مجلـــــة العلـــــوم التربـــــوية والدراســـــات الإنســـــانيـــة
ISSN: 2709-0302 (online)

Humanities and Educational Sciences Journal
ISSN: 2617-5908 (print)

الأداء التفاضلي للبنود وتأثيره على الأداء التفاضلي للاختبار باستخدام نماذج نظرية الاستجابة للمفردة والتحليل العاملي التوكيدي متعدد المجموعات(*)

د/ ربيع عبده أحمد رشوان
أستاذ علم النفس التربوي المشارك
قسم علم النفس -كلية التربية-جامعة القصيم
المملكة العربية السعودية

*) تاريخ تسليم البحث 14/11/2020. تاريخ قبوله للنشر 20/11/2020.
*) موقع المجلة: http://hesj.org/ojs/index.php/hesj/index

الأداء التفاضلي للبنود وتأثيره على الأداء التفاضلي للاختبار باستخدام نماذج نظرية الاستجابة للمفردة والتحليل العاملي التوكيدي متعدد المجموعات

د. ربيع عبده أحمد رشوان

أستاذ علم النفس التربوي المشارك
قسم علم النفس بكلية التربية- جامعة القصيم
المملكة العربية السعودية

ملخص البحث:

هدف البحث الحالي للتعرف على الأداء التفاضلي لبنود اختبار واطسون وجليسر "النسخة القصيرة" للتفكير الناقد في ضوء التخصص الأكاديمي باستخدام النموذج البارامتري ثنائي المعلمة ونموذج Ramsay للانحدار اللابارامتري الممهد، والكشف عن تأثير الأداء التفاضلي للبنود على الأداء التفاضلي للاختبار باستخدام التحليل العاملي التوكيدي متعدد المجموعات، وتم تطبيق الاختبار على 396 طالب من طلاب جامعة القصيم (198 من التخصصات العلمية، 198 من التخصصات الأدبية) وباستخدام برنامجي IRTPRO وTestGraf توصلت الدراسة لعدد من النتائج منها:

- للعديد من بنود الاختبار أداء تفاضلي لصالح التخصصات العلمية وبعض منها له أداء تفاضلي لصالح التخصصات الأدبية.
- باستخدام التحليل العاملي التوكيدي متعدد المجموعات أكدت النتائج تأثير الأداء التفاضلي للبنود على الأداء التفاضلي للاختبار ككل، وتم تفسير النتائج في ضوء محتوى بنود الاختبار والاختلاف في الأسس العلمية القائمة عليها النماذج البارامترية واللابارامترية في نظرية الاستجابة للمفردة الاختبارية.

الكلمات المفتاحية: الأداء التفاضلي للبند، النموذج البارامتري ثنائي المعلمة، نموذج الانحدار اللابارامتري الممهد، التحليل العاملي التوكيدي متعدد المجموعات، اختبار واطسون وجليسر للتفكير الناقد.

Differential Item Function and its Impact on the Differential Test Function Using Item Response Theory Models and Multiple Group Confirmatory Factor Analysis

Dr. Rabie A. A. Rashwan, PhD

Associate Professor of Educational Psychology
Faculty of Education; Qassim University (KSA)

Abstract:

The Current Research Aimed to Identify The Differential Item Function for Watson-Glaser Critical Thinking Test "Short Version" Using Tow Parameter Logistic Model, Nonparametric Smoothing Regression Model and Its Impact on The Differential Test Function by Multiple Group Confirmatory Factor Analysis, The Sample Consisted of 396 Student of Qassim University In KSA (198 Scientific Disciplines, 198 Literary Disciplines), by Using IRTPRO and TestGraf Programs The Following Results Were Reached:

- Many of The Test Item Has Differential Function Favoring The Students Majoring in Scientific Disciplines, and Some of Item Favoring The Students Majoring in Literary Disciplines.
- By Using Multiple Group Confirmatory Factor Analysis, Results Confirmed The Impact of Differential Item Function on The Differential Test Function As A Whole, The Results Interpreted in The Light of The Items Content and The Differences of The Scientific Basis for The Parametric and Nonparametric Item Response Theory Models.

Key Word: Differential Item Function, Tow Parameter Logistic Model, Nonparametric Smoothing Regression Model, Multiple Group Confirmatory Factor Analysis, Critical Thinking Test.

مقدمة:

يعد الكشف عن درجة امتلاك الفرد لسمة ما وبالتالي التعرف على الفروق الفردية بين الأفراد أو المجموعات في تلك السمة من أهم استخدامات الاختبارات والمقاييس النفسية، وعليه فإن عدم تكافؤ أداء بنود الاختبار بالنسبة للأفراد من مجموعات مختلفة يؤدى إلى نتائج مضللة لا يمكن الاعتماد عليها في وصف تلك الفروق.

ولذا فإن معايير بناء وتطوير الاختبارات الصادرة من المؤسسات والجمعيات العالمية ذات الصلة تنص صراحة وتحُث المهتمين بتطوير وبناء أدوات القياس على تطوير بنود متحررة من التحيز للمجموعات المختلفة كالنوع أو العرق أو المستوى الاجتماعي الاقتصادي أو الإعاقة (Pae & Park, 2006, 475; Lee & Geisinger, 2016, 155)؛ ولكي يتميز الاختبار بالصدق والعدالة بالنسبة للمختبرين يجب أن تتساوي احتمالات الاستجابات المختلفة لأي بند من بنوده لكل المختبرين من نفس مستوى السمة، وإذا اختلفت احتمالات الاستجابة على البند للأفراد من نفس مستوى السمة فإن هذا البند يُظهر أداءً تفاضلياً (Andrich & Hagquist, 2012, 387; Magis & De Boeck, 2014, 714; Monahan & Ankenmann, 2010, 194; Woods, 2008, 511).

ويحدث الأداء التفاضلي للبند Differential Item Functioning (DIF) عندما يكون لمجموعة من المختبرين احتمالات مختلفة في الإجابة للبند عن مجموعة أخرى، وقد يرجع ذلك لأن البند يقيس شيء أخر بجانب السمة المقصودة وأن هذا البند قد يكون غير منصف لبعض المختبرين، وهو ما يهدد الصدق البنائي للاختبار (Walker, 2011, 365)؛ ومن هنا يمكن اعتبار أن تحديد البنود التي لها أداءً تفاضلياً وحذفها أو تعديلها يزيد من دقة التنبؤات حول السمة الكامنة (Meyer, 2014, 69; Thissen, 2001, 1)؛ فالأداء التفاضلي للبند يعد من الأمور الحرجة في بناء وتطوير الاختبارات، خاصة الاختبارات التي يُعتمد عليها في اتخاذ القرارات، مثل قرارات القبول واختيار الأفراد ومدى ملائمتهم للدراسة أو العمل، وفي الحكم على الكفاءة والمفاضلة بين الأفراد (Baghi & Ferrara, 1989, 1; Kalaycioglu & Berberoglu, 2011, 468)؛ ويؤكد البعض على أن دراسة الأداء التفاضلي لبنود أدوات القياس يعد أمراً حتمياً ويسهم في صدق المقارنة وتكافؤ القياس بين المجموعات المختلفة (Teresi & Fleishman, 2007, 41; De Boeck, Cho, Wilson, 2011, 583).

ففي حالة وجود بنود لها أداء تفاضلي في الاختبار، فإن المقارنة بين المجموعات المختلفة قد ينتج عنها وجود فروق زائفة بين تلك المجموعات، فأداء الفرد على الاختبار يجب أن يكون فقط ناتجاً لمستوى السمة الكامنة وليس شيئاً آخر خلاف ذلك؛ وفي مجال القدرات العقلية هناك حاجة ماسة لوجود أدلة على تكافؤ القياس بالنسبة للمجموعات المختلفة (Gomes, 2012, 276; Stump, Monahan & McHorney, 2005, 357).

والكشف عن الأداء التفاضلي للبنود يلقى اهتماماً متزايداً في الفترة الأخيرة، وهناك العديد من البحوث التي هدفت إلى الكشف عن الأداء التفاضلي للبنود، وذلك بالاعتماد على الطرق المتوافرة في نظرية القياس التقليدية ونظرية الاستجابة للمفردة، ولكن هناك ندرة في الدراسات التي حاولت التعرف على علاقة الأداء التفاضلي للبنود بالأداء التفاضلي للاختبار ككل بطريقة منهجية وخاصة في البيئة العربية وهو ما يستدعي المزيد من الجهود البحثية لتغطية تلك الفجوة.

والجزئية السابقة على قدر كبير من الأهمية في مجال تكافؤ القياس لأنها ستوضح كيفية التعامل مع البنود التي لها أداء تفاضلي، وكذلك من يستخدم الاختبار غالباً لا يتعامل مع بنوده الفرعية وإنما يتعامل مع الدرجة الكلية أو درجة الأبعاد وإذا لم يكن أداء البنود والاختبار متماثلاً في المجموعات المختلفة فإن المقارنات بين تلك المجموعات لا تعكس الفروق الحقيقية في مستوى السمة.

وفي ضوء الإجماع بين المختصين في مجال التربية وعلم النفس على أهمية تعليم وقياس مهارات التفكير الناقد وأن اختبار واطسون وجليسر يعد من أكثر الأدوات شيوعاً واستخداماً في قياس مهارات التفكير الناقد خاصة في البيئة العربية منذ ظهور الصورة الأولى للاختبار عام 1964 والصورة المعدلة عام 1980 (العتيبي، 2012، 147؛ منصور، علي، 2010، 149) تتضح أهمية التعرف على الأداء التفاضلي لبنود هذا الاختبار في ضوء المتغيرات الديموجرافية المختلفة كالتخصص الأكاديمي.

مشكلة الدراسة:

تعتمد دقة القياس وبالتالي دقة المقارنة بين الأفراد في المجموعات المختلفة على تكافؤ معنى البنود في الاختبار وبالتالي احتمالات الاستجابة عليها، والذي يسهم في إظهار الفروق الحقيقية في السمة بين المجموعات المختلفة، ومعظم الدراسات التي قارنت بين طلاب التخصصات العلمية وطلاب التخصصات الأدبية في التفكير الناقد ومهاراته الفرعية وخاصة في البيئة العربية أكدت على دلالة الفروق بين طلبة التخصصين وفي غالبية الدراسات كانت الفروق لصالح طلبة التخصصات العلمية (حلفاوي، 1997؛ الحموري والوهر، 1998؛ منصور وعلي، 2010؛ سعادة، 2009؛ العتيبي، 2012؛ الحدابي، 2012؛ الجاف وسلمان، 2012)، وحاولت الدراسات تفسير تلك الفروق والكشف عن أسبابها بالتركيز على طبيعة التخصص وأن التخصصات الأدبية لا تركز في مقرراتها وطرق التدريس المرتبطة بها على تنمية التفكير الناقد لدى طلابها، وهذا التفسير قد يكون غير واقعي، فالتخصصات الأدبية كاللغة العربية والعلوم الشرعية قائمة أساساً على النقد والتحليل والاستنتاج في الكثير من مقرراتها الدراسية.

فإذا كان الاختلاف في نوعية الخبرات التعليمية التي يتعرض لها الفرد له تأثير مختلف في نمو بعض المتغيرات المعرفية بصفة عامة والتي منها التفكير الناقد (سليمان، 2006، 125؛ French, Hand, Therrien & Vazquez, 2012, 201)؛ فليس من الضروري أن يؤدي التأثير المختلف لهذه الخبرات إلى اختلاف مستوى هذه الإمكانات العقلية، ولكن الاختلاف هنا قد يكون في محتوى التفكير؛ وأغفلت الدراسات السابقة إمكانية أن يكون لبعض بنود الاختبار المستخدم في قياس التفكير الناقد أداء تفاضلي ضد طلبة التخصصات الأدبية، وهو ما يعد المبرر الرئيس للدراسة الحالية والتي تحاول التأكد مما إذا كانت تلك الفروق ترجع لفروق حقيقية في السمة المقاسة – التفكير الناقد – أو أنها راجعة للأداء التفاضلي لبنود اختبار واطسون وجليسر للتفكير الناقد باعتباره أكثر الاختبارات استخداماً في قياس التفكير الناقد.

فالمقارنة بين المجموعات في البنيات النفسية المختلفة لكي يكون له معنى وتعكس الفروق الحقيقية بين تلك المجموعات لابد أولاً من التأكد من افتراض أن الأداة تقيس نفس البنية المقصودة في تلك المجموعات وهو ما يعبر عنه بتكافؤ القياس وفي حالة ثبوت أن البنود أو الاختبار لا يؤدي بطريقة متكافئة للمجموعات المختلفة فيجب تعديل تلك البنود أو حذفها (Price, 1999, 3; Milfont & Fischer, 2010, 112; Wolf, Zahner, Kostoris & Benjamin, 2014, 1)؛ ويتم دراسة تلك القضايا في إطار ما يعرف بالأداء التفاضلي للبند والأداء التفاضلي للاختبار؛ وفي إطار نظرية القياس الكلاسيكية تعتبر طرق دراسة الأداء التفاضلي وتكافؤ القياس طرق مقيدة بمجتمع أو عينة الدراسة ومن هنا يوجه للإجراءات التي تتم في إطار هذه النظرية نقداً مفاده أن هذه الإجراءات لا تتمتع بالكفاءة في التأكد من تكافؤ القياس، ومن هنا جاءت الإجراءات المعتمدة على نظرية الاستجابة للمفردة الاختبارية لتوفر إجراءات محكمة للتحقق من تكافؤ القياس للمجموعات المختلفة (Price, 1999, 3; Kalaycioglu & Berberoglu, 2011, 468).

ونظراً للاهتمام المتزايد بقضية الأداء التفاضلي للبنود ظهرت العديد من الطرق التي تُستخدم في تحقيق هذا الهدف والتي يمكن تصنيفها إلى مجموعتين، الأولى تعتمد على الدرجات الملاحظة Observed Score (غالبا ما تكون الدرجة الكلية في الاختبار) كمتغير للمزاوجة بين المجموعات

موضوع الدراسة ومن أكثر هذه الطرق شيوعاً طريقة مانتل هانزل Mantel-Haenszel وطريقة الانحدار اللوغاريتمي Logistic Regression وطريقة التحليل المتآني SIBTEST وطريقة Angoff's Delta Plot وطريقة التحليل العاملي المقيد Restricted factor analysis، أما المجموعة الثانية منها فتعتمد على البنية الكامنة أو نظرية الاستجابة للمفردة الاختبارية ومنها اختبار مربع كاي "للورد" Lord Chi-Squared Test واختبار نسبة الأرجحية Likelihood Ratio Test، والمساحة بين المنحنيات المميزة للبند والتي يمكن تقديرها باستخدام النماذج البارامترية أو اللابارامترية التي تنتمي لنظرية الاستجابة للمفردة (Magis & Facon, 2013, 294; Lai, Teresi & Gershon, 2005, 284; Magis & De Boeck, 2014, 714; Lee & Geisinger, 2016, 142)

ومن أهم مشكلات الطرق المعتمدة على الدرجة الملاحظة كمعيار للمزاوجة بين المجموعات موضوع المقارنة عدم دقة هذا المعيار حيث يدخل في حسابه درجات البنود التي يفترض أن لها أداءً تفاضلياً، وهو ما يؤثر على كفاءة هذه الطرق، ومن هنا يؤكد البعض على أن عدم تحرير معيار المزاوجة بين المجموعات من البنود التي لها أداءً تفاضلياً يُحدث تضخيم للخطأ من النوع الأول وخاصة في طريقة مانتل هانزل وطريقة الانحدار اللوغاريتمي (French & Maller, 2007, 374; Teresi & Fleishman, 2007, 35; Monahan & Ankenmann, 2010, 193; Magis & DeBoeck, 2012, 292; Khalid & Glas, 2014, 187)

وقد غفلت العديد من الدراسات السابقة عن معالجة هذه المشكلة خاصة في البيئة العربية، حيث أن عدم تنقية الدرجة الكلية من تأثير البنود التي لها أداءً تفاضلياً يسهم في ابتعاد الدرجات الملاحظة عن المستوى الحقيقي للسمة ويجعل التوزيع الحقيقي للسمة الكامنة في المجموعتين غير متماثل، وهذا الافتراض هو أهم الافتراضات التي تقوم عليها إجراءات الكشف عن الأداء التفاضلي للبنود (Magis & De Boeck, 2014, 715).

ويتأكد مما سبق أن الاعتماد على الطرق المتاحة في نظرية الاستجابة للمفردة الاختبارية في الكشف عن الأداء التفاضلي للبند يجعل النتائج أكثر دقة ومصداقية، حيث لا يكون هناك تأثير لمشكلة الدرجة الكلية كمعيار للمزاوجة بين المجموعات، فالاعتماد على مقارنة المنحنيات المميزة للبند أو مقارنة معالم البند تعد تحليلات غير مشروطة بالمقابلة بين المجموعات في مستوى السمة Unconditional Analysis نظراً لافتراض نماذج نظرية الاستجابة للمفردة الاختبارية صراحة أن المساحة بين المنحنيات المميزة للبند في المجموعتين تكون على نفس متصل توزيع السمة (Zumbo, 2007, 226-227)؛ وكما يذكر Thissen (2001, 2) "أن لنظرية الاستجابة للمفردة الاختبارية بالاعتماد على اختبار نسبة الأرجحية مزايا كثيرة Striking Advantages في الكشف عن البنود ذات الأداء التفاضلي" حيث تتضمن إجراءاتها اختبار مدى صحة الفروض عن اختلاف بارامترات البند في المجموعات موضوع المقارنة (الصعوبة، التمييز، التخمين) وكذلك لأن تلك الإجراءات لا تعتمد على مجموع درجات بنود الاختبار؛ حيث توفر تلك الإجراءات الضبط الكافي لمستوى السمة في المجموعتين والذي ينتج من تدريج وتعيير Calibration البند في المجموعتين على نفس التدريج وهو ما يجعل بارامترات البند متحررة من خصائص العينة، بعكس الطرق التقليدية المعتمدة على الدرجات الملاحظة (Baghi & Ferrara, 1989, 4-5; Kim & Oshima, 2012, 467; Santelices & Wilson, 2012, 6).

وعلى الرغم من التأكيدات السابقة بأهمية الاعتماد على الإجراءات المتوافرة في نظرية الاستجابة للمفردة في الكشف عن الأداء التفاضلي للبنود، إلا أنه قد تحدث مشكلة مفادها عدم التوافق بين الافتراض المبدئي عن توزيع السمة والتوزيع الحقيقي للسمة (Keller, 2000, 20)؛ وفي حالة حدوث ذلك التناقض فإن تقديرات معالم البند تكون متحيزة وغير دقيقة وهو ما يؤثر بالتالي على دقة الكشف عن الأداء التفاضلي للبنود (Woods, 2008, 515)؛ ومن النقد الموجه للدراسات المهتمة بالأداء التفاضلي أيضاً اعتمادها على عينات محدودة الحجم، على الرغم من أن غالبية الطرق

الإحصائية المستخدمة في الكشف عن الأداء التفاضلي تتطلب عينات كبيرة الحجم حتى نضمن الوصول إلى الخوارزميات الأمثل Optimization Algorithms ودقة تقدير للبارامترات أو تحقق صدق افتراضات هذه الإجراءات الإحصائية (Magis & Facon, 2013, 294).

وللتغلب على المشكلات السابقة ظهر ما يعرف بالنماذج اللابارامترية في نظرية الاستجابة للمفردة وهي نماذج موازية للنماذج البارامترية المعروفة وتستند إلى جميع شروطها ماعدا شكل توزيع السمة الكامنة حيث لا تفترض هذه النماذج اللابارامترية شكل مسبق لتوزيع السمة الكامنة كما هو الحال في النماذج البارامترية والتي تستند في غالبيتها إلى التوزيع الطبيعي أو اللوغاريتمي للسمة الكامنة.

فدقة طرق الكشف عن الأداء التفاضلي باستخدام النماذج البارامترية في نظرية الاستجابة للمفردة تتحدد بمدى دقة تقديرات السمة وبارامترات البند في ضوء الدالة الرياضية المعتمد عليها النموذج، وعلى الرغم من تميز الإجراءات في نظرية الاستجابة للمفردة من حيث الوضوح ومناسبة الدوال الرياضية المعتمدة عليها النظرية، إلا أنه وفي حالة انتهاك الافتراض القبلي بأن التوزيع الحقيقي للقدرة يناسب ما هو مفترض في تلك النماذج الرياضية فإن استخدام هذه النماذج في الوصول إلى دالة الاستجابة للبند يترتب عليه نتائج غير دقيقة (Duncan & MacEachern, 2008, 42; Keller, 2000, 19; Meijer, 2004, 8)؛ فالمشكلة هنا تكمن في مدى ملاءمة النموذج الرياضي فعلاً للتوزيع الحقيقي للقدرة موضوع الاهتمام؛ وقد يختلف ذلك التوزيع عن ما هو مفترض من قبل النموذج، بل وقد لا يمكن التنبؤ به مسبقاً (Basokcu & Ogretmen, 2014, 32).

وهنا تظهر أهمية المنحى اللابارامتري في نظرية الاستجابة للمفردة الاختبارية، والذي من أهم ميزاته عدم وجود نموذج رياضي محدد مسبقاً، وهو ما يسمح للباحث من اكتشاف شكل (صيغة) المنحنى المميز للبند وبالتالي دراسة خصائص هذا البند والحكم على مدى جودته.

ففي النماذج اللابارامترية يتم تدريج السمة أولاً ثم يتم تقدير المنحنى المميز للبند تبعاً لذلك التدريج بعكس ما هو متبع في النماذج البارامترية، ومن هنا تتميز النماذج اللابارامترية بأنها أقل تقيداً من النماذج البارامترية من حيث شكل توزيع السمة؛ حيث يؤكد (Drasgow, Levine, Williams, McLaughlin, Candell, 1989, 286) على أنه من الناحية التطبيقية في مجال القياس في العلوم الإنسانية والتربوية من الصعب جداً أن يحكم الباحث على نموذج رياضي بارامتري معين بأنه صحيح ويتناسب مع توزيع السمة الكامنة موضوع الاهتمام، علاوة على أن مؤشرات الملاءمة لهذا النموذج في حد ذاتها لا تخلو من المشاكل.

وفي ضوء ما سبق فهناك تأكيدات متزايدة على أن الاعتماد على التقدير اللابارامتري للمنحنى المميز للبند يفيد بدرجة أكبر منها في حالة المنحى البارامتري عند المقارنة بين المجموعات والكشف عن البنود غير الجيدة (Drasgow et al., 1989, 285; Meyer, 2014, 113-114; Basokcu & Ogretmen, 2014, 32; Nay, Strong, Nay, Beidel & Turner, 2007, 136)؛ وعلى الرغم من الميزات السابق للنماذج اللابارامترية، إلا أن الجهود البحثية القائمة على النماذج اللابارامترية في الكشف عن الأداء التفاضلي للبنود محدودة جداً مقارنة بالجهود البحثية في إطار النماذج البارامترية.

وفي الدراسة الحالية تم استخدام الإجراءات المعتمدة على النموذج البارامتري ثنائي المعلم ونموذج الانحدار اللابارامتري الممهد في الكشف عن البنود التي لها أداء تفاضلي وذلك بهدف المقارنة بين الطريقتين هذا من جهة ومن جهة أخرى الاستفادة من الميزات التي تتوافر في كلاً منهما في تحديد البنود التي لها أداء تفاضلي؛ فالجمع بين أكثر من طريقة في دراسة الأداء التفاضلي للبنود يوفر دليلاً أكثر دقة عن البنود ذات الأداء التفاضلي إذا ما اتفقت الطرق المختلفة على اعتبار أن البند له أداءً تفاضلياً (Fidalgo, Ferreres & Muniz, 2004, 24)؛ كذلك أكدت العديد من

الدراسات السابقة على أن الاعتماد على فحص التمثيل البياني للمنحنى اللابارامتري المميز للبند في المجموعتين المرجعية والبؤرية بجانب التقديرات الكمية البارامترية للأداء التفاضلي للبند يوفر أدلة أكثر مصداقية ودقة عن مشكلات البند خاصة في حالة صغر حجم العينة (Sueiro & Abad, 2011, 835).

ويتضح مما سبق أن الكشف عن البنود التي لها أداءً تفاضلياً في أداة القياس أمر على قدر كبير من الاهمية، وعلى الرغم من أن فكرة وجود بنود لها أداء تفاضلي تفترض ضمناً أن بنية السمة المقاسة تختلف من مجموعة لمجموعة أخرى خاصة في حالة زيادة نسبة البنود التي لها أداءً تفاضلياً في الأداة (Walker, 2011, 365-366)؛ إلا أن الأداء التفاضلي للبنود قد لا يظهر تأثيره على الأداء التفاضلي الكلي للاختبار نظراً لأن تأثير الأداء التفاضلي للبنود قد يلغي بعضه بعضا (Baker, Caison & Meade, 2007, 548).

فدراسة (Pae, 2004) أشارت نتائجها إلى وجود بعض البنود التي لها أداء تفاضلي في اختبار الاستماع والفهم القرائي كاختبارات فرعية من اختبار الاستعداد الدراسي لطلاب الجامعة الكوريين في ضوء التخصص، إلا أن التحليلات الخاصة بالمنحنيات المميزة للاختبار ككل أكدت على عدم دلالة الفروق في الدرجات المتوقعة؛ وتوصلت لنفس النتيجة تقريبا دراسة (Runnels, 2013) والتي هدفت للكشف عن الأداء التفاضلي لبنود اختبار اللغة الإنجليزية في ضوء التخصص لدى طلاب جامعة هيروشيما بينكيو Hiroshima Bunkyo الخاصة باليابان.

كذلك أشارت نتائج دراسة (Takala & Kaftandjieva, 2000) إلى وجود 11 بنداً لها أداءً تفاضلياً بالنسبة للنوع في اختبار المفردات في اللغة الفنلندية كلغة ثانية من أصل 40 بنداً باستخدام النموذج البارامتري أحادي المعلم، وعلى الرغم من ذلك أكدت نتائجها على عدم وجود فروق في الاختبار ككل في ضوء النوع؛ كذلك أشارت نتائج دراسة (Zumbo, 2003) والقائمة على البيانات المولدة باستخدام المحاكاة إلى أنه على الرغم من وجود 1-16 بنداً من 38 بنداً لها أداءً تفاضلياً إلا أن ذلك لم يكن له تأثير على الأداء التفاضلي للاختبار؛ وفي دراسة (Drasgow, 1987) أكدت نتائجها على أن بعض بنود اختبار الاستعداد لدراسة الرياضيات لها أداء تفاضلي وعلى الرغم من ذلك أكدت النتائج على عدم وجود تأثير لذلك على الأداء التفاضلي للاختبار ككل.

في حين أشارت نتائج دراسة Stump et al., (2005) إلى أن البنود التي لها أداء تفاضلي كان لها تأثير على الدرجة الكلية للاختبار؛ ودراسة (Church, Alvarez, Mai, French, Katigbak & Ortiz, 2011) أكدت على أن بعض بنود قائمة العوامل الخمسة الكبرى للشخصية لها أداءً تفاضلياً بالنسبة لطلاب الجامعة الأمريكيين والفلبينيين، وأكدت نتائجها على أن تلك البنود لها تأثير على أبعاد القائمة ولذلك ظهر أن لبعض أبعاد القائمة أداء تفاضلي؛ وأكدت نتائج دراسة (Sahin, French, Hand & Gunel, 2015) أنه على الرغم من أن 10% من بنود اختبار كورينيل للتفكير الناقد CCTT لها أداء تفاضلي بين طلاب المرحلة المتوسطة التركيين والأمريكيين إلا أن ذلك لم يكن له تأثير على الأداء التفاضلي للاختبار ككل؛ وفي فحص الأداء التفاضلي لنفس الاختبار السابق في ضوء النوع لدى طلبة الجامعة بالولايات المتحدة الأمريكية توصلت دراسة (French et al., 2012) إلى أن 4 بنود في الاختبار بنسبة 5.6% لها أداء تفاضلي ولكن اعتبر أن هذه النسبة لا تؤثر على متوسطي درجات المجموعتين وصدق المقارنة بين الذكور والإناث.

وعلى الرغم من منطقية تفسير الدراسات السابقة بأن تأثير البنود التي لها أداء تفاضلي في الاختبار يلغي بعضه بعضاً وبالتالي لا يظهر تأثيره على الاختبار ككل، إلا أن ذلك التفسير يمكن اعتباره صحيح من الناحية الإحصائية حيث لا تظهر فروق دالة إحصائياً بين المجموعة المرجعية والمجموعة البؤرية في الاختبار ككل، لكن السؤال الذي يطرح نفسه مفاده "هل البنية الكامنة المقاسة بواسطة الاختبار واحدة في المجموعتين؟".

وقد يكون السبب في عدم تأكد تأثير البنود التي لها أداء تفاضلي على الأداء التفاضلي للاختبار ككل مرجعة الطريقة التي تم بها الكشف عن هذا التأثير والتي انحصرت في الكثير من الدراسات السابقة في المقارنة بين متوسطي الدرجات الملاحظة بين المجموعتين موضوع المقارنة بعد حذف البنود التي لها أداء تفاضلي وهذه المقارنة لا تكشف عن التأثير الحقيقي للبنود ذات الأداء التفاضلي إذا كان مستوى المجموعتين متشابهاً فعلاً؛ كذلك قد يكون مرجعه صغر عدد البنود التي لها أداء تفاضلي مقارنة بالعدد الكلي لبنود الاختبار.

وفي الدراسة الحالية تم استخدام التحليل العاملي التوكيدي متعدد المجموعات في التعرف على مدى تشابه البنية الكامنة للسمة موضوع القياس من حيث مدى تماثل علاقة المتغيرات الملاحظة بالمتغير الكامن في المجموعتين موضوع المقارنة ومدى تماثل التباين والتغاير والبواقي وهذه الإجراءات من أكثر الإجراءات دقة في اكتشاف مدى تماثل نموذج القياس في المجموعات المختلفة (Abbott, 2007,12; Teresi & Fleishman, 2007, 37; Koh & Zumbo, 2008, 471; Milfont & Fischer, 2010, 112)؛

وفي ضوء ما سبق تحاول الدراسة الحالية المقارنة بين النماذج البارامترية واللابارامترية في نظرية الاستجابة للمفردة في الكشف عن الأداء التفاضلي لبنود اختبار التفكير الناقد لواطسون وجليسر (النسخة القصيرة) وفقاً للتخصص الأكاديمي، وعلاقة الأداء التفاضلي للبنود بالأداء التفاضلي الكلي للاختبار باستخدام التحليل العاملي التوكيدي متعدد المجموعات، وتتحدد مشكلة الدراسة الحالية في الأسئلة التالية:

1- ما قيم مؤشرات جودة مطابقة بنود اختبار التفكير الناقد لواطسون وجليسر "النسخة القصيرة" للنموذج البارامتري ثنائي المعلم؟
2- ما بنود اختبار واطسون وجليسر للتفكير الناقد "النسخة القصيرة" التي لها أداء تفاضلي وفقاً للتخصص باستخدام النموذج البارامتري ثنائي المعلم؟
3- ما بنود اختبار واطسون وجليسر للتفكير الناقد "النسخة القصيرة" التي لها أداء تفاضلي وفقاً للتخصص باستخدام نموذج الانحدار اللابارامتري الممهد؟
4- هل تتشابه البنية الكامنة المقاسة باستخدام اختبار واطسون وجليسر للتفكير الناقد "النسخة القصيرة" لدى طلبة التخصصات العلمية وطلبة التخصصات الأدبية باستخدام التحليل العاملي التوكيدي متعدد المجموعات؟

أهداف الدراسة: تهدف الدراسة الحالية إلى:

1- الكشف عن بنود النسخة القصيرة من اختبار واطسون وجليسر للتفكير الناقد التي لها أداء تفاضلي للتخصص الأكاديمي (علمي، أدبي) لدى طلاب الجامعة باستخدام النموذج البارامتري ثنائي المعلم، وباستخدام نموذج الانحدار اللابارامتري الممهد، والمقارنة بين النتائج المعتمدة على النماذج البارامترية واللابارامترية.
2- التعرف على مدى تماثل البنية الكامنة المقاسة باستخدام النسخة القصيرة من اختبار واطسون وجليسر للتفكير الناقد لدى طلاب التخصصات العلمية وطلاب التخصصات الأدبية باستخدام التحليل العاملي التوكيدي متعدد المجموعات.
3- تحديد الأسباب المحتملة لظهور الأداء التفاضلي لبعض بنود اختبار واطسون وجليسر للتفكير الناقد في ضوء التخصص (علمي، أدبي)، والتعرف على مستويات القدرة التي يظهر عندها الأداء التفاضلي للبنود.

أهمية الدراسة:

تتضح أهمية الدراسة الحالية من خلال الجانب النظري والتطبيقي للنتائج التي يمكن التوصل إليها، ويتمثل ذلك في الآتي:

أولاً: الأهمية النظرية:

1- تركز الدراسة الحالية على الأداء التفاضلي للبنود وللاختبار ككل والذي يعد من أهم المشكلات التي تهدد صدق أدوات القياس والقرارات المترتبة عليها.
2- استخدام أحد النماذج اللابارامترية في نظرية الاستجابة للمفردة وتقديمة بصورة مفصلة للبيئة العربية وتوجيه نظر المهتمين إلى استخداماته المختلفة، فهناك ندرة واضحة في الاهتمام بالنماذج اللابارامترية مقارنة بالنماذج البارامترية في نظرية الاستجابة للمفردة.
3- استخدام التحليل العاملي التوكيدي متعدد المجموعات وتفصيل لمحكات المقارنة الممكن الاعتماد عليها وهو ما يمكن الاستفادة منه في دراسات مستقبلية.
4- المقارنة بين نتائج تطبيق النماذج البارامترية واللابارامترية في الكشف عن الأداء التفاضلي للبنود وهو ما يسهم في إيضاح نقاط القوة والضعف في كل أسلوب.

ثانياً: الأهمية التطبيقية:

1- في ضوء النتائج المتوقعة للدراسة الحالية تتضح حقيقية الفروق بين طلاب التخصصات العلمية والأدبية في التفكير الناقد ومهاراته الفرعية، وهو ما يسهم في تصحيح العديد من المفاهيم المرتبطة بهذه المشكلة.
2- تقديم الدليل على مدى الحاجة لبناء اختبارات للتفكير الناقد تتصف بالعدالة والإنصاف بين المجموعات الديموجرافية المختلفة ومدى حاجة اختبارات التفكير الناقد الحالية للتطوير، وبصفة خاصة اختبار واطسون وجليسر للتفكير الناقد.
3- لفت انتباه الباحثين والتربويين لضرورة الاعتماد على أدوات قياس تتصف بالعدالة والإنصاف بين المختبرين، والتأكد من عدم تحيز بنود الادوات والذي يهدد صدق النتائج التي يتم التوصل إليها.
4- توجيه الباحثين لضرورة تضمين بيانات عن الأداء التفاضلي لبنود أدوات القياس المستخدمة في بحوثهم واعتبارها جزء من إجراءات التأكد من شروط أدوات البحث كالصدق والثبات، خاصة مع توافر العديد من إجراءات الكشف عن الأداء التفاضلي والمتاحة من خلال البرامج الإحصائية شائعة الاستخدام.

مصطلحات الدراسة:

- **الأداء التفاضلي للبند**: Differential Item Functioning (DIF)

يشير الأداء التفاضلي للبند إلى الفروق السيكومترية في أداء البند لدى مجموعتين مختلفتين، ويحدث حينما تُظهر مجموعتين متكافئتين في السمة احتمالات مختلفة للإجابة الصحيحة على نفس البند (Kline, 2004, 551; Pae & Park, 2006, 475; Woods, 2009, 42).

ويعرف إجرائياً في الدراسة الحالية بأنه اختلاف احتمالات الوصول للإجابة الصحيحة على بنود اختبار واطسون وجليسر للتفكير الناقد لدى طلبة التخصصات الأدبية عنها لدى طلبة التخصصات العلمية والتي سيتم الكشف عنها باستخدام النموذج البارامتري ثنائي المعلم ونموذج الانحدار اللابارامتري الممهد.

- **الأداء التفاضلي للاختبار:** Differential Test Functioning (DTF)

ويعني عدم تماثل الخصائص السيكومترية للاختبار في المجموعات المختلفة من نفس مستوى السمة، (Meade & Wright, 2012, 1016)؛ ويكون للاختبار أداء تفاضلي إذا اتضح أنه يقيس سمات كامنة مختلفة لدى المجموعات المختلفة التي يمكن أن يطبق عليها أو أنه يقيس نفس السمة ولكن بدرجات متفاوتة من الدقة (أبو مسلم، 2010، 198؛ Andrich & Hagquist, 2012, 387; Runnels, 2013, 1)؛ فمصطلح الأداء التفاضلي يستخدم ليشير إلى عدم تماثل بنية السمة

المقاسة بواسطة أداة القياس والذي يسهم في زيادة أو نقصان درجات أحد المجموعات موضوع المقارنة، رغم تماثلهم في مستوى السمة (Tan & Gierl, 2005, 3).

ويعرف إجرائياً في الدراسة الحالية بأنه عدم تماثل البنية الكامنة المقاسة باختبار واطسون وجليسر للتفكير الناقد "النسخة القصيرة" لدى طلبة التخصصات العلمية وطلبة التخصصات الأدبية ويتم الكشف عنه باستخدام التحليل العاملي التوكيدي متعدد المجموعات.

- **التفكير الناقد:** Critical Thinking

مفهوم واطسون وجليسر للتفكر الناقد يتمثل في أن التفكير الناقد هو "المحاولة المستمرة لاختبار الحقائق والآراء في ضوء الأدلة بدلاً من القفز إلى النتائج، ويتضمن ذلك معرفة طرق البحث المنطقي التي تساعد في تحديد قيمة مختلف الأدلة والوصول إلى نتائج سليمة واختبار صحة النتائج وتقويم المناقشات بطريقة موضوعية خاصة (الحموري، الوهر، 1998، 116؛ الجاف وسليمان، 2012، 184؛238 Sahin et al., 2015,)؛ وفي الدراسة الحالية تم تبني مفهوم التفكير الناقد لواطسون وجليسر كتعريف إجرائي، ويتحدد بالدرجة التي يحصل عليها الطالب في الاختبار المستخدم في الدراسة الحالية وأبعاده الفرعية.

- **المجموعة المرجعية والمجموعة البؤرية** Reference & Focal Groups

الدراسات المهتمة بالكشف عن الأداء التفاضلي للبنود ترتكز على دراسة الفروق في احتمال الإجابة الصحيحة للبند بين مجموعتين متكافئتين في مستوى السمة (Schmitt & Dorans, 1988, 5; Gotzmann, 2002, 5) هما:

- المجموعة المرجعية أو مجموعة الأساس: وهي التي تستخدم لتقدير الاحتمال المشروط للأداء الصحيح على البند عند كل مستويات السمة، وهي المجموعة التي يكون الأداء التفاضلي لصالحها، ويطلق عليها أحياناً مجموعة الأغلبية، وفي الدراسة الحالية تمثل تلك المجموعة طلبة التخصصات العلمية.
- المجموعة البؤرية: وهي المجموعة المدروسة والتي يتم الكشف عن الأداء التفاضلي ضدها، وتسمى في بعض الأحيان مجموعة الأقلية، وفي الدراسة الحالية تمثل تلك المجموعة طلبة التخصصات الأدبية.

الإطار النظري:

أولاً: الأداء التفاضلي:

يعرف Abbott (2007, 8) الأداء التفاضلي بأنه "احتمال اختلاف أداء المجموعات المختلفة من المختبرين من نفس مستوى السمة على البند"؛ وبصفة عامة يُعبر عن الأداء التفاضلي بأنه "الحكم على مدى تغير القيم التقديرية لبارامترات البند من مجموعة لأخرى من المختبرين"، ويجاب فيه عن تساؤل مفاده "هل يؤدى البند بنفس الطريقة لأي مجموعة من المختبرين" (Zumbo, 2007, 223; French & Maller, 2007, 373)؛ فالأداء التفاضلي يصف البنود التي تختلف في طريقة قياسها للبنية المستهدفة عندما يقدّم الاختبار إلى مجموعات ذات خلفيات ديموجرافية مختلفة من نفس مستوى السمة (Ariffin, Idris & Ishak, 2010, 2).

ويضيف Liu (2011, 1) أن الأداء التفاضلي للبند يصف البنود التي تحمل في خصائصها ما يساعد أحد المجموعات في التوصل للإجابة الصحيحة أكثر من الأخرى؛ بينما يركز Meade & Wright (2012, 1016) في تعريفه للأداء التفاضلي على الخصائص السيكومترية لأداة القياس ومدى تماثل هذه الخصائص في المجموعات المختلفة من نفس مستوى السمة، ومدى تماثلها لنفس المجموعة وفقاً لعامل الزمن كالدراسات الطولية، ومدى تماثلها باختلاف صيغة الإجابة كالاختبارات الورقية والاختبار عبر الإنترنت.

ويمكن التمييز بين شكلين للأداء التفاضلي (Lai et al., 2005, 285; Ozdemir, 2015, 2076) هما:

-الأداء التفاضلي المنتظم Uniform DIF: ويحدث عندما تظهر احتمالات مختلفة في صعوبة البند لمجموعتين من الأفراد في كل مستويات السمة، بمعنى أن أداء إحدى المجموعتين على بند ما يكون أفضل من المجموعة الأخرى عند كل مستويات السمة.

-الأداء التفاضلي غير المنتظم: Non-Uniform ويحدث عندما تظهر احتمالات مختلفة في صعوبة البند لمجموعتين من الأفراد عند مستوى معين من السمة، بمعنى أن أداء إحدى المجموعتين على بند ما يكون أفضل من الأخرى ويختلف الفرق بين المجموعتين باختلاف مستوى السمة.

وفي ضوء ما أشار إليه Finch & French (2007, 566) يمكن التمييز بين نوعين من الأداء التفاضلي غير المنتظم هما:

-الأداء التفاضلي غير المنتظم المرتب Non-Uniform Ordinal DIF ويحدث عندما يكون البند أصعب بالنسبة لإحدى المجموعتين عند كل مستويات السمة ولكن الفروق في صعوبة البند ليست ثابتة عند كل مستويات السمة (يختلف الفرق في صعوبة البند باختلاف مستوى السمة).

-الأداء التفاضلي غير المنتظم غير المرتب Non-Uniform Disordinal DIF ويحدث عندما يكون البند أصعب لإحدى المجموعتين عند مستوى معين من السمة وأسهل لنفس المجموعة عند مستوى أخر للسمة، بمعنى أخر هناك تفاعل بين مستوى السمة والأداء على البند.

وهنا يتضح التشابه بين الأداء التفاضلي المنتظم والأداء التفاضلي غير المنتظم المرتب ولكن الفرق بينهما يكمن في أن الأداء التفاضلي المنتظم يكون فيه تمييز البند ثابت عند كل مستويات السمة بمعنى توازي المنحنى المميز للبند في المجموعتين عند كل مستويات السمة لكن في الأداء التفاضلي غير المنتظم المرتب لا يكون المنحنيان متوازيان ويكون تميز البند مختلف في المجموعتين (Walker, 2011, 368-369).

وقد ينتج عن الأداء التفاضلي للبنود أداء تفاضلي للاختبار ككل Differential Test Functioning (DTF)، وهنا لا يمكن اعتبار أن الاختبار يقيس نفس السمة في حالة المجموعات المختلفة؛ وفي حالة زيادة عدد البنود التي لها أداء تفاضلي فإنه لا يكون للدرجات النهائية في الاختبار نفس المعنى في المجموعات المختلفة (Runnels, 2013, 1)؛ ففي هذه الحالة يكون هناك زيادة أو نقصان في تقدير السمة لأفراد إحدى المجموعات مقارنة بالأخرى (LeBouthillier, Thibodeau, Alberts, Hadjistavropoulos & Asmundson, 2015, 385; Lee & Geisinger, 2016, 142)؛ ومن هذا يتضح أن الأداء التفاضلي للاختبار يحدث عندما تختلف الدرجات المتوقعة على المستوى الكلي للاختبار بالنسبة لمجموعتين من المختبرين، أو عندما لا نستطيع الإقرار بعدم تغير القياس باستخدام الاختبار Measurement Invariance of A Test بالنسبة لمجموعتين من المختبرين من نفس مستوى السمة.

ويعد الأداء التفاضلي للبند شرطاً أساسياً لاعتباره متحيزاً لكنه شرط غير كافٍ، بمعنى أنه إذا أظهر البند أداءً تفاضلياً فإنه لابد من إجراءات إضافية للحكم بأن ذلك البند متحيزاً، كتحكيم البند وتقييمه تجريبياً، للبحث عن سبب ظهور الأداء التفاضلي؛ فتحديد البنود التي لها أداء تفاضلي يعد أول خطوة من خطوات الكشف عن تحيز البنود وبالتالي تحيز الاختبار ككل؛ حيث يعد الأداء التفاضلي للبند بمثابة المؤشر الإحصائي لتحيزه (Liu, 2011, 10; Meyer, 2014, 71)؛ وفي ضوء ما ذكره Schmitt & Dorans (1988, 1) يتضح أن دراسات الأداء التفاضلي لها هدفين أساسين الأول منهما يتمثل في الكشف عن البنود التي لها أداء تفاضلي والثاني يتمثل في تحديد الأسباب التي قد تكون كامنة خلف ذلك؛ وهذه الأسباب غالباً ما تتعلق بمحتوى وصياغة البند وتعليمات التطبيق والسياق الذي يتم فيه التطبيق والعوامل الثقافية والخبرة واللغة (Stump et al., 2005, 373; Meade & Fetzer, 2009, 738; Tay, Huang & Vermunt, 2016, 37)؛ وهو ما تم مراعاته في تفسير نتائج الدراسة الحالية حيث كان هناك اهتمام بتحديد الأسباب المحتملة للأداء التفاضلي للبنود.

ومما سبق يمكن استنتاج أهمية الكشف عن البنود التي لها أداء تفاضلي وما قد يترتب على ذلك من مشكلات في عملية القياس منها:

-قد يكون الأداء التفاضلي دليل على تحيز البنود وهو ما يسهم في عدم عدالة القياس بشكل كبير.

-يؤثر الأداء التفاضلي للبنود على خصائص الاختبار السيكومترية (الصدق، الثبات، الموضوعية).

-يؤثر الأداء التفاضلي للبنود سلباً على الصدق التنبؤي للاختبار.

ويمكن الكشف عن الأداء التفاضلي للبنود في ضوء نظرية الاستجابة للمفردة باستخدام النماذج البارامترية اعتماداً على البرامج الإحصائية المستخدمة في هذا الصدد كبرنامج BILOG-MG في حالة البنود ثنائية الاستجابة وبرنامج MULTILOG-MG في حالة البنود متعددة الاستجابات كذلك يمكن الاعتماد على استخدام النماذج اللابارامترية اعتماداً على البرامج الإحصائية المستخدمة في هذا الصدد كبرنامج TestGraf كما أنه يمكن الكشف عن الأداء التفاضلي للاختبار باستخدام التحليل العاملي التوكيدي متعدد المجموعات اعتماداً على البرامج الإحصائية المستخدمة في هذا الصدد كبرنامج AMOS/و Lisrel؛ (Stump et al., 2005, 362; Cho, Martin, Conger & Widaman, 2010, 158; Pae & Park, 2006, 477) وهو ما سيتم عرضه فيما يأتي.

ثانياً: الأداء التفاضلي باستخدام النموذج البارامتري ثنائي المعلم:

ظهرت نظرية الاستجابة للمفردة الاختبارية كرد فعل للنقد المتزايد لنظرية القياس الكلاسيكية، حيث ظهرت الحاجة إلى نموذج قياس تستقل فيه تقديرات بارامترات بنود الاختبار عن المختبرين، وكذلك تحديد قدرات الأفراد باستقلالية عن خصائص البنود، وهناك عدة افتراضات أساسية للنظرية ينبغي تحققها حتى يمكن تطبيق أحد نماذج النظرية وحتى تكون النتائج أكثر مصداقية ومن أهم هذه الافتراضات:

1- أحادية البعد Unidimensional: ويعني أن احتمال الاستجابة للبند يعزا لسمة كامنة واحدة مسيطرة؛ وهنا يتم الاسترشاد بما أشار إليه Reckase في 1979 فيكفي وجود عامل أساسي مسيطر على الأداء، والذي يمكن التأكد منه باستخدام التحليل العاملي ووجود عامل أساسي يفسر نسبة كبيرة من التباين في الأداء أو من خلال كبر النسبة بين قيمة الجذر الكامن للعامل الأول والعامل التالي له (Baghi & Ferrara, 1989, 7).

2- الاستقلال الموضعي Local Independence: ويعني أن احتمال الإجابة الصحيحة للبند لا يتأثر بالإجابة على باقي البنود، بمعنى الاستقلال الإحصائي للاستجابة على البنود عند مستوى معين من السمة*(Kline, 2004, 552)*؛ ويمكن اعتبار أن تحقق أحادية البعد يتحقق معه بالتبعية شرط الاستقلال الموضعي، ولذلك يطلق على هذا الافتراض أحياناً مسمى الافتراض الخفي (Molenaar, 2003, 192)؛ ولا يعني افتراض الاستقلال الموضعي أن تكون المفردات غير مرتبطة وإنما يمكن الحصول على ارتباط موجب بين أزواج المفردات حيثما كان هناك تباين في القدرة التي تقيسها هذه المفردات، ولكي يتحقق افتراض الاستقلال الموضعي يجب أن يكون نمط استجابة الفرد مساوياً لحاصل ضرب احتمالات حصوله على درجة معينة في كل بند من البنود ويتحقق ذلك إذا كانت البنود تقيس سمة كامنة واحدة (علام، 2005، 63-64).

3- المنحنى المميز للبند أو منحنى خصائص البند Item Characteristic Curve: يمثل هذا المنحنى العلاقة بين احتمالية الإجابة الصحيحة للبند والسمة الكامنة، ويعبر عنه من خلال دالة الترجيح اللوغاريتمي والتي يستخدم فيها عدد من البارامترات تصف دالة الاستجابة للبند Item Response Function، والمنحنى المميز للبند هو اللبنة الأساسية في نظرية الاستجابة للمفردة (Baker, 2001, 7; Lee, 2007, 121).

وفي ضوء الاتجاهات المختلفة في تقدير المنحني المميز للبند، يمكن تصنيف نماذج نظرية الاستجابة للمفردة إلى نماذج بارامترية Parametric Item Response Theory (PIRT) تستند إلى دالة الترجيح اللوغاريتمي أو دالة التوزيع الاعتدالي التراكمي في تقدير المنحنى المميز للبند ونماذج

لابارامترية Nonparametric Item Response Theory (NIRT) لا يفترض فيها مسبقاً أي شكل معين للمنحنى (Lee, Wollack & Douglas, 2009, 182; Takano, Tsunoda & Muraki, 2014, 1).

ففي محاولة دراسة العلاقة الاحتمالية بين مستوى السمة الكامنة لدى الفرد ومعالم البند المتمثلة في صعوبة البند وقدرته على التمييز بين المستويات المختلفة من السمة وإمكانية التخمين في الوصول للإجابة الصحيحة والتي يرمز لها على الترتيب بالرموز $(c,\ a,\ b)$، ظهرت النماذج الرياضية الاحتمالية في نظرية الاستجابة للمفردة وهي عبارة عن دوال لوغاريتمية لاحتمال الوصول للإجابة الصحيحة للبند، وفي حالة النموذج البارامتري ثلاثي المعلم وبافتراض أن θ تمثل السمة الكامنة فإن احتمال الإجابة الصحيحة على البند (i) يمكن التعبير عنها بدالة الترجيح اللوغاريتمي التالية (Price, 1999, 6; Kang & Cohen, 2007, 331; Ozdemir, 2015, 2076):

$$P_i(\theta) = c_i + (1 - c_i)\ \frac{e^{Da_i(\theta - b_i)}}{1 + e^{Da_i(\theta - b_i)}}$$

حيث أن $(D = 1.702)$ تمثل عامل التدريج والذي يجعل قيم الدالة اللوغاريتمية تقترب من دالة التوزيع الاعتدالي التراكمي، وفي النموذج البارامتري ثنائي المعلم والذي يفترض فيه أن إمكانية التخمين تساوي صفر (Lee, Wollack & Douglas, 2004, 9; Lee, 2007, 126; Bechger & Maris, 2015, 335) تكون الدالة اللوغاريتمية لاحتمال الإجابة الصحيحة للبند كالتالي:

$$P_i(\theta) = \frac{e^{Da_i(\theta - b_i)}}{1 + e^{Da_i(\theta - b_i)}}$$

ولذا يعد النموذج البارامتري ثنائي المعلم حالة خاصة من النموذج الثلاثي وفيه قيمة معلم التخمين (c) تساوي صفر لجميع البنود، وفي حالة اعتبار أن جميع البنود تميز بنفس القدر بين المختبرين $(a=1)$ فإنه يمكن استنتاج معادلة نموذج راش أحادي المعلم (Santelices & Wilson, 2012, 11).

ومن الفوائد الأساسية والتي تميز نظرية الاستجابة للمفردة عن النظرية الكلاسيكية في القياس ما يعرف بعدم تغير القيم التقديرية للبارامترات والذي يعبر عنه البعض بتحرير تقدير السمة Sample-Freed وتحرير تقدير بارامترات البنود Item-Freed، فبارامترات البند تستقل إحصائياً عن توزيع السمة وكذلك بارامترات الأفراد تستقل إحصائياً عن توزيع البنود (Pastsula & Gessaroli, 1995, 5; Luciano, Ayuso-Mateos, Aguado, Fernandez, Serrano-Blanco, Roca & Haro, 2010, 2)، وهو ما يعني إمكانية تغير توزيع قدرات الأفراد دون أي تأثير على المنحنى المميز للبند، وبالنسبة لتوزيع البنود يمكن تغير توزيع البنود في الاختبار دون أي تأثير على قدرات المختبرين؛ وما سبق يسهم في تميز طرق الكشف عن الأداء التفاضلي في ضوء نظرية الاستجابة للمفردة عن غيرها، حيث تستند المقارنة بين المجموعات هنا على ما يسميه علام (2005، 63) مقارنات تتصف بعدم التغير في البارامترات المقدرة؛ وبالتالي فإن عدم تطابق معالم البند في المجموعات المختلفة عند مستوى معين من السمة يعني أن احتمالات الإجابة على البند لا تكون متساوية لتلك المجموعات، وبالتالي فإن البند يؤدي بطريقة مختلفة في كل مجموعة.

وفي الدراسة الحالية تم الاعتماد على اختبار النسبة الاحتمالية IRT-LRT باستخدام النموذج البارامتري ثنائي المعلم في الكشف عن البنود التي لها أداء تفاضلي، حيث يتطلب تطبيق النموذج البارامتري ثلاثي المعلم عينات كبيرة عن العينة المتوفرة والمطبق عليها الاختبار في الدراسة الحالية؛ ويعد اختبار النسبة الاحتمالية من أكثر الطرق استخداماً للكشف عن الأداء التفاضلي للبنود في إطار نظرية الاستجابة للمفردة (Kim & Oshima, 2012, 459; Santelices & Wilson, 2012, 8,9; Wang, Tay, & Drasgow, 2013, 317؛

اختبار النسبة الاحتمالية Likelihood Ratio Test (LRT) لمقارنة معالم البند:

في ضوء فكرة أن نقاط المنحنى المميز للبند عند كل مستوى من مستويات السمة تعبر عن الاحتمال المشروط للاستجابة الصحيحة عند قيمة معينة للسمة وفي حالة الأداء المختلف للبند في المجموعة البؤرية مقارنة بالمجموعة المرجعية فإن البند يكون له أداءً تفاضلياً، والعكس صحيح، ونظراً لاعتماد المنحنى المميز للبند على معالم البند فإنه يمكن الكشف عن الأداء التفاضلي للبند في ضوء ملاحظة Lord في 1977، 1980 بأنه من الممكن معالجة هذه القضية بمقارنة معالم البند في المجموعتين (Scientific Software International, 2015, 99)؛ ويستند ذلك على فكرة أن تماثل المنحنى المميز للبند لأية مجموعتين من المختبرين يحدث فقط في حالة تماثل معالم البند في المجموعتين، وهنا يتم اختبار مدى صحة فرضية "تساوي معالم البند بالنسبة للمجموعات المختلفة التي يقدّم لها الاختبار" (Wells, Cohen & Patton, 2009, 311; Woods, 2009, 43)، وفي حالة رفض هذه الفرضية يكون للبند أداءً تفاضلياً، ففي حالة النموذج ثنائي المعلم يكون الفرض الصفري الذي يتم اختباره لاكتشاف البنود ذات الأداء التفاضلي باعتبار أن المجموعتين البؤرية والمرجعية هما (R), (F) وأن صعوبة البند وتمييزه هما A, B على النحو التالي:

$$H_0: A_R = A_F, B_R = B_F$$

وفكرة الاعتماد على النسبة الاحتمالية في الكشف عن دلالة الفرق بين معلميّ البند في المجموعتين البؤرية والمرجعية نبعت من أفكار لورد *Lord* السابقة واستخدامه لمربع كاي في المقارنة بين معالم البند للمجموعات المختلفة (Bechger & Maris, 2015, 319)؛ وهنا يتم الاعتماد على اختبار النسبة الاحتمالية في اختبار مدى صحة الفرض الصفري السابق؛ وتتميز هذه الطريقة بأنها تختبر الأداء التفاضلي للبند بطريقة مباشرة كما تُمكن من اكتشاف الأداء التفاضلي المرتبط بصعوبة البند والتمييز ونسبة التخمين، كذلك تعبر عن الأداء التفاضلي بوحدات كمية واضحة المعنى، كما أنها لا تعتمد على المجموع الكلي للدرجات (Thissen, 2001, 2-4).

ويعد اختبار النسبة الاحتمالية LRT من الطرق شائعة الاستخدام للمقارنة بين النماذج المتداخلة Nested Models والتي يتم فيها اشتقاق نموذج من نموذج بافتراض قيود معينة، ويرمز للنسبة الاحتمالية بالرمز G^2 والتي يتم حسابها من خلال المعادلة:

$$G^2 = -2Log\left(\frac{Likelihood[A]}{Likelihood[B]}\right)$$

وهو ما يساوي:

$$G^2 = -2(LogLikelihood[A] - LogLikelihood[B])$$

حيث أن A, B هما النموذجان المتداخلان موضوع المقارنة، والمعادلة الأخيرة توضح أن النسبة الاحتمالية تمثل ضعف الفرق بين الترجيح اللوغاريتمي لجودة مطابقة كل نموذج بإشارة سالبة، حيث تعتبر قيمة الأرجحية Likelihood مؤشراً لجودة مطابقة البيانات للنموذج البارامتري في نظرية الاستجابة للمفردة ولتقريب وتبسيط الحسابات يتم حساب لوغاريتم Log هذه القيمة، وهذه القيمة الأخيرة سالبة دائماً ولذلك يتم ضربها في إشارة سالبة، وضعف الناتج يأخذ توزيع مربع كاي، عند درجة حرية تساوي الفرق بين عدد البارامترات التي يتم تقديرها في كل نموذج من النموذجين (Wang et al., 2013, 319; Whittaker, Chang & Dodd, 2012, 162).

واستخدام النسبة الاحتمالية للكشف عن البنود ذات الأداء التفاضلي يتم في عدد من الخطوات كالآتي: (Thissen, 2001, 7-11; Kang & Cohen, 2007, 332; Finch & French, 2007, 569-570; Woods, 2008, 512; Rivas, Stark & Chernyshenko, 2009, 252; Cho et al., 2010,159; Meade & Wright, 2012, 1017; Wang et al., 2013, 319)

الخطوة الأولى: حساب مؤشر جودة المطابقة للنموذج المقيد بتساوي بارامترات جميع البنود في المجموعتين (البؤرية، والمرجعية)، وينتج من هذه الخطوة قيمة الترجيح اللوغاريتمي لتساوي بارامترات جميع البنود في المجموعتين Log likelihood All Equal وتعد هذه القيمة بمثابة القيمة المعيارية لكل المقارنات التالية.

الخطوة الثانية: بعد تحديد البنود التي سيتم الاعتماد عليها كجزع مشترك لوضع بارامترات البنود في المجموعتين على نفس التدريج والتي تم تحديدها في الدراسة الحالية بأنها "كل البنود ماعدا البند" all of the other items ويتم حساب مؤشر جودة المطابقة للنموذج المقيد بتساوي بارامترات جميع البنود ماعدا البند موضوع الاهتمام وذلك لمقارنة بارامترات البند في المجموعتين، وينتج عن هذه الخطوة قيمة الترجيح اللوغاريتمي لتساوي بارامترات جميع البنود ماعدا البند (I) موضوع الاهتمام Log Likelihood Item (I) not equal.

الخطوة الثالثة: يتم فيها حساب قيمة الفرق بين الترجيح اللوغاريتمي في الخطوتين السابقتين (*-2log likelihood*) والذي يأخذ توزيع مربع كاي كما في المعادلة التالية:

$$G^2(total) = -2(Log\ likelihood\ All\ Equal - Log\ Likelihood\ Item\ (I)\ not\ equal)$$

وهذه القيمة هي التي يتم تقييم الأداء التفاضلي الكلي للبند بناءً على دلالتها وذلك عند درجة حرية تساوي الفرق بين عدد البارامترات التي يتم تقديرها في كل من النموذجين (2 في حالة النموذج البارامتري ثنائي المعلم) اعتماداً على فكرة أن النموذج الأقل تقييداً سيكون أفضل في جودة المطابقة، وفي حالة دلالة قيمة مربع كاي للأداء التفاضلي الكلي للبند يتم حساب قيمة الترجيح اللوغاريتمي لجودة مطابقة النموذج الذي يُفترض فيه تساوي بارامترات جميع البنود في المجموعتين البؤرية والمرجعية ما عدا البند موضوع التحليل (*I*) وبالنسبة لهذا البند وفي حالة النموذج ثنائي المعلم يتم وضع قيد بتساوي معلم التمييز (a_1) في المجموعتين ثم يتم الآتي:

–للكشف عن الأداء التفاضلي لمعلم التمييز (اختلاف تمييز البند في المجموعتين) يتم حساب:

$$G^2(a_1) = -2(Log\ likelihood\ Item\ (a_1)\ Equal - Log\ Likelihood\ Item\ (I)\ not\ equal)$$

–للكشف عن الأداء التفاضلي لمعلم الصعوبة b المشروط بتساوي معلم التمييز (اختلاف صعوبة البند في المجموعتين) يتم حساب:

$$G^2(b/a) = -2(Log\ likelihood\ All\ Equal - Log\ likelihood\ Item\ (a_1)\ Equal)$$

ويتم الحكم على الأداء التفاضلي بالنسبة لمعلم الصعوبة أو معلم التمييز في ضوء دلالة مربع كاي في الحالتين الأخيرتين عند درجة حرية تساوي (1) في حالة النموذج ثنائي المعلم، وفي حالة دلالة مربع كاي في أي حالة من الحالتين السابقتين يتم الحكم بوجود أداء تفاضلي للبند سواء فيما يتعلق بمعلم الصعوبة أو معلم التمييز أو الاثنين معاً (Whittaker et al., 2012, 162)؛ ومن الواضح أن الأداء التفاضلي الكلي للبند يتحدد بالأداء التفاضلي لمعلمي البند (الصعوبة، التمييز) والذي يعّبر عنه كالتالي:

$$G^2(total) = G^2(a) + G^2(b/a)$$

ثالثاً: الأداء التفاضلي باستخدام نموذج الانحدار اللابارامتري الممهد:

يعد نموذج الاستجابة للمفردة نموذجاً لابارامترياً في حالة عدم وجود نموذج رياضي يحدد شكل دالة الاستجابة للبند مسبقاً وعدم وجود افتراضات مسبقة تتعلق بتوزيع السمة الكامنة مع تحقق باقي افتراضات نظرية الاستجابة للمفردة الاختبارية المتمثلة في أحادية البعد، والاستقلال الموضعي، والعلاقة التزايدية المطردة بين احتمال النجاح في الاستجابة للبند والسمة الكامنة (Sijtsma & Molenaar, 1987, 79; Sijtsma, 2003, 184).

وانبثقت فكرة النماذج اللابارامترية من أن النماذج البارامترية اللوغاريتمية في تقديرها للمنحنى المميز للبند لا تتسم بالمرونة الكافية نظراً لافتراضها لشكل معين مسبق لهذا المنحنى (*Douglas,*

(Keller, 2000, 19; Lee, 2007, 121; Zheng, Gierl & Cui, 2010, 542)؛ ولذا يعتقد العديد من الباحثين في مجال القياس النفسي أن المنحنى المميز للبند لا يتم نمذجته دائماً بطريقة جيدة إذا اعتمدنا على تلك النماذج البارامترية (1997, 7)؛ فالنماذج البارامترية شائعة الاستخدام نظراً لسهولة التعامل مع دالة الترجيح اللوغاريتمي، ومع ذلك ربما لا تؤدي للوصول إلى الصورة الصحيحة لدالة الاستجابة للبند (Duncan & MacEachern, 2008, 45) .

ومشكلة النماذج البارامترية في نظرية الاستجابة للمفردة أنها تتقيد بشكل معين مفترض مسبقاً لتوزيع السمة، وبناء على ذلك يتم تدريج السمة، ومن هنا ظهرت النماذج اللابارامترية في نظرية الاستجابة للمفردة والتي لا يتم التقيد فيها بشكل معين لتوزيع السمة أو المنحنى المميز للبند، وإنما يتم استخدام الانحدار اللابارامتري في تدريج السمة أولاً ثم تقدير المنحنى المميز للبنود بناء على ذلك التدريج مما يجعل شكل المنحنى معبّراً فعلاً عن دالة الاستجابة للبند، ويسهم في التغلب على مشكلة التقيد بشكل مسبق لتوزيع السمة في النماذج البارامترية.

فدالة الاستجابة للبند في النماذج البارامترية دالة تزايديه بالنسبة للسمة وشكلها محدد ويتخذ شكل دالة الترجيح اللوغاريتمي أو دالة التوزيع الاعتدالي التراكمي (Baker, 2001, 21; Wang, 2012, 2-4)؛ بينما النماذج اللابارامترية لا تتقيد بشكل معين لهذه الدالة وهو ما يوفر لهذه النماذج المرونة الكافية لملاءمة النموذج للبيانات موضوع المعالجة (Molenaar, 2003, 191)؛ ولا يعنى ذلك أن النماذج اللابارامترية لا تفترض أن المنحنى المميز للبند دالة ليست تزايديه، ولكن تتميز بقبولها لأي شكل لهذه الدالة، وبالتالي تتوافر معلومات ذات قيمة عن علاقة الأداء على البند والسمة بينما النماذج البارامترية تتقيد بهذا الشرط وفي حالة الإخلال به يعد البند غير ملاءم للنموذج (Lee, 2007, 122)؛ فالفائدة من النموذج البارامتري لا تتحقق في حالة عدم التوافق بين الشكل المتوقع للمنحنى المميز للبند والبيانات موضوع المعالجة (Sueiro & Abad, 2011, 835).

ويتميز الاعتماد على النماذج اللابارامترية لنظرية الاستجابة للمفردة بالبساطة في التطبيق والتفسير، وعدم تغير التقديرات حتى في حالة صغر حجم العينات (Penfield, Giacobbi & Myers, 2007, 454; Ownby & Waldrop-Valverde, 2013, 2)؛ ومن أشهر النماذج اللابارامترية في تقدير المنحنى المميز للبند Nonparametric Estimation of ICC وأكثرها استخداماً ما توصل إليه Ramsay في 1991 وتبنيه لفكرة الانحدار اللابارامتري في تدريج السمة ثم تقدير المنحنى المميز للبند وانتشر النموذج تحت مسمى الانحدار اللابارامتري الممهد Nonparametric Smoothing regression (Ramsay, 1991, 611-630;1998, 367; Ramsay & Silverman, 1997, 51-56; Molenaar , 2001, 296; Lee et al., 2004, 5; Lee, 2007, 122; Sueiro & Abad, 2011, 835; Takano et al., 2014, 2).

وكما يذكر Ramsay (2000, 58-60); Xu, Douglas & Lee (2011, 244-245); Zheng et al., (2010, 544); Witarsa (2003, 19)؛ أنه بافتراض عينة عشوائية تتكون من N من المختبرين حيث ($j= 1, 2, 3, 4, \dots N$) وأن عدد بنود الاختبار n حيث ($i= 1, 2, 3, 4, \dots$) فإن تقدير المنحنى المميز للبند (i) في هذه الحالة $P_i(\theta)$ هو متوسط وزن متّجه الإجابة $\{Y_i\} = \{Y_{1i}, Y_{2i}, Y_{3i}, \dots Y_{Ni}\}$ وهو ما يمكن التعبير عنه بالمعادلة:

$$P_i(\theta) = \sum_{j=1}^{N} w_j Y_{ji}$$

حيث Y_{ji} تمثل استجابات الفرد (j) على البند (i) ، والأوزان w_j تحدد بطريقة معينة وتكون موجبة وتصل لأقصى قيمة لها عندما $\theta = \theta_j$ وتتجه لتساوي الصفر كلما زاد الفرق المطلق بين θ, θ_j والأوزان w_j لابد وأن تحقق شرطين هما w_j أكبر من الصفر ومجموعها يساوي (1)، ويمكن التعبير عما سبق بالمعادلة*:

* لكيفية اشتقاق المعادلة والأساس الرياضي والنظري لها يمكن الرجوع إلى (Douglas, 1997; Ramsay, 1991)

$$P_i(\theta) = \frac{\sum_{j=1}^{N} K\left(\frac{\theta_j - \theta}{h}\right) Y_{ji}}{\sum_{j=1}^{N} K\left(\frac{\theta_j - \theta}{h}\right)}$$

حيث θ نقطة على متصل السمة الكامنة والتي تعبر عن المستوى الحقيقي لقدرة الفرد و θ_j هو تقدير السمة للفرد (j) بناءً على أدائه على البنود، K تعرف بمقدر الوسط الموضعي Local Averaging Estimator أو الكيرنل Kernel وهو طريقة من طرق التقدير بحيث يتوافق الوسط الموضعي مع البيانات ويتميز بدرجة من التمهيد Smoothing بحيث يعطى معظم الأوزان للمختبرين الذين تقترب الدرجات المشاهدة لهم من مستوى السمة ويقلل من الأوزان للمختبرين ممن تبتعد القيم المشاهدة لهم عن مستوى السمة، بينما يعرف h بالنطاق المتّسع Bandwidth أو بارامتر التمهيد Smoothing Parameter والذي يؤثر في توزيع الأوزان ويضبط تمهيد المنحنى (Ramsay, 1998, 369; Wang, 2012, 15; Meyer, 2014, 113; Guo & Sinharay, 2010, 3).

ويمكن توضيح فكرة الوسط الموضعي كالتالي: لنفرض أن (n) هي عدد بنود الاختبار وتم تقدير عدد (n) من القيم التي تُعبر عن قدرة الفرد (j) حيث أن $(\theta_j = \theta_1, \theta_2, \theta_3 \ldots\ldots \theta_n)$ وأن استجابات هذا الفرد على بنود الاختبار $(Y_i = y_1, y_2, y_3, \ldots\ldots\ldots\ldots y_n)$، وباعتبار أن تقديرات السمة هي المتغير المستقل والاستجابات هي المتغير التابع، وأن θ هي القيمة الحقيقية المستهدف تقديرها لقدرة الفرد والتي يمكن تقديرها في هذه الحالة بواسطة معادلة الانحدار اللابارامتري السابقة $P(\theta)$ حتى لو لم تكن القيمة المستهدفة من بين القيم الفعلية لتقديرات المتغير المستقل θ_j وذلك عن طريق حساب متوسط القيم Y_i المرتبطة بالقيم θ_j باعتبار أن قيم θ_j قريبة جدا من قيم θ الحقيقية المستهدفة وهذه الطريقة هي ما يعرف بالوسط الموضعي (Zheng et al., 2010, 543).

فنموذج الانحدار اللابارامتري الممهد - مثل باقي النماذج اللابارامترية - يعتمد على فكرة الوسط الموضعي حيث أن دالة الاستجابة للبند يمكن الحصول عليها من خلال المتوسطات الموزونة لمتغير الاستجابة Y لكل المختبرين ممن يقترب مستوى السمة لهم من المستوى (θ)، وتقديرات أوزان الاستجابة تتحدد بواسطة مُقدر الوسط الموضعي K في نموذج الانحدار اللابارامتري الممهد والذي يعد من أهم عوامل دقة تقدير المنحنى المميز للبند، وقيمته دائماً أكبر من أو تساوي الصفر؛ فبافتراض أن الفرق بين القيمة المشاهدة للقدرة والقيمة الحقيقية لها للفرد (j) يساوي (u) حيث $(u = \theta - \theta_j)$ فإن $K(u)$ تقترب من الصفر كلما انحرفت قيمة (u) عن الصفر وتساوي (1) في حالة كانت (u) تساوي صفر بمعنى أن أقصى قيمة يمكن أن يصل لها الوسط الموضعي هي $k(0)$ وغالباً ما يتم الاعتماد على دالة التوزيع الاعتدالي لتحديد الوسط الموضعي (Ramsay & Silverman, 1997, 51; Ferrando , 2004, 99; Meijer, 2004, 4;) ويعبر عنها كما يلي:

$$K(u) = \frac{1}{\sqrt{2\pi}} exp^{\left(\frac{-u^2}{2}\right)}$$

والوصول لأفضل منحنى يُعبر عن الدالة المميزة للبند من خلال النموذج يعتمد على اقتراب أوزان الاستجابة وعدم تشتتها ولتحقيق ذلك يجب اختيار بارامتر التمهيد h بدقة حتى لا يؤدي لتشتيت النقاط المقدرة للمنحنى؛ وهنا يتضح دور بارامتر التمهيد h والذي يسهم في ضبط التوازن بين تحيز التقديرات والتباين فيها (Xu et al., 2011, 245)؛ فكلما قلت قيمة بارامتر التمهيد يقل التحيز ويزداد التباين في تقديرات الدالة عند كل نقطة من المنحنى والعكس صحيح، وبصفة عامة يتم اختيار بارامتر التمهيد بحث يقلل من متوسط مربعات الخطأ إلى أقل قدر ممكن (Ramsay & Silverman, 1997, 51; Douglas & Cohen, 2001, 236)؛ ويفضل أن تكون قيمة بارامتر

التمهيد قريبة من ($N^{-0.2}$) والقيمة الافتراضية في برنامج TestGraf المستخدم في الدراسة الحالية لبارامتر التمهيد تساوي ($1.1N^{-0.2}$) (Ramsay, 2000, 64).

وتأكد من نتائج الدراسات السابقة فاعلية ودقة التقديرات اللابارامترية لدالة الاستجابة للبند وتقديرات السمة الكامنة بالاعتماد على نموذج تحليل الانحدار اللابارامتري الممهد للدرجات الكلية في الاختبار خاصة في حالة زيادة عدد بنود الاختبار (Pastsula & Gessaroli, 1995; Douglas & Cohen, 2001;Zheng et al., 2010)؛ وأن تلك التحليلات تتميز بالسرعة والدقة، وأثبتت التحليلات القائمة على المحاكاة كفاءة ودقة النموذج في تقدير المنحني المميز للبند (Ramsay, 1991; Keller, 2000; Ferrando, 2004; Lee et al., 2009; Sueiro & Abad, 2011; Gierl & Bolt, 2001; Luciano et al., 2010; Guo & Sinharay, 2010; Khan, Lewis & Lindenmayer, 2011).

-المساحة بين المنحنيات اللابارامترية المميزة للبند كمؤشر للأداء التفاضلي:

يتم هنا الاعتماد على حساب المساحة بين المنحنى اللابارامتري المميز للبند في المجموعة المرجعية وفي المجموعة البؤرية، وقد اقترح Ramsay (1991; 2000) مؤشر بيتا β للتعبير عن المساحة بين المنحنيين في المجموعتين المرجعية والبؤرية، وهذا المؤشر تعبر قيمته عن الفروق بين منحنيات الدرجات المتوقعة الموزونة Weighted Expected Grade Curves في كلا من المجموعة المرجعية والبؤرية من نفس مستوى السمة، ويتم حسابها باستخدام المعادلة التالية:

$$\beta_{Fim}(\theta) = \sum_{q=1}^{Q} P_{Fq}\left[P_{im}^{(F)}(\theta) - P_{im}^{(R)}(\theta)\right]$$

حيث أن P_{Fq} تعبر عن نسبة أفراد المجموعة البؤرية ممن كانت درجاتهم المشاهدة تساوي θ_q، وQ تعبر عن المستويات التي يمكن تقسيم الدرجات المشاهدة لها (يتم تقسيم المجموع الكلي للدرجات إلى فئات صغيرة وتحدد نسبة المختبرين ممن توصلوا للاستجابة في كل فئة)، بينما $P_{im}^{(R)}(\theta)$, $P_{im}^{(F)}(\theta)$ هي قيم المنحنى المميز للبند (i) في المجموعتين المرجعية والبؤرية، و(m) تعبر عن الاختيارات في حالة البنود متعددة الاستجابات حيث تتم المقارنة بين المساحات في جميع الاختيارات الممكنة كإجابات محتملة للسؤال (Gotzmann, 2002, 11)؛ ويمكن حساب الخطأ المعياري في تقدير قيمة β عن طريق التباين في حساب قيمها، والذي يحسب من العلاقة التالية باعتبار أن g تعبر عن المجموعة والتي ممكن أن تكون قيمتها صفر في حالة العينة الكلية:

$$Var[\beta_{img}] = \sum_{q=1}^{Q} P_{q0}^2\left\{Var\left[P_{img}(\theta_q) + Var[P_{im0}(\theta_q)]\right]\right\}$$

وفي هذه الحالة يكون الخطأ المعياري $S.E$ في تقدير β هو الجذر التربيعي للمقدار السابق، وبقسمة β على الخطأ المعياري في تقديرها نتوصل لمؤشر لاختبار الدلالة الإحصائية لقيمة β والذي يتبع التوزيع الاعتدالي المعياري (Z) وبذلك يمكن قبول أو رفض الفرض الصفري الخاص بالأداء التفاضلي للبند (Basokcu & Ogretmen, 2014, 33; Witarsa, 2003, 24)؛ وأكدت نتائج دراسة (Gotzmann, 2002) على انخفاض نسبة الخطأ من النوع الأول في الكشف عن الأداء التفاضلي بالاعتماد على تحليل الانحدار اللابارامتري الممهد باستخدام TestGraf حتى في حالة العينات صغيرة الحجم ووجود فروق في مستوى السمة بين المجموعات موضوع المقارنة.

ويمكن الاستناد إلى المحكات التي حددها Roussos & Stout كما يشير Tan & Gierl (2005, 12) في تحديد حجم الأداء التفاضلي للبند في حالة دلالة β حيث يعد البند له أداء تفاضلي ضعيف إذا كانت قيمة β المطلقة أقل من 0.059 وله أداء تفاضلي متوسط إذا كانت القيمة أكبر من أو تساوي 0.059 وأقل من 0.088 وله أداء تفاضلي كبير إذا كانت القيمة أكبر من أو تساوي

0.088؛ وأشارت نتائج دراسة (Witarsa, 2003) أنه في حالة انخفاض حجم أحد المجموعتين عن 200 يمكن الاعتماد على القيمة 0.069 كدرجة فاصلة للحكم بأن البند له أداءً تفاضلياً أم لا، حيث أكدت النتائج على أن الاعتماد على القيمة 0.059 كمحك للحكم على البند بأن له أداءً تفاضلياً أسهم في زيادة الخطأ من النوع الأول، وهو ما تم مراعاته في الدراسة الحالية.

بينما أكدت دراسة (Zheng et al., 2010) أنه عند الاعتماد على التوزيع الاعتدالي المعياري في الحكم على الأداء التفاضلي للبنود في حالة استخدام نموذج الانحدار اللابارامتري الممهد يفضل أن يكون هناك تقارب في مستوى السمة بين المجموعات موضوع المقارنة خاصة في حالة انخفاض حجم العينة حيث أن عدم التكافؤ في توزيع السمة للمجموعتين المرجعية والبؤرية يؤدي إلى تضخيم نسبة الخطأ من النوع الأول في هذه الحالة؛ وهو ما تم مراعاته في الدراسة الحالية عند استخدام هذا النموذج.

واستخدمت الإجراءات السابقة في دراسة الأداء التفاضلي لقائمة قياس الاضطراب والمصممة بواسطة منظمة الصحة العالمية في ضوء النوع (Luciano et al., 2010)؛ وفي دراسة الأداء التفاضلي في تحصيل اللغة الإنجليزية في ضوء النوع (Gierl & Bolt, 2001)؛ وفي دراسة الأداء التفاضلي لاختبار الفهم القرائي كبُعد من أبعاد اختبار محو الأمية الصحية في ضوء السن (Ownby et al., 2013)؛ وفي دراسة (Nay et al., 2007) للأداء التفاضلي لقائمة الخوف والقلق الاجتماعي في ضوء النوع.

رابعاً: الأداء التفاضلي للاختبار DTF باستخدام التحليل العاملي التوكيدي متعدد المجموعات:

يعد التحليل العاملي التوكيدي متعدد المجموعات Multiple Group Confirmatory Factor Analysis من أكثر الطرق شيوعاً في الكشف عن عدم تغير نموذج القياس على مستوى الاختبار ككل باختلاف المجموعات، وهنا يكون التركيز على مدى تشابه علاقة المتغيرات الملاحظة (المقاسة) والمتغير أو المتغيرات الكامنة في المجموعات موضوع المقارنة، وبمعنى أدق مدى تشابه الدرجات المتوقعة في الاختبار للأفراد من نفس مستوى المتغير الكامن ومن مجموعات فرعية مختلفة (Abbott, 2007,12; Teresi & Fleishman, 2007, 37; Koh & Zumbo, 2008, 471; Milfont & Fischer, 2010, 112, Gomes, 2012, 276)؛ ففي حالة الكشف عن عدم تغير تقديرات بارامترات البنود في المجموعات المختلفة يمكن استخدام نظرية الاستجابة للمفردة الاختبارية بنماذجها المختلفة، وفي حالة الكشف عن عدم تغير تقديرات البارامترات على مستوى الدرجة الكلية أو درجات الأبعاد يفيد جداً في هذا الشأن التحليل العاملي التوكيدي متعدد المجموعات، وفي هذه الحالة وفي حالة البنود ذات الاستجابات الثنائية يفضل التعامل مع درجات الأبعاد باعتبارها متغير متصل عن التعامل مع درجات البنود خاصة في حالة استخدام طريقة الأرجحية القصوى في التحليل (Raju, Laffittr & Byrne, 2002, 517, 524).

وأكثر البارامترات التي يتم الكشف عن مدى تماثلها في المجموعات عند استخدام التحليل العاملي التوكيدي متعدد المجموعات هي تشبعات العوامل، وتباين وتغاير العوامل، وأخطاء القياس أو البواقي في المجموعتين، ويتم ذلك وفق عدة خطوات تتمثل في تحديد مؤشرات جودة المطابقة لعدد من النماذج المختلفة في درجة القيود المفترضة بتماثل بارامترات نموذج القياس في المجموعتين.

وفي الدراسة الحالية تمت المقارنة بين أربعة نماذج متداخلة Nested Models تتمثل في النموذج المعياري غير المقيد Unconstrained والذي لا توضع فيه أي قيود بتماثل بارامترات القياس في المجموعتين باعتباره النموذج المعياري للمقارنة Baseline for Comparison، والنموذج المقيد بتماثل التشبعات (أوزان القياس) Measurement Weights، والنموذج المقيد بتماثل التشبعات وتماثل التغاير Covariance of the Latent Variable، والنموذج المقيد بتماثل التشبعات وتماثل التباين والتغاير وتماثل البواقي Measurement Residuals، وتختلف تلك النماذج فيما بينها في ضوء تعقيد النموذج وعدد القيود لتنتهي بنموذج التماثل التام والذي يفترض

أكبر عدد من القيود لتماثل نموذج القياس في المجموعتين؛ ويتم الحكم على عدم تغير نموذج القياس في المجموعتين بناءً على دلالة الفرق بين قيمة مربع كاي في النماذج المتداخلة موضوع المقارنة وفي حالة عدم دلالة الفرق بين قيمتي مربع كاي عند درجة حرية تساوي الفرق بين درجتي حرية النموذجين يمكن الحكم بعدم تغير تقديرات بارامترات القياس في المجموعتين موضوع المقارنة (Kline, 2005, 295-296; Koh & Zumbo, 2008, 473; Byrne, 2010, 221, Milfont & Fischer, 2010, 117,118).

وفي ضوء ما أكده Church et al. (2011, 1073) من تأثر قيمة مربع كاي للفرق بين النماذج المتداخلة بحجم العينة؛ وما توصلت إليه نتائج دراسة (Cheung & Rensvold, 2002, 251) سيتم كذلك الاعتماد على الفرق بين مؤشر المطابقة المقارنة CFI في تقييم تماثل نموذج القياس في المجموعتين البؤرية، حيث أثبتت نتائج تلك الدراسة أن هذا الفرق له من المنعة والضلاعة Robustness ما يجعله لا يتأثر بحجم العينة كما في حالة الفرق بين قيمتي مربع كاي ولا يتأثر كذلك بجودة المطابقة الكلية للنماذج موضوع المقارنة، وإذا كانت قيمة هذا الفرق أقل من أو تساوي (0.01) يمكن الحكم بتماثل نموذج القياس في المجموعات موضوع المقارنة.

وكل نموذج من النماذج المقيدة الثلاثة –النموذج المعياري غير مقيد- له أهمية خاصة في توضيح مدى تماثل نموذج القياس، فالنموذج الثاني المقيد بتساوي التشبعات الهدف منه هو الكشف عن عدم تغير ميزان القياس Metric Invariance في المجموعتين، حيث يتم التأكد من تشبع العوامل على نفس البنية الكامنة ومدى اختلاف تشبعات العوامل، فمن الممكن أن تتشبع البنود أو الأبعاد على نفس البنية الكامنة وعلى الرغم من ذلك تختلف قوة العلاقة بين البنود أو الأبعاد والبنية الكامنة وهو ما يدل على اختلاف تمثيل البنود أو الأبعاد للبنية الكامنة ولذلك أهمية كبيرة في دلالة الفروق بين المجموعات في الدرجات الملاحظة للبنية الكامنة موضوع القياس، وفي النموذج الثالث يكون الهدف هو الكشف عن عدم تغير العلاقات المتبادلة بين متغيرات النموذج أو ما يعرف بتماثل مصفوفة التغاير في المجموعتين، بينما في النموذج الرابع يكون الهدف هو الكشف عن عدم تغير البواقي في نموذج القياس للمجموعتين والذي يعبر عن مدى قياس البند أو البعد للبنية الكامنة بنفس القدر من خطأ القياس، فعند استخدام درجات الأبعاد الفرعية للمقياس كمؤشرات للبنية الكامنة موضوع الاهتمام يكون من الأمور الجوهرية الكشف عن عدم تغير البواقي باعتبار أن تباين الدرجات الملاحظة هو مجموع تباين الدرجات الحقيقية وتباين الخطأ (Raju et al., 2002, 519; Schulz, 2005, 6;Wolf et al., 2014, 4; Cheung & Rensvold, 2002, 235-237; Church et al., 2011, 1069)؛ والإجراءات السابقة تستخدم في الكشف عن الاختلافات الممكنة في بارامترات نموذج القياس المفترض لقياس المتغيرات الكامنة بواسطة المقاييس والاختبارات النفسية وتسهم في الكشف عن مدى تغير البارامترات من مجموعة لمجموعة أخرى.

وفي حالة جودة مطابقة النموذج المعياري غير المقيد فإن المقارنة بين النموذج المعياري والنموذج المقيد بتساوي التشبعات، تعد أهم المقارنات الثنائية بين النماذج المتداخلة، فإذا ثبت عدم تماثل التشبعات في المجموعتين لا يكون للعبارات أو الأبعاد نفس المعنى في المجموعتين (Wolf et al., 2014, 5)؛ وعليه لا يكون للمقارنة بين المجموعتين في الدرجات الملاحظة في هذه الأبعاد أو العبارات معنى حقيقي، وهناك من يفترض أنه إذا ثبت عدم تماثل تشبعات عبارات أو أبعاد المقياس على المتغيرات الكامنة المفترضة في نموذج القياس لا يكون لباقي المقارنات الثنائية بين النماذج أهمية؛ فالبنية الكامنة غير صريحة وهي التي يتم المقارنة بين المجموعات فيها ولذا يجب أن يكون هناك عدم تغير في شكل علاقتها بالمؤشرات الصريحة لها في المجموعتين (Milfont & Fischer, 2010, 112)؛ وأكدت العديد من الدراسات السابقة فاعلية استخدام التحليل العاملي التوكيدي متعدد المجموعات في الكشف عن عدم تغير نموذج القياس عبر مجموعات مختلفة سواء على مستوى البنود أو الأبعاد الفرعية (Schuls, 2005; Abbott, 2007; Church et al., 2011; Raju et al., 2002).

إجراءات الدراسة:

الأدوات:

استُخدم في الدراسة الحالية اختبار واطسون-جليسر للتفكير الناقد "الصورة القصيرة" (WGCT-SF)، فقد ظهرت أول نسخة من هذا الاختبار في عام 1964 وتكونت من صورتين متكافئتين هما الصورة ZM والصورة YM وتكونت كل منهما من 100 سؤال تقيس خمسة مهارات فرعية للتفكير الناقد، ثم صدرت النسخة المعدّلة من الاختبار في عام 1980 وتم فيها تعديل صياغة بعض الأسئلة وحذف بعضها وبذلك تكونت النسخة المطوّرة من صورتين متكافئتين هما الصورة *A* والصورة *B* وتكونت كل منهما من 80 سؤالاً، وفي عام 1994 صدرت النسخة القصيرة من الاختبار اعتماداً على تطوير الصورة *A* وتكونت من 16 سيناريو يتبعها 40 سؤالاً تستغرق الإجابة عليها حوالي 30 دقيقة وقد تُرجم الاختبار إلى العربية في دراسة (العتيبي، 2012)؛ وفي دراسة (المبدل، 2010) وفي دراسة (السكري، 2010) مع اختلاف صياغة بعض السيناريوهات في الاختبار؛ وفي الدراسة الحالية تم استخدام النسخة المترجمة من قبل (العتيبي، 2012)؛ وتتوزع عبارات الاختبار على خمس مهارات فرعية للتفكير الناقد هي:

الاستنتاج Inference**:** ويشير إلى القدرة على استخلاص نتيجة من حقائق معينة وإدراك صحة النتيجة أو خطئها في ضوء الحقائق المعطاة، ويتكون من موقفين يُفترض صحتهما، ويلي كل منهما أربعة استنتاجات مرتبطة بالموقف السابق ويطلب من المفحوص الحكم على كل استنتاج إذا ما كان صحيحاً أم خاطئاً، وعبارات هذا البعد هي العبارات (1، 2، 3، 4، 5، 6، 7) وتتراوح درجته ما بين الدرجة صفر والدرجة 7.

تمييز الافتراضات Recognition of Assumptions*:* ويشير إلى القدرة على التمييز بين درجة صدق معلومات محددة أو عدم صدقها والتمييز بين الحقيقة والرأي والغرض من المعلومات المعطاة، ويتكون من ثلاث عبارات يتبع كلاً منها عدة افتراضات مقترحة وبعض هذه الافتراضات يتوافق مع الحقائق الواردة في العبارة وبعضها لا يتوافق مع الافتراضات الواردة في العبارة، وعبارات هذا البعد هي (8، 9، 10، 11، 12، 13، 14، 15) وتتراوح درجته ما بين الدرجة صفر والدرجة 8.

الاستنباط Deduction**:** ويشير إلى القدرة على تقرير ما إذا كانت استنتاجات معينة تترتب بالضرورة منطقياً على المعلومات الواردة في المقدمات أم لا، ويتكون من أربع عبارات يلي كلاً منها عدة نتائج مقترحة بعضها يترتب منطقياً على المقدمات وبعضها لا يترتب بالضرورة عليها، وإن كان صادقاً بوجه عام، وعبارات هذا البعد هي (16، 17، 18، 19، 20، 21، 22، 23، 24) وتتراوح درجته ما بين الدرجة صفر والدرجة 9.

التفسير Interpretation**:** ويعني القدرة على تحديد المشكلة، والتعرف على التفسيرات المنطقية وتقرير ما إذا كانت التعميمات والنتائج المبنية على معلومات معينة مقبولة أم لا، ويتكون من عبارتين يتبع كل منها عدة تفسيرات مقترحة بعضها يترتب منطقياً على العبارة وبعضها لا يترتب وإن كان صحيحاً بوجه عام، وعبارات هذا البعد هي (25، 26، 27، 28، 29، 30، 31) وتتراوح درجته ما بين الدرجة صفر والدرجة 7.

تقويم الحجج Evaluation of Arguments**:** ويعني القدرة على تقويم الفكرة وقبولها أو رفضها والتمييز بين المصادر الأساسية والثانوية والحجج القوية والضعيفة وإصدار الحكم على مدى كفاية المعلومات، ويتكون من خمس مسائل يتبع كل منها عدة حجج بعضها يمثل حججاً قوية وهامة تتصل مباشرة بالمسألة المطروحة وبعضها يمثل حججاً ضعيفة لا تتصل مباشرة بالمسألة المطروحة وإن كانت لها أهمية كبيرة، وعبارات هذا البعد هي (32، 33، 34، 35، 36، 37، 38، 39، 40) وتتراوح درجته ما بين الدرجة صفر والدرجة 9.

ويجيب الفرد على أسئلة البعد الأول بالاختيار من بين خمس بدائل (صادق تماماً، محتمل، بيانات ناقصة، محتمل خطؤه، خاطئ تماماً)، وأسئلة البعد الثاني بالاختيار من بديلين (وارد، غير

وارد)، وأسئلة البعد الثالث بالاختيار من بديلين (صحيحة، غير صحيحة)، وأسئلة البعد الرابع بالاختيار من بديلين (مترتبة، غير مترتبة)، وأسئلة البعد الخامس بالاختيار من بديلين (قوية، ضعيفة) وتتراوح الدرجة الكلية على الاختبار ما بين الدرجة صفر والدرجة 40.

وتم التأكد من صدق وثبات الاختبار في نسخته العربية في دراسة (العتيبي، 2012) وذلك بتطبيقه على 400 طالب بجامعة الملك سعود؛ وفي دراسة (المبدل، 2010) بتطبيقه على 84 طالب بالمرحلة الثانوية بمدينة الرياض؛ وفي دراسة (المالكي، 2012) بتطبيقه على 55 طالب بجامعة أم القري؛ وفي دراسة (السكري، 2010) بتطبيقه على 503 طالب بجامعة المنوفية بجمهورية مصر العربية.

مجتمع الدراسة وعينتها:

أ- مجتمع الدراسة: تمثل مجتمع الدراسة الحالية في جميع طلاب جامعة القصيم بجميع تخصصاتهم العلمية والأدبية ومستوياتهم الدراسية المختلفة، في الفصل الدراسي الثاني من العام الجامعي 1435/ 1436هـ، 2015 /2016م.

ب- عينة الدراسة: تم تطبيق اختبار واطسون وجليسر للتفكير الناقد على 405 طالب من الطلاب الذكور بجامعة القصيم بالمملكة العربية السعودية منهم 207 طالب بالتخصصات الأدبية و198 طالب بالتخصصات العلمية، وبعد فحص استجابات الطلاب وحذف الاستجابات التامة والصفرية لأن ذلك من شروط تطبيق نماذج نظرية الاستجابة للمفردة تم استبعاد 9 من طلاب التخصصات الأدبية كانت استجاباتهم صفرية (استجاباتهم خاطئة على جميع البنود) وبالتالي تكونت عينة الدراسة الحالية في صورتها النهائية من 396 طالب بالتخصصات العلمية والأدبية بالمستويات الأول والرابع والسابع والجدول التالي يوضح توزيع الطلاب المشاركون في الدراسة على الكليات والتخصصات المختلفة:

جدول (2): توزيع الطلاب عينة الدراسة في ضوء المستوى والتخصص

التخصص	الكلية	المستوى الأول	المستوى الرابع	المستوى السابع	المجموع	المجموع الكلي
أدبية	كلية الشريعة والدراسات الاسلامية	52	21	9	82	198
	كلية اللغة العربية والعلوم الاجتماعية	×××	10	21	31	
	كلية التربية	×××	33	52	85	
علمية	كلية العلوم	39	29	41	109	198
	كلية العلوم الطبية التطبيقية	21	23	9	53	
	كلية الحاسبات والمعلومات	11	10	15	36	
المجموع الكلي		129	126	147	396	

البرامج الإحصائية التي تم استخدامها:

-برنامج الحزمة الإحصائية في العلوم الاجتماعية SPSS V.21 وتم استخدامه في التأكد من الشروط اللازمة لتطبيق النموذج البارامتري ثنائي المعلم.

-برنامج AMOS V.21 وتم استخدامه للتأكد من عدم تغير نموذج القياس لدى طلاب التخصصات العلمية والأدبية بالاعتماد على التحليل العاملي التوكيدي متعدد المجموعات.

-برنامج IRTPRO تم استخدامه في التحليلات الخاصة بالنموذج البارامتري ثنائي المعلم 2PL وفي حساب المؤشرات التي في ضوئها الكشف عن الأداء التفاضلي لبنود الاختبار.

-برنامج TestGraf تم استخدامه في التحليلات الخاصة بنموذج تحليل الانحدار اللابارامتري الممهد.

حدود الدراسة: تتحدد نتائج الدراسة الحالية في ضوء التالي:

- **الحدود الموضوعية:** حيث تقتصر نتائج الدراسة الحالية على الكشف عن الأداء التفاضلي في ضوء متغير التخصص الأكاديمي باستخدام اختبار النسبة الاحتمالية بالاعتماد على النموذج البارامتري ثنائي المعلم، ونموذج Ramsay للانحدار اللابارامتري الممهد والمقارنة بين نتائج النموذجين، وتأثير الأداء التفاضلي للبنود على الأداء التفاضلي للاختبار ككل باستخدام التحليل العاملي التوكيدي متعدد المجموعات.
- **الحدود البشرية:** تتحدد نتائج الدراسة الحالية بالعينة التي تم تطبيق الاختبار عليها وهم الطلاب الذكور بجامعة القصيم بالمملكة العربية السعودية بالمستوى الأول والرابع والسابع من كليات (الشريعة والدراسات الإسلامية، اللغة العربية والدراسات الاجتماعية، التربية) كمجموعة بؤرية، وكليات (العلوم، العلوم الطبية التطبيقية، الحاسبات والمعلومات) كمجموعة مرجعية.
- **الحدود الزمنية:** تم تطبيق الاختبار في الفصل الدراسي الثاني من العام الجامعي 2016/2015م.
- **الأدوات:** تتحدد نتائج الدراسة الحالية بالاختبار المستخدم وهو اختبار واطسون وجليسر للتفكير الناقد النسخة القصيرة (ترجمة: العتيبي، 2012).

التحقق من الافتراضات الأساسية لنماذج نظرية الاستجابة للمفردة:

1- **تم التأكد من عدم وجود بيانات تامة أو صفرية** (بنود تمت الإجابة عليها إجابة صحيحة من قبل جميع الأفراد، الأفراد الذين أجابوا عن جميع الأسئلة إجابة صحيحة أو إجابة خاطئة) وفي ضوء تلك الخطوة تم حذف 9 طلاب من طلاب التخصصات الأدبية كانت إجابتهم صفرية.

2- التأكد من افتراض أحادية البعد:

تم التأكد من افتراض أحادية البعد باستخدام تحليل المكونات الرئيسية وتم الاعتماد على محك كيزر ومحك المتوسط الجزئي الأقل Minimum Partial Average ليفيلكر Velicer في تحديد عدد المكونات الرئيسية للاختبار كما هو موضح بالجدول التالي [تم الاكتفاء بعرض الجذور الكامنة التي تزيد عن أو تساوي (1)]:

جدول (3): تحليل المكونات الرئيسية ومحك المتوسط الجزئي الأقل لتحديد مكونات اختبار التفكير الناقد

المكونات	الجذر الكامن	نسبة التباين المفسر	التباين المفسر التراكمي	المكونات	الجذر الكامن	التباين المفسر	التباين المفسر التراكمي
1	6.851	17.129	17.129	8	1.255	3.137	44.582
2	2.047	5.118	22.247	9	1.197	2.993	47.576
3	1.812	4.530	26.777	10	1.142	2.856	50.432
4	1.636	4.089	30.866	11	1.089	2.723	53.155
5	1.521	3.804	34.669	12	1.071	2.676	55.832
6	1.393	3.483	38.152	13	1.056	2.639	58.471
7	1.317	3.293	41.446	14	1.020	2.550	61.021

محك المتوسط الجزئي الأقل

المكونات	متوسط مربعات معاملات الارتباط الجزئي	المكونات	متوسط مربعات معاملات الارتباط الجزئي
الكلي	0.027	3	0.007
1	0.007 (أقل متوسط)	4	0.008
2	0.007	5	0.008

والجدول السابق يتضح منه أن المكون الرئيسي الأول جذره الكامن 6.851 بينما الجذر الكامن للعامل الثاني 2.047 مما يعني أن الجذر الكامن للمكون الرئيس الأول يزيد عن ثلاثة أضعاف الجذر الكامن للمكون الرئيسي الثاني كما يلاحظ أن هناك تقارب بين الجذور الكامنة للمكونات التالية للمكون الأول، ونظراً لانخفاض نسبة التباين المفسر بواسطة المكون الأول بالاعتماد على محك كيزر، قام الباحث بالاعتماد على محك المتوسط الجزئي الأقل في تحديد عدد المكونات الرئيسية ويلاحظ من الجدول السابق أن متوسط مربعات معاملات الارتباط الجزئي بين بنود الاختبار تساوي

0.027 وبعد عزل تأثير المكون الأول أصبحت هذه القيمة مساوية 0.007 وبعد عزل تأثير المكون الثاني أصبحت هذه القيمة مساوية 0.007 مما يعني أن أقل متوسط لمربعات معاملات الارتباط الجزئي يتم الحصول عليه في حالة تشبع بنود الاختبار على مكون واحد فقط، حيث أن عدد المكونات التي يتم الإبقاء عليها كناتج لتحليل المكونات الرئيسية بالاعتماد على هذا المحك يتحدد بالنقطة التي يتم عندها الوصول إلى أقل متوسط لمربع معاملات الارتباط الجزئي (Velicer, Eaton & Fava, 2000, 54)؛ وهو ما يؤكد فرض أحادية البعد للاختبار.

وبتحقق افتراض أحادية البعد يمكن استنتاج تحقق فرض الاستقلال الموضعي بشكل ضمني والجدول التالي يوضح بعض المؤشرات الإحصائية لاستجابات أفراد عينة الدراسة على الاختبار:

جدول (4): المتوسط والانحراف المعياري والالتواء والتفرطح والثبات لاستجابات أفراد العينة على الاختبار

المجموعة	الالتواء	التفرطح	ثبات ألفا-كرونباخ	ثبات سبيرمان وبراون
علمي	0.079	0.819-	0.807	0.781
أدبي	0.292	0.786-	0.872	0.794
العينة الكلية	0.753-	0.005	0.873	0.805

ويتضح من الجدول السابق أن درجات الاختبار لها معاملات ثبات مقبولة، ويتضح كذلك أن معاملات الالتواء والتفرطح منخفضة مما يؤكد اقتراب توزيع الدرجات من التوزيع الاعتدالي، وذلك في حالة طلاب التخصصات العلمية أو طلاب التخصصات الأدبية أو العينة ككل.

نتائج الدراسة:

أولاً: نتائج السؤال الأول:

ينص السؤال الأول للدراسة الحالية على "ما مؤشرات جودة مطابقة بنود اختبار التفكير الناقد لواطسون وجليسر "النسخة القصيرة" للنموذج البارامتري ثنائي المعلم؟

تمثل الهدف من هذا السؤال في التأكد من مدى ملاءمة بنود الاختبار موضوع الدراسة للنموذج ثنائي المعلم والذي يعد شرطاً لاستخدام اختبار النسبة الاحتمالية في دراسة الأداء التفاضلي للبنود، وللإجابة عن هذا السؤال تم استخدم برنامج IRTPRO في التعرف على مَعلمي الصعوبة والتمييز لبنود الاختبار بوحدة اللوجيت وكذلك الأخطاء المعيارية في تقدير مَعلمي البند وإحصاءات جودة مطابقة البنود في ضوء النموذج البارامتري ثنائي المعلم فكانت النتائج كما هي موضحة بجدول(5):

جدول (5): معالم البنود (الصعوبة *b* والتمييز *a*) والأخطاء المعيارية في تقديرها ومؤشرات جودة المطابقة وفقاً للنموذج البارامتري ثنائي المعلم.

البنود	*Slope (a)*	*S.E*	*Threshold (b)*	*S.E*	x^2	*d.f*	*Prob.*
1	0.319	0.113	0.358	0.337	35.565	31	0.261
2	0.949	0.146	0.164	0.124	22.610	29	0.795
3	1.198	0.169	0.211	0.104	20.583	25	0.716
4	1.511	0.200	0.202	0.090	20.953	25	0.696
5	1.444	0.190	0.079	0.093	19.984	24	0.699
6	1.185	0.167	0.235	0.105	22.404	27	0.717
7	1.088	0.157	0.176	0.112	24.830	28	0.638
8	1.110	0.160	0.312	0.112	34.865	28	0.252
9	0.868	0.142	0.336	0.135	16.751	29	0.966
10	0.752	0.134	0.311	0.152	27.970	29	0.521
11	0.771	0.134	0.321	0.150	17.583	29	0.953
12	0.932	0.148	0.504	0.135	22.454	29	0.801
13	0.729	0.134	0.587	0.173	29.112	29	0.461
14	1.027	0.154	0.461	0.124	24.974	28	0.630
15	0.727	0.132	0.398	0.161	33.118	30	0.317

البنود	Slope (a)	S.E	Threshold (b)	S.E	x^2	d.f	Prob.
16	0.798	0.135	0.284	0.144	24.509	29	0.704
17	0.630	0.127	0.446	0.187	24.402	30	0.754
18	0.866	0.141	0.364	0.137	22.583	29	0.796
19	0.931	0.145	0.282	0.127	22.263	29	0.810
20	0.944	0.148	0.461	0.131	25.388	28	0.608
21	0.884	0.143	0.386	0.135	15.576	27	0.961
22	0.936	0.149	0.503	0.134	29.553	29	0.438
23	0.939	0.147	0.319	0.127	17.922	29	0.946
24	1.044	0.155	0.335	0.118	22.437	28	0.761
25	0.708	0.132	0.487	0.169	17.411	29	0.956
26	1.120	0.162	0.299	0.110	22.560	28	0.755
27	0.822	0.140	0.522	0.151	20.483	28	0.847
28	1.000	0.151	0.319	0.121	26.261	29	0.613
29	0.840	0.141	0.443	0.143	26.163	28	0.565
30	0.794	0.136	0.387	0.148	23.683	29	0.745
31	1.032	0.154	0.374	0.120	21.598	29	0.837
32	1.055	0.155	0.309	0.116	31.663	28	0.288
33	0.927	0.146	0.218	0.126	29.250	29	0.454
34	0.923	0.146	0.336	0.128	22.382	29	0.805
35	1.045	0.154	0.168	0.114	21.875	28	0.788
36	0.843	0.141	0.273	0.136	28.347	29	0.501
37	0.813	0.137	0.009	0.138	25.733	28	0.589
38	0.772	0.136	0.276	0.147	32.296	29	0.306
39	0.792	0.135	0.052	0.141	18.045	28	0.925
40	0.821	0.137	0.075-	0.138	24.772	28	0.641

التمييز *Slope*، الصعوبة *Threshold*، الخطأ المعياري *S.E*، مربع كا x^2 ، درجات الحرية *d.f*

يتضح من الجدول السابق أن:

- قيم معلم التمييز تراوحت ما بين 0.319 في حالة البند (1) وبين 1.511 في حالة البند (4) ويلاحظ على قيم معلم تمييز البنود أن جميع البنود تمييز تمييزاً عالياً بين مستويات السمة المختلفة وبلغ معامل التمييز الكلي للاختبار 0.922 وهو معامل تمييز مرتفع.
- قيم معلم الصعوبة تراوحت ما بين -0.075 في حالة البند 40 والذي يعد أسهل بنود الاختبار و0.587 في حالة البند 13 والذي يعد أصعب البنود ويلاحظ أن جميع قيم معلم الصعوبة موجبة ما عدا البند 40 وبالتالي تكون صعوبات 39 بند من بنود الاختبار على يمين متصل السمة، وكذلك بلغت قيمة معلم الصعوبة الكلي للاختبار 0.311 وهو ما يؤكد على أن صعوبة البنود ما بين المتوسطة والمرتفعة.
- امتدت الأخطاء المعيارية في تقدير معلم التمييز ما بين 0.113 و0.200 وفي حالة معلم الصعوبة امتدت ما بين 0.090 و0.187 لجميع البنود ما عدا البند رقم (1) والذي كان الخطأ المعياري في تقدير صعوبته مساوياً 0.337 وهو ما يعني أن الأخطاء المعيارية في تقدير معالم البنود مقبولة إلى حد كبير.
- جاءت جميع قيم مؤشر مربع كاي كمؤشر لجودة مطابقة البنود غير دالة إحصائياً وهو ما يؤكد على مطابقة جميع البنود للنموذج البارامتري ثنائي المعلم.

ثانياً: نتائج السؤال الثاني:

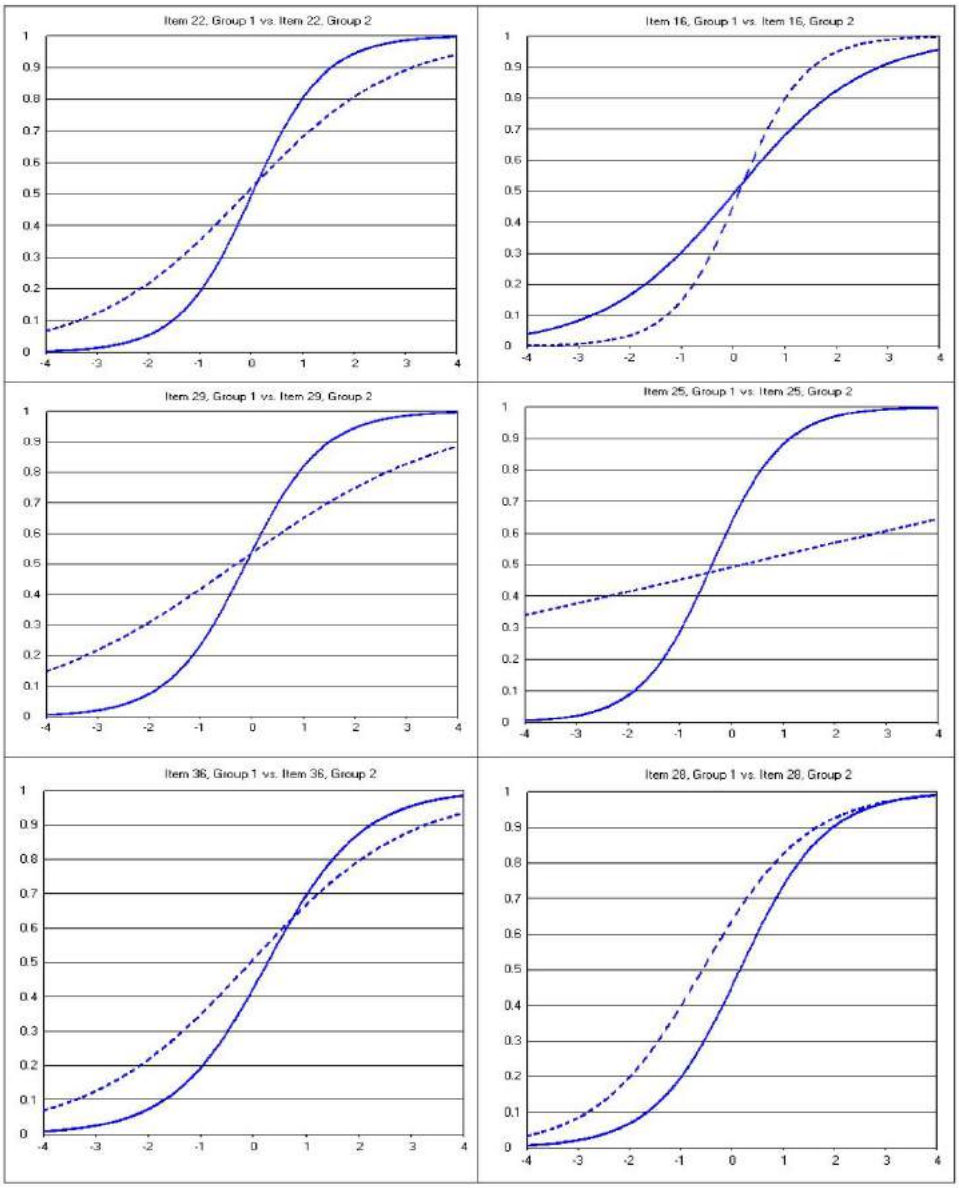
ينص السؤال الثاني للدراسة الحالية على "ما بنود اختبار واطسون وجليسر للتفكير الناقد "النسخة القصيرة" التي لها أداءً تفاضلياً بالنسبة للتخصص باستخدام النموذج البارامتري ثنائي المعلم؟".

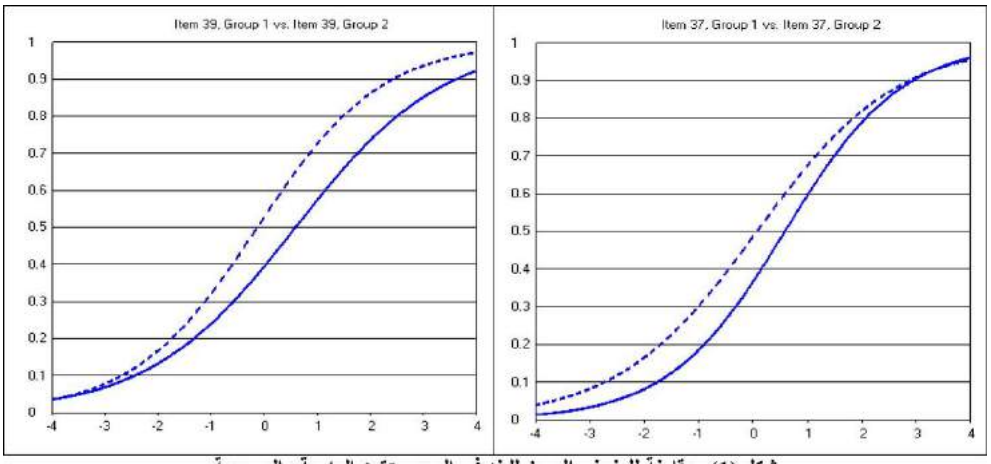
تم استخدام برنامج IRTPRO للكشف عن الأداء التفاضلي لبنود الاختبار وتم في البداية حساب قيمة النسبة الاحتمالية للمقارنة بين النموذج ذي المجموعة الواحدة (العينة ككل) والنموذج ذي المجموعتين، فبلغت قيمة النسبة الاحتمالية G^2 203.741، [تساوي (الفرق بين G^2 في حالة العينة الكلية (19965.551) وفي حالة المجموعتين (19761.809)] وهي قيمة دالة إحصائياً عند مستوى 0.001 ودرجة حرية 39 والذي يؤكد أن جودة المطابقة في حالة النموذج ذي المجموعتين كانت أفضل منها في حالة المجموعة الواحدة، ويؤكد على أن الاختبار والبنود المكونة له تحتاج إلى الفحص الإحصائي لمعرفة ما إذا كان الاختبار وبنوده يؤدى بطريقة مختلفة بالنسبة للمجموعتين أم لا؛ وجاءت مؤشرات الأداء التفاضلي للبنود كما هي موضحة بجدول (6):

جدول (6): البنود التي لها أداء تفاضلي بناء على مؤشر النسبة الاحتمالية في ضوء النموذج البارامتري ثنائي المعلم.

م	المجموعة البؤرية (تخصص أدبي)				المجموعة المرجعية (تخصص علمي)			مربع كا للفروق بين معالم البنود		
	Slope	S.E	Threshold	S.E	Slope	S.E	Threshold	total	Slope	Threshold
1	0.213	0.113	0.127-	0.146	0.443	0.173	0.172	*6.314	1.225	*4.089
2	0.702	0.161	0.281	0.155	0.804	0.211	0.304	0.028	0.010	0.019
3	0.768	0.173	0.399	0.168	1.182	0.265	0.128	2.126	1.719	0.407
4	1.154	0.238	0.318	0.174	1.404	0.298	0.546	1.183	0.567	0.616
5	1.104	0.233	0.677	0.176	1.106	0.254	0.441	0.515	0.001	0.513
6	1.027	0.217	0.706	0.162	1.084	0.236	0.031	*6.872	0.004	**6.868
7	0.843	0.191	0.176	0.159	1.000	0.227	0.224	1.820	0.301	1.519
8	0.779	0.186	0.241	0.156	0.896	0.214	0.054	0.511	0.189	0.322
9	0.570	0.148	0.003	0.156	0.904	0.214	0.169	2.718	1.717	1.001
10	0.651	0.163	0.148	0.146	0.430	0.173	0.177	0.899	0.899	0.000
11	0.500	0.144	0.070	0.152	0.751	0.197	0.095	1.101	1.101	0.009
12	0.688	0.181	0.066	0.151	0.727	0.194	0.002	0.099	0.000	0.099
13	0.635	0.168	0.171	0.146	0.502	0.176	0.110	1.212	0.295	0.917
14	0.859	0.212	0.021	0.150	0.681	0.188	0.121	0.603	0.397	0.206
15	0.587	0.168	0.043	0.146	0.489	0.174	0.112	0.199	0.199	0.000
16	0.464	0.149	0.051	0.157	0.916	0.211	0.126	*7.611	**6.830	0.781
17	0.556	0.155	0.224	0.144	0.342	0.166	0.130	0.901	0.901	0.000
18	0.506	0.152	0.073	0.155	0.858	0.204	0.020	1.913	1.913	0.000
19	0.598	0.168	0.160	0.152	0.727	0.193	0.099	0.333	0.333	0.000
20	0.594	0.170	0.141	0.157	0.935	0.211	0.070	2.645	1.644	1.001
21	0.511	0.154	0.049	0.154	0.827	0.201	0.039	1.601	1.599	0.000
22	0.836	0.216	0.142	0.144	0.398	0.169	0.016	*6.854	*5.582	1.272
23	0.759	0.202	0.110	0.148	0.574	0.180	0.229	0.614	0.512	0.102
24	0.884	0.225	0.112	0.149	0.635	0.185	0.258	0.900	0.722	0.188
25	0.856	0.220	0.020	0.142	0.092	0.161	0.318	**9.307	**7.924	1.383
26	1.101	0.280	0.110	0.148	0.585	0.180	0.501	3.451	2.418	1.033
27	0.791	0.216	0.393	0.147	0.520	0.177	0.259-	*6.945	0.123	**6.822
28	0.713	0.210	0.330	0.150	0.578	0.182	0.115-	*6.991	0.255	**6.736
29	0.790	0.230	0.082	0.143	0.279	0.164	0.091	*7.738	**6.928	0.810
30	0.448	0.155	0.025	0.153	0.804	0.198	0.032-	2.001	2.001	0.000
31	0.738	0.218	0.177	0.149	0.616	0.183	0.038	1.108	0.208	0.900
32	0.626	0.195	0.173	0.158	0.947	0.215	0.024	1.312	1.210	0.102
33	0.646	0.194	0.205	0.152	0.711	0.193	0.222	0.100	0.100	0.000
34	0.712	0.212	0.089	0.148	0.617	0.183	0.184	0.222	0.111	0.111
35	0.650	0.204	0.301	0.157	0.811	0.202	0.099	1.351	0.349	1.002
36	0.657	0.205	0.274	0.145	0.387	0.170	0.053-	*6.611	1.001	*5.610
37	0.549	0.186	0.588	0.153	0.460	0.180	0.073	**10.117	1.215	**8.902
38	0.489	0.176	0.240	0.148	0.559	0.180	0.037-	1.219	0.112	1.107
39	0.426	0.154	0.593	0.153	0.505	0.182	0.127-	**9.973	1.548	**8.425
40	0.553	0.181	0.044-	0.152	0.478	0.179	0.070	1.710	0.103	1.607

يتضح من الجدول السابق أن البنود (1، 6، 27، 28، 36، 37، 39) لها أداء تفاضلي في الصعوبة أما البنود (16، 22، 25، 29) فلها أداء تفاضلي في التمييز حيث كان البند (1) أصعب بالنسبة لطلاب التخصصات العلمية؛ بينما البنود (6، 27، 28، 36، 37، 39) أصعب بالنسبة لطلاب التخصصات الأدبية، وكانت البنود (22، 25، 29) تُميّز أكثر بين المستويات المختلفة في السمة لدى طلاب التخصصات الأدبية، بعكس البند (16) والذي يُميّز أكثر بين المستويات المختلفة في السمة لدى طلاب التخصصات العلمية، والأشكال التالية توضح المنحنيات المميزة لبعض البنود ذات الأداء التفاضلي في المجموعتين البؤرية والمرجعية:





شكل (1): مقارنة للمنحنى المميز للبند في المجموعتين البؤرية والمرجعية

مجموعة مرجعية (تخصصات علمية) ——— مجموعة بؤرية (تخصصات أدبية)

ويتضح من الشكل السابق أنه يمكن اعتبار الأداء التفاضلي للبنود (16، 22، 25، 29) غير منتظم غير مرتب، حيث أن البنود (22، 25، 29) تعد أصعب بالنسبة للتخصصات الأدبية عند المستوى المنخفض للسمة ثم ينعكس الوضع عند المستوى المرتفع للسمة وتصبح أصعب بالنسبة للتخصصات العلمية، بينما الوضع معكوس بالنسبة للبند (16) والذي يعد أصعب للتخصصات العلمية عند المستوى المنخفض للسمة ثم يصبح أصعب بالنسبة للتخصصات الأدبية عند المستوى المرتفع للسمة، بينما الأداء التفاضلي للبنود (1، 6، 27، 28، 36، 37، 39) قد يكون منتظم أو غير منتظم مرتب حسب مقدار الفرق في معلم التمييز في المجموعتين

ويلاحظ أنه في حالة الأداء التفاضلي غير المنتظم غير المرتب أن المنحنيات تتقاطع ولا تتركز الفروق في المساحة بين المنحنيين في منطقة واحدة، بينما في حالة البنود ذات الأداء التفاضلي المنتظم أو غير المنتظم المرتب (الفروق في معلم الصعوبة) لا تتقاطع المنحنيات كما هو الحال في البنود (28، 37، 38) أو تتقاطع ويتركز الجزء الأكبر من الفرق في المساحة بين المنحنيين في منطقة واحدة كما في حالة البند 36.

ثالثاً: نتائج السؤال الثالث:

ينص السؤال الثالث للدراسة الحالية على "ما بنود اختبار واطسون وجليسر للتفكير الناقد "النسخة القصيرة" التي لها أداءً تفاضلياً بالنسبة للتخصص باستخدام نموذج تحليل الانحدار اللابارامتري الممهد؟".

تتطلب التحليلات اللابارامترية التي تهدف للتعرف على الأداء التفاضلي للبنود -على العكس من التحليلات البارامترية- التأكد أولاً من تكافؤ المجموعات في مستوى السمة المقاسة (Ramsay, 2000, 68; Khan et al., 2011, 7) ففي النماذج البارامترية لنظرية الاستجابة للمفردة الاختبارية ليس من الضروري تماثل المجموعات في توزيع السمة، ولكن في النماذج اللابارامترية من الضروري التأكد من عدم اختلاف توزيع السمة في المجموعتين حيث أن الاستدلال على السمة الكامنة يتم بواسطة ترتيب الدرجة الكلية للمختبرين (Molenaar, 2003, 193)؛ ولتحقيق هذا الشرط تم استخدام اختبار "ت" للمجموعات المستقلة Independent Samples T-Test بعد ترتيب أفراد عينة الدراسة الكلية وحذف بعض الأفراد المنخفضين في الدرجة الكلية وكان غالبيتهم من مجموعة التخصصات الأدبية وحذف بعض الأفراد من مرتفعي السمة وكان غالبيتهم من مجموعة التخصصات العلمية وكانت نتائج تكافؤ المجموعتين كما هي موضحة بالجدول التالي:

جدول (8) تكافؤ مجموعتي الدراسة في الدرجة الكلية للتفكير الناقد

المجموعة	العدد	المتوسط	الانحراف المعياري	قيمة ت
أدبي	143	17.419	6.164	1.473
علمي	142	18.451	5.643	

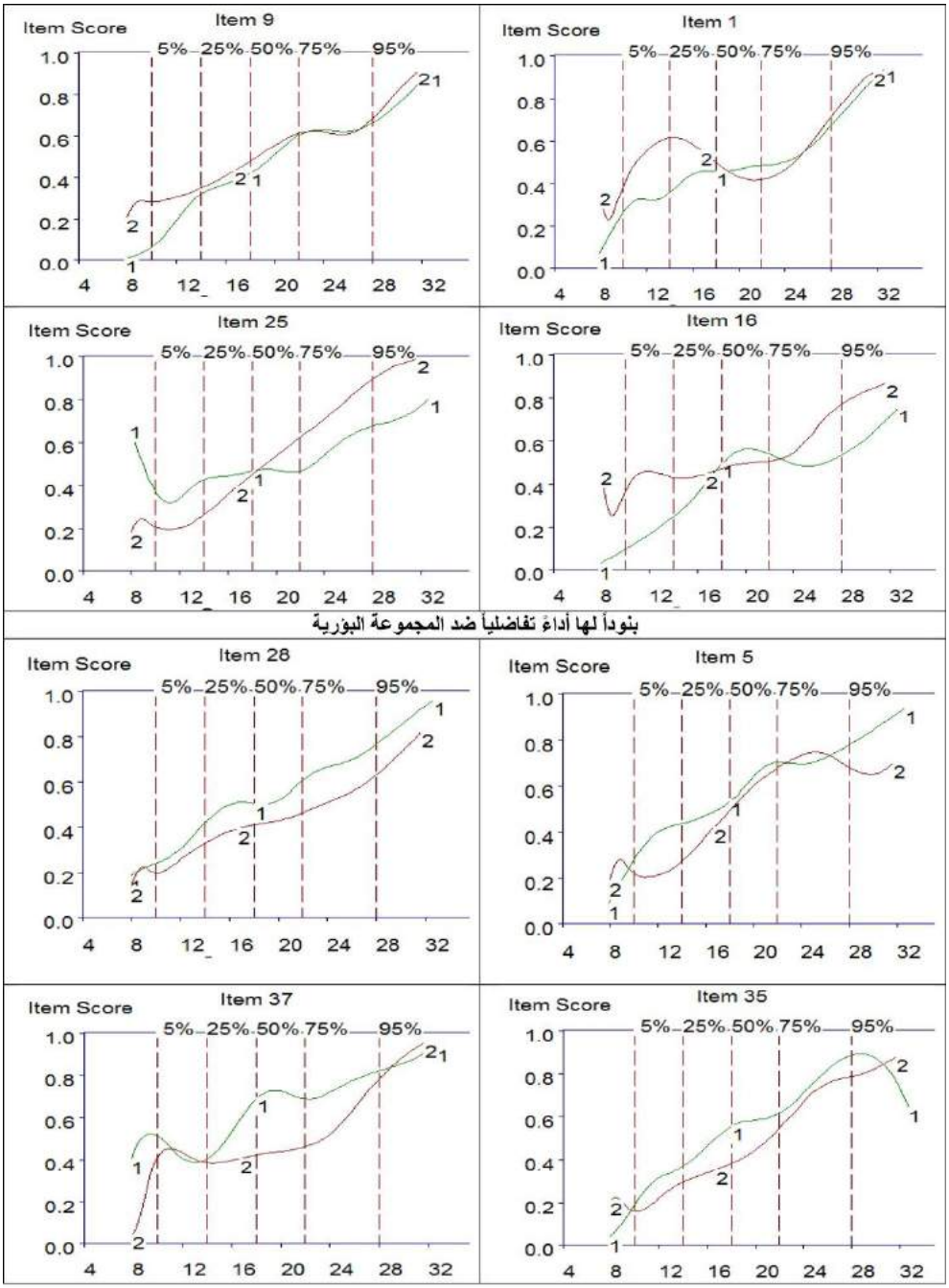
وبلغت قيمة "ت" 1.473 وهي غير دالة إحصائياً بدرجات حرية 283؛ ويتأكد من ذلك التكافؤ بين مجموعتي الدراسة في الدرجة الكلية للاختبار، وتم بعد ذلك استخدام برنامج TestGraf في حساب مؤشر β للفرق في المساحة بين المنحنى اللابارامتري المميز للبند في المجموعة البؤرية والمنحنى اللابارامتري المميز للبند في المجموعة المرجعية، وكذلك الخطأ المعياري في حساب مؤشر β ومن ثم حساب قيمة Z والتي تم بناءً عليها تقييم الأداء التفاضلي لبنود الاختبار، وهو ما يتضح من جدول (9) الآتي:

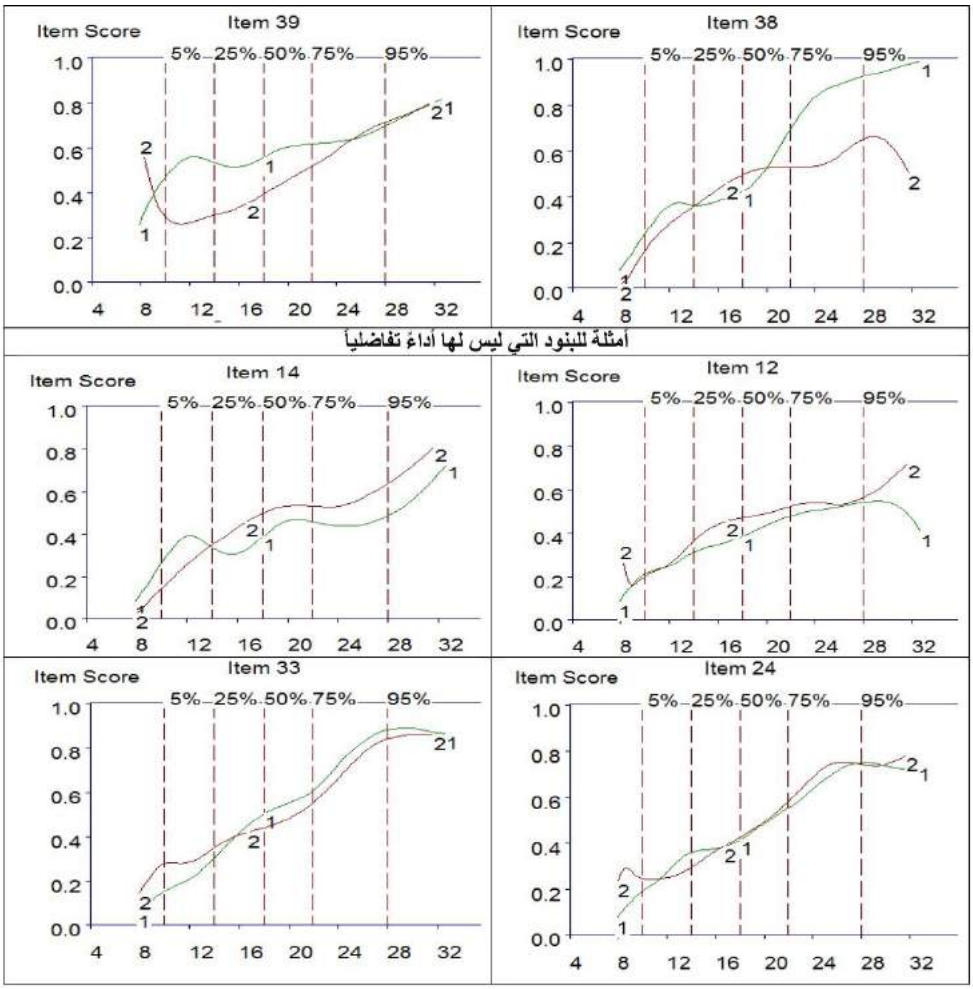
جدول (9): البنود التي لها أداءً تفاضلياً بناء على مؤشر β باستخدام نموذج الانحدار اللابارامتري الممهد

البنود	β	S. E.	Z	Prob.	البنود	β	S. E.	Z	Prob.
1	0.104	0.035	2.971	0.002	21	0.007	0.035	0.200	0.842
2	0.051	0.035	1.457	0.073	22	0.077	0.034	2.265	0.012
3	0.000	0.035	0.000	1.000	23	0.027	0.035	0.771	0.441
4	0.048	0.034	1.412	0.079	24	0.015	0.034	0.441	0.659
5	0.068-	0.034	2.000-	0.023	25	0.086	0.034	2.529	0.011
6	0.095-	0.035	2.714-	0.003	26	0.003	0.034	0.088	0.929
7	0.050	0.035	1.429	0.077	27	0.105-	0.034	3.088-	1.000
8	0.036-	0.035	1.029-	0.152	28	0.084-	0.034	2.471-	0.014
9	0.084	0.035	2.400	0.016	29	0.056-	0.034	1.647-	0.099
10	0.055-	0.035	1.571-	0.116	30	0.031	0.035	0.886	0.376
11	0.000	0.034	0.000	1.000	31	0.111-	0.033	3.364-	0.001
12	0.041	0.034	1.206	0.228	32	0.028	0.034	0.824	0.409
13	0.055	0.035	1.571	0.116	33	0.018	0.035	0.514	0.607
14	0.009	0.034	0.265	0.791	34	0.035	0.035	1.000	0.317
15	0.028	0.035	0.800	0.424	35	0.072-	0.034	2.118-	0.034
16	0.143	0.035	4.086	0.000	36	0.071-	0.034	2.088-	0.037
17	0.000	0.036	0.000	1.000	37	0.138-	0.035	3.943-	0.000
18	0.013	0.034	0.382	0.703	38	0.069-	0.034	2.029-	0.043
19	0.010	0.034	0.294	0.769	39	0.146-	0.035	4.171-	0.000
20	0.117-	0.035	3.343-	0.001	40	0.074-	0.035	2.114-	0.035

يتضح من جدول (9) أن عدد البنود التي لها أداءً تفاضلياً في ضوء المساحة بين المنحنى اللابارامتري المميز للبند في المجموعتين البؤرية والمرجعية عددها 17 بند، منها 5 بنود لها أداء تفاضلي ضد المجموعة المرجعية (التخصصات العليمة) وهي البنود (1، 9، 16، 22، 25) مما يعني أنها أسهل بالنسبة لمجموعة التخصصات الأدبية أو أنها ذات تمييز أقل بالنسبة للتخصصات العلمية وإن كان الاحتمال الثاني هو الأقرب للتحقق، ومنها 12 بنداً لها أداءً تفاضلياً ضد المجموعة البؤرية (التخصصات الأدبية) وهي البنود (5، 6، 20، 27، 28، 31، 35، 36، 37، 38، 39، 40)، مما يعني أنها أسهل بالنسبة لمجموعة التخصصات العلمية أو ذات تمييز أقل بالنسبة

للتخصصات الأدبية، ويوضح الشكل الآتي أمثلة للمقارنة بين المنحنى اللابارامتري في المجموعتين المرجعية (منحنى 1) والبؤرية (منحنى 2) للبنود ذات الأداء التفاضلي في الاختبار:





أمثلة للبنود التي ليس لها أداءً تفاضلياً

وفي ضوء محكات الحكم على درجة الأداء التفاضلي التي تم تحديدها من قبل Roussos & Stout ونتائج دراسة (Witarsa, 2003) يمكن اعتبار أن الأداء التفاضلي للبنود (1، 6، 9، 16، 20، 22، 25، 27، 28، 31، 35، 36، 37، 38، 39، 40) كبيراً من حيث حجم التأثير بينما الأداء التفاضلي للبند (5) متوسطاً حيث لم تتعدى قيمة β لهذا البند القيمة 0.069، ويوضح جدول (11) الاتفاق والاختلاف بين النموذج البارامتري ثنائي المعلم ونموذج الانحدار اللابارامتري الممهد في الكشف عن الأداء التفاضلي للبنود:

جدول (11): بنود الاختبار التي لها أداءً تفاضلياً في ضوء النموذج البارامتري ثنائي المعلم ونموذج الانحدار اللابارامتري الممهد

طريقة التقدير	البنود التي لها أداءً تفاضلياً ضد المجموعة المرجعية	البنود التي لها أداءً تفاضلياً ضد المجموعة البؤرية	المجموع
اختبار نسبة الأرجحية باستخدام النموذج البارامتري ثنائي المعلم	(1) صعوبة أعلى	(6، 27، 28، 36، 37، 39) صعوبة أعلى	11
	(22، 25، 29) تمييز أقل	(16) تمييز أقل	
المساحة بين المنحنى اللابارامتري المميز للبند	(1، 9، 16، 22، 25)	(5، 6، 20، 27، 28، 31، 35، 36، 37، 38، 39، 40)	17
البنود المتفق عليها	1، 22، 6، 16، 25، 27، 28، 36، 37، 39		10

ويلاحظ أن نسبة البنود التي تم تحديدها على أنها ذات أداء تفاضلي اعتماداً على اختبار نسبة الأرجحية باستخدام النموذج البارامتري ثنائي المعلم بلغت 27.5% من إجمالي عدد بنود الاختبار، بينما نسبة البنود التي تم تحديدها على أنها ذات أداء تفاضلي اعتماداً على نموذج الانحدار اللابارامتري الممهد بلغت 42.5% من إجمالي عدد بنود الاختبار، ونسبة الاتفاق بين الطريقتين بلغت حوالي 91% في ضوء عدد البنود التي لها أداءً تفاضلياً باستخدام النموذج البارامتري ثنائي المعلم، وبلغت حوالي 59% في ضوء عدد البنود التي لها أداءً تفاضلياً باستخدام نموذج الانحدار اللابارامتري الممهد، ويلاحظ من تفحص المنحنيات اللابارامترية للبنود ذات الأداء التفاضلي أن من هذه البنود ما له أداءً تفاضلياً منتظماً تقريباً كالبند (35) ومنها ما له أداءً تفاضلياً غير منتظماً مرتباً كالبنود (9، 28، 37، 39) ومنها ما له أداءً تفاضلياً غير منتظماً غير مرتباً كالبندين (16، 25) فعلى سبيل المثال في حالة البنود التي تُظهر أداءً تفاضلياً ضد التخصصات العلمية نلاحظ:

- البند 1: نجد أن المنحنى في الجزء الممثل للأفراد منخفضي السمة في المجموعتين المرجعية والبؤرية يظهر أنه أسهل بالنسبة للتخصصات الأدبية حتى مستوى السمة المقابل تقريباً للدرجة 16 كدرجة متوقعة في الاختبار وبعدها يكاد يكون هناك تماثل في صعوبة البند في المجموعتين.
- البند 9: نجد أن المنحنى في الجزء الممثل للأفراد منخفضي السمة يعبر عن سهولة البند بالنسبة لأفراد التخصصات الأدبية وحتى مستوى السمة المقابل تقريباً للدرجة 16 كدرجة متوقعة في الاختبار تبدأ صعوبة البند في التماثل في المجموعتين.
- البند 16: نجد أن المنحنى في الجزء الممثل للأفراد منخفضي السمة أسهل للتخصصات الأدبية حتى الدرجة 16 تقريباً وفي الجزء الممثل للمستوى المتوسط من السمة يكاد يكون هناك تماثل في صعوبة البند، ولكن عند مستوى السمة المقابل تقريباً للدرجة 24 كدرجة متوقعة في الاختبار تبدأ يظهر الفرق في صعوبة البند حيث تزداد صعوبته مرة أخرى بالنسبة للتخصصات العلمية.
- البند 25: نجد أن البند في الجزء الممثل للأفراد منخفضي السمة أسهل لطلاب التخصصات العلمية ثم ينعكس الأمر في الجزء الممثل للأفراد مرتفعي السمة ليصبح البند أسهل لطلاب التخصصات الأدبية.

وما سبق يؤكد أن طلاب التخصصات العلمية عند المستوى المنخفض للسمة بصفة خاصة يؤدون بطريقة غير متوقعة على البنود التي ظهر أن لها أداءً تفاضلياً ضد التخصصات العلمية مقارنة بطلاب التخصصات الأدبية من نفس مستوى السمة، خاصة البنود (1، 9، 16، 22) أما في حالة البند (25) فإن طلاب التخصصات العلمية ذوي المستوى المرتفع من السمة هم من يؤدون بطريقة غير متوقعة مقارنة بطلاب التخصصات الأدبية حيث يبدو البند لهم أكثر صعوبة من طلاب التخصصات الأدبية.

أما في حالة البنود التي لها أداءً تفاضلياً ضد طلاب التخصصات الأدبية فنلاحظ على سبيل المثال أن المنحنى المميز في حالة:

- البند 5: يلاحظ أن المنحنى في الجزء الممثل للأفراد منخفضي السمة أصعب لطلاب التخصصات الأدبية عنه بالنسبة للتخصصات العلمية، وفي المستوى المتوسط للسمة بداية من الدرجة 16 إلى الدرجة 27 تقريباً كدرجات متوقعة في الاختبار لا يظهر أن للبند أداءً تفاضلياً، ثم تزداد صعوبة البند لطلاب التخصصات الأدبية مرة أخرى عند المستوى المرتفع من السمة.
- البند 28: يلاحظ أن البند أصعب لطلاب التخصصات الأدبية عند كل مستويات السمة تقريباً، ويمكن اعتباره من أكثر البنود ذات الأداء التفاضلي المنتظم ضد طلاب التخصصات الأدبية.
- البند 35: يلاحظ أن البند أصعب لطلاب التخصصات الأدبية عند كل مستويات السمة تقريباً، ويمكن اعتبار أن البند له أداءً تفاضلياً منتظماً ضد طلاب التخصصات الأدبية.

-البند 37: يلاحظ على هذا البند أنه صعب جداً بالنسبة لطلاب التخصصات الأدبية منخفضي مستوى السمة، ويظهر الأداء التفاضلي للبند تحديداً في المستوى المتوسط للسمة بين الدرجة 12 والدرجة 26 كدرجات متوقعة في الاختبار، بينما لا يُظهر البند أداءً تفاضلياً عن المستويات المرتفعة من السمة.

-البند 38: يلاحظ أنه عند الدرجة 18 تقريباً كدرجة متوقعة في الاختبار تبدأ صعوبة البند تزداد بالنسبة لطلاب التخصصات الأدبية، والعكس بالنسبة للبند 39 والذي يظهر أنه أصعب لطلاب التخصصات الأدبية منخفضي ومتوسطي السمة وبعد الدرجة 22 كدرجة متوقعة في الاختبار يظهر أن هناك تشابه في صعوبة البند في مجموعتي الدراسة.

وعلى الرغم من أن البند (29) والذي ظهر أن له أداءً تفاضلياً ضد المجموعة المرجعية من حيث التمييز في ضوء النموذج البارامتري ثنائي المعلم إلا أنه لم يتم تشخيصه بأن له أداءً تفاضلياً بناءً على نموذج الانحدار اللابارامتري، كذلك هناك (6) بنود تم اعتبارها ذات أداء تفاضلي في ضوء النموذج اللابارامتري فقط.

ويتأكد مما سبق أن الأداء التفاضلي لبنود الاختبار الحالي ليس لها شكل موحد؛ كذلك يتضح أن النموذج اللابارامتري أكثر حساسية في الكشف عن الأداء التفاضلي للبنود، خاصة في حالة زيادة صعوبة البند في إحدى المجموعتين، أما في حالة الاختلاف في معلم التمييز والتشابه في صعوبة البند فإن قوة النموذج اللابارامتري تقل حيث لم يتمكن النموذج من تشخيص البند (29) والذي يتحيز في التمييز ضد التخصصات العلمية.

رابعاً: نتائج السؤال الرابع:

ينص السؤال الرابع للدراسة الحالية على "هل تتشابه البنية الكامنة المقاسة باستخدام اختبار واطسون وجليسر للتفكير الناقد "النسخة القصيرة" لدى طلاب التخصصات العلمية وطلاب التخصصات الأدبية باستخدام التحليل العاملي التوكيدي متعدد المجموعات؟

للتعرف على مدى تكافؤ نموذج القياس بالنسبة للمجموعتين المرجعية (التخصصات العلمية) والبؤرية (التخصصات الأدبية) ومدى التطابق بين البنية الكامنة المقاسة في المجموعتين باستخدام اختبار واطسون وجليسر للتفكير الناقد، تم استخدام التحليل العاملي التوكيدي متعدد المجموعات للمقارنة بين مؤشرات جودة المطابقة للنماذج المتداخلة الأربعة كما اتضح من الإطار النظري، وللحكم على جودة النماذج التي تم تحليلها تم الاعتماد على أن أفضل النماذج هو الذي تتوافر له أفضل قيم لأكبر عدد من مؤشرات جودة المطابقة وتم في الدراسة الحالية الاعتماد على مؤشرات جودة المطابقة التالي (في: عامر، 2004):

(أ) النسبة بين قيمة مربع كاي χ^2 ودرجات الحرية *df*: فإذا كانت أقل من 5 تدل على قبول النموذج ولكن إذا كانت أقل من 2 تدل على أن النموذج مطابق تماماً للبيانات.

(ب) مؤشرات المطابقة المطلقة Absolute Fit Indexes: والاهتمام هنا منصباً على مقارنة مصفوفة التغاير للعينة بالمصفوفة المحللة فعلاً التي تم استهلاكها من قبل النموذج ومن هذه المؤشرات:

- مؤشر حسن المطابقة Goodness of Fit Index (GFI): ويقيس مقدار التباين في المصفوفة المحللة عن طريق النموذج وتتراوح قيمته بين (صفر، 1) والقيمة المرتفعة تشير إلى تطابق أفضل للنموذج، وطور هذا المؤشر ليتحرر من تعقيد النموذج إلى ما يعرف بمؤشر حسن المطابقة المصحح بدرجات الحرية *(AGFI)*.
- مؤشر جذر متوسط مربع خطأ التقريب: Root Mean Square Error of Approximation (RMSEA) وإذا ساوت قيمته 0.05 فأقل دل ذلك على أن النموذج يطابق تماماً البيانات وإذا كانت قيمته محصورة بين 0.05، 0.08 دل ذلك على مطابقة النموذج بدرجة كبيرة أما إذا زادت قيمته عن 0.08 فيتم رفض النموذج.

(ج) مؤشرات المطابقة المتزايدة Incremental Fit Indexes

وهي تعتمد في تقديرها على مقارنة النموذج المفترض مع النموذج الصفري Null Model ومن هذه المؤشرات مؤشر المطابقة المعياري Normed Fit Index (NFI) ومؤشر المطابقة المقارن Comparative Fit Index (CFI) ومؤشر توكر لويس Tucker-Lewis Index (TLI) ومؤشر المطابقة المتزايد Incremental Fit Index (IFI)، وتتراوح قيمة هذه المؤشرات بين (صفر، 1) وتشير القيمة المرتفعة بين هذا المدى إلى تطابق أفضل للنموذج مع بيانات العينة.

جدول (12): مؤشرات جودة المطابقة والمقارنة المعيارية بين النماذج المفترضة لتساوي بارامترات القياس في المجموعتين البؤرية والمرجعية.

مؤشرات جودة المطابقة		النماذج			
		الأول (غير المقيد)	الثاني (المقيد بتماثل التشبعات)	الثالث (المقيد بتماثل التشبعات والتغاير)	الرابع (نموذج التماثل التام)
X^2		24.716	36.408	42.499	70.423
d.f		10	14	15	20
Prob.		0.006	0.001	0.000	0.000
$X^2/d.f$		2.472	2.601	2.833	3.521
GFI		0.974	0.963	0.958	0.932
AGFI		0.923	0.921	0.917	0.898
NFI		0.943	0.915	0.901	0.836
IFI		0.965	0.946	0.934	0.877
TLI		0.928	0.922	0.911	0.877
CFI		0.964	0.935	0.913	0.877
RMSEA		0.051	0.064	0.068	0.080
المقارنة بين النماذج المتداخلة nested Models	النموذج الأول هو الصحيح	ΔX^2	11.692	17.783	45.707
		Δ *d.f*	4	5	10
		Prob.	0.020	0.003	0.000
		ΔCFI	0.029	0.051	0.087
	النموذج الثاني هو الصحيح	ΔX^2		6.091	34.015
		Δ *d.f*		1	6
		Prob.		0.014	0.000
		ΔCFI		0.022	0.058
	النموذج الثالث هو الصحيح	ΔX^2			27.924
		Δ *d.f*			5
		Prob.			0.000
		ΔCFI			0.036

ويتضح من الجدول السابق أن مؤشرات جودة المطابقة للنموذج غير المقيد تؤكد أن النموذج البنائي المفترض للتفكير الناقد كما يقاس باختبار واطسون وجليسر النسخة القصيرة ملائم بدرجة كبيرة لبيانات المجموعتين أثناء التحليل المتآني لبيانات المجموعتين، وكانت أفضل مؤشرات لجودة المطابقة في حالة هذا النموذج (غير المقيد)، وأسوأ هذه النماذج من حيث مؤشرات جودة المطابقة هو النموذج كامل القيود (نموذج التماثل التام)، بمعنى أن أفضل تصور لنموذج القياس كان في حالة المجموعات المنفصلة وبدون أي قيود على تماثل البارامترات في المجموعتين، وهو ما تؤكده المقارنة بين النماذج المتداخلة حيث يتضح أن:

– بالنسبة للنموذج الثاني المقيد بتساوي التشبعات كان الفرق في قيمة مربع كاي في النموذج الثاني والنموذج غير المقيد مساوياً 11.692 وهي قيمة دالة عند مستوى 0.01 ودرجة حرية 4، وكانت قيمة ΔCFI (الفرق في مؤشر المطابقة المقارن) أكبر من 0.01 وهو ما يؤكد على عدم صحة

افتراض تساوي تشبعات العوامل على المتغير الكامن في المجموعتين، وأن بعض هذه العوامل يختلف تشبعها من مجموعة لأخرى.

-بالنسبة للنموذج الثالث المقيد بتساوي التشبعات وتغاير المتغير الكامن في المجموعتين، كان الفرق في قيمة مربع كاي في النموذج غير المقيد والنموذج الثالث مساوياً 17.783 وهي قيمة دالة عند مستوى 0.01 ودرجة حرية 5، وكذلك كان الفرق بين قيمة مربع كاي في النموذج الثاني المقيد بتساوي التشبعات فقط والنموذج الثالث مساوياً 6.091 وهي قيمة دالة عند مستوى 0.05 ودرجة حرية 1، وكانت قيمة ΔCFI في حالة المقارنة بين النموذج الثالث والنموذج غير المقيد وفي حالة المقارنة بين النموذج الثالث والنموذج الثاني أكبر من 0.01 وهو ما يؤكد على عدم صحة افتراض تساوي تغاير المتغير الكامن في المجموعتين، ويؤكد على أن مصفوفة التغاير تختلف من مجموعة لأخرى.

-بالنسبة للنموذج الرابع والمقيد بتساوي التشبعات والتغاير والبواقي كان الفرق بين قيمة مربع كاي في النموذج الأول غير المقيد والنموذج الرابع مساوياً 45.707 وهي قيمة دالة عند مستوى 0.01 ودرجة حرية 10، وكذلك كان الفرق بين قيمة مربع كاي في النموذج الرابع والنموذج الثاني مساوياً 34.015 وهي قيمة دالة عند مستوى 0.01 ودرجة حرية 6، كذلك كان الفرق بين قيمة مربع كاي في النموذج الرابع والنموذج الثالث مساوياً 27.925 وهي قيمة دالة إحصائياً عند مستوى 0.01 ودرجة حرية 5، وكذلك كانت قيمة ΔCFI في حالة المقارنة بين النموذج الرابع والنموذج الأول الثاني والثالث أكبر من 0.01 ويتأكد من ذلك عدم صحة افتراض تساوي البواقي في المجموعتين.

ومن الواضح أن زيادة عدد القيود على النموذج يؤدى إلى زيادة الفروق بينه وبين النموذج غير المقيد، مما يعنى عدم صحة افتراض تساوي التشبعات والتغاير والبواقي في المجموعتين، والنتائج في مجملها هنا تؤكد على عدم صحة افتراض تساوي بارامترات نموذج القياس في المجموعتين، وان البنية الكامنة المقاسة باستخدام الاختبار تختلف لدى طلاب التخصصات العلمية عنها لدى طلاب التخصصات الأدبية.

ووجود فروق في بارامترات القياس في المجموعتين (التخصصات العلمية، التخصصات الأدبية) يؤكد على اختلاف شكل العلاقة بين المتغيرات الكامنة ودرجات المتغيرات الملاحظة (الأبعاد) وهو ما يؤكد أن الفروق بين المجموعتين في متوسط الدرجات الملاحظة لا تعكس الفروق الحقيقية في مستوى السمة الكامنة.

ويلاحظ كذلك زيادة الاختلاف في نموذج القياس في المجموعتين في ضوء البواقي مما يعنى أن نسبة الخطأ في قياس البنية الكامنة في المجموعتين غير متماثلة وأن أحد المجموعتين يتم قياس السمة لديها بنسبة خطأ أكبر من الأخرى وأن الدرجات الملاحظة في الاختبار لا تعبر عن السمة الكامنة بدرجة كبيرة، وفي ضوء نتائج الأسئلة السابقة للدراسة الحالية يمكن استنتاج أن هذه المجموعة هي مجموعة التخصصات الأدبية.

مناقشة النتائج:

1- تأكد من النتائج الحالية أن بعض بنود اختبار واطسون وجليسر للتفكير الناقد لها أداء تفاضلي للتخصص الأكاديمي وهو ما يمكن تفسيره في ضوء التالي:

قد تتأثر الإجابة على تلك البنود بسمات أخرى بجانب السمة الرئيسية التي صمم الاختبار لقياسها، وهو ما يتفق مع ما أشار إليه صلاح علام (2005، 63) حيث يرى لورد Lord أن افتراض أحادية البعد في الاختبارات العقلية وبعض المقاييس الوجدانية لا يعد افتراضاً صحيحاً كلياً؛ فعلى الرغم من اعتماد التحليلات الخاصة بالكشف عن الأداء التفاضلي للبنود على المزاوجة بين المجموعتين في السمة الكامنة المقاسة إلا أن المجموعتين قد تكونان مختلفتين في سمات أخرى لا

يقيسها البند، ومن هنا يمكن استنتاج أن اختلاف المجموعتين في السمات الأخرى المؤثرة في الإجابة على البند قد يسهم في ظهور الأداء التفاضلي للبند.

ولذلك قد يكون الأداء التفاضلي ناتج عن الاختلافات بين طلاب التخصصات العلمية والأدبية في طريقة (استراتيجية) الاستجابة للبند على سبيل المثال، ومعنى البند لدى الفرد، وارتباط محتوى البند بخبرات الفرد السابقة، وسماته الشخصية، وهذا لا يتضح إلا بفحص البنود التي ظهر أن لها أداءً تفاضلياً؛ فكما يذكر Penfield et al., (2007, 453) و Ariffin et al., (2010, 2) فحص محتوى البنود التي ظهر أن لها أداءً تفاضلياً يعد من أفضل الطرق في تحديد أسباب الأداء التفاضلي للبند؛ ويعد ذلك نوعاً من التحليلات الكيفية المكملة للنتائج الكمية للدراسة، وهو ما يمكن توضيحه في الأمثلة التالية:

- البند 16: ينتمي لمهارات الاستنباط ونصه "الشخص الذي لا يواظب على قراءة الأبراج لا يصدق العرافين" والبند يعد نتيجة منطقية للسيناريو المرتبط به والذي نصه "الأشخاص الذين يؤمنون بالخرافات يواظبون على قراءة الأبراج بالصحف المحلية، بعض الناس لا يصدقون العرافين، إذا ....)" أقرب الاحتمالات لأداء هذا البند ضد طلاب التخصصات العلمية هو غموض صياغة البند خاصة مع تدني مستوى الطالب في الفهم اللفظي، فالبند هنا يركز أكثر على تقييم فهم الطالب للنص المقروء من قدرته على الاستنباط وهو ما يسهم في زيادة احتمال الإجابة الخاطئة للبند بالنسبة لطلاب التخصصات العلمية.
- البند 25: ينتمي لمهارات التفسير ونصه "في عام 1902م لم تكن تنتج شركة الفولاذ الأمريكية أقل من 66% من المنتج المحلي الكلي للفولاذ" وهي عبارة مترتبة على السيناريو المرتبط بهذا البند والذي ذكر فيه "أن تلك الشركة كانت تنتج ضعف كمية الفولاذ الذي ينتجه كل منافسيها" وبالتالي يمكن الاستنتاج من السيناريو أن الشركة كانت تنتج حوالي 66% من الفولاذ مقارنة بباقي الشركات وبذلك فالبند تفسير صحيح مترتب على السيناريو، ولكن بالنظر لصيغة البند نجد أن طلاب التخصصات العلمية يحتمل أن يستجيبوا لهذا البند استجابة خاطئة، نظراً لظهور كلمة "أقل" ومع الوضع في الاعتبار الزمن المحدد للإجابة، وأن الاستجابة لهذا البند ترتبط أكثر بالفهم اللغوي واتقان لقاعدة "نفى النفي إثبات" تتأكد احتمالية تفوق طلاب التخصصات الأدبية في الإجابة الصحيحة للبند، فالبند لطلاب التخصصات العلمية يبدو صحيح من الوهلة الأولى خاصة لمن لا يلاحظ كلمة "أقل" في صياغة البند.
- البند 27: ينتمي لمهارات التفسير ونصه "تنتج الآن شركة الفولاذ الأمريكية فولاذا أقل مما كانت تنتجه في عام 1902م" وهي عبارة غير واردة في السيناريو الخاص بها ولذا يجب اعتبارها خاطئة وغير مترتبة على السيناريو الخاص بها، ولكن ظهور جملة "واليوم تنتج شركة الفولاذ الأمريكية حوالي 20% من الفولاذ المنتج في أمريكا" في نهاية السيناريو والذي بدأ بتضخيم انتاج الشركة وتفوقها في الإنتاج على منافسيها يجعل طلاب التخصصات الأدبية أكثر احتمالاً للاستجابة للبند بأنه وارد في السيناريو ومترتب عليه، فالإجابة لهذا البند تحتاج من الفرد القدرة على الربط بين المقدمات والنواتج وكذلك مستوى أعلى من إدراك العلاقات خاصة المرتبطة بالنسب الكمية والتي يتفوق فيها طلاب التخصصات العلمية على طلاب التخصصات الأدبية.
- البند 28: ينتمي لمهارة التفسير ونصه "معالجة الدكتور عبدالله لزيد مكنت زيد من التحسن" وهي عبارة وارده في السيناريو المرتبط بها ومترتبة عليها ولكن طلاب التخصصات الأدبية كانوا أميل للاستجابة لهذا البند بأنه غير وارد في السيناريو وقد تتأثر استجابة طلاب التخصصات الأدبية لهذا البند بطبيعة التخصص حيث أن معظم هؤلاء الطلاب من تخصصات الشريعة واللغة العربية والتربية ممن يميلون أكثر لرفض الاعتقاد بان الطبيب سبب مباشر في الشفاء بعكس طلاب التخصصات العلمية (التخصصات الطبية والتمريض والعلوم الصحية والزراعة والطب البيطرى ....) أكثر ميلاً للاستجابة لهذا البند بأنه صحيح وأنه وارد في السيناريو.

-البندان 36، 37: ينتمي البندان لمهارة تقويم الحجج والإجابة لهما هي "حجة ضعيفة"؛ والسيناريو هنا يعرض لعدالة حكم القضاء بين الأغنياء والفقراء المتخاصمين، والبند 36 نصه "لا، لأنه من المرجح أكثر أن يحل الأغنياء قضاياهم خارج المحكمة" والبند 37 نصه "لا، فمعظم القضاة يتعاطفون مع الفقراء أكثر من الأغنياء ويؤثر تعاطفهم على نتائج حكمهم"، وهنا طلاب التخصصات الأدبية أميل لاعتبار هذه الحجج قوية ومبرراً لعدم العدالة في حكم القضاة بين الأغنياء والفقراء المتخاصمين، وهي استجابة خاطئة في ضوء مفتاح التصحيح، بينما طلاب التخصصات العلمية أميل للاستجابة بأن هذه الحجج ضعيفة وهي استجابة صحيحة، ومن الواضح جداً أن الاستجابة لهاذين البندين لا ترتبط بالتفكير الناقد وإنما ترتبط بخبرات الفرد السابقة ومشاعره ومعتقداته، وتعاطفه مع الحدث وقد تكون هذه النواحي مرتفعة لدى طلاب التخصصات الأدبية.

-البند 39: ينتمي لمهارة تقويم الحجج، ونصه "لا، لأن بعض الشباب يملك ثروة تغنية عن العمل بوجه عام" وهو إجابة عن السؤال المطروح في السيناريو المرتبط بهذا السؤال والمتمثل في "هل يجب على جميع الشباب أن يلتحقوا بالعمل في الدوائر الحكومية"، وهنا أيضا زيادة احتمالية إجابة طلاب التخصصات الأدبية لهذه الحجة بأنها حجة قوية، يمكن أن يكون له ما يبرره فمعظم الخريجين من التخصصات الأدبية يعاني من عدم وجود فرصة للالتحاق بالوظيفية الحكومية ويستغرق البحث عن الوظيفة فترة طويلة نوعاً ما، وهنا قد تكون هذه الحجة منطقيه من وجهة نظرهم، أما طلاب التخصصات العلمية وخاصة في بيئة المملكة العربية السعودية فإنهم يلتحقوا بالعمل فور التخرج نظراً لندرة هذه التخصصات، وهنا يكون لدى طلاب التخصصات العلمية اعتقاد بأحقيتهم في الوظيفة الحكومية، وهذا التفسير يتأكد من صيغة البند 40 والذي ظهر أن له أداءً تفاضلياً بناء على النموذج اللابارامتري فصيغة هذا البند تتمثل في "لا، لأن فرص العمل في الدوائر الحكومية محدودة" فهنا يميل طلاب التخصصات الأدبية للاستجابة لهذا البند بأنه حجة ضعيفة، وذلك كنوع من اللوم أو الشعور بالحق في الالتحاق بالوظيفة بينما طلاب التخصصات العلمية أكثر ميلا للاستجابة للبند بأنه حجة قوية.

ويتضح مما سبق أن معظم البنود التي لها أداءً تفاضلياً تحتاج إلى إعادة صياغة أو تستبدل ببنود لا ترتبط الاستجابة لها بخبرات الفرد ومعتقداته، فالتفكير الناقد من أهم القدرات العقلية بل وهناك من يعتبره محصلة النشاط العقلي للفرد، ولذا فيجب أن تكون البنود من نوع "الأداء الأقصى"، فعلى الرغم من أن درجات بنود الاختبار موضوع الدراسة الحالية هي (صفر، 1) إلا أن العديد من البنود يغلب عليها نمط "الاستجابة المميزة" وهو ما يتناقض مع مفهوم التفكير الناقد، ويتأكد ذلك من أن العديد من بنود الاختبار قد تختلف الإجابة عليها باختلاف شخصية الفرد وخبراته وعادته ومعتقداته وظروفه التي يمر بها والتي تختلف من شخص لأخر ومن تخصص لأخر، أكثر من تأثرها بمستوى التفكير الناقد.

وفي تفسير الأداء التفاضلي في ضوء محتوى البند والسمات الثانوية المرتبطة بالإجابة الصحيحة، يجب التفريق بين العوامل الثانوية الجوهرية التي تعد عاملاً مساعداً في السمة المراد قياسها وتدخل في تحديد مستوى الفرد في السمة وبين العوامل الثانوية الخارجية والتي لا يهدف الاختبار قياسها، ففي حالة العوامل الثانوية من النوع الأول فإن الفروق في الأداء على البند بين المجموعات قد تُعبر عن فروق فعلية في السمة، أما في الحالة الثانية فإن الفروق في الأداء على البند يمكن اعتبارها أداءً تفاضلياً للبند، وهو ما يتفق مع ما أشار إليه Penfield et al., (2007, 453) حيث يؤكد على أن ظهور الأداء التفاضلي للبنود يمكن أن يكون مؤكداً للصدق البنائي للاختبار وذلك في حالة كون العوامل الثانوية التي قد تكون مسؤولة عن الفروق في الأداء على البند ثابتة في كل البنود التي لها أداءً تفاضلياً وكونها جوهرية في تحديد مستوى السمة، أما في حالة كون هذه العوامل مختلفة من بند لبند وليست جوهرية في تحديد مستوى السمة فإن الأداء التفاضلي في هذه الحالة يكون مهدداً لصدق الاختبار.

فلا يختلف أحد على علاقة التفكير الناقد بالعديد من الإمكانيات العقلية الأخرى والتي قد تكون محددة لمستوى السمة لدى الفرد وليست هناك مشكلة في كون هذه السمات هي المسببة للأداء التفاضلي للبند ففي هذه الحالة يعكس البند فروقاً حقيقية في السمة المستهدفة؛ ولكن اتضح من السابق أن الإجابة لبعض البنود التي لها أداءً تفاضلياً تتأثر بعوامل ثانوية ليست جوهرية في تحديد مستوى السمة؛ وهو ما يتفق مع ما أكدت عليه نتائج دراسات (Gierl & Bolt, 2001; Liu, 2011) من أن أهم مسببات الأداء التفاضلي للبنود محتوى البند واختلاف معناه من مجموعة لمجموعة وكذلك المتغيرات المعرفية المختلفة المطلوبة للإجابة على البند وعدم تكافؤ هذه المتغيرات بين المجموعات.

وكذلك يتفق مع ما أكد عليه (De Boeck et al., 2011, 584) وأطلق عليه اسم الأداء التفاضلي الكامن، حيث أرجع الأداء التفاضلي لاختلاف المجموعتين في بعض الخصائص الأخرى المرتبطة بالسمة المستهدفة والتي تجعل أحد المجموعتين لها الأفضلية في الإجابة الصحيحة عن البند، نظراً لأن التحليلات الإحصائية يتم فيها فقط تحييد مستوى السمة موضوع القياس في المجموعتين؛ وهو نفس ما أكد عليه Abbott (2007, 9); Price (1999,7) حيث اعتبر أن وجود بعض البنود التي تتأثر الإجابة عليها بعوامل ثانوية لا يهدف الاختبار قياسها يسهم في زيادة احتمال ظهور الأداء التفاضلي لتلك البنود؛ وكذلك يتفق مع ما أشار إليه (Tay et al., 2016, 23) من أن اعتماد التحليلات الخاصة بالكشف عن الأداء التفاضلي على المقارنة بين المجموعتين بعد تحييد مستوى السمة المقاسة قد يتأثر بعوامل ثانوية مرتبطة بالسمة ولم يتم تحييدها في المجموعتين.

2- تأكد من النتائج الحالية اختلاف نتائج الكشف عن الأداء التفاضلي للبنود باستخدام اختبار النسبة الاحتمالية في النموذج البارامتري ثنائي المعلم عنها في حالة مؤشر المساحة بين المنحنيات اللابارامترية في نموذج الانحدار اللابارامتري الممهد والذي يمكن تفسيره في ضوء التالي:

صغر حجم كل مجموعة من مجموعتي الدراسة (المرجعية والبؤرية) ففي هذه الحالة قد يكون توزيع السمة الكامنة في المجموعتين ملتوياً لدرجة ما – معامل الالتواء يعتمد على الدرجات الملاحظة وليس السمة الكامنة – وهو ما يؤثر على تقدير معالم البند في حالة الاعتماد على النماذج البارامترية والتي يتطلب تطبيقها مجموعات كبيرة الحجم، وفي هذه الحالة يمكن اعتبار النماذج اللابارامترية أكثر منعة وضلاعة في الأداء، فاعتماد النماذج البارامترية في الكشف عن الأداء التفاضلي على النسبة الاحتمالية وفي حالة صغر حجم العينة يبتعد توزيع النسبة الاحتمالية G^2 عن توزيع مربع كاي والذي قد ينتج عنه عدم دقة في أداء مؤشر النسبة الاحتمالية في هذه الحالة.

كذلك يلاحظ من نتائج الدراسة الحالية أن نسبة عدد البنود التي لها أداءً تفاضلياً بلغ في حالة الاعتماد على اختبار النسبة الاحتمالية حوالي 27.5% من عدد البنود وفي ضوء عدم وجود دراسات سابقة في حدود اطلاع الباحث يمكن الاعتماد عليها في تحديد بنود الاختبار المؤكد عدم وجود أداء تفاضلي لها لتستخدم كجزع مشترك لتعيير البنود في المجموعتين تم الاعتماد على طريقة "كل البنود ماعدا البند" All Others as Anchors كجزع مشترك، وفي هذه الحالة من الممكن أن يتأثر أداء النسبة الاحتمالية بالبنود التي لها أداءً تفاضلياً، فوجود هذه البنود في عملية التعيير ووضع بارامترات البنود في المجموعتين على نفس التدريج قد يتسبب في عدم دقة هذه التقديرات والذي قد ينعكس سلباً على أداء اختبار النسبة الاحتمالية وهو ما يتفق مع نتائج الدراسات السابقة التي تؤكد على أن اختبار النسبة الاحتمالية يعد من أكثر الطرق دقة في الكشف عن الأداء التفاضلي للبند حتى في حالة الأداء التفاضلي البسيط وصغر حجم العينة بشرط أن يتم استخدام بنوداً ليس لها أداءً تفاضلياً كجزع مشترك DIF-Free Anchor لتعيير البنود (Woods, 2009; Rivas et al., 2009؛ Khalid & Glas, 2014) وكذلك تتفق مع نتائج دراسة (Meade & Wright, 2012) والتي أكدت على أن وجود بنوداً لها أداءً تفاضلياً في الجزع المشترك يزيد من نسبة الخطأ من النوع الثاني لاختبار النسبة الاحتمالية وهو ما يعنى عدم قدرة الاختبار على الكشف عن بعض البنود ذات الأداء التفاضلي.

كذلك يمكن تفسير النتائج الحالية في ضوء أن العديد من البنود كان الأداء التفاضلي لها غير منتظم وفي هذه الحالة فإن النماذج البارامترية المعتمدة على اختبار النسبة الاحتمالية وتوزيع مربع كاي تميل للتقليل من الخطأ من النوع الأول –عدم الكشف عن بنود يفترض أن لها أداءً تفاضلياً– ويؤكد ذلك التفسير ما أشارت إليه نتائج دراسة (Wang et al., 2013) والتي أكدت على انخفاض نسبة الخطأ من النوع الأول عند استخدام اختبار النسبة الاحتمالية في حالة الأداء التفاضلي غير المنتظم؛ وفي ضوء ذلك يمكن استنتاج أن الاعتماد على مؤشر المساحة بين المنحنيات اللابارامترية المميزة للبند أكثر حساسية (قوة الاختبار) للأداء التفاضلي غير المنتظم والمنتظم البسيط.

وعلى الرغم من أن التفسيرات السابقة لا تتفق مع افتراض أن الاعتماد على النماذج البارامترية يسهم كثيراً في ضبط تأثير عدم تماثل توزيع السمة في المجموعتين على الخطأ من النوع الأول والخطأ من النوع الثاني في الكشف عن الأداء التفاضلي للبنود (Walker, 2011, 366)؛ لكن قد يكون ذلك مشروطاً بكبر حجم العينة.

والنتائج السابقة تؤكد في مجملها على تمتع مؤشر المساحة بين المنحنيات اللابارامترية المميزة للبند بقوة مرتفعة في اكتشاف البنود التي يُفترض فعلاً أن لها أداءً تفاضلياً في حالة عدم تحقق شروط النماذج البارامترية؛ ويتماشى ذلك الاستنتاج مع ما أكده (Douglas & Cohen, 2001) أنه في حالة اختلاف نتائج تطبيق النماذج البارامترية والنماذج اللابارامترية فإن النتائج المعتمدة على النماذج اللابارامترية هي الأقرب للصواب؛ وما أكدته نتائج دراسة (Lee et al., 2004) من أن تقدير المنحنى المميز للبند باستخدام النماذج اللابارامترية وبصفة خاصة نموذج Ramsay يتميز بالدقة مقارنة بالنموذج البارامتري ثنائي المعلم والذي تزيد دقته بزيادة حجم العينة.

ويعنى ما سبق أن النماذج البارامترية أكثر منعة من حيث نسبة الخطأ من النوع الأول وتميل التقديرات فيها إلى قبول الفرض الصفري، بعكس التقديرات اللابارامترية والتي تميل لأن تكون أكثر منعة للخطأ من النوع الثاني حيث تميل التقديرات فيها لرفض الفرض الصفري، وهو ما يتفق بدرجة كبيرة مع نتائج دراسة (Ozdemir, 2015)؛ ودراسة (Rivas et al., 2009) والتي أكدت على أن نسبة الخطأ من النوع الأول لاختبار النسبة الاحتمالية في الكشف عن الأداء التفاضلي للبنود منخفض جداً.

3- تأكد من النتائج الحالية تأثير البنود التي لها أداءً تفاضلياً على الأداء التفاضلي للاختبار ككل وأن هذه البنود تؤثر على تماثل البنية الكامنة المقاسة باستخدام الاختبار لدى طلاب التخصصات العلمية وطلاب التخصصات الأدبية وهو ما يمكن تفسيره في التالي:

زيادة عدد البنود التي لها أداءً تفاضلياً في الاختبار والتي بلغت في حالة النموذج البارامتري 27.5% من عدد البنود وبلغت في حالة النموذج اللابارامتري ما نسبته 42.5% من عدد البنود وهو ما يتفق مع نتائج العديد من الدراسات التي أشارت إلى أنه في حالة زيادة عدد البنود التي لها أداءً تفاضلياً لا يكون للدرجات النهائية في الاختبار نفس المعنى في المجموعات المختلفة ويتسبب في تحيز تقدير السمة (LeBouthillier et al., 2015; Walker, 2011; Stump et al., 2005; Church et al., 2011).

كذلك يمكن تفسير تلك النتيجة في ضوء أن الأداء التفاضلي لمعظم البنود كان كبيراً في ضوء قيمة β كمؤشر للأداء التفاضلي وهو ما يفسر تأثير الأداء التفاضلي للبنود على الأداء التفاضلي الكلي للاختبار.

واختلاف النتيجة الحالية المتعلقة بتأثير البنود التي لها أداءً تفاضلياً على الأداء التفاضلي الكلي للاختبار مع بعض نتائج الدراسات السابقة والتي أكدت على عدم تأثير البنود التي لها أداءً تفاضلياً على الأداء التفاضلي الكلي للاختبار كدراسات (Baker et al., 2007; Pae, 2004; Takala & Kaftandjieva, 2000; Zumbo, 2003; Sahin et al., 2015; French et al., 2012)؛ يمكن تفسيره في ضوء الأسلوب الإحصائي المستخدم في التحقق من مدى تماثل البنية الكامنة المقاسة في المجموعتين البؤرية والمرجعية، حيث تمثل الإجراءات المعتمدة على التحليل العاملي

التوكيدي أكثر الإجراءات دقة في اكتشاف عدم تغير نموذج القياس في المجموعات المختلفة (Abbott, 2007,12; Teresi & Fleishman, 2007, 37; Koh & Zumbo, 2008, 471; Milfont & Fischer, 2010, 112; Wolf et al., 2014; Byrne, 2010؛

التوصيات والمقترحات:

- تأكد من نتائج الدراسة الحالية أن بعض بنود اختبار واطسون وجليسر للتفكير الناقد "النسخة القصيرة" لها أداءً تفاضلياً، وبالتالي فهذه البنود تحتاج إلى إعادة صياغة أو حذفها من الاختبار حتى نزيد من العدالة في المقارنة بين المجموعات المختلفة في التفكير الناقد باستخدام ذلك الاختبار.
- قد يكون من المناسب إجراء نفس الدراسة الحالية باستخدام النموذج البارامتري ثلاثي المعلم والاعتماد على عينات كبيرة الحجم حيث أن الكشف عن تأثير التخمين قد يوضح العديد من الأمور التي تسهم في فهم كيفية الاستجابة لبنود اختبار واطسون وجليسر للتفكير الناقد خاصة وأن العديد من بنود الاختبار الاستجابات لها ثنائية مما يزيد من نسبة التخمين في الإجابة لها.
- النتائج السابقة والمتعلقة بمقارنة أداء النموذج البارامتري ثنائي المعلم والنموذج اللابارامتري تحتاج إلى مزيداً من الجهود البحثية لتأكيدها.

المراجع:

- كاظم، أمينة محمد (1988). ***استخدام نموذج راش في بناء اختبار تحصيلي في علم النفس وتحقيق التفسير الموضوعي للنتائج***. الكويت: مطبوعات جامعة الكويت.
- كاظم، أمينة محمد (1994). تدريج ومعايرة المقاييس. في: عماد الدين اسماعيل، ناهد رمزي، ليلى كرم، أمينة كاظم، هدى شريف (محررون): ***معايير نمو طفل ما قبل المدرسة***، المجلد (2)، الدراسة النفسية، القاهرة: المجلس القومى للطفولة والأمومة، ص ص 114-232.
- سعادة، جودة احمد (2009). أثر عدد من المتغيرات في اكتساب طلبة الجامعة لمهارات التفكير الناقد. ***دراسات***، العلوم التربوية، عمادة البحث العلمي، الجامعة الأردنية، 36، 205-256.
- العتيبي، خالد ناهس (2012). الخصائص السيكومترية للصورة القصيرة من اختبار واطسون – جليسر للتفكير الناقد (WGCT-SF): دراسة على عينة من الطلاب/المعلمين. ***المجلة العربية للتربية***، تونس، 32(2)، 180-143.
- الحدابي، داود عبدالملك (2012). مستوى التفكير الناقد لدى طلبة جامعة العلوم والتكنولوجيا اليمنية. ***المؤتمر العلمي العربي التاسع لرعاية الموهوبين والمتفوقين***، المجلس العربي للموهوبين والمتفوقين، (10-11 نوفمبر).
- الجاف، رشدي على وسليمان، خديجة حسين (2012). مستوى التفكير الناقد لدى طلبة الجامعة وعلاقته ببعض المتغيرات. ***العلوم التربوية والنفسية***، العراق، العدد 87، 177-242.
- سليمان، السيد عبدالحميد (2006). التفكير الناقد وعلاقته بالذكاء والدافع للإنجاز وموضع الضبط ونوع التعليم لدى عينة من طلاب المدارس الثانوية. ***دراسات تربوية واجتماعية***، جامعة حلون، 2(3)، 145-119.
- المبدل، عبدالمحسن رشيد (2010). ***المكونات الإيجابية للبيئة الصفية في ضوء نظرية موراي وعلاقتها بمهارات التفكير الناقد***. رسالة دكتوراه غير منشورة، كلية العلوم الاجتماعية، جامعة الإمام محمد بن سعود الإسلامية.
- عامر، عبدالناصر السيد (2004). أداء مؤشرات حسن المطابقة لتقويم نموذج المعادلة البنائية. ***المجلة المصرية للدراسات النفسية***، تصدرها الجمعية المصرية للدراسات النفسية، 45(14)، 105-157.

- منصور، علي وعلي، لينا (2010). مستويات التفكير الناقد وعلاقتها بمتغيري الجنس والفرع الدراسي، دراسة ميدانية على عينة من طلبة الصف الثاني الثانوي في مدارس محافظة دمشق الرسمية. ***مجلة جامعة تشرين للبحوث والدراسات العلمية***، سلسلة الآداب والعلوم الإنسانية، 32(4)، 141-156.
- السكر، عماد الدين محمد (2010). بعض الخصائص السيكومترية لاختبار واطسون-جليسر للتفكير الناقد الصورة "S" لدى عينة من طلاب الجامعة. ***المجلة المصرية للدراسات النفسية***، الجمعية المصرية لعلم النفس، 20(67)، 1-25.
- المالكي، فهد عبدالله (2012). ***نمذجة العلاقات بين مداخل تعلم الإحصاء ومهارات التفكير الناقد والتحصيل الأكاديمي لدى طلاب جامعة أم القرى***. رسالة ماجستير غير منشورة، كلية التربية، جامعة أم القرى.
- أبومسلم، مايسة فاضل (2010). الكشف عن تحيز بنود الاختبار وفقاً لمدخل نظرية الاستجابة للمفردة (IRT) ومدخل مانتل هانزل (M-H) دراسة مقارنة. ***مجلة كلية التربية***، جامعة الإسكندرية، 20(2)، 188-236.
- حلفاوي، مسعف (1997). **اشتقاق معايير الأداء على مقياس التفكر الناقد لطلبة البكالوريوس في الجامعات الحكومية الأردنية**. رسالة ماجستير غير منشورة. عمان: الجامعة الأردنية.
- الحموري، هند والوهر، محمود (1998). تطور القدرة على التفكير الناقد وعلاقة ذلك بالمستوى العمري والجنس وفرع الدراسة. ***دراسات***، العلوم التربوية، عمادة البحث العلمي، الجامعة الأردنية، 25(1)، 112-126.

- Abbott, M. (2007). A Confirmatory Approach to Differential Item Functioning on an ESL Reading Assessment. ***Language Testing***, 24(1), 7–36.
- Andrich, D. & Hagquist, H. (2012). Real and Artificial Differential Item Functioning. ***Journal of Educational and Behavioral Statistics***, 37(3), 387–416.
- Ariffin, S.; Idris, R. & Ishak, N. (2010). Differential Item Functioning in Malaysian Generic Skills Instrument (MyGSI). ***Journal Pendidikan Malaysia***, 35(1), 1-10.
- Baghi, H. & Ferrara, S. (1989). A Comparison of IRT, Delta-Plot, and Mantel-Haenszel Techniques for Detecting Differential Item Functioning Across Subpopulations in the Maryland Test of Citizenship Skills. ***Paper Presented at the Annual Meeting of the American Educational Research Association***, San Francisco, (28 March).
- Baker, F. (2001). ***The Basics of Item Response Theory***. ERIC Clearinghouse on Assessment and Evaluation, USA, ISBN 1-886047-03-0.
- Baker, B.; Caison, A. & Meade, A. (2007). Assessing Gender-Related Differential Item Functioning and Predictive Validity with the Institutional Integration Scale. ***Educational and Psychological Measurement***, 67(3), 545-559.
- Basokcu, O. & Ogretmen, T. (2014). Comparison of Parametric and Nonparametric Item Response Techniques in Determining Differential Item Functioning in Polytomous Scale. ***American Journal of Theoretical and Applied Statistics***, 3(2), 31-38.

- Bechger, T. & Maris, G. (2015). A Statistical Test for Differential Item Pair Functioning. ***Psychometrika***, 80(2), 317–340.
- Baker, B.; Caison, A. & Meade, A. (2007). Assessing Gender-Related Differential Item Functioning and Predictive Validity with the Institutional Integration Scale. ***Educational and Psychological Measurement***, 67(3), 545-559.
- Byrne, B. (2010). ***Structural Equation Modeling with AMOS: Basic Concepts, Applications, and Programming***. 2nd ed., New York: Routledge, Taylor & Francis Group.
- Cheung, G. & Rensvold, R. (2002). Evaluating Goodness-of-Fit Indexes for Testing Measurement Invariance. ***Structural Equation Modeling***, 9(2), 233–255.
- Cho, Y.; Martin, M.; Conger, R. & Widaman, K. (2010). Differential Item Functioning on Antisocial Behavior Scale Items for Adolescents and Young Adults from Single-Parent and Two-Parent Families. ***Journal of Psychopathology Behavioral Assessment***, 32(2), 157–168.
- Church, A.; Alvarez, J.; Mai, N.; French, B.; Katigabk, M. & Ortiz, F. (2011). Are Cross-Cultural Comparisons of Personality Profiles Meaningful? Differential Item and Facet Functioning in the Revised NEO Personality Inventory. ***Journal of Personality and Social Psychology***, 101(5), 1068–1089.
- DeBoeck, P.; Cho, S. & Wilson, M. (2011). Explanatory Secondary Dimension Modeling of Latent Differential Item Functioning**.** ***Applied Psychological Measurement***, 35(8), 583–603.
- Douglas, J. & Cohen, A. (2001). Nonparametric Item Response Function Estimation for Assessing Parametric Model Fit. ***Applied Psychological Measurement***, 25(3), 234–243.
- Douglas, L. (1997). Joint Consistency of Nonparametric Item Characteristic Curve and Ability Estimation. ***Psychometrika***, 62(1), 7-28.
- Drasgow, F. (1987). Study of the Measurement Bias of Two Standardized Psychological Tests. ***Journal of Applied Psychology,*** 72(1), 19–29.
- Drasgow, F.; Levine, M.; Williams, B.; McLaughlin, M. & Candell, G. (1989). Modeling Incorrect Responses to Multiple-Choice Items with Multilinear Formula Score Theory. ***Applied Psychological Measurement***, 13(3), 285-299.
- Duncan, K. & MacEachern, S. (2008). Nonparametric Bayesian Modelling for Item Response. ***Statistical Modelling***, 8(1), 41-66.
- Ferrando, P. (2004). Kernel-Smoothing Estimation of Item Characteristic Functions for Continuous Personality Items: An Empirical Comparison with the Linear and the Continuous-Response Models. ***Applied Psychological Measurement***, 28(2), 95–109.

- Fidalgo, A.; Ferreres, D. & Muniz, J. (2004). Liberal and Conservative Differential Item Functioning Detection Using Mantel-Haenszl and SIBTEST: Implication for Type 1 and Type 11 Error Rates. ***The Journal of Experimental Education***, 73(1), 23-39.
- Finch, W. & French, B. (2007). Detection of Crossing Differential Item Functioning a Comparison of Four Methods. ***Educational and Psychological Measurement***, 67(4), 565-582.
- French, B. & Maller, S. (2007). Iterative Purification and Effect Size Use with Logistic Regression for Differential Item Functioning Detection. ***Educational and Psychological Measurement***, 67(3), 373-393.
- French, B.; Hand, B.; Therrien, W. & Vazquez, J. (2012). Detection of Sex Differential Item Functioning in the Cornell Critical Thinking Test. ***European Journal of Psychological Assessment***, 28(3), 201–207.
- Gierl, M. & Bolt, D. (2001). Illustrating the Use of Nonparametric Regression to Assess Differential Item and Bundle Functioning Among Multiple Groups. ***International Journal of Testing***, 1(3&4), 249–270.
- Gomes, R. (2012). Parent Ratings of ADHD Symptoms: Generalized Partial Credit Model Analysis of Differential Item Functioning Across Gender. ***Journal of Attention Disorders***, 16(4), 276-283.
- Gotzmann, A. (2002). The Effect of Large Ability Differences on Type I Error and Power Rates Using SIBTEST and TESTGRAF DIF Detection Procedures. ***Paper Prepared at the Annual Meeting of the American Educational Research Association***, New Orleans, LA, (1-5 April).
- Guo, H. & Sinharay, S. (2010). Measurement Error in Nonparametric Item Response Curve Estimation. ***Paper Presented at the Annual Meeting of the National Council on Measurement in Education (NCME)***, Denver, (29 April-3 May).
- Kalaycioglu, D. & Berberoglu, G. (2011). Differential Item Functioning Analysis of the Science and Mathematics Items in the University Entrance Examinations in Turkey. ***Journal of Psychoeducational Assessment***, 29(5), 467-478.
- Kang, T. & Cohen, A. (2007). IRT Model Selection Methods for Dichotomous Items. ***Applied Psychological Measurement***, 31(4), 331–358.
- Keller, L. (2000). ***Ability Estimation Procedures in Computerized Adaptive Testing.*** Technical Report. New York: American Institute of Certified Public Accountants.
- Khalid, M. & Glas, C. (2014). A Scale Purification Procedure for Evaluation of Differential Item Functioning. ***Measurement***, 50, 186-197.
- Khan, A.; Lewis, C. & Lindenmayer, J. (2011). Use of Non-Parametric Item Response Theory to Develop a Shortened Version of the Positive and Negative Syndrome Scale (PANSS). ***Bio Med Central Psychiatry***, 11(178), 1-23.

- Kim, J. & Oshima, T. (2012). Effect of Multiple Testing Adjustment in Differential Item Functioning Detection. ***Educational and Psychological Measurement***, 73(3), 458–470.
- Kline, T. (2004). Gender and Language Differences on The Test of Workplace Essential Skills: Using Overall Mean Scores and Item-Level Differential Item Functioning Analyses. ***Educational and Psychological Measurement***, 64(3), 549-559.
- Kline, R. (2005). Principles and practice of ***Structural Equation Modeling***. (2nd ed.), New York: The Guilford Press.
- Koh, K. & Zumbo, B. (2008). Multi-Group Confirmatory Factor Analysis for Testing Measurement Invariance in Mixed Item Format Data. ***Journal of Modern Applied Statistical Methods***, 7(2), 470-477.
- Lai, J.; Teresi, J. & Gershon, R. (2005). Procedures for the Analysis of Differential Item Functioning (DIF) For Small Sample Sizes. ***Evaluation & the Health Professions***, 28(3), 283-294.
- LeBouthillier, D.; Thibodeau, M.; Alberts, N.; Hadjistavropoulos, H. & Asmundson, G. (2015). Do People with and without Medical Conditions Respond Similarly to the Short Health Anxiety Inventory? An Assessment of Differential Item Functioning Using Item Response Theory. ***Journal of Psychosomatic Research***, 78(4), 384–390.
- Lee, Y.; Wollack, J. & Douglas, J. (2004). On the Use of Nonparametric ICC Estimation Techniques For Checking Parametric Model Fit. ***Paper presented at the annual conference of the American Educational Research Association***, San Diego, CA, (27 March).
- Lee, Y. (2007). A Comparison of Methods for Nonparametric Estimation of Item Characteristic Curves for Binary Items. ***Applied Psychological Measurement***, 31 (2), 121–134.
- Lee, Y.; Wollacl, J. & Douglas, J. (2009). On the Use of Nonparametric Item Characteristic Curve Estimation Techniques for Checking Parametric Model Fit. ***Educational and Psychological Measurement***, 69(2), 181-197.
- Lee, H. & Geisinger, K. (2016). The Matching Criterion Purification for Differential Item Functioning Analyses in a Large-Scale Assessment. ***Educational and Psychological Measurement***, 76(1), 141-163.
- Liu, Q. (2011). ***Item Purification in Differential Item Functioning Using Generalized Linear Mixed Models***. PH.D Dissertation, Florida State University, College of Education.
- Luciano, J.; Ayuso-Mateos, J.; Aguado, J.; Fernandez, A.; Serrano-Blanco, A.; Roca, M. & Haro, J. (2010). The 12-item World Health Organization Disability Assessment Schedule II (WHO-DAS II): A Nonparametric Item Response Analysis. ***BMC Medical Research Methodology***, 10(45), 1-9.

- Magis, D. & DeBoeck, P. (2012). A Robust Outlier Approach to Prevent Type I Error Inflation in Differential Item Functioning. ***Educational and Psychological Measurement***, 72 (2), 291-311.
- Magis, D. & Facon, B. (2013). Item Purification Does Not Always Improve DIF Detection: A Counterexample With Angoff's Delta Plot. ***Educational and Psychological Measurement***, 73(2), 293–311.
- Magis, D. & DeBoeck, P. (2014). Type I Error Inflation in DIF Identification With Mantel–Haenszel: An Explanation and a Solution. ***Educational and Psychological Measurement***, 74(4), 713–728.
- Meade, A. & Fetzer, M. (2009). Test Bias, Differential Prediction, and a Revised Approach for Determining the Suitability of a Predictor in a Selection Context. ***Organizational Research Methods***, 12(4), 738-761.
- Meade, A. & Wright, N. (2012). Solving the Measurement Invariance Anchor Item Problem in Item Response Theory. ***Journal of Applied Psychology***, 97(5), 1016–1031.
- Meijer, R. (2004). ***Investigating the Quality of Items in CAT Using Nonparametric IRT***. Law School Admission Council Computerized Testing Report, Newtown, Law School Admission Council, Inc.
- Meyer, J. (2014). ***Applied Measurement with JMetrik***. New York: Routledge, Taylor & Francis Group.
- Milfont, T. & Fischer, R. (2010). Testing Measurement Invariance across Groups: Applications in Cross Cultural Research. ***International Journal of Psychological Research***, 3(1), 111-121.
- Molenaar, I. (2001). Thirty Years of Nonparametric Item Response Theory. ***Applied Psychological Measurement***, 25(3), 295–299.
- Molenaar, I. (2003). Groups of Persons and Groups of Items in Nonparametric Item Response Theory. In: H. Yanai; A. Okada; K. Shigemasu; Y. Kano & J. Meulman (Eds.). ***New Developments in Psychometrics***, (Pp 191-198), Japan: Springer.
- Monahan, P. & Ankenmann, R. (2010). Alternative Matching Scores to Control Type I Error of the Mantel–Haenszel Procedure for DIF in Dichotomously Scored Items Conforming to 3PL IRT and Nonparametric 4PBCB Models. ***Applied Psychological Measurement***, 34(3), 193–210.
- Nay, R.; Strong, D.; Nay, W.; Beidel, D. & Turner, S. (2007). Development of an Abbreviated Social Phobia and Anxiety Inventory (SPAI) Using Item Response Theory: The SPAI-23. ***Psychological Assessment***, 19(1), 133–145.
- Ownby, R. & Waldrop-Valverde, D. (2013). Differential Item Functioning Related to Age in the Reading Subtest of the Test of Functional Health Literacy in Adults. ***Journal of Aging Research***, Open Access Journal, Article ID 654589, 6 pages.

- Ozdemir, B. (2015). A Comparison of IRT-Based Methods for Examining Differential Item Functioning in TIMSS 2011 Mathematics Subtest. ***Procedia-Social and Behavioral Sciences***, 174, 2075 – 2083.
- Pae, T. & Park, G. (2006). Examining the Relationship between Differential Item Functioning and Differential Test Functioning. ***Language Testing***, 23 (4), 475–496.
- Pae, T. (2004). DIF for Examinees with Different Academic Backgrounds. ***Language Testing***, 21(1), 53–73.
- Pastsula, L. & Gessaroli, M. (1995). A Comparison of Item Parameter Estimates and ICCs Produced with TESTGRAF and BILOG under Different Test Lengths and Sample Sizes. ***Paper presented at the Annual Meeting of the National Council on Measurement in Education***, San Francisco, CA, April.
- Penfield, R.; Giacobbi, P. & Myers, N. (2007). Using the Cumulative Common Log-Odds Ratio to Identify Differential Item Functioning of Rating Scale Items in the Exercise and Sport Sciences. ***Research Quarterly for Exercise and Sport***, 78(5), 451-464.
- Price, L. (1999). Differential Functioning of Items and Tests Versus the Mantel-Haenszel Technique for Detecting Differential Item Functioning in a Translated Test. ***Paper Presented at the Annual Meeting of the American Alliance of Health, Physical Education, Recreation, and Dance***, Boston, (12-16 April).
- Raju, N.; Laffittr, L. & Byrne, B. (2002). Measurement Equivalence: A Comparison of Methods Based on Confirmatory Factor Analysis and Item Response Theory. ***Journal of Applied Psychology***, 87(3), 517–529.
- Ramsay, J. & Silverman, B. (1997). ***Functional Data Analysis***. New York: Springer.
- Ramsay, J. (1991). Kernel Smoothing Approaches to Nonparametric Item Characteristic Curve Estimation. ***Psychometrika***, 56(4), 611-630.
- Ramsay, J. (2000). ***TestGraf Manual.*** Montreal, Quebec, Canada: McGill University.
- Ramsay, J. (1998). Estimating Smooth Monotone Functions. ***Journal of the Royal Statistical Society***, 60(2), 365–375.
- Rivas, G; Stark, S. & Chernyshenko, O. (2009). The Effects of Referent Item Parameters on Differential Item Functioning Detection Using the Free Baseline Likelihood Ratio Test. ***Applied Psychological Measurement***, 33(4), 251-265.
- Runnels, J. (2013). Measuring Differential Item and Test Functioning Across Academic Disciplines. ***Language Testing in Asia***, 3(9), 1-11.
- Sahin, H.; French, B.; Hand, B. & Gunel, M. (2015). Detection of Differential Item Functioning in the Cornell Critical Thinking Test

between Turkish and United States Students. ***European Journal of Psychological Assessment***, 31(4), 238–246.

- Santelices, M. & Wilson, M. (2012). On the Relationship Between Differential Item Functioning and Item Difficulty: An Issue of Methods? Item Response Theory Approach to Differential Item Functioning. ***Educational and Psychological Measurement,*** 72(1), 5–36.
- Schmitt, A. & Dorans, N. (1988). ***Differential Item Functioning for Minority Examinees on the SAT***. Research Report, Educational Testing Service, Princeton, New Jersey.
- Schulz, W. (2005). Testing Parameter Invariance for Questionnaire Indices Using Confirmatory Factor Analysis and Item Response Theory. ***Paper Presented at the Annual Meetings of the American Educational Research Association***, San Francisco, (7-11 April).
- Scientific Software International (2015). ***IRTPRO: User's Guide***. USA: Scientific Software International, Inc.
- Sijtsma, K. & Molenaar, I. (1987). Reliability of Test Scores in Nonparametric Item Response Theory. ***Psychometrika***, 52(1), 79-97.
- Sijtsma, K. (2003). Developments in Practical Nonparametric IRT Scale Analysis. In: H. Yanai; A. Okada; K. Shigemasu; Y. Kano & J. Meulman (Eds.). ***New Developments in Psychometrics***, (Pp 183-190), Japan: Springer.
- Stump, T.; Monahan P. & Mchorney, C. (2005). Differential Item Functioning in the Short Portable Mental Status Questionnaire. ***Research on Aging***, 27(3), 355-384.
- Sueiro, M. & Abad, F. (2011). Assessing Goodness of Fit in Item Response Theory with Nonparametric Models: A Comparison of Posterior Probabilities and Kernel-Smoothing Approaches. ***Educational and Psychological Measurement***, 71(5), 834–848.
- Takala, S. & Kaftandjieva, F. (2000). Test Fairness: A DIF Analysis of an L2 Vocabulary test. ***Language Testing*** 17(3), 323–40.
- Takano, Y; Tsunoda, S. & Muraki, M. (2014). ***Mathematical Optimization Models for Nonparametric Item Response Theory***. Technical Report, Institute of Technology, Department of Industrial Engineering and Management, Tokyo
- Tan, X. & Gierl, M. (2005). Using Global and Local DIF Analyses to Assess DIF across Language. ***Paper presented at the annual conference of the National Council on Measurement and Evaluation***, Montreal, Quebec, Abril.
- Tay, L.; Huang, Q. & Vermunt, J. (2016). Item Response Theory with Covariates (IRT-C): Assessing Item Recovery and Differential Item

Functioning for the Three-Parameter Logistic Model. ***Educational and Psychological Measurement***, 76(1), 22–42.

- Teresi, J. & Fleishman, J. (2007). Differential Item Functioning and Health Assessment. ***Quality Life Research***, 16(1), 33-42.
- Thissen, D. (2001). IRTLRDIF v.2.0b: Software for the Computation of the Statistics Involved in Item Response Theory Likelihood-Ratio Tests for Differential Item Functioning. Retrieved 25/1/2016, from ***http://www.unc.edu/~dthissen/dl.html***.
- Velicer, W.; Eaton, C. & Fava, J. (2000). Construct Explication through Factor or Component Analysis: A Review and Evaluation of Alternative Procedures for Determining the Number of Factors or Components. In R. D. Goffin & E. Helmes (Eds.). ***Problems and Solutions in Human Assessment***. (Pp. 41-71). New York: Kluwer Academic/Plenum
- Walker, C. (2011). What's the DIF? Why Differential Item Functioning Analyses Are an Important Part of Instrument Development and Validation. ***Journal of Psychoeducational Assessment***, 29(4), 364–376.
- Wang, W. (2012). ***Are All Item Response Functions Monotonically Increasing***?. PH.D Dissertation, department of psychology and Research in Education and the Graduate Faculty, University of Kansas.
- Wang, W.; Tay, L. & Drasgow, F. (2013). Detecting Differential Item Functioning of Polytomous Items for an Ideal Point Response Process. ***Applied Psychological Measurement***, 37(4), 316–335.
- Wells, C.; Cohen, A. & Patton, J. (2009). A Range-Null Hypothesis Approach for Testing DIF under the Rasch Model. ***International Journal of Testing***, 9 (4), 310–332.
- Whittaker, T.; Chang, W. & Dodd, B. (2012). The Performance of IRT Model Selection Methods with Mixed-Format Tests. ***Applied Psychological Measurement***, 36(3), 159–180.
- Witarsa, P. (2003). ***Nonparametric Item Response Modeling for Identifying Differential Item Functioning in the Moderate-to Small Scale Testing Context.*** PH.D Dissertation, Department of Educational and Counselling Psychology and Special Education, University of British Columbia.
- Wolf, R.; Zahner, D.; Kostoris, F. & Benjamin, R. (2014). ***A Case Study of an International Performance-Based Assessment of Critical Thinking Skills***. New York: Council for Aid to Education.
- Woods, C. (2008). Likelihood-Ratio DIF Testing: Effects of Non-Normality. ***Applied Psychological Measurement***, 32(7), 511–526.
- Woods, C. (2009). Empirical Selection of Anchors for Tests of Differential Item Functioning. ***Applied Psychological Measurement***, 33(1), 42–57.

- Xu, X.; Douglas, J. & Lee, Y. (2011). Linking With Nonparametric IRT Models. In A. A. Von Davier (ed.), ***Statistical Models for Test Equating, Scaling, and Linking***, (Pp. 243-258), New York: Springer.
- Zheng, Y.; Gierl, M. & Cui, Y. (2010). Using Cochran's Z Statistic to Test the Kernel-Smoothed Item Response Function Differences between Focal and Reference Groups. ***Educational and Psychological Measurement***, 70(4), 541–556.
- Zumbo, B (2007). Three Generations of DIF Analyses: Considering Where It Has Been, Where It Is Now, and Where It Is Going. ***Language Assessment Quarterly***, 4(2), 223–233
- Zumbo, B. (2003). Does Item-Level DIF Manifest Itself in Scale-Level Analyses? Implications for Translating Language Tests. ***Language Testing***, 20(2), 136–47.