Dirasat: *Educational Sciences*, Volume 50, No. 2, Supplement 1, 2023

Deanship of Scientific Research

The Impact of Score Distribution Shape on National Test Results: Evaluating the Accuracy of Estimating Item Information Function and Test Reliability for Students' Scores

Amjad Salem Mahmoud Kreshan [1] ****, Jihad Mohamed Mahmoud Al-Anati*** [2]

[1] Ministry of Education, Amman, Jordan.

[2] Department of Educational Psychology, College of Educational Sciences, University of Jordan.

Received: 9/4/2022
Revised: 24/5/2022
Accepted: 14/6/2022
Published: 15/7/2023

* *Corresponding author:*
amjad.ag@hotmail.com

Citation: Kreshan, A. S. M., & Al-Anati, J. M. M. (2023). The Impact of Score Distribution Shape on National Test Results: Evaluating the Accuracy of Estimating Item Information Function and Test Reliability for Students' Scores. *Dirasat: Educational Sciences*, *50*(2 -S1), 102–116.
https://doi.org/10.35516/edu.v50i2 -S1.998

© 2023 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

Objectives: This study aimed to examine how the distribution of students' scores on national tests affects the accuracy of estimating the item information function and test reliability, utilizing the principles of Item Response Theory.

Methods: The study employed a descriptive comparative approach. A random sample of 12,000 male and female students from the eighth grade, who had taken the Arabic, English, Mathematics, and Science papers of The National Exam for Controlling the Quality of Education during the 2018/2019 academic year, was selected. Based on the skewness of their score distributions (positive, normal, negative), the students were divided into 12 groups, each containing 1,000 male and female students. The statistical program BILOG_MG3 was used to analyze the data, applying the Three-Parameter Model of Item Response Theory to extract the item information function and its standard error. Additionally, the theoretical and experimental reliability coefficients were determined based on the distribution of the data.

Results: The findings revealed that scores with positive skewness had the highest mean standard error of the estimate for the item information function in national tests, and both the theoretical and empirical reliability of these tests were lower. Statistically significant differences were observed in the mean errors between data with a normal distribution and data with positive or negative skewness in the subjects of Arabic language, English language, and mathematics, while no significant differences were found in the Science subject.

Conclusions: Based on the study's results, it is recommended to consider the test information function as a criterion when designing national tests.

Keywords: Item response theory, item information function, test reliability, distribution shape of data.

أثر شكل التوزيع لدرجات الطلبة على الاختبارات الوطنية في دقة تقدير دالة معلومات الفقرة وثبات الاختبار

أمجد سالم محمود كريشان [1]****، جهاد محمد محمود العناتي*** [2]

[1] وزارة التربية والتعليم، عمان، الأردن.

[2] قسم علم النفس التربوي، كلية العلوم التربوية، الجامعة الأردنية.

ملخّص

الأهداف: هدفت الدراسة إلى التحقق من أثر شكل التوزيع لدرجات الطلبة على الاختبارات الوطنية، في دقة تقدير دالة معلومات الفقرة، وثبات الاختبار في ضوء نظرية الاستجابة للفقرة.

المنهجية: اتّبعت الدراسة المنهج الوصفي المقارن، ولتحقيق هدف الدراسة تم اختيار عينة عشوائية مكونة من (12000) طالب وطالبة من مجتمع الدراسة الذي شمل جميع طلبة الصف الثامن الأساسي، الذين تقدموا للصورة الورقية للاختبار الوطني لضبط نوعية التعليم، في مباحث اللغة العربية، واللغة الإنجليزية، والرياضيات، والعلوم للعام الدراسي (2018/ 2019)، توزعت العينة على (12) مجموعة باختلاف شكل الالتواء (موجب، معتدل، سالب) بواقع (1000) طالب وطالبة لكل مجموعة. تم استخدام البرنامج الإحصائي (BILOG_MG3)لتحليل البيانات في ضوء نظرية الاستجابة للفقرة حسب النموذج ثلاثي المعلمة، حيث تم استخراج دالة معلومات الفقرة والخطأ المعياري لها باختلاف شكل التوزيع للبيانات، كما تم استخراج معاملات الثبات النظري والتجريبي.

النتائج: أظهرت النتائج أن قيم متوسط الخطأ المعياري للتقدير لدالة معلومات فقرات الاختبارات الوطنية كانت الأعلى في البيانات ذات الالتواء الموجب، و كانت قيم الثبات النظري والتجريبي للاختبارات الوطنية منخفضة في البيانات ذات الالتواء الموجب كما كانت الفروق بين متوسطات الخطأ للبيانات ذات التوزيع المعتدل مع البيانات موجبة وسالبة الالتواء دالة إحصائيًا في مواد اللغة العربية، واللغة الانجليزية، والرياضيات، في حين كانت الفروق غير دالة في مادة العلوم.

الخلاصة: توصي الدراسة بالاعتماد على دالة معلومات الاختبار كمعيار عند بناء الاختبارات الوطنية.

الكلمات الدالة :: نظرية الاستجابة للفقرة، دالة معلومات الفقرة، ثبات الاختبار، شكل التوزيع للبيانات.

المقدمة

تعد الاختبارات التحصيلية جزءًا مهمًا من العملية التعليمية، فمن خلالها يتم الحكم على مدى تحقق الأهداف التربوية، والكشف عن جوانب القوة والضعف في أداء الطلبة تمهيدًا لوضع الحلول المناسبة والخطط العلاجية لتحسين تعلم الطلبة، ومن هنا شرعت إدارة الامتحانات والاختبارات بوزارة التربية والتعليم في الاردن منذ العام (2000) بإعداد الاختبار الوطني لضبط نوعية التعليم، بهدف تزويد متخذي القرار بمعلومات عن جودة التعليم؛ لتساعدهم على إتخاذ القرارات التربوية الملائمة لتطوير العملية التدريسية.

وتستخدم الدرجات على الاختبارات لاتخاذ قرارات مهمة، لذلك لا بد من التحقق من جودتها، وثباتها، وخلوها من الأخطاء العشوائية ما أمكن، حتى تكون الاستنتاجات التي نشتقها منها صحيحة، فعلى سبيل المثال تستخدم الدرجات لقبول الطلاب في المدرسة أو رفضهم؛ لتسجيل الطلاب في مقرر أو اجتيازهم أو رسوبهم في الامتحان. وفي هذه التطبيقات تكون المخاطر كبيرة بالنسبة للأفراد والمؤسسات المعنية، وبالتالي يجب أن تفي الاختبارات بمعايير الجودة لضمان اتخاذ قرارات صحيحة. (Oosterwijk, van & Sijtsma, 2019).

وعلى الرغم من أن تطبيقات نظرية الاختبار الكلاسيكية (Classical Test Theory: CTT) قد تم استخدامها لبناء الاختبارات التحصيلية وتطويرها، إلا أن الإفادة منها كانت محدودة، وذلك لاعتماد خصائص الفقرات على عينة المفحوصين الذين تمت اشتقاق الخصائص السيكومترية للاختبار من خلال نتائجهم، مما يؤدي إلى تغير في خصائص الفقرات من عينة إلى أخرى، لذلك اتجه خبراء القياس إلى نماذج نظرية الاستجابة للفقرة (Item Response Theory: IRT)، لمواجهة جوانب القصور في نظرية الاختبار الكلاسيكية (CTT) (علام، 2005)، (Embretson & Reise, 2000).

وتتميز نظرية الاستجابة للفقرة بوصف فقرات الاختبار بطريقة مستقلة عن المفحوصين الذين اختُبروا بها، ووصف المفحوص بطريقة مستقلة عن أي عينة من الفقرات التي أختُبر بها، وهو ما عبر عنه لورد (Lord, 1980) بخاصية اللاتغاير (Invariance)، ومن بين الفوائد العديدة لنظرية الاستجابة للفقرة (IRT) هي دالة المعلومات للفقرة والمتمثلة في كمية المعلومات المقدمة حول كل فقرة عبر مناطق مختلفة على متصل القدرة (θ). وتلعب دالة المعلومات في نظرية الاستجابة للفقرة دورًا مهمًا في وصف الفقرات وتوجيه اختيار الفقرات ومقارنة الكفاءة بين الاختبارات المختلفة (Hambleton, Swaminathan &Rogers , 1991). وبالإضافة إلى ذلك، يمكن استخدام دالة معلومات الفقرة لايجاد كمية المعلومات التي يوفرها الاختبار، بحيث تساهم جميع فقرات الاختبار بشكل مستقل في الشكل العام لدالة معلومات الاختبار مع تقدير الأخطاء في كل مستوى من مستويات القدرة (Doran, 2005).

وتلعب دالة معلومات الاختبار دورًا هامًا في نظرية الاستجابة للفقرة (IRT) لتقدير معامل الثبات، فيعد الحصول على مقياس لثبات درجات الاختبار في سياق نظرية الاستجابة للفقرة مفيدًا؛ لأنه يوفر مؤشرًا على الاتساق العام لدرجات الاختبار في قياس السمة الكامنة. ونظرًا لأن الهدف من الاختبار في نظرية الاستجابة للفقرة هو قياس السمة الكامنة لكل شخص يقوم بإجراء الاختبار، فمن المهم أخذ مقياس شامل لموثوقية (ثبات) درجات الاختبار عبر طيف التوزيع الكامن في الاعتبار (Yuan & Bentler, 2002).

وتستند نظرية الاستجابة للفقرة إلى افتراضات قوية لا بد من تحققها، وأن أي انتهاك لهذه الافتراضات يؤثر بشكل سلبي على دقة نتائج تقديرات المعالم للأفراد والفقرات، كما أنها تعتمد على أحجام عينات كبيرة لتقدير معالم القدرة والفقرة، وبالتالي مع زيادة حجم العينة يقترب شكل التوزيع للبيانات من التوزيع الطبيعي، ولكن هل تتأثر دقة هذه التقديرات مع تغير شكل التوزيع للبيانات؛ من هنا جاءت هذه الدراسة للتحقق من أثر الابتعاد عن التوزيع الاعتدالي لبيانات الاختبارات الوطنية في دقة تقدير دالة معلومات الفقرة وثبات الاختبار.

مشكلة الدراسة وأسئلتها:

نظرًا لأهمية الموضوعية والدقة في نتائج الاختبارات التي تستخدم للوقوف على مستوى تحصيل الطلبة، ولأهمية القرارات التي تُبنى عليها، فيجب أن يتم بناء هذه الاختبارات بصورة جيدة حتى نضمن الوصول إلى نتائج صادقة وثابتة. وترتبط دقة القياس بكمية المعلومات التي تساعم بها الفقرة في الاختبار، فالقيمة العظمى لدالة المعلومات تقابل أقل قيمة للخطأ المعياري، وبالتالي تؤدي إلى مستوى ثبات عالٍ للاختبار.

ويرتبط تقدير دالة المعلومات الفقرة، وثبات الاختبار، بكل من معالم الفقرة والقدرة، وتشير بعض الدراسات إلى وجود أثر لشكل توزيع القدرة على معالم الفقرة كما في دراسة علوان وجاسم (Alwan& Jasim, 2022)، ودراسة ساس وشميت ووكر (Sass, Schmitt & Walker, 2008) التي أظهرت إلى أن مستوى الأخطاء المعيارية لمعالم الفقرى في التوزيعات الملتوية أكبر مقارنة بالأخطاء المعيارية المحسوبة في حالة التوزيع الطبيعي للقدرة. كما اشارت دراسة نغواني (Ngwane, 2021) إلى أن مستوى الخطأ المعياري لتقدير معالم الفقرة كان منخفضًا في حالة التوزيع الطبيعي للقدرة، وبما أن دالة المعلومات للفقرة يتأثر تقديرها بكل من معالم الفقرة والقدرة، كان لا بد من دراسة دقة تقديرها في ضوء شكل التوزيع للبيانات، لذلك جاءت هذه الدراسة للتحقق من مدى تأثر دالة المعلومات للفقرة ومعامل الثبات باختلاف شكل التوزيع للدرجات على الاختبار، ولذلك ستحاول هذه الدراسة الإجابة عن السؤالين الآتيين:

1.هل تختلف دقة تقديرات معلومات الفقرات باختلاف شكل التوزيع للبيانات عند مستوى الدلالة(0.05)؟

2.هل تختلف تقديرات ثبات الاختبار وفق الطرق القائمة على نظرية الاستجابة للفقرة باختلاف شكل التوزيع للبيانات عند مستوى الدلالة (0.05)؟

أهمية الدراسة وهدفها:

تُعد دقة تقدير معالم الفقرة ودالة معلومات الفقرة من القضايا المهمة لنظرية الاستجابة للفقرة، كما يُعتبر معامل ثبات الاختبار من الخصائص السيكومترية المهمة التي يجب التحقق منها عند إعداد الاختبار؛ فارتفاع ثبات الاختبار يدل على انخفاض أخطاء القياس، وبالتالي تقترب الدرجة المقدرة من الدرجة الحقيقة، مما يجعلنا أكثر دقة في اتخاذ القرارت، ومن خلال الاطلاع على الدراسات التي اهتمت بدراسة تأثير المتغيرات على دالة المعلومات، وثبات الاختبار في ضوء نظرية الاستجابة للفقرة، نجد أنها تناولت العديد من المتغيرات مثل: حجم العينة، وطول الاختبار، وعدد المموهات، في حين نجد قلة من الدراسات التي تناولت متغير شكل التوزيع للدرجات، ومن هنا تبرز أهمية هذه الدراسة للتوصل إلى أحد العوامل الذي قد يؤثر على دقة تقدير دالة المعلومات للفقرة وثبات الاختبار، وبناء على ذلك هدفت هذه الدراسة إلى الكشف عن أثر شكل التوزيع لدرجات الاختبار في دقة تقدير دالة معلومات الفقرة وثبات الاختبار.

محددات الدراسة:

1. ستقتصر أداة الدراسة على نتائج الاختبارات الوطنية.
2. ستقتصر هذه الدراسة على فقرات ثنائية التدريج (0، 1)، ذات أربعة بدائل.
3. ستقتصر طرق الثبات المستندة إلى نظرية الاستجابة للفقرة على النموذج ثلاثي المعلمة.

متغيرات الدراسة:

المتغير المستقل: **شكل التوزيع للبيانات، وله ثلاثة مستويات: (التواء موجب، معتدل، التواء سالب).**

المتغير التابع: **الخطأ المعياري لدالة معلومات الفقرة، معامل الثبات التجريبي، معامل الثبات النظري.**

الإطار النظري والدراسات السابقة:

تعتمد نظرية الاستجابة للفقرة على المفردات التي يتكون منها الاختبار، وفي ظل هذه النظرية فإن كل فقرة من فقرات الاختبار تقيس القدرة أو السمة الكامنة الأساسية، ونتيجة لذلك فإن كمية المعلومات التي تعتمد على الفقرة يمكن حسابها عند أي مستوى من مستويات القدرة (θ)، والتي تسمى دالة معلومات الفقرة ويرمز لها بالرمز $I_i(\theta)$. وتقدم كل فقرة من فقرات الاختبار معلومات عن قدرة المفحوصين، لكن كمية هذه المعلومات تعتمد على مدى تطابق صعوبة الفقرة مع قدرة المفحوص، حيث تقل كمية المعلومات كلما ابتعد مستوى القدرة عن صعوبة الفقرة(Baker, 2001).

وتساعد دالة المعلومات للفقرة في بناء الاختبارات بحيث تجعل دقة القياس أكبر ما يمكن، ومن أهم ميزات دالة معلومات الفقرة، أنه يمكن معرفة درجة إسهام كل فقرة من الفقرات في كمية المعلومات الكلية للاختبار، وذلك لأن إسهام كل فقرة يكون مستقلًا عن إسهامات الفقرات الأخرى، وكلما كانت دالة المعلومات للفقرة أعلى، كلما كان ذلك مؤشرًا على جودة الفقرة، وبالتالي يتم تقدير القدرة للمفحوص بدقة أكبر (علام، 2005). وتمثل المعادلة (1) الصيغة الرياضية لدالة المعلومات للفقرة (Crocker & Algina, 1968).

$$I_i(\theta) = \frac{[P_i^{\backslash}(\theta)]^2}{[P_i(\theta)][Q_i(\theta)]} \dots\dots\dots\dots\dots\dots\dots\dots (1)$$

حيث أن:

$I_i(\theta)$: دالة المعلومات للفقرة (i) عند مستوى القدرة (θ)

$P_i(\theta)$: احتمال أن يجيب المفحوص ذو القدرة (θ)عن الفقرة $(\boldsymbol{i})$ إجابة صحيحة.

$Q_i(\theta)$: احتمال أن يجيب المفحوص ذو القدرة (θ)عن الفقرة $(\boldsymbol{i})$ إجابة خاطئة.

$P_i^{\backslash}(\theta)$: المشتقة الأولى لاحتمال الإجابة الصحيحة عن الفقرة $(\boldsymbol{i})$.

وتتأثر دالة المعلومات بمعالم الفقرة (الصعوبة، التمييز، التخمين)، حيث يتم حساب دالة المعلومات للفقرة في النموذج ثلاثي المعلمة حسب المعادلة (2) (Hambleton & Swaminathan, 1985).

$$I_i(\theta) = \frac{D^2 a_i^2 [Q_i(\theta)][P_i(\theta) - C_i]^2}{[P_i(\theta)][1 - C_i]^2} \dots\dots\dots\dots\dots\dots\dots\dots (2)$$

حيث أن:

$I_i(\theta)$: دالة المعلومات للفقرة $(\boldsymbol{i})$ عند مستوى القدرة $(\boldsymbol{\theta})$

$P_i(\theta)$: احتمال أن يجيب المفحوص ذو القدرة (θ)عن الفقرة $(\boldsymbol{i})$ إجابة صحيحة.

$Q_i(\theta)$: احتمال أن يجيب المفحوص ذو القدرة (θ)عن الفقرة $(\boldsymbol{i})$ إجابة خاطئة.

D : ثابت التدريج ويساوي 1.7

a_i : معلمة التمييز للفقرة (i).

c_i : معلمة التخمين للفقرة (i).

نلاحظ من المعادلة السابقة أن دالة المعلومات للفقرة تزداد كلما اقتربت معلمة التخمين من الصفر، وكذلك تتناسب دالة المعلومات طرديًا مع مربع التمييز للفقرة.

ويرتبط مفهوم الثبات في نظرية الاستجابة للفقرة، بدالة معلومات الاختبار والتي تمثل مجموع دوال المعلومات للفقرات المكونة للاختبار عند مستوى القدرة (θ)، والذي يعطى حسب المعادلة (3) (Hambleton & Swaminathan, 1985).

$$I(\theta) = \sum_{i=1}^{n} \frac{[P_i^{\backslash}(\theta)]^2}{[P_i(\theta)][Q_i(\theta)]} \dots\dots\dots\dots\dots\dots\dots\dots (3)$$

حيث أن:

$I(\theta)$: دالة المعلومات للاختبار عند مستوى القدرة $(\boldsymbol{\theta})$

ويرتبط مقدار المعلومات التي يزودنا بها الاختبار عند مستوى القدرة (θ)، عكسيًا مع مستوى الخطأ المعياري في التقدير، وهذا يعني أن الخطأ المعياري في التقدير يكون أقل ما يمكن عند مستويات القدرة التي تناظر أقصى معلومات (Hambleton & Swaminathan, 1985). والذي يعطى حسب المعادلة (4)

$$SE(\theta) = \frac{1}{\sqrt{I(\theta)}} \dots\dots\dots\dots\dots\dots\dots\dots (4)$$

حيث أن:

$SE(\theta)$: الخطأ المعياري في التقدير.

$I(\theta)$: دالة المعلومات للاختبار عند مستوى القدرة $(\boldsymbol{\theta})$

ويرتبط الخطأ المعياري في تقدير القدرة للمفحوص عكسيًا مع ثبات الاختبار، فالاختبار ذو الثبات المرتفع يقدم كمية معلومات كبيرة بأقل قدر من الخطأ المعياري للتقدير (Embretson & Reise, 2000).

وعلى الرغم من أن مفهوم معامل الثبات لم يتم تحديده بشكل جيد في إطار نظرية الاستجابة للفقرة، إلا أن العديد من الدراسات قدمت طرقًا لحساب الثبات، بالاعتماد على دالة معلومات الاختبار وتوزيع قدرة الأفراد، وفي الدراسة الحالية تم استخراج معاملات الثبات وفق نظرية الاستجابة للفقرة وفق النموذج ثلاثي المعلمة من خلال المرحلة الثالثة (phase3) من مراحل التحليل حسب البرنامج الإحصائي (BILOG-MG) من خلال الطريقتين التاليتين:

1. الثبات الإمبريقي (التجريبي) (Empirical Reliability):

يعتمد تقدير الثبات التجريبي على تقدير درجات القدرة للمفحوصين، من خلال قسمة تباين الدرجات الحقيقية (True Score Variance)على تباين الدرجة الملاحظة (Score Variance)، الذي يتم تقديره من خلال تقدير درجات القدرة للمفحوصين، ويتم تقدير تباين الدرجات الحقيقية من خلال طرح تباين الخطأ (Error Variance) من تباين الدرجة الملاحظة(Zimowski, Muraki, Mislevy & Bock, 2003).

2. الثبات النظري (Theoretical Reliability):

تعتمد قيمة الثبات النظري على معالم الفقرات التي يتم تقديرها، والصيغة الرياضية لحساب معامل الثبات النظري تعطى حسب المعادلة (5) (الصرايرة، 2016):

$$\text{Theoretical Reliability} = \frac{1}{1+\sigma^2} \dots\dots\dots\dots\dots\dots\dots\dots (5)$$

حيث تشير σ^2 إلى تباين الخطأ، ويتم حساب الخطأ المعياري للتقدير (SE) في ضوء النموذج ثلاثي المعلمة من خلال المعادلة (6) (Zimowski, et al, 2003):

$$S.E.(\hat{\theta}) = \left\{ 1/D^2 \sum_{j=1}^{n} a_j^2 \frac{1-P_j(\hat{\theta})}{P_j(\hat{\theta})} \left(\frac{P_j(\hat{\theta}) - g_j}{1-g_j} \right)^2 \right\}^{\frac{1}{2}} \dots\dots\dots\dots\dots\dots\dots\dots (6)$$

حيث أن g : اقتران الكثافة الاحتمالية لتوزيع المجتمع.

الدراسات السابقة

أجرى هامبلتون وجونز (Hambleton & Jones, 1994) دراسة بهدف الكشف عن أثر احتواء الاختبار على فقرات ذات معامل تمييز مرتفع على دقة تقدير دالة المعلومات، في ضوء حجم العينة ونسبة الفقرات المسحوبة من بنك الأسئلة، ولتحقيق هدف الدراسة قام الباحثان بتوليد بيانات خاصة بـ (80) فقرة شكلت النواة الأساسية للبنك المستخدم في الدراسة، بحيث يتم تكوين اختبارات مسحوبة من البنك بنسب 20% 40% ، في حين كان متغير حجم العينة بمستويين (200) مفحوص و (500) مفحوص، وأظهرت النتائج أنّ لكلا المتغيرين أثر على دقة تقدير دالة المعلومات، مع مستوى أثر أكبر لحجم العينة، كما أشارت النتائج إلى أنّ احتواء الاختبار على فقرات ذات معاملات تمييز عالية، تؤدي إلى تضخيم قيم دالة المعلومات للفقرة.

كما قام كل من ساس وشميت ووكر (Sass, Schmitt & Walker, 2008) بدراسة هدفت إلى التحقق من أثر شكل توزيع القدرة على معالم الفقرة، حيث تم توليد بيانات خاصة بـ (30) فقرة، بأحجام عينات (500) مفحوص و (1000) مفحوص، وأظهرت النتائج وجود فروق في متوسط الأخطاء المعيارية لتقدير معالم الفقرات باختلاف شكل التوزيع للقدرة، حيث كان مستوى الأخطاء المعيارية في التوزيعات الملتوية أكبر مقارنة بالأخطاء المعيارية المحسوبة في حالة التوزيع الطبيعي للقدرة.

وأجرى الرقب (2011) دراسة هدفت إلى التحقق من أثر الابتعاد عن التوزيع الطبيعي لدرجات الطلبة على اختبار قدرة عقلية في تقدير معالم فقراته، ولتحقيق هدف الدراسة استخدم الباحث اختبار اوتيس لينون (Otis – Lennon) للقدرة العقلية، وتم تطبيق الاختبار على مجموعة طلبة من الصفوف من الخامس إلى الحادي عشر بهدف تشكيل عينات مختلفة في درجة الإلتواء بأحجام تراوحت بين (509) إلى (542)، وتوصلت النتائج إلى وجود تحيز في تقدير معالم الفقرة ودالة المعلومات والخطأ المعياري عندما تكون درجة الالتواء أكبر من (2).

وهدفت دراسة بني عطا والشريفين (2012) إلى التحقق من أثر اختلاف شكل توزيع القدرة على معالم الفقرة ودالة المعلومات للاختبار، حيث قام الباحثان بتوليد اختبار مكون من (60) وفق افتراضات النموذج اللوجستي ثلاثي المعلمة، بحجم عينة (1000) مفحوص لكل توزيع (التواء نحو اليمين، التواء نحو اليسار، طبيعي، ثنائي المنوال)، وأسفرت النتائج على وجود فروق ذات دلالة إحصائية في معالم الصعوبة والتمييز باختلاف شكل توزيع القدرة، كما أظهرت النتائج بأن أكبر كمية للمعلومات وأعلى معامل ثبات جاءت للاختبار المعاير من التوزيع ثنائي المنوال، في حين كانت أقل كمية للمعلومات للاختبار المعاير بالتوزيع الطبيعي.

وقامت الحواري (2015) بدراسة هدفت إلى الكشف عن أثر طول الاختبار، وشكل توزيع القدرة في تقديرات معلمة الصعوبة ودالة المعلومات للفقرة وفق نموذج راش، ولتحقيق هدف الدراسة قامت الباحثة بتوليد بيانات خاصة بـ (1000) مفحوص، بطول اختبار (30) فقرة و (60) فقرة، لتشكيل توزيعات قدرة (التواء سالب، طبيعي، التواء موجب)، وأظهرت النتائج وجود فروق في تقدير معلمة الصعوبة لصالح الاختبار المكون من (30) فقرة، وأن كمية المعلومات التي قدمها الاختبار عندما كان توزيع القدرة ذو التواء موجب كانت هي الأعلى.

وهدفت دراسة محمود (2017) إلى التحقق من أثر شكل توزيع القدرة على دقة تقدير معلم الصعوبة في نموذج راش، حيث تم توليد بيانات خاصة بـ (60) فقرة ، في ضوء متغيري شكل التوزيع (اعتدالي، موجب الالتواء، سالب الالتواء، منتظم) وحجم العينة (100 , 200 , 300 , 400 , 500) مفحوص، وأظهرت النتائج أن تقدير معلمة الصعوبة في البيانات التي تتبع التوزيع المنتظم كان أكثر دقة، وأن مستوى الخطأ المعياري كان الأعلى في حالة التوزيعات الملتوية.

وهدفت دراسة نغواني (Ngwane, 2021) إلى التحقق من أثر شكل توزيع المعالم على دالة معلومات الاختبار، حيث تم توليد بيانات في ضوء متغيرات شكل التوزيع (طبيعي، منتظم)، وعدد الفقرات (50 , 100)، وحجم العينة (100 , 500 , 1000) مفحوص، وأظهرت النتائج أن دالة معلومات الاختبار تزداد مع زيادة طول الاختبار، ووجود فرق دال إحصائياً في دالة معلومات الاختبار بين البيانات التي تتبع التوزيع المنتظم والتوزيع الطبيعي ولصالح التوزيع الطبيعي.

وأجرى الصبح (2021) دراسة هدفت إلى المقارنة بين دقة تقدير المعالم باختلاف طول الاختبار وشكل التوزيع لمعلمة القدرة (طبيعي، التواء موجب، التواء سالب)، ولتحقيق هدف الدراسة استخدم الباحث بيانات الاختبار الوطني لمادة العلوم، بالاضافة إلى توليد بيانات خاصة بـ (25) فقرة ، و (50) فقرة ، وأشارت نتائج الدراسة إلى عدم وجود فروق عملية في دقة التقديرات باختلاف طبيعة البيانات، وحجم الاختبار، وشكل التوزيع لمعلمة القدرة.

وهدفت دراسة علوان وجاسم (Alwan& Jasim, 2022) إلى التحقق من أثر شكل توزيع القدرة على معالم فقرات اختبار للقدرة العقلية، حيث قام الباحثان باستخدام اختبار مكون من (78) فقرة ، بحجم عينة (1000) طالب وطالبة من المرحلة المتوسطة ، وأسفرت النتائج على وجود أثر للتوزيع موجب الالتواء على دقة معالم الصعوبة والتمييز والتخمين.

يتضح مما سبق ومن خلال مراجعة الدراسات السابقة نلاحظ أن غالبيتها اعتمدت على بيانات مولدة (غير حقيقية) باستخدام برنامج (WINGEN)،

والدراسة التي اعتمدت على بيانات حقيقية استخدمت بيانات اختبار قدرة عقلية لعينات من فئات عمرية فيها تفاوت كبير، كما أنها اعتمدت على توزيعات ملتوية فقط دون مقارنتها مع التوزيع الطبيعي، بالاضافة إلى أن بعض الدراسات السابقة لم تحدد درجة الالتواء المستخدمة، ولم توضح إجراءات تشكيل التوزيعات أو حتى التأكد من تحقق اعتدالية التوزيع في العينات التي أُعتبرت أنها تتبع التوزيع الطبيعي. علاوة على ذلك نجد أن الدراسات ركزت على شكل التوزيع لقدرة الأفراد على معالم الفقرات والخطأ المعياري، وليس شكل التوزيع للدرجات الخام للمفحوصين قبل معايرتها وفق نظرية الاستجابة للفقرة، وهذا ما يميز الدراسة الحالية باستخدامها لبيانات حقيقية واعتماد شكل التوزيع للدرجات الخام للمفحوصين.

الطريقة والاجراءات

منهج الدراسة

اتّبعت الدراسة المنهج الوصفي المقارن، حيث تم وصف النتائج المتعلقة بمعاملات الثبات ودالة المعلومات لفقرات الاختبارات الوطنية في ضوء شكل التوزيع للبيانات، ومن ثم المقارنة بينها.

مجتمع الدراسة

تكون مجتمع الدراسة من جميع طلبة الصف الثامن الأساسي من الذكور والإناث المنتظمين في مدارس المملكة الأردنية الهاشمية الحكومية، والخاصة، والمدارس التابعة لوكالة الغوث، والمدارس التابعة لمديرية التعليم والثقافة العسكرية للعام الدّراسيّ (2018/ 2019)، ممن تقدموا للصورة الورقية للاختبار الوطني لضبط نوعية التعليم في مباحث اللغة العربية واللغة الإنجليزية والرياضيات والعلوم والبالغ عددهم حسب إحصائيات وزارة التربية والتعليم (92419) طالبًا وطالبة.

عينة الدراسة

تم اختيار عينة عشوائية مؤلفة من(12000) طالبًا وطالبة، وذلك لتشكيل (12) مجموعة فرعية باختلاف شكل الالتواء (موجب، معتدل، سالب)، والمادة الدراسية (اللغة عربية، الرياضيات، اللغة الإنجليزية، العلوم)، بواقع (1000) طالب وطالبة لكل مجموعة.

أدوات الدراسة

لتحقيق هدف الدراسة والمتمثل في التحقق من مدى تأثر دالة المعلومات للفقرة ومعامل الثبات باختلاف شكل التوزيع للدرجات على الاختبار، تم استخدام نتائج الاختبارات الوطنية لضبط نوعية التعليم المعد من قبل إدارة الامتحانات والاختبارات في وزارة التربية والتعليم في الأردن للمباحث الأساسية الأربعة:

- اختبار اللغة العربية: تكون الاختبار من (40) فقرة من نوع (الاختيار من متعدد) موزعة على محاور: القراءة، القواعد والتطبيقات اللغوية، والكتابة.
- اختبار اللغة الإنجليزية تكون الاختبار من (36) فقرة من نوع (الاختيار من متعدد) موزعة على محاور: القراءة، الكتابة، والمحادثة.
- اختبار الرياضيات: تكون الاختبار من (40) فقرة من نوع (الاختيار من متعدد) موزعة على محاور: الأعداد والعمليات عليها، الجبر، القياس، الهندسة، والإحصاء والاحتمالات.
- اختبار العلوم: تكون الاختبار من (42) فقرة من نوع (الاختيار من متعدد) موزعة على محاور: المادة، الكائنات الحية وبيئاتها، الوراثة، الطاقة، القوة والحركة، علوم الأرض والفضاء، علوم البيئة.

صدق وثبات الأدوات:

قامت وزارة التربية والتعليم بتشكيل لجان؛ لكتابة فقرات الاختبارات، حيث تم التّحقق من صدق المحتوى للاختبار من خلال تحديد مهارات التعلم الأساسية التي سيقيسها الاختبار، والتي تعتمد على نتاجات التّعلم للصّفوف (الخامس، والسادس، والسابع، والثامن) في مباحث اللغة العربية، واللغة الإنجليزية، والرياضيات، والعلوم، ومن ثّم تم إعداد جدول المواصفات لكلّ اختبار، وبناء تجمع من الفقرات الاختبارية. وتم تطبيق الفقرات على عينة استطلاعية من الطلبة، للتحقق من مقروئيتها وخصائصها السيكومترية، حيث تم استخراج معاملات الصعوبة والتمييز وفاعلية المموهات، واستخراج معاملات ثبات الاختبارات باستخدام معادلة ألفا كرونباخ، والجدول (1) يوضح معاملات ثبات الصورة الورقية للاختبار الوطني لضبط نوعية التعليم (إدارة الامتحانات والاختبارات، 2019).

جدول 1. معاملات الثبات للصورة الورقية للاختبار الوطني لضبط نوعية التعليم.

الاختبار	معامل الثبات
اللغة العربية	0.87
اللغة الإنجليزية	0.91
الرياضيات	0.86
العلوم	0.84

التحقق من جودة مطابقة البيانات لافتراضات نظرية الاستجابة للفقرة:

تم التحقق من الافتراضات الخاصة بالنظرية كما يلي:

أولًا: افتراض أحادية البعد

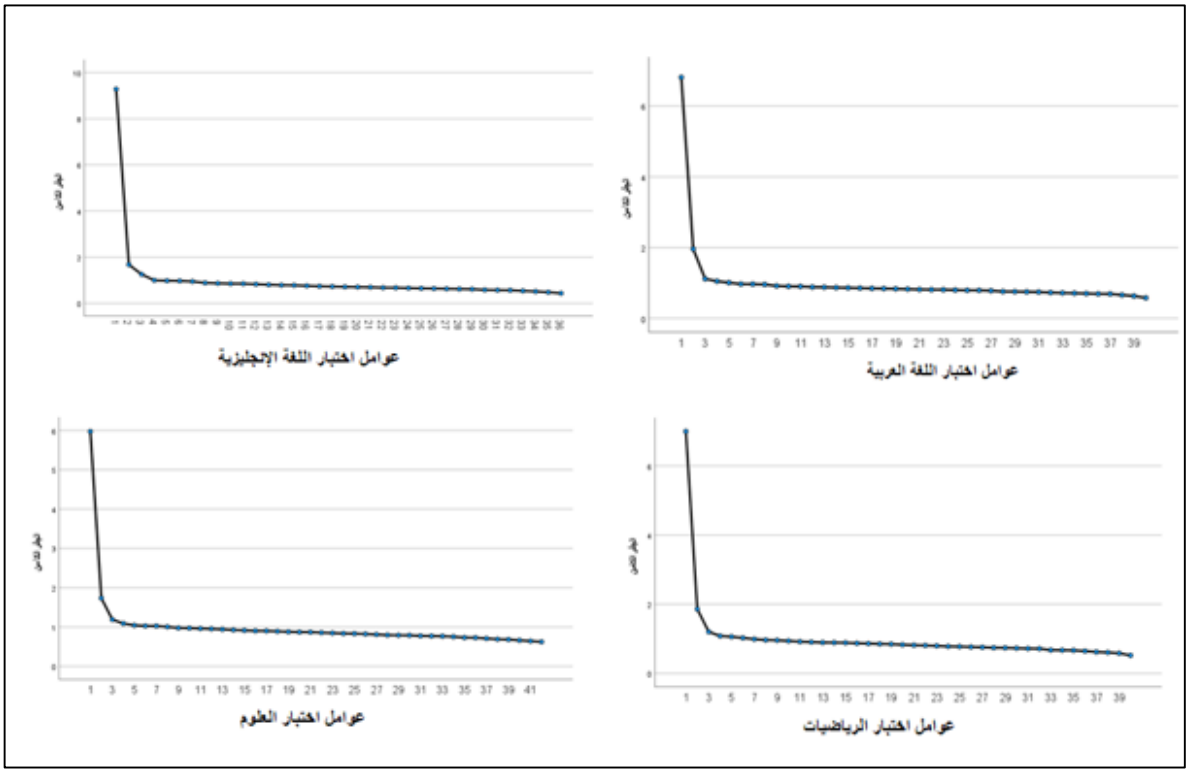
للتحقق من افتراض أحادية البعد لبيانات الاختبارات الأربعة، تم استخدام الرزمة الإحصائية للعلوم الاجتماعية (SPSS)؛ لإجراء التّحليل العاملي الاستكشافي (Exploratory Factor Analysis)، باستخدام تحليل المكونات الأساسية (Principal Component Analysis)، مع التدوير المتعامد (Varimax)، حيث حسبت قيم الجذور الكامنة (Eigen Values)، ونسبة التباين المفسر (Explained Variance) لكل عامل من العوامل، وفي التّحليل العاملي تعتمد أحادية البعد على أن نسبة الجذر الكامن للعامل الأول إلى الجذر الكامن للعامل الثاني لا تقل عن (2). ويوضح الجدول (2) نتائج هذا التحليل:

جدول 2. قيم الجذور الكامنة ونسب التباين المفسر لعوامل الاختبارات.

المادة	العامل	الجذر الكامن	نسبة التباين المفسر	نسبة التباين المفسر التراكمي	نسبة الجذر الكامن للعامل الأول إلى الجذر الكامن للعامل الثاني
اللغة العربية	1	6.81	17.04	17.04	
	2	1.96	4.90	21.94	
	3	1.12	2.79	24.73	3.47
	4	1.05	2.62	27.35	
	5	1.01	2.52	29.88	
اللغة الإنجليزية	1	9.28	25.79	25.79	
	2	1.68	4.66	30.45	5.52
	3	1.26	3.49	33.94	
الرياضيات	1	7.00	17.50	17.50	
	2	1.86	4.65	22.15	
	3	1.21	3.01	25.16	
	4	1.08	2.71	27.86	3.76
	5	1.06	2.65	30.52	
	6	1.03	2.57	33.08	
العلوم	1	5.97	14.22	14.22	
	2	1.73	4.13	18.35	
	3	1.19	2.84	21.19	
	4	1.09	2.60	23.79	
	5	1.05	2.49	26.27	3.45
	6	1.03	2.45	28.72	
	7	1.03	2.45	31.17	
	8	1.01	2.40	33.56	

يتضح من الجدول (2) أن اختبار اللغة العربية أفرز خمسة عوامل قيمة الجذر الكامن لكل منها تزيد عن (1)، وقد فسّر العامل الأول (17.04%) من تباين الأداء على الاختبار، في حين أفرز اختبار اللغة الإنجليزية ثلاثة عوامل، فسّر العامل الأول منها (25.79%) من تباين الأداء على الاختبار، وأفرز اختبار الرياضيات ستة عوامل قيمة الجذر الكامن لكل منها تزيد عن (1)، وبلغت نسبة التباين المفسر للعامل الأول (17.50%)، كما أظهرت النتائج أن اختبار العلوم أفرز ثمانية عوامل، فسّر العامل الأول منها (14.22%) من تباين الأداء على الاختبار، وباستخراج النسبة بين الجذر الكامن للعامل الأول إلى الجذر الكامن للعامل الثاني لجميع الاختبارات نجد أنها أكبر من القيمة (2)، الأمر الذي يعد مؤشرًا على تحقق افتراض أحادية البعد في الأداء على هذه الاختبارات.

كما تم استخدام التمثيل البياني (Scree Plot) لقيم الجذور الكامنة للعوامل المكونة للاختبارات، كما هو موضح بالشكل (1).



الشكل(1): التمثيل البياني لقيم الجذور الكامنة للعوامل المكونة للاختبارات.

يُلاحظ من الشكل (1) أن الجذر الكامن للعامل الأول لجميع الاختبارات يتميز بشكل واضح عن الجذور الكامنة لبقية العوامل، وأن الفرق بين الجذر الكامن للعامل الأول والجذر الكامن للعامل الثاني كبيرًا، إذا ما تمت مقارنته بالفرق بين الجذر الكامن للعامل الثاني والجذر الكامن للعامل الثالث، كما يُلاحظ أن هناك تحولًا في ميل المنحنى عند العامل الثاني ويليه استقرار لقيمة الجذر الكامن لبقية العوامل، مما يشير إلى تحقق افتراض أحادية البعد في تفسير الأداء.

ثانيًا: افتراض الاستقلال الموضعي (Local Independent Assumption)

يشير هامبلتون وسوامينثان (Hambleton & Swaminathan, 1985) إلى أن هذا الافتراض يكافئ افتراض أحادية البعد، بمعنى أنه إذا تحقق افتراض أحادية البعد، فإن افتراض الاستقلال الموضعي يتحقق أيضًا، وكما تبين من التحليل العاملي فقد تحقق افتراض أحادية البعد للاختبارات الأربعة، وعليه تحقق افتراض الاستقلال الموضعي.

ثالثًا: افتراض منحنى خصائص الفقرة

يمثل منحنى خصائص الفقرة دالة رياضية اطرادية تربط بين قدرة الفرد (θ) واحتمال الإجابة الصحيحة عن الفقرة ($P(\theta)$، وتم التحقق من هذا الافتراض من خلال رسم منحنى الخصائص للفقرة (ICC) لكل فقرة من فقرات الاختبارات، وأظهرت منحنيات الفقرات تحقق خاصية الاطرادية.

رابعًا: افتراض التحرر من السرعة

يشير هذا الافتراض إلى أن إخفاق الطلبة في الإجابة عن بعض الفقرات يُعزى إلى انخفاض عامل القدرة لديهم، وليس لعامل السرعة. وقد قام معدّو الاختبار في وزارة التربية والتعليم بتقدير الزمن المناسب للاختبار بعد التجريب الأولي له، حيث تم تحديد زمن الاختبار (ساعة ونصف) وهو وقت كافٍ للطلبة للقيام بالإجابة على جميع الأسئلة.

كما تم التحقق من نسبة الطلبة الذين أجابوا على جميع فقرات الاختبار حيث أشار هامبلتون وروجرز (Hambleton & Rogers, 1991) أنه إذا أجاب (75%) من الطلبة عن جميع فقرات الاختبار، وإذا كان (80%) من فقرات الاختبار قد تم الإجابة عنها من قبل المفحوصين فإن السرعة لا تعتبر عاملًا مهمًا في الأداء على الاختبار، وبعد مراجعة إجابات الطلبة على الاختبارات الأربعة نجد أن نسبة الطلبة الذين أجابوا على جميع الفقرات تفوق هذه النسبة، مما يشير إلى أنه لم يكن لعامل السرعة دور في الإجابة عن فقرات اختبارات اللغة العربية، واللغة الإنجليزية، والرياضيات، والعلوم.

التحقق من جودة مطابقة الأفراد للنموذج ثلاثي المعلمة:

بعد فرز البيانات الخاصة بكل اختبار تم استبعاد الإجابات غير المكتملة، واختيار جميع الأفراد الذين قاموا بالمحاولة للإجابة عن جميع الفقرات،

كما تم حذف إجابات كل طالب أخفق في الإجابة عن جميع فقرات الاختبار، وحذف إجابات كل طالب نجح في الإجابة عن جميع فقرات الاختبار، وللتحقق من مطابقة الأفراد للنموذج ثلاثي المعلمة، تم استخدام برنامج (BILOG-MG)، ومن خلال مؤشر مربع كاي(x^2) حيث يعد الفرد غير مطابق للنموذج إذا كانت الاحتمالية أقل من (0.05)، ويوضح الجدول (3) توزيع الطلبة بعد استبعاد الأفراد غير المطابقين للنموذج.

جدول 3. توزيع الطلبة الطلبة بعد استبعاد الأفراد غير المطابقين للنموذج بحسب المادة الدراسية.

المادة الدراسية	عدد الطلبة
اللغة العربية	18740
اللغة الإنجليزية	16271
الرياضيات	16151
العلوم	16810
المجموع	67972

التحقق من جودة مطابقة الفقرات للنموذج ثلاثي المعلمة:

للتحقق من مطابقة الفقرات للنموذج ثلاثي المعلمة، تم استخدام برنامج (BILOG-MG)، ومن خلال مؤشر مربع كاي(x^2) حيث تعد الفقرة غير مطابقة للنموذج إذا كانت الاحتمالية أقل من (0.05)، ويشير هامبلتون وسواميناثان(Hambleton & Swaminathan, 1985) إلى أنّ قيمة الإحصائي مربع كاي(x^2) لمطابقة الفقرات تتأثر كثيرًا بكبر حجم العينة، وتصبح دالة إحصائيًا وتُفسر على أن الفقرات لا تطابق النموذج، مما يضفي الشك على هذه القيم (علام، 2005). ونظرًا لحساسية الاختبار الإحصائي مربع كاي(x^2) لحجم العينة، وبالرجوع للدراسات السابقة نجد أن معظمها تعتمد على عينة مكونة من (1000) مفحوص للتحقق من مطابقة الفقرات للنموذج ثلاثي المعلمة لكل اختبار، وبناء على ذلك فقد تم التحقق من مطابقة الفقرات على عينة عشوائية مكونة من (1000) طالب لكل اختبار.

وأسفرت النتائج عن عدم مطابقة سبع فقرات من اختبار اللغة العربية مع توقعات النموذج ثلاثي المعلمة، وهي: (5, 8, 13, 16, 20, 31, 38)، وعدم مطابقة ست فقرات من اختبار اللغة الإنجليزية مع توقعات النموذج ثلاثي المعلم، وهي: (10, 15, 23, 25, 29, 30)، وعدم مطابقة سبع فقرات من اختبار الرياضيات مع توقعات النموذج ثلاثي المعلم، وهي: (1, 4, 8, 14, 17, 19, 39)، وعدم مطابقة خمس فقرات من اختبار العلوم مع توقعات النموذج ثلاثي المعلم، وهي: (4, 16, 17, 19, 40)، حيث كانت القيمة الاحتمالية للاحصائي (x^2) لهذه الفقرات أقل من (0.05)، مما يشير إلى أن الاستجابات الملاحظة عليها تبتعد عن الاستجابات التي يتوقعها النموذج، وبالتالي تم حذف هذه الفقرات.

وبناء على نتائج مطابقة الفقرات للنموذج ثلاثي المعلمة، فقد تكونت أداة الدراسة من (33) فقرة من فقرات الاختبار الوطني لضبط نوعية التعليم لمبحث اللغة العربية، و(30) فقرة من فقرات الاختبار الوطني لضبط نوعية التعليم لمبحث اللغة الإنجليزية، و (33) فقرة من فقرات الاختبار الوطني لضبط نوعية التعليم لمبحث الرياضيات، و(37) فقرة من فقرات الاختبار الوطني لضبط نوعية التعليم لمبحث العلومالتي تم استخدامها كأساس للإجابة عن أسئلة الدراسة.

تشكيل التوزيعات بعد الانتهاء من مطابقة الأفراد والفقرات.

بعد حذف الأفراد غير المطابقين والفقرات غير المطابقة لتوقعات النموذج ثلاثي المعلم، تم تشكيل التوزيعات المطلوبة للدراسة حسب شكل الالتوء (موجب، اعتدالي، سالب)، وذلك لكل مبحث من المباحث الأربعة، حيث تم تشيكل (12) ملفًا للبيانات بحيث يحتوي كل ملف على استجابات (1000) طالبًا وطالبة موزعة على النحو التالي:

- بيانات ذات التواء موجب بمعامل يقترب من (1).
- بيانات ذات التواء سالب بمعامل يقترب من (1-).
- بيانات معتدلة الالتواء بمعامل يقترب من (0).

وفيما يلي توضيح الإحصاءات الوصفية للبيانات باختلاف شكل التوزيع

جدول 4. الاحصاءات الوصفية لدرجات الطلبة باختلاف المبحث وشكل التوزيع.

المادة	شكل التوزيع	الوسط الحسابي	الوسيط	المنوال	الانحراف المعياري
اللغة العربية	موجب	11.94	10	8	6.08
	معتدل	16.08	16	16	5.70
	التواء سالب	21.11	23	25	6.07
اللغة الإنجليزية	التواء موجب	11.23	10	8	5.58
	معتدل	16.07	16	16	5.52
	التواء سالب	22.51	23	28	5.13
الرياضيات	التواء موجب	11.74	10	8	6.81
	معتدل	15.77	16	16	5.88
	التواء سالب	21.02	22	22	6.68
العلوم	التواء موجب	14.88	14	12	6.69
	معتدل	18.76	19	19	6.11
	التواء سالب	23.38	25	28	6.56

يتضح من الجدول (4) أن قيم المتوسط الحسابي والوسيط والمنوال لدرجات الطلبة على الاختبارات تزداد كلما اتجهنا من مستوى الالتواء الموجب إلى الالتواء السالب، وأن نتائج مقاييس النزعة المركزية كانت مؤشرات جيدة لأغراض تشكيل التوزيعات المطلوبة للدراسة، وللتأكد من تحقق شكل التوزيع للبيانات، تم الاعتماد على القيمة المعيارية للتفلطح Z_k، والقيمة المعيارية للالتواء Z_{S_k} لفحص التوزيع الطبيعي كما هو موضح بالمعادلتين (7) و (8) ، بحيث إذا وقعت القيم المعيارية للتفلطح والالتواء ما بين القيمتين (1.96-) و (1.96) يكون التوزيع مماثلًا للتوزيع الطبيعي تقريبًا (Corder & Foreman, 2014).

$$z_K = \frac{K-0}{SE_K} \dots\dots\dots\dots\dots\dots\dots\dots (7)$$

$$z_{S_k} = \frac{S_k-0}{SE_{S_k}} \dots\dots\dots\dots\dots\dots\dots\dots (8)$$

حيث أن:

K: معامل التفلطح.

SE_k : الخطأ المعياري لمعامل التفلطح.

S_k: معامل الالتواء.

SE_{S_k} : الخطأ المعياري لمعامل الالتواء.

جدول 5. معاملات الالتواء والتفلطح والقيم المعيارية للتحقق من اعتدالية التوزيع.

المادة	شكل التوزيع	معامل الالتواء	الخطأ المعياري للالتواء	Z_k	معامل التفلطح	الخطأ المعياري للتفلطح	Z_{S_k}
اللغة العربية	موجب	1.09	0.077	**14.16**	0.72	0.155	**4.65**
	معتدل	0.03	0.077	**0.39**	-0.29	0.155	**-1.87**
	التواء سالب	-1.03	0.077	**-13.37**	0.41	0.155	**2.65**
اللغة الانجليزية	التواء موجب	1.13	0.077	**14.68**	0.89	0.155	**5.74**
	معتدل	-0.07	0.077	**-0.91**	-0.28	0.155	**-1.81**
	التواء سالب	-0.96	0.077	**-12.47**	0.70	0.155	**4.52**
الرياضيات	التواء موجب	1.06	0.077	**13.77**	0.32	0.155	**2.06**
	معتدل	0.05	0.077	**0.65**	-0.29	0.155	**-1.87**

المادة	شكل التوزيع	معامل الالتواء	الخطأ المعياري للالتواء	Z_k	معامل التفلطح	الخطأ المعياري للتفلطح	Z_{S_k}
	التواء سالب	-0.99	0.077	**-12.86**	0.60	0.155	**3.87**
العلوم	التواء موجب	1.02	0.077	**13.25**	0.61	0.155	**3.94**
	معتدل	0.08	0.077	**1.04**	-0.21	0.155	**-1.35**
	التواء سالب	-1.01	0.077	**-13.12**	0.74	0.155	**4.77**

أشارت النتائج المبينة في الجدول (5) أن جميع القيم المعيارية للتفلطح والالتواء الخاصة بالتوزيع المعتدل وقعت ما بين القيمتين (1.96-) و (1.96) مما يشير إلى أن التوزيع مماثلًا للتوزيع الطبيعي تقريبًا، في حين كانت القيم المعيارية للتوزيع ذو الالتواء الموجب والسالب خارج هذه الفترة.

نتائج الدراسة وتفسيرها.

للإجابة عن سؤال الدراسة الأول الذي ينص على " هل تختلف دقة تقديرات معلومات الفقرات باختلاف شكل التوزيع للبيانات عند مستوى الدلالة(0.05)؟" تم استخراج المتوسطات الحسابية للأخطاء المعيارية الخاصة بتقدير دالة المعلومات للفقرات باختلاف شكل التوزيع، وذلك لجميع الاختبارت كما هو موضح بالجدول (6).

جدول 6. المتوسطات الحسابية للأخطاء المعيارية الخاصة بتقدير دالة المعلومات.

المادة	شكل التوزيع	متوسط الخطأ المعياري للتقدير
اللغة العربية	موجب	0.390
	معتدل	0.175
	التواء سالب	0.070
اللغة الإنجليزية	التواء موجب	0.459
	معتدل	0.154
	التواء سالب	0.103
الرياضيات	التواء موجب	0.477
	معتدل	0.376
	التواء سالب	0.108
العلوم	التواء موجب	0.556
	معتدل	0.403
	التواء سالب	0.094

يتضح من الجدول (6) أن قيم متوسط الخطأ المعياري للتقدير لدالة معلومات الفقرة، تنخفض كلما اتجهنا من مستوى الالتواء الموجب إلى الالتواء السالب، حيث تراوحت قيم متوسطات الخطأ للتوزيعات ذات الالتواء الموجب بين (0.390) إلى (0.556)، في حين كانت تتراوح بين (0.154) إلى (0.403) في التوزيعات المعتدلة، وتراوحت قيم متوسطات الخطأ المعياري للتوزيعات ذات الالتواء السالب بين (0.070) إلى (0.108)، وللتحقق من دلالة الفروق بين المتوسطات الحسابية للخطأ المعياري للتقدير، تم استخدام تحليل التباين الأحادي للقياسات المتكررة (one way ANOVA for repeated measures) والنتائج موضحة بالجدول (7).

جدول 7. نتائج تحليل التباين للقياسات المتكررة للفروق في متوسط الخطأ لتقدير دالة معلومات الفقرة باختلاف شكل التوزيع للبيانات.

المادة	مصدر التباين	مجموع مربعات التباين	درجات الحرية	متوسط مربعات التباين	قيمة f	مستوى الدلالة
اللغة العربية	شكل التوزيع	1.75	2	0.88	22	0.000
	الخطأ	2.36	64	0.04		
اللغة الإنجليزية	شكل التوزيع	2.23	2	1.12	25.16	0.000
	الخطأ	2.57	58	0.04		

المادة	مصدر التباين	مجموع مربعات التباين	درجات الحرية	متوسط مربعات التباين	قيمة f	مستوى الدلالة
الرياضيات	شكل التوزيع	2.40	2	1.20	7.05	0.002
	الخطأ	11.15	64	0.17		
العلوم	شكل التوزيع	4.11	2	2.06	6.06	0.004
	الخطأ	24.46	72	0.34		

يتضح من الجدول (7) أن جميع قيم f الخاصة بفحص دلالة الفروق في متوسط الخطأ لتقدير دالة معلومات الفقرة باختلاف شكل التوزيع للبيانات، كانت دالة إحصائيًا حيث كان مستوى الدلالة لجميع الاختبارات أقل من مستوى الدلالة (α= 0.05)، وللتحقق من موقع دلالة الفروق في المتوسطات بين أي مجموعتين، تم إجراء المقارنات البعدية بأسلوب أقل فرق دال (Least Significant Difference: LSD) كما هو موضح بالجدول (8).

جدول 8. نتائج المقارنات البعدية بطريقة (LSD) للفروق في متوسط الخطأ لتقدير دالة معلومات الفقرة باختلاف شكل التوزيع للبيانات.

المادة	شكل التوزيع		الفرق المطلق بين المتوسطين	مستوى الدلالة
اللغة العربية	التواء موجب	معتدل	0.214	0.000
	التواء موجب	التواء سالب	0.320	0.000
	معتدل	التواء سالب	0.106	0.000
اللغة الإنجليزية	التواء موجب	معتدل	0.305	0.000
	التواء موجب	التواء سالب	0.356	0.000
	معتدل	التواء سالب	0.051	0.055
الرياضيات	التواء موجب	معتدل	0.101	0.000
	التواء موجب	التواء سالب	0.369	0.000
	معتدل	التواء سالب	0.268	0.000
العلوم	التواء موجب	معتدل	0.153	0.134
	التواء موجب	التواء سالب	0.463	0.002
	معتدل	التواء سالب	0.309	0.060

يتضح من الجدول (8) أن الفروق بين متوسطات الخطأ لتقدير دالة معلومات الفقرة بين البيانات ذات التوزيع موجب الالتواء والتوزيع سالب الالتواء، كانت ذات دلالة إحصائية في جميع المواد، ولصالح التوزيع سالب الالتواء، كما كانت الفروق بين متوسطات الخطأ للبيانات ذات التوزيع المعتدل مع البيانات موجبة وسالبة الالتواء دالة إحصائيًا في مواد اللغة العربية، واللغة الإنجليزية، والرياضيات، في حين كانت الفروق غير دالة في مادة العلوم.

وللإجابة عن سؤال الدراسة الثاني الذي ينص على: "هل تختلف تقديرات ثبات الاختبار وفق الطرق القائمة على نظرية الاستجابة للفقرة باختلاف شكل التوزيع للبيانات عند مستوى الدلالة (0.05)؟" تم استخراج قيم معاملات الثبات النظري والتجريبي باستخدام برنامج (BILOG-MG) والنتائج موضحة بالجدول (9).

جدول 9. نتائج معامل الثبات النظري والتجريبي للاختبارات باختلاف شكل التوزيع للبيانات.

المادة	نوع معامل الثبات	شكل التوزيع			قيمة x^2	مستوى الدلالة
		التواء موجب	معتدل	التواء سالب		
اللغة العربية	النظري	0.787	0.835	0.867	51.62	0.000
	التجريبي	0.737	0.805	0.846	66.45	0.000
اللغة الانجليزية	النظري	0.769	0.846	0.856	57.95	0.000
	التجريبي	0.714	0.822	0.833	75.78	0.000
الرياضيات	النظري	0.798	0.845	0.884	72.09	0.000

المادة	نوع معامل الثبات	شكل التوزيع			قيمة x^2	مستوى الدلالة
		التواء موجب	معتدل	التواء سالب		
	التجريبي	0.764	0.818	0.872	89.12	0.000
العلوم	النظري	0.825	0.856	0.876	27.94	0.000
	التجريبي	0.798	0.835	0.859	38.34	0.000

يتضح من الجدول (9) أن معاملات الثبات ترتفع كلما اتجهنا من مستوى الالتواء الموجب إلى الالتواء السالب، حيث تراوحت قيم معاملات الثبات النظري للتوزيعات ذات الالتواء الموجب بين (0.769) إلى (0.825)، في حين كانت تتراوح بين (0.835) إلى (0.856) في التوزيعات المعتدلة، وتراوحت قيم معاملات الثبات النظري للتوزيعات ذات الالتواء السالب بين (0.856) إلى (0.884)، تراوحت قيم معاملات الثبات التجريبي للتوزيعات ذات الالتواء الموجب بين (0.714) إلى (0.798)، في حين كانت تتراوح بين (0.805) إلى (0.835) في التوزيعات المعتدلة، وتراوحت قيم معاملات الثبات التجريبي للتوزيعات ذات الالتواء السالب بين (0.833) إلى (0.872)، وللتحقق من دلالة الفروق بين معاملات الثبات، تم استخدام الحزمة الإحصائية كوكرون (Cocron) وهي حزمة جزيئة ضمن برمجية R (R-Package)، والتي تستخدم اختبار مربع كاي (x^2) للكشف عن الفروق في معاملات الثبات بالاعتماد على ما وصل له فلدت (Feldet, 1980) حول المقارنة بين معاملات الثبات، حيث أظهرت نتائج اختبار مربع كاي (x^2) أن الفروق بين معاملات الارتباط النظري باختلاف شكل التوزيع للبيانات كانت ذات دلالة إحصائية في جميع المواد، كما أشارت النتائج أن الفروق بين معاملات الارتباط التجريبي باختلاف شكل التوزيع للبيانات كانت ذات دلالة إحصائية في جميع المواد، وللتحقق من موقع دلالة الفروق في معاملات الثبات بين أي مجموعتين، تم إعادة استخدام اختبار مربع كاي (x^2) وذلك لكل معاملين على حدة كما هو موضح بالجدولين (10) و (11).

جدول 10. نتائج المقارنات الثنائية للفروق في قيم معامل الثبات النظري باختلاف شكل التوزيع للبيانات.

المادة	شكل التوزيع	معامل الثبات	شكل التوزيع	معامل الثبات	قيمة x^2	مستوى الدلالة
اللغة العربية	التواء موجب	0.787	معتدل	0.835	15.28	0.000
	التواء موجب	0.787	التواء سالب	0.867	51.59	0.000
	معتدل	0.835	التواء سالب	0.867	10.90	0.000
اللغة الإنجليزية	التواء موجب	0.769	معتدل	0.846	38.12	0.000
	التواء موجب	0.769	التواء سالب	0.856	51.64	0.000
	معتدل	0.846	التواء سالب	0.856	1.05	0.301
الرياضيات	التواء موجب	0.798	معتدل	0.845	16.43	0.000
	التواء موجب	0.798	التواء سالب	0.884	71.30	0.000
	معتدل	0.845	التواء سالب	0.884	19.67	0.000
العلوم	التواء موجب	0.825	معتدل	0.856	8.97	0.003
	التواء موجب	0.825	التواء سالب	0.876	27.93	0.000
	معتدل	0.856	التواء سالب	0.876	5.28	0.021

يتضح من الجدول (10) أن الفروق بين معاملات الثبات النظري للبيانات ذات التوزيع موجب الالتواء والتوزيع سالب الالتواء، كانت ذات دلالة إحصائية في جميع المواد، ولصالح التوزيع سالب الالتواء، وأن الفروق بين معاملات الثبات النظري للبيانات ذات التوزيع موجب الالتواء والتوزيع المعتدل كانت ذات دلالة إحصائية في جميع المواد، ولصالح التوزيع المعتدل، كما كانت الفروق بين معاملات النظري للبيانات ذات التوزيع المعتدل مع البيانات سالبة الالتواء دالة إحصائيًا في مواد اللغة العربية، والرياضيات، والعلوم، ولصالح التوزيع سالب الالتواء.

جدول 11. نتائج المقارنات الثنائية للفروق في قيم معامل الثبات التجريبي باختلاف شكل التوزيع للبيانات.

المادة	شكل التوزيع	معامل الثبات	شكل التوزيع	معامل الثبات	قيمة x^2	مستوى الدلالة
اللغة العربية	التواء موجب	0.737	معتدل	0.805	20.94	0.000
	التواء موجب	0.737	التواء سالب	0.846	66.45	0.000
	معتدل	0.805	التواء سالب	0.846	13.06	0.000
اللغة الإنجليزية	التواء موجب	0.714	معتدل	0.822	51.99	0.000
	التواء موجب	0.714	التواء سالب	0.833	66.73	0.000
	معتدل	0.822	التواء سالب	0.833	0.95	0.329
الرياضيات	التواء موجب	0.764	معتدل	0.818	15.82	0.000
	التواء موجب	0.764	التواء سالب	0.872	86.48	0.000
	معتدل	0.818	التواء سالب	0.872	28.95	0.000
العلوم	التواء موجب	0.798	معتدل	0.835	9.62	0.002
	التواء موجب	0.798	التواء سالب	0.859	30.29	0.000
	معتدل	0.835	التواء سالب	0.859	5.81	0.016

يتضح من الجدول (11) أن الفروق بين معاملات الثبات التجريبي للبيانات ذات التوزيع موجب الالتواء والتوزيع سالب الالتواء، كانت ذات دلالة إحصائية في جميع المواد، ولصالح التوزيع سالب الالتواء، وأن الفروق بين معاملات الثبات التجريبي للبيانات ذات التوزيع موجب الالتواء والتوزيع المعتدل، كانت ذات دلالة إحصائية في جميع المواد، ولصالح التوزيع المعتدل، كما كانت الفروق بين معاملات التجريبي للبيانات ذات التوزيع المعتدل مع البيانات سالبة الالتواء دالة إحصائيًا في مواد اللغة العربية والرياضيات والعلوم ولصالح التوزيع سالب الالتواء.

اتفقت نتائج الدراسة الحالية مع نتائج دراسة كل من ساس وشميت ووكر (Sass, Schmitt & Walker, 2008)؛ علوان وجاسم (Alwan& Jasim, 2022)؛ الرقب (2011)؛ ودراسة محمود (2017)، التي أظهرت نتائجها على وجود أثر للالتواء على دقة تقدير دالة معلومات الاختبار. ويمكن تفسير نتيجة الدراسة الحالية بالاعتماد على مستوى الطلبة، حيث أنه في حالة البيانات ذات الالتواء الموجب كان مستوى الأداء على الاختبارات منخفضًا بسبب كون المفحوصين من الطلبة ذوي القدرات المتدنية، مما يؤدي إلى إنخفاض درجات الأفراد على الفقرات وتشتتها، ويعد هذا مؤشرًا إلى أن مستوى صعوبة الفقرات لا يتوافق مع قدرات المفحوصين، مما يجعل احتمالية اعتماد الطلبة على التخمين للإجابة على فقرات الاختبار الوطني أكبر، والذي يؤدي إلى زيادة قيمة الخطأ المعياري، وبالتالي ينخفض معامل الثبات. بينما في حالة البيانات ذات الالتواء السالب، كان مستوى الأداء على الاختبارات مرتفعًا، مما يشير إلى أن مستوى صعوبة الأسئلة متوافق أو أدنى من قدرات المفحوصين، وبالتالي تكون إحتمالية إعتماد الطلبة على التخمين للإجابة عن فقرات الاختبار الوطني متدنية، والذي يؤدي إلى إنخفاض قيمة الخطأ المعياري، وبالتالي يزداد معامل الثبات.

ويمكن تبرير هذه النتيجة إحصائياً من خلال المعادلة رقم **(2)** التي وردت بالاطار النظري للدراسة، التي تبين طريقة حساب دالة المعلومات للفقرة في النموذج ثلاثي المعلمة، حيث يتضح من خلال المعادلة أن دالة المعلومات للفقرة تزداد كلما اقتربت معلمة التخمين من الصفر.

كما يمكن تفسير ما سبق في ضوء ما أشار إليه هامبلتون وسوامينثان (Hambleton & Swaminathan, 1985) إلى أنه من المتوقع أن يقدم الاختبار ذو المستوى الصعب تقديرات أكثر دقة عند مستويات القدرة المرتفعة، في حين من المتوقع أن يقدم الاختبار السهل تقديرات أكثر دقة عند مستويات القدرة المتدنية. أي أن الفقرة تقدّم أقصى معلومات عندما تتساوى معلمة الصعوبة مع معلمة القدرة ($b= \theta$)، مع افتراض اقتراب معلمة التخمين من الصفر.

وفي ضوء نتائج الدراسة الحالية توصي الدراسة بما يلي:

1.تجنب استخدام عينات ذات توزيعات ملتوية عند بناء الاختبارات في ضوء نظرية الاستجابة للفقرة.

2.الاعتماد على دالة معلومات الاختبار كمعيار عند بناء الاختبارات الوطنية.

الدراسات المقترحة:

1.إجراء دراسة لتحديد درجة القطع للالتواء، التي يبدأ عندها التأثير على الخطأ المعياري لدالة المعلومات.

2.إجراء دراسة لمقارنة أثر شكل التوزيع على دالة المعلومات والثبات في ضوء النموذج اللوجستي المستخدم (أحادي المعلمة، ثنائي المعلمة، ثلاثي المعلمة).

3.إجراء دراسة للتحقق من أثر شكل التوزيع على الثبات باستخدام النماذج متعددة التدريج (Polytomous Models).

المصادر والمراجع

إدارة الامتحانات والاختبارات. (2019). *التقرير الإحصائي للاختبار الوطني لضبط نوعية التعليم في الأردن*. عمان، الأردن: وزارة التربية والتعليم.

بني عطا، ز.، والشريفين، ن. (2012). أثر اختلاف شكل توزيع القدرة على معالم الفقرة ودالة معلومات الاختبار. *المجلة الاردنية في العلوم التربوية*، 8(2)، 151- 166.

الحواري، أ. (2015). أثر طول الاختبار وشكل توزيع القدرة في تقديرات قدرة الأفراد وفق نموذج راش في نظرية استجابة الفقرة. *مجلة جامعة النجاح للأبحاث و العلوم الإنسانية*، 29(8)، 1463- 1488.

الرقب، ي. (2011). *أثر الابتعاد عن التوزيع الطبيعي لدرجات الأفراد على اختبار قدرة عقلية في تقديرات معالم فقراته، رسالة دكتوراة غير منشورة*، جامعة اليرموك، إربد، الأردن.

الصبح، ع. (2021). *المقارنة بين دقة تقدير القدرة بإختلاف طول الإختبار وشكل توزيع معلمة القدرة تبعاً للنموذج اللوجستي ثلاثي المعلمة بإستخدام بيانات حقيقية ومولدة. مجلة جرش للبحوث والدراسات*، 22(2)، 1219- 1250.

الصرايرة، ر. (2016). أثر عدد نقاط التربيع في القيمة القصوى لدالة معلوماتية الفقرة والثبات الامبريقي والنظري للاختبار. دراسات: العلوم التربوية، 43(1)، 89- 101.

علام، ص. (2005). *نماذج الاستجابة للمفردة الاختبارية أحادية البعد ومتعددة الأبعاد وتطبيقاتها في القياس النفسي والتربوي*. القاهرة: دار الفكر العربي.

محمود، س. (2017). أثر شكل توزيع القدرة على ملائمة المفردات ودقة تقدير معلم الصعوبة في نموذج راش. *دراسات عربية في التربية وعلم النفس*، 2(86)، 543- 571.

References

Alwan, A. M., & Jasim, K. J. (2022). The Effect of the Difference in the Distribution of the Level of Ability that is Skeweded Positive for the Parameters of the Items of the Mental Ability Test According to the Item Response Theory. *International Journal of Early Childhood Special Education*, 14(1), 1150-1160.

Baker, F. B. (2001). *The Basic Item Response Theory*. ERIC.

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. Holt, Rinehart and Winston, 6277 Sea Harbor Drive, Orlando, FL 32887.

Corder, W. & Foreman, I. (2014). *Nonparametric statistics A Step-by-Step Approach*. (2nd ed.). Wiley.

Doran, H. C. (2005). The information function for the one-parameter logistic model: Is it reliability. *Educational and Psychological Measurement*, *65*(5), 665-675.

Embretson, S. E., & Reise, S. P. (2013). *Item response theory*. Psychology Press.

Feldt, L. S. (1980). A test of the hypothesis that Cronbach's alpha reliability coefficient is the same for two tests administered to the same sample. *Psychometrika*, 45(1), 99–105.

Hambleton, R. K., & Jones, R. W. (1994). Item Parameter Estimation Errors and Their Influence on Test Information Functions. *Applied Measurement in Education*, 7(3), 171–186.

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

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory* (Vol. 2). Sage.

Lord, F.M. (1980). *Application of Item Response Theory to Practical Testing Problems*. Hillsdale, NJ: Erlbaum.

Ngwane, F. F (2021). Effects of Item Parameter Distributions on Test Information Function. *International Journal of Mathematics And its Applications*, 9(1),41–50.

Oosterwijk, P. R., van der Ark, L. A., & Sijtsma, K. (2019). Using confidence intervals for assessing reliability of real tests. *Assessment*, *26*(7), 1207-1216.

Sass, D. A., Schmitt, T. A., & Walker, C. M. (2008). Estimating Non-Normal Latent Trait Distributions within Item Response Theory Using True and Estimated Item Parameters. *Applied Measurement in Education*, 21(1), 65–88.

Yuan, K.-H., & Bentler, P. M. (2002). On robustness of the normal-theory based asymptotic distributions of three reliability coefficient estimates. *Psychometrika*, 67, 251-259.

Zimowski, M. F., Muraki, E., Mislevy, R. J., & Bock, R. D. (2003). BILOG-MG [computer software]. *Lincolnwood, IL: Scientific Software International*.