Dirasat: *Educational Sciences*, Volume 48, No. 3, 2021

Deanship of Scientific Research

Effect of Test Dimensionality and Strength of their Relationship on Statistical Properties and Standard Errors of Person Fit Indices

Rashid Al-mehrzi, Yousef Abu Shindi
College of Education, Sultan Qaboos University, Oman.

Received: 8/6/2020
Revised: 6/9/2020
Accepted: 27/9/2020
Published: 1/9/2021

Citation: Al-mehrzi, R., & Abu Shindi, Y. (2021). Effect of Test Dimensionality and Strength of their Relationship on Statistical Properties and Standard Errors of Person Fit Indices. *Dirasat: Educational Sciences*, *48*(3), 161-173. Retrieved from https://dsr.ju.edu.jo/djournals/index.php/Edu/article/view/2865

© 2021 DSR Publishers/ The University of Jordan.

This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY-NC) license https://creativecommons.org/licenses/by-nc/4.0/

Abstract

The study aims to examine the effect of test dimensionality and the strength of their relationship on statistical properties and standard errors of three-person fit indices (Wright, Drasgow, Almehrizi) using seven simulated data sets of 1000 subjects. These data sets result from two factors: Number of dimensions (three levels: one, two, three) and relationship strength among dimensions (three levels: 0.0, 0.4,.8). Results revealed that Drasgow and Almehrizi indices showed the highest percentages of aberrant responses which were close to expected rates (5%) whereas Wright index showed the lowest. Descriptive statistics showed that the Almehrizi index was the closest to the descriptive statistics of standard normal distribution. Also, the result showed that test dimensionality and strength of interrelationship among dimensions affected the standard errors of both the mean and variance of the Wright index, whereas did not affect the standard errors of both the mean and variance of the Almehrizi index and Drasgow index. Almehrizi index was the most consistent index, especially with variance.

Keywords: Wright index, drasgow index, almehrizi index, simulated data, nonfit patterns.

تأثير أبعاد الاختباروقوة العلاقة بينها في الخصائص الإحصائية والأخطاء المعيارية لمؤشرات مطابقة الفرد

راشد المحرزي، يوسف ابوشندي
جامعة السلطان قابوس.

ملخّص

في حين أشارت إلى وجود فروق على مستوى مجال المعلمين وعلى مستوى الأداة ككل، تعزى لمتغير الجنس، ولصالح هدفت الدراسة إلى تفحص تأثير أبعاد الاختبار وقوة العلاقة بين الأبعاد في الخصائص الإحصائية والأخطاء المعيارية لثلاثة من مؤشرات مطابقة الفرد (رايت، ودرازجو، والمحرزي)، وذلك من خلال توليد استجابات سبع مجموعات عشوائيًا بحيث احتوت كل مجموعة على 1000مفحوص.نتجت مجموعات البيانات من خلال التحكم بمستويات عاملين: عدد أبعاد الاختبار: (بعد واحد، بعدان، ثلاثة ابعاد)، وقوة الارتباط بين الأبعاد(0.0)، 0.4، .0.8) توصلت نتائج الدراسة إلى أن مؤشري درازجو والمحرزي كشفا عن أعلى نسبة أنماط غير مطابقة وقريبة من القيم المتوقعة (5%)، بينما كان مؤشر رايت أقلها. وبيّنت الإحصاءات الوصفية أن خصائص مؤشر المحرزي الأكثر توزعت على نحو طبيعي. وكذلك بينت النتائج أن بعدية الاختبار وقوة العلاقة الارتباطية بين الأبعاد أثرت في الخطأ المعياري لمؤشر رايت؛ ولم تؤثر في الخطا المعياري لمؤشري درازجو والمحرزي. وقد تفوق مؤشر المحرزي في مستوى استقرار قيم التباين.

الكلمات الدالة : مؤشر رايت، مؤشر درازجو، مؤشر المحرزي، بيانات مولدة، أنماط غير مطابقة.

المقدمة

تعتمد تقديرات نظرية استجابة الفقرة على تحقق افتراضاتها في البيانات الاختبارية، وعلى مطابقة أنماط استجابة المفحوصين للنماذج الرياضية المستخدمه؛ وتتمثل افتراضات النظرية في الاستقلال الموضعي بين فقرات الاختبار، وأحادية البعد (في النماذج احادية البعد)، والدالة الوتيرية المتزايدة بين قدرات الأفراد واحتمال استجابتهم على نحو صحيح على الفقرة، والتحرر من السرعة.

ويشير افتراض أحادية البعد إلى أن فقرات الاختبار تقيس سمة واحدة عند الفرد، وتتطلب نماذج استجابة الفقرة أأحادية البعد تحقق هذا الافتراض في البيانات الاختبارية قبل تطبيقها. ولكن يصعب في الكثير من المواقف تفسير أداء الفرد على فقرات الاختبار من خلال سمة واحدة بسبب تأثير العديد من العوامل المعرفية والوجدانية والبيئية في هذا الأداء. ويكتفي البعض بوجود عامل سائد Dominant Factor في الاختبار للدلالة على أحادية البعد (Hambleton& Swaminathan, 1985)، وذلك بالاعتماد على طرق التحليل العاملي الاستكشافي الخطي التي أكد هاتي (Hattie,1985) أنها طرق ضعيفة للحكم على تحقق افتراض أحادية البعد، وبرر ذلك بأن النماذج الرياضية في نظرية استجابة الفقرة هي علاقات غير خطية. ولأن السلوك الانساني يرتبط بالعديد من المكونات، لذا فإن التعامل معه يتطلب منظورًا متعدد الأبعاد؛ وعليه عكف المختصون في القياس الحديث (Mckinley& Way, 1992; Reckase, 1985) على اشتقاق نماذج رياضية متعددة الأبعاد قادرة على تفسير التفاعل بين معالم قدرات المفحوصين وفقرات الاختبارات.

ويؤثر عدم تحقق افتراضات نظرية استجابة الفقرة في البيانات الاختبارية على نحو مباشر على تقدير المعالم؛ حيث أشارت الدراسات مثل: (النعيمي، 2007؛ المقصقص، 2008؛ ابوشندي والمحرزي وعماره، 2018؛Chen & Wang,2007; Reese,1999; Yen,1993) إلى تأثير انتهاك افتراض الاستقلال الموضعي في تقدير معالم الفقرات، وبيّنت الدراسات (Zenisky, Hambleton & Sireci, 2002; Thompson & Pommerich ,1996) تأثر خصائص الاختبار كالثبات ودالة المعلومات ايضًا في انتهاك هذا الافتراض. وكذلك يؤثر عدم تحقق افتراض أحادية البعد في خصائص الاختبار فقد بيّنت الدراسات (ابوشندي، 2008؛ ابوشندي وعماره والمحرزي، 2018، Ackerman, 1994; Reckase, 2000; Way, Ansley & Forsyth, 1988) تاثر تقديرات معالم الفقرات وقدرات المفحوصين وخصائص الاختبار في إجراء تقديرات أحادية بعد لبيانات اختبارات متعددة الأبعاد.

ويجب على الفاحص التحقق من مطابقة أنماط استجابة الأفراد للنموذج الرياضي إلى جانب فحص تحقق افتراضات النظرية قبل البدء بتقدير معالم الفقرات والمفحوصين أو تطبيق أحد نماذج استجابة الفقرة لتقدير إحتمال استجابة الفرد على نحو صحيح على الفقرة، والتعامل مع الأنماط غير المطابقة قبل الشروع بتقدير معالم الفقرات والمفحوصين وتطبيق النموذج.

ويعزو ياسوناس وبيل وديفيد (Iasonas, Bill & David, 2000) أسباب حدوث أنماط استجابة المفحوصين غير المطابقة للنموذج إلى: خصائص المفحوص الشخصية (التخمين، الغش، القلق، مستوى الدافعية، التباطؤ، التكاسل)، وخصائص تربوية (سوء فهم المفحوص للفقرة، الإبداع الاستثنائي، الاختلاف في أساليب التدريس)، وكذلك يعزوها إلى انتهاك واحد من افتراضات نماذج استجابة الفقرة (أحادية البعد، والتحرر من السرعة)، وعوامل خارجية (خطأ تعبئة ورقة الاختبار، المرض)، وخصائص المفحوصين الديمغرافية (الجنس، والعرق، والتحيز الثقافي) (عوده، المحرزي، وابوشندي، 2019).

وشرع المختصون بتطوير مؤشرات للكشف عن أنماط استجابة الأفراد غير المطابقة للنموذج وصلت إلى ما يقارب 36 مؤشر وتم تصنيفها إلى: مؤشرات إحصائية مبنية على البواقي، ومؤشرات إحصائية مبنية على الأرجحية العظمى، ومؤشرات إحصائية مبنية على التباين المشترك (جراح، 2009).

ومن مؤشرات مطابقة الفرد المبنية على البواقي التي شاع استخدامها في الأدب السيكومتري مؤشري مربع المطابقة الكلي (الموزون وغير الموزون)(UMS_j) لرايت (Wright, 1977). وتتمثل معادلة مؤشر رايت غير الموزون في:

$$UMS_j = \frac{1}{n-1}\sum_{i=1}^{n}\frac{(y_{ij}-p_{ij})^2}{p_{ij}q_{ij}}$$

n: عدد الفقرات، y_{ij}: درجة الفرد i الملاحظة على الفقرة j، p_{ij}: احتمال استجابة الفرد i استجابة صحيحة على الفقرة j، q_{ij}: احتمال استجابة الفرد i استجابة خاطئة على الفقرة j.

ويعد مؤشر لوغاريتم الارجحية المعياري (L_z) الذي اشتقّه درازجو وليفين ووليم (Drasgow, Levin & William, 1985) من الأمثلة على المؤشرات المبنية على الأرجحية العظمى:

$$L_z = \frac{l_0 - E(l_0)}{\sqrt{Var(l_0)}}$$

$$l_0 = \sum_{i=1}^{n} y_{ij} \ln p_{ij} + (1 - y_{ij}) \ln q_{ij},$$

$$E(l_0) = \sum_{i=1}^{n} p_{ij} \ln p_{ij} + q_{ij} \ln q_{ij},$$

$$Var(l_0) = \sum_{i=1}^{n} p_{ij} q_{ij} \left(\ln \frac{p_{ij}}{q_{ij}} \right)^2.$$

ويمكن تبسيط هذه المعادلة لتكون كالآتي:

$$L_z = \frac{\sum_{i=1}^{n} (y_{ij} - p_{ij}) ln \frac{p_{ij}}{q_{ij}}}{\sqrt{\sum_{i=1}^{n} p_{ij} q_{ij} \left(ln \frac{p_{ij}}{q_{ij}} \right)^2}}.$$

وقدّم المحرزي (Al-mehrzi, 2010) نسختين من مؤشر جديد لمطابقة الفرد (الموزون وغير الموزون) يعتمد على البواقي ويعد تطويرًا لمؤشري رايت (Wright, 1977) وعلاجًا للمشكلات الإحصائية التي تعانيا منها، ويقدر مؤشر المحرزي الموزون(WSR) بالمعادلة:

$$\text{WSR} = \frac{SR_j - E(SR_j)}{\sqrt{Var(SR_j)}}$$

$$SR_j = \sum_{i=1}^{n} (y_{ij} - p_{ij})^2,$$

$$E(SR_j) = \sum_{i=1}^{n} p_{ij} q_{ij},$$

$$Var(SR_j) = \sum_{i=1}^{n} p_{ij} q_{ij} (p_{ij} - q_{ij})^2.$$

ويمكن تبسيط هذه المعادلة لتكون كالآتي:

$$WSR = \frac{\sum_{i=1}^{n} (p_{ij} - y_{ij})(p_{ij} - 0.5)}{\sqrt{\sum_{i=1}^{n} p_{ij} q_{ij} (p_{ij} - 0.5)^2}}.$$

اهتمت الدراسات بمؤشرات مطابقة الفرد؛ فمنها ما توجه نحو اشتقاق مؤشرات مناسبة لفحص المطابقة، ومنها ما ركز على المقارنة بين المؤشرات تحت ظروف معينة، ومنها ما تفحص تأثير المتغيرات الأخرى في دقة هذه المؤشرات. ومن الدراسات التي قارنت بين المؤشرات ما أجراه المحرزي (Al-Mehrzi, 2010) بالاعتماد على بيانات مولدة للمقارنة بين موشري المحرزي (الموزون وغير الموزون) ومؤشري رايت في أوضاع مختلفة نتجت من اختلاف مستويات صعوبة الفقرات وتمييزها وتخمينها وطول الاختبار وحجم العينة. توصلت النتائج إلى تفوق مؤشري المحرزي على مؤشري رايت في الخصائص الإحصائية واقترابها من خصائص التوزيع الطبيعي عند الأوضاع المختلفة للبيانات. كما أظهر مؤشرا المحرزي قيما للخطأ من النوع الأول قريبة من مستوى الدلالة ٥٪ بينما لم يكن ذات النتيجة مع مؤشري رايت.

وقارن لي واوليجنيك (LI & Olijnike, 1997) بين خمسة مؤشرات من بينها مؤشر درازجو باستخدام بيانات مولدة حسب نموذج راش مع اختلاف كل من: طول الاختبار (30، 60 فقرة) وبعدية الاختبار (أحادي البعد، ثنائي البعد)، ودرجة عدم المطابقة (مرتفعة، منخفضة). وبيّنت النتائج: 1. عدم دلالة الارتباط بين المؤشرات وتقدير قدرة المفحوصين، 2. انحراف التوزيع العيني للمؤشرات عن التوزيع الطبيعي، 3. عند استخدام المعيار المعدل لتحديد عدم المطابقة أظهرت أربعة من المؤشرات سلوكًا متقاربًا في تحديد حالات عدم المطابقة باختلاف بعدية الاختبار ودرجة عدم المطابقة وطول الاختبار 4. كشف مؤشر درازجو عن عدد حالات عدم مطابقة أكثر في حالة الدرجات العالية الزائفة في الاختبار ثنائي البعد، 5.

نسبة الخطأ من النوع الأول لكل مؤشر كانت أقل من مستوى 0.05 المفترض، 6. جميع مؤشرات راش كانت حساسة للدرجات المرتفعة الزائفة أكثر من الدرجات المنخفضة الزائفة في حالة الاختبارات ثنائية البعد ولم تختلف في حالة أحادية البعد، ٧. قدرة المؤشرات على الكشف زادت بزيادة طول الاختبار.

وقارن درازجو (Drasgow, 1982) أداء مؤشر الارجحية العظمى (L_0) الذي اشتقه ليفين وروبين (Levine & Rubin, 1979) في الكشف عن أنماط الاستجابة غير المطابقة بين النموذجين أحادي وثلاثي المعلم باستخدام بيانات اختبار(Graduate Record Examinations (GRE. وخلصت الدراسة الى أن المؤشر يكشف عن أنماط استجابة غير مطابقة أكثر عند الدرجات المرتفعة، ونسبة الأنماط غير المطابقة كانت أكبر في النموذج ثلاثي المعلم.

وتقصى روجرز وهاتي (Rogers & Hattie, 1987) معدل كشف مؤشري رايت بمطابقة الفرد الداخلي الخام (W)، والخارجي الخام (U) عن أنماط الاستجابة غير المطابقة عند مستويات مختلفة من كل من: تخمين الفقرة (مستويين)، تمييز الفقرة (3 مستويات)، بعدية الاختبار (أحادي، ثنائي). وبيّنت النتائج أن مؤشر رايت للمطابقة الخارجية يتأثر في بعدية الاختبار وتخمين الفقرات، وبينما لا يتأثر المؤشر الداخلي بأي من التمييز، والتخمين، والبعدية.

وقارن جراح (2009) بين ثلاثة مؤشرات: رايت وماسترز (Wright & Masters, 1982) (Z_w)، ومؤشر هولين ودرازجو وبارسونز (Hulin, Drasgow & Parsons, 1983) (L_Z)، ومؤشر لين وتاتسوكا (Linn & Tatsouka, 1983) ($ECI4_z$) باستخدام بيانات اختبار TIMSS، وقارن بين المؤشرات في أوضاع مختلفة نتجت من اختلاف كل من طول الاختبار (3 مستويات: 23، 36، 43)، ونموذج استجابة الفقرة (3 مستويات: أحادي المعلم، وثنائي المعالم، وثلاثي المعالم). واعتمد للمقارنة بين المؤشرات على نسبة أنماط الاستجابة غير المطابقة التي يكشفها المؤشر، وخاصية اللاتغير لمعلمة القدرة، والمقارنة بين خصائص توزيع المؤشرات عند ثلاثة مستويات لقدرة المفحوصين (منخفض، متوسط، مرتفع). توصلت الدراسة إلى أن مؤشر لين وتاتسوكا كشف عن أعلى نسبة من الأنماط وفي المستويات المختلفة للنموذج وطول الاختبارات، وتلاه مؤشر هولين ودرازجو وبارسونز. وعند المقارنة بين قوة المؤشرات في الكشف عن حالات عدم المطابقة بالاعتماد على خاصية اللاتغير، توصلت الدراسة إلى أن مؤشر لين وتاتسوكا كان أفضلها ثم مؤشر هولين ودرازجو وبارسونز وأخيرًا مؤشر رايت وماستر. وعند المقارنة حسب خاصية التوزيع الطبيعي تبيّن أن مؤشر لين وتانسوكا كان الأفضل عند ذوي القدرة العالية، ومؤشر رايت كان الأفضل عند ذوي القدرة المتوسطة. وعند ذوي القدرة المنخفضة كان مؤشر لين وتاتسوكا الأفضل للنموذجين الأحادي والثلاثي، ومؤشر رايت وماستر للنموذج الثنائي. كما أظهر مؤشر لين وتانسوكا ميلًا للتوزيع الطبيعي عند الحالات المختلفة للنموذج وطول الاختبار، وتشابهت خصائص المؤشرين الآخرين في آن لهما التواءً سالبًا وتفلطحًا موجبًا.

وقارن لوبيز ومونتيسينوس (Lopez & Montesinos, 2005) بين خمسة من مؤشرات مطابقة الفرد تحت مستويات مختلفة من متغيرات: حجم العينة (100، 250، 500، 1000)، ومدى صعوبة الفقرات ((-2،2)، (-1،1))، وشكل توزيع القدرة (طبيعي، ومنتظم)، وطول الاختبار (15 فقرة، 30 فقرة). وأجريت تقديرات المعالم باستخدام نموذج راش لبيانات تم توليدها باستخدام النموذج ثلاثي المعلم. توصلت الدراسة إلى ابتعاد الخصائص الإحصائية لمؤشري (L_z, Z_w) عن الخصائص المفترضة عند استخدام القيم المقدرة للمعالم بدلًا عن الحقيقية وتشابهت خصائصهما تحت جميع الظروف، واقتربت خصائصهما من التوزيع الطبيعي عند المدى الضيق لصعوبة الفقرات وعند زيادة طول الاختبار. واقتربت الخصائص الاحصائية لمؤشر ($ECI4_z$)4 Extended Caution index من التوزيع الطبيعي. كما توصلت الدراسة إلى أن اعلى معدل كشف كان لمؤشري: $ECI4_z$, $ECI2_z$), Extended Caution index (بين 5%-10%).

ويجد المتتبع والمراجع للأدب السيكومتري المتعلق بنظرية استجابة الفقرة الاهتمام الواضح بمؤشرات مطابقة الفرد والمقارنة بينها وتأثرها في عوامل قد تتواجد في البيانات الاختبارية، مثل افتراض أحادية البعد الذي كثيرًا لا يتحقق، لا سيما أن الدراسات بيّنت أثر انتهاك هذا الافتراض في تقدير معالم الفقرات كدراسة ابوشندي (2008). ومن المؤمل أن تقدم هذه الدراسة صورة واضحة حول أي من مؤشرات مطابقة الفرد أكثر تحملاً لعدم تحقق افتراض أحادية البعد وبوجود ارتباط بين أبعاد الاختبار، بما قد يساعد الباحثين وطلبة الدراسات العليا المتخصصين في نظرية استجابة الفقرة.

مشكلة الدراسة

بيّنت العديد من الدراسات أن مؤشرات مطابقة الفرد تتأثر في عوامل تتعلق بالمفحوصين والفقرات والاختبار؛ حيث توصل بعضها إلى أن المؤشرات تتأثر في طول الاختبار واختلاف معالم فقراته (Al-mehrizi, 2010; Deng & Torre, 2008; Li & Olijnike, 1997; Meijer & Van, 1999; Raise & Due, 1991؛ جراح، 2009)، وبعضها توصل إلى أن المؤشرات تتأثر في حجم عينة المفحوصين ومواصفات قدراتهم (Lopez & Montestinos, 2005; Deng & Torre, 2008؛ حمادنه، 2015). وتفحصت دراسة لي وليجنك (Li & Olijnike, 1997) تأثير بعدية الاختبار (احادي، ثنائي) في خمسة من مؤشرات مطابقة الفرد المعيارية (L_z, $ECI4_z$, $ECI2_z$, UT, WT).

وتحاول هذه الدراسة تفحص الخصائص الإحصائية لمؤشر المحرزي (Al-Mehrzi, 2010) ومقارنته مع مؤشرين من مؤشرات مطابقة الفرد: رايت (Wright, 1977)، ودرازجو (Drasgow, Levine& William, 1985) لكونه مؤشرًا جديدًا وذلك تحت ثلاث حالات من بعدية الاختبار (أحادي، ثنائي، متعدد)، وثلاثة درجات لقوة العلاقة الارتباطية بين الأبعاد (0.0، 0.4، 0.8). ويعد عدم تحقق افتراض أحادية البعد في البيانات الاختبارية من الأسباب التي تؤدي إلى وجود أنماط استجابة غير مطابقة.

أهداف الدراسة

تحاول الدراسة تحقيق الأهداف الآتية:

1. تفحص أثر عدد أبعاد الاختبار والعلاقة بينها في عدد ونسبة الأنماط غير المطابقة التي تكشفها مؤشرات مطابقة الفرد (درازجو، ورايت، والمحرزي).
2. تفحص أثر عدد أبعاد الاختبار والعلاقة بينها في الخصائص الإحصائية لمؤشرات مطابقة الفرد (درازجو، ورايت، والمحرزي).
3. تفحص أثر عدد أبعاد الاختبار والعلاقة بينها في الأخطاء المعيارية للوسط الحسابي وللتباين لمؤشرات مطابقة الفرد (درازجو، ورايت، والمحرزي).

أسئلة الدراسة

تحديدًا حاولت الدراسة الإجابة عن التساؤلات الآتية:

1. ما تأثير عدد أبعاد الاختبار والعلاقة بينها في عدد ونسبة الأنماط غير المطابقة التي تكشفها مؤشرات مطابقة الفرد (درازجو، ورايت، والمحرزي)؟
2. ما تأثير عدد أبعاد الاختبار والعلاقة بينها في الخصائص الإحصائية لمؤشرات مطابقة الفرد (درازجو، ورايت، والمحرزي)؟
3. ما تأثير عدد أبعاد الاختبار والعلاقة بينها في الأخطاء المعيارية للوسط الحسابي وللتباين لمؤشرات مطابقة الفرد (درازجو، ورايت، والمحرزي)؟

مصطلحات الدراسة

تعدد الأبعاد داخل الفقرات التعويضي: وجود سمتين أو أكثر يقيسها الاختبار، بحيث تقيس كل فقرة من فقراته سمتين أو أكثر أيضًا، ويعوض النقص في مستوى القدرة عند المفحوص من قدرة أخرى مرتفعة.

المؤشر الإحصائي لمطابقة الفرد: مؤشر يحدد البعد بين استجابات المفحوص الفعلية والاستجابات المفترضة من خلال النموذج، وتقارن قيم الإحصائي بقيم حرجة للوصول إلى الأنماط المطابقة وغير المطابقة.

قوة الارتباط بين الأبعاد: معامل الارتباط بين متجهات قدرة المفحوصين التي تعبر عن كل سمة تقيسها فقرة الاختبار، ففي حالة ثنائية البعد توجد قيمة واحدة للارتباط بين البعدين، وفي حالة تعدد الأبعاد توجد قيم متساوية أو مختلفة للارتباط بين جميع الأبعاد. وفي هذه الدراسة استخدمت ثلاث قيم للارتباط بين الأبعاد (0.0، 0.4، 0.8).

محددات الدراسة

اقتصرت الدراسة على:

- بيانات مولدة تواءم نماذج متعددة أبعاد تعويضية ثنائية الاستجابة.
- ثلاث حالات من بعدية البيانات، وثلاث حالات من قوة العلاقة الارتباطية بين الأبعاد.
- ثلاثة مؤشرات لمطابقة الفرد: رايت، درازجو، المحرزي.
- عشر عينات عند كل حالة.
- برنامج BILOG-MG في تقدير معالم الفقرات وقدرة المفحوصين، وعلى أكسل في حساب قيم مؤشرات مطابقة الفرد.

المنهجية والإجراءات

عينات الدراسة

تكونت عينة الدراسة من 10 عينات عشوائية لعدد سبع مجموعات من بيانات مولدة عشوائيًا باستخدام برمجية RESGEN4 لموراكي (Muraki, 2000) بحيث احتوت كل مجموعة على استجابات 1000 مفحوص على اختبار يحوي 20 فقرة. نتجت مجموعات البيانات من خلال التحكم بمستويات متغيرين: بعدية البيانات (3 حالات: أحادي البعد، ثنائي البعد، ثلاثي البعد)، وقوة الارتباط بين الأبعاد (3 درجات: 0.0، 0.4، 0.8).

إجراءات توليد البيانات

تتطلب برمجية توليد البيانات تزويدها بقيم ابتدائية لمعالم صعوبة وتمييز الفقرات؛ وفي هذه الدراسة استخدم النموذج متعدد الأبعاد داخل الفقرات التعويضي كما الآتي:

- **حالة أحادية البعد:** تختلف فقرات الاختبار في ما بينها بالتحكم في معلمة الصعوبة ومعلمة التمييز. وجرى اختيار قيم مختلفة لمعلمة صعوبة الفقرة لاختبار مكون من 20 فقرة وكذلك قيم مختلفة لمعلمة تمييز واحدة لكل فقرة.
- **حالة ثنائية البعد:** تختلف فقرات الاختبار في ما بينها بالتحكم في معلمة الصعوبة ومعلمتين للتمييز (واحدة لكل بعد). استخدمت معالم حالة أحادية البعد السابقة بالإضافة لقيم مختلفة لمعلمة تمييز ثانية للفقرات، فنتج اختبار مكون من 20 فقرة ثنائية البعد.
- **حالة ثلاثية البعد:** تختلف فقرات الاختبار في ما بينها بالتحكم في معلمة الصعوبة وثلاثة معالم للتمييز (واحدة لكل بعد). أضيفت معلمة تمييز ثالثة لكل فقرة على حالة ثنائية البعد السابقة. وضمن هذا الإجراء توحيد القيم الابتدائية لمعلمة صعوبة الفقرات لجميع حالات بعدية البيانات، وكذلك قيمة معلمة التمييز على البعد الأول.

أنتجت عملية توليد البيانات سبع مجموعات من استجابات الأفراد على الاختبارات (مجموعة واحدة على اختبار أحادي البعد، وثلاث مجموعات على اختبار ثنائي البعد، وثلاث مجموعات على اختبار ثلاثي البعد) واحتوت كل مجموعة على 1000 مفحوص. جرى اختيار معالم الفقرات الابتدائية بطريقة عشوائية ضمن مدى معلمتي الصعوبة والتمييز؛ حيث تم تقسيم مدى معلمة الصعوبة)-2، +2)الى 60 جزء متساوي على نحو متسق، ومن ثم تم الاختيار العشوائي لعشرين معلمة. تم استخدامها للفقرات أحادية وثنائية وثلاثية البعد. أما عن معلمة التمييز فقد تم ايضًا تقسيم المدى)0.19، 1.79)الى 60 جزء متساوي على نحو متسق ومن ثم اختيار ثلاث عينات منها على نحو عشوائي.

وفي ما يتعلق بقدرات المفحوصين الابتدائية تتطلب برمجية توليد البيانات خصائص توزيع معلمة القدرة (استخدم التوزيع الطبيعي في هذه الدراسة بمتوسط حسابي صفر وانحراف معياري 1)، ومعامل الارتباط بين القدرات المولدة (معلمة واحدة للقدرة في حالة احادية البعد، ومعلمتين في حالة ثنائية البعد، وثلاث في حالة ثلاثية البعد)، واستخدمت ثلاث قيم لمعاملات الارتباط بين قدرات الأفراد (0.0، 0.4، 0.8).

تم تكرار توليد هذه المجموعات السبع بنفس الطريقة عشر مرات لتمثل 10 عينات عشوائية عند كل حالة من الحالات السبع بمجموع 70 عينة وبعدد 70000 مفحوص.

الأساليب الإحصائية

تم إخضاع بيانات كل عينة لكل مجموعة من المجموعات السبع من البيانات للإجراءات الإحصائية الآتية:

1. معايرة البيانات باستخدام برنامج BILOG-MG لتقدير معالم الصعوبة والتمييز، وقدرات 1000 مفحوص على افتراض أن البيانات أحادية البعد باستخدام النموذج ثنائي المعلم.
2. حساب احتمال الإجابة الصحيحة لكل فقرة باستخدام القيم المقدرة لمعالم الصعوبة والتمييز والقيم المقدرة لقدرات الأفراد.
3. توظيف احتمال الاجابة الصحيحة لكل فقرة والاستجابة الفعلية لها في حساب المؤشرات الثلاثة لمطابقة الفرد: درازجو، ورايت، والمحرزي.
4. حساب الإحصاءات الوصفية لكل مؤشر من المؤشرات الثلاثة لمطابقة الفرد عبر 1000 مفحوص: أقل قيمة، وأكبر قيمة، والوسط الحسابي، والانحراف المعياري، وعدد الأفراد الذين يظهرون أنماط استجابة غير مطابقة.
5. اختبار ت للعينة الواحدة لفحص الدلالة الإحصائية للفروق بين المتوسطات الحسابية للمؤشرات الناتجة من البيانات والمتوسطات الحسابية النظرية.
6. اختبار تحليل التباين الأحادي لتفحص الدلالة الإحصائية للفروق بين المجموعات السبع الناتجة من اختلاف بعدية الاختبار وقوة الارتباط بين الأبعاد في قيم الوسط الحسابي لمؤشرات مطابقة الفرد.

نتائج الدراسة ومناقشتها

نتائج السؤال الأول:" ما تأثير عدد أبعاد الاختبار والعلاقة بينها في نسبة الأنماط غير المطابقة التي تكشفها مؤشرات مطابقة الفرد (رايت، ودرازجو، والمحرزي)؟"

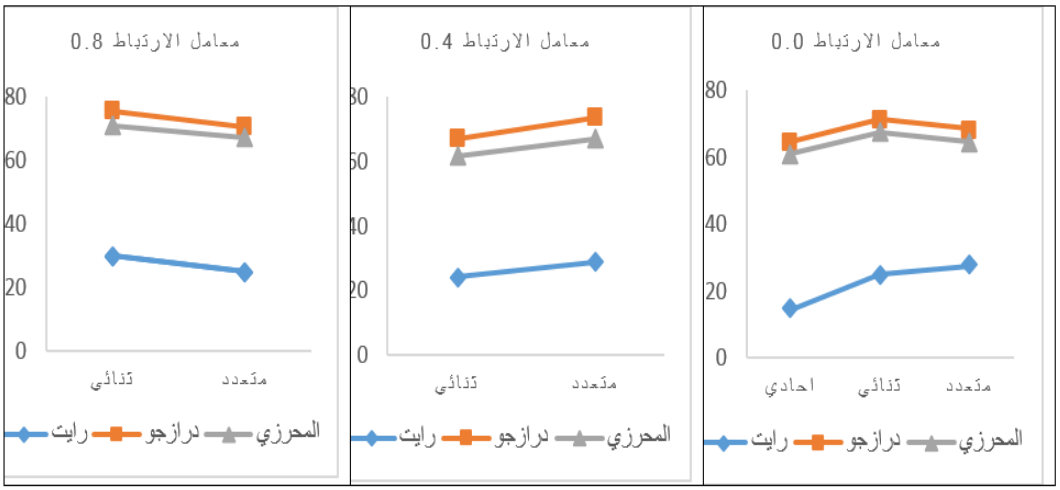
لإجابة هذا السؤال تم حساب المتوسط الحسابي لنسبة أنماط الاستجابة غير المطابقة التي كشفها كل مؤشر في العينات العشر لكل مجموعة من المجموعات السبع الناتجة من عدد الأبعاد وقوة العلاقة بينها التي تم عرضها في الجدول 1 والشكل 1.

الجدول 1: نسبة أنماط الاستجابة غير المطابقة التي كشفتها المؤشرات عند حالات أبعاد الاختبار وقوة الارتباط بينها

بعدية الاختبار	قوة الارتباط	رايت		درازجو		المحرزي	
		العدد	النسبة	العدد	النسبة	العدد	النسبة
احادي	-	15	1.5	65	6.5	61	6.1
ثنائي	0	25	2.5	71	7.1	68	6.8
	0.4	24	2.4	67	6.7	62	6.2
	0.8	30	3.0	76	7.6	71	7.1
ثلاثي	0	28	2.8	68	6.8	65	6.5
	0.4	29	2.9	74	7.4	67	6.7
	0.8	25	2.5	70	7.0	67	6.7

تظهر النتائج في الجدول 1 أن أعلى نسبة أنماط غير مطابقة كشف عنها مؤشر درازجو، ومن ثم مؤشر المحرزي وكان قريبًا من مؤشر درازجو، بينما أظهر مؤشر رايت أقل نسبة عند جميع حالات بعدية البيانات وقوة الارتباط بين الأبعاد. وقد وافقت هذه النتائج دراستي عوده (2018) و Li & Olijnike (1997)، التي توصلت إلى أن مؤشر درازجو كشف عن أعلى نسبة أنماط استجابة غير مطابقة للنموذج، وقد يعزى ذلك إلى أن مؤشري درازجو والمحرزي يقومان على تحويل متغير متصل إلى توزيع طبيعي (l_0) في درازجو و(SR_j) في المحرزي)، بينما مؤشر رايت يقوم على تحويل متغير منفصل (الاستجابة الثنائية للفرد على المفردة) إلى توزيع طبيعي ثم تحويله الى متغير مربع كاي. ولم تتغير نسبة أنماط الاستجابة غير المطابقة بانتهاك افتراض أحادية البعد عند فحصها باستخدام مؤشري درازجو والمحرزي. وكذلك لم يؤثر اختلاف قوة الارتباط بين الأبعاد في هذين المؤشرين.

ويلاحظ أيضًا أن نسبة الأنماط التي كشفها مؤشر المحرزي هي الأقرب الى نسبة الخطأ من النوع الأول 5% مقارنة بالمؤشرين الآخرين، فكانت النسبة لمؤشر رايت أقل من 5%، وكانت النسبة لمؤشر درازجو أكبر من 5%. وظهرت هذه النتائج عند جميع حالات بعدية البيانات وقوة الارتباط بين الأبعاد.



شكل 1 عدد الأنماط غير المطابقة للنموذج باختلاف بعدية الاختبار وقوة الارتباط بين الأبعاد

نتائج السؤال الثاني " ما تأثير عدد أبعاد الاختبار والعلاقة بينها في الخصائص الإحصائية لمؤشرات مطابقة الفرد (رايت، ودرازجو، والمحرزي)؟"

لإجابة هذه السؤال تم حساب الإحصاءات الوصفية لقيم المؤشرات عبر العينات العشر لكل مجموعة من المجموعات السبع الناتجة من الحالات الثلاث لبعدية الاختبار (أحادي البعد، ثنائي البعد، ثلاثي البعد)، والحالات الثلاث لقوة الارتباط بين الأبعاد (0.0، 0.4، 0.8) كما يظهر في الجداول 2، و3، و4.

أ- مؤشر رايت

تظهر النتائج في الجدول 2 اختلاف ظاهري في الوسط الحسابي والانحراف المعياري لمؤشر رايت عن الوسط الحسابي والانحراف المعياري وفقًا

للتوزيع التكراري المتوقع لمؤشر رايت (توزيع مربع كاي بمتوسط قدره 1 وانحراف معياري قدره 1.414) وذلك في جميع المجموعات السبع للبيانات. حيث أظهرت قيم الوسط الحسابي لمؤشر رايت اختلافًا بسيطًا عن الواحد الصحيح، وظهر الاختلاف الأكبر في حالة العلاقة القوية بين الأبعاد الثلاثة (0.8)، حيث كانت قيمة الوسط الحسابي 0.88. وفي المقابل جاء الانحراف المعياري لمؤشر رايت مرتفعًا ومختلفًا بقدر كبير عن القيمة المتوقعة لتوزيع مربع كاي في جميع حالات البعدية وقوة العلاقة بينها، فقد وصل إلى 3.43 في حالة ثنائي البعد ومعامل ارتباط 0.8. ولفحص الدلالة الإحصائية لهذه الفروق استخدم اختبار ت للعينة الواحدة One-sample t-test للمقارنة بين المتوسطات الحسابية لقيم المؤشر عند الحالات المختلفة والمتوسط النظري لتوزيع مربع كاي (1)، وأظهرت النتائج فروقًا دالةً احصائيًا عند حالة ثلاثي البعد وقوة الارتباط 0.8. وتتوافق هذه النتيجة جزئيًا مع ما توصلت اليه دراسات (جراح، 2009؛ عوده، 2018) التي توصلت إلى أن الإحصاءات الوصفية لمؤشر رايت في النموذج ثنائي المعلم لم تختلف عن التوزيع المفترض للمؤشر(توزيع مربع كاي).

ويضاف اختلاف توزيع مؤشر رايت باختلاف عدد أبعاد الأختبار وارتفاع قوة الارتباط بين الأبعاد عن التوزيع المفترض إلى مجموعة قضايا توصلت لها الدراسات السابقة حول فائدة هذا المؤشر؛ ومنها ما توصلت إليه دراسات (Hambleton et al, 1987; Smith, 1982; Waller, 1981) من أن مؤشر رايت حساس لاختلاف طول الاختبار، وحجم عينة المفحوصين، ويحتاج إلى عينات كبيرة الحجم، وهذا يتنافى مع خصائص توزيع مربع كاي الذي هو حساس لزيادة حجم العينة.

الجدول 2: المتوسطات الحسابية والانحرافات المعيارية لقيم مؤشر رايت تبعًا لاختلاف عدد الأبعاد المكونة للاختبار وقوة العلاقة بينها

بعدية الاختبار	قوة الارتباط	الوسط الحسابي	الانحراف المعياري	أقل قيمة	أكبر قيمة	قيمة ت	الدلالة الاحصائية
احادي	-	1.00	1.22	0.14	25.53	0.08	0.941
ثنائي	0.0	0.97	1.44	0.09	26.49	-2.01	0.075
	0.4	0.98	2.86	0.06	73.07	-0.59	0.572
	0.8	1.04	3.43	0.06	85.57	1.47	0.177
ثلاثي	0.0	0.96	2.47	0.06	55.21	-1.99	0.078
	0.4	0.93	3.34	0.03	78.42	-1.79	0.108
	0.8	0.88	3.15	0.02	67.25	-2.90	0.018

ب- مؤشر درازجو

تظهر النتائج في الجدول 3 على نحو واضح عدم اختلاف الوسط الحسابي والانحراف المعياري لمؤشر درازجو عن القيم المناظرة في التوزيع التكراري المتوقع (توزيع طبيعي بوسط حسابي صفر وانحراف معياري واحد) باختلاف بعدية البيانات (أحادي، ثنائي، متعدد)، وباختلاف قوة العلاقة الارتباطية بين الأبعاد. وللتحقق من ذلك استخدم اختبار ت للعينة الواحدة للمقارنة بين الوسط الحسابي للمؤشر عند الحالات المتعددة والوسط الحسابي النظري(0) وأكدت النتائج أن الفروق غير دالة إحصائيًا. وتوافق هذه النتيجة ما جاء في عدد من الدراسات (عودة والمحرزي وابوشندي، 2019؛ جراح، 2009؛ Drasgow, Levin& Williams, 1985) من حيث اقتراب توزيع درازجو من الطبيعي وخاصة في الاختبارات القصيرة (23 فقرة) والنموذج ثنائي المعلم، وتوافق أيضًا دراسة (Molennar& Hoijtink as cited in Meijer& Sijtsma, 1994) من حيث إن قيمة تباين مؤشر درازجو ينحرف عن المتوقع (1) في الاختبارات القصيرة. وتتفق نتائج الدراسة مع نتيجة دراسة جراح (2009) من حيث إن توزيع مؤشر رايت يقترب من الطبيعي في حالة النموذج ثنائي المعلم وكذلك أوصت دراسة حمادنة (2015) باستخدام هذا المؤشر في للكشف عن أنماط الاستجابة غير المطابقة للنموذج بغض النظر عن النموذج اللوجستي (ثنائي المعلم، وثلاثي المعلم)، وعند اختلاف مستويات قدرات المفحوصين.

الجدول 3: المتوسطات الحسابية والانحرافات المعيارية لقيم مؤشر درازجو تبعًا لاختلاف عدد الأبعاد المكونة للاختبار وقوة العلاقة بينها

بعدية الاختبار	قوة الارتباط	الوسط الحسابي	الانحراف المعياري	أقل قيمة	أكبر قيمة	قيمة ت	الدلالة الاحصائية
احادي	-	0.01-	0.99	4.45-	1.88	-0.86	0.411
ثنائي	0.0	0.00	1.00	4.56-	1.72	-0.32	0.760
	0.4	0.03	0.97	4.57-	1.70	1.48	0.174
	0.8	0.02-	1.02	5.43-	1.62	-1.51	0.165
ثلاثي	0.0	0.02	0.99	4.85-	1.52	1.12	0.292
	0.4	0.01	1.00	5.23-	1.46	1.14	0.283
	0.8	0.00	0.99	5.51-	1.46	0.41	0.692

ج- مؤشر المحرزي

تظهر النتائج في الجدول 4 على نحو واضح عدم اختلاف الوسط الحسابي والانحراف المعياري لمؤشر المحرزي عن القيم المناظرة في التوزيع التكراري المتوقع (توزيع طبيعي بوسط حسابي صفر وانحراف معياري واحد) باختلاف بعدية البيانات (احادي، ثنائي، متعدد)، وباختلاف قوة العلاقة الارتباطية بين الأبعاد. وللتحقق من ذلك استخدم اختبار ت للعينة الواحدة للمقارنة بين الوسط الحسابي للمؤشر عند الحالات المتعددة والوسط الحسابي النظري(0)، وأكدت النتائج أن الفروق غير دالة إحصائيًا. ووافقت هذه النتيجة دراسة المحرزي (Al-mehrzi, 2010; Al-mahrazi, 2004) التي أكدت التوزيع الطبيعي لمؤشر المحرزي وخاصة في البيانات المولدة. وكذلك دلل المحرزي (2003) على أن مؤشره يعمل على توحيد تحويل مربع البواقي المعيارية بين اجابة المفحوص الملاحظة واحتمال استجابته على الفقرة، وهذا بدوره يقبل أي قيمة بين الصفر والواحد، وكذلك تقوم على أساس توحيد مربع البواقي المعيارية لاستجابات الفرد على فقرات الاختبار بعد الحصول على المتوسط الحسابي المتوقع وهذا موحد افتراضًا، و على نحو مختصر تظهر أهمية مؤشر المحرزي في أنه يقدم المتوسطات لتحسين منهج البواقي المعيارية في سياق نظرية استجابة الفقرة.

الجدول 4: المتوسطات الحسابية والانحرافات المعيارية لقيم مؤشر المحرزي تبعًا لاختلاف عدد الأبعاد المكونة للاختبار وقوة العلاقة بينها

بعدية الاختبار	قوة الارتباط	الوسط الحسابي	الانحراف المعياري	أقل قيمة	أكبر قيمة	قيمة ت	الدلالة الاحصائية
احادي	-	0.01	0.99	2.06-	4.12	0.99	0.346
ثنائي	0.0	0.00	1.00	1.89-	4.10	0.10	0.924
	0.4	0.04-	0.98	1.87-	4.24	-1.79	0.107
	0.8	0.02	1.01	1.79-	4.22	1.46	0.178
ثلاثي	0.0	0.02-	1.00	1.78-	4.66	-2.20	0.055
	0.4	0.02-	1.00	1.73-	4.48	-2.30	0.053
	0.8	0.00	0.99	1.71-	4.61	-0.15	0.887

ولتفحص الدلالة الإحصائية للفروق بين المجموعات السبع الناتجة من اختلاف بعدية الاختبار وقوة الارتباط بين الأبعاد في قيم الوسط الحسابي لمؤشرات مطابقة الفرد، استخدم تحليل التباين الأحادي. وتُظهر نتائجه في الجدول 5 أن اختلاف بعدية الاختبار والعلاقة بين الأبعاد تؤثر و على نحو دال احصائيًا (α= 0.05) في قيم مؤشرات مطابقة الفرد (رايت ودرازجو والمحرزي)، وعند استخدام اختبار شيفية للمقارنات البعدية أظهرت النتائج أن الفروق الثنائية (بين كل حالتين) لم تكن دالة احصائيًا عند مستوى α= 0.05 الا بين حالتي الثنائي وقوة الارتباط 0.4 والمتعدد وقوة ارتباط 0.8 لمؤشر رايت(Mean Differences=0.154, sig=0.032)، وهذا يؤشر على أن الفروق بين الحالات على نحو عام هي فروق تراكمية فقط ولا توجد فروق ثنائية دالة احصائيًا.

الجدول 5: تحليل التباين الأحادي لقيم الوسط الحسابي لمؤشرات مطابقة الفرد وفقًا لحالات بعدية الاختبار وقوة الارتباط بينها

		مجموع المربعات	درجات الحرية	متوسط المربعات	ف	الدلالة لاحصائية	مربع ايتا
رايت	بين المجموعات	0.150	6	0.025	3.15	0.009	0.231
	داخل المجموعات	0.501	63	0.008			
	الكلي	0.652	69				
درازجو	بين المجموعات	0.015	6	0.003	2.41	0.037	0.187
	داخل المجموعات	0.066	63	0.001			
	الكلي	0.081	69				
المحرزي	بين المجموعات	0.023	6	0.004	3.52	0.005	0.251
	داخل المجموعات	0.068	63	0.001			
	الكلي	0.091	69				

ثالثاً: نتائج السؤال الثالث:"ما تأثير عدد أبعاد الاختبار والعلاقة بينها في الأخطاء المعيارية للوسط الحسابي والتباين لمؤشرات مطابقة الفرد (رايت، ودرازجو، والمحرزي)؟"

لإجابة هذه السؤال تم حساب الخطأ المعياري للوسط الحسابي والتباين لكل مؤشر من مؤشرات مطابقة الفرد عبر العينات العشر باستخدام المعادلة:

$$SE = \sqrt{\frac{1}{10}\sum(\theta_s - \bar{\theta})^2}$$

حيث θ_s قيمة المعلم (الوسط الحسابي والتباين) لكل عينة، $\bar{\theta}$ المتوسط الحسابي للمعلم عبر العينات.

وتبيّن قيم الأخطاء المعيارية للوسط الحسابي لمؤشرات مطابقة الفرد في الجدول 6 عبر العينات العشر في المجموعات السبع الناتجة من اختلاف بعدية الاختبار وقوة العلاقة بين الأبعاد أن مؤشر رايت أظهر أخطاء معيارية أكبر من مؤشري درازجو والمحرزي في جميع المجموعات السبع. كما أظهرت الأخطاء المعيارية لمؤشر رايت أنها تزداد بزيادة عدد أبعاد الاختبار وزيادة قوة الارتباط بين هذه الأبعاد. فتراوحت قيم الخطأ المعياري لمؤشر رايت من 0.047 في حالة البعد الواحد الى 0.122 في حالة ثلاثة أبعاد بقوة علاقة بمقدار 0.8.

الجدول 6: الأخطاء المعيارية للوسط الحسابي لمؤشرات مطابقة الفرد عند حالات بعدية الاختباروقوة الارتباط بين الأبعاد

المؤشر	البعدية/قوة الارتباط	0.0	0.4	0.8
رايت	احادي	0.047	-	-
	ثنائي	0.050	0.080	0.075
	ثلاثي	0.063	0.121	0.122
درازجو	احادي	0.034	-	-
	ثنائي	0.031	0.034	0.031
	ثلاثي	0.015	0.027	0.036
المحرزي	احادي	0.035	-	-
	ثنائي	0.032	0.034	0.030
	ثلاثي	0.019	0.031	0.035

وأظهرت النتائج تقارب مستوى الأخطاء المعيارية للأوساط الحسابية لمؤشري درازجو والحرزي، فقد كانت قيم الأخطاء المعيارية للمؤشرين صغيرة وقريبة من الصفر عبر جميع المجموعات السبع، ولم تتأثر الأخطاء المعيارية في تغير بعدية الاختبار (ثنائية وثلاثية البعد) وقوة العلاقة بينها. فقد تراوحت قيم الأخطاء المعيارية لمؤشر درازجو بين 0.015 و 0.036، وتراوحت قيم الأخطاء المعيارية لمؤشر المحرزي بين 0.019 و0.035.

ويظهر الجدول 7 الأخطاء المعيارية للتباين لكل مؤشر من مؤشرات مطابقة الفرد عبر العينات العشر في المجموعات السبع الناتجة من اختلاف بعدية الاختبار وقوة العلاقة بين الأبعاد. وأظهر مؤشر المحرزي أصغر الأخطاء المعيارية للتباين بينما أظهر مؤشر رايت قيمًا كبيرةً للأخطاء المعيارية للتباين. فقد تراوحت قيم الأخطاء المعيارية للتباين لمؤشر المحرزي بين 0.033 و 0.058، وتراوحت قيم الأخطاء المعيارية للتباين لمؤشر درازجو بين 0.047 و 0.081، وقيم الأخطاء المعيارية للتباين لمؤشر رايت بين 1.022 و 17.601. وتشير هذه النتائج إلى أن قيم التباين لمؤشر رايت غير مستقرة ويقل استقرارها عند زيادة انتهاك بعدية الاختبار وزيادة قوة العلاقة بين الأبعاد. بينما تشير هذه النتائج إلى استقرار قيم التباين لمؤشري المحرزي ودرازجو (مؤشر المحرزي أكثر استقرارًا) سواءً عند تحقق افتراض أحادية البعد أو انتهاكه وعند مستويات مختلفة من قوة الارتباط بين الأبعاد.

الجدول 7: الأخطاء المعيارية لتباين مؤشرات مطابقة الفرد عند حالات بعدية الاختباروقوة الارتباط بين الأبعاد

المؤشر	البعدية/قوة الارتباط	0.0	0.4	0.8
رايت	أحادي	1.022	-	-
	ثنائي	1.000	14.502	8.008
	ثلاثي	6.019	17.601	8.406
درازجو	أحادي	0.069	-	-
	ثنائي	0.066	0.081	0.074
	ثلاثي	0.055	0.047	0.078
المحرزي	أحادي	0.04	-	-
	ثنائي	0.045	0.054	0.058
	ثلاثي	0.033	0.048	0.055

الاستنتاجات والتوصيات

هدفت الدراسة إلى تفحص تأثير أبعاد الاختبار وقوة العلاقة بينها في الخصائص الإحصائية والاخطاء المعيارية لثلاثة من مؤشرات مطابقة الفرد (رايت، ودرازجو، والمحرزي) من خلال بيانات مولدة. وخلصت نتائج الدراسة إلى أن أعلى نسبة أنماط غير مطابقة كشف عنها مؤشر درازجو، ومن ثم مؤشر المحرزي وأقلها مؤشر رايت مع ملاحظة التقارب الكبير بين نسبة مؤشري درازجو والمحرزي. وجاءت الخصائص الإحصائية الوصفية لمؤشر المحرزي الأقرب إلى التوزيع الطبيعي، وكذلك مؤشر درازجو مع تغير بسيط في انحرافه المعياري، وتغيرت على نحو واضح خصائص مؤشر رايت الإحصائية عن خصائص توزيع مربع كاي باختلاف بعدية الاختبار وقوة الارتباط بين الأبعاد. وجاءت قيم الأخطاء المعيارية لتؤكد على وجود تأثير واضح لبعدية الاختبار والعلاقة بين الأبعاد في نتائج مؤشر رايت، ولم يتأثر مؤشري درازجو والمحرزي الا بفروق بسيطة بين حالات اختلاف بعدية الاختبار. وفي ضوء هذه النتائج توصي الدراسة بالآتي:

- استخدام مؤشري المحرزي ودرازجو للكشف عن أنماط استجابة الأفراد غير المطابقة للنموذج اللوجستي ثنائي المعلم عند عدم تحقق افتراض أحادية البعد في البيانات الاختبارية.
- استخدام مؤشري المحرزي ودرازجو للكشف عن أنماط استجابة الأفراد غير المطابقة للنموذج اللوجستي ثنائي المعلم للنموذج عند وجود ارتباط بين أبعاد الاختبار.
- إجراء دراسات مقارنة بين هذه المؤشرات ومؤشرات أخرى لمطابقة الفرد.
- إجراء دراسات مقارنة بين هذه المؤشرات باستخدام بيانات حقيقية.

المصادر والمراجع

أبوشندي، ي. (2008). تأثير تعدد الأبعاد للاختبار والعلاقة بينها في تقديرات معالم فقراته: دراسة محاكاة، *رسالة دكتوراه غير منشورة، جامعة اليرموك، الأردن*.

أبوشندي، ي.، المحرزي، ر. وعماره، ا. (2018). دقة تقدير العلامات الحقيقية عند درجات مختلفة للارتباط الموضعي بين فقرات الاختبار في توزيعات مختلفة للقدرة. *مجلة العلوم التربوية والنفسية، البحرين*، 19(3)، 466-491.

جراح، ب. (2009). مقارنة مؤشرات مطابقة الشخص لنماذج استجابة الفقرة باستخدام بيانات فعلية، *رسالة دكتوراه غير منشورة، جامعة اليرموك، الأردن*.

حمادنة، ا. (2015). تقصي فاعلية الإحصائي (LZ*new) في الكشف عن أنماط استجابة المفحوص غير المطابقة وفق نظرية استجابة الفقرة. *مجلة العلوم التربوية والنفسية*، 16(3)، 565- 593.

عوده، ا.، المحرزي، ر.، وابوشندي، ي. (2019). طريقة مطورة لتفسير أنماط الاستجابات غير المطابقة للفرد في الاختبارات ومقارنة خمسة مؤشرات مطابقة الفرد. *مجلة جامعة الشارقة للعلوم الاجتماعية والانسانية*.

المقصقص، م. (2008). مقارنة طرق للكشف عن افتراض الاستقلال الموضعي لفقرات متعددة الأبعاد عند مستويات مختلفة للقدرة: دراسة محاكاة، *رسالة دكتوراه غير منشوره، جامعة اليرموك، الأردن*.

النعيمي، ع. (2007). أثر انتهاك افتراض الاستقلال الموضعي على التقديرات المختلفة لنظرية استجابة الفقرة، *رسالة دكتوراة غير منشورة، جامعة اليرموك، اربد، الأردن*.

References

Abu shindi, Y.(2008). The influence of test dimensionality and correlation between them on items parameters: A simulation study, *Unpublished doctoral dissertation, Yarmouk University, Jordan*.

Abu Shindi, Y., Al-mehrzi, R., & Omara, E.(2018). The Estimation Accuracy of True Scores at Different Degrees of Local Dependence among Test Items for Different Ability distributions. *Journal of educational and psychological Science-Bahrain*, 19(3), 466-491.

Ackerman, T.(1994). Using multidimensional item response theory to understand what items and tests are measuring. *Applied measurement in education*, 7(4), 255-278.

Al-Mahrazi, R. (2004). Investigating a new modification of the residual-based person fit index and its relationship with other

indices in dichotomous item response theory, *Unpublished dissertation, University of Iowa, Iowa City*.

Al-Mehrzi, R. (2010). Comparing among new residual-fit and wright's Indices for dichotomous three-parameter IRT model with standardized tests. *Journal of Educational and Psychological Studies, Sultan Qaboos University*, 4(2), 14-26

Al-Mqasqus, M.(2008). Comparing Methods for Detecting Local Independence of Multidimensional Items at Different Ability Levels: A Simulation Study, *Unpublished doctoral dissertation, Yarmouk University, Jordan*.

Alnuami, E.(2007). The effect of Local independent on item response theory estimation, *Unpublished doctoral dissertation, Yarmouk University, Jordan*.

Chen, C. & Wang, W. (2007). Effect of ignoring item interaction on item parameter estimation and detection of interacting items. *Applied Psychological Measurement*, (31), 388-411.

Deng, W., & Torre, J. (2008). Improving person fit assessment by correcting the ability estimate and its reference distribution. *Journal of Educational Measurement*, 45(2), 159-177.

Drasgow, F. (1982). Choice of the test models for appropriateness measurement. *Applied Psychological Measurement*, (6), 297-308.

Drasgow, F., Levine, M.V., & Williams, E.A. (1985). Appropriateness measurement with polytomous item response models and standardized indices. *British Journal of Mathematical and statistical psychology*, (38), 67-86.

Hambleton, R., & Swaminathan, H. (1985). *Item Response Theory: Principles and applications*. Boston: Kluwer-Nijhoff.

Hambleton, R., Swaminathan, H., Cook, L., Eignor, D., & Gifford, J.(1987). Developments in Latent trait theory: Models, technical issues, and applications. *Review of Educational Research*, (48), 467-510.

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

Hulin, C., Drasgow, F., & Parsons, C. (1983). *Item Response Theory: Application to Psychological Measurement*. Homewood, II: Irwin.

Iasonas, L., Bill, B., & David, W.(2000). The consistency of examinee misfit across tests on the same subject and across subject: the case of the KS2 mathematics and science National Curriculum tests in England.

Jarrah, B.(2009). Comparative of Person fit indices in item response theory models by actual Data, *Unpublished doctoral dissertation, Yarmouk University, Jordan*.

Levine, M., & Rubin, D. (1979). Measuring the appropriateness of multiple - choice test scores. *Journal of Educational Statistics*, (4), 269-290.

Li, M. F., & Olijnike, s. (1997). The power of rasch person-fit statistics in detecting unusual person patterns. *Applied Psychological Measurement*, (21), 215-231.

Linn, R., & TatsukaK. (1983). Indications for detecting unusual patterns: Links between two general approaches and potential application. *Applied psychological measurement*, 7, 81-96.

Lopez, A., & Montesinos, H. (2005). Fitting rasch model using appropriateness Measure Statistics. *The Spanish Journal of Psychology*, (8), 100-110.

McKinley, R. L., & Way, W. D. (1992). The feasibility of modeling secondary TOEFL ability dimensions using multidimensional IRT models. *ETS Research Report Series*, (1), i-22.

Meijer, R. R., & Sijtsma, K. (1994). Detection of aberrant item score patterns: A review of recent developments. *Applied Measurement in Education*, *8*(3), 261-272.

Meijer, R., & Van, K. (1999). The Null distribution of the person-fit statistics for conventional and adaptive tests. *Applied psychological Measurement*, 23, 327-345.

Muraki, E. (2000). *RESGEN: Item response generator*. Princeton, NJ: Educational Testing Service.

Odeh, A., Almehrzi, R., & Abu shindig, Y. (2019).An Improved method to interpret persons aberrant response pattern in tests and acomparative of five person fit indices. *Alshariqa university for social and humanities Science*.

Raise, S., & Due, A. (1991). The influence of test characteristic on the detection of aberrant response patterns. *Applied psychological Measurement*, (15), 217-226.

Reckase, M. (2000). *Multidimensional Item Response Theory*. New York: Springer.

Reckase, M. D. (1985). The difficulty of test items that measure more than one ability. *Applied Psychological Measurement*, 9(4), 401-412.

Reese, L. M. (1999). A Classical Test Theory Perspective on LSAT Local Item Dependence. LSAC Research Report Series. Statistical Report.

Rogers, H., & Hattie, J.(1987). A Monte Carlo Investigation of several person and item fit statistics for item response models. *Applied Psychological Measurement*, (11), 47-57.

Smith, R.(1982). Detecting measurement distribution with the rash model, *Unpublished doctoral dissertation, University of Chicago*.

Thompson, T. D., & Pommerich, M. (1996). Examining the Sources and Effects of Local Dependence.

Waller, M. (1981). A procedure for comparing logistic latent trait models. *Journal of Educational Measurement*, (18), 119-125.

Way, W. D., Ansley, T. N., & Forsyth, R. A. (1988). The comparative effects of compensatory and non-compensatory two-dimensional data on unidimensional IRT estimation. *Applied psychological Measurement*, 12(3), 239-252.

Wright, B. D. (1977). Solving measurement problems with the Rasch model. *Journal of Education Measurement*, (114), 96-115.

Wright, B. D., & Masters, G. N. (1982). *Rating scale analysis: Rasch measurement*. Chicago: Mesa Press.

Yen, W. (1993). Scaling performance assessments: strategies for managing local item dependence. *Journal of Educational Measurement*, 30(3), 187–213.

Zenisky, A. L., Hambleton, R. K., & Sireci, S. G. (2001). Effects of Local Item Dependence on the Validity of IRT Item, Test, and Ability Statistics. MCAT Monograph.