المقارنة بين النموذجين ثلاثي ورباعي المعلمة وفقاً لنظرية الاستجابة للمفردة في تدريج اختبار تحصيلي

إعداد

د. أحمد سمير مجاهد أبو الحسن

أستاذ علم النفس التربوي المساعد

كلية التربية- جامعة الزقازيق

د. إيمان إبراهيم محمد سليم نافع

مدرس علم النفس التربوي

كلية التربية- جامعة الزقازيق

المقارنة بين النموذجين ثلاثي ورباعي المعلمة وفقاً لنظرية الاستجابة للمفردة في تدريج اختبار تحصيلي

د. أحمد سمير مجاهد أبو الحسن د. إيمان إبراهيم محمد سليم نافع

الملخص:

هدف البحث إلى تقييم مدى ملاءمة النموذج ثلاثي المعلمة (3PL) والنموذج رباعي المعلمة (4PL) كنموذجين من نماذج نظرية الاستجابة للمفردة (IRT) لاختبار تحصيلي في أحد المقررات العامة. تم تحليل بيانات الاختبار لـ ١١٦٨ طالباً من طلاب الفرقة الثانية كلية التربية جامعة الزقازيق باستخدام مؤشرات ملاءمة المفردات، وملاءمة النموذج الكلية، بالإضافة إلى فحص ملاءمة أنماط استجابات الأفراد باستخدام إحصاءة (zh). كما تم تقدير قدرة الأفراد باستخدام عدة طرق (MLE, MAP, EAP) ومقارنة دقة القياس، وأخيراً، تم تقييم دالة معلومات الاختبار الكلية (TIF) والخطأ المعياري في القياس (SEM)، والثبات التجريبي والهامشي. وأظهرت النتائج أن النموذج ثلاثي المعلمة حقق ملاءمة أفضل للمفردات بشكل فردي وللنموذج ككل مقارنة بالنموذج رباعي المعلمة، حيث تفوق في غالبية مؤشرات الملاءمة الكلية (AIC, BIC, SRMSR, logLik). وعلى الرغم من أن كلا النموذجين أظهرا ملاءمة عالية لأنماط استجابات الأفراد، إلا أن النموذج رباعي المعلمة أظهر عددًا أقل قليلاً من الأفراد ذوي أنماط الاستجابة غير الملائمة. وبالنسبة لتقدير قدرة الأفراد، أثبتت طريقتي MAP وEAP دقة قياس عالية جداً لكلا النموذجين وتفوقتا على طريقة MLE التي واجهت مشكلات في الاستقرار. كما لوحظ أن النموذج رباعي المعلمة يميل إلى تقدير متوسط قدرة أعلى للأفراد مقارنة بالنموذج ثلاثي المعلمة. وفي جانب معلومات الاختبار والثبات، قدم النموذج ثلاثي المعلمة دالة معلومات اختبار أعلى في نطاق أوسع من القدرة، وحقق قيم ثبات تجريبي وهامشي أعلى بشكل ملحوظ مقارنة بالنموذج رباعي المعلمة. وتُشير هذه النتائج إلى أن التعقيد الإضافي للنموذج رباعي المعلمة لم يُقدم تحسينات جوهرية تُبرر استخدامه في هذا الاختبار، بل أثر سلباً على تقديرات الثبات.

الكلمات المفتاحية: نظرية الاستجابة للمفردة، النموذج ثلاثي المعلمة (3PL)، النموذج رباعي المعلمة (4PL)، ملاءمة المفردة، ملاءمة النموذج، ملاءمة الأفراد، دالة معلومات الاختبار، الثبات الهامشي والتجريبي.

Comparing Three- and Four-Parameter Models in Achievement Test Scaling Using Item Response Theory

Abstract: This study aimed to evaluate the fit of the Three-Parameter Logistic Model (3PLM) and Four-Parameter Logistic Model (4PLM) models of Item Response Theory (IRT) for an achievement test in a general course. Test data from 1168 second-year students at the Faculty of Education, Zagazig University, were analyzed using item fit indices, and overall model fit indices. person fit was also examined using the zh statistic. Furthermore, person ability was estimated using Maximum Likelihood Estimation (MLE), Maximum A Posteriori (MAP), and Expected A Posteriori (EAP) methods, with a comparison of measurement precision. Finally, the overall Test Information Function (TIF), Standard Error of Measurement (SEM), and empirical and marginal reliability were assessed. Results indicated that the 3PL model achieved a better fit for individual items and the overall model compared to the 4PL model, excelling in most global fit indices (AIC, BIC, SRMSR, logLik). Although both models showed high fit for individual response patterns, the 4PL model identified slightly fewer individuals with misfit response patterns. For person ability estimation, both MAP and EAP methods demonstrated very high and identical measurement precision for both models, outperforming the unstable MLE method. It was also observed that the 4PL model tended to estimate a higher average person ability compared to the 3PL model. Regarding test information and reliability, the 3PL model provided a higher test information function across a wider range of ability and achieved significantly higher empirical and marginal reliability values than the 4PL model. These findings suggest that the added complexity of the 4PL model did not yield substantial improvements to justify its use for this test, and in fact, negatively impacted reliability estimates.

Keywords: Item Response Theory (IRT), Three-Parameter Logistic Model (3PLM), Four-Parameter Logistic Model (4PLM), Item Fit, Model Fit, Person Fit, Test Information Function, Marginal Reliability, empirical reliability.

مقدمة:

تُعد الاختبارات التحصيلية أداة محورية في العملية التعليمية، فهي لا تُمكن من تقييم مدى اكتساب الطلاب للمعرفة والمهارات فحسب، بل تُقدم أيضاً معلومات قيّمة للمربين والمخططين التربويين لتحسين المناهج وطرق التدريس. ولضمان دقة وفاعلية هذه الاختبارات، يكتسب التحليل السيكومتري لخصائص المفردات والاختبار ككل أهمية قصوى. في هذا السياق، برزت نظرية الاستجابة للمفردة (Item Response Theory- IRT) كإطار إحصائي متقدم يُتيح تقديراً أكثر دقة لخصائص المفردات وقدرات الأفراد، متجاوزةً بذلك بعض القيود التي تفرضها النظرية الكلاسيكية للاختبار.(Classical Test Theory- CTT)

تُقدم نماذج IRT المختلفة، مثل النموذج ثلاثي المعلمة (3PL) والنموذج رباعي المعلمة(4PL)، مستويات متفاوتة من التعقيد في نمذجة العلاقة بين قدرة الفرد واحتمالية استجابته الصحيحة لمفردة معينة. يُضيف النموذج ثلاثي المعلمة معلمة التخمين (g) إلى معلمتي التمييز (a) والصعوبة(b) ، مما يجعله مناسباً للمفردات متعددة الخيارات حيث يُمكن للتخمين أن يؤثر على الأداء (محمد محمد جاد محمد, ٢٠١٦). في المقابل، يُوسع النموذج رباعي المعلمة هذا الإطار بإضافة معلمة رابعة تُعرف باسم "الإهمال" أو "الحد الأقصى للأداء(u) "، والتي تحاول نمذجة احتمال عدم وصول الأداء إلى الكمال حتى بالنسبة للأفراد ذوي القدرات العالية جداً، وذلك بسبب عوامل مثل عدم الانتباه أو أخطاء الإهمال (Loken & Rulison, 2010; Pardede et al., 2023)

مع تعدد هذه النماذج، يبرز التحدي البحثي في تحديد النموذج الأكثر ملاءمة للبيانات الفعلية لاختبار معين. فالاختيار غير المناسب للنموذج قد يُفضي إلى تقديرات غير دقيقة لمعلمات المفردات وقدرات الأفراد، وبالتالي يؤثر على دقة القياس وثباته (DoĞRuÖZ & Akin Arikan, 2020) تتضارب نتائج الدراسات السابقة حول تفضيل نموذج على آخر، حيث تُشير بعضها إلى تفوق النماذج الأكثر بساطة مثل النموذج ثلاثي المعلمة في سياقات معينة (Büyükkıdık & İnal, 2023)، بينما تُفضل أخرى النموذج رباعي المعلمة لنمذجته سلوكيات الاستجابة المعقدة، خاصةً عند الأطراف العليا من منحنى القدرة (Erdemir & Önen, 2019; Yalçın, 2018)؛ هذا التباين يُبرز ضرورة إجراء دراسات تقييمية تُقارن أداء هذه النماذج على بيانات اختبارات محددة.

يستهدف هذا البحث سد هذه الفجوة من خلال فحص مدى ملاءمة النموذج الثلاثي

والرباعي المعلمة لاختبار تحصيلي في أحد فصول مادة الاتجاهات الحديثة والمعاصرة في علم النفس.

مشكلة البحث وأسئلته:

انطلاقًا من الأهمية المتزايدة للتحليل السيكومتري الحديث في تقييم الأدوات التربوية، وتحديداً استخدام نماذج نظرية الاستجابة للمفردة، يُركز هذا البحث على اختبار تحصيلي في أحد فصول مادة الاتجاهات الحديثة والمعاصرة في علم النفس. ولضمان أن النتائج المستخلصة من هذا الاختبار دقيقة وموثوقة وقابلة للتفسير بشكل سليم، من الضروري إجراء تحليل سيكومتري شامل يُحدد النموذج الإحصائي الأمثل الذي يصف خصائص المفردات وأنماط استجابات الأفراد على هذا الاختبار.

يُعد اختيار النموذج الأنسب من بين النماذج المتوفرة في نظرية الاستجابة للمفردة خطوة مهمة، حيث تتضمن هذه النماذج فروضاً مختلفة حول كيفية تفاعل القدرة الكامنة للفرد مع خصائص المفردات. فالنماذج الأكثر بساطة قد تكون كافية إذا كانت البيانات لا تُبرر النماذج الأكثر تعقيداً، بينما قد تُقدم النماذج الأكثر تعقيداً وصفاً أكثر شمولية لأنماط الاستجابة غير العادية. إن عدم تحديد النموذج الأمثل قد يؤدي إلى تقديرات متحيزة لمعلمات المفردات وقدرات الأفراد؛ وبالتالي يُقلل من دقة القياس الكلية للاختبار وثباته. في ظل التباين في نتائج الدراسات السابقة حول هذا الموضوع، يصبح من الضروري تقييم أداء كل من النموذج ثلاثي المعلمة والنموذج رباعي المعلمة على بيانات الاختبار التحصيلي المعد.

تقدم نماذج نظرية الاستجابة للمفردة (IRT)، كالنموذج ثلاثي المعلمة (3PL) ورباعي المعلمة (4PL)، مستويات متفاوتة لنمذجة العلاقة بين قدرة الفرد واستجابته الصحيحة. يضيف نموذج 3PL معلمة التخمين (g) لمعلمتي التمييز والصعوبة، مما يناسب الاختبارات متعددة الخيارات. بينما يوسع نموذج 4PL هذا بإضافة معلمة الإهمال (u)، لتمثيل احتمال عدم الوصول للأداء المثالي حتى لذوي القدرات العالية جدًا بسبب الإهمال أو عدم الانتباه (Loken & Rulison, 2010; Pardede et al., 2023)

التحدي البحثي يكمن في اختيار النموذج الأنسب لبيانات اختبار معين؛ فالخطأ في الاختيار قد يؤدي لتقديرات غير دقيقة لقدرات الأفراد ومعلمات المفردات، مؤثرًا على دقة القياس وثباته. تتباين الدراسات السابقة في تفضيل نموذج على آخر، فبعضها يفضل 3PL لبساطته،

وبعضها يفضل 4PL لنمذجته السلوكيات المعقدة، خاصة في مستويات القدرة العليا. هذا التباين يؤكد أهمية إجراء دراسات تقييمية لمقارنة أداء هذه النماذج على بيانات اختبارات محددة (Büyükkıdık & İnal, 2023) & (Erdemir & Önen, 2019; Yalçın, 2018) & (DoĞRuÖZ & Akin Arikan, 2020).

وعليه، يسعى البحث لمعرفة أفضلية النموذج ثلاثي المعلمة أو النموذج رباعي المعلمة في ملاءمة مفردات الاختبار التحصيلي المعد، وأيهما يُحقق ملاءمة أفضل للنموذج ككل؟، ومدى ملاءمة أنماط استجابات الأفراد لكل نموذج، ومدى اختلاف تقديرات قدرة الأفراد ودقة القياس بين النموذجين وطرق التقدير المختلفة، وأخيراً، مدى تأثير اختيار النموذج على دالة معلومات الاختبار الكلية ودقة القياس وثباته، وتهدف هذه الأسئلة إلى تقديم رؤية شاملة تُسهم في فهم أعمق لخصائص الاختبار التحصيلي المعد وتُوجه الممارسات المستقبلية في التحليل السيكومتري للاختبارات التي سيتم بناؤها لاحقاً.

وعليه يمكن تحديد مشكلة الدراسة في الأسئلة التالية:

السؤال الأول: " ما أفضل نموذج (النموذج ثلاثي المعلم– النموذج رباعي المعلم) والذي يحقق ملاءمة لمفردات الاختبار وفقاً لمحك المقارنة (the signed chi-squared (s-χ2) statistics)؟ "

السؤال الثاني: " ما أفضل نموذج (النموذج الثلاثي المعلم –النموذج الرباعي المعلم) والذي يحقق ملاءمة بشكل عام للاختبار؟ "

السؤال الثالث: "ما مدى ملاءمة نمط استجابات الأفراد لمفردات الاختبار في ضوء كل من النموذج الثلاثي المعلم والنموذج الرباعي المعلم وفقاً للمحك الإحصائي (zh statistics)؟"

السؤال الرابع: "هل تختلف منحنيات خصائص المفردة (Item Characteristic Curves - ICCs) ودوال معلومات المفردة (Item Information Functions - IIFs) وكذلك دقة القياس لكل مفردة من مفردات الاختبار في ضوء كل من النموذج الثلاثي المعلم والنموذج الرباعي المعلم؟"

السؤال الخامس: "هل يختلف تقدير قدرة (معلمة) الأفراد باستخدام طرق التقدير (maximum likelihood (ML), maximum a posteriori (MAP), and expected a posteriori (EAP)) وكذلك دقة القياس في تقدير قدرة الأفراد، في ضوء كل من النموذج

الثلاثي المعلم والنموذج الرباعي المعلم للاختبار؟"

السؤال السادس: "هل تختلف دالة معلومات الاختبار الكلية (Test Information Function - TIF) والخطأ المعياري في القياس (Standard Error of Measurement - SEM)، إلى جانب تقديرات الثبات التجريبي والهامشي (Marginal Reliability)، في ضوء كل من النموذج الثلاثي المعلم والنموذج الرباعي المعلم للاختبار؟"

أهداف البحث:

يهدف البحث الحالي إلى:

١. تحديد أفضل نموذج (النموذج ثلاثي المعلمة – النموذج رباعي المعلمة) الذي يحقق ملاءمة لمفردات الاختبار التحصيلي وفقاً لمحك الملاءمة الإحصائية (the signed chi-squared (s-χ2) statistics).

٢. الكشف عن أفضل نموذج (النموذج الثلاثي المعلم – النموذج الرباعي المعلم) الذي يحقق ملاءمة بشكل عام للاختبار التحصيلي ككل.

٣. التحقق من مدى ملاءمة نمط استجابات الأفراد لمفردات الاختبار في ضوء كل من النموذج الثلاثي والرباعي المعلم وفقاً للمحك الإحصائي (zh statistics).

٤. مقارنة منحنيات خصائص المفردة (ICCs) ودوال معلومات المفردة (IIFs) وكذلك دقة القياس لكل مفردة من مفردات الاختبار في ضوء كل من النموذج الثلاثي والرباعي المعلم.

٥. مقارنة تقديرات قدرة (معلمة) الأفراد باستخدام طرق التقدير المختلفة (maximum likelihood (ML)، maximum a posteriori (MAP)، و expected a posteriori (EAP)) وكذلك دقة القياس في تقدير قدرة الأفراد، في ضوء كل من النموذج الثلاثي والرباعي المعلم للاختبار.

٦. مقارنة دالة معلومات الاختبار الكلية (TIF) والخطأ المعياري في القياس (SEM)، إلى جانب تقديرات الثبات التجريبي والهامشي (Marginal Reliability)، في ضوء كل من النموذج الثلاثي والرباعي المعلم للاختبار.

أهمية البحث:

أولاً: الأهمية النظرية:

١. إثراء الأدبيات العلمية : يُعد هذا البحث إضافة نوعية للأدبيات المتخصصة في مجال القياس النفسي والتربوي، وخصوصًا في ميدان نظرية الاستجابة للمفردة (IRT)، من خلال تقديم مقارنة معمقة ومنهجية بين النموذجين ثلاثي ورباعي المعلمة.

٢. تعميق الفهم النظري للنماذج : يسهم البحث في تعميق الفهم النظري لخصائص وتطبيقات كل من النموذجين ثلاثي ورباعي المعلمة، وتسليط الضوء على الفروق الدقيقة بينهما، مما يدعم اتخاذ القرارات المستنيرة عند اختيار النموذج الأنسب لتدريج الاختبارات.

٣. تسليط الضوء على مؤشرات الملاءمة : يبرز البحث أهمية استخدام محكات إحصائية (مثل $s-\chi 2$ و zh statistics) لتقييم ملاءمة المفردات وأنماط استجابات الأفراد؛ مما يعزز من الممارسات الجيدة في تقييم جودة النماذج.

ثانياً: الأهمية التطبيقية:

١. دعم صانعي الاختبارات والمتخصصين في القياس: يقدم البحث دليلاً عملياً للمختصين في بناء وتدريج الاختبارات التحصيلية؛ حيث يمكنهم الاستفادة من النتائج في اختيار النموذج الأكثر كفاءة ودقة لبياناتهم؛ مما يؤدي إلى اختبارات ذات جودة أعلى.

٢. تحسين دقة القياس: من خلال تحديد النموذج الأفضل، يسهم البحث في تحسين دقة تقدير قدرات الأفراد وتوفير معلومات أكثر موثوقية حول أداء الطلاب، وهو أمر مهم في عمليات التقييم والتشخيص واتخاذ القرارات التعليمية.

٣. تعزيز كفاءة أدوات القياس: يمكن لنتائج البحث أن تساعد في بناء اختبارات تحصيلية تتمتع بدالة معلومات اختبار (TIF) أفضل وخطأ معياري في القياس (SEM) أقل؛ مما يضمن قياسًا أكثر فعالية وكفاءة عبر مختلف مستويات القدرة.

٤. توجيه الدراسات المستقبلية: تفتح نتائج هذا البحث آفاقًا لدراسات مستقبلية لاستكشاف جوانب أخرى من المقارنات بين نماذج IRT، أو تطبيق هذه النماذج في سياقات تعليمية مختلفة، أو باستخدام أنواع أخرى من الاختبارات.

مصطلحات البحث:

سيتضمن هذا البحث تعريفاً لعدد من المصطلحات الأساسية وهي:

نظرية الاستجابة للمفردة (Item Response Theory - IRT): هي إطار نظري للقياس النفسي والتربوي يصف العلاقة الاحتمالية بين استجابة الفرد على المفردة وقدرته الكامنة (السمة الكامنة). تفترض أن أداء الفرد على المفردة يعتمد فقط على قدرته وخصائص المفردة، وتوفر تقديرات للقدرة مستقلة عن المفردات المستخدمة (Embretson & Reise, 2013)، والتعريف الإجرائي عبارة عن الإطار الإحصائي المستخدم لنمذجة استجابات الأفراد على مفردات الاختبار التحصيلي قيد البحث، ويُطبق من خلال النموذجين ثلاثي ورباعي المعلمة لتقدير معالم المفردات وقدرات الأفراد.

النموذج الثلاثي المعلمة (Three-Parameter Logistic Model - 3PL): هو أحد نماذج نظرية الاستجابة للمفردة الذي يفترض أن احتمال إجابة الفرد الصحيحة على المفردة يعتمد على ثلاث معالم: معلمة الصعوبة (b)، معلمة التمييز (a)، ومعلمة التخمين العشوائي (c) (Petersen, 2024) & (DoĞRuÖZ & Akin Arikan, 2020)، والتعريف الإجرائي عبارة عن النموذج الإحصائي الذي سيتم تطبيقه على بيانات الاختبار التحصيلي لتقدير معالم الصعوبة والتمييز والتخمين لكل مفردة، وسيتم مقارنته بالنموذج رباعي المعلمة.

النموذج الرباعي المعلمة (Four-Parameter Logistic Model - 4PL): هو نموذج متقدم من نماذج نظرية الاستجابة للمفردة، يضيف معلمة رابعة (الحد الأقصى للإجابة الصحيحة، أو عدم اكتمال الإجابة أو معلمة الإهمال أو اللامبالاة – d) إلى معالم النموذج الثلاثي المعلمة (الصعوبة، التمييز، التخمين). يُستخدم هذا النموذج عندما يُفترض أن الأداء الأقصى على المفردة قد لا يصل إلى ١٠٠٪ حتى مع القدرة العالية جداً (Battauz, 2020, p. 296) & (DoĞRuÖZ & Akin Arikan, 2020)، والتعريف الإجرائي عبارة عن النموذج الإحصائي الذي سيتم تطبيقه على بيانات الاختبار التحصيلي لتقدير معالم الصعوبة والتمييز والتخمين والإهمال لكل مفردة، وسيتم مقارنته بالنموذج ثلاثي المعلمة

الاختبارات التحصيلية (Achievement Tests): هي أدوات قياس مصممة لتقييم مدى اكتساب الفرد للمعارف والمهارات في مجال تعليمي محدد بعد فترة تعلم معينة (Embretson & Reise, 2013)، والتعريف الإجرائي عبارة عن الاختبار الذي ستُطبق عليه نماذج نظرية

الاستجابة للمفردة (ثلاثي ورباعي المعلمة) لتقدير معالم مفرداته وقدرات الأفراد (وهو اختبار تحصيلي في أحد فصول مادة الاتجاهات الحديثة والمعاصرة في علم النفس).

ملاءمة المفردة (Item Fit): تشير إلى مدى توافق الأداء الفعلي للأفراد على مفردة معينة مع ما يتوقعه النموذج النظري لتلك المفردة (Chalmers, 2012)، والتعريف الإجرائي عبارة عن قياسها باستخدام المحك الإحصائي (the signed chi-squared (s-χ2) statistics)، وستشير إلى مدى اتفاق بيانات المفردة مع خصائص المفردة المقدرة من قبل النموذج.

ملاءمة النموذج (Model Fit): تشير إلى مدى توافق البيانات الكلية (بيانات الاختبار) مع الافتراضات الكلية للنموذج الإحصائي المطبق (Chalmers, 2012)، والتعريف الإجرائي عبارة عن تقييمها بشكل عام للاختبار ككل، وتُستخدم مؤشرات إحصائية لتقييمها (مثل AIC، BIC).

منحنى خصائص المفردة (Item Characteristic Curve - ICC): رسم بياني يوضح العلاقة الاحتمالية بين مستوى قدرة الفرد (السمة الكامنة) واحتمالية إجابته الصحيحة على مفردة معينة (Embretson & Reise, 2013)، والتعريف الإجرائي عبارة عن التمثيل البياني الذي سيتم استخدامه لمقارنة كيفية تمثيل كل من النموذجين (ثلاثي ورباعي المعلمة) لخصائص المفردات.

دالة معلومات المفردة (Item Information Function - IIF): دالة تصف مدى مساهمة المفردة في دقة قياس قدرة الأفراد عند مستويات قدرة مختلفة (Chalmers, 2012; Embretson & Reise, 2013; Han & Hambleton, 2014)، والتعريف الإجرائي عبارة عن الدالة التي سيتم استخدامها لمقارنة دقة المعلومات التي تقدمها كل مفردة تحت النموذجين (ثلاثي ورباعي المعلمة).

دالة معلومات الاختبار (Test Information Function - TIF): دالة تمثل مجموع دوال معلومات المفردات في اختبار معين، وتشير إلى مدى دقة الاختبار الكلية في قياس قدرة الأفراد عند مستويات قدرة مختلفة (Chalmers, 2012; Embretson & Reise, 2013; Han & Hambleton, 2014)، والتعريف الإجرائي عبارة عن الدالة التي سيتم استخدامها لمقارنة دقة القياس الكلية للاختبار تحت النموذجين (ثلاثي ورباعي المعلمة).

الخطأ المعياري في القياس (Standard Error of Measurement - SEM): هو

تقدير للخطأ العشوائي الموجود في الدرجة المقاسة، ويعكس عدم دقة القياس عند مستوى معين من القدرة في نظرية IRT، يختلف باختلاف مستوى القدرة (Embretson & Reise, 2013; Han & Hambleton, 2014)، والتعريف الإجرائي عبارة عن المقياس الذي سيتم استخدامه لتقدير مدى عدم دقة قياس قدرات الأفراد تحت كل من النموذجين، وهو معكوس دالة معلومات الاختبار.

الثبات التجريبي والهامشي (Empirical and Marginal Reliability): يشير الثبات التجريبي إلى مدى اتساق أو استقرار درجات الاختبار عبر تطبيقات مختلفة أو عبر مفردات مختلفة، ويشير الثبات الهامشي إلى تقدير الثبات الكلي للاختبار بناءً على جميع المفردات (Embretson & Reise, 2013)، والتعريف الإجرائي عبارة عن المؤشرات التي سيتم استخدامها لتقييم مدى موثوقية الاختبار الكلية عند تطبيق كل من النموذجين (ثلاثي ورباعي المعلمة).

تقدير قدرة الأفراد (Ability Estimation Methods): وهي عبارة عن الطرق الإحصائية التي سيتم استخدامها لتقدير قدرات الأفراد عند تطبيق النموذج الثلاثي والرباعي المعلمة، وسيتم مقارنة دقتها (Baker, 2001; Doğruöz & Arıkan, 2020; Embretson & Reise, 2013; Han & Hambleton, 2014)، وهي:

maximum likelihood (ML): طريقة لتقدير قدرة الفرد بناءً على احتمالية استجاباته.

maximum a posteriori (MAP): طريقة تقدير تأخذ في الاعتبار كلاً من احتمالية الاستجابات والمعلومات المسبقة عن توزيع القدرة.

expected a posteriori (EAP): طريقة تقدير تعتمد على متوسط توزيع القدرة.

الإطار النظري:

في هذا الجزء سيتم عرض إطار نظري عن متغيرات البحث:

أولاً: نظرية الاستجابة للمفردة Item Response Theory (IRT):

يذكر (Zanon et al., 2016, p. 2) أنه تم تطوير نظرية الاستجابة للمفردة في البداية في الخمسينيات والستينيات من القرن الماضي على يد فريدريك لورد وغيره من علماء القياس

النفسي الذين كان هدفهم تطوير طريقة قادرة على تقييم المجيبين دون الاعتماد على نفس المفردات المدرجة في الاختبار، ولذلك طورت IRT من نظرية القياس الكلاسيكية بهدف التغلب على العديد من قيودها وعيوبها، وهي نظرية إحصائية تتألف من مجموعة متنوعة من النماذج الرياضية التي لها الخصائص التالية:

- التنبؤ بدرجات الشخص بناءً على قدراته أو سماته الكامنة.
- إقامة علاقة بين أداء الفرد للمفردة ومجموعة السمات التي تكمن وراء أداء المفردة من خلال دالة تسمى "منحنى خصائص المفردة"، وهذه الخصائص ممكنة لأن نماذج IRT توفر ثبات معلمة المفردة والقدرة لمفردات الاختبار والأشخاص وذلك عندما يتناسب نموذج IRT محل الاهتمام فعليًا مع بيانات الاختبار المتاحة، بمعنى آخر ستحتفظ نفس المفردات المستخدمة في عينات مختلفة بخصائصها الإحصائية (على سبيل المثال، الصعوبة والتمييز)، ولن تعتمد درجات الأشخاص التي تمثل القدرة أو السمات الكامنة في بنية معينة على مفردات الاختبار المحددة التي تم إجراؤها عليهم.

المفاهيم المرتبطة بنظرية الاستجابة للمفردة:

تشير (فردوس بنت عبدالله الخوتانى, ٢٠٢٣, p. 364) أنه يوجد أربعة مفاهيم مرتبطة بالنظرية الحديثة في القياس وهما:

١- معلمة القدرة (θ): ويرمز لها بالرمز اليوناني θ، ويمثل المفهوم الكمي للسمة، ويمكن التصور بأن لكل فرد قيمة رقمية من القدرة ودرجة اختبارية التي تضع الفرد في موقعه علي مقياس متصل من القدرة، وأن مستوى القدرة التي يتمتع به أي فرد يتحدد باحتمال إجابته عن أي مفردة ويزداد هذا الاحتمال بزيادة القدرة ويقل بانخفاضها، وتقع قيم القدرة بين المجال ($-\infty$، ∞)، ويمكن حصرها ضمن مجال المدى وعلي الرغم من ذلك يمكن أن تكون هناك قيم خارج هذا المدى.

٢- معلمة صعوبة المفردة (β): وهي تصف موقع دالة المفردة علي طول مقياس القدرة، وإن صعوبة المفردة في إطار نماذج الاستجابة للمفردات الاختبارية ليست نسبة عدد الأفراد الذين يجيبون عنها صحيحة كما هو الحال في النماذج التقليدية، وإنما هي نقطة علي متصل السمة الكامنة التي نتوقع عندها احتمال الإجابة الصحيحة علي المفردة (٠.٠٥).

٣- معلمة تميز المفردة (a): وهي تصف مقدار قدرة مفردة ما على تمييز الأفراد المفحوصين

ذوى القدرة دون موقع المفردة وأولئك الأفراد ذوو القدرة فوق موقع المفردة، وهذه الخاصية تعكس شدة انحدار المفردة، ويعبر عن التمييز بشدة انحدار المنحنى في الجزء الأوسط.

٤- معلمة التخمين (c): ويسمى بالخط التقاربي الأدنى للمنحنى المميز للمفردة، ويمثل احتمال الإجابة الصحيحة للأفراد من ذوى القدرة المنخفضة، والهدف من إضافة هذه المعلمة في النموذج الثلاثي هو محاولة مراعاة عدم مطابقة المنحنيات المميزة للمفردة عند الطرف الأدنى لمتصل السمة، حيث يكون التخمين أحد المتغيرات العشوائية التي تؤثر في أداء الفرد علي مفردة ما.

ومما سبق يمكن القول أنه يوجد أربع معلمات للمفردة في نظرية الاستجابة للمفردة ، معلمة القدرة والتي يرمز لها بالرمز (θ) وهى تمثل المفهوم الكمي للسمة، معلمة صعوبة المفردة والتي يرمز لها بالرمز (β) والتي تشير إلي توقع احتمال الإجابة الصحيحة علي المفردة (٠.٠٥)، معلمة تمييز المفردة والتي يرمز لها بالرمز (a) وهى تصف مدى قدرة المفردة علي التمييز ذوى القدرات المرتفعة وذوى القدرات المنخفضة، معلمة التخمين والتي يرمز لها بالرمز (c) وهى تصف احتمال الإجابة الصحيحة علي مفردات الاختبار عن طريق التخمين من قبل الأفراد ذوى القدرات المنخفضة.

نماذج نظرية الاستجابة للمفردة (IRT)

يوضح (Thomas, 2019, p. 1443) أن معظم نماذج نظرية الاستجابة للمفردة جزءًا من إطار النموذج الخطي المعمم، وتستخدم دوال الارتباط لربط مصطلح خطي بناءً على معلمات الشخص والمفردة بيانات الاستجابة المنفصلة، أي أنه يجب استخدام دالة رياضية للسماح للتنبؤات من النموذج الخطي بالتعيين على مصطلحات الاستجابة غير الخطية (على سبيل المثال، سجل احتمالات الدقة)، وإن دالتي الارتباط الأكثر استخدامًا هما (اللوغاريتم — معكوس دالة التوزيع التراكمي للتوزيع اللوجستي — والاحتمال — معكوس دالة التوزيع التراكمي للتوزيع العادي)، إن الاختيار بين هذه الدوال له تأثير ضئيل على تنبؤات النموذج، حيث يمكن استخدام ثابت القياس لتحقيق قياسات مكافئة تقريبًا، ويمكن توسيع وظائف النموذج لتشمل بيانات الاستجابة الفئوية والاسمية المطلوبة، بالإضافة إلى العديد من نماذج الدوال الأخرى.

ويذكر (Aybek & Demirtasli, 2017, p. 476) أنه عند فحص نماذج نظرية

الاستجابة للمفردة أحادية البعد، يمكن ملاحظة أنه تم إجراء تصنيف للمفردات ثنائية ومتعددة الاستجابة، وتختلف النماذج التي تم تطويرها للعناصر الثنائية تبعًا لصعوبة المفردة (b)، وتمييز المفردة (a)، ومعلمات التخمين الزائف (c)، في النموذج اللوغاريتمي أحادي المعلمة يتم استخدام معلمة صعوبة العنصر فقط، حيث تشير المعلمة b لأحد المفردات إلى احتمال أن يكون احتمال الإجابة الصحيحة على مستوى القدرة/السمة هو ٠.٥، وبمعنى آخر يُفترض في هذا النموذج أن احتمال استجابة الفرد بشكل صحيح لعنصر ما يتأثر فقط بمستوى صعوبة تلك المفردة، أما في النموذج اللوغاريتمي ثنائي المعلمة تعتمد احتمالية الاستجابة بشكل صحيح لمفردة ما على معلمة الصعوبة (b) بالإضافة إلى معلمة التمييز (a)، أما في النموذج اللوجستي ثلاثي المعلمات يتم أخذ احتمالية الإجابة الصحيحة في الاعتبار مع المعلمتين a وb، ويتم تضمين المعلمة c، فالمفردات الثنائية التي تم تسجيلها (صحيحة – خاطئة) تأخذ (١– ٠) علي الترتيب، بينما يمكن تسجيل المفردات متعددة الدرجات جزئيًا أو في درجات مثل ١ – ٢ – ٣ – ٤ – ٥.

ويري (Thomas, 2019, p. 1444) أنه يمكن تمييز نماذج نظرية الاستجابة للمفردة إلي نوعين من النماذج وهما:

١- نموذج راش (Rash Model): في هذا النوع من النموذج تتطلب أن تكون معلمات التمييز متساوية عبر المفردات، وتشتمل نموذج راش الكلاسيكي أو نموذج المعلمة الواحدة لاستجابات المفردات ثنائية الاستجابة (على سبيل المثال، الصحيح مقابل غير الصحيح أو الصواب مقابل الخطأ)، وكلاً من نموذج التقدير الجزئي (partial credit model) ونموذج مقياس التصنيف (rating scale model) لاستجابات المفردات الفئوية المطلوبة (المرتبة بشكل قاطع).

٢- نماذج غير راشية (non Rash type Models): وفي هذا النوع من النماذج تسمح بتنوع تمييز المفردات، وتشتمل نماذج النوع غير راش على: –

- نموذج ثنائي المعلمات (أي معلمات الصعوبة والتمييز)
- نموذج ثلاثي المعلمات (أي معلمات الصعوبة والتمييز والتخمين) لاستجابات العناصر الثنائية
- نموذج التقدير الجزئي المعمم والاستجابة المتدرجة.

- نموذج لاستجابات المفردات الفئوية.

ويشير (Aybek & Demirtasli, 2017, p. 475) إلي أنه إذا كان مستوى قدرة/سمة الفرد مشابهًا لخصائص المفردة، فمن الممكن مقارنة قدرة/سمة الفرد مع خصائص المفردة وتقدير احتمالية استجابة الفرد لها باستخدام نظرية الاستجابة للمفردة (IRT)، والأسباب الرئيسية لذلك هي الخصائص التالية لهذه النظرية وهى:

١- معلمات المفردة مستقلة عن مستوى قدرة المجموعة.

٢- قدرة الفرد مستقلة عن عينة المادة المأخوذة.

٣- يتم تقدير الخطأ وفقًا لمستوى قدرة/سمة الفرد.

٤- يتم تحديد درجات المادة على مقياس.

٥- يمكن تطوير اختبارات مميزة لجميع مستويات القدرة.

٦- يمكن تقدير احتمالية استجابة الفرد بشكل صحيح لعنصر ما.

ويذكر كل من (فردوس بنت عبدالله الخوتانى, ٢٠٢٣, p. 366) و (Zanon et al., 2016, p. 3) إلي وجود ثلاثة نماذج IRT شائعة للبيانات الثنائية هي النماذج اللوجستية ذات المعلمة الواحدة والثنائية والثلاثية، وإنهم يميزون أنفسهم عن عدد المعلمات المقدرة:

١- النموذج اللوغاريتمي أحادي المعلمة (1PL): باستخدام هذا النموذج من الممكن تقدير احتمالية قيام شخص ما بالإجابة على العنصر (ذو الصعوبة، b) بشكل صحيح.

٢- النموذج اللوغاريتمي ثنائي المعلمة (2PL): ويهدف إلى تقدير احتمالية الإجابة الصحيحة، ولكنه يسمح أيضًا بتقدير التمييز للبند (a). كلما كانت قيمة "a" أكبر، كلما زاد ميل المنحنى المميز للمفرد (ICC) وزاد التمييز الذي توفره المفردة.

٣- النموذج اللوغاريتمي ثلاثي المعلمات (3PL): ويقوم هذا النموذج بتقدير المعلمتين الأخريين الموصوفتين (a وb) وكذلك احتمالية التخمين (المعلمة "c"). وفي بعض الأحيان يحصل الطلاب على الإجابات الصحيحة عن طريق التخمين على بدائل المفردة، وإن إدراج المعامل c يسمح للنموذج بتقدير تأثير التخمين على صعوبة المادة وعلى الدرجات النهائية للممتحنين.

ويشير (Yiğiter & Boduroğlu, 2024, p. 121) أن حجم العينة المستخدم يختلف

من نموذج لنموذج كالتالي:

١- نموذج راش أو نموذج 1PLM: حجم العينة يبلغ ١٠٠ أو ٢٠٠

٢- نموذج 2PLM: حجم العينة المطلوب ٥٠٠ أو أكثر

٣- نموذج 3PLM: حجم العينة المطلوب ١٠٠٠ أو أكثر

افتراضات نظرية الاستجابة للمفردة:

يذكر (Yiğiter & Boduroğlu, 2024, pp. 121-123) أنه يوجد خمسة افتراضيات أساسية النموذج IRT، ويمكن عرضهم كالتالي:-

١- أحادية البعد (Unidimensionality): البعد الأحادي يعني أن هناك نوعًا واحدًا فقط من القدرات يؤثر على أداء الفرد في الاختبار، وبعبارة أخرى هو خاصية مشتركة تجمع بين الفقرات في أداة القياس، و يُعد افتراضًا أساسيًا في نماذج IRT الأحادية البعد، ولكي تُعتبر مجموعة من الفقرات أحادية البعد يجب أن يكون لديها خاصية مشتركة وتباين مشترك يفسر التباين بين الأفراد، وإذا تم انتهاك افتراض البعد الأحادي لن تتطابق البنية الكامنة متعددة الأبعاد مع نموذج IRT الأحادي البعد، وقد تكون الدرجات المستخلصة باستخدام IRT الأحادي البعد منحازة، ومن الصعب تحقيق البعد الأحادي النقي في الواقع العملي لأن الأفراد قد يعتمدون على أكثر من سمة أثناء الإجابة، وعندما تكون السمة المقاسة متعددة الأبعاد.

٢- الاستقلالية المحلية (Local Independence): وتعني أن احتمالية إجابة الفرد على فقرة معينة لا تعتمد على سلوكه في الإجابة على الفقرات الأخرى، وبعبارة أخرى عند تثبيت القدرة التي تؤثر على الأداء في الاختبار تكون استجابات الأفراد للفقرات مستقلة إحصائيًا عن بعضها البعض، يُعد هذا الافتراض أساس دالة الاحتمال التي تعتمد عليها IRT، وفشله يؤدي إلى تقديرات غير دقيقة، وتُستخدم دالة الاحتمال القصوى (Maximum Likelihood Function) لتقدير القدرة بضرب احتمالات استجابات الفرد للفقرات، لكي يكون حاصل ضرب احتمالات الأحداث صحيحًا يجب أن تكون الأحداث مستقلة، وفي حال عدم تحقق الاستقلالية المحلية: يتم المبالغة في تقدير معلومات الاختبار ومعلمات تمييز الفقرات، و تصبح معلمات القدرة متحيزة بشدة، ويتم تقدير معلمات صعوبة الفقرات بشكل غير دقيق، ويوجد مجموعة من العوامل المؤثرة على الاستقلالية المحلية: الفقرات

ذات الجذور المشتركة: قد تؤثر الإجابة على فقرة معينة على استجابات فقرات أخرى بنفس الجذر، والسلوكيات مثل الغش أو التعب: يمكن أن تؤثر على الاستجابات، واختلاف ممارسات الطلاب: أو كون الاختبار يعتمد على السرعة، ولحل لانتهاك الاستقلالية المحلية: يتم استبعاد إحدى الفقرات المتأثرة بالاعتماد المحلي، تجميع الفقرات ذات الاعتماد المحلي: واستخدام نماذج IRT متعددة الفئات، استخدام نماذج نظرية استجابة الفقرات المتجمعة (Testlet Response Theory) والتي تأخذ بعين الاعتبار تجميع الفقرات.

٣- التوافق العام للنموذج (Overall Model Fit): حيث إن دراسة التوافق العام للنموذج يعتبر ذات أهمية كبيرة مثل أي دراسة نمذجة، يجب قياس مدى توافق النموذج مع البيانات، فالتقديرات والاستنتاجات المستخلصة من نموذج غير مناسب تكون غير صالحة، ويتم فيه مقارنة توزيع الدرجات الملاحظة مع التوزيع المتوقع وفقًا للنموذج.

٤- توافق الفقرات (Item Fit): لضمان دقة التقديرات المستخلصة من النموذج يجب أن تتوافق الفقرات مع النموذج، حيث أن عدم توافق الفقرات يؤدي إلى تقديرات متحيزة لمعلمات القدرة، وتصنيف غير عادل للمختبرين، ومعادلة غير دقيقة للدرجات.

٥- عدم تأثير الزمن (Non-Speediness Test): يفترض أن يتم إجراء الاختبار في ظروف غير مسرَّعة، فإذا لم يُجب المختبر على بعض الفقرات فيجب أن يكون السبب هو مستوى قدرته غير الكافي وليس بسبب عدم وصوله إلى الفقرات أو انتهاء الوقت المخصص للاختبار، ويُشار أحيانًا إلى هذا الافتراض ضمن افتراض البعد الأحادي، وعندما يؤثر الزمن على الأداء في الاختبار يصبح الأداء متأثرًا ببعدين على الأقل: القدرة المقاسة والسرعة، مما يؤدي إلى تأثير سلبي على البعد الأحادي، ويفترض أن يكون لدى المختبرين وقت كافٍ للإجابة على الفقرات التي يعتقدون أنهم قادرون على الإجابة عنها.

ويشير (Aybek & Demirtasli, 2017, p. 477) إلي أنه يجب استيفاء بعض الافتراضات من أجل الحصول على نتائج صحيحة باستخدام نماذج نظرية الاستجابة للمفردة والتي تم تطويرها لكل من المفردات الثنائية والمتعددة، وهذه الافتراضات هي:

١- أحادية البعد: إن افتراض أحادية البعد هو افتراض يجب تحقيقه عند العمل مع نماذج IRT أحادية البعد، وبناء على ذلك فإن كل مفردة من مفردات الاختبار يجب أن يكون قد اجتمع معًا لقياس سمة اختبارها باستخدام تقنيات تحليل العوامل الاستكشافية أو التأكيدية.

٢- الاستقلال المحلي: ويمكن تلخيص افتراض الاستقلال المحلي بأن الاستجابة لمفردة ما لا تؤثر على الاستجابة لمفردة أخرى، بالإضافة إلى ذلك فإن توفير افتراض أحادية البعد يوفر أيضًا دليلاً على الاستقلال المحلي.

٣- ملاءمة بيانات النموذج: هو أن البيانات المرصودة متوافقة مع منحنى خاصية المفردة.

بينما يذكر (Zanon et al., 2016, pp. 2-3) أن هناك افتراضين أساسيين يتعلقان بجميع تطبيقات IRT وهما كالتالي:

١- أحادية البعد: من المفترض أن يتم قياس بنية كامنة واحدة فقط من خلال مجموعة من العناصر في الاختبار، ولذلك فمن المتوقع وجود عامل مهيمن يفسر معظم درجات تباين الأداة المتوقعة، فبعض التقنيات الشائعة المستخدمة للتحقق من أحادية البعد هي؛ نسبة التباين (٢٠٪ أو أكثر من التباين في العامل الأول)، أو التحليل الموازي أو التحليل العاملي التأكيدي (اختبار فرضية عامل واحد).

٢- شكل المنحنيات المميزة للمفردة (ICC): حيث إن ICC المحدد بالنموذج يعكس العلاقات بين السمات الكامنة واستجابات المفردة، وعندما تكون مفردات الاختبار ذات درجات ثنائية (على سبيل المثال، صواب-خطأ، أو نعم-لا)، تميل ICCs إلى أن تكون منحنيات على شكل حرف S وتتطابق مع احتمالات إجابة الممتحنين بنجاح على مفردة الاختبار كدالة لقدراتهم، و يتم نقل مفردات الاختبار السهلة إلى اليسار على مقياس قياس السمة، ويتم نقل العناصر الصعبة إلى الطرف الأيمن من مقياس القياس، فالعناصر المميزة لها منحدرات أعلى من العناصر المميزة الأقل

الطرق الإحصائية المستخدمة في تحقيق افتراضيات نموذج IRT:

يشير(Yiğiter & Boduroğlu, 2024, pp. 121-123) أنه يوجد مجموعة من الطرق الإحصائية المستخدمة في كل افتراض من افتراضيات نماذج IRT، ويمكن عرضها كالتالي:-

١- أحادية البعد : لاختبار افتراض أحادية البعد يستخدم التحليل العاملي (Factor Analysis): يُستخدم بشكل شائع لاختبار البعد الأحادي، حيث يحاول التحليل العاملي تفسير العلاقات بين استجابات الفقرات باستخدام عوامل أقل، وهناك نوعان رئيسيان: التحليل العاملي الاستكشافي (EFA): يحدد البناء العاملي المحتمل بناءً على استجابات

الأفراد، والتحليل العاملي التوكيدي (CFA): ويختبر الفرضيات حول العلاقات المحددة مسبقًا، ويوجد طرق أخرى طرق إضافية لاختبار أحادية البعد.

٢- الاستقلال المحلي: يُفحص من خلال العلاقة بين الفقرات باستخدام مصفوفة البواقي، التي تمثل الفرق بين المصفوفة المُلاحظة والمصفوفة الناتجة عن النموذج، ومن الأساليب المستخدمة مؤشر Q3، مربع G2 (أو $\chi 2$)، والارتباط بين البواقي، وJSI.

٣- التوافق العام للنموذج (Overall Model Fit): يُستخدم منحنى خصائص الفقرة (ICC)لتقييم توافق الفقرات، حيث يظهر التوافق عندما تكون الفروقات بين المنحنى المقدر والقيم الملاحظة صغيرة، و معايير التوافق: مؤشرات معيار المعلومات -Log Likelihood) Test [-2*LL]، Akaike Information Criteria [AIC].، Bayesian Information Criteria [BIC])، حيث إن انخفاض هذه الإحصائيات يشير إلى توافق أفضل للنموذج مع البيانات، مؤشرات أخرى مؤشرات مثل ($\chi 2$/sd, GFI, AGFI, CFI, NFI, TLI, SRMR, RMSEA) تُستخدم لتحديد أفضل نموذج ملائم للبيانات.

٤- توافق الفقرات (Item Fit): ويوجد نوعين من مؤشرات توافق الفقرات: الأول: المؤشرات التقليدية (مثل OUTFIT وINFIT (لنماذج راش و 1PL)، ومؤشر $\chi 2$ Bock's، مؤشر Yen's Q1، ومؤشر G2) ، الثاني: المؤشرات البديلة: (مثل S-$\chi 2$، والنسبة المعدلة $\chi 2$/df).

٥- اختبار عدم تأثير الزمن (Non-Speediness Test): يوجد عدة طرق لاختبار افتراض عدم تأثير الزمن: الأول: مقارنة النتائج بين حالتين: تطبيق نفس صيغة الاختبار على نفس المجموعة، مرة مع تحديد زمن ومرة أخرى دون تحديد زمن، وتفحص العلاقة بين الدرجات في الحالتين، الثاني: نسبة التباين: يتم حساب نسبة التباين بين عدد الفقرات التي تُركت دون إجابة وعدد الفقرات التي أُجيبت بشكل خاطئ، فإذا كانت النسبة قريبة من الصفر يكون الاختبار غير متأثر بالزمن، وإذا كانت النسبة قريبة من الواحد يكون الاختبار متأثرًا بالزمن، الثالث: معدل الإجابة على الفقرات: يتم تقييم معدلات الوصول إلى الفقرات وإعطاء إجابات صحيحة أو خاطئة، ومن المتوقع أن تكون نسبة الإجابات على الفقرات عالية، الرابع: تقييم الفقرات غير المُجابة: يجب أن يُكمل ٨٠٪ من المختبرين الاختبار بالوصول إلى جميع الفقرات، ويجب أن يصل كل مختبر إلى ٧٥٪ على الأقل من

الفقرات.

الفرق بين نظرية الاستجابة للمفردة (IRT) ونظرية الاختبار الكلاسيكية (CTT):

يشير كل من (Thomas, 2019, p. 1445) و (Yiğiter & Boduroğlu, 2024, p. 119) إلي أنه عادة ما يتم تمييز نماذج CTT عن نماذج IRT بناءً على قوة الافتراضات، حيث تعتمد نماذج CTT عادةً على افتراضات ضعيفة في حين تميل نماذج IRT إلى الاعتماد على افتراضات قوية، فالنماذج الضعيفة لا تضع افتراضات محددة فيما يتعلق بالشكل الوظيفي للنتيجة المرصودة، أو النتيجة الحقيقية [معلمة الشخص الكامن]، أو توزيعات درجة الخطأ، حيث تتمتع الافتراضات الضعيفة بميزة جعل النموذج قابلاً للتطبيق على نطاق واسع، ولكنها قد تحد أيضًا من فائدته، على عكس IRT، وتعتمد معلمات المفردة والشخص في CTT على العينة، ونتيجة لذلك فإن تقديرات CTT لصعوبة المفردة (أي النسبة الصحيحة) والتمييز بين المفردات (أي ارتباط الدرجات الإجمالية للعنصر) وكذلك تقديرات الاحتمالية تعتمد على العينة التي تم الحصول عليها منها، وفي بعض الأحيان تجسد نماذج CTT في بعض الأحيان افتراضات قوية وفي هذه الحالة يصبح التمييز بين CTT وIRT أقل وضوحًا، بينما تعتبر نماذج IRT قوية لأنها تضع افتراضات مقيدة مثل عدد القدرات الكامنة أو السمات المقاسة وشكل الارتباط (أو الوظيفة) المتعلق باستجابات المفردات للقدرة الكامنة والارتباطات الشرطية بين الأخطاء، ويجب استيفاء هذه الافتراضات للوثوق في الاستنتاجات المشتقة من تحليل IRT، وبالتالي يمكن استخلاص أن الفرق الأساسي بين CTT و:IRT هي أن CTT يركز على مستوى الدرجة الإجمالية للاختبار، بينما IRT يركز على مستوى الفقرات الفردية ويُقدم تقديرات أكثر دقة للقدرة والخصائص، وفي الواقع إن أحد أسباب قيام علماء القياس النفسي بتطوير العديد من نماذج IRT هو أن هناك حاجة إلى قدر كبير من المرونة للباحثين التطبيقيين للعثور على النموذج الصحيح والافتراضات اللازمة لاختبار معين.

ويذكر(Zanon et al., 2016, p. 2) أنه يوجد أربعة قيود وعيوب على CTT وهى كالتالي:-

١- أن تفسير خصائص المستجيب يعتمد على الاختبار المستخدم: سيبدو المستجيبون أكثر ذكاءً، إذا تم إجراء اختبار أسهل، لكنهم سيبدون أقل ذكاءً، إذا تمت الإجابة على اختبار أكثر صعوبة.

٢- أن خصائص الاختبار تعتمد على العينة: نفس الاختبار الذي يتم إجراؤه على مجموعة من الطلاب ذوي القدرات العالية ومجموعة أخرى من الطلاب ذوي القدرات المنخفضة سوف ينتج مفردات ذات مستويات مختلفة من الصعوبة، فعلى سبيل المثال في العينة الأولى ستظهر صعوبة المفردات أقل من صعوبة المجموعة الثانية، وتشير أوجه القصور هذه إلى أنه لا يمكن إجراء خصائص الاختبار إلا في نفس السياق (العينة)، وبمجرد أن تعتمد معلمات الاختبار على السمات الكامنة للأشخاص والعكس صحيح ستتغير خصائص العنصر والاختبار عندما يجيب أشخاص آخرون (عينات ذات مستويات مختلفة من السمات الكامنة) على الاختبار.

٣- تفترض نظرية CTT أن أخطاء القياس متساوية لجميع الأشخاص: وهذا يمثل مشكلة لأن الأشخاص ذوي مستويات مختلفة من القدرة سيظهرون مستويات مختلفة من الخطأ (التخمين) في اختبار يقيم الذكاء أو أي بناء آخر.

٤- لا تسمح بالتنبؤات الدقيقة حول النتائج المحتملة للمستجيب أو لعينة على عنصر ما وذلك باستخدام درجات قدراتهم فقط: وبالتالي يمكن معالجة كل هذه النقاط بشكل فعال باستخدام IRT، إذا كان من الممكن بالطبع العثور على نموذج IRT يناسب بيانات الاختبار.

النموذج اللوغاريتمي رباعي المعلمات (The four-parameter logistic item (4PL) response model:

يعتبر النموذج اللوغاريتمي الرباعي (4PL) هو أداة رياضية تُستخدم بشكل شائع في علوم الحياة، ويعتمد النموذج على أربع معلمات رئيسية تساعد في تحديد سلوك الاستجابة، ويتيح هذا النموذج تقدير الاستجابة عند القدرات المنخفضة والعالية على حد سواء، مما يجعله أداة مهمة في التحليل الإحصائي.

ويشير (Zheng et al., 2021, p. 1) إلي أنه تم ذكر نموذج الاستجابة للعناصر اللوجستية ذات الأربعة معاملات (4PL) لأول مرة علي يد McDonald ماكدونالد عام ١٩٦٧ واقترحه رسميًا بارتون ولورد Barton and Lord عام ١٩٨١، وعلى الرغم من ذلك فقد حظي لسنوات عديدة باهتمام ضئيل للغاية في الأدبيات بسبب الشكوك حول فائدته، وفضلاً عن الصعوبات الفنية المتعلقة بتقدير المعاملات، ومع ذلك فقد أعاد مجتمع القياس النفسي مؤخرًا إحياء اهتمامه بنموذج الاستجابة للمفردة اللوجستية ذات الأربعة معلمات بعد ثلاثة عقود

من الإهمال.

ويعرف (Battauz, 2020, p. 296) النموذج اللوغاريتمي رباعي المعاملات بأنه نموذج من نماذج نظرية استجابة للمفردة للمفردات ثنائية الاستجابة، ويحد من احتمالية إعطاء استجابة إيجابية لمفردة ما في نطاق محدود، بحيث لا يكون لدى الأشخاص الذين لديهم سمة كامنة احتمالية قريبة من الصفر أو الواحد.

أهمية النموذج اللوغاريتمي رباعي المعلمات 4PL:

يري (Zheng et al., 2021, p. 1) أنه ظهرت تطبيقات 4PL في الأدبيات كدليل داعم لفائدتها:-

١. حيث زعمت العديد من الدراسات أن معلمة الإهمال في 4PL يمكن تقديره لتفسير الأخطاء المبكرة غير المقصودة لتحسين تقدير القدرة.

٢. وفي علم النفس لوحظ أن هناك قيمة في نمذجة المقاربات السفلية والعلوية لوظائف استجابة العنصر في الحالات التي قد يتردد فيها الأشخاص ذوو المستويات الأعلى من الأمراض النفسية في الإبلاغ عن المواقف والسلوكيات و/أو الخبرات التي قد تشير إلى تحيز المرغوبية الاجتماعية.

٣. وفي التعليم، أظهرت الدراسات الحديثة انتشار وتأثير معلمة الإهمال في التقييمات التعليمية واسعة النطاق منخفضة المخاطر.

ويضيف (Jian et al., 2021, pp. 2-3) أن النموذج اللوغاريتمي رباعي المعلمات قد يكون مناسبًا بشكل فريد لتوصيف بيانات علم النفس المرضي في الاختبارات غير التكيفية، حيث وجُد أن الأخطاء العرضية التي يرتكبها الأفراد ذوو القدرات العالية نسبيًا في البنود السهلة تؤدي إلى تحيزات أكبر نسبيًا في الاختبارات غير التكيفية أو اختبارات الورقة والقلم، فعن طريق استخدام هذا 4PL يمكن تقليل هذه التحيزات بشكل فعال، بالإضافة إلي أن الانحياز وجذر متوسط مربع الخطأ (RMSE) للطلاب ذوي القدرات العالية يمكن تقليلهما باستخدام نموذج 4PL، وأنه يقدّم تقديرًا أكثر قوة للقدرة مقارنة بنموذج 3PL في الاختبارات التكيفية.

ويشير (Meng et al., 2020, p. 52) إلي أنه يجب إعادة النظر في استخدام هذا النموذج نظراً لأهمية كبيرة في تحسين دقة التقدير في الاختبارات المحوسبة، ويكون أكثر ملاءمة لقياس سمات الأمراض النفسية مقارنة بنموذج اللوجستي ذي الثلاث معلمات (3PL)

أو النموذج ثنائي المعلمات (2PL) وذلك لأن وجود مفحوص يمتلك سمة عالية لكنه متردد في الإفصاح عن مواقفه هو أمر شائع جدًا في قياسات الأمراض النفسية، وتحليل التوزيع التقاربي للقدرات، ويستخدم طرق تقدير بايزية وMML مثل الموجودة في حزمة mirt في R.

ويذكر (Świst, 2015, p. 82) أن تم اعتماد 4PL عمليًا في الاختبارات التكيفية المحوسبة (CAT)، حيث استُخدم لتقليل تأثير الأخطاء المبكرة للطلاب على تقدير مستواهم، وأثبت فاعلية أكبر مقارنةً بنموذج 3PL، كما أن 4PL قدّر مستوى القدرة بدقة حيث منع انخفاض التقدير المبدئي الناتج عن إجابات خاطئة على أول عنصرين في الاختبار، بالإضافة إلي أن نموذج 4PL يسمح بتقدير أكثر قوة للقدرة من خلال إعطاء أوزان مختلفة لدالة الاحتمالية اللوغاريتمية؛ حيث تُخفف تأثيرات الإجابات الشاذة في عملية التقدير.

ويضيف (Battauz, 2020, p. 297) أن نموذج (4PL) يُعد خيارًا أكثر مرونة في حالة الاستجابات الثنائية، حيث يلتقط العلاقة بين الاستجابات والمتغير الكامن مع السماح بوجود بعض العشوائية، فعلى سبيل المثال قد يكون للأشخاص ذوي السمات المنخفضة جدًا احتمال غير صفري للاستجابة الإيجابية، والعكس صحيح أيضًا للأشخاص ذوي السمات العالية جدًا، فالتخمين في اختبارات متعددة الخيارات مثال شائع على ضرورة نمذجة هذا السلوك عند مستويات القدرة المنخفضة، بينما قد يفشل الأشخاص ذوو القدرات العالية في تقديم إجابة صحيحة بسبب الإهمال أو التعب.

ومما سبق يري الباحثان أن النموذج اللوغاريتمي رباعي المعلمات له أهمية عديدة حيث أن معلمة الإهمال أو اللامبالاة التي تمت إضافتها في هذا النموذج تساعد في توصيف بيانات علم النفس المرضى وتفسر الأخطاء غير المقصودة والتي بدورها تعمل علي تحسين القدرة وتستخدم على نطاق واسع في التعليم، حيث يقدر مستوى القدرة بدقة، كما أنها تساعد في تبسيط العلاقات وتحويل العلاقات غير الخطية بين المتغيرات إلى علاقات خطية، وتقدم مرونة إضافية في التحكم بمعدل الانحدار عبر المعلمة الرابعة، وتقلل من تأثير القيم المتطرفة، وتساعد في تحسين التوزيع الطبيعي للبيانات، بالإضافة إلي فاعليتها الكبيرة في إنشاء الاختبارات المحوسبة.

المعادلات الرياضية لنماذج IRT (3PL, 4pL)

ويشير (Han & Hambleton, 2014, p. 12) إلي أن عالم الرياضيات الدنماركي

جورج راش قدم منهجًا مختلفًا في IRT في الخمسينيات، حيث استخدمت الدالة اللوغاريتمية التي تم الحصول عليها من المنحنى المميز للمفردة دالة ogive الطبيعي (Normal ogive function)، ويذكر (Ronald K Hambleton et al., 1991, p. 12) أن النموذج اللوغاريتمي ذو المعلمة الواحدة أحد النماذج الأكثر استخدامًا في نماذج IRT، ويتفق كل من (Ronald K Hambleton et al., 1991, p. 13) و (Embretson & Reise, 2013, p. 69) و(Baker, 2001, p. 25) و(Kolen et al., 2014, p. 175) أن في هذا النموذج تكون معلمة التمييز (a) لجميع العناصر هي نفسها ومعلمة التخمين الزائف (c) تعتبر صفرًا، ومع ذلك تختلف معلمات الصعوبة للمفردات في الاختبار (b) وفقًا للمفردة.

ويذكر (Ronald K Hambleton et al., 1991, p. 13) أن لورد عام ١٩٥٢ قام بتطوير نموذج اللوغاريتمي ثنائي المعلمات لأول مرة بناءً على التوزيع الطبيعي التراكمي (التوزيع الأوجيفي الطبيعي Normal ogive function)، وقد استبدل بيرنبوم عام ١٩٦٨ دالة الأوجيفي الطبيعي ذات المعاملين كشكل من أشكال الدالة المميزة للمفردة، ويوضح كل من (Baker, 2001, p. 22) & (Jian et al., 2021, p. 2) أن معامل الميل أو التمييز في النموذج اللوغاريتمي ثنائي المعلمات (2PL)، نظريًا يقع في النطاق (−∞، +∞) ولكن في التطبيقات يأخذ قيمًا في النطاق من ٢.٨٠ إلى +٢.٨٠، ويري (Ronald K. Hambleton et al., 1991) أن عادةً ما يحصل معامل (التمييز) على قيمة بين ٠ و٢، وعندما تقترب قيمة المعامل من ٢، يزداد التمييز، ويوضح (DeMars, 2010, p. 5) أن القيم الأعلى للمعامل تشير إلى تمييز أعلى في IRT مثل CTT

١- النموذج اللوغاريتمي ثلاثي المعلمات(3PL):

يوضح (Baker, 2001, p. 28) أن العالم Birnbaum عام 1968 قام بإضافة معامل التخمين الزائف إلى النموذج اللوغاريتمي ثنائي المعلمة، وبالتالي تم الحصول علي النموذج اللوغاريتمي ثلاثي المعلمة 3PLM، حيث يشير (Han & Hambleton, 2014, p. 13) أن النموذج اللوغاريتمي ثلاثي المعلمات يسمح بأن يكون المقارب الأدنى لمنحنى المميز للمفردة مختلفًا عن الصفر، و هذا النموذج مناسب حتى لو كان الأشخاص الذين تم اختبارهم على مستوى كفاءة منخفض إلى حد ما، فعلى سبيل المثال عندما يجيبون على مفردة اختيار من متعدد مع التخمين.

ويذكر (Jian et al., 2021, p. 2) أنه فيما يتعلق بتخمين الاستجابات في الاختبارات تم اقتراح نموذجًا لوغاريتمياً ثلاثي المعلمات (3PL) والذي تم التعبير عنه على النحو التالي:-

$$Pvi = Ci + (1 - Ci)\frac{\exp(ai(\theta v - bi))}{1 + \exp(\ ai\ (\theta v - bi))}$$

Ci تشير إلي تخمين المفردة i، ai تشير إلى تمييز المفردة i، θv تشير إلى قدرة الفرد v، bi تشير إلى صعوبة المفردة i، Pvi تشير إلى احتمال إجابة المفحوص v على مفردة i

ويري (Ronald K Hambleton et al., 1991, p. 17) و (Hambleton & Swaminathan, 2013, pp. 37-38) أن المعامل المضاف لـ Ci يمثل (مستوى الفرصة الزائفة (التخمين)) في النموذج احتمالية المستجيبين بمستوى قدرة منخفض للإجابة بشكل صحيح على العنصر ويوفر منحنيات خصائص العنصر مع مقارب منخفض مختلف عن الصفر.

ويوضح (Baker, 2001, p. 25) و (Han & Hambleton, 2014, p. 13) أن 2PLهو الإصدار الخاص من 3PL وذلك عندما يكون $Ci = 0$ ، وأن نموذج راش هو الإصدار الخاص من 2PL وذلك عندما يكون $ai = 1$.

ويذكر(Baker, 2001, pp. 28-29) أنه من الناحية النظرية يحصل معامل التخمين الزائف Ci على قيم في النطاق [٠, ١]، ولكن من الناحية العملية يُذكر أن قيم Ci أعلى من ٠.٣٥، وأن النطاق خارج [٠, ٠.٣٥] تكون غير مقبولة ، ويشير (DeMars, 2010, p. 21) إلي أن المقارب الأدنى أو معامل Ci يأخذ قيمًا بين ٠ و١ من الناحية النظرية، فإنه يأخذ عادةً قيمًا بين ٠ و٠.٣ في البيانات الحقيقية.

٢- النموذج اللوغاريتمي رباعي المعلمات(4PL):

يذكر (Büyükkıdık & İnal, 2023, p. 831) أنه قام بارتون ولورد عام ١٩٨١ بتطوير نموذج 4PL بإضافة احتمالية حصول المستجيبين رفيعي المستوى لأخطاء عند الإجابة على المفردة السهلة، أي معامل di المقابل للمقاربة العليا لنموذج 3PL .

واستكشف (Barton & Lord, 1981, p. 7) ما إذا كان تغيير المتقارب العلوي (the upper asymptote parameter of item) يمكن أن يحسن من تسجيل النتائج في الاختبارات القياسي، وبالتالي قاما بإضافة معامل رابع di على نموذج المتقارب العلوي، مما أدى إلى خفض المقارب العلوي إلى أقل من ١:

$$Pvi = Ci + (di - Ci)\frac{\exp(ai(\theta v - bi))}{1 + \exp(\ ai\ (\theta v - bi))}$$

حيث تشير di إلي معامل المقارب العلوي للعنصر i أو معامل الإهمال، وعندما تكون $di = 1$ فإن 4PL يصبح 3PL، ومعني ذلك أن 3PL هي حالة خاصة من 4PL.

ويوضح (Pavlech & Martinková, 2024, p. 2) أن Ci هي معامل التخمين الزائف في سياق الاختبار التعليمي، والذي يتوافق مع احتمال الإجابة على العنصر بشكل صحيح دون امتلاك المعرفة المعنية (بسبب التخمين أو لأسباب أخرى)، أما في سياق القياس النفسي والصحي يكون المعامل Ci هو معامل التظاهر والذي يتوافق مع احتمال تأييد عنصر ما دون امتلاك الصفة، بينما يطلق علي di اسم معامل عدم الانتباه أو الانزلاق أو اللامبالاة أو الإهمال في سياق القياس التعليمي، ومعامل الإخفاء في سياق التقييم النفسي أو المتعلق بالصحة، وقيمة هذا المعامل في النطاق [٠، ١] من الناحية النظرية، وتشير حقيقة أن معامل di أقل بكثير من ١ إلى أن المستجيبين ذوي مستويات القدرة العالية هم أكثر عرضة للإجابة على المفردة بشكل غير صحيح بسبب الإهمال وأسباب مماثلة.

ويري (Meng et al., 2020, p. 52) أن معامل di ناتج عن عدم قدرة الممتحن ذي القدرة العالية على اجتياز فقرة سهلة، كما أنه يتم تسمية di بمعلمة الإهمال (Inattention Parameter) تُستخدم هذه المعلمة لحساب الأخطاء غير المقصودة مثل الإجابة الخاطئة بسبب التعب أو النقر العرضي، ويتم تحديد قيمة قصوى تعسفية تصل إلى ٠.٨، وهذه المعلمة نادرة الاستخدام في تقييمات الصحة النفسية.

ويشير (Büyükkıdık & İnal, 2023, pp. 827-828) إلي أن النماذج اللوغاريتمية (PL) ذات المعلمات ١ و٢ و٣ هي الأكثر استخدامًا، ومن الممكن أيضًا إجراء تقديرات بناءً على 4PL، والتي تنتج معلمة المقارب العلوي (upper asymptote parameter)، حيث أنه تم إنشاء هذا النموذج بواسطة Barton and Lord عام ١٩٨١ مع إضافة معلمة di إلى 3PL، ففي هذا النموذج (4PL) يأخذ المستجيبون ذوو القدرة العالية في الاعتبار إمكانية ارتكاب خطأ في الإجابة على عنصر سهل، ومع إضافة المقارب العلوي بقيمة أقل من ١ فإنه يضمن أن المستجيب ذو مستوى القدرة العالية لا يتغير بشكل كبير في مقياس القدرة إذا استجاب بشكل غير صحيح لعنصر سهل.

الأساليب الإحصائية المستخدمة في تقدير معلمات نموذج 4PL:

يذكر (Meng et al., 2020, p. 52) أنه من خلال مراجعة أدبيات البحث لمعرفة الأساليب الإحصائية المستخدمة في النموذج 4PL تم التوصل إلي النتائج التالية:

١- تم استخدام التوزيع التقاربي لتقدير القدرة ضمن نموذج 4PL.

٢- تم اقتراح عدة طرق لتقدير معلمات هذا النموذج مثل منهجية بايزية تعتمد على تقنية ماركوف تشين مونتي كارلو (MCMC) لتقدير معلمات النموذج.

٣- طريقة الاحتمالية العظمى الهامشية (MML) لاسترجاع النموذج باستخدام حزمة R المسماة mirt.

٤- تطبيق أسلوب Bayes Modal (BM) لتقدير معلمات نموذج 4PL كما هو مطبّق في حزمة mirt.

وبالمقارنة مع طريقة التقدير البايزية التي تعتمد على خوارزمية MCMC، فإن طريقة MML تحتاج وقتًا حسابيًا أقل، لكنها قد لا تكون مستقرة وقد تنتج عنها قيم غير منطقية في كثير من الأحيان، وللتغلب على هذا العيب تم اقتراح سلوب Bayes Modal (BM) لتقدير معلمات نموذج 3PL، ويمكن اعتبار هذا الأسلوب نوعًا من تقدير الاحتمال الأقصى المرجح الهامشي (MMAP)؛ إذ أنه يستخدم هدف تحسين موسع يشمل دالة الاحتمال مع بعض المعتقدات السابقة (priors) حول معلمات المفردات، وتُستخدم هذه المعتقدات لمنع ظهور تقديرات غير منطقية، وفي الواقع يمكن اعتبار التقدير بأسلوب BM بمثابة تنظيم لطريقة MML، أي أن MML هي حالة خاصة من BM عندما نفترض أن التوزيعات السابقة للمعلمات هي توزيعات متجانسة، ومؤخرًا تم تطبيق تقدير BM وهو مطبّق في حزمة mirt على نموذج 4PL.

ويشير (Jian et al., 2021, p. 3) إلي أنه تم استخدام نهجًا بايزيًا لاسترجاع تقديرات المعلمات بنجاح لكل من العناصر والمفحوصين و تحديدًا للمعلمتين Ci و di في نموذج 4PL ، كما أنه أداة Gibbs sampler، وهي حزمة R مفتوحة المصدر مبنية على بيئة OpenBUGS لتقدير النموذج، واستخدام برنامج WINSTEPS لتقدير النهاية العليا في نموذج 4PL، ويمكن اعتبار هذه النهاية تقريبًا لمعلمة di.

وأضاف (Battauz, 2020, p. 276) طريقة تقدير لنموذج (4PL) تعتمد على تقدير

الاحتمالية العظمى المعاقبة (Penalized Maximum Likelihood Estimation)، وعلى عكس تقدير الانحدار الخطي، حيث يُدخل التقدير المعاقب دائمًا انحيازًا من أجل تقليل تباين التقديرات، فإن العقوبة المُضافة في هذا النموذج لا تؤدي في كثير من الحالات إلى زيادة الانحياز، بل تؤدي دائمًا إلى تقليل جذر متوسط مربع الخطأ (RMSE)، ومن الجدير بالذكر أن مقدر المربعات الصغرى (OLS) المستخدم في الانحدار الخطي يعطي تقديرات غير متحيزة للمعاملات، بينما تقدير الاحتمالية العظمى (MLE) في النماذج غير الخطية يُعرف بأنه متسق لكنه منحاز عند استخدام عينات محدودة.

ماذا يحدث عندما يتم تطبيق نماذج أبسط من IRT على بيانات تحتوي على معلمة di فعالة؟

يشير (Świst, 2015, p. 83) إلي أنه عند تطبيق نموذج ثلاثي المعلمات 3PL على مثل هذه البيانات انخفضت قيم معلمات التمييز وتحول مستوى صعوبة البنود إلى الأعلى (بما يقارب ٠.٥ انحراف معياري)، وأنه عند مقارنة معلمات 4PL بمعلمات 3PL لوحظ أن معلمات التخمين الزائفة Ci الناتجة عن 3PL كانت تحتوي علي جذر متوسط مربع الخطأ (RMSE) أعلى وارتباطًا أقل مع الدرجة الحقيقية، وعند تطبيق نموذج 2PL حدث انحراف بمتوسط ٠.٥ في معلمات التمييز وانتقلت معلمات الصعوبة إلى الصفر، ورغم هذا التغير بقيت الارتباطات بين قيم القدرة (θ) الناتجة عن النماذج الثلاثة عالية جدًا (r = 0.98)، فرغم أن استخدام النماذج الأبسط لا يؤدي إلى تحيّز كبير في المعلمات، إلا أنه يؤثر على دالة معلومات الاختبار (TIF) ونمذجة الأخطاء المعيارية حيث أنه عند مستويات قدرة منخفضة قلّلت 3PL من مستوى المعلومات في TIF، وعند مستويات القدرة العالية بالغت في تقديرها، وكما أثرت 3PL على حساب الأخطاء المعيارية، إذ لم تكن كبيرة بما يكفي لدى أصحاب القدرات المنخفضة (ما أدى إلى فواصل ثقة واسعة جدًا)، أما 2PL فقد أظهر مستوى معلومات أعلى من 4PL، لكن كانت دقته الأعلى فقط في منتصف مقياس السمة الكامنة، في حين لم يُظهر دقة كافية في الأطراف.

نقد نموذج 4PL:

يذكر (Ronald K Hambleton et al., 1991, pp. 48-49) أن 4PL لم يكن له أي قيمة عملية، ولذلك توقف البحث في الخصائص السيكومترية وتطبيق 4PL بسبب مشاكل في التقدير (حيث يُعتبر تقدير المتقارب الأدنى إشكاليًا) ويُنظر إليه على أنه عديم الفائدة، ولكن

يري (Świst, 2015, p. 82) أنه بعد مرور ٢٥ عامًا، عاد الاهتمام بالنموذج، فقد أُدخلت طرق بايزية في الحسابات تُسرّع وتُسهّل في عملية التقدير، وعلى الرغم من ذلك فلا تُحل جميع المشكلات المفاهيمية المتعلقة بـ 4PL.

ويرى (Świst, 2015, p. 83) أنه يُعاب على 4PL أنه لا يوجد إثبات رسمي لهويته حيث أنه أكثر تعقيدًا من الناحية الحسابية وبالتالي فإن إثبات تحديد هويته (أو غيابه) يصبح أكثر صعوبة، كما أن تفسير معلمة di قد يكون مربكًا والتي تم افتراضها أنها ناتجة عن إهمال الطالب أو التوتر، وإذا كان هذا الافتراض صحيحًا فإن معلمة di يجب أن تُقدّر حسب كل شخص، مما يعني أن النموذج يجب أن يُعاد صياغته ليتنبأ بالإجابات الصحيحة بشكل فردي، ولكن حتى في حال تم تقدير معلمة di بشكل فردي فإن النموذج لا يفسّر سبب حدوث هذا السلوك، وخاصة في الاختبارات المصيرية حيث يُفترض أن الجميع لديهم دافع للإجابة بشكل صحيح، وأسباب ظهور معلمة di يمكن أن تكون متعددة منها الإهمال أو التوتر، ضيق الوقت (اختبار سريع)، الإجابة بأسلوب إبداعي لا يتطابق مع مفتاح التصحيح، ومع ذلك يبقى تحديد السبب وراء ظهور di أمرًا غامضًا وقد يؤدي إلى نتائج غير واقعية، وربما يكون نموذج 4PL بمعلمة di مخصص للفرد أكثر تعقيدًا، بل وحتى غير قابل للتحديد، ولذلك ومن أجل تبسيط التحليل يُفترض أن di تشير فقط إلى وجود أخطاء في كتابة المفردات وليس عمليات معرفية خاصة بالطلاب المتفوقين.

ويشير (Meng et al., 2020, p. 52) إلي وجود صعوبة في تقدير المعلمات، وغياب الأدلة الداعمة لضرورة استخدام النموذج تُعد من الأسباب المحتملة لعدم استخدام هذا النموذج، وبالإضافة إلي قصور نهج بارتون ولورد في أن جميع مفردات الاختبار تشترك في معامل متقارب علوي مشترك، ولم يُقدّر بارتون ولورد المعامل الرابع، بل زودا النموذج بقيم ثابتة لـ di، وقد أظهرت دراسات حديثة أن معامل المتقارب العلوي يختلف في معظم الحالات بين مفردات الاختبار.

ويضيف (Battauz, 2020, p. 297) إن تقدير معلمات هذا النموذج لا يزال مهمة صعبة، وهذا ما يفسّر تجاهله الطويل، فبعض الدراسات الحديثة لجأت إلى التقدير البايزي لمعالجة هذه الصعوبة، بينما اعتمدت دراسات أخرى على نماذج خليط وتوزيعات مسبقة للمعلمات.

ملائمة نموذج IRT المستخدم للبيانات:

يشير (Zhao & Hambleton, 2017, p. 2) إلي أنه في نظرية الاستجابة للمفردات (IRT) تم اقتراح طرق وأساليب متعددة للكشف عن عدم توافق النموذج، وتلخص هذه المقاييس لملاءمة النموذج عادة الفجوة بين القيم الملاحظة والقيم المتوقعة وفقًا لنموذجIRT سواء على مستوى المفردة أو الاختبار من خلال اختبارات لها دلالة إحصائية و/أو تمثيلات بيانية، ورغم عدم التوصل إلى اتفاق عام بشأن أفضل الأساليب للكشف عن عدم التوافق، فإن الملاحظة الأكثر أهمية هي أن الدراسات نادرًا ما تُقيّم تأثير عدم توافق نموذج IRT من خلال التركيز على عواقب استخدام مفردات غير متوافقة أو الأخطاء الإحصائية في تقدير المفردات المرتبطة بها، وبمعنى آخر هل يؤثر مقدار عدم الملاءمة الملحوظ على التطبيق المستهدف؟، وفي النهاية يجب النظر إلى عواقب عدم الملاءمة عند اتخاذ قرار بشأن جدوى استخدام نموذج معين في سياقات محددة. إذا كانت العواقب طفيفة، فقد يكون من المقبول التغاضي عن عدم الملاءمة في النموذج.

عواقب عدم توافق نموذج IRT للبيانات:

يري (Hambleton & Han, 2005, p. 58) رغم أهمية تقييم عواقب عدم توافق IRT لاتخاذ قرارات عملية، إلا أن هذا المجال البحثي لم يحظَ بالاهتمام الكافي الذي يستحقه في أدبيات ملاءمة IRT.

وتوصلت نتائج بعض الدراسات التي توفر أمثلة جيدة لتقييم عواقب عدم توافق IRT من خلال فحص مدى توافق القرارات المتخذة بناءً على تضمين/استبعاد المفردات غير المتوافقة (عدم توافق المفردات) أو الأشخاص غير المتوافقين (عدم توافق الأشخاص) للتالي:

- درس (Sinharay & Haberman, 2014, p. 23) الأهمية العملية لعدم توافق النموذج باستخدام مجموعات بيانات تجريبية مختلفة، وقد توصلوا إلى أن عدم التوافق لم يكن دائمًا ذا أهمية عملية، رغم وجود أدلة على عدم توافق عدد كبير من المفردات.
- وحقق (Zhao, 2017, p. 555) في التأثير العملي لعدم توافق المفردات باستخدام بنوك مفردات نظام قياس نتائج المرضى (PROMIS) للاكتئاب وتداخل الألم، وأشار إلى أن عدم توافق المفردات كان له تأثير طفيف على تقديرات الدرجات وتصنيفات الشدة في العينة المدروسة.

- أما (Meijer & Tendeiro, 2015, p. 1) فقد حللا مجموعتين من البيانات التجريبية وفحصا تأثير إزالة المفردات غير المتوافقة وأنماط درجات المفردات غير المتوافقة على ترتيب المتقدمين للاختبار وفقًا لدرجات كفاءتهم، ووجدا أن تأثير الإزالة يختلف حسب نموذج IRT المستخدم.

ويذكر (Zhao & Hambleton, 2017, p. 2) أنه في الممارسة العملية قد يكون من غير الممكن أحيانًا ببساطة إزالة المفردات غير المتوافقة، وحتى إذا أشارت اختبارات الدلالة إلى أن المفردة غير متوافقة فقد تظل بحاجة إلى الاحتفاظ بها ضمن الاختبار لأسباب مثل تحقيق توازن في تغطية المحتوى أو لاستخدامات تحليلية تعتمد على IRT مثل المعادلة المستندة إلى IRT، وتحليل الفروق بين المفردات وتقديرات القدرة في بيئة اختبار تكيفية، وعلاوة على ذلك عند حذف المفردات غير المتوافقة من المحتمل أن تتأثر ثبات وصدق درجات الاختبار، وبالمثل من غير العملي إزالة الأشخاص غير المتوافقين خاصة في التحليلات على مستوى الأفراد حيث تُستخدم تقديرات معلمة القدرة لاتخاذ القرارات.

ويري (Tendeiro & Meijer, 2015, p. 3) أنه يوجد نوع ثالث من عدم التوافق، بالإضافة إلى مستوى المفردة ومستوى الشخص، يحدث على مستوى الاختبار ويُعرف أيضًا باسم سوء تحديد النموذج. تقييم عدم التوافق بشكل شامل على مستوى الاختبار ليس نادرًا ويعد ذا قيمة عملية كبيرة في برامج التقييم التشغيلي (مثل التقييمات التعليمية على مستوى الولاية أو الوطن). كما هو الحال مع جميع النماذج الإحصائية تقريبًا، نادرًا ما يوجد نموذج يتوافق بشكل مثالي فالملاءمة دائمًا مسألة درجة.

ومن بين الدراسات القليلة التي تناولت عواقب عدم الملاءمة على مستوى الاختبار، أجرى (Lu, 2006, p. 28) دراسات محاكاة لفحص تأثير عدم الملاءمة على اختلاف الدرجات في تطبيقات المعادلة مع بيانات ثنائية، وخلص إلى أن النموذج اللوجستي ثنائي المعلمة والنموذج اللوجستي ثلاثي المعلمة أديا بشكل متساوٍ، بينما أدى النموذج اللوجستي أحادي المعلمة إلى أخطاء كبيرة في تحويل المعادلة.

ويشير (Zhao & Hambleton, 2017, p. 2) عمليًا إلي أنه قد يكون النموذج الذي لا يظهر توافقًا جيدًا قابلًا للاستخدام في تطبيق معين، وقد يُفضل النموذج غير المتوافق (أو المحدد بشكل سيئ) لأسباب عملية مثل البساطة وتوافر البرمجيات والكفاءة من حيث التكلفة والوقت، ومع إضافة المزيد من المعلمات إلى النموذج، يميل النموذج الأكثر تعقيدًا في التسلسل

الهرمي للنماذج إلى تحقيق توافق أفضل مقارنة بالنموذج الأبسط، ولكن هذه المعلمات قد تكون أقل استقرارًا عند تكرار التحليل، وبمعنى آخر يتضمن قرار اختيار النموذج موازنة بين تعقيد النموذج/استقرار المعلمات ودرجة توافق النموذج مع البيانات.

ومما سبق يستنتج الباحثان أنه من خلال الاطلاع علي البحوث والدراسات التي تناولت عدم ملائمة البيانات للنموذج أنه يوجد ثلاثة أنوع من عدم التوافق هي؛ عدم توافق النموذج للمفردات وعدم توافق النموذج للأفراد وعدم توافق النموذج للاختبار، كما أنه نستخلص أنه ليس من الشرط أنه عندما يكون النموذج غير ملائم للبيانات أنه يتم استبعاد النموذج لأنه في بعض الأحيان يقوم النموذج بحذف مفردات تكون مهمة وأساسية لواضعي الاختبار ويجب الإبقاء عليها في ضوء رؤيته لأهمية هذه المفردات ولخدمتها للاختبار ولقياس مهام معينة أو لتغطيتها لجزء هام في المحتوى أو أن حذفها قد يؤثر علي ثبات وصدق الاختبار، وبالتالي فان اختيار النموذج يتضمن موازنة بين تعقيد النموذج واستقرار المعلمات ودرجة توافق النموذج مع البيانات.

بحوث ودراسات سابقة:

في هذا الجزء سيتم عرض البحوث والدراسات التي تناولت موضوعات لها صلة وثيقة بالبحث الحالي:

دراسة (Loken & Rulison, 2010) هدفت إلى استكشاف نموذج نظرية الاستجابة للمفردة رباعي المعلم، في الاختبارات التعليمية، يبدو أن الحاجة إلى نمذجة المقارب الأعلى (أقل من ١) (النموذج الرباعي المعلم) أقل ضرورة. في اختبار المخاطر العالية (high-stakes testing)، يُفترض أن الطالب يزيد من بذله لمجهوده الذهني إلى الحد الأقصى، وبالتالي إذا تمكن من الإجابة على السؤال بشكل صحيح، فسيفعل ذلك. ومع ذلك، فقد وجد أن خفض المقاربة العلوية يقلل من التحيز في خوارزميات الاختبار التكيفي المحوسب عندما يستجيب الطلاب ذوي القدرة المرتفعة على الأسئلة الخطأ في وقت مبكر من الاختبار. ونحن نعتقد أنه سيكون هناك سياقات اختبارات تعليمية أخرى حيث قد تكون مفيدة، بما في ذلك استجابات النمذجة في مواقف اختبارية غير المرتفعة المخاطر (مثل اختبارات الممارسة (practice tests) أو التطبيقية أو الاختبارات التي يطبقها التلميذ على بعض الموضوعات كتطبيق أ كممارسة على درس ما أو مواد الدراسة التكيفية المحوسبة (computer adaptive study materials)، واستخدمت المنهج البايزي (Bayesian approach) لاستعادة تقديرات

المعلمات بنجاح للمفردات والمستجيبين. بالنسبة للبيانات التي تم إنشاؤها باستخدام النموذج رباعي المعلم، يتم تحسين الملاءمة الإجمالية عند استخدام المعلمات الأربعة بدلاً من المعلمات الثلاثية أو الثنائية. علاوة على ذلك، على الرغم من أن درجات السمات المقدرة في ظل النماذج المختلفة ترتبط بشكل مثالي تقريبًا، إلا أن الاستدلالات في الأطراف العالية والمنخفضة على متصل السمات تتعرض للخطر، مع تغطية رديئة لفترات الثقة عند استخدام النموذج الخطأ. نظهر أيضًا في مثال تجريبي على المعلمات الأربع يمكن أن تسفر عن رؤى جديدة في خصائص مقياس الجنوح (delinquency scale) المستخدم على نطاق واسع.

دراسة (Świst, 2015) هدفت إلي استكشاف ما إذا كان من الممكن استخدام نموذج 4PL لاكتشاف العيوب في صياغة الأسئلة (والتي قد تُدخل تباينًا غير ذي صلة بالبنية المقاسة داخل الاختبار)، وكان عدد العينة (٣٩٣٨٣٧) طالباً لعام ٢٠١٢ و (٣٧٩٧٥٢) طالباً لعام ٢٠١٣ و (٣٦٢٧٥٥) لعام ٢٠١٤، وقد تم تنفيذ التحليل على مرحلتين: تحليل نوعي: تقييم مدى التزام الأسئلة بإرشادات صياغة الأسئلة المعتمدة، حليل كمي: وفيت تم تطبيق نموذج 4PL ومقارنة نتائجه مع نتائج التحليل النوعي – للتحقق مما إذا كانت نفس الأسئلة تم تحديدها كمعيبة، وتم أيضًا استخدام نماذج IRT الأخرى (3PL و2PL) لاختبار صحة النتائج، وتم استخدام أوراق الامتحانات الوطنية البولندية في القراءة والكتابة كمادة تحليلية وتحديدًا جزء القراءة الذي يتضمن عادة ٢٠ سؤال اختيار من متعدد ويتطلب تحليلًا نقديًا للنصوص مثل المقالات أو مقتطفات من كتب أو قصائد، وتصميم هذه المفردات يجب أن يتم بحذر شديد، لأن أي تفسير خاطئ للسؤال قد يؤثر على اختيار الإجابة وبالتالي يؤدي إلى تقدير خاطئ لقدرة الطالب، وتوصلت النتائج إلى أنه يمكن اكتشاف العيوب في الأسئلة سواء من خلال التحليل النوعي أو باستخدام نموذج 4PL وحتى من خلال النماذج الأبسط مثل 2PL.

دراسة (محمد محمد جاد محمد, ٢٠١٦) هدفت إلي المقارنة بين النماذج اللوغاريتمية الثلاث لنظرية الاستجابة للمفردة (أحادي/ الثنائي/ الثلاثي) البارامتر في تدريج بنك للأسئلة الموضوعية في مادة الرياضيات للصف الخامس الابتدائي، وتكونت العينة من (١٥٠٠) تلميذاً وتلميذة في الصف الخامس الابتدائي، قام الباحث بعمل (١٠) صور اختبارية من نوع الاختيار من متعدد في مادة الرياضيات للصف الخامس الابتدائي في وحدة "الكسور والعمليات المختلفة عليها"، وتم استخدام برنامج (Parscale) الإصدار الرابع وبرنامج (spss) في إجراء

المعاملات الإحصائية اللازمة لنظرية الاستجابة للمفردة، وتوصلت النتائج إلي أن النموذج اللوغاريتمي الثلاثي المعالم هو أكثر النماذج ملائمة للبيانات وأكثرها دقة في تقدير وتدريج صعوبة المفردة.

دراسة (Waller & Feuerstahler, 2017) في هذه الدراسة، استكشفنا استعادة معلمات المفردات والأفراد لنموذج الأربعة معلمات في أكثر من ٢٤,٠٠٠ مجموعة بيانات حقيقية، واقعية، ومثالية. في التحليلات الأولى، قمنا بتطبيق نموذج الأربع معلمات وثلاثة نماذج بديلة على بيانات من ثلاثة مقاييس لعوامل نموذج الشخصية المتعددة الأوجه للمراهقين في ولاية مينيسوتا باستخدام التقدير البايزي النمط (Bayesian modal estimation (BME)). أشارت نتائجنا إلى أن النموذج الرباعي المعلم يناسب هذه المقاييس أفضل من نماذج نظرية الاستجابة للعناصر (IRT) الأبسط. بعد ذلك، باستخدام تقديرات المعلمات من تحليلات البيانات الحقيقية هذه، قمنا بتقدير معلمات العناصر للنموذج الرباعي المعلم في ٦,٠٠٠ مجموعة بيانات واقعية لتحديد متطلبات الحد الأدنى لحجم العينة لاستعادة دقيقة لمعلمات المفردات والأفراد. باستخدام تصميم عاملي يجمع بين مستويات منفصلة من معلمات المفردات، وحجم العينة، وطول الاختبار، قمنا أيضًا بتطبيق النموذج الرباعي المعلم على ١٨,٠٠٠ مجموعة بيانات مثالية إضافية لتوسيع نتائج استعادة المعلمات لدينا. أظهرت نتائجنا المجمعة أن معلمات المفردات في النموذج الرباعي المعلم ودوال المعلمات (مثل دوال استجابة المفردة) يمكن تقديرها بدقة باستخدام التقدير البايزي النمط (Bayesian modal estimation (BME)) في عينات متوسطة إلى كبيرة ($N \geqslant 5,000$) ويمكن تقدير معلمات الأفراد بدقة في عينات أصغر ($N \geqslant 1,000$).

دراسة (Yalçın, 2018) هدفت الدراسة إلي تحديد أفضل نموذج [IRT (Rasch) ، 2PL، 3PL، 4PL و IRT المختلط (2PL و 3PL)] للاختبار الفرعي للعلوم والتكنولوجيا لامتحان الانتقال من التعليم الأساسي إلى التعليم الثانوي (TEOG)، والتنبؤ بمعلمات المفردة في ظل أفضل نموذج، وتكونت العينة من ٥٠٠٠ طالب تم اختيارهم عشوائيًا من الطلاب الذين شاركوا في امتحان TEOG لعام ٢٠١٥، وتم إجراء التحليلات على ١٧ مفردة متعددة الاختيارات في اختبار TEOG الفرعي للعلوم والتكنولوجيا، وتم اجراء التحليلات الإحصائية، وتوصلت النتائج إلي أن نموذج MixIRT مع معاملين وثلاث فئات كامنة لديه أفضل قيم ملاءمة لبيانات النموذج.

دراسة (Erdemir & Önen, 2019) هدفت إلي مقارنة تقديرات المفردات ومعلمات القدرة لـ 1PL و2PL و3PL و4PL، وتكونت العينة من ١٥٠٠ طالب، تم أخذ الاختبار الفرعي التركي SBS (امتحان القبول بالمدارس الثانوية) لعام ٢٠١٢ من وزارة التعليم، وتم إجراء التحليلات الإحصائية، وتوصلت النتائج إلي أن 4PL كان النموذج الأكثر ملاءمة من 1PL و2PL و3PL لهذه الدراسة.

دراسة (Doğruöz & Arıkan, 2020) هدفت الدراسة إلي مقارنة طرق تقدير القدرات المختلفة بناءً على نظرية الاستجابة للمفردة 3PL و 4PL، وتكونت العينة من مجموعة ٤٠٠ طالب من طلاب الصف الثامن في عام ٢٠١٥-٢٠١٦ تم اختيارهم عشوائيًا من الطلاب المشاركين في امتحان TEOG، تم تطبيق ٢٠ مفردة ضمن الاختبار الفرعي للرياضيات لامتحان TEOG (الانتقال الوطني من التعليم الابتدائي إلى الثانوي)، تم حساب تقديرات القدرة وقيم الخطأ المعياري لهذه التقديرات بناءً على البيانات حيث تمت مقارنة هذه التقديرات من خلال تحليل التباين ثنائي الاتجاه (ANOVA) للقياسات المتكررة وفقًا لنتائج البحث، وتوصلت النتائج إلي أن نموذج طريقة تقدير WLE كان أفضل خوارزمية لمعلمات قدرة 4PL، تبين أن نموذج العنصر اللوجستي رباعي المعلمات (4PL) يناسب بشكل أفضل، من حيث طرق تقدير القدرة، حيث كانت دقة تقدير الاحتمالية المرجحة (WLE) أعلى من طريقة Maximum A Posteriori (MAP) وطريقة Expected A Posteriori (EAP)، وأعطى نموذج تقدير القدرة WLE وMAP قيم خطأ معيارية أقل مقارنة بنموذج 4PLM و 3PLM على التوالي، وتم حساب أعلى قيمة لمعامل الموثوقية الهامشي لنموذج 3PLM باستخدام التقديرات التي تم إجراؤها وفقًا لـ MAP بينما تم استخدام التقديرات التي تم إجراؤها وفقًا لـ WLE لنموذج 4PL.

دراسة (Kalkan, 2022) هدفت إلي مقارنة طرق التقدير لنموذج نظرية الاستجابة للمفردة اللوجستية رباعي المعلمات، وكذلك تقييم الأداء لطرق expectation-maximization (EM) و(Quasi-Monte Carlo EM (QMCEM وطريقة تقدير ميتروبوليس-هاستينجز روبنز-مونرو(Metropolis-Hastings Robbins-(MH-RM Monro لمعلمات المفردات في نموذج نظرية الاستجابة للمفردة ل 4PL في ظل ظروف الدراسة، بما في ذلك عدد العوامل والارتباط بين العوامل وطول الاختبار، تم إعادة الأخذ في الاعتبار نموذج نظرية استجابة المفردة ذات الأربعة المعلمات في أدبيات البحوث بسبب التقدم

في برامج النمذجة الإحصائية والتطورات الحديثة في تقدير معلمات النموذج الرباعي المعلم. قيمت دراسة المحاكاة الحالية أداء (expectation-maximization (EM), Quasi-Monte Carlo EM (QMCEM), and Metropolis-Hastings Robbins-Monro (MH-RM) estimation methods) لمعلمات المفردة في النموذج الرباعي المعلم تحت المعالجة وفقاً لظروف الدراسة، بما في ذلك عدد العوامل، والعلاقة بين العوامل، وطول الاختبار. أشارت النتائج إلى أنه لا توجد طريقة يجب التوصية بها كأفضل طريقة من بين خوارزميات التقدير الثلاثة لتقدير معلمات المفردة وفقاً للنموذج الرباعي المعلم بدقة في جميع ظروف الدراسة. ومع ذلك، يمكن اقتراح استخدام خوارزمية (Metropolis-Hastings Robbins-Monro (MH-RM)) لتقدير بارامترات المفردة عندما يكون عدد العوامل ٢ أو ٣. بالإضافة إلى ذلك، قد يُنصح بتفضيل أطوال الاختبار الطويلة بدلاً من أطوال الاختبار الأقصر (ن = ٢٤) نظرًا لأن الثلاث خوارزميات توفر تقديرات معلمة للمفردة أفضل في أطوال الاختبارات الطويلة (ن = ٤٨).

دراسة (Pardede et al., 2023) تم استخدام ثلاثة نماذج شائعة لوصف خصائص عناصر الاختبار وتقدير قدرة الممتحنين بموجب نموذج IRT ثنائية الاستجابة، وهي النماذج اللوجستية ذات المعلمة الواحدة والثنائية والثلاثية. المعلمات المكونة من ثلاثة عناصر هي القوة التمييزية، والصعوبة، والتخمين الزائف. في تطوير نموذج IRT ثنائية الاستجابة، تم اقتراح معلمة الإهمال أو الخط المقارب العلوي، والتي تشكل نموذجًا لوجستيًا مكونًا من أربع معلمات (4PL) لاستيعاب حالة حيث يعطي الممتحن ذو القدرة العالية استجابة غير صحيحة لعنصر الاختبار عندما يكون / هي يجب أن تكون قادرًا على الاستجابة لعنصر الاختبار بشكل صحيح. ومع ذلك، لم يتم قبول معلمة الإهمال ونموذج 4PL على نطاق واسع واستخدامها بسبب عدة عوامل، ولا يزال فهم الناس لتلك المعلمة واستراتيجيات تقديرها غير كافٍ. لذلك، تهدف هذه الدراسة إلى تسليط الضوء على الأفكار التي يقوم عليها نموذج 4PL، ومعنى معلمة الإهمال، والاستراتيجيات المستخدمة لتقدير تلك المعلمة بناءً على الأدبيات الموجودة. تم بعد ذلك توسيع نطاق تركيز هذه الدراسة لعرض أمثلة عملية لتقدير معلمات العناصر والأشخاص باستخدام نموذج 4PL باستخدام البيانات التجريبية حول استجابات ١٠٠٠ طالب من جامعة إندونيسيا المفتوحة (Universitas Terbuka) على ٢١ من ٣٠ عنصرًا متعدد الاختيارات على اختبار اللغة الإنجليزية للأعمال، اختبار الورقة والقلم. قمنا بشكل أساسي بتحليل البيانات التجريبية باستخدام حزمة "mirt" في RStudio. نقدم نتائج التحليل بشكل

متماسك بحيث يكون لدى مستخدمي IRT فهم كافٍ لنموذج 4PL ومعلمة الإهمال، ويمكنهم تقدير معلمات العنصر والشخص ضمن نموذج 4PL.

دراسة (Büyükkıdık & İnal, 2023) هدفت الدراسة إلي هدفت هذه الدراسة إلى التحقيق في تقدير معلمات نظرية استجابة العنصر (IRT) وثباتها وكذلك فحص افتراضات نظرية استجابة العنصر (IRT) أولاً وإجراء تقديرات المفردات والقدرة باستخدام 1PL و2PL و3PL و4PL للبيانات الثنائية، وتكونت العينة من ٧٥٠٠ طالب كمجموعة فرعية ذات عينة واحدة و٣٧٥٠ طالبًا موزعين على مجموعتين فرعيتين، تم تطبيق اختبار فرعي للرياضيات لعام ٢٠١٧ (أبريل) لامتحان الانتقال من التعليم الابتدائي إلى الثانوي (TPSEE)، وتم عمل التحليلات الإحصائية، وتوصلت النتائج إلي أنه عند فحص ملاءمة بيانات النموذج وجد أنه تم الحصول على أفضل ملاءمة باستخدام 3PL في جميع الظروف، ولوحظ أن معلمات المفردات لم تختلف بشكل كبير مع تغير العينة، وأن المعلمات a وb اختلفت وفقًا لنماذج IRT المختلفة، في حين أن هناك فرقًا جزئيًا بين معلمات القدرة مع تغير العينات، وهناك أيضًا اختلافات مع تغير النماذج، وقد لوحظت اختلافات طفيفة بين معلمات القدرة التي تم الحصول عليها من خلال طرق تقدير القدرة (EAP) و(MAP)، وكانت معاملات الثبات الهامشية متشابهة في جميع الظروف. يوصى بأن يقوم الباحثون بتقدير المعلمات التي لديها أفضل ملاءمة لبيانات النموذج من 3PL أو 4PL لتوفير مزيد من المعلومات أثناء إجراء التحليل في IRT.

تعليق عام على الدراسات والبحوث السابقة:

تُظهر الدراسات السابقة التي تم استعراضها اهتمامًا بحثيًا متزايدًا بنظرية الاستجابة للمفردة (IRT)، وتحديدًا المقارنة بين نماذجها المختلفة، مع تركيز ملحوظ على النموذج رباعي المعلمة (4PL) الذي بدأ يكتسب قبولًا أكبر بفضل التطورات في برامج التحليل الإحصائي وطرق التقدير. تتفق غالبية هذه الدراسات على أن النموذج رباعي المعلمة يقدم مرونة إضافية من خلال تضمين معلمة "الحد الأقصى للأداء" أو "الإهمال" (d)، مما يسمح بنمذجة أكثر دقة لاستجابات الأفراد ذوي القدرات العالية الذين قد لا يجيبون بشكل صحيح دائمًا، أو لوجود مشكلات في صياغة المفردات تؤثر على أداء الأفراد ذوي القدرة المرتفعة (النخبة).

يُلاحظ أن بعض الدراسات، مثل دراسة (Pardede et al., 2023)، ركزت على إلقاء الضوء على الأفكار الكامنة وراء نموذج 4PL ومعنى معلمة الإهمال، وتقديم أمثلة عملية

لتقدير المعلمات، مما يعكس الحاجة إلى فهم أعمق لهذا النموذج. كما أن دراسات أخرى مثل (Kalkan, 2022; Loken & Rulison, 2010; Waller & Feuerstahler, 2017) استكشفت طرقًا مختلفة لتقدير المعلمات في نموذج 4PLM، وناقشت متطلبات حجم العينة ودقة استعادة المعلمات؛ مما يشير إلى التحديات المنهجية المرتبطة بتطبيق هذا النموذج الأكثر تعقيدًا.

وعلى الرغم من أن بعض الدراسات مثل (محمد محمد جاد محمد, ٢٠١٦) وجدت أن نموذج 3PL هو الأفضل ملاءمة لبياناتها، إلا أن دراسات أخرى مثل (Erdemir & Önen, 2019; Yalçın, 2018) أشارت إلى أن 4PL أو النماذج المختلطة (MixIRT) قد توفر ملاءمة أفضل في سياقات معينة، مما يؤكد أن اختيار النموذج الأمثل يعتمد بشكل كبير على طبيعة البيانات وخصائص الاختبار. كما تُبرز دراسة (DoĞRuÖZ & Akin Arikan, 2020) أهمية مقارنة طرق تقدير القدرات المختلفة وتأثيرها على دقة القياس تحت نماذج 3PL و4PL.

وبشكل عام، تعكس هذه الدراسات الجدل المستمر حول الحاجة إلى استخدام نماذج IRT الأكثر تعقيدًا (مثل 4PL) مقابل النماذج الأبسط (مثل 3PL)، وتؤكد على أن الملاءمة الأفضل قد لا تعني بالضرورة الأداء الأفضل في جميع الظروف. كما أنها تسلط الضوء على أهمية الأساليب الإحصائية المتقدمة (مثل المنهج البايزي) وقوة برامج التحليل الحديثة في التعامل مع هذه النماذج. وتشترك هذه الدراسات بشكل مباشر مع البحث الحالي من خلال توفير سياق نظري وتطبيقي غني للمقارنة بين نموذجي 3PL و4PL في تدريج الاختبارات التحصيلية، مما يبرز أهمية البحث الحالي في إضافة رؤى جديدة لهذا المجال.

منهج البحث:

يتبع البحث الحالي المنهج الوصفي المقارن التحليلي، يهدف هذا المنهج إلى وصف وتحليل الفروق بين نموذجي نظرية الاستجابة للمفردة (ثلاثي ورباعي المعلمة) من حيث ملاءمة المفردات والاختبار الكلي، وتقدير معالم المفردات والقدرات، ودقة القياس، وخصائص الاختبار. يتم ذلك من خلال تطبيق النموذجين على بيانات اختبار تحصيلي واحد ومقارنة النتائج إحصائياً.

حدود البحث:

يقتصر البحث الحالي على مقارنة النموذجين ثلاثي ورباعي المعلمة لنظرية الاستجابة للمفردة. وتشمل حدوده الزمنية (عامين دراسيين متتاليين ٢٠٢٢/٢٠٢٣- ٢٠٢٣/٢٠٢٤) والمكانية (كلية التربية- جامعة الزقازيق). كما يقتصر البحث على الاختبار التحصيلي المستخدم في جمع البيانات، والعينة التي تم تطبيق الاختبار عليها.

عينة البحث:

يقتصر البحث على العينة النهائية نظراً لطبيعته التي اقتضت التطبيق للاختبار التحصيلي وتكونت عينة البحث النهائية من (١١٦٨) فرداً، تم اشتقاقهم من طلاب كلية التربية- جامعة الزقازيق ممن يدرسون مادة الاتجاهات الحديثة والمعاصرة في علم النفس، موزعين على عامين دراسيين متتاليين؛ وهم طلاب جميع شعب تعليم أساسي الفرقة الثانية للعام الجامعي ٢٠٢٢/٢٠٢٣م (علوم (٢٢٨)- لغة عربية (٦٣)- لغة إنجليزية (١٨٩)- مواد اجتماعية (١٩٩))، وكذلك طلاب جميع شعب تعليم أساسي الفرقة الثانية للعام الجامعي ٢٠٢٣/٢٠٢٤م (علوم (١٧٧)- لغة عربية (٤٥)- لغة إنجليزية (١٢٢)- مواد اجتماعية (١٤٠))، وعدد الطلاب الذين لم يحددوا شعبهم (٥) طلاب؛ لتكون عينة البحث النهائية من (١١٦٨) فرداً.

أدوات البحث:

يستخدم البحث الحالي اختبار تحصيلي مكون من (١٠٠) مفردة ثنائية الاستجابة (منها (٥٠) مفردة اختيار من متعدد من أربعة بدائل، (٥٠) مفردة صواب وخطأ)، وسيتم توضيحها تفصيلياً في إجراءات البحث.

إجراءات البحث:

الأساليب الإحصائية وحزم البرامج المستخدمة:

لتحقيق أهداف البحث واختبار فروضه، سيتم استخدام مجموعة من الأساليب الإحصائية المتقدمة المستندة إلى نظرية الاستجابة للمفردة. وسيتم تحليل البيانات وتقدير معالم النموذجين ثلاثي ورباعي المعلمة باستخدام حزم متخصصة في برنامج R (حزمة mirt)، بالإضافة إلى برامج إحصائية أخرى مثل SPSS لمعالجة البيانات الأولية وتحليلها.

نتائج البحث وتفسيرها ومناقشتها:

تحليل البيانات:

نصف في هذا القسم كيف تم إجراء التحليل الكمي للبيانات التجريبية حول درجات استجابة الطلاب في اختبار تحصيلي في مادة "الاتجاهات الحديثة (المعاصرة) في علم النفس التربوي"؛ والتي قررها أساتذة المادة على طلاب الفرقة الثانية لشعب التعليم الأساسي أدبي وعلمي وفقاً للائحة (٢٠٢٠) ووفقاً للائحة المعدلة النهائية التي طُبقت على طلاب كلية التربية على مستوى الجمهورية؛ حيث طُلب من الباحثين باعتبار أنهما المكلفان بتدريس المادة؛ طُلب منهما إعداد كتاب المقرر ونبعت فكرة هذا البحث أثناء إعداد أسئلة عن هذا المقرر وكل فصل فيه، ثم فكر الباحثان في كيفية تدريج مفردات الاختبار التحصيلي وفقاً لأحدث التقنيات الموجودة، وجاء نموذج الرباعي المعلمة وفقاً لنظرية الاستجابة للمفردة على رأس هذه النماذج الخاصة بالنظرية الحديثة لما يقدمه من معلمة حديثة هي معلمة الإهمال أو معلمة اللامبالاة والتي يحتاج الباحثون إلى معرفتها وتفسيرها؛ لعلها تضيف شيئاً حديثاً إلى أدبيات القياس النفسي، للكشف عن معلمات المفردات والأشخاص بناءً على النموذج الرباعي المعلمة وفقاً لنظرية الاستجابة للمفردة والمقارنة بينه وبين النموذج الثلاثي المعلمة وفقاً لمحكات المقارنة المستخدمة في أدبيات القياس النفسي الحديثة؛ والجدول (١) يوضح عدد فقرات الاختبارات التحصيلية بالنسبة لكل فصل ومسمى الفصل.

جدول (١) يوضح عدد فقرات كل اختبار تحصيلي على كل فصل من فصول الكتاب المقرر "الاتجاهات الحديثة (المعاصرة) في علم النفس التربوي"

م	مسمى الفصل	عدد فقرات الاختبار	النوع
١	سلوك رفض المدرسة	١٠٠	ثنائية الاستجابة (٠ أو ١)
٢	البيئة الأسرية	٨١	ثنائية الاستجابة (٠ أو ١)
٣	المهارات الاجتماعية	١٠٠	ثنائية الاستجابة (٠ أو ١)
٤	المخاوف المرتبطة بكوفيد-١٩	١٠٠	ثنائية الاستجابة (٠ أو ١)
٥	العوامل الخمسة الكبرى في الشخصية	١٠٠	ثنائية الاستجابة (٠ أو ١)

ملحوظة: الفصل الأول هو الفصل الذي سيتم عرض نتائجه وفقاً لكل من النموذجين ثلاثي ورباعي المعلمة.

فيما يلي عرض للخطوات المتبعة للوصول إلى النتائج الخاصة بالدراسة، وسيتم إجراء

التحليل التقليدي (الكلاسيكي) على الفصل الأول، ويتم إجراء التحليل القائم على نظرية السمات الكامنة (الاستجابة للمفردة) للتحقق من استيفاء الافتراضات الثلاثة الرئيسية لنموذج IRT على الفصل الأول. كما أنه سيتم الإجابة عن أسئلة الدراسة استناداً إلى الفصل الأول.

التحليل التقليدي (الكلاسيكي) بالنسبة لاختبار الفصل الأول "سلوك رفض المدرسة":

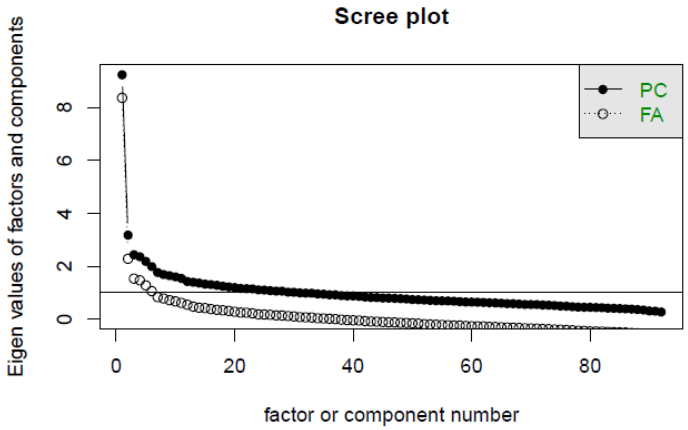
تم تحليل مفردات الاختبار (١٠٠) مفردة ثنائية الاستجابة (منها (٥٠) مفردة اختيار من متعدد من أربعة بدائل، (٥٠) مفردة صواب وخطأ) باستخدام النظرية التقليدية في القياس (القياس الكلاسيكي) ((Classical Test Theory (CTT) مع التركيز على اكتشاف مفردات الاختبار التي فشلت في تمييز الطلاب ذوي القدرات العالية عن الطلاب ذوي القدرات المنخفضة في الاختبار كما يشير إلى ذلك (معامل الارتباط الثنائي الأصيل) (Point Biserial Correlation) لهذه المفردات وهذه المفردات تعتبر مفردات لابد من حذفها. وتم إجراء هذا التحليل في RStudio (Team, 2025)، وهي بيئة تطوير متكاملة (IDE) لـ R، باستخدام حزمة (Willse, 2018) "CTT". وتم استخدام برنامج R للتحقق من خصائص الاختبار ومفرداته بناءً على النظرية التقليدية في القياس (وكانت درجة ثبات الاختبار بناءً على ألفا كرونباخ (٠.٨٤٢)، والخطأ المعياري في القياس (SEM) (٣.٣٩٤)، وقوة تمييز المفردة بناءً على معامل الارتباط الثنائي الأصيل أسفرت عن وجود ثماني مفردات سالبة وهي أرقام (٣٦-٥٦-٦٢-٦٥-٦٦-٦٨-٨٧-٩٧)) وبالتالي، باستثناء هذه المفردات الثمان في التحليل اللاحق، (وأصبحت درجة ثبات الاختبار بناءً على ألفا كرونباخ (٠.٨٧٩)، والخطأ المعياري في القياس (SEM) (٣.١١٧))، وتكوَّن الاختبار في صورته النهائية التي ستدخل في التحليل الحديث وفقاً للنموذجين الثلاثي والرباعي المعلمة من (٩٢) مفردة ثنائية الاستجابة (منها (٤٩) مفردة اختيار من متعدد من أربعة بدائل، (٤٣) مفردة صواب وخطأ).

التحليل القائم على نظرية السمات الكامنة (الاستجابة للمفردة) بالنسبة لاختبار الفصل الأول "سلوك رفض المدرسة":

استيفاء الافتراضات الثلاثة الرئيسية لنموذج IRT

لضمان دقة تقديرات معلمات المفردات والأفراد بموجب IRT، سيتم التحقق من استيفاء الافتراضات الثلاثة الرئيسية لنموذج IRT، وهي أحادية البعد، واللاتباين (ثبات) المعلمة، والاستقلال الموضعي.

تم دراسة افتراض أحادية البعد من خلال تحليل العوامل والمكونات الرئيسية كما بالشكل (١) باستخدام الحزمة "psych" package (Revelle, 2023) ومن خلال النظر في المؤشرات والقواعد المختلفة في تحديد عدد العوامل المهيمنة التي لخصها (Hattie, 1985) و (Slocum-Gori & Zumbo, 2011)، وقد تم إجراء التحليل العاملي للمحور الرئيسي "principal axis factor analysis" بتدوير الفاريماكس "varimax rotation"؛ لتقييم البنية الكامنة "underlying structure" لمفردات الاختبار، وتم اختبار افتراضات التحليل العاملي من خلال "اختبار بارتليت (مربع كاي التقريبية) و كايزر – ماير – أولكن" " KMO and Bartlett's Test"؛ وكان قيمة اختبار بارتليت و كايزر – ماير – أولكن تساوي (٠.٨٣) وهي أعلى من (٠.٧٠)، وكانت قيمة اختبار بارتليت (مربع كاي التقريبية) (٢٠٩٤٠.٨٩) وهي دالة عند (٠.٠١)؛ مما يدل على تحقق افتراضات التحليل العاملي (Leech et al., 2005, p. 80).



شكل (١) تمثيل الجذور الكامنة من خلال المكونات الرئيسية وتحليل العوامل

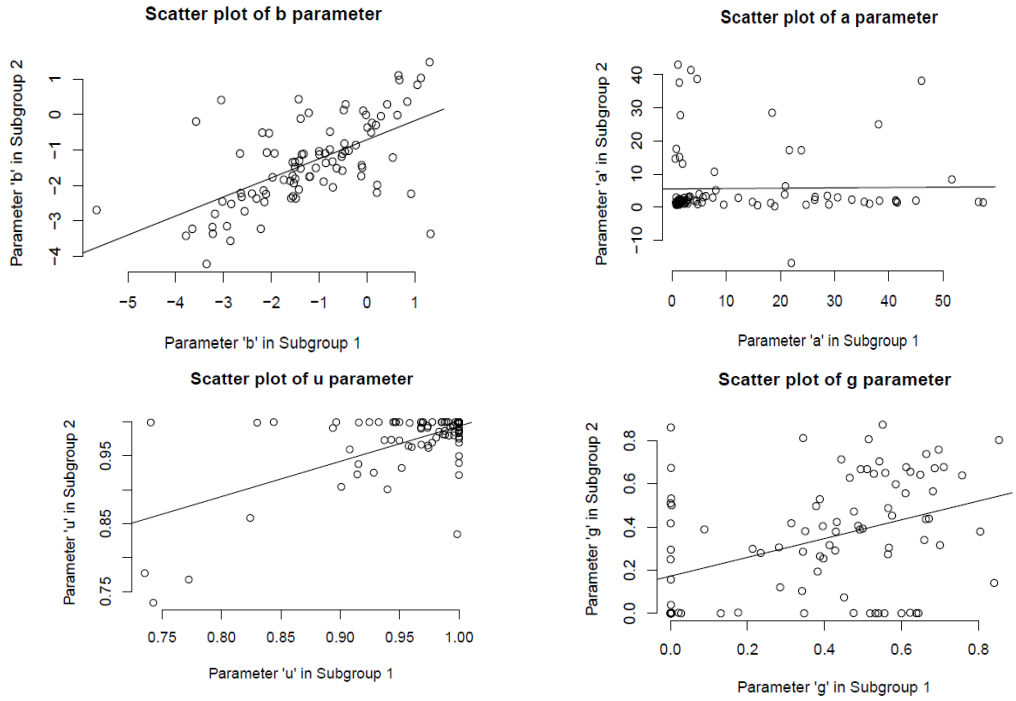
تم دراسة اللاتباين في المعلمات (ثبات المعلمة) باستخدام معامل ارتباط بيرسون لإثبات أن تقديرات معلمات المفردات والأفراد للمجموعتين الفرعيتين متطابقة أو لا تختلف كثيرًا

الافتراض الثاني الأساسي لـ IRT هو ثبات المعلمة؛ فهناك مستويان من المعلمات التي نحتاجها لإثبات أنهما لا يتغيران، وهما معلمة المفردة ومعلمة الفرد. ويشير ثبات معلمة المفردة إلى أن معلمات مفردة الاختبار، التي يعتمد فيها عدد معلمات المفردة على نموذج IRT المناسب، لا تتأثر بتنوع أو توزيع قدرات الممتحنين (Hambleton & Swaminathan, 1985; Ronald K. Hambleton et al., 1991). وهذا يعني أن نتائج تقدير المعلمة

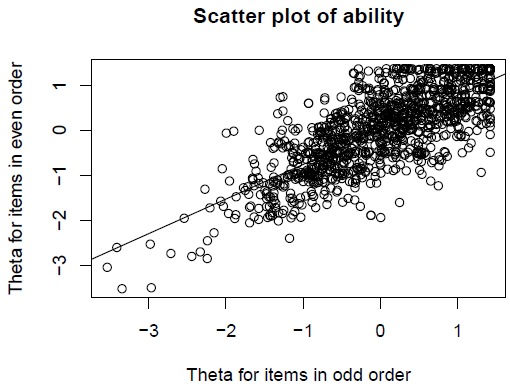
لمفردة ما لن تتغير، أو تميل إلى أن تكون متطابقة، أو لا تختلف كثيرًا عندما يتم إعطاء المفردة إلى مجموعات فرعية مختلفة في مجموعة من الأفراد أو سياقات مختلفة (Ronald K. Hambleton et al., 1991; Rupp & Zumbo, 2004). ولأنه سيتم استخدام النموذجين ثلاثي ورباعي المعلمات في هذا البحث؛ فسيتم تقديم دليل على استيفاء افتراض ثبات معلمات المفردة بناءً على أربع معلمات لتشمل أكبر نموذج من حيث عدد المعلمات، وهي (التمييز والصعوبة والتخمين وعدم الاهتمام). حصلنا على دليل على ثبات المعلمات الأربعة من خلال تشكيل مجموعتين فرعيتين تم تشكيلهما بشكل عشوائي من بيانات الممتحنين في هذه الدراسة بنسبة ٠.٥ لكل مجموعة فرعية بحيث تم الحصول على مجموعتين فرعيتين بنفس الحجم أو تقريبًا نفس الحجم. قمنا بعد ذلك بإجراء تحليل الارتباط على نتائج تقدير معلمة المفردة من كل مجموعة فرعية للتحقق من قوة العلاقة بين نتيجتي تقدير المعلمة. يوضح تحليل ارتباط بيرسون وجود علاقة إيجابية قوية (Schober et al., 2018) بين تقديرات معلمة التمييز a (r(90) = 0.016 ، p = 0.877)، وتقديرات معلمة الصعوبة b (r(90) = 0.607 b، p = 0.000)، تقديرات معلمة التخمين g (r(90) = 0.416، p = 0.000)، وتقديرات معلمة الإهمال u (r(90) = 0.621، p = 0.000) المقدرة من المجموعة الفرعية الأولى وتلك المقدرة من المجموعة الفرعية الثانية. والشكل (٢) مخطط التشتت الذي يوضح توزيع النتائج المقدرة لكل معلمة مفردة مقدرة من المجموعتين الفرعيتين. وبالتالي، تم استيفاء افتراض الثبات (اللاتباين) على معلمات المفردات.

باستخدام نفس خطوات ثبات معلمة المفردة، تشير ثبات معلمة الفرد (القدرة) إلى اتساق تقديرات قدرة الممتحن التي تم الحصول عليها من مجموعتين فرعيتين من مفردات الاختبار، حيث يمكن أن يتم تكوين مجموعتين فرعيتين من مفردات الاختبار ويتم تشكيلهما بناءً على شروط معينة (Ronald K. Hambleton et al., 1991; Rupp & Zumbo, 2004) الشرط الذي استخدمناه في هذه الدراسة لتكوين مجموعتين فرعيتين من مفردات الاختبار هو ترتيب مفردات الاختبار، حيث تتكون المجموعة الفرعية الأولى من مفردات الاختبار ذات الرتبة الفردية وتتكون المجموعة الفرعية الثانية من مفردات الاختبار ذات الرتبة الزوجية. بافتراض أن المفردات تتبع النموذج رباعي المعلمة وباستخدام expected a posteriori (EAP)، وهي طريقة تقدير درجة العامل الافتراضية في حزمة "mirt"، لتقدير قدرات الممتحنين، تم الحصول على توزيع قدرات الممتحنين المقدرة بناءً على المفردات ذات الترتيب الفردي وعلى المفردات ذات الترتيب الزوجي كما يظهر بالشكل (٣). يُظهر تحليل ارتباط

بيرسون أن هناك علاقة إيجابية قوية (Schober et al., 2018) بين تقديرات قدرة الممتحن المقدرة من مجموعة فرعية من المفردات بترتيب فردي وتلك المقدرة من مجموعة فرعية من المفردات بترتيب زوجي ($r(1166) = 0.757$، $p = 0.000$). ومن ثم، فقد تم أيضًا استيفاء افتراض اللاتباين (الثبات) على معلمة الفرد أو القدرة (θ).



شكل (٢) مخطط تشتت توزيع بارامترات (التمييز والصعوبة والتخمين وعدم الاهتمام) المقدرة من مجموعتين فرعيتين مختلفتين



شكل (٣) مخطط تشتت توزيع قدرات الممتحنين المقدرة من مجموعة المفردات فردية الرتبة ومجموعة المفردات زوجية الرتبة

وباستخدام حزمة (Chalmers, 2012) "mirt" من الممكن اكتشاف الاستقلال

الموضعي لزوج من المفردات بأربعة إحصائيات، وهي (Yen's Q3) (Yen, 1984) ؛ حيث أن قيمة الإرباعي الثالث تساوي (٠.٠٢١)؛ مما يعني أنه على الأقل ٧٥٪ من قيمة تلك الإحصاءة أقل من هذه القيمة؛ مما يعنى أن القيمة المطلقة للمعامل معظمها تقل عن (٠.٢)؛ مما يدل على توفر شرط استقلالية الموضع أو المحل، (Chen & Thissen, 1997) ($\chi 2$)، (likelihood-ratio (G2)) (Chen & Thissen, 1997)؛ حيث أن قيمة الإرباعي الثالث للإحصاءتين تساوي (٠.٠٢٠)؛ مما يعني أنه على الأقل ٧٥٪ من قيمة الإحصاءتين أقل من هذه القيمة؛ مما يعنى أن القيمة المطلقة للمعامل معظمها تقل عن (٠.٣)؛ مما يدل على توفر شرط استقلالية الموضع أو المحل، (jack-knife slope index (JSI)) (Edwards et al., 2018)؛ حيث أن قيمة الإرباعي الثالث للإحصاءة تساوي (٠.٠٦٢)؛ مما يعني أنه على الأقل ٧٥٪ من قيمة الإحصاءة أقل من هذه القيمة كما بالجدول (٢)؛ مما يعنى أن القيمة المطلقة للمعامل معظمها تقل عن (٠.٧٣٨)؛ مما يدل على توفر شرط استقلالية الموضع أو المحل.

جدول (٢) ملخص إحصاءات الكشف عن الاستقلال الموضعي

الإحصاءة	أقل درجة	الإرباعي الأول	الوسيط	المتوسط	الإرباعي الثالث	أعلى درجة
Yen's Q3	-٠.١٩١	-٠.٠٤٣	-٠.٠١٣	-٠.٠٠٧	٠.٠٢١	٠.٥٣٣
Pearson's $\chi 2$	-٠.١٤٤	-٠.٠٣١	-٠.٠٠٨	-٠.٠٠٣	٠.٠٢٠	٠.٣٤٤
Likelihood-ratio G2	-٠.١٤٨	-٠.٠٣٢	-٠.٠٠٨	-٠.٠٠٤	٠.٠٢٠	٠.٣٥٢
Summed JSI	-٠.٨٤٦	-٠.٠٧٠	-٠.٠١١	٠.٠٠٤	٠.٠٦٢	٢.٩٦٨

JSI = Jack-knife slope index

السؤال الأول: ما أفضل نموذج (النموذج ثلاثي المعلم- النموذج رباعي المعلم) والذي يحقق ملاءمة لمفردات الاختبار التحصيلي وفقاً لمحك المقارنة (the signed chi-squared ($\chi 2$) statistics)؟

للإجابة عن هذا السؤال؛ تم التحليل على اختبار الفصل الأول "سلوك رفض المدرسة": تم اختبار مدى ملاءمة استجابات الطلاب لمفردات الاختبار باستخدام نموذجين من نماذج الاستجابة للمفردة هي (النموذج ثلاثي المعلم- النموذج رباعي المعلم). تم إجراء هذا الاختبار، الذي يسمى اختبار ملاءمة المفردة أو اختبار مدى ملاءمة المفردات (the item fit test or

the goodness-of-fit test of items)، باستخدام حزمة (Chalmers, 2012) "mirt"، حيث مع هذه الحزمة، يتم تقييم اختبار ملاءمة المفردة افتراضيًا باستخدام الإحصاءة (the signed chi-squared (s-χ2) statistics) والمقترحة من قبل (Orlando & Thissen, 2000, 2003)، حيث يُقال إن المفردة تتناسب مع النموذج قيد التحقق عندما تكون الدلالة الإحصائية لاختبار الملاءمة سابق الذكر غير دالة (أي: الدلالة أكبر من أو تساوي ٠.٠٥). وكلما زاد عدد المفردات التي تناسب النموذج، يكون النموذج أكثر تفضيلاً من النماذج الأخرى لتقدير معلمات المفردة والأفراد، والجدول (٣) يوضح أن النموذج الثلاثي المعلم يحقق ملاءمة لعدد أكبر من المفردات (٨٧) مفردة بنسبة مئوية (٩٤.٥٧٪) عن النموذج الرباعي المعلم.

جدول (٣) ملخص تحليل ملائمة المفردات

النموذج المختبر	عدد المفردات الملائمة للنموذج	النسبة المئوية
النموذج ثلاثي المعلم	٨٧	% ٩٤.٥٧
النموذج رباعي المعلم	٦١	% ٦٦.٣٠

جدول (٤) التحليل التفصيلي لملائمة المفردات وفقاً للنموذجين الثلاثي والرباعي

م	المفردة	النموذج ثلاثي المعلم			النموذج رباعي المعلم		
		s-χ2	الدلالة	الملاءمة مع النموذج	s-χ2	الدلالة	الملاءمة مع النموذج
1	q1	50.566	0.054	ملائمة	51.088	0.049	غير ملائمة
2	q2	35.696	0.097	ملائمة	34.733	0.072	ملائمة
3	q3	29.795	0.124	ملائمة	30.535	0.082	ملائمة
4	q4	51.078	0.093	ملائمة	109.19	0	غير ملائمة
5	q5	27.235	0.246	ملائمة	22.261	0.271	ملائمة
6	q6	19.633	0.877	ملائمة	20.118	0.826	ملائمة
7	q7	46.715	0.072	ملائمة	50.469	0.044	غير ملائمة
8	q8	21.35	0.377	ملائمة	28.313	0.102	ملائمة
9	q9	49.002	0.109	ملائمة	54.795	0.023	غير ملائمة
10	q10	33.779	0.382	ملائمة	35.266	0.273	ملائمة
11	q11	29.237	0.78	ملائمة	38.096	0.374	ملائمة
12	q12	42.052	0.11	ملائمة	45.748	0.069	ملائمة
13	q13	8.39	0.678	ملائمة	8.475	0.583	ملائمة
14	q14	16.552	0.957	ملائمة	18.819	0.844	ملائمة
15	q15	32.21	0.604	ملائمة	33.963	0.421	ملائمة

م	المفردة	النموذج ثلاثي المعلم			النموذج رباعي المعلم		
		s-χ2	الدلالة	الملاءمة مع النموذج	s-χ2	الدلالة	الملاءمة مع النموذج
16	q16	61.575	0.005	غير ملائمة	61.695	0.005	غير ملائمة
17	q17	21.433	0.124	ملائمة	25.689	0.139	ملائمة
18	q18	19.671	0.291	ملائمة	32.581	0.019	غير ملائمة
19	q19	28.355	0.292	ملائمة	30.164	0.218	ملائمة
20	q20	38.194	0.326	ملائمة	40.311	0.211	ملائمة
21	q21	45.734	0.069	ملائمة	50.605	0.033	غير ملائمة
22	q22	32.274	0.308	ملائمة	33.916	0.204	ملائمة
23	q23	54.729	0.018	غير ملائمة	54.159	0.015	غير ملائمة
24	q24	52.598	0.058	ملائمة	67.096	0.001	غير ملائمة
25	q25	29.503	0.642	ملائمة	34.483	0.397	ملائمة
26	q26	26.986	0.76	ملائمة	33.171	0.362	ملائمة
27	q27	35.597	0.303	ملائمة	35.828	0.293	ملائمة
28	q28	34.441	0.635	ملائمة	34.022	0.563	ملائمة
29	q29	29.341	0.738	ملائمة	29.045	0.664	ملائمة
30	q30	3.542	0.472	ملائمة	5.554	0.475	ملائمة
31	q31	37.942	0.518	ملائمة	104.6	0	غير ملائمة
32	q32	40.193	0.331	ملائمة	55.894	0.018	غير ملائمة
33	q33	21.869	0.081	ملائمة	24.805	0.037	غير ملائمة
34	q34	48.549	0.079	ملائمة	43.478	0.105	ملائمة
35	q35	32.941	0.615	ملائمة	40.82	0.23	ملائمة
36	q37	32.71	0.626	ملائمة	35.464	0.494	ملائمة
37	q38	41.91	0.165	ملائمة	45.794	0.127	ملائمة
38	q39	38.15	0.21	ملائمة	39.146	0.18	ملائمة
39	q40	25.783	0.872	ملائمة	26.761	0.729	ملائمة
40	q41	26.985	0.719	ملائمة	29.807	0.476	ملائمة
41	q42	47.138	0.066	ملائمة	50.368	0.035	غير ملائمة
42	q43	36.403	0.357	ملائمة	38.932	0.257	ملائمة
43	q44	33.38	0.352	ملائمة	35.862	0.336	ملائمة
44	q45	53.44	0.018	غير ملائمة	61.783	0.002	غير ملائمة
45	q46	37.388	0.199	ملائمة	36.738	0.259	ملائمة
46	q47	45.64	0.088	ملائمة	50.232	0.028	غير ملائمة

م	المفردة	النموذج ثلاثي المعلم			النموذج رباعي المعلم		
		s-χ2	الدلالة	الملاءمة مع النموذج	s-χ2	الدلالة	الملاءمة مع النموذج
47	q48	45.813	0.068	ملائمة	45.318	0.06	ملائمة
48	q49	25.508	0.853	ملائمة	29.074	0.708	ملائمة
49	q50	36.38	0.232	ملائمة	33.04	0.321	ملائمة
50	q51	44.19	0.137	ملائمة	49.466	0.067	ملائمة
51	q52	37.37	0.2	ملائمة	42.103	0.16	ملائمة
52	q53	50.648	0.053	ملائمة	55.019	0.022	غير ملائمة
53	q54	47.886	0.058	ملائمة	55.662	0.008	غير ملائمة
54	q55	9.881	0.98	ملائمة	20.203	0.57	ملائمة
55	q57	46.302	0.117	ملائمة	414.72	0	غير ملائمة
56	q58	30.658	0.76	ملائمة	51.5	0.045	غير ملائمة
57	q59	29.091	0.662	ملائمة	31.872	0.473	ملائمة
58	q60	36.807	0.297	ملائمة	44.347	0.16	ملائمة
59	q61	44.018	0.169	ملائمة	51.08	0.049	غير ملائمة
60	q63	45.158	0.198	ملائمة	66.602	0.001	غير ملائمة
61	q64	30.078	0.817	ملائمة	49.504	0.066	ملائمة
62	q67	36.282	0.318	ملائمة	39.98	0.222	ملائمة
63	q69	24.931	0.896	ملائمة	31.696	0.673	ملائمة
64	q70	32.719	0.625	ملائمة	40.938	0.226	ملائمة
65	q71	53.38	0.024	غير ملائمة	54.726	0.014	غير ملائمة
66	q72	35.913	0.473	ملائمة	39.648	0.311	ملائمة
67	q73	37.674	0.484	ملائمة	81.348	0	غير ملائمة
68	q74	34.429	0.496	ملائمة	42.94	0.198	ملائمة
69	q75	45.632	0.13	ملائمة	48.132	0.055	ملائمة
70	q76	22.433	0.611	ملائمة	26.025	0.517	ملائمة
71	q77	40.901	0.344	ملائمة	71.585	0.001	غير ملائمة
72	q78	31.092	0.611	ملائمة	40.848	0.195	ملائمة
73	q79	49.023	0.089	ملائمة	56.211	0.017	غير ملائمة
74	q80	37.009	0.211	ملائمة	36.64	0.223	ملائمة
75	q81	39.271	0.097	ملائمة	39.943	0.067	ملائمة
76	q82	35.766	0.385	ملائمة	63.912	0.001	غير ملائمة
77	q83	41.111	0.13	ملائمة	40.586	0.238	ملائمة

م	المفردة	النموذج ثلاثي المعلم			النموذج رباعي المعلم		
		s-χ2	الدلالة	الملاءمة مع النموذج	s-χ2	الدلالة	الملاءمة مع النموذج
78	q84	22.022	0.142	ملائمة	26.824	0.061	ملائمة
79	q85	39.104	0.251	ملائمة	40.198	0.182	ملائمة
80	q86	52.021	0.032	غير ملائمة	59.432	0.002	غير ملائمة
81	q88	47.171	0.066	ملائمة	47.742	0.074	ملائمة
82	q89	28.324	0.699	ملائمة	29.251	0.606	ملائمة
83	q90	17.994	0.843	ملائمة	19.784	0.758	ملائمة
84	q91	45.673	0.13	ملائمة	67.295	0.001	غير ملائمة
85	q92	44.661	0.053	ملائمة	44.293	0.045	غير ملائمة
86	q93	23.154	0.873	ملائمة	23.212	0.841	ملائمة
87	q94	35.724	0.482	ملائمة	44.597	0.154	ملائمة
88	q95	43.111	0.112	ملائمة	42.689	0.098	ملائمة
89	q96	34.921	0.246	ملائمة	34.55	0.259	ملائمة
90	q98	47.565	0.094	ملائمة	50.356	0.035	غير ملائمة
91	q99	50.643	0.082	ملائمة	74.904	0	غير ملائمة
92	q100	23.316	0.327	ملائمة	25.112	0.292	ملائمة

النتائج الملخصة في الجدول (٤) تُظهر بوضوح أن النموذج ثلاثي المعلمة (3PL) يتفوق على النموذج رباعي المعلمة (4PL) من حيث عدد المفردات التي حققت ملاءمة للنموذج. فقد حقق النموذج ثلاثي المعلمة ملاءمة لـ ٨٧ مفردة من أصل ٩٢ مفردة (بنسبة ٩٤.٥٧٪)، بينما حقق النموذج رباعي المعلمة ملاءمة لـ ٦١ مفردة فقط (بنسبة ٦٦.٣٠٪).

وبالنظر إلى الجدول التفصيلي (٤)، يمكن ملاحظة أن: النموذج الثلاثي المعلمة يُظهر ملاءمة ممتازة لغالبية المفردات. فمعظم المفردات أظهرت قيم دلالة إحصائية (p-value) أكبر من ٠.٠٥، مما يعني أن استجابات الطلاب على هذه المفردات تتوافق بشكل جيد مع الافتراضات التي يضعها النموذج 3PL. المفردات القليلة التي لم تتلاءم مع هذا النموذج (مثل q16, q23, q44, q65, q80) تُشكل نسبة ضئيلة جدًا من إجمالي المفردات.

النموذج الرباعي المعلمة (4PL) يواجه صعوبة أكبر في تحقيق ملاءمة لعدد كبير من المفردات. يتضح هذا من العدد الكبير للمفردات التي أظهرت قيم دلالة إحصائية أقل من

٠.٠٥ (مثل q1, q4, q7, q9, q18, q21, q24, q31, q32, q33, q41, q46, q52, q53, q55, q56, q59, q60, q65, q67, q71, q73, q76, q80, q84, q85, q90, q91). هذه النتائج تشير إلى أن إضافة معلمة "الحد الأقصى للأداء أو معلمة الإهمال أو اللامبالاة" (d) في النموذج 4PL لم يحسن بالضرورة ملاءمة هذه المفردات، بل في كثير من الحالات أدت إلى تدهور الملاءمة مقارنة بالنموذج 3PL. هذا يمكن أن يحدث إذا كانت البيانات لا تظهر بالفعل "خطأ إهمال" (lapse error) أو "سقف أداء أقل من ١" بشكل كبير، أو إذا كان النموذج 4PL أكثر تعقيداً من اللازم لتفسير التباين في هذه المفردات، مما يجعله أقل استقراراً في التقدير لمفردات معينة.

كما يمكن ملاحظة أنه توجد مفردات معينة (مثل q16, q23, q44, q65, q80) لم تتلاءم مع أي من النموذجين؛ مما قد يشير إلى وجود مشكلات أساسية في هذه المفردات نفسها بالنسبة للممتحنين، والتي لا يمكن لأي من النموذجين التعامل معها بفعالية.

وبناءً على نتائج ملاءمة المفردات وفقاً لمحك statistics s−χ2، يُعد النموذج ثلاثي المعلمة (3PL) هو الأفضل في تحقيق ملاءمة لمفردات اختبار "سلوك رفض المدرسة". هذا يشير إلى أن البيانات الفعلية لاستجابات الطلاب على هذا الاختبار تُوصف بشكل أكثر كفاءة ودقة بواسطة النموذج 3PL، الذي يركز على معلمات الصعوبة والتمييز والتخمين العشوائي. كما أن استخدام النموذج 4PL، الذي يتضمن معلمة إضافية، لم يُظهر تحسناً في ملاءمة المفردات، بل على العكس، أدى إلى عدم ملاءمة عدد أكبر من المفردات. هذا قد يعني أن معلمة "الحد الأقصى للأداء أو معلمة الإهمال أو اللامبالاة" ليست ضرورية أو أنها لا تعكس بشكل دقيق خصائص استجابات الطلاب على هذا الاختبار المحدد.

وتُشير نتائج السؤال الأول بوضوح إلى أن النموذج ثلاثي المعلمة يُعد الأفضل في تحقيق ملاءمة لمفردات اختبار "سلوك رفض المدرسة"، حيث حقق ملاءمة لـ ٨٧ مفردة (٩٤.٥٧٪)، متفوقًا بذلك على النموذج رباعي المعلمة الذي حقق ملاءمة لـ ٦١ مفردة فقط (٦٦.٣٠٪). هذه النتيجة تتوافق بشكل مباشر مع ما توصلت إليه بعض الدراسات السابقة، مثل دراسة (محمد محمد جاد محمد, ٢٠١٦) التي وجدت أن النموذج اللوغاريتمي الثلاثي المعالم كان "أكثر النماذج ملائمة للبيانات وأكثرها دقة" في تدريج بنك أسئلة الرياضيات للصف الخامس الابتدائي. كما تدعمها جزئيًا دراسة (Büyükkıdık & İnal, 2023) التي وجدت أن النموذج 3PL حقق أفضل ملاءمة للبيانات في جميع الظروف التي اختبرتها.

إن تفوق النموذج ثلاثي المعلمة في ملاءمة المفردات في هذه الدراسة يوحي بأن البيانات لا تُظهر الحاجة الماسة لمعلمة "الحد الأقصى للأداء" ('upper asymptote 'd) التي يضيفها النموذج رباعي المعلمة. هذا يعني أن المفردات في اختبار "سلوك رفض المدرسة" مصممة بطريقة لا تتسبب في إجابات خاطئة متكررة من قِبل الطلاب ذوي القدرات العالية جدًا، أو أن أي "أخطاء إهمال" (lapse errors) أو عدم وصول الأداء إلى الكمال حتى عند القدرات العالية، إن وجدت، ليست كبيرة بما يكفي لتبرير التعقيد الإضافي للنموذج 4PL.

وعلى النقيض من ذلك، فإن العديد من الدراسات السابقة مثل (DoĞRuÖZ & Akin Arikan, 2020; Erdemir & Önen, 2019; Loken & Rulison, 2010; Yalçın, 2018) قد أشارت إلى أن النموذج الرباعي المعلمة أو النماذج المختلطة قد تُوفر ملاءمة أفضل في سياقات معينة. على سبيل المثال، تناولت دراسة (Pardede et al., 2023) أهمية معلمة الإهمال (d) في النموذج 4PL لاستيعاب حالات يخطئ فيها الممتحنون ذوو القدرة العالية. كما أشار (Loken & Rulison, 2010) إلى أن الحاجة إلى نمذجة المقارب الأعلى (أقل من ١) قد تكون أقل ضرورة في الاختبارات عالية المخاطر (high-stakes testing) حيث يُفترض بذل أقصى جهد، والباحثان يفسران نتيجة بحثهما من هذا القبيل فقد كانت الاختبارات تمثل أهمية للطلاب حيث أن نتيجة الاختبارات ستندرج ضمن أعمال السنة، بينما قد تكون مفيدة في سياقات أخرى مثل اختبارات الممارسة (practice tests) أو الاختبارات التي لا تكون فيها المخاطر عالية. تُشير نتائج دراستنا إلى أن اختبار "سلوك رفض المدرسة" قد يندرج تحت فئة الاختبارات التي لا تستفيد بشكل كبير من إضافة معلمة d أو أن طبيعة المفردات نفسها لا تتطلبها.

إن عدم ملاءمة عدد كبير من المفردات لنموذج 4PL في هذا البحث، رغم محاولته تقديم وصف أكثر شمولية، يُشير إلى أن تعقيد النموذج قد يكون غير ضروري للبيانات المعنية، وقد يؤدي إلى تقديرات أقل استقرارًا أو ملاءمة أسوأ إذا لم تكن البيانات تدعم المعلمة الإضافية بشكل كافٍ. هذا يؤكد على أهمية مبدأ "الاقتصاد في النماذج"، حيث يُفضل النموذج الأبسط القادر على تفسير البيانات بفعالية. وعلى ذلك فإن اختيار النموذج الأمثل يعتمد بشكل كبير على الخصائص الفعلية للبيانات وطبيعة الاختبار، وفي سياق اختبار "سلوك رفض المدرسة"، يبدو أن النموذج ثلاثي المعلمة يقدم التوازن الأفضل بين الكفاءة والملاءمة.

السؤال الثاني: ما أفضل نموذج (النموذج الثلاثي المعلم –النموذج الرباعي المعلم) والذي يحقق ملاءمة بشكل عام للاختبار التحصيلي؟

يهدف السؤال الثاني إلى تحديد النموذج الأفضل (ثلاثي المعلمة أم رباعي المعلمة) الذي يحقق ملاءمة شاملة للاختبار التحصيلي ككل، وذلك بالاعتماد على مجموعة من مؤشرات ملاءمة النموذج الكلية (M2, AIC, BIC, SRMSR, RMSEA, logLik).

ويُعد تقييم ملاءمة النموذج خطوة حاسمة في تحليل بيانات نظرية الاستجابة للمفردة (IRT)، حيث يساعد على تحديد مدى تمثيل النموذج الإحصائي للبيانات التجريبية بشكل فعال. ويعتمد هذا التقييم على مجموعة من المؤشرات، التي تُقسم عادةً إلى: مقاييس قائمة على اللوغاريتم الاحتمالي ومعايير معلوماتية، بالإضافة إلى مؤشرات الملاءمة المطلقة والمتبقية. الهدف الأساسي هو اختيار النموذج الأكثر اقتصادًا الذي يحافظ على قدرة تفسيرية جيدة للبيانات:

مقاييس قائمة على اللوغاريتم الاحتمالي ومعايير معلوماتية:

• المؤشر LogLik (اللوغاريتم الاحتمالي – Log-Likelihood) يمثل قيمة اللوغاريتم الاحتمالي للنموذج، حيث تشير القيم الأعلى إلى ملاءمة أفضل للبيانات. ومع ذلك، لا يوجد محك قاطع لـ LogLik بحد ذاته للقبول أو الرفض المطلق. تُستخدم قيم LogLik بشكل أساسي للمقارنة بين النماذج المتداخلة (Nested Models) من خلال اختبار نسبة الاحتمال (Likelihood Ratio Test) (Loken & Rulison, 2010).،

• المؤشر AIC (معيار أكايكي للمعلومات – Akaike Information Criterion): تُفضل النماذج ذات قيم AIC الأقل. يُستخدم هذا المعيار للمقارنة النسبية بين النماذج، سواء كانت متداخلة أو غير متداخلة. يعاقب AIC النموذج على تعقيده (عدد المعلمات المستخدمة)، مما يشجع على اختيار النموذج الأكثر بساطة الذي لا يزال يقدم ملاءمة جيدة للبيانات (Akaike, 1974)

• المؤشر BIC (معيار بايز للمعلومات – Bayesian Information Criterion): على غرار AIC، تُفضل النماذج ذات قيم BIC الأقل. يفرض BIC عقوبة أكبر على تعقيد النموذج مقارنة بـ AIC، خاصة مع أحجام العينات الكبيرة، مما يجعله أكثر تفضيلًا للنماذج البسيطة (Schwarz, 1978).

مؤشرات الملاءمة المطلقة والمتبقية

• SRMSR (الجذر التربيعي المعياري لمتوسط المربعات المتبقية – Standardized Root Mean Square Residual): يقيس هذا المؤشر متوسط الفرق المعياري بين الارتباطات المرصودة والمتوقعة للبيانات. تشير القيم الأقل إلى ملاءمة أفضل للنموذج. تُعتبر القيم أقل من ٠.٠٨ مؤشرًا جيدًا للملاءمة، بينما تشير بعض المراجع إلى أن القيم أقل من ٠.٠٥ تدل على ملاءمة ممتازة (Hu & Bentler, 1999).

• RMSEA (جذر متوسط مربع خطأ التقريب – Root Mean Square Error of Approximation) [٥٪, ٩٥٪ CI]: يقيم هذا المؤشر مدى ملاءمة النموذج في المجتمع، وليس فقط في العينة المدروسة، ويُقدم غالبًا مع فاصل ثقة (Confidence Interval). تُشير القيم أقل من ٠.٠٥ إلى ملاءمة جيدة جدًا للنموذج، بينما تُعتبر القيم بين ٠.٠٥ و٠.٠٨ مقبولة. تُشير القيم أكبر من ٠.١٠ إلى ملاءمة ضعيفة (Browne & Cudeck, 1993).

• M2 (df): هو إحصاء اختبار ملاءمة النموذج الكلي، والذي يتبع توزيع مربع كاي ($\chi 2$). تُشير القيم غير الدالة إحصائيًا (p-value > 0.05) إلى أن النموذج لا يختلف بشكل كبير عن البيانات المرصودة، مما يدل على ملاءمة جيدة. يجب دائمًا أخذ درجات الحرية (df) في الاعتبار عند تفسير قيمة M2 (Maydeu-Olivares & Joe, 2006).

ولا ينبغي الاعتماد على مؤشر واحد فقط عند تقييم ملاءمة نموذج .IRT بدلاً من ذلك، يجب النظر إلى مجموعة المؤشرات بشكل متكامل لاتخاذ قرار مستنير حول النموذج الأكثر ملاءمة للبيانات، مع الموازنة بين بساطة النموذج وقدرته التفسيرية.

وللإجابة عن السؤال الثاني؛ تم التحليل على اختبار الفصل الأول "سلوك رفض المدرسة"؛ تم اختبار مدى ملاءمة النموذج ككل (المعروف أيضًا باسم اختبار مدى ملاءمة النموذج) (goodness-of-fit test of the model) لفحص النموذج الذي يناسب البيانات في ضوء كل من النموذج الثلاثي المعلم والنموذج الرباعي المعلم. وتم إجراء اختبار ملاءمة النموذج باستخدام عدد من مؤشرات ملاءمة النموذج، مثل M2، ومعيار المعلومات (Akaike information criterion (AIC))، ومعيار المعلومات (Akaike information criterion (AIC))، والجذر المعياري لمتوسط مربعات البواقي (standardized root mean square residual (SRMSR))، وكذلك (log-likelihood).

جدول (٥) مؤشرات ملاءمة النموذج

النموذج المختبر	Model-Fit Indices مؤشرات ملاءمة النموذج					
	M2 (df)	RMSEA [5%, 95% CI]	SRMSR	AIC	BIC	logLik
النموذج ثلاثي المعلم	١٠١٨٠.١٨ (4002)**	٠.٠٣٦٣٧١٠٩ [0.03548092, 0.03723167]	٠.٠٥٤	٧٤٠١٠.٤٠	٧٥٤٠٧.٨٠	– ٣٦٧٢٩.٢٠
النموذج رباعي المعلم	٦٤٣٦.٣١٧ (3910)**	٠.٠٢٣٥٢٩٩ [0.02249054, 0.02454122]	٠.٠٦٥١	٧٦١٧٢.٠٩	٧٨٠٣٥.٢٩	– ٣٧٧١٨.٠٥

Note: RMSEA = Root mean square error of approximation, CI = Confidence interval, SRMSR = Standardized root mean square residual, AIC = Akaike information criterion, BIC = Bayesian information criterion, logLik = Log-likelihood, and df = the degrees of freedom

من جدول (٥) ووفقاً للمحكات المذكورة سلفاً؛ فإننا نجد أنه وفقاً للمؤشرين (M2, RMSEA) فإن النموذج رباعي المعلمة (4PLM) هو الذي يحقق القيم الأقل ومن ثم فإنه هو الملائم للبيانات، أما وفقاً للمؤشرات (SRMSR, AIC, BIC) فإن النموذج ثلاثي المعلمة (3PLM) هو الذي يحقق القيم الأقل ومن ثم فإنه هو الملائم للبيانات، أما وفقاً للمؤشر (logLik) فإن النموذج ثلاثي المعلمة (3PLM) هو الذي يحقق القيمة الأعلى ومن ثم فإنه هو الملائم للبيانات.

تشير غالبية مؤشرات الملاءمة (logLik, AIC, BIC, SRMSR) إلى أن النموذج ثلاثي المعلمة (3PL) يحقق ملاءمة أفضل بشكل عام للاختبار ككل. فقيم logLik الأعلى، وقيم AIC وBIC وSRMSR الأقل للنموذج 3PL مقارنة بالنموذج 4PL، وهي تُعد مؤشرات قوية على تفوق النموذج 4PL في تفسير البيانات بفاعلية أكبر مع تعقيد أقل. هذه المؤشرات تُفضل النماذج الأكثر بساطة (أقل معلمات) إذا كانت لا تُضحي بشكل كبير بقدرتها على تفسير البيانات.

من ناحية أخرى، يُظهر النموذج رباعي المعلمة (4PL) قيمًا أفضل لمؤشري M2 وRMSEA. قيمة RMSEA البالغة ٠.٠٢٣٥٢٩٩ للنموذج 4PL تُعتبر ممتازة (أقل من

٠.٠٥)، بينما قيمة RMSEA للنموذج 3PL (٠.٠٣٦٣٧١٠٩) تُعتبر جيدة جداً أيضاً. تفوق 4PL في M2 وRMSEA قد يشير إلى أن النموذج 4PL يقلل من مقدار الخطأ المتبقي في نموذج معين بشكل أكثر فعالية من 3PL، خاصة فيما يتعلق بالبواقي المعيارية.

وعلى الرغم من وجود تضارب، إلا أن الأغلبية الكبيرة من المؤشرات الرئيسية (logLik, AIC, BIC, SRMSR) تميل بشكل واضح لصالح النموذج ثلاثي المعلمة (3PL) كنموذج يحقق ملاءمة أفضل بشكل عام للاختبار التحصيلي. تفضيل مؤشري AIC وBIC للنموذج 3PL يشير إلى أن التعقيد الإضافي للنموذج 4PL (المعلمة الرابعة) لا يبرر التحسين في الملاءمة الذي قد يقدمه في مؤشرات مثل RMSEA وM٢. هذا يعني أن النموذج 3PL يقدم تفسيراً أكثر اقتصاداً (أقل معلمات) للبيانات مع الحفاظ على مستوى عالٍ من الملاءمة، مما يجعله الخيار المفضل لتطبيق نظرية الاستجابة للمفردة على هذا الاختبار تحديداً.

وتُشير نتائج السؤال الثاني إلى أن النموذج ثلاثي المعلمة (PL٣) يُعد الأفضل في تحقيق ملاءمة شاملة للاختبار ككل بناءً على غالبية مؤشرات الملاءمة الرئيسية (logLik, AIC, BIC, SRMSR)، على الرغم من أن النموذج رباعي المعلمة (4PL) قد أظهر تفوقًا في مؤشري M2 وRMSEA.

وتُقدم هذه النتائج دليلًا إضافيًا يدعم اختيار النموذج الأبسط (3PL) عندما يكون كافياً لتفسير بيانات الاختبار. ففي حين أن بعض الدراسات مثل (Erdemir & Önen, 2019; Yalçın, 2018) وجدت أن 4PLM أو نماذج مختلطة كانت الأفضل ملاءمة لبياناتها الكلية، إلا أن تفوق 3PL في مؤشرات مثل AIC و BIC يُعزز مبدأ الاقتصاد في النماذج. هذه المؤشرات تُكافئ النماذج التي تُفسر البيانات بشكل جيد بأقل عدد ممكن من المعلمات، مما يعني أن التعقيد الإضافي للنموذج 4PL (بسبب معلمة d) لم يُقدم تحسينًا كافيًا في ملاءمة النموذج الكلية لكي يُفضل على النموذج 3PL في هذا الاختبار تحديداً. هذا يتوافق مع دراسة (محمد محمد جاد محمد, ٢٠١٦) التي رجحت النموذج ثلاثي المعالم كونه الأكثر ملاءمة ودقة في تدريج بنك الأسئلة. كما أن دراسة (Büyükkıdık & İnal, 2023) وجدت أن 3PLM قدم أفضل ملاءمة لبيانات النموذج في جميع الظروف التي اختبرتها، مما يعزز هذا الاتجاه.

ومن جانب آخر، يُشير تفوق النموذج 4PL في مؤشري M2 وRMSEA إلى أن هذا النموذج قد يكون أكثر دقة في تقليل الأخطاء المتبقية والتنبؤ بالبواقي، وهو ما قد تُشير إليه دراسات مثل (Loken & Rulison, 2010) التي وجدت أن 4PL يحسن الملاءمة الإجمالية

للبيانات التي تم إنشاؤها باستخدام 4PL. ومع ذلك، فإن تفضيل AIC وBIC للنموذج 3PL يعني أن هذا التحسين الطفيف في الملاءمة لا يبرر التكلفة المرتبطة بزيادة عدد المعلمات وتقديرها، خاصة إذا كانت هذه المعلمة (الحد الأقصى للأداء– معلمة الإهمال) لا تُسهم بشكل جوهري في وصف بيانات الاختبار. هذا يعني أن الاختبار قيد الدراسة قد لا يُظهر مشكلات "سقف الأداء" أو "أخطاء الإهمال" عند الطلاب ذوي القدرة العالية بشكل يستدعي نمذجتها بواسطة معلمة d.

وبشكل عام، تُؤكد هذه النتائج على أن اختيار النموذج الأمثل في نظرية الاستجابة للمفردة ليس قرارًا أحادي الجانب يعتمد على مؤشر واحد. بل يتطلب تقييمًا شاملًا لعدة مؤشرات، مع الأخذ في الاعتبار مبدأ الاقتصاد والتعقيد النظري للنموذج. في سياق اختبار "سلوك رفض المدرسة"، تُشير النتائج إلى أن النموذج 3PL يُقدم تفسيراً أكثر كفاءة وملاءمة للبيانات، مما يجعله الخيار العملي الأمثل.

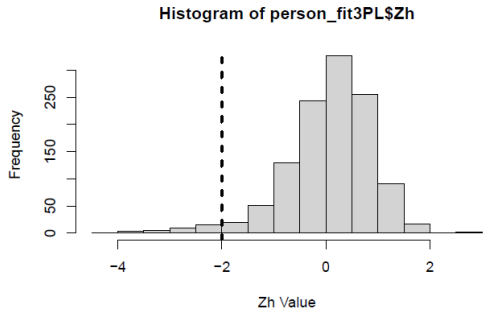
السؤال الثالث: ما مدى ملاءمة نمط استجابات الأفراد لمفردات الاختبار في ضوء كل من النموذج الثلاثي المعلم والنموذج الرباعي المعلم وفقاً للمحك الإحصائي (zh statistics)؟

للإجابة عن هذا السؤال؛ تم التحليل على اختبار الفصل الأول "سلوك رفض المدرسة" وتم استخدام نتائج اختبارات ملاءمة المفردات وملاءمة النموذج في اختبار ملاءمة الأفراد لفحص مدى ملاءمة نمط الاستجابات التي يقدمها كل فرد لمفردات الاختبار وفقاً للنموذج الثلاثي المعلم والنموذج الرباعي المعلم باستخدام الإحصاءة (zh statistics) (Drasgow et al., 1985)

تم بعد ذلك استخدام تقديرات معلمات المفردة التي تم الحصول عليها للحصول على تقدير معلمات الطالب أو الفرد. قبل تقديم المزيد من التقارير، نتائج تقدير معلمة الفرد، قمنا بالتحقق من مدى جودة نمط الاستجابة الذي قدمه الفرد للمفردات الـ ٩٢ الواردة في الاختبار مع نمط الاستجابة المتوقع بناءً على التقديرات وفقًا للنموذجين الثلاثي والرباعي المعلمة من خلال تحليل تناسب الفرد. كان الغرض من التحليل هو معرفة مدى دقة تقدير قدرة الفرد وفقا للنموذجين المستخدمين، فضلا عن كونه كاشفا مبكرا لوجود نمط استجابة غير عادي أو غير مناسب يظهره الفرد وخصائصه. لقد أجرينا تحليلًا مناسبًا للفرد باستخدام إحصائيات zh، وهو مؤشر مناسب موحد يقدمه (Drasgow et al., 1985) والتي تشكل قيمها عادةً توزيعًا غير طبيعي (Felt et al., 2017). باستخدام هذه الإحصاءة، يتم اكتشاف فرد غير ملائم عندما

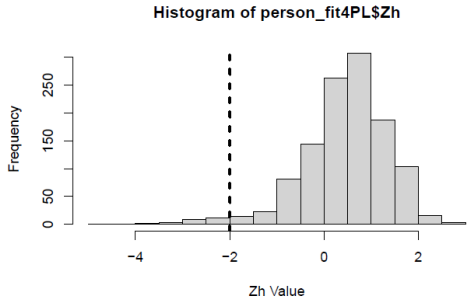
يكون zh أقل من (-٢) (Desjardins & Bulut, 2018; Felt et al., 2017; Merino-Soto et al., 2023).

يوضح التحليل وفقاً للنموذج الثلاثي المعلم أنه من بين (١١٦٨) فرداً، هناك إجمالي (٣٤) فرداً بنسبة (٢.٩١١ %) فرداً تكون قيمة zh أقل من (-٢)؛ مما يشير إلى أن أنماط الاستجابة على مفردات الاختبار الـ (٩٢) الحالية التي قدموها غير متوافقة مع أنماط الاستجابة التي قدمها الطلاب الآخرون البالغ عددهم (١١٣٤) (٩٧.٠٨٩ %) كأنماط الاستجابة المتوقعة وفقًا للنموذج الثلاثي المعلم كما بالشكل (٤).



شكل (٤) توزيع قيم zh للنموذج ثلاثي المعلمة

يوضح التحليل وفقاً للنموذج الرباعي المعلم أنه من بين (١١٦٨) فرداً، هناك إجمالي (٢٧) فرداً بنسبة (٢.٣١٢٪) فرداً تكون قيمة zh أقل من (-٢)؛ مما يشير إلى أن أنماط الاستجابة على عناصر الاختبار الـ (٩٢) الحالية التي قدموها غير متوافقة مع أنماط الاستجابة التي قدمها الطلاب الآخرون البالغ عددهم (١١٤١) (٩٧.٦٨٨ %) كأنماط الاستجابة المتوقعة وفقًا للنموذج الرباعي المعلم كما بالشكل (٥).



شكل (٥) توزيع قيم zh للنموذج رباعي المعلمة

كما ذكر (Felt et al., 2017)؛ تشمل الأسباب المحتملة للعدد الصغير من أنماط الاستجابة غير الملائمة هنا سلوك التخمين، وسلوك الغش، ونقص الحافز، وتشتيت الطلاب

بشيء ما أثناء الاستجابة لمفردات الاختبار.

بعد أن تم إجراء اختبارات ملاءمة المفردات، وملاءمة النموذج، وملاءمة الأفراد، تم تقدير معلمات المفردة باستخدام حزمة "mirt" (Chalmers, 2012)، باستخدام هذه الحزمة، تم إجراء تقدير معلمات المفردات باستخدام خوارزمية (the fixed quadrature expectation-maximization (EM)) (Chalmers, 2012) كطريقة افتراضية.

تُشير هذه النتائج إلى أن كلا النموذجين (3PL و4PL) يُظهران مستوى مرتفعًا جدًا من ملاءمة نمط استجابات الأفراد، حيث إن الغالبية العظمى من الأفراد (أكثر من ٩٧٪ لكل نموذج) قدموا أنماط استجابة متوافقة مع التوقعات. هذا يُعد مؤشرًا إيجابيًا للغاية على جودة الاختبار وانسجام استجابات الطلاب مع افتراضات نماذج نظرية الاستجابة للمفردة.

وعلى الرغم من أن كلا النموذجين أظهرا ملاءمة عالية، إلا أن النموذج رباعي المعلمة (4PL) أظهر عددًا أقل من الأفراد ذوي أنماط الاستجابة غير الملائمة (٢٧ فردًا مقابل ٣٤ فردًا للنموذج 3PL). هذا يعني أن النموذج 4PL كان أفضل قليلاً في استيعاب أو تفسير تلك الحالات الشاذة من الاستجابات الفردية التي قد لا تتوافق مع الأنماط المتوقعة. يُمكن تفسير هذا التحسن الطفيف في ملاءمة الأفراد تحت النموذج 4PL بأنه، حتى وإن لم تُحسن معلمة "الحد الأقصى للأداء" (d) ملاءمة المفردات بشكل عام أو ملاءمة النموذج ككل (كما أظهرت نتائج السؤالين الأول والثاني)، فإنها قد تُساعد في نمذجة بعض الحالات الفردية التي تُظهر "أخطاء إهمال" أو "تخمين" أو سلوكيات غير مثالية بشكل أفضل، مما يؤدي إلى تقليل عدد أنماط الاستجابة غير الملائمة.

والأسباب المحتملة لأنماط الاستجابة غير الملائمة، كما ذكرها (Felt et al., 2017)، تشمل سلوك التخمين، والغش، ونقص الحافز، وتشتيت الانتباه. كون هذه النسبة صغيرة جدًا في كلا النموذجين (أقل من ٣٪) يعكس أن هذه المشكلات السلوكية (أو المشكلات في الاختبار التي تسبب استجابات غير متوقعة) ليست سائدة في عينة البحث، أو أن النماذج قادرة على التعامل مع جزء كبير منها.

وأخيراً تُؤكد النتائج أن كلا النموذجين مناسبان جداً لبيانات الأفراد. ومع ذلك، فإن النموذج 4PL يُقدم تحسناً طفيفاً ولكنه ملحوظ في تقليل عدد الأفراد الذين تُعتبر أنماط استجاباتهم غير ملائمة، مما قد يُشير إلى قدرته على التقاط بعض الفروق الدقيقة في سلوك

الاستجابة الفردي التي لا يستطيع النموذج 3PL التعامل معها بالكامل.

كما تُعد ملاءمة الأفراد مؤشرًا حاسماً على جودة القياس، حيث تُظهر ما إذا كانت استجابات الفرد متسقة داخليًا وتتوافق مع النموذج النظري لقدرته. إن النتائج التي تُظهر ملاءمة عالية لمعظم الأفراد في كلا النموذجين هي مؤشر إيجابي على أن اختبار "سلوك رفض المدرسة" يعمل بشكل جيد على مستوى الفرد، وأن استجابات الطلاب ليست عشوائية أو غير متوقعة بشكل كبير.

ويُمكن تفسير تفوق النموذج 4PL في هذا الجانب (على الرغم من أنه كان أقل ملاءمة في المفردات والنموذج الكلي كما تمت الإشارة في الأسئلة السابقة) من زوايا متعددة على سبيل المثال، ناقشت دراسة (Pardede et al., 2023) أهمية معلمة الإهمال (d) في النموذج 4PL لاستيعاب حالات يخطئ فيها الممتحنون ذوو القدرة العالية. وفي حين أن هذه المعلمة قد لا تكون ضرورية لكل مفردة على حدة أو لكل الاختبار ككل، فإنها قد تُساعد في نمذجة سلوكيات فردية معينة قد تنحرف عن الأنماط المتوقعة بشكل طفيف. قد يكون هذا الانحراف ناتجاً عن عوامل غير معرفية (مثل الإرهاق، أو قلة التركيز في بعض المفردات، أو حتى سلوكيات غش بسيطة غير مكتشفة)، والتي يُمكن للنموذج 4PL أن يتعامل معها بمرونة أكبر.

ويُؤكد (Felt et al., 2017) على أن الأنماط غير الملائمة قد تُعزى إلى سلوك التخمين، أو الغش، أو نقص الحافز، أو تشتيت الطلاب. تفوق 4PL هنا يُشير إلى أن النموذج بمرونته الإضافية ربما يكون قد استوعب بعض هذه الانحرافات البسيطة بشكل أفضل من 3PL، مما أدى إلى تصنيف عدد أقل من الأفراد على أن أنماط استجاباتهم غير ملائمة. هذا يُبرز قيمة 4PL في تحليل الاستجابات الفردية الدقيقة، حتى لو لم يكن هو النموذج الأفضل للملاءمة الكلية للمفردات أو للاختبار ككل.

وبشكل عام، تُسلط هذه النتائج الضوء على طبيعة الاختيار المعقدة للنموذج في IRT ففي حين أن النموذج 3PL قد يكون أكثر اقتصادًا وملائمةً على مستوى المفردة والاختبار الكلي (كما أظهرت الأسئلة السابقة)، فإن النموذج 4PL قد يُقدم تحسينًا في فهم ومعالجة الأنماط الشاذة لاستجابات الأفراد، مما يُعزز من دقة تقدير القدرة الفردية، خاصة في الحالات التي يُشتبه فيها بوجود عوامل تؤثر على الأداء بخلاف القدرة الكامنة.

السؤال الرابع: "هل تختلف منحنيات خصائص المفردة (Item Characteristic Curves - ICCs) ودوال معلومات المفردة (Item Information Functions - IIFs) وكذلك دقة القياس لكل مفردة من مفردات الاختبار في ضوء كل من النموذج الثلاثي المعلم والنموذج الرباعي المعلم؟"

للإجابة عن هذا السؤال؛ تم التحليل على اختبار الفصل الأول "سلوك رفض المدرسة": تم تقدير معلمات المفردات في ضوء كل من النموذج الثلاثي المعلم (التمييز– الصعوبة– التخمين) والنموذج الرباعي المعلم (التمييز– الصعوبة– التخمين– الإهمال)، وكانت كما بالجدول (٦).

جدول (٦) ملخص إحصاءات تقديرات معلمات المفردات وفقاً للنموذجين الثلاثي المعلم (التمييز– الصعوبة– التخمين) والرباعي المعلم (التمييز– الصعوبة– التخمين– الإهمال)

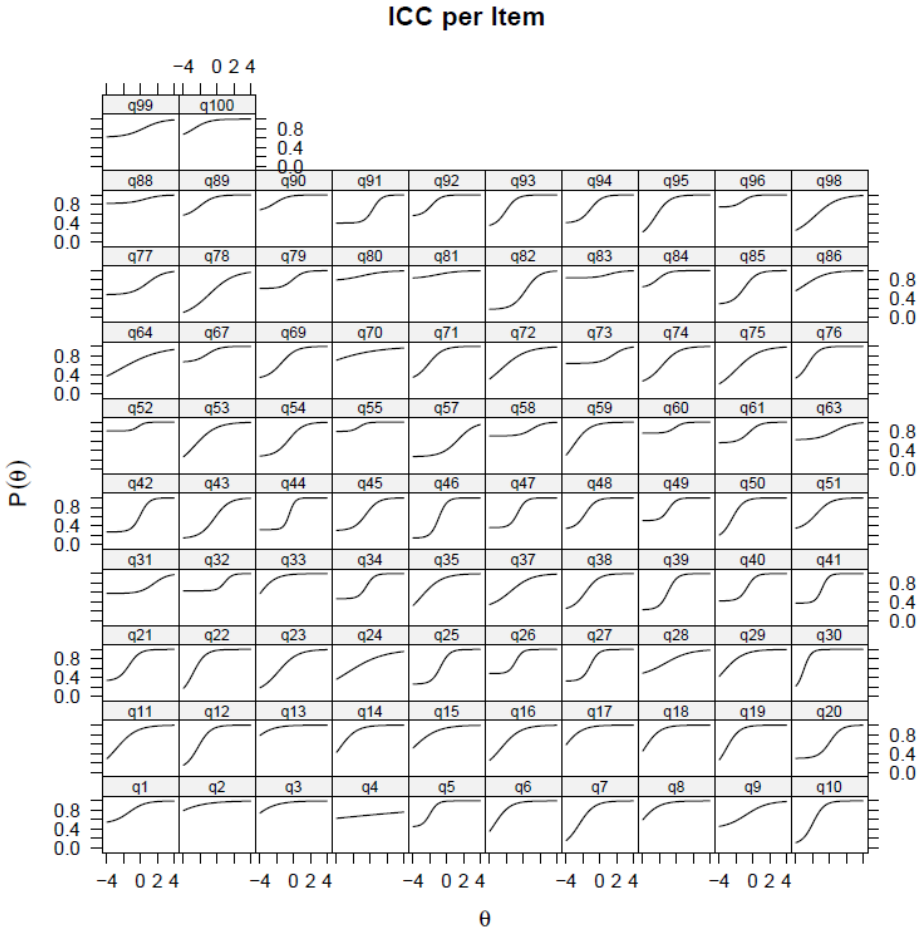
م	المفردة	النموذج ثلاثي المعلم			النموذج رباعي المعلم			
		a	b	g	a	b	g	u
1	q1	1.038	-1.353	0.520	0.464	-2.847	0.147	0.998
2	q2	0.500	-6.612	0.031	0.389	-7.154	0.137	1.000
3	q3	0.692	-5.488	0.015	0.542	-5.799	0.137	1.000
4	q4	0.084	-8.072	0.106	1.366	-2.010	0.271	0.748
5	q5	2.182	-1.986	0.447	1.417	-2.401	0.143	1.000
6	q6	1.079	-3.408	0.012	0.832	-3.348	0.133	1.000
7	q7	1.100	-2.455	0.002	0.612	-2.806	0.123	0.999
8	q8	0.890	-4.409	0.015	0.715	-4.399	0.138	1.000
9	q9	0.774	-0.787	0.412	0.625	-1.900	0.157	0.925
10	q10	1.506	-1.937	0.072	1.126	-1.701	0.144	0.999
11	q11	0.820	-2.874	0.003	0.488	-3.020	0.122	0.997
12	q12	1.321	-2.262	0.063	0.920	-2.216	0.136	1.000
13	q13	0.835	-5.127	0.217	0.670	-5.716	0.143	1.000
14	q14	0.978	-3.649	0.008	0.745	-3.665	0.134	1.000
15	q15	0.675	-4.068	0.008	0.306	-6.214	0.129	1.000
16	q16	0.857	-2.624	0.022	0.519	-2.556	0.135	0.997
17	q17	0.940	-4.342	0.010	0.722	-4.505	0.138	1.000
18	q18	1.075	-3.796	0.004	0.820	-3.874	0.134	1.000
19	q19	1.266	-3.162	0.003	1.014	-2.980	0.129	1.000
20	q20	1.449	0.073	0.292	1.323	0.283	0.260	0.861
21	q21	1.544	-1.356	0.328	0.802	-1.644	0.173	0.997
22	q22	1.367	-2.833	0.002	0.989	-2.826	0.125	1.000
23	q23	0.873	-1.792	0.065	0.875	-1.391	0.168	0.929

النموذج رباعي المعلم				النموذج ثلاثي المعلم			المفردة	م
u	g	b	a	g	b	a		
0.890	0.147	-1.986	0.722	0.011	-2.704	0.453	q24	24
0.944	0.249	-0.462	1.565	0.261	-0.728	1.835	q25	25
0.998	0.205	-1.203	1.220	0.486	-0.904	2.411	q26	26
0.997	0.168	-1.175	1.164	0.326	-1.116	2.016	q27	27
0.948	0.146	-2.743	0.439	0.414	-1.296	0.676	q28	28
1.000	0.131	-3.784	0.573	0.015	-3.599	0.821	q29	29
1.000	0.155	-2.956	1.697	0.133	-3.056	2.321	q30	30
0.752	0.320	-1.574	1.244	0.577	1.593	1.206	q31	31
0.896	0.154	-1.462	0.574	0.635	0.927	2.331	q32	32
1.000	0.137	-4.308	0.784	0.009	-4.293	0.977	q33	33
0.956	0.225	-0.748	1.154	0.466	-0.439	2.115	q34	34
0.999	0.127	-3.304	0.470	0.006	-3.054	0.807	q35	35
0.938	0.154	-1.870	0.602	0.256	-1.468	0.795	q37	36
0.997	0.148	-2.082	0.597	0.220	-1.675	1.244	q38	37
0.974	0.205	-0.727	1.378	0.232	-0.976	1.848	q39	38
0.976	0.228	-0.757	1.130	0.420	-0.664	1.986	q40	39
0.995	0.230	-0.804	1.449	0.373	-0.875	2.504	q41	40
0.872	0.253	0.254	1.566	0.276	-0.011	1.923	q42	41
0.847	0.228	0.172	1.428	0.136	-0.310	1.128	q43	42
0.951	0.268	-0.272	1.951	0.325	-0.497	2.945	q44	43
0.913	0.229	-0.468	1.267	0.304	-0.482	1.357	q45	44
0.962	0.250	-0.467	1.771	0.145	-0.948	2.048	q46	45
0.965	0.209	-0.673	1.304	0.366	-0.644	2.118	q47	46
1.000	0.147	-1.852	1.104	0.333	-1.708	1.575	q48	47
0.999	0.174	-1.648	0.863	0.515	-0.929	2.094	q49	48
1.000	0.133	-2.432	1.085	0.128	-2.458	1.473	q50	49
0.992	0.162	-2.057	0.585	0.315	-1.554	1.122	q51	50
1.000	0.135	-4.289	0.414	0.813	-0.261	2.500	q52	51
0.984	0.127	-2.743	0.470	0.003	-2.652	0.774	q53	52
0.904	0.202	-0.572	1.130	0.266	-0.535	1.170	q54	53
1.000	0.144	-3.332	0.945	0.801	-1.276	2.199	q55	54
0.180	0.180	-1.961	-2.310	0.264	1.393	0.991	q57	55
0.888	0.177	-2.745	0.569	0.706	1.047	1.424	q58	56
1.000	0.129	-3.038	0.774	0.004	-3.140	1.008	q59	57
0.999	0.139	-2.843	0.474	0.765	-0.157	1.959	q60	58
0.985	0.175	-1.747	0.566	0.561	-0.463	1.495	q61	59
0.923	0.162	-2.056	0.563	0.623	0.652	0.971	q63	60
0.869	0.154	-1.865	0.652	0.037	-2.406	0.409	q64	61
1.000	0.139	-3.479	0.523	0.666	-1.161	1.345	q67	62

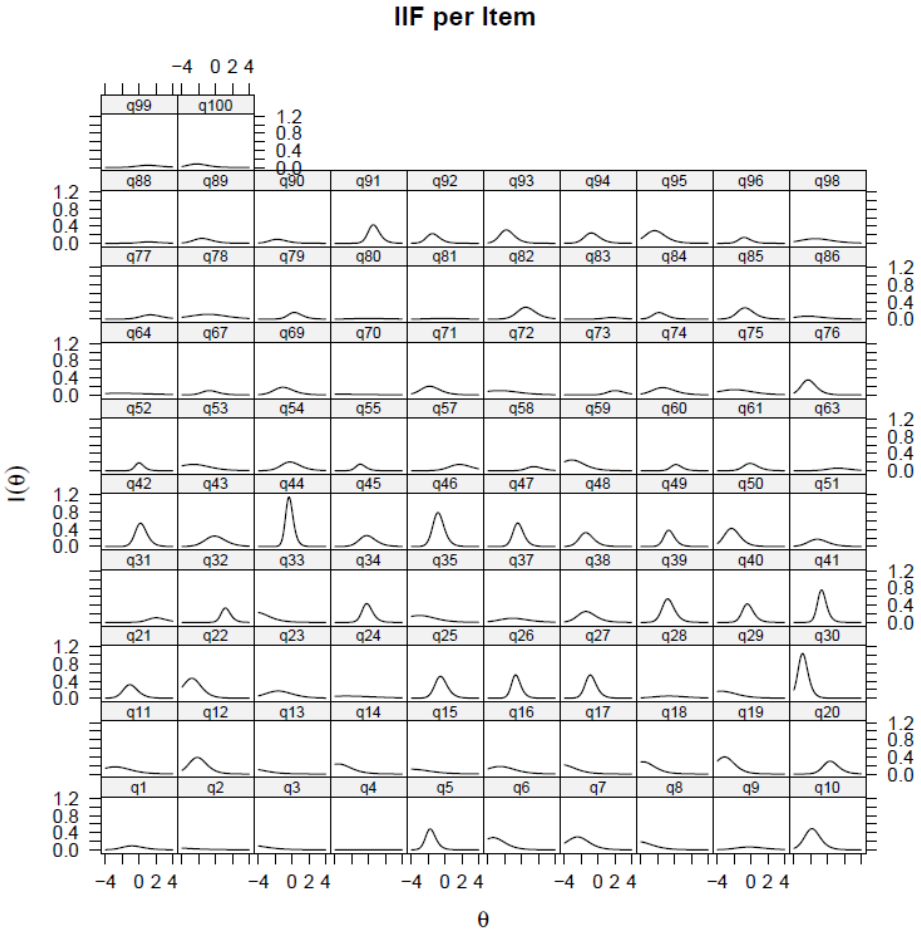
النموذج رباعي المعلم				النموذج ثلاثي المعلم			المفردة	م
u	g	b	a	g	b	a		
0.984	0.173	-1.831	0.602	0.305	-1.415	1.134	q69	63
0.998	0.141	-4.292	0.331	0.104	-6.274	0.314	q70	64
0.999	0.143	-2.825	0.635	0.269	-2.133	1.168	q71	65
0.952	0.149	-2.574	0.474	0.012	-2.679	0.648	q72	66
0.778	0.258	-2.275	0.899	0.638	1.605	1.292	q73	67
0.974	0.173	-1.716	0.630	0.192	-1.674	0.997	q74	68
0.905	0.156	-1.518	0.847	0.011	-1.987	0.708	q75	69
1.000	0.141	-2.804	1.052	0.254	-2.534	1.512	q76	70
0.948	0.181	-0.804	0.422	0.488	0.889	1.039	q77	71
0.822	0.180	-0.248	1.063	0.003	-0.895	0.675	q78	72
0.982	0.173	-1.871	0.514	0.620	-0.079	1.552	q79	73
1.000	0.135	-9.942	0.186	0.780	-0.941	0.721	q80	74
1.000	0.136	-10.47	0.224	0.828	-1.125	0.807	q81	75
0.788	0.256	0.468	2.326	0.174	0.310	1.242	q82	76
0.998	0.144	-5.012	0.272	0.846	1.116	1.247	q83	77
1.000	0.146	-3.309	1.025	0.637	-2.203	1.576	q84	78
0.948	0.230	-0.805	1.206	0.283	-0.986	1.350	q85	79
1.000	0.132	-5.475	0.348	0.390	-2.907	0.770	q86	80
0.998	0.144	-4.642	0.297	0.817	0.513	1.064	q88	81
1.000	0.136	-3.387	0.682	0.523	-2.054	1.161	q89	82
1.000	0.139	-3.953	0.749	0.649	-2.224	1.262	q90	83
0.908	0.223	0.107	0.924	0.401	0.342	1.950	q91	84
1.000	0.141	-2.802	0.963	0.549	-1.815	1.689	q92	85
1.000	0.139	-2.453	0.987	0.320	-2.061	1.523	q93	86
0.993	0.181	-1.758	0.619	0.401	-1.091	1.461	q94	87
1.000	0.134	-2.681	0.759	0.078	-2.491	1.173	q95	88
1.000	0.139	-3.164	0.793	0.741	-1.155	1.802	q96	89
0.889	0.174	-1.352	1.016	0.110	-1.690	0.732	q98	90
0.915	0.160	-2.302	0.579	0.616	0.426	0.883	q99	91
1.000	0.139	-4.112	0.749	0.596	-2.745	1.091	q100	92

Note. a = Item discrimination parameter, b = Item difficulty parameter, g = Pseudo-guessing, guessing, or pseudo-chance-level parameter, and u = Carelessness, inattention, or slipping parameter

كما تم رسم منحنى خصائص المفردة (ICC) ودالة معلومات المفردة (IIF) لكل مفردة من مفردات الاختبار قيد التحقيق في هذه الدراسة في ضوء النموذج الثلاثي المعلم وكانت كما بالشكلين (٦) & (٧).

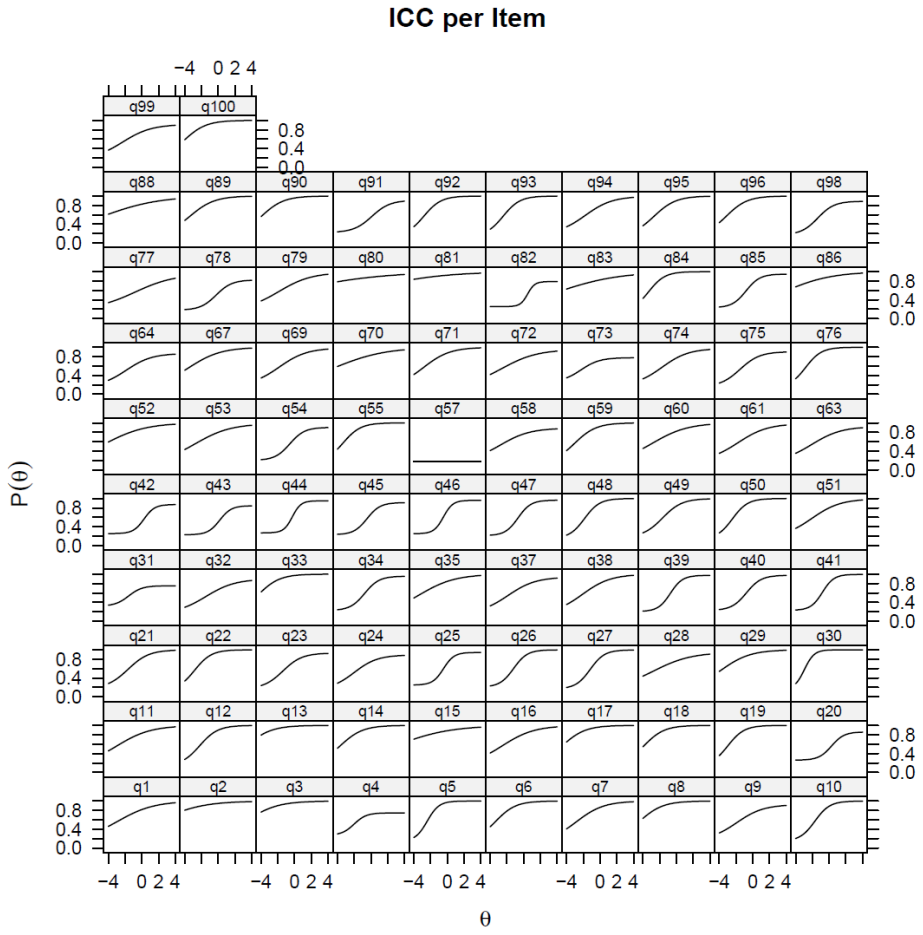


شكل (٦) المنحنيات المميزة للمفردات ال (٩٢) وفقاً للنموذج ثلاثي المعلمة

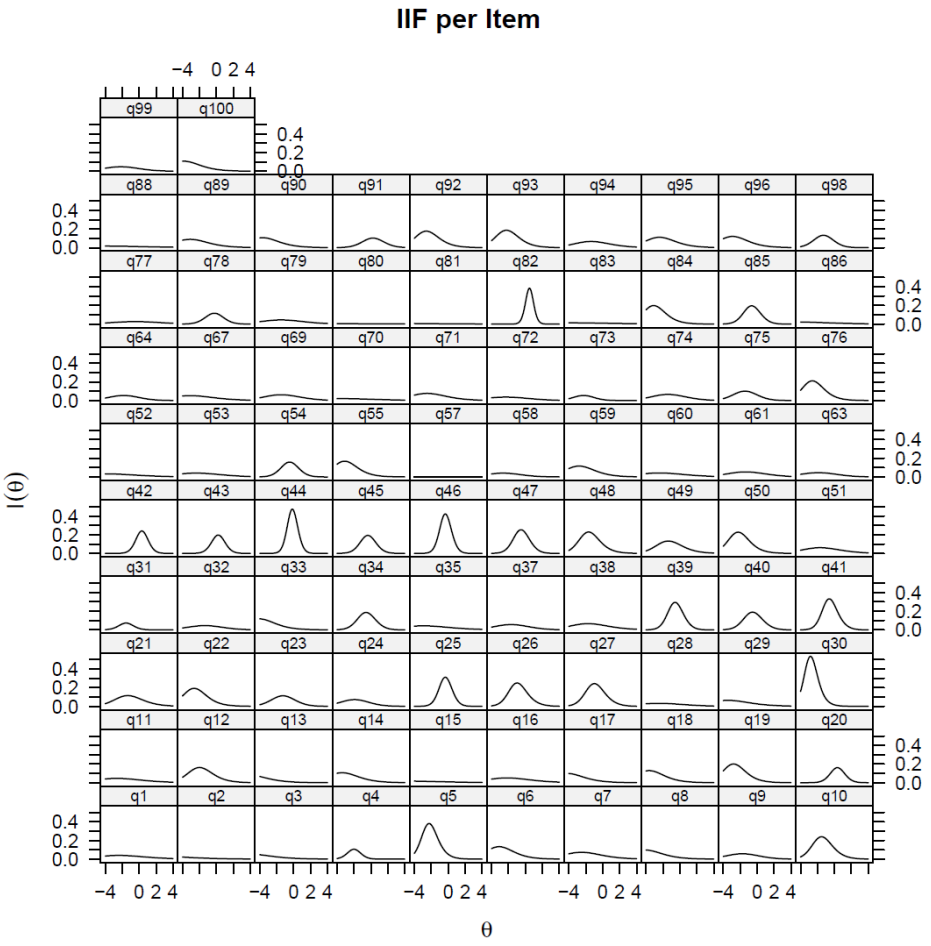


شكل (٧) دوال المعلومات للمفردات ال (٩٢) وفقاً للنموذج ثلاثي المعلمة

وتم رسم منحنى خصائص المفردة (ICC) ودالة معلومات المفردة (IIF) لكل مفردة من مفردات الاختبار قيد التحقيق في هذه الدراسة في ضوء النموذج الرباعي المعلم وكانت كما بالشكلين (٨) & (٩).



شكل (٨) المنحنيات المميزة للمفردات ال (٩٢) وفقاً للنموذج رباعي المعلمة



شكل (٩) دوال المعلومات للمفردات ال (٩٢) وفقاً للنموذج رباعي المعلمة

تحليل تقديرات معلمات المفردات من الجدول:(6)

- معلمة التمييز (a): بشكل عام، تبدو قيم a في النموذج 3PL أعلى في كثير من المفردات مقارنة بالنموذج 4PL (على سبيل المثال، q1، q5، q26، q40، q43). القيم الأعلى لـ a في 3PL تعني أن هذه المفردات تميز بشكل أفضل بين الطلاب ذوي القدرات المختلفة عند مستوى الصعوبة. في المقابل، فإن قيم a في 4PL قد تكون أقل في بعض المفردات مما يشير إلى أن النموذج يحاول تفسير بعض التباين من خلال المعلمة u وهي معلمة الإهمال بدلاً من التمييز.
- معلمة الصعوبة (b): تظهر قيم b اختلافات كبيرة بين النموذجين لبعض المفردات. على سبيل المثال، q1 وq٤ وq31 تظهر قيم b أكثر سلبية (أسهل) بشكل ملحوظ في 3PL مقارنة بـ 4PL، أو العكس لبعض المفردات. هذا يعني أن إدراج معلمة u في 4PL قد يؤثر

على تقدير صعوبة المفردة، حيث يمكن أن تعوض معلمة u عن بعض ما كان يُعزى في 3PL إلى صعوبة المفردة أو تمييزها.

- معلمة التخمين (g): تظهر قيم g (التخمين) في النموذج 3PL غالبًا ما تكون أعلى من قيم g في النموذج 4PL. وهذا أمر متوقع، حيث إن النموذج 4PL يُقدم معلمة إضافية (u) للتعامل مع "الحد الأقصى للأداء" أو "الإهمال"، مما قد يُقلل من الحاجة إلى تفسير بعض الأداء المنخفض على أنه مجرد تخمين عشوائي.
- معلمة الإهمال (u) في 4PL: هذه المعلمة تُشير إلى الاحتمال الأقصى للإجابة الصحيحة حتى بالنسبة للطلاب ذوي القدرة اللانهائية. نلاحظ أن قيم u في النموذج 4PL تتراوح من قيم منخفضة (مثل ٠.١٨٠ لـ q55) إلى قيم قريبة جدًا من ١.٠ (مثل ١.٠٠٠ لعدد كبير من المفردات). القيم المنخفضة لـ u (على سبيل المثال، q4: 0.748، q31: 0.752، q55: 0.180) تُشير إلى أن هذه المفردات قد تُظهر سلوك "سقف الأداء" (أي حتى الأذكياء قد يخطئون)، أو أن هناك عوامل تؤثر على إجابات الطلاب ذوي القدرة العالية. أما القيم القريبة من ١.٠٠٠ فتشير إلى أن معلمة u لم تُقدم فرقًا جوهريًا عن النموذج 3PL لتلك المفردات، وأن الاحتمال الأقصى للإجابة الصحيحة هو ١ (أي لا يوجد إهمال يُذكر).

مقارنة منحنيات خصائص المفردة (ICCs):

- ICCs في 3PL (6): تُظهر هذه المنحنيات بشكل عام (S-شكل)، حيث تبدأ من مستوى التخمين (g) وترتفع لتصل إلى ١.٠٠٠ عند القدرات العالية. الفروقات في قيم a وb بين المفردات ستؤثر على انحدار المنحنى وموقعه على مقياس القدرة.
- ICCs في 4PL (7): تُظهر هذه المنحنيات أيضًا بشكل عام (S-شكل)، ولكنها تختلف عن 3PL في نقطتين رئيسيتين:

١. المقارب العلوي :لن تصل المنحنيات بالضرورة إلى ١.٠٠٠ عند القدرات العالية، بل ستقترب من قيمة u(معلمة الإهمال) إذا كانت قيمتها أقل من ١.٠٠٠. هذا يعني أن هناك "سقفًا" للأداء حتى للطلاب ذوي القدرات العالية جدًا.

٢. تأثير المعلمة u وg: قد تُغير قيم u وg من شكل المنحنى بشكل طفيف، خاصة في الأطراف، لتعكس هذه الظواهر بشكل أفضل.

مقارنة دوال معلومات المفردة (IIFs):

IIFs في 3PL شكل (٨): تُظهر هذه الدوال مقدار المعلومات التي تُقدمها كل مفردة عند مستويات قدرة مختلفة. تُظهر المفردات ذات التمييز العالي (a) ذروة معلوماتية أعلى وأكثر حدة. ذروة المعلومات تقع عادةً بالقرب من قيمة صعوبة المفردة (b).

IIFs في 4PL شكل (٩): تُظهر هذه الدوال أيضًا ذروات معلوماتية، ولكن شكلها وموقعها قد يتأثران بقيمتي g و u. إذا كانت المفردة ذات قيمة u منخفضة، فقد تُقدم معلومات أقل عند مستويات القدرة العالية جدًا مقارنة بنظيرتها في 3PL. بمعنى آخر، ستقلل معلمة الإهمال (u) من المعلومات المقدمة من المفردة عند مستويات القدرة العالية لأنها تُشير إلى أن الأداء هناك لن يكون كاملاً حتى لأعلى القدرات.

دقة القياس لكل مفردة:

تُشير دقة القياس (أو خطأ القياس المعياري – Standard Error of Measurement) إلى مدى موثوقية الدرجة المقدرة للفرد من المفردة. دالة معلومات المفردة هي مقلوب مربع الخطأ المعياري، كما أن النموذج الذي يقدم دالة معلومات مفردة أعلى (أكثر حدة) عند مدى معين من القدرة، يُعتبر أكثر دقة في قياس تلك القدرة، وبما أن العديد من المفردات في النموذج 4PL تُظهر قيمًا لـ u أقل من ١، فهذا يعني أن هذه المفردات قد تُقدم معلومات أقل دقة (أو أقل) عند المستويات العليا من القدرة مقارنة بالنموذج 3PL، الذي يفترض أن الطلاب ذوي القدرة العالية جدًا يُجيبون بشكل صحيح تمامًا. على العكس، قد تُقدم المفردات في 4PL معلومات أفضل عند المستويات المنخفضة إذا كانت معلمة g أقل بشكل كبير من 3PL، مما يُقلل من تأثير التخمين.

ومما سبق تختلف منحنيات ICCs و IIFs ودقة القياس بين النموذجين نتيجة لاختلاف تقدير المعلمات، لا سيما إدخال معلمة u في 4PL. بينما قد يُقدم النموذج 3PL دقة قياس أعلى عند مستويات القدرة العليا في المفردات التي يُفترض فيها الوصول إلى أداء مثالي، فإن النموذج 4PL يُقدم وصفًا أكثر واقعية للمفردات التي قد لا يصل فيها الأداء إلى ١٠٠٪ حتى لأكثر الطلاب قدرة، مما يؤثر على شكل IIFs ودقة القياس عند تلك المستويات.

السؤال الخامس: "هل يختلف تقدير قدرة (معلمة) الأفراد باستخدام طرق التقدير (maximum likelihood (ML), maximum a posteriori (MAP), and expected a posteriori (EAP)) وكذلك دقة القياس في تقدير قدرة الأفراد، في ضوء كل من النموذج الثلاثي المعلم والنموذج الرباعي المعلم للاختبار؟"

للإجابة عن هذا السؤال؛ تم التحليل على اختبار الفصل الأول "سلوك رفض المدرسة": تم تقدير قدرة أو معلمة الأفراد بناءً على ثلاث طرق متوفرة في حزمة "mirt"، هي (maximum likely-hood (ML))، (maximum a posteriori (MAP))، (expected a posteriori (EAP)).

جدول (٧) ملخص تقديرات بارامترات الأفراد للنموذجين الثلاثي والرباعي المعلم

النموذج	الطريقة	M (SD)	Min.	Max.	Skewness	Kurtosis	SE*
النموذج ثلاثي المعلم	MLE	Inf (NaN)	-7.08	Inf	NaN	NaN	NaN
	MAP	-0.02 (0.91)	-4.44	٢.٠٧	-0.58	١.٤٨	٠.٠٣
	EAP	0.00 (0.94)	-4.49	٢.١٦	-0.55	١.٣٥	٠.٠٣
النموذج رباعي المعلم	MLE	Inf (NaN)	-4.62	Inf	NaN	NaN	NaN
	MAP	0.56 (1.0)	-3.68	٢.١٧	-0.76	٠.٧٨	٠.٠٣
	EAP	0.63 (1.05)	-3.72	٢.٢٧	-0.76	٠.٦٢	٠.٠٣

Note: MLE = Maximum likelihood estimation, MAP = Maximum a posteriori, EAP = Expected a posteriori, and SE* = Mean of standard error of person parameter estimate, NaN = undefined value, inf = infinity

بعد تقديم نتائج التحليل المناسب للفرد، نقدم نتائج إحصائيات الوصف لتقدير معلمة الفرد بموجب النموذج الثلاثي المعلم والنموذج الرباعي المعلم كما بالجدول (٧). كانت هناك ثلاث طرق اعتدنا عليها لتقدير قدرة الفرد، وهي ((maximum likely-hood (ML))، (maximum a posteriori (MAP))، (expected a posteriori (EAP)). من بين الطرق الثلاث، اتضح أن طريقة (EAP) و (MAP) قد أسفرت عن نفس النتائج في الإحصاءات الوصفية لتقدير معلمة الفرد. إن متوسط القدرة المقدرة لـ (١١٦٨) طالب تم

تحليلهم في هذه الدراسة يقترب من (٠.٠٠) بالنسبة للنموذج ثلاثي المعلمة، ويقترب من (٠.٠٦) بالنسبة للنموذج رباعي المعلمة. استنادًا إلى متوسط الخطأ المعياري لتقدير معلمة الفرد (*SE)؛ فإن القدرة المقدرة للأفراد على الاستجابة للمفردات الـ (٩٢) الواردة في الاختبار والتي تم الحصول عليها بواسطة طريقة (EAP) و (MAP) هي الأكثر دقة مقارنةً بطريقة (MLE) بالنسبة للنموذج ثلاثي ورباعي المعلمة.

أظهرت الدراسات التي قام بها (DoĞRuÖZ & Akin Arikan, 2020) أن طريقة (MAP) هي الأكثر دقة لقدرة الطلاب من طريقة (EAP) و (MLE). على الرغم من أن عددًا من الدراسات (على سبيل المثال، (DoĞRuÖZ & Akin Arikan, 2020; Wechsler et al., 2018) أشاروا إلى أنه لا يوجد فرق كبير في نتائج تقدير قدرات الأفراد عند تقديرها بموجب النموذج رباعي المعلم وعند تقديرها بموجب النماذج الأكثر بساطة نموذج IRT (على سبيل المثال ، 2PLM أو 3PLM)، ويذكر (Pardede et al., 2023, p. 104) أنه لا يمكن إنكار أن استخدام النموذج رباعي المعلم يمكن أن يستوعب الاستجابات الشاذة التي يظهرها عدد قليل من المتقدمين في الاختبار العالي القدرات وزيادة دقة تقدير معلمات الفرد (DoĞRuÖZ & Akin Arikan, 2020; Liao et al., 2012; Loken & Rulison, 2010; Primi et al., 2018; Rulison & Loken, 2009)

تُشير النتائج إلى أن طريقتي MAP و EAP تُقدمان تقديرات دقيقة وموثوقة لقدرة الأفراد في كلا النموذجين 3PL و 4PL، مع تفوق واضح على MLE بينما تُقدم هذه الطرق دقة قياس متطابقة (متوسط الخطأ المعياري ٠.٠٣)، إلا أن متوسط تقدير القدرة يختلف بين النموذجين، حيث يُقدر 4PL متوسط القدرة بكونه أعلى من 3PL هذا يُشير إلى أن معلمة الإهمال (u) في 4PL تُحدث فرقًا في كيفية تفسير النموذج لأداء الأفراد وتُؤثر على تقديرات قدرتهم النهائية.

وتُقدم نتائج السؤال الخامس رؤى مهمة حول دقة تقدير قدرة الأفراد وتأثير اختيار النموذج وطريقة التقدير. تُشير النتائج بوضوح إلى عدم ملاءمة طريقة تقدير الاحتمالية القصوى (MLE) في هذا السياق، حيث أسفرت عن تقديرات غير مستقرة (قيم لانهائية)، وهو ما يُعد مشكلة معروفة وموثقة في الأدبيات عند التعامل مع الاستجابات المتطرفة (مثل الأداء الكامل أو الصفري) (Thissen & Wainer, 2001). هذا يُعزز التوجه العام في تطبيقات نظرية الاستجابة للمفردة نحو استخدام الطرق البايزية مثل EAP وMAP.

وفيما يتعلق بطرق التقدير البايزية، تُظهر النتائج أن طريقتي MAP وEAP أسفرتا عن دقة قياس عالية جداً ومتطابقة تقريباً (متوسط الخطأ المعياري ٠.٠٠٣) لكلا النموذجين (3PL و4PL). هذا يُشير إلى موثوقية عالية في تقدير قدرة الأفراد باستخدام هاتين الطريقتين في هذا الاختبار. وعلى الرغم من أن دراسة (DoĞRuÖZ & Akin Arikan, 2020) تشير إلى أن MAP هي الأكثر دقة لقدرة الطلاب من EAP و MLE، فإن نتائج دراستنا تُظهر أن كلا من MAP و EAP كانتا على قدم المساواة في الدقة العالية (متوسط الخطأ المعياري ٠.٠٠٣) وتفوقت كلتاهما بوضوح على MLE.

النقطة الأكثر أهمية في هذه النتائج تكمن في الاختلاف الملحوظ في متوسط تقدير القدرة بين النموذج ثلاثي المعلمة (متوسط يقترب من ٠.٠٠٠) والنموذج رباعي المعلمة (متوسط يقترب من ٠.٦٠). فعلى الرغم من أن بعض الدراسات مثل (DoĞRuÖZ & Akin Arikan, 2020; Wechsler et al., 2018) ذكرت أنه "لا يوجد فرق كبير في نتائج تقدير قدرات الأفراد عند تقديرها بموجب النموذج رباعي المعلم وعند تقديرها بموجب النماذج الأكثر بساطة"، فإن نتائجنا تُظهر اختلافًا في القيم المطلقة للمتوسطات. يُمكن تفسير هذا الاختلاف بأن معلمة "الإهمال" (u) في النموذج 4PL تُقدم تفسيرًا مختلفًا لسلوك الاستجابة. فعندما تسمح معلمة u باحتمال أن يُجيب الفرد ذو القدرة العالية إجابة خاطئة (مما يُشير إلى الإهمال أو التشتت بدلاً من نقص القدرة)، فإن النموذج قد يُعيد تقدير قدرة هؤلاء الأفراد الكامنة بشكل أعلى مما لو كان يُفسر تلك الأخطاء على أنها نقص في القدرة (كما في 3PL). هذا يتوافق مع ما ذكره (DoĞRuÖZ & Akin Arikan, 2020; Liao et al., 2012; Loken & Rulison, 2010; Pardede et al., 2023) من أن استخدام النموذج 4PL "يمكن أن يستوعب الاستجابات الشاذة التي يظهرها عدد قليل من المتقدمين في الاختبار العالي القدرات وزيادة دقة تقدير معلمات الفرد". بينما قد لا يعني ذلك بالضرورة زيادة في "دقة" التقدير بالمعنى الإحصائي للخطأ المعياري، فإنه يعني أن النموذج يُقدم تفسيراً مختلفاً لبعض أنماط الاستجابات، مما قد يؤدي إلى تقدير قيمة قدرة أعلى للأفراد الذين يُظهرون مثل هذه الأنماط.

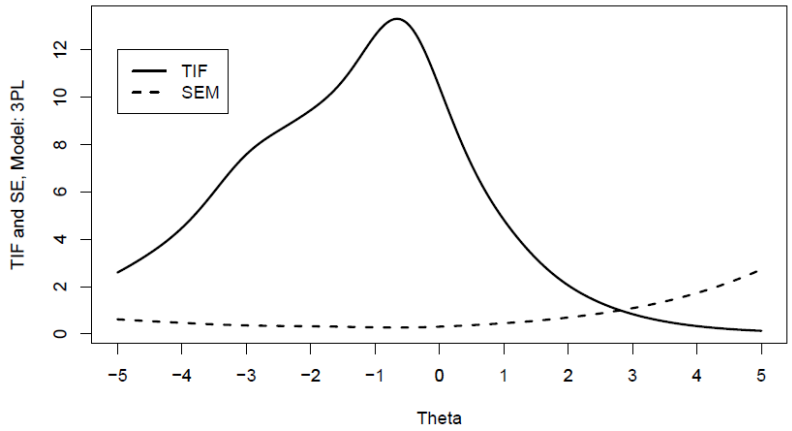
وبشكل عام، تُؤكد هذه النتائج على أهمية اختيار طريقة تقدير مناسبة (مثل MAP أو EAP) لتجنب مشكلات MLE، وتُسلط الضوء على أن إضافة المعلمة الرابعة (u) في 4PL، حتى وإن لم تُحسن بالضرورة ملاءمة المفردة أو النموذج الكلي بشكل جذري في كل الجوانب (كما نوقش في الأسئلة السابقة)، فإنها تُؤثر على كيفية تفسير النموذج لأداء الأفراد ويمكن أن

تُغير من متوسط تقديرات القدرة، مما يُعطي منظورًا مختلفًا للأداء الكامن.

السؤال السادس: "هل تختلف دالة معلومات الاختبار الكلية (Test Information Function - TIF) والخطأ المعياري في القياس (Standard Error of Measurement - SEM)، إلى جانب تقديرات الثبات التجريبي والهامشي (Marginal Reliability)، في ضوء كل من النموذج الثلاثي المعلم والنموذج الرباعي المعلم للاختبار؟"

للإجابة عن هذا السؤال؛ تم التحليل على اختبار الفصل الأول "سلوك رفض المدرسة": لقد تم تقديم النتائج المتعلقة بدالة معلومات الاختبار (TIF) والخطأ المعياري في القياس (SEM) إلى جانب تقديرات الثبات التجريبي والهامشي، في ضوء كل من النموذج الثلاثي المعلم والنموذج الرباعي المعلم كما يلي:

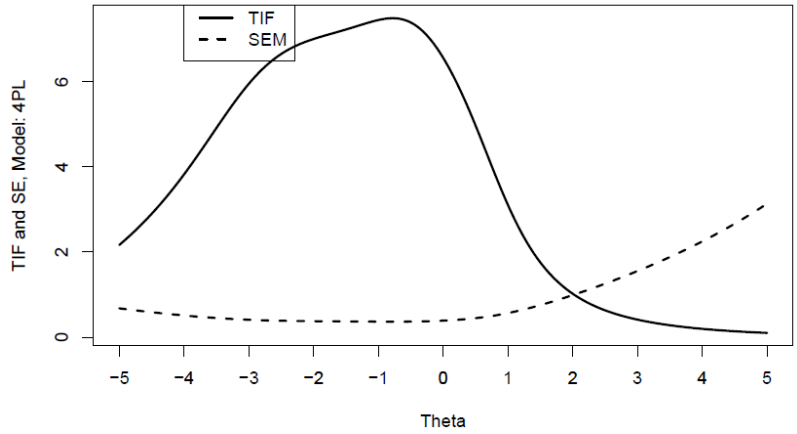
يمكننا الحصول على كمية المعلومات التي يوفرها الاختبار، والتي يمثلها TIF، عن طريق إضافة كمية المعلومات المقدمة من ال (٩٢) مفردة (Hambleton & Swaminathan, 1985; Ronald K. Hambleton et al., 1991). وبنفس طريقة تفسير IIF ومن خلال تحويل التركيز إلى نطاق أوسع؛ ليشمل الاختبار ككل؛ فإن الاختبار الذي يتكون من (٩٢) مفردة سيؤيد معلومات دقيقة حول مستوى قدرة الطلاب عند إدارته إلى أولئك الذين يكون مستوى قدرتهم يقع بين حوالي (- ٦) إلى حوالي (٢.٨). علاوة على ذلك، سيوفر هذا الاختبار أكبر قدر من المعلومات، مما يعني أنه سيؤدي إلى التقدير الأكثر دقة لقدرة الطالب، عندما يتم إجراء الاختبار للطلاب الذين يكون مستوى قدرتهم حوالي (- ٠.٥)، وذلك في ضوء النموذج الثلاثي المعلم في كما بالشكل (١٠).



شكل (١٠) دالة معلومات الاختبار (TIF) والخطأ المعياري للقياس (SEM) للنموذج الثلاثي

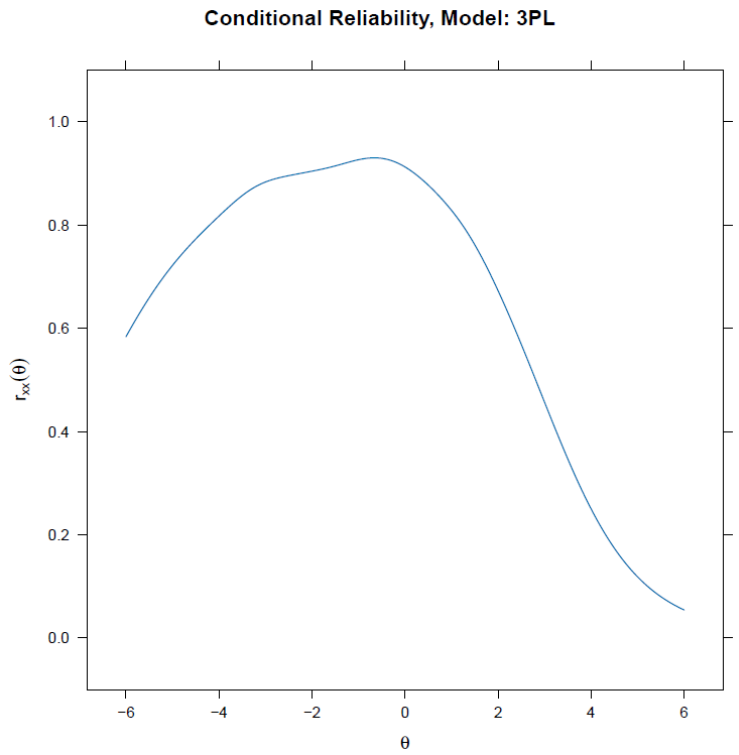
المعلم

كما أن الاختبار الذي يتكون من (٩٢) مفردة سيؤيد معلومات دقيقة حول مستوى قدرة الطلاب عند إدارته إلى أولئك الذين يكون مستوى قدرتهم يقع بين حوالي (- ٦) إلى حوالي (٢). علاوة على ذلك، سيوفر هذا الاختبار أكبر قدر من المعلومات، مما يعني أنه سيؤدي إلى التقدير الأكثر دقة لقدرة الطالب، عندما يتم إجراء الاختبار للطلاب الذين يكون مستوى قدرتهم حوالي (- ٠.٥)، وذلك في ضوء النموذج الرباعي المعلم في كما بالشكل (١١).



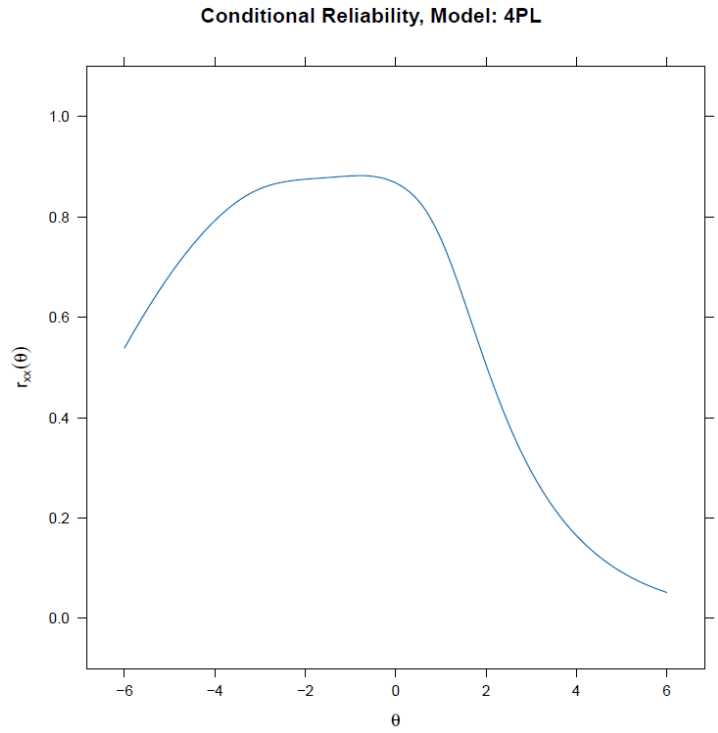
شكل (١١) دالة معلومات الاختبار (TIF) والخطأ المعياري للقياس (SEM) للنموذج الرباعي المعلم

تقديرات الثبات التجريبي يساوي (٠.٨٨٢٠٥٤) وتقديرات الثبات الهامشي يساوي (٠.٨٧٩٨١٣٨) كما بالشكل (١٢) في حالة النموذج ثلاثي المعلمة



شكل (١٢) تقديرات الثبات الهامشي في حالة النموذج ثلاثي المعلمة

تقديرات الثبات التجريبي يساوي (٠.٨٢٤٥٧٧٢) وتقديرات الثبات الهامشي يساوي (٠.٨٢٢٣٣٩١) كما بالشكل (١٣) في حالة النموذج رباعي المعلمة



شكل (١٣) تقديرات الثبات الهامشي في حالة النموذج رباعي المعلمة

دالة معلومات الاختبار (TIF) والخطأ المعياري في القياس (SEM):

النموذج ثلاثي المعلمة (3PL): يوفر الاختبار معلومات دقيقة لمستويات القدرة التي تتراوح بين حوالي (−٦) إلى (٢.٨)، ويُقدم الاختبار أكبر قدر من المعلومات (مما يعني أعلى دقة قياس وأقل خطأ معياري) عند مستوى قدرة حوالي (−٠.٥). هذا يُشير إلى أن الاختبار مصمم بشكل مثالي لقياس الطلاب الذين يقعون ضمن هذا النطاق من القدرة (أقل قليلاً من المتوسط).

النموذج رباعي المعلمة (4PL): يوفر الاختبار معلومات دقيقة لمستويات القدرة التي تتراوح بين حوالي (−٦) إلى (٢). نلاحظ أن النطاق العلوي للمعلومات الدقيقة يقل قليلاً مقارنة بـ 3PL (٢ بدلاً من ٢.٨)، ويُقدم الاختبار أكبر قدر من المعلومات (أعلى دقة) أيضًا عند مستوى قدرة حوالي (−٠.٥)، مما يُشير إلى تمركز ذروة المعلومات في نفس النطاق لنموذج 3PL.

المقارنة بين TIFs والـ SEMs:

بشكل عام، عندما تُقدم دالة معلومات الاختبار (TIF) معلومات أعلى في منطقة معينة،

فإن الخطأ المعياري في القياس (SEM) يكون أقل في تلك المنطقة، ويبدو أن كلا النموذجين يُقدمان معلومات جيدة في المدى السلبي من القدرة (حول −٠.٥)، وهو النطاق الذي يقع فيه متوسط صعوبة المفردات.

الفرق في النطاق العلوي لـ TIF (٢.٨ في 3PL مقابل ٢ في 4PL) يعني أن النموذج 3PL قد يُقدم معلومات أكثر (وبالتالي دقة أعلى وخطأ معياري أقل) للطلاب ذوي القدرات الأعلى قليلاً مقارنة بالنموذج 4PL. هذا متوقع لأن النموذج 4PL يُقدم معلمة الحد الأقصى للأداء (u) التي يمكن أن تُخفض المعلومات المقدمة عند مستويات القدرة العالية جداً إذا كانت المفردات لا تُمكن الطلاب ذوي القدرة العالية من تحقيق أداء مثالي.

تقديرات الثبات التجريبي والهامشي (Marginal Reliability):

النموذج ثلاثي المعلمة (3PL): الثبات التجريبي (Empirical Reliability) = 0.882054، والثبات الهامشي (Marginal Reliability) = 0.8798138

النموذج رباعي المعلمة (4PL): الثبات التجريبي = ٠.٨٢٤٥٧٧٢، والثبات الهامشي = ٠.٨٢٢٣٣٩١

ومما سبق نجد أن النموذج ثلاثي المعلمة (3PL) يُظهر قيم ثبات (تجريبي وهامشي) أعلى بشكل ملحوظ من النموذج رباعي المعلمة (4PL). فالفارق يقترب من ٠.٠٦ (حوالي ٠.٨٨ مقابل ٠.٨٢)، وهو فارق جوهري في تقديرات الثبات.

وتُشير النتائج إلى أن النموذج 3PL يُقدم معلومات أكثر دقة في نطاق أوسع من القدرة (خاصة في المستويات الأعلى قليلاً)، ويُحقق ثباتًا أعلى بكثير مقارنةً بالنموذج 4PL. على الرغم من أن كلا النموذجين يُقدمان ذروة معلوماتية عند مستوى القدرة (−٠.٥)، فإن الثبات العام للاختبار يكون أفضل عند استخدام النموذج 3PL. هذا يُعزز النتائج السابقة (من السؤالين الأول والثاني) التي فضلت النموذج 3PL من حيث ملاءمة المفردة والنموذج الكلي.

وتُسلط نتائج السؤال السادس الضوء على دالة معلومات الاختبار الكلية (TIF)، والخطأ المعياري في القياس (SEM)، وتقديرات الثبات (التجريبي والهامشي) لكل من النموذج ثلاثي المعلمة (3PL) والنموذج رباعي المعلمة (4PL) وتُظهر النتائج أن النموذج 3PL يتفوق على 4PL في تقديم معلومات دقيقة في نطاق أوسع من القدرة (خاصة في المستويات العليا) ويُحقق قيم ثبات أعلى بشكل ملحوظ.

كما تُعد دالة معلومات الاختبار (TIF) مؤشرًا حيويًا لكفاءة الاختبار في قياس مستويات القدرة المختلفة. فالاختبار الذي يُقدم معلومات عالية في مدى معين، يعني أنه يُقدم قياساً أكثر دقة (أقل خطأ معياري) في ذلك المدى. والنتائج التي تُشير إلى أن كلا النموذجين يُركزان أعلى معلومات الاختبار عند مستوى قدرة حوالي (−٠.٥) تُشير إلى أن الاختبار مصمم بشكل جيد لقياس الطلاب الذين يقعون في هذا النطاق، وهو ما يتوافق مع ممارسات تصميم الاختبارات التحصيلية التي تهدف إلى تمييز الطلاب حول متوسط القدرة.

ومع ذلك، فإن تفوق النموذج 3PL في توفير معلومات دقيقة في نطاق أوسع من القدرة (وصولاً إلى ٢.٨ في TIF مقارنة بـ ٢ في 4PL) أمر جدير بالملاحظة. هذا الاختلاف يمكن تفسيره في ضوء معلمة الإهمال (u) في النموذج 4PL. فمعلمة u تُحدد الاحتمال الأقصى للإجابة الصحيحة حتى للطلاب ذوي القدرات العالية جدًا (حيث تُشير قيمة u الأقل من ١ إلى أن الأداء المثالي ليس مضمونًا). إذا كانت المفردات تُظهر قيم u منخفضة، فهذا يعني أن حتى الطلاب ذوي القدرات العالية قد يُخطئون بسبب الإهمال أو الغش أو عوامل أخرى غير القدرة. هذا السلوك، عندما يتم نمذجته، يمكن أن يُقلل من المعلومات التي تُقدمها المفردة (وبالتالي الاختبار) عند مستويات القدرة العالية جدًا، لأن التمييز بين القدرات العالية يصبح أقل وضوحًا إذا كان هناك "سقف" للأداء. هذا يتوافق مع مبدأ أن النماذج التي تُضيف معلمات إضافية قد لا تُحسن دائمًا من معلومات الاختبار الكلية إذا كانت هذه المعلمات لا تُفسر تباينًا جوهريًا أو إذا كانت تقديراتها غير مستقرة (Chalmers, 2012).

كما أن الأكثر أهمية هو تفوق النموذج 3PL في تقديرات الثبات (حوالي ٠.٨٨ مقابل ٠.٨٢ للنموذج 4PL). تُعد قيم الثبات (التجريبي والهامشي) مؤشرًا على اتساق ودقة درجات الاختبار ككل. قيم الثبات الأعلى تُشير إلى أن الاختبار يُقدم قياسًا أكثر استقرارًا وموثوقية. هذا الفارق الجوهري في الثبات يُعزز النتائج السابقة من الأسئلة الأول والثاني التي فضلت النموذج 3PL من حيث ملاءمة المفردة والنموذج الكلي. غالبًا ما تُشير قيم الثبات المرتفعة إلى أن النموذج يُفسر التباين في الأداء بشكل جيد دون إضافة تعقيد غير ضروري. تُدعم هذه النتيجة الدراسات التي تشير إلى أن النماذج الأكثر بساطة (مثل 3PL) قد تكون أكثر استقرارًا وفاعلية عندما لا تُبرر خصائص البيانات تعقيد النماذج الأكثر تعقيدًا (مثل 4PL)، خاصةً إذا كانت معلمة الإهمال لا تُقدم إضافة معلوماتية كبيرة (Loken & Rulison, 2010).

وبشكل عام، تُشير هذه النتائج إلى أن النموذج 3PL يُقدم أداءً قياسياً أكثر قوة وفعالية

لهذا الاختبار، سواء من حيث نطاق المعلومات المقدمة أو من حيث مستويات الثبات، مما يجعله الخيار الأفضل لتطبيقات هذا الاختبار.

التوصيات والبحوث المقترحة:

في ضوء النتائج التي توصل إليها هذا البحث حول ملاءمة نماذج نظرية الاستجابة للمفردة (IRT) الثلاثي (3PL) والرباعي (4PL) لاختبار تحصيلي في إحدى وحدات مادة الاتجاهات الحديثة والمعاصرة في علم النفس "سلوك رفض المدرسة"، يُمكن تقديم التوصيات والبحوث المقترحة التالية:

أولاً: التوصيات:

١. الاعتماد على النموذج ثلاثي المعلمة :(3PL) نظراً لتفوق النموذج 3PL في غالبية مؤشرات ملاءمة المفردة والنموذج الكلي (السؤالان الأول والثاني)، وتقديمه لثبات أعلى للاختبار ككل (السؤال السادس)، يُوصى بالاعتماد على هذا النموذج في التحليلات الإحصائية المستقبلية والتقييمات الخاصة بالاختبارات التحصيلية مثل "سلوك رفض المدرسة" أو اختبارات مشابهة في السياق التعليمي الحالي. يُقدم هذا النموذج التوازن الأمثل بين البساطة والقدرة التفسيرية لبيانات الاختبار.

٢. استخدام طرق التقدير البايزية (MAP وEAP): لتجنب مشكلات عدم الاستقرار التي تُظهرها طريقة MLE، يُوصى بالاستمرار في استخدام طريقتي MAP وEAP لتقدير قدرة الأفراد، نظراً لدقتهما العالية وموثوقيتهما (السؤال الخامس).

٣. مراجعة المفردات غير الملائمة :على الرغم من الأداء الجيد للنموذج 3PL، يجب مراجعة المفردات التي لم تتلاءم مع النموذج في السؤال الأول (٥.٤٣٪ من المفردات) لتقييم صياغتها، أو محتواها، أو مدى تمييزها، بهدف تحسين جودة الاختبار.

ثانياً: البحوث المقترحة:

١. دراسة معمقة لتأثير معلمة الإهمال :(u) على الرغم من أن النموذج 4PL لم يتفوق بشكل عام، إلا أنه أظهر تحسناً طفيفاً في ملاءمة الأفراد وتقليل الأنماط غير الملائمة (السؤال الثالث)، وأثر على متوسط تقدير القدرة (السؤال الخامس). يُقترح إجراء دراسات مستقبلية تُركز بشكل أعمق على معلمة الإهمال(u) ، وتأثيرها على فئات معينة من المفردات أو الأفراد (مثل ذوي القدرات العالية)، ومتى يُمكن أن تُبرر هذه المعلمة إدراجها في النموذج.

٢. تحليل مقارن في سياقات ثقافية وتعليمية مختلفة :بما أن اختيار النموذج يعتمد على خصائص البيانات، يُقترح تكرار هذه الدراسة على اختبارات تحصيلية أخرى وفي سياقات تعليمية مختلفة (مثل المراحل الدراسية المختلفة أو أنواع المدارس) لتعميم النتائج أو تحديد الظروف التي قد يُفضل فيها نموذج على آخر.

٣. تحليل مصادر الأنماط غير الملائمة للاستجابة :بما أن هناك نسبة صغيرة من الأفراد أظهرت أنماط استجابة غير ملائمة، يُمكن إجراء بحث نوعي أو كمي لتحديد الأسباب الكامنة وراء هذه الأنماط (مثل سلوك التخمين، نقص الحافز، تشتت الانتباه، أو حتى الغش) وتطوير استراتيجيات للحد منها.

٤. تأثير حجم العينة على تقدير المعلمات: على الرغم من أن الدراسة لم تتناول هذا الجانب بشكل مباشر، إلا أن حجم العينة يؤثر على استقرار تقدير المعلمات، خاصة في النماذج الأكثر تعقيدًا. يُمكن استكشاف تأثير أحجام العينات المختلفة على دقة واستقرار تقديرات 3PL و4PL في دراسات محاكاة مستقبلية.

المراجع:

فردوس بنت عبدالله الخوتانى. (٢٠٢٣). بناء بنك أسئلة في مقرر الرياضيات "٤" للصف الثاني الثانوى في ضوء نظرية الاستجابة للمفردة وفق النموذج اللوغاريتمى ثلاثي المعلمة. المؤتمر الثامن لتعليم وتعلم الرياضيات : تعليم وتعلم في ضوء المتغيرات الدولية –*بحوث وتجارب متميزة ورؤي مستقبلية*, جدة: جامعة الأعمال والتكنولوجيا.

محمد محمد جاد محمد. (٢٠١٦). دراسة سيكومترية في دقة بناء بنك أسئلة في مادة الرياضيات باستخدام بعض نماذج نظرية الاستجابة للمفردة. *عالم التربية*, ٥٤(١٧), ٢٦٣ – ٢٦٩.

Akaike, H. (1974). A new look at the statistical model identification. IEEE Transactions on Automatic Control, 19(6), 716-723.

Aybek, E. C., & Demirtasli, R. N. (2017). Computerized Adaptive Test (CAT) Applications and Item Response Theory Models for Polytomous Items. *International Journal of Research in Education and Science*, *3*(2), 475-487 .

Baker, F. B. (2001). *The basics of item response theory*. ERIC .

Barton, M. A., & Lord, F. M. (1981). An upper asymptote for the three-parameter logistic item-response model. *ETS Research Report Series*, *1981*(1), i-8 .

Battauz, M. (2020). Regularized estimation of the four-parameter logistic model. *Psych*, *2*(4), 269-278 .

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), Testing structural equation models (pp. 136-162). Sage.

Büyükkıdık, S., & İnal, H. (2023). An Investigation of Item Response Theory Parameter Estimations and Reliability in Multiple Groups. *Gazi Üniversitesi Gazi Eğitim Fakültesi Dergisi*, *43*(2), 825-855 .

Chalmers, R. P. (2012). mirt: A multidimensional item response theory package for the R environment. *Journal of statistical Software*, *48*, 1-29 .

Chen, W.-H., & Thissen, D. (1997). Local dependence indexes for item pairs using item response theory. *Journal of Educational and Behavioral Statistics*, *22*(3), 265-289 .

DeMars, C. (2010). *Item response theory*. Oxford University Press .

Desjardins, C. D., & Bulut, O. (2018). *Handbook of educational measurement and psychometrics using R*. CRC Press .

DoĞRuÖZ, E., & Akin Arikan, Ç. (2020). 3 ve 4PL Madde Tepki Kuramı Modellerine Göre Farklı Yetenek Kestirim Yöntemlerinin Karşılaştırılması. *Pamukkale University Journal of Education*, 1-20. https://doi.org/10.9779/pauefd.585774

Doğruöz, E., & Arıkan, Ç. A. (2020). Comparison of different ability estimation methods based on 3 and 4PL item response theory. *Pamukkale Üniversitesi Eğitim Fakültesi Dergisi*(50), 50-69 .

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 .

Edwards, M. C., Houts, C. R., & Cai, L. (2018). A diagnostic procedure to detect departures from local independence in item response theory models. *Psychological methods*, *23*(1), 138 .

Embretson, S. E., & Reise, S. P. (2013). *Item response theory for psychologists*. Psychology Press .

Erdemir, A., & Önen, E. (2019). BİR, İKİ, ÜÇ VE DÖRT PARAMETRELİ LOJİSTİK MADDE TEPKİ KURAMI MODELLERİNİN KARŞILAŞTIRILMASI. *e-Turkish Studies (elektronik)*, *14* .(١)

Felt, J. M., Castaneda, R., Tiemensma, J., & Depaoli, S. (2017). Using person fit statistics to detect outliers in survey research. *Frontiers in psychology*, *8*, 1-9 .

Hambleton, R., & Han, N. (2005). Assessing the fit of IRT models to educational and psychological test data: A five step plan and several graphical displays. *Advances in health outcomes research methods, measurement, statistical analysis, and clinical applications*, 57-78 .

Hambleton, R. K., & Swaminathan, H. (19 .(٨٥*Item response theory : principles and applications*. Kluwer-Nijhoff Pub. ; Distributors for North America, Kluwer Boston .

Hambleton, R. K., & Swaminathan, H. (2013). *Item response theory: Principles and applications*. Springer Science & Business Media .

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

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. Sage Publications, Inc .

Han, K., & Hambleton, R. (2014). User's manual for WINGEN 3: windows software that generates IRT model parameters and item responses (Center for Educational Assessment Report No. 642). *Amherst, MA: University of Massachusetts* .

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

Jian, X., Buyun, D., & Yuanping, D. (2021). The robust estimation of examinee ability based on the four-parameter logistic model when guessing and carelessness responses exist. *Plos one*, *16*(4), e0250268 .

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Structural Equation Modeling: *A Multidisciplinary Journal*, 6(1), 1-55.

Kalkan, Ö. K. (2022). The comparison of estimation methods for the four-parameter logistic item response theory model. *Measurement: Interdisciplinary Research and Perspectives*, *20*(2), 73-90 .

Kolen, M. J., Brennan, R. L., Kolen, M. J., & Brennan, R. L. (2014). *Linking*. Springer .

Leech, N. L., Barrett, K. C., & Morgan, G. A. (2005). *SPSS for intermediate statistics: Use and interpretation, 2nd ed*. Lawrence Erlbaum Associates Publishers .

Liao, W.-W., Ho, R.-G ,.Yen, Y.-C., & Cheng, H.-C. (2012). The four-parameter logistic item response theory model as a robust method of

estimating ability despite aberrant responses. *Social Behavior and Personality: an international journal*, *40*(10), 1679-1694 .

Loken, E., & Rulison, K. L. (2010). Estimation of a four-parameter item response theory model. *Br J Math Stat Psychol*, *63*(Pt 3), 509-525. https://doi.org/10.1348/000711009X474502

Lu, Y. (2006). *Assessing fit of item response theory models*. University of Massachusetts Amherst .

Maydeu-Olivares, A., & Joe, H. (2006). Limited information goodness-of-fit statistics for item response models. Psychometrika, 71(4), 725-742.

Meijer, R. R., & Tendeiro, J. N. (2015). The effect of item and person misfit on selection decisions: An empirical study .

Meng, X., Xu, G., Zhang, J., & Tao, J. (2020). Marginalized maximum a posteriori estimation for the four-parameter logistic model under a mixture modelling framework. *British Journal of Mathematical and Statistical Psychology*, *73*, 51-82 .

Merino-Soto, C., Angulo-Ramos, M., Rovira-Millán, L. V., & Rosario-Hernández, E. (2023). Psychometric properties of the generalized anxiety disorder-7 (GAD-7) in a sample of workers. *Frontiers in psychiatry*, *14*, 999242 .

Orlando, M., & Thissen, D. (2000). Likelihood-based item-fit indices for dichotomous item response theory models. *Applied Psychological Measurement*, *24*(1), 50-64 .

Orlando, M & ,.Thissen, D. (2003). Further investigation of the performance of S-X2: An item fit index for use with dichotomous item response theory models. *Applied Psychological Measurement*, *27*(4), 289-298 .

Pardede, T., Santoso, A., Diki, D., Retnawati, H., Rafi ,I., Apino, E., & Rosyada, M. N. (2023). Gaining a deeper understanding of the meaning of the carelessness parameter in the 4PL IRT model and strategies for estimating it. *REID (Research and Evaluation in Education)*, *9*(1), 86-117. https://doi.org/10.21831/reid.v9i1.63230

Pavlech, J., & Martinková, P. (2024). Three-and four-parameter item response model in factor analysis framework. *arXiv preprint arXiv:2407.04071* .

Petersen, I. T. (2024). *Principles of psychological assessment: With applied examples in R* .Chapman and Hall/CRC .

Primi, R., Nakano, T. d. C., & Wechsler, S. M. (2018). Using four-parameter item response theory to model human figure drawings. *Revista Avaliação Psicológica*. https://doi.org/10.15689/ap.2018.1704.7.07

Revelle, W. (2023). *psych :Procedures for psychological, psychometric, and personality research (R package version 2.3.3)*. In

Rulison, K. L., & Loken, E. (2009). I've fallen and I can't get up: Can high-ability students recover from early mistakes in CAT? *Applied Psychological Measurement*, *33*(2), 83-101 .

Rupp, A. A., & Zumbo, B. D. (2004). A Note on How to Quantify and Report Whether Irt Parameter Invariance Holds: When Pearson Correlations are Not Enough. *Educational and Psychological Measurement*, *64*(4), 588-599. https://doi.org/10.1177/0013164403261051

Schober, P., Boer, C., & Schwarte, L. A. (2018). Correlation Coefficients: Appropriate Use and Interpretation. *Anesthesia & Analgesia*, *126*(5), 1763-1768. https://doi.org/10.1213/ane.0000000000002864

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics, 6*(2), 461-464.

Sinharay, S., & Haberman, S .J. (2014). How often is the misfit of item response theory models practically significant? *Educational Measurement: Issues and Practice*, *33*(1), 23-35 .

Slocum-Gori, S. L., & Zumbo, B. D. (2011). Assessing the unidimensionality of psychological scales: Using multiple criteria from factor analysis. *Social indicators research*, *102*, 443-461 .

Świst, K. (2015). Item analysis and evaluation using a four-parameter logistic model. *Edukacja*, *134*(3), 79-99 .

Team, R. S. (2025). *RStudio: integrated development environment for R*. In

Tendeiro, J. N., & Meijer, R. R. (2015). How serious is IRT misfit for practical decision-making ?

Thissen, D., & Wainer, H. (2001). An overview of test scoring. *Test scoring*, 13-32 .

Thomas, M. L. (2019). Advances in applications of item response theory to clinical assessment. *Psychological assessment*, *31*(12), 1442 .

Waller, N. G., & Feuerstahler, L. (2017). Bayesian modal estimation of the four-parameter item response model in real, realistic, and idealized data sets. *Multivariate behavioral research*, *52*(3), 350-370 .

Wechsler, S. M., de Cassia Nakano, T., & Primi, R. (2018). Using four-parameter item response theory to model human figure drawings. *Avaliaçao Psicologica: Interamerican Journal of Psychological Assessment*, *17*(4), 473-4 .٨٣

Willse, J. T. (2018). *CTT: Classical Test Theory Functions.(Version 2.3. 3.)[Computer software]*. In

Yalçın, S. (2018). Data fit comparison of mixture item response theory models and traditional models. *International Journal of Assessment Tools in Education*, *5*(2), 301-313 .

Yen, W. M. (1984). Effects of local item dependence on the fit and equating performance of the three-parameter logistic model. *Applied Psychological Measurement*, *8*(2), 125-145 .

Yiğiter, M. S., & Boduroğlu, E. (2024). Item Response Theory Assumptions: A Comprehensive Review of Studies with Document Analysis. *International Journal of Educational Studies and Policy*, *5*(2), 119-138 .

Zanon, C., Hutz, C. S., Yoo, H. H., & Hambleton, R. K. (2016). An application of item response theory to psychological test development. *Psicologia: Reflexão e Crítica*, *29* .

Zhao, Y. (2017). Impact of IRT item misfit on score estimates and severity classifications: an examination of PROMIS depression and pain interference item banks. *Quality of life research*, *26*, 555-564 .

Zhao, Y., & Hambleton, R. K. (2017). Practical consequences of item response theory model misfit in the context of test equating with mixed-format test data. *Frontiers in psychology*, *8*, 484 .

Zheng, C., Guo, S., & Kern, J. L. (2021). Fast Bayesian estimation for the four-parameter logistic model (4PLM). *SAGE Open*, *11*(4), 21582440211052556 .