

مؤشرات عدالة اختبار الكفاية اللغوية لمجمع اللغة العربية بدلالة الأداء التفاضلي للفقرات

أ.د. أحمد سليمان عودهⁱⁱ

تاريخ القبول

2024/8/3

بيداء بسام البطرانⁱ

تاريخ الاستلام

2024/7/3

ملخص

هدفت هذه الدراسة للتحقق من مؤشرات عدالة اختبار الكفاية اللغوية بدلالة الأداء التفاضلي للفقرات، تبعاً لمتغيري النوع الاجتماعي والنموذج. تكونت عينة الدراسة من (1161) من المتقدمين لاختبار الكفاية اللغوية للعام 2024/2023، المعد من قبل مجمع اللغة العربية الأردني وذلك من خلال استخدام المنهج الوصفي المقارن في تحليل نتائج اختبار الكفاية اللغوية، ولذلك تم اعتماد الخصائص السيكومترية (معاملات الصعوبة، معاملات التمييز) للنماذج المختارة للدراسة. وأشارت النتائج إلى عدم وجود فروق تعزى لأثر النموذجين (4، 6) في معاملات الصعوبة ومعاملات التمييز. وأظهرت النتائج أيضاً هناك (13) فقرة تظهر أداءً تفاضلياً، وفي النموذجين (4، 6) هناك (43) فقرة أظهرت أداءً تفاضلياً، وتوصي الدراسة بضرورة التحقق من مؤشرات عدالة الاختبارات الوطنية والاختبارات واسعة النطاق بالطرائق التي تشمل جوانب الاختبار بأكمله سواء أكانت النوعية أم الكمية.

الكلمات المفتاحية: عدالة الاختبارات، اختبار الكفاية اللغوية، الخصائص السيكومترية، الأداء التفاضلي للفقرات.

ⁱ جامعة اليرموك

ⁱⁱ جامعة اليرموك

Indicators of the Fairness of the Arabic Language Academy's Language Proficiency Test Based on the Differential Performance of the Items

Abstract

This study aimed at investigating the indicators of fairness of the language proficiency test (LPT) based on the differential performance of the items, in terms of the two variables (gender and model). The sample of the study consisted of (1161) applicants for the language proficiency test for the year 2023/2024. The (LPT) was prepared by the Jordanian Arabic Language Academy. The descriptive-comparative approach to analyzing the results of the language proficiency test was used. Therefore, the psychometric properties (difficulty coefficients, discrimination coefficients) of the models of the study were adopted.

The results showed that there were no statistically significant differences attributed to the effect of the two models (4&6) on the difficulty coefficients, and discrimination coefficients. The results of the study also showed that there were (13) items that showed differential performance, and (43) items showed differential performance in models (4 & 6). Therefore, the study recommended the necessity of investigating the indicators of fairness of the national tests as well as large-scale tests, using ways and methods that encompass all aspects of the test, whether qualitative or quantitative.

Keywords: Fairness of tests, language proficiency test, psychometric properties, differential performance of items.

مقدمة

نظراً لأهمية الاختبارات النفسية والتربوية في اتخاذ القرارات، ولأهمية هذه القرارات المترتبة على نتائج تطبيق الاختبارات المختلفة، ونظراً للحاجة إلى تكرار تطبيق الاختبار على المفحوصين أنفسهم أو على مجموعات متعددة من المفحوصين، وخوفاً من تعرض سرية الاختبارات للخطر، تبرز أهمية عملية تعدد صور الاختبارات، مما يعني ضرورة التحقق من عدالة الاختبار لمعرفة العوامل المؤثرة فيه، ولضمان نجاح تطبيق الاختبار ضمن متغيرات مختلفة، وذلك بعد بناء الاختبار وإدارته وتقويمه، حيث يُعد الاختبار من أهم الأدوات التي تعكس سلوك المتعلم إن كان في المدارس أو الجامعات أو العيادات أو الدوائر الرسمية وغير الرسمية، ويستخدم في عدة مجالات؛ كالاختبار والتصنيف والتقويم والإرشاد والبحث العلمي، وتهدف الاختبارات التحصيلية إلى تقييم معارف المتعلم ومهاراته المكتسبة، ويكون الاختبار عادلاً وصادقاً بالدرجة التي تكون الاستنتاجات المبنية عليه مناسبة وذات دلالة وفائدة.

ولقد تركزت جهود العديد من الباحثين ومطوري الاختبارات على بناء اختبارات ذات خصائص سيكومترية مقبولة للفقرات، تتمثل في الصعوبة والتمييز والتخمين، والصدق والثبات، ونظراً لتأثر خصائص الفقرات والاختبار بخصائص المجموعة التي يطبق عليها، مما يؤدي إلى اختلاف الأداء التفاضلي للفقرات باختلاف المجموعة، فقد سعى الباحثون إلى تطوير الاختبارات بحيث تتمتع بخصائص سيكومترية مناسبة لا تتأثر بعينة المفحوصين، بالإضافة إلى خلوها من خصائص معينة قد يجعلها تتأثر بعوامل ذات علاقة بعينة الدراسة؛ كالنوع الاجتماعي والعرق ومكان الإقامة واللغة والمستوى الاقتصادي والاجتماعي، والتي قد تؤثر على القرارات المتخذة بناءً عليها (الشرفين، 2018).

ويجب أن يتوافر الصدق في الاختبارات المختلفة، كونه يتعلق بالدرجة التي تتوافر فيها أدلة عملية ونظرية تدعم تفسير درجات الاختبار بناءً على الاستخدام الذي وُجدت من أجله، وهذا يتطلب معلومات تتعلق بثبات الاختبار وتطبيقه وطريقة توزيع الدرجات، وعملية معادلتها، والتحقق من عدالة الاختبار بالنسبة لكافة الطلبة بغض النظر عن النوع الاجتماعي أو العرق أو الثقافة (Bert & Stoneberg, 2004).

وترى زيدان وبوجردة (2017) أن الصدق من أهم المفاهيم التي يجب وضعها بعين الاعتبار عند إعداد الاختبار، وله عدة أنواع؛ وهي: صدق المحتوى، الصدق المرتبط بمحك، وصدق التكوين، والصدق التمييزي، ويقصد بالصدق جمع الأدلة التي تدعم الاستدلالات والاستنتاجات التي سيتوصل إليها الباحث من خلال جمع البيانات.

أما ثبات الاختبار، فيقصد به مدى استقرار نتائجه في حال تم تطبيقه على عينة من الأفراد في موقفين مختلفين، ويمكن اعتبار الاختبار ثابتاً؛ من خلال الحصول على النتائج نفسها مهما تم إعادة تطبيق الاختبار، واحتمال وجود أخطاء في القياس؛ أي أن درجة الفرد في الاختبار تساوي مجموع الدرجة الحقيقية والدرجة الخطأ، ويمكن اختبار ثبات الاختبار عن طريق التطبيق وإعادة التطبيق، أو طريقة الصورتين المتكافئتين، أو طريقة التجزئة النصفية، أو طريقة الاتساق الداخلي (بوحفص، 2011).

وبعد تطبيق الاختبارات يتم تحليل نتائجها، وتقديم تقارير بخصوص ذلك، وتقديمها للمسؤولين على المستويات كافة لاتخاذ القرارات المناسبة، ولكي تكون هذه القرارات صحيحة، يجب أن تعتمد على بيانات صادقة، وذلك يرجع إلى أن تتمتع الاختبارات بخصائص معينة؛ كالصدق والثبات والعدالة، بالإضافة إلى خاصية استقرار القياس، ويقصد بانتهاك خاصية استقرار

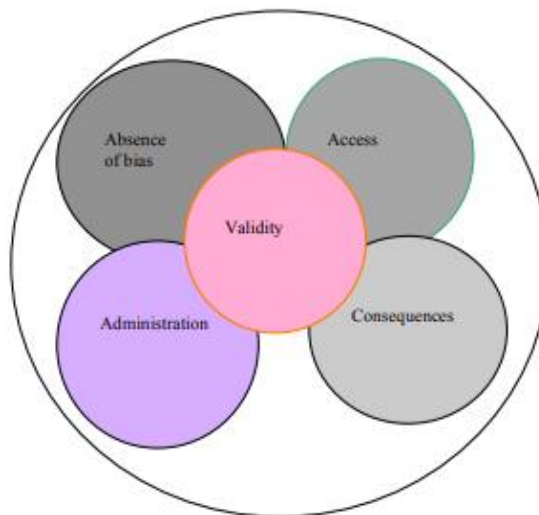
القياس على مستوى الفقرة باسم الأداء التفاضلي للفقرة، ويُعدّ الاختبار الذي يتضمن مثل هذه الفقرات متحيزاً غير محقق لمبدأ العدالة بين المفحوصين، (Komboz & Strobl & Zeileis, 2018).

وُعدّ عدالة الاختبار من أقل الشروط السيكمترية التي يسعى معدو الاختبارات في الوطن العربي إلى التحقق منها بشكل مباشر، وذلك إذا ما تمت مقارنتها بشروط سيكمترية أخرى؛ كالصدق والثبات وصعوبة الفقرات وقدرتها على التمييز وفاعلية المشتتات، ولكن في الوقت ذاته حظي موضوع عدالة الاختبار باهتمام كبير لدى المتخصصين في الدول الأجنبية، ونادوا بضرورة التحقق من عدالة الاختبارات المختلفة (عبدالوهاب، 2014).

ويُعدّ مفهوم عدالة الاختبار من المفاهيم المعقدة والمثيرة للجدل، لدرجة أنه لم يتم تقديم تعريف متفق عليه إلى الآن، بالإضافة إلى ذلك، فقد تم اقتراح بعض النماذج لعدالة الاختبار، ولكن لا يوجد أي منها يفسح المجال بسهولة للتحقيق العملي في مفهوم العدالة، وقد لوحظ أنه يمكن تحديد عدالة الاختبار بشكل موضوعي من خلال فحص الخصائص السيكمترية لنتائج الاختبار، لا سيما ما إذا كان الاختبار متحيزاً لصالح أو ضد مجموعة من المتقدمين للاختبار؛ فإذا كشفت نتائج الاختبار عن وجود تحيز، فسيتم اعتبار الاختبار غير عادلاً (Kunnan, 2018).

وبعد ذلك قدم كونان عام 2010 إطاراً مفاهيمياً يتضمن المنهج النوعي والكمي في جمع البيانات للتوصل إلى مؤشرات عدالة الاختبار، حيث كان ذلك الإطار ينظر إلى النظام الكامل لممارسة الاختبار وليس الاختبار نفسه، وأشار كونان أنّ هناك جوانب متعددة لعدالة الاختبار تتضمن استخدامات الاختبار المتعددة (الأغراض المقصودة، والأغراض غير المقصودة)، والعديد من أصحاب المصلحة في عملية الاختبار (المتقدمون للاختبار، مستخدمو الاختبار، والمعلمون، وأصحاب العمل)، وخطوات متعددة في عملية تطوير الاختبار (تصميم الاختبار، وتطويره)، والإدارة والاستخدام.

ويعرض الشكل (1) نموذج عدالة الاختبار لكونان (2010) ضمن دائرة الاختبارات، وممارستها:



وفي هذا النموذج، يرى كونان أن العدالة كمفهوم شامل تتضمن خمسة جوانب للاختبار؛ وهي: الصدق، وغياب التحيز في الاختبار، والوصول المتكافئ للتعليم والاختبار، وإدارة الاختبار، والعواقب الاجتماعية، ولقد تم توسيع هذا النموذج لاحقاً عام (2010)، وأشار إلى أن هذا النموذج يأخذ في الاعتبار نظام الاختبار بأكمله، وليس الاختبار نفسه فقط، وكان أكثر شمولاً من النماذج الأخرى الموجودة (Rezai, 2022).

ولقد حدد متخصصو القياس عدالة الاختبار بطرائق مختلفة، وقد لا توضح هذه الطرائق مفهوم عدالة الاختبار والجوانب المرتبطة به إلى الحد الذي يمكن للممارسين فيه ملاحظة العدالة في سياقات الاختبار الفعلية؛ فقد يحتاجون إلى أنموذج ملموس يمكن من خلاله ملاحظة العدالة وتطبيقه في مواقف الاختبار، ومن بين جميع النماذج المتاحة المقترحة لعدالة الاختبار، يمكن اعتبار النموذج الذي قدمه كونان العديد من دراساته، واعتبر أساساً منطقياً لإطار عدالة الاختبار، وقدم مجموعة من المؤشرات العامة والمؤشرات الفرعية وكانت تستند جميعها على المنهج النوعي في جمع البيانات وتحليلها بعيداً عن المنهج الكمي (Kunnan, 2004).

وأشار الوابل وتعلب (2023) إلى أن التحقق من عدالة الاختبار (Fairness test) خطوة مهمة جداً لتقييم قدرات المفحوصين في السمة التي يقيسها الاختبار، وذلك من خلال استخدام الأساليب والطرائق الإحصائية المناسبة لتحليل الأداء التفاضلي (DIF)، وقبل التحليل، يتم افتراض أن المفحوصين متساوون في القدرة، ويجب أن لا يتأثر أدائهم على الفقرات بمتغير آخر غير القدرة.

يُعدّ فحص الأداء التفاضلي للفقرة (DIF Differential Item Functioning) جزءاً مهماً في تحليل فقرات الاختبار، إذ إنّ خلو الاختبار وفقراته من الأداء التفاضلي من أهم الشروط التي يجب توافرها في الاختبار لنشره واستخدامه، ويُعدّ الأداء التفاضلي من أهم قواعد النشر للاختبارات واعتمادها، وإذا استخدمت الاختبارات في اتخاذ القرارات، لا بدّ من الحصول على دالة تجريبية واضحة من أجل دعم الإدعاءات بأن فقرات الاختبار لا تعكس أداءً تفاضلياً للفقرة عبر المجموعات المختلفة من المفحوصين (Fidalgo & Madeira, 2008).

ويظهر الأداء التفاضلي للفقرة عندما يختلف أداء المفحوصين المتساوين في مستوى القدرة، والذين ينتمون إلى مجموعات مختلفة، ونظراً لاحتمالية أن يؤثر الأداء التفاضلي على صدق الاختبار، فالكشف عن الأداء التفاضلي لفقرات يُعدّ من الأمور الضرورية للقياس الصادق والمتحور من التحيز (Park, 2010).

ويعدّ الأداء التفاضلي للفقرة دليلاً على صدق البنية الداخلية للاختبار، فهو يشير إلى الاختلافات في الأداء على الفقرة عبر مجموعات مختلفة من الأفراد، والكشف عن الفقرات ذات الأداء التفاضلي لا يكفي من أجل إدعاء عدم العدالة في الاختبار، فعادلة الاختبار تعتمد على كيفية استخدامه، وتحيز الاختبار هو فرق منتظم في الدرجة الكلية للاختبار ضد مجموعة معينة (Camilli, 2006).

ويرى العجلوني والسوالمه (2016) أن الكشف عن الأداء التفاضلي للفقرة يعدّ ضرورياً للقياس الصادق والمتحور من التحيز في الفقرات. لذا ظهرت العديد من الطرائق الإحصائية للكشف عن الأداء التفاضلي للفقرة، منها ما يعتمد على نظرية القياس الكلاسيكية، ومنها ما يعتمد على نظرية الاستجابة، ومنها ما يعتمد على التحليل العاملي، والإحصائي مربع كاي، ولكن أكثر الطرائق استخداماً ضمن هذه المجموعة هي: طريقة مانتل – هانزل (Mantel-Haenszel)، وطريقتي مربع كاي (Chi-Square Statistic) للاستجابات الصحيحة،

ومربع كاي للاستجابات الصحيحة والخطأ، وطريقة الانحدار اللوجستي (LogisticRegression Methods) (علام، 2005).

وتعدّ طريقة مانتل هانزل من أكثر الطرائق نظرية القياس الكلاسيكية شيوعاً في الكشف عن تحيز الفقرة، وذلك لسهولة حساباتها وإجراءاتها، وإحدى طرائق الكشف عن الأداء التقاضي للفقرة، وذلك بسبب سهولة تطبيقها، حيث تقوم على مبدأ مقارنة متوسطات المجموعات في الأداء على الاختبار ككل، أو على فقراته المتكافئة، وذلك باستقصاء التحيز بين مجموعتين (المجموعة المرجعية (R) والمجموعة المستهدفة (G) وهي المجموعة التي يعتقد بوجود فقرات تحيز ضدها) (Fidalgo & Madeira, 2008)، ويتم بناء الفرضية الصفرية من خلالها، بحيث تنص على عدم وجود فروق دالة إحصائية في نسب الإجابة الصحيحة بين المجموعتين المرجعية والمستهدفة في مستوى القدرة الواحدة (Kamata & Vaughn, 2004).

ويرى الباحثان أنّ اختبارات الكفاية اللغوية كغيرها من الاختبارات التي تلعب دوراً إيجابياً في نجاح العملية التعليمية، والتي يلاحظ تطورها بشكل كبير في الفترة الأخيرة نتيجة التقدم العلمي والتكنولوجي القائم على تحقيق نواتج تعليمية ناجحة، بالإضافة إلى إنشاء جيل من المعلمين يتقنون اللغة بشكل جيد يجعلهم قادرين على التواصل مع الطلبة بشكل فعال.

ففي السنوات الأخيرة، لعبت اختبارات الكفاية اللغوية دوراً مهماً بشكل متزايد في السياقات التعليمية أو الاجتماعية؛ حيث يتم اتخاذ مجموعة من القرارات المهمة والحاسمة (كالتعليم والتعيين، والعمل، والهجرة، والمواطنة) بناءً على نتائج هذا الاختبار الذي يُعد من المعايير الأساسية لتحديد المستويات والقدرات اللغوية للأفراد في ضوء خبراتهم وتجاربهم السابقة، والتي تتراوح من القبول إلى المستوى الأعلى. وبالتالي، ينبغي إيلاء اهتمام شامل للنظر في مراحل بناء اختبارات الكفاءة اللغوية، والتأكد من صدقها وثباتها وفعاليتها من أجل إصدار أحكام عادلة تخلو بالحد الأدنى من الأخطاء (Taylor, 2009).

ولأنّ الكفاية اللغوية مهمة بشكل كبير في مختلف المجالات، وخاصة المجال العلمي، اهتمت الدول العربية وأولها المملكة الأردنية الهاشمية بوضع اختبارات تهدف لقياس مستوى الكفاية اللغوية للمعلمين الجدد، حيث ألزم رئيس ديوان الخدمة المدنية وزير التربية والتعليم بعدم تعيين أي شخص دون اجتياز اختبار الكفاية اللغوية، وذلك ضمن مفهوم الاستحقاق والجدارة من جهة، وضمان معايير العدالة وتكافؤ الفرص من جهة أخرى (مجمع اللغة العربية الأردني، 2018).

واهتمت العديد من الدول بوضع اختبارات لقياس مستوى الكفاية اللغوية لدى المعلمين الجدد، ومن أهم هذه الدول المملكة الأردنية الهاشمية، حيث نصت المادة (10) من قانون حماية اللغة العربية رقم (35) لسنة 2015: "لا يعين معلم في التعليم العام أو عضو هيئة تدريس في التعليم العالي أو مديع أو معدّ أو محرر في أي مؤسسة إعلامية إلا إذا اجتاز امتحان الكفاية في اللغة العربية...". ويشتمل الاختبار على أربعة أقسام، وهي: الفهم والاستيعاب، والتراكيب اللغوية، والنحوية، والمعجم والدلالة (مجمع اللغة العربية الأردني، 2018).

وبعد البحث والاطلاع، وجد الباحثان العديد من الدراسات التي تناولت متغيراً أو أكثر من متغيرات الدراسة، كدراسة كارمي (Karami, 2011) التي هدفت الكشف عن الأداء التقاضي ل فقرات

وقام يو وآخرون (2019) بدراسة في الولايات المتحدة الأمريكية هدفت التعرف إلى كيفية استخدام تقييم تكافؤ الدرجات (SEA) لتقييم عدالة درجات الاختبارات المُبلغ عنها من التقييمات المخصصة للمتقدمين للاختبار من خلفيات ثقافية ولغوية وتعليمية متنوعة. ولتحقيق أهداف الدراسة، تم تحديد المجموعات الفرعية من خلال خصائص خلفية المتقدم للاختبار التي أظهرت الدراسة أنها مرتبطة بالأداء في اختبارات اللغة. أظهرت النتائج التجريبية أن الطرائق الإحصائية والقياسية النفسية المستخدمة في استخراج درجات الاختبار لم تتأثر بشدة بالمجموعات الفرعية من المتقدمين للاختبار التي تم اشتقاق الدرجات منها، وتوفر هذه النتيجة دليلاً يدعم قابلية المقارنة ومعنى درجات الاختبار عبر مختلف مجموعات المتقدمين للاختبار التي تمت دراستها، وبالتالي يعكس مدى عدالة درجات الاختبارات.

أجرى العسيري والزهراي (2020) دراسة في السعودية هدفت إلى بناء اختبار محكي المرجع لقياس الكفايات اللغوية لدى طلبة المرحلة الثانوية في مدينة الرياض، والتحقق من خصائصه السيكومترية. ولتحقيق أهداف الدراسة، تم استخدام المنهج الوصفي المسحي، وبناء الاختبار مروراً بأربع مراحل (مرحلة التحليل، مرحلة البناء، مرحلة التجريب، مرحلة الإخراج)، وبعد المرور بهذه المراحل، تم بناء اختبار محكي المرجع لقياس الكفايات اللغوية، يتكون من (40) مفردة اختبارية تقيس كفايتين لغويتين وهي (الكفاية النحوية، والكفاية الإملائية). تكونت عينة الدراسة من (983) طالباً من المرحلة الثانوية. أظهرت نتائج الدراسة أن الاختبار يتمتع بصدق جيد، وثبات جيد، وذلك وفقاً لمعامل ليفجستون، حيث وصل إلى درجة (85%).

وأجرى مقدم ونصير زاده (2022) دراسة في إيران هدفت التحقق من عدالة اختبار الفهم القرائي في إطار عدالة اختبار كونان، وذلك باعتباره النموذج الأكثر شمولاً المتاح للكشف عن عدالة الاختبار. ولتحقيق أهداف الدراسة، تم استخدام اختبار الفهم القرائي كما تم إجراء المقابلات بغرض تحليل البيانات. تكونت عينة الدراسة من (300) طالباً درسوا دورة اللغة الإنجليزية العامة، و(3) مطوري الاختبار، و(7) مسؤولين جامعيين عن إجراء الاختبار. للتوصل إلى النتائج تم استخدام التحليل العاملي الاستكشافي، واختبار (t)، وتحليل التباين الأحادي (ANOVA)، وطريقة مانتل – هانزل (Mantel- Haenszel). أظهرت نتائج الدراسة عدالة اختبار الفهم القرائي، وأن الاختبار يتمتع بدلالات صدق وثبات جعلته مناسباً للاستخدام.

وأجرى صباح (2020) دراسة في الجزائر هدفت الوقوف على اختبارات الكفاءة اللغوية الإلكترونية بصورة عامة، واختبارات الكفاءة اللغوية الإلكترونية للغة العربية المتوافرة على شبكة الإنترنت على قرار اختبار العين، واختبار الكاب، وتحليلها من حيث المهارات اللغوية الأربع (القراءة، الاستماع، التحدث والكتابة، المفردات والتراكيب اللغوية)، ومستويات التقييم ونطاق الدرجات، ومقارنة هذه الاختبارات فيما بينها من جهة، ومقارنتها باختبار اللغة الإلكترونية TOEFL جهة أخرى بالاعتماد على مجموعة من المعايير المختارة منها: طبيعة الاختبار، الجمهور المستهدف، المهارات اللغوية الأربع. أظهرت نتائج الدراسة وجود تشابه كبير بين الاختبارات، وكان تقييم مهارتي الكتابة والتحدث الأصعب، وتمحورت الاختلافات حول عدد مستويات الكفاءة اللغوية، فهي متعددة أكثر في اختبار التوفل، يليه اختبار الكاب العربي، وأخيراً يأتي اختبار العين.

وأجرى مرادكاسيموفا (2021) دراسة في أوزبكستان هدفت الكشف عن مستوى معرفة معلمي اللغة الإنجليزية في معاهد التعليم العالي بتقييم المهارات اللغوية. ولتحقيق أهداف الدراسة، تم استخدام المنهج الكمي النوعي، وبناء اختبار لتقييم المهارات اللغوية، بالإضافة إلى إجراء المقابلات مع المعلمين. تكونت عينة الدراسة من (103) معلماً من معلمي اللغة الإنجليزية. أظهرت نتائج الدراسة أن لدى المعلمين نقصاً في المعرفة بتقييم المهارات اللغوية، وأنهم لم يشتركوا في دورات تدريبية في التقييم اللغوي، وأنهم بحاجة إلى التدريب على المعرفة بتقييم المادة الدراسية، وأساليب تقييمها.

وقام دوستي وصفا (Doosti & Safa, 2021) بدراسة في إيران هدفت التحقق من عدالة اختبار تقييم اللغة الشفوية، والكشف عن أثر التدريب بالعدالة، والأخذ بعين الاعتبار توقعات المتقدمين للاختبار. ولتحقيق أهداف الدراسة، تم استخدام اختبار تقييم اللغة الشفوية (اللغة الإنجليزية كلغة أجنبية) على مرحلتين (قبل وبعد التدريب)، واستبانة حول عدالة اختبار تقييم اللغة الشفوية. تكونت عينة الدراسة من (31) متعلماً للغة الإنجليزية كلغة أجنبية. أظهرت نتائج الدراسة أن التدريب زاد من الموثوقية بين المتقدمين للاختبار، وبالتالي عزز ذلك عدالة القرارات على أساس درجات الاختبار، وأن توقعات المتقدمين للاختبار فيما يتعلق بعدالة الاختبار زادت من تصوراتهم العامة حول التقييم العادل للاختبار.

وقام (الخوالدة وبنو ياسين، 2022) بدراسة في الأردن هدفت الكشف عن مستوى معرفة معلمي اللغة العربية بتقييم المهارات اللغوية (الاستماع، التحدث، القراءة، الكتابة)، وما إذا كان ذلك يختلف باختلاف متغيري المؤهل العلمي وعدد سنوات الخبرة. ولتحقيق أهداف الدراسة، تم استخدام اختبار المعرفة بتقييم المهارات اللغوية، مكونة من أربعة اختبارات فرعية للمهارات (القراءة، الكتابة، الاستماع، التحدث). تكونت عينة الدراسة من (205) معلماً ومعلمة. أظهرت نتائج الدراسة أن مستوى معرفة معلمي اللغة العربية بتقييم المهارات اللغوية متدنٍ في الاختبار ككل، وفي المجالات جميعها، إذ جاءت المجالات حسب مستوى معرفتهم بتقييمها مرتبة كالآتي: القراءة، الكتابة، التحدث، الاستماع، وأشارت النتائج إلى عدم وجود فروق في تقديرات معلمي اللغة العربية لمستوى معرفتهم بتقييم المهارات اللغوية تعزى لمتغيري المؤهل العلمي وعدد سنوات الخبرة.

وأجرى هينجير وديبيلاك وستروبل (2022) دراسة في سويسرا هدفت الكشف عن الأداء التقاضي باستخدام طريقة أشجار راش بمقياس توقف معتمد على حجم الأثر. ولتحقيق أهداف الدراسة، تم جميع البيانات من المستجيبين والبحث عن نقاط الانقسام المثلى في المتغيرات المشتركة، وتحديد المجموعات الفرعية للمستجيبين. أظهرت نتائج الدراسة أن طريقة أشجار راش تستخدم اختبارات الدلالة الإحصائية، وبالتالي، من المرجح أن تقوم أشجار راش باعتبار تأثيرات الأداء التقاضي الصغيرة أنها مهمة وذات دلالة في العينات الأكبر، وهذا بدوره يؤدي إلى أشجار أكبر، تُقسم العينة إلى مجموعات فرعية أكثر، وقد يكون مرغوباً فيه أكثر هو الطريقة المدفوعة بحجم الأثر بدلاً من حجم العينة، ومن أجل تحقيق ذلك، تم تنفيذ معيار توقف إضافي معتمد على حجم الأثر.

بعد الاطلاع على الدراسات السابقة، يتضح بأن بعضها تناولت موضوع عدالة الاختبار، كدراسة يو وآخرون (2019) التي تضمنت عدالة اختبار إجادة اللغة الإنجليزية، ودراسة مقدم ونصير زاده (2020) والتي تناولت عدالة اختبار الفهم القرائي، ودراسة دوسي وصفا (2021) التي تناولت عدالة اختبار تقييم اللغة الشفوية. ولكن على الرغم من ذلك وحسب—

اطلاع الباحثان- لا يوجد دراسات عربية بشكل عام، وأردنية بشكل خاص بحثت في عدالة الاختبار. وقد أفاد الباحثان من الدراسات السابقة في التعرف إلى مفهوم عدالة الاختبار، وطرائق لتحقيق من ذلك، بالإضافة إلى الاطلاع على ما قد ترتبط به العدالة مع متغيرات أخرى؛ كتقييم تكافؤ الدرجات، وذلك كما ورد في دراسة يو وآخرون (2019). وتناولت بعض الدراسات الكفائية اللغوية، كدراسة (العسيري والزهراني، 2020)، ودراسة (صباح، 2022)، بينما تناولت دراسات أخرى المهارات اللغوية؛ كدراسة مرادكاسيموف، (و) (الخالدة وبنو ياسين، 2022)، ولم تتناول أي من الدراسات السابقة اختبار الكفائية اللغوية، وهذا ما يميز الدراسة الحالية عن الدراسات السابقة. ويلاحظ أن على الرغم من أن غالبية الدراسات السابقة أجريت في البيئات العربية، وأن دراسة (الخالدة وبنو ياسين، 2022) أجريت في البيئة الأردنية، وتضمنت المهارات اللغوية، إلا أنها لم تتناول اختبار الكفائية اللغوية الذي هدفت الدراسة الحالية للتحقق من عدالته. وقد أفاد الباحثان من الدراسات السابقة في التعرف إلى أهمية الكفائية اللغوية، والمهارات اللغوية لدى المتقدمين للاختبار، بالإضافة إلى الاستفادة من كيفية صياغة مشكلة الدراسة وأهميتها.

ويرى الباحثان أن لأهمية اختبار الكفائية اللغوية الذي تم طرحه ليكون النجاح فيه شرطاً في تعيين فإنه يجب أن تتوفر فيه كافة الخصائص السيكمترية التي تجعله من الاختبارات الوطنية التي تسعى المملكة الأردنية الهاشمية لتطبيقه بالشكل الصحيح والخروج بنتائج تساعد في اتخاذ القرارات المتعلقة بالتعيين، ومن أهم هذه الخصائص السيكمترية الصدق والثبات، والعدالة، وقد لوحظ عدم وجود دراسات وأبحاث -على حد علم الباحثين- تناولت موضوع عدالة اختبار الكفائية اللغوية، لذا جاءت الرغبة لدى الباحثين في إجراء هذه الدراسة؛ لتكون انطلاقة لدراسات أخرى في هذا المجال.

مشكلة الدراسة وأسئلتها

نظراً لأهمية الاختبارات في العملية التعليمية في مختلف المراحل، وتأثيرها بالعديد من العوامل الخارجية، فإن ذلك يتطلب من المختصين العمل على فحص كل ما يتعلق بهذه الاختبارات لجعلها أعلى جودة وعدالة وموثوقية، وذلك لتحقيق أهدافها بأفضل شكل ممكن. فعلى الرغم من الاهتمام الكبير في السنوات الأخيرة بالاختبارات التربوية والكشف عن عدالتها وفق متغيرات مختلفة، إلا أن هناك القليل من الأدبيات -على حد علم الباحثين- المتعلقة بالكشف عن عدالة اختبار الكفائية اللغوية، الذي يعقده مجمع اللغة العربية.

إذا كان من الضروري الكشف عن عدالة اختبار الكفائية اللغوية لما لها من أهمية كبيرة في تحقيق الأهداف التي يسعى الاختبار لتحقيقها، وبالتالي ينعكس إيجاباً على القرارات التي قد تتخذها الجهات المسؤولة فيما يتعلق بالمرشحين للتعيين بناءً على نتائج اختبار الكفائية اللغوية. وفي سياق آخر، وبالرجوع إلى الدراسات السابقة، يتضح قلة الدراسات العربية التي تناولت عدالة الاختبار والعوامل المؤثرة فيها -في حدود إطلاع الباحثين- وخاصة في البيئة الأردنية، وربما يعود هذا الأمر إلى عدم معرفة بعض الباحثين فيما يتعلق بمفهوم عدالة الاختبار، أو عدم معرفة البعض الآخر بخطوات وإجراءات التحقق من عدالة الاختبار، لذا هدفت هذه الدراسة الكشف عن مؤشرات عدالة اختبار الكفائية اللغوية بدلالة الأداء التفاضلي للفقرات وفقاً لمتغيرات مختارة. وبالتحديد تكمن مشكلة الدراسة في الإجابة عن السؤالين التاليين:

- السؤال الأول: الخصائص السيكومترية (معاملات الصعوبة، معاملات التمييز) لاختبار الكفائية اللغوية؟
- السؤال الثاني: هل تظهر فروقات اختبار الكفائية اللغوية أداءً تفضلياً تبعاً لمتغير النوع الاجتماعي والنموذج؟

أهداف الدراسة

تهدف هذه الدراسة التعرف إلى مؤشرات الصدق والثبات لاختبار الكفائية اللغوية وكذلك بمعاملات الصعوبة والتمييز لفقرات اختبار الكفائية اللغوية، ومعرفة الفقرات ذات الأداء التفاضلي لاختبار الكفائية اللغوية تبعاً لمتغيرات الدراسة (النموذج، النوع الاجتماعي)، وهل تحدث هذه الفقرات فرقاً جوهرياً تجعل من اختبار الكفائية متحيزاً لفئة معينة، وبالتالي يكون غير عادل لباقي الفئات.

أهمية الدراسة

تظهر أهمية الدراسة بأنها تتناول اختبار الكفائية اللغوية، كونه أحد أهم الاختبارات المعتمدة في المملكة الأردنية الهاشمية لأغراض التعيين في الدوائر والمؤسسات الحكومية. وتكمن أهمية هذه الدراسة في البحث عن مؤشرات عدالة اختبار الكفائية اللغوية، وفق منهجية علمية تستند إلى النظرية التقليدية في القياس ونظرية استجابة الفقرة، ونمذجة المعادلة البنائية. كما تظهر أهمية الدراسة من خلال التعرف إلى البناء العاملي لاختبار الكفائية اللغوية عبر اختلاف (النموذج، والنوع الاجتماعي)، وبالتالي يتوقع أن يتم في ضوء ذلك اتخاذ قرارات أكثر عدالة عندما يمنح الشخص الذي يجتاز هذا الاختبار فرصة للتعيين في أي من وزارات العمل في المملكة الأردنية الهاشمية.

التعريفات الاصطلاحية والإجرائية

اشتملت الدراسة على المصطلحات الآتية:

عدالة الاختبار: هي غياب التحيز، والمعاملة العادلة لجميع المتقدمين للاختبار. **مؤشرات عدالة الاختبار:** هي الأدلة التي تؤكد أن الاختبار يقدم معاملة متكافئة لجميع المتقدمين له، ويعتبر العديد من المتخصصين في القياس عدالة الاختبار مجالاً واسعاً، ويشمل الجودة في تصميم الاختبار، وإدارته والاستجابة عليه والتغطية المناسبة للمواد المرتبطة به، والعمل الكافي لضمان التحقق من صدق بناء الاختبار، وتكافؤ الفرص في التعلم والوصول إلى الاختبار، والعناصر التي تقيس فقط مستوى القدرة والكفاءة من دون أن تتأثر بالعوامل المتعلقة بخلفية المتقدمين للاختبار؛ أي جميع الأدلة المرتبطة بنظام الاختبار بأكمله. **التعريف الإجرائي:** جمع الأدلة المتعلقة بصدق الاختبار، والأدلة المتعلقة بغياب التحيز عن اختبار الكفائية اللغوية. **الكفائية اللغوية:** هي المقدرة اللغوية العامة المركبة، التي تتكامل فيها فروع اللغة العربية ومهاراتها، ويتم توظيفها في التفكير والتعبير، والتواصل مع الآخرين، والتحصيل الدراسي (السبع، 2017). **التعريف الإجرائي:** يُقصد بها في هذه الدراسة الاستخدام الصحيح لمهارات اللغة العربية (القراءة والكتابة والاستماع) من قبل المتقدمين للاختبار.

اختبار الكفائية اللغوية: هو اختبار يتم وضعه من قبل مجمع اللغة العربية الأردني، يهدف إلى قياس مستوى مهارات اللغة العربية لدى المعلمين الجدد، ويتم اعتماد الصياغات النهائية للأسئلة وأجوبتها النهائية، ويُعقد الاختبار مرة واحدة شهرياً على الأقل، ويكون الحد الأدنى لعلامة

النجاح في الاختبار (50%) (مجمع اللغة العربية الأردني، 2019). التعريف الإجرائي: يُقصد به في هذه الدراسة اختبار الكفاية اللغوية المنعقد من قبل مجمع اللغة العربية الأردني والذي يخضع له

الأداء التفاضلي: هو مؤشر إحصائي يتم استخدامه للتعبير عن الفروق في احتمال الاستجابة الصحيحة على الفقرة بين مجموعتين فرعيتين من المفحوصين، ممن لهم مستوى القدرة نفسه (العمرى وشطناوى، 2016). **التعريف الإجرائي:** يُقصد به في هذه الدراسة مؤشر إحصائي يستخدم للتعبير عن الفروق في احتمال الاستجابة الصحيحة على الفقرة بين مجموعتين فرعيتين من المختبرين لمجتمع إحصائي ما، ممن لهم نفس مستوى القدرة، تسمى إحداها بالمجموعة المرجعية وتكون ذات الحجم الأكبر، فيما تسمى الأخرى بالمجموعة المستهدفة. **طريقة مانتل - هانزل:** طريقة الكشف عن الفقرات التي لها أداءً تفاضلياً وفق مؤشر قيمة مانتل هانزل ودرجات الحرية ومستوى الخطأ، وذلك من خلال مقارنة أداء المجموعتين المرجعية والمستهدفة **التعريف الإجرائي:** يُقصد به مقارنة أداء المعلمين الجدد في اختبار الكفاية اللغوية، للكشف عن الأداء التفاضلي لفقرات الاختبار.

حدود الدراسة

اقتصرت الدراسة الحالية على مؤشرات عدالة اختبار الكفاية اللغوية التي يمكن التوصل إليها بطرائق حصائية كمية (كمؤشرات الصدق Validity، التحيز Bais)، واستنتجت مؤشرات عدالة اختبار الكفاية اللغوية التي يمكن التوصل إليها بطرائق حصائية نوعية (مؤشرات الوصول الإدارة والعواقب الاجتماعية).

الطريقة والإجراءات

منهج الدراسة

تم استخدام المنهج الوصفي التحليلي المقارن للتحقق من عدالة اختبار الكفاية اللغوية، بدلالة الأداء التفاضلي للفقرات وفقاً لمتغيرات النموذج والنوع الاجتماعي، وقد تم اعتماد نموذجين من نماذج اختبار الكفاية اللغوية، وذلك من أصل (11) نموذجاً.

مجتمع الدراسة وعينتها

تألف مجتمع الدراسة من جميع المتقدمين لاختبار الكفاية اللغوية لعام 2024/2023، وعددهم (4942)، وتم اختيار عينة ممثلة لمجتمع الدراسة بالطريقة الطباقية العشوائية من مجتمع الدراسة الكلي، وتم الأخذ بعين الاعتبار متغيرات الدراسة (النموذج، النوع الاجتماعي) عند اختيار العينة لضمان تمثيل عينة الدراسة لمجتمع الدراسة، وبلغ عدد العينة (1161)، شكلوا (23.5%) من مجتمع الدراسة.

أداة الدراسة

أداة الدراسة باختبار الكفاية اللغوية، والمطروح من قبل مجمع اللغة العربية الأردني، ويتضمن أربعة أقسام؛ وهي: الفهم والاستيعاب، التراكيب اللغوية، التراكيب النحوية، المعجم والدلالة، ويطلب الاختبار شرطاً للتعين من كل معلم في وزارة التربية والتعليم أو عضو هيئة تدريس في الجامعات أو معد أو محرر أو مذيع في وسائل الإعلام، وعلامة النجاح (50%) لكافة التخصصات، وهو اختبار إلكتروني، مدته ساعتان، ونوع الأسئلة اختيار من متعدد، ولا يجوز

أن يُعقد عن بُعد، وعند التسجيل يتحمل المتقدم مسؤولية بياناته المدخلة عند التسجيل للاختبار، خاصةً رقم الهاتف ومكان السكن ويعطى المتقدم عدد مفتوح من مرات التقدم للاختبار إلى أن يصل إلى علامة النجاح المحددة.

يتكون امتحان الكفاية في اللغة العربية من أربعة مستويات، ويقع في ستين سؤال اختيار من متعدد، المستوى الأول (الفهم والاستيعاب)، وهو عبارة عن نص قطعة وعليها عشرة أسئلة نسبتهم من الامتحان (38%). أما المستوى الثاني فهو يمثل تراكيب لغوية، وهذا المستوى قواعد اللغة العربية بشكل عام وبعض الأخطاء اللغوية، يتكون عادةً من اثنين وعشرين سؤالاً، صيغة السؤال (اختر العبارة الصحيحة لغوياً)، نسبة كل سؤال (1.2) من العلامات نسبة هذا المستوى (5.26%).

والمستوى الثالث يتضمن تركيبين لغويين، مضمون هذا المستوى قواعد اللغة العربية بشكل عام، يتكون عادةً من واحد وعشرين سؤالاً، صيغة السؤال (اختر العبارة التي ورد فيها خطأ لغوي أو نحوي أو صرفي أو إملائي)، نسبة كل سؤال (1.26)، نسبة هذا المستوى (26.5%). أما المستوى الرابع فهو يتعلق بالمعجم والدلالة، مضمون هذا المستوى معاني الكلمات وبعض علامات الترقيم، صيغة السؤال (املأ الفراغ في الكلمة المناسبة)، نسبة هذا المستوى (9%)، عدد الأسئلة في هذا المستوى كما يلي: إذا كان عددها (6) أسئلة، فإن كل سؤال يساوي (1.5) علامة، إذا كان عدد الأسئلة (5)، فإن كل سؤال يساوي (1.8)، إذا كان عدد الأسئلة (4) فإن كل سؤال يساوي (2.25)، إذا كان عدد الأسئلة (7)، فإن كل سؤال يساوي (1.28)، إذا كان عدد الأسئلة (8) فإن كل سؤال يساوي (1.1)، وبالتالي ستكون علامات الامتحان كالتالي: $(100 = 9 + 26.5 + 0.38)$.

النسب أعلاه تُقسم على عدد الأسئلة في كل مستوى، وبالتالي تحصل على نسبة سؤال.

إجراءات تنفيذ الدراسة

- الحصول على كتاب تسهيل مهمة موجه من جامعة اليرموك إلى مجمع اللغة العربية، الذي قام بدورة بالموافقة على منح الباحثين جميع البيانات اللازمة للبحث، علماً بأن كل الدرجات بالتفاصيل والمواصفات المشار إليها وجميع البيانات المطلوبة متوفرة في وزارة التربية والتعليم - قسم الامتحانات بقاعدة بيانات مخصصة للاختبار لصالح مجمع اللغة العربية.
- الحصول على كتاب تسهيل مهمة موجه من جامعة اليرموك إلى وزارة التربية والتعليم.
- قسم الامتحانات، وتم تزويد الباحثين بالبيانات اللازمة على شكل ملف اكسيل.
- إدخال البيانات لبرمجية SPSS، مع مراعاة المحافظة على درجات كل نموذج.
- تنقية البيانات وحذف المتغيرات والبيانات التي ليست لها علاقة بالدراسة.
- اختيار نموذجين من نماذج اختبار الكفاية اللغوية من أصل (11) نموذج - أعدت لهذا الاختبار - بطريقة عشوائية.
- تقسيم البيانات كاملة إلى مجموعتين حسب متغير النوع الاجتماعي، مجموعة ذكور ومجموعة إناث، وتخزينها.
- إنشاء ملف البيانات وتدقيقه.

متغيرات الدراسة التصنيفية

أولاً: المتغيرات المستقلة:

- نموذج الاختبار، وله فئتان (نموذج 4، نموذج 6)
- النوع الاجتماعي، وله فئتان (ذكر، أنثى).
- ثانياً: المتغير التابع
- مؤشرات عدالة اختبار الكفائية اللغوية
- الأداء التفاضلي.

نتائج الدراسة ومناقشتها

أولاً: النتائج المتعلقة بالسؤال الأول ومناقشتها: "ما الخصائص السيكومترية (معاملات الصعوبة، معاملات التمييز) لاختبار الكفائية اللغوية؟" قبل الإجابة عن هذا السؤال، قام الباحثان بحساب معاملات ارتباط الفقرة بالمجال الذي تنتمي إليه، ومعاملات ارتباط بويونت بايسيرال لفقرات اختبار الكفائية اللغوية، حيث لوحظ أن معاملات الارتباط بين مجالات النماذج تقترب من بعضها البعض، وجميعها ذات دلالة إحصائية. وللتحقق من الخصائص السيكومترية للاختبار، تم حساب معامل الصعوبة ومعامل التمييز لفقرات اختبار الكفائية اللغوية كما في جدول (1):

جدول (1): معاملات الصعوبة ومعاملات التمييز لفقرات اختبار الكفائية اللغوية حسب النموذجين (4، 6)

6		4		النماذج	6		4		النماذج
a	B	a	B	رقم الفقرة	a	b	A	B	رقم الفقرة
0.42	0.38	0.22	0.25	31	0.16	0.93	0.10	0.90	1
0.20	0.28	0.14	0.44	32	0.12	0.94	0.09	0.80	2
0.11	0.69	0.46	0.56	33	0.07	0.87	0.22	0.83	3
0.29	0.52	0.39	0.63	34	0.04	0.96	0.14	0.26	4
0.14	0.34	0.37	0.56	35	0.14	0.94	0.10	0.52	5
0.23	0.28	0.22	0.35	36	0.14	0.96	0.16	0.83	6
0.34	0.61	0.28	0.51	37	0.10	0.87	0.19	0.90	7
0.38	0.50	0.47	0.59	38	0.08	0.95	-0.01	0.40	8
0.26	0.58	0.43	0.44	39	0.11	0.96	0.08	0.82	9
0.17	0.38	0.30	0.45	40	0.16	0.95	0.13	0.92	10
0.19	0.28	0.27	0.53	41	0.03	0.13	0.45	0.63	11
0.32	0.60	0.47	0.47	42	0.27	0.62	0.47	0.58	12
0.21	0.50	0.50	0.52	43	0.17	0.49	0.33	0.50	13
0.33	0.61	0.20	0.50	44	0.29	0.71	0.43	0.45	14
0.24	0.54	0.33	0.69	45	0.27	0.45	0.33	0.52	15
0.36	0.55	0.23	0.83	46	0.35	0.59	0.39	0.61	16
0.24	0.63	0.10	0.23	47	0.33	0.52	0.26	0.58	17
0.29	0.52	0.05	0.47	48	0.34	0.51	-0.10	0.08	18
0.13	0.51	0.36	0.54	49	0.05	0.38	0.39	0.65	19
0.29	0.34	0.42	0.31	50	0.26	0.75	0.11	0.36	20
0.28	0.56	0.32	0.57	51	0.36	0.59	0.11	0.53	21
0.20	0.16	0.18	0.72	52	0.35	0.28	0.43	0.38	22
0.23	0.65	0.38	0.55	53	0.33	0.67	0.34	0.77	23
0.29	0.40	0.32	0.52	54	0.40	0.71	0.25	0.56	24

النماذج	4	6	النماذج	4	6	النماذج	4	6
25	0.02	0.51	55	0.37	0.89	0.18	0.78	0.25
26	0.27	0.42	56	0.26	0.97	0.04	0.18	0.21
27	0.27	0.35	57	0.25	0.98	0.18	0.43	0.22
28	0.16	0.43	58	0.01	0.87	0.03	0.98	0.13
29	0.30	0.38	59	0.25	0.96	0.09	0.98	0.12
30	0.17	0.54	60	0.07	0.93	0.16	0.57	0.25
المتوسط الحسابي								
	0.68	0.17	0.58	0.25	0.58	0.22	0.58	0.22

يتضح من البيانات الواردة في الجدول (1) أن معاملات الصعوبة لفقرات اختبار الكفاية اللغوية للنموذج (4) تراوحت بين 0.08 و 0.98، وللنموذج 6 تراوحت بين 0.13 و 0.98، وتشير النتائج إلى وجود فروق ظاهرية بين المتوسطات الحسابية لمعاملات صعوبة الفقرات بحسب النموذجين (4،6)؛ وبهدف الكشف عن الدلالة الإحصائية لهذه الفروق، فقد تم استخدام تحليل التباين الأحادي (One Way ANOVA) للمتوسطات الحسابية لمعاملات صعوبة الفقرات بحسب النموذجين (4، 6). والجدول (2) يبين ذلك.

جدول (2): تحليل التباين الأحادي لأثر (النموذجين) على معاملات صعوبة فقرات اختبار الكفاية اللغوية

مصدر التباين	مجموع المربعات	درجات الحرية	متوسط المربعات	قيمة ف	الدلالة الإحصائية
بين المجموعات	0.023	3	0.008	0.172	0.915
داخل المجموعات	10.647	236	0.045		
الكل	10.670	239			

يتبين من جدول (2) عدم وجود فروق ذات دلالة إحصائية ($\alpha=0.05$) تعزى لأثر النموذج في معاملات الصعوبة. ويمكن عزو هذه النتيجة إلى استخدام مجمع اللغة العربية النماذج المتعلقة باختبار الكفاية اللغوية بشكل صحيح ومناسب للمتقدمين للاختبار، بالإضافة إلى أن الأسئلة في النموذجين تم وضعها من قبل متخصصين في اللغة العربية، وقد تمت مراعاة الصعوبة والسهولة في طرح الأسئلة، جعل غالبية المتقدمين على دراية في كيفية الإجابة الصحيحة على معظم الأسئلة في هذه النماذج، وهذا أدى إلى عدم وجود فروق في معاملات الصعوبة. ويرى الباحثان أن مجمع اللغة العربية أولى اختبار الكفاية اللغوية اهتماماً كبيراً كون نتيجته ترتبط ارتباطاً وثيقاً بقرارات حاسمة تتعلق بالمتقدمين للاختبار، وبناءً على نتائجه يتم تعيين المعلم الجديد. وقد أوضح تايلو (Taylor, 2009) أن اختبارات الكفاية اللغوية تلعب دوراً مهماً بشكل متزايد في السياقات التعليمية أو الاجتماعية؛ حيث يتم اتخاذ مجموعة من القرارات المهمة والحاسمة (كالتعليم والتعيين، والعمل، والهجرة، والمواطنة) بناءً على نتائج هذا الاختبار الذي يُعد من المعايير الأساسية لتحديد المستويات والقدرات اللغوية للأفراد في ضوء خبراتهم وتجاربهم السابقة، والتي تتراوح من القبول إلى المستوى الأعلى. وبالتالي، ينبغي إيلاء اهتمام شامل للنظر في مراحل بناء اختبارات الكفاءة اللغوية، والتأكد من صدقها وثباتها وفعاليتها من أجل إصدار أحكام عادلة تخلو بالحد الأدنى من الأخطاء.

أما معاملات التمييز لفقرات اختبار الكفاية اللغوية، فقد أظهرت نتائج جدول (1) أن معاملات التمييز لفقرات الاختبار تراوحت بين 0.02 و 0.02 ووفقاً للنموذج (6) تراوحت بين 0.01 و 0.42 وتشير النتائج إلى وجود فروق ظاهرية بين المتوسطات الحسابية لمعاملات التمييز لفقرات بحسب النموذجين (4، 6)؛ وبهدف الكشف عن الدلالة الإحصائية لهذه الفروق، فقد تم استخدام تحليل التباين الأحادي (One Way ANOVA) للمتوسطات الحسابية لمعاملات تمييز الفقرات بحسب النموذجين (4، 6) والجدول (3) يبين ذلك.

جدول (3): تحليل التباين الأحادي لأثر (النماذج) على معاملات التمييز فقرات اختبار الكفاية اللغوية

مصدر التباين	مجموع المربعات	درجات الحرية	متوسط المربعات	قيمة ف	الدلالة الإحصائية
بين المجموعات	0.121	3	0.040	2.479	0.062
داخل المجموعات	3.849	236	0.016		
الكل	3.970	239			

يتبين من جدول (3) عدم وجود فروق ذات دلالة إحصائية ($\alpha=0.05$) تعزى لأثر النموذجين (4، 6). ويمكن تفسير هذه النتيجة في ضوء اهتمام مجمع اللغة العربية في اقتناء أسئلة الاختبار بكل دقة ومصداقية وتراعي الفروق الفردية بين المتقدمين للاختبار، وتميز القوي من الضعيف في اللغة العربية ومهاراتها المختلفة، وهذا يجعل المتقدمين الأقوياء في اللغة العربية قادرين على الإجابة بكل سهولة عن الأسئلة، بغض النظر عن عدم قدرة المتقدمين للاختبار الضعاف في اللغة العربية على الإجابة بسهولة عن لأسئلة نفسها في النموذج نفسه. وقد أشارت ويبرج (Wiberg, 2007) إلى أن بناء الاختبار بعد التخطيط له وتطوير فقراته وتدقيقها وتحليلها بأساليب مختلفة يتيح انتقاء أفضل الفقرات لتضمينها في النسخة النهائية من الاختبار، ومن ثم كتابة الفقرات بدقة لضمان الحصول على أسئلة غير غامضة وواضحة للمتعلم، فصدق الاختبار يعتمد على جودة الفقرات. وتم حساب معامل كرونباخ ألفا، ومعامل التجزئة النصفية؛ للتحقق من خاصية الثبات لاختبار الكفاية اللغوية للنموذجين، وكذلك تم حساب معامل ألفا الطبقي باستخدام المعادلة (1)، حيث كانت النتائج كما في الجدول (4):

$$strata \rho_{xx}' = 1 - \frac{\sum \sigma^2_{x_i}(1-\alpha\rho_{xx})}{\sigma^2_x} \dots\dots\dots (1)$$

جدول (4): قيم معاملات كرونباخ ألفا وفق النموذجين (4، 6)

رقم النموذج	معامل كرونباخ ألفا	معامل التجزئة النصفية	ألفا الطبقي
4	.833	.760	0.907
6	.808	.799	0.898

يظهر من النتائج الواردة في جدول (4) أن معامل الثبات بلغ (.833) باستخدام النموذج 4، وفيما يتعلق بالنموذج 6 فقد بلغ معامل الثبات (.808)، وتشير جميعها إلى ثبات الاختبار بدرجة مرتفعة؛ حسب ما أشار إليه فيلدت وبرنان (Feldt & Bernnan, 1989)؛ إلى أن

قيمته مقبولة عند (70). وما فوق. ويرى الباحثان أن هذه النتيجة تشير إلى ما يتميز به اختبار الكفاية اللغوية من ثبات تجعله صالحاً ومناسباً للهدف الذي وضع من أجله، وهو الكشف عن المهارات والقدرات اللغوية لدى المتقدمين للاختبار الذين ينتظرون وظيفة المستقبل. واتفقت هذه النتيجة مع نتيجة دراسة العسيري والزهراي (2020)، التي أظهرت أن الاختبار المحكي المرجع لقياس الكفايات اللغوية يتمتع بصدق وثبات جيدين، وذلك وفقاً لمعامل ليفجستون، حيث وصل إلى درجة (85%).

النتائج المتعلقة بالسؤال الثاني ومناقشتها: "هل تظهر فقرات اختبار الكفاية اللغوية أداءً تفاضلياً تبعاً لمتغيرات (النموذج، النوع الاجتماعي)؟"
للإجابة عن السؤال الثاني، تم إيجاد المتوسطات الحسابية والانحرافات المعيارية، لفقرات الاختبار تبعاً للمتغيرات (النوع الاجتماعي، النموذج)، كما هو موضح في جدول (5).

جدول (5): المتوسطات الحسابية والانحرافات المعيارية لفقرات الاختبار تبعاً للمتغيرات (النوع الاجتماعي، النموذج)

الانحراف المعياري	المتوسط الحسابي	التصنيف	المتغير
7.66	32.69	ذكر	النوع الاجتماعي
8.16	36.91	أنثى	
8.16	34.95	4	النموذج
7.50	34.69	6	

يظهر من النتائج في جدول (5) وجود اختلافات بين الأوساط الحسابية والانحرافات المعيارية تبعاً لمتغيرات الدراسة (النوع الاجتماعي، النموذج)، وللكشف عن الأداء التفاضلي بين الذكور والاناث، تم إيجاد اختبار ماننل هانزل، وكانت النتائج كما في الجدول (6).

جدول (6): نتائج اختبار ماننل هانزل حسب متغير الدراسة النوع الاجتماعي

رقم الفقرة	ماننل هانزل	الدلالة	نسبة الأرجحية	هل يوجد أداء تفاضلي	رقم الفقرة	ماننل هانزل	الدلالة	نسبة الأرجحية	هل يوجد أداء تفاضلي
1	0.791	0.374	0.658	لا	31	0.055	0.815	0.846	لا
2	0.143	0.705	0.792	لا	32	0.115	0.734	0.859	لا
3	0.042	0.837	0.812	لا	33	0.293	0.588	0.870	لا
4	0.176	0.675	0.787	لا	34	4.860	0.027	0.646	نعم
5	4.795	0.029	1.028	نعم	35	0.002	0.963	0.828	لا
6	2.868	0.090	0.580	لا	36	0.057	0.811	0.848	لا
7	0.427	0.513	0.701	لا	37	0.313	0.576	0.769	لا
8	0.064	0.800	0.786	لا	38	4.695	0.030	1.028	نعم
9	0.214	0.643	0.777	لا	39	0.241	0.624	0.867	لا
10	0.210	0.647	0.696	لا	40	3.056	0.080	0.696	لا

رقم الفقرة	مانتل هانزل	الدلالة	نسبة الأرجحية	هل يوجد أداء تفاضلي	رقم الفقرة	مانتل هانزل	الدلالة	نسبة الأرجحية	هل يوجد أداء تفاضلي
11	5.616	0.018	1.045	نعم	41	1.294	0.255	0.728	لا
12	4.236	0.040	0.671	نعم	42	0.055	0.814	0.795	لا
13	0.475	0.491	0.886	لا	43	1.357	0.244	0.730	لا
14	0.032	0.858	0.842	لا	44	2.324	0.127	0.963	لا
15	3.915	0.048	0.669	نعم	45	0.035	0.852	0.846	لا
16	0.041	0.839	0.799	لا	46	0.044	0.833	0.852	لا
17	0.663	0.415	0.761	لا	47	15.292	0.000	1.213	نعم
18	1.591	0.207	0.937	لا	48	15.158	0.000	1.206	نعم
19	0.597	0.440	0.762	لا	49	4.590	0.032	1.023	نعم
20	0.000	0.985	0.823	لا	50	7.015	0.008	0.630	نعم
21	2.163	0.141	0.713	لا	51	0.165	0.685	0.863	لا
22	3.590	0.058	0.678	لا	52	0.103	0.749	0.855	لا
23	0.022	0.881	0.815	لا	53	1.525	0.217	0.721	لا
24	0.842	0.359	0.741	لا	54	0.566	0.452	0.759	لا
25	0.224	0.636	0.787	لا	55	0.379	0.538	0.760	لا
26	12.968	0.000	1.180	نعم	56	0.757	0.384	0.893	لا
27	0.686	0.408	0.748	لا	57	0.258	0.611	0.848	لا
28	4.658	0.031	0.664	نعم	58	1.076	0.300	0.891	لا
29	1.071	0.301	0.917	لا	59	0.080	0.777	0.682	لا
30	5.429	0.020	0.659	نعم	60	10.284	0.001	1.164	نعم

يتبين من جدول (6) أن هناك (13) فقرة تظهر أداء تفاضلياً باستخدام مستوى دلالة (0.05) وهي تشكل ما نسبته (21.67%) من فقرات الاختبار. أي بمعنى آخر ذات أداء تفاضلي لصالح الإناث أو الذكور، مما يعني أن لدى نوع اجتماعي معين مهارات لغوية أعلى من النوع الاجتماعي الآخر، تتمثل في استخراج الدلالات والمعاني واستخدام المعجم، بالإضافة إلى معرفة القواعد والتراكيب اللغوية، والتراكيب النحوية وما يتعلق بالنحو، والفهم والاستيعاب أيضاً.

ويمكن تفسير هذه النتيجة إلى الدقة في إعداد مفردات اختبار الكفاية اللغوية، ومراجعتها وفق معايير دقيقة؛ كمطابقتها لمستويات الطلبة، ومناسبتها للمساقات الدراسية الجامعية، والدقة في صياغة المموهات، بحيث لا يكون لها دور واضح في إظهار الأداء التفاضلي لنوع اجتماعي على حساب نوع اجتماعي آخر، ومراعاة الفروق الفردية الخاصة بثقافة كل فرد، أضف إلى

ذلك الاختلاف في الكلمات المستخدمة في مفردات الاختبار، وملاءمتها لكل من الذكور والإناث.

ويظهر الأداء التفاضلي للفقرة عندما يختلف أداء المفحوصين المتساوين (المتقدمين للاختبار) في مستوى المقدرة، والذين ينتمون إلى مجموعات مختلفة، بغض النظر عن أي متغير معين؛ كالتنوع الاجتماعي، وهذا ما أشار إليه بارك (2010, Park).
وللكشف عن الأداء التفاضلي وفق متغير النموذج (4، 6)، تم إيجاد اختبار ماننل هانزل، وكانت النتائج كما في الجدول (7).

جدول (7): نتائج اختبار ماننل هانزل حسب متغير النموذج (4، 6)

رقم الفقرة	ماننل هانزل	الدلالة	نسبة الأرجحية	هل يوجد أداء تفاضلي	رقم الفقرة	ماننل هانزل	الدلالة	نسبة الأرجحية	هل يوجد أداء تفاضلي
1	4.550	0.033	1.059	نعم	31	30.237	0.000	1.641	نعم
2	38.826	0.000	2.358	نعم	32	28.067	0.000	0.386	نعم
3	2.183	0.140	0.930	لا	33	21.390	0.000	1.408	نعم
4	546.028	0.000	36.224	نعم	34	12.683	0.000	0.476	نعم
5	227.355	0.000	9.793	نعم	35	58.679	0.000	0.286	نعم
6	44.084	0.000	2.816	نعم	36	5.928	0.015	0.544	نعم
7	3.013	0.083	0.481	لا	37	14.688	0.000	1.284	نعم
8	372.785	0.000	20.535	نعم	38	6.882	0.009	0.528	نعم
9	42.613	0.000	2.771	نعم	39	29.231	0.000	1.583	نعم
10	2.555	0.110	0.940	لا	40	4.543	0.033	0.590	نعم
11	310.167	0.000	0.060	نعم	41	68.556	0.000	0.250	نعم
12	3.759	0.053	1.004	لا	42	28.618	0.000	1.596	نعم
13	0.271	0.603	0.723	لا	43	0.123	0.726	0.734	لا
14	102.364	0.000	3.204	نعم	44	17.385	0.000	1.341	نعم
15	7.330	0.007	0.544	نعم	45	29.934	0.000	0.369	نعم
16	0.274	0.601	0.709	لا	46	106.974	0.000	0.161	نعم
17	3.300	0.069	0.608	لا	47	182.548	0.000	4.551	نعم
18	259.978	0.000	7.903	نعم	48	2.553	0.110	0.962	لا
19	85.926	0.000	0.243	نعم	49	0.661	0.416	0.703	لا
20	174.810	0.000	4.364	نعم	50	2.014	0.156	0.934	لا
21	3.116	0.078	0.983	لا	51	0.003	0.959	0.773	لا
22	7.675	0.006	0.499	نعم	52	355.446	0.000	0.045	نعم
23	14.372	0.000	0.426	نعم	53	17.259	0.000	1.349	نعم
24	29.736	0.000	1.623	نعم	54	16.733	0.000	0.449	نعم
25	7.370	0.007	1.102	نعم	55	26.190	0.000	0.280	نعم
26	0.590	0.442	0.697	لا	56	742.201	0.000	0.002	نعم
27	0.031	0.861	0.745	لا	57	438.743	0.000	0.003	نعم
28	6.944	0.008	1.095	نعم	58	47.198	0.000	4.030	نعم
29	1.554	0.213	0.654	لا	59	1.771	0.183	0.836	لا
30	19.995	0.000	0.440	نعم	60	198.280	0.000	0.058	نعم

يتبين من جدول (7) أن هناك 43 فقرات تظهر أداء تفاضلياً باستخدام مستوى دلالة (0.05) وهي تشكل ما نسبته 71.67% من فقرات الاختبار. ويمكن عزو هذه النتيجة إلى استخدام فقرات غير مألوفة ومعروفة لدى أفراد عينة الدراسة، قد تكون ذات معنى مزدوج أو غير واضح لديهم، وخاصة مع وجود بعض الفروق الفردية المتعلقة بالثقافة اللغوية، وطبيعة

المهارات اللغوية التي يمتلكها كل فرد، بالإضافة إلى اختلاف مكان السكن الذي قد يؤثر على بعض المصطلحات والمفاهيم اللغوية. وقد اشار ميكينغ (Mikyung, 2001) إلى أن ظهور أداء تفاضلي لفقرة ما من حيث اللغة قد يكون سببه استخدام فقرات غير مألوفة لدى أفراد إحدى المجموعات في مضمون الفقرة، وهذه الفقرات لها معان مختلفة وفق مناطق معينة.

التوصيات

بناءً على النتائج توصي الدراسة

- استخدام طرائق تستند إلى نظرية استجابة المفردة للكشف عن الأداء التفاضلي لفقرات اختبار الكفاية اللغوية.
- إجراء معادلة لجميع الصور الاختبارية الفرعية لاختبار الكفاية اللغوية لضمان تحقق مؤشر العدالة للاختبار.
- تطوير إطار مرجعي عام قبل الشروع في بناء اختبار الكفاية اللغوية، وتوضيح مواصفاته والمخرجات المتوقعة في كل مستوى للمهارات اللغوية، وجميع المتطلبات اللغوية.
- وضع أسس واضحة ودقيقة للتقييم، وذلك من خلال توضيح الإجراءات التي يجب اتباعها في كل مرحلة من مراحل الاختبار، تضمن صدقه وثباته.
- تكوين لجان بناء اختبار الكفاية اللغوية، تتكون من مختصين في اللغة العربية، ومختصين في القياس والتقويم، ومختصين أيضاً في تقنيات تعليم اللغات وتقييمها.

أولاً: المراجع العربية

- بوحفص، عبد الكريم (2011). أسس ومناهج البحث في علم النفس. الجزائر: ديوان المطبوعات الجامعية.
- الخوالدة، محمد وبني ياسين، رشا (2022). مستوى معرفة معلمي اللغة العربية في الأردن بتقييم المهارات اللغوية. المجلة الدولية للبحوث النفسية والتربوية، 1(2)، 206-225.
- زيدان، جميلة وبوجردة، محمد (2017). الخصائص السيكومترية لأدوات القياس النفسي والتربوي والاجتماعي. مجلة السراج في التربية وقضايا المجتمع، 1، 209-216.
- السبع، سعاد (2017). اختبار الكفاية اللغوية في مهارات اللغة العربية اللازمة للالتحاق بالدراسة الجامعية. المجلة الدولية للبحوث التربوية، جامعة الإمارات، 41(4)، 187-227.
- الشريفين، نضال (2018). أثر نوع الأداء التفاضلي للفقرات على الخصائص السيكومترية للفقرات والاختبار وفق النماذج المعلمية والنماذج اللامعلمية لنظرية الاستجابة للفقرات. دراسات، العلوم التربوية، 45(4)، ملحق (2)، 605-632.
- صباح، براهيم (2020). دراسة تحليلية لعينة من اختبارات الكفاءة اللغوية الإلكترونية للغة العربية ومقارنتها بمثيلاتها للغة الإنجليزية – التوفل نموذجاً. اللسانيات، 2(26)، 228-251.
- عبد الوهاب، محمد (2014). الكشف عن الأداء التفاضلي على مفردات مقياس أساليب المعاملة الوالدية بين الآباء والأمهات وبين الأبناء من الجنسين. دراسات عربية في التربية وعلم النفس، السعودية، 4(45)، 155-178.
- العجلوني، جهاد والسوالمه، يوسف (2016). فحص الأداء التفاضلي لمموهات فقرات اختبار الرياضيات في التقييم الوطني لاقتصاد المعرفة وفقاً لجنس الطالب وموقع المدرسة. مجلة جامعة النجاح للأبحاث (العلوم الإنسانية)، 30(11)، 2265-2286.
- العسيري، إبراهيم والزهراني، محمد (2020). بناء اختبار تحصيلي محكي المرجع لقياس الكفايات اللغوية – دراسة وصفية مسحية على طلاب المرحلة الثانوية بمدينة الرياض - . مجلة العلوم التربوية والنفسية، 4(40)، 179-194.
- علام، صلاح الدين (2005). نماذج الاستجابة للمفردة الاختبارية أحادية البعد ومتعددة الأبعاد وتطبيقاتها في القياس النفسي والتربوي. القاهرة: دار الفكر العربي.
- العمرى، حسان وشطناوي، عبد الله (2016). الأداء التفاضلي لفقرات الاختبار الوطني لضبط نوعية التعليم في الرياضيات للصف العاشر تبعاً لمتغير الجنس. مجلة جامعة النجاح للأبحاث (العلوم الإنسانية)، 30(8)، 1530-1554.
- مجمع اللغة العربية الأردني (2018). توضيحات مهمة بشأن امتحان الكفاية في اللغة العربية. عمان، الأردن.
- مجمع اللغة العربية الأردني (2019). نظام امتحان الكفاية في اللغة العربية. عمان، الأردن.
- الوابل، وديان وتعلب، صبرين (2023). الأداء التفاضلي لفقرات اختبار الذكاء الثلاثي تبعاً لمتغيري الجنس والتخصص لدى طلاب جامعة قصيم. مجلة الآداب للدراسات النفسية والتربوية، جامعة دمار، 5(1)، 59-102.

ثانياً: المراجع الأجنبية

- Bert, D. & Stoneberg, J. (2004). A Study of Gender-Based and Ethnic-Based Differential Item Functioning (DIF) in the Spring 2003 Idaho Standards Achievement Tests Applying the Simultaneous Bias Test (SIBTEST) and the Mantel-Haenszel Chi Square Test. Graduate Certificate Program in Large-Scale Education Assessment, The University of Maryland (UM) Measurement.
- Camilli, G. (2006). Test fairness. In R. L. Brennan (Ed.) Educational Measurement (4th ed., Vol. 4, pp. 221- 256). Westport: American Council on Education & Praeger Publishers.
- Doosti, M. & Safa, M. (2021). Fairness in Oral Language Assessment: Training Raters and Considering Examinees' Expectations. Tabaran Institute of Higher Education ISSN 2476-5880 International Journal of Language Testing, 11(2), 64-90.
- Feldt, L. & Brennan, R. (1989). Reliability. In Educational Measurement (3rd). My current thoughts on coefficient alpha and successor procedures (CSE Report 643). Los Angeles, Ca: University Macmillan Publishing Company.
- Fidalgo, A. & Madeira, J. (2008). Generalized Mantel-Haenszel methods differential item functioning in polytomous items. Journal of Educational Measurement, (33)8, 315-332. <https://doi.org/10.1177/0013164408315265>
- Henninger, M., Debelak, R. & Strobl, C. (2022). A New Stopping Criterion for Rasch Trees Based on the Mantel-Haenszel Effect Size Measure for Differential Item Functioning. Educational and Psychological Measurement, 83(1), 1-32, <https://doi.org/10.1177/00131644221077135>
- Kamata, A. & Vaughn, B. (2004). An Introduction to item functioning analysis. A Contemporary Journal, 2(2), 49-62.
- Komboz, B., Strobl, C. & Zeileis, A. (2018). Treebased global model tests for polytomous Rasch models. Educational and Psychological Measurement, 78(1), 128-166, DOI: 10.1177/0013164416664394.
- Kunnan, A. (2004). Test fairness. In M. Milanovic and C. Weir (Eds.), European language testing in a global context (pp.27-48). Cambridge: Cambridge University Press.
- Kunnan, A. (2018). Assessing Languages for Specific Purposes. In Presentation at the ALTE Cluj Meeting. Available online at: [https://www. Alte. Org/resources/Documents/ALTE% 20talk% 20on% 20LSP, % 20April \(Vol. 202018\)](https://www.Alte.Org/resources/Documents/ALTE%20talk%20on%20LSP,%20April(Vol.202018)).
- Mikyung, K. (2001). Detecting DIF across the different language groups in a speaking test. Language Testing, 18 (1), 89-114.
- Moghadam, M. & Nasirzadeh, F. (2020). The application of Kunnan's test fairness framework (TFF) on a reading comprehension test. Language Testing in Asia, 10(7), 1-21
- Muradkasimova, K. (2021). The importance of developing teachers' assessment literacy competence Philology Matters

Park, C. (2010). Examining the relationship between deferential item functioning and differential test functioning. *Language Testing* 23(4), 475 - 496, <https://doi.org/10.1191/0265532206lt338oa>.

Taylor, L. (2009). .Developing assessment literacyAnnual Review of Applied Linguistics, 29(7), 21-36.

Wiberg, M. (2007). Measuring and detecting differential item functioning in criterion-referenced licensing test: A theoretic comparison of methods. UMBA University.

Yoo, H., Manna, V., Monfils, L. & Oh, H. (2019). Measuring English language proficiency across subgroups: Using score equity assessment to evaluate test fairness.

Sega Journals, 36(2), 289-309, <https://doi.org/10.1177/0265532218776040>.