptive and Explanatory Item Response Models. In Paul De Boeck, Mark Wilson. (Eds.), Explanatory Item Response Models: *A Generalized Linear and Nonlinear Approach*, 343- 373, New York: Springer-Verlag014662161879866. doi:10.1177/0146621618798666

Cavanagh, R. F., & Romanoski, J. T. (2006). Rating scale instruments and measurement. *Learning Environments Research*, 9(3), 273-289.

De La Torre, J., & Deng, W. (2008). Improving Person-Fit Assessment by Correcting the Ability Estimate and Its Reference Distribution. *Journal of Educational Measurement*, 45(2), 159-177.

Emons, W. H. M. (2008). Nonparametric person-fit analysis of polytomous item scores. *Applied Psychological Measurement*, 32(3), 224–247. https://doi.org/10.1177/0146621607302479

Garson, G. D. (2009). Structural Equation Modeling. Retrieved on 5th October 2014 from http://faculty.chass.ncsu.edu/garson/PA765/structur.htm.

Hambleton , R.K., & Swaminathan, H. (1985). Item ResponseTheory : Principles and applications. Bosten, MA: kluwer-Nijhoff .Publishing

Hambleton , R.K., Swaminathan, H., & Rogers, H. J. (1991).(Fundamentals of item response theory (Vol.2).Sage

Hemker, B. T., Sijtsma, K., Molenaar, I. W., & Junker, B. W. (1996). Polytomous IRT models and monotone likelihood ratio of the total score. Psychometrika, 61,679-693.

Hui-Ju, W. (2011). Anxiety and reading comprehension performance in English as a foreign language. *Asia EFL Jornal*,13(2)273- 307.

Junker, B & Sijtsma, K. (2001). Cognitive Assessment Models With Few Assumptions & Connections With Nonparametric Item Response Theory, *Applied Psychological Measurement*, 25, 258-272

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

Molenaar, I. W. (1991). A weighted loevinger h-coefficient extending mokken scaling to multicategory items. *Psychologische Instituten der Rijksuniversiteit Groningen*

Meijer, R. R., Molenaar, L. 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. (1995). Detection of aberrant item score patterns: A review of recent developments. *Applied Measurement in Education*, 8(3), 261-272.

Meijer, R. R. (1996). Person-Fit Research: An Introduction. *Applied Measurement in Education, 9(1), 3–8*doi:10.1207/s15324818ame0901_2

Meijer, R. R., Muijtjens, A. M. M., & van der Vlueten, C. P. M. (1996). Nonparametric Person-Fit Research: Some Theoretical issues an Empirical Example. *Applied Measurement in Education, 9(1), 77–89.* doi:10.1207/s15324818ame0901_7

Meijer, R. R. & Sijtsma, K. (1999). A review of methods for evaluating the fit of item score patterns on a test (No. 99-01). The Netherlands: University of Twente.

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

Meijer, R. R., & Tendeiro, J. N. (2014). The Use of Person-Fit Scores in High-Stakes Educational Testing: How to Use Them and What They Tell Us. LSAC RESEARCH REPORT SERIES, pp 1-22.

Nering, M. L., & Ostini, R. (Eds.). (2011). *Handbook of polytomous item response theory models. Taylor & Francis*

Rudner, L. M. (1983). Individual assessment accuracy. *Journal of Educational Measurement,* 20(3), 207–219. JSTOR

Reise, S. P. & Flannery, W. P. (1996). Assessing person-fit on measures of typical performance. *Applied "Psychological Measurement,* 9(1), 9-26.

Reise, S. P., & Yu, J. (1990). Parameter recovery in the graded response model using MULTILOG. *Journal of educational Measurement,* 27(2), 133-144.

Rauch, W. A., Schweizer, K., & Moosbrugger, H. (2008). An IRT analysis of the personal optimism scale. *European Journal of Psychological Assessment,* 24(1), 49-56

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph Supplement,* 17, 1- 100.

Samejima, F. (1997).*Graded response model. In handbook of modern item response theory* (PP.85-100). Springer New York

Sijtsma, K. (1998). Methodology review: Nonparametric IRT approaches to the analysis of dichotomous item scores. *Applied Psychological Measurement,* 22(1), 3-31.

Schreiber, J. B., Nora, A., Stage, F. K., Barlow, E. A., & King, J. (2006). Reporting structural equation modeling and confirmatory factor analysis results: A review. The Journal of educational research, 99(6), 323-338.

Sijtsma, K., Emons, W. H., Bouwmeester, S., Nyklícek, I., & Roorda, L. D. (2008). Nonparametric IRT analysis of quality of life scales and its application to the

world health organization quality of life scale (Whoqol-Bref). Quality of Life Research: An International Journal of Quality of Life Aspects of Treatment, Care and Rehabilitation, 17(2), 275–290. http://dx.doi.org/10.1007/ s11136-007-9281-6

Şengül Avşar, A., & Tavşancıl, E. (2017). Examination of polytomous items' psychometric properties according to nonparametric item response theory models in different test conditions. *Educational Sciences: Theory & Practice*, 17(2). doi: 10.12738/estp.2017.2.0246

Tang, K. L., & Eignor, D. R. (1997). Concurrent calibration of dichotomously and

polytomously scored TOEFL items using IRT models. *ETS Research Report Series*, 1997(1), i-98

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

Turner, M. L. (2018). The detection and impact of low cognitive effort survey responses [Ph.D., University of Colorado at Boulder].

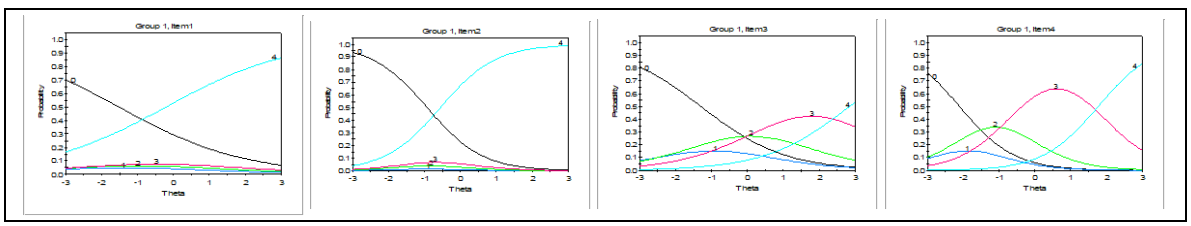
Ullman, J. B. (2001). Structural equation modeling. In B. G. Tabachnick & L. S. Fidell (2001). Using Multivariate Statistics (4th ed& pp 653- 771). Needham Heights, MA: Allyn & Bacon.

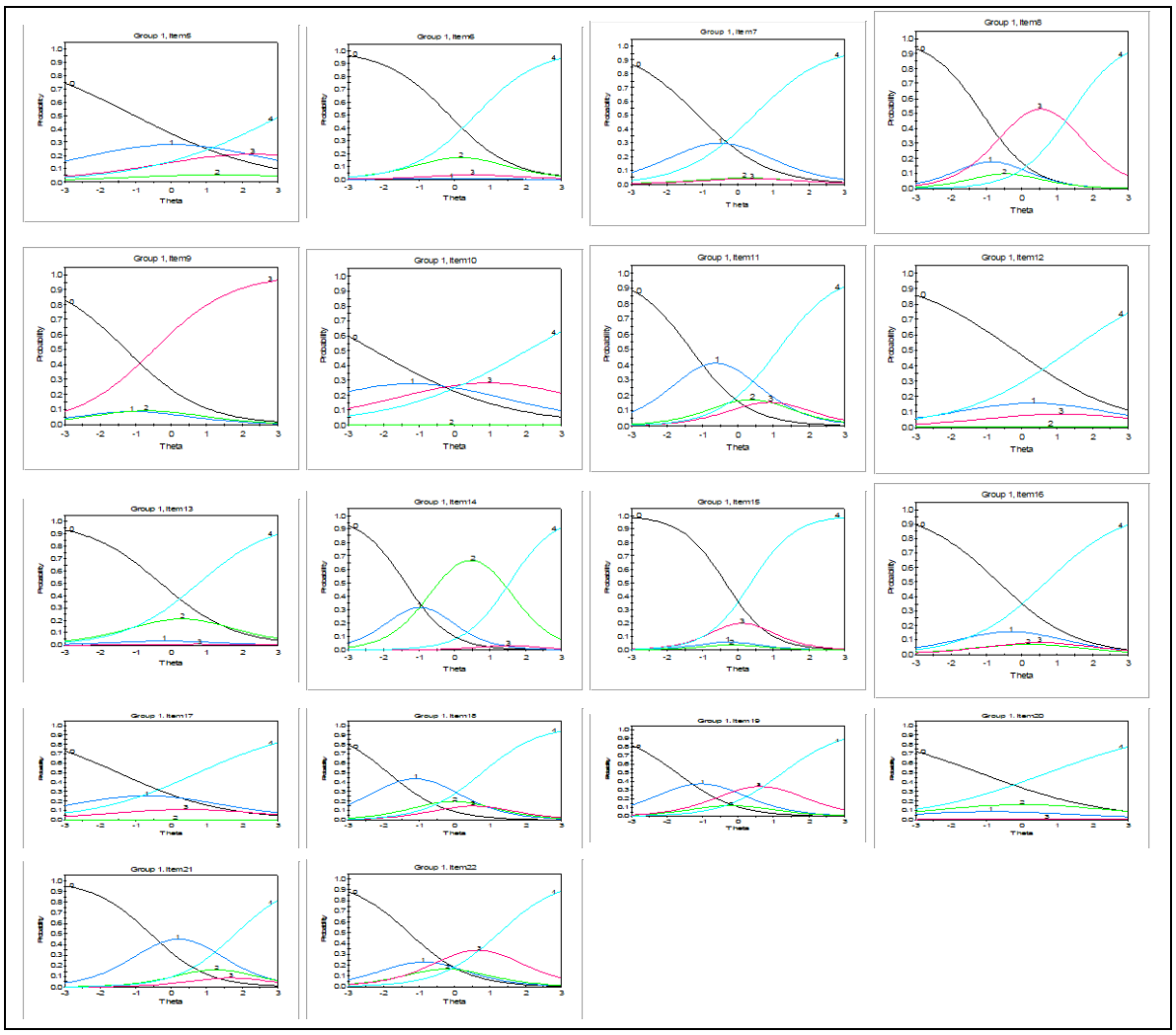
Van der Flier, H. (1980). Vergelijkbaarheid van individuele testprestatie [Comparability of individual test performance]. Lisse, The Netherlands: Swets & Zeitlinge

الملاحق

ملحق أ

منحنى خصائص الفقرة لمقياس قلق القراءة باللغة الانجليزية بصورته الأصلية (التدريج الخماسي) .




Group 1, Item5
Group 1, Item6
Group 1, Item7
Group 1, Item8
Group 1, Item9
Group 1, Item10
Group 1, Item11
Group 1, Item12
Group 1, Item13
Group 1, Item14
Group 1, Item15
Group 1, Item16
Group 1, Item17
Group 1, Item18
Group 1, Item19
Group 1, Item20
Group 1, Item21
Group 1, Item22
Probability
Theta