Jordan Journal of Applied Science-Humanities Series

Volume 31
Issue 1 *31-1*

Article 6

2022

The Effect of Missing Data Handling Methods and its Volume on Detecting Differential Item Functioning of Test Items

Iman Mustafa Ghazou
Yarmouk University, Irbid, Jordan., eman.ghazo1982@gmail.com

Follow this and additional works at: https://digitalcommons.aaru.edu.jo/jjoas-h

Part of the Education Commons

Recommended Citation

Ghazou, Iman Mustafa (2022) "The Effect of Missing Data Handling Methods and its Volume on Detecting Differential Item Functioning of Test Items," *Jordan Journal of Applied Science-Humanities Series*: Vol. 31: Iss. 1, Article 6.
Available at: https://digitalcommons.aaru.edu.jo/jjoas-h/vol31/iss1/6

This Article is brought to you for free and open access by Arab Journals Platform. It has been accepted for inclusion in Jordan Journal of Applied Science-Humanities Series by an authorized editor. The journal is hosted on Digital Commons, an Elsevier platform. For more information, please contact rakan@aaru.edu.jo, marah@aaru.edu.jo, u.murad@aaru.edu.jo.

VOL 31 NO 1 ISSN: 2708-9126

Jordan Journal of Applied Science

Humanities Series

Applied science Private University

The Effect of Missing Data Handling Methods and its Volume on Detecting Differential Item Functioning of Test Items

أثر طرق معالجة البيانات المفقودة وحجمها في الكشف عن الأداء التفاضلي لفقرات الاختبار

Iman Mustafa Ghazwa[1*], Amal Ahmed Al Zoubi.[2]
Yarmouk University, Irbid, Jordan.[12]

ARTICLE INFO

Article history:
Received 19 Dec 2020
Accepted 23 Feb 2021
Published 01 Apr 2022

https://doi.org/10.35192/jjoas-h.v31i1.309

***Corresponding author** at Yarmouk University, Irbid, Jordan.
Iman Mustafa Ghazwa.
Email: eman.ghazo1982@gmail.com.

Keywords:
missing data
Generalized Mantel Haenszel Test
(TIMSS, 2015)
Differential Item Functioning
Missing Data Handling Methods

الكلمات المفتاحية:
البيانات المفقودة
اختبار مانتل هانزل المعمم
اختبار تيمس لعام ٢٠١٥
الأداء التفاضلي للفقرة، طرق معالجة
البيانات المفقودة

ABSTRACT

The study aimed at investigating the effect of different methods for handling missing data and its volume on detecting differential item functioning (DIF) for items of test according to the gender variable using the generalized Mantel Haenszel test (GMH). To achieve the aims of the study, the responses of Jordanian eighth grade students to the items of the Mathematics test in the International Study (TIMSS, 2015) were obtained. The sample of the study consisted of 2261 students, who responded to four test booklets (1, 7, 9, 13). Several methods were used for handling missing data (missing data as nominal, corrected item mean substitution, expectation- maximization algorithm, multiple imputation, and chained equations), GMH was also used to detect DIF depending on the gender variable after each treatment method. The effect of the volume of missing data (5%, 10%) on the percentages of questions that showed DIF was also shown. The results indicated that there were no statistically significant differences at (α=0.05) between the ratios of items that showed DIF across different missing data processing methods. In contrast, the five methods agreed on the existence of a statistically significant relationship at (α=0.05) between the presence of DIF in favor of females in the field of algebra and in favor of males in the field of numbers. As for the volume of missing data, the findings revealed that there were no differences in the percentage of questions that showed DIF in favor of males or females due to the volume of missing data of 5% or 10%.

هدفت الدراسة الحالية إلى التحقق من أثر الطرق المختلفة لمعالجة البيانات المفقودة وحجمها في الكشف عن الأداء التفاضلي للفقرة (DIF) تبعًا لمتغير الجنس باستخدام اختبار مانتل هانزل المعمم (GMH). ولتحقيق أهداف الدراسة، تم الحصول على استجابات طلبة الصف الثامن الأساسي الأردنيين على فقرات اختبار الرياضيات في الدراسة الدولية (TIMSS, ٢٠١٥)، وبلغت عينة الدراسة ٢٢٦١طالبًا ممن استجابوا على أربع كراسات اختبارية (١، ٧، ٩، ١٣). تم استخدام عدة طرق لمعالجة البيانات المفقودة (من دون تعويض، والوسط المصحح للفقرة، وتعظيم التوقعات، والتعويض المتعدد، والمعادلات المتسلسلة)، وتم استخدام اختبارGMH للكشف عن DIF تبعًا لمتغير الجنس بعد كل طريقة معالجة. كما تم بيان أثر حجم البيانات المفقودة (٥%، ١٠%) في نسب الأسئلة التي تُظهرDIF. أشارت النتائج إلى عدم وجود فروق ذات دلالة إحصائية عند(α=٠,٠٥) بين نسب الفقرات التي أظهرت DIFعبر مختلف طرق معالجة البيانات المفقودة، في المقابل اتفقت الطرق الخمسة على وجود علاقة دالة إحصائيًا عند (α= ٠,٠٥) بين وجود DIF لصالح الإناث في مجال الجبر ولصالح الذكور في مجال الأعداد. أما بالنسبة لحجم الفقد فقد أظهرت النتائج عدم وجود فروق في نسب الأسئلة التي أظهرتDIF لصالح الذكور أو الإناث ترجع إلى نسب الفقد ٥% أو١٠%.

المقدمة

حظي موضوع قياس سماتالأفراد وخصائصهم العقلية بالبحث والاهتمام منذ القدم، ويُستخدم القياس في الأوضاع التربوية لِيخُدم أغراضًا متعددَة من التخطيط والرصد والتصنيف والتنبؤ، وتؤثر نتائج القياس في اتخاذ القرارات التربوية، وتجويد عمليتي التَعلّم والتعليم.وتُعدُّ دقة القياس من العناصر الأساسية في الاختبارات كونها تسمح بتفسيراتٍ صحيحةٍ لدرجاتِ الاختبارات (Kane, ١٩٩٦)، ويجب أن تدعم الأدلة التجريبية صدق الاستنتاجات حول درجات الاختبارات، ومن المشاكل التي يمكن أن تُهدد صدق نتائج الاختبارات تباين أداء المفحوصين من نفس مستوى القدرة على نفس فقرات الاختبار باختلاف المجموعة التي ينتمون إليها، وإحدى الإجراءات المستخدمة لتقييم صدق الاختبارات وعدالتها هي دراسة الأداء التفاضلي لفقراتها(Ankenmannet al., ١٩٩٩) .

يحدثُ الأداءُ التفاضليُ للفقرة (Differential Item Functioning, [DIF]) عندما لا يحصل المفحوصين الأعضاء في مجموعة فرعية على احتمال للإجابة على الفقرة بشكل صحيح مساوٍ لاحتمال الإجابة على نفس الفقرة لأعضاء آخرين في مجموعة أُخرى من نفس القدرة، ويُشيرُ الأداء التفاضلي أيضًا إلى أن البنية الداخلية لفقرات الاختبار (كالصعوبة والتمييز) ليست متساوية للمفحوصين في المجموعات المختلفة من نفس القدرة (Woods, ٢٠٠٨)، والاختلاف غير المتوقع في أداء المفحوصين من نفس مستوى القدرة على نفس الفقرة، من الممكن أن يُسبب استنتاجات خاطئة حول السمة لبعض المفحوصين، وبالتالي يُؤثر في صدق الاختبار، وقد يكون اكتشاف الأداء التفاضلي للفقرة معقدًا بسبب وجود مشكلة واسعة الانتشار في البحوث التربوية وهي مشكلة البيانات المفقودة، التي تؤثر في نتائج الكشف عن الأداء التفاضلي لفقرات الاختبارات (Bernhard et al., ١٩٩٨). إن وجود البيانات المفقودة والطرق المستخدمة للتعامل معها يمكن أن يكون لهما تأثير ملحوظ على قدرة التقنيات الشائعة الاستخدام لتحديد وجود الأداء التفاضلي للفقرة بدقة. ففي بعض الحالات، أشارت دراسات إلى تضخيم معدلات الخطأ من النوع الأول لهذه الطرق بحيث تم تحديد الفقرات عن طريق الخطأ على أنها تعرض أداء تفاضلي عندما لا ينبغي أن تكون كذلك، بينما في حالات أخرى، كانت القوة الإحصائية لاكتشاف الأداء التفاضلي بوجود بيانات مفقودة منخفضة (Robitzsch & Rupp, ٢٠٠٩). وبالنظر إلى هذه النتائج، يجد متخصصو القياس أن البيانات المفقودة قد تؤثر بشكل كبير على القدرة على تحديد الفقرات التي تبدي أداء تفاضلي بدقة. ونظرًا لأهمية اكتشاف الأداء التفاضلي الدقيق في تطوير تقييمات عادلة وصادقة، فإنه لا يمكن التغاضي عن تأثير البيانات المفقودة (Finch, ٢٠١١). ويُعرّف الأداء التفاضلي للفقرة (DIF) على أنه اختلاف احتمال استجابة المفحوصين ممن هم في نفس مستوى القدرة على الفقرة استجابة صحيحة نتيجة انتمائهم لمجموعات مختلفة (Hambleton & Rogers, ١٩٨٩).

تعتمد الطرق الإحصائية للكشف عن (DIF)على تحديد الطلاب بنفس مستوى القدرة، حيثُ أن تحديد مستويات قدرة الطلاب هو أحد أهداف درجات الاختبار، والنظريات المُهتمة بدرجات الاختبار هي نظرية الاستجابة للفقرة([Item Response Theory, [IRT) والنظرية الكلاسيكية للاختبار([Classical Test Theory, [CTT])، لذلك تستند بعض الطرق للكشف عن(DIF) إلى (CTT)مثل: طريقة مانتل-هانزل، وطريقة الصعوبة المحولة للفقرة وطريقة الانحدار اللوجستي، والطرق الأُخرى تستند إلى (IRT) مثل: الطرق القائمة على المقارنة بين المنحنيات المميزة للفقرة، والطرق القائمة على المقارنة بين معالم الفقرات في المجموعات (Hambleton & Swaminathan, ١٩٨٥).

لكل طرق الكشف عنDIF مزايا وعيوب متباينة في ظل الظروف المختلفة، وللكشف عن DIFبالطرق المناسبة دور مهم في إعداد اختبارات صادقة وغير متحيزة، وبالتالي اتخاذ قرارات دقيقة. لذا فإن تحديد الطريقة أو الطرق الأكثر دقة في ظل ظروف معينة يوفر الفرصة للحصول على قرارات دقيقة. كما أن الطرق نفسها تتأثر بالعديد من المتغيرات مثل نسبة عدد الفقرات التي تبدي DIF، طول الاختبار، مستوىDIF، حجم العينة، شكلDIF في الفقرات، والبيانات المفقودة (Missing data) والتي قد تسبب في إعطاء هذه الطرق نتائج خاطئة في حالة عدم القدرة على التعامل معها بالطرق الملائمة (Selvi & Alici, ٢٠١٨). وأشار بانكز (Banks, ٢٠١٥) إلى أن العديد من أساليب الكشف عن DIF، بما في ذلك مانتل هانزل، وطريقة الانحدار اللوجستي، وطريقة تحيز الفقرة المتزامن، ليست قادرة على التعامل مع البيانات المفقودة.

عرَّف ليتل وروبن (Little& Rubin, ٢٠١٩, p.٤) البيانات المفقودة "بالقيم غير الملحوظة، والتي قد تكون ذات معنى للتحليل إذا لوحظت، ومعنى ذلك أن القيمة المفقودة تُخفي قيمة ذات معنى". وفي سياق البحوث التربوية، تنشأ البيانات المفقودة عندما لا يجيب المفحوص عن فقرة ما، أو ينسى الإجابة عليها، أو لا يستطيع الوصول إليها بسبب نفاذ الوقت من الممتحن، وقد تنشأ بسبب تخطي الممتحن لها بقصد العودة إليها، أو يتخطاها لأنه لا يعرف الإجابة ولا يرغب في التخمين(Emenogu, ٢٠٠٦).

توجّه الباحثون والمهتمون بقضايا القياس النفسي والتربوي إلى استقصاءِ أثر العديد من الطرق الإحصائية لمعالجةِ البيانات المفقودة؛ لمساعدة الباحثين في معالجة مشكلة البيانات المفقودة في بحوثهم بما يكفل تحسين مُخرجات الأساليب الإحصائية، والوصول إلى استنتاجات صحيحة وموثوقة قابلة للتعميم حول الظاهرة المدروسة في مختلف المجالات.

ولتحديد الطريقة الأفضل للتعامل مع البيانات المفقودة لابد من فهم نمط وآلية فقد هذه البيانات، ومُعدل البيانات المفقودة، ولذلك يتم التحقق ما إذا كانت هذه البيانات المفقودة تتوزع بالكامل بشكل عشوائي، أو لها نمط معين، ونسبة الفقد في البيانات (Little & Rubin, ٢٠١٩).

مُعدل (حجم) البيانات المفقودة Rate of Missing Date

يُقصد بمُعدل البيانات المفقودة نسبة عدد القيم المفقودة إلى العدد الكلي للبيانات. يُشير جراهام (Graham, ٢٠٠٩) إلى أنه عندما تزداد نسبة البيانات المفقودة في البيانات عن٥% فإن تأثيرها على نتائج الدراسات يزداد، كما أشارت نتائج دراسات أخرى (Tabachnick & Fidell, ٢٠٠٧; Schafer, ١٩٩٩) إلى أن معدل ٥% أو أقل من البيانات المفقودة في عينة كبيرة سيكون غير مُهم. من ناحية أخرى، أشار بينت (Bennet, ٢٠٠١) إلى أن تقديرات المعلمات من خلال التحليل الإحصائي من المرجح أن تكون متحيزة عندما يكون أكثر من ١٠% من البيانات مفقودة، كما أشار بانكز (Banks, ٢٠١٥) إلى تضخم في معدلات الخطأ من النوع الأول عندما تكون نسبة البيانات المفقودة ١٠% وأكثر.

الأنماط المختلفة للبيانات المفقودة The Different Patterns of Missing Data

يُشير نمط البيانات المفقودة إلى ترتيب البيانات الملاحظة والمفقودة في مصفوفة البيانات، فهو يصف ببساطة موقع البيانات المفقودة ولا يوضح سبب فقد البيانات، ويُشير الأدب السابق إلى ستة أنماط لفقد البيانات وهي: ١) نمط وحيد المتغير، ٢) نمط وحدة عدم الاستجابة، ٣) النمط الوتيري، ٤) النمط العام للبيانات المفقودة، ٥) نمط البيانات المفقودة المخطط لها، ٦) نمط المتغير الكامن (Enders, ٢٠١٠).

آلية فقد البيانات Missing Date Mechanism

يشير مفهوم آلية فقد البيانات إلى الكيفية التي تمت من خلالها عملية فقد بعض القيم من بعض متغيرات الدراسة (Enders, ٢٠١٠). وقد وصف روبن (Rubin, ١٩٧٦) ثلاث آليات مُختلفة لفقد البيانات:

١-الفقد العشوائي بالكامل Missing Completely at Random (MCAR)

٢- الفقد العشوائي (Missing at Random (MAR

٣- الفقد غير العشوائي(Missing Not at Random (MNAR

طرق التعامل مع البيانات المفقودة Missing Data Handling Methods

وفقًا لِلتيل وروبين(Little& Rubin, ٢٠١٩) يمكن تصنيف معظم طرق معالجة البيانات المفقودة إلى أربع فئات عامة كما في جدول ١:

جدول ١ طرق معالجة البيانات المفقودة

طرق معالجة البيانات المفقودة	أشكال الطريقة	أنواع الطريقة
الإجراءات المعتمدة على البيانات الكاملة	حذف الحالة	تحليل الحالة الكاملة تحليل الحالة المتوفرة
	طريقة الخطأ أو التعويض بصفر طريقة غير الموجود	
إجراءات الترجيح (الوزن)	حذف البيانات التي تتضمن قيمًا مفقودة، وإعطاء أوزان للبيانات المستجابة سواءً أكانت خاطئة أم صحيحة	يتم إيجاد قيم لهذه الأوزان من خلال النماذج الرياضية، أو من خلال الاحتمالات
الإجراءات المستندة إلى التعويض	طرق النمذجة الصريحة	التعويض من خلال الوسط الحسابي حساب قيمة تعويضية من خلال الوسط المصحح للفقرة التعويض من خلال الانحدار التعويض من خلال الانحدار العشوائي
	طرق النمذجة الضمنية	التعويض باستخدام أقرب الأشخاص الإحلال التعويض بطريقة الطرق المُركبة
الإجراءات القائمة على نموذج		خوارزمية تعظيم التوقعات حساب قيم تعويضية متعددة التعويض بواسطة المعادلات المتسلسلة

مشكلة الدراسة وأسئلتها

تُعدُّ ظاهرة البيانات المفقودة من الظواهر الشائعة في البحوث النفسية والتربوية؛ فكثيرًا ما نجد دراسات تتعامل مع عينات في التصاميم البحثية، وخاصة الاختبارات إلاَّ وفيها نسبة بيانات مفقودة (Peugh & Enders, ٢٠٠٤)، وعلى الرغم من أن البيانات المفقودة لا تكون عادةً محور أي دراسة مُعطاة، إلا أنها تُمثل مشكلة منتشرة يُصادفها الباحثون بشكل مُتكرر في بحوثهم التجريبية (Schafer & Graham, ٢٠٠٢).

تُعد البيانات المفقودة مُعضلة؛ لأن الأساليب الإحصائية تفترض وجود بيانات كاملة لجميع الحالات ولجميع المتغيرات الموجودة في الدراسة والمُدرجة في التحليل، وينبغي على الباحثين قبل إجراء التحليلات الإحصائية تحديد كيفية معالجة البيانات المفقودة؛ لأن غالبية التقنيات الإحصائية ليست فاعلة وقوية مع البيانات المفقودة (Allison, ٢٠٠١)، وإذا تجاهل التعامل معها (أي ترك افتراضات البرامج مُستمرة)، تظهر مشاكل عديدة، فهي تُقلص من حجم العينة، وتُضعف القوة الإحصائية، كما تؤدي إلى تقدير خاطئ للخطأ المعياري، وزيادة معدلات الخطأ من النوع الأول، وتقديرات أقل كفاءة لمعالم الفقرات والأفراد، وتحيز نتائج الاختبارات والمقاييس المستخدمة (Bori, ٢٠١٣).

ومع زيادة الاهتمام بصدق نتائج الاختبارات وتعميمها، ولأهمية القرارات التي تُبنى عليها، وأيضًا زيادة الاهتمام بمعايير الجودة التي يتم تقييم البحوث العلمية على أساسها، زاد الاهتمام بقضية التعامل مع البيانات المفقودة، وطرق مُعالجتها، وخاصة مع التطور في البرامج الإحصائية الذي أدى إلى تنوع الأساليب الإحصائية التي يمكن من خلالها التعامل معالبيانات المفقودة، لمساعدة الباحثين في اختيار أفضل الطرق للتعامل مع المُشكلة؛ لإثراء نتائج التحليلات الإحصائية، والوصول إلى نتائج صادقة، قابلة للتعميم حول الظاهرة المدروسة، وما زالت الدراسات تتباين في نتائجها حول أفضل الطرق وأكثرها فاعلية للتعامل مع البيانات المفقودة (بني عواد والمومني، ٢٠١٨).

ولما أظهرت نتائج الدراسات التي تناولت أثر طرق التعامل مع البيانات المفقودة على دقة الكشف عن DIF لفقرات الاختبار، بأن لهذه الطرق أثرًا على فاعلية طرق الكشف عنDIF متمثلة بالقوة الإحصائية ومعدلات الخطأ من النوع الأول سواءً بالزيادة أو بالنقصان (Finch, ٢٠١١; Garrett, ٢٠٠٩). من هنا جاءت الدراسة الحالية لاستكمال البحث في هذا المجال من خلال تقصي فاعلية بعض طرق معالجة البيانات المفقودة وحجمها في الكشف عن DIF لفقرات اختبار في ضوء متغير الجنس. وتهدف هذه الدراسة لتقصي فاعلية كل من خوارزمية تعظيم التوقعات(Expectation-Maximization Algorithm (EM، والتعويض المتعدد(Multiple Imputation (MI، والتعويض المتعدد باستخدام المعادلات المتسلسلة Multiple Imputation by Chained Equations (MICE)، والتعويض باستخدام الوسط المُصحح للفقرة Corrected Item Mean Substitution (CIM) ، وطريقة اعتبار القيمة المفقودة كقيمة أسمية Missing Data As Nominal (MD)، في الكشف عن DIF للفقرات في ضوء متغير الجنس، باستخدام اختبار مانتل -هانزل المعمم (Generalized Mantel Haenszel test, [GMH])وفقًا للطرق الكلاسيكية CTT، باستخدام بيانات حقيقية، حيث تم الاستعانة ببيانات اختبار تيمس الدولي TIMISS, (٢٠١٥) في الرياضيات لطلبة الصف الثامن الأساسي الأردنيين، وذلك في ضوء تفوق الإناث على الذكور في الرياضيات في جميع الدورات (١٩٩٩, ٢٠٠٣, ٢٠٠٧, ٢٠١١, ٢٠١٥) (المركز الوطني لتنمية الموارد البشرية، ٢٠١٧)؛ مما شكّل ظاهرة تحتاج إلى التبصر والتمعن بالأسباب المسؤولة عن هذا التفوق، التي يمكن أن يكون من ضمنها عدم جدية المفحوصين في الإجابة عن فقرات الاختبار، وترك فقرات من دون إجابة، مما أدى إلى ظهور بيانات مفقودة في بيانات الاختبار، أثّرت بدقة وموضوعية نتائج الاختبار، لذا جاءت ضرورة هذه الدراسة بالاستعانة بهذا الاختبار في محاولة لمعرفة بعض الأسباب التي أثرت بدقة وموضوعية نتائج الاختبار، وبناءً على ما تقدم فإن هذه الدراسة تهدف إلى دراسة أثر طرق معالجة البيانات المفقودة وحجمها في الكشف عن DIF لفقرات اختبار تيمس (TIMSS, ٢٠١٥) وفقًا لاختبار GMH، وذلك من خلال الإجابة عن الأسئلة الآتية:

١: هل توجد فروق ذات دلالة إحصائية تُعزى إلى متغير الجنس عند (α= ٠,٠٥) بين عدد الأسئلة التي تبديDIF لأسئلة اختبار الرياضيات (TIMSS, ٢٠١٥)، باستخدام اختبار (GMH)، باختلاف طرق معالجة البيانات المفقودة (CIM، EM، MI، MICE، MD)؟

٢: هل توجد فروق ذات دلالة إحصائية بين مجالات بُعد المحتوى لأسئلة اختبار الرياضيات (TIMSS, ٢٠١٥)، التي تبدي DIFباستخدام اختبار(GMH)، عند (α= ٠,٠٥) باختلاف طرق معالجة البيانات المفقودة (CIM، EM، MI، MICE، MD)؟

٣: هل توجد فروق ذات دلالة إحصائية عند (α=٠,٠٥) بين نسب الأسئلة التي تبدي DIFلأسئلة اختبار الرياضيات (TIMSS, ٢٠١٥)، باختلاف حجم البيانات المفقودة (٥%، ١٠%)؟

أهمية الدراسة

تكتسب الدراسة الحالية أهميتها النظرية من خلال محاولتها تحديد أثر طرق التعامل مع البيانات المفقودة في الكشف عن DIFللفقرات، ومقارنة ما توصلت إليه نتائج الدراسة الحالية المبنية على بيانات حقيقية بنتائج الدراسات المبنية على بيانات مولدة من حيث أثرها في الكشف عن DIF، مما يسهل على الباحثين معرفة أثرها في الكشف عن DIF، وتتيح لهم المجال بالاستفادة من هذه الطرق، لضمان صدق نتائج الاختبارات وعدالتها وعدم تحيزها؛ نظرًا لأهمية القرارات المبنية على هذه النتائج المنبثقة عنها.

وتكمن أهمية هذه الدارسة أيضًا باختيار اختبار تيمس (TIMSS, ٢٠١٥)في الرياضيات للكشف عن أثر التعامل مع البيانات المفقودة في أداء الطلبة في هذا الاختبار، نظرًا لِأهمية هذا الاختبار ونتائجه للأردن، فوجود البيانات المفقودة قد يؤدي إلى اختلافات كبيرة في نتائج التحليلات الإحصائية التي تجري بالاعتماد على بيانات هذا الاختبار، ومن بينها الكشف عن DIFلفقرات الاختبار وفقًا لمتغير الجنس، وأيضًا وجود حجم البيانات الحقيقية الكبير الذي لا يمكن توفيره بتطبيق اختبارات مُعدة من قِبل الباحث.

كما يتوقع أن تُسهم الدراسة الحالية في دعم القاعدة النظرية للبحوث المُتعلقة بتحليل وتفسير النتائج لأثر أسلوب التعامل مع البيانات المفقودة بالكشف عن DIFالتي تُعد من القضايا المهمة، وما يترتب عليها من قرارات للوصول إلى نتائج أكثر دقة وكفاءة في التقويم التربوي.

أما من الناحية العملية فيتوقع أن تُسهم الدراسة في ما يلي:

أن تعود بالفائدة على الباحثين والمسؤولين عن الاختبارات وصانعي القرارات في وزارة التربية والتعليم ووزارة التعليم العالي والجامعات والمُهتمين بالاختبارات الوطنية والدولية بتصور أوضح ومعلومات جديدة حول كيفية وأثر طرق التعامل مع البيانات المفقودة، وتحسين الإحصائيات المتعلقة بأداء الطلبة ومستوياتهم، للوصول إلى نتائج أكثر دقة وكفاءة في التقييمات التربوية.

مصطلحات الدراسة وتعريفاتها الإجرائية

البيانات المفقودة: تمثلت البيانات المفقودة في هذه الدراسة بالفقرات المتروكة فارغة من دون إجابات، حيثُ تم التعامل مع الفقد على أنه نوع واحد بغض النظر عن موقع حدوث الفقد.

طرق معالجة البيانات المفقودة: هي طرق إحصائية يتم من خلالها إحلال البيانات المفقودة بقيم معينة في مصفوفة بيانات الاختبار للحصول على مصفوفة مكتملة من البيانات القابلة للتحليل بواسطة برامج إحصائية، وفي هذه الدراسة كانت طرق معالجة البيانات المفقودة من خلال (MD، وMICE، وEM، وMI، وCIM)، حيثُ:

⇐ **MD:** في هذه الطريقة تم استخدام البرنامج الإحصائي (SPSS, Version ٢٥)، بحيثُ تم أولًا تهيئة مصفوفة البيانات واستبدال الخلايا الفارغة بالعدد (٩) من الخيارTransform، ثم تعريف البرنامج بأن هذه القيمة مفقودة (Missing)، ويعاملها كقيمة أسمية Nominalولا يتم إدخالها في التحليلات الإحصائية.

⇐ **MICE:** طريقة حساب قيمة تعويضية متعددة، وتُعد من طرق التعويض متعددة المتغيرات التي تعتمد على أسلوب التوصيف الشرطي الكامل Fully Conditional Specification (FCS)، تستخدم لتعويض البيانات المفقودة باستخدام الحزمة الإحصائية (MICE, ٢,٩)وهي مُختصر لـ Chained Multivariate Imputation by Equations ضمن البرنامج الإحصائي (R, Version ٣,٦,٣).

⇐ **EM:** طريقة لحساب قيمة تعويضية بتعظيم التوقعات، تتضمن عمليات تكرارية لتقدير القيم المفقودة، تدور بين خطوة التوقع وخطوة التعظيم باستخدام البرنامج الإحصائي(SPSS, Version ٢٥).

⇐ **MI:** طريقة حساب قيمة تعويضية متعددة، يتم فيها استبدال كل قيمة مفقودة بالوسط الحسابي لمجموعة من القيم المختارة عشوائيًا بمقدار عدد محدد من المرات _وتم اعتماد ١٠ مرات للتعويض_ باستخدام البرنامج الإحصائي(SPSS, Version ٢٥).

⇐ **CIM:** طريقة التعويض باستخدام الوسط المُصحح للفقرة، تم استخدام البرنامج الإحصائي(SPSS, Version٢٥) وكتابة أكواد لحساب تلك الطريقة اشتقت من الكود الذي قام حبشي (٢٠٢٠) بكتابته، ويعتمد البرنامج على تصحيح متوسط الفقرة بمتوسط الفرد على بقية الفقرات.

حجم البيانات المفقودة: يُقصد بها نسبة عدد الاستجابات المفقودة على الفقرة الواحدة قياسًا إلى عدد الاستجابات الكلية، ويتم في هذه الدراسة التعبير عنها بالنسبة المئوية للاستجابات المفقودة، وتم اعتماد معيار حجم الفقد (٥%، ١٠%).

الأداء التفاضلي للفقرة (DIF): مفهوم إحصائي يستدل عليه من دالة مشتقة إحصائيًا للتعبير عن الفروق في احتمالات الاستجابة الصحيحة للفقرة بين مجموعتين ممن هم في نفس مستوى القدرة (Gruijter & Kamp, ٢٠٠٥). وتم في هذه الدراسة الكشف عن DIF باستخدام اختبار (GMH).

اختبار مانتل – هانزل المعمم (GMH): اختبار إحصائي لفحص DIF للفقرات الثنائية الاستجابة أو للفقرات متعددة الاستجابة استنادًا للنظرية الكلاسيكية للاختبار. وهو نوعان أحدهما للبيانات الاسمية متعددة الاستجابة، والآخر للبيانات الرتبية متعددة الاستجابة، وفي هذه الدراسة تم استخدام اختبار مانتل للبيانات الرتبية متعددة الاستجابة، وتم استخدام البرنامج الإحصائي (DIFAS, ٥,٠) للكشف عن DIF باستخدام هذا الإجراء.

اختبار تيمس الدولي (TIMSS) Trends in International Mathematics and Science Study: مجموعة اختبارات عالمية تعقد بصورة دورية منتظمة كل أربع سنوات على المستوى الدولي والعالمي، تقيس مستوى أداء الطلبة، وتقييم مدى تحصيلهم في مادتي الرياضيات والعلوم في الصفين الرابع والثامن (المركز الوطني لتنمية الموارد البشرية، ٢٠١٧)، وتم في هذه الدراسة اعتماد اختبار تيمس الدولي لعام ٢٠١٥م لمادة الرياضيات ولطلبة الصف الثامن الأردنيين.

مُحددات الدراسة

اقتصرت الدراسة الحالية على استخدام اختبار GMHللكشف عن DIFالمنتظم، ودراسة بُعد واحد من أبعاد DIF وهو الجنس، واقتصرت أداة الدراسة على الكراسات (الأولى، والسابعة، والتاسعة، والثالثة عشر) من اختبار الرياضيات في اختبار تيمس الدولي لعام ٢٠١٥م ولطلبة الصف الثامن الأساسي الأردنيين.

الدراسات السابقة

بالنظر إلى الدراسات والأبحاث التي اهتمت بطرق التعامل مع البيانات المفقودة، هناك العديد من الدراسات تناولت طرق التعامل مع البيانات المفقودة وأثرها على دقة تقدير معالم الفقرات وقدرات الأفراد (بني عواد، ٢٠١٠؛ الزعبي، ٢٠١٣)، ومنها ما درس أثرها على دقة تقدير معالم معادلة الانحدار البسيط (اللصاصمة، ٢٠١٦)، ومنها ما درس أثرها على دقة معادلة الاختبار(عاشور، ٢٠١٨)، وعدد قليل من الدراسات الأجنبية تناولت طرق التعامل مع البيانات المفقودة وأثرها في الكشف عن الأداء التفاضلي للفقرات، فقد هدفت دراسة إمنجو (Emenogu, ٢٠٠٦) إلى تقصّي أثر ثلاث طرق للتعامل مع البيانات المفقودة وهي: طريقة تحليل الحالة الكاملة، وتحليل الحالة المتوفرة، وطريقة اعتبار الإجابة المفقودة خاطئة، في الكشف عن DIF، باستخدام طريقة مانتل هانزل (Mantel Haenszel test [MH])، حيثُ تم استخدام بيانات حقيقية مستمدة من بيانات استجابات الطلبة على٤١ فقرة ثنائية الاستجابة من اختبار تيمس(TIMMS,١٩٩٥) في الرياضيات والعلوم، المقدم باللغة الفرنسية والإنجليزية، وأظهرت نتائج الدراسة اختلاف في عدد ونوع الفقرات التي تبدي أداءً تفاضليًا عند تغير طريقة التعامل مع البيانات المفقودة، وأظهرت ٨٠% من الفقرات نفس القرار فيما إذا كانت الفقرة تُظهر DIF أم لا، وحددت جميع معالجات البيانات المفقودة نفس الحجم والاتجاه لـ ٣٣% من فقرات DIF. ولفتت هذه الدراسة الانتباه إلى التناقض المحتمل الذي قد ينتج إذا تم التعامل مع البيانات المفقودة بشكل مختلف في تحليلات DIF- MH. ومن ناحية أخرى أجرت جاريت (Garrett, ٢٠٠٩) دراسة هدفت إلى تقصي أثر طريقة (التعويض المتعددMI، والتعويض من خلال الوسط الحسابي لاستجابات المفحوص([Within-Person Mean Substitution [PMS) على معدلات الخطأ من النوع الأول والقوة الإحصائية للكشف عن (DIF) للفقرات متعددة الاستجابة في ظل نموذج الاستجابة الجزئية(PCM)، باستخدام طريقتي(MH) وطريقة الانحدار اللوجستي الرتبي (OLR) للكشف عن (DIF) عندما تكون آلية فقد البيانات(MCAR) باستخدام بيانات مولدة لـ ٢٠ فقرة، وأظهرت نتائج الدراسة ارتفاع معدلات الخطأ من النوع الأول عبر طرق الكشف عن (DIF) لطريقة التعويض (PMS)، وزادت معدلات القوة الإحصائية لطريقتي التعويض بزيادة حجم الأداء التفاضلي، لكنها انخفضت بزيادة نسبة البيانات المفقودة. تم التوصية بطريقة MI؛ لسيطرتها بشكل أفضل على معدلات الخطأ من النوع الأول والقوة الإحصائية عبر طرق الكشف عن (DIF).

وكما أجرى روبتزش وروب (Robitzch& Rupp, ٢٠٠٩)دراسة باستخدام بيانات مولدة للتحقق من أثر خمس طرق للتعامل مع البيانات المفقودة: (الحذف، والتعويض بصفر، والتعويض الثنائي، والتعويض الثنائي المعدّل، والتعويض المتعدد باستخدام المعادلات المتسلسلة) على طريقتين للكشف عن DIFالمنتظم وهما الانحدار اللوجستي (Logistic Regression (LR و (MH)، تم إجراء تحليلات التحيز والجذر التربيعي لمتوسط الخطأ للتحقق من دقة المعلمات المقدرة لإجراء MH، وإجراء LR، من حيث كمية DIF التي تمت معالجته، وأظهرت نتائج الدراسة أن التفاعل بين آليات فقد البيانات، وطرق التعويض عن البيانات المفقودة، وحجم الأثر هي الأكثر تأثيرًا في تفسير التباين في الأداء التفاضلي للفقرات. كما أظهرت النتائج إلى أن الطريقة الخاطئة للتعويض عن البيانات المفقودة من الممكن أن تزيد معدلات الخطأ من النوع الأول والثاني. وسَعتْ دراسة فينش(Finch, ٢٠١١) البحث في أثر ثلاث طرق للتعويض عن البيانات المفقودة وهي: MI، وحذف الحالة Listwise Deletion (LD)، والتعويض بصفر Zero Imputation (ZI) ، على معدلات الخطأ من النوع الأول والقوة الإحصائية للكشف عن DIF المنتظم، باستخدام الطرق (LR، MH، وطريقة تحيز الفقرة المُتزامن (Simultaneausly Bias Test, SIBTEST باستخدام بيانات مولدة لـ ٤٠ فقرة ثنائية الاستجابة وفق النموذج اللوجستي ثلاثي المعلمة، أظهرت نتائج الدراسة نتائج مماثلة عبر طرق الكشف عن DIF للفقرات تحت تأثير ظروف الدراسة وتفاعلاتها، كما أظهرت طريقة ZI أنها الأقل قوة في الكشف عن DIFمقارنة بالطرق الأخرى، وتشير نتائج هذه الدراسة إلى أن القدرة على اكتشاف DIF بشكل صحيح عندما يكون موجودًا (أو الامتناع بشكل صحيح عن اكتشافه عندما لا يكون موجودًا) بوجود بيانات مفقودة متشابهة سواء استخدم الباحث MI أو LD. وقدّم بانكز (Banks, ٢٠١٥) دراسة عرّف بها الباحثين بموضوع البيانات المفقودة في سياق(DIF)، حيثُ تم استعراض تسع دراسات، حددت كل هذه الدراسات تأثير طرق المعالجة للبيانات المفقودة على نتائج الكشف عن DIF في ظل ظروف مختلفة. تضمن الاكتشاف الأكثر أهمية في هذه المراجعة بأن استخدام طريقة (ZI) يمكن أن يؤدي إلى تضخم الخطأ من النوع الأول، لذا يُنصح الباحثين بتجنب هذه الطريقة. وأظهرت نتائج الدراسات أن طريقة MH وطريقة LR كانت أكثر إجراءات DIF الموصى بها شيوعًا. وفي دراسة سيلفي و أليجي (Selvi& Alici, ٢٠١٨) هدفت إلى دراسة تأثير طرق معالجة البيانات المفقودة المختلفة بالكشف عن DIFبطريقة MH وطريقة التوحيد القياسي وطريقة اختبار نسبة الأرجحية. تم الحصول على بيانات ١٠٤٦ مرشحًا الذين أجابوا على اختبار مهارات التعلم الأساسية المكون من ٨٠ فقرة ثنائية الاستجابة، تراوحت نسبة البيانات المفقودة بين(٠,٣% - ١٠%)، وباستخدام طرق مختلفة لمعالجة البيانات المفقودة (التعويض من خلال الانحدار، وEM)، ومن ثم تنفيذ تحليلات DIF بالطرق المعتمدة. أظهرت النتائج اختلاف في عدد الفقرات التي أظهرت DIFباختلاف أساليب معالجة البيانات المفقودة، عبر طرق الكشف عن DIF. وأشارت الدراسة أن أساليب الكشف عن DIF المستندة إلى CTT أكثر اتساقًا داخلياً من الأساليب المستندة إلى IRT.من خلال استعراض الدراسات السابقة، وجدت الباحثتان أن معظم الدراسات أهتمت بالبحث عن مشكلة البيانات المفقودة، وما زالت الدراسات البحثية حول هذا الموضوع مستمرة، ويوجد في الأدبيات العديد من الدراسات المتعلقة بنقاط القوة والضعف لطرق التعامل مع البيانات المفقودة، ومع ذلك لم تجد الباحثتان_ في حدود اطلاعهما_ دراسات تناولت مشكلة البيانات المفقودة في سياق DIFولبيانات حقيقية، وقد لوحظ أن معظم الدراسات الأجنبية قد أُجريت على مجموعة البيانات التي تم الحصول عليها باستخدام طريقة المحاكاة، ولوحظ أن هذه الدراسات تقريبًا قد أُجريت على طرق كشف DIF المُستخدمة بشكل مُتكرر مثل: طريقة الانحدار اللوجستي الخطي، وطريقة تحيز الفقرة المتزامن، وطريقة التوحيد القياسي، وطريقة مانتل هانزل للفقرات الثنائية الاستجابة، وطريقة الأرجحية القصوى. وعلى هذا الأساس جاءت مشكلة الدراسة تبحث في أثر طرق مختلفة لمعالجة البيانات المفقودة وحجوم مختلفة للبيانات المفقودة في الكشف عن DIF لفقرات اختبار في ضوء CTT، وباستخدام بيانات حقيقية، وتحديد مدى التوافق بين الطرق المختلفة لمعالجة البيانات المفقودة. وتم اعتماد اختبار GMH للكشف عن DIFلأنه بسيط من الناحية المفاهيمية، وسهل نسبيًا للتطبيق، ويقدم اختبارًا للدلالة الإحصائية ويوفر تقديرًا لحجم الأثر المستخدم مع نسبة الارجحية. ومع ذلك، فهي ليست فعالة لـ DIF غير المنتظم(Osterlind& Everson, ٢٠٠٩) . أيضًا، ليست هناك حاجة لأحجام كبيرة من العينات للإجراء الإحصائي GMH، وتتميز بقوة إحصائية عالية وتحكم جيد بالخطأ من النوع الأول في عينات من ١٠٠شخص لكل مجموعة. على الأرجح، هذه هي الأسباب التي تجعل GMH يحتل المرتبة الأولى في الكشف عن الفقرات ذات الأداء التفاضلي (Ibrahim, ٢٠١٣).

الطريقة والإجراءات

تم تطبيق هذه الدراسة على بيانات الاختبار الدولي تيمس (TIMSS, ٢٠١٥) لمبحث الرياضيات لطلبة الصف الثامن الأردنيين، حيثُ يُعقد هذا الاختبار بصورةٍ دوريةٍ كل أربع سنوات منذُ عام ١٩٩٥م، ويهدفُ إلى دراسةِ فعاليةِ المناهج المطبقة، وطرق تدريسها، والتطبيق العملي لها، وتقييم التحصيل، وتوفير المعلومات لتحسين تعليم الرياضيات والعلوم وتعلمهما.

مجتمع الدراسة

تكون مجتمع الدراسة من استجابات جميع طلبة الصف الثامن الأردنيين المشاركين في الاختبار الدولي تيمس الذي تم إجراءه في عام ٢٠١٥م، وبلغَ عدد الطلبة الأردنيين في الصف الثامن الذين شاركوا في الدراسة ٨٦١٧طالبًا وطالبةً، موزعين على٢٥٢ مدرسة.

أداة الدراسة

تم تطبيق هذه الدراسة بالاعتماد على بيانات (TIMSS, ٢٠١٥)لمادة الرياضيات لطلبة الصف الثامن الأردنيين، حيثُ يغطي اختبار تيمس للرياضيات في دورة ٢٠١٥م بُعدين اثنين هما: بُعد المحتوى ويتضمن أربعة مجالات وهي: الأعداد Numbers (N)، والجبر Algebra (A)، والهندسة Geometry(G) ، والبيانات والاحتمالات Data & Chance(DC) ، وبُعد المجالات المعرفية ويتضمن ثلاثة مجالات وهي: المعرفة Knowing(KN) ، والتطبيق Applying(AP) ، والتفكير Reasoning(RS). وبلغ عدد الأسئلة ٢١٢ سؤالًا، توزعت على مجالات المحتوى الأربعة، وكان ٥١% من أسئلة الاختبار من نوع الاختيار من مُتعدد (Multiple Choice, MC)، أما الأسئلة الأخرى جاءت من النوع المقالي (Constructed Response, CR).

عينة الدراسة

تألفت عينة الدراسة التي طُبّقت عليها المعالجات الإحصائية من استجابات ٢٢٦١طالبًا وطالبةً من طلبة الصف الثامن، موزعة على أربع كراسات اختبارية كما يوضحها جدول ٢.

جدول ٢ توزيع عينة الدراسة حسب (رقم الكراسة الاختبارية، وعدد الطلبة المستجيبين لكل كراسة، والجنس، وعدد الأسئلة في كل كراسة اختبارية)

رقم الكراسة الاختبارية	عدد المستجيبين	عدد الإناث	عدد الذكور	عدد الأسئلة
١	٥٦٨	٣٠٥	٢٦٣	٣١
٧	٥٥٤	٣٠٤	٢٥٠	٣٢
٩	٥٦٦	٣١٢	٢٥٤	٢٨
١٣	٥٧٣	٣١٥	٢٥٨	٢٩
المجموع	٢٢٦١	١٢٣٦	١٠٢٥	١٢٠

وقد تم اعتماد هذه الكراسات الأربعة في الدراسة الحالية؛ لتوافر فيها نسب فقد مختلفة تتراوح بين ٤,٢٢% و١٠,٤٢%على مستوى الكراسة الاختبارية، حيثُ كان متوسط نسبة الفقد في الكراسة الأولى: ١٠,٤٢%، والسابعة٨,٠٢%، والتاسعة٥,٩٤%، والثالثة عشر ٤,٢٢%).

الإجراءات والمعالجات الإحصائية للدراسة

١-تحديد حجم البيانات المفقودة في استجابات الطلبة عن أسئلة الاختبار في الكراسات المعتمدة باستخدام الحزمة الإحصائية nanair ضمن البرنامج الإحصائي R. version ٣,٦,٣، وتراوحت نسبة الفقد للكراسة الأولى بين (١,٤% - ٤٦,٣%)، والسابعة بين (١,١% - ٢٥,٦%)، والتاسعة بين (١,٢% - ١٧,٥%)، والثالثة عشر بين (١% - ١٢,٤%).

٢- تحديد نمط البيانات المفقودة باستخدام البرنامج الإحصائي (SPSS)، حيث تم تحديدها للكراسات الأربعة المعتمدة في الدراسة، وكان نمط البيانات المفقودة للكراسات الأربعة هو النمط العام للبيانات المفقودة (النمط الاعتباطي).

٣- تحديد آلية البيانات المفقودة فيما إذا كانت آلية الفقد العشوائية بالكامل أم لا، وللتحقق من ذلك استخدمت الباحثتان اختبار ليتل Little's Test للفقد العشوائي بالكامل باستخدام البرنامج الإحصائي (SPSS)، وأشارت النتائج إلى أن الفقد ليس عشوائيًا بالكامل MCARللكراستين الأولى والتاسعة، وأن هناك احتمالًا أن يكون الفقد عشوائيًا MAR أو فقد غير عشوائي MNAR، فقد كانت قيمة مربع كاي دالة عند (=α٠,٠٠١)، وكما أشارت النتائج أن آلية الفقد في الكراستين السابعة والثالثة عشر هي فقد عشوائي بالكامل MCARفقد كانت قيمة اختبار ليتل غير دالة إحصائيًا عند (α= ٠,٠٠١).

٤- استخدام طرق معالجة البيانات المفقودة المعتمدة في الدراسة، كلٍ على حدة، للحصول على مصفوفة بيانات كاملة، لتشغيل البرنامج الإحصائي (DIFAS, ٥,٠) للكشف عن DIF، وتم اعتماد الطرق الآتية لمعالجة البيانات المفقودة (MD ،MICE ،MI ،EM ،CIM).

٥- إجراء اختبار GMHباستخدام البرنامج الإحصائي (DIFAS, ٥,٠) على جميع ملفات الكراسات الأربعة المعتمدة في الدراسة، للكشف عن الفقرات التي تبديDIF بين الذكور والإناث.

نتائج الدراسة ومناقشتها

النتائج المتعلقة بالسؤال الأول ونصه: "هل توجد فروق ذات دلالة إحصائية تُعزى إلى متغير الجنس عند (α= ٠,٠٥) بين عدد الأسئلة التي تبدي DIF لأسئلة اختبار الرياضيات (TIMSS, ٢٠١٥)، باستخدام اختبار (GMH)، باختلاف طرق معالجة البيانات المفقودة (MD ،MICE ،MI ،EM ،CIM)"؟

تم إيجاد عدد الأسئلة التي تبدي DIFباستخدام اختبار ((GMHباختلاف طرق معالجة البيانات المفقودة للكراسات الأربعة المعتمدة في الدراسة، ويعرض جدول ٣ أعداد ونسب الأسئلة التي أظهرت DIF عبر مختلف طرق المعالجة.

جدول ٣أعداد ونسب الأسئلة التي أظهرت أداءً تفاضليًا عبر مختلف طرق معالجة البيانات المفقودة

طرق معالجة البيانات المفقودة	لصالح الذكور		لصالح الإناث		لا يوجد أداء تفاضلي	
	العدد	النسبة	العدد	النسبة	العدد	النسبة
الوسط المصحح للفقرة CIM	١٧	%١٤,٢	١٦	%١٣,٣	٨٧	%٧٢,٥
تعظيم التوقعات EM	١٦	%١٣,٣	١٩	%١٥,٨	٨٥	%٧٠,٨
من دون تعويض MD	١٢	%١٠	١٢	%١٠	٩٦	%٨٠
التعويض المتعددMI	١٧	%١٤,٢	٢٢	%١٨,٣	٨١	%٦٧,٥
المعادلات المتسلسلةMICE	١٦	%١٣,٣	١٧	%١٤,٢	٨٧	%٧٢,٥

يُبين جدول ٣أن هناك تقاربًا في عدد الأسئلة التي أظهرت DIF عبر مختلف الكراسات، فقد كانت أعلى عدد أسئلة لصالح الإناث في ٢٢ سؤال عند استخدام طريقة MI، بينما كان أقل عدد أسئلة لصالح الإناث والذكور ١٢سؤال في حالة MD. كما يتضح من جدول ٣توافق تقريبًا بين الطريقتين MICE وCIM في عدد الأسئلة التي أظهرت DIF، كما تطابقت عدد الأسئلة التي أظهرت عدم وجود DIF عند استخدام كلتا الطريقتين. تم حساب الفروق بين نسب الأداء التفاضلي للأسئلة باستخدام اختبار كوكران Cochran Q Testباستخدام الحزمة الإحصائية RVAideMemoire v ٠,٩-٧٨، حيث أن الطرق المختلفة لمعالجة البيانات المفقودة استخدمت مع نفس البيانات،

وأظهرت النتائج وجود فروق ذات دلالة إحصائية عند (α= ٠,٠١)، فقد كانت القيمة الاحتمالية للخطأ تساوي٠,٠٠١٣٧١٨٤، ولهذا فقد تم إجراء مقارنات ثنائية باستخدام اختبار مكنيمار McNemar Chi-squared test لعينتين مرتبطتين باستخدام الحزمة الإحصائيةRcompanion v٢,٣,٧، وقد أظهر جدول ٤ النتائج، مع عمل تعديل لمستوى القيمة الاحتمالية للخطأ_ وقد تم اعتماد تعديل بونفيروني Bonferroni_ لتجنب زيادة المخاطرة في الوقوع في خطأ من النوع الأول.

جدول ٤ المقارنات الثنائية بين مختلف طرق معالجة البيانات المفقودة

الطريقة الأولى	الطريقة الثانية	القيمة الاحتمالية للخطأ	القيمة الاحتمالية للخطأ المعدلة
CIM	EM	٠,٦١٧٠٠	١,٠٠٠٠
	MD	٠,٠٦٦٥٠	٠,٦٦٥٠
	MI	٠,١٤٩٠٠	١,٠٠٠٠
	MICE	١,٠٠٠٠٠	١,٠٠٠٠
EM	MD	٠,٠٣٧١٠	٠,٣٧١٠
	MI	٠,٣٤٣٠٠	١,٠٠٠٠
	MICE	٠,٦٨٣٠٠	١,٠٠٠٠
MD	MI	٠,٠٠٥١١	٠,٠٥١١
	MICE	٠,٠٦٦٥٠	٠,٦٦٥٠
MI	MICE	٠,١٤٩٠٠	١,٠٠٠٠

يتضح من جدول ٤ وجود فروق في نسب الأداء التفاضلي للأسئلة بين طريقة EM وMD عند (α= ٠,٠٥)، إلا إن الفرق لم يظهر بعد تعديل مستوى الدلالة لتجنب الوقوع في الخطأ من النوع الأول، كما يتضح وجود فروق في نسب الأداء التفاضلي للأسئلة بين طريقة MI وMD، إلا إن الفرق لم يظهر بعد تعديل مستوى الدلالة، وعلى الجانب الأخر فقد كانت جميع المقارناتالثانية الأخرى غير دالة إحصائيًا قبل أو بعد تعديل مستوى الدلالة. ويمكن تفسير عدم وجود فروق ذات دلالة إحصائية بين عدد الأسئلة التي تبدي أداءً تفاضليًا باختلاف طرق معالجة البيانات المفقودة إلى توافق نتائج طرق المعالجة، وللتحقق من ذلك تم إيجاد توزيع تكراري لعدد مرات الاتفاق بين مختلف طرق معالجة البيانات المفقودة حول الأداء التفاضلي لأسئلة جميع الكراسات، ويوضح جدول٥ توزيع تكراري لعدد مرات الاتفاق بين مختلف طرق معالجة البيانات المفقودة حول الأداء التفاضلي لأسئلة جميع الكراسات.

جدول ٥ توزيع تكراري لعدد مرات الاتفاق بين مختلف طرق معالجة البيانات المفقودة حول الأداء التفاضلي لأسئلة جميع الكراسات

		عدد مرات الاتفاق					
الأداء التفاضلي		٠	١	٢	٣	٤	٥
لصالح الذكور	التكرار	١٠٠	٢	١	٢	٦	٩
	النسبة المئوية%	٨٣,٣	١,٦٧	٠,٨٣	١,٦٧	٥	٧,٥
لصالح الإناث	التكرار	٩٣	٩	٢	١	٦	٩
	النسبة المئوية%	٧٧,٥	٧,٥	١,٦٧	٠,٨٣	٥	٧,٥
لا يوجد أداء تفاضلي	التكرار	٢٤	١٢	٣	٣	٨	٧٠
	النسبة المئوية%	٢٠	١٠	٢,٥	٢,٥	٦,٦٧	٥٨,٣٣

يتضح من جدول ٥ توافق الطرق الخمس في معالجة البيانات المفقودة عبر جميع الكراسات على إبداءDIF في ٩ أسئلة لصالح الذكور و٩ أسئلة لصالح الإناث و٧٠ سؤال لم تُبدDIF، كما تم تحديد توزيع تكراري لعدد مرات الاتفاق بين طرق التعويض الأربعة، ويوضح جدول ٦ توزيع تكراري لعدد مرات الاتفاق بين طرق التعويض الأربعة عن البيانات المفقودة حول الأداء التفاضلي لأسئلة جميع الكراسات.

جدول ٦ توزيع تكراري لعدد مرات الاتفاق بين مختلف طرق التعويض عن البيانات المفقودة حول الأداء التفاضلي لأسئلة جميع الكراسات

الأداء التفاضلي		عدد مرات الاتفاق				
		٠	١	٢	٣	٤
لصالح الذكور	التكرار	١٠١	٢	١	١	١٥
	النسبة المئوية%	٨٤,١٧	١,٦٧	٠,٨٣	٠,٨٣	١٢,٥
لصالح الإناث	التكرار	٩٦	٦	٢	١	١٥
	النسبة المئوية%	٨٠	٥	١,٦٧	٠,٨٣	١٢,٥
لا يوجد أداء تفاضلي	التكرار	٣٠	٢	٣	٨	٧٧
	النسبة المئوية%	٢٥	١,٦٧	٢,٥	٦,٦٧	٦٤,١٧

يتضح من جدول ٦ أن هناك اتفاقًا بين طرق التعويض الأربعة على وجود ١٥ سؤال تبدي DIF لصالح الذكور في أسئلة الكراسات الأربعة مجتمعة وعددها ١٢٠ سؤالًا، ووجود ١٥ سؤالًا أخرى تبدي DIF لصالح الإناث، بينما هناك ٧٧ سؤالًا تتفق طرق التعويض الأربعة على أنها لا تؤدي بصورة تفاضلية لأي من الجنسين. وبمقارنة جدول ٥ مع جدول ٦ يلاحظ أن طريقة MD قللت عدد مرات الاتفاق بين طرق المعالجة بشكل ملحوظ. ويوضح جدول ٧ تكرار ونسبة الاتفاق بين كل زوج من طرق معالجة البيانات المفقودة لجميع الكراسات الاختبارية.

جدول ٧ تكرار ونسبة الاتفاق بين كل زوج من طرق معالجة البيانات المفقودة لجميع الكراسات الاختبارية

	MD	خوارزميات التعويض			
		CIM	EM	MI	MICE
MD		*(٨٤,٢)١٠١	(٨٠,٨)٩٧	(٧٩,٢)٩٥	(٨٤,٢)١٠١
CIM			(٩٦,٧)١١٦	(٩٠)١٠٨	(١٠٠)١٢٠
EM				(٩١,٧)١١٠	(٩٦,٧)١١٦
MI					(٩٠)١٠٨
MICE					

( )* النسبة المئوية % للاتفاق بين كل زوج من طرق معالجة البيانات المفقودة لجميع الكراسات الاختبارية

يتضح من جدول ٧ أنه عبر جميع الكراسات الاختبارية حققت طريقة MICE اتفاق تام مع طريقة CIM بنسبة ١٠٠%، بينما حققت طريقة MI أعلى نسبة اتفاق مع EM بنسبة ٩١,٧%، وحققت طريقة MD أقل نسبة اتفاق مع MI في ٩٥ سؤال من أصل ١٢٠ سؤال بنسبة ٧٩,٢%. ويرجع سبب توافق الطريقتين MICE وCIM إلى اعتماد كلتا الطريقتين على صعوبة الفقرة وقدرة الفرد بنفس الوقت في تعويض البيانات المفقودة، بينما حققت طريقة المعالجة(MD) والتعامل مع القيمة المفقودة كقيمة أسمية أقل نسبة اتفاق مع جميع الطرق الأخرى؛ نظرًا لعدم إدخال الخلايا التي فيها قيم مفقودة في التحليلات الإحصائية وعددها ٤٨٩٩ من أصل ٦٨٩٢١ بنسبة ٧,١١%؛ مما أعطى عدد أسئلة تبدئ DIF أقل من طرق التعويض الأربعة؛ لأنه في تحليلاتGMH للكشف عن DIF يتم الاعتماد على الدرجة الكلية للفرد كمعيار للمطابقة، وعندما تم تجاهل الخلايا ذات القيم المفقودة وعدم إدخالها في التحليلات الإحصائية في اختبار GMH أثر ذلك في الكشف عن DIF وأعطى نسبة أسئلة تبدئ DIF أقل من طرق التعويض الأربعة .

النتائج المتعلقة بالسؤال الثاني ونصه:"هل توجد فروق ذات دلالة إحصائية بين مجالات بُعد المحتوى لأسئلة اختبار الرياضيات (TIMSS, ٢٠١٥)، التي تبدي DIFباستخدام اختبارGMHعند (α= ٠,٠٥) باختلاف طرق معالجة البيانات المفقودة (MD ،MICE ،MI ،EM ،CIM)"؟

تم اختبار الفروق في أعداد الأسئلة التي أظهرت DIFعبر مجالات المحتوى الأربعة داخل كل طريقة من طرق المعالجة، ويُبين جدول ٨الأعداد والنسب المئوية للأسئلة وفقًا لبُعد المحتوى والأداء التفاضلي باختلاف طرق معالجة البيانات المفقودة

جدول ٨ الأعداد والنسب المئوية للأسئلة وفقًا لبُعد المحتوى والأداء التفاضلي باختلاف طرق معالجة البيانات المفقودة

طرق معالجة البيانات المفقودة	بعد المحتوى	لصالح الإناث		لصالح الذكور		لا يوجد أداء تفاضلي	
		العدد	النسبة	العدد	النسبة	العدد	النسبة
MD	A الجبر	٦	%١٧,٦	٠	%٠	٢٨	%٨٢,٤
	DCالبيانات والاحتمالات	٠	%٠	١	%٤,٢	٢٣	%٩٥,٨
	Gالهندسة	٢	%٨	٢	%٨	٢١	%٨٤
	Nالأعداد	٤	%١٠,٨	٩	%٢٤,٣	٢٤	%٦٤,٩
CIM	A الجبر	٩	%٢٦,٥	٠	%٠	٢٥	%٧٣,٥
	DCالبيانات والاحتمالات	٣	%١٢,٥	١	%٤,٢	٢٠	%٨٣,٣
	Gالهندسة	٢	%٨	٢	%٨	٢١	%٨٤
	Nالأعداد	٢	%٥,٤	١٤	%٣٧,٨	٢١	%٥٦,٨
EM	A الجبر	١١	%٣٢,٤	٠	%٠	٢٣	%٦٧,٦
	DCالبيانات والاحتمالات	٤	%١٦,٧	١	%٤,٢	١٩	%٧٩,٢
	Gالهندسة	٢	%٨	٢	%٨	٢١	%٨٤
	Nالأعداد	٢	%٥,٤	١٣	%٣٥,١	٢٢	%٥٩,٥
MI	A الجبر	١٤	%٤١,٢	٠	%٠	٢٠	%٥٨,٨
	DCالبيانات والاحتمالات	٢	%٨,٣	١	%٤,٢	٢١	%٨٧,٥
	Gالهندسة	٤	%١٦	٣	%١٢	١٨	%٧٢
	Nالأعداد	٢	%٥,٤	١٣	%٣٥,١	٢٢	%٥٩,٥
MICE	A الجبر	١١	%٣٢,٤	٠	%٠	٢٣	%٦٧,٦
	DCالبيانات والاحتمالات	٢	%٨,٣	١	%٤,٢	٢١	%٨٧,٥
	Gالهندسة	٢	%٨	٢	%٨	٢١	%٨٤
	Nالأعداد	٢	%٥,٤	١٣	%٣٥,١	٢٢	%٥٩,٥

يتضح من جدول ٨ اتساق القرارات حول إبداء DIFلصالح الإناث في مجال الجبر، ولصالح الذكور في مجال الأعداد. ولاختبار دلالة الفروق بين أعداد الأسئلة التي أظهرت DIF، تم استخدام اختبار فيشر Fisher Exact Testباستخدام الحزمة الإحصائية stats v٣,٦,٢، وقد أظهرت النتائج وجود فروق ذات دلالة إحصائية عند(α= ٠,٠١)، فقد كانت قيمة p-value تساوي ٠,٠٠٤. ولتحديد مصدر تلك الفروق تم إجراء مقارنات ثنائية باستخدام اختبار مكنيمار، وجاءت النتائج كما يوضحها جدول ٩.

جدول ٩ المقارنات البعدية للأداء التفاضلي لبُعد المحتوى باختلاف طرق معالجة البيانات المفقودة باستخدام اختبار مكنيمار

		DIF	
الطريقة المحتوى	الذكور: الإناث	الذكور: لا يوجد DIF	الإناث: لا يوجد DIF
الجبر: البيانات والاحتمالات	٠,٣٢١	٠,٧٣٣	٠,٣٢١
: الهندسة	٠,٣٢١	٠,٣٩٧	٠,٧٣٣
MD : الأعداد	٠,٠٩٨	٠,٠٤٧	١
البيانات والاحتمالات: الهندسة	١	٠,٧٨٢	٠,٧٣٣
: الأعداد	١	٠,٢٠٢	٠,٣٢١
الأعداد: الهندسة	٠,٧٨٢	٠,٣٢١	٠,٨١٣
الجبر: البيانات والاحتمالات	٠,٤٨٧	٠,٦٣٢	٠,٤٨٧
: الهندسة	٠,٢٣١	٠,٤٠٤	٠,٣٤
CIM : الأعداد	٠،٠٠٠١	٠,٠٠١	٠,٣٤
البيانات والاحتمالات: الهندسة	١	١	١
: الأعداد	٠,١١٥	٠,٠٢٧	١
الأعداد: الهندسة	٠,٣٤	٠,٠٦٧	١
الجبر: البيانات والاحتمالات	٠,٥١١	٠,٦٢٩	٠,٤٢٩
: الهندسة	٠,١٢٩	٠,٦٢٩	٠,١٢٩
EM : الأعداد	٠,٠٠٠١	٠,٠٠٧	٠,١٢٩
البيانات والاحتمالات: الهندسة	٠,٦٢٩	١	٠,٧٤٨
: الأعداد	٠,٠٦٣	٠,٠٥٩	٠,٦٢٤
الأعداد: الهندسة	٠,٣٥٧	٠,١٠٧	١
الجبر: البيانات والاحتمالات	٠,٢٨٩	١	٠,٠٢٩
: الهندسة	٠,٠٧٩	٠,٣٤٨	٠,١٧٤
MI : الأعداد	٠,٠٠٠١	٠,٠١٧	٠,٠٢٩
البيانات والاحتمالات: الهندسة	١	٠,٤٧٨	٠,٤٩٧
: الأعداد	٠,١٩٤	٠,٠٢٩	١
الأعداد: الهندسة	٠,١٣٨	٠,١٧٤	٠,٤٩٧
الجبر: البيانات والاحتمالات	٠,٣٥١	٠,٦٧٧	٠,١٢٩
: الهندسة	٠,١٢٩	٠,٦٧٧	٠,١٢٩
MICE : الأعداد	٠,٠٠٠١	٠,٠٠٧	٠,١٢٩
البيانات والاحتمالات: الهندسة	١	١	١
: الأعداد	٠,٢١٦	٠,٠٣١	١
الأعداد: الهندسة	٠,٣٢١	٠,١٢٩	١

يظهر جدول ٩أن الطرق الخمسة لمعالجة البيانات المفقودة اتفقت على وجود فروق دالة إحصائيًا عند مستوى الدلالة (α=٠,٠٥) بين وجود DIF لصالح الإناث أو لصالح الذكور في مجال الجبر والأعداد، كما اتفقت طرق التعويض الأربعة بوجود فروق في مجالي الجبر والأعداد والأداء التفاضلي للذكور وعدم وجود DIF، واتفقت ثلاث طرق تعويض (MICE, MI, CIM) على وجود فروق دالة إحصائيًا بين مجال البيانات والاحتمالات مقابل الأعداد والأداء التفاضلي لصالح الذكور وعدم وجود DIF، وبالرجوع إلى جدول ٨ نجد اتفاق بوجود DIF لصالح الإناث في مجال الجبر، ولصالح الذكور في مجال الأعداد، وهذا يتفق مع نتائج المركز الوطني لتنمية الموارد البشرية لبيانات تيمس لعام ٢٠١٥ م، وكذلك مع دراسة عنابي ودودين(Innabi& Dodeen, ٢٠١٨)، و دراسة عبد العزيز وآخرون(Abedalaziz et al., ٢٠١٤).

النتائج المتعلقة بالسؤال الثالث ونصه:" هل توجد فروق ذات دلالة إحصائية عند (α=٠,٠٥) بين عدد الأسئلة التي تبدي DIF لأسئلة اختبار الرياضيات(TIMSS, ٢٠١٥)، باختلاف حجم البيانات المفقودة (٥%، ١٠%)"؟

تم إيجاد عدد الأسئلة ونسبها المئوية وفقًا لحجم الفقد (٥%: منخفض: حجم البيانات المفقودة أقل من ٥%، مرتفع: ٥% فأكثر) والأداء التفاضلي عدة مرات باختلاف طرق معالجة البيانات المفقودة(MD، CIM، EM، MI، MICE)، وللتأكد من العلاقة بين نسب الفقد والأداء التفاضلي للأسئلة تم استخدام اختبار فيشر، ويوضح جدول ١٠ ذلك.

جدول١٠ أعداد الأسئلة ونسبها وفقًا لحجم الفقد ٥% والأداء التفاضلي ودلالة اختبار فيشر باختلاف طرق معالجة البيانات المفقودة

طرق معالجة البيانات المفقودة	عدد الأسئلة ونسبها	لصالح الإناث		لصالح الذكور		القيمة الاحتمالية للخطأ
		حجم الفقد٥%		حجم الفقد٥%		
		منخفض	مرتفع	منخفض	مرتفع	
MD	عدد الأسئلة	٧	٥	٤	٨	٠,٤٩٧
	النسبة المئوية %	١٢,٥	٧,٨	٧,١	١٢,٥	
CIM	عدد الأسئلة	٧	٩	٨	٩	٠,٩٦٨
	النسبة المئوية %	١٢,٥	١٤,١	١٤,٣	١٤,١	
EM	عدد الأسئلة	٨	١١	٨	٨	٠,٨٨٠
	النسبة المئوية %	١٤,٣	١٧,٢	١٤,٣	١٢,٥	
MI	عدد الأسئلة	١١	١١	٧	١٠	٠,٨٩٠
	النسبة المئوية %	١٩,٦	١٧,٢	١٢,٥	١٥,٦	
MICE	عدد الأسئلة	٧	١٠	٧	٩	٠,٨٧٦
	النسبة المئوية %	١٢,٥	١٥,٦	١٢,٥	١٤,١	

وقد أظهرت نتائج جدول ١٠ عدم وجود فروق بين أعداد الأسئلة التي أظهرت DIF لصالح الذكور أو الإناث ترجع إلى نسب الفقد أقل من ٥% مقابل ٥% فأكثر، فقد كانت قيمة اختبار فيشر غير دالة (أكبر من ٠,٠٥) باختلاف طرق معالجة البيانات المفقودة. وقد جاءت هذه النتائج متوافقة مع دراسة تاباتشنيك وفيديل(Tabachnick & Fidell, ٢٠٠٧) ، ودراسة شيفر (Schafer, ١٩٩٩) والتي تشير إلى أن معدل ٥%أو أقل من البيانات المفقودة في عينة كبيرة سيكون غير مُهم.

كما تم إيجاد عدد الأسئلة ونسبها المئوية وفقًا لحجم الفقد (**١٠%:** منخفض: حجم البيانات المفقودة أقل من ١٠%، مرتفع: ١٠%فأكثر) والأداء التفاضلي للأسئلة عدة مرات باختلاف طرق معالجة البيانات المفقودة(MD، CIM، EM، MI، MICE)، وللتأكد من الفروق بين نسب الفقد والأداء التفاضلي للأسئلة تم استخدام اختبار فيشر ويوضح جدول١١ ذلك.

جدول ١١ أعداد الأسئلة ونسبها المئوية وفقًا لحجم الفقد١٠% والأداء التفاضلي ودلالة اختبار فيشر باختلاف طرق معالجة البيانات المفقودة

طرق معالجة البيانات المفقودة	عدد الأسئلة ونسبها	لصالح الإناث		لصالح الذكور		القيمة الاحتمالية للخطأ
		حجم الفقد١٠%		حجم الفقد١٠%		
		منخفض	مرتفع	منخفض	مرتفع	
MD	عدد الأسئلة	٩	٣	٩	٣	٠,٥٠٧
	النسبة المئوية %	٩,٢	١٣,٦	٩,٢	١٣,٦	
CIM	عدد الأسئلة	١٠	٦	١٣	٤	٠,٠٦٢٨
	النسبة المئوية %	١٠,٢	٢٧,٣	١٣,٣	١٨,٢	
EM	عدد الأسئلة	١٢	٧	١٢	٤	*٠,٠٣٣٤
	النسبة المئوية %	١٢,٢	٣١,٨	١٢,٢	١٨,٢	
MI	عدد الأسئلة	١٦	٦	١٤	٣	٠,٥٢٠
	النسبة المئوية %	١٦,٣	٢٧,٣	١٤,٣	١٣,٦	
MICE	عدد الأسئلة	١٢	٥	١٢	٤	٠,٢١٨٦
	النسبة المئوية %	١٢,٢	٢٢,٧	١٢,٢	١٨,٢	

* دال إحصائياً عند (α= ٠,٠٥)

تم حساب اختبار فيشر وأظهرت النتائج وجود فروق بين نسب الفقد والأداء التفاضلي للأسئلة وفق طريقة EM، فقد كانت القيمة الاحتمالية للخطأ ٠,٠٣٣٤ وهي أصغر من ٠,٠٥، ولتحديد مصدر تلك الفروق تم إجراء مقارنات ثنائية بين الأداء التفاضلي للأسئلة ونسب الفقد، وقد أشارت النتائج إلى عدم وجود فروق ذات دلالة إحصائية عند (α=٠,٠٥)للأداء التفاضلي للأسئلة لصالح الذكور أو الإناث ترجع إلى نسب الفقد أقل من ١٠% مقابل ١٠% فأكثر. فقد كانت جميع مستويات الدلالة أكبر من ٠,٠٥. مما يدل على عدم وجود تأثير لنسب الفقد على الأداء التفاضلي للأسئلة لصالح الذكور أو الإناث. لذلك تشير النتائج إلى عدم وجود فروق في أعداد الأسئلة التي أظهرت DIF لصالح الذكور أو الإناث ترجع إلى نسب الفقد أقل من ١٠% مقابل ١٠% وأكثر.

التوصيات

في ضوء ما توصلت إليه الدراسة من نتائج توصي الباحثتان فيما يلي:

١. الاهتمام بعملية تقدير البيانات المفقودة والكشف عنها بمجموعة البيانات الإحصائية المختلفة لأهميتها وأثرها على النتائج في البحوث والدراسات المستقبلية دون بتر هذه القيم.

٢. إجراء دراسات مستقبلية أخرى عن أثر أسلوب معالجة البيانات المفقودة في اختبارات أخرى، وفي متغيرات أخرى، وعلى مقاييس الاتجاهات، واستخدام معالجات إحصائية مختلفة.

٣. على الباحثين تخصيص جزء من بحوثهم ودراساتهم عن كيفية التعامل مع البيانات المفقودة في مجريات بحثهم.

٤. دعوة الباحثين والمؤسسات التربوية والتربويين المتخصصين والمسؤولين عن الاختبارات والمهتمين بها ومطوريها وصانعي القرارات والجهات المسؤولة إلى الاهتمام أكثر بالاختبارات الوطنية والدولية واختبارات المستوى والقبول في الجامعات ومزاولة المهن والاهتمام بطرائق الكشف عن البيانات المفقودة فيها وأسلوب معالجتها.

٥. تحفيز الطلبة والمعلمين وإدارات المدارس على الاهتمام بالأداء في الاختبارات الوطنية والدولية من خلال اعتبار نتائجها جزءً له اعتباره في نظام تقييم أداء الطلبة؛ مما يقلل من احتمال وجود البيانات المفقودة، فالوقاية أهم من العلاج.

المراجع

- بني عواد، علي. (٢٠١٠). مقارنة طرق التعامل مع البيانات المفقودة في تقدير معالم الفقرات وقدرات الأفراد.
- رسالة دكتوراه غير منشورة. جامعة اليرموك، اربد، الأردن.
- بني عواد، علي والمومني، خالد. (٢٠١٨). إدراك أعضاء هيئة التدريس في جامعة الملك فيصل لأهمية طرق التعامل مع البيانات المفقودة في الاستبانات البحثية واختبارات الطلبة. مجلة جامعة النجاح للأبحاث (العلوم الإنسانية) ، ٣٢(٦)، ١- ٣٢.
- الحبشي، محمد. (٢٠٢٠، ٦ تشرين أول). التعويض عن القيم المفقودة في الاستبيانات باستخدام متوسط العبارة المصحح. [فيديو]. يوتيوب. https://youtu.be/c٢GioPM٢G٤٠
- الزعبي، عمر. (٢٠١٣). أثر نسبة البيانات المفقودة وطريقة التعويض عنها في دقة تقدير معالم الفقرات والأفراد.
- رسالة دكتوراه غير منشورة. جامعة اليرموك، اربد، الأردن.
- عاشور، أحمد. (٢٠١٨). أثر نسبة البيانات المفقودة في الجذع المشترك وطريقة التعامل معها على دقة المعادلة العمودية.
- رسالة دكتوراه غير منشورة جامعة اليرموك اربد الأردن.
- اللصاصمة، عمران. (٢٠١٦). أثر نسبة البيانات المفقودة وطريقة معالجتها في دقة تقدير معالم معادلة الانحدار البسيط.
- رسالة ماجستير غير منشورة .جامعة مؤته، الكرك، الأردن.
- المركز الوطني لتنمية الموارد البشرية. (٢٠١٧). التقرير الوطني الأردني عن الدراسة الدولية للرياضيات والعلوم ٢٠١٥ (*TIMSS*,٢٠١٥). عمان: وحدة المتابعة والتقييم.
- Abedalaziz, N., Leng, C., & Alahmadi, A. (2014). Detecting a gender-related differential item functioning using transformed item difficulty. *Malaysian Online Journal of Educational Sciences,* 2*(*1), 16–22.
- Allison, P. (2001). Missing data. Thousand Oaks Sage Publications
- Ankenmann, R., Witt, E., & Dunbar, S. (1999). An Investigation of the power of the likelihood ratio goodness of fit statistic in detecting differential item functioning. *Journal of Educational Measurement,* 36(4), 277-300.
- Banks, K. (2015). An introduction to missing data in the context of differential item functioning. *Practical Assessment, Research, and Evaluation*, 20(12), 1-10.
- Bennett, D. (2001). How can I deal with missing data in my study? *Australian and New Zealand Journal of Public Health,* 25, 464–469.
- Bernhard, J., Cella, D., Coates, A., Fallowfield, L., Ganz, P., Moinpour, C., & Hürny, C. (1998). Missing quality of life data in cancer clinical trials: serious problems and challenges. Statistics in Medicine, 17(5-7), 517-532.
- Bori, S. (2013, May). *Dealing with missing data: key assumptions and methods for applied analysis* (Technical Report No. 4). Boston University, 1-19.
- Emenogu, B. (2006). *The effect of missing data treatment on Mantel*-Haenszel DIF detection Enders C 2010). *Applied missing data analysis.* Guilford Press
- Finch, H. (2011). The use of multiple imputation for missing data in uniform DIF analysis: power and type I error rates. *Applied Measurement in* Education, 24(4), 281-301.
- Garrett, P. (2009). *A monte carlo study investigating missing data, differential item functioning, and effect size* Unpublished doctoral dissertation.Georgia State University, Atlanta. https://scholarworks.gsu.edu/eps_diss.
- Graham, J. (2009). Missing data analysis: Making it work in the real world. Annual Review of Psychology, 60, 549-576.
- Gruijter, D. & Kamp, L. (2005). *Statistical test theory for education and psychology (*1st. *ed.)* .Chapman & Hall/CRC Statistics in the Social and Behavioral Sciences.
- Hambleton, R., & Rogers, H. (1989). Detecting potentially biased test items: Comparison of IRT area and Mantel- Haenszel methods. *Applied Measurement in Education,*2(4), 313–334.

- Innabi, H., & Dodeen, H. (2018). Gender differences in mathematics achievement in Jordan: A differential item functioning analysis of the2015 TIMSS. School Science and Mathematics Association, 118(3), 127-137.
- Kane, M. (1996). The precision of measurements. *Applied Measurement in Education,* 9(4), 355-379.
- Little, R., & Rubin, D. (2019). *Statistical analysis with missing data* (3rd ed.). John Wiley& Sons.
- Osterlind, S. & Everson, H. (2009) .*Differential item functioning*. Sage Publications, Inc.
- Peugh, J., & Enders, C. (2004). Missing data in educational research: A review of reporting practices and suggestions for improvement. Review of Educational Research, 74(4), 525-556.
- Robitzsch, A., & Rupp, A. (2009). Impact of missing data on the detection of differential item functioning: The case of Mantel-Haenszel and logistic regression analysis. Educational and Psychological Measurement, 69(1), 18-34.
- Rubin, D. (1976). Inference and missing data. *Biometrika,* 63, 581–592.
- Schafer, J. (1999). Multiple imputation: A primer. Statistical Methods in Medical Research, 8(1), 3–15.
- Schafer, J., & Graham, J. (2002). Missing data: Our view of the state of the art. American Psychological Association, 7(2), 147-177.
- Selvi, H., &Alıcı, D. (2018). Investigating the impact of missing data handling methods on the detection of differential item functioning. International Journal of Assessment Tools in Education, 5(1), 1-14.
- Tabachnick, B., & Fidell, L. (2007). *Using multivariate statistics (*5th*ed.).* Allyn& Bacon Pearson Education
- Woods, C. (2008). Likelihood-ratio DIF testing: Effects of nonnormality. Applied Psychological Measurement, 32(7), 511-526.