التّحليل البعديّ للدّراسات التي قارنت بين طريقتي نسبة الأرجحيّة ومانتل- هانزل للكشف عن الأداء التفاضليّ للفقرات

كوثر عبد القادر بشارات و يوسف محمد السوالمة*

Doi: //10.47015/20.3.1

تاريخ تسلم البحث: 2022/12/7 تاريخ قبوله: 2023/1/19

ملخص: هدفت الدّراسة إلى مراجعة الدّراسات التي قارنت بين طريقتي نسبة الأرجحيّة ومانتل هانزل للكشف عن الأداء التفاضليّ للفقرات DIF التي أجيزت في الفترة الممتدة من 1980 إلى 2021م، وإخضاعها للتّحليل البعديّ. وقد بلغ عدد الدّراسات التي حققت شروط ومحكّات اختيار عينة الدّراسة (30) دراسة. وجمعت المعلومات والبيانات الخاصة بتلك الدّراسات، وحللت باستخدام البرنامج (CMA) Comprehensive Meta-Analysis الخاصّ بالتّحليل البعديّ. وقد أشارت النتائج إلى تفوّق طريقة نسبة الأرجحيّة في التحكّم في الخطأ من النوع الأوّل، بوسط حجم أثر يساوي (0.044) وفقًا لنموذج الأثر الثّابت، و(0.030) وفقًا لنموذج الأثر العشوائيّ. وتفوّقت طريقة نسبة الأرجحيّة في قوّة الاختبار الإحصائيّ، بوسط حجم أثـــر يســـاوي (0.068-) وفقًـــا لنمـــوذج الأثـــر الثّابـــت، و(0.111-) وفقًا لنموذج الأثر العشوائيّ. وفيما يتعلّق بنسبة الفقرات ذات الأداء التفاضليّ فقد تفوّقت طريقة مانتل هانزل بوسط حجم أثر (0.116) وفقًا لنموذج الأثر الثابت، و(0.013) وفقًا لنموذج الأثر العشوائيّ.

(الكلمات المفتاحية: التّحليل البعديّ، طريقة نسبة الأرجحيّة، طريقة مانتل هانزل، الأداء التفاضليّ للفقرة)

مقدمة: تركّز جهود الباحثين في مجال بناء الاختبارات على التحقق من دلالات الصّدق والثّبات والفاعليّة للفقرات من حيث مستوى صعوبتها وتمييزها لكي تصبح أدوات يمكن الثّقة بها لوصف السّمات المراد قياسها، إلا أنّ كثيرًا من الباحثين قد لا يهتمّون بخصائص أخرى للاختبارات، ومن بينها التحيّز الذي لا يقلّ أهميّة عن الصّدق والثّبات في بناء الاختبارات وتطويرها (Berk, 1982).

ومؤخرًا زاد الاهتمام بالبحوث المتعلّقة بالأداء التفاضليّ للفقرات في المقاييس النّفسيّة والتّربويّة، والذي تمثّل في أحد جوانبه بوجود طرق إحصائيّة عديدة للكشف عن الأداء التفاضليّ، إلا أنّ هذه الطرق لم تؤدِّ إلى نتائج متسقة وثابتة، وهذا يمكن أن يفسر التّأثير المحدود لدراسات الأداء التفاضليّ على تطوير الاختبارات، ولذلك من المهم التّدقيق بالدّراسات المتعلّقة بالأداء التفاضليّ لما يترتّب عليها من آثار تتعلّق بعدم تحقق العدالة في التّقييمات التربويّة (Huang & Han, 2012).

وقد اقترح علماء القياس طرق وأساليب إحصائيّة عديدة للكشف عن تحيّز الفقرات في جميع أشكال الاختبارات، بهدف تنقية الاختبارات من التحيّز ولتحقيق مبدأ العدالة والمساواة بين المجموعات المختلفة (Camilli & Shepard, 1994). بعضها يعتمد على نظرية القياس الكلاسيكيّة، مثل طريقة الصعوبة المحوّلة وطريقة مانتل- هانزل وطريقة الانحدار اللوجستيّ، ويعتمد بعضها الآخر على نظريّة الاستجابة للفقرة، سواء كانت طرق مَعلميّة أو لامَعلميّة (Park, 2010).

Meta-analysis of Studies Comparing the Likelihood Ratio and the Mantel-Haenszel Methods to Detect Differential Item Functioning

Kuther Abd Al-Qader Bsharat and Yousef Mohammad Al-Sawalmeh, *Yarmouk University, Jordan.*

Abstract: This study aimed at reviewing the studies that compared the Likelihood Ratio (LR) and Mantel-Haenszel (MH) methods for detecting differential item functioning (DIF), which were approved in the period from 1980 to 2021. (30) studies fulfilled the conditions and criteria to be in the study sample. The information and data from those studies were collected and analyzed by using a special program (Comprehensive Meta-Analysis: CMA) for meta-analysis. The results indicated the superiority of the LR method in controlling the type I error rate, with a mean effect size of (0.044) according to the fixed effect model, and (0.03) according to the random effect model. The LR method was superior in the power of the statistical test, with a mean effect size of(-0.068) according to the fixed effect model and (-0.111) according to the random effect model. With regard to the percentage of DIF items, the MH method was superior, with a mean effect size of (0.116) according to the fixed effect model, and (0.013) according to the random effect model.

(**Keywords**: Meta-analysis, Likelihood Ratio Method, Mantel-Haenszel Method, Differential Item Functioning)

وتعدّ طريقة مانتل-هانزل (Mantel-Haenszel) من أكثر الطرق المستخدمة على نطاق واسع للكشف عن الأداء التفاضليّ للفقرة. وتقوم على مقارنة الأداء لمجموعتين من الأفراد على كلّ فقرة حيث يتمّ تقسيم المجتمع إلى مجموعتين فرعيتين، بحيث تسمّى المجموعة الأولى مجموعة الأغلبية أو المجموعة المرجعيّة (Reference group)، وتسمّى المجموعة الثانية مجموعة الأقليّة أو المجموعة المستهدفة (Focal group) (Michaelides, 2008). ولتطبيق هذه الطريقة يتمّ تنظيم البيانات في مصفوفة ثنائيّة البعد (2×2) كما في الجدول (1).

* جامعة اليرموك، الأردن.

© حقوق الطبع محفوظة لجامعة اليرموك، إربد، الأردن، 2024.

الجدول (1)

مصفوفة ثنائية البعد توضح توزيع إجابات أفراد المجموعتين المرجعية والمستهدفة على الفقرة.

الإجابة / المجموعة	الصحيحة	الخاطئة	المجموع
المرجعية (r)	A_J	B_J	n_{rJ}
المستهدفة (f)	C_J	D_J	n_{fJ}
المجموع	m_{1J}	m_{0J}	T_J

وتشير الرموز الواردة في الجدول السابق إلى ما يلي:

C_J , A_J: عدد الأفراد في المجموعتين المرجعيّة والمستهدفة على التّوالي الذين أجابوا عن الفقرة إجابة صحيحة عند مستوى القدرة (J).

D_J , B_J: عدد الأفراد في المجموعتين المرجعيّة والمستهدفة على التوالي الذين أجابوا عن الفقرة إجابة خاطئة عند مستوى القدرة (J).

m_{1J}: عدد أفراد المجموعتين الذين أجابوا عن الفقرة إجابة صحيحة عند مستوى القدرة (J).

m_{0J}: عدد أفراد المجموعتين الذين أجابوا عن الفقرة إجابة غير صحيحة عند مستوى القدرة (J).

n_{rJ}: عدد الأفراد في المجموعة المرجعيّة عند مستوى القدرة (J).

n_{fJ}: عدد الأفراد في المجموعة المستهدفة عند مستوى القدرة (J).

ويتمّ تمثيل نسبة الأرجحيّة بالنسبة لإجراء مانتل-هانزل MH للفقرة i عند مستوى القدرة j بالرمز α وبقيمة محتملة تتراوح من 0 إلى ∞. تشير القيمة 1.0 إلى عدم وجود DIF وبالتالي أداء متشابه لكلتا المجموعتين. وتشير القيم التي تزيد على 1.0 إلى تفوّق المجموعة المرجعيّة في الأداء أو أنها وجدت الفقرة أقلّ صعوبة من المجموعة المستهدفة. من ناحية أخرى، إذا كانت القيمة التي تمّ الحصول عليها أقلّ من 1.0، فهذا مؤشر على أنّ الفقرة كانت أقلّ صعوبة بالنّسبة للمجموعة المستهدفة. وباستخدام البيانات من الجدول (1) تحسب قيمة α كما يلي:

$$\alpha = \frac{A_k D_k}{B_k C_k} \quad \ldots\ldots\ldots\ldots (1)$$

ويمكن تقدير قيمة مجتمعية لقيمة α لتعكس نسبة الأرجحيّة الشائعة عبر جميع مستويات القدرة k. يُشار إلى مقدر نسبة الأرجحيّة الشائعة بالرمز αMH ويمكن حسابها بالمعادلة التّالية:

$$\alpha MH = \frac{\sum_k {}^{A_k D_k}/_{N_k}}{\sum_k {}^{B_k C_k}/_{N_k}} \quad \ldots\ldots\ldots (2)$$

حيث تمثّل N_k الحجم الإجماليّ للعينة عند مستوى القدرةk. ويتمّ معايرة αMH من خلال التحويل اللوغاريتميّ الطبيعيّ الذي يجعل الوسط الحسابيّ حول 0. ويشار للقيمة المعايرة بالرمز MHD-DIF وتحسب على النحو التّالي:

$$\text{MHD-DIF} = -2.35 \ln(\alpha MH) \quad \ldots\ldots (3)$$

وبالتالي فإنّ القيمة 0 تشير إلى عدم وجود DIF. وتشير القيم الأقلّ من 0 إلى ميزة للمجموعة المرجعيّة بينما تشير القيم التي تزيد على 0 إلى ميزة للمجموعة المستهدفة.

ونظرًا لعدم قدرة طرق الكشف عن الأداء التفاضليّ للفقرة القائمة على النظرية التقليديّة على التّمييز بين الفروق الناتجة عن خصائص الفقرة والتي تسبب الأداء التفاضليّ والفروق الحقيقيّة الناتجة عن تأثير الفقرة (Embreston & Reise, 2000)، فقد ظهرت مجموعة من الطرق القائمة على نظريّة الاستجابة للفقرة للكشف عن الأداء التفاضليّ للفقرات. ومن أبرز هذه الطّرق طريقة نسبة الأرجحيّة (Likelihood Ratio).

وقد لاحظ ثيسن وآخرون (Thissen et al., 1988) أنّ اختبار نسبة الأرجحيّة يقارن ملاءمة نموذجين، الأول النّموذج القاعديّ (Compact model) ويحتوي على جميع الفقرات ما عدا الفقرة التي يراد الكشف عن أدائها التفاضليّ وتسمّى مجموعة فقرات الجذع المشترك (Set of items in an anchor)، والنّموذج الثاني ويسمّى النّموذج المعزّز (Augmented model) ويتكوّن من مجموعة فقرات الجذع المشترك مضافًا إليها الفقرة المراد الكشف عن أدائها التفاضليّ، ويتمّ ذلك من خلال إيجاد المؤشر الإحصائيّ لاختبار نسبة الأرجحية (G) المعطى وفقًا للمعادلة التّالية:

$$G = -2logLC - (-2logLA) = -2\log LC + 2logLA \ldots . (4)$$

حيث LC هو الاحتماليّة للنّموذج القاعديّ، وLA هو الاحتماليّة للنّموذج المعزّز.

ويستخدم اختبار نسبة الأرجحيّة لاختبار الفرضيّة الصّفريّة التي تنصّ على عدم اختلاف منحنى خصائص الفقرة للمجموعتين، وذلك من خلال اختبار الفرق بين معالمها، ولاختبار المعالم يقوم اختبار نسبة الأرجحيّة بمقارنة تلك المعالم قبل وبعد إضافة الفقرة المراد الكشف عنها (أي بين النّموذجين القاعديّ والمعزّز)، فإذا تمّ الكشف عن فرق بعد إضافة الفقرة، فإنّ ذلك غالبًا ما يكون بسبب التفاعل بين عضويّة المجموعة والفقرة. وتختبر طريقة نسبة الأرجحيّة بصورة ملاءمة الفقرة باستخدام جميع الفقرات الأخرى كمجموعة جذع مشترك، ويقارن المؤشر الإحصائيّ (G) قبل وبعد إضافة الفقرة المراد الكشف عن أدائها التفاضليّ، وقد أثبتت دراسات المحاكاة قوة القاعدة النظريّة لطريقة نسبة الأرجحيّة في الكشف عن الأداء التفاضليّ المنتظم وغير المنتظم وفق طريقة نظريّة الاستجابة للفقرة المَعلميّة (Thissen et al., 1988).

وحاولت العديد من الدّراسات السّابقة الكشف عن الأداء التفاضليّ لفقرات الاختبار في مختلف التّخصصات، وقد قامت معظم هذه الدّراسات بإجراء مقارنات بين طريقتين أو أكثر. ولقد أجريت العديد من الدّراسات الأجنبيّة والعربيّة التي تناولت طرق الكشف عن الأداء التفاضليّ للفقرات باستخدام طريقتي مانتل-هانزل ونسبة

الأرجحيّة. ومما لا شكّ فيه، فإنّه يوجد تباين كبير في نتائج تلك الدّراسات. فقد قارن يلدريم وجيراي (Yildirim & Giray, 2007) ثلاث طرق (نسبة الأرجحيّة، ومانتل-هانزل، والتّحليل العامليّ التوكيديّ) للكشف عن الأداء التفاضليّ لمتغيّر اللّغة في فقرات اختبار الرّياضيات في الدّراسة الدوليّة بيزا (2003)، وقد توصلا إلى اتفاق الطرق الثّلاث في الكشف عن الأداء التفاضليّ.

و ركّز انكينمان وآخرون (Ankenmann et al., 1999) على تحديد مستويات الخطأ من النّوع الأوّل لطريقتي نسبة الأرجحيّة ومانتل-هانزل عند استخدامهما للكشف عن الأداء التفاضليّ للفقرات، وبينا أنّ القوّة الإحصائيّة للطّريقتين تتأثّر بحجم العينة وأنّ طريقة نسبة الأرجحيّة تفتقر إلى الاتّساق في الكشف عن الأداء التفاضليّ عندما يكون حجم العينة لكلّ من المجموعتين المرجعيّة والمستهدفة أقلّ من (500).

وتوصّل النفيعي (Al-Nafi'i, 2008) إلى أنّ طريقة نسبة الأرجحيّة اكثر فاعليّة في الكشف عن الفقرات ذات الأداء التفاضليّ من طريقة مانتل هانزل، فقد حافظت على معدلات الخطأ من النّوع الأوّل، و أنّ معدلات الخطأ من النّوع الأوّل تختلف باختلاف حجم العينة في الطريقتين. وبحثت أماندا (Amanda, 2008) في الخطأ من النوع الأوّل وقوّة الاختبار في طرق الكشف عن الأداء التفاضليّ للفقرات، وبيّنت أنّ هناك ارتفاعًا في قيم الخطأ من النّوع الأوّل لطريقة نسبة الأرجحيّة، بينما كانت قيمها في المدى المقبول لطريقة مانتل هانزل، كما بينت أنّ قوّة الاختبار كانت أعلى في طريقة نسبة الأرجحيّة وبالأخصّ في الأداء التفاضليّ غير المنتظم. وقد أكّد اللبدي (Al-Labadi, 2008) النّتيجة نفسها عند تقييم فاعليّة ومقارنة طريقة مانتل هانزل وطريقة نسبة الأرجحيّة.

وأكّد كيكلك (Keklik, 2012) على أنّ طريقة نسبة الأرجحيّة أفضل من طريقة مانتل هانزل في التّحكّم بمعدل الخطأ من النّوع الأوّل في التّوزيعات المختلفة للسّمة، لكن في التّوزيعات الطّبيعيّة تكون طريقة مانتل هانزل أكثر ضبطًا لمعدل الخطأ من النّوع الأوّل. وأشار كوندراتك وجرودنيوسك (Kondratek & Grudniewska, 2013) إلى أنّ طريقة مانتل هانزل أكثر قوّة في الكشف عن الأداء التفاضليّ المنتظم للفقرة من طريقة نسبة الأرجحيّة، على الرّغم من أنّ طريقة نسبة الأرجحيّة كانت أقوى في الكشف عن الأداء التفاضليّ غير المنتظم من طريقة مانتل هانزل. وبين الغامدي (Alghamdi, 2017) أن معدلات الخطأ من النّوع الأوّل تكون منخفضة لكلتا الطريقتين عندما تكون تقديرات القدرة متماثلة عبر المجموعات، وعند وجود اختلافات في القدرة بين المجموعات يكون معدل الخطأ من النّوع الأوّل لطريقة مانتل هانزل أقلّ منه لطريقة نسبة الأرجحيّة.

وفيما يتعلّق بنسبة الفقرات ذات الأداء التفاضليّ، فقد أشار كرير (Creer, 2004) إلى أنّ طريقة نسبة الأرجحيّة تكشف عن فقرات ذات أداء تفاضليّ في اختبار (SATV) بنسب أعلى من ما تكشفه طريقة مانتل- هانزل. وكشف مبارك (Mubarak, 2010) عن درجة مقبولة من التّوافق بين طريقتي مانتل- هانزل ونسبة الأرجحيّة في الكشف عن الأداء التفاضليّ في فقرات اختبار العلوم في الدّراسة الدّوليّة بيزا (2006). وأكّد أبو مسلم (Abu Muslim, 2010) على التّوافق بين الطّريقتين في الكشف عن تحيّز فقرات اختبار القبول في برنامج الدّبلوم العام في التّربية بمعهد الدّراسات والبحوث التّربويّة.

وهكذا فإنّ هناك نتائج متباينة، فبعضها أكّد على فعاليّة طريقة مانتل- هانزل، وبعضها الآخر أكّد على فعاليّة طريقة نسبة الأرجحيّة، بينما أشارت نتائج بعضها إلى وجود اتّفاق بين الطريقتين في الكشف عن الأداء التفاضليّ، وبالتّالي فإنّ محاولة التّوصّل إلى أيّ الطرق أفضل في الكشف عن الأداء التفاضليّ للفقرات من خلال التّحليل البعدي للدّراسات السابقة تمثّل فكرة جديرة بالاهتمام تسعى الدّراسة الحاليّة إلى البحث فيها.

وتشهد ميادين البحث العلميّ زيادة مطّردة في عدد الأبحاث المتباينة في نتائجها حول المشكلة نفسها، مما يوجد حالة من الشّكّ حول تلك النتائج ويزيد من صعوبة اتّخاذ القرارات ذات العلاقة، الأمر الذي يتطلّب إيجاد طريقة يمكن من خلالها الحصول على صورة مركبة شاملة لمجموعة النتائج المتناثرة في الدّراسات المختلفة (Pigott, 2012). لذلك ظهرت العديد من الطّرق للتّخلص من حالة الشّكّ تلك، وذلك بالعمل على جمع وترميز المعلومات من الدّراسات السّابقة (Petitti, 1994)، ويعدّ أسلوب التّحليل البعديّ Meta –analysis من أكثر تلك الأساليب فعاليّة وموضوعيّة لدمج وتلخيص نتائج البحوث السّابقة (Gay, 2000).

وقد أشار كوليك وكوليك المذكور في العتوم 2005 (Al-Kulik & Kulik Autoom, 2005) إلى أربع خطوات أساسيّة يتمّ من خلالها إجراء التّحليل البعديّ وتتمثل هذه الخطوات بمراجعة الدّراسات السّابقة ذات العلاقة بموضوع البحث وفق محكّات محدّدة، وتحديد مخرجات كلّ دراسة بطريقة كميّة، وترميز البيانات الخاصّة بتلك الدّراسات، وتلخيصها وتحليلها إحصائيًا وتفسيرها. وتُصنّف النّماذج الإحصائيّة للتّحليل البعديّ في فئتين أساسيتين هما (1) نماذج الأثر الثّابت التي تفترض أنّ هناك قيمة واحدة مشتركة لحجم الأثر لجميع الدّراسات التي اختيرت عيناتها من المجتمع نفسه، وأنّ اختلاف حجم الأثر في أيّ دراسة يعود إلى كيفيّة اختيار عينتها، و(2) نماذج الأثر العشوائيّ التي تفترض بأنّ عينة كلّ دراسة قد تمّ سحبها من مجتمع مختلف، ولذلك فإنّه يمكن التعامل مع العينات المختلفة على أنّها مجموعات جزئيّة من مجتمعات مختلفة ومنفصلة لكلّ منها وسطه الحسابيّ الخاصّ، لذلك فإنّ مصادر تباين حجم الأثر في هذه الحالة هما: تباين داخل الدّراسة الواحدة وتباين بين الدّراسات (Gioacchino, 2005). وعلى الرّغم من توافر الكثير من الدّراسات التي تتحقّق من فاعليّة طرق الكشف عن الأداء التّفاضليّ للفقرات فإنّ تباين نتائجها يتطلّب تركيب تلك النتائج وتوليفها معًا من أجل الوصول إلى حقائق

ومفاهيم وتعميمات تساهم في بناء العلم، وإجراء تقييم أكثر شمولًا عن طريق الجمع بين نتائج هذه الدّراسات (Omar, 2013). لذلك جاءت هذه الدّراسة لتلخيص فعاليّة طريقتي نسبة الأرجحيّة ومانتل-هانزل في الكشف عن الأداء التفاضليّ للفقرة في الدّراسات السّابقة باستخدام أسلوب التّحليل البعديّ، من أجل مساعدة الباحثين للوصول إلى المعلومات الكميّة حول الطّريقة المناسبة.

مشكلة الدّراسة

في كثير من الدّراسات قد تكون النتائج دالة إحصائيًا، ولكنّها في الوقت نفسه ليست ذات أهميّة، بمعنى أنّها لا تقدّم معلومات مهمة ومفيدة وتفتقر للدّلالة العمليّة. ويعطي بعض الباحثين الدّراسات الدّالة عند مستوى الدّلالة ($\alpha = 0.01$) أهميّة أكبر من الدّراسات الدّالة عند مستوى الدّلالة ($\alpha = 0.05$) دون اعتبار لحجم العينة ودلالة الفرق أو العلاقة في المجال العلميّ المبحوث. فقد يرفض الباحث فرضية صفريّة ليس بسبب أنّ هناك أثرًا للمعالجة، بل لزيادة حجم العينة. وهذا يحتمّ الاهتمام بالدّلالة العمليّة وقوّة الاختبار الإحصائيّ المستخدم لاتّخاذ قرار حيال الظّاهرة المدروسة (Al-Sharifeen, 2017).

وتنبع مشكلة الدّراسة الحاليّة من خلال وجود تباين في نتائج الدّراسات السّابقة التي اهتمت بالمقارنة بين طريقتي مانتل-هانزل ونسبة الأرجحيّة في الكشف عن الأداء التفاضليّ للفقرات. لذلك قد يساعد استخدام أسلوب التّحليل البعديّ في تحديد أيّهما أنسب وفقًا لمعايير محدّدة تتعلّق بالخطأ من النّوع الأوّل وقوّة الاختبار وحساسيته. كما أن قلّة استخدام أسلوب التّحليل البعديّ في دراسات الأداء التّفاضليّ في البيئة العربيّة كان محفزًا لإجراء هذه الدّراسة من أجل توفير ملخص كميّ شامل حول فاعليّة الطريقتين. ويتلخّص الغرض الرّئيس من هذه الدّراسة بمقارنة فاعليّة طريقة نسبة الأرجحيّة وطريقة مانتل-هانزل للكشف عن الأداء التفاضليّ DIF باستخدام التّحليل البعديّ. وتتحدّد مشكلة الدّراسة بالأسئلة التّالية:

1- ما دلالة حجم الأثر الكليّ للدّراسات التي اهتمت بمقارنة الخطأ من النّوع الأول بين طريقتي نسبة الأرجحيّة ومانتل هانزل في الكشف عن الأداء التفاضليّ للفقرة؟

2- ما دلالة حجم الأثر الكلّيّ للدّراسات التي اهتمت بمقارنة قوة الاختبار الإحصائيّ بين طريقتي نسبة الأرجحيّة ومانتل هانزل في الكشف عن الأداء التفاضليّ للفقرة؟

3- ما دلالة حجم الأثر الكلّيّ للدّراسات التي اهتمت بمقارنة نسبة الفقرات ذات الأداء التفاضليّ بين طريقتي نسبة الأرجحيّة ومانتل هانزل في الكشف عن الأداء التفاضليّ للفقرة؟

أهميّة الدّراسة

تنبع أهميّة هذه الدّراسة في استخدامها أسلوب التّحليل البعديّ في مقارنة فاعليّة طريقتي نسبة الأرجحيّة ومانتل- هانزل في الكشف عن الأداء التفاضليّ للفقرات. وهي بمثابة مرجع يساعد الباحثين في توظيف التّحليل البعديّ كأسلوب لدمج وتوليف نتائج مجموعة من الدّراسات المتعلّقة بمشكلة بحثيّة مشتركة، خاصّة في ضوء قلّة الدراسات العربيّة في هذا المجال. كما تساعد في إعطاء صورة أوضح وأشمل وأدقّ للمقارنة بين الطريقتين من حيث القدرة على ضبط معدل الخطأ من النّوع الأوّل، وقوّة الاختبار الإحصائيّ، وحساسيته للفقرات ذات الأداء التفاضليّ لاختيار الأنسب منهما. وقد تفتح الباب أمام فرص بحثيّة أخرى في هذا المجال أو مجالات أخرى.

التعريفات الاصطلاحيّة

التّحليل البعديّ: هو أسلوب إحصائيّ لتجميع وتلخيص نتائج مجموعة من الدّراسات التي بحثت المشكلة ذاتها، من خلال الوصول إلى قيمة تمثّل معدل النّتائج التي ظهرت في الدّراسات السابقة.

الأداء التفاضليّ للفقرة: هو دالّة مشتقة إحصائيًا للتعبير عن الفروق في الاستجابة على الفقرة بين مجموعتين من المفحوصين ممن هم في نفس المستوى من القدرة.

حجم الأثر: هو مقياس كميّ لحجم الأثر التجريبيّ، ويعني حجم الأثر الكبير أن لنتائج البحث دلالة (أهميّة) عمليّة، ويتميّز بعدم تأثره بحجم العينة. ومقاييس الأثر إمّا أنّ تكون مقاييس نسبة (risk ratio, odds ratio) أو مقاييس فرق (mean difference, risk difference).

نسبة الفقرات ذات الأداء التّفاضليّ: هي عبارة عن عدد الفقرات ذات الأداء التّفاضليّ مقسومة على العدد الكلّيّ للفقرات في 100%.

طريقة نسبة الأرجحيّة: إحدى الطرق الإحصائيّة القائمة على نموذج الاستجابة للفقرة التي تستخدم للكشف عن الأداء التّفاضليّ للفقرة، تقوم على مقارنة نموذجين الأوّل يسمى القاعديّ، والثّاني يسمى المعزّز، من خلال إيجاد المؤشر الإحصائيّ لاختبار نسبة الأرجحيّة (G).

طريقة مانتل هانزل: طريقة للكشف عن الأداء التّفاضليّ للفقرة من خلال مقارنة نسبة الإجابات الصّحيحة مقابل الإجابات غير الصّحيحة على فقرة اختبار معين بحسب انتماء المفحوص للمجموعة المستهدفة أو المرجعيّة.

معدل الخطأ من النّوع الأوّل: هو النّسبة المئويّة لعدد مرات الكشف عن الأداء التّفاضليّ للفقرة بشكل خاطئ ويقارن بالقيمة الاسميّة (α=0.05).

الطّريقة والإجراءات

مجتمع الدّراسة وعينتها

تكون مجتمع الدّراسة من جميع الدّراسات ورسائل الماجستير والأطروحات الجامعيّة التي بحثت في موضوع مقارنة فاعليّة طريقة نسبة الأرجحيّة وطريقة مانتل-هانزل للكشف عن الأداء التّفاضليّ للفقرة DIF، للفترة الزمنيّة الممتدة من 1980 إلى 2021م. وتكوّنت عينة الدّراسة من (30) دراسة أوفت بمحكّات ومعايير الاختيار للعينة. ويبين الجدول (2) الدّراسات وتوزيعها حسب معايير المقارنة بين الطريقتين.

الجدول (2)

توزيع الدراسات وفقاً لمعايير المقارنة بين طريقتي نسبة الأرجحية ومانتل هانزل

معيار المقارنة	عدد الدراسات	النسبة المئوية
معدل الخطأ من النوع الأول	14	35%
قوة الاختبار الإحصائي	8	20%
نسبة الفقرات ذات الأداء التفاضلي	18	45%
المجموع	*40	100%

* بعض الدراسات تتضمن اكثر من معيار.

أداة الدّراسة وصدقها وثباتها

تمّ تطوير أداة خاصة لتجميع المعلومات والبيانات الخاصة بالدّراسات المشمولة في العينة، بعد الاطّلاع على عدد من الدّراسات السّابقة (Aljughaiman & Ayoub, 2013; Gellin, 2003; Kim, 2016) وملاحظة الكيفيّة التي كتبت فيها نتائجها والبيانات التي تمّ استخدامها لحساب قيم حجم الأثر، وتضمّنت الأداة بصورتها الأوليّة معلومات عامّة تتعلّق باسم الباحث، وسنة النشر، ونوع العينة، ونوع الأداء التّفاضليّ للفقرة DIF ومعلومات خاصّة تتعلّق بالخطأ من النّوع الأوّل، وقوّة الاختبار الإحصائيّ، ونسبة الفقرات ذات الأداء التّفاضليّ. وللتأكّد من الصّدق الظّاهريّ لأداة الدّراسة تمّ عرض الصورة الأوليّة على مجموعة من المحكمين وعددهم (5)، ممن يحملون درجة الدكتوراة في القياس والتّقويم، حيث طلب من كلّ مُحكّم إبداء رأيه في محتوى الأداة، واقتراح أيّ تعديلات تتعلّق بإدراج المعلومات أو حذفها، وقد تمّ الأخذ بالملاحظات التي تمّ تقديمها، وبناءً على ذلك تمّ إعداد الصّورة النهائيّة للأداة. وقد قام الباحثان بصورة مستقلة بمراجعة الدّراسات المشمولة بالعينة وتفريغ البيانات، وقد كان هناك اتّفاق شبه تامّ بينهما. وهذا مؤشر على وجود ثبات مرتفع بين الباحثين في تجميع المعلومات وتفريغها. ويبين الجدول (3) المعلومات الأساسيّة المتعلّقة بالدّراسات المشمولة في عينة الدّراسة.

الجدول (3)

المعلومات الأساسية المتعلقة بالدراسات المشمولة في عينة الدراسة.

المؤلف	السنة	طبيعة البيانات	حجم العينة	طول الاختبار	نوع العينة	نوع الأداء التفاضلي	توزيع القدرة
Ankenmann, Witt and Dunbar	1999	محاكاة	صغير، متوسط، كبير	متوسط	ذكور وإناث	منتظم	متماثل
Kim	2000	محاكاة	متوسط	متوسط	ذكور وإناث	منتظم	متماثل
Creer	2004	محاكاة	كبير	متوسط	ذكور وإناث	منتظم وغير منتظم	متماثل
Pae	2004	محاكاة	كبير	قصير	ذكور وإناث	منتظم وغير منتظم	متماثل
Finch	2005	محاكاة	صغير	متوسط	ذكور وإناث	منتظم وغير منتظم	متماثل
Yildirim and Giray	2007	محاكاة	متوسط	طويل	ذكور وإناث	منتظم وغير منتظم	متماثل
Al-labadi	2008	محاكاة	صغير، متوسط، كبير	متوسط	ذكور واناث	منتظم وغير منتظم	متماثل وغير متماثل
Al-Nafi'i	2008	محاكاة	صغير، متوسط، كبير	متوسط	ذكور وإناث	منتظم وغير منتظم	متماثل، وغير متماثل
Amanda	2008	حقيقية	صغير، متوسط، كبير	طويل	ذكور وإناث	منتظم وغير منتظم	متماثل
Yildirim & Berberoğlu	2009	حقيقية	متوسط	متوسط	ذكور وإناث	منتظم وغير منتظم	متماثل
Mubarak	2010	حقيقية	كبير	طويل	ذكور وإناث	منتظم وغير منتظم	متماثل
Abu Muslim	2010	محاكاة	متوسط	متوسط	ذكور وإناث	منتظم وغير منتظم	متماثل
Pei & Li	2010	محاكاة	صغير	متوسط	ذكور وإناث	منتظم وغير منتظم	متماثل وغير متماثل
Paek and Wilson	2011	محاكاة	صغير، متوسط، كبير	متوسط	ذكور وإناث	منتظم وغير منتظم	متماثل
Çepni	2011	محاكاة	صغير	متوسط	ذكور وإناث	منتظم وغير منتظم	متماثل

المؤلف	السنة	طبيعة البيانات	حجم العينة	طول الاختبار	نوع العينة	نوع الأداء التفاضلي	توزيع القدرة
Atalay, Gök, Kelecioğlu and Arsan	2012	محاكاة	صغير، متوسط	متوسط	ذكور وإناث	منظم وغير منتظم	غير متماثل
Keklik	2012	محاكاة	متوسط	متوسط	ذكور وإناث	منتظم	متماثل وغير متماثل
Kondratek and Grudniewska	2013	محاكاة	متوسط	قصير	ذكور وإناث	منتظم وغير منتظم	متماثل
Kabasakal et al.	2014	محاكاة	صغير، متوسط، كبير	قصير، متوسط، طويل	ذكور وإناث	منتظم وغير منتظم	متماثل
Price	2014	محاكاة	صغير، متوسط	متوسط	ذكور وإناث	منتظم وغير منتظم	متماثل
Al-Dawy	2015	محاكاة	متوسط	متوسط	ذكور واناث	منتظم وغير منتظم	متماثل
Demirtasli & Ulutas	2015	حقيقية	متوسط	متوسط	ذكور واناث	منتظم وغير منتظم	غير متماثل
Al-Bashabsha	2016	محاكاة	كبير	طويل	ذكور وإناث	منتظم وغير منتظم	متماثل
Alghamdi	2017	محاكاة	صغير، متوسط، كبير	متوسط	ذكور وإناث	منتظم	متماثل وغير متماثل
Ardic & Gelbal	2017	حقيقية	متوسط	متوسط	ذكور وإناث	منتظم وغير منتظم	متماثل
Selvi & Alıcı	2018	حقيقية	متوسط	طويل	ذكور وإناث	منتظم وغير منتظم	متماثل
Gao	2019	محاكاة	صغير، متوسط، كبير	قصير، متوسط	ذكور وإناث	منتظم وغير منتظم	متماثل
Sapmaz	2019	حقيقية	كبير	متوسط	ذكور وإناث	منتظم وغير منتظم	متماثل
Uyar	2020	حقيقية	متوسط	متوسط	ذكور وإناث	منتظم وغير منتظم	متماثل
Cepni & Kelecioglu	2021	حقيقية	كبير	متوسط	ذكور وإناث	منتظم وغير منتظم	متماثل

إجراءات الدّراسة

أولًا: البحث في المصادر المختلفة (مكتبة الجامعة وقواعد البيانات المختلفة) عن الدّراسات والرّسائل الجامعيّة التي قارنت بين طريقتي نسبة الأرجحيّة ومانتل هانزل في الكشف عن الأداء التّفاضليّ للفقرة التي تمّ العمل عليها ونشرها خلال الفترة الزمنيّة من 1980 إلى 2021.

ثانيًا: حصر الدّراسات التي يتوفّر فيها معلومات كافية تتعلّق بحساب حجم الأثر فيما يتعلّق بالخطأ من النّوع الأوّل، أو قوّة الاختبار الإحصائيّ، أو نسبة الفقرات ذات الأداء التفاضليّ.

ثالثًا: رصد كافة المعلومات ذات العلاقة وترميزها بحسب ما تطلبه أداة الدّراسة والتّحقق من موثوقيّة تلك المعلومات.

رابعًا: استخدام برنامج Comprehensive Meta- Analysis في التّحليل البعديّ لتلك الدّراسات بحسب كلّ من نموذج الأثر الثّابت Fixed Effect Model ونموذج الأثر العشوائيّ Random Effect Model. وقد تمّ استخدام المؤشرات الإحصائيّة التّالية:

1- مؤشر حجم الأثر بدلالة فرق المخاطر (Risk difference) لكلّ من الخطأ من النّوع الأوّل وقوّة الاختبار الإحصائيّ ونسبة الأداء التّفاضليّ للفقرة لطريقتي نسبة الأرجحيّة ومانتل- هانزل.

2- فترات الثّقة (CI) 95% لقيم حجم الأثر لكلّ دراسة.

3- اختبار Q، وهو الاختبار التّقليديّ لعدم التّجانس في التّحليل البعديّ. بالاعتماد على توزيع مربع كاي، فإنّه يعطي احتمالات تشير القيم المرتفعة منها إلى تباينات كبيرة عبر الدّراسات المختلفة.

4- مؤشرات تحليل التّحيّز I^2 & Funnel Plot.

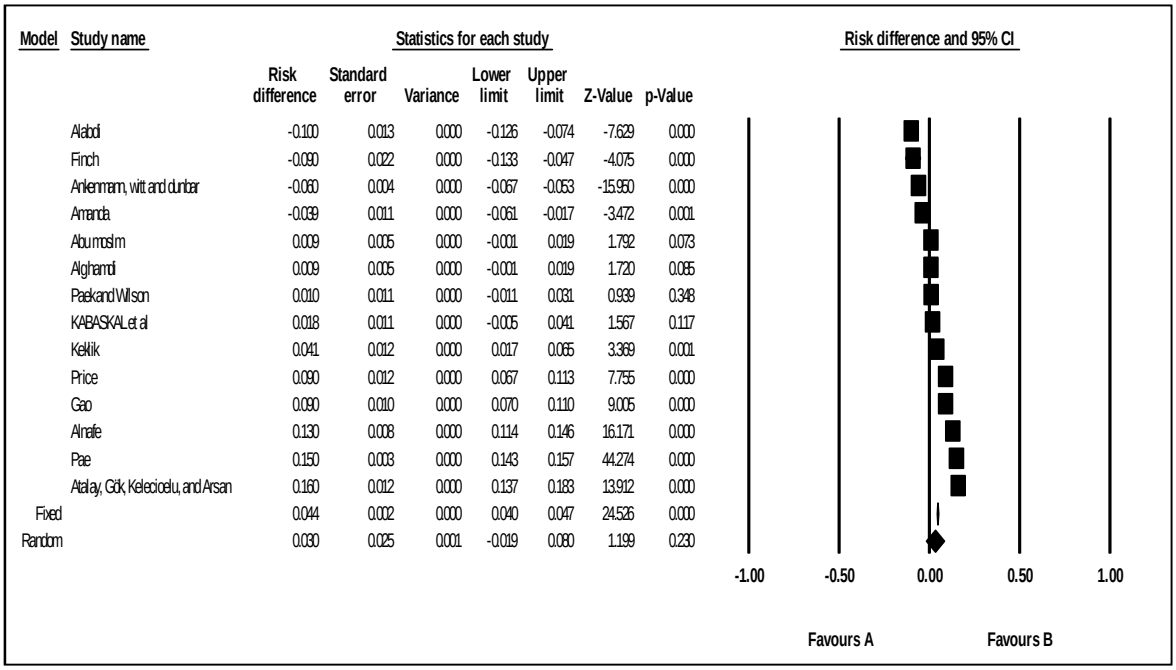
النّتائج ومناقشتها

أولاً: النتائج المتعلقة بالسؤال الأوّل، والذي نص على: *"ما دلالة حجم الأثر الكلّيّ للدّراسات التي اهتمّت بمقارنة الخطأ من النّوع الأول بين طريقتي نسبة الأرجحيّة ومانتل- هانزل في الكشف عن الأداء التفاضليّ للفقرة؟"*

للإجابة عن هذا السّؤال، تمّ إجراء التّحليل البعديّ للتعرّف على حجم الأثر باستخدام مؤشر (Risk Difference) للدّراسات التي اهتمّت بمقارنة الخطأ من النّوع الأوّل بين طريقتي نسبة الأرجحيّة ومانتل- هانزل في الكشف عن الأداء التّفاضليّ للفقرة. ويظهر الشكل (1) نتائج التّحليل البعديّ لدراسات معدل الخطأ من النّوع الأوّل بين طريقتي نسبة الأرجحيّة ومانتل- هانزل في الكشف عن الأداء التّفاضليّ. كما يبيّن الجدول (4) وسط حجوم الأثر وفترة الثّقة وفقًا لنموذجي الأثر الثّابت والعشوائيّ، ومؤشرات التّجانس لمعدل الخطأ من النّوع الأوّل.

الشكل (1)

نتائج التحليل البعدي لدراسات معدل الخطأ من النوع الأول بين طريقتي نسبة الأرجحية ومانتل هانزل في الكشف عن الأداء التفاضلي.



الجدول (4)

وسط حجوم الأثر وفقاً لنموذجي الأثر الثابت والعشوائي ومؤشرات التجانس لمعدل الخطأ من النوع الأول

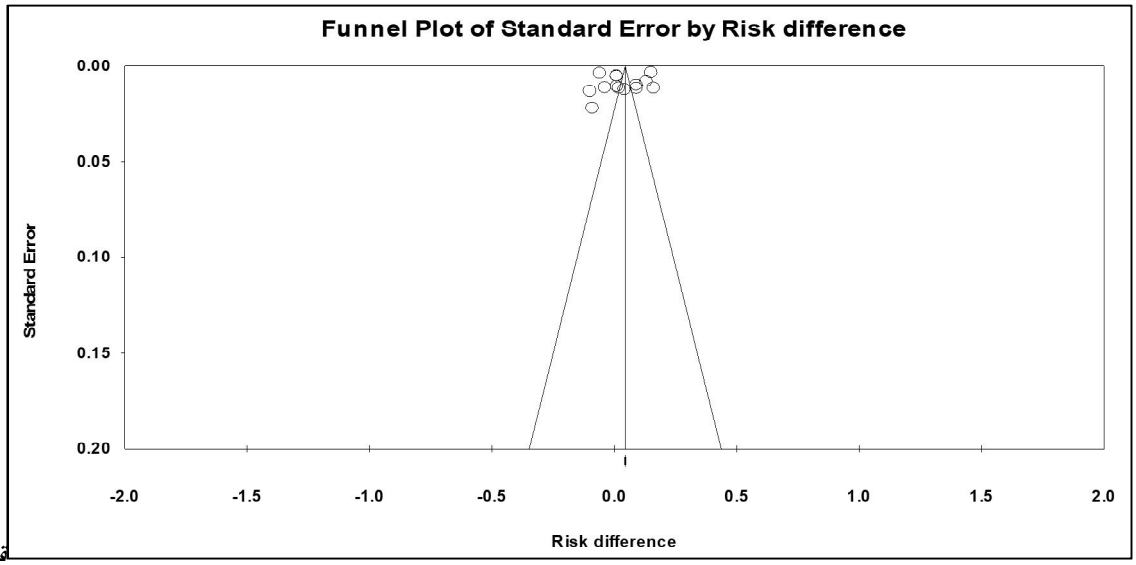
النموذج	عدد حجوم الأثر	وسط حجوم الأثر	الخطأ المعياري	اختبار (Q)	درجات الحرية	P- value	اختبار (I^2)
الأثر الثابت	14	0.044	0.002	2317.0	13	0.00	99.4
الأثر العشوائي	14	0.030	0.025				

ويظهر من الشكل (1) أنّ قيم حجم الأثر للدّراسات المختلفة تتراوح بين (0.10-) و(0.16)، وقد كانت موجبة في (10) دراسات تتفوّق فيها طريقة نسبة الأرجحيّة (مشار لها في الشكل بالحرف B) على طريقة مانتل هانزل (مشار لها في الشكل بالحرف A) في التحكّم في معدل الخطأ من النّوع الأوّل، وقد كانت سالبة في أربع دراسات تتفوّق فيها طريقة مانتل هانزل على طريقة نسبة الأرجحيّة في التحكّم بمعدل الخطأ من النّوع الأوّل. ويتّضح من الجدول (4) أنّ وسط حجم الأثر الكلّيّ للدّراسات التي بحثت في معدل الخطأ من النّوع الأوّل يساوي (0.044) وفق نموذج الأثر الثّابت، و(0.030) وفق نموذج الأثر العشوائيّ. ويشير ذلك إلى تفوّق طريقة نسبة الأرجحيّة على طريقة مانتل هانزل في التّحكّم في الخطأ من النّوع الأوّل. وفيما يتعلّق بعدم التّجانس في الدّراسات التي تناولت الخطأ من النّوع الأوّل، يبيّن اختبار Q أن الدّراسات كانت غير متجانسة في نتائجها حيث بلغت قيمة الاختبار (2317) وهي دالة إحصائيًا عند مستوى الدّلالة $\alpha=0.05$، كما يبيّن مؤشر I^2 أنّ نسبة عدم التّجانس تساوي 99.4%، وهذا يعني أنّ حجوم الأثر متباينة بشكل واضح لدراسات معدل الخطأ من النّوع الأوّل، وبناءً عليه فإنّ ثمة ضرورة لإجراء تحليل خاصّ يتناول ظروف تلك الدّراسات وما تضمّنته من متغيّرات.

ولتحليل التّحيّز في الدراسات تمّ استخدام طريقة فحص شكل انتشار القمع Funnel Plot، كما يظهره الشكل (2)، والذي يمثّل العلاقة بين قيم حجم الأثر والأخطاء المعياريّة لها.

الشكل (2)

حجم الأثر والأخطاء المعيارية للدراسات التي تناولت معدل الخطأ من النوع الأول بين طريقتي نسبة الأرجحية ومانتل هانزل في الكشف عن الأداء التفاضلي.



قيم

يبيّن الشكل (2) أنّ معظم الدّراسات قد تركّزت في الجزء الأعلى(حيث تكون قيم الخطأ المعياريّ قريبة من الصّفر) خارج القمع بشكل متماثل تقريبًا حول الخط العموديّ الذي يمثّل وسط قيم حجم الأثر، ويشير ذلك إلى عدم وجود التّحيّز. ولكون هذا التفحّص البصريّ للشّكل ذا صبغة ذاتية، فقد تمّ استخدام اختبار ايجر Egger`s regression test of the intercept لتحليل التّحيّز كما هو موضح في الجدول (5).

الجدول (5)

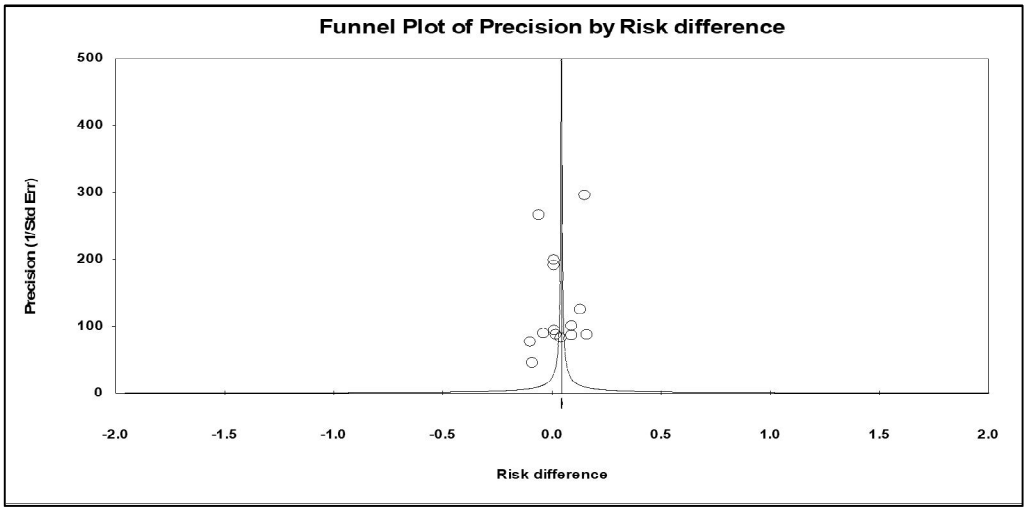
اختبار ايجر لتحليل التحيز Egger`s regression test of the intercept.

الإحصائي	القيمة
نقطة التقاطع	-3.0792
الخطأ المعياري	7.485
اختبار t	0.4115
درجة الحرية	12
قيمة p (1-tailed)	0.3439
قيمة p (2-tailed)	0.6879

يبين الجدول (5) أن قيمة اختبار ايجر بلغت (t = 0.4115, p=0.34)، وتشير هذه النتيجة إلى عدم وجود التّحيّز في الدّراسات المستخدمة في التّحليل. وقد تمّ تمثيل قيم حجم الأثر مع مؤشر دقّة تقديرها (معكوس قيمة الخطأ المعياريّ) باستخدام شكل انتشار القمع كما هو موضح في الشكل (3).

الشكل (3)

قيم حجم الأثر وما يقابلها من مؤشر دقة تقديرها.



ويلاحظ من الشكل (3) أنّ معظم قيم حجم الأثر للدّراسات قد توزّعت على جانبي الخطّ العموديّ الذي يمثّل وسط حجم الأثر في المجتمع، وهذا يشير إلى عدم وجود التّحيّز بصورة مؤثّرة وقد كانت معظم القيم لمؤشّرات الدّقّة تساوي أو تزيد على 100. وللاستدلال على أهميّة التّحيّز الموجود ومدى تأثيره في نتائج التّحليل البعديّ، فقد تمّ استخدام طريقة التقليم والتّعبئة Trim and Fill (Borenstein et al., 2009) لقياس أثر التّحيّز المحتمل على نتائج التّحليل البعديّ، وقد تبين عدم حدوث أيّ تغيير بعد إجراء عمليّة التّقليم (استبعاد قيم حجم الأثر الكبرى المتطرفة من الجانب الملتوي للشكل وهو غالبًا الجانب الأيمن) والتّعبئة (إدخال قيم حجم الأثر الصّغيرة للدّراسات المفقودة على الجانب الإيسر في العادة)، وهذا مؤشّر على عدم وجود أثر للتّحيّز في النّتائج (Steenbergen & Olszewski, 2016).

وتتّفق هذه النّتيجة مع ما توصّلت إليه كلّ من دراسة النفيعي (Al-Nafi'i, 2008) ودراسة كيكلك (Keklik, 2012) التي بيّنت نتائجهما أنّ طريقة نسبة الأرجحيّة أفضل من طريقة مانتل-هانزل في التحكّم بمعدل الخطأ من النّوع الأوّل وبشكل خاصّ في التّوزيعات غير الطّبيعيّة للسّمة. وقد تعزى هذه النّتيجة إلى اعتماد طريقة مانتل- هانزل على العلامة الخام (عدد الإجابات الصّحيحة) كإحصائيّ كافي للقدرة، ممّا يضعف قدرتها على ضبط الخطأ من النوع الأوّل عند اختلاف توزيعات القدرة للمجموعات. ويفيد الاعتماد على العلامات الخام فقط لإظهار التكافؤ في الدّرجة الكليّة ولا يفيد في إظهار التكافؤ البنائيّ أو المتريّ، لذلك لا يوفّر معلومات عن العلاقة بين القدرة للفرد والدّرجة الكليّة له في كلّ من المجموعة المرجعيّة والمجموعة المستهدفة. وهذا يتّفق مع نتائج دراسة جوليرا وآخرين (Guilera et al., 2013) التي أكّدت على ضرورة أخذ التباين في توزيع القدرة بالاعتبار عند استخدام طريقة مانتل- هانزل في الكشف عن الأداء التّفاضليّ للفقرات. وتتعارض نتائج الدّراسة مع نتائج دراسات كل من اللبدي (Al-Labadi, 2008) وامانداAmanda, 2008)) وبرايس (Price, 2014) وفينش (Finch, 2005) والغامدي (Alghamdi, 2017) التي أظهرت أنّ معدل الخطأ من النّوع الأوّل كان مُرضيًا وقريبًا من القيمة الأسميّة 0.05 لطريقة مانتل هانزل بينما كان متضخمًا لطريقة نسبة الأرجحيّة. وبينت الدّراسة الحاليّة أنّ هناك عدم تجانس مرتفع جدًا في الدّراسات التي اهتمّت بمعدل الخطأ من النّوع الأوّل ممّا يشير إلى التباين الكبير في المتغيّرات الوسيطة عبر تلك الدّراسات حيث من الممكن أن يتفاوت تأثير تلك المتغيّرات عبر الطريقتين.

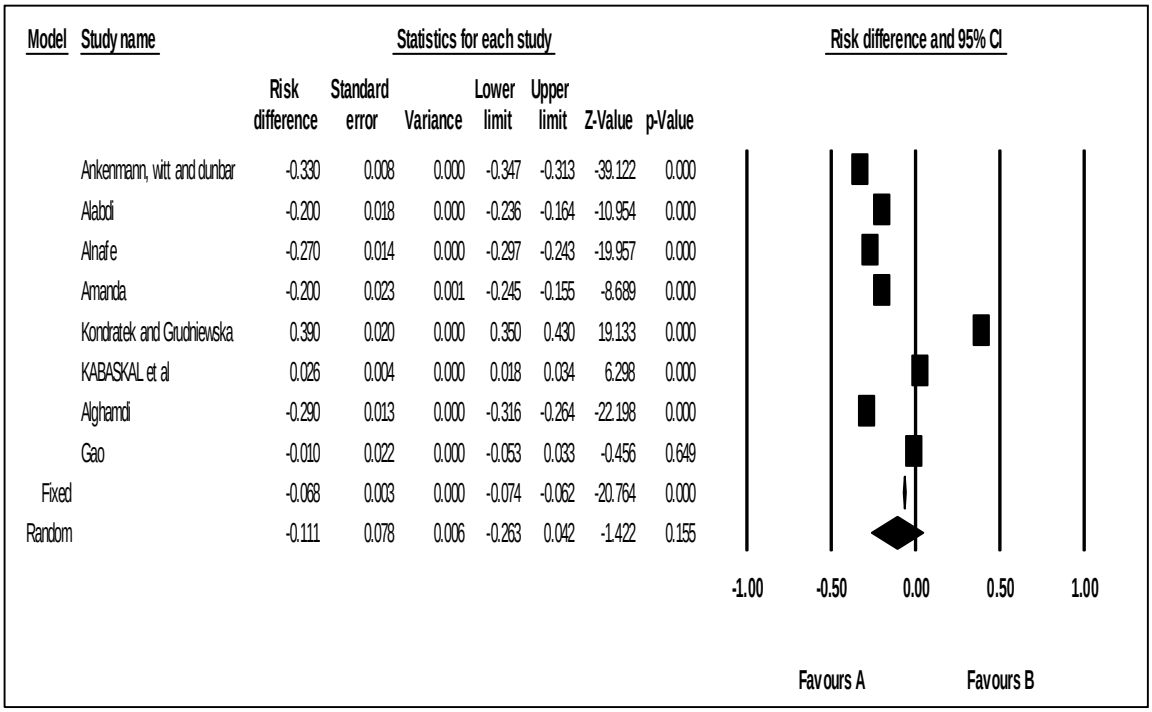
وتجدر الإشارة إلى أنّ هذه النّتيجة ذات طبيعة احتماليّة وليست يقينيّة كونها تعتمد على 14 دراسة فقط. وإذا تمّ النّظر إلى هذه الدّراسات باعتبارها عينة عشوائيّة من مجتمع الدّراسات التي اهتمّت بالمقارنة بين الطّريقتين في ضوء درجة ضبطهما للخطأ من النّوع الأوّل فإنّ الوسط الحسابيّ لحجم الأثر لهذه الدّراسات وفق نموذج الأثر العشوائيّ يساوي 0.03 وهو غير دال إحصائيًا عند مستوى الدلالة الإحصائيّة 0.05، (Z = 1.199, sig=0.23). وهذا يشير إلى أنّ الطّريقتين متكافئتين من النّاحية العمليّة في ضبطهما للخطأ من النّوع الأوّل.

ثانيًا: النتائج المتعلقة بالسؤال الثّاني، والذي نص على: *"ما دلالة حجم الأثر الكلّيّ للدّراسات التي اهتمّت بمقارنة قوّة الاختبار الإحصائيّ بين طريقتي نسبة الأرجحيّة ومانتل هانزل في الكشف عن الأداء التّفاضليّ للفقرة؟"*

يبيّن الشكل (4) نتائج التّحليل البعديّ لدراسات قوّة الاختبار الإحصائيّ بين طريقتي نسبة الأرجحيّة ومانتل- هانزل في الكشف عن الأداء التّفاضليّ. كما يبيّن جدول (6) وسط حجوم الأثر وفقًا لنموذجي الأثر الثّابت والعشوائيّ ومؤشّرات التّجانس لقوّة الاختبار الإحصائيّ.

الشكل (4)

نتائج التحليل البعدي لدراسات قوة الاختبار الإحصائي بين طريقتي نسبة الأرجحية ومانتل هانزل في الكشف عن الأداء التفاضلي



الجدول (6)

وسط حجوم الأثر وفقاً لنموذجي الأثر الثابت والعشوائي ومؤشرات التجانس لقوة الاختبار الإحصائي

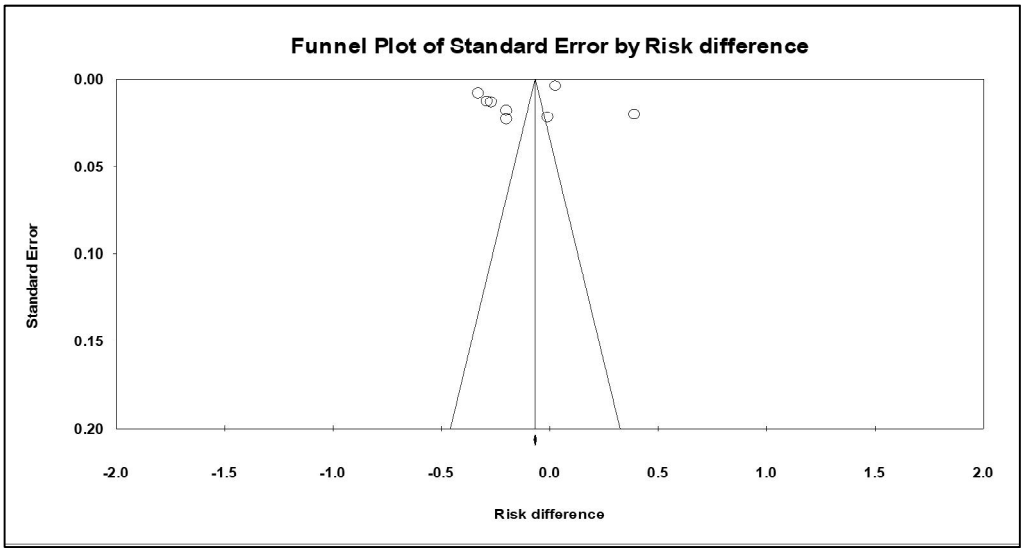
النموذج	عدد حجوم الأثر	وسط حجوم الأثر	الخطأ المعياري	مؤشرات التجانس: اختبار (Q)	درجات الحرية	P- value	اختبار (I^2)
الأثر الثابت	8	-0.063	0.003	2591.8	7	0.00	99.7
الأثر العشوائي	8	-0.111	0.078				

يظهر من الشكل (4) أن قيم حجم الأثر للدراسات المختلفة تتراوح بين (0.33-) و(0.39)، وأن القيم سالبة في (6) دراسات تتفوّق فيها طريقة نسبة الأرجحيّة(مشار لها في الشكل بالحرفA) على طريقة مانتل هانزل (مشار لها في الشكل بالحرفB في قوة الاختبار الإحصائيّ؛ وقد كانت القيم موجبة في دراستين تتفوّق فيهما طريقة مانتل هانزل على طريقة نسبة الأرجحيّة في قوّة الاختبار الإحصائيّ. ويتّضح من الجدول (6) أنّ وسط حجم الأثر الكلّيّ للدّراسات التي بحثت في قوّة الاختبار الإحصائيّ يساوي (0.068-) وفق نموذج الأثر الثّابت، و(0.111-) وفق نموذج الأثر العشوائيّ، ويشير كلاهما إلى تفوّق طريقة نسبة الأرجحيّة على طريقة مانتل هانزل في قوة الاختبار الإحصائيّ. وفيما يتعلّق بعدم التّجانس في الدّراسات التي تناولت قوّة الاختبار الإحصائيّ، يبيّن اختبار Q أنّ الدّراسات كانت غير متجانسة في نتائجها حيث بلغت قيمة الاختبار 2591.8 وهي دالّة إحصائيًا عند مستوى الدّلالة α=0.05، كما يبيّن مؤشر I^2 أنّ نسبة عدم التّجانس تساوي 99.7%، وهذا يعني أنّ حجوم الأثر متباينة بشكل واضح لدراسات قوّة الاختبار الإحصائيّ، وبناءً عليه فإنّ ثمة ضرورة لإجراء تحليل خاصّ يتناول ظروف تلك الدّراسات وما تضمّنته من متغيّرات.

ويبيّن الشكل (5) قيم حجم الأثر والأخطاء المعياريّة للدّراسات التي تناولت قوّة الاختبار الإحصائيّ بين طريقتي نسبة الأرجحيّة ومانتل هانزل في الكشف عن الأداء التّفاضليّ. وقد تركّزت معظم الدّراسات في الجزء الأعلى(حيث يكون الخطأ المعياريّ قريب من الصّفر) خارج القمع بشكل متماثل حول الخط العموديّ الذي يمثّل وسط قيم حجم الأثر، الأمر الذي قد يشير إلى عدم وجود التحيز. وقد كشفت نتائج اختبار ايجر Egger`s regression test of the intercept عن عدم وجود التّحيّز؛ إذ بلغت قيمة اختبار ايجر (t =0.875, p=0.21).

الشكل (5)

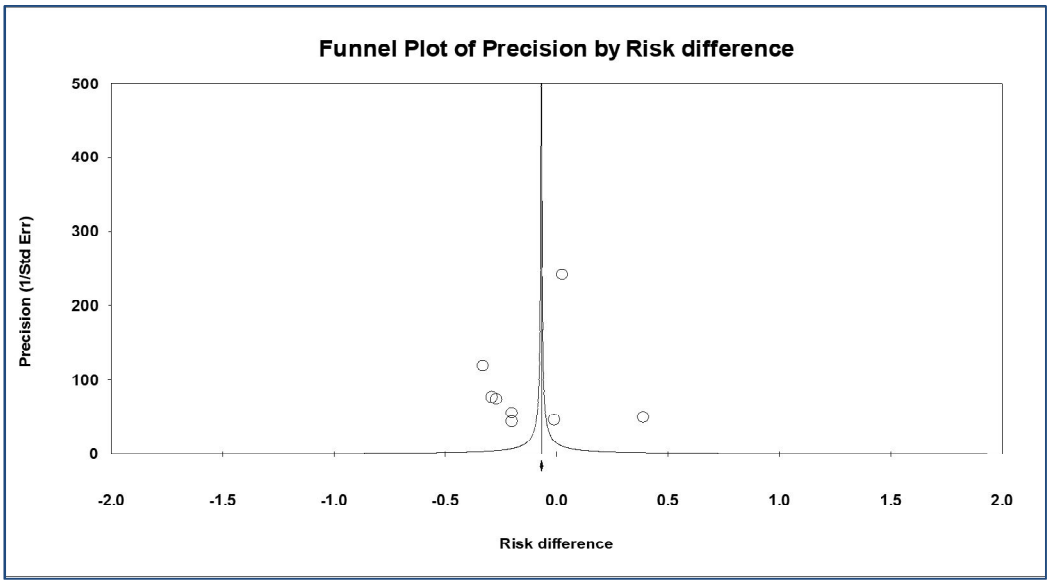
قيم حجم الأثر والأخطاء المعيارية للدراسات التي تناولت قوة الاختبار الإحصائي بين طريقتي نسبة الأرجحية ومانتل هانزل في الكشف عن الأداء التفاضلي.



ويبيّن الشكل (6) أنّ قيم حجم الأثر للدّراسات قد توزّعت على جانبي الخط العموديّ، وهذا يشير إلى عدم وجود التّحيّز. وباستخدام طريقة التقليم والتّعبئة، فقد تبيّن عدم حدوث أيّ تغيير، وهذا يؤكّد عدم وجود أثر للتّحيّز في النتائج (Steenbergen & Olszewski, 2016).

الشكل (6)

قيم حجم الأثر وما يقابلها من مؤشر دقة تقديره.



وتتّفق هذه النّتيجة مع نتائج دراسات كلّ من امائدا (Amanda, 2008) والنفيعي (2008) واللبدي (2008) التي أظهرت أنّ طريقة نسبة الأرجحيّة أقوى من طريقة مانتل هانزل في الكشف عن الأداء التّفاضليّ للفقرات. وتتعارض مع نتائج دراسة كوندراتك وجرودنيوسك (Kondratek & Grudniewska, 2013) التي أشارت إلى أنّ طريقة مانتل هانزل أكثر قوّة في الكشف عن الأداء التّفاضليّ للفقرة المنتظم من طريقة نسبة الأرجحيّة، على الرّغم من أنّ نسبة الأرجحيّة كانت أقوى في الكشف عن الأداء التّفاضليّ غير المنتظم. وقد تعزى مثل هذه النتيجة إلى أنّ المتغيّرات والظّروف في الدّراسات السّابقة أثرت بصورة مختلفة في الطّريقتين.

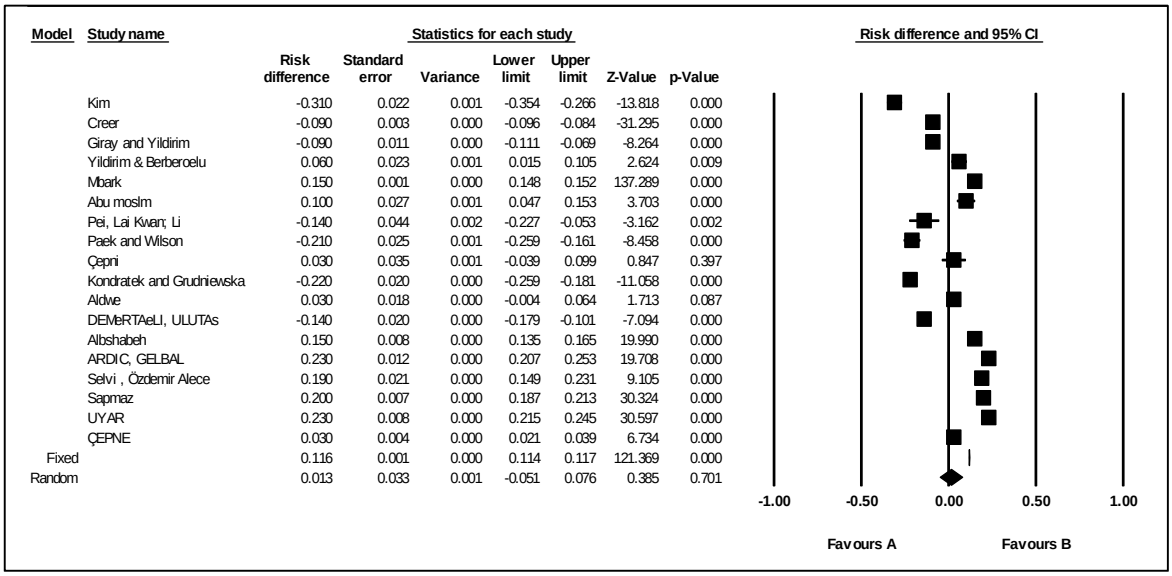
وتجدر الإشارة إلى أنّ هذه النّتيجة اعتمدت على 8 دراسات فقط. وإذا تمّ النّظر إلى هذه الدّراسات باعتبارها عينة عشوائيّة من مجتمع الدّراسات التي اهتمّت بالمقارنة بين الطّريقتين في ضوء درجة قوة الاختبار الإحصائيّ فإن الوسط الحسابيّ لحجم الأثر لهذه الدّراسات وفق نموذج الأثر العشوائيّ يساوي 0.111- وهو غير دال إحصائيًا عند مستوى الدّلالة الإحصائيّة 0.05، حيث أن (Z =-1.422, sig=0.155). وهذا يشير إلى أنّ الطّريقتين متكافئتين من النّاحية العمليّة في قوّة الاختبار الإحصائيّ

ثالثًا: النتائج المتعلقة بالسؤال الثالث، والذي نص على: *"ما دلالة حجم الأثر الكلّيّ للدّراسات التي اهتمّت بمقارنة نسبة الفقرات ذات الأداء التّفاضليّ بين طريقتي نسبة الأرجحية ومانتل هانزل في الكشف عن الأداء التّفاضليّ للفقرة؟"*

يبيّن الشّكل (7) نتائج التّحليل البعديّ للدّراسات التي تناولت نسبة الأداء التّفاضليّ بين طريقتي نسبة الأرجحيّة ومانتل هانزل في الكشف عن الأداء التّفاضليّ للفقرة. كما يبيّن الجدول (7) وسط حجوم الأثر وفترة الثقة وفقًا لنموذجي الأثر الثّابت والعشوائيّ ومؤشّرات التّجانس لنسبة الأداء التّفاضليّ للفقرة.

الشكل (7)

نتائج التحليل البعدي لدراسات نسبة الأداء التفاضلي للفقرات بين طريقتي نسبة الأرجحية ومانتل هانزل في الكشف عن الأداء التفاضلي.



الجدول (7)

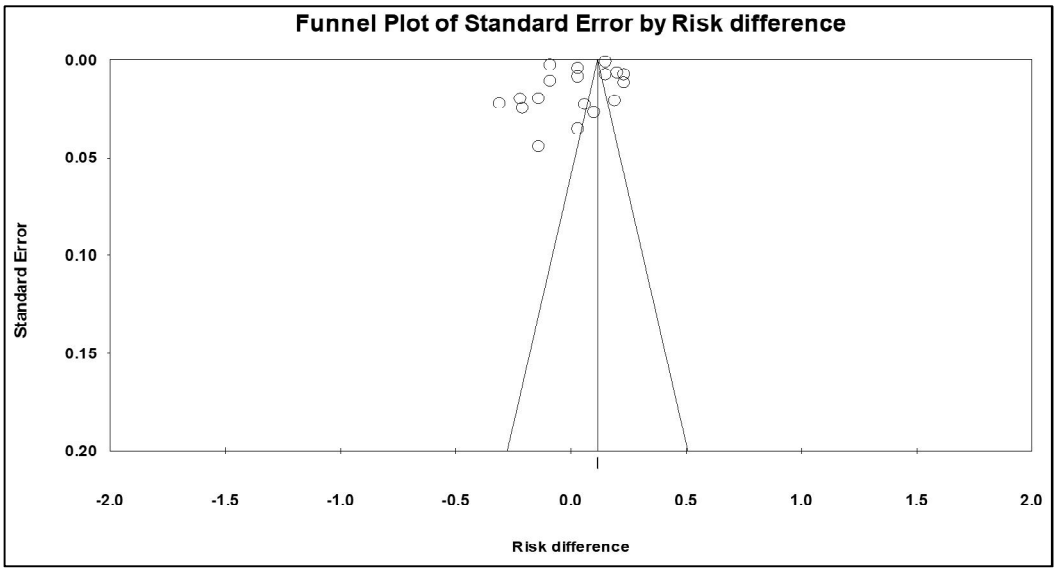
وسط حجوم الأثر وفقاً لنموذجي الأثر الثابت والعشوائي ومؤشرات التجانس لنسبة الأداء التفاضلي للفقرة.

			مؤشرات التجانس			
النموذج	عدد حجوم الأثر	متوسط حجوم الأثر	الخطأ المعياري	اختبار (Q)	درجات الحرية	P- value	اختبار (I^2)
الأثر الثابت	18	0.116	0.001	8406.7	17	0.00	99.8
الأثر العشوائي	18	0.013	0.033				

ويظهر من الشّكل (7) أن قيم حجم الأثر للدّراسات المختلفة تتراوح بين (0.31-) و(0.23)، وقد كانت موجبة في (11) دراسة تتفوّق فيها طريقة مانتل هانزل (مشار لها في الشّكل بالحرف B) على طريقة نسبة الأرجحيّة (مشار لها في الشكل بالحرف A) في نسبة الفقرات ذات الأداء التفاضليّ، وسالبة في (7) دراسات تتفوّق فيها طريقة نسبة الأرجحيّة على طريقة مانتل هانزل. ويتّضح من الجدول (7) أن وسط حجم الأثر الكلّيّ للدراسات التي بحثت في نسبة الأداء التّفاضليّ للفقرة يساوي (0.116) وفق نموذج الأثر الثّابت، و(0.013) وفق نموذج الأثر العشوائيّ، ويشير كلاهما إلى تفوّق طريقة مانتل هانزل في نسبة ما تكشفه من فقرات ذات أداء تفاضليّ. ويبيّن اختبار Q أنّ الدّراسات غير متجانسة في نتائجها حيث بلغت قيمة الاختبار 8406.7 وهي دالّة إحصائيًا عند مستوى الدّلالة α=0.05، كما يبيّن مؤشّر I^2 أنّ نسبة عدم التّجانس تساوي 99.8%، وهذا يعني أنّ حجوم الأثر متباينة بشكل واضح لدراسات نسبة الأداء التّفاضليّ، وبناءً عليه فان ثمة ضرورة لإجراء تحليل خاصّ يتناول ظروف تلك الدّراسات وما تضمّنته من متغيّرات. ويبيّن الشكل (8) قيم حجم الأثر والأخطاء المعياريّة في تلك الدّراسات.

الشكل (8)

قيم حجم الأثر والأخطاء المعيارية للدراسات التي تناولت نسبة الأداء التفاضلي للفقرة بين طريقتي نسبة الأرجحية ومانتل هانزل في الكشف عن الأداء التفاضلي.



ويلاحظ من الشكل (8) أنّ معظم الدّراسات قد تركّزت في الجزء الأعلى (حيث يقترب الخطأ المعياريّ من صفر) خارج القمع بشكل غير متماثل حول الخط العموديّ، ويشير ذلك إلى احتماليّة وجود التّحيّز. وقد كشفت نتائج اختبار ايجر Egger`s regression test of the intercept عدم وجود التّحيّز؛ إذ بلغت قيمة اختبار ايجر (t =1.458, p=0.08)، وتشير هذه النّتيجة إلى عدم وجود التّحيّز في ما هو متاح من الدّراسات.

ولقياس أثر التّحيّز المحتمل على نتائج التّحليل البعديّ، فقد تمّ استخدام طريقة التقليم والتّعبئة، وذلك من خلال تقدير وتضمين قيم حجم الأثر الصّغيرة للدّراسات التي يحتمل أنّه لم يتم تضمينها في التّحليل البعديّ، ويبيّن الجدول (8) النّتائج الإحصائيّة لهذه الطّريقة.

الجدول (8)

القيم المقدرة لحجم الأثر لنموذجي الأثار الثابتة والعشوائية وتجانس التباين بحسب طريقة التقليم والتعبئة Duval and Tweedie`s Trim and Fill.

	الدراسات المحذوفة	حجم الأثر الثابت	حجم الأثر العشوائي	Q
القيم الملاحظة	-	0.116	0.013	8406.7
القيم المعدلة	6	0.120	0.115	9843.6

ويلاحظ من الجدول (8) أنّ حجم الأثر قد غيّر قيمته من (0.116) إلى (0.120) بعد التّعديل. وهذا مؤشّر لوجود التّحيّز الذي تزداد شدته بزيادة الفرق بين حجم الأثر الملاحظ Observed وحجم الأثر المعدل Adjusted ممّا يترتّب عليه تغير الحكم المتعلّق بدرجة فعاليّة المعالجة (Steenbergen & Olszewski, 2016).

وقد تمّ تحديد عدد الدّراسات التي يلزم إضافتها إلى التّحليل البعديّ من أجل إعادة ضبط معدل حجم الأثر بعدما تغيّرت قيمته باستخدام طريقة التقليم والتّعبئة السّابقة، " Classic fail- safe N" (Borenstein et al., 2009) كما هو موضّح في الجدول (9).

الجدول (9)

عدد الدراسات التي لم تتح لها فرصة الظهور في العينة Classic fail- safe N.

الإحصائي	القيمة
قيمة z للدراسات الملاحظة	42.305
قيمة p للدراسات الملاحظة	0.000
الفا	0.05
الذيل	2
قيمة z لالفا	1.96
عدد الدراسات الملاحظة	18
عدد الدراسات المفقودة التي تجعل قيمة p اكبر من الفا	8369

حيث تشير النّتائج إلى أنّ عدد الدّراسات التي يلزم إضافتها لإرجاع قيمة الإحصائيّ Z للدّراسات الملاحظة من 42.3 إلى قيمته الأسميّة هو 8369 دراسة، وهذا عدد كبير جدًا مقارنة بعدد الدّراسات التي تضمّنها التّحليل البعديّ، وهذا يعني ان مؤشّرات التّحيّز التي ظهرت في مؤشّرات شكل القمع لم تكن على درجة كبيرة من الأهميّة وبالتالي يمكن القول بعدم وجود التّحيّز في نتائج التّحليل البعديّ في هذا البحث.

وتتّفق نتائج هذه الدّراسة مع نتائج دراسات كلّ من مبارك (Mubarak, 2010) وأبو مسلم (Abu Muslim, 2010) التي أظهرت تفوّق طريقة مانتل هانزل على طريقة نسبة الأرجحيّة في الكشف عن نسبة أعلى من الفقرات ذات الأداء التّفاضليّ. وتتعارض مع نتائج دراسات كل من كرير (Creer, 2004) وباي (Pae, 2004) التي أشارت إلى أنّ طريقة نسبة الأرجحيّة أفضل من طريقة مانتل هانزل في الكشف عن نسبة أعلى من الفقرات ذات الأداء التّفاضليّ، وتضعف طريقة مانتل هانزل في الكشف عن الأداء التّفاضليّ من النّوع غير المنتظم.

وتجدر الإشارة إلى أنّ هذه النّتيجة اعتمدت على 18 دراسة فقط. وإذا تمّ النّظر إلى هذه الدّراسات باعتبارها عينة عشوائيّة من مجتمع الدّراسات التي اهتمّت بالمقارنة بين الطّريقتين في ضوء نسبة الأداء التّفاضليّ في الفقرات فإن الوسط الحسابيّ لحجم الأثر لهذه الدّراسات وفق نموذج الأثر العشوائيّ يساوي 0.013 وهو غير دالّ إحصائيًا عند مستوى الدّلالة الإحصائيّة 0.05 (Z = 0.033, sig=0.701). وهذا يشير إلى أنّ الطّريقتين متكافئتين من النّاحية العمليّة في نسبة الأداء التّفاضليّ.

التّوصيات

بناءً على النّتائج التي تمّ التّوصّل إليها توصي الدّراسة بما يلي:

- إجراء التّحليل البعديّ لطرق أخرى في الكشف عن الأداء التّفاضليّ للفقرة مثل طريقة الاختبار المتزامن وطريقة الانحدار اللوجستيّ.
- يمكن استخدام أيّ من الطّريقتين في تحليلات الأداء التّفاضليّ للفقرات، حيث لا توجد فروق دالّة إحصائيًا بين الطريقتين سواء في ضبط الخطأ من النّوع الأوّل أو قوّة الاختبارات الإحصائيّة أو في نسبة الأداء التّفاضليّ للفقرات.

References

Abu Muslim, Maysa. (2010). Detection of test item bias according to the IRT and Mantel-Hansel (HM) approaches: a comparative study. *Journal of the College of Education.* 20(2), 189-237.

Al-Autoom, S. (2005). *A meta-analysis of the effect of several instructional strategies on students achievement in science*. Unpublished Master Thesis, Yarmouk University.

Al-Bashabsha, K. (2016). *Detection of the differential performance of the gender variable in the international PISA test for the year 2012*. Unpublished Master's Thesis, Mutah University.

Al-Dawy, A. (2015). Investigating the differential performance of the item according to gender in the Organizing Ability Questionnaire. Education (Al-Azhar): *Refereed Scientific Journal for Educational Psychological and Social Research*, 34(162) 1, 11-65.

Alghamdi, A. (2017). *Impact of Unbalanced Designs on the Performance of Parametric and Nonparametric DIF Procedures: A Comparison of Mantel Haenszel, Logistic Regression, SIBTEST, and IRTLR Procedures.* Ph.D. thesis.Florida State University.

Aljughaiman, A. & Ayoub, A. (2013). Evaluating the effects of the Oasis Enrichment Model on gifted education: A meta-analysis study. *Talent Development & Excellence.* 5(1), 99-113.

Al-Labadi, Nizar. (2008). *Comparison of four methods for detecting differential item accuracy: a simulation study*. Unpublished Ph.D. Thesis, University of Jordan.

Al-Nafi'i, Abdul Rahman. (2008). *Comparing the effectiveness of a number of statistical methods to detect the differential performance of test items and its effect on type one error and test power*. Unpublished Ph.D. Thesis, Umm Al-Qura University.

Al-Sharifeen, Nidal. (2017). Meta-analysis of research published in the Jordanian Journal of Educational Sciences: Type I error and Test Power. *Journal of the Association of Arab Universities for Education and Psychology*, 15(3), 130-170.

Amanda, M. (2008). *Methods for identifying differential item and test functioning: an investigation of type I error rates and power.* Unpublished Doctoral Dissertation, James Madison University.

Ankenmann, R. D., Witt, E. A. & Dunbar, S. B. (1999). An investigation of the power of the likelihood ratio goodness-of-fit statistic in detecting differential item functioning. *Journal of Educational Measurement,* 36(4), 277-300.

Ardic, E. O. & Gelbal, S. (2017). Cross-Group Equivalence of Interest and Motivation Items in PISA 2012 Turkey Sample. *Eurasian Journal of Educational Research,* 17(68), 221-238.

Berk, R. A. (Ed). (1982). *Handbook of methods for detecting test bias.* Johns Hopkins University press.

Borenstein, M., Hedges, L., Higgins. J. & Rothstein, H. (2009). *Introduction to meta-analysis,* 1st edition John Wiley & Sons. Ltd.

Camilli, G. & Shepard, L. (1994). *Methods for Identifying biased test item.* Thousand Oaks, CA: sage.

Çepni, Z. (2011). *Differential item functioning analysis using SIBTEST, Mantel Haenszel, logistic regression and item Response Theory Methods*. Unpublished Doctoral Dissertation. Hacettepe University, Ankara. Turkey.

Çepni, Z. & Kelecioglu, H. (2021). Detecting differential item functioning using SIBTEST, MH, LR and IRT methods. *Journal of Measurement and Evaluation in Education and Psychology,* 12(3), 267-285.

Creer, T. G. (2004*). Detection of differential item functioning (DIF) on the SATV: A comparison of four methods*: *Mantel-Haenszel, logistic regression, simultaneous item bias, and likelihood ratio test.* Ph.D. Dissertation, University of Houston.

Demirtasli, N. & Ulutas, S. (2015). A Study on Detecting of Differential Item Functioning of PISA 2006 Science Literacy Items in Turkish and American Samples. *Eurasian Journal of Educational Research,* 58, 41-60.

Embreston, S. E. & Reise, S (2000). *Item response theory for psychologists.* Lawrence Erlbum Associated

Finch, H. (2005). The MIMIC model as a method for detecting DIF: Comparison with Mantel-Haenszel, SIBTEST, and the IRT likelihood ratio. *Applied Psychological Measurement,* 29(4), 278-295.

Gao, X. (2019). *A comparison of six DIF detection methods*. Master Thesis, University of Connecticut

Gay, L. (2000). *Educational research: Competencies for analysis and application.* Prentice Hall.

Gellin, A. (2003). The effect of undergraduate student involvement on critical thinking: A meta-analysis of the literature 1991-2000. *Journal of College Student Development,* 44(6). ProQuest Central. Pg. 746.

Gioacchino, L. (2005). *Meta-analysis in medical research: The handbook for the understanding and practice of meta-analysis,* Blackwell.

Guilera, G., Gómez-Benito, J., Hidalgo, M. D. & Sánchez-Meca, J. (2013). Type I error and statistical power of the Mantel-Haenszel procedure for detecting DIF: A meta-analysis. *Psychological Methods,* 18(4), 553.

Huang, J. & Han, T. (2012). Revisiting differential item functioning implications for fairness investigation. *International Journal of Education,* 4(2), 74-86.

Kabasakal, K., Arsan, N., Gök, B. & Kelecioglu, H. (2014). Comparing Performances (Type I Error and Power) of IRT Likelihood Ratio SIBTEST and Mantel-Haenszel Methods in the Determination of Differential Item Functioning. *Educational Sciences: Theory and Practice,* 14(6), 2186-2193.

Keklik, B. (2012). Determination of leadership style preferred in health institutions: Example of a private hospital. *Afyon Kocatepe University Journal of Economics and Administrative Sciences,* 14(1), 73-93.

Kim, M. (2016). A meta-analysis of the effects of enrichment programs on gifted students. *Gifted Child Quarterly,* 60(2), 102-116.

Kim, S. H. (2000). *An Investigation of the Likelihood Ratio Test, the Mantel Test, and the Generalized Mantel-Haenszel Test of DIF*. Paper presented at the Annual Meeting of the American Educational Research Association (New Orleans, LA, April 24-28, 2000).

Kondratek, B. & Grudniewska, M. (2013). Comparison of Mantel–Haenszel test with IRT procedures for DIF detection and effect size estimation for dichotomous items. *Edukacja,* 2, 34-55.

Michaelides, M. P. (2008). A mantel-haenszel procedure in test equating. *Practical Assessment Research & Evaluation,* 13(7), 4-6.

Mubarak, W. (2010*). Differential functioning of science test items in the international study Pisa 2006*. Unpublished Ph.D. Dissertation, Yarmouk University.

Omar, M. (2013). Indicators of dimensional analysis of the results of some scientific productions in the field of improving the condition of children with autism disorder in the Arab world in the period 1989-2013. *Journal of Special Education and Rehabilitation,* 1(1), 1-74.

Pae, T. I. (2003). Parameter invariance for males and females on an English listening comprehension test. *English Teaching,* 58(2), 143-158.

Paek, I. & Wilson, M. (2011). Formulating the Rasch differential item functioning model under the marginal maximum likelihood estimation context and its comparison with Mantel–Haenszel procedure in short test and small sample conditions. *Educational and Psychological Measurement,* 71(6), 1023-1046.

Park, C. (2010). Examining the relationship between differential item functioning and differential test functioning. *Language Testing,* 23(4). 475-496.

Pei, L. K. & Li, J. (2010). Effects of unequal ability variances on the performance of logistic regression, Mantel-Haenszel, SIBTEST IRT, and IRT likelihood ratio for DIF detection. *Applied Psychological Measurement,* 34(6), 453-456.

Petitti, D. (1994). *Meta-analysis, decision analysis, and cost-effectiveness analysis methods for quantitative synthesis in medicine.* Prentice Hall.

Pigott, T. (2012). *Advances in meta-analysis. Springer Science & Business Media.*

Price, E. A. (2014). *Item discrimination, model-data fit, and type I error rates in DIF detection using lord's ch. 2, the likelihood ratio test, and the mantel-haenszel procedure*. Unpublished Ph.D. Dissertation, Ohio University.

Sapmaz, Z. M. (2019). *Detection of Gender-Related Differential Item Functioning (DIF) in the Mathematics Subtests in Turkey*. University of South Florida.

Selvi, H. & Alici, D. Ö. (2018). Investigating the impact of missing data handling methods on the detection of differential item functioning. *International Journal of Assessment Tools in Education,* 5(1), 1-14.

Steenbergen-Hu, S. & Olszewski,-Kubilius, P. (2016). How to conduct a good meta-analysis in gifted education. *Gifted Child Quarterly,* 60(2), 134-154. Doi:10. 1177/0016986216629545.

Thissen, D., Steinberg, L. & Wainer, H. (1988). *Use of item response theory in the study of group differences in trace lines*. In: H. Wainer & H. Braun (Eds.), *Test Validity*. Erlbaum, pp. 147-169.

Uyar, S. (2020). Latent Class Approach to Detect Differential Item Functioning: PISA 2015 Science Sample. *Eurasian Journal of Educational Research,* 88, 179-198.

Yildirim, H. H. & Giray, B. (2006). *The differential item functioning (DIF) analysis of mathematics items in the international assessment programs*. Unpublished doctoral thesis). Middle East Technical University, Ankara, Turkey.

Yildirim, H. H. & Berberoğlu, G. (2009). Judgmental and statistical DIF analyses of the PISA-2003 mathematics literacy items. *International Journal of Testing,* 9(2), 108-121.