مجلة كلية التربية – جامعة بورسعيد العدد (٥٢) أكتوبر ٢٠٢٥م
الترقيم الدولي للنسخة المطبوعة: ٥٣١٩ - ٢٠٩٠ الترقيم الدولي للنسخة الالكترونية: ٣٢٦٨-٢٦٨٢
الموقع الالكترونى : website : https://jftp.journals.ekb.eg

بناء بنك أسئلة لمادة "الاتجاهات المعاصرة في علم النفس" باستخدام المعايرة المتزامنة في ضوء نظرية الاستجابة للمفردة

د. أحمد سمير مجاهد أبو الحسن
أستاذ علم النفس التربوي المساعد
قسم علم النفس التربوي
كلية التربية – جامعة الزقازيق

د. إيمان إبراهيم محمد سليم نافع
مدرس علم النفس التربوي
قسم علم النفس التربوي
كلية التربية – جامعة الزقازيق

تاريـــخ استلام البحـث : ٢٠٢٥/٦/٢٣
تاريـــخ قبـول البحـث : ٢٠٢٥/٧/٢
البريد الالكتروني للباحث : emannafea2020@gmail.com

DOI: JFTP-2507-1500

المستخلص

يهدف البحث إلى تقييم فعالية المعايرة المتزامنة في بناء بنك أسئلة لمقرر "الاتجاهات المعاصرة في علم النفس" ضمن نماذج نظرية الاستجابة للمفردة أحادية البعد (راش، أحادي، ثنائي، ثلاثي، رباعي المعلمة)، وتحديد النموذج الأنسب. كما يهدف إلى الكشف عن الخصائص السيكومترية النهائية للبنك بناءً على النموذج المختار بعد تنقية المفردات والأفراد، تكونت العينة من ١١٦٥ طالبًا وطالبةً من جامعة الزقازيق، أتموا الإجابة على ٤٨١ مفردة أولية. استخدم البحث المنهج الوصفي التحليلي، متضمنًا مرحلتين: تحليلاً تقليدياً للمفردات (CTT) لحذف ١٩ مفردة سلبية، ليصبح البنك ٤٦٢ مفردة، وتحليلاً باستخدام نظرية الاستجابة للمفردة (IRT) مع المعايرة المتزامنة، أظهرت المعايرة المتزامنة فعالية كبيرة مع نماذج راش، أحادي وثنائي المعلمة، بينما واجهت صعوبات حسابية مع النموذجين الثلاثي والرباعي المعلمة بسبب ضخامة البيانات. عند مقارنة النماذج الناجحة، تبين أن النموذج ثنائي المعلمة (2PLM) هو الأنسب؛ حيث أظهر ملاءمة لـ ٣٥١ مفردة (٧٦٪)، وتفوق في مؤشرات ملاءمة النموذج الكلية (AIC, BIC, HQ, logLik). يُعزى هذا التفوق إلى قدرة النموذج ثنائي المعلمة على تقدير معلمتي الصعوبة والتمييز لكل مفردة، مما يوفر وصفاً أكثر دقة لخصائص المفردات، أظهرت الخصائص السيكومترية النهائية للبنك، بعد تنقية المفردات والأفراد، أن قيم معامل التمييز المرتفعة تشير إلى قدرة البنك على التمييز بوضوح بين الأفراد ذوي القدرات المختلفة. هذه النتائج تؤكد أن بنك الأسئلة المُعَد يتمتع بخصائص سيكومترية قوية تضمن دقة وموثوقية القياس في التقييمات التربوية والنفسية.

الكلمات المفتاحية: بنك الأسئلة، نظرية الاستجابة للمفردة (IRT)، المعايرة المتزامنة، معالم المفردة، النموذج ثنائي المعلمة.

Building an item bank for "Contemporary Trends in Psychology" using concurrent calibration according to Item Response Theory

ABSTRACT

This research aims to evaluate the effectiveness of concurrent calibration in constructing an item bank for the "Contemporary Trends in Psychology" course within unidimensional Item Response Theory (IRT) models (Rasch, 1PL, 2PL, 3PL, 4PL), and to identify the most suitable model. It also seeks to reveal the final psychometric properties of the item bank based on the selected model after purifying items and persons. The sample consisted of 1165 students from Zagazig University, who completed 481 initial items. The research employed a descriptive analytical approach, comprising two stages: classical item analysis (CTT) to remove 19 negative items, resulting in 462 items, and an Item Response Theory (IRT) analysis with concurrent calibration. Concurrent calibration showed significant effectiveness with Rasch, 1PL, and 2PL models, but faced computational difficulties with 3PL and 4PL models due to large data size. Comparing the successful models, the Two-Parameter Logistic Model (2PLM) was identified as the most suitable; it showed a good fit for 351 items (76%) and excelled in overall model fit indicators (AIC, BIC, HQ, logLik). This superiority is attributed to the 2PLM's ability to estimate both difficulty and discrimination parameters for each item, providing a more accurate description of item characteristics. The final psychometric properties of the item bank, after item and person purification, indicated that high discrimination values signify the bank's clear ability to differentiate between individuals of varying abilities. These results confirm that the developed item bank possesses strong psychometric properties, ensuring the accuracy and reliability of measurement in educational and psychological assessments.

KEYWORDS: Item Bank, Item Response Theory (IRT), Concurrent Calibration, Item Parameters, Two-Parameter Logistic Model.

مقدمة:

تُعد الاختبارات من أدوات التقويم المهمة في العملية التعليمية، إذ تُستخدم بأشكال متعددة لقياس تحصيل الطلاب في المواد الأكاديمية المختلفة. تساعد نتائج الاختبارات في تحديد مدى تحقق الأهداف التعليمية، ومقارنة الطلاب ببعضهم البعض، وتشخيص نقاط القوة والضعف لديهم، أو اختيار الأشخاص لمهام مختلفة، ومعرفة الأنشطة التعليمية المتنوعة التي يقدمها المعلم والتي تُسهم في رفع مستوى تحصيل الطلاب. ولذلك، حرص المشرفون التربويون على أن تكون هذه الاختبارات ذات كفاءة عالية في عملية القياس والتقييم، فهي تُعد موقفًا تربويًا يتوقع أن يتضمن أسئلة تقيس الأهداف الأساسية، ويغطي أهم أجزاء المحتوى الأكاديمي، ويتوقع من الطالب أن يفكر بعمق حيال هذا الموقف مقارنة بمواقف أخرى. وبالتالي، فإن الأفكار والمعلومات المتضمنة في الاختبار تحظى بنصيب مميز من التفكير، مما يزيد من احتمالية الاحتفاظ بها.

ويُشير (Falani et al, 2020, p. 546) إلى أن مفردات الاختبار تأخذ عدة أشكال يمكن استخدامها في تقويم تحصيل الطلاب، فبعضها يتطلب من المفحوص اختيار الإجابة من بين بدائل متعددة مثل اختبارات الاختيار من متعدد والصواب والخطأ والمطابقة، بينما يتطلب البعض الآخر تقديم الإجابة بكلماته الخاصة، مثل اختبارات المقال والإجابة القصيرة والإكمال.

وعند تطبيق نماذج نظرية الاستجابة للمفردة (IRT) في بناء وتطبيق الاختبارات، تُعد دقة التقييم عنصرًا جوهريًا في تحديد ثبات الاختبار. ومع ذلك، فإن الممارسة الشائعة في التحقق من صحة وتقييم الاختبارات غالبًا ما تكتفي بالإبلاغ عن معلومات حول التقييم دون توضيح كيفية اشتقاق هذه المعلومات للوصول إلى مستوى مقبول من الدقة (Tan, 2024, p. 33)، وتُعد العلاقة بين نماذج IRT والثبات والمعلومات التي يقدمها التقييم عناصر حاسمة، إلا أنها نادرًا ما تُناقش بتفصيل لربطها بعملية التقييم والبناء. وبالإضافة إلى ذلك، فإن الجمع بين نماذج نظرية الاستجابة للمفردة (راش، 1PL، 2PL، 3PL، 4PL) يعتبر أمرًا مهمًا لفهم العلاقة بين معلمات المفردة الأربعة (التمييز والصعوبة والتخمين والإهمال) ودالة المعلومات، كما أنها تساعد في انتقاء المفردات عند إنشاء بنك الأسئلة.

ويذكر (Hambelton & Swaminthan, 1985, p. 257) أنه اعتمادًا على الغرض المقصود من الاختبار، يمكن استخلاص المفردات ذات الخصائص المرغوبة من بنك الأسئلة واستخدامها لإنشاء اختبار ذي خصائص إحصائية مهمة. وعلى الرغم من استخدام الإحصائيات الكلاسيكية للمفردات (صعوبة المفردة والتمييز) لهذا الغرض، إلا أنها ذات قيمة محدودة لوصف المفردات الموجودة في البنك لأن هذه الإحصائيات تعتمد على مجموعة معينة تُستخدم في عملية

معايرة المفردات. ومع ذلك، لا تحتوي معلمات مفردة نموذج الاستجابة للمفردة على هذا القيد، وبالتالي، فهي ذات استخدام أكبر بكثير في وصف مفردات الاختبار في بنك الأسئلة.

ويُشير (Lee & Lee, 2018, p. 644) إلى وجود ثلاثة أنواع من المعايرة: المعايرة المنفصلة، والمعايرة المتزامنة، والمعايرة الثابتة. ورغم أن الدراسات التي تقارن الطرق الثلاث جميعها في الوقت نفسه قليلة، إلا أن الأداء النسبي لهذه الطرق يقدم نتائج متباينة. ومع ذلك، إذا لم يكن النموذج مناسبًا (مثلًا: وجود تعددية الأبعاد)، فإن المعايرة المنفصلة تقدم نتائج أفضل. ويرى (Joo et al., 2017, p. 83) أن المعايرة المتزامنة باستخدام مفردات الربط أثبتت فعاليتها كبديل فعال للمعايرة المنفصلة وربط النتائج عند بناء وتطوير بنوك أسئلة كبيرة بفضل دقتها في تقدير المعلمات، والتي تُعد ضرورية لدعم الاختبار المستمر، فالمعايرة المتزامنة هي عبارة عن تقدير معلمات المفردات بشكل مشترك باستخدام مجموعة بيانات موحدة، مع اعتبار استجابات المفردات الفريدة لكل مجموعة مفقودة للمفحوصين الذين لم يستلموها.

ويُشير (Kim, 2019, pp. 512-513) إلى وجود صعوبتين رئيسيتين في تقدير معلمات المفردات في نظرية استجابة المفردة .(IRT) أولاً، تتطلب طريقة التقدير بأقصى احتمال هامشي (MMLE) التي تُنفذ عبر خوارزمية التوقع والتعظيم، أو النهج البايزي الهامشي تقدير معلمات المفردات من خلال تهميش (أو تكامل) معلمات الأشخاص، ويتطلب هذا النوع من التقدير توصيف معلمات الأشخاص من خلال توزيع احتمالي، غير أن تحديد هذا التوزيع بدقة يُعد أمرًا صعبًا. ثانياً: غياب نظام إحداثي معياري، وبشكل أكثر تحديدًا، في كلٍّ من النموذج أحادي البُعد (UIRT) والنموذج متعدد الأبعاد (MIRT)، فإن موقع نقطة الأصل ووحدة القياس على كل محور من محاور الإحداثيات تكون اختيارية، كما أن اتجاه المحاور في نموذج MIRT أيضًا يُعد اختياريًا. ولمعالجة هذه الحالات من عدم التعيين، تقوم برامج معايرة IRT عادةً بتحديد توزيع المتغيرات الكامنة ليكون توزيعًا طبيعيًا معياريًا أحادي أو متعدد المتغيرات، وذلك اعتمادًا على عدد الأبعاد المحددة من قِبل المستخدم في نموذج IRT، وعمومًا، تميل المعايرة المتزامنة إلى أن تكون أدق من المعايرة المنفصلة إذا اقترنت بطرق المنحنيات المميزة وكان النموذج مناسبًا، ولذلك سوف يستخدم الباحثان المعايرة المتزامنة.

مشكلة البحث وأسئلته:

تُعد الاختبارات التحصيلية حجر الزاوية في العملية التعليمية والتربوية، فهي الأداة الرئيسة لتقييم مدى تحقيق الأهداف التعليمية، وقياس تحصيل الطلاب في مختلف المراحل الدراسية. إن الاهتمام بجودة هذه الاختبارات لا ينعكس فقط على دقة القياس، بل يمتد ليشمل كافة مكونات العملية التعليمية من أهداف ومحتوى وأساليب تدريسية، مما يؤثر في النهاية على الطالب الذي يُعد محور هذه العملية وهدفها الأساسي. فالاختبارات الجيدة تُمكن المعلمين من فهم نقاط القوة والضعف لدى

الطلاب، وتُوفر لهم التغذية الراجعة اللازمة لتحسين أدائهم التعليمي، وتساعد في اتخاذ قرارات تعليمية وتربوية مستنيرة.

وعلى الرغم من الأهمية المحورية للاختبارات التحصيلية، إلا أن عملية إعدادها وتطويرها غالبًا ما تواجه تحديات تتعلق بضمان دقة وموضوعية القياس. فكثيرًا ما تعتمد الممارسات التقليدية في بناء الاختبارات على نظرية القياس الكلاسيكية (Classical Test Theory - CTT)، والتي على الرغم من بساطتها وفوائدها، إلا أنها تُقدم تقديرات لخصائص المفردات (الأسئلة) والأفراد التي تعتمد بشكل كبير على العينة التي طُبق عليها الاختبار. هذا القيد يُقلل من قابلية تعميم النتائج ويُعيق بناء بنوك أسئلة مرنة ودقيقة يمكن الاستفادة منها في تصميم اختبارات متنوعة ومُعدلة وفقًا لاحتياجات القياس المختلفة. تبرز الحاجة المُلحة إلى تجاوز هذه القيود لضمان أن تكون المفردات الاختبارية ذات قدرة عالية على التمييز بين مستويات وقدرات الطلاب المختلفة، وأن تُقدم معلومات قياسية دقيقة حول مدى إتقانهم للمادة الدراسية.

تُقدم نظرية الاستجابة للمفردة (Item Response Theory - IRT) حلًا متقدمًا لهذه التحديات، حيث تُمكن من تقدير خصائص المفردات (مثل الصعوبة والتمييز) وخصائص الأفراد (مثل القدرة أو السمة المقاسة) بشكل مستقل عن بعضها البعض، ومعتمدًا على مقياس واحد ومشترك. هذا التطور يُتيح بناء بنوك أسئلة (Item Banks) ضخمة، تُخزن فيها المفردات مع معالمها المُعايرة بدقة، مما يُسهل عملية سحب المفردات وتصميم اختبارات مُختلفة (مثل الاختبارات التكيفية المحوسبة) تُلبي متطلبات قياسية محددة وتُقدم قياسًا أكثر كفاءة ودقة.

وفي سياق مقرر "الاتجاهات المعاصرة في علم النفس"، الذي يتميز بمفاهيمه الحديثة والمتغيرة باستمرار، تزداد أهمية وجود أداة تقييم مرنة ودقيقة تُواكب هذه التطورات. فإعداد كتاب مقرر جديد يتطلب بالضرورة تطوير أدوات تقييم مُبتكرة تُعكس محتوى المقرر وتُقيمه بفاعلية. من هذا المنطلق، نبعت فكرة البحث الحالي أثناء عملية إعداد أسئلة لكل فصل من فصول هذا المقرر، مما أدى إلى التفكير في كيفية تدريج هذه الاختبارات الخمسة وفقًا لأحدث التقنيات السيكومترية المتاحة، وصولًا إلى إنشاء بنك أسئلة مُعاير بدقة من خلال المعايرة المتزامنة من خلال اقترانها بطرق المنحنيات المميزة واستخدام النموذج المناسب.

يُسعى هذا البحث إلى استكشاف مدى فعالية تطبيق المعايرة المتزامنة (Concurrent Calibration)، وهي إحدى الأساليب المتقدمة في IRT التي تتيح معايرة مجموعة كبيرة من المفردات في وقت واحد، وذلك ضمن إطار نماذج نظرية الاستجابة للمفردة أحادية البعد. على الرغم من أن نماذج IRT مثل النموذج أحادي المعلمة (1PL)، والنموذج ثنائي المعلمة (2PL)، والنموذج ثلاثي المعلمة (3PL)، والنموذج رباعي المعلمة (4PL) تُوفر أدوات قوية للتحليل، إلا أن تطبيقها

العملي قد يواجه تحديات، خاصة مع النماذج الأكثر تعقيدًا التي تتطلب بيانات أكثر كثافة وربما لا تتناسب مع جميع أنواع البيانات. لذلك، فإن هذا البحث لا يهدف فقط إلى بناء بنك أسئلة، بل يتعداه إلى تقييم أداء هذه النماذج المختلفة وتحديد النموذج الأنسب لبيانات مقرر "الاتجاهات المعاصرة في علم النفس."

وبالنظر إلى ما سبق، يحاول البحث الإجابة عن الأسئلة التالية:

١. ما مدى فعالية استخدام المعايرة المتزامنة في بناء بنك أسئلة لمقرر "الاتجاهات المعاصرة في علم النفس" ضمن نماذج نظرية الاستجابة للمفردة أحادية البعد (راش، أحادي المعلمة، ثنائي المعلمة، ثلاثي المعلمة، رباعي المعلمة)؟

٢. ما هو النموذج الأنسب من بين نماذج نظرية الاستجابة للمفردة أحادية البعد التي أتمت التحليل بنجاح (راش، أحادي المعلمة، ثنائي المعلمة) لبيانات بنك أسئلة مقرر "الاتجاهات المعاصرة في علم النفس"؟

٣. ما هي الخصائص السيكومترية النهائية لبنك أسئلة مقرر "الاتجاهات المعاصرة في علم النفس" في ضوء النموذج المختار (النموذج ثنائي المعلمة)، بعد تنقية المفردات والأفراد؟

أهداف البحث:

يهدف البحث الحالي إلى:

١. تقييم فعالية استخدام المعايرة المتزامنة في بناء بنك أسئلة لمقرر "الاتجاهات المعاصرة في علم النفس التربوي" ضمن النماذج المختلفة لنظرية الاستجابة للمفردة أحادية البعد (راش، أحادي المعلمة، ثنائي المعلمة، ثلاثي المعلمة، رباعي المعلمة)، مع تحديد أي النماذج لم تتمكن من إكمال عملية التدريج بنجاح.

٢. تحديد النموذج الأنسب من بين نماذج نظرية الاستجابة للمفردة أحادية البعد التي أتمت التحليل بنجاح (راش، أحادي المعلمة، ثنائي المعلمة) لبيانات بنك أسئلة مقرر "الاتجاهات المعاصرة في علم النفس التربوي."

٣. الكشف عن الخصائص السيكومترية النهائية لبنك أسئلة مقرر "الاتجاهات المعاصرة في علم النفس التربوي" في ضوء النموذج المختار (النموذج ثنائي المعلمة)، وذلك بعد تنقية المفردات والأفراد.

أهمية البحث:

تتمثل أهمية هذا البحث في مدى إسهامه في إثراء المعرفة من الجانبين النظري والتطبيقي، حيث يهدف إلى تقديم معيار جديد لزيادة دقة أدوات القياس (الاختبارات). ونظرًا لأن الاختبارات التحصيلية تُعد من أكثر أدوات القياس شيوعًا وانتشارًا في الجامعات والمؤسسات التعليمية بجميع مراحلها، فإن

جودتها تؤثر بشكل مباشر على جودة العملية التعليمية ونتائجها، وتكمن أهمية البحث الحالي في كونه سيسهم في:

١. **تحسين دقة أدوات القياس: من خلال توفير منهجية واضحة ومُفصلة لاستخدام نظرية الاستجابة للمفردة (IRT) والمعايرة المتزامنة، سيُقدم البحث نموذجًا عمليًا لانتقاء أفضل المفردات التي تُحقق أعلى درجات الصدق والثبات للاختبارات التحصيلية، وذلك بناءً على دالة المعلومات وقدرتها على التمييز بين قدرات الطلاب المختلفة.**
٢. **بناء بنوك أسئلة عالية الجودة: ستُسفر نتائج البحث عن بناء بنك أسئلة مُعاير سيكومتريًا لمقرر "الاتجاهات المعاصرة في علم النفس"، يتمتع بخصائص سيكومترية عالية (كما سيُظهرها النموذج ثنائي المعلمة المختار)، مما سيُعد مرجعًا للقائمين على تطوير الاختبارات.**
٣. **توجيه الباحثين والمختصين: سيُساعد البحث في توجيه الباحثين والمختصين في مجال القياس النفسي والتربوي نحو اختيار المفردات التي تُحقق الصدق والثبات لنتائج الاختبارات، وسيُقدم لهم إطارًا لإنشاء بنوك أسئلة مُشابهة لمقررات أخرى.**
٤. **دعم اتخاذ القرارات التعليمية: تُعطي بنوك الأسئلة المُعايرة صورة دقيقة وصادقة عن مستوى إتقان الطلاب للمهارات والمعارف المكتسبة، وهو ما يُعد أساسًا للقرارات التربوية المتعلقة بالطلاب، وقرارات الترفيع الأكاديمي، وتقييم مستوى خريجي الجامعات، مما ينعكس إيجابيًا على قدرتهم التنافسية في سوق العمل.**
٥. **تسهيل عملية بناء الاختبارات: سيُسهم البحث في مساعدة القائمين على تطوير الاختبارات في المؤسسات التعليمية من خلال توفير بنك أسئلة جاهز ومُعاير، مما يُسهل عليهم سرعة إنشاء اختبارات تحصيلية متنوعة ومُحكمة في أي وقت، ويُعزز الاهتمام ببنية الاختبارات وتطويرها المستمر.**

مصطلحات البحث:

بنك الأسئلة (Item Bank)

هو مجموعة كبيرة من الأسئلة أو المفردات التي تمت معايرتها سيكومتريًا وتخزينها إلكترونيًا، مما يسمح بسحب مفردات متنوعة لبناء اختبارات مختلفة الأغراض والمستويات بكفاءة ودقة (Hambelton & Swaminthan, 1985, p. 257) وقد ظهرت فكرته في عام ١٩٦٦ وفقًا للنظرية الكلاسيكية (CTT) ونظرية الاستجابة للمفردة (IRT)، وحظي باهتمام كبير من خبراء القياس والتقويم (ماجد عبده ناجى وعبد المحسن بن رشيد المبدل، ٢٠٢٢، ص ٥٥٢)، وتُبرز أهميته في بناء اختبارات لقياس الأهداف التربوية المنشودة بسهولة، وتقليل الوقت والجهد المستغرق في بناء

الاختبارات، بالإضافة إلى المرونة في القياس من خلال تشكيل اختبارات سريعة وسهلة (ماجد عبده ناجي وعبد المحسن بن رشيد المبدل، ٢٠٢٢، ص ٥٥٩-٥٦٠)

نظرية الاستجابة للمفردة (Item Response Theory - IRT)

هي إطار قياسي حديث يُستخدم لتقدير خصائص المفردات (مثل الصعوبة والتمييز) وخصائص الأفراد (مثل القدرة أو السمة) بشكل مستقل عن بعضها البعض وعن العينة التي طُبق عليها الاختبار، وذلك بالاستناد إلى منحنى خصائص المفردة (Item Characteristic Curve - ICC) وتُعتبر هذه النظرية الأساس الذي تستند إليه معايرة المفردات وبناء بنك الأسئلة في البحث الحالي، حيث إن معلمات مفردة نموذج الاستجابة للمفردة ذات استخدام أكبر بكثير في وصف مفردات الاختبار في بنك الأسئلة مقارنةً بالإحصائيات الكلاسيكية (Hambelton & Swaminthan, 1985, p. 257)

المعايرة المتزامنة (Concurrent Calibration)

هي منهجية تطبيقية ضمن إطار IRT ، تتضمن عملية تقدير واحدة فقط لمعلمات جميع المفردات في بنك الأسئلة بشكل متزامن، من خلال دمج بيانات النماذج المختلفة. (Lee, 2011, p. 12)، وتُنتج تقديرات معلمات المفردات على نفس المقياس الأساسي دون الحاجة إلى عملية تحويل مقياس إضافية (Shen et al., 2024, p. 553) & (Lee & Lee, 2018, pp. 641-643). وستتكون المصفوفة الخاصة ببنك الأسئلة من صفوف تمثل جميع الأفراد ومن أعمدة تمثل جميع المفردات (أجاب عنها جميع الطلاب تقريباً بلا استثناء) وستتم معالجتها بصورة متزامنة، وتُعد هذه الطريقة المنهجية المحددة التي ستستخدم لمعايرة بنك الأسئلة في البحث الحالي.

معالم المفردة (Item Parameters)

هي القيم الإحصائية التي تصف خصائص كل مفردة داخل نموذج IRT، وهي نتائج أساسية لعملية المعايرة، تشمل هذه المعالم:

- **صعوبة المفردة (Difficulty Parameter / b-parameter): تُشير إلى مستوى القدرة الذي يمتلكه الفرد لكي تكون لديه فرصة ٥٠٪ للإجابة الصحيحة على المفردة، ومعرفتها ضرورية لتفسير نتائج السؤال.(Choppin, 1976, p. 218)**
- **تمييز المفردة (Discrimination Parameter / a-parameter): تُشير إلى مدى قدرة المفردة على التمييز بين الأفراد ذوي القدرات المختلفة (الأفراد ذوي القدرة العالية مقابل ذوي القدرة المنخفضة).(Choppin, 1976, p. 218) وتُعد هذه المعالم ضرورية "للتحقق من معاملات الثبات والصدق والصعوبة والتمييز" في المفردات (وليد محمد فرج الله، ٢٠١٨، ص ٤٦٣).**

نماذج نظرية الاستجابة للمفردة أحادية البعد (Unidimensional IRT Models)

هي نماذج IRT التي تفترض أن الاختبار يقيس بُعدًا واحدًا أو سمة كامنة واحدة. هذه النماذج هي التي ستُطبق ويُقارَن بينها لتحديد الأنسب لبيانات البحث، وتُساهم في "تحقيق الموضوعية في تقويم تحصيل الطلاب" (ماجد عبده ناجي وعبد المحسن بن رشيد المبدل، ٢٠٢٢، ص ٥٥٩). تشمل هذه النماذج:

- نموذج راش :(Rasch Model) نموذج أحادي المعلمة يُقدر الصعوبة فقط، ويفترض تمييزًا متساويًا الواحد الصحيح (١) ومعامل تخمين صفرًا لجميع المفردات.
- نموذج أحادي المعلمة :(1PLM) يُشير أحيانًا إلى نموذج راش أو نسخة معدلة منه تُقدر الصعوبة وتفترض تمييزًا ثابتًا غير الواحد الصحيح (غير ١) ومعامل تخمين صفرًا لجميع المفردات.
- نموذج ثنائي المعلمة :(2PLM) يُقدر صعوبة المفردة وتمييزها.

مطابقة النموذج للبيانات (Model-Data Fit)

هي عملية إحصائية لتقييم مدى توافق استجابات المفحوصين الفعلية مع ما يتنبأ به النموذج النظري لـIRT ، وتُستخدم مؤشرات إحصائية لتقييم هذه المطابقة. هذا المعيار حاسم لتحديد النموذج الأنسب لبنك الأسئلة، حيث إن "تحليل النتائج وتحديث الأسئلة" و"تقييم أداء السؤال" هي عمليات ضرورية لضمان جودة المفردات وتوافقها (Cohen & Swerdlik, 2018, p. 269)؛ (Haladyna, 2004, p. 14-18)

ملاءمة المفردة (Item Fit)

تُشير إلى مدى توافق استجابات المفحوصين على مفردة معينة مع ما يتنبأ به نموذج IRT لهذه المفردة. المفردات التي لا تتناسب جيدًا قد تحتاج إلى المراجعة أو الحذف. تُعد ملاءمة المفردة جزءًا أساسيًا من عملية تنقية المفردات بعد التحليل الكلاسيكي، وتضمن جودتها في البنك، حيث يتم "استبعاد وحذف المفردات غير المناسبة إحصائياً والإبقاء علي الصالح منها" في مرحلة التجريب (وليد محمد فرج الله، ٢٠١٨، ص ٤٦٣). وهذا يضمن أن بنك الأسئلة "يعزز فعالية القياسات لأنه يمكن تحسين مفردات الاختبار من حيث الثبات والصدق.(Bahrom et al., 2025, p. 11) "

ملاءمة الشخص (Person Fit)

تُشير إلى مدى توافق نمط استجابات الفرد مع ما يتنبأ به نموذج IRT لقدرته. الأفراد الذين لا تتناسب استجاباتهم (قد يكونون أجابوا بشكل عشوائي، أو ارتكبوا أخطاء كبيرة، أو كانت لديهم سمة كامنة مختلفة) قد يتم تنقية بياناتهم لضمان دقة تقدير معالم المفردات والأفراد. ويُشير اهتمام البحث بـ

"تحليل أداء الطلاب" و"التعرف على نتائج التعلم لكل طالب على حدة" إلى أهمية فهم استجابات الأفراد (ماجد عبده ناجي وعبد المحسن بن رشيد المبدل، ٢٠٢٢، ص ٥٦٠).

دالة المعلومات الاختبار (Test information function) (TIF): وهي علاقة رياضية تُعبر عن مجموعة من دوال المعلومات لجميع مفردات الاختبار (ALKursheh, Al-zboon & AlNasraween, 2022, p. 1114).

الإطار النظري:

أولاً: بنوك الأسئلة Item Bank

يري (ماجد عبده ناجى و عبد المحسن بن رشيد المبدل، ٢٠٢٢، ص ٥٥٢) أن فكرة إنشاء بنوك الأسئلة قد ظهرت في عام ١٩٦٦ في شتى المجالات المعرفية وفقاً للنظرية الكلاسكية (CTT) ونظرية الاستجابة للمفردة (IRT)، حيث حظيت باهتمام كبير من قبل خبراء القياس والتقويم في العالم وخاصة بنوك الأسئلة التي يتم اعدادها وفقاً لنظرية الاستجابة للمفردة .

تعريف بنك الأسئلة:

يعرفه (Choppin, 1976, p. 216) بأنه "مجموعة من مفردات الاختبار المنظمة والمفهرسة بطريقة مماثلة للكتب الموجودة في المكتبة، ولكن أيضًا مع بيانات معايرة حول خصائص القياس الخاصة بها".

ويعرفه (Hambelton & Swaminthan, 1985, p. 255) أنه "عدد كبير من المفردات التي تتوافق مع الأهداف أو المهارات أو المهام، والتي لها خصائص سيكومترية معلومة، والتي يتم وضعها في مكان آمن بحيث يسهل الإضافة لها والسحب منها، ويمكن لمطوري الاختبارات استخدامها لبناء الاختبارات على أساس "حسب الحاجة".

ويعرفه (Millman & Greene, 1989, p. 29) بأنه " نظام حاسوبي يُخزن العناصر الاختبارية مع بياناتها الإحصائية، ما يسمح بإنشاء اختبارات مُخصصة بشكل سريع ودقيق".

ويعرفه (Haladyna, 2004, p. 56) " أنه مجموعة منظمة من المفردات الاختبارية المصممة لقياس أهداف تعليمية أو مهارات معينة وتُستخدم لإعداد اختبارات متنوعة حسب الحاجة".

ويعرفه (Wainer et al., 2000, p. 112) "مجموعة من الأسئلة المترابطة التي تخضع لمعايير محددة في التصميم وتُستخدم لتقديم اختبارات قابلة للتكيف أو ثابتة".

ويعرفه (Downing & Haladyna, 2006, p. 268) بأنه "نظام يُركز على إنشاء مجموعة من المفردات التي يتم تحليلها إحصائيًا بهدف تحسين جودة القياس وتوفير اختبارات متنوعة".

ويعرفه (Cohen & Swerdlik, 2018, p. 243) أنه "عبارة عن مجموعة كبيرة نسبيًا وسهلة الوصول من أسئلة الاختبار."

ويعرفه (زياد أحمد العبد الله، ٢٠١٢، ص ٩٦) مجموعة كبيرة نسبياً من المفردات والتي تتميز بخصائص سيكومترية معينة تم تدريجها بوحدة محددة بهدف تشكيل قياس له صفر واحد مشترك قادر على قياس أكبر قدر ممكن من المتغير المراد قياسه.

ومما سبق يعرف الباحثان بنك الأسئلة بأنه " مجموعة كبيرة من مفردات الاختبار الجيدة التي يتم تحليل جودتها ولها خصائص سيكومترية معينة، و يتم تخزينها بشكل منهجي في جهاز كمبيوتر بحيث يساعد المعلمين وواضعى الاختبار في الحصول علي نسخة اختبارية عند الحاجة إليها وذلك لمساعدتهم في تقييم للطلاب وقياس إنجازاتهم أو قدراتهم."

أنواع بنوك الأسئلة

١- بنك الأسئلة التقليدي (Traditional item bank)

يذكر (توفيق علي سالم و سوسن حسن أبو الهادي،٢٠١٩، ص ١٧٦) أن بنك الأسئلة التقليدى هو البنك الذي يحتوي على مجموعة من الأسئلة المعدة مسبقًا، ويحتوي هذا النوع على أسئلة متنوعة تغطي الموضوعات المختلفة، وفي هذا النوع من البنك تكون عمليات تخزين فقرات الأسئلة، واستدعائها عند الحاجة، واختيار فقرات منها وكذا القيام بعمليات تطوير للفقرات المخزنة وإضافة فقرات جديدة تتم بطريقة يدوية، وهي طريقة معقدة تتطلب خبرة وجهد ووقت وتكلفة مستمرة، كما أنها قد لا تكون دقيقة في العمليات والنتائج وأقل أمانا في الحفاظ على السرية.

٢- بنك الأسئلة التكيفي (Adaptive item Banks)

يوضح (Winer et al., 2000, p. 119) أنه تم تصميم هذا النوع من البنك يتم تعديل مستوى صعوبة الأسئلة بناء علي أداء الطالب، بحيث يتغير اختبار الطالب بناءً على أدائه في الأسئلة السابقة، فالأسئلة التي يتم تقديمها للطلاب تتناسب مع مستواهم مما يسمح بتقييم دقيق للمهارات والمعرفة بناءً على استجابة الطالب.

٣- بنك الأسئلة المحوسب (Computerized Item Banks)

يرى كل من (Bergstrom &Gershon, 1995, p. 187) و (Chuesathuchon & Waugh, 2008, p. 108) و(DeVellis, 2017, p. 211) و (Downing & Haladyna, 2011, p. 262 أن هذا النوع من البنوك يتضمن استخدام برامج كمبيوتر متخصصة في تنظيم الأسئلة وتصنيفها وتخزينها واجراء الإحصائيات المرتبطة بها، وفيه يتم إضافة وحذف الأسئلة بسهولة، ويساعد بنك الأسئلة المحوسب المنظم جيدًا والمُحافظ عليه جيدًا فى تسهيل وتعزيز بناء الاختبارات الورقية والقلمية والاختبارات المحوسبة، ويمكن فرز المفردات وتصفيتها لتمكين مراجعتها بسهولة من قبل خبراء المحتوى والموظفين النفسيين.

٤- بنك الأسئلة الإلكتروني (Electronic Item Banks)

يشير (Baker & Kim, 2004, p. 25) و(توفيق علي سالم وسوسن حسن أبو الهادي، ٢٠١٩، ص ١٧٦) إلى أن هذا النوع يدار عبر الانترنت والمنصات ويعتمد على أنظمة إلكترونية وبرامج لتخزين الأسئلة وإدارتها، مما يساعد واضعي الامتحان أو الإداريين إضافة الأسئلة وتعديلها، وفيه يتم الوصول إلى الأسئلة من أي مكان عن طريق الإنترنت، ويتسم بالدقة والسرعة والسهولة والأمان، وقدي مكلف نوعا ما في مرحلة التصميم، ولكن هذه الكلفة ليست مستمرة عند الاستخدام كما في الطريقة التقليدية.

تصنيفات بنوك الأسئلة

التصنيف الأول

أشار كل من (توفيق علي سالم وسوسن حسن أبو الهادي، ٢٠١٩، ص ١٧٢) و (Kolen & Brennan, 2013, p. 276) أنه تم تصنيف بنوك الأسئلة للنوعين:

١- بنك الأسئلة المفتوح: ويشير بنك الأسئلة المفتوح إلى مجموعة من الأسئلة التي يتم تنظيمها وتخزينها بشكل يسمح للمستخدمين سواء كانوا معلمين أو طلابًا، بإنشاء اختبارات متنوعة باستخدام الأسئلة المتاحة، وفي هذا النوع من البنوك لا توجد قيود على كيفية استخدام الأسئلة أو على الشخص الذي ينشئ الاختبارات مما يسمح بمرونة أكبر في تخصيص الأسئلة، ويتميز هذا النوع من البنوك بقدرته على توسيع قاعدة الأسئلة بشكل مستمر من خلال إضافة أسئلة جديدة أو تعديل الأسئلة الحالية، ويستخدم هذا النظام في الولايات المتحدة الأمريكية، والهدف منه هو استخدام الأسئلة في عملية التقويم التكويني(البنائي)وفي التقويم التشخيصي.

٢- بنك الأسئلة المغلق: ففي هذا النوع تكون الأسئلة معدة مسبقًا من قبل جهة معينة (مثل المعلمين أو فرق تطوير المناهج) ولا يمكن تعديلها أو إضافتها من قبل المستخدمين الآخرين، ويتم تحديد الأسئلة مسبقًا في هذا النوع من البنوك وتخزينها في قاعدة بيانات مغلقة حيث تتميز بالسرية، حيث لا يستطيع المعلمون أو الممتحنون إضافة أسئلة جديدة أو تعديل الموجودة، مما يحد من مرونة هذا النوع مقارنة ببنك الأسئلة المفتوح، و يستخدم هذا النظام في إنجلترا وأستراليا ومصر حيث يكون الهدف منه هو استخدام الأسئلة الموضوعية المقننة والمخزونة بالبنك في التقويم النهائي فقط سواء كان في نهاية الفصل الدراسي أو نهاية العام .

التصنيف الثاني

١- تصنيف بنك الأسئلة حسب الاستخدام

يوضح (Choppin, 1976, p. 218) و(Nicol & Macfarlane-Dick, 2006, p. 204-206) وأنه تم تصنيف بنوك الأسئلة علي حسب الاستخدام إلي نوعين كالتالي:-

- **بنك الأسئلة التكويني: يستخدم لقياس التقدم التدريجي للطلاب خلال العملية التعليمية، حيث تساعد الأسئلة في هذا النوع من البنوك في تحديد نقاط القوة والضعف للطلاب أثناء دراستهم.**
- **بنك الأسئلة النهائي: يحتوي على أسئلة تستخدم في اختبارات التقييم النهائي لقياس مدى إلمام الطالب بالمحتوى الدراسي بعد إتمامه.**

٢- تصنيف بنك الأسئلة حسب الوظيفة

يشير(Wilson, 2005, p. 96-99) إلي أنه تم تقسيم بنك الأسئلة علي حسب الوظيفة إلي نوعين كالتالي:-

- **بنك الأسئلة التحليلي: يُستخدم لتقييم الفهم العميق والتحليل والتفكير النقدي للمفاهيم والمعلومات، حيث يحتوي علي أسئلة تتطلب تحليل محتوى معقد وتفسير نتائج معينة.**
- **بنك الأسئلة الوصفي: يُستخدم لتقييم المعرفة الأساسية والمباشرة حول موضوعات معينة، وعادةً ما يتضمن علي أسئلة بسيطة تُقيّم الحفظ أو الاستيعاب السطحي للمعلومات، فهي تهدف إلى قياس قدرة الطالب على تذكر واسترجاع المعلومات بطريقة وصفية، فالأسئلة تكون مباشرة وتستدعي تذكر المعلومات والحقائق والمفاهيم دون الحاجة إلى تحليل عميق.**

٣- بنك الأسئلة حسب تكنولوجيا المعلومات

يذكر (Baker & Kim, 2004, p. 25-35) و (Weiss, 2013, p. 10) و (Friyatmi et al., 2020, p. 294) أنه تم تصنيف بنك الأسئلة علي حسب تكنولوجيا المعلومات إلي نوعين كالتالي:-

- **بنك الأسئلة الإلكتروني: يتم تخزينه وإدارته باستخدام أنظمة تكنولوجيا المعلومات، حيث يمكن الوصول إليه إلكترونيًا وتقديمه في شكل اختبارات على الإنترنت، ويتم اختيار الاختبار بناء علي مستوى الصعوبة، ويشمل علي أسئلة مثل (الاختيار من متعدد – صح وخطأ- المزاوجة)، وفيه يتم التصحيح آلياً .**
- **بنك الأسئلة الورقي: لا يعتمد على التكنولوجيا يتم تحضيره بشكل يدوى ويتم تخزين الأسئلة في شكل ورقي ويتم استخدامها في الاختبارات التقليدية، ويحتوى علي أسئلة متنوعة ويستخدم بشكل منتشر في التعليم، وفيه يتم التصحيح يدوياً.**

٤- بنك الأسئلة حسب نوع السؤال

ذكر(Downing & Haladyna, 2006, p. 271) و (Cohen & Swerdlik, 2018, p. 242-243) أنه تم تصنيف بنوك الأسئلة حسب نوع السؤال إلي عدة أنواع كالتالي:

- **بنك الأسئلة الاختيارية: يحتوي على أسئلة اختيار من متعدد، ويقوم الطالب باختيار إجابة واحدة صحيحة من بين عدة اختيارات.**
- **بنك الأسئلة الصواب والخطأ: يتضمن أسئلة تتطلب من الطالب تحديد ما إذا كانت العبارة صحيحة أو خاطئة.**
- **بنك الأسئلة التوصيلية: يتطلب من الطلاب توصيل المعلومات أو الإجابة على أسئلة تتطلب الربط بين عناصر مختلفة.**
- **بنك الأسئلة المقالي: يقوم الطالب بتقديم إجابة مفصلة.**
- **بنك الأسئلة المقالي القصير: ويتطلب تقديم إجابة مختصرة حول موضوع معين.**
- **بنك الأسئلة ملء الفراغ: وفيه يقوم الطالب باكمال الجملة أو العبارة عن طريق اكمال الفراغ بالاجابة المطلوبة والناقصة.**

أهمية واستخدامات بنوك الأسئلة:

يري كل من (Hambelton & Swaminthan, 1985, p. 255-256) و (Hamad & Alnuzaili, 2020, p. 644) و (ماجد عبده ناجي وعبد المحسن بن رشيد المبدل، ٢٠٢٢، ص ٥٥٩ -٥٦٠) و (Friyatmi et al., 2020, p. 294)) أنه وحد أهمية كبيرة لبنوك الأسئلة وهى:-

١- بناء اختبارات لقياس الأهداف التربوية المنشودة بسهولة.

٢- إنتاج اختبارات بالعدد المطلوب من مفردات الاختبار لكل هدف.

٣- إذا كان بنك الأسئلة يتكون من مفردات صالحة للمحتوى وسليمة من الناحية الفنية، فعادةً ما تكون جودة الاختبار أفضل مما يمكن لمطوري الاختبار إنتاجه إذا قاموا بإعداد مفردات الاختبار بأنفسهم.

٤- تساعد المعلمين في بناء الاختبارات مرجعية المعيار.

٥- تقليل الوقت والجهد المستغرق في بناء الاختبارات.

٦- تحقيق الموضوعية في تقويم تحصيل الطلاب.

٧- تتيح إجراء عمل اختبارات متكافئة، وسحب صور متوازية من الاختبارات.

٨- توفر وقت المعلم في بناء الاختبارات التحصيلية وتكلفة تجميع الاختبارات.

٩- تقليل أثر القلق والتوتر والخوف المصاحبة للاختبارات التقليدية ومشكلات الأمن والسرية عند وضع الاختبارات.

١٠- تقليل الغش بين الطلاب إذا كان الاختبار الكترونياً.

١١- أمكانية التقويم المستمر خلال العام الدارسي بسهولة عن طريق الاستعانة بأسئلة البنك.

١٢- اتاحة الفرصة لقياس عمليات التفكير العليا وذلك لاتساع مدى المحتوى الذي تغطيه البنوك والمستويات المعرفية التي تشملها.

١٣- المرونة في القياس من خلال تشكيل اختبارات في أي وقت بسرعة وسهولة.

١٤- إمكانية التطوير المستمر للتعليم والتعلم، لتحقيق الأهداف التربوية المنشودة.

١٥- تعريف المعملين بطريقة تحليل المحتوى وصياغة الأهداف السلوكية واعداد جداول المواصفات.

١٦- تطبيق سياسة اللامركزية في برنامج الاختبارات الوطنية دون التأثير على نتائج الاختبارات.

١٧- يعزز أمان الاختبارات وذلك عن طريق زيادة عدد المفردات المخزنة في بنك الأسئلة.

خطوات بناء بنك الأسئلة

يشير كل من (Cohen & Swerdlik, 2018, p. 268- 269) أنه هناك مجموعة من الخطوات يجب اتباعها لبنتاء بنك أسئلة كالتالي:-

١- تحديد أهداف الاختبار والمعايير: حيث يتم تحديد الهدف والغرض من الاختبار بدقة، فيجب أن يكون هناك وضوح بشأن ما سيتم قياسه من خلال الاختبار مثل معرفة الطالب بالمحتوى أو قدراته التطبيقية أو مهاراته التحليلية وهذا يساعد في تحديد نوع الأسئلة التي يجب أن تكون في البنك.

٢- تصميم أنواع الأسئلة: يجب تصميم الأسئلة التي سيتم استخدامها، ويمكن أن تتنوع الأسئلة في بنك الأسئلة مثل الأسئلة متعددة الخيارات وأسئلة الصواب والخطأ والأسئلة المقالية وأسئلة التوصيل، ويجب أن تكون الأسئلة متنوعة لتغطية جميع جوانب المحتوى وتقييم مستويات مختلفة من الفهم.

٣- كتابة الأسئلة: يجب أن تكون الأسئلة واضحة ومفهومة وغير مبهمة، حيث يمكن استخدام معايير محددة لكتابة الأسئلة مثل التأكد من أن الإجابات لا تتداخل مع بعضها البعض (أي أن كل سؤال يكون مستقل عن باقي الأسئلة) وأن الأسئلة تكون متناسبة مع مستوى الطلاب المستهدف، ومن الأفضل أن تتم مراجعة الأسئلة من قبل متخصصين لضمان جودتها.

٤- مراجعة الأسئلة وتقييم جودتها: من الضروري مراجعة الأسئلة بعناية للتأكد من أنها تقيس الأهداف المحددة بشكل صحيح، ويجب أن يتم فحص الأسئلة من حيث الموضوعية والدقة اللغوية والوضوح، ويمكن أيضًا اختبار الأسئلة على مجموعة صغيرة من الطلاب للحصول على ملاحظات حول دقتها وفعاليتها.

٥- تصنيف الأسئلة: بعد كتابة الأسئلة يجب تصنيفها بناءً على موضوعاتها ومستوى الصعوبة، مما يساعد ذلك في تحديد كيف سيتم توزيع الأسئلة في الاختبارات المختلفة وضمان التوازن بين المواضيع.

٦- تخزين الأسئلة في قاعدة بيانات: يجب تخزين الأسئلة في قاعدة بيانات أو نظام إلكتروني وذلك من أجل الوصول إليها بسهولة واستخدامها عند الحاجة، ويمكن استخدام أدوات إدارة الاختبارات الإلكترونية لإنشاء قاعدة بيانات تتضمن الأسئلة مع إجاباتها الصحيحة ومعلومات إضافية.

٧- اختبار الأسئلة علي عينة تجريبية: من المفيد اختبار الأسئلة في بيئة تعليمية حقيقية علي مجموعة من الطلاب، حيث يمكن استخدام الأسئلة كجزء من اختبارات تجريبية لجمع البيانات حول كيفية أداء الطلاب عند الإجابة عليها.

٨- تحليل النتائج وتحديث الأسئلة: بعد إجراء الاختبارات يجب تحليل نتائج الطلاب لتحديد الأسئلة مستوى سهولة وصعوبة الأسئلة، وبناءً على هذه التحليلات يمكن تعديل أو استبعاد الأسئلة غير المناسبة أو التي لم تقدم المعلومات المطلوبة.

٩- استخدام الأسئلة في الاختبارات المستقبلية: بعد إجراء التعديلات اللازمة يمكن استخدام الأسئلة في الاختبارات المستقبلية، ومن المهم أن يتم تحديث بنك الأسئلة بشكل دوري لضمان أن الاختبارات تظل دقيقة وملائمة مع مرور الوقت.

بينما يوضح (توفيق علي سالم و سوسن حسن أبو الهادي، ٢٠١٩، ص ١٧٦) أنه هناك يوجد مجموعة من المراحل والخطوات والأنشطة الهامة التي يجب اتباعها عند بناء بنوك الاسئلة وهي ستة مراحل كالتالي:- مرحلة التخطيط والتنسيق، مرحلة التدريب والإعداد، مرحلة تصميم الأسئلة وبنائها، مرحلة التجريب والتعيير، مرحلة التصنيف والتخزين، مرحلة الاستخدام والتقييم للبنك بصورة عملية.

في حين يشير (وليد محمد فرج الله، ٢٠١٨، ص٤٦٣) إلي أنه يوجد ثلاث مراحل لبناء بنوك الأسئلة وهي كالتالي:-

١- إعداد مفردات بنك الأسئلة : وتعتبر من أهم مراحل إعداد بنوك الأسئلة وتستغرق وقتاً طويلاً، وتتطلب خبرة ومهارة عالية في تأليف وكتابة المفردات.

٢- التجريب الإستطلاعي للمفردات: وفيه يتطلب تجريب المفردات الاختبارية استطلاعياً وقتاً طويلاً، وذلك بسبب العدد الكبير الذى يحتويه بنك الأسئلة من مفردات، ولذلك يمكن تقسيم البنك لمجموعة من الاختبارات لا يتجاوز عدد مفرداتها ٥٠ مفردة حتي يتم تجنب عوامل الإرهاق والملل التي يمكن أن تؤثر سلبياً علي استجابات أفراد عينة البحث، وفي هذه المرحلة يتم حساب الخصائص السيكومترية لكل مفردة، وبناء علي ذلك يتم استبعاد وحذف المفردات غير المناسبة إحصائياً والإبقاء علي الصالح منها.

٣- التحقق من صدق المفردات وتدرجها: في هذه المرحلة يتم التحقق من معاملات الثبات والصدق والصعوبة والتميز والتخمين.

واستفاد الباحثان من الخطوات السابقة في بناء بنك أسئلة إلكتروني لمقرر الإتجاهات المعاصرة في علم النفس، وسوف يتم تناوله بالتفصيل في إجراءات البحث

الخطوات المهمة التي يجب اتباعها في تطوير بنك الأسئلة

يذكر (Haladyna, 2004, p. 14-18) أن هناك مجموعة من الخطوات المهمة التي يجب اتباعها عند تطوير بنك الأسئلة ، والتفاصيل المطلوبة لإنشاء بنك أسئلة يحتوي على أسئلة تشغيلية جاهزة للاستخدام في الاختبارات المستقبلية:

١- وضع خطة لتطوير الأسئلة: وجود خطة جيدة يعد أمرًا أساسيًا لإنشاء بنك أسئلة والمحافظة عليه، ويجب أن تحتوي الخطة على الجدول الزمني والموارد المطلوبة والأشخاص المسؤولين، ويعتبر تطوير الأسئلة أحد التكاليف الرئيسية لأي برنامج اختبار، وتتطلب وقتًا وجهدًا.

٢- إنشاء جدول زمني لتطوير الأسئلة: يجب أن يكون الجدول الزمني واقعيًا ويتضمن قائمة بالمهام والأشخاص المسؤولين عن تنفيذ كل مهمة، ففي بعض الأحيان قد تكون الجداول الزمنية غير واقعية، فتطوير بنك أسئلة قوي قد يستغرق أكثر من عام أو عامين اعتمادًا على الموارد المتاحة.

٣- إجراء جرد للأسئلة الموجودة في بنك الأسئلة: تُظهر مواصفات الاختبار (جدول المواصفات) للمطورين عدد الأسئلة التي يجب اختيارها للاختبار، أنواع المحتوى الذي يتم اختباره، وأنواع السلوكيات المعرفية المطلوبة من المتقدمين للاختبار عند الإجابة على كل سؤال، حيث يتم اختيار الأسئلة بناءً على هذه المواصفات وعوامل تقنية أخرى مثل مستوى صعوبة السؤال وقدرته التمييزية، و يجب توثيق مواصفات الاختبار بما في ذلك المنطق الذي تم تطويرها على أساسه والعملية التي تم اتباعها، يجب أن تحدد مواصفات الاختبار محتوى الاختبار، عدد الأسئلة المقترح، صيغ الأسئلة المطلوبة، الخصائص النفسية الإحصائية المرغوبة للأسئلة، وترتيب الأسئلة والأقسام، ويجب أيضًا

تحديد وقت الاختبار، الإرشادات الموجهة لمنشئي الاختبار، الإجراءات المستخدمة لإدارة الاختبار وتصحيحه، وأي معلومات أخرى ذات صلة.

٤- تحديد عدد الأسئلة المطلوبة في كل مجال من المجالات: من خلال معرفة عدد الأسئلة المطلوبة للاختبار والشروط الأخرى التي تؤثر على تصميم الاختبار، يمكن لمطورى الاختبار تحديد عدد الأسئلة التي يجب تطويرها، ويُفضل أن يحتوي البنك على ٢٥٠٪ من عدد الأسئلة المطلوبة لأي اختبار، ولكن هذا التقدير قد يختلف بناءً على الظروف، يُعد الجرد الطريقة الرئيسية التي نحدد بها الأسئلة اللازمة لضمان توفر إمدادات كافية من الأسئلة لتلبية الاحتياجات المستقبلية.

٥- تحديد وجذب خبراء مؤهلين في الموضوعات المختلفة لتطوير أسئلة جديدة: تعتمد جودة الأسئلة بشكل مباشر على مهارة وخبرة كتاب الأسئلة، ولا يمكن لأي قدر من التحرير أو المراجعات تحسين الأسئلة التي كتبت بشكل سيئ، ولهذا السبب يعد توظيف كتاب الأسئلة خطوة حاسمة في عملية تطوير الأسئلة، ويجب أن يكون كتاب الأسئلة خبراء و متخصصين في المجال الذي ستُسند إليهم الأسئلة فيه، فعليهم التحقق من صحة كل سؤال ومن وجود إجابة صحيحة واحدة فقط.

٦- إعداد دليل لكتابة الأسئلة: يجب إعداد دليل لكتابة الأسئلة وتقديمه لجميع كتاب الأسئلة، وينبغي أن يكون الدليل محددًا حول جميع الجوانب المهمة لكتابة الأسئلة، ويوضح لكتاب الأسئلة أي تنسيقات للأسئلة يجب استخدامها وأيها يجب تجنبها، كما يجب أن يحتوي على العديد من الأمثلة لأسئلة نموذجية، وينبغي تقديم إرشادات لكتابة الأسئلة.

٧- توزيع الدليل على كتاب الأسئلة.

٨- إجراء تدريب على كتابة الأسئلة لهؤلاء الكتاب: يجب أن يُشرك جميع كتاب الأسئلة في التدريب على كتابة الأسئلة.

٩- تخصيص المهام لكتاب الأسئلة بناءً على الجرد وتقييم الاحتياجات: يجب اختيار كل كاتب أسئلة بناءً على خبرته في موضوع معين، وبالنظر إلى المخزون يجب تكليف كل كاتب أسئلة بتطوير أسئلة قد تساهم في تحسين بنك الأسئلة وفي النهاية إدراجها في الاختبارات المستقبلية.

١٠- إجراء مراجعات: عند صياغة الأسئلة عادة ما تخضع لعدة مراجعات تكاملية، وتهدف إلى الاحتفاظ بالسؤال وتعديله أو استبعاده، ويتم إجراءها ا بواسطة موظفين مختلفين حسب طبيعة المراجعة، وتعتبر المراجعة من قبل الخبراء لتقييم جودة السؤال هي من أهم المراجعات.

١١- إجراء اختبار ميداني للأسئلة التي اجتازت المراجعات وتقييم أداء السؤال: يتم تجربة الأسئلة في اختبار تشغيلي، ومن المهم تقييم السؤال بالنسبة للأسئلة الأخرى في الاختبار، وبناء علي النتائج يمكن حذف أو إدراج السؤال في البنك لاستخدامه في الاختبارات المستقبلية.

١٢-وضع الأسئلة التي اجتازت التقييم في بنك الأسئلة التشغيل.

الطرق المختلفة لتخزين مفردات بنوك الأسئلة

يذكر (Bergstrom &Gershon, 1995, p. 188) و (توفيق علي سالم وسوسن حسن أبو الهادي، ٢٠١٩، ص ١٧٦-١٧٧) أن تخزين فقرات بنوك الأسئلة يعد جزءًا مهمًا من إدارة هذه البنوك. حيث يتطلب تخزين الأسئلة بشكل منظم وفعال لضمان سهولة الوصول إليها واستخدامها بشكل فعال، وهناك عدة طرق لتخزين فقرات الأسئلة:-

١- التخزين اليدوي (ورقيًا): وهى طريقة تقليدية لتخزين الأسئلة، وتشمل حفظ الأسئلة في ملفات أو سجلات مصنفة وفقًا للمحتوى أو الصعوبة، ورغم بساطة هذه الطريقة إلا أنها غير فعالة في التعامل مع كميات كبيرة من الأسئلة أو في الحفاظ على تحديثها بشكل دوري، كما أنه كلما زاد عدد البيانات التي تتم معالجتها يدويًا، زادت فرصة حدوث أخطاء.

٢- التخزين الآلي: وتتم باستخدام برامج حاسوبية مناسبة تؤدي العمليات المطلوبة بطريقة آلية، بحيث تتسم بالدقة والسرعة والسهولة والأمان من أجل الحصول علي أعلي جودة في العمليات والنتائج، وتلك البرامج يمكن تصميمها من قبل خبرات تقنية جيدة في ضوء رؤية خبراء في بناء بنوك الأسئلة وفق نظريات محددة، بحيث تحقق عملية الأداء في البرنامج المواصفات المطلوبة ، لذا فإن هذه العملية قد تكون مكلفة نوعا ما في مرحلة التصميم، ولكن هذه الكلفة ليست مستمرة عند الاستخدام كما في الطريقة التقليدية، ويمكن أيضا استخدام برامج جاهزة غير مجانية.

٣- طريقة الدمج: وتشمل استخدام الطريقتين الآلية والتقليدية لتنفيذ عمليات إدارة بنك الاسئلة، فمثلا قد يتم تخزين واستدعاء الأسئلة وانتقاء فقرات معينة بطريقة حاسوبية بينما يتم تنفيذ الاختبار بالطريقة التقليدية (ورقياً).

تنظيم بنك الأسئلة

يشير (Choppin, 1976, p. 218) إلي أنه يجب معرفة كيف يجب تنظيم بنك الأسئلة؟، وما البيانات التي ينبغي تخزينها مع الأسئلة؟ ، حيث أنه من المعتاد تسجيل بعض المعلومات حول ما يُفترض أن يقيسه كل سؤال حتى في بنوك الأسئلة غير المنظمة، ويمكننا هنا التمييز بين نوعين مختلفين تمامًا من بنوك الأسئلة:

١- النوع الأول :يحتوي على أسئلة تقيس جوانب مختلفة من الإنجاز، كل سؤال يركز على مهمة معيارية معينة، حيث يُظهر أداؤها إتقان هدف معين، وهذه الأسئلة تُستخدم وتُفسر بشكل فردي.

٢- النوع الثاني :يتضمن عددًا كبيرًا من الأسئلة التي تزعم قياس نفس البُعد، وغالبًا ما يتم تعريف هذا البُعد بشكل عام، مثل "الإنجاز في الهندسة" أو "معرفة مفردات اللغة الألمانية" أو "فهم المبادئ العلمية"، في هذه الحالات يُتوقع استخراج مجموعات من الأسئلة من البنك لتكوين اختبار مؤقت يهدف إلى قياس السمة المطلوبة بدقة نسبية.

ويجب أن يحتوي البنك على معلومات حول ما يُفترض أن يقيسه كل سؤال، لكن هذا ليس كافيًا بمفرده. إذا أردنا تفسير النتائج المستخلصة من سؤال معين، فيجب معرفة مدى صعوبة السؤال ومدى قدرته على التمييز بين الأفراد ذوي القدرات المختلفة.

مميزات وخصائص بنك الأسئلة

يري كل من(Choppin, 1976, p. 217) و (Chuesathuchon & Waugh, 2008, p. 106-107) و (Downing & Haladyna, 2006, p. 22) و (Cohen & Swerdlik, 2018, p. 243) و(وليد محمد فرج الله، ٢٠١٨، ص٤٦٤) و(توفيق علي سالم و سوسن حسن أبو الهادي، ٢٠١٩، ص ١٧٣- ١٧٤) و(ماجد عبده ناجي وعبدالمحسن بن رشيد المبدل، ٢٠٢٢، ص ٥٦٠- ٥٦١) و (Bahrom etal., 2025, p. 11), أن هناك مجموعة من المميزات والخصائص والفوائد الاختبارية لبنك الأسئلة و ويمكن تجميعها كالتالي:-

١- المرونة والتنوع: بنك الأسئلة يوفر تنوعًا كبيرًا في أنواع الأسئلة مثل الأسئلة متعددة الخيارات والأسئلة المقالية والصواب والخطأ، وهذا يسمح بتصميم اختبارات تتناسب مع الأهداف التعليمية المختلفة ومتطلبات الطلاب المتنوعة، كما تتمثل المرونة في إضافة أو حذف أو تعديل لفقرات البنك بشكل مستمر، وتعد هذه العممية ضرورية للمحافظة علي دقة تمثيل فقرات البنك للموضوع المستهدف للقياس.

٢- الموضوعية: وهذا هو الهدف الأساسي لبنوك الأسئلة، ويعتمد تحقيق هذا الهدف على مجموعة من العوامل التي يمكن أن تؤثر على بناء وتكوين بنك الأسئلة مثل فلسفة القياس ومدى الدقة في صياغة الأسئلة وتحليلها وتحديد مواصفاتها الإحصائية.

٣- تحقيق العدالة والمساواة: من خلال إعداد اختبارات متكافئة بدقة وموضوعية، وهذا يساعد في خلق اختبارات متوازنة وتقدم فرصة عادلة لجميع الطلاب بغض النظر عن مستوى فهمهم.

٤- إمكانية التخصيص: بنوك الأسئلة تساعد في توفير أسئلة مناسبة في مادة دراسية معينة، وبالتالي تستفيد منه المؤسسة التعليمية وتعده نواه لبنوك أسئلة أكثر شمولا في المادة المحددة والمواد الدراسية الأخرى، وبالتالي سرعة تجهز الأسئلة في أي وقت.

٥- تحليل أداء الطلاب: بنك الأسئلة يهتم بالتعرف على نتائج التعلم لكل طالب على حدة وليس التعرف على الدرجة الكلية لنتائج الاختبار، مما يوفر إمكانيات لتحليل أداء الطلاب بشكل أكثر دقة، وذلك عن طريق تحليل كل سؤال ومعرفة مدى فعاليته في قياس المعرفة أو المهارات المطلوبة، مما يساعد في تحسين جودة الأسئلة وضمان استخدام الأسئلة الجيدة أكثر من مرة، وبالتالي ضمان مستوى جيد من الاختبارات بصفة مستمرة.

٦- الإقتصاد توفير الوقت والجهد : استخدام بنك الأسئلة يساعد في إعداد الاختبارات بسرعة وبتكلفة منخفضة حيث يتم استخدام الأسئلة المخزنة مسبقًا، وهذا يساعد في توفير وقت المعلمين والممتحنين ويزيد من كفاءة إدارة الاختبارات، كما أنه ممكن أن تظل فاعلية البنك لسنوات وهذا يوفر كثيراً من الوقت والجهد والتكلفة.

٧- الإتساق: ويشير إلى استخدام نظام تدريج موحد لجميع فقرات البنك يجعل النظام البنكي على درجة عالية من الاتساق والتماسك، ويسمح بتكوين اختبارات عدة، يصعب الحصول عليه في قوائم الأسئلة أو مجموعات الاختبارات المقننة.

٨- التحديث المستمر: يمكن تحديث بنك الأسئلة بشكل دوري لضمان أسئلة جديدة أو تعديل الأسئلة الحالية بناءً على ملاحظات الطلاب أو التغييرات في المناهج، وهذا يضمن أن الأسئلة تظل ملائمة ومتجددة، ويتم ذلك عن طريق تدريب المعلمين على عمل أسئلة جديدة مشابهة لأسئلة بنك الأسئلة.

٩- التعلم الذاتي: حيث أنه يساعد الطلاب علي التعلم الذاتي وفق سرعته وإمكانياته.

١٠- تقويم جميع نتائج التعلم المنشودة من تدريس محتوى المقرر.

١١- تحقيق أهداف القياس المنشودة من كل مقرر: بنوك الأسئلة تساعد المعلمين في اختيار مفردات الاختبار الجيدة التي تلبي أهداف القياس والمحتوى من بنك الأسئلة لتناسب قدرات طلابهم في كل مجال من مجالات الاختبار.

١٢- إنشاء اختبارات تكيفية: إن بنوك الأسئلة تسمح بإنشاء اختبار يتكيف مع أي مجموعة من الطلاب الذين لديهم قدرات تعلم مختلفة وللطلاب ذوي الإعاقات.

١٣- ضمان أمان وسرية الاختبار: في بنوك الأسئلة يتم ضمان الأمان، لأن هناك الكثير من المفردات في البنك من غير المرجح أن يتذكر الطلاب الذين يؤدون الاختبار جميع المفردات من اختبار واحد أو أكثر، وبالتالي يمكن أن تحمي بنوك الأسئلة تسرب المفردات على الأقل إلى حد كبير.

١٤- تعزيز فعالية القياسات: يعزز بنك الأسئلة المتطور جيدًا من فعالية القياسات لأنه يمكن تحسين مفردات الاختبار من حيث الثبات والصدق لتلبية المعايير التعليمية العليا، وبالتالي يضمن هذا دقة وثبات القياس.

عيوب بنوك الأسئلة

يشير كل من (Hiscox & Brzezinski, 1980, p. 5-6) و (وليد محمد فرج الله، ٢٠١٨، ص٤٦٤) إلي أنه علي الرغم من أهمية ومميزات بنوك الأسئلة إلا أنه يتخللها بعض العيوب كالتالي:-

١- تكرار استخدام الأسئلة: استخدام نفس الأسئلة أكثر من مرة في اختبارات متعددة مما يؤدي إلى معرفة الأسئلة وإجاباتها بين الطلاب مما يقلل من مصداقية نتائج الاختبارات ويؤثر على نزاهتها ويقل من سرية الأسئلة المسحوبة.

٢- التحيز في الأسئلة: بنك الأسئلة قد يتضمن أسئلة غير محايدة أو متحيزة نحو ثقافة معينة، فالأسئلة المتحيزة قد تؤثر على نتائج الطلاب الذين لا ينتمون إلى نفس الخلفية الثقافية أو التعليمية مما يؤدي إلى عدم العدالة في التقييم.

٣- محدودية التقييم الشامل: قد لا يمكن لبنك الأسئلة أن يغطي جميع جوانب المحتوى أو المهارات المطلوبة بشكل كامل، رغم تنوع الأسئلة إلا أن بعض الموضوعات قد لا تُغطى بشكل كافٍ مما يؤدي إلى تقييم غير كامل للمعرفة.

٤- إجهاد واضعي الإمتحان في التحديث المستمر: يواجه واضعي الإمتحان صعوبة في تحديث بنك الأسئلة بشكل دوري لضمان جودتها وملاءمتها، حيث أنه إذا لم يتم تحديث الأسئلة بانتظام قد تصبح قديمة أو غير ملائمة للمحتوى الحالي.

٥- ارتكاب أخطاء في صياغة الأسئلة: إنشاء أسئلة عالية الجودة يتطلب وقتًا وجهدًا كبيرين، وقد يؤدي إنشاء الأسئلة إلى ارتكاب أخطاء في الصياغة أو في الإجابات الصحيحة مما يقلل من دقة الاختبارات.

٦- الاعتماد على نوع واحد من الأسئلة: في بعض الأحيان قد يحتوي بنك الأسئلة على نوع واحد من الأسئلة (مثل أسئلة الاختيارات المتعددة) مما يؤدي إلى عدم تنوع التقييمات، وهذا يحد من القدرة على قياس مجموعة واسعة من المهارات المعرفية أو التحليلية.

ثالثاً: المعايرة المتزامنة(Concurrent calibration)

يعرفها (Lee, 2011, p. 12) بأنها "العملية التي يتم فيها تقدير المعلمات لجميع المجموعات في نفس الوقت باستخدام برنامج واحد، ويتم اعتبار المفردات غير المشتركة كعناصر غير مُعتمدة، ويتم استخدام برامج مثل BILOG-MG و ICL و XCALIBRE 4.0 لتنفيذها".

ويعرفها (Kim, 2019, p. 516) بأنها "تقدير معلمات مفردات لعدة نماذج اختبار مقدمة لمجموعات غير مكافئة".

وفي البحث الحالي يقصد بها الباحثان تدريج جميع فقرات البنك التي أجاب عنها جميع الطلاب مرة واحدة باستخدام النموذج الذي يحقق أفضل ملائمة.

شروط(متطلبات) المعايرة

يشير(Shen et al., 2024, p. 554) إلي أنه لتحقيق معايرة دقيقة من الضروري وجود عدد كبير من الممتحنين خلال مرحلة الاختبار الأولي ، كما أن معايرة بنك الأسئلة تتطلب اختبارًا تمهيديًا يُطبق جميع المفردات المحتملة على المتقدمين المختارين باستخدام تصميم اختبار غير مكتمل يُخصص مفردات مختلفة لأفراد مختلفين، وبالتالي يؤدي هذا بطبيعة الحال إلى مجموعة بيانات استجابة غير مكتملة، حيث تقع نسب الاستجابة المفقودة ضمن نطاق ٤٠٪-٧٠٪ ، وللتغلب على تحدي حجم العينة الصغير، تم اقترح استراتيجيات معايرة مستمرة تسمح بمعايرة بنك الأسئلة دون معلومات أولية عن معاملات المفردات بالتزامن مع بدء التشغيل الفعلي للبنك، وتعتمد فعالية هذه الاستراتيجيات على التصميم الدقيق وربط دورات الاختبار المتعددة، وإلا فقد تُسبب مشاكل في تقدير القدرة.

أنواع المعايرة

يوضح (Lee &Lee, 2018, pp. 641- 643) أنه يوجد ثلاثة أنواع من المعايرة وهي كالتالي:

١- المعايرة المنفصلة (Separate calibration): في المعايرة المنفصلة، تتم معايرة لنموذجين النموذج القديم (Form O) والنموذج الجديد (Form N) بشكل منفصل، كلٌ منهما يعتمد على المقياس القياسي، وبما أن المجموعتين اللتين تجريان الاختبارين تنتميان لمجتمعين مختلفين، فلن تكون تقديرات معلمات البنود للنموذجين على نفس المقياس، وتقدم نتائج الربط للمعايرة

المنفصلة معلومات تشخيصية مفيدة عند مقارنة المجموعتين من تقديرات معلمات المفردات المشتركة للتحقق مما إذا كانت المفردات المشتركة تتصرف بشكل متماثل في إدارتين مختلفتين.

٢- المعايرة المتزامنة (Concurrent calibration): تتضمن المعايرة المتزامنة تنفيذ عملية تقدير واحدة فقط لنظرية استجابة المفردة (IRT) من خلال دمج بيانات النموذج القديم (Form O) والنموذج الجديد (Form N)، بعد ذلك، تُجرى المعايرة بحيث يُحدد أحد المجموعات (مثل مجموعة النموذج القديم) كمجموعة مرجعية تُستخدم كمقياس أساسي، حيث يتم تثبيت المتوسط والانحراف المعياري عند ٠ و١، ويمكن أن يختلف المتوسط والانحراف المعياري لمجموعة النموذج الجديد عن المقياس الأساسي ويتم تقديرهما بحرية نسبةً إلى المقياس الأساسي القياسي، وتنتج المعايرة المتزامنة مجموعة واحدة فقط من تقديرات المعلمات للمفردات المشتركة، وبعد تنفيذ هذه المعايرة المتزامنة لمرة واحدة، ستكون جميع تقديرات معلمات المفردات في النموذجين على نفس المقياس الأساسي دون الحاجة إلى عملية تحويل مقياس إضافية.

٣- المعايرة الثابتة (Fixed calibration): تتضمن طريقة المعايرة الثابتة خطوتين: أولًا، تتم معايرة بيانات النموذج القديم (Form O) للحصول على تقديرات معلمات المفردات في النموذج القديم، بما في ذلك المفردات المشتركة، وبالاحتفاظ بالمفردة المشتركة ثابتةً عند القيم التي تم الحصول عليها من المعايرة السابقة باستخدام بيانات النموذج القديم، تتم معايرة بيانات النموذج الجديد (Form N) ، وبما أن تقديرات معلمات المفردات المشتركة موجودة بالفعل على المقياس الأساسي للنموذج القديم، فسيتم وضع تقديرات معلمات المفردات غير المشتركة في النموذج الجديد على نفس المقياس الأساسي للنموذج القديم بعد المعايرة، وتُعتبر المعايرة الثابتة مفيدة بشكل خاص عندما يكون من الضروري وضع تقديرات معلمات المفردات للمفردات المجربة (Pretest Items) على نفس المقياس الذي وُضعت عليه البنود التشغيلية (Operational Items)التي سبق معايرتها ووُضعت على مقياس محدد مسبقًا ضمن بنك الأسئلة، وفي هذه الحالة، تُثبَّت تقديرات معلمات المفردات التشغيلية عند تقدير معلمات المفردات المجربة بحيث توضع تلقائيًا على مقياس بنك الأسئلة.

وبالتالي يري الباحثان أن الفرق بين الثلاث أنواع من المعايرات هي أن المعايرة المتزامنة توفّر تقديرات أكثر استقرارًا ودقةً إذا توافرت بيانات النموذجين معًا وكان الافتراض الأحادي البُعد مناسبًا، وأن المعايرة المنفصلة تُستخدم عندما تكون بيانات النموذج القديم غير متاحة بالكامل وقت معايرة النموذج الجديد، وهي مفيدة لتحليل تشخيصي لسلوك المفردات المشتركة، وأن

المعايرة الثابتة هي الأنسب عندما يكون الهدف هو وضع بنود جديدة أو تجريبية على نفس المقياس المرجعي لبنك بنود قائم دون تعديل المعلمات السابقة للمفردات التشغيلية، وبناء علي ما سبق سوف يستخدم الباحثين المعايرة المتزامنة وذلك بسبب دقة نتائجها وخاصةً عند استخدام النموذج المناسب.

طرق المعايرة المتزامنة

يشير(Kim & Cohen, 1998, p. 132) إلي أنه يمكن إجراء المعايرة المتزامنة باستخدام طريقتين تقدير مختلفتين:

١- طريقة التقدير الأقصى المشترك (JMLE) : تُستخدم فيها مجموعة من البرامج مثل (LOGISTو MICROCALE)، وتتضمن المعايرة المتزامنة في الشكل الأبسط تقدير المفردات ومعلمات القدرة معًا في مجموعة واحدة، دون تقسيمها إلى مجموعات فرعية، ودون تثبيت معلمات المفردات أو الأشخاص لمجموعة معينة أو تركها مفقودة يمكن تنفيذ اختلاف لهذا أيضًا، حيث يتم تثبيت معلمات المفردات المشتركة بين مجموعتين من خلال مجموعة ربط، ويتم معايرة المفردات المتبقية في مجموعات الاختبار الخاصة بها، وتشير نتائج الدراسات التي أجريت في سياق JMLE إلى أن المعايرة المتزامنة مع عينات كبيرة (مثلاً، ١٠٠٠ لنموذج المعلمة الثلاثية) توفر نتائج أكثر استقرارًا، من حيث ارتفاع الارتباطات بين التقديرات المرتبطة وتقديرات العينات الكبيرة، والتحيز الأصغر، وفرق الجذر التربيعي لمتوسط المربعات(RMSDs) مع المعلمات التي تم توليدها، وبالمثل، عند استخدام أعداد أقل من المفردات المشتركة (على سبيل المثال، أقل من ١٠) في اختبار الربط فإن المعايرة المتزامنة توفر نتائج معادلة أكثر استقرارًا، ويُعد استخدام مجموعات صغيرة نسبيًا من المفردات المشتركة في اختبار الربط مسألة مهمة، خاصة بالنسبة للاختبارات القصيرة.

٢-طريقة التقدير الأقصى المرجح باستخدام الطريقة البايزية (MMLE): وفي سياق MMLE يمكن تحقيق معادلة المفردات المشترك عبر اختبارات متعددة باستخدام برنامجي BILOG و MULTILOG، وفي حالة المعايرة المتزامنة لتقديرات معلمات المفردات باستخدام BILOG، يفترض التهميش أن توزيع قدرة المفحوصين هو إما توزيع غاوسي (Gaussian) أو توزيع عشوائي يتم الحصول عليه بشكل مشترك مع تقديرات معلمات المفردات، وقد ثبت أن التحديد المناسب لتوزيع المفحوصين هو أمر حاسم بالنسبة لدقة التقدير، وعندما تكون هناك مجموعتان من الممتحنين فإن الخيارات الافتراضية في MULTILOG تفترض توزيعًا طبيعيًا موحدًا لمجموعة الأساس، ويتم الحصول على متوسط مجموعة الهدف تجريبيًا مع معلمات المفردات، ويتم تحديد الانحراف المعياري (SD) كوحدة واحدة ، ويتمثل الإختلاف الأساسي بين BILOG و MULTILOG في ظل

المعايرة المتزامنة هو أن MULTILOG يسمح بتحديد معلمات قدرات المفحوصين المختلفة، وهذا مفيد إذا كانت هناك مجموعتان أو أكثر من الممتحنين، وإذا كان توزيع المجموعة المستهدفة مخلفاً عن توزيع المجموعة فإن التهميش في دالة الاحتمال تحت افتراض وجود توزيعين مختلفين سيكون أكثر ملاءمة، ويعتبر شكل توزيع قدرات المفحوصين إحدى القضايا غير المحسومة في المعايرة المتزامنة باستخدام MMLE.

أساليب للتعامل مع البيانات المفقودة في معايرة المفردات

يوضح (Shen et al., 2024, p. 554) أن العديد من الدراسات طورت أساليب للتعامل مع البيانات المفقودة أثناء معايرة معلمات المفردات، وتشمل أساليب لا تعتمد على الإدخال وأخرى تعتمد على الإدخال، ومع ذلك، قد تُسبب طرق الإدخال هذه أخطاء في تحديد النموذج وانحيازًا كبيرًا، وخاصةً عندما تكون البيانات متفرقة جدًا وحجم العينة محدودًا.

ويذكر (Schminkey et al., 2016, pp. 287- 288) أنه يوجد مجموعة من الطرق الشائعة للتعامل مع البيانات المفقودة وهي كالتالي:

١- حذف البيانات(Deletion): يمكن أن يكون الحذف إما على مستوى القوائم (Listwise) أو أزواج المتغيرات (Pairwise)

أ- الحذف القائم على القوائم: وفيه ببساطة يستبعد جميع الحالات التي تحتوي على قيم مفقودة، مما يقلل حجم العينة وبالتالي القوة الإحصائية، وكما أنه يُحدث تحيزاً كبيرًا، إذ قد يكون المشاركون ذوو البيانات الكاملة مختلفين تمامًا عن المشاركين ذوي البيانات الناقصة، ويعتبر الحذف القائم على القوائم إشكاليًا بشكل خاص إذا كان لدى العديد من المشاركين بيانات مفقودة في خلايا قليلة، كما أن هذا الأسلوب غير مجدٍ إطلاقًا إذا كان لدى كل حالة بعض البيانات المفقودة.

ب- الحذف القائم على أزواج المتغيرات: فيه يُتخذ قرار الاحتفاظ بالحالة أو استبعادها من التحليل بناءً على وجود البيانات لأزواج متغيرات معينة، فإذا كانت إحدى المتغيرات في التحليل الارتباطي مفقودة لدى المشارك، يُستبعد من المقارنة الزوجية المعنية ولكنه قد يظل مشمولًا في مقارنات أخرى تتوفر فيها القيم لكلا المتغيرين، ولا يمكن استخدام الحذف الزوجي عند تقدير الخطأ المعياري (SE) لأن حجم العينة لا يكون ثابتًا عبر جميع المتغيرات، ولهذا السبب، فإن أي محاولة لتقدير الخطأ المعياري ستكون مبالغًا فيها لبعض المتغيرات ومُقللة لغيرها، مما قد يُدخل تحيزاً شديدًا في تقديرات المعلمات للإجراءات الوصفية لكامل المجتمع.

٢- الترجيح (Weighting): وتستخدم هذه الطريقة غالبًا لتعويض احتمالات الاختيار غير المتساوية للعينة، كما يُستخدم عندما يكون القلق الأساسي بشأن البيانات المفقودة هو التحيز

(الانحياز) الذي تُدخله، فقد يحدث ذلك إذا وُجدت فروق كبيرة بين الفئات الفرعية في معلمة معينة، ويمكن تطبيق انحدار لوجستي مرجّح على المشاركين ذوي القيم المفقودة للحصول على تقدير معلمة أكثر تمثيلًا مما لو حُذفت الحالات ذات القيم المفقودة، ويمكن أن يزيل الترجيح بعض تحيز عدم الاستجابة وقد يمكّن من التوصل إلى استنتاجات أكثر تمثيلًا، لكن هناك مخاطرة بحدوث نتائج غير متسقة لأن الترجيح يتحدد بواسطة متغيرات أخرى، وليس بواسطة المعلمة موضع الاهتمام، ويمكن أن تكون تقنيات مثل الترجيح ضمن طرق التقدير باستخدام المعادلات التقديرية المعممة(GEE) –سواء باستخدام طرق خاصة بالملاحظة أو خاصة بالمشارك– مفيدة عند التعامل مع البيانات المفقودة بسبب المشاركين المفقودين في المتابعة أو المستبعدين من العينة أثناء الدراسة، ومع ذلك، فإن هذه التقنيات لا تكون مجدية إذا لم تكن البيانات مفقودة في نهاية الدراسة فقط، بل بشكل متقطع طوال فترة الدراسة.

٣- الإكمال (Imputation): هو عملية استبدال القيم المفقودة بتقديرات لها، ويمكن أن يكون الإكمال بسيطًا أو معقدًا:

أ- <u>الإكمال البسيط:</u> وفيه يتم إكمال الوسط حيث يتم استبدال القيمة المفقودة بمتوسط المتغير لجميع المشاركين، ومع ذلك فإن هذا الأسلوب يقلل من التباين ويزيد من خطر تقدير المعلمات بشكل غير دقيق، ويمكن أيضًا أن يؤدي إلى تقليل الخطأ المعياري (SE) وإعطاء انطباع مُزيف بدقة التقديرات.

ب-<u>الإكمال المعقد:</u> مثل الإكمال المتعدد(MI) ، ففي الإكمال المتعدد يتم تقدير قيم مفقودة متعددة لنفس الخلية المفقودة، مما يولّد عدة نسخ مكتملة من مجموعة البيانات، ويتم تحليل هذه النسخ على حدة، ثم تُدمج النتائج لاستخلاص تقديرات نهائية تأخذ في الحسبان عدم اليقين المرتبط بالقيم المفقودة، ويشتهر الإكمال المتعدد بدقته النسبية ومرونته مع البيانات MCARو MAR ، ومع ذلك، فهو يتطلب افتراضات قوية حول نماذج البيانات ويحتاج إلى معرفة وخبرة إحصائية كافية من الباحث.

استراتيجية معايرة لبنوك الأسئلة

يري (Shen et al., 2024, pp. 554-555) أنه يوجد استراتيجية لمعايرة لبنوك الأسئلة مُصممة خصيصًا للتقييمات صغيرة النطاق، وتعتمد هذه الاستراتيجية على خطوتين:

١-تصميمات معايرة غير مكتملة (incomplete calibration designs): وتساعد في جمع البيانات بكفاءة خلال مرحلة الاختبار الأولي لبنك الأسئلة. بعد ذلك، نستخدم طريقة إكمال المصفوفة

أحادية البت، وهي ممارسة تُخصص مفردات مختلفة لمختلف المتقدمين للاختبار، ويساعد استخدام تصميم غير مكتمل تقليل تأثيرات الإرهاق على المتقدمين للاختبار وضمان صلاحية بيانات الاستجابة، ويتكون تصميم الاختبار المبدئي من مكونين أساسيين:

أ- تصميم متعدد الأشكال: الذي يضمن ربط معلمات المفردات بالقياس نفسه، مثل تصميم المجموعة غير المتكافئة مع الاختبار الثابت (NEAT)

ب- تصميم الاختبار المستهدف : وهى استراتيجية تخصيص أشكال الاختبار للمتقدمين، حيث يتم استخدام المعرفة السابقة عن المفردات والمتقدمين لتكوين أشكال الاختبار وتخصيصها للمتقدمين، وتتعلق المعرفة السابقة عن المتقدمين بمستويات قدرتهم التي يتم تحديدها غالبًا بواسطة عوامل مثل درجاتهم الأكاديمية أو أعمارهم، وبالمثل تتعلق المعرفة السابقة بالمفردات عادةً بصعوبتها، والتي تُستند غالبًا إلى تقييمات الخبراء في المحتوى أو المعلمين، ومن خلال تخصيص مجموعات معينة من المتقدمين لمجموعات محددة من المفردات يمكن لتصاميم الاختبار المستهدف تحسين كفاءة معايرة بنك العناصر، وهذه التصاميم تسهل تحقيق تقديرات دقيقة لمعلمات المفردات حتى مع أحجام عينات أصغ، كما أنه في هذا النوع من التصميم عادة ما تُترتب مجموعات المفردات بترتيب تصاعدي للصعوبة، وتُخصص أشكال الاختبار للمتقدمين بناءً على مستويات قدرتهم، من الأدنى إلى الأعلى.

٢-طريقة إكمال المصفوفة أحادية البت (1-bit matrix completion method): وفيها يتم إدخال الاستجابات المفقودة الثنائية الناتجة عن هذا التصميم غير المكتمل، وتُستخدم المصفوفة المكتملة الناتجة لتقدير معاملات المفردات، فمصطلح "إكمال المصفوفة ذات البت الواحد" مشتق من مجال نظريات الاستشعار المضغوط ذو البت الواحد، وتصبح مشكلة إتمام المصفوفة أكثر تعقيدًا عندما تكون المصفوفة الملاحظة ثنائية.

ثالثاً: دالة المعلومات (information function)

يذكر (Al-zboon et al., 2021, p. 891) أن دالة المعلومات تُعد أحد المفاهيم الأساسية التي تلعب دوراً مهماً في نظرية استجابة المفردة (IRT)، إذ تُحدد دالة المعلومات مقدار المعلومات الذي توفره المفردة أو الاختبار ككل عند تقدير قدرة الأفراد أو المفحوصين، ومن خلالها يمكن تحديد الخطأ المعياري في التقدير، وتكتسب دالة المعلومات التي توفرها المفردة أو الاختبار أهمية كبيرة في إطار نظرية استجابة الفقرة، إذ يمكن اختيار المفردات بناءً على دالة معلومات المفردة، ونظراً لأن معلومات المفردة تختلف حسب مستويات القدرة المختلفة، مما يتيح إمكانية اختيار المفردات التي توفر دقة قياس عالية عند نقطة معينة على متصل السمة (القدرة) من جهة ومن جهة أخرى فإن المفردات

التي تمتاز بدرجة تمييز عالية تساهم في تقديم معلومات أكثر عن قدرة المفحوص، مما يُفضي إلى تحقيق دقة أعلى في التقدير.

أنواع دالة المعلومات:

يشير (ALKursheh et al., 2022, p. 1116) إلي أن هناك نوعين من دوال المعلومات وهي كالتالي:

١- دالة المعلومات الاختبار (Test information function)(TIF) : وتُعد دالة معلومات الاختبار (TIF) من الأفكار الأساسية التي تلعب دورًا مهمًا في نظرية استجابة الفقرة (IRT) حيث تُحدد TIF كمية المعلومات التي يقدمها الاختبار أو المفردة ككل عند تقدير قدرة الأشخاص أو المفحوصين، والتي يمكن من خلالها تحديد الخطأ المعياري في التقدير، ونظرًا لأنه يتم استخدام جميع مفردات الاختبار في تقدير قدرة المفحوص، فإن قيمة المعلومات التي يتم الحصول عليها من خلال الاختبار يمكن تحديدها لكل مستوى من مستويات القدرة، ويُعد الاختبار مجموعة من المفردات؛ وبالتالي يمكن الحصول على دالة معلومات الاختبار لمستوى قدرة معين من خلال جمع دوال معلومات جميع المفردات، ويمكن التعبير عن قيمة دالة معلومات الاختبار المحسوبة عند كل مستوى من مستويات القدرة (Θ) من خلال الآتي:

$$\mathbf{I(\theta)= = \sum_{i=1}^{N} Ii(\theta)}$$

I(θ) تشير إلي دالة معلومات الاختبار، Ii(θ) تشير إلي دالة معلومات المفردة، N تشير إلي عدد مفردات الاختبار

وبناء علي المعادلة السابقة إن دالة معلومات الاختبار ككل تكون أكبر بكثير من دالة معلومات المفردات الفردية، وبناءً عليه، فإن الاختبار الكامل يقيس قدرة الشخص بدقة وشمولية أكثر من كل فقرة على حدة، والنقطة المهمة هي أن وجود عدد أكبر من المفردات في الاختبار يزيد من دالة معلومات الاختبار، و لذلك يمكننا عمومًا القول إن الاختبارات الأطول تقيس قدرة المفحوص بدقة أكبر من الاختبارات الأقصر.

ويعرفها (Al-zboon & Maan, 2022, p. 55) بأنها مجموع دوال معلومات المفردات عند مستوى معين من القدرة، إذ إنها لا تعتمد على عينة المفحوصين، في حين يري (Al-zboon et al., 2021, p. 889) بأنها مجموع دوال العناصر التي تشكل الاختبار.

ويضيف (Hambleton & Swaminathan, 2013, p. 104) أنه لفهم خصائص دقة التقييم، يجب دمج معلومات المفردات لصياغة تقييم متكامل، وبما أن المفردات مستقلة محليًا، يمكن جمع المعلومات لكل مفردة عند كل مستوى قدرة θ وفق المعادلة السابقة، يساعد ذلك في اختيار الفقرات بما يتوافق مع طبيعة الفئة المستهدفة، مثل استخدام فقرات سهلة لتحديد الطلاب ذي القدرات المنخفضة.

٢- دالة معلومات الفقرة IIF (Item Information Function): تُمثل مدى دقة المفردة في قياس القدرة (θ)، وكلما زادت المعلومات التي توفرها المفردة عند مستوى معين من القدرة، كانت الفقرة أكثر فاعلية في التمييز بين المفحوصين في هذا المستوى، ويجب التنويه إلى أن المعنى الرياضي لقيمة معلومات المفردة يعتمد على المنحنى المحدد المستخدم.

ويضيف (Al-zboon & Maan, 2022, p. 55) أنها تُعد من أهم الإحصاءات في نظرية القياس الحديثة، حيث يمكن من خلالها تحديد الخطأ المعياري في التقدير بالاعتماد على الاحتمال الأقصى لمعلمة القدرة، ويكون التباين في تقدير القدرة مساوياً لعكس دالة المعلومات، كما أن مصفوفة التباين والتغاير للتقديرات تساوي معكوس مصفوفة معلومات التقديرات الخاصة بمعلمات المفردة

دالة معلومات المفردة في ضوء نماذج نظرية الاستجابة للمفردة

يذكر (Tan, 2024, p. 34) أنه في سياق نظرية الاستجابة للمفردة IRT تشير الدقة إلى "معلومات" المفردة أوالاختبار؛ فالأولى تتعلق بمستوى دقة المفردة، والثانية تتعلق بدقة التقييم، وتُعرف المعلومات في IRT بأنها المعكوس للخطأ المعياري للتقدير، أي التغير حول القيمة المقدَّرة للمعلمة، كما هو موضح في المعادلة التالية:

$$SE(\theta)= \frac{1}{\sqrt{I(\theta)}}$$

حيث I(θ) هي دالة معلومات الفقرة المعطاة القدرة θ، و SEهو الخطأ المعياري.

تشير قيمة SE إلى مستوى عدم التأكد في التقدير، وهو ما يقيس مباشرةً معلومات الفقرة عند قيمة θفكلما زادت المعلومات عن فقرة، زادت دقتها والعكس صحيح.

ويري (Kalkan, 2022, p. 74) و (Zheng et al., 2021, p. 3) و (Tan, 2024, p. 33) إلي أنه قد أصبح فحص العلاقة بين معلمات نماذج IRT الأربعة والمعلومات أمرًا مركزيًا، وقد أدى إدراج معلمة "اللامبالاة (inattention) " أو الإهمال في التطبيقات العملية، إلى توجيه مفهوم

المعلومات نحو سياق أكثر ارتباطًا بالواقع، مما جعله أكثر قيمة ودقة في تحديد دقة التقييم، وفيما يلي دالة المعلومات للمفردة في نماذج نظرية الاستجابة للمفردة

نماذج IRT	دالة المعلومات للمفردة $I_i(\theta)$
1PL	$D^2 p_i(\theta) q_i(\theta)$
2PL	$D^2 a^2 p_i(\theta) q_i(\theta)$

$p_i(\theta)$ تشير إلى احتمال الإجابة الصحيحة،$q_i(\theta)$ تشير إلى احتمال الإجابة الخطأ، D ثابت تحويل (عادةً = ١.٧)، a: معامل التمييز.

ويذكر (Tan, 2024, p. 33-34) أن الجمع بين النماذج الأربعة في التقييم (1PL، 2PL، 3PL، 4PL) أصبح أمرًا شائعًا أكثر في الممارسة العملية، ويتطلب هذا المنهج التكاملي لفهم العلاقة بين معلمات الفقرة الأربع والمعلومات: أولاً: فهم العلاقة بين معلمات المفردة الأربعة والمعلومات، ثانياً: فهم التأثيرات المجمّعة والمتعاكسة الناتجة عن إدراج جميع معلمات المفردات الأربعة في التقييم، حيث يمكن أن يؤدي هذا الجمع إلى تأثيرات معاكسة في حجم المعلومات، لذلك يجب فهم كيفية حدوث هذا الجمع وتأثيره، ثالثاً: قد يستغرق فحص المعلومات وقتًا طويلًا إذا لم تكن هناك أدوات برمجية واضحة وسهلة لتحديد مستوى المعلومات، رابعًا: يصبح وجود برمجية يمكنها إجراء تحليلات الحساسية بطريقة مرنة خيارًا عمليًا مناسبًا.

ويوضح (Tan, 2024, p. 34-36) دالة معلومات المفردة في كل نموذج علي حدى كالتالي:

١- دالة معلومات المفردة في نموذج 1PL: دالة معلومات المفردة في نموذج 1PL هي ناتج احتمال الحصول على إجابة صحيحة وخاطئة معًا $(p_i(\theta)\times q_i(\theta))$ لذلك تبلغ القيمة العظمى للمعلومات عندما $p_i(\theta)=q_i(\theta)=0.5$، وتكون قيمتها $I_i(\theta)=0.25D^2$.

٢- دالة معلومات المفردة في نموذج 2PL: يلعب معامل التمييز a دورًا محوريًا في تحديد معلومات المفردة، حيث يظهر في الصيغة على هيئة a^2_i ، مما يعني أن زيادته تؤدي إلى زيادة معلومات المفردة وتصبح القيمة العظمى أعلى، و معنى ذلك أنه كلما زادت قيمة معلمة التمييز ضاق انتشار المعلومات.

٣- دالة معلومات المفردة في نموذج 3PL: كلما ارتفع التخمين انخفضت المعلومات، فعلى سبيل المثال، عندما $c=0$ تكون المعلومات في أعلى مستوياتها.

٤-دالة معلومات المفردة في نموذج 4PL: كلما قلَّ (معامل اللامبالاه / الإهمال) d قلَّت المعلومات، وتحولت القيمة العظمى نحو قيم قدرة أقل (θ أصغر).

مما سبق يستنتج الباحثان أن القيمة العظمى لدالة المعلومات في النموذج أحادى المعلمة القيمة العظمى لدالة المعلومات تحدث عندما تكون القدرة (θ) مساوية للصعوبة(b) (b=θ)، أما في

النموذج ثنائي المعلمة كلما زاد معامل التمييز (a) فانه تزداد القيمة العظمى لدالة المعلومات وتصبح أكثر حدة، أما في النموذج ثلاثي المعلمة كلما قل معامل التخمين كلما زادت دالة المعلومات والقيمة العظمى تتحرك بعيداً عن معلمة الصعوبة (b) تلقائياً، أما في النموذج رباعي المعلمات كلما كبر معامل الإهمال أو عدم الإهتمام أو اللامبالاه (d) كلما ازدادت دالة المعلومات، وتنحرف القيمة العظمى لدالة المعلومات نحو اليسار عند قيم قدرة (θ)أقل.

أهمية دالة معلومات المفردة

يذكر (Al-zboon & Maan, 2022, p. 56) أن دالة معلومات المفردة لها استخدمات هامة كالتالي:

١- إمكانية تحديد مدى مساهمة كل مفردة في دالة معلومات الاختبار دون الاعتماد على المفردات الأخرى في الاختبار، فإذا توفرت فكرة جيدة عن قدرات مجموعة المفحوصين، فإنه يمكن اختيار المفردات التي تعظم كمية المعلومات التي يقدمها الاختبار إلى الحد الذي تتوزع فيه قدرات المفحوصين.

٢-تُستخدم في اختيار المفردات عند بناء الاختبارات بالاستناد إلى نظرية القياس الحديثة، والتي تفترض أن معلومات المفردة تختلف باختلاف مستويات السمة، وبناء عليه يمكن اختيار المفردات التي توفر دقة قياس عالية عند نقطة معينة على متصل السمة، بالإضافة إلى ذلك، فإن المفردات التي تحتوي على قيم تمييز كبيرة توفر معلومات أكبر عن قدرة المفحوصين، وبالتالي يمكن اختيار مفردات الاختبار بناءً على كمية المعلومات التي تقدمها المفردات لمجمل معلومات الاختبار.

ملاحظات حول إعداد اختبار باستخدام دوال المعلومات:

يذكر (Boone et al., 2020, p. 44) أنه يمكن استخدام دالة معلومات الاختبار لتقييم مدى جودة اختبار معين من حيث بناء اختبار مرجعي المعايير، حيث أنه من خلال إرشادات دالة معلومات الاختبار في Winsteps Help Tab تم ملاحظة أن قمة المنحنى يجب أن تقع عند نقطة القطع للاختبار، ويعتبر ذلك منطقي لأنه إذا أردنا اتخاذ قرار بالرسوب أو النجاح ، وعلي سبيل المثال كمطورى اختبار نريد الحصول علي أقصي قدر من المعلومات عند نقطة القطع، فأينما نضع نقطة القطع، يمكننا رؤية المفردات القريبة منها، وهذا يسمح لنا برؤية المفردات التي تساعدنا في التمييز بين المفحوصين الذين تكون مقاييسهم قريبة من نقطة القطع، ويمكننا أيضًا تحديد المفردات التي لا تساعد كثيرًا في قياس المفحوصين بالقرب من نقطة القطع، ويجب أن تكون هذه المفردات (التي لا تساعدنا من منظور القياس) هي المفردات البعيدة عن نقطة القطع.

ويرى (Tan, 2024, p. 42) أنه نظرًا لأن دالتي معلومات المفردة والاختبار ترتبطان ارتباطًا وثيقًا بمعلمات النماذج الأربعة لنظرية استجابة المفردة (IRT)، فإنه يوجد مجموعة من التوصيات والإرشادات العامة التي تلخص نهج المعلومات عند إعداد اختبار:

١- تحديد سمات المفحوصين: ويعتبر مهماً لاختيار المفردات في الاختبار لتحقيق هدف التقييم، فمثلاً اختيار قائمة من المفردات شديدة الصعوبة يجب أن يتماشى مع مجموعة الطلاب ذوي القدرة العالية الذين سيخضعون للاختبار، أي أنه يجب أن يتطابق الهدف مع المفردات التي يتم اختيارها.

٢- بما أن الاهتمام الأساسي في نموذج 1PL ينصب فقط على صعوبة المفردة، فإن كثافة معلومات المفردة عبر مدى القدرة الكلي تصبح الشاغل الرئيسي عند تشكيل اختبار باستخدام 1PL

٣- يُعد معامل التمييز من المعاملات القوية التي تمنح معلومات عالية للمفردة، وذلك من خلال تجميع قائمة من المفردات ذات معاملات تمييز مرتفعة، وبناء عليه سيتم الحصول على دالة معلومات اختبارية مرتفعة للاختبار.

٤- استخدام مفردات ذات معاملات تخمين عالية ومعاملات إهمال منخفضة يؤدي إلى تقليل معلومات الاختبار، ومع ذلك قد تكون هناك مقايضة ممكنة بين مقدار المعلومات وشمول مفردة معينة لتحقيق غرض محدد. فعلى سبيل المثال، فحص مستوى المرغوبية الاجتماعية باستخدام مفردات ذات معامل تخمين مرتفع سيؤدي إلى انخفاض المعلومات، ولكنه سيمكن من التحقق مما إذا كان المفحوص يقدم إجابة ذات طابع اجتماعي مرغوب، وإذا كان أحد أهداف الاختبار هو عدم التغاضي عن المفحوصين ذوي القدرة العالية الذين يجيبون بشكل خاطئ على مفردة فإن استخدام نموذج 4PL مع معامل إهمال سيكون مناسبًا

٥- الحصول على نقطة قطع باستخدام دالة معلومات اختبارية هو أحد القدرات الفعلية لنظرية IRT حيث يمكن استخدام دالة المعلومات في بناء اختبار يركز على منح المنح الدراسية لعدد محدود من الطلاب، من خلال إدراج قائمة من المفرات ذات معاملات تمييز مرتفعة بالقرب من مستوى القدرة المطلوب فإنه يسمح بتمييز المرشحين الذين يمتلكون قدرة أعلى قليلاً من العتبة المستهدفة للتأهل للحصول على المنحة، وبشكل مشابه، فإن إنشاء دالة معلومات اختبارية تستهدف مستويات قدرة محددة مثل درجة النجاح، يمكن استخدام نفس الاستراتيجية فيها.

٦- قد يكون من الضروري أن يتضمن بنك الأسئلة فقرات من نماذج 3PL و 4PL لفحص خصائص المفردات من حيث التخمين والإهمال.

٧- طول الاختبار: الاختبار الأطول بشكل عام أكثر دقة في قياس قدرة المفحوص من الاختبار الأقصر، لأن دالة معلومات الاختبار عادةً ما تكون أعلى بسبب طول الاختبار، ومع ذلك قد يؤدي الاختبار الطويل إلى الإرهاق بسبب تكرار المحتوى، كما إن نموذج IRT والقيم المقدّرة للمفردات ليسا العاملين الحاسمين الوحيدين في تحديد طول الاختبار المطلوب للوصول إلى مستوى معلومات مستهدف، بل قد تكون هناك اعتبارات أخرى ذات أهمية كذلك.

الدراسات السابقة:

أولاً: الدراسات التى تناولت بنوك الأسئلة:

دراسة (Bahrom et al., 2025) هدفت إلى توضيح كيفية تطوير بنك أسئلة باستخدام تحليل راش، وتقييم مدى صدق وثبات ورقة الامتحان النهائية لدورة إحصاء تعتمد على بنك الأسئلة كأساس لها. حيث تم تُقسّم أسئلة مقرر الإحصاء، والتي تتكون من ٧ أسئلة، إلى ٢١ سؤالًا بناءً على جدول مواصفات الاختبار (TST)، وتم تسجيل نتائج تحليل راش في واجهة بنك أسئلة مُصممة باستخدام برنامج إكسل، تُسهّل واجهة بنك الأسئلة الوصول إلى سهولة الوصول لمجموعة كبيرة ومتنوعة من الأسئلة المُختبرة مسبقًا، مما يسمح بإنشاء أوراق امتحان متنوعة ومتوازنة، ويُعدّ بنك الأسئلة المُطوّر جيدًا عونًا كبيرًا لمُعدّي الامتحانات، إذ يُسهّل عملية إنشاء الاختبارات ويسرعها ويزيد من كفاءتها، مما يُؤدي إلى أوراق أسئلة امتحان عالية الجودة.

دراسة (keung Yim et al., 2024) هدفت الدراسة إلي تقييم الخصائص السيكومترية لبنك الأسئلة لتقييم مهارات فهم المقروء لدى طلاب المدارس الابتدائية، حيت تم تطوير مجموعة مكونة من ٤٦ فقرة قراءة من الصف الأول إلى السادس الابتدائي، بالإضافة إلى ٥٢٢ بندًا من نوع الاختيار من متعدد والإجابة القصيرة، وذلك بالاستناد إلى إطار تقييم "التقدم في دراسة القراءة الدولية" (PIRLS)، وقد تم اختبار هذه الفقرات ميدانيًا في ٢٧ مدرسة في سنغافورة بمشاركة ٩٨٣٤ طالبًا تتراوح أعمارهم بين ٧ و١٣ عامًا، وقد تم تقييم أربع عمليات فهم رئيسية وفقًا لما حدده إطار PIRLS، وهي: التركيز على واسترجاع المعلومات المصرح بها صراحة، واستنتاج المعلومات المباشرة، وتفسير ودمج الأفكار والمعلومات، وتقييم ونقد المحتوى والعناصر النصية، وتم استخدام تحليل "راش" (Rasch) لفحص أنماط استجابات الطلاب للأصناف من حيث: ملاءمة النموذج والبنود، والأداء التفاضلي للبنود (DIF) فيما يتعلق بنوع الجنس ومنصة الاختبار المستخدمة، والاعتماد المحلي للبنود (LID) ضمن وبين الفقرات، ومشكلات المشتتات في خيارات أسئلة الاختيار من متعدد، وأظهرت النتائج أن البيانات تناسب نموذج راش أحادي البعد عبر جميع مستويات الاختبار مع ثبات

داخلي جيد، وتمثلت المشكلات السيكومترية التي تم تحديدها في البنود أساسًا في المشتتات غير الفعالة والاعتماد المحلي بين البنود.

دراسة (فردوس بنت عبدالله الخوتانى ، ٢٠٢٣) هدفت الدراسة إلي بناء بنك أسئلة في مقرر الرياضيات (٤) للصف الثاني الثانوى في ضوء نظرية الاستجابة للمفردة وفق النموذج اللوغاريتمى ثلاثي المعلمة، وتكونت العينة من (٦٠٣) من طلاب الصف الثانى الثانوى، تم بناء نموذجي اختبارين متكافئين ضمن النسب الموزونة لكل هدف من أهداف مقرر الرياضيات، وتم استخدام المنهج الوصفي التحليلي حيث تم استخدام برنامج (Spss) من أجل التحقق من الخصائص السيكوموترية لأداوات الدراسة وكما تم استخدام برنامج (Xcaliber) من أجل التحقق من افتراضيات نظرية الاستجابة للمفردة وتقدير معالم الفقرات والخطأ المعيارى للفقرات ومطابقة الفقرات للنموذج اللوغاريتمى ثلاثي المعلمة، وتوصلت النتائج إلي أنه تم حذف عدد (١) من الفقرات غير المطابقة في الصورة الأولي وعدد (٣) من الفقرات غير المطابقة في الصورة الثانية، وتكون بنك الأسئلة في صورته النهائية من (٣٦) فقرة من كلا صورتي الاختبار تتمتع بجودة مطابقة عالية للنموذج ثلاثي المعلمة.

دراسة (Perez & Padrones, 2022) هدفت الدراسة إلي بناء ومعايرة اختبار آلي (اختيار من متعدد بالورقة والقلم) في مادة الرياضيات باستخدام بنك أسئلة وفق النموذج اللوغاريتمى ثلاثي المعلمة في نظرية الاستجابة للمفردة، وتكونت عينة البحث من (٨١٩) طالباً من طلاب جامعة ولاية ميشيغان، تم تقديم اختبار في مادة الرياضيات من نوع الاختيار من متعدد، ,تم تحليل البيانات تحت نموذج اللوغاريتمى ثلاثي المعلمة باستخدام الحزم الإحصائية R وتحديداً الحزمة ltm، حيث تم تقدير معلمات المفردة مثل الصعوبة والتمييز والتخمين باستخدام تقدير الأرجحية العظمى الهامشية (MMLE) نظرًا لأن حزمة ltm تستخدم MMLE من بين إجراءات التقدير المتاحة، تم استخدام مربع كاي لتقييم مدى ملاءمة النموذج للبيانات، وتوصلت النتائج إلي عدم وجود مفردات سهلة في الاختبار التمهيدى الثاني، وتبين أن ثلاثة وأربعين بالمائة (٤٣٪) من المفردات كانت صعبة، وتم تصنيف ٢٩٪ من المفردات على أنها صعبة للغاية، وكانت معظم المفردات شديدة التمييز حيث وجدت أن معظم الفقرات تميز بشكل كبير بين المجموعات بين الممتحنين ذوي القدرة المنخفضة والمرتفعة في الإجابة على تلك المفردات، أما بالنسبة لمعلمة التخمين فإن أكثر من ٥٠٪ منها كانت لها قيم منخفضة وهذا يعني أن معظم المفردات أعطت فرصة ضئيلة للممتحنين ذوي القدرة المنخفضة للإجابة بشكل صحيح عن طريق التخمين، وبالتالي نجح البحث في بناء اختبار آلي باستخدام المفردات التي تمت معايرتها حيث أنه تم تضمين ٤٦٪ فقط من مفردات الاختبار بسبب عدم ملاءمة المفردات الأخرى للنموذج اللوغاريتمى ثلاثي المعلمة.

دراسة (ماجد عبده ناجي وعبدالمحسن بن رشيد المبدل، ٢٠٢٢) هدفت الدراسة إلي تطوير بنك أسئلة لأساسيات القياس والتقويم لطلاب كلية التربية جامعة إب وفق النموذج اللوغاريتمى ثلاثي المعلمة في نظرية الاستجابة للمفردة، وتكونت العينة من (٤٩٠) طالباً وطالبة من طلبة كلية التربية جامعة إب، وتم عمل أربع صور إختبارية من خلال تحليل محتوى الكتاب المقرر وفي ضوء جدول المواصفات، وتم تحليل الإجابات علي الصور الاختبارية الأربع باستخدام برنامج (Xcaliber4.2) وفق النموذج اللوغاريتمى ثلاثي المعلم، وتوصلت النتائج فيما يتعلق بتقدير معالم الفقرات وفق النموذج اللوغاريتمى ثلاثي المعلم بعد التدريج أن قيم معلم التمييز تراوحت مابين (٠,٢٦٢-١,٩٦٦)، وكان متوسط معلم التمييز(٠,٨٩٥)، وتراوحت قيم معلم الصعوبة مابين (-١,١٢٣ إلي ١,٥٣٩) وكان متوسط معلم الصعوبة (٠,٣٥٦)، وتراوحت قيم معلم التخمين ما بين (٠,٠١٢-٠,٣٨٠) وكان متوسط معلم التخمين (٠,٢٥١)، وتم تخزين (٢٧٦) فقرة وتم تخزين المعلومات لكل فقرة.

دراسة (توفيق علي عالم و سوسن حسن أبوالهادى، ٢٠١٩) هدفت إلي بناء بنك أسئلة في مادة الرياضيات للصف التاسع الأساسي في ضوء نظرية استجابة الفقرة باستخدام النموذج اللوغاريتمى ثلاثي المعلمة، والتحقق من جودة توظيف برنامج حاسوبي لإدارة فقرات بنك الأسئلة من حيث التخزين والاستدعاء واختيار فقرات منه لتشكيل اختبار جديد، وتكونت العينة من ١٢٠٠ طالب وطالبة من طلاب الصف التاسع من التعليم الأساسي موزعين علي بعض المناطق التعليمية لأمانة العاصمة صنعاء، تم عمل ستة نماذج اختبارية متكافئة ويحتوى كل نموذج علي ٤٠ فقرة منهم ٦ فقرات مكررة في كل نموذج، وتم استخدام برنامج (Bilog MG3) للتأكد من افتراضيات النظرية، وتوصلت نتائج التحليل فيما يتعلق بنتائج مطابقة الفقرات أن ٤ فقرات من أصل ٢١٠ فقرة لم تطابقي النموذج الثلاثي المعلمة، بينما نتائج المطابقة للأفراد فإن ٤ أفراد من أصل ١٢٠٠ لم يطابق النموذج ثلاثي المعلمة.

دراسة (وليد محمد خليفة فرج الله، ٢٠١٨) هدفت إلي بناء بنك أسئلة إلكتروني في مقرر جغرافيا البحار والمحيطات وقياس أثره علي تنمية الأعماق المعرفية وخفض قلق الإختبارلدى منخفضات التحصيل بقسم الجغرافيا كلية الآداب جامعة بيشه، وتكونت عينة البحث من (٣٠) طالبة من طالبات المستوى الخامس، وتم استخدام المنهج شبه التجريبي تصميم القياس القبلي- البعدى لمجموعتين إحداهما تجريبية والآخر ضابطة، وتمت المعالجة الإحصائية باستخدام برنامج (Spss)، وتوصلت النتائج إلي فاعلية استخدام بنك الأسئلة في خفض قلق الاختبار وتنمية مستويات العمق المعرفي.

دراسة (Hung et al., 2017) هدفت الدراسة إلي تقييم الأداء لبنك أسئلة الخاصة بالطرف العلوى ضمن مجموعة مرضى العظام في الأطراف العلوية، وتكونت العينة من (١١٩٧) مريضاً بالغاً

يقدمون إلى عيادة العظام المتخصصة في حالات اليد والأطراف العلوية، وتم فحص البيانات باستخدام الأساليب الاتقليدية ونموذج راش، وتوصلت النتائج إلي أن بنك الأسئلة كان مناسبا لنموذج راش مما يشير إلى أن العناصر قادرة على تقديم معلومات حول مجالات متميزة وقابلة للقياس في PF.

دراسة (هبه الجمل وآخرون، ٢٠١٧) هدفت إلي تدريج بنك أسئلة باستخدام النموذج اللوغاريتمى الثلاثى المعلم في مادة الرياضيات للصف الثالث الإبتدائي والكشف عن مدى اختلاف التقديرات باختلاف الاختبارات المسحوبة من البنك (سهل وصعب)، كما تهدف إلي الكشف عن مدى اختلاف دالة معلومات الاختبار باختلاف مستوى العينة (منخفضة القدرة – مرتفعة القدرة)باستخدام النموذج اللوغاريتمى ثلاثي المعلم، وتكونت العينة من (١٦٦٨) تلميذاً وتلميذة بالصف الثالث الإبتدائي، تم تدريج بنك أسئلة مكون من (٢٢٥) فقرة، وتم استخدام برنامج (Spss 20) و(Winsteps) و(Bilog MG٣) للتأكد من افتراضيات النظرية ، وتوصلت النتائج إلي اختلاف تقديرات الأفراد باختلاف الاختبارات المسحوبة من البنك (سهل – صعب) وكذلك لا تختلف دالة معلومات الاختبار باختلاف مستوى العينة (منخفضة القدرة – مرتفعة القدرة) باستخدام النموذج ثلاثي المعلم.

دراسة (محمد جاد محمد، ٢٠١٦) هدفت إلي المقارنة بين النماذج اللوغاريتمية الثلاث لنظرية الاستجابة للمفردة (أحادى/ الثنائئ/ الثلاثي) البارامتر في تدريج بنك للأسئلة الموضوعية في مادة الرياضيات للصف الخامس الإبتدائي، وتكونت العينة من (١٥٠٠) تلميذاً وتلميذة في الصف الخامس الإبتدائي، قام الباحث بعمل (١٠) صور اختبارية من نوع الاختيار من متعدد في مادة الرياضيات للصف الخامس الإبتدائي في وحدة "الكسور والعمليات المختلفة عليها"، وتم استخدام برنامج (Parscale) الإصدار الرابع وبرنامج (Spss) في إجراء المعاملات الإحصائية اللازمة لنظرية الإستجابة للمفردة، وتوصلت النتائج إلي أن النموذج اللوغاريتمى الثلاثي المعالم هو أكثر النماذج ملائمة للبيانات وأكثرها دقة في تقدير وتدريج صعوبة المفردة.

ثانياً: الدراسات التي تناولت دالة المعلومات:

دراسة (Tan, 2024) هدفت إلى تقديم دليل استراتيجي عملي لاختيار المفردات لبنك الأسئلة وتوضيح العلاقة بين دوال معلومات المفردة والاختبار ومعلمات نماذج IRT الأربعة، و عرض صيغ R البرمجية لتوليد دوال المعلومات والاختبار، واستخدام المنهج البياني لرسم منحنيات معلومات المفردة والاختبار، وتوصلت النتائج إلي إلي أن معلمة الصعوبة تشير إلى موقع القيمة العظمى في دالة معلومات المفردة، وأنه كلما زادت قيمة معامل التمييز، زادت المعلومات وازداد ارتفاع وضيق منحنى المعلومات، وكلما قلت قيمة معامل التخمين، زادت المعلومات، وكلما زادت قيمة معامل الإهمال، زادت المعلومات، وتوصلت أيضاً إلي النتائج إلي أن نهج المعلومات ليس الهدف الوحيد في بناء التقييمات

إذ غالبًا ما تتجاوز عملية اختيار مفردات الاختبار حدود الدقة لتشمل نطاق المادة العلمية، والاعتبارات المتعلقة بالمحتوى، ومستوى الأهداف، وبناء المشتتات، وصيغة الاختبار، وغرض التقييم، وخصائص المفحوصين، ومدى تأثر الخصائص السيكومترية للاختبار بطريقة التصحيح، ولذلك يمكن دمج تقييم التقييم مع هذه العوامل مع أخذ الدقة بعين الاعتبار.

دراسة (ALKursheh et al., 2022) هدفت إلي مقارنة تأثير شكلين من أشكال مفردات الاختبار (الاختيار من متعدد والإكمال) على تقدير قدرة الشخص، ومعلمات المفردة، ودالة معلومات الاختبار (TIF)، وتم إنشاء اختبار في الرياضيات (١) بصيغتين: الاختيار من متعدد والإكمال، وتكون الاختبار في صورته النهائية من (٣١) مفردة، وتم تطبيق الاختبار على (٣٠٠) طالب وطالبة في جامعة تبوك، وقد تم تحليل استجابات المفحوصين باستخدام برنامج BILOG–MG 3 لكل صيغة من صيغ المفردات وفقًا لنموذج لوجستي ثنائي المعلمة، وتوصلت النتائج إلي وجود فروق ذات دلالة إحصائية عند مستوى الدلالة (α=0.05) بين متوسطات الخطأ المعياري في تقدير معلمات المفردات (الصعوبة والتمييز) ترجع إلى شكل مفردة الاختبار لصالح مفردات الإكمال، كما أظهرت النتائج وجود فروق ذات دلالة إحصائية عند مستوى (α=0.05) بين متوسطات الخطأ المعياري في تقدير قدرة الشخص تُعزى إلى شكل مفردة الاختبار لصالح مفردات الإكمال، وتوصلت أيضاً إلي وجود فروق ذات دلالة إحصائية (α=0.05) بين متوسطات الخطأ المعياري لدالة معلومات الاختبار تُعزى إلى شكل مفردة الاختبار لصالح مفردات الإكمال.

دراسة (Al-zboon & Maan, 2022) هدفت إلي تحديد أثر عدد المشتتات في مفردات الاختيار من متعدد على الخصائص السيكومترية للمفردات ودالة معلومات المفردة وفق نموذج لوجستي ثنائي المعلمة في نظرية استجابة للمفردة (IRT)، وللإجابة على أسئلة الدراسة، تم بناء اختبار تحصيلي يغطي كافة محاور مقرر الاختبارات المدرسية ويقيس مدى معرفة المعلمين ببناء الاختبارات المدرسية، وتكون الاختبار من (٤٢) مفردة وجاء في ثلاثة نماذج مختلفة تختلف في عدد البدائل، وتكونت عينة الدراسة من ٣٥٦ معلمًا ومعلمة من محافظة معان ولواء وادي السير وتم اختيارها بطريقة عشوائية بسيطة، وتم استخدام البرنامج الإحصائي (SPSS) وبرنامج (Bilog-MG3) لتحليل استجابات العينة، و توصلت النتائج إلي وجود فروق دالة إحصائيًا بين متوسطات قيم دالة معلومات المفردات لصالح نموذج الاختبار ذي الأربع مشتتات مقارنة بنموذجي الثلاث والخمس مشتتات، وكما أظهرت النتائج وجود فروق دالة إحصائيًا لصالح نموذج الاختبار ذي الخمس مشتتات مقارنة بنموذج الثلاث مشتتات، وكذلك أظهرت النتائج عدم وجود فروق في متوسط تقديرات معاملات صعوبة المفردات تُعزى إلى عدد المشتتات في المفردة. في المقابل، أُشير إلى وجود فروق في متوسط تقديرات معاملات التمييز للمفردات بين نموذجي الثلاث والأربع مشتتات لصالح النموذج ذو

الأربع مشتتات، ووجود فروق دالة إحصائيًا لصالح النموذج ذو الخمس مشتتات عند المقارنة بنموذج الثلاث مشتتات.

دراسة (Al-zboon, 2020) هدفت إلي تحديد أثر طول اختبار الاختيار من متعدد على تقدير معلمات المفردات ودالة معلومات الاختبار وفقًا للنموذج اللوجستي ثلاثي المعلمات في نظرية الاستجابة للمفردات، ولتحقيق أهداف الدراسة اعتمد الباحث برنامج (WINGEN3) لتوليد البيانات، حيث تم توليد معلمات المفردات (الصعوبة، والتمييز، والتخمين) لثلاثة نماذج من الاختبارات تتكون من (٢٠، ٤٠، ٦٠) مفردة، واستخدمت معلمات المفردات الناتجة في توليد القدرات لـ (٤٠٠٠) مفحوص باستخدام النموذج اللوجستي ثلاثي المعلمات، كما تم استخدام برنامج (BILOG-MG3) لإيجاد معلمات المفردات ودالة معلومات النماذج الثلاثة، وأظهرت نتائج الدراسة وجود فروق ذات دلالة إحصائية عند مستوى الدلالة ($\alpha = 0.05$) بين المتوسطات الحسابية لتقديرات معلمات المفردات تعزى إلى طول الاختبار ولصالح الاختبار الأطول في عدد المفردات، وكما أظهرت النتائج أن دالة معلومات الاختبار تزداد بزيادة عدد المفردات.

تعليق عام على الدراسات السابقة:

تُقدم الدراسات السابقة استعراضًا شاملاً ومهمًا للأدبيات المتعلقة ببناء بنوك الأسئلة وتطبيق نظرية الاستجابة للمفردة (IRT)، بالإضافة إلى دراسات متخصصة في دالة معلومات الاختبار والمفردة. تتنوع هذه الدراسات في أهدافها ومنهجياتها ومجالات تطبيقها، لكنها تتفق جميعًا على أهمية استخدام المقاييس السيكومترية الحديثة لضمان جودة أدوات القياس.

تُبرز الدراسات المتعلقة ببنوك الأسئلة (مثل Bahrom et al., 2025؛ Yim, Lye & Koh, 2024؛ فردوس بنت عبد الله الخوتاني، ٢٠٢٣؛ Perez & Padrones, 2022؛ ماجد عبده ناجي وعبد المحسن بن رشيد المبدل، ٢٠٢٢؛ توفيق علي عالم وسوسن حسن أبو الهادي، ٢٠١٩؛ وليد محمد خليفة فرج الله، ٢٠١٨؛ Hang et al., 2017) الاتجاه المتزايد نحو تطوير بنوك أسئلة مُعايرة بدقة باستخدام نماذج IRT المختلفة (مثل نموذج راش، النموذج ثلاثي المعلمة). تُشدد هذه الدراسات على أن بنوك الأسئلة هذه تُسهم في تسهيل بناء اختبارات متنوعة ومتوازنة، وتحسين كفاءة عملية التقييم، وتوفير أدوات قياس عالية الجودة. كما تتناول بعضها التحديات العملية مثل مطابقة المفردات للنموذج المقترح، وفعالية المشتتات، والاعتماد المحلي بين البنود، وكيفية التعامل مع المفردات غير المطابقة.

من جانب آخر، تُسلط الدراسات التي ركزت على دالة معلومات المفردة والاختبار (مثل Tan, 2024؛ ALKursheh, Al-zboon & AlNasraween, 2022؛ Al-zboon & Maan, 2022؛ محمد جاد محمد، ٢٠١٦) الضوء على الجانب الحيوي لدوال المعلومات

في فهم جودة المفردات والاختبارات. تُظهر هذه الدراسات كيف تتأثر دالة المعلومات بمعالم المفردة (صعوبة، تمييز، تخمين)، وكيف يمكن استخدامها لتقييم كفاءة القياس عند مستويات قدرة مختلفة. كما تُناقش تأثير عوامل مثل شكل المفردة (اختيار من متعدد مقابل إكمال)، وعدد المشتتات، وطول الاختبار على الخصائص السيكومترية ودالة المعلومات، مما يوفر رؤى قيمة لتصميم الاختبارات المثلى. دراسة محمد جاد محمد (٢٠١٦) بارزة في تأكيدها على أن النموذج اللوغاريتمي ثلاثي المعالم غالبًا ما يكون الأنسب والأكثر دقة في تدريج صعوبة المفردة.

بشكل عام، تُعزز هذه الدراسات أهمية استخدام IRT في بناء أدوات قياس نفسية وتعليمية موثوقة وعادلة. تُظهر جميعها التزامًا بتطبيق المنهجية الكمية وتحليل البيانات بدقة لتقييم الخصائص السيكومترية للمفردات والاختبارات. تتراوح المجالات التطبيقية من الرياضيات والإحصاء إلى مهارات فهم المقروء والجغرافيا، مما يدل على اتساع نطاق تطبيق IRT وفوائدها في مختلف التخصصات.

تبرز هذه الدراسات الحاجة المستمرة إلى تطوير بنوك أسئلة مُعايرة بعناية، مع الأخذ في الاعتبار أدق التفاصيل السيكومترية. والبحث الحالي ينسجم تمامًا مع هذا التوجه المعاصر، حيث يسعى إلى بناء بنك أسئلة لمقرر "الاتجاهات المعاصرة في علم النفس" باستخدام المعايرة المتزامنة ونماذج IRT المتعددة، مما سيضيف قيمة كبيرة للأدبيات القائمة من خلال تطبيق هذه التقنيات المتقدمة في سياق تخصصي محدد وهو علم النفس التربوي، وهو ما يميزه عن العديد من الدراسات السابقة التي ركزت على مجالات أخرى أو نماذج أقل شمولية في IRT .

إجراءات البحث:

يتطلب تحقيق أهداف البحث الإجابة عن تساؤلاته من خلال منهجية دقيقة ومنظمة. لذا، سيعرض هذا القسم أولًا إجراءات البحث التي تشمل المنهج المتبع، وحدود البحث، ومجتمع البحث وعينته، وأدوات الدراسة المستخدمة، بالإضافة إلى الخطوات التفصيلية لبناء وتطبيق البنك. ثانيًا، سيتم توضيح الأساليب الإحصائية وحزم البرامج التي أُعتمد عليها في تحليل البيانات، بدءًا من التحليل التقليدي وصولًا إلى نماذج نظرية الاستجابة للمفردة المتقدمة، وذلك لضمان الدقة والموثوقية في النتائج. أخيرًا، سيتناول القسم الأخير نتائج البحث وتفسيرها ومناقشتها، حيث سيتم عرض الاستنتاجات المستخلصة من التحليلات الإحصائية، مع ربطها بالأدبيات النظرية والدراسات السابقة، مما يُسهم في تقديم رؤى معمقة حول جودة بنك الأسئلة المُنشأ وخصائص مفرداته في ضوء المعايرة المتزامنة ونظرية الاستجابة للمفردة.

منهج البحث:

اعتمد البحث الحالي المنهج الوصفي التحليلي، بهدف بناء بنك أسئلة لمقرر "الاتجاهات المعاصرة في علم النفس التربوي" ومعايرته باستخدام أحدث التقنيات المتاحة في القياس النفسي.

شملت المنهجية جمع البيانات الكمية من استجابات الطلاب على خمسة اختبارات تحصيلية، كل منها يمثل فصلًا من فصول المقرر. تم تحليل هذه البيانات على مرحلتين: أولاً، تم إجراء تحليل تقليدي (كلاسيكي) للمفردات باستخدام نظرية القياس الكلاسيكية (CTT) بهدف تنقية بنك الأسئلة من المفردات ذات التمييز الضعيف. ثانيًا، تم تطبيق التحليل القائم على نظرية الاستجابة للمفردة (IRT) على جميع المفردات المتبقية معًا (معايرة متزامنة) للكشف عن معالم المفردات والأفراد بناءً على النموذج الأنسب من نماذج الاستجابة للمفردة أحادية البعد (راش، أحادي المعلمة، ثنائي المعلمة، ثلاثي المعلمة، رباعي المعلمة)، وذلك للإجابة عن أسئلة الدراسة وتحقيق فكرة إنشاء بنك أسئلة يضيف للقياس النفسي.

حدود البحث:

تحددت حدود البحث بالآتي:

- **الحدود الزمانية: تم جمع البيانات وتطبيق الاختبارات على الطلاب خلال فترات زمنية مختلفة على عامين دراسيين مختلفين (٢٠٢٢/٢٠٢٣& ٢٠٢٤/٢٠٢٥م)، حيث طُبق كل اختبار في نهاية الفصل المحدد له، وذلك خلال الفترة التي توافق تطبيق لائحة ٢٠٢٠ ووفقاً للائحة المعدلة النهائية التي طُبقت على طلاب كلية التربية على مستوى الجمهورية**
- **الحدود المكانية: كلية التربية – جامعة الزقازيق.**
- **الحدود البشرية (العينة): اقتصر البحث على طلاب الفرقتين الثانية والثالثة من شعبتي التعليم الأساسي وعلم النفس التربوي.**
- **الحدود الموضوعية:**
  - **اقتصر البحث على مفردات الاختبارات التحصيلية لمادة "الاتجاهات الحديثة (المعاصرة) في علم النفس."**
  - **تم التعامل مع المفردات ثنائية الاستجابة (٠ أو ١) فقط، والتي تشمل مفردات الاختيار من متعدد ومفردات الصواب والخطأ.**
  - **تم تطبيق التحليل باستخدام النظرية الكلاسيكية للاختبار (CTT) ونظرية الاستجابة للمفردة (IRT)، مع التركيز بشكل خاص على نهج المعايرة المتزامنة، واستكشاف نماذج IRT أحادية البعد المتعددة (راش، أحادي، ثنائي، ثلاثي، رباعي المعلمة).**

عينة البحث النهائية:

نظراً لطبيعة البحث فقد تم الاكتفاء بالعينة النهائية والتي تكونت من (١١٦٥) طالبًا وطالبة من الفرقتين الثانية والثالثة، من شعبتي التعليم الأساسي وعلم النفس التربوي بكلية التربية– جامعة

الزقازيق. قام جميع الطلاب بالإجابة على جميع الأسئلة البالغ عددها (٤٨١) مفردة في الصورة الأولية للبنك، وذلك على فترات زمنية متفرقة وفقًا لوقت تطبيق كل اختبار في نهاية كل فصل من فصول الكتاب المقرر إنشاء بنك أسئلة له.

أدوات البحث:

تمثلت أداة البحث الرئيسية في بنك الأسئلة التحصيلي الذي تم تطويره لمقرر "الاتجاهات الحديثة (المعاصرة) في علم النفس". تكون هذا البنك في صورته الأولية من (481) مفردة ثنائية الاستجابة، مقسمة على خمسة اختبارات تغطي فصول الكتاب المقرر، كما يوضح الجدول (١):

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

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)

تفاصيل بنك الأسئلة وعملية تنقيحه:

- **توزعت المفردات الأصلية وعددها (٤٨١) مفردة بواقع ((٢٥٠) مفردة اختيار من متعدد من أربعة بدائل، و(٢٣١) مفردة صواب وخطأ).**
- **خضع بنك الأسئلة لتحليل أولي باستخدام النظرية التقليدية في القياس (CTT) في برنامج RStudio (Team, 2025) وحزمة "CTT" (Willse, 2018)**
- **أظهر هذا التحليل أن درجة ثبات البنك (ألفا كرونباخ) بلغت (0.980) والخطأ المعياري في القياس (SEM)كان (7.538)**
- **تم تحديد (19) مفردة سالبة بناءً على معامل الارتباط الثنائي الأصيل (Point Biserial Correlation)، مما يشير إلى ضعف تمييزها وفشلها في التفريق بين الطلاب ذوي القدرات العالية والمنخفضة، وتم حذف هذه المفردات.**
- **بعد حذف المفردات الـ (١٩)، أصبح بنك الأسئلة في صورته النهائية التي ستدخل التحليل الحديث وفقًا لنظرية الاستجابة للمفردة مكونًا من (462) مفردة ثنائية الاستجابة، منها (٢٤٨) مفردة اختيار من متعدد من أربعة بدائل، و(٢١٤) مفردة صواب وخطأ. وقد تحسنت خصائصه السيكومترية ليصبح ثباته (ألفا كرونباخ) (0.982) والخطأ المعياري في القياس(7.263) (SEM) .**

الأساليب الإحصائية وحزم البرامج المستخدمة:

لتحقيق أهداف البحث، تم الاعتماد على منهجية تحليلية ثنائية الشق، شملت التحليل التقليدي للمفردات وتطبيق نماذج نظرية الاستجابة للمفردة (IRT)، وذلك باستخدام بيئة RStudio الإحصائية المتطورة.

أولاً: التحليل التقليدي (الكلاسيكي) للمفردات : تم تحليل (٤٨١) مفردة ثنائية الاستجابة (٢٥٠ اختيار من متعدد، ٢٣١ صواب وخطأ) في بنك الأسئلة مبدئيًا باستخدام النظرية التقليدية في القياس (CTT) تركز هذا التحليل على تقييم الخصائص السيكومترية الأولية للمفردات، وتحديدًا اكتشاف المفردات ضعيفة التمييز. تم الاستدلال على ذلك من خلال معامل الارتباط الثنائي الأصيل (Point Biserial Correlation) أسفر التحليل عن وجود ١٩ مفردة ذات قيم ارتباط سالبة، والتي تم حذفها لاحقًا. أُجري هذا التحليل في RStudio (Team, 2025) باستخدام حزمة "CTT" (Willse, 2018) بعد حذف المفردات، أصبح العدد النهائي للمفردات (٤٦٢) مفردة (٢٤٨ اختيار من متعدد، ٢١٤ صواب وخطأ). تحسنت خصائص البنك بشكل طفيف بعد الحذف؛ حيث بلغت درجة ثباته بناءً على ألفا كرونباخ (0.982)، وانخفض الخطأ المعياري في القياس (SEM) إلى ٧.٢٦٣)؛ مما يؤكد جودة المفردات التي استمرت في التحليل.

ثانياً: المعايرة باستخدام نظرية الاستجابة للمفردة (IRT):

بعد التنقية الأولية، تم الانتقال إلى مرحلة المعايرة باستخدام نظرية الاستجابة للمفردة، وتحديدًا من خلال المعايرة المتزامنة. تم محاولة معايرة بنك الأسئلة باستخدام عدة نماذج من IRT وهي (نموذج راش، النموذج الأحادي المعلمة، النموذج الثنائي المعلمة، النموذج الثلاثي المعلمة، والنموذج الرباعي المعلمة). واجهت عملية التحليل تحديات مع النموذجين الثلاثي والرباعي المعلمة، حيث لم يكملا التحليل بسبب مشكلات حسابية تتعلق بضخامة البيانات (على سبيل المثال: Error in solve.default و "cannot allocate vector of size 84.5 Gb" لمؤشر M2). في المقابل، أتمت نماذج راش، الأحادي المعلمة، والثنائي المعلمة التحليل بكفاءة.

تم تقييم ملاءمة هذه النماذج للبيانات على مستويين :ملاءمة المفردة (Item Fit) وملاءمة النموذج (Model Fit)، على الرغم من عدم ظهور نتائج مؤشر M2 لضخامة البيانات. كما تم إجراء تحليل التباين (ANOVA) بين النماذج الثلاثة التي أتمت التحليل لتحديد الأنسب. وبناءً على هذه التحليلات، تم تحديد النموذج الثنائي المعلمة (2PLM) باعتباره النموذج الأكثر ملاءمة لبنك الأسئلة.

ثالثاً: التحليل النهائي بالنموذج الثنائي المعلمة:

بعد اختيار النموذج الثنائي المعلمة، تم حذف المفردات والأفراد الذين لم يُظهروا ملاءمة كافية لهذا النموذج، تضمنت المخرجات النهائية تقدير معالم المفردات (Item Parameter

(Estimation)، ورسوم بيانية لـ منحنى خصائص المفردة(ICC) ، دالة معلومات المفردة (IIF)، ودالة معلومات الاختبار (TIF) مع الأخطاء المعيارية (SE) كما شملت النتائج اختبارات ملاءمة الشخص (Person Fit Test)، ورسوم بيانية للمدرج التكراري لملاءمة الشخص، ونتائج تقدير معالم الأفراد باستخدام طرق (MLE- MAP- EAP)، ولأسباب تتعلق بطول البحث وضخم النتائج وصعوبة عرضها تم الاكتفاء بعرض أهم النتائج.

نتائج البحث وتفسيرها ومناقشتها:

نتائج السؤال الأول "ما مدى فعالية استخدام المعايرة المتزامنة في بناء بنك أسئلة لمقرر "الاتجاهات المعاصرة في علم النفس" ضمن نماذج نظرية الاستجابة للمفردة أحادية البعد (راش، أحادي المعلمة، ثنائي المعلمة، ثلاثي المعلمة، رباعي المعلمة)؟"

أظهر استخدام المعايرة المتزامنة فعالية كبيرة في بناء بنك أسئلة لمقرر "الاتجاهات المعاصرة في علم" ضمن إطار نظرية الاستجابة للمفردة أحادية البعد، وذلك للنماذج المناسبة لخصائص البيانات المتوفرة.

قبل البدء بتحليلات نظرية الاستجابة للمفردة، تم تجهيز بنك الأسئلة من خلال مرحلة التحليل التقليدي (CTT) في هذه المرحلة، تم تحليل ٤٨١ مفردة ثنائية الاستجابة (٢٥٠ اختيار من متعدد، ٢٣١ صواب وخطأ). تم حذف ١٩ مفردة منها بناءً على معامل الارتباط الثنائي الأصيل (Point Biserial Correlation) السالب، والتي تُشير إلى مفردات فشلت في التمييز بين الطلاب ذوي القدرات العالية والمنخفضة. وبذلك، أصبح بنك الأسئلة في صورته النهائية المُعدة للتحليل الحديث مكونًا من 462 مفردة ثنائية الاستجابة (٢٤٨ مفردة اختيار من متعدد من أربعة بدائل، و٢١٤ مفردة صواب وخطأ). أظهر هذا البنك المُعدل خصائص ثبات ممتازة بناءً على التحليل الكلاسيكي CTT، حيث بلغت درجة ثباته بناءً على ألفا كرونباخ (0.982)، وكان الخطأ المعياري في القياس (SEM) (7.263).

وبعد هذه التنقية، تم الانتقال إلى مرحلة المعايرة المتزامنة باستخدام نماذج نظرية الاستجابة للمفردة. تم إجراء جميع هذه التحليلات في بيئة RStudio (Team, 2025) بالاستعانة بحزمة "mirt" (Chalmers, 2012)، وهي حزمة متخصصة وقوية لنمذجة IRT تم تطبيق المعايرة المتزامنة لكل نموذج على حدة، على النحو التالي:

- نموذج راش (Rasch Model): تم تطبيقه بنجاح باستخدام دالة mirt بتحديد النموذج "Rasch".
- النموذج الأحادي المعلمة (1PLM): تم تطبيقه بنجاح باستخدام دالة mirt بتحديد النموذج 1PL

- **النموذج الثنائي المعلمة (2PLM): تم تطبيقه بنجاح باستخدام دالة mirt بتحديد النموذج 2PL**
- **النموذج الثلاثي المعلمة (3PLM): تم محاولة تطبيقه باستخدام دالة mirt بتحديد النموذج 3PL، لكنه لم يكمل التحليل.**
- **النموذج الرباعي المعلمة (4PLM): تم محاولة تطبيقه باستخدام دالة mirt بتحديد النموذج 4PL، لكنه أيضًا لم يكمل التحليل.**

ولقد واجهت المعايرة المتزامنة تحديات مع النموذجين الثلاثي والرباعي المعلمة؛ حيث لم يكملا التحليل بنجاح. يُعزى ذلك إلى المشكلات الحسابية المرتبطة بضخامة حجم البيانات، والتي تجاوزت القدرات الحاسوبية المتاحة لإجراء العمليات المعقدة المطلوبة لهذه النماذج، كما دلت رسائل الخطأ التي تشير إلى مشكلات في تخصيص الذاكرة (cannot allocate vector of size 84.5 Gb) ووجود أنظمة غير قابلة للحل حسابيًا .(system is computationally singular) هذا يشير إلى أن فعالية المعايرة المتزامنة، وإن كانت عالية، قد تكون مقيدة بالقدرات الحاسوبية وحجم البيانات عند تطبيقها على نماذج IRT الأكثر تعقيدًا.

وفي المقابل، أتمت نماذج راش، الأحادي المعلمة، والثنائي المعلمة التحليل بكفاءة باستخدام المعايرة المتزامنة في حزمة mirt تم تقييم ملاءمة المفردة (Item Fit) وملاءمة النموذج (Model Fit) لهذه النماذج، على الرغم من عدم إخراج مؤشر M2 لضخامته (cannot allocate vector of size 84.5 Gb) .

ومما سبق، أثبتت المعايرة المتزامنة، المطبقة عبر حزمة mirt، فعاليتها في بناء بنك الأسئلة لمقرر "الاتجاهات المعاصرة في علم النفس" من خلال النجاح في معايرة المفردات ضمن نماذج راش، أحادي وثنائي المعلمة. هذا سمح بتقدير دقيق لمعالم المفردات واستخراج دوال معلومات المفردة والاختبار. ورغم القيود الحاسوبية التي حالت دون إتمام التحليل للنماذج الأكثر تعقيدًا (3PL & 4 PL)، إلا أن اختيار النموذج الثنائي المعلمة وتطبيقه بكفاءة ضمن إطار المعايرة المتزامنة قدَّم بنك أسئلة مُعايرًا بجودة عالية، وجاهزًا للاستخدام مع خصائص سيكومترية موثوقة.

وأظهرت النتائج فعالية ملحوظة لاستخدام المعايرة المتزامنة في بناء بنك الأسئلة، خاصةً مع النماذج الأقل تعقيدًا في نظرية الاستجابة للمفردة .(IRT) فقد نجحت المعايرة المتزامنة، التي أُجريت باستخدام حزمة mirt، في إتمام تقدير معالم المفردات بكفاءة لكل من نموذج راش، النموذج أحادي المعلمة، والنموذج ثنائي المعلمة .هذا يعكس قدرة هذه المنهجية على معالجة البيانات واستخلاص الخصائص السيكومترية الأساسية للمفردات، وهو ما يتوافق مع ما ذكره (Joo et al., 2017, p. 83)

بأن المعايرة المتزامنة تُظهر فعاليتها في بناء وتطوير بنوك أسئلة كبيرة بفضل دقتها في تقدير المعلمات. هذه الدقة ضرورية لدعم الاختبارات المستمرة وتوفير بيانات قياس موثوقة.

ومع ذلك، كشفت النتائج عن قيود عملية تتعلق بتطبيق المعايرة المتزامنة على النماذج الأكثر تعقيدًا. فقد واجه النموذجان الثلاثي والرباعي المعلمة صعوبات في إكمال التحليل، وهو ما يُعزى إلى المشكلات الحسابية المرتبطة بضخامة حجم البيانات ومتطلباتها المعالجة. وتُشير رسائل الخطأ التي ظهرت إلى تحديات في تخصيص الذاكرة والتعامل مع الأنظمة المعقدة حسابيًا، مما يؤكد أن فعالية المعايرة المتزامنة، رغم دقتها، قد تكون مقيدة بالموارد الحاسوبية وحجم البيانات عند تطبيق نماذج IRT ذات المعلمات المتعددة. هذه التحديات ليست فريدة من نوعها في هذا البحث، حيث يُقر (Kim, 2019, pp. 512-513) بوجود صعوبات عامة في تقدير معلمات IRT، لا سيما تلك المرتبطة بتوصيف المتغيرات الكامنة وغياب نظام إحداثي معياري.

وعلى الرغم من هذه القيود، فإن العملية قادت إلى اختيار النموذج الثنائي المعلمة (2PLM) باعتباره الأنسب لبنك الأسئلة، بناءً على تحليلات ملاءمة المفردة والنموذج، وهذا الاختيار يتماشى مع فكرة أن المعايرة المتزامنة تميل إلى أن تكون أدق عندما يكون النموذج مناسبًا للبيانات، كما أوضح (Lee & Lee, 2018, p. 644) وبالتالي، فإن فعالية المعايرة المتزامنة هنا لا تكمن فقط في قدرتها على إجراء التحليل، بل في توجيهها نحو النموذج الأكثر ملاءمة الذي يُمكن من بناء بنك أسئلة ذي خصائص سيكومترية عالية الجودة. إن استخدام نماذج IRT، مثل 2PLM، يوفر وصفًا أكثر تفصيلاً للمفردات مقارنة بالإحصائيات الكلاسيكية، مما يعزز من قيمة بنك الأسئلة (Hambelton & Swaminthan, 1985, p. 257) ويُسهم في دقة التقييم.(Tan, 2024, p. 33)

نتائج السؤال الثاني "ما هو النموذج الأنسب من بين نماذج نظرية الاستجابة للمفردة أحادية البعد التي أتمت التحليل بنجاح (راش، أحادي المعلمة، ثنائي المعلمة) لبيانات بنك أسئلة مقرر "الاتجاهات المعاصرة في علم النفس"؟"

لتحديد النموذج الأنسب من بين النماذج التي أتمت التحليل بنجاح (راش، أحادي المعلمة، وثنائي المعلمة)، تم الاعتماد على عدة مؤشرات إحصائية لتقييم ملاءمة النموذج للبيانات، بما في ذلك عدد المفردات الملائمة بناءً على اختبار مربع كاي ($\chi 2$)، ومؤشرات معيار معلومات أكايكي (AIC)، ومعيار معلومات بايزي (BIC)، ومعيار معلومات هانان–كوين (HQ)، بالإضافة إلى قيمة الأرجحية (logLik) وقيم مربع كاي ($\chi 2$) الكلية مع درجات الحرية وقيمة الاحتمالية (p-value)

١. ملاءمة المفردات (Item Fit) بناءً على اختبار مربع كاي($\chi 2$):

- **نموذج راش (Rasch): أظهر ٢٣٣ مفردة ملاءمة جيدة للنموذج.**
- **النموذج أحادي المعلم (1PL): أظهر ٢٣٢ مفردة ملاءمة جيدة للنموذج.**

- **النموذج ثنائي المعلمة (2PL): أظهر ٣٥١ مفردة ملاءمة جيدة للنموذج.**

جدول (٢) ملخص تحليل ملائمة المفردات للنماذج التي أتمت التحليل بنجاح

النموذج المختبر	عدد المفردات الملائمة للنموذج	النسبة المئوية
نموذج راش	٢٣٣	% ٥٠,٤
النموذج أحادي المعلمة	٢٣٢	% ٥٠,٢
النموذج ثنائي المعلمة	٣٥١	% ٧٦

تُشير هذه النتائج بجدول (٢) بوضوح إلى أن النموذج ثنائي المعلمة هو الأكثر ملاءمة للمفردات، حيث إن عدد المفردات التي أظهرت ملاءمة جيدة له (٣٥١ مفردة) يفوق بكثير العدد في النموذجين الآخرين. هذا يعني أن هذا النموذج يصف استجابات المفحوصين على عدد أكبر من المفردات بشكل أفضل مقارنة بنموذجي راش وأحادي المعلمة.

٢. ملاءمة النموذج الكلية (Overall Model Fit) بناءً علىAIC ، BIC، Log-likelihood:

جدول (٣) ملاءمة النموذج الكلية (Overall Model Fit)

النموذج	مؤشر الملاءمة							
	AIC	SABIC	HQ	BIC	logLik	χ2	df	p
راش	394543.2	395415.5	395427.1	396886.2	-196809	-	-	-
أحادي المعلم	394541.8	395414.1	395425.7	396884.8	-196808	1.404	0	NaN
ثنائي المعلم	384101.7	385842.6	385865.6	388777.6	-191127	11362.09	461	0

عند مقارنة النماذج، تُفضل القيم الأقل لمؤشرات AIC، BIC، وHQ، والقيم الأعلى للأرجحية (logLik)، لأنها تُشير إلى نموذج أكثر اقتصادًا وأفضل ملاءمة للبيانات.

تُظهر النتائج بجدول (٣) بوضوح أن النموذج ثنائي المعلمة (2PL) يتفوق بشكل كبير على النموذجين الآخرين في جميع هذه المؤشرات:

- **لديه أقل قيم (384101.7) AIC، (385842.6) SABIC، (385865.6) HQ، و BIC (388777.6)**
- **لديه أعلى قيمة للأرجحية (−١٩١١٢٦.٨)، مما يُشير إلى أنه يفسر البيانات بشكل أفضل.**

علاوة على ذلك، تُظهر قيمة مربع كاي (χ2) الكبيرة والدالة إحصائياً (١١٣٦٢.٠٩٣ عند ٤٦١ درجة حرية و p=0) أن النموذج ثنائي المعلمة يُقدم تحسناً دالاً إحصائياً في ملاءمة البيانات مقارنةً بالنموذجين الأبسط عن طريق إجراء تحليل التباين (ANOVA) (مع الأخذ في الاعتبار أن قيمة p=0 تعني دلالة عالية جداً، وتشير إلى أن التحسن في الملاءمة ليس مجرد صدفة).

تحديد النموذج الأنسب:

بناءً على مقاييس ملاءمة المفردات وملاءمة النموذج الكلية، يتضح أن النموذج الثنائي المعلمة (2PLM) هو النموذج الأنسب لبيانات بنك أسئلة مقرر "الاتجاهات المعاصرة في علم النفس " فهو لا يصف استجابات عدد أكبر من المفردات بشكل جيد فحسب، بل يُقدم أيضًا ملاءمة

إحصائية كلية أفضل بكثير للبيانات مقارنةً بنموذجي راش وأحادي المعلمة. هذا يؤكد أن الأخذ بعين الاعتبار كل من معلمة الصعوبة ومعلمة التمييز (التي يقدرها النموذج الثنائي المعلمة) ضروري لوصف خصائص مفردات هذا البنك بدقة أكبر وتقديم قياس أكثر فعالية. هذا يتوافق مع الأدبيات التي تُشير إلى أن النموذج 2PLM غالبًا ما يوفر تمثيلاً أكثر واقعية للبيانات في الاختبارات التي تتباين فيها قوة تمييز المفردات.

إن اختيار النموذج الثنائي المعلمة (2PLM) باعتباره الأنسب لبيانات بنك أسئلة مقرر "الاتجاهات المعاصرة في علم النفس" يمثل نتيجة محورية تتسق مع العديد من الدراسات السابقة في مجال القياس النفسي والتربوي. وتُشير النتائج بوضوح إلى تفوق النموذج الثنائي المعلمة على نموذجي راش وأحادي المعلمة، سواء من حيث عدد المفردات الملائمة فرديًا أو من خلال مؤشرات ملاءمة النموذج الكلية (AIC, BIC, logLik)

يُعزى هذا التفوق إلى قدرة النموذج الثنائي المعلمة على تقدير معلمتين للمفردة هما الصعوبة (b-parameter) والتمييز (a-parameter) ففي حين يفترض نموذج راش والنموذج أحادي المعلمة أن جميع المفردات لها قوة تمييز متساوية (عادةً ما تُفترض عند قيمة معينة أو تُجمد عند ١)، يسمح النموذج الثنائي المعلمة لكل مفردة بامتلاك معامل تمييز خاص بها. هذه المرونة الإضافية تُعد حاسمة، خاصةً في بنوك الأسئلة الكبيرة التي تحتوي على مفردات متنوعة. إن الاختلافات في قدرة المفردات على التمييز بين الأفراد ذوي القدرات العالية والمنخفضة هي سمة طبيعية لأي اختبار جيد، وبالتالي فإن نموذجًا يأخذ هذه الاختلافات في الاعتبار يُقدم تمثيلاً أكثر واقعية ودقة لخصائص المفردات. وتتفق هذه النتيجة مع ما يؤكده (Hambelton & Swaminthan, 1985, p. 257) بأن معلمات مفردة نموذج الاستجابة للمفردة (خاصة تلك التي تُقدر التمييز) ذات استخدام أكبر بكثير في وصف مفردات الاختبار في بنك الأسئلة مقارنة بالإحصائيات الكلاسيكية.

كما يعزز اختيار النموذج الثنائي المعلمة من جودة القياس الكلية للبنك. فقدرة النموذج على تقديم ملاءمة إحصائية أفضل للبيانات الكلية، كما دلت عليها القيم الأقل لمؤشرات AIC و BIC والقيمة الأعلى لـlogLik ، تُشير إلى أن النموذج 2PLM يفسر التباين في استجابات الطلاب بشكل أكثر كفاءة. هذا يضمن أن التقديرات الناتجة عن البنك (سواء لخصائص المفردات أو قدرات الأفراد) ستكون أكثر دقة وموثوقية. يرتبط هذا بشكل مباشر بأهمية دقة التقييم في تحديد ثبات الاختبار، كما ذكر (Tan, 2024, p. 33) فعندما يصف النموذج البيانات بشكل أفضل، تُصبح المعلومات المستقاة منه أكثر فائدة في تحسين جودة القياس.

وعلى الرغم من أننا لم نتمكن من تحليل النماذج الأكثر تعقيدًا (الثلاثي والرباعي) بسبب قيود حسابية، فإن اختيار النموذج 2PLM يمثل حلاً عمليًا ومناسبًا يوفر توازنًا ممتازًا بين بساطة

النموذج وقدرته على تمثيل البيانات المعقدة. هذا القرار لا يعكس فقط النتائج الإحصائية، بل أيضًا الاعتبارات التطبيقية عند التعامل مع بنوك أسئلة ضخمة. بشكل عام، تُظهر هذه النتائج أن المعايرة المتزامنة، عند تطبيقها مع النموذج المناسب، تُسهم بفعالية في بناء بنك أسئلة ذي خصائص سيكومترية قوية يمكن الاعتماد عليها في تقييم تحصيل الطلاب في مقرر الاتجاهات المعاصرة في علم النفس.

نتائج السؤال الثالث " ما هي الخصائص السيكومترية النهائية لبنك أسئلة مقرر "الاتجاهات المعاصرة في علم النفس" في ضوء النموذج المختار (النموذج ثنائي المعلمة)، بعد تنقية المفردات والأفراد؟"

في إطار السعي نحو تطوير أدوات تقييم دقيقة وموثوقة في المجال التربوي والنفسي، يبرز تحليل الخصائص السيكومترية لبنوك الأسئلة كخطوة جوهرية لضمان جودة الاختبارات وقدرتها على قياس السمات المستهدفة بدقة. وفي هذا السياق، يأتي هذا السؤال لاستكشاف الخصائص النهائية لبنك أسئلة مقرر *"الاتجاهات المعاصرة في علم النفس "*بعد تطبيق النموذج ثنائي المعلمة (2PL) من نظرية الاستجابة للفقرة (IRT)، والذي يُعد أحد النماذج الأكثر ملاءمةً لتحليل المفردات ذات الإجابة الثنائية (١/٠).

تتمثل أهمية هذا التحليل في كشفه عن قدرة البنك على التمييز بين الأفراد ذوي المستويات المختلفة من الكفاءة، وتحديد مدى توازن صعوبة مفرداته، فضلًا عن تقييم كفاءته الإحصائية عبر مؤشرات مثل دالة المعلومات (TIF) والخطأ المعياري للقياس (SEM) وقد شملت عملية التنقية استبعاد الفقرات ذات التمييز الضعيف والمفردات غير الملائمة للنموذج، وكذلك الأفراد ذوي الاستجابات غير المتسقة، مما يعزز مصداقية النتائج.

سنعرض الإجابة تفصيلًا للخصائص السيكومترية النهائية للبنك بعد التنقية، مع التركيز على دلالات معاملات التمييز والصعوبة، وتوزيع القدرات، وجودة القياس على مستوى الاختبار ككل، مدعومةً بالرسوم البيانية التوضيحية. هذه الخطوة تُعتبر حاسمة في تحديد مدى صلاحية البنك للاستخدام في التقييمات النفسية والتربوية، وتقديم توصيات لتطويره مستقبلًا.

جدول (٤) نتائج تقدير معلمات العناصر لنموذج IRT ثنائي المعلمة (2PL) المفردات من (Q1 إلى Q58_2)

المفردات	التمييز	الصعوبة	المفردات	التمييز	الصعوبة	المفردات	التمييز	الصعوبة
Q1	0.455	-4.862	Q51	0.515	-4.139	Q4_2	0.955	-2.865
Q2	0.328	-10.550	Q53	0.503	-4.213	Q8_2	1.450	-3.776
Q3	0.469	-8.488	Q54	0.692	-1.636	Q9_2	1.535	-2.863
Q6	0.641	-5.616	Q57	0.466	0.523	Q11_2	0.071	-9.335
Q7	0.637	-4.148	Q59	0.414	-7.681	Q13_2	1.120	-2.291
Q8	0.519	-7.633	Q60	0.383	-6.301	Q15_2	1.351	-1.238
Q9	0.230	-6.205	Q61	0.415	-4.309	Q17_2	1.004	-3.057
Q11	0.324	-7.237	Q63	0.224	-5.515	Q18_2	0.736	-1.963
Q12	0.538	-5.046	Q64	0.381	-2.743	Q19_2	0.648	-2.750
Q14	0.518	-6.517	Q67	0.541	-5.044	Q21_2	1.324	-2.885
Q15	0.436	-6.262	Q69	0.655	-3.071	Q22_2	1.282	-3.003
Q16	0.504	-4.464	Q70	0.084	-26.125	Q23_2	0.934	-3.732
Q17	0.657	-6.419	Q71	0.677	-4.036	Q24_2	1.196	-2.894
Q18	0.628	-6.765	Q72	0.251	-6.953	Q25_2	1.222	0.146
Q19	0.832	-4.915	Q73	0.089	-9.994	Q27_2	1.247	-3.593
Q20	0.523	-1.257	Q74	0.564	-3.395	Q28_2	1.347	-3.355
Q21	0.581	-3.975	Q75	0.363	-3.921	Q29_2	1.342	-3.363
Q22	0.611	-5.708	Q76	0.690	-5.743	Q32_2	1.424	-1.033
Q23	0.799	-2.140	Q77	0.417	-1.644	Q33_2	0.978	-2.916
Q24	0.182	-7.043	Q78	0.346	-1.830	Q34_2	1.029	-2.073
Q25	0.905	-1.809	Q79	0.306	-5.277	Q35_2	0.970	-3.070
Q27	0.727	-3.107	Q80	0.133	-19.388	Q37_2	1.343	-2.877
Q28	0.205	-7.804	Q81	0.153	-20.690	Q38_2	1.082	-0.543
Q29	0.491	-6.074	Q82	0.551	-0.287	Q40_2	0.796	-3.239
Q30	0.997	-6.129	Q83	-0.050	45.324	Q42_2	1.508	-1.718
Q31	0.089	-7.796	Q84	0.473	-9.730	Q43_2	1.116	-3.487
Q32	0.350	-2.846	Q85	0.507	-3.297	Q44_2	1.364	-2.586
Q33	0.666	-6.189	Q86	0.119	-23.485	Q45_2	0.883	-2.850
Q35	0.438	-5.662	Q89	0.368	-8.455	Q46_2	1.518	-1.223
Q36	0.120	-2.452	Q90	0.508	-8.078	Q47_2	1.061	-1.228
Q37	0.524	-2.938	Q92	0.383	-9.519	Q48_2	1.474	-1.593
Q39	0.915	-2.051	Q93	0.655	-4.976	Q50_2	1.194	-0.469
Q41	0.613	-3.360	Q94	0.826	-2.646	Q51_2	1.621	-3.684
Q42	1.263	-0.588	Q95	0.765	-3.958	Q52_2	1.309	-3.208
Q43	0.717	-0.920	Q96	0.424	-7.972	Q53_2	1.394	-2.278
Q46	0.740	-2.384	Q99	0.186	-7.218	Q54_2	0.740	-3.860
Q47	0.854	-2.005	Q100	0.546	-7.663	Q55_2	1.564	-3.059
Q48	0.705	-4.007	Q1_2	1.246	-3.433	Q56_2	1.118	-3.208
Q49	0.568	-4.054	Q2_2	1.648	-2.881	Q57_2	1.640	-3.416
Q50	0.635	-5.307	Q3_2	1.117	-1.491	Q58_2	1.030	-0.394

ملحوظة: ترقيم المفردات (Q1, Q2,…Q100) للاختبار الأول، (Q1_2, Q2_2,…Q100_2) للاختبار الثاني، (Q1_3, Q2_3,…Q100_3) للاختبار الثالث، (Q1_4, Q2_4,…Q100_4) للاختبار الرابع، (Q1_5, Q2_5,…Q100_5) للاختبار الخامس، والمفردات غير الموجودة هي المفردات المحذوفة.

جدول (٥) نتائج تقدير معلمات العناصر لنموذج IRT ثنائي المعلمة (2PL) المفردات من (Q59_2 إلى Q38_4)

المفردات	التمييز	الصعوبة	المفردات	التمييز	الصعوبة	المفردات	التمييز	الصعوبة
Q59_2	1.550	-3.039	Q28_3	1.881	-2.647	Q82_3	1.579	-0.317
Q60_2	1.430	-3.105	Q29_3	0.958	-0.534	Q83_3	1.005	-2.341
Q61_2	1.423	-2.736	Q30_3	0.134	0.002	Q84_3	1.200	-3.157
Q62_2	1.173	-3.485	Q31_3	0.684	-1.421	Q85_3	1.381	-2.005
Q63_2	0.745	-2.352	Q32_3	1.273	-2.859	Q86_3	1.023	-3.295
Q66_2	1.073	-3.564	Q33_3	1.808	-2.326	Q88_3	1.588	-1.188
Q68_2	0.672	-0.487	Q34_3	1.263	-2.509	Q89_3	0.977	-2.969
Q70_2	1.504	-1.063	Q36_3	0.863	-1.946	Q91_3	0.680	-1.546
Q71_2	1.344	-2.095	Q37_3	1.424	-3.223	Q92_3	1.052	-1.171
Q72_2	0.903	-4.164	Q40_3	1.957	-1.071	Q93_3	0.897	-3.460
Q73_2	0.860	-2.751	Q41_3	1.834	-2.628	Q94_3	0.690	-3.784
Q74_2	0.754	-1.966	Q43_3	0.857	-2.279	Q95_3	0.986	-2.979
Q75_2	1.157	-3.330	Q44_3	1.250	-0.691	Q96_3	0.806	-2.584
Q77_2	0.735	-3.976	Q45_3	0.979	-1.412	Q97_3	0.819	-0.516
Q78_2	1.396	-1.749	Q46_3	1.500	-1.740	Q98_3	0.700	-0.374
Q80_2	0.663	-4.249	Q47_3	1.222	-1.586	Q99_3	0.791	-3.207
Q81_2	1.205	-3.778	Q48_3	0.745	-2.224	Q100_3	1.030	-3.107
Q2_3	0.703	-3.742	Q49_3	0.695	-2.082	Q4_4	3.413	-1.017
Q3_3	1.439	-1.690	Q50_3	1.180	-2.382	Q5_4	2.997	-1.041
Q4_3	1.128	-2.820	Q51_3	0.451	-4.923	Q6_4	2.880	-1.290
Q5_3	0.568	-3.193	Q52_3	1.541	-2.884	Q7_4	1.473	-1.158
Q6_3	0.413	0.670	Q54_3	1.353	-2.752	Q9_4	1.892	-1.365
Q7_3	1.574	-2.549	Q56_3	1.361	-2.538	Q10_4	1.403	-1.274
Q8_3	1.627	-2.672	Q59_3	1.360	-1.898	Q11_4	1.107	-0.021
Q9_3	0.930	0.177	Q60_3	1.165	-2.973	Q16_4	2.607	-1.379
Q10_3	1.312	-3.269	Q64_3	1.473	-0.958	Q17_4	3.109	-1.521
Q12_3	0.281	-0.397	Q65_3	0.950	-0.683	Q19_4	2.816	-0.040
Q13_3	1.188	-3.157	Q66_3	0.943	-3.361	Q21_4	0.360	-0.759
Q14_3	1.359	-3.186	Q67_3	1.173	-2.769	Q22_4	0.972	0.153
Q15_3	0.839	-2.178	Q68_3	0.645	-0.077	Q23_4	3.181	-1.355
Q16_3	1.174	-1.997	Q69_3	1.017	-1.213	Q24_4	2.159	-1.592
Q17_3	1.196	-3.091	Q71_3	0.675	-3.897	Q25_4	1.192	-1.443
Q18_3	1.261	-2.241	Q72_3	1.055	-2.152	Q27_4	1.664	-1.470
Q19_3	1.455	-2.301	Q73_3	1.216	-1.673	Q28_4	2.678	-1.320
Q21_3	0.901	-0.497	Q74_3	1.284	-3.256	Q30_4	2.710	-1.054
Q22_3	1.520	-2.331	Q76_3	1.063	-2.446	Q32_4	2.526	-1.512
Q23_3	0.907	-2.744	Q77_3	0.665	-1.154	Q34_4	3.037	-0.323
Q24_3	0.890	-2.441	Q78_3	1.154	-3.255	Q35_4	1.240	-1.035
Q25_3	0.871	-0.567	Q80_3	0.945	-0.407	Q36_4	1.041	-1.310
Q27_3	1.624	-1.462	Q81_3	1.235	-2.336	Q38_4	1.751	-0.277

ملحوظة: ترقيم المفردات (Q1, Q2,…Q100) للاختبار الأول، (Q1_2, Q2_2,…Q100_2) للاختبار الثاني، (Q1_3, Q2_3,…Q100_3) للاختبار الثالث، (Q1_4, Q2_4,…Q100_4) للاختبار الرابع، (Q1_5, Q2_5,…Q100_5) للاختبار الخامس، والمفردات غير الموجودة هي المفردات المحذوفة.

جدول (٦) نتائج تقدير معلمات العناصر لنموذج IRT ثنائي المعلمة (2PL) المفردات من (Q39_4 إلى Q100_5)

المفردات	التمييز	الصعوبة	المفردات	التمييز	الصعوبة	المفردات	التمييز	الصعوبة
Q39_4	1.722	-0.335	Q5_5	0.959	-1.179	Q57_5	1.232	-1.403
Q40_4	3.077	-1.519	Q7_5	0.515	-1.190	Q58_5	1.294	-2.618
Q42_4	0.591	-2.001	Q8_5	1.235	-1.973	Q60_5	0.352	-1.574
Q43_4	1.139	-1.679	Q9_5	1.472	-1.845	Q61_5	1.882	-1.596
Q44_4	2.164	-0.165	Q10_5	1.464	-2.213	Q62_5	0.705	-3.616
Q45_4	1.234	-0.730	Q13_5	1.955	-2.234	Q63_5	1.070	-2.481
Q46_4	1.402	-1.259	Q14_5	0.979	-0.889	Q64_5	0.521	-3.031
Q47_4	2.669	-1.352	Q15_5	1.106	-1.876	Q66_5	1.074	-2.004
Q48_4	2.792	-1.343	Q16_5	1.151	-1.693	Q67_5	1.141	-2.812
Q49_4	1.497	-1.583	Q18_5	0.966	-1.873	Q68_5	0.585	-0.587
Q50_4	2.658	-1.492	Q19_5	1.650	-2.489	Q70_5	1.006	-2.215
Q51_4	2.021	-1.406	Q20_5	2.269	-2.104	Q72_5	1.093	-3.083
Q55_4	2.719	-1.575	Q21_5	1.364	-2.756	Q73_5	1.288	-2.969
Q56_4	2.079	-1.720	Q22_5	1.219	-0.153	Q74_5	0.594	-3.142
Q58_4	0.344	1.145	Q23_5	1.808	-1.821	Q75_5	0.807	-0.872
Q62_4	2.195	-1.083	Q24_5	2.211	-2.080	Q77_5	1.598	-0.812
Q65_4	1.511	-1.348	Q25_5	2.322	-1.964	Q78_5	1.543	-1.260
Q68_4	1.289	-0.314	Q26_5	2.419	-1.937	Q80_5	0.435	-2.396
Q69_4	2.169	-0.892	Q27_5	1.635	-2.051	Q83_5	1.326	-0.430
Q71_4	2.131	-1.390	Q28_5	1.472	-2.649	Q85_5	1.216	-2.711
Q72_4	1.155	0.070	Q29_5	1.234	-2.277	Q88_5	0.655	-0.963
Q74_4	1.678	-1.286	Q30_5	0.381	-2.966	Q89_5	0.660	-1.261
Q75_4	1.267	-0.844	Q31_5	0.548	-2.830	Q90_5	0.200	-5.110
Q76_4	2.082	-1.182	Q34_5	1.778	-1.840	Q93_5	1.302	-2.676
Q79_4	1.532	-1.206	Q35_5	1.414	-1.353	Q94_5	1.229	-1.822
Q85_4	1.310	-0.017	Q37_5	1.226	-2.550	Q95_5	1.317	-1.604
Q86_4	1.558	-1.440	Q38_5	1.199	-2.249	Q96_5	0.498	-2.414
Q87_4	2.542	-0.578	Q39_5	1.418	-2.417	Q97_5	1.189	-2.050
Q88_4	2.259	-1.038	Q40_5	1.326	-1.987	Q98_5	0.514	-3.730
Q90_4	1.504	-1.210	Q41_5	0.737	-2.352	Q99_5	1.016	-1.606
Q91_4	1.910	-1.245	Q42_5	0.219	-1.497	Q100_5	0.235	-5.639
Q93_4	1.274	-1.445	Q43_5	0.593	-1.968	-	-	-
Q94_4	1.563	-0.173	Q44_5	1.035	-1.863	-	-	-
Q96_4	1.439	-0.973	Q45_5	0.606	-2.345	-	-	-
Q97_4	2.719	-1.557	Q46_5	0.870	-2.058	-	-	-
Q98_4	0.870	-1.395	Q47_5	1.977	-2.510	-	-	-
Q100_4	1.466	-1.267	Q48_5	1.199	-2.829	-	-	-
Q1_5	0.610	-1.712	Q53_5	0.940	-1.638	-	-	-
Q2_5	1.182	-2.097	Q55_5	0.359	-0.575	-	-	-
Q4_5	2.576	-2.352	Q56_5	1.213	-2.528	-	-	-

ملحوظة: ترقيم المفردات (Q1, Q2,...Q100) للاختبار الأول، (Q1_2, Q2_2,...Q100_2) للاختبار الثاني، (Q1_3, Q2_3,...Q100_3) للاختبار الثالث، (Q1_4, Q2_4,...Q100_4) للاختبار الرابع، (Q1_5, Q2_5,...Q100_5) للاختبار الخامس، والمفردات غير الموجودة هي المفردات المحذوفة.

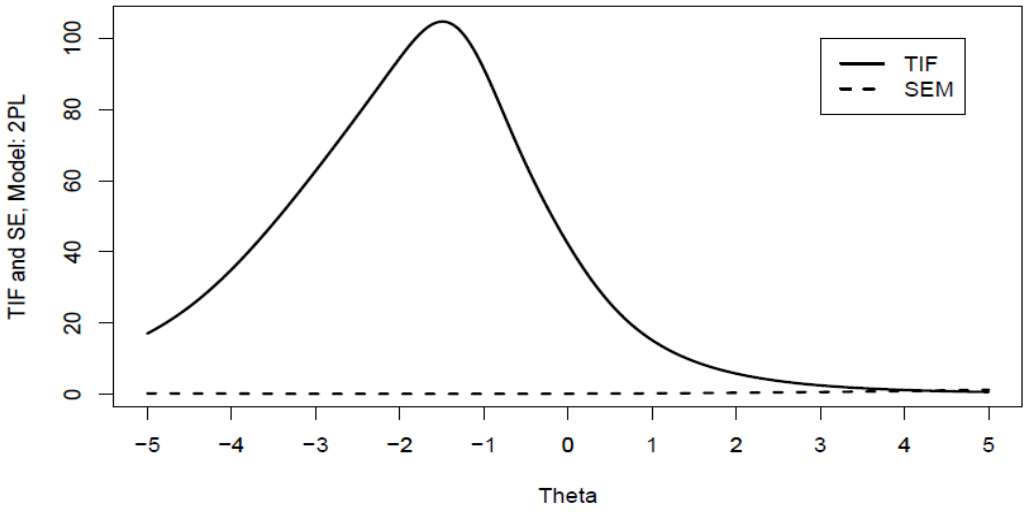
تقدم النتائج بالجداول (٤- ٥- ٦) تقدير معلمات المفردات لنموذج نظرية الاستجابة للعنصر ثنائي المعلمة (2PL IRT Model) بعد حذف المفردات غير الصالحة وكذلك الأفراد وإعادة التدريج مرة أخرى، وهو ما يُعد خطوة أساسية في بناء وتطوير الاختبارات المقننة. هذه النتائج، التي تشمل

معلمات التمييز (a) والصعوبة (b) لكل مفردة، تُمكننا من فهم خصائص كل مفردة وكيف تتفاعل مع قدرة الأفراد.

معلمة التمييز (a): وتُشير قيم "a" إلى مدى قدرة العنصر على التمييز بين الأفراد ذوي القدرات المختلفة. العناصر ذات قيم "a" المرتفعة (عادةً ما تكون أكبر من ١.٠) تُعد مميزة جدًا، مما يعني أنها تُميز بوضوح بين الأفراد ذوي القدرة العالية والمنخفضة. على سبيل المثال، العنصر Q4_4 (a = 3.41) يُظهر قدرة تمييزية عالية جدًا، بينما العنصر Q70 (a = 0.084) يُظهر قدرة تمييزية ضعيفة جدًا، وقد يُشير ذلك إلى أنه لا يساهم بشكل فعال في التمييز بين الأفراد. العناصر ذات التمييز المنخفض جدًا (مثل Q83 بقيمة سالبة -٠.٠٤٩) تُشير إلى أن الأفراد ذوي القدرة الأقل لديهم فرصة أكبر للإجابة الصحيحة، مما يستدعي مراجعتها أو حذفها أيضاً، وهو ما يفسر سبب حذف المفردات غير الصالحة مسبقًا، وقد تكون المفردة تلك أظهرت قيمة سالبة نظراً لإعادة عملية التدريج مرة أخرى.

معلمة الصعوبة (b): تُشير قيم "b" إلى مستوى القدرة اللازم للإجابة على العنصر بشكل صحيح بنسبة ٥٠٪. القيم السلبية الكبيرة (مثل Q2: b = -10.55) تُشير إلى عناصر سهلة جدًا، حيث يمكن للأفراد ذوي القدرة المنخفضة الإجابة عليها. على النقيض، القيم الإيجابية (مثل Q57: b = 0.523) تُشير إلى عناصر أكثر صعوبة. نطاق قيم "b" هنا يتراوح من شديد السلبية إلى إيجابية طفيفة، مما يُشير إلى أن معظم عناصر البنك تميل إلى أن تكون سهلة نسبيًا. هذا التوزيع للصعوبة يجب أن يتوافق مع الغرض من البنك والمستوى المستهدف للمفحوصين. إذا كان البنك يستهدف مستوى قدرة مرتفع، فإن وجود العديد من العناصر السهلة جدًا قد يحد من قدرته على قياس القدرات العليا بدقة. هذا التنوع في الصعوبة، بعد إعادة التدريج، يُعد أمرًا حيويًا لضمان تغطية جيدة لمدى القدرة المقاسة؛ وعلى ذلك فالبنك يحتاج إلى إضافة بنود تتميز بالصعوبة وهو الأمر الذي قد يحدث في السنوات المقبلة وإضافتها إلى البنك وعمل التدريج مرة أخرى لضمان تحديث مفردات البنك بصورة دورية، وهو ما يحقق الهدف من إنشاء البنك والحاجة لتطويره بصورة مستمرة.

وبشكل عام، تُقدم هذه النتائج نظرة شاملة على جودة العناصر المتبقية في البنك بعد عملية التنقية، مما يسمح بتحليل أعمق لأداء البنك الكلي وصلاحيته.



شكل (١)

ومن الشكل بياني (١) منحنى معلومات الاختبار (TIF) والخطأ المعياري للقياس (SEM) في نفس المقياس لنموذج IRT ثنائي المعلمة (2PL) يتضح الآتي:

- **المحور الأفقي (Theta): يمثل مستوى القدرة (theta) للمفحوصين، ويتراوح من −٥ إلى +٥ .يُعتبر مستوى القدرة (صفر) نقطة متوسطة، والقيم السالبة تشير إلى قدرة أقل من المتوسط، بينما القيم الموجبة تشير إلى قدرة أعلى من المتوسط .**
- **المحور الرأسي (TIF and SE): يمثل قيمة معلومات الاختبار (Total Information Function - TIF) والخطأ المعياري للقياس (Standard Error of Measurement - SEM).**

تحليل المنحنيات:

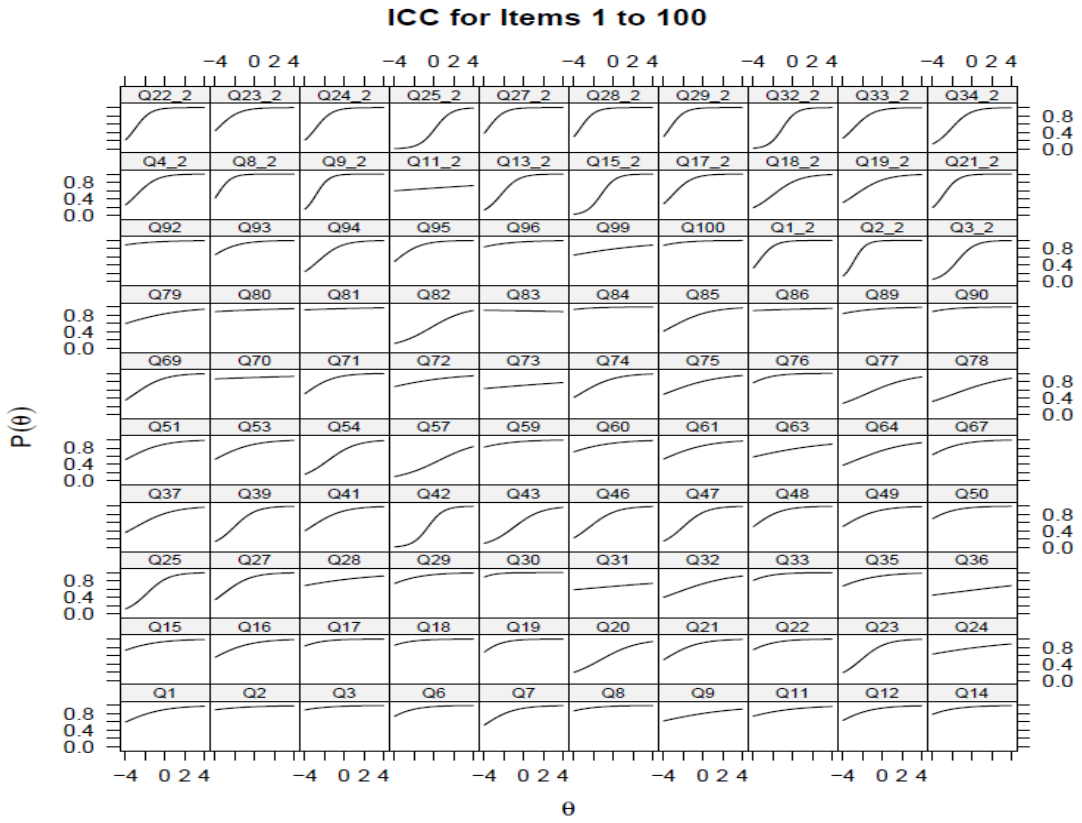
منحنى معلومات الاختبار (TIF – الخط المتصل):

- **يُظهر هذا المنحنى أن الاختبار يوفر أقصى قدر من المعلومات (أعلى دقة قياس) عند مستوى قدرة يقع تقريبًا بين −٢ و −١ على مقياس Theta**
- **تصل قمة منحنى المعلومات إلى حوالي ١٠٠ عند هذه النقطة .**
- **كلما ابتعدنا عن هذه النقطة (سواء نحو القدرة المنخفضة جدًا أو العالية جدًا)، تتناقص معلومات الاختبار بشكل كبير .هذا يعني أن الاختبار يكون أقل دقة في قياس الأفراد ذوي القدرات المنخفضة جدًا (أقل من −٣) والعالية جدًا (أعلى من ٢) .**
- **هذا النمط يتسق مع الاختبارات التي تركز على تمييز الأفراد في نطاق معين من القدرة، والذي يبدو أنه يميل قليلاً نحو القدرات الأقل من المتوسط في هذا الاختبار .**

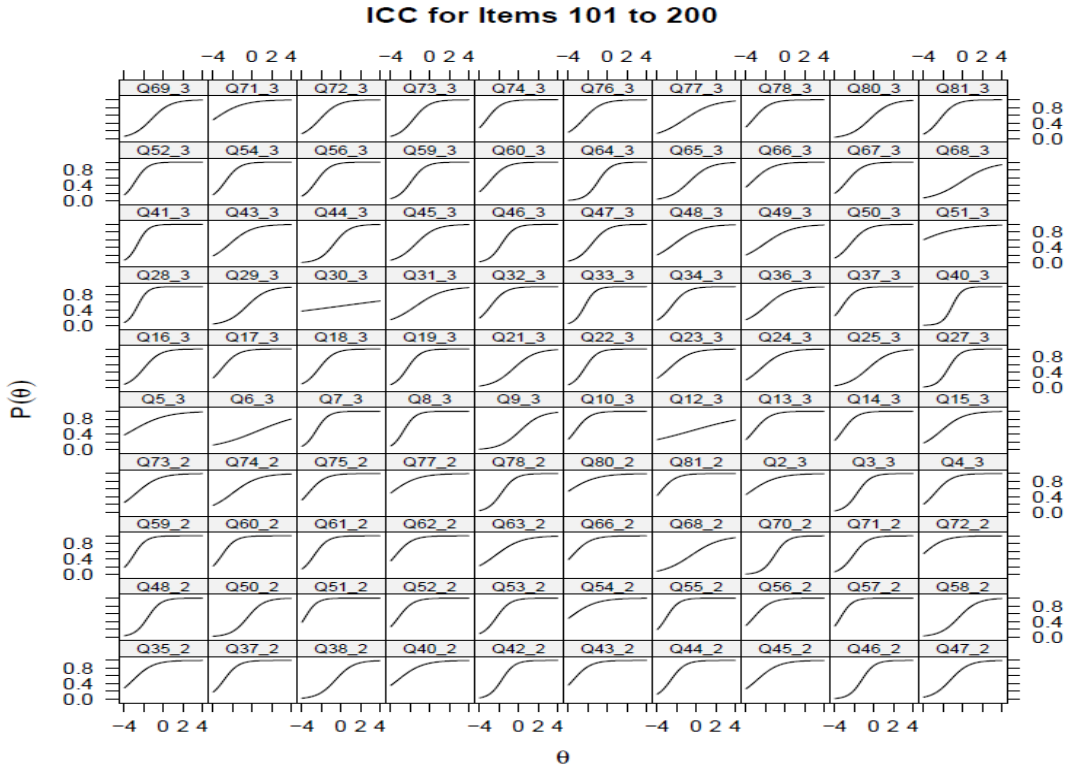
منحنى الخطأ المعياري للقياس (SEM – الخط المتقطع):

- **يُمثل هذا المنحنى الخطأ المعياري المرتبط بتقدير قدرة الفرد عند مستويات مختلفة من Theta**
- **العلاقة بين TIF و SEMهي علاقة عكسية: حيث تكون معلومات الاختبار عالية، يكون الخطأ المعياري منخفضًا، والعكس صحيح .**
- **يُظهر الرسم أن الخطأ المعياري للقياس منخفض جدًا (قريب من الصفر) في النطاق الذي يوفر فيه الاختبار أقصى معلومات (تقريبًا بين −٢ و −١) .**
- **يزداد الخطأ المعياري بشكل طفيف كلما ابتعدنا عن هذا النطاق، ولكنه يظل منخفضًا جدًا عبر معظم مقياس Theta**
- **يشير انخفاض قيم SEM بشكل عام إلى أن تقديرات القدرة التي يوفرها الاختبار موثوقة إلى حد كبير في نطاق القدرة الذي يغطيه الاختبار بشكل جيد .**

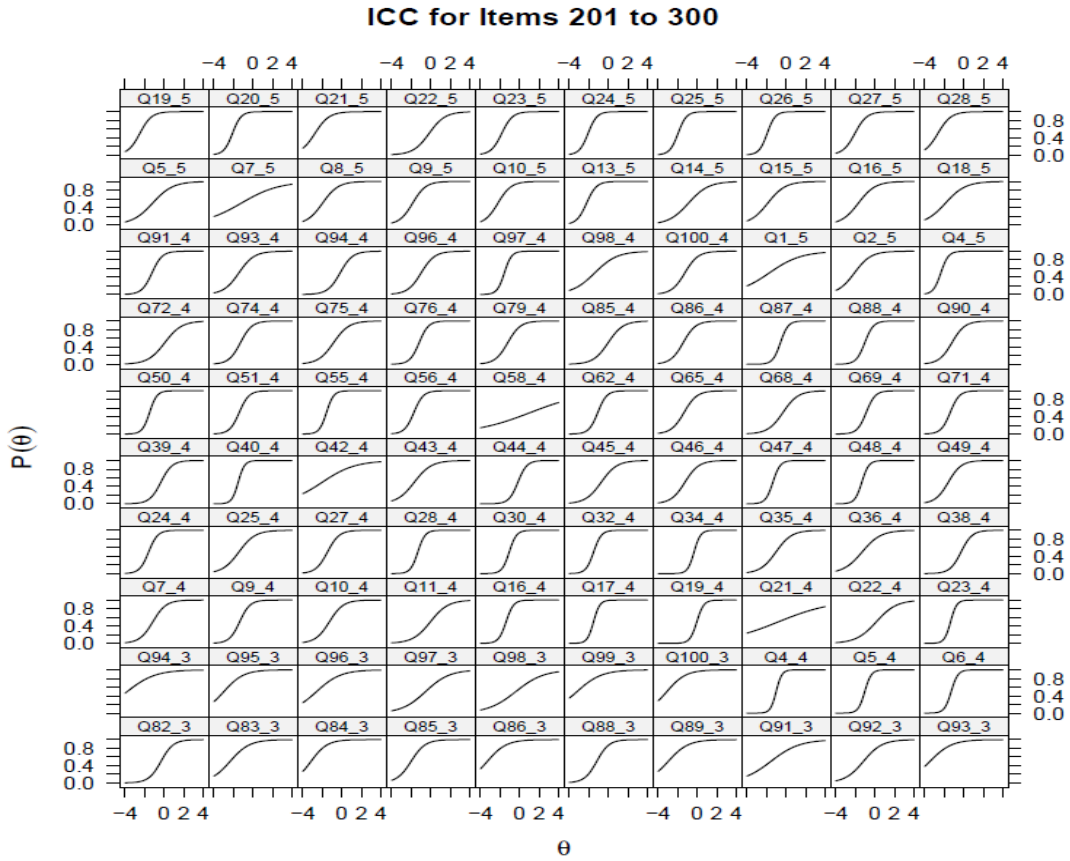
ويُشير الشكل (١) إلى أن هذا الاختبار فعال للغاية في قياس الأفراد الذين تقع قدراتهم ضمن نطاق معين (متمركز حول −١.٥)؛ حيث يوفر أقصى دقة قياس وأقل خطأ معياري .خارج هذا النطاق، تقل دقة القياس، مما قد يعني أن الاختبار قد لا يكون مناسبًا تمامًا للأفراد ذوي القدرات شديدة الانخفاض أو شديدة الارتفاع .



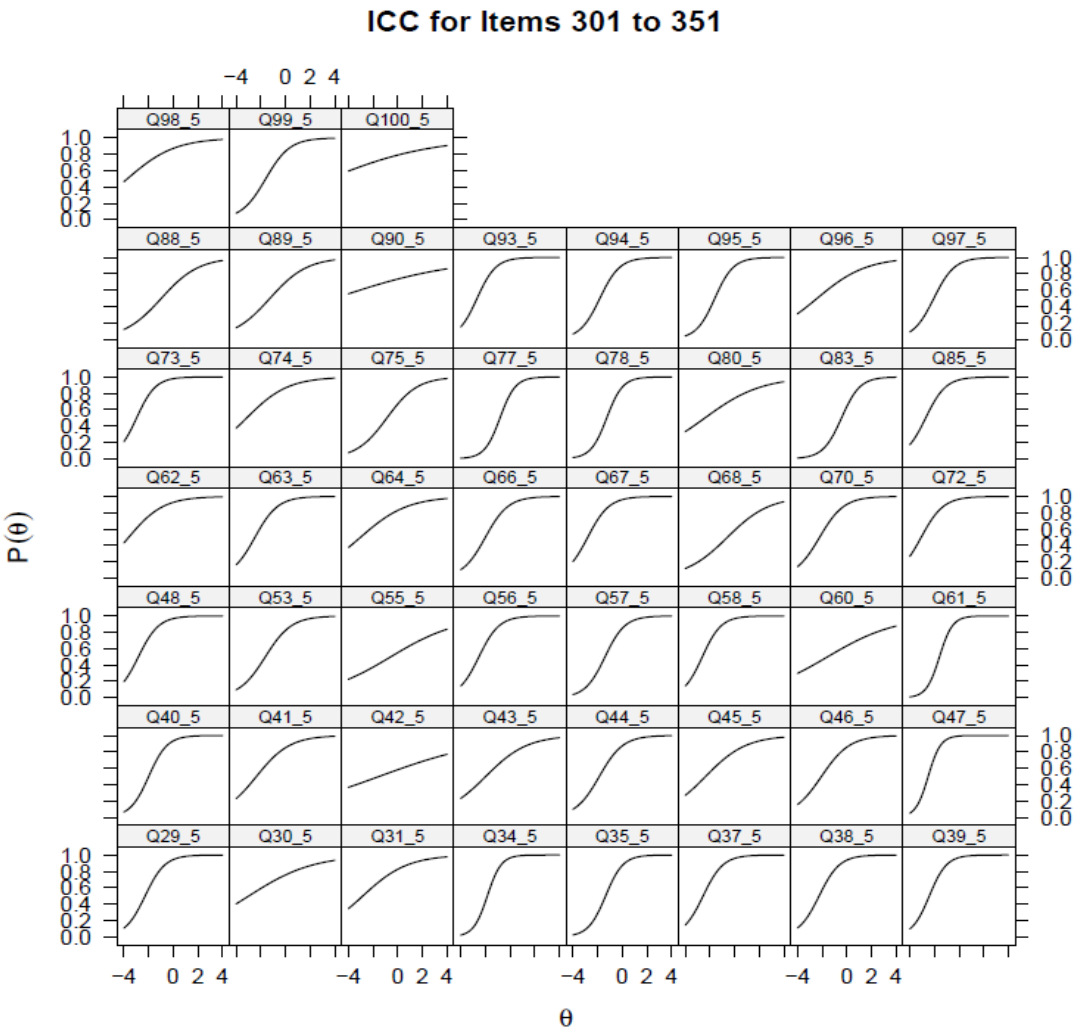
شكل (٢) منحنيات ICC للعناصر ١ إلى ١٠٠



شكل (٣) منحنيات ICC للعناصر ١٠١ إلى ٢٠٠



شكل (٤) منحنيات ICC للعناصر ٢٠١ إلى ٣٠٠



شكل (٥) منحنيات ICC للعناصر ٣٠١ إلى ٣٥١

ملحوظة: ترقيم المفردات (Q1, Q2,...Q100) للاختبار الأول، (Q1_2, Q2_2,...Q100_2) للاختبار الثاني، (Q1_3, Q2_3,...Q100_3) للاختبار الثالث، (Q1_4, Q2_4,...Q100_4) للاختبار الرابع، (Q1_5, Q2_5,...Q100_5) للاختبار الخامس، والمفردات غير الموجودة هي المفردات المحذوفة.

المنحنيات المميزة للعنصر (ICC) لمجموعات العناصر المختلفة:

تُعد المنحنيات المميزة للعنصر (ICC) من الرسوم البيانية المحورية في نظرية الاستجابة للعنصر(IRT) ، حيث تُجسد العلاقة بين مستوى القدرة الكامنة للفرد (Theta) واحتمال إجابته على عنصر ما بشكل صحيح. في النموذج ثنائي المعلمة(2PL) ، تُحدد هذه العلاقة من خلال معلمتي التمييز (a) والصعوبة (b) لكل عنصر. يُقدم هذا التحليل نظرة عامة على خصائص منحنيات ICC لمجموعات مختلفة من العناصر، مع الأخذ في الاعتبار أن كل مجموعة تُشير إلى جزء من البنك ومجموعها هو مجموع أسئلة البنك.

١.منحنيات ICC للعناصر ١ إلى ١٠٠ و١٠١ إلى ٢٠٠ شكل (٢ & ٣)

تُمثل هذه المجموعة الجزء الأول من البنك، وتُقدم رؤى مبدئية حول خصائص المفردات. بالنظر إلى هذا النطاق من العناصر، نُلاحظ نمطاً واضحاً. غالبية منحنيات ICC لهذه العناصر تميل إلى أن تكون مُزاحة نحو الجانب السلبي من مقياس القدرة (Theta)، مما يُشير إلى أن معظم هذه العناصر تُعد سهلة نسبياً. تقع نقاط المنتصف (التي تُمثل معلمة الصعوبة 'b') لهذه المنحنيات عند قيم Theta سالبة كبيرة. على سبيل المثال، نجد عناصر مثل Q2 بصعوبة −١٠.٥٥ وQ٣ بصعوبة −٨.٤٩.

وفيما يتعلق بالتمييز (انحدار المنحنى)، تُظهر بعض هذه العناصر انحداراً معتدلاً إلى حاداً، مما يعني أنها تُميز بشكل جيد بين الأفراد في نطاق صعوبتها. ومع ذلك، فإن التركيز الشديد لمعظم هذه العناصر على نطاق القدرات المنخفضة جداً يُشير إلى أن هذا الجزء من البنك قد يكون فعالاً بشكل أساسي في التمييز بين الأفراد ذوي القدرات المتدنية وتحديد أولئك الذين يمتلكون الحد الأدنى من الكفاءة. قد لا تُقدم هذه المجموعة من العناصر معلومات كافية أو دقيقة للأفراد ذوي القدرات المتوسطة أو العالية.

٢.منحنيات ICC للعناصر ٢٠١ إلى ٣٠٠ شكل (٤)

من المُرجح أن تُظهر هذه المجموعة تحولاً نحو صعوبات أكثر اعتدالاً .عادةً، في الاختبارات الطويلة أو متعددة الأقسام، تُوزع الصعوبات عبر الأجزاء لضمان تغطية واسعة لمدى القدرة. بناءً على الأنماط التي بدأت تظهر في العناصر اللاحقة، يُتوقع أن تُظهر منحنيات ICC في هذا النطاق نقاط منتصف أقرب إلى الصفر على مقياس Theta

وهذا يعني أن هذه العناصر أكثر صعوبة من المجموعة الأولى، وتُساهم بشكل أكبر في قياس والتمييز بين الأفراد ذوي القدرات المتوسطة .قد تُظهر هذه العناصر أيضاً تمييزاً جيداً (منحدرات حادة)، مما يُعزز من قدرتها على تقديم معلومات دقيقة في النطاق المتوسط من القدرة. هذا التحول ضروري لضمان أن الاختبار لا يُقلل من دقة قياس الأفراد الأكثر قدرة، وهو مبدأ تدعمه الدراسات التي تُشدد على أهمية توزيع الصعوبة لتغطية مدى القدرة المستهدف بكفاءة (Hambleton & Swaminathan, 2013).

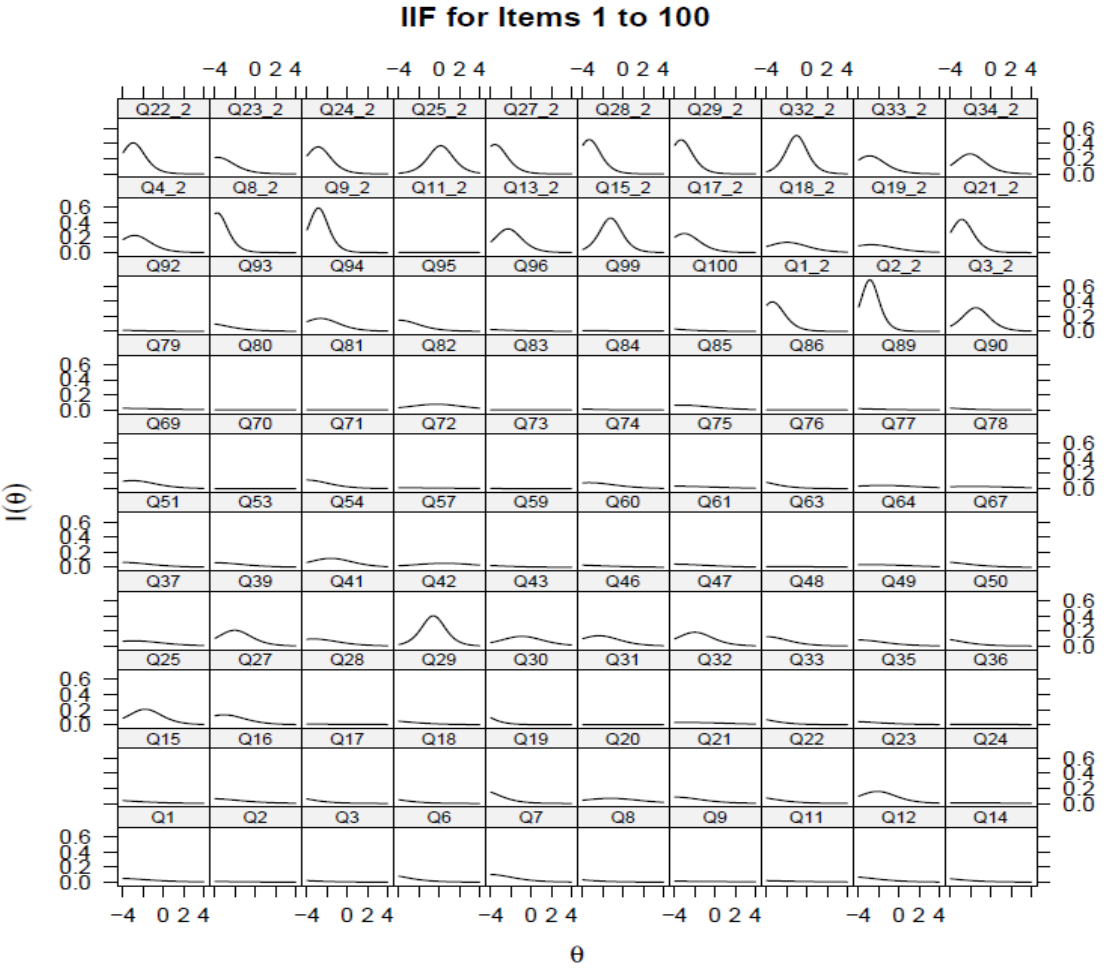
٣.منحنيات ICC للعناصر ٣٠١ إلى ٣٥١ شكل (٥)

تُمثل هذه المجموعة الجزء الأخير من البنك، وبناءً على منطق تصميم الاختبارات وتطورها بعد عمليات التنقية وإعادة التدريج، يُتوقع أن تُظهر هذه العناصر أوسع نطاق من الصعوبات، بما في ذلك العناصر الأكثر صعوبة في الاختبار. قد تُظهر منحنيات ICC في هذا النطاق نقاط منتصف (الصعوبة 'b') التي تتوزع عبر مقياس Theta، من القيم السالبة إلى القيم الموجبة.

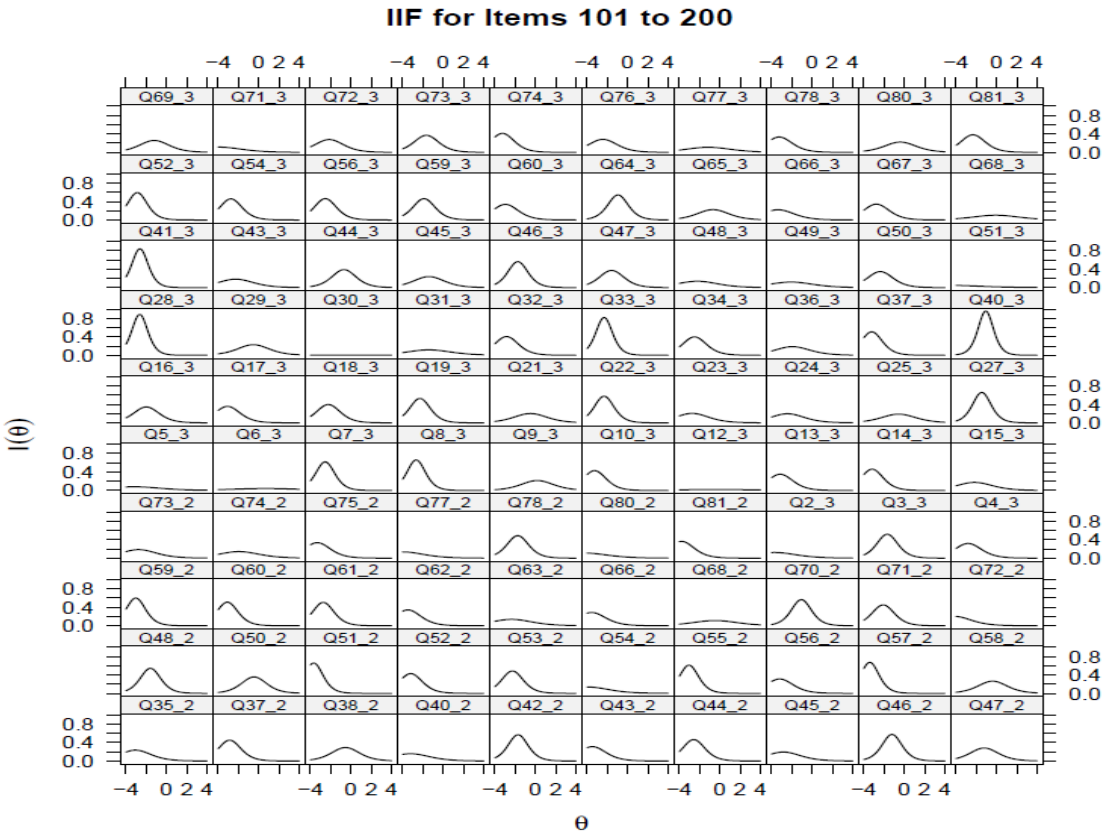
وتُشير المنحنيات التي تمتد إلى الجانب الموجب من مقياس Theta إلى أن الاختبار أصبح يمتلك عناصر قادرة على تحديد الفروق الدقيقة بين الأفراد ذوي القدرات العليا .وجود عناصر صعبة يُعزز "سقف القياس(Ceiling Effect) "، مما يضمن أن الأفراد الأكثر قدرة لا يحصلون ببساطة على درجات كاملة دون أن يتم قياس قدراتهم الحقيقية بشكل دقيق.

بالإضافة إلى ذلك، تُؤكد فإن عملية مراجعة العناصر وتعديلها أو حذفها بناءً على خصائص IRTتُعد أمراً بالغ الأهمية لضمان صلاحية الاختبار وموثوقيته. تُشير هذه المنحنيات إلى أن البنك قد خضع لهذه العملية بنجاح؛ مما أدى إلى مجموعة من العناصر ذات جودة قياسية أفضل، قادرة على توفير تقديرات دقيقة للقدرة عبر نطاق أكثر شمولاً من الأفراد.

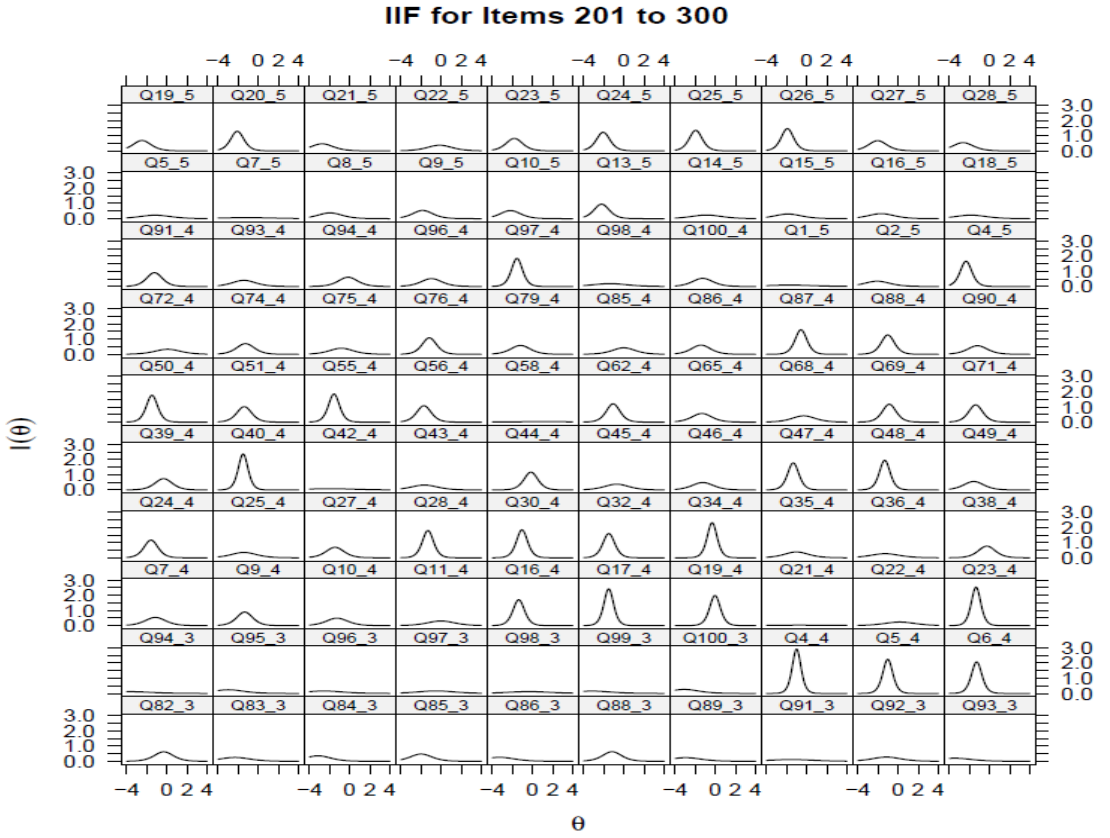
ولعل هذا التطور من الأجزاء المعروضة يعكس انتقال المعلومات المقدمة من الفصل الأول إلى الفصل الأخير من السهولة إلى الصعوبة النسبية؛ حيث إن المفردات مرتبة وفقاً لتقديمها للمفحوص وليس وفقاً لصعوبتها.



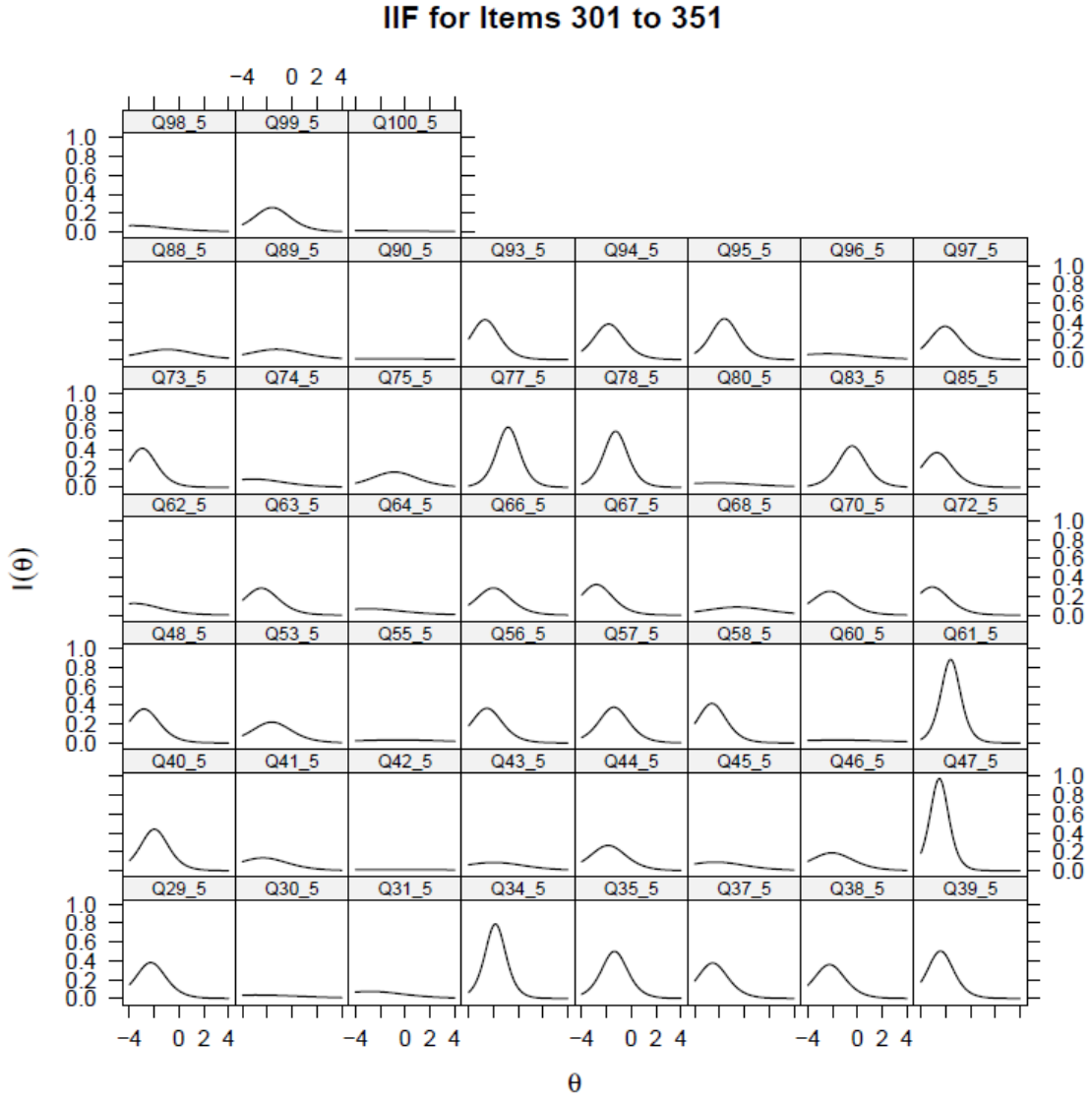
شكل (٦) منحنيات IFF للعناصر ١ إلى ١٠٠



شكل (٧) منحنيات IFF للعناصر ١٠١ إلى ٢٠٠



شكل (٨) منحنيات IFF للعناصر ٢٠١ إلى ٣٠٠



شكل (٩) منحنيات IFF للعناصر ٣٠١ إلى ٣٥١

ملحوظة: ترقيم المفردات (Q1, Q2,...Q100) للاختبار الأول، (Q1_2, Q2_2,...Q100_2) للاختبار الثاني، (Q1_3, Q2_3,...Q100_3) للاختبار الثالث، (Q1_4, Q2_4,...Q100_4) للاختبار الرابع، (Q1_5, Q2_5,...Q100_5) للاختبار الخامس، والمفردات غير الموجودة هي المفردات المحذوفة.

منحنيات معلومات العنصر (IIF) لمجموعات العناصر المختلفة:

تُعد منحنيات معلومات العنصر (Item Information Functions - IIFs) مكونًا حيويًا في نظرية الاستجابة للعنصر (IRT)، حيث تُوضح مدى دقة العنصر في قياس مستوى قدرة المفحوص (Theta) عند نقاط مختلفة على مقياس القدرة. تُشير ذروة منحنى IIF إلى مستوى القدرة الذي يُقدم فيه العنصر أقصى قدر من المعلومات، وبالتالي يُقلل من الخطأ المعياري للقياس (SEM) في تلك النقطة. يُقدم هذا التحليل نظرة عامة على خصائص IIFs لمجموعات مختلفة من العناصر، مع الأخذ في الاعتبار أن كل مجموعة تُشير إلى جزء من البنك.

١.منحنيات IIF للعناصر ١ إلى ١٠٠ و١٠١ إلى ٢٠٠ شكل (٦ & ٧)

تُشكل هذه المجموعة الجزء الأول من البنك وتُقدم لمحة عن جودة المعلومات المُقدمة. بالنظر إلى IIFs لهذه العناصر نُلاحظ نمطًا واضحًا: غالبية هذه العناصر تُقدم أقصى معلوماتها في نطاق القدرات المنخفضة جدًا (قيم Theta السلبية الكبيرة). على سبيل المثال، العنصر Q2 يُقدم معلومات قصوى عند Theta = -10.55، وQ٣ عند -٨.٤٩، وQ٨ عند -٧.٦٣.

ويُشير هذا التركيز الشديد للمعلومات في الذيل الأيسر من مقياس القدرة إلى أن هذا الجزء من الاختبار فعال للغاية في التمييز بين الأفراد ذوي القدرات المنخفضة جدًا . وهذا يُمكن أن يعني أن الاختبار يُقدم معلومات قليلة أو غير دقيقة للأفراد ذوي القدرات المتوسطة أو العالية، مما يؤدي إلى ارتفاع الخطأ المعياري في تقدير قدراتهم.

٢.منحنيات IIF للعناصر ٢٠١ إلى ٣٠٠ شكل (٨)

يُمكن أن نُلاحظ أن ذروات IIF لهذه العناصر تقع عند قيم Theta أقرب إلى -٢ أو -١، أو حتى حول الصفر. تُشير الدراسات في IRT إلى أن الاختبار الفعال يجب أن يُوفر معلومات كافية عبر نطاق القدرة ذي الصلة ليُقدم تقديرات موثوقة.

٣.منحنيات IIF للعناصر ٣٠١ إلى ٣٥١ شكل (٩)

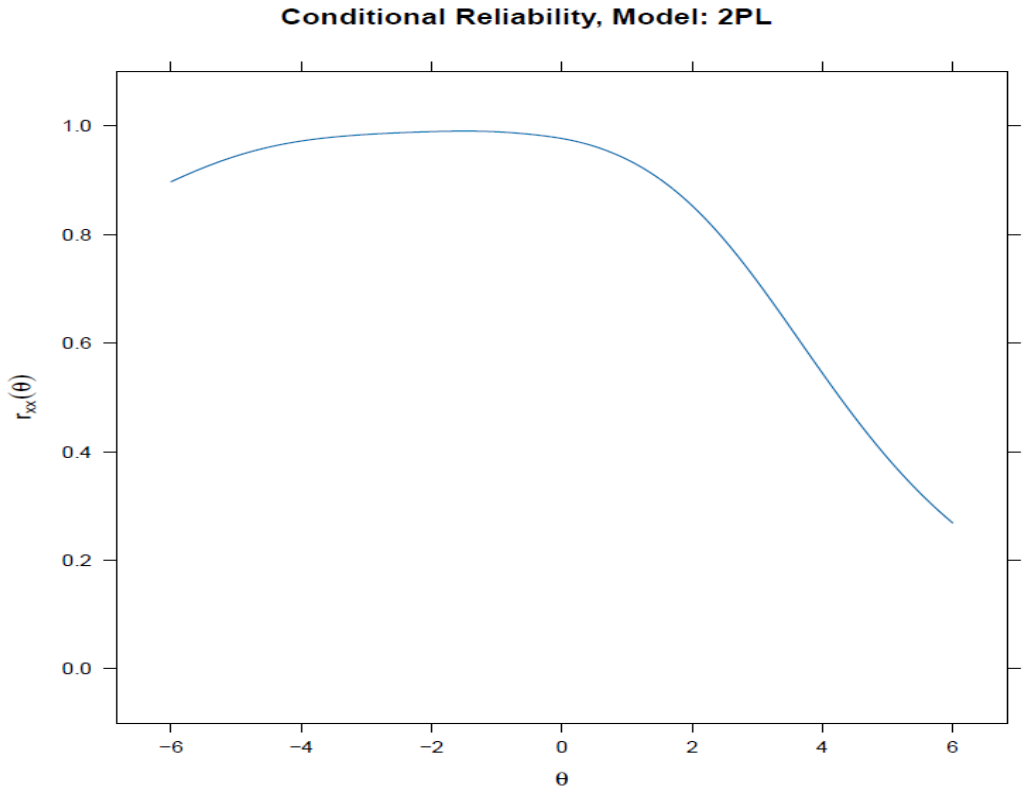
تُمثل هذه المجموعة الجزء الأخير من البنك، ومن المُتوقع أن تُظهر IIFs لهذه العناصر توزيعًا أوسع للمعلومات، بما في ذلك العناصر التي تُقدم معلومات قيمة للأفراد ذوي القدرات العالية. من المُرجح أن تُظهر هذه المجموعة بعض ذروات IIF عند قيم Theta موجبة أو قريبة جداً من الصفر، بالإضافة إلى بعض العناصر ذات الصعوبة المتوسطة.

وكذلك فإن وجود عناصر تُقدم معلومات قصوى عند Theta موجبة (مثل Q6_3 عند +٠.٦٧ أو Q68_3 عند -٠.٠٧، وهي أمثلة تُشير إلى تحسن في التغطية) يُشير إلى أن الاختبار أصبح يمتلك القدرة على التمييز بفعالية بين الأفراد ذوي القدرات العالية. هذا يُساعد على تجنب "تأثير السقف" (Ceiling Effect)، حيث لا يتمكن الاختبار من قياس الفروق بين الأفراد الأكثر قدرة لأن جميع العناصر سهلة جدًا بالنسبة لهم. تُشدد الدراسات السابقة في تصميم الاختبارات على أن الاختبار الذي يُقدم معلومات جيدة في جميع نطاقات القدرة المستهدفة هو اختبار أكثر صلاحية وموثوقية، حيث يُقلل من الخطأ المعياري للقياس (SEM) عبر طيف أوسع من القدرات (Hambleton et al., 1991).

كما أن الاختبارات عالية الجودة يجب أن تُوفر معلومات كافية لتمييز الأفراد عبر جميع مستويات القدرة ذات الأهمية. يُمكن القول إن التطور الملحوظ في توزيع ذروات IIF من الجزء الأول إلى الأجزاء اللاحقة من الاختبار يُشير إلى نجاح في تحقيق هذا الهدف، حيث لم يعد الاختبار يُركز

فقط على القدرات المنخفضة، بل أصبح قادراً على تقديم معلومات دقيقة للأفراد ذوي القدرات المتوسطة والعالية أيضاً. هذا يُعزز من الصلاحية البنائية للاختبار ويزيد من فائدته التطبيقية في سياقات مختلفة تتطلب تقييم مدى أوسع من الكفاءات.

ولعل هذا التطور من الأجزاء المعروضة يعكس انتقال المعلومات المقدمة من الفصل الأول إلى الفصل الأخير من السهولة إلى الصعوبة النسبية؛ حيث إن المفردات مرتبة وفقاً لتقديمها للمفحوص وليس وفقاً لصعوبتها.



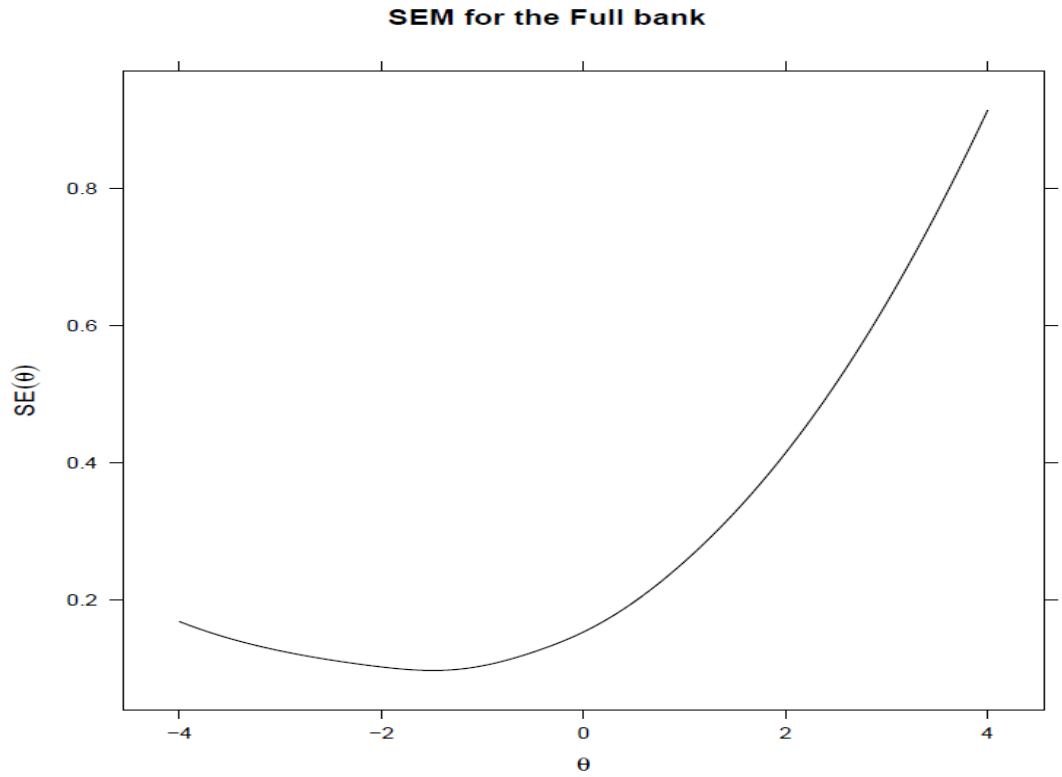
شكل (١٠) مُوثوقية الاختبار الشرطية (Conditional Reliability) لنموذج 2PL

يُقدم هذا الشكل البياني منحنى الموثوقية الشرطية للاختبار (Conditional Reliability, R(θ)) كدالة لمستوى القدرة (θ). في سياق نظرية الاستجابة للعنصر (IRT)، لا تُفترض الموثوقية كثابت واحد للاختبار بأكمله، بل كقيمة تتغير باختلاف مستوى قدرة المفحوصين. يُعد هذا المفهوم أكثر دقة من الموثوقية الكلاسيكية (مثل معامل ألفا كرونباخ) لأنه يُقدم تقديرًا لدقة القياس عند نقاط محددة على مقياس القدرة.

كما يُظهر المنحنى ارتفاعًا في الموثوقية في نطاق القدرات المتوسطة (حول θ=٠) وانخفاضًا تدريجيًا كلما ابتعدنا عن هذا النطاق نحو القدرات المنخفضة جدًا أو العالية جدًا (أي عند قيم θ السلبية أو الموجبة الكبيرة). على سبيل المثال، تُشير الذروة العالية للموثوقية (تقترب من ٠.٩٥) عند θ حوالي −٠.٥ إلى أن الاختبار يُقدم تقديرات دقيقة وموثوقة للغاية للأفراد الذين يمتلكون هذه

المستويات من القدرة. مع الابتعاد عن هذه النقطة، تنخفض الموثوقية بشكل ملحوظ، لتصل إلى حوالي ٠.٦ عند θ=−٤ وإلى حوالي ٠.٢ عند θ=+٤.

كما يُمكن ربط شكل هذا المنحنى بشكل مباشر بدالة معلومات الاختبار (Test Information Function - TIF). حيث تُوجد علاقة عكسية بين الخطأ المعياري للقياس (SEM) ودالة المعلومات، وعلاقة مباشرة بين دالة المعلومات والموثوقية الشرطية. كلما زادت معلومات الاختبار عند مستوى قدرة معين (ذروة TIF)، كلما انخفض الخطأ المعياري للقياس، وبالتالي ارتفعت الموثوقية الشرطية عند تلك النقطة. تُشير هذه العلاقة إلى أن تصميم البنك، وخصائص عناصره (خاصة الصعوبات)، تُؤثر بشكل مباشر على مدى دقة القياس عبر مقياس القدرة. تُظهر البيانات أن البنك يُقدم أفضل دقة له في النطاق الذي تتركز فيه معظم عناصر البنك ذات الصعوبة المتوسطة.



شكل (١١) الخطأ المعياري للقياس للبنك ككل (Standard Error of Measurement (for the Full Test Bank

وأيضاً يُقدم الشكل البياني (١١) منحنى الخطأ المعياري للقياس (Standard Error of Measurement - SEM) كدالة لمستوى قدرة المفحوص (θ). في سياق نظرية الاستجابة للعنصر (IRT)، لا يُعد الخطأ المعياري قيمة ثابتة للاختبار بأكمله، بل يتغير بناءً على مستوى القدرة الذي يتم

قياسه. يُعد هذا المفهوم أكثر دقة من الخطأ المعياري الكلاسيكي، لأنه يُقدم تقييمًا لمدى الدقة في تقدير قدرة الأفراد عند نقاط مختلفة على مقياس القدرة.

كما يُظهر المنحنى أن أقل قيمة للخطأ المعياري للقياس تقع في نطاق القدرات المتوسطة، وتحديداً حول θ تقترب من الصفر (حوالي −٠.٥). عند هذه النقطة، يكون الخطأ المعياري منخفضًا للغاية (أقل من ٠.٢)، مما يُشير إلى أن تقديرات القدرة في هذا النطاق هي الأكثر دقة وموثوقية. كلما ابتعدنا عن هذا النطاق، سواء نحو القدرات المنخفضة جدًا (قيم θ السالبة الكبيرة) أو القدرات العالية جدًا (قيم θ الموجبة الكبيرة)، يزداد الخطأ المعياري للقياس بشكل ملحوظ. على سبيل المثال، يرتفع الخطأ المعياري إلى حوالي ٠.٨ عند θ=٤ وإلى حوالي ٠.٦ عند θ=−٤.

كما يُشير هذا النمط إلى أن البنك الكلي مُصمم لتقديم أفضل دقة قياس للأفراد ذوي القدرات المتوسطة، مما يعني أن معظم العناصر ذات الصعوبة المتوسطة تتركز في هذا البنك. كلما زاد الخطأ المعياري، قلَّت دقة تقدير القدرة في ذلك النطاق، مما يعني أن البنك أقل كفاءة في التمييز بين الأفراد في الأطراف القصوى لمقياس القدرة. يُمكن ربط هذا المنحنى مباشرة بدالة معلومات الاختبار (Test Information Function - TIF)، حيث توجد علاقة عكسية؛ فكلما ارتفعت دالة معلومات الاختبار عند نقطة معينة، انخفض الخطأ المعياري للقياس عندها، مما يُعزز دقة التقدير.

التوصيات والبحوث المقترحة:

بناءً على التحليل السيكومتري لنموذج الاستجابة للمفردة ثنائي المعلمة (2PL)، والذي أظهر فعالية في بناء بنك الأسئلة بعد عمليات التنقية، ومع الأخذ في الاعتبار ضرورة تمكين المعلمين، تُقدم التوصيات والبحوث المقترحة التالية:

التوصيات:

١. **<u>عقد ورش عمل مكثفة للمعلمين حول تصميم المفردات المتباينة والتركيز على:</u>**

- **أهمية تنويع أشكال المفردات: توعية المعلمين بأن الاختبار الجيد لا يعتمد فقط على المحتوى، بل على تنوع أشكال المفردات (مثل الاختيار من متعدد، المطابقة، المقال القصير، المسائل التطبيقية) وقدرتها على قياس مستويات معرفية مختلفة (تذكر، فهم، تطبيق، تحليل، تقييم). هذا التنوع يُسهم في تغطية أوسع لمدى القدرة الحقيقي للطلاب.**
- **أهمية دالة معلومات العنصر: تدريب المعلمين على مفهوم دالة معلومات العنصر (IIF) وكيف تُشير إلى فعالية العنصر في قياس مستويات قدرة معينة. شرح كيفية استخدام هذه الدالة لانتقاء المفردات التي تُقدم أقصى معلومات في نطاقات القدرة المطلوبة (على سبيل المثال، عناصر سهلة للتشخيص، وعناصر متوسطة لقياس التحصيل العام، وعناصر صعبة للمتميزين).**

٢. برامج تدريبية لإنشاء بنوك الأسئلة باستخدام مبادئ IRT:

- **تطوير برامج تدريبية عملية للمعلمين تُمكنهم من بناء وتطوير بنوك أسئلة لمقرراتهم الدراسية. يجب أن تُركز هذه البرامج على المبادئ الأساسية لنظرية الاستجابة للمفردة (IRT) وكيفية استخدامها في: أ) صياغة مفردات ذات خصائص محددة (صعوبة وتمييز مستهدفين)، ب) تحليل البيانات الأولية لتقدير معلمات المفردات، ج) تفسير منحنيات ICC وIIF، د) اتخاذ قرارات مستنيرة بشأن الاحتفاظ بالمفردات أو تعديلها أو حذفها.**
- **الهدف هو تمكين المعلمين من إنشاء اختبارات تحصيلية ذات أهداف مختلفة (مثلاً: اختبارات تشخيصية، اختبارات تحديد مستوى، اختبارات تحصيل نهائية) من خلال تجميع مجموعات فرعية من بنك الأسئلة تُناسب الهدف المحدد ومستوى قدرة الطلاب المستهدف.**

البحوث المقترحة:

١.دراسة تقييم فعالية برامج تدريب المعلمين: إجراء دراسة بحثية لتقييم أثر برامج تدريب المعلمين المقترحة على جودة المفردات التي يُنشئونها وقدرتهم على بناء بنوك أسئلة فعالة، بالإضافة إلى تقييم مدى زيادة وعيهم بمفاهيم IRT.

٢.تحليل تجريبي لتأثير تنويع أشكال المفردات على الخصائص السيكومترية: تصميم دراسة تجريبية تُقارن الخصائص السيكومترية (ICC، IIF، TIF، SEM) لبنوك أسئلة تحتوي على تنوع كبير في أشكال المفردات مقابل بنوك أخرى ذات أشكال محدودة، لتقديم أدلة تجريبية قوية تدعم التوصية بتنويع أشكال المفردات.

٣.تطوير دليل إرشادي عملي للمعلمين: البحث في تطوير دليل إرشادي مُبسط وعملي للمعلمين لإنشاء بنوك أسئلة باستخدام IRT، مع التركيز على الأمثلة التطبيقية لمقررات مثل "الاتجاهات المعاصرة في علم النفس"

المراجع

أولاً: المراجع العربية:

توفيق علي عالم، وسوسن حسن أبوهادي (٢٠١٩). بناء بنك أسئلة في مادة الرياضيات للصف التاسع الأساسي في ضوء نظرية استجابة الفقرة والتحقق من جودة توظيف برنامج حاسوبي لإدارته، مجلة أبحاث، ع١٥، ١٦٠-١٩٧. مسترجع من http://search.mandumah.com/Record /10405907

زياد أحمد العبدالله (٢٠٢١). أثر اختلاف حجم العينة علي تقدير مفردات بنك الأسئلة. مجلة بابل- العلوم الإنسانية، مج٢٩، ع١٠، ٩٤-١١٤. مسترجع من http://search.mandumah.com/Record /1259379

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، ٣٥٨-٣٨٦. مسترجع من http://search.mandumah.com/Record/1414091

ماجد عبده قاسم ناجي، و عبدالمحسن بن رشيد المبدل (٢٠٢٢). تطوير بنك أسئلة في أساسيات القياس والتقويم باستخدام النموذج اللوغاريتمى. العلوم التربوية، مج٣٠، ع٤، ٥٤٧-٥٩٩. مسترجع من http://search.mandumah.com/Record /1403651

محمد جاد محمد (٢٠١٦). دراسة سيكومترية في دقة بناء بنك أسئلة في مادة الرياضيات باستخدام بعض نماذج نظرية الاستجابة للمفردة. عالم التربية، مج ١٧،ع٥٤، ٢٦٣- ٢٦٩. مسترجع من http://search.mandumah.com/Record/826138

هبه محمد إبراهيم الجمل ، شيرين محمد أحمد دسوقي، حسين حسن حسين طاحون، وهبه إبراهيم محمد الناغي (٢٠١٧). تدريج بنك أسئلة مادة الرياضيات لتلاميذ الصف الثالث الإبتدائي باستخدام النموذج اللوغاريتمى الثلاثي المعلم.مجلة كلية التربية، ع٢١، ٨٣٧- ٨٦٧. مسترجع من http://search.mandumah.com/Record/883314

وليد محمد خليفة فرج الله (٢٠١٨). أثر استخدام بنك أسئلة الكتروني في تدريس الجغرافيا على تنمية الأعماق المعرفية وخفض قلق الاختبار لدى الطالبات منخفضات التحصيل

بقسم الجغرافيا، مجلة العلوم التربوية، ع٣٥، ٤٥١-٤٩٥. مسترجع من http://search.mandumah.com/Record /949324

ثانياً: المراجع الأجنبية

ALKursheh, T. O., Al-zboon, H. S., & AlNasraween, M. E. S. (2022). The Effect of Item Form on Estimating Person's Ability, Item Parameters, and Information Function According to Item Response Theory (IRT). *International Journal of Instruction*, 15(3), 1111-1130. https://doi.org/10.29333/iji.2022.15359a.

Al-Zboon, H. S. A. (2020). The Effect of the Multiple-Choice Test Length on Estimating the Item Parameters and the Test Information Function According to the Three-Parameter Logistic Model in the Item Response Theory. *Journal of Economics and Sustainable Development,* 11(24), 42-50. DOI: 10.7176/JEP/11-24-05.

Al-zboon, H. S. A., & Maan, J. (2022). The Impact of the Number of Distractors in Multiple-choice Test Items on the Psychometric Characteristics of the Items and Item Information Function According to the Two-Parameter Logistic Model in the Item Response Theory. *JEP*, 13(13), 53-64. https://doi.org/10.7176/JEP/13-13-07.

Al-zboon, H. S., Alnasraween, M. E. S., & Alkursheh, T. O. (2021). The Effect of the Percentage of Missing Data on Estimating the Standard Error of the Items' Parameters and the Test Information Function According to the Three-Parameter Logistic Model in the Item Response Theory. *Ilkogretim Online*, 20(1), 887-898. https://doi.org/10.17051/ilkonline.2021.01.82.

Bahrom, S., Mat Yaacob, S. N., Happas, A., Saidin, N., Mohd Nawi, M. N., & Shaharom, N. (2025). Construction of an Item Bank Using Rasch Analysis. *EDUCATUM Journal of Science, Mathematics and Technology,* 12(1), 10–14. https://doi.org/10.37134/ejsmt.vol12.1.2.2025.

Baker, F. B., & Kim, S. H. (2004). *Item response theory: Parameter estimation techniques.* CRC Press.

Bergstrom, B. A., & Gershon, R. C. (1995). *8. Item Banking* . Licensure Testing: Purposes,Procedures, and Practices. 13. https://digitalcommons.unl.edu/buroslicensure/13.

Boone, W. J., Staver, J. R., Boone, W. J., & Staver, J. R. (2020). Test information function (TIF). ***Advances in Rasch analyses in the human sciences,*** **39-55. https://doi.org/10.1007/978-3-030-43420-5.**

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

Choppin, B. (1976). ***Developments in item banking*****. In** ***First European Contact Workshop,*** **Windsor, UK.**

Chuesathuchon, C., & Waugh, R. F. (2008). Item banking with rasch measurement: an example for primary mathematics in Thailand. ***International Conference*****. Sustainability in Higher Education: Directions for Change, Edith Cowan University, Perth Western Australia. Pp104-117. http://ro.ecu.edu.au/ceducom/8.**

Cohen, R. J., & Swerdlik, M. E. (2018). ***Psychological testing and assessment: An introduction to tests and measurement (9th ed.).*** **McGraw-Hill Education.**

DeVellis, R. F (2017). ***Scale development: Theory and applications (4th ed)*****. Sage publications.**

Downing, S. M., & Haladyna, T. M. (2006). ***Handbook of Test Development*****. Mahwah, NJ: Lawrence Erlbaum Associates.**

Downing, S. M., & Haladyna, T. M. (2011). Computerized Item Banking. In ***Handbook of Test Development*** **(pp. 275-300). Routledge.**

Falani, I., Akbar, M., & Naga, D. S. (2020). The Precision of Students' AbilityEstimation on Combinations of Item Response Theory Models. ***International Journal of Instruction,*** **13(4), 545-558. https://doi.org/10.29333/iji.2020.13434a.**

Friyatmi., Mardapi, D., Haryamto., & Rahmi, E. (2020). The Development of Computerized Economics Item Banking for Classroom and School-Based Assessment. ***European Journal of Educational Research, 9*****(1), 293-303.**

Haladyna, T. M. (2004). ***Developing and validating multiple-choice test items(3th ed)*****. Mahwah, NJ: Lawrence Erlbaum Associates.**

Hamad, M. M., & Alnuzaili, E. S. (2020). Steps of Designing a Personal Questions Bank in a Pedagogical Way. ***Theory & Practice in***

***Language Studies (TPLS)*, 10(6), 638-646. DOI: http://dx.doi.org/10.17507/tpls.1006.03.**

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

Hambleton, R. K., & Swaminathan, H. (2013). *Item response theory principles and applications.* Springer Science.

Hiscox, M., & Brzezinski, E. *A guide to item banking*. Portland, OR: Northwest. https://creativecommons.org/licenses/by/4.0/.

Hung, M., Voss, M., Bounsanga, J & Crum, A. (2017). Examination of the PROMIS upper extremity item bank. *Journal of Hand Therapy*, 30 (4), 485-490. doi: 10.1016/j.jht.2016.10.008.

Joo, S. H., Lee, P., & Stark, S. (2017). Evaluating anchor-item designs for concurrent calibration with the GGUM. *Applied psychological measurement*, 41(2), 83-96. https://doi.org/10.1177/0146621616673997

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. https://doi.org/10.1080/15366367.2021.1897398.

Keung Yim, L. W., Lye, C. Y., & Koh, P. W. (2024). A psychometric evaluation of an item bank for an English readingcomprehension tool using Rasch analysis. *REID (Research and Evaluation in Education)*, 10(1), 18-34. https://doi.org/10.21831/reid.v10i1.65284

Kim, K. Y. (2019). A comparison of the separate and concurrent calibration methods for the full-information bifactor model. *Applied Psychological Measurement*, 43(7), 512-526. https://doi.org/10.1177/0146621618813095

Kim, S. H., & Cohen, A. S. (1998). A comparison of linking and concurrent calibration under item response theory. *Applied psychological measurement, 22*(2), 131-143.

Kolen, M. J., & Brennan, R. L. (2013). *Test Equating, Scaling, and Linking: Methods and Practices (3th ed)*, Springer. DOI:10.1007/978-1-4939-0317-7.

Lee, J. E. (2011). Full-metric concurrent calibration for the development of CAT item banks. *Unpublished master's thesis*, University of Hanyang in Seoul, South Korea .

Lee, W. C., & Lee, G. (2018). IRT linking and equating. ***The Wiley handbook of psychometric testing: A multidisciplinary reference on survey, scale and test development,*** **639-673. https://doi.org/10.1002/9781118489772.ch21.**

Millman, J., & Greene, J. (1989). The specifications and development of tests of achievement and ability. ***In: R. L. Linn (Ed.),*** **Educational measurement (3rd Edition). Phoenix, AZ: American Council on Education.**

Nicol, D. J., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. ***Studies in higher education, 31*****(2), 199-218. http://dx.doi.org/10.1080/03075070600572090.**

Perez, J. E & Padrones, W. (2022). Implementation of a test constructor utilizing a calibrated item bank using 3PL-IRT model. ***Procedia Computer Science,*** **197, 495–502. https://doi.org/10.1016/j.procs.2021.12.166.**

Schminkey, D. L., von Oertzen, T., & Bullock, L. (2016). Handling missing data with multilevel structural equation modeling and full information maximum likelihood techniques. ***Research in Nursing & Health,*** **39(4), 286-297. https://doi.org/10.1002/nur.21724.**

Shen, Y., Wang, S., & Xiao, H. (2024). A two-step item bank calibration strategy based on 1-bit matrix completion for small-scale computerized adaptive testing. ***British Journal of Mathematical and Statistical Psychology,*** **77(3), 553-612. DOI: 10.1111/bmsp.12340.**

Tan, T. K. (2024). Evaluating assessment via item response theory utilizing information function with R. ***The Quantitative Methods for Psychology,*** **20(1), 33-49. DOI:10.20982/tqmp.20.1.p033.**

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

Wainer, H., Dorans, N. J., Flaugher, R., Green, B. F., & Mislevy, R. J. (2000). ***Computerized adaptive testing: A primer.*** **Routledge. https://doi.org/10.4324/9781410605931.**

Weiss, D. J. (2013). Item banking, test development, and test delivery. In K. F. Geisinger (Ed.), ***The APA Handbook on Testing and Assessment.*** **Washington, WA: American Psychological Association.**

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

Wilson, M. (2005). *Constructing measures: An item response modeling approach.* Lawrence Erlbaum Associates. https://doi.org/10.4324/9781410611697.

Zheng, C., Guo, S., & Kern, J. L. (2021). *Fast Bayesian estimation for the four-parameter logistic model (4PLM).* SAGE Open, 11(4), 1-13. https://doi.org/10.1177/21582440211052556.