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

Deanship of Scientific Research

Measurement Invariance in the Graduate Record Examination: A Language-Based Analysis

Abd-Alelah Naji AL-Khawaldeh[1]*, Yousef Mohammad Sawalmeh[2]

[1] Ministry of Education, Amman, Jordan

[2] Department of Counseling and Educational Psychology, Faculty of Education, Yarmouk University, Irbid, Jordan

Received: 3/1/2022
Revised: 29/1/2022
Accepted: 7/2/2022
Published: 15/6/2023

* *Corresponding author:*
abdkhf18@gmail.com

Citation: Al-Khawaldeh, A.-A. N. ., & Sawalmeh, Y. M. . (2023). Measurement Invariance in the Graduate Record Examination: A Language-Based Analysis. *Dirasat: Educational Sciences*, *50*(2): 104-116. https://doi.org/10.35516/edu.v50i2.305

© 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: The study aimed to examine the measurement invariance in the GRE test for quantitative reasoning administered in both Arabic and English.

Method: A cluster random sample of 541 male and female MA students was selected from the University of Jordan, Yarmouk University, and Mutah University. Of these, 273 students (both male and female) took the test in Arabic, while 268 students (both male and female) took it in English. The data were analyzed using both confirmatory factor analysis (CFA) and multi-group conformity factor analysis (MGCFA) through the statistical program AMOS 24. The fit of the data to the factorial model was assessed using four indicators: χ^2/df, CFI, RMSEA, SRMR.

Results: The results indicated that the data fit the four-factor model (arithmetic, algebra, geometry, and data analysis) for both the Arabic and English versions of the exam. However, the results also showed that while configural invariance was achieved, metric and scalar invariances were not achieved based on the language of the exam.

Conclusions: The quantitative reasoning exam presented in both English and Arabic versions had the same structure as assumed in the factorial model. However, the loadings on the factors and the intersection between the factors differed between the two versions. Preference was given to the Arabic version of the exam because it demonstrated a good fit with the factorial structure model, and proficiency in the English language did not significantly influence performance. Therefore, using the Arabic version with students whose mother tongue is Arabic allows for fair and unbiased comparisons between them.

Keywords: Graduate record examination (gre) test, quantitative reasoning, measurement invariance, confirmatory factor analysis, multi-group confirmatory factor analysis.

اللاتغير في القياس في امتحان تقييم الخريجين GRE باختلاف لغة الامتحان

عبدالإله ناجي الخوالدة*[1]، يوسف محمد سوالمة[2]

[1] وزارة التربية والتعليم، عمان، الأردن.

[2] قسم علم النفس الإرشادي والتربوي، كلية التربية، جامعة اليرموك، الأردن.

ملخّص

الأهداف: هدفت الدراسة إلى فحص اللاتغير في القياس في امتحان GRE الخاص بالاستدلال الكمي المقدم باللغتين العربية والإنجليزية.

المنهجية: تكونت عينة الدراسة من (541) طالباً وطالبة من طلبة الماجستير في الجامعة الأردنية وجامعة اليرموك وجامعة مؤته، تم اختيارهم بالطريقة العنقودية العشوائية؛ تقدم (273) طالباً وطالبة للامتحان بصورته العربية، و(268) للامتحان بصورته الإنجليزية. وقد تم تحليل استجاباتهم على (20) فقرة في كل من صورتي الامتحان، باستخدام التحليل العاملي التوكيدي (CFA)والتحليل العاملي التوكيدي متعدد المجموعات (MGCFA) في البرنامج الإحصائي .(AMOS 24) وقد تم الحكم على مطابقة البيانات للنموذج العاملي باستخدام أربعة مؤشرات، وهي: χ^2/df, CFI, RMSEA, SRMR.

النتائج: أظهرت النتائج مطابقة البيانات لنموذج العوامل الأربعة (الحساب، والجبر، والهندسة، وتحليل البيانات) لكل من الامتحان المقدم باللغة العربية، واللغة الإنجليزية. كما أظهرت النتائج تحقق اللاتغير التكويني، وعدم تحقق كل من اللاتغير المتري والعددي بحسب لغة الامتحان.

الخلاصة: امتحان الاستدلال الكمي المقدم باللغتين الإنجليزية والعربية لهما نفس شكل البناء العاملي المفترض، لكن لم يكن لهما نفس التشبعات على العوامل، والتقاطعات بين العوامل. وقد كانت الأفضلية لصورة الامتحان المقدم باللغة العربية؛ لأنها تحقق مطابقة أفضل مع النموذج العاملي للامتحان، و لأن درجة اتقان اللغة الإنجليزية لا تمثل عامل مؤثر في الأداء عليها، لذلك فإن استخدامها مع الطلبة الذين تكون لغتهم الأم اللغة العربية يحقق مقارنات عادلةٍ وصادقةٍ بينهم.

الكلمات الدالة :امتحان تقييم الخريجين (GRE) ، الاستدلال الكمي، اللاتغير في القياس، التحليل العاملي التوكيدي، التحليل العاملي التوكيدي متعدد المجموعات.

المقدمة

يتطلب أي نوع أو مستوى من التعليم استعدادات وقدرات معينة، وقد أصبحت الحاجة إلى قياس هذه القدرات والاستعدادات ضرورة ملحة؛ فكثير من مؤسسات التعليم العالي في مختلف دول العالم، تستخدم بعض اختبارات الاستعداد لأغراض القبول أو تقديم المنح الدراسية. ومع التطورات الكبيرة والسريعة في الأساليب والبرامج الإحصائية، ازداد الاهتمام بالتحقق من الخصائص السيكومترية للاختبارات، وبالأخص الاختبارات العالمية، التي يزداد عدد الاشخاص المتأثرين بالقرارات المبنية على نتائجها نتيجةً لاختلاف لغة الفرد عن لغة الاختبار.

ويعرّف ريفيز (Reeves, 2002) الاستعداد بأنه قدرة فطرية أو مكتسبة على تعلم عمل، أو القابلية للقيام بنشاط معين. ويعرّف اختبار الاستعداد بأنه أداة لقياس قابلية وإمكانية الفرد للقيام بسلوك غير مرتبط بتعلم أو تدريب معين؛ وذلك بهدف التصنيف أو الانتقاء للالتحاق ببرنامج معين (جروان، 2008). ويطلق مصطلح اختبارات الاستعداد الدراسي Scholastic Aptitude على الاختبارات التي تقيس مجموعة من القدرات الضرورية للنجاح الدراسي (Marlene et al., 1978). .

ويعد امتحان تقييم الخريجين (GRE) واحدا من اختبارات الاستعداد. وهو امتحان محوسب ومقنن ومعتمد، يمثل الأداء عليه شرطاً للقبول في العديد من كليات الدراسات العليا في الولايات المتحدة الأمريكية وكندا، وفي الكثير من البلدان الناطقة باللغة الإنجليزية، ولبرامج العمل في مختلف أنحاء العالم، وتشرف على إعداده وتطويره، وتطبيقه مؤسسة خدمة الاختبارات التربوية Educational Testing Service (Wendler et al, 2014; [ETS], 2013, 2014).

ويهدف امتحان (GRE) لتقييم الطلاب المتقدمين لمختلف برامج الدراسات العليا، من خلال تقييم مختلف المعارف والمهارات، كالحساب، والجبر، وتحليل البيانات، ومخزون المفردات للطالب الجامعي، كما تقيس القدرة على تحليل وتقييم النصوص المكتوبة، ومهارات التفكير الناقد وحل المشكلات؛ إذ أن الحصول على علامة عالية في الامتحان يعطي فرصة أكبر للقبول (ETS, 2013, 2014; Wendler, et al, 2014) .

ويتكون امتحان (GER) من ثلاثة امتحانات فرعية، هي: الكتابة التحليلية (Analytical Writing)، الاستدلال اللفظي/ اللغوي (Verbal Reasoning)، الاستدلال الكمي/المنطقي (Quantitative Reasoning) (ETS, 2013, 2014). وتشير مؤسسة خدمات الاختبارات التربوية (ETS, 2015) إلى أنه من الممكن: أن يكون الامتحان متاحاً بأكثر من لغة، على أن يقيس البناء نفسه، ويتحقق ذلك عند توافر بيانات ومعلومات كافية عن صدق نتائج الاختبار باللغات المتنوعة. أو من الممكن توفير مترجم فوري يجيد اللغة الإنجليزية واللغة المستهدفة، ولديه معرفة بالاختبار.

إن العديد من أدوات القياس والاختبارات التي تم إعدادها في أوروبا، والولايات المتحدة الأمريكية، تم نقلها وتطبيقها على الثقافات الأخرى دون أن يتم تطويرها وتقنينها وتكييفها (Cheung, 2012). وشهد مجال تطوير وتقنين الاختبارات باللغات والثقافات المختلفة تقدماً ملحوظاً في السنوات الـ25 الماضية، نتيجة انتقال تطبيقات علم النفس عبر الثقافات المختلفة، واستخدام الامتحانات في الاعتماد، وزيادة الاهتمام بدراسات المقارنة عبر الدول، والتركيز على الموضوعية والإنصاف والعدالة في الاختبارات، والأخذ بالاعتبار السماح للمتقدمين لتلك الاختبارات باختيار اللغة التي يتم تقديم الاختبار بها، مثل اختبار القبول في الجامعات (International Test Commission [ITC], 2017).

ويشير بوتشر (Butcher, 2006) إلى أن الفقرات نفسها قد يكون لها معانٍ مختلفةٍ من ثقافةٍ إلى أخرى. الأمر الذي يدعو إلى استكشاف الاختلافات الثقافية التي يمكن أن تؤثر في درجات الاختبارات؛ بحيث لا نكتفي بنقل الاختبار وترجمته وتطبيقه مباشرةً.

وبما أن عملية تطوير الاختبارات تنطوي على القيام بتعديلات على الاختبار، وذلك لجعل الاختبار مناسباً للبيئة الثقافية الجديدة، الأمر الذي يثير تساؤلاً مهماً حول مدى تكافؤ الصورة الأصلية مع الصورة المطورة؟ وينبثق من هذا التساؤل مجموعة من التساؤلات الفرعية. هل الصورة المطورة تقيس البناء المفاهيمي نفسه؟ وبنفس الطريقة؟ (كريش، 2015). ولا بد من تقديم أدلة على صدق وثبات الاختبارات التي يجري تطويرها بلغات أو ثقافات أخرى؛ لأن الأدلة على صدق الاختبار وثباته بلغة أو ثقافة معينة، قد لا تنطبق على الاختبار نفسه بلغة أو ثقافة أخرى، كما أنه يجب توفير أنواع مختلفة من الأدلة على صدق الاختبار (AERA, APA & NCME, 2014).

ويعرف فاندنبرغ ولانس (Vandenberg and Lance, 2000) اللاتغير في القياس (Measurement Invariance (MI، أو تكافؤ القياس (Measurement Equivalence) أو اللاتغير العاملي (Factor Invariance) بأنه خاصية إحصائية للقياس تشير إلى أن البنية العاملية نفسها يتم قياسها عبر مجموعات محددة، فعلى سبيل المثال يمكن استخدام اللاتغير في القياس للتحقق مما إذا تم تفسير إجراء معين بالطريقة نفسها من المستجيبين على اختلاف جنسهم أو خلفيتهم الثقافية. ويميز فان دي فيجفر وتانزر (Van de, Vijver, and Tanzer, 2004) بين ثلاث فئات لمصادر التحيز التي تؤثر على القياس عبر المجموعات: تحيز البناء (مثل التداخل غير الكامل في تعريفات السمة المقاسة عبر المجموعات، والملاءمة التفاضلية للسمة المقاسة عبر المجموعات)، وتحيز الأسلوب (مثل التأثيرات التفاضلية لأنماط الاستجابة، وتأثيرات المجموعة المرجعية، والتمييز التفاضلي للعينات)، وتحيز الفقرة (مثل الترجمة الضعيفة، وصدق البناء التفاضلي). وبناء على ذلك، فقد أكد فشر وبورتنجا (Fisher & Poortinga, 2018) في عملية فحص اللاتغير في القياس على استخدام تصميم يستبعد التفسيرات بدلالة المتغيرات الدخيلة أو المختلطة (confounding variables) التي

تنتج التحيز وعدم التكافؤ في القياس.

ويشير هورن ومكاردل (Horn & Mcardle, 1992) إلى أن السؤال العام عن اللاتغير في القياس هو ما إذا كانت عمليات القياس في ظل ظروف مختلفة تعطي قياسات للسمة نفسها أم لا. وقد أكدا أنه في غياب دراسة اللاتغير في القياس أو عندما يتوفر دليل على عدم تحقق اللاتغير في القياس فإنه لا يمكن تفسير الفروق في نتائج عملية القياس بين الأفراد والجماعات في تلك الظروف بشكل واضح. ويعني اللاتغير في القياس أن استخدام المقياس نفسه في مجموعات مختلفة (مثل بلدان مختلفة أو أزمان مختلفة أو ظروف مختلفة) يقيس البناء نفسه وبالطريقة نفسها (Daridov et al., 2014). وفي هذا إشارة واضحة على أن الاختلاف بين الظروف المختلفة لا يقتصر على صفة خاصة بمجموعات الأفراد. فقد تكون لها علاقة بنمط الاستجابة لفقرات الامتحان (استجابة ورقية أو الكترونية) أو لغة الامتحان أو وقت التطبيق أو أي متغير دخيل آخر.

يرى كيم وساساكي (Kim & Sasaki, 2017) أنه من المهم أخذ اختلافات اللغة في الاعتبار عند عمل مقارنات بين مجموعات ثقافية مختلفة في المقاييس المقدمة لتلك المجموعات. ويتم ذلك في بعض الامتحانات الدولية (امتحان PISA مثلا) من خلال عمل صور متعددة للامتحان تختلف عن بعضها في اللغة فقط حيث يتم ترجمة النسخة الإنجليزية إلى اللغات المختلفة. ويقدم الامتحان باللغة العربية للعرب وباللغة الإنجليزية في الدول التي تتكلم الإنجليزية وباللغة التركية في تركيا وهكذا لبقية اللغات. ويتم استخدام النتائج للدول المختلفة في دراسة اللاتغير في تلك الامتحانات (Sekercioglu & Kogar, 2018). لكن هناك امتحانات دولية (امتحان GRE) تقدم لكل المفحوصين باللغة الإنجليزية دون مراعاة للغتهم الأم. أي لدينا مجموعات من المفحوصين الذين يختلفون في اللغة. في هذه الحالة، هل ما يقيسه الامتحان لدى العرب يختلف عما يقيسه لدى الذين لغتهم هي الإنجليزية؟

لإجابة هذا السؤال لا بد من دراسة اللاتغير في امتحان GRE بنسخته الإنجليزية المطبقة. وهذا يتطلب وجود مجموعتين بأعداد مناسبة (مجموعة عربية ومجموعة لغتها الأم هي الإنجليزية) وهذا ما لا يمكن توفيره عمليا لأي مرة تطبيق للامتحان خاصة ما يتعلق بالمجموعة العربية. ويرى كيم وساكس (Kim & Sasaki, 2017) أنه يمكن التغلب على هذه المشكلة من خلال التركيز على مجموعتين ثنائيتي اللغة واستخدام نسختين مختلفتين في اللغة لنفس المقياس بحيث تطبق أحد النسختين على المجموعة الأولى وتطبق النسخة الأخرى على المجموعة الثانية. هذا وقد سبق أن درس سوسا وآخرون (Sousa et al., 2012) اللاتغير في القياس بين نسختين (اسبانية وإنجليزية) لاستبانة جودة الحياة لدى الأطفال الذين لديهم أزمة(ربو) من خلال تطبيق كل منهما على عينة من أصول اسبانية. ويُعد التحقق من تكافؤ بنية الاختبار عبر إصداراته باللغات المتعددة، الأساس لإجراء المقارنات الصحيحة سواء على مستوى الأفراد أو المجموعات. وينبغي توفير الأدلة على التكافؤ من خلال البيانات التجريبية، باستخدام عينات من المفحوصين ثنائيي اللغة (ITC, 2017).

ويصنف اللاتغير في القياس إلى ثلاثة مستويات هرمية؛ يمثل كل مستوى منها جانباً واحداً من العلاقة بين البناء الكامن والمشاهدات الملاحظة، ويكون السابق فيها شرطاً للاحق، وهي مرتبه من الأقل تعقيداً إلى الأكثر تعقيداً، كالآتي:

أولاً: اللاتغير التكويني (Configural Invariance): ويعني أن البناء العاملي للاختبار (عدد العوامل ونمط العلاقات بين العوامل والمؤشرات) هو نفسه عبر المجموعات المختلفة في التحليل العاملي التوكيدي متعدد المجموعات (MGCFA).

ثانياً: اللاتغير المتري (Metric Invariance): ويعني أن التشبعات للعوامل متشابهة عبر المجموعات؛ إي أنه نموذج تكويني مع تساوي أوزان المؤشرات عبر المجموعات، ويشير إلى ثبات كمي لنموذج القياس للبناء الكامن عبر المجموعات.

ثالثاً: اللاتغير العددي (Scalar Invariance): تكون قيم الأوساط والتباينات متشابهة أيضاً عبر المجموعات، ويشير إلى ثبات مستويات المتوسط عبر المجموعات المختلفة، وهو بذلك يتجاوز حاجز المستوى المتري بافتراضات التكافؤ عبر المجموعات (Ariely & Davidov, 2012; Sinkovics et al., 2016; Steenkamp & Baumgartner, 1998).

ويعتبر كلاين (Kline, 2015) التحليل العاملي التوكيدي متعدد المجموعات (M-GCFA) أحد أكثر الاستراتيجيات شيوعاً لفحص اللاتغير في القياس (MI)، فهنالك العديد من الدراسات التي استخدم فيها التحليل العامل التوكيدي لتقييم إذا ما كان البناء العاملي للاختبار الأصلي هو نفسه في الثقافات واللغات المختلفة. ويشير إلى أهم مؤشرات المطابقة التي أثبتت الدراسات التقويمية تفوقها على المؤشرات الأخرى، بالإضافة إلى مؤشر مربع كاي، وهي: الجذر لوسط مربع خطأ التقريب(Root Mean Square Error of Approximation) (RMSEA)، ومؤشر المطابقة المقارن (CFI) (Comparative Fit Index)، والجذر المعياري لوسط مربع البواقي (Root Mean Square Residual)(SRMR).

وتتحقق المطابقة عندما تكون قيمة مؤشر (χ^2/df) لا تزيد على 3 (Marsh & Hocevar. 1985)، وقيمة مؤشر (CFI) لا تقل عن 0.80 (Hair, et al., 1995)، وقيمة مؤشر (RMSEA) لا تزيد على 0.080، (Brown and Cudeck. 1993)، وقيمة مؤشر (SRMR) لا تزيد على 0.090 (Hu and Bentler. 1999).

ويجري التحقق من خاصية اللاتغير المتري والعددي اعتماداً على الفرق في قيمة مؤشرات المطابقة، ويفضل أن يقل الفرق في قيمة مؤشر المطابقة المقارن (ΔCFI) عن 0.01، والفرق في قيمة مؤشر جذر وسط مربع الخطأ التقريبي (ΔRMSEA) عن 0.015، والفرق في قيمة مؤشر الجذر المعياري لوسط البواقي (ΔSRMR) عن 0.03 (Chen, 2007).

وقد حظيت دراسة البناء العاملي لكثير من الاختبارات الدولية والوطنية باهتمام الدراسات. فقد أجرى كونيسل وآخرون (Kuncel, et al., 2010) دراسة للتحقق من التفاضل في التنبؤ باستخدام امتحانات (GRE) بأداء الطلاب المسجلين في برنامج الماجستير، والدكتوراه، وتكونت عينة الدراسة من (10000) طالب وطالبة. وخلصت الدراسة بشكل عام إلى أن امتحانات (GRE) صالحة للتنبؤ بالنجاح بكل من برنامج الماجستير والدكتوراه، وهذا دليل على الصدق التنبؤي لها، ويمكن استخدام درجات الامتحانات لاتخاذ قرار القبول. كما وأجرى بروير وآخرون (Broer, et al., 2005) تقييما لعدة طرائق في الكشف عن (DIF) في امتحان الكتابة التحليلية وفقاً للعرق واللغة. وتم استخدم استجابات 397806 من المتقدمين للامتحان بين (2002-2003)، وقد تم استخدام ثلاثة أساليب، هي: اختبار مانتل هانزل، والانحدار اللوجستي، واحصائي القياس المعياري لتحليل الأداء التفاضلي في الاستجابات المتعددة (المتدرجة)، وقد أظهرت النتائج أنه لم يتم العثور على قيم (DIF) كبيرة؛ وكان هنالك اتفاق عالٍ بين الطرق الثلاثة. كما أظهرت قيماً متوسطة عند المقارنة بين الذين اعتبروا اللغة الإنجليزية لغتهم المفضلة والذين لم يعتبروها لغتهم المفضلة، ولصالح الذين كانت اللغة الإنجليزية لغتهم المفضلة. وتوفر النتائج دعماً إضافياً لاستمرار مثل هذه المراجعات كجزء من عملية تطوير الاختبار.

وتحقق هي وآخرون (He, et al., 2014) من مستوى تكافؤ القياس في اختبارات العلوم الوطنية للمرحلة الأساسية (KS2) للعام 2010م و2011م وفق مجموعة من المتغيرات: منها اللغة الإنجليزية (لغة أولى أو ثانية)، تم استخدام التحليل العامل الاستكشافي، ونموذج راش لتحليل بيانات الفقرات للكشف عن الأداء التفاضلي (DIF) للتحقق من طريقة عمل فقرات الاختبارات عبر المجموعات الفرعية. وأظهرت النتائج أن الطلبة الذين يتحدثون اللغة الإنجليزية كلغة ثانية كان أدائهم أقل من الذين يتحدثون اللغة الإنجليزية كلغة أولى، كما كشف التحليل العاملي الاستكشافي نمط مماثل عبر المجموعات الفرعية والعينة الإجمالية في العامين، كما أظهرت مستوى عالٍ من اللاتغير في القياس في معالم الفقرات عبر المجموعات.

وعلى الصعيد العربي فقد أجرى تساوسيز(Tsaousis, 2015) دراسة للتحقق من البناء العاملي للاختبار التحصيلي المعد من قبل المركز الوطني للقياس والتقويم (SAAT) للقبول في التعليم العالي في المملكة العربية السعودية ، والذي تقدم له (63380) طالباً وطالبة، حيث تم التحقق من ثلاثة نماذج عاملية للاختبار، فقد أظهرت مطابقة للنماذج مع البيانات، ووفَّرت أدلة على صدق البناء العاملي للاختبار، كما أظهرت النتائج تحقق اللاتغير في القياس عبر الجنس.

وفي الأردن قام السمكي (2017) بفحص اللاتغير في القياس عبر متغير موقع المدرسة وجنس الطالب، لاختبار تمس (TIMSS) لعام 2011م في الرياضيات والعلوم للصف الثامن الأساسي في الأردن، باستخدام التحليل العاملي التوكيدي (CFA) والتحليل العاملي التوكيدي متعدد المجموعات (MGCFA)، في برنامج (AMOS 22). أظهرت النتائج ملاءمة نموذج العوامل الأربعة المفترض لاختبار الرياضيات، ونموذج العوامل الأربعة المفترض لاختبار العلوم، كما أظهرت تحقق لاتغير تكويني، ومتري، وعددي في فقرات اختبار كل من الرياضيات والعلوم عبر متغير موقع المدرسة (مدينة، ريف)، ومتغير جنس الطالب (ذكر، أنثى)، ووجود أفضلية للإناث على الذكور، ولمدارس المدن على مدارس الريف. كما أجرت أبو عواد (2018) دراسة لاستقصاء دلالات الصدق والثبات لمقياس الأساليب المعرفية لدى طلبة الجامعة الأردنية، والذي تم تعريبه للبيئة الأردنية. وقد كشفت نتائج التحليل العاملي أربعة عوامل تفسر ما نسبته (45.7%) من التباين، وأظهرت وجود اختلاف بين البناء العاملي للصورة الأردنية للمقياس، والبناء العاملي للصورة الأجنبية الأصلية.

وهكذا فإن الدراسات السابقة تبين أن اختلاف لغة الامتحان عن لغة المتقدم تؤثر سلبا على أدائه، فالذين يتحدثون اللغة الإنجليزية لغة ثانية كان أداؤهم أقل في الامتحانات المقدمة في اللغة الإنجليزية من الطلبة الذين يتحدثون اللغة الإنجليزية لغة أولى (He, et al., 2014). وقد تحقق اللاتغير في القياس للاختبارات التي تقدم باللغة نفسها عبر مجموعة من المتغيرات (السمكي،2017 ;Tsaousis, 2015)، في حين لا يتحقق اللاتغير في القياس عند تقديم المقياس بلغة جديدة تختلف عن لغته الأصلية (أبو عواد، 2016). لذلك جاءت الدراسة الحالية لفحص اللاتغير في القياس في امتحان(GRE) وفق متغير اللغة(العربية، الإنجليزية) حيث لم يسبق دراستها _ في حد علم الباحثين.

مشكلة الدراسة

برزت مشكلة هذه الدراسة من أن امتحانات تقييم الخريجين (GRE) تقدم باللغة الإنجليزية لخريجي الجامعات العربية، حيث تعد شرطاً للقبول في برامج الدراسات العليا في العديد من الجامعات في الولايات المتحدة الأمريكية وكندا، وفي الكثير من البلدان الناطقة باللغة الإنجليزية، وكذلك للعمل في برامج مختلفة في أنحاء العالم، وقد بلغ عدد الأردنيين المتقدمين لامتحانات (GRE) (475)، أما العرب فقد بلغ عددهم تقريباً (8476) خلال الفترة الزمنية 2013-2016م (ETS, 2016)؛ وما يزال يتقدم لها الكثير من العرب الراغبين باستكمال دراستهم في الدول الغربية بنسختها الأصلية باللغة الإنجليزية، أي بلغة تختلف عن لغتهم الأم . الأمر الذي قد يجعل هذه الاختبارات متحيزة ضدهم. ليشكل ذلك دافعاً لدراسة فيما إذا كان لاختلاف اللغة التي يقدم بها امتحان الاستدلال الكمي (GRE) أثر على اللاتغير في القياس (MI)؛ فامتحان الاستدلال الكمي يقيس مهارات كمية أساسية من المفترض أنها لا تتأثر باللغة، كما أن الدرجات عليه لها دور مهم في قبول الطلبة في التخصصات والمجالات التي تتطلب مهارات رياضية أو في الحصول

على منحة دراسية. ولهذا أهمية في بناء نسخ معربة لأغراض القبول في الجامعات العربية أو حث مؤسسة خدمات الاختبارات التربوية Educational Testing Services (ETS) على تقديم الاختبار بلغات مختلفة لتجنب ما قد يلحقه من تحيز بسبب اللغة خاصة، وأنه تم الإشارة فيما سبق أن لديها مثل هذا التوجه.

لذلك، ظهرت الحاجة لدراسة اللاتغير في القياس (MI) مع اختلاف اللغة التي يقدم بها الامتحان عن اللغة الأم للمتقدم له، وذلك للتوصل إلى تفسيرات واضحة ومنطقية للنتائج في امتحان الاستدلال الكمي وفق متغير اللغة، لذا سعت الدراسة للإجابة عن الأسئلة الآتية:

1. ما مدى مطابقة البيانات لنموذج العوامل الأربعة المقترح لمحتوى امتحان الاستدلال الكمي (الحساب، والجبر، والهندسة، وتحليل البيانات) من امتحانات (GRE)، للصورة المقدمة باللغة العربية؟
2. ما مدى مطابقة البيانات لنموذج العوامل الأربعة المقترح لمحتوى امتحان الاستدلال الكمي (الحساب، والجبر، والهندسة، وتحليل البيانات) من امتحانات (GRE)، للصورة المقدمة باللغة الإنجليزية؟
3. هل يتحقق اللاتغير في القياس لامتحان الاستدلال الكمي من امتحانات (GRE)، باختلاف اللغة التي يقدم بها؟

أهمية الدراسة

تبرز أهمية الدراسة في الكشف على خاصية اللاتغير في القياس لامتحان الاستدلال الكمي في اختبارات تقييم الخريجين (GRE) وفق متغير اللغة التي يقدم فيها، وذلك بتوفير أدلة على الصدق البنائي لامتحان تقييم الخرجين (GRE) باختلاف اللغة التي يقدم بها من خلال التحقق من البنية العاملية للامتحان في الأردن؛ نظراً لأنّ نتائجه تستخدم لاتخاذ قرارات بالغة الأهمية، تتمثل في الحصول على قبول ومنح للدراسات العليا.

وتشكل الدراسة محاولة تساعد الجامعات الأردنية للنظر في تطوير مثل تلك الامتحانات واعتمادها محكاً للقبول في برامج الدراسات العليا التي تقدمها. كما ستوفر نتائج ومعلومات يمكن استخدامها والاستفادة منها من مؤسسة خدمات الاختبارات التربوية (ETS) في تطوير الاختبار، وتفسير النتائج وبناء صور من الاختبار بلغات متعددة مراعاة للمتقدمين له من ثقافات مختلفة.

مصطلحات الدراسة

- **اللاتغير في القياس (Measurement Invariance):** يعرف بأنه خاصية إحصائية للقياس تشير إلى أن البنية العاملية نفسها يتم قياسها عبر بعض المجموعات (Vandenberg & Lance, 2000).
- **امتحان الاستدلال الكمي (Quantitative Reasoning):** هو أحد امتحانات (GRE)، تقييم المهارات الكمية الأساسية، والتي تشمل: الحساب، الجبر، الهندسة، وتحليل البيانات.
- **صدق البناء الداخلي:** الأدلة على اتساق البيانات المستمدة من المقياس مع الإطار المفاهيمي للمكونات البنائية للمقياس، والعلاقات فيما بينها.

محددات الدراسة

اقتصرت الدراسة على امتحان الاستدلال الكمي (Quantitative Reasoning) من امتحانات (GRE). وعلى عينة عشوائية عنقودية من طلبة الماجستير في الجامعة الأردنية، وجامعة اليرموك، وجامعة مؤتة.

منهجية الدراسة

اتبع المنهج الوصفي التحليلي من خلال جمع البيانات، والمعلومات، وتحليلها، وتفسيرها بأدق الطرائق العلمية؛ وذلك لتوفير أدلة على مدى تحقق خاصية اللاتغير في القياس بحسب اللغة التي يقدم بها امتحان الاستدلال الكمي كأحد امتحانات (GRE)، الذي يقدم في الأصل باللغة الإنجليزية، التي تختلف عن اللغة الأم لكثير من المتقدمين للامتحان.

مجتمع الدراسة

تكون مجتمع الدراسة من جميع طلبة الماجستير في جامعة اليرموك، والجامعة الأردنية، وجامعة مؤتة خلال الفصل الدراسي الأول للعام 2021/2020، والبالغ عددهم (11186) طالباً وطالبةً. منهم (4467) من الذكور، (6719) من الإناث.

عينة الدراسة

تم الحصول على استجابة (565) طالب وطالبة من أفراد مجتمع الدراسة، تم اختيارهم بالطريقة العشوائية العنقودية، حيث تم اختيار (27) شعبة من الكليات الإنسانية، و(27) شعبة من الكليات العلمية، ثم تم توزيع الامتحان على الطلبة عشوائياً داخل تلك الشعب؛ بحيث يعطى طالب الصورة العربية من الامتحان، والذي يليه الصورة الانجليزية من الامتحان، وبعد حجم العينة مناسب لإجراء التحليلات الإحصائية المستخدمة، كان منها (283) استجابة على الامتحان باللغة العربية، و(282) استجابة على الامتحان باللغة الإنجليزية, ويبين الجدول (1) توزيع أفراد العينة حسب لغة الامتحان، وجنس الطالب.

الجدول (1): توزيع أفراد العينة حسب لغة الامتحان والجنس.

لغة الامتحان	الجنس		المجموع
	ذكور	إناث	
اللغة العربية	96	187	283
اللغة الإنجليزية	105	177	282
المجموع	201	364	565

أداة الدراسة

استُخدم في الدراسة القسم الأول من امتحان الاستدلال الكمي لكونه أحد نماذج امتحانات (GRE) المنشورة (Varsity Tutors, 2019) ، و يهدف إلى قياس القدرة على الاستدلال الكمي من خلال: قياس القدرات الرياضية الأساسية، عن طريق فهم المبادئ الرياضية، والقدرة على حلّ المسائل بالطرائق التي تم دراستها في علم الحساب، والجبر، والهندسة، وتحليل البيانات. ويتكون الامتحان بصورته الأصلية من 20 فقرة، يخصص لها 30 دقيقة، ضمن أربعة أنواع من الأسئلة:

- أسئلة المقارنة الكمية: تتطلّب القيام بالمقارنة بين كميتين (الكمية أ والكمية ب) ومن ثمّ تحديد أي من الجمل الأربعة الموجودة تصف المقارنة، وتمثل ما مجموعه 7 فقرات؛ أي ما نسبته 35% من الفقرات.
- أسئلة الاختيار من متعدّد – جواب واحد صحيح: تتطلّب القيام باختيار جواب واحد صحيح من قائمة من خمسة اختيارات، وتمثل ما مجموعه 9 فقرات؛ أي ما نسبته 45% من الفقرات.
- أسئلة الاختيارات المتعدّدة – جواب واحد صحيح أو أكثر: تتطلّب القيام باختيار جواب واحد أو أكثر من بين قائمة من الاختيارات. قد يحدّد السؤال عدد الاختيارات الصحيحة، وتمثل ما مجموعه فقرتان؛ أي ما نسبته 10% من الفقرات.
- أسئلة إدخال الأرقام: تتطلّب القيام إما بإدخال عدد صحيح أو عشري في صندوق واحد للإجابة، أو بإدخال كسر في صندوقين منفصلين – واحد للبسط (الرقم فوق خط الكسر)، وواحد للمقام (الرقم تحت خط الكسر)، وتمثل ما مجموعه فقرتان؛ أي ما نسبته 10% من الفقرات.

ويظهر كل سؤال إما بشكل منفصل أو كجزء من مجموعة من الأسئلة التي تعتمد على نفس البيانات المعروضة في جداول، أو صور أو أشكال أخرى لعرض البيانات.

وتتوزع الأسئلة على أربعة أبعاد (الحساب، الجبر، الهندسة، وتحليل البيانات)، والفقرات في بعد الحساب هي (2, 8، 11، 13، 15)، وفي بعد الجبر هي (1، 5، 12، 18، 20)، وفي بعد الهندسة هي (3، 6، 10، 16، 19)، وفي بعد تحليل البيانات هي (4، 7، 9، 14، 17).

تعريب الامتحان

تم ترجمة الامتحان وتعليماته من اللغة الإنجليزية إلى اللغة العربية، بمساعدة متخصصين في الترجمة والرياضيات، ثم تم إعادة الترجمة (الترجمة العكسية) من اللغة العربية إلى اللغة الإنجليزية للتحقق من دقة الترجمة. تم عرض الترجمة على محكمين ثنائيي اللغة متخصصين في اللغة العربية، واللغة الإنجليزية، والرياضيات، والهندسة؛ للتحقق من دقة الترجمة، وسلامة الصياغة اللغوية ووضوحها، تم تعديل بعض الأسماء التي لا تتناسب مع الثقافة والبيئة الأردنية.

متغيرات الدراسة

اللغة التي يقدم بها الاختبار، ولها فئتان (اللغة العربية، اللغة الإنجليزية).

المعالجة الإحصائية

للإجابة عن السؤالين الأول والثاني تم تحديد المشاهدات المتطرفة (Outliers) باستخدام طريقة (Mahalanobis distance)، وحذفها من التحليل. كما تم التحقق من افتراض التوزيع الطبيعي أحادي المتغيرات(Univariate Normality)، من خلال مؤشري التفلطح (Kurtosis)، والالتواء (Skewness)؛ فالقيم المطلقة لمعامل الالتواء التي تزيد على 2، والقيمة المطلقة لمعامل التفلطح التي تزيد على 7، تمثل مؤشراً على عدم تحقق افتراض التوزيع الطبيعي (West, et al., 1995)، وحذف المتغيرات التي لا تحقق ذلك الافتراض، ثم تم التحقق من افتراض التوزيع الطبيعي متعدد المتغيرات (Multivariate Normality) لاستجابات الأفراد من خلال اختبار ماردیا (Mardia's Test)، وذلك باستخدام برنامج (SPSS)، ثم أجرى التحليل العاملي التوكيدي (CFA) باستخدام برنامج (AMOS).

للإجابة عن السؤال الثالث، أجري التحليل العاملي التوكيدي متعدد المجموعات(Multi-group Confirmatory Factor Analysis (MGCFA باستخدام برنامج أموس (AMOS). والحكم على مطابقة البيانات للبناء العاملي باستخدام المؤشرات الآتية: مربع كاي/ درجات الحرية (χ^2/df)،

ومؤشر الجذر لوسط مربع خطأ التقريب (RMSEA)، ومؤشر المطابقة المقارن (CFI)، ومؤشر الجذر المعياري لوسط البواقي (SRMR).

نتائج الدراسة ومناقشتها

السؤال الأول: ما مدى مطابقة البيانات لنموذج العوامل الأربعة المقترح لمحتوى امتحان الاستدلال الكمي (الحساب، والجبر، والهندسة، وتحليل البيانات) من امتحانات (GRE)، للصورة المقدمة باللغة العربية؟

تم تحديد الاستجابات المتطرفة (Outliers) باستخدام طريقة (Mahalanobis distance) ومعالجتها بالحذف؛ حيث تم حذف 10 استجابات، كانت قيم مربع مسافة ماهالانوبس لها دالة إحصائياً عند مستوى الدلالة α = 0.01. ثم تم التحقق من افتراض التوزيع الطبيعي أحادي المتغيرات (Univariate Normality)، لاستجابات أفراد العينة على كل فقرة من فقرات امتحان الاستدلال الكمي المقدم باللغة العربية، وتبين أن الفقرات تحقق افتراض التوزيع الطبيعي. كما تم التحقق من افتراض التوزيع الطبيعي متعدد المتغيرات (Multivariate Normality) لاستجابات أفراد العينة على فقرات الامتحان المقدم باللغة العربية، باستخدام اختبار ماردیا (Mardia's Test)، ويبين الجدول (2) النتائج.

الجدول (2): نتائج اختبار (Mardia's Test) للتحقق من افتراض التوزيع الطبيعي للاستجابات على فقرات الامتحان المقدم باللغة العربية.

اختبار (Mardia's Test)	الإلتواء Skewness	التفلطح Kurtosis
قيمة الاختبار	87.0414	444.0714
χ^2	3974.8918	13.2052
sig	0.000	0.000

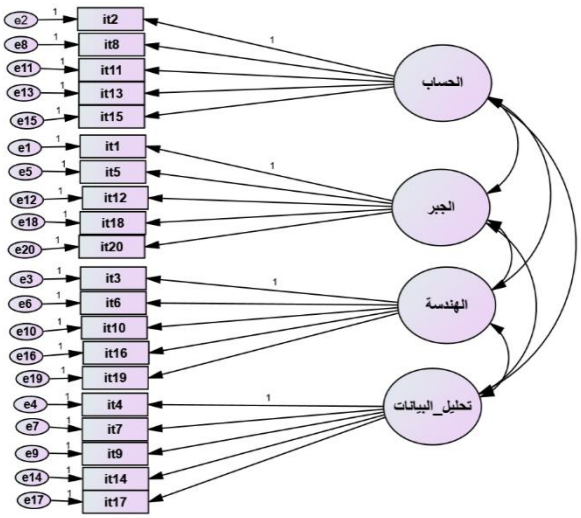
يتبين من الجدول (2) عدم تحقق افتراض التوزيع الطبيعي متعدد المتغيرات، ولمعالجة ذلك تم استخدام طريقة (BOOTSTRAP). وللتحقق من مدى مطابقة البيانات للنموذج المقترح تم استخدام التحليل العاملي التوكيدي، ويبين الجدول (3) النتائج.

الجدول (3): نتائج التحليل العاملي التوكيدي لفقرات الامتحان المقدم باللغة العربية.

المؤشر الإحصائي	χ^2	Df	χ^2/df	CFI	RMSEA	SRMR	sig
القيمة	*228.429	164	1.393	0.964	0.038	0.059	0.001

*** دالة إحصائياً عند مستوى الدلالة الإحصائية (α=0.05)**

يتبين من الجدول (3) وجود دلالة إحصائية لقيمة (χ^2)، وكانت قيم مؤشرات χ^2/df, CFI, RMSEA, SRMR تقع ضمن المدى المقبول لجودة المطابقة (Hu & Bentler, 1999; Marsh & Hocevar. 1985). وبتفحص قيم مؤشرات التعديل (Modification Indices)، تبين عدم وجود ارتباطات بين الأخطاء في النموذج المقترح المبين في الشكل (1).



الشكل (1): النموذج العاملي للامتحان المقدم باللغة العربية.

وتدل قيم هذه المؤشرات على مطابقة البيانات لنموذج العوامل الأربعة (الحساب، الجبر، الهندسة، تحليل البيانات) للامتحان المقدم باللغة العربية. وتشير هذه النتيجة إلى أن نموذج العوامل الأربعة يمثل وبشكلٍ واضح المكونات التي أعد امتحان الاستدلال الكمي بصورته الأصلية لقياسها، و ينبغي معرفتها عند هذا المستوى التعليمي؛ وقد تعزى هذه النتيجة إلى الطبيعة الكمية للامتحان التي يتوقع أن لا تختلف المفاهيم والعمليات فيها باللغة العربية وللأردنيين عنها في اللغة الإنجليزية ولغير الأردنيين. كما تعزى إلى دقة ترجمة الامتحان وتطبيقه وتصحيحه كما أوضحت دراسة .....، وإلى وجود ارتباطات موجبة دالةً إحصائياً بين أبعاد الامتحان الأربعة المفترضة المبينة في الجدول (4).

الجدول (4): معاملات الارتباط بين الأبعاد الأربعة لامتحان الاستدلال الكمي المقدم باللغة العربية

البعد	الحساب	الجبر	الهندسة
الجبر	0.579**		
الهندسة	0.352**	0.470**	
تحليل البيانات	0.437**	0.361**	0.398**

**** دالة إحصائياً عند مستوى الدلالة الإحصائية (α=0.01)**

السؤال الثاني: ما مدى مطابقة البيانات لنموذج العوامل الأربعة المقترح لمحتوى امتحان الاستدلال الكمي (الحساب والجبر والهندسة وتحليل البيانات) من امتحانات (GRE)، للصورة المقدمة باللغة الإنجليزية؟

تم تحديد الاستجابات المتطرفة (Outliers) باستخدام طريقة (Mahalanobis distance)، ومعالجتها بالحذف؛ حيث تم حذف 14 استجابة، كانت قيم مربع مسافة ماهالانوبس لها دالة إحصائياً عند مستوى الدلالة α =0.01. ثم تم التحقق من افتراض التوزيع الطبيعي أحادي المتغيرات (Univariate Normality)، لاستجابات أفراد العينة على كل فقرة من فقرات امتحان الاستدلال الكمي المقدم باللغة الإنجليزية، وتبين أن الفقرات تحقق افتراض التوزيع الطبيعي. كما تم التحقق من افتراض التوزيع الطبيعي متعدد المتغيرات (Multivariate Normality) لاستجابات أفراد العينة على فقرات الامتحان المقدم باللغة الإنجليزية، باستخدام اختبار ماردیا(Mardia's Test)، المبين في الجدول (5).

الجدول (5): نتائج اختبار(Mardia's Test) للتحقق من افتراض التوزيع الطبيعي متعدد المتغيرات لاستجابات أفراد العينة على فقرات الامتحان المقدم باللغة الإنجليزية.

اختبار(Mardia's Test)	الإلتواء Skewness	التفلطح Kurtosis
قيمة الاختبار	74.9842	470.9951
χ^2	3349.2934	8.5524
sig	0.000	0.000

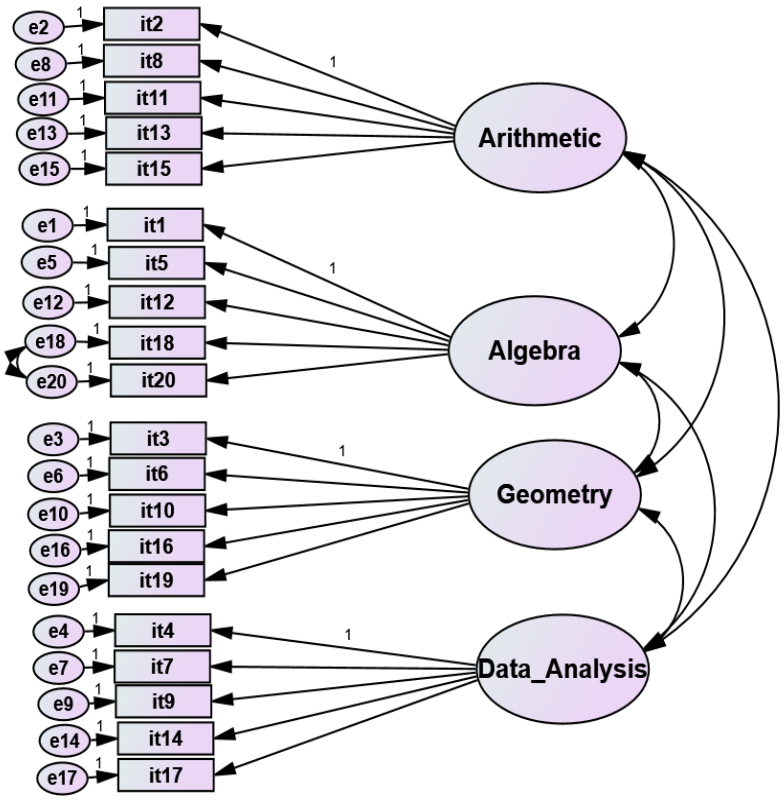
يتبين من الجدول (5) عدم تحقق افتراض التوزيع الطبيعي متعدد المتغيرات، ولمعالجة ذلك تم استخدام طريقة (BOOTSTRAP). وللتحقق من مدى مطابقة البيانات لنموذج العوامل الأربعة المقترح تم استخدام التحليل العاملي التوكيدي المبينة نتائجه في الجدول (6).

الجدول (6): ***نتائج التحليل العاملي التوكيدي لفقرات الامتحان المقدم باللغة الإنجليزية.***

المؤشر الإحصائي	χ^2	Df	χ^2/df	CFI	RMSEA	SRMR	sig
القيمة	*427.773	164	2.608	0.845	0.078	0.088	0.000

*** دالة إحصائياً عند مستوى الدلالة الإحصائية (α=0.05)**

يتبين من الجدول (6) وجود دلالة إحصائية لقيمة (χ^2)، وكانت قيم مؤشرات CFI, RMSEA, SRMR, χ^2/df تقع ضمن المدى المقبول لجودة المطابقة، وبتفحص مؤشرات التعديل (Modification Indices)، تبين أنه يمكن إجراء تعديل على الارتباط بين الأخطاء في النموذج المقترح، لتحسن من مؤشرات مطابقة النموذج المبين في الشكل(2).



الشكل (2): النموذج العاملي المعدل للامتحان المقدم باللغة الإنجليزية.

تم التحقق من جودة النموذج المعدل باستخدام التحليل العاملي التوكيدي، ويبين الجدول (7) النتائج.

الجدول (7): نتائج مطابقة بيانات الامتحان المقدم باللغة الإنجليزية للنموذج بعد إجراء التعديلات المتاحة.

المؤشر الإحصائي	χ^2	Df	χ^2/df	CFI	RMSEA	SRMR	sig
القيمة	*306.309	163	1.879	0.916	0.057	0.076	0.000

*** دالة إحصائياً عند مستوى الدلالة الإحصائية (α=0.05)**

يتبين من الجدول (7) أن جميع المؤشرات الإحصائية تشير إلى تحقق مطابقة أفضل للبيانات مع النموذج بعد إجراء التعديل، فقد ظهر تحسن على جميع المؤشرات، وتدل قيم هذه المؤشرات على مطابقة البيانات لنموذج العوامل الأربعة (الحساب، الجبر، الهندسة، تحليل البيانات) للامتحان المقدم باللغة الإنجليزية. وتشير هذه النتيجة إلى أن هذه العوامل الأربعة تمثل وبشكلٍ واضح المكونات التي يقيسها امتحان الاستدلال الكمي، و ينبغي معرفتها عند هذا المستوى التعليمي؛ أي أن البنية العاملية للامتحان في المجتمع الأردني لا تختلف عن البنية العاملية المستهدفة من قبل مؤسسة خدمات الاختبارات التربوية (ETS). ويؤكد تلك النتيجة وجود ارتباطات موجبة دالةً إحصائياً بين أبعاد الامتحان الأربعة المفترضة، كما يبينها الجدول (8).

الجدول (8): معاملات الارتباط بين الأبعاد الأربعة لامتحان الاستدلال الكمي المقدم باللغة الإنجليزية

البعد	الحساب	الجبر	الهندسة
الجبر	0.384**		
الهندسة	0.224**	0.201**	
تحليل البيانات	0.302**	0.279**	0.264**

**** دالة إحصائياً عند مستوى الدلالة الإحصائية (α=0.01)**

السؤال الثالث: هل يتحقق اللاتغير في القياس لامتحان الاستدلال الكمي من امتحانات (GRE)، باختلاف اللغة التي يقدم بها؟

تم التحقق من مدى مطابقة بيانات الامتحان في العينة الكلية للطلبة بغض النظر عن اللغة التي قدم بها الامتحان (اللغة العربية والإنجليزية) للنموذج المقترح باستخدام التحليل العامل التوكيدي، ويبين الجدول (9) نتائج التحليل العاملي التوكيدي.

الجدول (9): نتائج التحليل العاملي التوكيدي، لبيانات امتحان الاستدلال الكمي المقدم باللغة العربية واللغة الإنجليزية مجتمعة.

المؤشر الإحصائي	χ^2	Df	χ^2/df	CFI	RMSEA	SRMR	sig
القيمة	*448.299	164	2.734	0.906	0.057	0.063	0.000

*** دالة إحصائياً عند مستوى الدلالة الإحصائية (α=0.05)**

يتبين من الجدول (9) وجود دلالة إحصائية لقيمة (χ^2)، وكانت قيم مؤشرات χ^2/df, CFI, RMSEA, SRMR تقع ضمن المدى المقبول لجودة المطابقة، وبتفحص مؤشرات التعديل (Modification Indices)، تبين أنه يمكن إجراء تعديل على الارتباط بين الأخطاء في النموذج المقترح، لتحسن من مؤشرات مطابقة النموذج، وهي في بعد الجبر بين (e18) و(e20)، وفي بعد تحليل البيانات بين (e14) و(e17). وتم التحقق من جودة النموذج المعدل باستخدام التحليل العاملي التوكيدي، ويبين الجدول (10) نتائج ملاءمة النموذج بعد إجراء التعديلات المتاحة.

الجدول (10): نتائج التحليل العاملي التوكيدي، لبيانات امتحان الاستدلال الكمي المقدم باللغة العربية واللغة الإنجليزية مجتمعة بعد إجراء التعديلات المتاحة.

المؤشر الإحصائي	χ^2	Df	χ^2/df	CFI	RMSEA	SRMR	sig
القيمة	*319.051	162	1.969	0.948	0.042	0.055	0.000

*** دالة إحصائياً عند مستوى الدلالة الإحصائية (α=0.01)**

يتبين من الجدول (10) أن جميع المؤشرات الإحصائية تشير إلى تحقق مطابقة أفضل للبيانات مع النموذج بعد إجراء التعديل، فقد ظهر تحسن على جميع المؤشرات، وتدل قيم هذه المؤشرات على مطابقة البيانات مجتمعة لنموذج العوامل الأربعة (الحساب، الجبر، الهندسة، تحليل البيانات).

أولاً: اللاتغير التكويني

يبين الجدول (11) نتائج مطابقة بيانات كل من الامتحان المقدم باللغة العربية والامتحان المقدم باللغة الإنجليزية.

الجدول (11): نتائج مطابقة البيانات للنموذج وفق متغير اللغة.

المؤشر الإحصائي	χ^2	Df	χ^2/df	CFI	RMSEA	SRMR	Sig
العربية	*228.429	164	1.393	0.964	0.038	0.059	0.001
الإنجليزية	*427.772	164	2.608	0.845	0.078	0.088	0.000

*** دالة إحصائياً عند مستوى الدلالة الإحصائية (α=0.05)**

يتبين من الجدول (11) وجود دلالة إحصائية لقيمة (χ^2)، وكانت قيم مؤشرات

χ^2/df, CFI, RMSEA, SRMR تقع ضمن المدى المقبول لجودة المطابقة ، وذلك دليل على مطابقة النموذج لكل من اللغة العربية واللغة الإنجليزية. وبعد تحقق مطابقة النموذج المقترح لمتغير اللغة (العربية، الإنجليزية)، تم استخدام التحليل العاملي التوكيدي متعدد المجموعات مع البيانات مجتمعة لاختبار خاصية اللاتغير التكويني وفق متغير لغة الامتحان؛ دون فرض قيود (No Constraints) ، ويبين الجدول (12) النتائج.

الجدول (12): *نتائج اختبار خاصية اللاتغير التكويني للامتحان وفق متغير اللغة.*

المؤشر الإحصائي	χ^2	Df	χ^2/df	CFI	RMSEA	SRMR	Sig
اللغة	*624.205	328	1.903	0.915	0.041	0.059	0.000

*** دالة إحصائياً عند مستوى الدلالة الإحصائية (α=0.05)**

يتبين من الجدول (12) وجود دلالة إحصائية لقيمة (χ^2)، وكانت قيم مؤشراتχ^2/df ,CFI, RMSEA, SRMR تقع ضمن المدى المقبول لجودة المطابقة ، وتدل قيم تلك المؤشرات مجتمعة على تحقق خاصية اللاتغير التكويني بشكل جيد وفق متغير اللغة. وهذا يعني أن النسختين تتضمنان نفس العدد من الأبعاد وبنفس مسمياتها.

ثانياً: اللاتغير المتري

تم فحص تحقق خاصية اللاتغير المتري في فقرات الامتحان وفق متغير لغة الامتحان؛ وذلك بفرض قيود على النموذج المراد فحصه، ويبين الجدول (13) النتائج. ويبين الجدول 14 التغير في تلك النتائج عنها في اللاتغير التكويني.

الجدول (13): نتائج نموذج اللاتغير المتري للامتحان وفق متغير اللغة.

المؤشر الإحصائي / النموذج	χ^2	Df	χ^2/df	CFI	RMSEA	SRMR	Sig
المتري	*765.361	348	2.199	0.880	0.064	0.089	0.000

*** دالة إحصائياً عند مستوى الدلالة الإحصائية (α=0.05)**

الجدول (14): نتائج اختبار تحقق خاصية اللاتغير المتري للامتحان مقارنة باللاتغير التكويني وفق متغير اللغة.

نموذج اللاتغير المتري	الفرق في					Sig
	Df	χ^2	CFI	RMSEA	SRMR	
لجميع الفقرات	20	*141.156	-0.035	0.023	0.030	0.000

*** دالة إحصائياً عند مستوى الدلالة الإحصائية (α=0.05)**

يتبين من الجدول (13) أن البيانات تناسب النموذج المتري، ويبين الجدول (14) أن قيم الفرق في المؤشرات (ΔRMSE, ΔSRMR, ΔCFI) بين النموذج التكويني والنموذج المتري يدل على عدم تحقق خاصية اللاتغير المتري في الامتحان وفق متغير لغة الاختبار، بالإضافة إلى أن الاختبار الإحصائي Chi) (Square Difference Test للفرق في قيمة χ^2 بين نموذج اللاتغير التكويني، ونموذج اللاتغير المتري لجميع الفقرات دالة إحصائياً، وهذا دليل إحصائي آخر على عدم تحقق خاصية اللاتغير المتري؛ أي أن فرض قيد تساوي تشعبات الفقرات على العوامل وفق متغير اللغة يؤثر بصورة دالة إحصائياً على مطابقة البيانات للنموذج. وهذا يدلل على اختلاف قيم التشبع للفقرات في النسخة العربية عن قيم التشبع لها في النسخة الإنجليزية.

ثالثاً: اللاتغير العددي

نتيجةً لعدم تحقق خاصية اللاتغير المتري؛ التي تعني أن المفهوم غير متكافئ عبر المجموعات، وبما أن اللاتغير في القياس يصنف إلى ثلاث مستويات هرمية السابق فيها شرطاً للاحق، فإنه لا معنى من مواصلة اختبار خاصية اللاتغير العددي.

كشفت النتائج عن تحقق اللاتغير التكويني فقط، في حين لم يتحقق اللاتغير المتري والعددي في القياس للبناء العاملي المفترض لامتحان الاستدلال الكمي وفق متغير اللغة؛ كدليل على أن امتحان الاستدلال الكمي المقدم باللغة الإنجليزية والعربية لهما نفس شكل البنية العاملية المفترضة، لكن لم يكن لهما نفس التشعبات على العوامل، والتقاطعات بين العوامل. وبما أن اللاتغير في القياس يعتبر بمثابة متطلب سابق قبل إجراء المقارنات بين الأوساط؛ إذ أن الاستنتاجات المبنية على تحليلات الفروق بين المتوسطات قد تكون متحيزة أو غير دقيقة إذا كانت القياسات ليس لها المعنى نفسه عبر تلك المجموعات المختلفة؛ فالفروق الموجودة بين أداء الطلبة على الامتحانين لا تعزى فقط للاختلاف في القدرة على الاستدلال الكمي، بل تعزى كذلك إلى اختلاف اللغة التي قدم فيها الامتحان. ويؤكد ذلك أن المجموعتين متكافئتين في القدرة من خلال الاختيار العشوائي، ومع ذلك وجدت فروق ذات دلالة إحصائية بين الوسط الحسابي للصورة العربية (10.9) والصورة الإنجليزية (7.17) لصالح الصورة العربية، وهذا يتفق مع دراسة هي وآخرون (He, et al., 2014) ودراسة بروير وآخرون (Broer, et al., 2005)، التي أكدت أن أداء الطلبة الذين تكون لغتهم الأولى مختلفة عن لغة الاختبار أقل من نظرائهم الذين تكون لغتهم الأولى هي لغة الامتحان، مما يعني أن الامتحان بالصورة الإنجليزية قد يكون متحيزاً ضد الطلبة الأردنيين بشكل خاص والطلبة العرب بشكل عام؛ لأن لغتهم الأم هي اللغة العربية، وأغلبهم تعلم المهارات التي يقيسها الامتحان باللغة العربية في المدرسة، بالإضافة إلى أن معظم الطلبة لا يتقنون اللغة الإنجليزية بشكلٍ كافٍ، وعليه فإن الامتحان بصورته الإنجليزية قد يكون غير نقي، بحيث يقيس بعد اللغة الإنجليزية بالإضافة إلى الاستدلال الكمي، وهذا يتفق مع ما أشار إليه فينهوفين (Veenhoven, 1996) بأن اختلاف تفسير أفراد عينة الدراسة لبعض المفردات في فقرات الامتحان يعود إلى اختلاف اللغة أو الاختلافات الثقافية بين الشعوب. وبما أن الامتحان المقدم باللغة العربية

يحقق مطابقة أفضل مع النموذج العاملي للامتحان، ولا تمثل درجة اتقان اللغة الإنجليزية عامل مؤثر فيه؛ فإن استخدامه مع الطلبة الذين تكون لغتهم الأم هي اللغة العربية، قد يعكس الاختلاف في القدرة على الاستدلال الكمي بشكلٍ أفضل، ويحقق مقارنات عادلةٍ وصادقةٍ بين الطلبة، الذين تكون لغتهم الأم هي اللغة العربية. كما يتفق هذا مع ما أشارت إليه مؤسسة خدمات الاختبارات التربوية (ETS) إلى أنه واحدة من الخيارات الممكنة، أن يكون الاختبار متاحاً بأكثر من لغة، بشرط أن يقيس نفس البناء، مع توفر بيانات ومعلومات كافية عن الصدق بتلك اللغات (ETS, 2015).

التوصيات

- استخدام امتحان الاستدلال الكمي باللغة العربية.
- إجراء دراسات حول القدرة التنبؤية لامتحان الاستدلال الكمي باللغة العربية.
- اعتماد الجامعات الأردنية على مثل تلك الامتحانات كأحد محكّات القبول في برامج الدراسات العليا.
- التعاون مع المؤسسات التي تطور امتحانات قبول والاستفادة من تجاربها، ونقل مثل تلك الامتحانات للبيئة العربية.

المصادر والمراجع

أبو عواد، ف. (2018). دلالات الصدق والثبات لمقياس الأساليب المعرفية_ دراسة سيكومترية على طلبة الجامعات الأردنية. *مجلة جامعة النجاح للأبحاث، العلوم الإنسانية، 32*(12)، 2190-2218.

جروان، ف. (2008). *الموهبة والتفوق والإبداع*. دار الفكر.

السمكي، ل. (2017). *اللاتغير في القياس لاختبارTIMSS 2011 في الرياضيات والعلوم عبر موقع المدرسة وجنس الطالب. أطروحة دكتوراه غير منشورة*، جامعة اليرموك.

كريش، أ. (2015). أهمية التحليل العاملي التوكيدي المتعدد المجموعات كأسلوب للقياس في ميدان علم النفس الجنائي. *المؤتمر العلمي الدولي الأول: علم النفس وقضايا المجتمع الراهنة*، جامعة الجزائر 2، 276-287.

References

Piotr, S. (2015). American Educational Research Association, American Psychological Association, National Council on Measurement in Education, Standards for educational and psychological testing, American Educational Research Association, Washington 2014, 240 s. Kwartalnik Pedagogiczny, (60 (4 (238), 201-203.

Ariely, G., & Davidov, E. (2012). Assessment of measurement equivalence with cross-national and longitudinal surveys in political science. *European Political Science*, *11*(3), 363-377.

Broer, M. (2005). Ensuring the fairness of GRE writing prompts: Assessing differential difficulty. *ETS Research Report Series*, (1), i-41.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. *Testing structural equation models*, 136-162.

Butcher, J. N. (2006). *MMPI-2: A practitioner's guide*. American Psychological Association.

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement in variance. *Structural equation modeling: a multidisciplinary journal*, *14*(3), 464-504.

Cheung, G. W. (2012). Mainstreaming culture in psychology. *American Psychologist*, *67*, 721-730.

Davidov, E., Meuleman, B., Crecyuch, J., Schmidt, P., & Billiet, J. (2014). Measurement equivalence in cross- national research. *Annual Review of Sociology*, *40*(1), 55-75.

Educational Testing Service. (2013*). A snapshot of the individual who took the GRE revised General Test*. *Educational Testing Service*.

Educational Testing Service. (2014). *MBA programs that accept the GRE revised General. Educational Testing Service*. http://www.ets.org/gre/revised_general/about/mba/programs?WT.ac=rx42

Educational Testing Service. (2015*). ETS standards for quality and fairness*. Educational Testing Service. http:// www.ets.org/gre/guide

Educational Testing Service. (2016*). A Snapshot of the Individuals Who Took the GRE General Test.* Educational Testing Service. http:// www.ets.org/gre/pdf/snapshot_test_taker_data_2016.pdf

Fischer, R. & Poortinga, H. (2018). Addressing methodical challenges in culture comparative research. *Journal of Cross-Cultural Psychology*, *49*(5), 691-712.

Hair, J. F., Anderson, R. E., Tatham, R. L., & Black, W. C. (1995). *Multivariate Data Analysis with Reading.* (4th ed.). Plentice-Hall.

He, Q., Anwyll, S., Glanville, M., &Opposs, D. (2014). An investigation of measurement invariance of the Key Stage 2 National Curriculum science sampling test in England. *Research Papers in Education, 29*(2), 211-239.

Horn, J. & Mcardle, J. (1992). Apractical and theoretical guide to measurement invariance in aging research. *Experimental Aging Research, 18*(3), 117-144.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural equation modeling: a multidisciplinary journal*, *6*(1), 1-55.

International Test Commission. (2017). The ITC guidelines for translating and adapting tests. http:// www.InTestCom.org

Kim, H., & Sasaki, J. Y. (2017). Intercultural similarities and differences in personality development. In Personality Development Across the Lifespan (pp. 419-434). Academic Press.

Kline, R. B. (2015). *Principles and practice of structural equation modeling*. Guilford Press.

Kuncel, N. R., Wee, S., Serafin, L., & Hezlett, S. A. (2010). The validity of the Graduate Record Examination for master's and doctoral programs: A meta-analytic investigation. *Educational and Psychological Measurement*, *70*(2), 340-352.

Marlene, E., Henerson, L. M. & Carol, T. G. (1978). *How to Measure Attitudes*. (7th ed.). Sage Publications.

Marsh, H. W., & Hocevar, D. (1985). Application of confirmatory factor analysis to the study of self-concept: First-and higher order factor models and their in variance across groups. *Psychological bulletin*, *97*(3), 562.

Reeves, J. (2002). Aptitude assessment for career and educational guidance. *The Work Suite.*

Sekercioglu, G., & Kogar, H. (2018). The Examination of Measurement Invariance and Differential Item Functioning of PISA 2015 Cognitive Tests in Terms of the Commonly Used Languages. *Novitas-ROYAL (Research on Youth and Language)*, *12*(2), 152-172.

Sinkovics, R. R., Henseler, J., Ringle., C.M., & Sarstedt, M. (2016). Testing measurement invariance of composites using partial least squares. *International marketing review*.

Sousa, K., West, S., Moser, S., Harris, J. & Cook, S. (2012). Establishing measurement invariance: English and Spanish Pediatric Asthma Quality of Life Questionnaire. *Nurse Research, 61*(3), 171-180.

Steenkamp, J. B. E., & Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of consumer research*, *25*(1), 78-90.

Tsaousis, I. (2015). *Factorial invariance and latent mean differences of scores on SAAT across gen*der. http://www.qiyas.sa/MAndAssesment/Researches/Pages/Researches.aspx

Wendler, C., Bridgeman, B., & Ezzo, C. (2014). The Research Foundation for the GRE revised General Test: A compendium of studies. *Educational Testing Service.* http://www.ets.org/gre/compendium

West, S. G., Finch, J. F., & Curran, P. J. (1995). Structural equation models with nonnormal variables: Problems and remedies.

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational research method*, 3(1), 4-70.

Van de Vijver, F., & Tanzer, N. K. (2004). Bias and equivalence in cross-cultural assessment: An overview. *European Review of Applied Psychology, 54*(2), 119-135.

Varsity Tutors. (2019*). GRE Study Guide: Lessons, Strategies, and Diagnostic Tests*. Varsity Tutors LLC. https://www.google.com/url?sa=t&scource=web&rct=j&url=https://varsitytutors.com/varsity-tutors-gre-book-first-edition.pdf&ved=2ahUKEwiatO-5ocj0AhU77uAKHTgqCLsQFnoECDIQAQ&usg=AOvVaw2BBD-opn4tBmwDAb4oQPGE

Veenhoven, R. (1996). Developments in satisfaction-research. *Social Indicators Research, 37*(1), 1-46.