صلاحية تفسير نتائج الاختبارات النفسية في ظل مفهوم الأداء التفاضلي للبنود

عبد العزيز بوسالم: أستاذ التعليم العالي

أحمد كريش: أستاذ مساعد "ب"

كلية العلوم الاجتماعية والانسانية، جامعة البليدة 2

ملخص:

نتناول في هذا المقال دور تأثير التفاضلي للبند على صلاحية تفسير نتائج الاختبارات النفسية على اعتباره مفهوما ظهر حديثا في قاموس القياس النفسي وبناء الاختبارات على المستوى العالمي، وسوف نوضح مصادر تحيز البنود وتأثيره السلبي على صدق تفسير نتائج مختلف الاختبارات النفسية والقرارات المتخذة بشأن الأفراد، ومن أجل ذلك وبالتالي يصبح الاختبار غير صالحا ويزيد احتمال تحيزه مما يهدد صدق الاستدلال بدرجاته ويقلل من الثقة والدقة في التفسيرات المترتبة عليها. وللتعامل مع مشكلة التحيز طوّر الباحثون بعض الطرق الإحصائية للكشف عن خصائص أداء بنود الاختبار ومدى اختلافها حسب انتماء الأفراد إلى مجموعات معينة، وهو ما أصبح يعرف بالأداء التفاضلي للبند، الذي يعد من الطرق الحديثة لتحليل بنود الاختبار، ومن الأدلة المهمة على صدق الاستدلال من درجات الاختبار وصولا إلى التأويلات النفسية والاجتماعية للنتائج.

الكلمات المفتاحية: الأداء التفاضلي - البند - البنية العاملية - التفسير - الصدق الثقافي.

Abstract:

This article provides an overview of one of the most recent and sophisticated methods of item analysisis Differential Item Functioning (DIF) and gives a glance of the sources and explanation of its occurrence in the items of a test. The DIF gives a strong evidence for the fairness of the test and the validity of inferences based on the test scores.

The DIF occurs when examinees with equal latent trait, but from different groups for example (male and female), have an unequal probability of item endorsement.

Key words: Factor Analysis, Differential Item Functioning. Statistical Properties

مقدمة:

عندما نقوم بتطبيق اختبار ما على مجموعات من الأفراد غير متجانسة نتحصل على تباين في الدرجات مما يطرح تساؤلا حول هذا التباين في الدرجات هل هو راجع إلى اختلاف في مستويات السمة المقاسة التي يقيسها الاختبار عند هؤلاء الأفراد ،أم أن هذه الاختلافات راجعة إلى خصائص أخرى في البنود ليس لها علاقة بهذه السمة الكامنة وبالتالي تصبح البنود تفضل أفراد بسبب انتمائهم إلى مجموعة معينة مثل الجنس (ذكور وإناث)، أو العمر (أكبر من 60 سنة وأصغر من 60 سنة)،أو اللغة (العربية والإنجليزية). وبالتالي يصبح الاختبار غير عادلا واحتمال أن يكون متحيزا وهذا ما يهدد صدق الاستدلال بدرجاته ويقلل من الثقة والدقة في التفسيرات المترتبة عليها. وللتعامل مع هذا الموضوع طور الباحثين بعض الطرق الإحصائية للكشف عن خصائص أداء بنود الاختبار هل يختلف باختلاف انتماء الأفراد إلى مجموعات معينة، وهو ما أصبح يعرف الآن بالأداء التفاضلي للبند، التي تعد من الطرق الحديثة لتحليل بنود الاختبار، ومن الأدلة المهمة على صدق الاستدلال من درجات الاختبار.

1- مفهوم الأداء التفاضلي للبند:

يعرف شن (Chen)[1] الأداء التفاضلي للبند بأنه الحالة التي يكون فيها أفراد من مجموعات مختلفة (العمر، الجنس، العرق، التعليم، المهنة) على نفس المستوى من السمة الكامنة، لهم احتمال مختلف لإعطاء استجابة معينة على بند معين. ويذكر هاملتون(Hambleton,1991)[2] بأن مصطلح الأداء التفاضلي للبند يشير إلى الاختلافات في الخصائص الإحصائية (Statistical Properties) لبند معين بين مجموعات من المفحوصين لهم قدرة متساوية وغالبا ما تكون المجموعات مرجعية (Reference group) والتي تحتوي على الأغلبية. ومجموعة مركزية (Focal group) التي تحتوي على الأقلية أو الفئة المدروسة.

2- الأداء التفاضلي للبند والبنية العاملية للاختبار:

أنه قبل القيام بتحليل الأداء التفاضلي للبند في بنود الاختبار يجب أولا أن نتأكد من أن البنية العاملية للاختبار متكافئة عبر المجموعات، لأنه افتراض أساسي للقيام بالتحليل، فإذا لم يتوفر هذا الافتراض فيمكن أن يختلف أداء البند أساسا بسبب اختلاف البنية العاملية للاختبار عبر المجموعات وليس بسبب احتواء البند على الأداء التفاضلي للبند. وهذا ما أشار إليه كلوسر (Clauser)[3] بأن تحليل الأداء التفاضلي للبند صمم لتحديد البنود التي يكون أدائها مختلفا بالنسبة إلى محك محدد، فإذا كانت كل بنود الاختبار تعطي الأفضلية لمجموعة على أخرى فإن تحليل الأداء التفاضلي للبند يكون من دون فائدة.

إن كلام كلوسر يحمل نفس المعنى لكلام تريزي فإذا كانت جميع بنود الاختبار التي تشكل البنية العاملية التي يقيسها الاختبار غير متكافئة أساسا بسب عدم ثبات القياس بين المجموعات، فتحليل الأداء التفاضلي للبند في هذه الحالة يعتبر غير ممكن لأن المحك الذي نقوم بواسطته مطابقة المجموعات وهو غالبا ما يكون نفس المستوى من السمة الكامنة غير ممكن، لأن اختلاف البنية العاملية يعني وجود اختبارين يقيسان هذه السمة الكامنة بطريقتين مختلفتين وبالتالي لا يمكن مقارنة الدرجات الكلية التي تحصل عليها المفحوصين، وبالتالي الاختلاف موجود أولا وأساسا بين المجموعتين في الاختبار ككل ولا داعي للبحث عنه بواسطة تحليل الأداء التفاضلي للبند. لأن اختلاف الكل يعني اختلاف الجزء.يشير بودجال (Budgell)[4] إلى ضرورة التحقق من أحادية البعد للاختبار، لتحديد مدى تكوين البنود لمجموعات أحادية البعد، وذلك للقيام بتحليل الأداء التفاضلي للبند، ويمكننا القيام بذلك بواسطة إجراء التحليل العاملي الاستكشافي. لكي يكون اختبار ما أحادي البعد يجب أن يكون هناك عامل كامن واحد له أكبر جذر كامن، وتليه عوامل بجذور كامنة صغيرة، وهذا يعني وجود عامل مسيطر يفسر أكبر نسبة تباين. ويضع ريكازي (Reckase)[5] محك الجذر الكامن لقياس أحادية البعد تعتمد على نسبة التباين على الأقل (20%). وهذا يعني أنه لافتراض أحادية البعد للاختبار معين يجب أن يكون هناك عامل كامن مسيطر يفسر على الأقل (20%) من التباين الملاحظ.

3- الأداء التفاضلي للبند وصدق درجات الاختبار

يعد الأداء التفاضلي للبند من الأخطاء المنتظمة في الاختبارات النفسية، وهو من مهددات صدق درجات الاختبار، لأنه يحدث عندما يكون الأفراد من نفس مستوى السمة

المقاسة يجيبون إجابات مختلفة على البند نفسه اعتمادا على انتمائهم لمجموعة معينة، وهذا يعني أن الإجابة على البند تخضع إلى متغير غير التكوين الفرضي الذي يسعى الاختبار إلى قياسه، فمثلا في اختبار يقيس الإكتئاب هناك بند يسأل عن عدد مرات البكاء في اليوم، إن هذا البند قد يظهر تأييدا من النساء أكثر من الرجال رغم أن لهم نفس المستوى من الاكتئاب، وبالتالي يحتوي هذا البند على الأداء التفاضلي للبند راجع إلى متغير الجنس، لأن النساء عادة أكثر بكاءا من الرجال.غير أن هاملتون(Hambleton, 2005)[6] يشير إلى أن تفسير نتائج تحليل الأداء التفاضلي للبند ليكون ذو معنى يفترض مسبقا جودة الأدلة على صدق المفهوم (التكوين الفرضي) والصدق التنبؤي لدرجات الاختبار الذي نقوم بدراسته. كما أن نتائج تحليل الأداء التفاضلي للبند تعتبر من أشكال أدلة الصدق التي يجمعها الباحث.ومع النظرية الحديثة للصدق يشير زومبو (Zumbo,1999)[7] إلى أننا لم نعد نتحدث عن أنواع للصدق بل أصبحنا نكتفي بنوع واحد، وصار الصدق عملية جمع أكبر عدد من الأدلة على صلاحية تأويل أو تفسير درجات الاختبار، وهو عملية مستمرة، ولقد أصبح من الضروري التأكد من عدم تحيز البنود بواسطة الأداء التفاضلي للبند كمؤشر أساسي للتحقق من صدق درجات الاختبار.

4- الأداء التفاضلي وتحيز البند:

يذكر كل من (Holland, Thayer 1988)[8] بأن بداية الاهتمام بمفهوم الأداء التفاضلي للبند تعود إلى الستينات من القرن الماضي في مجال الاختبارات التحصيلية، وكان يشار إليه في بادئ الأمر على أنه تحيز البند (item bias). ويكون البند متحيزا عندما يفضل (يكون أكثر سهولة للإجابة عليه إجابة صحيحة، أو أكثر تأييدا) مجموعة من المفحوصين، ولكن يكون ضد (صعب الإجابة عليه، أو أقل تأييدا) مجموعة أخرى، بسبب بعض خصائص البند التي ليس لها صلة مباشرة بغرض الاختبار(قياس السمة الكامنة). تذكر إليس (Ellis, and Raju,2003)[9] أنه في أواخر الثمانينات حل مصطلح الأداء التفاضلي للبند محل تحيز البند، ويحتمل أن يعود هذا التغيير إلى سبب سياسي ولغوي أكثر منه نفسي قياسي، لأن مصطلح تحيز البند يحمل معه دلالة سلبية وهو يرتبط عموما مع مفهوم غير عادل وله طابع سياسي واجتماعي أكثر منه ارتباطا بمصطلحات القياس النفسي، وبذلك تيقن علماء القياس النفسي أنه من المفيد التفريق بين المصطلحات التقنية للقياس النفسي وتلك المصطلحات المشحونة سياسيا واجتماعيا، وبذلك تم تغيير

المصطلح.فتحليل الأداء التفاضلي للبند يعطي إمكانية فحص الاختلافات بين المجموعات في أداء الاختبار والكشف عن ما إذا كانت هذه الاختلافات تشير إلى تحيز البند أم لا. فبنود الاختبار تكون متحيزة بسبب أنها تحتوي على مصادر صعوبة ليست ذات صلة، أو تكون دخيلة على التكوين الفرضي (السمة) التي تهدف البنود لقياسه، وهذا ما يطلق عليه تحيز الاختبار (Test Bias) الذي كان موضوع بحوث كثيرة، وتحول إلى موضوع الأداء التفاضلي للبند، حيث أصبح المعيار الجديد لتحليل تحيز الاختبارات. وحسب شن (Chen)[10] فقد ظهر مصطلح الأداء التفاضلي للبند بعد إدراك أن الأداء المختلف للبند لا يظهر دائما بسبب وضعية متحيزة (تفضيل مجموعة على مجموعة)، بل يكون أداء البند مختلفا اعتمادا على الانتماء لمجموعة معينة. فالأداء التفاضلي للبند يعتبر ضروري للدلالة على تحيز البند، لكن يعتبر غير كاف للجزم بذلك، ففي حالة احتواء أحد البنود على الأداء التفاضلي فهذا يعتبر غير كاف للحكم على البند بأنه متحيز، بل على الباحث مواصلة التحليل بالقيام بتحليل المحتوى (Content Analysis) والتقييم التجريبي (Empirical Evaluation) لتحديد إذا ما كان حقيقة البند متحيز، والاستعانة بالمتخصصين في ذلك. فأهمية التفريق بين تحيز البند أو الأداء التفاضلي للبند من جهة، والمحتوى غير المناسب للبند من جهة أخرى إجراء ضروري تماشيا مع الفكرة القائلة بأن محتوى البند المزعج يؤثر في المجموعات على حد سواء ولا تتأثر مجموعة دون الأخرى.

5- تحيز البند وتأثيره على صلاحية تفسير نتائج الاختبار النفسي:

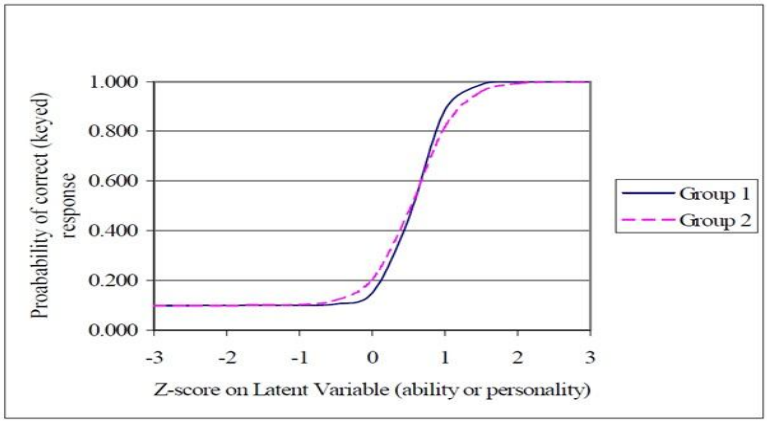
عندما يتباين المفحوصين في مختلف المجموعات في مستويات القدرة المراد قياسها، فهذه الاختلافات في الأداء على الاختبار تكون متوقعة، وهذه النتيجة يمكن أن تفسر بتأثير البند وليس بتحيزه. فتأثير البند يشير إلى اختلاف حقيقي بين الأفراد راجع إلى الاختلاف في مستويات السمة الكامنة وليس بسبب الانتماء إلى مجموعة معينة. فاحتواء بند ما على الأداء التفاضلي يمكن إرجاعه إلى سببين رئيسيين، وهما إما تحيز البند أو تأثير البند، والحكم في الحالتين يعود إلى المختصين في القياس النفسي. لكن هولند (Holland)[11] يوضح أن اختلاف أداء البند يشير إلى تأثير البند في حالة عدم القيام بمطابقة (Unmatched) المفحوصين على محك المطابقة الذي يتم اختياره، مما يجعل هذا الاختلاف راجع إلى تباين الأفراد في السمة الكامنة.

نتيجة لهذه الأسباب وغيرها أصبح نقل الاختبار من بيئة وتطبيقه في بيئة أخرى مختلفة يتطلب الاهتمام باعتبارات سيكولوجية، قياسية، ثقافية أولا، وهذا ما ذهبت

إليه الجمعية العالمية للقياس النفسي في دليلها الصادر عام (1996)[12] والمحرر من طرف عديد علماء النفس والقياس المهتمين بتكييف الاختبارات للاستعمال عبر الثقافي منهم (Hambleton,Bollwark,Van de Vijver) والذي حدد بعض القواعد الواجب احترامها عند محاولة نقل اختبار من ثقافة وتطبيقه في أخرى، قصد تقليل حجم تأثير الأداء التفاضلي للبنود الناتج عن الفروق الثقافية، بداية من تقديم معلومات دقيقة عن خصائص الأفراد المطبق عليهم ومختلف المراحل المعتمدة في إعداد الاختبار في صورته الجديدة ومراحل التقنيين، معادلة الجوانب المقيسة مفاهيميا ووظيفيا، تعليمات الاختبار وطرق تطبيقه، تدريب الفاحصين وتهيئة البيئة المناسبة للتطبيق. ليتم في الأخير توثيق كل هذه الإجراءات المتبعة مع الاختبار للتدليل على سلامة المنهج المتبع ومنه الحديث عن عدم تحيز محتوى البنود لثقافة معينة دون الأخرى ومنه مصداقية النتائج أو صدق تفسيرها.

6- أنواع الأداء التفاضلي للبند:

إن الأداء التفاضلي للبند كعامل من العوامل المهددة لصدق الاختبار النفسي يقاس من الناحية النظرية عن طريق مقارنة منحنى خصائص البند (Item Characteristic Curve ICC). والذي يمثل العلاقة بين أداء المفحوص على البند ومتصل السمة الكامنة لمختلف المجموعات على نفس البند. فإذا كان منحنى خصائص البند لكل من مجموعة أطفال في بريطانيا ومجموعة أطفال في الجزائر متطابق تماما، أو قريب جدا من التطابق، فيمكننا القول بأن هذا البند لا يحتوي على الأداء التفاضلي الناتج عن العامل الثقافي، حسب ما يوضحه الشكل رقم (01).

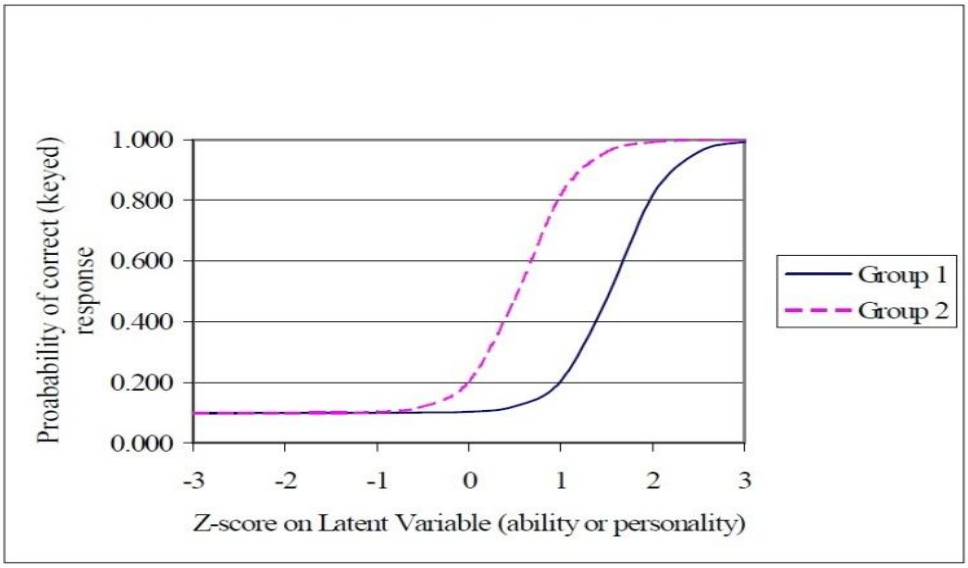


الشكل رقم (01) مثال لبند لا يحتوي على الأداء التفاضلي

إن منحنى خصائص البند يمكن تمييزه بواسطة معلمين، الأول هو الصعوبة (b) أو العتبة (تأييد البند)، وهذا البرامتر يعكس وضعية المنحنى على المحور -س- (انتقال المنحنى إلى اليمين أو إلى اليسار) أين احتمال الإجابة الصحيحة على البند أو تأييده تكون (0.5)، والبرامتر الثاني هو التمييز (a) ويعكس انحدار المنحنى، فالمنحنيات المنبسطة والمسطحة حيث يكون المنحدر قليل، تدل على نقص قدرة البند على تمييز الفروق بين الأفراد في المستويات المختلفة للسمة الكامنة. فإذا كان هناك اختلاف واضح لمنحنى خصائص البند لكل مجموعة يمكن القول بأن هذا البند يحتوي على الأداء التفاضلي والذي ينقسم إلى نوعين:

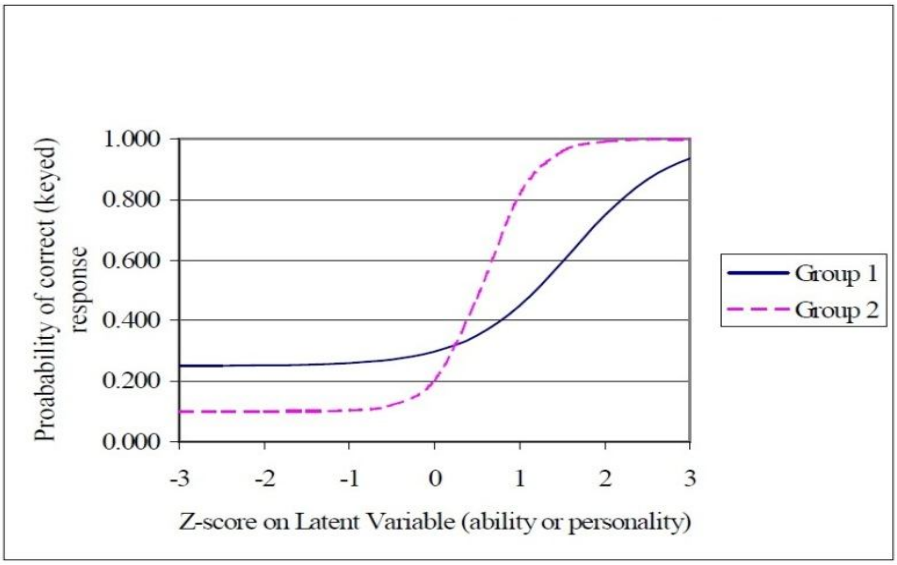
1-6 الأداء التفاضلي المنتظم (Uniform DIF):

تشير تيريزي وفلنشمان (Teresi, and Fleishman 2007)[13] إلى أن الأداء التفاضلي للبند المنتظم يحدث إذا كان احتمال الإجابة متسق، حيث يكون أعلى باستمرار أو أقل باستمرار في واحدة من المجموعات التي نقارنها على جميع مستويات السمة المقاسة. والنوع المنتظم يحدث عندما تكون أحد المجموعتين المرجعية أو المركزية مفضلة عبر كل أبعاد السمة المقاسة. وهذا التفضيل هو منتظم تقريبا عبر جميع بنود الاختبار. ففي هذا النوع هناك احتمال الإجابة الصحيحة على البند يكون أكبر لمجموعة معينة بانتظام على جميع مستويات القدرة، وفي مصطلحات نظرية الاستجابة لبنود الاختبار (IRT) والتي تركز على مدى ملائمة عملية تأويل درجات الأختبار لأغراض معينة؛ أي أن صدق بنود الاختبارات النفسية ينظر إليه حسب طبيعة الأدلة والبيانات (éléments de preuve) التي تقدم للدلالة على مدى دقة تأويل درجات أداء المفحوصين على اختبار معين، ومنه فالأداء التفاضلي المنتظم يشير إلى حالة توازي منحنيات خصائص البند (الشكل رقم 02). حيث يكون المنحنيان متوازيان للمجموعتين لكن يشيران إلى اختلاف في مستوى صعوبة البند (تأييد البند) بين المجموعتين.



الشكل رقم (02) مثال لبند يحتوي على الأداء التفاضلي المنتظم.

6-2 الأداء التفاضلي غير المنتظم (Nonuniform DIF): إن الأداء التفاضلي غير المنتظم للبند يحدث عندما يكون احتمال الإجابة في اتجاهات مختلفة في المجموعات التي نقارنها على مستويات مختلفة من السمة المقاسة، مثل احتمال الاستجابة يكون عالي عند الإناث أكثر من الذكور في المستويات العليا في الاكتئاب ويكون منخفض عند الإناث من الذكور في المستويات المنخفضة من الاكتئاب. كما يشير كلوسر (Clauser, 1998)[14] الى أن النوع غير المنتظم يحدث عندما يكون تفضيل أحد المجموعتين المرجعية أو المركزية يزيد أو ينقص أو يتحول في المستويات الأخرى. وفي إطار نظرية الاستجابة للمفردة الأداء التفاضلي غير المنتظم للبند يعني عدم توازي منحنيات خصائص البند، أي أن المنحنيين يتقاطعان(الشكل رقم 03). وفي هذه الحالة المنحنيان يتقاطعان وهذا يدل على اختلاف في قيمة الانحدار (Slope) للمنحيين معا، مما يشير إلى وجود اختلاف في القدرة التمييزية للبند بين المجموعتين، وفي هذا النوع من الأداء التفاضلي للبند هناك تفاعل بين مستويات القدرة وخصائص المجموعة.



الشكل رقم (03) مثال لبند يحتوي على الأداء التفاضلي للبند غير المنتظم.

7- خطوات إجراء تحليل الأداء التفاضلي للبند:

للكشف عن الأداء التفاضلي للبند، يفترض بأن مجموعة بنود الاختبار أحادية البعد، وهناك مجموعة من الاختبارات تقيس أبعاد متعددة تكون على شكل عوامل كامنة تؤثر في إجابة الأفراد على بنود معينة، فأول خطوة عملية للكشف عن الأداء التفاضلي هي فحص الأبعاد الكامنة لمجموعة من البنود، على الأقل نقوم بإجراء التحليل العاملي الاستكشافي لكل مجموعة لوحدها، وينبغي أن تظهر كل مجموعة نفس العدد من العوامل الكامنة، وهذا يعني ضرورة تكافؤ البنية العاملية للاختبار عند كل المجموعات حتى نشرع في إجراء تحليل الأداء التفاضلي للبند إحصائيا.

الخطوة الأولى: فحص مدى توفر أحادية البعد لكل مجموعة، وتحديد فيما إذا كان هناك نفس العدد من العوامل الكامنة.

الخطوة الثانية: اختيار الطريقة المناسبة للتحليل وهي:

أ- الطرق القائمة على نظرية الاستجابة للمفردة عند توفر افتراض أحادية البعد، أو الطرق البرمترية الأخرى.

ب- الطرق اللابرمترية، هذه الطرق ليس لها القدرة على الكشف على النوع غير المنتظم من الأداء التفاضلي.

الخطوة الثالثة:

أ-إجراء التحليل باستخدام إحدى الطرق الإحصائية البرمترية أواللابرمترية.

ب- القيام بعملية التنقية،فبعد التحليل الأول الذي يوصلنا إلى حذف البنود التي تحتوي على الأداء التفاضلي ثم حساب الدرجة الكلية لكل فرد انطلاقا من البنود التي لا تحتوي على الأداء التفاضلي.

الخطوة الرابعة: بعد مطابقة المفحوصين حسب الدرجة الكلية الجديدة نقوم بإعادة التحليل للكشف عن الأداء التفاضلي للبنود.

الخطوة الخامسة: وفق كل طريقة مستخدمة نقوم باستخدام كل من الدلالة الإحصائية، وحجم الأثر أو حجم الأداء التفاضلي للبند لتحديد القيم الكبيرة والبنود التي تحتوي فعلا على الأداء التفاضلي.

الخطوة السادسة: فحص تأثير الأداء التفاضلي للبند من حيث اختلافات المجموعة في الدرجة الكلية المتوقعة وفي متوسطات العامل الكامن مع أو من دون البنود التي تحتوي على الأداء التفاضلي.

8- طرق الكشف عن الأداء التفاضلي للبند:

حسب نورتقيت (Noortgate, 2005)[15] غالبا ما نستخدم تحليل الأداء التفاضلي للبند للتحقق مما إذا كان بند معين لا يفضل المجموعة المرجعية أو الأغلبية (مثل الذكور بمقابل الإناث، أو البيض بمقابل السود، المتعلمين بمقابل الأميين، الأصحاء بمقابل المرضى، الجزائريين مقابل الفرنسيين) ويذكر نوعين من الطرق للكشف عن احتمال تحيز الاختبار وهي الطرق التحكيمية التي تعتمد على المحكمين، والطرق الإحصائية التي تعتمد على الأداء التفاضلي للبند.وكل من الطرق التحكيمية والإحصائية لهما أهمية في التحقق من صلاحية البنود ولا يمكن لطريقة أن تكون في مكان الأخرى في عملية التحليل.

8-1 الطرق التحكيمية:

تعتمد الطرق التحكيمية على رأي المتخصصين وخبرتهم للكشف عن البنود التي تحتوي على عبارات غير مناسبة أو غير عادلة للمجموعات المختلفة مثل الجنس (بين الذكور

والإناث)، ويمكن مراجعة المحتوى بواسطة المحكمين لاكتشاف تحيز البند المحتمل.فاللجوء إلى الخبراء (المحكمين) لفحص محتوى البنود يكون قبل إجراء الطرق الإحصائية للكشف عن الأداء التفاضلي للبند. وبعد القيام بالتحليل لتوليد الفرضيات حول المصادر المحتملة لهذا الأداء التفاضلي للبند. إلا أن هناك حيرة للباحثين بسبب صعوبة أو استحالة تقديم تفسيرات مناسبة لهذه المصادر، إلا أن هذا الإجراء يقدم تفسيرا بعديا يجب التحقق منه بواسطة دراسات إحصائية لاحقة.

8-2 الطرق الإحصائية :

يشير كلوسر (Clauser, 1998)[16] الى أن هناك العشرات من الطرق الإحصائية لتحليل الأداء التفاضلي للبند التي تم ذكرها في أدبيات القياس النفسي والتربوي، واعتمادا على القوة النظرية ونتائج العديد من الدراسات المقارنة هناك عدد قليل من هذه الطرق الذي يعد مفضلا من طرف الباحثين. كل هذه الطرق توفر المقارنة لأداء بند تحت الدراسة بعد مطابقة (Matching) المفحوصين على السمة المقاسة. وفي جميع الطرق الإحصائية للكشف عن الأداء التفاضلي للبند يجب تحديد المجموعات أولا على حسب كل متغير (مثل السن أكبر من 60 وأقل من 60 ، أو الجنس ذكور وإناث، أو المهنة عامل طبيب). ومن الإجرائي تحديد مجموعة معينة كمجموعة مرجعية والأخرى كمجموعة مركزية ثم نطابق الأفراد في كل مجموعة حسب محك المطابقة الذي كثيرا ما يكون الدرجة الكلية للاختبار وهي الدرجة الملاحظة، أو مستويات السمة الكامنة وهي الدرجة الكامنة.ولقد أشارت ويبرق(Wiberg, 2007)[17] أن استخدام الدرجة الملاحظة كمحك مطابقة تعتمد على الافتراض القوي بأن السمة الكامنة هي ممثلة جيدا بواسطة هذه الدرجة الملاحظة، وهناك أدلة تظهر أن الدرجة الكامنة هي قياس أدق للسمة الكامنة عند المفحوصين ويمكن التخلص من مشكل استخدام الدرجة الملاحظة بحذف البنود المتحيزة وإعادة التحليل مرة أخرى وهذا ما يسمى بعملية التنقية (Purification) وهي عملية تكرارية للوصول إلى البنود غير المتحيزة.

9- تصنيف طرق الكشف عن الأداء التفاضلي للبند :

تضع ويبرق (Wiberg, 2007)[18] تصنيف لأشهر الطرق (14 طريقة) حسب طبيعتها، نكتفي في هذا المقال بذكر أربعة طرق.

الجدول رقم: (01) تصنيف طرق الكشف عن الأداء التفاضلي للبند حسب طبيعتها

الطريقة	برمترية/لابرمترية	د.ك / س.ك	نوع البنود	خ / ق DIF	م / غ.م
منتل-هينزل	لابرمترية	د.ك	ثنائي ورتبي	خ / ق	م
اختبار تحيز البند المتزامن	لابرمترية	س.ك	ثنائي ورتبي	خ / ق	م / غ.م
الإنحدار اللوجيستي	برمترية	د.ك	ثنائي ورتبي	خ / ق	م / غ.م
نسبة الأرجحية- نظرية الاستجابة للمفردة	برمترية	س.ك	ثنائي ورتبي	خ / ق	م / غ.م

1- د.ك / س.ك: الدرجة الكلية أو السمة الكامنة. 2- خ / قDIF : إمكانية تطبيق اختبار الدلالة الإحصائية و/أو قياس حجم DIF. 3- م / غ.م : إمكانية الكشف عن النوع المنتظم وغير المنتظم.

هذا التصنيف يقسم الطرق إلى طرق برامترية، وهي طرق تفترض نموذجا إحصائيا محددا (مثل نموذج الانحدار) وتقوم بفحص الأداء التفاضلي للبند من حيث تقديرات البرامترات المحددة للنموذج ويمكن تصورها كالتنبؤ بالاستجابة على البند من الانتماء لمجموعة معينة. والطرق اللابرمترية لا تفترض وجود نموذج إحصائي محدد. وعليه فالطرق البرمترية أكثر قوة في الكشف عن الأداء التفاضلي للبند من الطرق اللابرمترية، لكن تتطلب افتراضات أكثر يجب توفرها للحصول على نتائج دقيقة. كما أن هناك بعض الطرق تعتمد على الدرجة الكلية للاختبار لمطابقة المفحوصين في المجموعة المرجعية والمركزية، وطرق أخرى تعتمد على المتغير الكامن كمحك للمطابقة، والطرق التي تستخدم مع البنود الثنائية والرتبية، ومن الطرق التي يعتمد على النظرية الكلاسيكية للقياس باستخدام الدرجة الكلية للاختبار كمحك مطابقة، وأخرى تعتمد على النظرية الحديثة في القياس أو النمذجة بالمعادلة البنائية (Structural Equation Modeling) لاستخدامها السمة الكامنة كمحك مطابقة للمفحوصين في المجموعات للقيام بالتحاليل المطلوبة.

وأخيرا عندما نتحدث عن اختيار الطريقة الأفضل، يمكن القول بأن الطرق التي تقوم على النظرية الحديثة للقياس (نظرية الاستجابة للمفردة) تعد عالية الكفاءة في الكشف عن الأداء التفاضلي للبند سواء للبنود الثنائية أو الرتبية، بالإضافة إلى القدرة على الكشف عن

النوعين المنتظم وغير المنتظم. لكن هذا بتوفر الافتراضات الصارمة لنماذج نظرية الاستجابة للمفردة. فعدم توفر هذه الافتراضات،مثل أحادية البعد، وإعتدالية التوزيع وخطية العلاقات، مع وجود متغير كامن واحد يفسر أكبر نسبة من التباين، وتوفر حجم عينة كبير جدا سيؤدي ذلك إلى نتائج غير دقيقة.فهناك بعض نماذج نظرية الاستجابة للمفردة، مثل ثلاثية البرامتر تتطلب عينة أكبر من 1000 لكل مجموعة وهذا عددا كبيرا مقارنة بالطرق الأخرى التي تتطلب عينة في حدود(200) فرد لكل مجموعة. وبالتالي يفضل استخدام الطرق الأخرى وخاصة طريقة الانحدار اللوجستي التي تعد من أحسن الطرق للتحقق من تحيز بنود الاختبارات.

10- تفسير الأداء التفاضلي للبند:

هناك العديد من الدراسات التي استخدمت الطرق المختلفة للكشف عن الأداء التفاضلي للبند باستخدام متغيرات مختلفة لتكوين المجموعة المرجعية والمجموعة المركزية مثل الجنس (Gierl, 1999; Francisco, 2004; Lyons-Thomas, 2014)، واللغة (Budgell,1995 ;Orlando, 2002; Kwakkenbos, 2013)، وكذلك العرق، كما أن بعض الدراسات توصلت إلى نتائج تشير إلى تأثير عوامل أخرى الأداء التفاضلي حسب (Sideridis, G and Tsaousis, I, 2013)[19] حيث اشتملت على تحليل عدة متغيرات في آن واحد. إلا أن العديد من الدراسات لم تذكر مصادر هذا الأداء التفاضلي للبند وأسباب وجوده في البنود، إلا بعض الدراسات مثل دراسة روسوس (Roussos,1996)[20] التي تناولت اللغة تحدثت عن حدوث الأداء التفاضلي للبند بسبب مشاكل لغوية ناجمة عن سوء الترجمة، والاختلاف في الكلمات والتعابير حسب اللغة والثقافة. حيث يمكن أن يحدث تحيز البند بسبب تغير المعنى بالترجمة. وقدم مثال عن هذا المشكل، ففي اختبار مترجم من اللغة الانجليزية إلى السويدية كان هناك بند على هذا الشكل:

في أي الأماكن الأكثر احتمالا أن يعيش طائر بقدم مكفف (webbedfeet).

أ- في الجبال. ب- في الغابة .ج- في البحر .ء- في الصحراء

إن عبارة (webbedfeet) عندما تم ترجمتها من اللغة الانجليزية إلى اللغة السويدية أصبحت تعني قدم السباحة (Swimmingfeet)، وبالتالي أصبحت هذه الترجمة تعطي تلميحا صريحا بالمكان الذي يعيش فيه هذا الطائر، وهذه يعطي الأفضلية للمفحوصين السويديين للإجابة على هذا السؤال. وبالتالي الاختلافات بين المجموعتين لا ترجع إلى اختلاف حقيقي في القدرة

وإنما إلى تحيز هذا البند بإعطائه الأفضلية لمجموعة المستجيبين الذين يتحدثون السويدية على حساب الذين يتحدثون الإنجليزية، وبالتالي يصبح هذا البند غير عادل ومتحيز ثقافيا.

قد يساعد تحليل الأداء التفاضلي للبند على تحسين نوعية الاختبار، ويزيد من عدالته (fairness). والسؤال الذي نطرحه هنا هو هل وجود الأداء التفاضلي في بند ما يعني أن البند غير عادل (Unfair)؟ يجيب بون (Boone, 2014. P 275)[21] بنعم ولا في آن واحد، فمن منظور القياس النفسي هذا يعني أن البند يقيس السمة الكامنة بطريقة مختلفة عند المجموعات المقارن بينها، غير أن هذا لا يعني بالضرورة بأن البند غير عادل للمجموعات الفرعية المختلفة من المفحوصين (كأن يكون البند أكثر سهولة للفهم من طرف مجموعة على حساب أخرى). وبالإضافة إلى ذلك فإن احتواء بند على الأداء التفاضلي لا يعني بالضرورة أنه بند غير مناسب، لذلك علينا محاولة معرفة مصدر التحيز.غير أن هناك تطور بسيط في مجال تحديد المصادر والأسباب الموضوعية، أو الخصائص الموضوعية المميزة للبنود التي تحتوي على الأداء التفاضلي. فمن الصعب اكتشاف أسباب هذا الأداء المختلف لبنود الاختبار، أو النقص المشترك بين البنود التي تحتوي على الأداء التفاضلي. فالأسباب الموضوعية المحتملة لحدوث الأداء التفاضلي للبند تبقى غير معروفة إلى حد كبير.ومن بين الأسباب المحتملة هو أن البند يقيس سمة كامنة غير التي يقيسها الاختبار.غير أن ليون وآخرون (Lyons et all, 2014)[22]قدموا بعض الأسباب الأخرى فبالاظافة إلى سوء الترجمة (في حالة ترجمة الاختبار من لغة أخرى) وهو الحاصل في جل الاختبارات غير المكيفة المتداولة في الممارسة السيكولوجية في الجزائر. ذكروا الأسباب العشوائية حيث يحدث الأداء التفاضلي للبند بمحض الصدفة، كما يمكن أن يكون هذا الأداء التفاضلي للبند زائف (Pseudo-DIF) خاصة عند استخدام الدرجة الكلية للاختبار كمحك مطابقة، وبما أن الدرجة الكلية هي مجموع درجات البنود، فإن اختبار الأداء التفاضلي للبند لا يكون مستقلا عنها، ويصبح زائفا وهذا ما يوضح بأن الدرجة الكلية هي غير مثالية للاستخدام كمحك مطابقة، غير أنها كثيرا ما تكون أحسن اختيار لعدم توفر البديل، وخاصة إذا كان الاختبار يحتوي على عدة كبير من البنود فمن الممكن تحديد مجموعة من البنود الجوهرية التي لا تحتوي على الأداء التفاضلي ونستخدم الدرجة الكلية التي نحصل عليها كمحك مطابقة لأفراد المجموعتين.

لقد قام هوانق (Huang, 2012)[23]بذكر بعض الدراسات التي حاولت إعطاء تفسيرا للأسباب الأكثر شيوعا لحدوث الأداء التفاضلي، منها ألفة المفحوص بمحتوى البنود،

وخبرته في مجال السمة، والثقافة، ومدى اهتمام المفحوصين بمحتوى الاختبار وردود فعلهم السلبية، والغضب، والاشمئزاز والخوف من محتوى الاختبار، جميع هذه الأسباب أعطيت لتفسير سبب حدوث الأداء التفاضلي في بنود الاختبار. غير أنه هناك القليل من الأدلة الإمبريقية التي تدعم هذه الافتراضات المقدمة حول الاختلاف بين المجموعات في الأداء على نفس البند، كما نشير إلى عدم توفر إطارا نظريا قويا لتفسير سبب حدوث الأداء التفاضلي للبنود. إلا أنه من الضروري اللجوء إلى التحليل التحكيمي للكشف عن هذه الأسباب الموضوعية. وفي الأخير نشير إلى أن تطوير طرق موضوعية لتحديد أسباب ومصادر الأداء التفاضلي للبند يعتبر من أكبر التحديات التي تواجه العاملين في ميدان القياس النفسي وبناء الاختبارات.

11- كيفية التعامل مع البنود التي تحتوي على الأداء التفاضلي:

يطرح تساؤلا مهما وهو ماذا يفعل الباحث في حالة توصله إلى وجود بنود تحتوي على الأداء التفاضلي؟ فهل يقوم بحذفها كلية من الاختبار؟ وهذا إجراء لاينصح به من طرف خبراء القياس، أم يقوم الباحث بالتحفظ عليه إلى غاية التحقق منه عن طريق المحكمين وعمل المزيد من التحاليل لمعرفة مدى صلاحيته. إن جزءا من الإجابة على هذه التساؤلات يتعلق بمدى أهمية القرارات التي نتخذها بشأن الأفراد وفق نتائج تطبيق هذا الاختبار الذي يتضمن بنودا متحيزة، أو مدى خطورة النتائج نفسيا واجتماعيا جراء اتخاذ قرار خاطئ، غير أن بعض الباحثين(Pastor and Richarde2003)[24] يتحدثون عن إمكانية إعادة صياغة البنود التي تحتوي على الأداء التفاضلي للبند، أو حذف تلك التي يكون حجمها كبير جدا، لكن قبل الحذف يجب الأخذ بعين الاعتبار تأثير هذا الحذف على صدق محتوى للاختبار، والفقدان المحتمل للمعلومات من قياس مستويات مختلفة من السمة الكامنة أو بعض مؤشراتها.على العموم حذف هذه البنود قد يكون غير مقبول مع الاختبارات التي لا تحتوي على عدد كبير من البنود.فتأثير وجود الأداء التفاضلي في بنود اختبار ما يختلف حسب حجم الأداء التفاضلي، وعدد البنود التي تحتوي عليه في الاختبار، والغرض من استخدام الاختبار وأهمية القرارات المترتبة عن نتائج الاختبار.

خلاصة:

يعتبر الأداء التفاضلي للبند من أهم الأدلة التي يجب على الباحث جمعها لتدعيم صلاحية درجات الاختبار الذي يسعى إلى استخدامه للتأويل الموضوعي للنتائج، ولكي تزيد ثقته في أن الفروق في الدرجات الملاحظة عند العينة يراجع إلى اختلافات حقيقية في السمة، وليس بسبب انتماء المفحوصين إلى مجموعات مختلفة (حسب الجنس، أو العمر، أو العرق، أو اللغة أو المهنة أو المستوى العلمي وغيرها)، وبالتالي يستطيع المقارنة بين هذه المجموعات الفرعية بكل عدل، ويعتبر الأداء التفاضلي للبند من الأخطاء المنتظمة التي تهدد صدق درجات الاختبار، فهناك الأداء المنتظم حيث أن البند يفضل إما المجموعة المرجعية أو المركزية في جميع مستويات السمة المقاسة، وهناك الأداء غير المنتظم حيث يفضل البند إحدى المجموعتين مثلا في المستوى المنخفض من السمة ويفضل الأخرى في المستوى المرتفع. وللكشف عن هذا الأداء التفاضلي هناك طرق تحكيمية وأخرى إحصائية، منها ما يعتمد على النظرية الكلاسيكية للقياس مثل طريقة الانحدار اللوجستي، ومنها ما يعتمد على النظرية الحديثة مثل طريقة نسبة الأرجحية التي تعتمد على السمة الكامنة كمحك للمطابقة، وأحدث الطرق هي تلك التي تعتمد على النمذجة بالمعادلة البنائية.

الهوامش:

1- **Chen, W-H :**(2014). Differences in Well-Being of Older Immigrants in Australia. In A.C. Michalos (ed.). Encyclopedia of Quality of Life and Well-Being Research. (PP 1611-1613). Springer Science+Business Media. DOI 10.1007/978-94-007-0753-5.

2- **Hambleton, R.K, Swaminathan, H and Rogers, D.J**: (1991). Fundamentals of Item Response Theory. Sage Publications.

3- **Clauser, B.E:** (1998). Using Statistical Procedures to Identify Differentially Functioning Test Items. National Council on Measurement in Education. [http://ncme.org/publications/items/].

4- **Budgell, G.R, Raju, N.S and Quartetti, D.A**: (1995). Analysis of Differential Item Functioning in Translated Assessment Instruments. Applied Psychological Measurement. 19, 4. 309-321.

5- **Reckase, M. D**:(1979). Unifactor latent trait models applied to multi-factor tests: Results and implications. Journal of Educational Statistics, 4, 207-230.

6- **Hambleton, R.K**: (2005). Issues, Designs, and Technical Guidelines for Adapting Tests Into Multiple Languages and Cultures. In Hambleton, R.K, Merenda, F.P and Spielberger, D.C (Eds). Adapting Educational and Psychological Tests for Cross-Cultural Assessment. (PP:3-38). LAWRENCE.

7- **Zumbo, B. D**: (1999). A Handbook on the Theory and Methods of Differential Item Functioning (DIF): Logistic Regression Modeling as a Unitary Framework for Binary and Likert-Type (Ordinal) Item Scores. Ottawa, ON: Directorate of Human Resources Research and Evaluation, Department of National Defense.

8- **Holland, P. W and Thayer:** (1988). Differential item functioning and the Mantel-Haenszel procedure. In H. Wainer & H. I. Braun (Eds.), Test validity (pp. 129-145). Hillsdale, NJ: Erlbaum.

9- **Ellis, B.B and Raju, N.S**: (2003). Test and Item Bias: What They Are, What They Aren't, and How to Detect Them. http://eric.ed.gov/?id=ED480042.

10- **Chen,** (2011). Sample Size in Differential Item Functioning: An Application of Hierarchical Linear Modeling. Educational Sciences: Theory & Practice. 11, 1. 284-288.

11- **Zumbo, B. D**: (1999). A Handbook on the Theory and Methods of Differential Item Functioning (DIF): Logistic Regression Modeling as a Unitary Framework for Binary and Likert-Type (Ordinal) Item Scores. Ottawa, ON: Directorate of Human Resources Research and Evaluation, Department of National Defense.

12- **Clauser, B.E**: (2004). Sex differential item functioning in the Raven's Advanced Progressive Matrices: evidence for bias. Personality and Individual Differences 36 1459–1470. Doi: 10.1016/S0191-8869 (03) 00241 -1.

13- **Noortgate, W.V and Boeck, P. D**: (2005). Assessing and Explaining Differential Item Functioning Using Logistic Mixed Models. Journal of Educational and Behavioral Statistics. 30, 4. 443-464.

14- **Clauser, B. E:** (2005). Assessing and Explaining Differential Item Functioning Using Logistic Mixed Models. Journal of Educational and Behavioral Statistics. 30, 4. 443-464.

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

16- **Clauser, B.E**: (1979). Unifactor latent trait models applied to multi-factor tests: Results and implications. Journal of Educational Statistics, 9, 124-58.

17- **Sideridis, g and Tsaousis, I**: (2013). DIF Analysis for Item and Test on the NCA Tests The General Ability Test (GAT) Art Major. National Center for Assessment in Higher Education. TR035-2013.

18- **Roussos, L. A**: (1996). Simulation studies of the effects of small sample size and studied item parameters on SIBTEST and Mantel-Haenszel type I error performance. Journal of Educational Measurement, 33, 215-230.

19- **Boone, W.J, Staver, J.R and Yale, M.S** :(2014). Rasch Analysis in the Human Sciences. Springer.
20- **Teresi, J.A and Fleishman, J.A**: (2007). Differential item functioning and health assessment. Quality of Life Research. 16:33–42.
21- **Huang, J and Han, T:** (2012). Revisiting Differential Item Functioning: Implications for Fairness Investigation. International Journal of Education. 4,2. 74-86.
22- **Fleishman, J.A, Spector, W.D and Altman, B.M**: (2002). Impact of Differential Item Functioning on Age and Gender Differences in Functional Disability. Journal of Gerontology: SOCIAL SCIENCES. 57B, 5. S275-S284.
23- **car, T**: (2011). Sample Size in Differential Item Functioning: An Application of Hierarchical Linear Modeling. Educational Sciences: Theory & Practice. 11, 1. 284-288.
24- **Pastor D.A and Richarde, R.G**: (2003) Investigation of Differential Item Functioning for African-American and Anglo-American College Students in the Modified Achievement Goal Questionnaire. Poster presented at the National Council of Measurement in Education's Graduate Student Poster Session, Chicago, April 2003.