International Journal for Research in Education

Volume 47
Issue 3 *Vol.(47), issue(3), August 2023*

Article 5

2023

A Comparison of Mantel-Haenszel and the Area Difference Between Item Characteristic Curves Index Techniques for Detecting Differential Item Functioning

Eqbal Darandari PhD
King Saud University, eqbal.darandari1@gmail.com

Muna A. Almeri
King Saud University, mona_almarri@hotmail.com

Follow this and additional works at: https://scholarworks.uaeu.ac.ae/ijre

Part of the Educational Assessment, Evaluation, and Research Commons

Recommended Citation

Darandari, E. Z., & Almeri, M.A. (2023). A comparison of Mantel-Haenszel and the area difference between item characteristic curves index techniques for detecting differential item functioning. International Journal for Research in Education, 47(3), 141-171. http://doi.org/10.36771/ijre.47.3.23-pp141-171

This Article is brought to you for free and open access by Scholarworks@UAEU. It has been accepted for inclusion in International Journal for Research in Education by an authorized editor of Scholarworks@UAEU. For more information, please contact j.education@uaeu.ac.ae.

UAEU كلية التربية College of Education
جامعة الإمارات العربية المتحدة United Arab Emirates University
IJRE

المجلة الدولية للأبحاث التربوية
International Journal for Research in Education

Vol. (47), issue (3) August 2023 - المجلد (47) العدد (3) أغسطس 2023

Manuscript No.: 1986

A Comparison of Mantel-Haenszel and the Area Difference Between Item Characteristic Curves Index Techniques for Detecting Differential Item Functioning

This article extracted from an unpublished master's thesis by the second author

المقارنة بين طريقتي مانتل هانزل ومؤشر فرق المساحة بين منحنيي خصائص الفقرة في الكشف عن الأداء التفاضلي للفقرة

هذا البحث مستل من رسالة ماجستير غير منشورة للباحثة الثانية

Recevied الاستلام	Jan 2022 يناير 2022	Accepted القبول	Jun 2022 يوليو 2022	Published النشر	August 2023 أغسطس 2023

DOI : http://doi.org/10.36771/ijre.47.3.23-pp141-171

Eqbal Z. Darandari, PhD
College of Education, King Saud University,
Saudi Arabia
Muna A. Almeri
College of Education, King Saud University,
Saudi Arabia

د./ إقبال زين العابدين درندري
كلية التربية، جامعة الملك سعود–
المملكة العربية السعودية
dr.e.darandari@gmail.com
أ./ منى علي المري
كلية التربية، جامعة الملك سعود–
المملكة العربية السعودية

حقوق النشر محفوظة للمجلة الدولية للأبحاث التربوية
ISSN : 2519-6146 (Print) - ISSN : 2519-6154 (Online)

المجلة الدولية للأبحاث التربوية - جامعة الإمارات العربية المتحدة
المجلد (47) العدد (3) أغسطس 2023

A Comparison of Mantel-Haenszel and the Area Difference Between Item Characteristic Curves Index Techniques for Detecting Differential Item Functioning

Abstract

The study is aimed at comparing between Mantel-Haenszel-MH chi-square and the area difference between three item characteristic curves Index -3ICCs by using Raju's index to detect the differential item functioning DIF in post-graduate general aptitude test according to gender variable. The study sample of (2280) students was distributed equally between male and female groups. The MH techniques revealed a uniform DIF in (43) items, equal to (45%) of the items in post-graduate general aptitude test. 3ICCs technique, revealed a uniform DIF based on gender in two items, equals to (2%) of item's test. MH technique revealed more efficiency in detecting differential item functioning in post-graduate general aptitude test compared to area difference between 3ICCs. The agreement rate was low between these two techniques. Based on these results, the researchers recommended to use more than one technique when detecting the differential item functioning.

Keywords: Differential Item Functioning, Mantel-Haenszel, The three-parameter logistic model, post-graduate general aptitude test.

المقارنة بين طريقتي مانتل هانزل ومؤشر فرق المساحة بين منحنيي خصائص الفقرة في الكشف عن الأداء التفاضلي للفقرة

مستخلص البحث

هدف البحث إلى المقارنة بين طريقتي مانتل هانزل ومؤشر فرق المساحة بين منحنيي خصائص الفقرة للنموذج اللوجستي ثلاثي المعلم في الكشف عن الأداء التفاضلي لفقرات اختبار القدرات العامة للجامعيين وفقاً لمتغير الجنس. وتكونت عينة البحث من (2280) طالباً وطالبةً موزعةً بالتساوي بين مجموعتي الذكور والإناث. كشفت نتائج التحليلات الإحصائية باستخدام طريقة مانتل-هانزل كاي تربيع عن وجود أداءٍ تفاضليٍّ منتظمٍ في (43) فقرة بنسبةٍ تُعادل (45%) من فقرات الاختبار وفقاً لمتغير الجنس. كما كشفت طريقة مؤشر قياس المساحة بين منحنيي خصائص الفقرة عن أداءٍ تفاضليٍّ منتظمٍ في فقرتين بنسبةٍ تعادل (2%) من فقرات الاختبار. أظهرت طريقة مانتل-هانزل فاعليةً أكبر في الكشف عن الأداء التفاضلي لفقرات اختبار القدرات العامة للجامعيين مقارنةً بطريقة مؤشر قياس المساحة بين منحنيي خصائص الفقرة، وكانت نسبة الاتفاق بين الطريقتين منخفضةً. وبناء على النتائج التي توصل إليها البحث، أوصت الباحثتان بضرورة استخدام أكثر من طريقة عند الكشف عن الأداء التفاضليّ للفقرة.

الكلمات المفتاحية: الأداء التفاضلي للفقرة، مانتل هانزل، النموذج اللوجستي ثلاثي المعالم، اختبار القدرات العامة للجامعيين.

المجلة الدولية للأبحاث التربوية - جامعة الإمارات العربية المتحدة
المجلد (47) العدد (3) أغسطس 2023

المقدمة

يعدُّ تحليل الأداء التفاضلي للفقرة جُزءاً من مراحل بناء الاختبار وصدقه؛ لأنه يضمن أن درجات الأفراد على الاختبار تعكس نفس البنية أو التكوين للأفراد المتكافئين في السمة المقاسةِ (Walker, 2011). فيما يُعدّ خلو الاختبار وفقراته من الأداء التفاضلي من أهم الشروط التي ينبغي توافرها في الاختبار قبل نشْره. وقد وضعته الجمعية الأمريكية للبحث التربوي (AERA)، والجمعية الأمريكية لعلم النفس (APA) ، والمجلس الوطني للقياس التربوي بأمريكا (NCME)، ضمن قواعد النشر للاختبارات فأشارت إلى أنه إذا استخدمت الاختبارات في اتخاذ القرارات فلا بد من الحصول على أدلةٍ تجريبيةٍ واضحةٍ ومُلزمةٍ من أجل دعم الادعاءات بأن فقرات الاختبار لا تعكس أداءً تفاضليًّا للفقرة عبر المجموعات الهامة من المختبرين، وأنه يجب على مطوِّري الاختبارات أن يشيروا إلى طبيعة الأدلة المتعلقة بملاءمة الاختبار للمجموعات ذات الخلفيات العرقية أو اللغوية المختلفة. كما يجب دراسة أداء المتلقين للاختبار من أعراقٍ مختلفةٍ وبأحجامِ عيناتٍ كافيةٍ للتأكد من أن الفروق في الأداء ترتبط بالمهارات التي يجرى تقييمها أكثر مِن كوْنها ترتبط بعوامل لا علاقةَ لها بالاختبار (American Educational Research Association، American Psychological Association & National Council on Measurement in Education, 2014).

استُخْدم مصطلح الأداء التفاضليّ للفقرة Differential Item Functioning (DIF) منذ بداية ثمانينيات القرن العشرين للدلالة على الطرق والمعالجات الإحصائية التي تستخدم للكشف عن تحيّز فقرات الاختبارات (Camilli & Shepard, 1994). ورغم أن مفهوم التحيز يحتمل تفسيراً إحصائيًّا إلا أنه ارتبط بدلالاتٍ ثقافيةٍ واجتماعيةٍ سلبيةٍ حول عدم العدالة والمساواة. لذلك حاول الباحثون والمهتمون بالقياس والتقويم إيجادَ مصطلحٍ يعطي دلالاتٍ سيكومترية أكثر، فاستخدم مفهوم الأداء التفاضليّ للفقرة (Ellis & Raju, 2004). فإذا ظهر أداء تفاضلي في فقرة فلا يعني ذلك بالضرورة أن الفقرة متحيزة حتى يمكن الحكم على الفقرة أنها متحيزة تعرض على هيئة من الخبراء لتقرر ما إذا كان مصدر الأداء التفاضلي ذا صلة، أو غير ذا صلة بالبعد الذي يقيسه الاختبار (دي إيالا، 2017/2009).

وترتكز طرق وأساليب الكشف عن الأداء التفاضلي للفقرة على مدخلين أساسين هما: النظرية التقليدية للاختبارات Classical Test Theory (CTT)، والنظرية الحديثة المعروفة بنظرية الاستجابة للفقرة Item Response Theory (IRT). وتختلف الطرق المستخدمة في الكشف عن الأداء التفاضليِّ للفقرة باختلاف أساسياتها النظرية، وإجراءاتها الإحصائية، كما تتباين هذه الطرق فيما بينها من حيث فاعليتها في الكشف عن الأداء التفاضلي، وقوة اختبارها الإحصائي، ونوع الأداء التفاضلي الذي تكشف عنه. وتتباين نتائج الدراسات حول مدى دقة طرق الكشف عن

الأداء التفاضلي (الأحمدي، 2020). فلا يمكن القول إن طريقة ما هي الأفضل على الإطلاق، أو تستطيع الكشف عن جميع فقرات الاختبار التي تبدي أداءً تفاضليًا. إلا أن هناك عددًا من الأساسيات التي يجب أخذها بعين الاعتبار عند اختيار أيٍّ من هذه الطرق منها: طبيعة البيانات، وتوزيع القدرة بين المجموعات الفرعية، وحجم العينة، وطول الاختبار، وأبعاده، ونوع الأداء التفاضلي المستهدف، وسهولة العمليات الحسابية، والبرامج الإحصائية المتوفرة (Benito et al., 2010).

مفهوم الأداء التفاضليّ للفقرة

الأداء التفاضليّ للفقرة دالةٌ مشتقةٌ إحصائيًا للتعبير عن الفرق في الاستجابة للفقرة بين مجموعتين عند مستوى قدرةٍ واحدٍ. ويعرف هامبلتون وروجر (Hambleton & Rogers, 1989) الأداء التفاضليّ للفقرة بأنه الفروق في احتمالات الإجابة الصحيحة للفقرة في المجموعات المختلفة المتساوية في القدرة. ويرى هامبلتون وسوامينثان وروجرز أن الأداء التفاضليّ للفقرة يظهر عندما تكون دوال الاستجابة على الفقرة "المنحنيات المميزة للفقرة" غير متشابهةٍ في المجموعات الفرعية المختلفة (Hambleton, Swaminathan & Rogers, 1991).

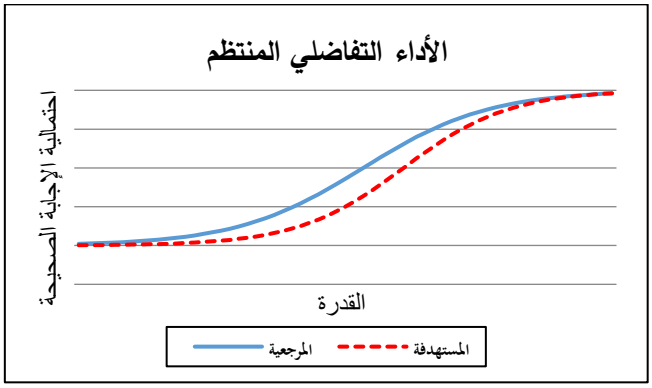
والبعض يستخدم مصطلح الأداء التفاضليّ للفقرة كمرادفٍ لمصطلح التحيز، ومثل هذا الاستخدام ليس له ما يبرره، فمصطلح الأداء التفاضليّ للفقرة يُستخدم لتحديد الفقرات التي يكون احتمال الإجابة الصحيحة عليها مختلفًا بين مجموعتين لهما نفس مستوى القدرة. وعند تفسير أسباب ذلك يمكن وصف الفقرات ذات الأداء التفاضليّ بأنها فقرات متحيزة لمجموعةٍ ما دون الأخرى لأسبابٍ لا تتعلق بقدرة الأفراد فقط. ويعدُّ الأداء التفاضليّ للفقرة شرطاً أساسياً لاعتبار الفقرة متحيزةً، لكنه شرطٌ غير كافٍ لذلك، فإذا أظهرت فقرة ما أداءً تفاضليًّا لمجموعةٍ ما، عندها لا بد من إجراءاتٍ إضافيةٍ تتمثل في عرض الفقرة على الخبراء لمراجعتها للحكم عليها بأنها متحيزة (Camilli & Shepared, 1994). وبذلك يمكن القول إن الكشف عن الأداء التفاضليّ للفقرة إجراءٌ إحصائيٌّ ضمن عددٍ من الإجراءات التي يتخذها معدّو ومُقيّمو الاختبار للحكم على تحيز فقراته.

وينقسم الأداء التفاضليّ إلى نوعين؛ أداءٌ تفاضليٌّ منتظمٌ "موحدٌ" Uniform Differential Item Functioning (UNDIF)، وأداءٌ تفاضليٌّ غير منتظمٌ "غير موحدٌ" Non Uniform Differential Item Functioning (Non-UNDIF). ويحدث الأداء التفاضليّ المنتظم عندما يكون الفرق بين منحني خصائص الفقرة (ICCs) لمجموعة الأفراد المرجعية والمستهدفة ثابتة عبر جميع مستويات القدرة كما هو موضحٌ في الشكل (1). ومن الطرق التي تستخدم للكشف عن هذا النمط طريقة مانتل- هانزل Mantel-Haenszel(MH)، والطريقة المعيارية Standardization Method، وطرق كاي تربيع Chi-square Methods.

المجلة الدولية للأبحاث التربوية - جامعة الإمارات العربية المتحدة
المجلد (47) العدد (3) أغسطس 2023

شكل 1

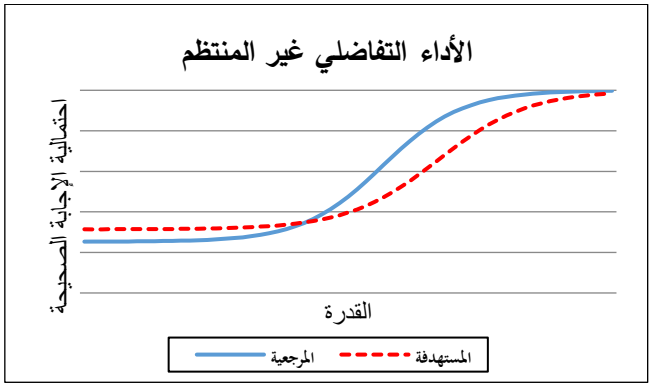
منحنيا خصائص الفقرة في نوع الأداء التفاضلي المنتظم



أما الأداء التفاضليّ غير المنتظم فيحدث عندما يكون الفرق بين منحنيي خصائص الفقرة (ICCs) لمجموعة الأفراد المرجعية والمستهدفة غير ثابتٍ عبر جميع مستويات القدرة (Walker, 2011)، كما هو موضحٌ في الشكل (2). ومن الطرق التي تكشف عن هذا النمط، بالإضافة لنمط الأداء التفاضليّ المنتظم، طريقة اختبار تحيز الفقرة المتزامن Simultaneously Item Bias Test (SIBTEST)، وطريقة الانحدار اللوجستي Logistic Regression (LR)، واختبار نسبة الأرجحية لنظرية الاستجابة للفقرة IRT Likelihood-Ratio Test (IRTLR)، وطرق مؤشر معلم الصعوبة b parameter index، وطرق منحنى خصائص الفقرة لنظرية الاستجابة الفقرة Item Characteristic Curve Methods، وطريقة لورد كاي تربيع Lord's Chi-Sqaure، والنماذج الخطية اللوغاريتمية Log linear models.

شكل 2

منحنيا خصائص الفقرة في نوع الأداء التفاضلي غير المنتظم



وتنقسم طرق الكشف عن الأداء التفاضليّ للفقرة من حيث مفاهيم كل طريقةٍ وإجراءاتها إلى قسمين رئيسيين حسب نظريتي القياس التقليدية والحديثة. وتعتبر طريقة مانتل-هانزل كاي تربيع التي قدمها مانتل وهانزل في عام 1959م (Mantel & Haenszel, 1959)، وطبقت من قبل هولند وثاير في عام 1988م من أشهر طرق النظرية التقليدية، وأكثرها استخداماً في الكشف عن الأداء التفاضليّ للفقرة، وذلك لسهولة حساباتها وإجراءاتها. وتقوم هذه الطريقة باستقصاء التحيز بين مجموعتين إحداهما تسمى المجموعة المرجعية والأخرى تسمى المجموعة المستهدفة وهي المجموعة التي يعتقد أنها تتأثر بالأداء التفاضليّ للفقرة، بمعنى أن الفقرات تتحيز ضدها. وتعتمد على الاختلاف في نسبة الإجابات الصحيحة والخاطئة، بين المجموعة المستهدفة والمرجعية في كل مستوى من مستويات القدرة باستخدام الدرجة الكلية للاختبار. وتستخدم طريقة مانتل-هانزل للكشف عن الأداء التفاضليّ في الفقرات ثنائية الاستجابة والفقرات متعددة الاستجابة. وتوفر طريقة مانتل-هانزل كلاً من الدلالة الإحصائية للاختبار، وتقدير حجم تأثير الأداء التفاضليّ للفقرة (Holland & Thayer, 1988). وتختبر طريقة مانتل هانزل الفرضية الصفرية التي تنصُّ على أنه لا توجد علاقةٌ بين عضوية الفرد في المجموعة وأدائه على الاختبار.

ويتم تحديد الأداء التفاضليّ للفقرة من خلال قيمة دلالة إحصائي مانتل-هانزل، فإذا كانت قيمة الاحتمالية Probability Value (P-Value) أقل من أو يساوي 0,05 فإن ذلك يعني أن الفقرة تظهر أداءً تفاضليًّا، أما إذا كانت أكبر فهذا يعني أن الفقرة لا تظهر أداءً تفاضليًّا. كما يُحدد الأداء التفاضليّ للفقرة من خلال قيمة نسبة الأرجحية التي تمتد من الصفر إلى ما لا نهاية موجب (+∞)، فإذا كانت قيمة نسبة الأرجحية تساوي الواحد الصحيح فالفقرة لا تظهر أداءً تفاضليًّا، وإنْ كانت أقل من الواحد الصحيح فالفقرة تظهر أداءً تفاضليًّا لصالح المجموعة المستهدفة، وإنْ كانت أكبر من الواحد الصحيح فالفقرة تظهر أداءً تفاضليًّا لصالح المجموعة المرجعية. واقترح مركز خدمات الاختبارات التربوية، قيمة دلتا مانتل-هانزل لتقدير حجم الأداء التفاضليّ. فينقسم حجم الأداء التفاضلي إلى ثلاثة أنواع هي: النوع A ويشير إلى وجود أداء تفاضليّ ضئيلٌ لا يكاد يذكر، والنوع B ويشير إلى وجود أداء تفاضليّ متوسط، والنوع C ويشير إلى وجود أداء تفاضليّ كبير (Zieky, 1993).

ولا تتطلب طريقة مانتل هانزل أحجام عينات كبيرة حيث تقترح عدد من الدراسات التجريبية على بيانات حقيقية ودراسات المحاكاة 200 فرد كحد أدنى لحجم العينة لكل من المجموعتين المرجعية والمستهدفة عند استخدام طريقة مانتل هانزل (Paek & Guo, 2011).

وساعد ظهور النظرية الحديثة للقياس في معالجة كثيرٍ من عيوب النظرية التقليدية، وذلك لتمتعها بعددٍ من الخصائص التي تميزها عن النظرية التقليدية وتجعلها تلبي احتياجات الباحثين ومطوري الاختبارات ومنها:

المجلة الدولية للأبحاث التربوية - جامعة الإمارات العربية المتحدة
المجلد (47) العدد (3) أغسطس 2023

- استقلال تقديرات الخصائص السيكومترية لفقرات الاختبار عن عينة الأفراد المستخدمة في تقدير هذه الخصائص.
- استقلال تقدير معلم قدرة الفرد عن الفقرات التي طبقت عليه.
- توفير مؤشرٍ إحصائيٍّ لتقدير درجة الدقة في قياس القدرة لكل فردٍ، وتقدير الدقة في قياس خصائص الفقرات أيضاً.

مما يوفر خاصية اللاتغاير في نظرية الاستجابة للفقرة والتي ساهمت في تقديم حلولٍ لعددٍ من المشاكل التي كانت تواجه معدي الاختبارات ومطوريها ومنها الكشف عن تحيز الفقرات. وترتكز نظرية الاستجابة للفقرة على مجموعةٍ من الافتراضات التي يجب توافرها في بيانات عينة الاختبار، وهي:

1. أحادية البعد Unidimensionality حيث تفترض نماذج الاستجابة للفقرة وجود قدرةٍ أو سمةٍ واحدةٍ تفسر أداء الفرد في الاختبار، مما يعني أن فقرات الاختبار ترتبط بسمةٍ أو خاصيةٍ محددةٍ تقيسها (Hambleton & Swaminathan, 1985). ويمكن التأكد من هذا الفرض من خلال التحليل العاملي حيث إن تفسير العامل الأول لأكبر قدرٍ من التباين الكلي يعتبر مؤشراً على أن الاختبار أحادي البعد (Hattie, 1985). ويمكن استخدام نموذج السمات متعددة الأبعاد عند مخالفة فقرات الاختبار لفرض أحادية البعد.

2. الاستقلال الموضعي أو المحلي Local Independency والذي يعني أن استجابات الأفراد على فقرات الاختبار تكون مستقلةً إحصائيًا عند مستوى قدرةٍ معينةٍ، أي أن استجابة الفرد لفقرة ما يجب ألا تؤثر سلباً أو إيجاباً على استجابته لفقرة أخرى. ويرى هامبلتون وسوامينثان أن الاستقلال الموضعي لا يعني أن الفقرات لا ترتبط مع بعضها البعض، بل على العكس إذ يجب أن ترتبط الفقرات بارتباطاتٍ عاليةٍ، وذلك لتحقيق افتراض أحادية البعد في الاختبار (Hambleton & Swaminathan, 1985).

3. منحنى خصائص الفقرة Item Characteristic Curve (ICC) وهو عبارةٌ عن وصفٍ للعلاقة بين احتمالية أداء الفرد على الفقرة وقدرته المقاسة بالاختبار من خلال اقترانٍ تراكمي صاعدٍ يطلق عليه منحنى خصائص الفقرة. ويوفر احتمالات إجابة الفرد إجابةً صحيحةً عن الفقرة في مستويات القدرة المختلفة، ويختلف شكل المنحنى باختلاف النموذج المستخدم، أحادي المعلم، أو ثنائي المعلم، أو ثلاثي المعلم.

4. عامل التحرر من السرعة في الإجابة Speededness Free تفترض نماذج نظرية الاستجابة للفقرة أن عامل السرعة لا يلعب دوراً في الإجابة عن الفقرة، وأن إخفاق الفرد

في الإجابة عن الفقرة في الاختبار يرجع إلى تدني قدرته، وليس إلى تأثير عامل السرعة في الأداء (Hambleton, Swaminathan & Rogers, 1991).

وتتعدد نماذج نظرية الاستجابة للفقرة، وتختلف باختلاف تحقق افتراضات النظرية المتعلقة ببيانات الاختبار. ومن أهم نماذج نظرية الاستجابة للفقرة أحادية البعد النماذج اللوجستية وهي:

1. النموذج اللوجستي أحادي المعلم One Parameter Logistic Model

هو أبسط نماذج نظرية الاستجابة للفقرة، إلا أنه أكثرها تعقيداً في عدد الافتراضات اللازم توفرها في البيانات، حيث يفترض أن تكون الاستجابة على الفقرات ثنائية إما (1) أو (صفر)، كما يفترض تساوي معاملات التمييز للفقرات، واقتراب معلم التخمين من الصفر، وبالتالي فإن الفقرات تتباين في صعوبتها لذلك يتم تقدير معلم صعوبة الفقرات Difficulty parameter فقط، بحيث تكون منحنيات خصائص الفقرات (ICC) متوازية.

2. النموذج اللوجستي ثنائي المعلم Two Parameter Logistic Model

ويقوم على تقدير معلمتي الصعوبة والتمييز للفقرة Discrimination parameter، بينما يفترض انعدام التخمين باقتراب قيمته من الصفر، مما يؤدي إلى تقاطع المنحنيات المميزة لفقرات الاختبار (ICC)، حيث يفترض أن الفقرات تختلف في درجة صعوبتها وقدرتها التمييزية في ضوء المستويات المختلفة للقدرة.

3. النموذج اللوجستي ثلاثي المعلم Three Parameter Logistic Model

ويُسمى بنموذج التخمين، حيث أضاف لورد معلماً ثالثاً أطلق عليه معلم الخط التقاربي الأدنى، وهو ما يعرف بمعلم التخمين Guessing parameter، حيث يكون التخمين عاملاً مؤثراً في الأداء مما يعني احتمال استجابة فردٍ من ذوي مستوى القدرة المنخفضة على الفقرة بشكلٍ صحيحٍ. ويتلاءم تطبيق النموذج اللوجستي ثلاثي المعلم مع الفقرات من نمط الاختيار من متعددٍ والصواب والخطأ. وهو النموذج المستخدم في هذا البحث (Hambleton, Swaminathan, & Rogers, 1991).

وعند اختيار أحد النماذج الثلاثة السابقة لابد من التحقق من مطابقة البيانات للنموذج المستخدم من خلال مطابقة معلم القدرة للأفراد، ومعالم الفقرات للنموذج المستخدم حتى يتمكن النموذج من التنبؤ بالدرجات الفعلية للاختبار. وتوصى الدراسات بأن يكون حجم العينة أكبر من 1000 فرد وذلك لتوفير درجة معقولة من الدقة في تقدير معالم الفقرات وللحد من المشكلات التي يتعرض لها أحياناً التدريج باستخدام النموذج اللوجستي ثلاثي المعلم (دي إيالا، 2017/2009).

المجلة الدولية للأبحاث التربوية - جامعة الإمارات العربية المتحدة
المجلد (47) العدد (3) أغسطس 2023

وتعتمد طرق الكشف عن الأداء التفاضليّ المبنية على إجراءات النظرية الحديثة على منحنى خصائص الفقرة في الحكم على وجود الأداء التفاضليّ من عدمه. ويعد مؤشر المساحة بين المنحنيات، وفحص الفروق بين المعلمات عبر المجموعات الفرعية، الأكثر استخداماً لكشف عن الأداء التفاضليّ في حال تبني النموذج اللوجستي ثلاثي المعلم لرسم منحنيات الفقرة. حيث تظهر الفقرة أداءً تفاضليًا عندما تكون منحنيات خصائصها غير متماثلة في كل مجموعةٍ من المجموعات الفرعية موضع المقارنة.

وتبدأ عملية الكشف عن الأداء التفاضليّ للفقرة بدلالة الفرق في المساحة بين منحنيي خصائص الفقرة كما أوردها هامبلتون وروجرز (Hambleton & Rogers, 1989)، بإيجاد قيمة نقطة القطع (Cut-Off Value) والتي تعتبر معياراً للحكم على وجود الأداء التفاضليّ من عدمه. ويمكن إيجادها باتباع الخطوات التالية:

1. تحديد المجموعة المرجعية والمجموعة المستهدفة من المجموعتين الفرعيتين حسب المتغير موضع البحث.
2. تقسيم المجموعة المرجعية إلى مجموعتين عشوائيتين متكافئتين.
3. تقدير معالم الفقرة لكل مجموعةٍ عشوائيةٍ على حدة حسب النموذج المستخدم.
4. حساب فرق المساحة بين منحنيي خصائص الفقرة للمجموعتين العشوائيتين.
5. تحديد نقطة القطع وتساوي أكبر قيمة لفرق المساحة بين منحنيي خصائص الفقرة للمجموعتين العشوائيتين وتعرف بالمساحة الحرجة.
6. تقدير منحنيات خصائص الفقرة ومعلم القدرة لكل مجموعة فرعية من مجموعتي البحث.
7. حساب فرق المساحة بين منحنيي خصائص الفقرة للمجموعتين الفرعيتين وتعرف بالمساحة الفعلية.
8. الكشف عن وجود أداءٍ تفاضليٍّ للفقرة من خلال المقارنة بين قيمة المساحة الفعلية والمساحة الحرجة، فإذا كانت المساحة الفعلية أكبر من المساحة الحرجة دلَّ ذلك على وجود أداءٍ تفاضليٍّ للفقرة، وإذا كانت المساحة الفعلية أصغر من المساحة الحرجة فذلك يعني عدم وجود أداءٍ تفاضليٍّ للفقرة.

وقدم راجو صيغة لحساب الفرق في المساحة بين منحنيي خصائص الفقرة تعرف بالمساحة الإشارية (Signed Area) (Raju,1988) حيث يتم حساب المساحة بالاعتماد على الفروق بين منحنيات الفقرة، وتدل على الأداء التفاضليّ المنتظم، والمساحة غير الإشارية

(Unsigned Area) وتدل على الأداء التفاضليّ غير المنتظم حيث يتم حساب المساحة بالاعتماد على القيمة المطلقة للفروق بين المنحنيات من خلال المعادلة التالية عند حساب المساحة بين منحنيات خصائص الفقرة في حالة استخدام النموذج اللوجستي ثلاثي المعلم:

$$Area = (1 - C)\left|\frac{2(a_2 - a_{1})}{Da_1a_2}\, In\left[1 + exp\left(\frac{Da_1\, a_2\, (b_2 - b_1)}{a_2 - a_1}\right)\right] - (b_2 - b_1)\right|$$

حيث:

Area: المساحة بين منحني خصائص الفقرة.

a_1: قيمة معلم التميز للمجموعة الأولى.

a_1: قيمة معلم التميز للمجموعة الثانية.

b_1: قيمة معلم الصعوبة للمجموعة الأولى.

b_2: قيمة معلم الصعوبة للمجموعة الثانية.

C: قيمة معلمة التخمين.

D: قيم ثابتة تساوى 1.7.

ويفترض لاستخدام هذه الصيغة تساوي معلم التخمين في المجموعتين (C= C1 =C2)، وفي حالة عدم تساوي معلمة التخمين للمجموعتين فإن المساحة الإشارية تتراوح بين (+∞) إلى (-∞)، والمساحة غير الإشارية تساوي (+∞). وفي حال عدم تساويها بين المجموعتين، يتم عادة تثبيت قيمة معلمة التخمين للمجموعتين عند (C=0.2) لحساب فرق المساحة بين منحني الفقرة، أو بحساب المتوسط الحسابي لمعلمتي التخمين للمجموعتين الفرعيتين. ويمكن استخدام المساحة لراجو كذلك في حال النموذج أحادي المعلمة وثنائي المعلمة (Raju ,1990).

ومن الدراسات التي قارنت بين طريقتي مانتل-هانزل، ومؤشر فرق المساحة بين منحني خصائص الفقرة الكشف عن الأداء التفاضليّ، دراسة هامبلتون وروجرز (Hambleton & Rogers, 1989) التي هدفت إلى المقارنة بين طريقتي مانتل-هانزل، وطريقة مؤشر فرق المساحة بين منحني خصائص الفقرة وفق النموذج اللوجستي ثلاثي المعلم، في الكشف عن الأداء التفاضلي لفقرات اختبار نيومكسيكو للكفاءة للصفوف العليا وفقاً لمتغير العرق والمعروفة بتحيزها مسبقاً. أظهرت نتائج تحليل (75) فقرة من أصل (150) فقرة لعينة مكونة من (1000) فرد، اتفاق نتائج الطريقتين في الكشف عن الأداء التفاضليّ مع تصنيف الفقرات إلى متحيزةٍ وغير متحيزةٍ بنسبةٍ تراوحت بين (75-80 %)، وتشابه نتائج الطريقتين في الكشف عن الأداء التفاضليّ المنتظم، بينما

كانت قدرة طريقة مؤشر فرق المساحة بين منحنيي خصائص الفقرة أعلى من طريقة مانتل-هانزل في الكشف عن الأداء التفاضليّ غير المنتظم.

وأجرى كل من باغهي وفيرارا (Baghi & Ferrara, 1989) بحثاً بهدف المقارنة بين طرق منحنى خصائص الفقرة (النموذج اللوجستي أحادي المعلم، والنموذج اللوجستي ثلاثي المعلم) وطرق النظرية التقليدية في القياس (تحويل صعوبة الفقرة، مانتل-هانزل) في الكشف عن الأداء التفاضليّ لفقرات اختبار ميرلاند لمهارات المواطنة Maryland Test of Citizenship Skills (MTCS) المكون من (45) فقرة، وفقاً لمتغيري العرق والجنس مع الأخذ بعين الاعتبار أثر اختلاف حجم العينة التي كانت (200-500-750-1000) فرداً. وأظهرت النتائج اتفاقاً عالياً بين الطرق المستخدمة في الكشف عن الأداء التفاضليّ للفقرة، حيث أظهرت نتائج طريقة مانتل-هانزل مطابقةً عاليةً لنتائج طريقتي منحنى خصائص الفقرة في الكشف عن الفقرات التي أظهرت أداءً تفاضليًّا. وكانت طريقة منحنى خصائص الفقرة أكثر فاعلية في الكشف عن الفقرات التي تبدي أداءً تفاضليًّا.

وهدفت دراسة سكاجز وليستز (Skaggs & Lissitz, 1992) إلى الكشف عن درجة اتساق عددٍ من طرق الكشف عن الأداء التفاضليّ للفقرة باختلاف طرق تطبيق الاختبار باستخدام (96) فقرة من فقرات اختبار الرياضيات التحصيلي. وتم التطبيق على (6600) طالباً وطالبةً. واستخدم الباحثان سبعة طرقٍ، أربعة منها تستند لإجراءات النظرية الحديثة في القياس، وهي: طريقة المساحة بين منحنيي خصائص الفقرة، وطريقة لورد كاي تربيع، وطريقة مجموع مربعات الفروق بين المعالم المختلفة للفقرة الإشارية Signed Sum of Squares، وطريقة مجموع مربعات الفروق بين المعالم المختلفة للفقرة غير الإشارية Unsigned Sum of Squares، وثلاثة طرقٍ تنتمي لنظرية القياس التقليدية، وهي: طريقة مانتل-هانزل المحولة إلى دلتا، وطريقة مانتل- هانزل باستخدام كاي تربيع، وطريقة كاي تربيع. وأظهرت النتائج باستخدام معامل الارتباط الرتبي كمؤشرٍ على استقرار الطرق أن طرق مجموع مربعات الفروق بين المعلم أكثر الطرق استقراراً، تليها طريقة مانتل-هانزل، بينما كانت طرق كاي تربيع، والمساحة بين منحنيي خصاص الفقرة، وطريقة لورد كاي تربيع أقلها استقراراً.

وفي دراسة لراجو ودراسجو وسلينند (Raju, Drasgow & Slind, 1993) بهدف المقارنة بين أربعٍ من طرق الكشف عن الأداء التفاضليّ للفقرة وفقًا لمتغير الجنس والعرق، وهي: طريقة المساحة الإشارية Signed Area، وطريقة المساحة غير الإشارية Unsigned Area، وطريقة لورد كاي تربيع التي تستند لنظرية الاستجابة للفقرة، وطريقة واحدة تستند لنظرية القياس التقليدية وهي طريقة مانتل-هانزل. واستخدمت فقرات الجزء اللفظي من اختبار جيتس ماكجانيت للقراءة Gates-MacGinitie Reading Tests (GMRT)، طبق على (839) طالباً وطالبةً. وأظهرت النتائج قدرة طريقة مانتل-هانزل على الكشف عن أكبر عددٍ من الفقرات التي تظهر أداءً تفاضليًّا،

كما أظهرت وجود ارتباط عالٍ بين طرق نظرية الاستجابة للفقرة الثلاثة وطريقة مانتل-هانزل في الكشف عن الأداء التفاضليّ وفقاً لمتغير الجنس فقط.

وأجرى شيو (Chiu, 2008) دراسةً بهدف المقارنة بين طريقتي مانتل-هانزل، والمساحة بين منحني خصائص الفقرة في الكشف عن الأداء التفاضليّ لفقرات اختبار إتقان اللغة الإنجليزية المكون من (84) فقرة، وذلك على عينةٍ حجمها (1088) من المتقنين للغة الإنجليزية وغير المتقنين لها. أظهرت النتائج قدرة أعلى لطريقة المساحة بين منحني خصائص الفقرة في الكشف عن الأداء التفاضلي لفقرات الاختبار بنسبة (21%) مقارنة بطريقة مانتل-هانزل التي كشفت عما نسبته (6%) من فقرات الاختبار. وكانت نسبة الاتفاق بين الطريقتين (13%).

ومن الدراسات العربية التي هدفت إلى المقارنة بين طرق الكشف عن الأداء التفاضليّ، دراسة للبستنجي (2004) للمقارنة بين طريقتين من طرق النظرية التقليدية للكشف عن الأداء التفاضليّ للفقرة (طريقة مانتل-هانزل، والانحدار اللوجستي)، وطريقتين من طرق النظرية الحديثة (طريقة مؤشر الفرق في المساحة لمنحنيي خصائص الفقرة ثنائي المعلم، وثلاثي المعلم) في اختبار يقيس ثلاث قدراتٍ عقليةٍ خاصةٍ (القدرة اللفظية، والرياضية، والمكانية) على البيئة الأردنية وفقاً لمتغير الجنس. بلغت عينة البحث (950) طالباً وطالبةً. أظهرت النتائج أن طريقة مانتل-هانزل أكثر فاعلية في الكشف عن الأداء التفاضلي، كما كانت نسبة الاتفاق بينها وبين الطرق الأخرى متدنية.

وهدفت دراسة العطيوي (2004) إلى الكشف عن الأداء التفاضليّ لمتغير الجنس لفقرات اختبار العلوم العامة التحصيلي في الأردن المكون من (60) فقرة، وطُبِق على عينةٍ مكونةٍ من (575) فرداً من الجنسين، باستخدام طريقتي مانتل-هانزل والمساحة بين منحنيات خصائص الفقرة أظهرت النتائج قدرة أعلى لطريقة المساحة بين منحنيي خصائص الفقرة في الكشف عن الأداء التفاضلي لفقرات الاختبار بنسبة (35%) مقارنة بطريقة مانتل-هانزل التي كشفت عما نسبته (18%) من فقرات الاختبار. وكانت نسبة الاتفاق بين الطريقتين (18%).

أما دراسة حمادنة (2007) فقد هدفت إلى مقارنة الأداء التفاضليّ لفقرات اختبار تحديد الكفاءة اللغوية في اللغة الإنجليزية في الجامعات الأردنية تبعًا لمتغير الجنس، والحقل الأكاديمي، باستخدام طريقتي مانتل-هانزل، وفرق المساحة بين منحنيات خصائص الفقرة في النموذج ثلاثي المعلم. وتكونت العينة من (1935) طالباً وطالبةً من ثلاث جامعاتٍ، أظهرت النتائج اتفاق طريقتي مانتل-هانزل ومؤشر فرق المساحة بين منحنيي خصائص الفقرة للنموذج اللوجستي ثلاثي المعلم في عدد الفقرات التي أظهرت أداءً تفاضليًّا منتظماً. واستطاعت طريقة مانتل-هانزل الكشف عن أكبر عددٍ من الفقرات التي أظهرت أداءً تفاضليًّا من النوع المنتظم، بينما كانت طريقة مؤشر فرق المساحة بين منحنيات خصائص الفقرة للنموذج اللوجستي ثلاثي المعلم أكثر فاعلية في الكشف

المجلة الدولية للأبحاث التربوية - جامعة الإمارات العربية المتحدة
المجلد (47) العدد (3) أغسطس 2023

عن الفقرات التي أظهرت أداءً تفاضليًا من نوعٍ غير منتظمٍ، كما كانت نسبة الاتفاق بين الطريقتين منخفضة وفقاً لمتغير الجنس.

ويُلاحظ مما سبق الاختلاف بين نتائج الدراسات التي هدفت إلى المقارنة بين طريقتي مانتل-هانزل وطريقة المساحة بين منحنيي خصائص الفقرة لنموذج الاستجابة اللوجستي ثلاثي المعلم من حيث فاعلية كل طريقة، ونسبة الاتفاق بين الطريقتين. فدراسة (هامبلتون وروجرز Hambleton & Rogers, 1989؛ باغهي وفيرارا Baghi & Ferrara, 1989؛ لراجو ودراسجو وسليند Raju, Drasgow & Slind, 1993) خلصت إلى نسبة اتفاقٍ عالية بين الطريقتين في الكشف عن الأداء التفاضليّ، بينما توصلت دراسة (البستنجي، 2004؛ العطيوي، 2004؛ حمادنة، 2007 شيو Chiu, 2008) إلى نسبة اتفاق منخفضة بين الطريقتين.

مشكلة البحث

تختلف نتائج دراسات الأداء التفاضليِّ باختلاف الطرق المستخدمة للكشف عنه، فالبعض يركز على استخدام طرقِ النظرية التقليدية في الكشف عن الأداء التفاضليِّ للفقرة لسهولة افتراضاتها وبساطتها، وإمكانية توفرها في البيانات الحقيقية وعبْر الدرجات الخام. بينما يُفضل آخرون استخدام طرقِ نظرية الاستجابة للفقرة لتغلبها على نقطة ضعفٍ رئيسةٍ متأصلةٍ في النظرية التقليدية بتحقيق خاصيةِ عدم التباين من خلال استقلال معلم الفرد عن معالمِ فقرات الاختبار. لذلك تُوصى الدراسات المتعلقة بالكشف عن الأداء التفاضليِّ لفقرات الاختبار باستخدام أكثر من طريقةٍ للكشف عن الأداء التفاضلي (Camilli & Shepard, 1994)، للجمع بين نقاط القوة لكل طريقةٍ والتقليل من نقاط الضعف، وغالباً ما يتم الجمع بين طريقتي مانتل-هانزل، وطريقة مؤشر فرق المساحة بين منحنيي خصائص الفقرة، عند الكشف عن الأداء التفاضلي والمقارنة بين نتائجهما، لسمعتهما في تحقيق نتائج متقاربةٍ (Rudner, Getson & Knight, 1980). لذلك تم الجمع بين هاتين الطريقتين للكشف عن الأداء التفاضليِّ لفقرات اختبار القدرات العامة للجامعيين. وتم اختيار النموذج اللوجستي ثلاثي المعلم لقوة القاعدة التنظيرية التي يعتمد عليها، ولِدقة نتائجه، كما أن تقديره لمعلمي التمييز والصعوبة بالإضافة لمعلم التخمين يسمح بأن تختلف فقرات الاختبار في صعوبتها وتمييزها وهو ما يلاحظ عادةً في بناء الاختبارات، مما يتيح فرصةً أكبر لمطابقة الأفراد لبيانات الاستجابة على فقرات الاختبار من نمط الاختيار من متعددٍ (Lord, 1980)، كفقرات اختبار القدرات العامة للجامعيين. وتتمحور مشكلة البحث في الإجابة عن التساؤل التالي:

- ما نسبة الاتفاق بين نتائج استخدام كلٍّ من طريقة مانتل-هانزل، وطريقة مؤشر فرق المساحة بين منحنيي خصائص الفقرة للنموذج اللوجستي ثلاثي المعلم في الكشف عن الأداء التفاضليّ لفقرات اختبار القدرات العامة للجامعيين لمتغير الجنس؟

أهداف البحث

يهدف البحث إلى تحقيقِ الآتي:

1. التعرف على ما إذا كانت فقرات اختبار القدرات العامة للجامعيين تظهر أداءً تفاضليًّا لمجموعةٍ دون أخرى وفقاً لجنسهم، ولصالح أيّ المجموعتين، إضافةً، لنوْع الأداء التفاضليِّ الظاهرِ.

2. التعرف على نسبة الاتفاق بين كلٍّ من طريقتي مانتل-هانزل، ومؤشر فرق المساحة بين منحنَيَي خصائص الفقرة للنموذج الاستجابة للفقرة اللوجستي ثلاثي المعلم في الكشف عن الأداء التفاضليِّ لفقرات اختبار القدرات العامة للجامعيين.

أهمية البحث

يستمد هذا البحث أهميته من تناوله لموضوع الأداء التفاضليِّ للفقرة الذي يُعتبر من أهم الخصائص السيكومترية للاختبار. كذلك من تناوله لاختبار القدرات العامة للجامعيين الذي يُعدّ معيارًا للقبول في الجامعات في المملكة العربية السعودية. وتسهم نتائج هذا البحث في الحصول على بياناتٍ حول الخصائص السيكومترية المتعلقة بالأداء التفاضليِّ لفقرات اختبار القدرات العامة للجامعيين. كما يفتح آفاقاً للمزيد من البحوث حول الأداء التفاضليّ للفقرة في اختبارات أخرى وفقًا لمتغيراتٍ وطرقٍ تجريبيةٍ وإحصائيةٍ مختلفةٍ. وتوفير أداةٍ للكشف عن الأداء التفاضليِّ بقياس المساحة بين منحنيَي خصائص الفقرة لراجو، ورسم منحنَيَي خصائص الفقرة.

مصطلحات البحث

الأداء التفاضليّ للفقرة Differential Item Functioning-DIF: هو دالةٌ مشتقةٌ عن الفروق في الاستجابة على الفقرة بين مجموعتين من الأفراد ممن هم في نفس المستوى من القدرة (Hambleton & Rogers, 1989). ويعرف إجرائياً بالدلالة الإحصائية لقيمة مانتل-هانزل كاي تربيع، ومؤشر فرق المساحة بين منحنَي خصائص الفقرة لمجموعتي المقارنة حسب متغير الجنس.

طريقة مانتل هانزل Mantel Haenszel-MH: إحدى طُرق كاي تربيع التي تستقصي الأداء التفاضليّ للفقرة بين مجموعتين فرعيتين من مجتمعٍ لبحث، إحداهما تسمى المجموعة المرجعية Reference Group، والأخرى تسمى المجموعة المستهدفة Focal Group وهي المجموعة التي يُعتقد أنها تتأثر بالأداء التفاضليّ للفقرة (Camilli & Shepard, 1994; Mantel & Haenszel, 1959).

المجلة الدولية للأبحاث التربوية - جامعة الإمارات العربية المتحدة
المجلد (47) العدد (3) أغسطس 2023

مؤشر فرق المساحة بين منحنيي خصائص الفقرة The area difference between item characteristic curves index: أحد طرق الكشف عن الأداء التفاضلي وفق نظرية الاستجابة للفقرة من خلال حساب المساحة بين منحنيي خصائص الفقرة لمجموعتي المقارنة بتطبيق معادلة راجو للنموذج اللوجستي ثلاثي المعلم (Raju, 1988).

حدود البحث

تقتصر نتائج هذا البحث على ما يلي:

1. استخدام طريقة مانتل-هانزل كاي تربيع التابعة للنظرية التقليدية، كذلك الاقتصار على نموذجٍ واحدٍ من نماذج نظرية الاستجابة للفقرة وهو النموذج اللوجستي ثلاثي المعلم، وطريقةٍ واحدةٍ من طرق الكشف عن الأداء التفاضليّ وفقاً لهذا النموذج، وهي طريقةُ مؤشرِ فرْق المساحة بين منحنيي خصائص الفقرة للنموذج اللوجستي ثلاثي المعلم.
2. تحليلِ بيانات العام الجامعي 1435/ 1436هـ التي تم توفيرها من المركز الوطني للقياس والتقويم في التعليم العالي في المملكة العربية السعودية.
3. استخدام نموذجٍ من نماذج اختبار القدرات العامة للجامعيين بنسخته المكونة من ثلاثة أقسامٍ هي: اللفظي، الرياضي، والمنطقي، الذي طُبق في البحث، حيث تجدر الإشارة إلى وجود نسخةٍ حديثةٍ من الاختبار تتكون من خمسة أجزاءٍ هي: اللفظي، الرياضي، التفكير الناقد، الأشكال، والمنطقي.
4. استخدام متغير الجنس (ذكور، إناث) للكشف عن الأداء التفاضليّ لفقرات اختبار القدرات العامة للجامعيين تبعاً لهما.
5. استخدام البرامج الحاسوبية التي اختيرت لاستخدامها في تنفيذ البحث ومعالجة البيانات، حيث تتحدد النتائج.

منهجية البحث وإجراءاته

مجتمع البحث وعينته

تكون مجتمع البحث من (3269) طالباً وطالبةً من الذين قدموا لاختبار القدرات العامة للجامعيين في المملكة العربية السعودية في التطبيق الثاني في الفترة 11-1436/1/15هـ من بيانات العام 1435/ 1436هـ. بينما تكونت عينة البحث من (2280) طالباً وطالبةً بعد استبعاد الحالات ذات البيانات الناقصة. وتوزعت عينة الدراسة بالتساوي بين مجموعتي الذكور والإناث بواقع (1140) لكل مجموعة.

أداة البحث

تمثلت أداة البحث في اختبار القدرات العامة للجامعيين الذي يقدمه المركز الوطني للقياس لخريجي المرحلة الجامعية باللغة العربية. ويتكون من (104) فقرة مقسمة على ثلاثة أجزاءٍ تقيس عدداً من القدرات: الأول: لفظيٌّ (لغويٌّ) يشتمل على (48) فقرة تقيس القدرة على القراءة بفهمٍ عميقٍ، والقدرة على تمييز البناء المنطقي للتعبيرات اللغوية. والثاني: كميٌّ (رياضيٌّ) يشتمل على (24) فقرة تقيس القدرة على حل المسائل المبنية على مفاهيم رياضيةٍ أساسيةٍ، والقدرة على الاستدلال، والقدرة على الاستنتاج. بينما الثالث: منطقيٌّ (استدلاليٌّ مكانيٌّ) يشتمل على (32) فقرة تقيس القدرة على إدراك العلاقات المنطقية، المكانية وغير المكانية، والقدرة على التحليل، والقدرة على الاستقراء، والقدرة على تفسير النتائج. وتقدم أسئلة الأجزاء الثلاثة للاختبار بشكلٍ متناوبٍ في ستة أقسامٍ يعطى كل منها (25 دقيقة). وجميع الأسئلة ثنائية الاستجابة، ومن نمط الاختيار من متعددٍ، من أربعة اختياراتٍ معطاة (المركز الوطني للقياس والتقويم في التعليم العالي، 1436).

الخصائص السيكومترية للاختبار: للتحقق من دلائل صدْق درجات الاختبار تم حساب معاملات الارتباط بين درجات الأفراد في كل جزءٍ من أجزاء اختبار القدرات العامة للجامعيين (اللفظي، والكمي، والمنطقي) ودرجاتهم الكلية على الاختبار. وكانت قيم معاملات الارتباط عاليةً حيث بلغت قيم معاملات الارتباط بين أجزاء اختبار القدرات العامة للجامعيين (اللفظي، والكمي، والمنطقي) والدرجة الكلية على الاختبار (0.88، و0.78، و0.82) على التوالي. مما يعطي دلائل على اتساق أبعاد الاختبار وصدْق البناء والتكوين له. وتم استخراج مؤشرات الثبات لدرجات اختبار القدرات العامة للجامعيين وفق نظريتي القياس التقليدية والحديثة بحساب قيمة معامل ألفا كرونباخ، والثبات التجريبي (Empirical Reliability). حيث بلغت قيمة ألفا كرونباخ لثبات درجات أجزاء الاختبار اللفظي والكمي، والمنطقي (0.76، 0.66، 0.60) على التوالي، ومعامل الثبات التجريبي للأجزاء الثلاثة على التوالي (0.79، 0.79، 0.70)، وللاختبار ككل (0.90). مما يدل على تمتع درجات اختبار القدرات العامة للجامعيين بدرجةٍ مقبولةٍ من الثبات.

إجراءات البحث

تم تحليل بيانات الاختبار باستخدام برنامج BILOG-MG3، وبناء على نتائج تحليل قيم معاملات التمييز لفقرات اختبار القدرات العامة للجامعيين تم حذف الفقرات التي يكون قيم معامل تمييزها سالبًا بناءً على ما هو معتمدٌ في البرنامج، حيث يقوم البرنامج باستبعاد الفقرات ذات معاملات التمييز السالبة من عمليات التحليل اللاحقة، وبناءً عليه تم استبعاد أربع فقراتٍ كانت قيم معامل تمييزها سالبةً من مجموع فقرات اختبار القدرات العامة للجامعيين. منها فقرتان من فقرات الجزء اللفظي وفقرة من كل من الجزء الكمي والمنطقي.

المجلة الدولية للأبحاث التربوية - جامعة الإمارات العربية المتحدة
المجلد (47) العدد (3) أغسطس 2023

وتمثلت إجراءات البحث في ثلاث خطواتٍ رئيسةٍ لتجهيز البيانات وتحليلها للخروج بنتائج البحث وهي:

أولاً/ التحقق من افتراضات النموذج اللوجستي ثلاثي المعلم:

تم التأكد من مطابقة بيانات اختبار القدرات العامة للجامعيين للنموذج اللوجستي ثلاثي المعلم من خلال التحقق من افتراضات النموذج، وهي كالآتي:

1. أحادية البعد

للتحقق من افتراض أحادية البعد تم استخدام التحليل العاملي لفقرات اختبار القدرات العامة للجامعيين باستخدام طريقة المكونات الرئيسية (Principal Component Analysis) ومِن ثَمَّ تدوير العوامل المتعامد بطريقة فاريمكس (Varimax)، واعتماد محك كايزر لاستخلاص العوامل (Kaiser)، وهي العوامل التي يكون الجذر الكامن لها أكبر من أو يساوي الواحد الصحيح.

جدول 1

قيمة الجذر الكامن ونسبة التباين المفسر للعوامل الناتجة عن التحليل العاملي لفقرات اختبار القدرات العامة للجامعيين

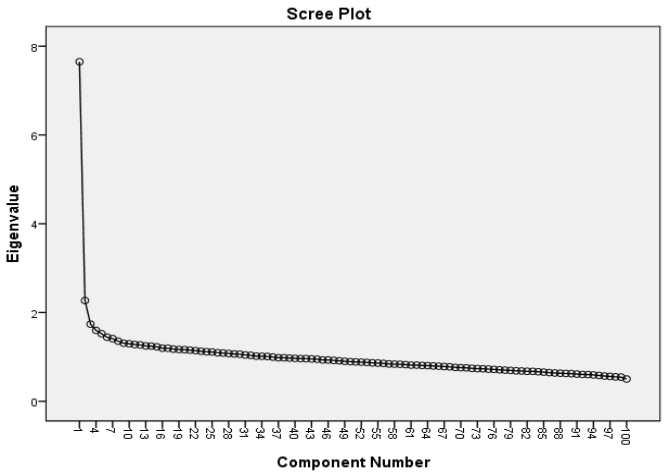
العامل	الجذر الكامن	نسبة التباين المفسر %	العامل	الجذر الكامن	نسبة التباين المفسر %	العامل	الجذر الكامن	نسبة التباين المفسر %
1	7.65	7.65	13	1.25	1.25	25	1.11	1.11
2	2.27	2.27	14	1.24	1.24	26	1.09	1.09
3	1.74	1.74	15	1.23	1.23	27	1.09	1.09
4	1.6	1.6	16	1.2	1.2	28	1.08	1.08
5	1.52	1.52	17	1.2	1.2	29	1.07	1.07
6	1.44	1.44	18	1.18	1.18	30	1.06	1.06
7	1.41	1.41	19	1.17	1.17	31	1.04	1.04
8	1.35	1.35	20	1.16	1.16	32	1.03	1.03
9	1.31	1.31	21	1.15	1.15	33	1.02	1.02
10	1.3	1.3	22	1.14	1.14	34	1.01	1.01
11	1.28	1.28	23	1.13	1.13	35	1.01	1.01
12	1.27	1.27	24	1.11	1.11			

يتضح من الجدول (1) أن نتائج التحليل العاملي لفقرات اختبار القدرات العامة للجامعيين أسفرت عن وجود (35) عاملاً الجذر الكامن لها أكبر من الواحد وفقاً لمحك كايزر، وأن نسبة التباين

المفسر للعامل الأول تساوي (7.65%). ويعتبر الاختبار أحادي البعد إذا كان نسبة ما يفسره العامل الأول كبيرة، وعادة ما تكون أكبر من 20%، إلا أن أنه ليس مستغرباً أن يفسر العامل الأول ما نسبته 10% من التباين الكلي، ولذلك فإنه طالما كان ذلك العامل أكبر بشكلٍ واضحٍ عن بقية العوامل الأخرى دلَّ ذلك على أحادية البعد لذلك الاختبار (Wiberg, 2007). ويتضح ذلك من خلال ما يعرف باختبار فرز العوامل Scree Plot الذي يوضحه الشكل رقم (3)، كما أن العامل الأول يفسر ما نسبته 7.65% وهو أكثر من ثلاثة أضعاف ما يفسره العامل الثاني من التباين، مما يشير إلى تحقق افتراض أحادية البعد للاختبار (Hatti, 1985).

شكل 3

التمثيل البياني لقيم الجذر الكامن للعوامل الناتجة عن تحليل فقرات الاختبار



كما تم التحقق من أحادية البعد من خلال تحليل البواقي Residual Analysis، باستخدام برنامج Normal Ogive Harmonic Analysis Robust Method (NOHARM)، حيث يتم فحص المواءمة بين عدد الأبعاد التي يتم تحديدها للاختبار والبيانات من خلال حساب مصفوفة البواقي Residual Matrix لإيجاد قيمة الجذر التربيعي لمتوسطات مربعات البواقي Root Mean Square of Residuals (RMSR). فإذا كانت قيمة RMSR قريبة من الصفر، وقيمة مؤشر تاناكا لحسن المطابقة Tanaka Index of Goodness of Fit (GFI) قريبة من الواحد يُعدُّ ذلك دليلاً على أحادية البعد (دي إيالا، 2017/2009). وبلغت قيمة الجذر التربيعي لمتوسطات مربعات البواقي لاختبار القدرات العامة للجامعيين (0.008). كما بلغت قيمة مؤشر تاناكا (0.91)، مما يشير إلى تحقق افتراض أحادية البعد في بيانات اختبار القدرات العامة للجامعيين.

2. الاستقلال الموضعي

تم التحقق من افتراض الاستقلال الموضعي لفقرات اختبار القدرات العامة للجامعيين من خلال فحص مصفوفة معاملات الارتباط للبواقي بين أزواج الفقرات من خلال برنامج نوهارم (NOHARM)، حيث كانت قيم معاملات الارتباط للبواقي صغيرةً وتقترب من الصفر مما يشير إلى تحقق افتراض الاستقلال الموضعي (دي إيالا، 2017/2009). كما يشير هامبلتون وسوامينثان إلى أن افتراض أحادية البعد يكافئ افتراض الاستقلال الموضعي ويعتبر متحقق بتحققه (Hambleton & Swaminathan, 1985).

3. التحرر من عامل السرعة

لا يُعتبر اختبار القدرات العامة للجامعيين اختبار سرعةٍ، ونتائج افتراض أحادية البعد تدعم ذلك إذ لو كانت السرعة في الإجابة تؤثر في أداء الأفراد على الاختبار فسيظهر بعدين للاختبار أحدهما سرعة الإجابة والثاني القدرة المراد قياسها (Hambleton & Swaminathan, 1985).

4. مطابقة البيانات للنموذج اللوجستي ثلاثي المعلم

تم التحقق من جودة مطابقة بيانات اختبار القدرات العامة للجامعيين للنموذج اللوجستي ثلاثي المعلم باستخدام برنامج Bilog-MG3، باستخراج دلالة قيمة مربع كاي لاختبار جودة المطابقة (Chi-Square Statistic Fit X^2)، عند مستوى دلالة (α=0.01). للتأكد من مدى مطابقة البيانات للنموذج المستخدم، وذلك من خلال الخطوات الثلاث التالية:

1. تم إدخال البيانات للبرنامج لفحص مطابقة الأفراد للنموذج المستخدم، حيث تم استبعاد (96) فرداً من العينة لعدم مطابقة بياناتهم للنموذج المستخدم.

2. تم إدخال البيانات مرةً ثانيةً للبرنامج لاختبار مدى مطابقة الفقرات للنموذج المستخدم، حيث تم استبعاد (5) فقرات من الجزء اللفظي من اختبار القدرات العامة للجامعيين، لعدم مطابقتها للنموذج اللوجستي ثلاثي المعلم. وبهذا يكون مجموع الفقرات التي تم استبعادها من عملية الكشف عن الأداء التفاضليّ لفقرات اختبار القدرات العامة للجامعيين باستخدام مؤشر فرق المساحة للنموذج اللوجستي ثلاثي المعلم (9) فقرات، منها (4) فقرات بسبب قيم معاملات تمييزها السالبة، و(5) فقرات لعدم مطابقتها لافتراضات النموذج اللوجستي ثلاثي المعلم.

3. بعد استبعاد كلٍّ من الأفراد والفقرات غير المطابقة للنموذج المستخدم، تم إجراء التحليل للمرة الثالثة للحصول على التقديرات النهائية لمعالم الفقرات (الصعوبة، والتمييز، والتخمين) وتقدير معلم قدرة الأفراد والخطأ المعياري لكل معلمٍ على النموذج المستخدم. وقد أصبح حجم عينة الاختبار (2184)، وعدد فقراته في صورته النهائية (95) فقرة.

ثانياً/ الكشف عن الأداء التفاضليّ باستخدام طريقة مانتل-هانزل:

تم تحليل البيانات باستخدام برنامج نظام تحليل الأداء التفاضليّ للفقرة Differential item functioning analysis system (DIFAS5.0) لاستخراج قيمة مانتل-هانزل كاي تربيع ودلالته الإحصائية.

ثالثاً/ الكشف عن الأداء التفاضليّ باستخدام طريقة مؤشر فرق المساحة بين منحنيي خصائص الفقرة للنموذج اللوجستي ثلاثي المعلم:

وذلك من خلال الخطوات التالية:

1. قسمت البيانات وفقاً لمتغير الجنس إلى مجموعتي (الذكور، والإناث). وتم تحليل بيانات كل مجموعة على مقياسٍ واحدٍ بحيث يساوي المتوسط الحسابي للقدرة صفراً، والانحراف المعياري يساوي واحداً. وتم تقدير معالم الفقرة (الصعوبة، والتمييز، والتخمين) بطريقة التقدير بالأرجحية القصوى الهامشية (MMLE)، للنموذج اللوجستي ثلاثي المعلم لحساب قيمة المساحة الفعلية باستخدام معادلة راجو.
2. تم تحديد المجموعة المرجعية من خلال مقارنة المتوسط الحسابي للمجموعتين، فكان المتوسط الحسابي لمجموعة الذكور (44.5) أقل من المتوسط الحسابي لمجموعة الإناث (44.7)، وبالتالي تعتبر مجموعة الذكور مجموعةً مرجعيةً ومجموعة الإناث مجموعةً مستهدفةً. ثم قسمت المجموعة المرجعية (الذكور) إلى عينتين عشوائيتين متكافئتين في القدرة ومتساويتين ما أمكن في العدد، وحساب قيمة المساحة بين منحنيي خصائص الفقرة للمجموعتين العشوائيتين. حيث اعتمدت أكبر قيمة كنقطة قطع (Cut-Off Value)، وتعرف بالمساحة الحرجة والتي تعتبر معياراً للحكم على وجود أداءٍ تفاضليٍّ للفقرة من عدمه.
3. تمت مقارنة المساحة الفعلية بالمساحة الحرجة في كل جزءٍ من أجزاء اختبار القدرات العامة للجامعيين، فإذا كانت قيمة المساحة الفعلية أكبر من قيمة المساحة الحرجة فإن ذلك يدل على وجود أداءٍ تفاضليٍّ للفقرة. أما إذا كانت قيمة المساحة الفعلية أقل من قيمة المساحة الحرجة فإن ذلك يعني عدم وجود أداءٍ تفاضليٍّ للفقرة.
4. تم رسم منحنيات خصائص فقرات اختبار القدرات العامة للجامعيين لمعرفة نوع الأداء التفاضليّ الظاهر باستخدام برنامج حساب ورسم الأداء التفاضليّ للفقرة بقياس المساحة لراجو.

نتائج البحث

بعد معالجة بيانات اختبار القدرات العامة للجامعيين للكشف عن الأداء التفاضليّ للفقرة وفقاً لمتغير الجنس، تم تحليل بيانات الاختبار واستخراج قيمة مانتل- هانزل كاي تربيع ودلالته

المجلة الدولية للأبحاث التربوية - جامعة الإمارات العربية المتحدة
المجلد (47) العدد (3) أغسطس 2023

الإحصائية، كما تم قياس مؤشر فرق المساحة بين منحنيي خصائص فقرات الاختبار بتطبيق معادلة راجو للنموذج اللوجستي ثلاثي المعلم لفقرات الجزء اللفظي، والجزء الكمي، والجزء المنطقي كما في جدول رقم (2، 3، 4،) على الترتيب. وللإجابة على تساؤل البحث والذي ينص على: "ما نسبة الاتفاق بين نتائج استخدام كلٍّ من طريقة مانتل-هانزل، وطريقة مؤشر فرق المساحة بين منحنيي خصائص الفقرة للنموذج اللوجستي ثلاثي المعلم في الكشف عن الأداء التفاضليّ لفقرات اختبار القدرات العامة للجامعيين لمتغير الجنس؟" تم استخراج عدد ونسبة فقرات اختبار القدرات العامة للجامعيين التي أظهرت أداءً تفاضليًّا في كلٍّ من طريقتي مانتل-هانزل وطريقة مؤشر فرق المساحة بين منحنيي خصائص الفقرة للنموذج اللوجستي ثلاثي المعلم، كما يظهر في جدول رقم (5).

جدول 2

قيم مانتل هانزل كاي تربيع ومؤشر فرق المساحة بين منحنيي خصائص فقرات الجزء اللفظي من اختبار القدرات العامة للجامعيين

الفقرة	مانتل-هانزل كاي تربيع	مؤشر فرق المساحة بين منحنيي خصائص الفقرة ***	الفقرة	مانتل-هانزل كاي تربيع	مؤشر فرق المساحة بين منحنيي خصائص الفقرة ***
1	**13.44	0.56	25	0	0.22
2	**8.76	0.5	26	0.73	0.07
3	**60.3	1	27	0.43	0.37
4	**12.01	1.61	28	1.56	0.24
5	3.2	0.29	29	1.02	0.11
6	2.83	0.15	30	**48.58	0.9
7	**24.53	1.15	31	0	0.09
8	0.98	0.2	32	0.53	0.12
9	0.21	0.9	33	**23.56	1.3
11	1.61	0.11	34	**15.53	1.49
12	0.05	0.55	36	**13.17	0.15
13	*4.79	0.68	37	0.32	0.3
14	0.47	0.16	38	0.04	0.23
15	*4.37	0.38	39	**19.32	0.7
16	0.48	0.1	40	**32.85	0.91
17	0.4	0.15	42	**38.58	0.59
18	0	0.04	43	**6.74	1.02
19	0.15	0.24	46	**74.09	0.72
20	2.23	0.34	47	1.39	0.19
21	1.67	0.24	48	0.42	0.04
23	1.58	0.48			

*الفقرة تظهر أداءً تفاضلياً عند $p < 0.05$

**الفقرة تظهر أداءً تفاضلياً عند $p < 0.01$

*** المساحة الحرجة=1.6

يظهر من جدول (2) أن مجموع فقرات الجزء اللفظي التي أظهرت أداءً تفاضليًّا منتظماً باستخدام طريقة مانتل- هانزل (16) فقرة بنسبةٍ تعادل 39% من أصل (41) فقرة، بينما لم تظهر أي فقرة أداءً تفاضليًّا باستخدام طريقة مؤشر فرق المساحة بتطبيق معادلة راجو للنموذج اللوجستي ثلاثي المعلم وفقاً لمتغير الجنس.

جدول 3

قيم مانتل هانزل كاي تربيع ومؤشر فرق المساحة بين منحنيي خصائص فقرات الجزء الكمي من اختبار القدرات العامة للجامعيين

الفقرة	مانتل-هانزل كاي تربيع	مؤشر فرق المساحة بين منحنيي خصائص الفقرة ***	الفقرة	مانتل-هانزل كاي تربيع	مؤشر فرق المساحة بين منحنيي خصائص الفقرة***
1	1.71	0.24	13	**22.5	0.52
2	*6.08	0.17	14	0.28	0.28
3	**6.8	0.08	15	**45.6	0.44
4	0.68	0.34	16	**10.6	0.33
5	2.6	3.72	17	0.29	0.16
6	**10.1	0.39	18	2.7	0.3
7	0.02	1.67	19	**14.6	0.1
8	0.16	0.28	21	0.19	0.11
9	*4.7	1.64	22	3.4	0.12
10	2.8	0.05	23	0.49	1.93
11	*4.9	0.02	24	2.2	0.67
12	0.01	0.09			

*الفقرة تظهر أداءً تفاضلياً عند $p < 0.05$

**الفقرة تظهر أداءً تفاضلياً عند $p < 0.01$

*** المساحة الحرجة=2.03

يظهر من جدول (3) أن مجموع فقرات الجزء الكمي التي أظهرت أداءً تفاضليًّا منتظماً باستخدام طريقة مانتل- هانزل (9) فقرات من أصل (23) فقرة أي ما نسبته 39%، بينما كشفت طريقة مؤشر فرق المساحة بتطبيق معادلة راجو للنموذج اللوجستي ثلاثي المعلم عن أداءٍ تفاضليٍّ من النوع المنتظم في فقرةٍ واحدةٍ ما يعادل 4.35%.

جدول 4

قيم مانتل هانزل كاي تربيع ومؤشر فرق المساحة بين منحنيي خصائص فقرات الجزء المنطقي من اختبار القدرات العامة للجامعيين

الفقرة	مانتل-هانزل كاي تربيع	مؤشر فرق المساحة بين منحنيي خصائص الفقرة ***	الفقرة	مانتل-هانزل كاي تربيع	مؤشر فرق المساحة بين منحنيي خصائص الفقرة ***
1	*3.94	0.38	18	**6.83	0.16
2	**35.92	0.33	19	**8.13	0.02

الفقرة	مانتل-هانزل كاي تربيع	مؤشر فرق المساحة بين منحنيي خصائص الفقرة ***	الفقرة	مانتل-هانزل كاي تربيع	مؤشر فرق المساحة بين منحنيي خصائص الفقرة ***
3	0.12	0.17	20	0.59	1.44
4	**10.98	0.86	21	**9.43	0.31
6	**175.37	2.53	22	0.04	0.27
7	2.3	0.31	23	0.06	0.09
8	3.59	0.58	24	**7.8	0.8
9	**14.37	0.28	25	**10.09	0.11
10	**46.82	1.13	26	**6.8	0.25
11	0.01	0.16	27	0.5	0.12
12	**8.06	0.6	28	**29.7	0.78
13	2.36	0.14	29	0.32	0.53
14	3.46	0.71	30	*5.54	0.22
15	0.7	0.16	31	**6.77	0.62
16	0.08	0.01	32	**16.78	0.67
17	*4.65	0.5			

*الفقرة تظهر أداءً تفاضلياً عند $p < 0.05$
**الفقرة تظهر أداءً تفاضلياً عند $p < 0.01$
*** المساحة الحرجة=1.6

يظهر من خلال جدول (4)، أن (18) فقرة من أصل (31) فقرة بنسبةٍ تُعادل 58% أظهرت أداءً تفاضليًّا منتظماً باستخدام طريقة مانتل- هانزل، بينما كشفت طريقة مؤشر فرق المساحة بتطبيق معادلة راجو للنموذج اللوجستي ثلاثي المعلم عن أداءٍ تفاضليٍّ من النوع المنتظم في فقرةٍ واحدةٍ بنسبةٍ تعادل 3.23%.

جدول 5

عدد ونسبة فقرات اختبار القدرات العامة للجامعيين التي أظهرت أداءً تفاضليًّا في كلٍّ من طريقتي مانتل-هانزل وطريقة مؤشر فرق المساحة بين منحنيي خصائص الفقرة وفقاً لمتغير الجنس

جزء الاختبار	طريقة مانتل-هانزل	طريقة مؤشر فرق المساحة بين منحنيي خصائص الفقرة	الطريقتان معاً
اللفظي (41 فقرة)	16 (39%) (1، 2، 3، 4، 7، 13، 15، 3، 33، 34، 36، 39، 40، 42، 43، 46)	-	-
الكمي (23 فقرة)	9 (39%) (2، 3، 6، 9، 11، 13، 15، 16، 19)	1 (4.3%) (5)	-
المنطقي (31 فقرة)	18 (58%) (1، 2، 4، 6، 9، 10، 12، 17، 18، 19، 21، 24، 25، 26، 28، 30، 31، 32)	1 (3.2%) (6)	1 (5.5%) (6)

يظهر من جدول (5) أن مجموع فقرات الجزء اللفظي التي أظهرت أداءً تفاضليًّا منتظماً باستخدام طريقة مانتل-هانزل (16) فقرة بنسبةٍ تعادل (39%) من أصل (41) فقرة، بينما لم تظهر أي فقرة أداءً تفاضليًّا باستخدام طريقة فرق المساحة بتطبيق معادلة راجو للنموذج اللوجستي ثلاثي المعلم وفقاً لمتغير الجنس.

ويُظهر كذلك أن مجموع فقرات الجزء الكمي التي أظهرت أداءً تفاضليًّا منتظماً باستخدام طريقة مانتل-هانزل (9) فقرات من أصل (23) فقرة أي ما نسبته (39%)، بينما كشفت طريقة مؤشر فرق المساحة بين منحني خصائص الفقرة للنموذج اللوجستي ثلاثي المعلم بتطبيق معادلة راجو عن أداءٍ تفاضليٍّ من النوع المنتظم في فقرةٍ واحدةٍ ما يعادل (4.35%)، كما يظهر من خلال الجدول عدم اتفاق الطريقتين في اعتبار أيٍّ من فقرات الجزء الكمي تُظهر أداءً تفاضليًّا لمتغير الجنس.

ويتبين من خلال الجدول رقم (2)، أن (18) فقرة من أصل (31) فقرة بنسبةٍ تُعادل (58%) أظهرت أداءً تفاضليًّا منتظماً باستخدام طريقة مانتل-هانزل، بينما كشفت طريقة مؤشر فرق المساحة بين منحني خصائص الفقرة عن أداءٍ تفاضليٍّ من النوع المنتظم في فقرةٍ واحدةٍ بنسبةٍ تعادل (3.23%)، واتفقت الطريقتان في الكشف عن فقرةٍ واحدةٍ من مجموع (18) فقرة أظهرت أداءً تفاضليًّا وفقاً لمتغير الجنس باستخدام الطريقتين، بنسبةٍ تعادل (5.5%).

وبناءً عليه فإن نسبة الاتفاق بين طريقتي مانتل هانزل كاي تربيع، وفرق المساحة بين منحني خصائص الفقرة منخفضةٌ حيث تساوي (2.3%) بواقع فقرةٍ واحدةٍ من أصل (44) فقرة أظهرت أداءً تفاضليًّا منتظماً في اختبار القدرات العامة للجامعيين باستخدام الطريقتين. كما أن طريقة مانتل-هانزل كاي تربيع أكثر فاعليةً في الكشف عن الفقرات التي تبدي أداءً تفاضليًّا مقارنةً بطريقة مؤشر فرق المساحة بين منحني خصائص الفقرة.

مناقشة النتائج

هدف هذا البحث إلى المقارنة بين طريقتي مانتل-هانزل، ومؤشر فرق المساحة بين منحنيات خصائص الفقرة للنموذج اللوجستي ثلاثي المعلم بتطبيق معادلة راجو إلى الكشف عن الأداء التفاضليّ لفقرات اختبار القدرات العامة للجامعيين وفقاً لمتغير الجنس حيث يستخدم اختبار القدرات العامة للجامعيين كمعيارٍ للقبول في برامج الدراسات العليا بالجامعات السعودية لذلك يعتبر ذا أهمية عالية. كشفت النتائج عن اتفاق الطريقتين في الكشف عن فقرةٍ واحدةٍ من فقرات الجزء المنطقي بنسبةٍ تُعادل (2.3%)، من مجموع (44) فقرة من فقرات اختبار القدرات العامة أظهرت أداءً تفاضليًّا وفقاً لمتغير الجنس باستخدام الطريقتين. وكانت نسبة الاتفاق منخفضةً بين

الطريقتين في الكشف عن الأداء التفاضليّ لفقرات اختبار القدرات العامة للجامعيين وفقاً لمتغير الجنس. واتفقت هذه النتيجة مع دراسات (البستنجي، 2004؛ العطيوي، 2004؛ حمادنة، 2007؛ شيو Chiu, 2008)، بينما اختلفت مع نتائج دراسات (هامبلتون وروجرز Hambleton & Rogers, 1989؛ باغهي وفيرارا Baghi & Ferrara, 1989؛ راجو ودراسجو وسليند Raju, Drasgow & Slind,1993) التي أشارت إلى نسبة اتفاقٍ عاليةٍ بين طريقتي مانتل- هانزل وطريقة مؤشر فرق المساحة بين منحنيي خصائص الفقرة.

كما أن عدد الفقرات التي أظهرت أداءً تفاضليًّا لفقرات اختبار القدرات العامة للجامعيين وفقاً لمتغير الجنس باستخدام طريقة مانتل-هانزل أكبر من عدد الفقرات التي أظهرت أداءً تفاضليًّا باستخدام طريقة مؤشر فرق المساحة بين منحنيي خصائص الفقرة، وذلك بواقع (47) فقرة باستخدام طريقة مانتل-هانزل، مقابل فقرتين باستخدام طريقة مؤشر فرق المساحة بين منحنيي خصائص الفقرة وفقاً لمتغير الجنس. واتفقت هذه النتيجة مع دراسات (سكاجز وليستز Skaggs & Lissitz, 1992؛ راجو ودراسجو وسليند Raju, Drasgow & Slind,1993؛ حمادنة، 2007) التي أشارت إلى أن طريقة مانتل- هانزل أكثر فاعلية في الكشف عن الأداء التفاضليّ المنتظم مقارنةً بطريقة مؤشر فرق المساحة بين منحنيي خصائص الفقرة. واختلفت مع دراسات (هامبلتون وروجرز Hambleton & Rogers, 1989؛ باغهي وفيرارا Baghi & Ferrara, 1989، العطيوي، 2004؛ شيو Chiu, 2008) التي توصلت إلى أن طريقة مؤشر فرق المساحة بين منحنيي خصائص الفقرة أكثر فاعلية في الكشف عن الأداء التفاضليّ مقارنةً بطريقة مانتل-هانزل.

ويمكن القول إن طريقة مؤشر فرق المساحة بين منحنيي خصائص الفقرة لم تكن ذات فاعليةٍ عاليةٍ في الكشف عن الأداء التفاضليّ لفقرات اختبار القدرات العامة للجامعيين، مقارنةً بطريقة مانتل-هانزل كاي تربيع. وربما يعود هذا الاختلاف بين الطريقتين إلى اختلاف المفاهيم والإجراءات التي تقوم عليها كل طريقةٍ. فبينما تقوم طريقة مانتل-هانزل على مفاهيم النظرية التقليدية في القياس، تقوم طريقة مؤشر فرق المساحة بين منحنيي خصائص الفقرة على مفاهيم النظرية الحديثة في القياس. كما أن عدم وجود قيمةٍ محددةٍ كمعيارٍ للحكم على وجود الأداء التفاضليّ للفقرة عند استخدام مؤشر فرق المساحة بين منحنيي خصائص الفقرة يؤدي للاعتماد على التحليل والاجتهادات النسبية، بعكس طريقة مانتل-هانزل التي تقوم على الدلالة الإحصائية لقيمة كاي تربيع عند الكشف عن وجود الأداء التفاضليّ للفقرة (المري، 2018).

وبناء على ما أشار إليه دي إيالا (2017/2009)، بأنه قد لا يكون للدلالة الإحصائية للاختبار المستخدم في الكشف عن الأداء التفاضلي أي معنى، لذلك يجب أخذ حجم الأداء التفاضلي الظاهر بعين الاعتبار عند التفسير والحكم على الفقرة.

بناءً على ذلك نشير إلى أن (43) فقرة، وهو ما نسبته (45%) من أصل (95) فقرة من اختبار القدرات العامة للجامعيين، التي أظهرت أداءً تفاضليًّا باستخدام طريقة مانتل-هانزل تبعاً لمتغير الجنس كان حجم الأداء التفاضليّ الظاهر في (26) فقرة من النوع (A) والذي يعبر عن أداءٍ تفاضليٍّ صغيرٍ ولا يكاد يُذكر، بينما كان حجم الأداء التفاضليّ في (10) فقرات من النوع (B) والذي يُعبر عن أداءٍ تفاضليٍّ متوسطٍ، وكان حجم الأداء التفاضليّ في (7) فقراتٍ من النوع (C) والذي يُعبر عن أداءٍ تفاضليٍّ كبيرٍ. فإذا ما أهملنا الفقرات التي أظهرت أداءً تفاضليًّا من النوع (A) يصبح عدد الفقرات التي أظهرت أداءً تفاضليًّا باستخدام طريقة مانتل-هانزل (17) أي ما نسبته (17.9%) فنلاحظ انخفاضًا كبيرًا في نسبة الفقرات التي أظهرت أداءً تفاضليًّا وفقاً لمبدأ حجم الأداء التفاضليّ ومعناه.

وختاماً يمكن القول بأنه لا يمكن ضمان أن أحد طرق الكشف عن الأداء التفاضليّ للفقرة قادرةٌ على كشف كافة الفقرات التي تُظهر أداءً تفاضليًّا. ويكمن حل هذه الإشكالية من خلال الجمع بين أكثر من طريقةٍ لمعالجة مشكلة عدم الاستقرار، ومعالجة أوجه القصور الموجودة في بعض الطرق. كما أن الجمع بين الطرق التحكيمية والطرق الإحصائية هو الأسلوب الأمثل في الكشف عن الأداء التفاضليّ للفقرة، رغم صعوبة التنفيذ من الناحية المادية والوقت (Camilli & Shepard, 1994).

ويتم ذلك من خلال عرض الفقرات التي تبدي أداء تفاضلياً على هيئة من الخبراء للكشف عن أسباب ظهور الأداء التفاضلي للفقرة من خلال فحص محتوى أو بنية أو صياغة الفقرة. وترجع مهمة التقصي حول أسباب الأداء التفاضليّ الظاهر في بعض فقرات اختبار القدرات العامة للجامعيين إلى القائمين على وضع الاختبار ومعدّيه من المتخصصين في المركز الوطني للقياس والتقويم بما يتناسب مع سياسة سرية محتوى الاختبار.

التوصيات والبحوث المقترحة

في ظل النتائج التي تم التوصل إليها نوصي بما يلي:

- إعادة النظر في فقرات اختبار القدرات العامة للجامعيين التي أظهرت أداءً تفاضليًّا لمتغيري الجنس، للتعرف على مصدر الأداء التفاضليّ الظاهر، واستبدالها بفقراتٍ خاليةٍ من الأداء التفاضليّ في حال وجود بدائل جيدةٍ تقيس نفس المهارة من بنك الأسئلة.
- اهتمام الباحثين بدراسة الأداء التفاضليّ للفقرة والاختبار، باعتبارها مؤشرًا مهمًّا من مؤشرات الخصائص السيكومترية لأي اختبارٍ.
- الاهتمام بنشر مؤشرات الأداء التفاضليّ للفقرة في أدلة الاختبارات والمقاييس المنشورة.

المجلة الدولية للأبحاث التربوية - جامعة الإمارات العربية المتحدة
المجلد (47) العدد (3) أغسطس 2023

- توظيف نتائج دراسات الأداء التفاضليّ في عملية مراجعة وتطوير الاختبارات والمقاييس.
- استخدام طرق أساليب متنوعة في دراسات الكشف عن الأداء التفاضليّ.
- المقارنة بين طرق النظرية التقليدية والحديثة الأخرى في الكشف عن الأداء التفاضلي لفقرات اختبار القدرات العامة لجامعيين.

تضارب المصالح

" أفادت الباحثتان بعدم وجود تضارب في المصالح فيما يتعلق بالبحث، والملكية الفكرية، ونشر هذا البحث"

المراجع

الأحمدي، شرف (2020). مقارنة طريقتي مانتل هانزل ومؤشر قياس المساحة الإشارية لراجو في الكشف عن الأداء التفاضلي لمفردات استبانة مؤشر أساليب التعلم. *مجلة البحث العلمي في التربية*، *10*(21)، 321-280.

البستنجي، محمود محمد (2004). *مقارنة فاعلية أربع طرق للكشف عن الأداء التفاضلي لمتغير الجنس في فقرات اختبار قدرات عقلية خاصة للفئة العمرية (15-16) سنة في الأردن* [رسالة دكتوراه غير منشورة]. جامعة عمان العربية للدراسات العليا.

حمادنة، إياد (2007). *الأداء التفاضلي لفقرات تحديد الكفاءة اللغوية في اللغة الإنجليزية في الجامعات الأردنية: دراسة مقارنة الأردن* [رسالة دكتوراه غير منشورة]. جامعة اليرموك.

دي إيالا، آر. جي (2017). *النظرية والتطبيق في نظرية الاستجابة للفقرة* (عبدالله الكيلاني وإسماعيل البرصان، مُترجم). دار جامعة الملك سعود للنشر. (العمل الأصلي نشر في 2009).

العطيوي، إيمان محمد (2004). *الكشف عن الأداء التفاضلي لمتغير الجنس لفقرات اختبار تحصيلي في العلوم العامة للصف الثامن الأساسي في مديرية تربية وتعليم عمان الرابعة الأردن* [رسالة ماجستير غير منشورة]. جامعة عمان العربية للدراسات العليا.

المركز الوطني للقياس والتقويم في التعليم العالي (1436). *دليل اختبار القدرات العامة للجامعيين*. الرياض.

المري، منى علي (2018). الأداء التفاضلي لفقرات اختبار القدرات العامة للجامعيين باستخدام مانتل-هانزل ونموذج الاستجابة للفقرة اللوجستي ثلاثي المعلم وفقاً لمتغيري الجنس والجامعة [رسالة ماجستير غير منشورة]. جامعة الملك سعود.

Al-Ahmadi, S. (2020). A comparison of the Mantle-Hansel and Raju's signed area measure index methods for detecting differential performance of item of the learning styles index questionnaire (in Arabic). *Journal of Scientific Research in Education, Ain Shams University*, *10*(21), 280-321.

Al-Bustanji, M. (2004). *Gender related differential item functioning in general science achievement test for English graders in Amman fourth eductional directorate* [Unpublished master's thesis]. (In Arabic) Amman Arab University.

Al-Etawi, I. (2004). *Comparing four procedures for detecting gender - based differential functioning of items of a specific mental abilities test for the 15 - 16-year age group in Jordan* [Unpublished master's thesis]. (In Arabic) Amman Arab University.

Al-meri, M. A. (2018). *Gender and university differential item functioning in general aptitude test for university graduates by Mantel-Haenszel (MH) and three parameters logistic model (3PLM*) [Unpublished master's thesis]. (In Arabic) King Saud University.

American Educational Research Association, American Psychological Association & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. American Psychological Association.

Baghi, H., & Ferrara, S. (1989, March). *A comparison of IRT, delta plot, and Mantel-Haenszel techniques for detecting differential item functioning across subpopulations in the Maryland test of citizenship skills* [Paper presentation]. The Annual Meeting of the American Educational Research Association, San Francisco.

Benito, J. G., Hidalgo, D., & Guilera, G. (2010). Bias in measurement instruments. Fair tests. *Papeles Del Psicologo*,*31*(1), 75-84.

Camilli, G., & Shepard, L. (1994). Methods for identifying biased test items. Sag.

Chiu, P. C. (2008). *The effect of English proficiency on mathematics performance: A comparison of item response theory-based area and Mantel-Haenszel methods* (Publication No. 1450675) [Doctoral dissertation, Wilmington University]. ProQuest Dissertations & Theses Global.

De Ayala, R. (2017). *The theory and practice of item response theory* (A. Al-Kaylani, & I. Al-Bursan, Trans.). (In Arabic) King Saud university press. (Original work published 2009).

Ellis, B. B., & Raju, N. S. (2004). Test and item bias: what they are, what aren't, and how to detect them: measuring up. In J. Wall and G. Walz (Eds.), *Measuring Up: Assessment Issues for Teachers, Counselors, and Administrators* (pp. 89-98). CAPS Press.

Hamadneh, I. (2007). *Differential item functioning in the English language proficiency test given at Jordanian university* [Unpublished doctoral dissertation]. (In Arabic) Yarmouk University.

Hambleton, R., & Rogers, J. (1989). Detecting potentially biased test item: comparison of IRT Area and Mantel- Haenzel methods, *Applied Measurement in Education*, *2*(4), 313-334.

Hambleton, R., & Swaminathan, H. (1985). *Item response theory, principles and applications*. Kluwer -Nijhoff, USA.

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

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

Holland, P. W., & Thayer, D. T. (1988). Differential item performance and the Mantel- Haenszel procedure. In H. Wainer and H. I. Braun (Eds.), *Test validity* (129-145). Lawrence Erlbaum Associates.

Lord, F. M. (1980). *Applications of item response theory to practical: testing problems*. Lawrence Erlbaum Associates.

Mantel, N., & Haenszel, W. (1959). Statistical aspects of analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute*, *22*(4), 719-748.

Paek, I., & Guo, H. (2011). Accuracy of dif estimates and power in unbalanced designs using the Mantel–Haenszel detection procedure. *Applied Psychological Measurement*, *35*(7), 518-535.

Raju, N. S. (1988). The area between two item characteristic curves. *Psychometrika, 53*(4), 495–502.

Raju, N. S. (1990). Determining the significance of estimated signed and unsigned areas between two item response functions. *Applied Psychological Measurement*, *14*(2), 197-207.

Raju, N., Drasgow, F., & Slind, J. (1993). An empirical comparison of the area methods, Lord's chi-square test, and the Mental-Haenszel technique for assessing differential item function. *Educational and Psychological Measurement, 53*(2), 301-314.

Rudner, L. M., Getson, P. R., & Knight, D. L. (1980). Biased item detection techniques. *Journal of Educational Statistics*,*3*(5), 213-233.

Skaggs, G., & Lissitz, R. (1992). The Consistency of detecting item bias across-different test administration: implications of another failure. *Journal of Educational Measurement*, *29*(3), 227-242.

The National Center for Assessment and Evaluation in Higher Education. (2015). *Post-graduate general aptitude test guide*. (In Arabic) Riyadh.

Walker, C. M. (2011). What's the DIF? Why differential item functioning analyses are an important part of instrument development and validation. *Journal of Psychoeducational Assessment*, *29*(4), 364-376.

Wiberg, M. (2007). *Measuring and detecting differential item functioning in criterion-referenced licensing test: a theoretic comparison of methods* [Unpublished doctoral dissertation]. Umea University.

Zieky, M. (1993). Practical questions use of DIF statistics in item development. In P. W. Holland & H. Wainer (Eds.), *Differential item functioning* (337-364). Lawrence Erlbaum.