الأداء التفاضلي للبند في الإختبارات النفسية، مصادره وتفسيره

The Differential Item Functioning in tests, sources and explanation

د.بوسالم عبد العزيز- جامعة البليدة٢- الجزائر

أ.كريش أحمد- جامعة البليدة٢ – الجزائر

مخبر القياس والدراسات النفسية

ملخص:

نتناول في هذا المقال دور الأداء التفاضلي للبند في الاختبارات النفسية على إعتباره مفهوما ظهر حديثا في قاموس القياس النفسي وبناء الاختبارات عالميا، كما نوضح مصادر الوقوع في تحيز البنود وتأثيره السلبي على تفسير نتائج مختلف الاختبارات النفسية ومنه القرارات المختلفة، وبالتالي يصبح الإختبار غير صالحا وإحتمال أن يكون متحيزا وهذا ما يهدد صدق الاستدلال بدرجاته ويقلل من الثقة والدقة في التفسيرات المترتبة عليها. وللتعامل مع هذا الموضوع طور الباحثون بعض الطرق الإحصائية للكشف عن خصائص أداء بنود الاختبار وهل يختلف باختلاف إنتماء الأفراد إلى مجموعات معينة، وهو ما أصبح يعرف بالأداء التفاضلي للبند، التي تعد من الطرق الحديثة لتحليل بنود الاختبار، ومن الأدلة المهمة على صدق الاستدلال من درجات الاختبار.

Abstract

This article provides an overview of one of the most recent and sophisticated methods of item analysis is 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.

مقدمة:

عندما نقوم بتطبيق إختبار ما على مجموعات من الأفراد غير متجانسة نتحصل على تباين في الدرجات مما يطرح تساؤلا حول هذا التباين في الدرجات هل هو راجع الى إختلاف في مستويات السمة المقاسة التي يقيسها الإختبارعند 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 الى مجموعة معينة مثل الجنس (ذكور وإناث)، أو العمر (أكبر من ٦٠ سنة وأصغر من ٦٠ سنة) ،أو اللغة (العربية والإنجليزية)....وبالتالي يصبح الإختبار غير عادلا وإحتمال 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 إنتماء 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.

١- مفهوم الأداء التفاضلي للبند:

يعرف شن (Chen, 2014 P 1611) الأداء التفاضلي للبند بأنه يشير الى الحالة أين يكون أفراد من مجموعات مختلفة (العمر، والجنس، والعرق، والتعليم، الثقافة) على نفس المستوى من السمة الكامنة، لهم احتمال مختلف لإعطاء إستجابة معينة على بند معين.

تذكر مالر (Maller, 2003) بأن مصطلح الأداء التفاضلي للبند يشير إلى الاختلافات في الخصائص الإحصائية (Statistical Properties) لبند ما بين مجموعات من المفحوصين لهم قدرة متساوية. غالبا ما تكون المجموعات هي مجموعة مرجعية (Reference group) التي تحتوي على الأغلبية،

ومجموعة مركزية (Focal group) التي تحتوي على الأقلية أو الفئة المدروسة.

٢- الأداء التفاضلي للبند والبنية العاملية للاختبار:

تذكر تيريزي (Teresi, 2008) 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، فاذا لم يتوفر هذا الافتراض فيمكن أن يختلف أداء البند أساسا بسبب اختلاف البنية العاملية للاختبار عبر المجموعات وليس بسبب إحتواء البند على الأداء التفاضلي للبند. وهذا ما أشار إليه كلوسر (Clauser, 1998) 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 الإختلاف موجود أولا وأساسا بين المجموعتين في الاختبار ككل ولا داعي للبحث عنه بواسطة تحليل الأداء التفاضلي للبند. لأن اختلاف الكل يعني إختلاف الجزء. يشير بودجال (Budgell, 1995) إلى ضرورة التحقق من أحادية البعد للإختبار، لتحديد مدى تكوين البنود لمجموعات أحادية البعد، وذلك للقيام بتحليل الأداء التفاضلي للبند، ويمكننا القيام بذلك بواسطة إجراء التحليل العاملي

الاستكشافي. لكي يكون إختبار 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. يضع ريكازي (Reckase, 1979) محك الجذر الكامن لقياس أحادية البعد تعتمد على نسبة التباين على الأقل (٢٠%). وهذا يعني أنه لافتراض أحادية البعد للإختبار يجب أن يكون هناك عامل كامن مسيطر يفسر على الأقل (٢٠%) من التباين الملاحظ.

٣- الأداء التفاضلي للبند وصدق درجات الاختبار:

يعد الأداء التفاضلي للبند من الأخطاء المنتظمة، وهو حسب شن (Chen, 2014) مهدد لصدق درجات الاختبار، لأنه يحدث عندما يكون الأفراد من نفس مستوى السمة المقاسة يجيبون إجابات مختلفة على نفس البند إعتمادا على إنتمائهم لمجموعة معينة، وهذا يعني أن الاجابة على البند تخضع إلى شيئ غير التكوين الفرضي الذي يسعى الاختبار إلى قياسه، فمثلا في إختبار 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 الى متغير الجنس، لأن النساء عادة أكثر بكاءا من الرجال. غير أن كلوسر (Clauser, 1998) يشير الى 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 واحد من أشكال أدلة الصدق التي يجمعها الباحث. ومع النظرية الحديثة للصدق يشير زومبو (Zumbo, 1999) أننا لم نعد نتحدث عن أنواع للصدق حيث نكتفي بنوع واحد، بل الصدق أصبح عملية جمع أكبر عدد من الأدلة على صدق درجات الإختبار، وهو عملية مستمرة، ولقد أصبح من الضروري التأكد من عدم تحيز البنود بواسطة الأداء التفاضلي للبند كشيئ أساسي للتحقق من صدق

درجات الاختبار. هذا يعني أن صدق درجات الاختبار تضمن بأن أداء البند المختلف بين المجموعات هو راجع حقيقة لمشكل حقيقي في البند، مما يعني وجود أخطاء منتظمة هي التي تؤدي الى هذه الاختلافات في أداء البند مع المجموعات المختلفة.

٤- الأداء التفاضلي للبند وتحيز البند:

يذكر الباحثون (Chen, 2014, Holland, 1988) بأن بداية الاهتمام بمفهوم الأداء التفاضلي للبند كانت منذ الستينات من القرن الماضي في مجال الإختبارات 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 بغرض الإختبار(قياس السمة الكامنة). تذكر إليس (Ellis, 2003) أنه في أواخر الثمانينات من القرن الماضي حل مصطلح الأداء التفاضلي للبند محل تحيز البند وأسباب هذا التغيير من المحتمل ما يكون بسبب السياسة واللغة أكثر من القياس النفسي، لأن مصطلح تحيز البند يحمل معه دلالة سلبية وهو يرتبط عموما مع مفهوم غير عادل، والممارسات التمييزية للاختبار أكثر منه تعريف يرتبط بمصطلحات القياس النفسي، وبذلك أحس علماء القياس النفسي أنه من المفيد التفريق بين المصطلحات التقنية للقياس النفسي عن تلك المصطلحات المشحونة سياسيا واجتماعيا، وبذلك تم تغيير المصطلح. وكما يذكر هوانق (Huang, 2012) أن تحليل الأداء التفاضلي للبند يعطي إمكانية فحص الاختلافات بين المجموعات في أداء الاختبار والكشف عن ما إذا كانت هذه الاختلافات تشير الى تحيز البند أو لا. ولقد أشار زومبو (Zumbo, 1999) الى أن بنود الاختبار تكون متحيزة بسبب أنها تحتوي على مصادر صعوبة ليست ذات صلة، أو تكون دخيلة على التكوين الفرضي التي تهدف البنود لقياسه،

وهذا ما يطلق عليه تحيز الاختبار (Test Bias) الذي كان موضوع بحوث كثيرة، وأصبح موضوع بحث تقنية حديثة هي الأداء التفاضلي للبند، حيث أصبحت المعيار الجديد لتحليل تحيز الاختبارات. وحسب شن (Chen, 2014) ظهر مصطلح الأداء التفاضلي للبند بعدما تم فهم أن الأداء المختلف للبند لا يظهر دائما بسبب وضعية متحيزة (تفضيل مجموعة على مجموعة)، بل يكون أداء البند مختلفا اعتمادا على الانتماء لمجموعة معينة. ولذلك أشار الباحثين (Hambleton, 1991; Sireci (2011); Zumbo, 1999) أن الأداء التفاضلي للبند يعتبر ضروري للدلالة على تحيز البند، لكن يعتبر غير كافي للجزم بذلك، فعند القيام بالتحليل ولم تحتوي أي البنود على الأداء التفاضلي للبند دل هذا على عدم وجود تحيز للبند، أما في حالة احتواء أحد البنود على الأداء التفاضلي للبند فهذا يعتبر غير كافي 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، والإستعانة بالمتخصصين (الخبراء) في ذلك. يعطي كلوسر (Clauser, 1998) مثالا عن اختبار في الرياضيات يتطلب مهارة القيام بالعمليات الحسابية ومهارة القراءة بفهم، فاذا تم المطابقة بين المفحوصين في المهارة الأولى وعدم تكافئهم في المهارة الثانية، في هذه الحالة هناك احتمال أن يكون التباين في الأداء راجع الى الأداء التفاضلي للبند ومع ذلك يمكن أن لا يكون البند متحيزا، إن الحكم على البند هل هو متحيز أم لا في هذه الحالة يعتمد على هل مهارة القراءة بفهم تعتبر ذات صلة بغرض الاختبار، وبالتالي الأداء التفاضلي للبند يعتبر ضروري لكن غير كافي كدليل لتحيز البند. ويضيف كلوسر في نفس المرجع على أهمية التفريق كذلك بين تحيز البند والأداء التفاضلي للبند من جهة، والمحتوى غير المناسب للبند من جهة أخرى، أو التي نحكم عليها أنها تحتوي على لغة مزعجة وجارحة، حيث أن في هذا النوع من البنود لا يضمن تحليل

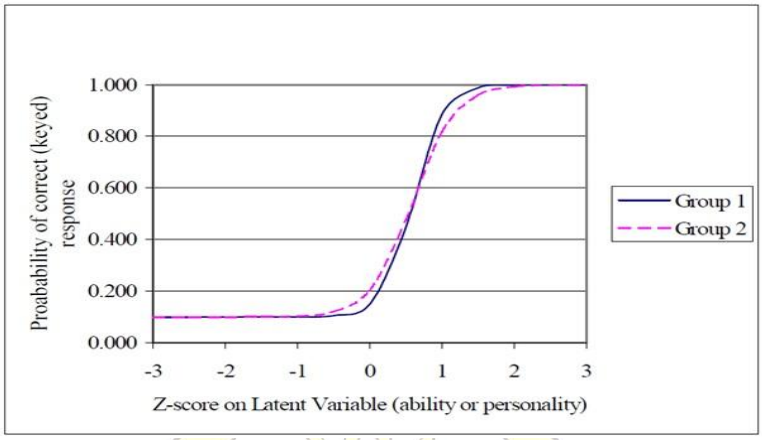
الأداء التفاضلي للبند مناسبة محتواها، وهذا راجع إلى الحقيقة البسيطة التي مفادها بأن محتوى البند المزعج يؤثر في المجموعات على حد سواء ولا تتأثر مجموعة فقط لوحدها على حساب الأخرى.

٥- تحيز البند وتأثير البند:

يذكر كلوسر (Clauser, 1998) أنه في بعض الظروف المفحوصين في مختلف المجموعات قد يختلفون حقيقة في القدرة، ففي هذه الحالة الاختلافات في الأداء تكون متوقعة، وهذه النتيجة ترجع الى تأثير البند وليس لتحيز البند. كما يؤكد زومبو (Zumbo, 1999) أن تأثير البند يشير الى اختلاف حقيقي بين الأفراد راجع الى الاختلاف في مستويات السمة الكامنة وليس بسبب الانتماء الى مجموعة معينة. كما يشير بعض الباحثين (Acar, 2011; Gierl, 1999; Huang, 2012) الى أن احتواء بند ما على الأداء التفاضلي للبند يرجع الى سببين هو إما إلى تحيز البند أو إلى تأثير البند، والحكم في هذه الحالة يرجع الى المتخصصين. لكن هولند (Holland, 1988) يوضح أن اختلاف أداء البند يشير الى تأثير البند في حالة عدم مطابقة (Unmatched) المفحوصين على محك المطابقة. مما يجعل هذا الاختلاف راجع الى اختلاف الأفراد في السمة الكامنة.

٦- أنواع الأداء التفاضلي للبند:

يذكر زومبو (Zumbo, 1999) أنه من الناحية النظرية يقاس الأداء التفاضلي للبند بواسطة مقارنة منحنى خصائص البند (Item Characteristic Curve ICC). الذي هو حسب همبلتون (Hambleton, 1991) يمثل العلاقة بين أدء المفحوص على البند ومتصل السمة الكامنة لمختلف المجموعات على بند معين. فإذا كان منحنى خصائص البند لكل مجموعة مثل (ذكور وإناث) متطابق تماما، أو قريب جدا من التطابق، يمكننا القول بأن هذا البند لا يحتوي على الأداء التفاضلي للبند أنظر الشكل رقم (٠١).



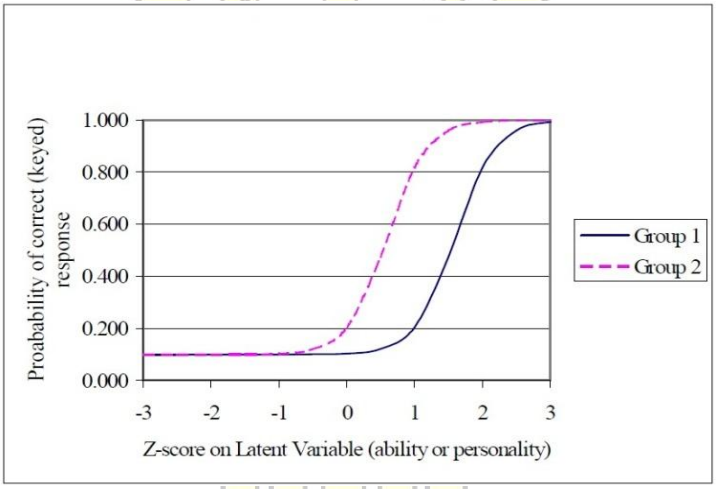
الشكل رقم (٠١) مثال لبند لا يحتوي على الأداء التفاضلي للبند.
(Zumbo, 1999)

وحسب تريزي (Teresi, 2007) فإن منحنى خصائص البند يمكن تمييزه بواسطة برمترين، البرمتر الأول هو الصعوبة (b) أو العتبة (تأييد البند)، وهذا البرمتر يعكس وضعية المنحنى على المحور س (انتقال المنحنى الى اليمين أو اليسار) أين احتمال الاجابة الصحيحة على البند أو تأييده تكون (٠.٥)، والبرمتر الثاني هو التمييز (a) ويعكس إنحدار المنحنى، فالمنحنيات المنبسطة والمسطحة حيث يكون المنحدر قليل، يدل على نقص قدرة البند لتمييز الأفراد على المستويات المختلفة للسمة الكامنة. فإذا كان هناك اختلاف واضح لمنحنى خصائص البند لكل مجموعة يمكن القول بأن هذا البند يحتوي على الأداء التفاضلي للبند. وهناك نوعين هما:

٦-١ الأداء التفاضلي للبند المنتظم (Uniform DIF):

تشير تيريزي وآخرون (Teresi, et al, 2009) أن الأداء التفاضلي للبند المنتظم يحدث إذا كان إحتمال الإجابة متسق، حيث يكون أعلى باستمرار أو أقل باستمرار في واحدة من المجموعات التي نقارنها على جميع مستويات السمة. كما يذكر الباحثين (Clauser, 1998; Ong, 2015) أن النوع المنتظم

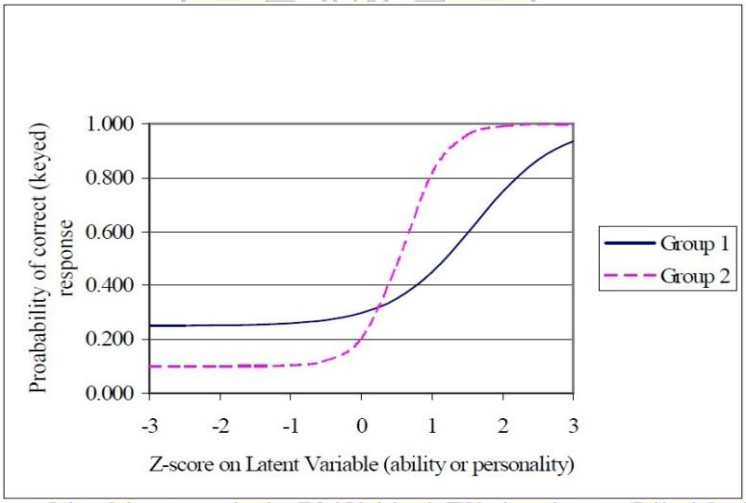
يحدث عندما تكون أحد المجموعتين المرجعية أو المركزية مفضلة عبر كل السمة المقاسة، وهذا التفضيل هو منتظم تقريبا عبر جميع نقاط الإختبار. كما يشير جياغل (Gierl, 1999) أن في النوع المنتظم إحتمال الإجابة الصحيحة على البند يكون أكبر لمجموعة معينة بانتظام على جميع مستويات القدرة، وفي مصطلحات نظرية الاستجابة للبند الأداء التفاضلي للبند المنتظم يعني توازي منحنيات خصائص البند أنظر الشكل رقم (٠٢). كما يذكر سيريديس (Sideridis, 2013) أنه في هذه الحالة المنحنيين يكونان متوازيين للمجموعتين لكن يشيران الى اختلاف في مستوى صعوبة البند (تأييد البند) بين المجموعتين.



الشكل رقم (٠٢) مثال لبند يحتوي على الأداء التفاضلي للبند المنتظم.
(Zumbo, 1999)

٢-٦ الأداء التفاضلي للبند غير المنتظم (Nonuniform DIF): تذكر تيريزي وآخرون (Teresi, et al, 2009) أن الأداء التفاضلي للبند غير المنتظم يحدث عندما يكون احتمال الاجابة في إتجاهات مختلفة في المجموعات التي نقارنها على مستويات مختلفة من السمة المقاسة، مثل احتمال الاستجابة يكون عالي عند الاناث من الذكور في المستويات العليا في الاكتئاب ويكون

منخفض عند الاناث من الذكور في المستويات المنخفضة من الاكتئاب. كما يشير كلوسر (Clauser, 1998) الى أن النوع غير المنتظم يرجع في حالة ما يكون تفضيل أحد المجموعتين المرجعية أو المركزية يزيد أو ينقص أو يتحول في المستويات الأخرى. كما يذكر جياغل (Gierl, 1999) أن في مصطلحات نظرية الاستجابة للمفردة الأداء التفاضلي للبند غير المنتظم يعني عدم توازي منحنيات خصائص البند، أي أن المنحنيين يتقاطعان. أنظر الشكل رقم (٠٣). ويذكر الباحثين (Ong, 2015; Herrera, 2008; Sideridis, 2013; Teresi, 2007) أنه في هذه الحالة المنحنيين يتقاطعان وهذا يدل على اختلاف في الانحدار (Slope) للمنحيين، مما يشير الى وجود اختلاف في القدرة التمييزية للبند بين المجموعتين، وفي هذا النوع من الأداء التفاضلي للبند هناك تفاعل بين مستويات القدرة والمجموعة.



الشكل رقم (٠٣) مثال لبند يحتوي على الأداء التفاضلي للبند غير المنتظم. (Zumbo, 1999)

٧- خطوات إجراء تحليل الأداء التفاضلي للبند:

تذكر تريزي (Teresi, 2007) أن في الكشف عن الأداء التفاضلي للبند، عادة يفترض بأن مجموعة البنود هي أحادية البعد ، والعديد من الاختبارات

تقيس عدة أبعاد التي تكون على شكل عوامل كامنة تؤثر في الإجابة على بنود معينة. إن أول خطوة للكشف عن الأداء التفاضلي للبند وهو فحص الأبعاد الكامنة لمجموعة من البنود، على الأقل نقوم بإجراء التحليل العاملي الإستكشافي لكل مجموعة على حدى وينبغي أن تظهر كل مجموعة نفس العدد من العوامل الكامنة، وهذا يعني ضرورة تكافؤ البنية العاملية للاختبار عند كل المجموعات. وفي مايلي الخطوات التي ينبغي اتباعها:

الخطوة الأولى: فحص أحادية البعد لكل مجموعة وتحديد فيما إذا كان هناك نفس العدد من العوامل الكامنة عند كل مجموعة.

الخطوة الثانية: إختيار الباحث لطريقة التحليل وهي:

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

ب-الطرق اللابرمترية، لكن يجب الوضع في الذهن أن هذه الطرق ليس لها القدرة على الكشف على النوع غير المنتظم من الأداء التفاضلي للبند.

الخطوة الثالثة: أ- إجراء التحليل باستخدام إحدى الطرق الاحصائية البرمترية أو اللابرمترية.

ب- القيام بعملية التنقية بعد التحليل الأول بحذف البنود التي تحتوي على الأداء التفاضلي للبند ثم حساب الدرجة الكلية من البنود التي لا تحتوي على الأداء التفاضلي للبند، تعتبر هذه العملية تكرارية.

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

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

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

٨- طرق الكشف عن الأداء التفاضلي للبند:

حسب نورتقيت (Noortgate, 2005) غالبا ما نستخدم تحليل الأداء التفاضلي للبند للتحقق مما إذا كان بند ما في إختبار لا يفضل المجموعة المرجعية أو الأغلبية (مثل الذكور ، أو البيض) بالمقارنة مع واحدة أو أكثر من المجموعات المركزية أو الأقلية (مثل الاناث، أو الهنود الحمر). يذكر بعض الباحثين (Wiberg, 2007; Zumbo, 1999) أن هناك نوعين من الطرق للكشف عن إحتمال تحيز الاختبار وهي الطرق التحكيمية التي تعتمد على المحكمين، والطرق الاحصائية التي تعتمد على الأداء التفاضلي للبند. كما يذكر كلوسر (Clauser, 1998) بأن الطرق التحكيمية والاحصائية يعتبر كليهما مهما ولا تستطيع طريقة أن تأخذ مكان الأخرى.

٨-١ 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)، ولقد تحدث جياغل (Gierl, 1999) عن مراجعة المحتوى بواسطة المحكمين لإكتشاف والقضاء على تحيز البند المحتمل. كما تشير إليس (Ellis, 2003) إلى أن اللجوء إلى الخبراء (المحكمين) لفحص محتوى البنود يكون قبل إجراء الطرق الاحصائية للكشف عن الأداء التفاضلي للبند. وأيضا يتم اللجوء اليهم بعد القيام بالتحليل لتوليد الفرضيات حول المصادر المحتملة لهذا الأداء التفاضلي للبند. إلا أن هناك إستياء للباحثين بسبب صعوبة

أو استحالة تقديم تفسيرات مناسبة لهذه المصادر، إلا أن هذا الإجراء يقدم تفسير بعدي يجب التحقق منه بواسطة دراسات لاحقة.

٨-٢ الطرق الاحصائية:

يشير كلوسر (Clauser, 1998) الى أن هناك العشرات من الطرق الاحصائية لتحليل الأداء التفاضلي للبند التي تم ذكرها في أدبيات القياس النفسي والتربوي، واعتمادا على القوة النظرية ونتائج العديد من الدراسات المقارنة هناك عدد صغير من هذه الطرق يعد مفضلا. كل هذه الطرق توفر المقارنة لأداء بند تحت الدراسة بعد مطابقة (Matching) المفحوصين على السمة المقاسة. كما يذكر شن (Chen, 2014) أن في جميع الطرق الاحصائية للكشف عن الأداء التفاضلي للبند يجب تحديد المجموعات أولا على حسب متغير معين (مثل السن أكبر من ٦٠ وأقل من ٦٠، أو الجنس ذكور وإناث، أو العرق السود والبيض)، ومن المتعارف عليه أن نحدد مجموعة 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) 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 هو بحذف البنود المتحيزة وإعادة التحليل وهذا ما يسمى بعملية التنقية (Purification) وهي عملية تكرارية .

٨- تصنيف طرق الكشف عن الأداء التفاضلي للبند:

تضع ويبرق (Wiberg, 2007) تصنيف لأشهر الطرق (١٤ طريقة) حسب طبيعتها نكتفي بذكر أربعة طرق كما هو موضح في الجدول رقم (٠١).

الجدول رقم: (٠١) 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

١– د.ك / س.ك: الدرجة الكلية أو السمة الكامنة. ٢– خ / قDIF : امكانية نن اختبار الدلالة الاحصائية و/أو قياس حجم DIF. ٣– م / غ.م : امكانية الكشف عن النوع المنتظم وغير المنتظم.

هذا التصنيف يقسم الطرق الى طرق برمترية، وهي حسب تريزي (Teresi, 2007) طرق تفترض نموذج إحصائي محدد (مثل نموذج الإنحدار) وتقوم بفحص الأداء التفاضلي للبند من حيث تقديرات البرمترات المحددة للنموذج ويمكن تصورها كالتنبؤ بالاستجابة على البند من الإنتماء لمجموعة معينة. والطرق اللابرمترية لا تفترض وجود نموذج احصائي محدد. إن الطرق البرمترية قد تكون 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 ما 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) لإستخدامها السمة الكامنة كمحك مطابقة للمفحوصين في المجموعات للقيام بالتحليل.

وأخيرا عندما نتحدث عن إختيار 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 (Hambleton, 1991) الصارمة لنماذج نظرية الاستجابة للمفردة. فعدم توفر هذه الافتراضات (Budgell,1995; Lyons-Thomas, 2014; Pastor, 2003) مثل أحادية البعد، وهو وجود متغير كامن واحد يفسر أكبر نسبة من التباين، وتوفر حجم عينة كبير جدا سيؤدي ذلك إلى نتائج غير دقيقة. يذكر كرامي (Karami, 2012) أن بعض نماذج نظرية الإستجابة للمفردة، مثل ثلاثية البرمتر تتطلب عينة تكون أكبر من ١٠٠٠ لكل مجموعة فهذا عدد كبير مقارنة بالطرق الأخرى التي تتطلب ٢٠٠ فرد لكل مجموعة. وبالتالي يفضل استخدام الطرق الأخرى وخاصة طريقة الانحدار اللوجستي التي تعد من أحسن الطرق حسب بعض الباحثين (Pastor, 2003; Wiberg, 2007; Zumbo, 1999).

٩- أسباب وتفسير 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 (Pastor, 2003; Molander, 2011; Emmert-Aronson, 2014)، كما أن بعض الدراسات (Sideridis, 2013a; 2013b) 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 بيترسون وآخرين (Petersen, et al, 2003) 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. ولقد قدم همبلتون (Hambleton, 2005) مثالا عن هذا المشكل. ففي اختبار مترجم من اللغة الانجليزية الى السويدية كان هناك بند على هذا الشكل:

في أي الأماكن الأكثر إحتمالا يعيش طائر بقدم مكفف (webbed feet)؟

أ- في الجبـــــال

ب-في الغابـــــــة

ج- في البحـــــــر

ء- في الصحراء

إن عبارة (webbed feet) عندما تم ترجمتها من الانجليزية الى السويدية أصبحت تعني قدم السباحة (Swimming feet)، وبالتالي أصبحت هذه الترجمة تعطي تلميحا صريحا بالمكان التي تعيش فيه هذه الطيور، وهذه يعطي الأفضلية للمفحوصين السويديين للإجابة على هذا السؤال. وبالتالي الإختلافات بين المجموعتين لا ترجع إلى إختلاف 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.

يذكر هوانق (Huang, 2012) أن تحليل الأداء التفاضلي للبند قد يساعد على تحسين نوعية الإختبار، 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) بنعم ولا في آن واحد، فمن منظور القياس هذا يعني أن البند يقيس السمة الكامنة بطريقة مختلفة عند المجموعتيين المقارنتين، غير أن هذا لا يعني بالضرورة بأن البند غير عادل للمجموعات الفرعية المختلفة من المفحوصين (كأن يكون البند أكثر سهولة للفهم لمجموعة على أخرى). وبالاضافة الى ذلك تذكر ويبرق (Wiberg, 2007) أن إحتواء بند على الأداء التفاضلي لا يعني بالضرورة أنه بند سيئ، لذلك علينا محاولة معرفة مصدر ذلك. غير أن هناك تطور بسيط في مجال تحديد المصادر والأسباب الموضوعية، أو الخصائص الموضوعية المميزة للبنود التي تحتوي على الأداء التفاضلي للبند. إنه من الصعب جدا إكتشاف أسباب هذا الأداء المختلف لبنود الاختبار، أو النقص المشترك بين البنود التي تحتوي على الأداء التفاضلي للبند. وحسب روسوس (Roussos, 1996) الأسباب الموضوعية المحتملة لحدوث الأداء التفاضلي للبند تبقى غير معروفة إلى حد كبير. كما أشارت ويبرق (Wiberg, 2007) أن من بين الأسباب المحتملة هو أن البند يقيس سمة كامنة غير التي يقيسها الإختبار. غير أن بيترسون وآخرون (Petersen, et al, 2003) قدموا بعض الأسباب الأخرى فبالاظافة الى سوء الترجمة، ذكروا الأسباب العشوائية حيث يحدث الأداء التفاضلي للبند بمحض الصدفة، كما يمكن أن يكون هذا الأداء التفاضلي للبند زائف (Pseudo-DIF) خاصة عند إستخدام الدرجة الكلية للاختبار كمحك مطابقة، وبما أن هذه الدرجة الكلية هي مجموع درجات البنود فإن إختبار الأداء التفاضلي للبند لا يكون مستقلا، ويصبح زائفا وهذا مايوضح بأن الدرجة الكلية هي غير مثالية للإستخدام كمحك مطابقة، غير أنها كثيرا ما تكون أحسن اختيار لعدم توفر البديل، وخاصة إذا كان الإختبار يحتوي على عدة بنود فمن الممكن تحديد مجموعة من البنود الجوهرية التي لا تحتوي على الأداء التفاضلي

للبند ونستخدم الدرجة الكلية التي نتحصل عليها منها كمحك مطابقة لأفراد المجموعتين.

لقد قام هوانق (Huang, 2012) بذكر بعض الدراسات التي أعطت بعض التفسيرات للأسباب الأكثر شيوعا لحدوث الأداء التفاضلي للبند ومنها: الألفة بمحتوى البنود، والخبرة، والثقافة، ومدى إهتمام المفحوصين بمحتوى الاختبار، وردود فعلهم السلبية والغضب والاشمئزاز والخوف من محتوى الاختبار، جميع هذه التفسيرات أعطيت لتفسير سبب حدوث الأداء التفاضلي في بنود الاختبار. غير أنه هناك القليل أو لا توجد أدلة إمبريقية تدعم هذه التخمينات حول الإختلافات في المجموعات في الجنس (بين الذكور والإناث)، والعرق (بين المفحوصين البيض والسود)، اللغة (بين الفرنسية والانجليزية)، وهكذا دواليك في تفاعلهم مع بنود الاختبار. ليس هناك إطار نظري قوي لتفسير سبب حدوث الأداء التفاضلي للبند، إلا أنه من الضروري اللجوء الى التحليل التحكيمي للكشف عن هذه الأسباب الموضوعية. وفي الأخير تذكر إليس (Ellis, 2003) أن تطوير طرق موضوعية لتحديد أسباب ومصادر الأداء التفاضلي للبند يعتبر من أكبر التحديات الراهنة في هذا المجال.

١٠- كيفية التعامل مع البنود التي تحتوي على الأداء التفاضلي للبند:

يطرح بعض الباحثين (Boone, 2014; Zumbo, 1999) تساؤلا مهما وهو: ماذا يفعل الباحث في حالة العثور على بند يحتوي على الأداء التفاضلي للبند؟ هل يقوم بحذف هذا البند؟ وهذا ما لا ينصح به، أم يقوم بالتحفظ عليه إلى غاية التحقق منه عن طريق المحكمين وعمل المزيد من الدراسات حول صلاحية البند. إن جزءا من الاجابة على هذه التساؤلات يتعلق بمدى أهمية القرارات التي نتخذها على أساس عملية القياس هذه، أو مدى خطورة الوقوع في الخطأ جراء إتخاذ قرار خاطئ. غير أن بعض الباحثين (Fleishman, 2002; Teresi, 2007; Wiberg, 2007) 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 المحتوى للإختبار، والفقدان المحتمل للمعلومات من قياس مستويات مختلفة من السمة الكامنة. على العموم حذف هذه البنود قد يكون غير مقبول مع الاختبارات التي لا تحتوي على عدد كبير من البنود. كما أشار أورلندو (Orlando, 2002) أن تأثير وجود الأداء التفاضلي للبند في بنود اختبار ما يختلف حسب حجم الأداء التفاضلي للبند، وعدد البنود التي تحتوي عليه في الإختبار، والغرض من إستخدام الإختبار.

١١- خاتمة

يعتبر الأداء التفاضلي للبند من أهم الأدلة التي يجب على الباحث جمعها ليدعم صدق درجات الاختبار الذي يسعى إلى استخدامه، لكي تزيد ثقته في أن الفروق في الدرجات الموجودة في العينة هو راجع إلى إختلافات حقيقية في السمة المقاسة، وليس بسبب إنتماء المفحوصين الى مجموعات مختلفة (حسب الجنس، أو العمر، أو العرق، أو اللغة وغيرها...)، وبالتالي يستطيع المقارنة بين هذه الجموعات الفرعية بكل عدل. يعتبر الأداء التفاضلي للبند من الأخطاء المنتظمة التي تهدد صدق درجات الاختبار وهو نوعين، المنتظم حيث أن البند يفضل إما المجموعة المرجعية أو المركزية في جميع مستويات السمة المقاسة، وهناك النوع غير المنتظم حيث يفضل البند إحدى المجموعتين مثلا في المستوى المنخفض من السمة المقاسة ويفضل الأخرى في المستوى العالي. وللكشف عن هذا الأداء التفاضلي للبند هناك طرق تحكيمية، وأخرى إحصائية منها ما يعتمد على النظرية الكلاسيكية للقياس مثل طريقة الانحدار اللوجستي، حيث تستخدم الدرجة الكلية كمحك للمطابقة وأخرى تعتمد على النظرية الحديثة، مثل طريقة نسبة الأرجحية التي تعتمد على السمة الكامنة كمحك للمطابقة، وأحدث الطرق

هي تلك التي تعتمد على النمذجة بالمعادلة البنائية. إن البند الذي يحتوي على الأداء التفاضلي للبند لا يعني بالضرورة أنه متحيز، فهذا لا يكفي مما يجعل من الضروري اللجوء إلى المحكمين لمعرفة المصادر والأسباب المحتملة لذلك. لقد أصبح تحليل الأداء التفاضلي للبند من الأدلة المهمة التي يجب على الباحث تقديمها في إطار التحقق من صلاحية الاختبارات، خاصة وأن هذا يزيد من ثقتنا في النتائج وبالتالي القرارات التي سنتخذها بناءا عليها لأننا قد أعطينا الدليل على صدق الاستدلال بدرجات الاختبار.

المراجع

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

2. Boone, W.J, Staver, J.R and Yale, M.S, (2014). ***Rasch Analysis in the Human Sciences***. Springer.

3. 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.

4. Chen, W-H and Revicki, D (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.

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

6. 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.

7. Emmert-Aronson B.O, Moore M.T.and Brown T.A, (2014). Differential Item Functioning of the Symptoms of Major Depression by Race and Ethnicity: An Item Response Theory Analysis***. Journal of sychopathology and Behavioral***

Assessment. Springer Science+Business Media New York. DOI 10.1007/s10862-014-9407-4.

8. 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.

9. Francisco J. Abad, F.J, Colom, R, Rebollo I and Escorial S, (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.

10. Gierl M, Khaliq, S.N and Boughton K, (1999). ***Gender Differential Item Functioning in Mathematics and Science: Prevalence and Policy Implications***. Paper Presented at the Symposium entitled "Improving Large-Scale Assessment in Education" at the Annual Meeting of the Canadian Society for the Study of Education. Sherbrooke, Québec, CANADA.

11. 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.

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

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

14. Huang, J and Han, T, (2012). Revisiting Differential Item Functioning: Implications for Fairness Investigation. ***International Journal of Education.*** 4,2. 74-86.

15. Karami, H, (2012). An Introduction to Differential Item Functioning. ***The International Journal of Educational and Psychological Assessment***. 11(2): 59-76.

16. Kwakkenbos L, Arthurs E, van den Hoogen FHJ, Hudson M, van Lankveld WGJM, et al. (2013) Cross-Language Measurement Equivalence of the Center for

Epidemiologic Studies Depression (CES-D) Scale in Systemic Sclerosis: A Comparison of Canadian and Dutch Patients. PLoS ONE 8(1): e53923. doi:10.1371/ journal.pone.0053923.

17. Lyons-Thomas J, Sandilands D and Ercikan K, (2014). Gender Differential Item Functioning in Mathematics in Four International Jurisdictions. ***Education and Science***. 39, 172. 20-32.

18. Maller, S.J, (2003). Best Practices in Detecting Bias in Nonverbal Tests. In MacCallum, R.S. (Eds). ***Handbook of Nonverbal Assessment***. (PP 23-48). Springer Science+Business Media, LLC.

19. Molander B, Holmström S and TakšiĆ V, (2011) South and North: DIF Analyses of University-Student Responses to the Emotional Skills and Competence Questionnaire. ***Psychological Topics***. 20, 3, 425-447.

20. 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.

21. Ong, M.L, Lu, L, Lee, S, and Cohen, A, (2015). A Comparison of the Hierarchical Generalized Linear Model, Multiple-Indicators Multiple-Causes, and the Item Response Theory-Likelihood Ratio Test for Detecting Differential Item Functioning. In Millsap, R.E, Bolt, D.M, Van der Ark, L.A and Wang, W-C. (Eds). ***Quantitative Psychology Research The 78th Annual Meeting of the Psychometric Society***. (PP 343-357). Springer.

22. Orlando M and Marshall G.N, (2002). Differential Item Functioning in a Spanish Translation of the PTSD Checklist: Detection and Evaluation of Impact. ***Psychological Assessment***. 14,1. 50-59.

23. 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.

24. Petersen, M.A Groenvold, M, Bjorner, J.B, Aaronson, N, Conroy, T, Cull, A, Fayers, P, Hjermstad, M, Sprangers, M and Sullivan, M, (2003). Use of differential

item functioning analysis to assess the equivalence of translations of a questionnaire. ***Quality of Life Research*** 12: 373–385.

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

26. Roussos, L. A., & Stout, W. F. (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.

27. Sideridis, g and Tsaousis, I, (2013a). 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.

28. Teresi, J.A and Fleishman, J.A, (2007). Differential item functioning and health assessment. ***Quality of Life Research***. 16:33–42.

29. Teresi, J.A. Ocepek-Welikson, K. Kleinman, M. Eimicke, J.P. Crane, P.K. Jones, R.N. Lai, J-H. Choi, S.W. Hays, R.D. Reeve, B.B. Reise, S.P. Pilkonis, P.A and Cella, D, (2009). Analysis of differential item functioning in the depression item bank from the Patient Reported Outcome Measurement Information System (PROMIS): An item response theory approach. ***Psychology Science Quarterly***. 51,2. 148-180.

30. Teresi, J.A. Ramirez, M. Lai, J-H and Silver, S, (2008). Occurrences and sources of Differential Item Functioning (DIF) in patient-reported outcome measures: Description of DIF methods, and review of measures of depression, quality of life and general health. ***Psychology Science Quarterly***. 50,4. 538-612.

31. Wiberg, M, (2007). **Measuring and detecting differential item functioning in criterion-referenced licensing test : A theoretic comparison of methods.** ***Institutionen f□r beteendevetenskapliga m□tningar, Ume□.***

32. 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.