MJHSS

Volume (41), Number (1), 2026.
ISSN: 1021-6804. On line ISSN: 3006-8665.

Mutah Journal of Humanities and Social Sciences

Assessing the Accuracy of Outcome-Based Assessment Decisions Using the Two-Parameter IRT Model (2PL): An Applied Framework for Concentrating Test Information at Critical Cut Scores

Sami Salameh Almassarweh *

sami.almassarweh@iu.edu.jo

Mohammad A. Abumaal*

Mohammad.abumaal@iu.edu.jo

Received:16 /12/2025 **Accepted: 8 / 3/2026**

Abstract:

This study aimed to examine the precision of assessment decisions in an Outcome-Based Education (OBE) context by calibrating a 40-item True/False test designed to measure five specific learning outcomes using the 2PL-IRT model. The study employed a descriptive-analytical approach using a census sample of 275 students at a private university in Jordan. The selection of the 2PL model was intended to ensure stable parameter estimation given the sample size, with a focus on the parameters of discrimination (a) and difficulty (b). The results indicated that the items had discrimination indices ranging from 1.18 to 2.10 and difficulty parameters from -1.10 to 1.60, reflecting their effectiveness in differentiating students with varying ability levels. However, the TIF function revealed that maximum measurement precision was centered at the average ability level ($\theta \approx 0$), rather than at the cut score required in OBE contexts. Fit analyses showed that most items conformed to the model assumptions, although items q10, q20, and q24 displayed significant p-value deviations. The study concludes that traditional tests, despite possessing sound psychometric properties, may functionally fail to support critical "mastery" decisions and therefore recommends redesigning the item difficulty distribution to align with decision thresholds rather than with the mean ability level.

Keywords: Outcome-Based Education (OBE) , Two-Parameter Logistic Model (2PL), Cut-Score.

* Department of Psychology, Faculty of Arts, Isra University, Jordan.

المجلد(41)، العدد (1)، 2026م.
ISSN: 1021-6804. On line ISSN: 3006-8665.
MJHSS
مجلة مؤتة للدراسات الإنسانية والاجتماعية

تقييم دقة قرارات التقييم المستندة إلى المخرجات باستخدام نموذج IRT ثنائي المعلَمة: إطار تطبيقي لتركيز معلومات الاختبار عند نقاط القرار الحرجة

سامي سلامه المصاروه *
sami.almassarweh@iu.edu.jo
محمد عبدالفتاح أبو معال*
Mohammad.abumaal@iu.edu.jo

تاريخ الاستلام:16/12/ 2025 **تاريخ القبول: 2026/3/8**

الملخص:

تهدف هذه الدراسة إلى فحص دقة قرارات التقييم في بيئة التعليم القائم على المخرجات (OBE) من خلال معايرة فقرات اختبار صح/خطأ مكون من 40 فقرة لقياس 5 مخرجات تعلم محددة باستخدام نموذج نظرية الاستجابة للمفردة ثنائية المعلمة (IRT 2PL).تم تطبيق الدراسة باستخدام المنهج الوصفي التحليلي على عينة كلية (Census Sample) مكونة من 275 طالبًا في إحدى الجامعات الخاصة في الأردن، اعتمدت الدراسة نموذج 2PL لضمان استقرار تقديرات المعالم نظرًا لحجم العينة، مع التركيز على معلمتي الصعوبة (b) والتمييز (a). وأظهرت النتائج أن الفقرات تمتعت بمعاملات تمييز تتراوح بين (1.18 – 2.10) ومعاملات صعوبة بين (−10.1 إلى 1.60)، مما يعكس قدرة الفقرات على التمييز بين الطلبة ذوي مستويات القدرة المختلفة. إلا أن دالة معلومات الاختبار (TIF) أظهرت تمركز الدقة القصوى عند مستوى القدرة المتوسط $(\theta \approx 0)$ وليس عند نقطة القطع (Cut-score) التي يتطلبها نظام OBE ، كما أظهرت نتائج تحليل الملاءمة أن معظم الفقرات تتوافق مع افتراضات النموذج، مع وجود بعض الفقرات(q10، q20، q24) التي أظهرت اختلافات في قيم p-value. تخلص الدراسة إلى أن الاختبارات التقليدية، وإن كانت تتمتع بخصائص سيكومترية جيدة، قد تفشل وظيفيًا في دعم قرارات الإتقان الحاسمة، وتوصي بإعادة هندسة توزيع صعوبة الفقرات لتتواءم مع نقاط اتخاذ القرار لا مع المتوسط الحسابي.

الكلمات المفتاحية: التعليم القائم على المخرجات (OBE)، نموذج ثنائي المعلمة (2PL)، نقطة القطع.

* قسم علم النفس، كلية الآداب، جامعة الإسراء، الأردن.

Mutah Journal of Humanities and Social Sciences, Volume (41), Number (1), 2026.
DOI: https://doi.org/10.35682/mjhss.v41i1

المقدمة:

في السنوات الأخيرة، حظيت نظرية الاستجابة للفقرة (IRT) باهتمام واسع في مجال القياس والتقويم التربوي، كونها تعالج ثغرات نظريـــة الاختبار التقليديـــة وتقـدم مقاييس أكثر دقـة لفحـص خصـــائـص فقرات الاختبار وقدرات الأفراد (Hambleton & Swaminathan, 1985). فبدلاً من افتراض تساوي جميع الفقرات في الصعوبة والمعلومات كما في نظرية الاختبار الكلاسيكية Classical Test Theory (CTT)، تسمح نماذج نظرية الاستجابة للمفردة (IRT) بمعايرة كل فقرة اختبارية اعتمادًا على مجموعة من المعلمات الإحصائية، تشمل معامل الصعوبة (b)، ومعامل التمييز (a)، ومعامل التخمين (c)، وذلك لتقدير مدى كفاءة الفقرة ودقتها في قياس السمة الكامنة لدى المفحوصين (Reise & Waller,2009).

وشهد حقل القياس التربوي تحولات جذرية تنتقل به من الأطر التقليدية التي تعتمد على الدرجة الخام، إلى نماذج احتمالية معقدة تسعـــى لتقدير القدرة الكامنة بدقة تتجاوز حـدود العينة والاختبار. وقد أشـار باحثون مثل هامبلتون وآخرون (Hambleton et al., 1991) إلى أن نظرية القياس الكلاسيكية (CTT) تعجز عن تقديم تقديرات مستقرة؛ لأنها تجعل خصائص الفقرات مرتبطة بالعينة، وتفترض ثبات خطأ القياس عبر جميع مستويات القدرة، وهو ما يتعارض مع الواقع المعرفي.

وتبرز أهمية هذا المنهج بوضـــوح فـــي الاختبارات عالية المخاطر أو القياسات المرتبطـــة بقرارات حاسمة، حيث إنّ كمية المعلومـات المتـوفـرة حـــول مستـوى قـدرة الطالـب عنـد نقطة القرار تحـدد دقة التقديـرات واتخـاذ القرار الصحيـح (van der Linden & Hambleton, 1997) .

من ناحية أخرى، نشأت فلسفة التعليم المرتكز على المخرجات Outcome-Based Education (OBE) حول مبدأ أن يتركز النظام التعليمي بأكمله على تحقيق مخرجات تعلم محددة، بحيث يصبح الطالب قادراً على أداء مهمات أو مهارات معينة بنهاية تجربته التعليمية(Spady, 1994). يقتضي هذا النهج تصميم المناهج وأساليب التقييم بحيث تحقق هذه المخرجات وتؤكدها. وبتبنيOBE ، تصبح القرارات التقييمية (مثل الحكم على نجاح الطالب أو رسوبه، وقياس مدى تحقيق مخرجات معينة) معتمدة بشكل أساسي على بيانات تقييمية دقيقة وصحيحة. بناءً عليه، فإنّ تحسين أدقّية قرارات التقييم المبنية على المخرجات يرتبط مباشرة بتحسين أدوات القياس وفقرات الاختبار التي تتناسب مع هذه المخرجات(Malan, 2000).

وفي المقابل، ينطلق التعليم المرتكز على المخرجات (OBE) من مبدأ أن التعلم الحقيقي هو ما يمكن إثباته، وليس ما يُفترض أنه تم تدريسه. ويؤكد Biggs & Tang (2011) أن محور OBE هو المواءمة البنّاءة(Constructive Alignment) ، حيث يجب أن يكون التقييم مرآة صادقة للمخرج التعليمي .إلا أن الإشكالية تكمن في أن معظم الممارسات الجامعية لا تزال تستخدم أدوات قياس تقليدية (Norm-referenced) لاتخاذ قرارات مرجعية المحك (Criterion-referenced) مثل "الإتقان" أو "التحقق.

تكمن أهمية البحث في أنه يقدم إطاراً يستفيد من قوة IRT في تركيز المعلومات الإحصائية عند مستويات القدرة الحرجة، بما يدعم القرارات التي تتخذ بناءً على تلك المستويات (Reise & Waller, 2009). وكما نلاحظ في الأدبيات، فإن الاعتماد على اختبارات مبنية على IRT يمكن أن يزيد من موثوقية ودقة استنتاجات القائمين على التقييم (على سبيل المثال، في الاختبارات الوطنية أو الجامعية) (Embretson & Reise, 2013).

بناءً على ما سبق، تسعى هذه الدراسة إلى سدّ الفجوة بين الصرامة السيكومترية التي تميز نظرية الاستجابة للمفردة (IRT) والفلسفة البراغماتية التي يقوم عليها نظام التعليم القائم على المخرجات (OBE)، وذلك من خلال فحص تجريبي لقدرة اختبار جامعي مُصمَّم لقياس مخرجات التعلم (PLOs) على توفير معلومات قياس دقيقة عند نقاط اتخاذ القرار الحرجة.

تأتي هذه الدراسة لتعالج السؤال الأساسي التالي: كيف يمكن تقوية القرارات التقييمية المرتكزة على المخرجات من خلال معايرة فقرات الاختبار باستخدام نموذج IRT ثنائية المَعلمة لتركيز معلومات الاختبار حول نقاط القرار؟ لتحقيق ذلك، سيتم طرح أسئلة بحث محددة تتعلق بخصائص فقرات الاختبار (الصعوبة والتمييز)، وعلاقتها بالمخرجات، ومدى تشتت معلومات الاختبار بالنسبة لمستويات قدرة الطلبة.

مشكلة الدراسة وأسئلتها:

تكمن مشكلة البحث في أن القرارات المستخلصة من اختبارات مخرجات التعلم (مثل قرار النجاح أو الرسوب في مخرج تعليمي معين) قد تكون عرضة للخطأ إذا لم تكن أدوات التقييم مصممة بدقة أو إذا لم تكن مليئة بالمعلومات في مناطق القرار الحاسمة. على سبيل المثال، إذا كان هناك فقرة اختبار يصعب تجاوزها ولكنها لا تساهم في تمييز الطلبة بشكل فعال، فإن وجودها قد لا يخدم الهدف المطلوب أو قد يقود إلى حكم خاطئ على قدرة الطلبة. ومن جهة أخرى، فقد يكون لبعض المخرجات متعددة الأبعاد (كمعرفة نظرية وتطبيق عملي) فقرات اختبار غير متوازنة الصعوبة، مما يشوش دقة تقدير مدى تحقيقها.

وفي ضوء التوجه المتزايد نحو تبنّي التعليم المستند إلى المخرجات(OBE) ، يواجه المتخصصون في القياس والتقويم تحدّيًا عمليًا يتمثل في كيفية بناء أو معايرة فقرات الاختبارات بحيث تُسهم في تعظيم معلوماتها حول مستويات القدرة الحرجة التي تعتمد عليها قرارات تحقق المخرج من عدمه. وتبرز أهمية هذا التحدّي على نحو خاص عند استخدام نماذج نظرية الاستجابة للمفردة (IRT) ثنائية المعلمة، نظرًا لقدرتها على إعادة توزيع معلومات الاختبار وتكثيفها حول نقاط القرار الحساسة، بما يتيح تقديرات أكثر دقة وموثوقية لأداء المتعلمين ويُحسّن جودة الحكم على تحقق المخرجات التعليمية.

وعليه، يبرز أمام المتخصصين في القياس والتقويم في إطار التعليم المستند إلى المخرجات (OBE) تساؤل عملي محوري: كيف يمكن بناء أو معايرة فقرات الاختبارات بحيث تتزايد معلوماتها حول مستويات القدرة الحرجة التي يُحسم بناءً عليها قرار تحقق المخرج من عدمه؟

أي بعبارة أخرى، هل يمكن استخدام IRT بنماذج 2PL لتركيز معلومات الاختبار حول المستوى الذي يفصل بين نجاح الطالب في مخرج التعلم أو فشله فيه؟

لتحقيق ذلك، يمكن تبني فرضية تفيد بأنّ: "معايرة فقرات الاختبارات وفق نماذج IRT وتعديلها حسب معاملات الصعوبة والتمييز يزيد المعلومات الإحصائية عند نقاط القرار ويحسن موثوقية ودقة التقييم المستند إلى النتائج.

لمعالجة هذه الفرضية، نحدد أسئلة البحث الرئيسة كالآتي :

1 – ما هي خصائص فقرات الاختبار (معاملات الصعوبة والتمييز) بعد معايرتها بمعايير IRT (2PL) ؟

2 – هل تدل هذه القيم على جودة الفقرات؟

3 – كيف يساهم توزيع معلومات الفقرات (Item Information) في مستويات القدرة المختلفة في تحسين دقة القرارات على مستوى المخرجات (مثل الاحتفاظ بالنظام السابق عند نقطة الفاصل)؟

أهداف الدراسة:

هدفت هذه الدراسة إلى:

1 – تحليل خصائص فقرات الاختبار المصمم لقياس مخرجات التعلم (PLOs) باستخدام نموذج نظرية الاستجابة للفقرة (IRT) ثنائية المَعلمة (2PL)، مع التركيز على معاملات الصعوبة(b) ، التمييز (a) .

2 – تقدير توزيع معلومات الفقرات ومجموع معلومات الاختبار (Item & Test Information Functions) حول نقاط القرار الحاسمة، مثل حد النجاح، لتقييم دقة التقديرات التربوية.

3 – استكشاف العلاقة بين نتائج الطلبة وتحقيق المخرجات التعليمية من خلال تحليل مدى مطابقة أداء الطلبة في الفقرات المرتبطة بكل مخرج تعليمي.

4– تقديم توصيات عملية لتحسين دقة القرارات التقييمية في نظام التعليم المرتكز على المخرجات (OBE) من خلال إعادة توجيه فقرات الاختبار لتعزيز التركيز المعلوماتي حول نقاط القرار.

أهمية الدراسة:

تتحدد هذه الدراسة من إسهامها المنهجي في تعزيز دقة وموثوقية التقييم التربوي من خلال توظيف إحدى نماذج نظرية الاستجابة للمفردة (IRT) ضمن إطار التعليم القائم على المخرجات (OBE)، بما يحدّ من احتمالية الوقوع في أخطاء القياس وانعكاساتها على قرارات النجاح أو الرسوب. كما تسهم نتائجها في تحسين تصميم الاختبارات التحصيلية عبر تمكين مطوّري أدوات القياس من معايرة فقرات الاختبار بطريقة أكثر عدالة وحساسية لقياس قدرات الطلبة عند مستويات القدرة المرتبطة بعتبات اتخاذ القرار الحاسمة. وتدعم الدراسة مواءمة عملية التقويم مع مخرجات التعلم من خلال توفير قرارات تقويمية مستندة إلى بيانات سيكومترية دقيقة تعكس بدرجة عالية من الموضوعية مدى تحقق تلك المخرجات. إضافة إلى ذلك، تتمثل أهميتها التطبيقية في إظهار إمكانية استخدام أدوات تحليل متقدمة مثل SPSS و jMetrik لتقدير معاملات نماذج IRT ضمن بيئة تحليل مألوفة للباحثين، دون الحاجة إلى برمجيات متخصصة معقدة. وعلى الصعيد العلمي، تسهم الدراسة في إثراء الأدبيات العربية في مجال القياس والتقويم التربوي عبر تقديم نموذج تطبيقي متكامل يربط بين IRT وOBE، بما يعزز جودة البحث الأكاديمي في مجال القياس النفسي الحديث.

مصطلحات الدراسة:

نظرية الاستجابة للفقرة(Item Response Theory – IRT): نظرية سيكومترية حديثة تُنمذج العلاقة الاحتمالية بين القدرة الكامنة للمفحوص (θ) واحتمال إجابته على الفقرة، اعتمادًا على معلمات الفقرة وليس خصائص العينة، بما يحقق استقلالية نسبية للقياس ودقة أعلى في تقدير القدرة وبناء الاختبارات (Crocker & Algina, 1986; Embretson & Reise, 2000; (Hambleton et al., 1991).

معامل الصعوبة(b): هو معْلَمة تمثل موقع الفقرة على متصل القدرة الكامنة (θ)، ويشير في نموذج ثنائي المعلمة (2PL) إلى مستوى القدرة الذي يكون عنده احتمال الاستجابة الصحيحة للفقرة مساويًا تقريبًا لـ (0.50)، في ظل ثبات معامل التمييز، بما يعكس مدى ملاءمة الفقرة لمستويات محددة من القدرة الكامنة (Lord, 1980; Hambleton et al., 1991; Embretson & Reise, 2000).

معامل التمييز(a): هو معْلَمة تعكس قدرة الفقرة على التمييز بين الأفراد ذوي المستويات المختلفة من القدرة الكامنة، ويُعبَّر عنه بدرجة انحدار منحنى خصائص الفقرة حول نقطة الصعوبة (b)، حيث تشير القيم المرتفعة له إلى حساسية أعلى للفقرة في التفريق بين المفحوصين القريبين من مستوى الصعوبة المحدد (الزغول، 2012؛ Birnbaum, 1968; Hambleton et al., 1991; De Ayala, 2022).

معامل التخمين(c): معْلَمة تمثل الحد الأدنى لاحتمال الإجابة الصحيحة الناتجة عن التخمين، ويُستخدم أساسًا في فقرات الاختيار من متعدد ضمن النموذج اللوجستي ثلاثي المعلمات، بما يحسن دقة نمذجة أداء ذوي القدرة المنخفضة (Birnbaum, 1968; Lord, 1980; عدس، 2010).

منحنى معلومات الفقرة(Item Information Function): دالة تبين مقدار المعلومات التي تقدمها الفقرة عند مستويات مختلفة من القدرة الكامنة، وتبلغ هذه المعلومات أقصاها قرب معامل الصعوبة، وتزداد بزيادة معامل التمييز (Embretson & Reise, 2000; Baker & Kim, 2017; أبو علام، 2011).

دالة معلومات الاختبار(Test Information Function): دالة تمثل مجموع معلومات فقرات الاختبار عبر متصل القدرة، وتُعد مؤشرًا مباشرًا على دقة القياس وكفاءة الاختبار عند مستويات القدرة المختلفة، حيث يرتبط ارتفاعها بانخفاض الخطأ المعياري لتقدير القدرة (علّام، 2007؛ Crocker & Algina, 1986).

التعليم المرتكز على المخرجات (OBE): هو نهج وفلسفة تعليمية يركز على تحديد مخرجات تعلم قابلة للقياس، بحيث يكون الطالب قادرًا على أداء مهام أو اكتساب مهارات محددة بنهاية التجربة التعليمية، مع توجيه التدريس والتقويم لضمان اتساق أدوات التقويم مع نواتج التعلم المستهدفة (Spady, 1994; Biggs & Tang, 2011; ملحم، 2013).

مواءمة الفقرات مع المخرجات (Constructive Alignment): تشير إلى الاتساق بين مخرجات التعلم، وأنشطة التدريس، وأدوات التقويم، بحيث تقيس الفقرات الاختبارية البناء النظري للمخرجات بدقة، مما يعزز صدق القياس وجودة التقويم (Biggs & Tang, 2011; أبو علام، 2011).

الإطار النظري:

ينطلق هذا البحث من مقاربة تكاملية تجمع بين الدقة السيكومترية لنظرية الاستجابة للفقرة (IRT) والمتطلبات الوظيفية لنظام التعليم القائم على المخرجات (OBE). وفيما يلي تأصيل نظري لمتغيرات الدراسة والعلاقة بينها:

أولاً: نظرية الاستجابة للفقرة (IRT) ونموذج المعلمتين(2PL)

تمثل نظرية الاستجابة للفقرة (IRT) الإطار الإحصائي الأنسب لمعالجة مشكلات القياس المرتبطة بقرارات الإتقان، كونها تحرر تقديرات القدرة (θ) من خصائص العينة، وتوفر خطأً معيارياً مشروطاً يختلف باختلاف مستوى القدرة، خلافاً للنظرية الكلاسيكية التي تفترض ثبات الخطأ (Hambleton & Swaminathan, 1985; Embretson & Reise, 2000).

وقد تبنت هذه الدراسة النموذج اللوجستي ثنائي المعلمة (2PL) كخيار استراتيجي يحقق الكفاءة الإحصائية؛ فهو يتجاوز قصور نموذج راش في افتراض تساوي التمييز، ويتجنب تعقيدات النموذج ثلاثي المعلمة التي تتطلب أحجام عينات ضخمة لتقدير التخمين بدقة. ويُعبر عن هذا النموذج رياضياً بالمعادلة التالية:

$$P_i(\theta) = \frac{e^{Da_i(\theta - b_i)}}{1 + e^{Da_i(\theta - b_i)}}$$

حيث: a_i : يمثل معلم تمييز الفقرة.

حيث يقوم نموذج (2PL) على معلَمَتِين رئيستين تحكمان دقة القرار(Lord, 1980; Hambleton et al., 1991):

1 – معلَمة التمييز (ai): وتمثل ميل المنحنى عند نقطة الانعطاف؛ وتعد المعلمة الأهم في سياق هذه الدراسة، لأن الفقرات ذات التمييز المرتفع توفر دقة قياسية أعلى ، وتعمل على تعظيم الفروق بين الطلبة المتقنين وغير المتقنين حول نقطة القطع (de Ayala, 2022).

2 – معلَمة الصعوبة (bi): وتحدد الموقع على متصل القدرة الذي تبلغ فيه الفقرة أقصى دقة لها. ومن منظور سيكومتري، يجب أن تتطابق صعوبة الفقرات مع نقطة القطع المطلوبة لضمان أدنى خطأ في التصنيف (van der Linden & Hambleton, 1997).

وتجدر الإشارة إلى أن اختيار نموذج (2PL) جاء بعد المفاضلة العلمية مع النماذج الأخرى؛ فهو يتفوق في دقته على نموذج راش أحادي المعلمة الذي يفترض تساوي التمييز (Bond & Fox, 2015 ; أبو علام، 2011)، كما يتجنب التعقيد الحسابي للنماذج المتقدمة كالنموذج رباعي المعلمة (4PL) الذي يتطلب تقدير الحد الأعلى للاحتمالية (Barton & Lord, 1981; Magis,2013). ويأتي هذا الاختيار منسجماً مع دراسات المقارنة التي أوصت بكفاءة (2PL) في أحجام

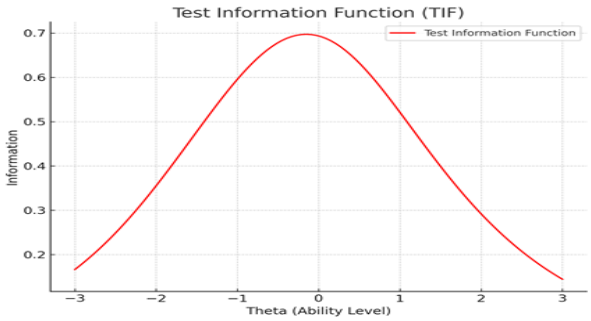
العينات المتوسطة (Wolkowitz, 2008)، مما يجعله الأنسب لخدمة مبادئ التعليم القائم على المخرجات، كالتصميم العكسي وتوسيع فرص الأداء (Wiggins & McTighe, 2005; Jackson, 2002).

ثانياً: دالة المعلومات (Information Function) ودقة القرار

تُعد دالة المعلومات المفهوم المحوري الذي يربط خصائص الفقرات بدقة القرارات التربوية. فبدلاً من الاعتماد على الثبات الكلي للاختبار، تتيح دالة معلومات الاختبار (TIF) تحديد مقدار الدقة عند كل نقطة من نقاط القدرة (Baker & Kim, 2017). وتُحسب دالة المعلومات للفقرة ($I_{i(\theta)}$) في نموذج (2PL) وفق المعادلة التالية:

$$I_{i(\theta)} = a_i^2 P_{i(\theta)[1 - P_{i(\theta)}]}$$

وتشير الأدبيات إلى أن الاختبار الفعال في اتخاذ القرارات الحاسمة (Mastery Decisions) هو الذي تتطابق فيه ذروة منحنى المعلومات (Peak Information) مع نقطة القطع (Cut-score) المعتمدة في المؤسسة التعليمية، مما يقلل من احتمالية الخطأ المعياري في التقدير (SEM) عند هذه المنطقة الحرجة، ويخفض بالتالي من معدلات التصنيف الخاطئ (النجاح الزائف أو الرسوب الزائف) (Embretson & Reise, 2000; Uesaka et al., 2022) كما في الشكل رقم (1).



شكل رقم (1) تمثيل نظري لمنحنى دالة معلومات الاختبار (TIF) وعلاقته بدقة التقدير.

ثالثاً: التعليم القائم على المخرجات (OBE) ومتطلبات القياس

يقوم نظام (OBE) على مبدأ التصميم العكسي، حيث تشتق التقييمات مباشرة من مخرجات تعلم (PLOs) محددة وقابلة للقياس (Spady, 1994; Biggs & Tang, 2011). وتكمن الإشكالية القياسية في هذا النظام في أنه يتطلب قرارات قاطعة (حقق/لم يحقق) بناءً على درجات الاختبار (Butler, 2018).

وهنا تبرز الفجوة التي تعالجها الدراسة؛ فمعظم الاختبارات التقليدية تُبنى لتميز الفروق الفردية حول المتوسط (Normal Distribution)، بينما يتطلب نظام (OBE) اختبارات تُبنى لتعظيم التمييز عند حدود الإتقان (Criterion-referenced points) (Malan, 2000).

رابعاً: التكامل الوظيفي(Functional Alignment)

تأسيساً على ما سبق، لا تهدف الدراسة إلى مجرد تطبيق نموذج(IRT) ، بل توظيفه كأداة ضبط جودة لقرارات (OBE). فاستخدام نماذج (IRT) لضبط خصائص الفقرات (خاصة التمييز والصعوبة) يضمن أن الأداة تمتلك حساسية قياس (Measurement Sensitivity) موجهة نحو مخرجات التعلم المستهدفة، مما يحقق المواءمة البنّاءة (Constructive Alignment) بين دقة الأداة سيكومترياً ووظيفتها التربوية في اتخاذ قرارات مصيرية (Lahner et al., 2020; Asbari & Novitasari, 2024).

كما اعتمدت هذه الدراسة على دمج التعليم المرتكز على المخرجات (OBE) مع نظرية الاستجابة للمفردة(IRT) ، حيث يركز OBE على تحديد مخرجات تعلم واضحة وقابلة للقياس وربطها بما نعلّمه ونقيّمه(Spady, 1994) ، ويتيح الدمج مع IRTتحسين دقة وعدالة التقييم ودعم القرارات التربوية المتعلقة بتحقيق هذه المخرجات.

التعليم المرتكز على المخرجات(OBE) :

التعليم المرتكز على المخرجات هو نهج تربوي يركّز على ما يجب أن يكون المتعلم قادرًا على فعله بنهاية العملية التعليمية، وليس فقط على المحتوى الذي تم تدريسه(Spady,1994).وفي هذا المنهج، تُحدد مخرجات تعلم واضحة في البداية، وتُصمم المناهج والتقييمات حول هذه المخرجات.

وتعتمد فلسفة OBE على عدد من المبادئ الجوهرية:

1 – وضوح التركيز :(Clarity of Focus) يجب تحديد مخرجات تعلم دقيقة وواضحة بحيث تكون مفهومة للمعلمين والطلبة (Spady,1994).

2 – التصميم العكسي :(Backward Design) يبدأ تصميم المنهج من تحديد المخرجات، ثم يُصمّم المحتوى والأنشطة التربوية والتقييمات لدعم هذه المخرجات ((Wiggins & McTighe, 2005).

3 – توقعات عالية :(High Expectations) الفرضية الأساسية أن جميع الطلبة قادرون على تحقيق المخرجات إذا توفرت لهم الفرص والدعم اللازم(Spady,1994).

4 – توسيع الفرص :(Expanded Opportunities) يمنح OBE الطلبة طرقًا متعددة لإظهار مخرجاتهم، سواء من خلال مشاريع، أو مهام أداء، أو اختبارات، أو أنشطة عملية(Jackson,2002).

وفي إطار OBE ، يمكن تصنيف مخرجات التعلم إلى:

1 – معرفية :(Knowledge) مثل الفهم، التذكر، التحليل.

2 – مهارية :(Skills) مثل حل المشكلات، التفكير النقدي، التطبيق.

3 – وجدانية :(Attitudes/Values) مثل الالتزام الأخلاقي، التعاون، المسؤولية.

4 – مخرجات مهنية/عملية: خاصة في التعليم المهني أو الجامعي الذي يتطلب كفاءات عملية محددة.

أما فيما يخص التقييم، فيعد جوهريًا لتحقيق العدالة والتأكد من أن المخرجات قد تحققت فعلًا. وعليه، يجب أن يكون مبنياً على معايير (criterion-referenced) وليس فقط مقارنة بين الطلبة. ويجب أن تكون أدوات التقييم مصمَّمة خصيصًا لتقيس المخرجات المحدّدة، سواء من خلال اختبارات ورقية، مشاريع، مهام أداء، أو ملاحظة أداء(Jackson,2002). وهنا يجب استخدام حدود قطع (cut-scores) لتحديد ما إذا كان الطالب قد "حقق" المخرج أم لا. هذه الحدود يجب أن تكون مدعومة ببيانات موثوقة ودقيقة لضمان قرارات عادلة.

وعلى الرغم من هذا كله، يواجه تطبيق OBE في المؤسسات التعليمية عددًا من التحديات، من أهمها ما يلي:

1– صياغة مخرجات دقيقة وقابلة للقياس، خاصة عندما تكون المخرجات غير معرفية أو معقدة.

2 – مقاومة التغيير من قبل المتخصصين وأعضاء هيئة التدريس في الجامعات، الذين قد يفضلون الاستمرار في استخدام الأساليب التقليدية في التدريس والتقييم.

3– ضمان العدالة في التقييم من خلال تصميم أدوات متنوعة متعددة الطرق لتمكين جميع الطلبة من إظهار مخرجاتهم.

4– تحديد نقاط القطع (cut-scores) بدقة واستخدامها في اتخاذ قرارات مثل الانتقال أو التخرج، ما يتطلب تحليلًا إحصائيًا ودعمًا بياناتيًا (.(Lahner, F. M., et al. (2020).

التكامل بين IRT وOBE:

على الرغم من محدودية الدراسات التي تطبق IRT داخل إطار OBE بشكل مباشر، فإن الأدبيات المعاصرة توفر دعمًا نظريًا وعمليًا لهذا التكامل. فعلى سبيل المثال، تُظهر دراسة (Uesaka et al ,2022) كيف يمكن استخدام IRT لتحليل استراتيجيات تعلم الطلبة كمخرجات تعلم استراتيجية، بالمقابل، يوضّح تطبيق OBE في التعليم العالي (Implementing OBE, 2024)مدى تركيز هذا الإطار على مخرجات التعلم والربط بينها وبين التقييم (Nguyen et al., 2024).كما أن المراجعة النظرية لـ IRT في القياس التربوي توضح القوة الإحصائية الكبيرة لنماذجها في تقدير قدرات الطلبة وتحليل الفقرات (van der Linden & Hambleton, 1997) ودليل التقييم في OBE يبيّن كيفية تصميم نظام تقييم يعتمد على مخرجات تعلم واضحة واتخاذ قرارات تقييمية بناءً عليها(Butler ,2018).

بناءً على هذه الأدبيات، فإن الجمع بين نظرية الاستجابة للفقرة (IRT) والتعليم المرتكز على المخرجات (OBE) يشكل إطارًا نظريًا متينًا لتطوير التقييم التربوي بطرق دقيقة وعادلة. إذ توفر IRT الأدوات الإحصائية اللازمة لتحليل الفقرات بدقة، وتقدير القدرة، وتصميم اختبارات موضوعة بعناية لتعزيز دقة القرار عند النقاط الحرجة. من جهة أخرى، يقدم OBE الفلسفة التربوية التي تضمن أن ما يتم تقييمه هو ما يجب أن يُعلّم حقًا، وأن المخرجات التعليمية هي الأساس في تصميم العملية التعليمية. بالتكامل بينهما يمكن تحقيق تقييم تربوي مترابط، وفعال، وعادل، ومبني على بيانات قوية تدعم اتخاذ قرارات تربوية موضوعية. وهنا يمكن استنتاج أن دمج IRT داخل إطار OBE يتبنى أساسا نظريًا قويًا، إذ يوفر IRT دقة إحصائية عالية في

تقدير القدرات وتحليل الفقرات، بينما يوفر OBE إطارًا منطقيًا لاتخاذ قرارات تعتمد على أساس ما إذا كان الطالب قد حقّق مخرجات معينة.

طرق الدمج:

1 – محاذاة الفقرات مع المخرجات(Item–Outcome Mapping) :

تُربط فقرات الاختبار مباشرة بمخرجات التعلم PLOs (أو مخرجات البرنامج)، بحيث يتم ضمان أن الفقرات تغطي مستويات القدرة المطلوبة لتحقيق تلك المخرجات. ثم يُستخدم تحليل IRT لتحديد الفقرات التي تقدم معلومات عالية حول تلك النقاط الحرجة من القدرة (مثل حدود القطع).

2 – تصميم دالة المعلومات موجهة للمخرجات:

في هذه الدراسة، تم استخدام دالة معلومات الاختبار (Test Information Function – TIF) ضمن إطار نظرية الاستجابة للمفردة (IRT) لتصميم الاختبارات بحيث تتركّز المعلومات حول مستويات القدرة المرتبطة بقرارات تعليمية مهمة، مثل حد النجاح أو مستوى الكفاءة المطلوب. ويُسهم هذا التركيز في زيادة موثوقية القرار وتقليل الخطأ المعياري لتقدير القدرة عند تلك النقاط الحرجة، مما يعزّز دقة وموضوعية تقييم مخرجات التعلم(de Ayala, 2022).

إلى جانب ذلك، تؤكد الأدبيات السيكومترية الحديثة أن تكوين بنوك الأسئلة من خلال استهداف معلومات عالية عند نقاط القرار يعد من الأساليب الفعّالة لدعم اتخاذ القرارات التعليمية الدقيقة (de Ayala, 2022; Embretson & Reise, 2000).

الدراسات السابقة:

تشير مراجعات الأدبيات إلى كثرة الدراسات التي تناولت نظرية الاستجابة للمفردة (IRT) وتطبيقاتها في تقييم القدرات الأكاديمية، إضافة إلى دراسات حول التعليم القائم على المخرجات (OBE). ومع ذلك، يلاحظ غياب الدراسات التي تدمج بشكل تكاملي نماذج IRT متعددة المعلمات مع OBE، حسب حدود معرفة الباحثين، مما يترك فجوة بحثية مهمة لتطوير أساليب تقدير دقة قرارات التقييم عند نقاط القطع الحرجة وفق مخرجات التعلم، لذا، سنستعرض في هذا القسم أهم الدراسات التي تتبنى كل من IRT وOBE على حدة، وذلك للوقوف على أبرز الطرق وأهم النتائج والتحديات التي يمكن الاستفادة منها في هذا البحث. فمثلا، تهدف دراسة المطيري (2025) إلى تقييم مدى دقة تقدير قدرات الأفراد بناءً على حجم العينة والنموذج المستخدم (ثلاثي أو رباعي المعلمة) وفقًا لنظرية الاستجابة للفقرة. وتشكلت عينة الدراسة من مجموعات متنوعة بأحجام مختلفة (200، 500، 800، 1100، 1400) تم فحصها افتراضيًا. ولتحقيق أهداف البحث، تم اعتماد المنهج المقارن، مع استخدام أساليب المحاكاة لتوليد البيانات وتحليلها ومقارنة النتائج. وقد أُنشئت البيانات وحُلّلت باستخدام برنامج (R) وحزمة النماذج متعددة الأبعاد ضمن إطار نظرية الاستجابة للفقرة متعددة الأبعاد (MIRT) .وتمت المقارنات من خلال المتوسطات، والانحرافات المعيارية، واختبار تحليل التباين بين المجموعات. وأظهرت النتائج تأثيرًا واضحًا لحجم العينة على دقة تقدير القدرات في كلا النموذجين (الثلاثي والرباعي)، لصالح أحجام العينة الأصغر، كما أظهرت تفوق النموذج الرباعي على الثلاثي في دقة تقدير قدرات الأفراد.

وهدفت دراسة (Al Ajmi et al, 2024) إلى فحص الخصائص السيكومترية لاختبار القدرة العددية باستخدام نموذج لوجستي ثلاثي المعايير (3PL) في إطار نظرية الاستجابة للفقرة (IRT). يتألف الاختبار من 30 فقرة ثنائية التدرج، وطُبّق على عينة

عنقودية متعددة المراحل مكوّنة من 2689 طالبًا وطالبة من الصفين الخامس والسادس في مدارس دول الخليج العربي. وقد أظهرت النتائج توافقًا مرتفعًا بين فقرات الاختبار ومتطلبات النموذج الثلاثي، مما يدعم ملاءمة نموذج IRT للاختبار. كما تحقق الاختبار من افتراض أحادية البعد (UD) والاستقلال المحلي، ممّا يعزّز من سلامته السيكومترية. وأظهرت النتائج أيضًا قدرة الاختبار على التمييز بين الممتحنين ذوي مستويات القدرة العددية المختلفة، ولا سيّما أصحاب القدرات المنخفضة والمتوسطة. إضافةً إلى ذلك، حقّق المقياس مستوى مرتفعًا من الموثوقية، حيث بلغ معامل الموثوقية الهامشي 0.83. وتشير هذه المعطيات إلى أهمية مواصلة البحث مستقبلاً بهدف تعزيز دقة الاختبار ورفع كفاءته.

وهدفت دراسة (Uesaka, Y et al, 2022)، إلى تحديد مستوى استخدام استراتيجيات التعلم لدى الطلبة (acquisition level)باستخدام تحليلIRT ، وتم اختيار عينة الدراسة حجمها (472) من خمس جامعات يابانية مختلفة الترتيب الأكاديمي من طلبة مقرر علم النفس التربوي، واستخدام المنهج الاستقصائي. وتوصلت الدراسة إلى أن كل العناصر تقريبًا لديها معامل تمييز 0.55 < (discrimination) ، ما يعني أن العناصر تميز جيدًا بين مستويات الاستخدام المختلفة، والطلبة في الجامعات ذات الأداء الأكاديمي العالي استخدموا استراتيجيات تفكير عميقة واستراتيجيات ميتا معرفية أكثر، لكن استخدموا استراتيجيات الموارد الخارجية (مثل المشاركة في برامج علمية أو قراءة مواد خارجية) بشكل أقل، وهناك علاقة إيجابية بين مستوى اكتساب الاستراتيجية θمن IRT وترتيب الجامعة (كنوع من مقياس الأداء الأكاديمي): الجامعات ذات الترتيب الأعلى كان لديها متوسط θأعلى.

أما دراسة (Asbari & Novitasari, 2024) بحثت في أثر التعليم القائم على النتائج (OBE) في تعزيز الإبداع والابتكار لدى المحاضرين في مؤسسات التعليم العالي، مع التركيز على تصميم المناهج، وأساليب التدريس، وممارسات التقييم. وباعتماد منهج وصفي نوعي يشمل مقابلات مع مصادر معلومات رئيسية، وملاحظات ميدانية في إحدى الجامعات بمدينة تانجيرانج في اندونيسيا، وكان هدفها استكشاف كيفية تأثير نموذج OBE على إبداع الأساتذة والابتكار في التعليم الجامعي، خاصة في تصميم المناهج، أساليب التدريس، والتقييم. وتم استخدام عينة قصدية من أساتذة الجامعة، وتشير النتائج إلى أن تنفيذ OBE دفع الأساتذة إلى تصميم مناهج أكثر تفاعلية وواقعية، مع تركيز على المشاريع والتقييم التعاوني، وشجع النموذج الأساتذة على استخدام أساليب تدريس مبتكرة وأكثر مركزية على الطلبة، والتقييم التقييم أصبح أكثر تنوعًا (مشروعات، تقييم جماعي، تقييم عملي)، بدلاً من الامتحانات التقليدية فقط.

أما دراسة (Nguyen et al., 2024) فكان الهدف منها هو دراسة تطبيق التعليم القائم على النتائج OBE في أربع جامعات في فيتنام ولاوس. واستُخدمت دراسة حالة متعددة، مع التأمل الشخصي وتحليل الوثائق، لتحديد الإنجازات والمعوقات. وتُظهر نتائج البحث أن تطبيق التعليم القائم على النتائج غيّر خطط الدروس والأنشطة الصفية، وعزز مشاركة الطلبة في عمليات التعليم والتعلم. ومع ذلك، واجه أساتذة الجامعات والطلبة تحديات عند تطبيقهم للتعليم القائم على النتائج في دروسهم. وتخلص الدراسة إلى أنه على الرغم من صعوبات التطبيق، فقد كان لنهج التعليم القائم على النتائج، الذي يركز على الطالب، تأثير إيجابي على تعلم الطلاب في المؤسسات محل الدراسة.

تعقيب على الدراسات السابقة:

توضح الدراسات السابقة تنوع التطبيقات النظرية والتطبيقية لنظرية الاستجابة للفقرة (IRT) والتعليم المرتكز على المخرجات (OBE) . فقد ركزت دراسات المطيري (2025) و Al Ajmi et al (2024) و Uesaka et al. (2022)على الجوانب القياسية والتحليل السيكومتري للبيانات، حيث أظهرت نتائجها فعالية IRT في تقدير القدرات وتحليل الخصائص السيكومترية للاختبارات، مع مراعاة تأثير حجم العينة ونموذج المعلَمة، وقدرة الفقرات على التمييز بين مستويات القدرات المختلفة. كما أظهرت الدراسات أن النماذج متعددة المعلمات توفر دقة أكبر في تقدير القدرات، وتحقق التوافق مع افتراضات أحادية البعد والاستقلال المحلي، مما يعزز من صلاحية وموثوقية أدوات القياس.

أما الدراسات المرتبطة بـ OBE مثل (Asbari & Novitasari, 2024)؛ Nguyen et al., 2024) فقد أظهرت أثرًا إيجابيًا لتطبيق هذا النهج على تحسين الممارسات التعليمية، حيث ساهم في تطوير مناهج تفاعلية، وأساليب تدريس مبتكرة، وتقييمات متنوعة، مع تعزيز مشاركة الطلبة في عمليات التعلم. ومع ذلك، أشارت النتائج إلى وجود تحديات عند تطبيق OBE على أرض الواقع، خاصة فيما يتعلق بالتكيف مع ممارسات التدريس التقليدية وقيود البيئة التعليمية.

بشكل عام، تقدم هذه الدراسات قاعدة قوية تربط بين التحليل السيكومتري الدقيق باستخدام IRT وبين تعزيز جودة التعليم والتعلم عبرOBE ، ما يشير إلى إمكانية دمج المنهجين لتحقيق تقييم أكثر دقة وكفاءة وفعالية في تطوير القدرات والمهارات التعليمية.

بناءً على ما سبق، فإن المقاربة البحثية للجمع بين نظرية الاستجابة للفقرة (IRT) والتعليم القائم على المخرجات (OBE) لا تستهدف دمجاً فلسفياً، بل تسعى إلى تحقيق تكامل منهجي ووظيفي (Methodological and Functional Integration). فبينما يختص إطار (OBE) بتحديد السمة الكامنة (Latent Trait) المستهدفة ومحكات الأداء(Performance Criteria) ، يوفر إطار (IRT) النماذج الاحتمالية (Probabilistic Models) اللازمة لتقدير دقة القياس الشرطية (Conditional Measurement Precision)، وتحديداً من خلال تعظيم قيم دالة معلومات الاختبار (TIF) وتقليل الخطأ المعياري عند نقاط القطع الحرجة. وعليه، فإن استثمار الخصائص السيكومترية (Psychometric Properties) المتقدمة لنماذج (IRT) في ضبط القرارات التصنيفية (Classification Decisions) داخل نظام (OBE) يمثل إطاراً تطبيقياً يحقق التجسير الوظيفي (Functional Bridging) بين النظرية والتطبيق، مما يضمن تقييماً يتسم بالكفاءة والعدالة في الحكم على تحقق النواتج التعليمية.

منهجية الدراسة وإجراءاتها:

التصميم والعينة:

اعتمدت الدراسة المنهج الوصفي التحليلي على عينة كلية (Census Sample) قوامها 275 طالبًا في جامعة الإسراء في الأردن. وتم اختيار العينة الكلية لتجنب أخطاء المعاينة العشوائية، ولتوفير أكبر قاعدة بيانات ممكنة لعملية المعايرة السيكومترية.

أداة البحث:

تم استخدام اختبار من نوع الصح والخطأ مكوّن من (40) فقرة، مصمم لقياس مخرجات التعلم المحددة(PLOs) . تم ربط كل فقرة بمخرج تعليمي محدد لضمان دقة القياس وتحليل العلاقة بين الفقرات وتحقيق المخرجات.

صُمم الاختبار لقياس التحصيل في مساق الاختبارات النفسية، حيث تم بناء الفقرات استناداً إلى مصفوفة مواصفات دقيقة تربط بين المحتوى وخمسة مخرجات تعلم محددة (PLOs) تمثل أبعاد القدرة المستهدفة، وهي:

مخرجات التعلم المستهدفة(PLOs)

1 –المخرج المعرفي(Knowledge): استيعاب المفاهيم الأساسية للاختبارات النفسية، والإلمام بخصائصها السيكومترية، وتصنيفاتها، ومجالات استخدامها المختلفة.

2 –المخرج المهاري(Skills): تطبيق الخطوات العملية لبناء الاختبارات النفسية، وإجراء التحليلات الإحصائية اللازمة للفقرات لاستخراج دلالات الصدق والثبات.

3 – مخرج اتخاذ القرار(Decision-Making): توظيف نتائج القياس النفسي والتربوي في اتخاذ قرارات تشخيصية وتوجيهية ملائمة ومبنية على الأدلة العلمية.

4 – المخرج الذهني(Cognitive Skills): تحليل البيانات الكمية المستخلصة من الاختبارات وتفسيرها، وتقييم جودة أدوات القياس في ضوء معايير القياس النفسي المعاصرة.

5– المهارات القابلة للنقل(Transferable Skills): توظيف مهارات التفكير الناقد في فحص ومراجعة أدوات القياس النفسي، وتحديد مدى ملاءمتها للسياق التطبيقي.

تبرير اختيار النموذج (Model Selection Rationale)

على الرغم من الطبيعة الاحتمالية لفقرات 'الصح/الخطأ' التي قد تستدعي نظرياً استخدام النموذج ثلاثي المعلمة (3PL) لضبط أثر التخمين، إلا أن محددات حجم العينة في هذه الدراسة (N=275) حتمت تبني النموذج ثنائي المعلمة (2PL) استناداً إلى مبدأ الكفاءة الإحصائية (Statistical Efficiency). حيث تشير الأدبيات السيكومترية إلى أن التقدير الدقيق والمستقر لمعلمة التخمين (c-parameter) يتطلب أحجام عينات كبيرة تتجاوز غالباً (1000) مفحوص لضمان تقارب النموذج (Model Convergence) وتجنب الأخطاء المعيارية المرتفعة في التقدير (Embretson & Reise, 2000; de Ayala, 2022). وعليه، فإن محاولة استخدام النموذج الثلاثي (3PL) مع حجم العينة الحالي قد تؤدي إلى تشويه دقة تقديرات معالم الصعوبة والتمييز؛ لذا عُدّ نموذج (2PL) الخيار الأكثر مواءمة (Parsimonious) لتحقيق الاستقرار التفسيري للمعلمات، مع اعتبار التباين غير المفسر الناتج عن التخمين جزءاً من خطأ القياس العشوائي الذي لا يهدد البنية العاملية للاختبار، وهو ما يخدم الهدف الرئيس للدراسة المتمثل في فحص دقة التمييز عند نقاط القطع لا نمذجة سلوك التخمين.

إجراءات التحقق من الافتراضات:

للتحقق من افتراض "أحادية البعد(Unidimensionality) "، تم إجراء تحليل عاملي. ونظراً لطبيعة البيانات الثنائية (1/0)، تم التعامل بحذر مع النتائج لتجنب "عوامل الصعوبة الوهمية (Difficulty Factors) " التي قد تنتج عن استخدام مصفوفات بيرسون التقليدية (Asbari & Novitasari, 2024). دعمت نتائج تحليل الملاءمة (Item Fit) التي أظهرت قيم

($p > 0.05$) لمعظم الفقرات (de Ayala, 2022) فرضية أن البعد المهيمن في البيانات هو قدرة الطالب، مما يبرر استخدام نماذج IRT .

"تعتمد جامعة الإسراء درجة 50 % كحد أدنى للنجاح (Cut-score) في هذا المساق. ومن منظور نظرية الاستجابة للمفردة (IRT)، وفي ظل توزيع صعوبة الفقرات المتوازن لهذا الاختبار، فإن الحصول على 50% من الدرجة الكلية يكافئ تقريبًا مستوى قدرة متوسط.($\theta \approx 0$) . وهذا يعني أن قرارات النجاح والرسوب تتمركز حول منتصف متصل القدرة".

الثبات:

تم حساب الثبات للمقياس بعدة طرق باستخدام برنامج JMetrik كما هو موضح في الجدول رقم (1).

جدول رقم (1) معامل الثبات.

الطريقة	قيمة معامل الثبات	فاصل الثقة 95%	الخطأ المعياري للقياس (SEM)
معامل جتمان(L2)	0.9682	(0.9626, 0.9734)	1.8204
معامل ألفا لكرونباخ	0.9673	(0.9615, 0.9726)	1.8459
معامل فيلدت–جيلمر	0.9680	(0.9623, 0.9732)	1.8263
معامل فيلدت–برينان	0.9680	(0.9623, 0.9732)	1.8261
معامل بيتا لراجو	0.9673	(0.9615, 0.9726)	1.8459

ويتضح من خلال نتائج الجدول رقم (1) أن أداة الدراسة تتمتع بدرجة ثبات مرتفعة، مما يدل على صلاحية المقياس وتطبيقه. إضافة إلى ذلك، جاء الاتساق الداخلي للأداة مرتفعًا جدًا كما تعكسه مؤشرات الثبات المتعددة، حيث بلغت قيمة كرونباخ ألفا = 0.9673 وقيمة Guttman L2 = 0.9682، وخطأ قياس قياسي منخفض (SEM ≈ 1.82–1.84) مما يدعم تجانس الفقرات واتساقها البنائي.

الصدق:

للتحقق من صدق البناء وملاءمة البيانات للتحليل العاملي، تم تطبيق اختباري KMO و Bartlett على مجموعات الفقرات المرتبطة بالمخرجات التعليمية، كما يبينه الجدول رقم (2).

DOI: https://doi.org/10.35682/mjhss.v41i1

جدول رقم (2) اختبار كفاية العينة (KMO) واختبار بارتليت لجميع المخرجات

KMO (Kaiser–Meyer–Olkin Measure of Sampling Adequacy)

المخرج	الفقرات	مقياس كايزر–ماير–أولكن لكفاية العينة(KMO)	قيمة كاي تربيع التقريبية(χ^2)	درجات الحرية	مستوى الدلالة الإحصائية
الأول	1–8	0.880	935.013	28	0.000
الثاني	9–16	0.873	807.887	28	0.000
الثالث	17–24	0.900	665.990	28	0.000
الرابع	25–32	0.880	829.977	28	0.000
الخامس	33–40	0.899	896.213	28	0.000

أظهرت نتائج التحليل السيكومتري كما في الجدول رقم (2) توفر أدلة قوية على صدق البناء للأداة؛ إذ كشفت مصفوفة الارتباطات بين الفقرات (N = 275) عن ارتباطات متوسطة ومناسبة تراوحت بين (0.30–0.60)، بما يعكس علاقة جوهرية ومتوازنة بين الفقرات والسمة الكامنة. كما جاءت معاملات التمييز مرتفعة (0.521–745.0)، مما يدل على قدرة الفقرات على التفريق بين مستويات القدرة المختلفة. وانسجامًا مع ذلك، أكد التحليل العاملي الاستكشافي لكل مجموعة فقرات تحقق بنية عامليه قوية؛ حيث تراوحت قيم KMO بين (0.88–0.90)، وكان اختبار Bartlett دالًا إحصائيًا في جميع الحالات(Sig = .000) ، مع تشاركيات مرتفعة (0.70–0.96) وأحمال عامليه قوية (0.60–0.90). كما أظهر Scree Plot الشكل رقم (2) عاملًا واحدًا مهيمنًا يفسر ما بين (47%–52%) من التباين، بما يدعم تحقق افتراض أحادية البعد ويبرر منهجيًا تطبيق نموذج نظرية الاستجابة للمفردة ثنائي المعلمة(2PL) ، مع ضمان صلاحية ودقة تقدير معاملي الصعوبة والتمييز.

ويوضح الجدول رقم (3) مقدار التباين الذي تفسره المكونات المستخرجة في التحليل العاملي، مما يساعد على فهم مدى مساهمة كل مكون في تفسير البيانات الكلية.

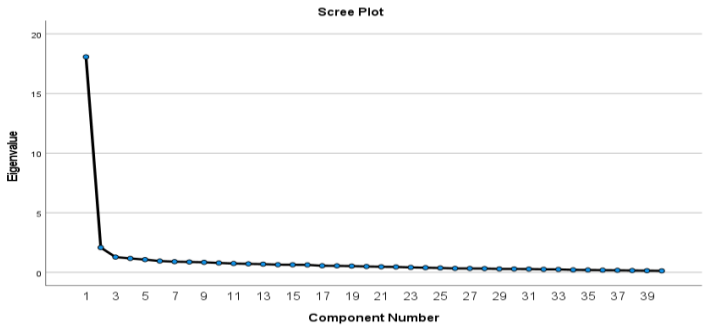
جدول رقم (3) مقدار التباين الذي تفسره المكونات المستخرجة في التحليل العاملي

العامل	إجمالي التباين المفسر						
	القيم الذاتية الأولية			مجاميع مربعات التشبعات بعد الاستخراج			مجاميع مربعات التشبعات بعد التدوير
	المجموع	نسبة التباين المفسر (%)	النسبة التراكمية للتباين المفسر (%)	المجموع	نسبة التباين المفسر (%)	النسبة التراكمية للتباين المفسر (%)	المجموع
1	18.072	45.179	45.179	18.072	45.179	45.179	6.085
2	2.083	5.207	50.387	2.083	5.207	50.387	5.637
3	1.285	3.213	53.600	1.285	3.213	53.600	4.893
4	1.174	2.934	56.534	1.174	2.934	56.534	4.576
5	1.075	2.688	59.222	1.075	2.688	59.222	2.497
6	.954	2.385	61.606				
7	.903	2.257	63.863				
8	.877	2.192	66.055				
9	.847	2.119	68.174				
10	.793	1.983	70.157				
11	.747	1.868	72.025				
12	.719	1.797	73.822				
13	.694	1.734	75.556				
14	.647	1.618	77.174				
15	.642	1.604	78.778				
16	.631	1.577	80.355				
17	.559	1.397	81.753				
18	.552	1.380	83.132				
19	.530	1.325	84.457				
20	.498	1.245	85.702				
21	.474	1.186	86.888				
22	.459	1.148	88.036				
23	.421	1.052	89.088				

DOI: https://doi.org/10.35682/mjhss.v41i1

24	.397	.992	90.080
25	.372	.930	91.011
26	.336	.841	91.852
27	.331	.828	92.680
28	.321	.804	93.484
29	.294	.734	94.218
30	.291	.727	94.945
31	.281	.704	95.649
32	.260	.650	96.300
33	.251	.627	96.927
34	.215	.537	97.464
35	.205	.512	97.976
36	.191	.478	98.454
37	.178	.445	98.899
38	.160	.401	99.299
39	.147	.368	99.667
40	.133	.333	100.000

تُظهر نتائج تحليل التباين الكلي المفسر Total Variance Explained كما في الجدول رقم (3) هيمنة واضحة للعامل الأول إذ بلغ القيمة الذاتية (Eigenvalue = 18.072) الذي يفسر 45.18% من التباين الكلي، في مقابل انخفاض حاد في القيم الذاتية للعوامل اللاحقة، حيث لم يضف العامل الثاني سوى 5.21% من التباين. ويؤكد Scree Plot كما في الشكل (2) هذا النمط بوضوح من خلال كسر حاد بعد العامل الأول يعقبه استواء تدريجي، بما يدل على وجود عامل واحد مهيمن. وعلى الرغم من تجاوز عدة عوامل لمعيار Kaiser ، فإن ضعف الإسهام التفسيري الإضافي يدعم اعتماد بنية أحادية البعد، ويبرر منهجيًا تطبيق نموذج IRT (2PL).



الشكل رقم (2) Scree Plot للمكونات المستخرجة في التحليل العاملي

DOI: https://doi.org/10.35682/mjhss.v41i1

وللتأكد من صدق المحتوى، عُرضت الأداة بصيغتها الأولية التي تضمنت (56) فقرة على خمسة خبراء متخصصين في القياس والتقويم والمناهج، بهدف التحقق من مدى ملاءمة الفقرات للمجال المعرفي والسمة الكامنة التي تهدف الأداة إلى قياسها. قام الخبراء بتقدير كل فقرة وفق معايير: وضوح الصياغة، ودقة المحتوى، وملاءمة مستوى الصعوبة، وانتماء الفقرة للبعد المقاس. وتم حساب نسبة الاتفاق البسيطة بين الخبراء حول صلاحية الفقرات، وبلغت 80%، وهي نسبة تُعدّ مرتفعة ومقبولة لأغراض البناء والاختبارات التربوية. وبناءً على ملاحظات الخبراء ومعايير التحكيم، تم حذف (16) فقرة غير مناسبة أو منخفضة الاتفاق، ليصبح عدد الفقرات في صورتها النهائية (40) فقرة تتمتع بدرجة عالية من الاتساق وملاءمة المحتوى.

وتم احتساب معامل ارتباط درجة كل فقرة بالدرجة الكلية للمقياس، والتي تراوحت قيمتها (0.521–0.745) وبهذا يمكن اعتبار ذلك أحد مؤشرات صدق البناء. والجدول (4) يبين معاملات ارتباط الفقرة مع الدرجة الكلية للمقياس.

جدول رقم (4) معاملات الارتباط للفقرات

إحصاءات الفقرات					
الفقرات	معامل الارتباط المصحح بين الفقرة والمجموع الكلي	الفقرات	معامل الارتباط المصحح بين الفقرة والمجموع الكلي	الفقرات	معامل الارتباط المصحح بين الفقرة والمجموع الكلي
q1	.638	q16	.663	q31	.587
q2	.642	q17	.695	q32	.679
q3	.704	q18	.619	q33	.704
q4	.540	q19	.658	q34	.609
q5	.676	q20	.594	q35	.651
q6	.678	q21	.521	q36	.641
q7	.634	q22	.638	q37	.716
q8	.716	q23	.652	q38	.740
q9	.673	q24	.615	q39	.657
q10	.690	q25	.704	q40	.584
q11	.564	q26	.630		
q12	.621	q27	.584		
q13	.575	q28	.745		
q14	.726	q29	.620		
q15	.660	q30	.683		

إجراءات جمع البيانات:

1 – تطبيق الاختبار على جميع أفراد العينة.

2 – تسجيل إجابات الطلبة لكل فقرة بدقة لضمان تكامل البيانات.

3 – ربط كل فقرة بالمخرج التعليمي الذي يقيسه لتحليل العلاقة بين أداء الطلبة وتحقيق المخرجات.

البرمجيات المستخدمة:

أ – JMetrik لتحليل فقرات الاختبار وتقدير معاملات نموذج 2PL

ب – برنامج SPSS الإصدار 27 لتحليل البيانات الإحصائية الأساسية واستكشاف العلاقة بين أداء الطلبة وتحقيق المخرجات.

نتائج البحث ومناقشتها:

السؤال الأول: ما هي خصائص فقرات الاختبار (معاملات الصعوبة والتمييز) بعد معايرتها بمعايير (2PL) IRT ؟

يعرض الجدول رقم (5) تقديرات معاملات الصعوبة (b) والتمييز (a) لفقرات الاختبار كما أسفر عنها نموذج الاستجابة للفقرة ثنائي المعلمة (2PL) باستخدام أسلوب التقدير الأقصى الهامشي(MMLE).

جدول رقم (5) معاملات الصعوبة والتمييز للفقرات

تقديرات معلَمات فقرات الاختبار							
رقم الفقرة	رقم المخرج	معامل التمييز	معامل الصعوبة	رقم الفقرة	رقم المخرج	معامل التمييز	معامل الصعوبة
q1	1	1.70	-1.10	q21	3	1.53	0.40
q2	1	1.45	-0.80	q22	3	1.99	0.80
q3	1	1.60	-0.40	q23	3	1.62	1.20
q4	1	1.80	0.00	q24	3	1.18	1.60
q5	1	1.55	0.40	q25	4	1.71	-1.10
q6	1	1.90	0.80	q26	4	1.33	-0.80
q7	1	1.65	1.20	q27	4	1.49	-0.40
q8	1	1.20	1.60	q28	4	1.88	0.00
q9	2	1.75	-1.10	q29	4	1.47	0.40
q10	2	1.30	-0.80	q30	4	1.92	0.80
q11	2	1.50	-0.40	q31	4	1.59	1.20
q12	2	1.85	0.00	q32	4	1.22	1.60
q13	2	1.40	0.40	q33	5	1.66	-1.10
q14	2	1.95	0.80	q34	5	1.38	-0.80
q15	2	1.60	1.20	q35	5	1.54	-0.40
q16	2	1.25	1.60	q36	5	1.81	0.00
q17	3	1.68	-1.10	q37	5	1.50	0.40
q18	3	1.42	-0.80	q38	5	2.10	0.80
q19	3	1.58	-0.40	q39	5	1.69	1.20
q20	3	1.82	0.00	q40	5	1.15	1.60

DOI: https://doi.org/10.35682/mjhss.v41i1

أظهرت نتائج تحليل فقرات الاختبار باستخدام نموذج الاستجابة للمفردة ثنائي المعلمة (2PL) كما في الجدول رقم (5) أن الفقرات تتمتع بخواص سيكومترية قوية؛ فقد تراوحت معاملات التمييز بين 1.18 و2.10، مما يدل على قدرتها العالية على التمييز بين المفحوصين ذوي المستويات المختلفة من القدرة، بينما تراوحت معاملات الصعوبة بين −1.10 و1.60، وهو مدى متوازن يشمل فقرات سهلة ومتوسطة وصعبة. كما خلت المعاملات من القيم المتطرفة، ما يعكس استقرار التقديرات وملاءمة النموذج للبيانات، ويعزز موثوقية الاختبار وصلاحيته للاستخدام البحثي والتربوي.

أظهرت النتائج تمتع الفقرات بقوة تمييزية عالية (1.18<) تجاوزت المعايير المرجعية لدراسة (Uesaka et al., 2022)، مما يؤكد كفاءة نماذج (IRT) في فرز القدرات كما أشارت (Al Ajmi et al., 2024)وخلافاً لتوصية المطيري (2025) باستخدام نماذج معقدة(3PL/4PL) ، أثبتت الدراسة تجريبياً أن نموذج (2PL) يحقق الكفاءة الإحصائية واستقرار التقدير مع العينات المتوسطة.

السؤال الثاني: هل تدل هذه القيم على جودة الفقرات؟

تم فحص ملاءمة الفقرات للنموذج(Item Fit) باستخدام إحصاء $S–X^2$، وقد أظهرت معظم الفقرات مستويات غير دالة إحصائيًا($p > 0.05$) ، مما يدل على اتساق سلوك الاستجابة مع المنطق الاحتمالي لنموذج 2PL، كم في الجدول رقم (6).

جدول رقم (6) احصاء $S–X^2$ لفحص ملاءمة الفقرة للنموذج

إحصائيات ملاءمة الفقرات							
رقم الفقرة	مؤشر الملاءمة كاي–تربيع المعياري للفقرة	درجة الحرية	مستوى الدلالة الإحصائية	رقم الفقرة	مؤشر الملاءمة كاي–تربيع المعياري للفقرة	درجة الحرية	مستوى الدلالة الإحصائية
q1	29.0031	28.0000	0.4124	q21	19.1149	28.0000	0.8946
q2	16.3149	28.0000	0.9609	q22	23.6204	29.0000	0.7474
q3	14.3680	28.0000	0.9844	q23	36.4838	28.0000	0.1307
q4	34.6885	27.0000	0.1469	q24	54.9736	28.0000	0.0017
q5	14.8222	27.0000	0.9719	q25	28.8915	29.0000	0.4707
q6	24.6502	29.0000	0.6963	q26	23.0596	29.0000	0.7738
q7	29.4307	27.0000	0.3403	q27	23.2913	29.0000	0.7630
q8	25.1853	27.0000	0.5641	q28	28.9294	28.0000	0.4161
q9	18.8057	27.0000	0.8770	q29	18.2250	28.0000	0.9203
q10	40.5782	27.0000	0.0452	q30	26.3862	29.0000	0.6048
q11	39.0518	27.0000	0.0627	q31	29.2320	29.0000	0.4530
q12	30.4132	27.0000	0.2959	q32	36.0391	28.0000	0.1416
q13	38.4293	28.0000	0.0905	q33	25.4901	28.0000	0.6011
q14	16.8559	28.0000	0.9514	q34	33.8129	28.0000	0.2071
q15	33.0607	29.0000	0.2753	q35	38.1348	28.0000	0.0959
q16	23.2446	27.0000	0.6718	q36	22.7076	27.0000	0.7006
q17	23.6218	27.0000	0.6512	q37	36.9239	28.0000	0.1205
q18	29.9056	29.0000	0.4187	q38	25.1049	28.0000	0.6221
q19	27.3774	28.0000	0.4978	q39	29.6068	29.0000	0.4338
q20	49.2401	28.0000	0.0079	q40	34.5183	29.0000	0.2208

أظهرت النتائج في الجدول رقم (6) فحص ملاءمة الفقرات باستخدام مؤشر الملاءمة كاي–تربيع المعياري للفقرة وفق نموذج الاستجابة للمفردة ثنائي المعلمة (2PL) أن معظم الفقرات ملائمة للنموذج، حيث كانت مستويات الدلالة الإحصائية غير دالة ($p > 0.05$)، مما يعكس اتساق أنماط استجابة المفحوصين مع افتراضات النموذج وجودة تقديرات معاملات الصعوبة والتمييز. في المقابل، أظهرت ثلاث فقرات فقط (q10، q20، q24) انخفاضًا في مستوى الدلالة، مما يستدعي مراجعتها لتحسين دقة القياس، بينما يعزز الأداء الجيد لبقية الفقرات ثقة الباحثين في صلاحية الاختبار ودقة تقديرات القدرة(θ).

أكدت إحصاءات حسن المطابقة امتثال معظم الفقرات لافتراضات النموذج، متسقةً مع (Al Ajmi et al., 2024) في اعتبار الملاءمة شرطاً لصدق البناء. كما التزمت الدراسة بمعايير (Swaminathan et al., 2006) باستبعاد الفقرات غير الملائمة لضمان 'الاستقلال الموضعي ودقة الاستدلالات السيكومترية.

السؤال الثالث: كيف يساهم توزيع معلومات الفقرات (Item Information) في مستويات القدرة المختلفة في تحسين دقة القرارات على مستوى المخرجات؟

للإجابة عن هذا السؤال، تم تحليل توزيع معلومات الفقرات عبر مستويات القدرة المختلفة (θ) باستخدام دالة معلومات الاختبار الكلي(Test Information Function – TIF) ، إلى جانب تحليل توزيع تقديرات القدرة للطلبة. ويهدف هذا التحليل إلى بيان مدى إسهام بنية المعلومات في الاختبار في تعزيز دقة القرارات التربوية المرتبطة بمخرجات التعلم عند مستويات القدرة المختلفة.

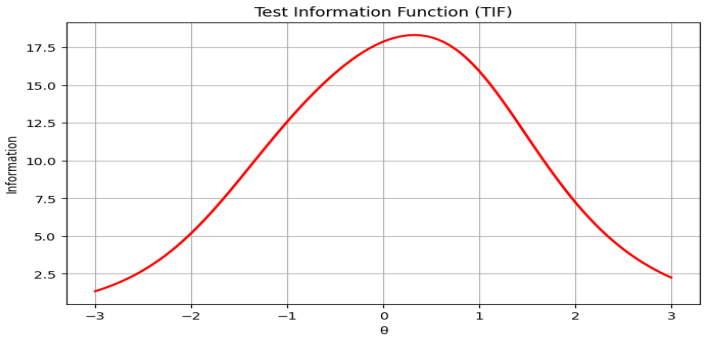
ولغايات فحص دقة القرارات التصنيفية(Classification Accuracy) ، تم تصنيف أفراد العينة إلى ثلاثة مستويات من القدرة(θ) استناداً إلى الدرجات المعيارية؛ حيث حُددت الفئة المتوسطة ضمن النطاق(Mean ±0.60 SD) ، أي الفترة الممتدة بين (0.60- و 0.60+) لوجيت، باعتبارها منطقة القرار الحرج (Critical Decision Zone) التي تتطلب أعلى درجات حساسية القياس للفصل بين الإتقان وعدمه، في حين أُدرج الطلبة خارج هذا النطاق ضمن فئتي القدرة المنخفضة والمرتفعة، وكما يوضح الجدول رقم (7).

جدول رقم (7) توزيع الطلبة حسب مستويات الأداء(θ)

المجموعة	عدد الطلبة	النسبة	نطاق θ
ضعيفون	82	%30	من −5.2 إلى −0.60
متوسطون	138	%50	من −0.60 إلى +0.60
مرتفعون	55	%20	من +0.60 إلى +2.5

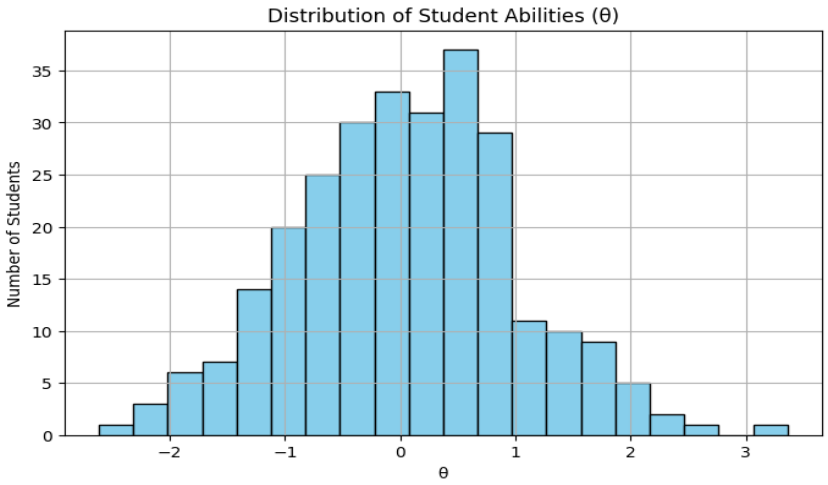
تستند مناقشة النتائج المتعلقة بهذا السؤال إلى المبدأ السيكومتري القائل بأن دقة القياس في نماذج الاستجابة للفقرة (IRT) ليست قيمة ثابتة، بل هي دالة مشروطة بمستوى القدرة تحكمها العلاقة العكسية بين المعلومات والخطأ المعياري ($SE(\theta)=1/\sqrt{I(\theta)}$). وقد كشف تحليل دالة المعلومات (TIF) الوارد في الشكل رقم (3) عن تمركز الكفاءة السيكومترية للاختبار عند الوسط ($\theta\approx0$)، وهو ما يحمل دلالات جوهرية لتباين دقة القرار يمكن تفسيرها في ضوء حالتين متمايزتين:

DOI: https://doi.org/10.35682/mjhss.v41i1



شكل رقم (3) منحنى معلومات الاختبار الكلي

في الحالة الأولى التي تمثل منطقة القرار الحرج (Cut-score zone) والمحددة بالمجال $\theta \in [-0.6, +0.6]$ ، في هذه المنطقة يبلغ منحنى المعلومات ذروته، مما يعني أن الخطأ المعياري للتقدير يكون في أدنى مستوياته. وهذا التمركز يمنح صانع القرار فترة ثقة (Confidence Interval) ضيقة جداً لتصنيف الطلبة الذين تقع درجاتهم على حافة النجاح والرسوب، مما يقلل جذرياً من احتمالية وقوع أخطاء التصنيف (Classification Errors) كالنجاح الزائف أو الرسوب الزائف، وتكتسب هذه الدقة أهمية مضاعفة نظراً لأن هذه الفئة تشكل النسبة الأكبر من العينة (50%) كما هو موضح في الجدول رقم (7) وتوزيع القدرات كما في الشكل رقم (4).



شكل رقم (4) Histogram θ توزيع القدرات

أما في الحالة الثانية التي تمثل الأطراف المتطرفة ($\theta > +2.0$ أو $\theta < -2.0$) فيُلاحظ انخفاض منحنى المعلومات كما في الشكل رقم (3) وارتفاع الخطأ المعياري نسبياً، ورغم أن هذا يشير سيكومترياً إلى دقة أقل في تحديد موقع القدرة بدقة متناهية،

إلا أن أثره الوظيفي على القرار التصنيفي (متقن/غير متقن) يبقى هامشياً وغير مؤثر؛ لأن المسافة الكبيرة التي تفصل مستوى الطالب عن نقطة القطع تقلل من مخاطر اتخاذ قرار خاطئ بحقه.

وتأسيساً على ما سبق، يحقق توزيع المعلومات الحالي مواءمة وظيفية (Functional Alignment) مع أهداف التعليم القائم على المخرجات (OBE)؛ حيث يركز الاختبار حساسيته القياسية لفرز الطلبة بدقة عند الحد الفاصل بدلاً من تشتيت الدقة عبر مستويات لا تؤثر في القرار النهائي. وبهذا تتميز الدراسة الحالية عن الأبحاث النوعية في (OBE) كدراستي (Asbari & Novitasari, 2024) و (Nguyen et al., 2024) بتقديمها تدقيقاً كمياً لدقة القرار، مؤكدةً مبدأ (van der Linden & Hambleton, 1997) حول ضرورة المواءمة الوظيفية بين ذروة المعلومات ونقطة القطع الحرجة.

الاستنتاجات والتوصيات:

على الرغم من دقة الاختبار عند نقطة النجاح (50%)، توصي الدراسة بتوسيع نطاق الصعوبة مستقبلاً ليشمل فقرات أصعب لتمييز الطلبة المتفوقين، حيث إن الدقة الحالية تخدم قرار (ناجح/راسب) بكفاءة عالية، إلا أنها تتراجع عند الانتقال إلى التمييز بين المستويات المتقدمة من الأداء.

خلصت الدراسة إلى أن تطبيق نموذج 2PL على اختبارات مخرجات التعلم كشف عن فجوة بين "تصميم الاختبار" و"هدف القرار". الاختبار الحالي يدعم قرارات الفرز العام، لكنه يحتاج لتطوير ليدعم قرارات الإتقان الدقيقة.

التوصيات التنفيذية:

1 – إعادة هندسة الصعوبة: عدم الاكتفاء بحساب صعوبة الفقرة، بل استهداف قيم صعوبة (b) تطابق نقطة النجاح المعتمدة في الجامعة مثلاً ($\theta = +0.5$) لتعظيم المعلومات عند تلك النقطة.

2 – التنويع في صيغ الفقرات: التقليل من الاعتماد على أسئلة "صح/خطأ" لتقليل أثر التخمين العشوائي الذي يشوه دقة القرار عند المستويات الدنيا من القدرة.

3 – تبني نماذج IRT كمعيار للجودة: اعتماد منحنى معلومات الاختبار (TIF) كوثيقة اعتماد رسمية لأي اختبار نهائي في الجامعة قبل استخدامه لاتخاذ قرارات النجاح والرسوب.

قائمة المراجع

أبو علام، رجاء محمود، *القياس والتقويم التربوي والنفسي*، ط 6، دار المسيرة، 2011.

الزغول، عماد عبد الرحمن، *القياس النفسي*، دار المسيرة، 2012.

عدس، عبد الرحمن، *القياس والتقويم في العملية التعليمية*، دار الفكر، 2010.

علّام، صلاح الدين محمود، *القياس والتقويم التربوي والنفسي*، دار الفكر العربي، 2007.

المطيري، عياد ركا عيادة، "أثر حجم العينة والنموذج على دقة تقدير بارامتر القدرة وفق نظرية الاستجابة للفقرة: دراسة محاكاة"، *مجلة العلوم التربوية والدراسات الإنسانية*، 48، 2025، ص 22–47.

ملحم، سامي محمد، *القياس والتقويم في التربية*، دار المسيرة، 2013.

References

Al Ajmi, M., Mustakim, S. S., Roslan, S., & Almehrizi, R. (2024). Psychometric characteristics of the numerical ability test for Gulf students. International Journal of Evaluation and Research in Education (IJERE), 13(4), 2552–2561. https://doi.org/10.11591/ijere.v13i4.27981

Asbari, M., & Novitasari, D. (2024). Outcome-based education model: Its impact and implications for lecturer creativity and innovation in higher education. International Journal of Social and Management Studies (IJOSMAS), 5(5), 22–31.

Baker, F. B., & Kim, S. H. (2017). The basics of item response theory using R. Springer. https://doi.org/10.1007/978-3-319-54205-8

Barton, M. A., & Lord, F. M. (1981). An upper asymptote for the three-parameter logistic item-response model (ETS Research Report No. RR-81-20). Educational Testing Service. https://doi.org/10.1002/j.2333-8504.1981.tb01255.x

Biggs, J., & Tang, C. (2011). Teaching for quality learning at university: What the student does (4th ed.). McGraw-Hill Education.

Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord & M. R. Novick (Eds.), Statistical theories of mental test scores (pp. 397–479). Addison-Wesley.

Bond, T. G., & Fox, C. M. (2015). Applying the Rasch model: Fundamental measurement in the human sciences (3rd ed.). Routledge. https://doi.org/10.4324/9781315814698

Butler, S. M. (2018). Outcome-based assessment: A framework for improving student learning. Stylus Publishing.

Crocker, L., & Algina, J. (1986). Introduction to classical and modern test theory. Holt, Rinehart and Winston.

de Ayala, R. J. (2022). The theory and practice of item response theory (2nd ed.). Guilford Press.

Embretson, S. E., & Reise, S. P. (2000). Item response theory for psychologists. Lawrence Erlbaum Associates. https://doi.org/10.4324/9781410605269

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

Hambleton, R. K., & Swaminathan, H. (1985). Item response theory: Principles and applications. Springer. https://doi.org/10.1007/978-94-017-1988-9

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

Jackson, N. (2002). Learning outcomes and curriculum development in higher education. Learning and Teaching Support Network.

Lahner, F.-M., Huber, A., & Müller, A. (2020). Standard setting and cut-score determination in educational and medical assessment: A systematic review. BMC Medical Education, 20(1), Article 289. https://doi.org/10.1186/s12909-020-02204-4

Lord, F. M. (1980). Applications of item response theory to practical testing problems. Lawrence Erlbaum Associates. https://doi.org/10.4324/9780203056615

Magis, D. (2013). A note on the item information function of the four-parameter logistic model. Applied Psychological Measurement, 37(3), 304–315. https://doi.org/10.1177/0146621613475471

Malan, S. P. T. (2000). The "new paradigm" of outcomes-based education in perspective. Journal of Family Ecology and Consumer Sciences, 28, 22–28. https://doi.org/10.4314/jfecs.v28i1.52788

Nguyen, C. H., Nong, H. H. T., Saynavong, N., & Nguyen, S. T. (2024). Implementing outcome-based education in higher education programs: A multiple case study in Vietnam and Laos. Vietnam Journal of Education, 8(2), 112–120. https://doi.org/10.52296/vje.2024.331

Reise, S. P., & Waller, N. G. (2009). Item response theory and clinical measurement. Annual Review of Clinical Psychology, 5, 27–48. https://doi.org/10.1146/annurev.clinpsy.032408.153553

Spady, W. G. (1994). Outcome-based education: Critical issues and answers. American Association of School Administrators.

Swaminathan, H., Hambleton, R. K., & Rogers, H. J. (2006). Assessing the fit of item response theory models. In C. R. Rao & S. Sinharay (Eds.), Handbook of statistics: Vol. 26. Psychometrics (pp. 683–718). Elsevier. https://doi.org/10.1016/S0169-7161(06)26020-7

Uesaka, Y., Suzuki, M., & Ichikawa, S. (2022). Analyzing students' learning strategies using item response theory: Toward assessment and instruction for self-regulated learning. Frontiers in Education, 7, Article 886345. https://doi.org/10.3389/feduc.2022.886345

van der Linden, W. J., & Hambleton, R. K. (Eds.). (1997). Handbook of modern item response theory. Springer. https://doi.org/10.1007/978-1-4757-2691-6

van der Linden, W. J., & Hambleton, R. K. (1997). Item response theory: Brief history, common models, and extensions. In W. J. van der Linden & R. K. Hambleton (Eds.), Handbook of modern item response theory (pp. 1–28). Springer. https://doi.org/10.1007/978-1-4757-2691-6_1

Wiggins, G., & McTighe, J. (2005). Understanding by design (Expanded 2nd ed.). Association for Supervision and Curriculum Development.

Wolkowitz, A. A. (2008). A comparison of classical test theory and item response theory methods for equating number-right scored to formula-scored assessments [Doctoral dissertation, University of Kansas]. KU ScholarWorks. http://hdl.handle.net/1808/4255