التكامل بين التحليل الإحصائي والبياني في انتقاء وتدريج مفردات الاختيار من متعدد باستخدام نظرية الاستجابة للمفردة

د/ شادية عبد العزيز مهتدي منتصر
مدرس علم النفس التربوي (قياس وتقويم)
كلية البنات للآداب والعلوم والتربية – جامعة عين شمس

ملخص الدراسة:

هدفت الدراسة إلى اختبار الاتساق بين محكات التحليل البياني والمحكات الإحصائية للملاءمة لنموذج راش في تقدير جودة المفردة الاختبارية، ودراسة أثر استخدام التحليل البياني والتحليل الإحصائي معا في انتقاء المفردات التي تحقق دقة القياس على خصائص كل من التدريج النهائي للاختبار والتحليل البياني النهائي لمفرداته. وتكونت عينة الدراسة من (١٩٩) طالبة من طالبات الفرق الرابعة التربوية بكلية البنات جامعة عين شمس، وتمثلت أداة الدراسة في مفردات الاختيار من متعدد (٣٠ مفردة) بأحد النماذج الاختبارية للامتحان النهائي في مقرر نظريات التعلم. وقد تم تطوير ملف أوامر Syntax لإجراء التحليل البياني للمفردات، كما استخدم برنامج Winsteps لتدريج الاختبار تبعا لنموذج راش.

وبعد إجراء المعالجات الإحصائية والتحليل الكيفي المناسب، توصلت الدراسة للنتائج التالية:

- يوجد اتساق وتكامل بين محكات التحليل البياني والمحكات الإحصائية للملاءمة لنموذج راش في تقييم جودة المفردة الاختبارية من نوع الاختيار من متعدد.
- إن حذف المفردات غير الملائمة وفقا لمحكات التحليل البياني قبل تدريج الاختبار تبعا لنموذج راش، لا يؤثر على خصائص الاختبار في تدريجه النهائي. كما لا يؤثر على تدرج الأفراد على المتغير موضوع القياس، في حين يؤدي إلى انخفاض نسبة الأفراد غير الملائمين، وارتفاع معامل ثبات تقديرات قدرة عينة التدريج النهائي، وارتفاع قيم الخطأ المعياري لهذه التقديرات، وتحسن الملاءمة التباعدية للأفراد وفق مؤشر ZSTD.
- إن حذف المفردات غير الملائمة لأسس القياس وفقا لكل من المحكات البيانية والإحصائية وحذف الأفراد غير الملائمين، يؤدي إلى تحسن نتائج التحليل البياني لباقي المفردات.

الكلمات المفتاحية: التحليل البياني للمفردات (GIA) – نظرية الاستجابة للمفردة – نموذج راش – مفردات الاختيار من متعدد – دقة القياس.

التكامل بين التحليل الإحصائي والبياني في انتقاء وتدريج مفردات الاختيار من متعدد باستخدام نظرية الاستجابة للمفردة

د/ شادية عبد العزيز مهتدي منتصر
مدرس علم النفس التربوي
كلية البنات جامعة عين شمس

مقدمة:

يفيد الواقع العملي بأن أغلب الاختبارات التحصيلية في مراحل التعليم المختلفة في مصر تعتمد في بنائها على النظرية التقليدية للقياس، ومن الملاحظ أن هناك اهتماما متزايدا باستخدام الاختبارات الموضوعية مثل الاختيار من متعدد، ولا سيما مع تزايد أعداد الطلاب، حيث أن هذا النوع من الاختبارات يمكن المعلم من سرعة التصحيح يدويا وآليا، مع تحقيق الدقة والموضوعية في حساب الدرجة، الأمر الذي يعد أحد متطلبات تحقيق جودة العملية التعليمية.

وقد تزامن هذا الاهتمام بالاختبارات الموضوعية مع نشاط حركة البحث العلمي في مجال القياس التربوي والنفسي وتأسيس مراكز متخصصة للقياس على المستوى القومي وعلى مستوى الجامعات تهتم بتطوير الاختبارات وفقا لأحدث أساليب القياس النفسي والتربوي . ويعد الهدف الأساس من هذه البحوث العلمية والتطبيقات العملية هو التوصل لأفضل الظروف والإجراءات التي تحقق دقة وموضوعية القياس.

وتتحدد دقة القياس النفسي والتربوي بعدد من العوامل يتمثل أولها في دقة بناء أدوات القياس. ودقة أداة القياس تعنى دقة كل مفردة من مفرداتها، وكذا دقة المقياس ككل، بما يؤدى إلى دقة تقدير السمة موضوع القياس، ومن ثم تحقيق أهداف القياس المتنوعة.

وقد كانت نظرية الاستجابة للمفردة هي ثمرة جهود العلماء من أجل تحقيق دقة القياس وموضوعيته، وذلك عندما قصرت النظرية التقليدية عن تحقيقها. ولذا فقد حظيت باهتمام الباحثين سواء في الدراسات النظرية أو التطبيقية. وقد تنوعت أهداف هذه الدراسات، فمنها ما اهتم باستخدام النظرية في تدريج المقاييس وتقنينها، ومنها ما أضاف إلى ذلك سحب نماذج اختباريه متعادلة القياس وتوظيفها، وثالثة اهتمت بمقارنتها بالنظرية التقليدية، ورابعة اهتمت بتطوير إجراءات التدريج من خلال دراسة العوامل المؤثرة في دقة القياس، أو من خلال طرح أساليب جديدة لتيسير تلك الإجراءات. ويمكن اعتبار الدراسة الحالية تنتمي للصنف الرابع من هذه

الدراسات.

وتتعدد نماذج نظرية الاستجابة للمفردة ومن أوسع تلك النماذج انتشارا واستخداما: نموذج راش. وقد استخدمت تلك النماذج في تدريج المقاييس سواء تلك التي تتضمن مفردات ثنائية الاستجابة مثل مفردات الاختيار من متعدد، أو المقاييس المتدرجة، أو مقاييس إنتاج الاستجابة.

وغنى عن البيان أن أساليب القياس – أيا كانت دقتها – لا يمكنها إنتاج أداة قياس جيدة إلا إذا أحسن إعداد بنود المقياس في مراحله الأولى وتم ضبط موقف القياس، وذلك قبل استخدام أسلوب القياس المناسب وتحليل البيانات وانتقاء البنود الجيدة في ضوء المحكات الإحصائية المرتبطة بهذا الأسلوب.

ولذا اهتم علماء القياس بتطوير أساليب وطرق تساعد الباحثين على إجراء المزيد من الفحص لمفردات المقياس بهدف الكشف عن خصائصه قبل تدريجه وفقا لأحد نماذج نظرية الاستجابة للمفردة، كما تساعد على تيسير فهم مخرجات التحليل الإحصائي، وفي هذا الإطار قدم باتنبرج ولاروس Batenburg & Laros (١٩٩٧) أسلوبا مبسطا يساعد الباحثين على فحص مفردات الاختيار من متعدد أطلقا عليه اسم التحليل البياني للمفردات Graphical Item Analysis (GIA) (Batenburg & Laros, 2002).

وقد استخدم التحليل البياني للمفردات GIA لأول مرة في مقاييس التحصيل لمشروع SAEB في البرازيل من قبل كل من باتنبرج ولاروس، ثم مشروع SSPE-SE في تركيا (Yurdugul & Batenburg, 2006 , P1). كما استخدم التحليل البياني للمفردات في مصر في المركز القومي للامتحانات والتقويم التربوي في عدد من المشروعات القومية والدولية حيث قام باتنبرج بتدريب فريق الباحثين المصريين على استخدامه. واستخدم أيضا في العديد من الدراسات العلمية.

وتهدف الدراسة الحالية إلى اختبار الاتساق بين محكات التحليل البياني والمحكات الإحصائية للملاءمة لنموذج راش في تقدير جودة المفردة الاختبارية، بالإضافة إلى دراسة أثر استخدام التحليل البياني والتحليل الإحصائي معا في انتقاء المفردات التي تحقق دقة القياس على خصائص كل من التدريج النهائي للاختبار والتحليل البياني النهائي لمفرداته.

مشكلة الدراسة:

في محاولة لتحقيق المزيد من دقة القياس النفسي والتربوي باستخدام نماذج نظرية الاستجابة للمفردة قدم باتنبرج ولاروس (١٩٩٧) أسلوب التحليل البياني للمفردات الاختبارية من نوع

الاختيار من متعدد. ويمكن استخدام التحليل البياني للمفردات في فحص مفتاح التصحيح والاطمئنان إلى دقة إدخال البيانات وكذا اختبار بنية وصعوبة وتمييز المفردة قبل تدريج الاختبار تبعا لنماذج الاستجابة للمفردة، مع مراعاة إعادة التحليل البياني بعد حذف المفردات غير الملائمة بيانيا(Batenburg & Laras,2002).

إلا أن أغلب الدراسات التطبيقية التي استفادت من التحليل البياني (مثل: Bahakta, Tennat, Horton, Lawton & Andrich, 2005 ؛ أمينة كاظم، شادية عبد العزيز، سوسن عبد الرحمن، ٢٠٠٦ ؛ منى ربيع، ٢٠٠٨؛ وليد مسعود، ٢٠١٠) قد ركزت على استخدامه في التحقق من مفتاح التصحيح ودقة إدخال البيانات، وفيما يتعلق بدقة بنية المفردة فقد استخدم التحليل البياني كمؤشر يساعد على دعم اتخاذ القرار بشأن المفردات التي يرفضها النموذج وتفسير نتائج التحليل الإحصائي بعد ذلك، مع الإشارة إلى ما يعكسه ذلك من اتساق بين نتائج كل من التحليل البياني والإحصائي.

والقليل من هذه الدراسات (مثل: أحلام الباز، رضا حجازي، ٢٠٠٢ ؛ مصطفى قاسم، ٢٠١٥) أضافت للاستخدامات السابقة حذف المفردات غير الملائمة وفقا لمحكات التحليل البياني وعدم إدخالها مع باقي المفردات في التحليل الإحصائي باستخدام أحد نماذج الاستجابة للمفردة.

فاستبقاء هذه المفردات قد يؤثر على إحصاءات ملاءمة باقي المفردات مما قد يؤدى إلى اتخاذ قرارات غير صائبة باستبعاد أو استبقاء بعض المفردات الأخرى.

كما أن هذه البحوث التطبيقية، التي استخدمت التحليل البياني للمفردات قد أجرته مرة واحدة قبل التحليل الإحصائي، وعادة ما ترفق نتائج التحليل البياني مع النتائج النهائية ضمن توصيف مفردات المقياس الكلي، وفي ضوء هذا التوصيف يتم سحب الاختبارات الفرعية المناسبة منه. وبالأخذ في الاعتبار أن عمليات التحليل الإحصائي بهدف تدريج مقياس أو بنك للأسئلة وفقا لأحد نماذج الاستجابة للمفردة تسفر عادة عن حذف بعض الحالات من الأفراد والمفردات التي لا تحقق شروط دقة القياس، فإنه من المحتمل أن تتغير معالم التحليل البياني لسائر مفردات المقياس بعد حذف تلك الحالات غير الملائمة.

ومن ثم يمكن تحديد مشكلة الدراسة الحالية في الأسئلة التالية:

١- هل تتسق محكات التحليل البياني مع المحكات الإحصائية للملاءمة لنموذج راش في تقييم جودة المفردة الاختبارية ؟

٢- هل يؤثر حذف المفردات غير الملائمة وفقا لمحكات التحليل البياني على إحصاءات المفردات

(نسبة المفردات المحذوفة – إحصاءات الملاءمة – تدرج المفردات – دقة التدرج – دالة معلومات الاختبار) في التدريج النهائي للاختبار التحصيلي تبعا لنموذج راش؟

٣- هل يؤثر حذف المفردات غير الملائمة وفقا لمحكات التحليل البياني على إحصاءات الأفراد (نسبة الأفراد المحذوفين – إحصاءات الملاءمة – تدرج الأفراد – دقة التدرج) في التدريج النهائي للاختبار التحصيلي تبعا لنموذج راش؟

٤- هل تتحسن نتائج التحليل البياني للمفردات بعد حذف الحالات (الأفراد والمفردات) غير الملائمة لأسس القياس الموضوعي وفقا لكل من المحكات البيانية والإحصائية؟

أهداف الدراسة:

- التعرف على درجة الاتساق بين محكات التحليل البياني والمحكات الإحصائية للملاءمة لنموذج راش في تقدير جودة المفردة الاختبارية.
- استكشاف أثر حذف المفردات غير الملائمة وفقا لمحكات التحليل البياني على بعض الخصائص الإحصائية للمفردات في التدريج النهائي للاختبار التحصيلي تبعا لنموذج راش.
- استكشاف أثر حذف المفردات غير الملائمة وفقا لمحكات التحليل البياني على بعض الخصائص الإحصائية للأفراد في التدريج النهائي للاختبار التحصيلي تبعا لنموذج راش.
- التعرف على ما إذا كانت نتائج التحليل البياني للمفردات تتحسن بعد حذف الأفراد والمفردات غير الملائمة لأسس القياس الموضوعي وفقا لكل من المحكات البيانية والإحصائية.

أهمية الدراسة:

تتمثل أهمية الدراسة الحالية في:

- ان دراسة فاعلية الاعتماد على كل من محكات التحليل البياني والمحكات الإحصائية للملاءمة في انتقاء المفردات الملائمة للقياس الموضوعي الدقيق، وكذا دراسة اثر حذف الأفراد والمفردات غير الملائمة على نتائج التحليل البياني للمفردات ومن ثم على استكمال توصيف مفردات المقياس، قد:
  - يلقى الضوء على بعض العوامل والإجراءات التي تحقق دقة القياس.
  - يثير بعض التساؤلات التي تحفز على إجراء بحوث علمية أخرى في هذا المجال.
  - يفيد في تطوير الإجراءات التي يتبعها الباحثون في تدريج الاختبارات وبنوك الأسئلة، في كل من: البحوث العلمية الأكاديمية، ومراكز القياس والتقويم النفسي والمهني والتربوي.
- قد توضح الدراسة اختلاف ملامح التحليل البياني للمفردات بعد عمليات التحليل الإحصائي وتدريج المقاييس وبنوك الأسئلة وفقا لنموذج راش، وقد يفيد ذلك الباحثين والمهتمين في

- التكامل بين التحليل الإحصائي والبياني في انتقاء وتدرج مفردات الاختيارمن متعدد مراكز القياس والتقويم في توصيفهم لمفردات المقاييس في صورتها النهائية بما يحقق دقة القياس والتقويم ومن ثم دقة اتخاذ القرار، حيث يتم سحب النماذج الاختبارية من المقياس الكلي وفقا لهذه المواصفات.
- قد تثير الدراسة اهتمام الأساتذة والقائمين على عملية التقويم التربوي بالجامعات باستخدام طريقة التحليل البياني للمفردات بهدف فحص مفردات الاختيار من متعدد، بما يساعد على إعادة صياغة بعض تلك المفردات على أساس علمي مبسط قبل استخدامها في امتحانات أخرى.

التعريف الإجرائي لمصطلحات الدراسة:

التحليل الإحصائي للمفردات: هو استخدام برنامج الحاسب الآلي WINSTEPS لإجراء سلسلة من المعالجات الإحصائية لاستجابات أفراد العينة على مفردات الاختبار التحصيلي المستخدم في الدراسة، بهدف تقييم مدى ملاءمتها لأسس القياس الموضوعي كما تتمثل في نموذج راش باستخدام المؤشرات الإحصائية للملاءمة، واستبعاد الأفراد غير الملائمين ثم إعادة تحليلها لتحديد وحذف المفردات غير الملائمة، ثم إعادة تحليلها مرة ثالثة بهدف إنشاء التدريج النهائي للاختبار. وقد تم الاعتماد على إحصاءي الملاءمة التقاربي والتباعدي لمؤشر ZSTD لانتقاء المفردات الملائمة لأسس القياس الموضوعي، حيث اعتبر تجاوز المدى (± ٢) محكا لعدم ملاءمة الأفراد، واعتبر تجاوز المدى (±٢,٥) محكا لعدم ملاءمة المفردات.

التحليل البياني للمفردات: هو إنشاء تمثيل بياني لكل مفردة من مفردات الاختبار التحصيلي المستخدم في الدراسة، يوضح العلاقة بين نسبة اختيار كل بديل من بدائل الإجابة على المفردة (البديل الصواب والمشتتات) والدرجة الكلية على الاختبار، ويتم تقييم جودة المفردة في ضوء المحكات التي قدمها (Batenburg & Laras,2002, P330) ومن ثم انتقاء المفردات الملائمة لأسس القياس، حيث تعد المفردة غير ملائمة في حالة:

١- تعدد النقاط التي لا يتحقق فيها شرط زيادة نسبة اختيار البديل الصواب مع زيادة الدرجة الكلية على الاختبار.

٢- تعدد النقاط التي لا يتحقق فيها شرط تناقص نسبة اختيار المشتت مع زيادة الدرجة الكلية على الاختبار(ويتضمن ذلك ضعف أو انعدام جاذبية المشتت).

٣- تعدد التقاطعات بين خط البديل الصواب مع خطوط المشتتات بعد نقطة بداية فترة التمييز.

٤- انخفاض ميل خط البديل الصواب.

تدريج الاختبار: يقصد به إنشاء ميزان لتدرج القدرة على التحصيل الدراسي لمادة نظريات التعلم

مقدرة بوحدة قياس معرفة، يتحدد عليه موقع كل مفردة من حيث صعوبتها، وذلك وفقا لنموذج راش الاحتمالي أحادى البعد . ويتم ذلك بعد استبعاد الأفراد والمفردات التي لا تحقق شروط دقة القياس من مصفوفة البيانات المستخدمة في إنشاء التدريج.

وتم استخدام وحدة المنف لتقدير كل من قدرة الفرد وصعوبة المفردة، وهي وحدة قياس محولة من وحدة اللوجيت باستخدام معادلة التحويل الخطي: منف = ٥٠ + ٥ × لوجيت.

مفردات الاختيار من متعدد: يقصد بها مفردات الجزء الموضوعي في الاختبار التحصيلي النهائي لمقرر نظريات التعلم، لكل مفردة أربعة بدائل للإجابة، واحدة منها فقط هي البديل الصواب للمفردة.

نظرية الاستجابة للمفردة: هي اتجاه معاصر في القياس النفسي والتربوي، تقوم على عدة شروط تساعد على تحقيق موضوعية القياس وعدالته وتكامل أهداف القياس. ويندرج تحت هذه النظرية عدد من النماذج الرياضية الاحتمالية التي تحدد العلاقة بين الاستجابات الملاحظة للأفراد على المفردة والقدرة أو الخاصية غير الملاحظة التي تحدد هذه الاستجابات. وتهتم الدراسة الحالية بأحد هذه النماذج وهو نموذج راش اللوغاريتمي أحادي المعلم.

ويتم توضيح المفاهيم النظرية للمصطلحات السابقة فيما يلي:

الإطار النظري والدراسات السابقة

أولا: الاختبار التحصيلي من نوع الاختيار من متعدد:

يعرف **الاختبار** بأنه إجراء منظم لقياس سمة ما من خلال عينة من السلوك. ويعرف **الاختبار التحصيلي** بأنه إجراء منظم لتحديد مقدار ما تعلمه الطالب (سامي ملحم، ٢٠١٥، ص ص٤٧ ، ٢٠٦)، كما يعرفه (Livingston & Dorans, 2004, P4) بأنه أداة قياس تصف كميا درجة التعلم تحت ظروف مقننة. ويلاحظ من ذلك أن اصطلاح "الاختبار" في أدبيات القياس التربوي يستخدم ليدل على أسلوب قياس كما يستخدم ليدل على أداة قياس كما في الدراسة الحالية.

وتتعدد أنواع الاختبارات التحصيلية ومن بينها الاختبارات الموضوعية التي تحقق موضوعية تقدير الدرجة على الاختبار بمعنى عدم اختلافها باختلاف القائم بالتصحيح، ومن أنواع هذه الاختبارات اختبارات الاختيار من متعدد، وهذه بدورها لها أنواعها المتعددة.

وتعد مفردات الاختيار من متعدد من أكثر أنواع المفردات الموضوعية استخداما في الاختبارات التحصيلية المقننة(صلاح علام، ٢٠١٥، ص٣٣٦). حيث تصلح لقياس العديد من

مخرجات التعلم في المستويات المعرفية المختلفة، كما أنها توفر الوقت والجهد في حساب الدرجة وتقلل فرصة التخمين، إلا أنها لا تصلح لقياس القدرة على تنظيم الأفكار والتعبير الكتابي والتأليف كما تتطلب وقتا وجهدا في إعدادها(ماهر صبري ومحب الرافعي، ٢٠٠٣، ص ص ٢٦٨-٢٦٩).

وتتكون كل مفردة من مفردات الاختبار التقليدي من نوع الاختيار من متعدد من متن أو جذر السؤال، ويتمثل في عبارة أو جملة، يليه عدد من بدائل الإجابة، يمثل أحدها الإجابة الصواب والبدائل الأخرى تسمى مشتتات. ولتحقيق جودة مفردات الاختيار من متعدد ينبغي مراعاة الشروط العامة والشروط الخاصة بالمتن والشروط الخاصة بالبديل الصواب والمشتتات. وقد تناولت أدبيات القياس والتقويم التربوي هذه الشروط **بالشرح والتفصيل**، مثل: (جابر عبد الحميد، ١٩٨٣، ص ص ٣٦٨-٣٧٨ ؛ Haladyna, 2004, pp116-118 ؛ رجاء أبو علام، ٢٠٠٥؛ سامي ملحم، ٢٠١٥، ص ص٢١٩-٢٢٤ ؛ أمينة كاظم، شادية عبد العزيز، رباب سيف، ٢٠١٦).

ويمكن للدراسة **تلخيص** أهم الشروط العامة لتحقيق جودة المفردة في: أن تقيس المفردة أحد الأهداف التعليمية للمقرر، وألا تعتمد إجابة إحدى المفردات على مفردة أخرى، وأن تكون المفردات ككل شاملة للمحتوى والمستويات المعرفية ومستويات الصعوبة المختلفة وذلك وفقا لجدول مواصفات الورقة الامتحانية، مع مراعاة دقة الصياغة ووضوح تعليمات الإجابة وتدريب المتعلمين عليها.

كما يمكن تلخيص أهم الشروط الخاصة بمتن المفردة في: أن تتضمن مشكلة محددة بدقة، وأن تخلو من الهاديات اللغوية والنحوية التي توحي بالإجابة الصواب على غير علم، وأن تتضمن الكلمات المكررة في البدائل.

أما الشروط المتعلقة بالبدائل فيمكن تلخيص أهمها في: أن يكون من بينها بديل واحد يمثل الإجابة الصواب أو الأكثر دقة، وأن تكون متجانسة في المحتوى والطول والبناء اللغوي، وأن تكون للمشتتات قوة جذب مناسبة بحيث تبدو جميع البدائل مقبولة منطقيا من قبل المفحوصين الذين لديهم المعرفة أو المهارة المراد قياسها، وأن يكون ترتيب البديل الصواب عشوائيا.

ويتطلب ذلك أن يكون معد الاختبار على وعي ودراية بكل من المحتوى وخصائص المفحوصين كي يتمكن من بناء المشتتات من خلال الأخطاء الشائعة للمتعلمين مثل الأخطاء في المفاهيم والتفسيرات الخطأ أو الناتجة عن الأخطاء الحسابية، بالإضافة إلى توفر القدرة على الابتكار وبذل الوقت والجهد (كروكر والجينا، ٢٠٠٩، ص١١١).

وتعد مرحلة وضع مفردات الاختبار من أهم مراحل بناء الاختبار الجيد، حيث أن المفردة الجيدة البناء والمناسبة للمتعلم، تحقق ملاءمة جيدة لأسس القياس الموضوعي الدقيق كما تتمثل في محكات الملاءمة لنماذج نظرية الاستجابة للمفردة.

ثانيا: نظرية الاستجابة للمفردة:

يعكس تاريخ القياس النفسي والتربوي الجهود التي بذلها علماء القياس بهدف تطوير أدوات دقيقة وموضوعية للقياس السلوكي. وقد أسفرت هذه الجهود عن مداخل عديدة للقياس والتقويم التربوي من أشهرها أساليب القياس جماعي المرجع ومحكي المرجع، التي تصنف عادة تحت ما أطلق عليه النظرية التقليدية في القياس، وأساليب القياس الموضوعي الذي تتمثل في نظرية الاستجابة للمفردة Item Response Theory (IRT) (أمينة كاظم، ٢٠٠٠ ، ص ٣٢٣).

وتتدرج الأساليب السابقة تحت مظلة المنحى السيكومتري في التقويم التربوي. وهناك المنحي السياقي الذي يركز علي التقويم في سياق شبيه بسياق عملية التعليم والتعلم، والمنحي التفريدي الذي يهتم بإعداد مواقف للتعلم والتقويم تناسب كل متعلم. وتتكامل هذه المداخل مع المنحى السيكومتري في تحقيق أهداف التقويم التربوي(صلاح علام، ٢٠٠٧، ص٣٢). وتهتم الدراسة الحالية بنظرية الاستجابة للمفردة لما تحققه من موضوعية القياس.

فقد طورت نظرية الاستجابة للمفردة لمحاولة التغلب على مشكلات أساليب القياس التقليدية، وقد تناولت مراجع عديدة هذه المشكلات بالشرح الوافي، مثل: (أمينة كاظم، ١٩٩٦، ص ص ٢٩٥-٣٠٦ ؛ Reckase ,2009, P 26 ؛ Kirishnan, 2013, PP9-10 ؛ Toland, 2014, P124؛ صلاح علام، ٢٠١٥، ص ص ٦٨٦-٦٨٨ ؛ Tay, Meade & Cao, 2015, PP4-6).

ويمكن للدراسة **تلخيص** ابرز هذه المشكلات في: الاعتماد علي الدرجة الخام الكلية كأساس للقياس، وعدم توفر وحدة قياس معرفة لكل من الأفراد والمفردات، وافتراض تساوي أخطاء القياس لجميع أفراد العينة، واعتمادية القياس على خصائص كل من الأفراد والمفردات، مما يعيق تحقيق بعض أهداف القياس مثل المقارنة بين الأداء السابق واللاحق للفرد،.والمقارنة بين الأفراد في حالة اختلاف مفردات الاختبار، والمقارنة بين تقديرات المفردات في حالة اختلاف عينة الطلاب التي أدتها، وبالتالي يصعب تحقيق موضوعية القياس والتقويم وعدالته.

ويعتمد القياس الموضوعي في نظامه المرجعي على تقدير علاقة احتمالية بين الأداء الملاحظ

للفرد في الاختبار أو المقياس والسمات أو القدرات التي تكمن وراء هذا الأداء وتفسره (ماهر صبري ومحب الرافعي ، ٢٠٠٣، ص ٥٩)، ولذا أطلق على النظرية في بدايتها اسم نظرية السمات الكامنة. ومن مسميات هذه النظرية أيضا: نظرية المنحنى المميز للمفردة (Suen, 1990, p85) ويعتبر اسم نظرية الاستجابة للمفردة هو الاسم الأكثر شيوعا واستخداما حاليا. وقد وضع لورد (١٩٦٨) أسسها النظرية، بينما وضع راش (١٩٦٠) أول نماذجها الرياضية. ثم توالت جهود العلماء في تطوير هذه النماذج وتطبيقها في القياس النفسي والتربوي حتى اليوم (صلاح الدين علام، ٢٠١٥، ص ص ٦٨٨-٦٨٩).

وتستخدم النماذج اللوغاريتمية في تحديد العلاقة بين أداء الفرد في الاختبار وبين السمة أو القدرة التي تكمن وراءه، وذلك في صورة دالة رياضية احتمالية. وقد تناولت كتابات عديدة تطور هذه النماذج وأنواعها ، مثل:(Embretson & Reise, 2000, PP65-122 ؛ Suen, 1990, PP88-93 ؛ Reckase, 2009, pp11-111). ويتم اختيار النموذج المناسب وفقا لهدف و طبيعة الاختبار، وإمكانية حساب التقديرات الخاصة بالفرد و المفردة، ومدى ملاءمة البيانات للنموذج (كروكر والجينا، ٢٠٠٩ ؛ ص ٤٧٨). ويعد نموذج راش الاحتمالي اللوغاريتمي أحادي المعلم من ابسط هذه النماذج وأكثرها استخداما.

نموذج راش:

وفقا لنموذج راش – في حالة الاختبارات المعرفية ثنائية الاستجابة- فإن احتمال أن يستجيب الفرد (صوابا أو خطأ) لمفردة ما هو دالة للفرق بين قدرة هذا الفرد و صعوبة المفردة. فكلما ارتفعت قدرة الفرد عن صعوبة المفردة ازداد احتمال حدوث الاستجابة الصحيحة للمفردة والعكس صحيح، وإذا تساوت قدرة الفرد مع صعوبة المفردة يتساوى احتمال حدوث الاستجابة الصحيحة مع احتمال حدوث الاستجابة الخطأ (أمينة كاظم،١٩٩٦أ، ص ص ٣١٣).

ويقوم نموذج راش على افتراضات إذا تحققت تتحقق مطالب القياس الموضوعي، وتتمثل في: **أحادية البعد**؛ وتعنى أن مفردات المقياس تكون متجانسة، أي أن كلا منها يتطلب في حله نفس التوليفة من الإجراءات و العمليات ولكنها تختلف فيما بينها من حيث درجة صعوبتها فقط (أمينة كاظم ، ١٩٩٤ ، ص ١١٤). ويتطلب افتراض أحادية البعد تحقق افتراضات **انعدام التخمين والتحرر من عامل السرعة** في الإجابة حتى لا تكون هناك عوامل أخرى تحدد الاستجابة سوى السمة المقاسة (Elliott, 1983, p61 ؛ Hambleton & Swaminthan, 1985, p30).

ويتكافأ افتراض أحادية البعد مع افتراض **الاستقلال الموضعي** للمفردة بمعني ألا تؤثر استجابة الفرد

لإجدى مفردات الاختبار على استجابته للمفردات الأخرى. وبذلك تكون استجابات الأفراد لمفردات الاختبار مستقلة إحصائياً، وهذا يعني وجود قدرة واحدة فقط هي التي تكمن وراء العلاقة بين مجموعة مفردات الاختبار (Hambleton, Swaminthan & Rogers , 1991, pp 1-12 ؛ Reckase, 2009, P 12).

ويرجح تحقق شرط أحادية البعد في حالة تدريج اختبار للتحصيل في مجال دراسي محدد حيث يكون الأداء المتعلق بالقدرات اللازمة للنجاح مرتبطا ارتباطا مرتفعا (صلاح علام ، ٢٠٠٥، ص ٢٨١). وغالباً ما يتم التحقق من هذا الشرط باستخدام التحليل العاملي.

ومن افتراضات النموذج أيضا تساوى قوى التمييز بحيث تكون جميعها متوسطة التمييز (Reckase , 2009, p21 ؛ Embretson & Reise, 2000, pp40-48)، وخطية القياس حيث يتدرج كل من الأفراد و المفردات على نفس المتصل تدرجا خطيا بوحدة قياس واحدة (أمينة كاظم ١٩٩٦أ، ص ٣٠٦).

الموضوعية في نموذج راش:

يتيح نموذج راش تدريج كل من قدرة الفرد و صعوبة المفردة على تدريج خطى ، بحيث يكون تقدير كل منها مستقلا عن مجموعة المفردات المستخدمة، ومستقلا عن عينة الأفراد التي تؤديها. ومن ثم فالموضوعية فى نموذج راش تعنى :

- موضوعية المقارنة بين نتيجة تفاعل قدرتي فردين مع مفردة مناسبة.
- موضوعية المقارنة بين صعوبتي مفردتين استجاب لهما فرد مناسب(أمينة كاظم، ١٩٩٦أ، ص ص ٣٢١-٣٢٢ ؛ Isaac & Micael , 1982 , p113).

وقد أثبتت دراسات عديدة تحقيق النموذج لاستقلالية القياس، ومن تلك الدراسات دراسة أمينة كاظم (١٩٩٦ب) وشادية عبد العزيز (١٩٩٧) ومحمد الشافعي (١٩٩٦) واعتدال غازي (٢٠٠٠) ومنى ربيع (٢٠٠٠)، وأمينة كاظم وآخرون (٢٠٠٦)، وغيرها. حيث تكافأت إحصائياً التقديرات المتناظرة لقدرات الأفراد، وكذا صعوبة المفردات بالرغم من اختلاف مستوى صعوبة مجموعة المفردات التي يتضمنها الاختبار.

إنشاء ميزان تدرج الأفراد والمفردات:

وفقاً لنموذج راش يمكن رد التقديرات الخاصة بكل من الفرد والمفردة إلى صفر تدريج واحد. وتقدر كل من قدرة الفرد وصعوبة المفردة بوحدة قياس واحدة هي وحدة اللوجيت Logit، وتعرفها أمينة كاظم (١٩٩٤) بأنها " قدرة الفرد على النجاح على المفردات التي تعبر نقطة صفر التدريج

عن صعوبتها، عندما يكون احتمال النجاح (٠,٧٣) (أمينة كاظم ، ١٩٩٤ ، ص ص١٢٦). وتحقق وحدة اللوجيت مميزات التدريج ذي الوحدات المتساوية، ولكنها تتضمن قياسات سالبة و صفرية، ولذا تحول خطيا إلى تدريجات جديدة تلائم أهداف القياس(Wright & Masters , 1982 , PP 192) مثل تدريج المنف المستخدم في الدراسة الحالية.

ويمكن حساب قدرة الفرد وصعوبة المفردة يدويا وباستخدام الحاسوب بحيث يتم التقدير وإعادة التقدير إلى أن تستقر التقديرات عبر مستويات القدرة والصعوبة. وتتوفر العديد من برامج الحاسب الآلي اللازمة لذلك، ومنها برنامج WINSTEPS من تطوير لياكر (٢٠٠٦). ومن أهم مميزات هذا البرنامج أنه يتيح تصحيح البيانات المدخلة، ويوفر إحصاءات للتحقق من أحادية البعد، ويوفر إحصاءات لتقييم ملاءمة كل من المفردات والأفراد، ويحسب تقديرا لثبات بارامترات الأفراد والمفردات، مع حساب الخطأ المعياري لقدرة كل فرد وصعوبة كل مفردة.

انتقاء الأفراد و المفردات وفقا للمحكات الإحصائية للملاءمة:

تتحقق افتراضات نموذج راش عندما تتطابق البيانات الملاحظة مع توقعات النموذج أي عند ملاءمة البيانات للنموذج. ونقص ملاءمة البيانات للنموذج قد يرجع إلى سوء ملاءمة المفردات، أو سوء ملاءمة الأفراد، أو كليهما. وتتيح برامج الحاسب الآلي الإحصاءات التي تساعد على استبعاد الحالات غير الملائمة لنموذج راش من الأفراد و المفردات. و يوفر برنامج WINSTEPS نوعين من هذه الإحصاءات، هما: هو إحصاء الملاءمة التقاربي (Infit)، وإحصاء الملاءمة التباعدي (Outfit).

ويختص إحصاء الملاءمة التقاربي للمفردة بأداء الأفراد الذين تقترب قدراتهم من مستوى صعوبة هذه المفردة. وبالمثل يختص إحصاء الملاءمة التقاربي للفرد بأداء هذا الفرد على مفردات تقترب صعوبتها من مستوى قدرته. **ويوفر برنامج Winsteps مخرجين لكل من إحصاءي الملاءمة التقاربي والتباعدي، هما:**

- **Mean-Square (MNSQ)** : وهو عبارة عن إحصاء (كا^2) مصححا من أثر حجم العينة، وقيمته المتوقعة تكون قريبة من الواحد الصحيح. وتعتبر MNSQ مؤشراً لحجم العشوائية في الاستجابات أو لحجم التهديد الموجه لصدق وسلامة القياس.
- **Z-standardized (ZSTD)** : ويوضح هذا الإحصاء الدلالة الإحصائية لإحصاء (كا^2)، والقيمة المتوقعة لهذا الإحصاء هي "صفر". ويشير**ZSTD** إلى ما إذا كانت البيانات تلائم النموذج أم لا (Linacre , 2006 , PP210-214).

ويذكر ليناكر (٢٠٠٦) أن إحصاء ZSTD غير حساس في حالة نقص حجم العينة عن (٣٠) مفحوص حيث تبدو كل البيانات ملائمة، بينما يكون مفرط الحساسية في حال زيادة حجم العينة عن (٣٠٠) مفحوص حيث تبدو كل البيانات غير ملائمة (في: أسماء عرفان، ٢٠١٠، ص ١٢٤). وقد تم الاعتماد على هذا الإحصاء لاختبار الملاءمة في الدراسة الحالية حيث أن حجم العينة يزيد عن (٣٠) ويقل عن (٣٠٠). كما تم الاعتماد على كل من إحصاءي ZSTD و MNSQ لمقارنة درجة الملاءمة في التدريج النهائي في حالتي التدريج.

وترجع سوء ملاءمة الفرد إلى أن الصعوبة النسبية للمفردات لدى هذا الفرد تختلف عن صعوبتها النسبية عند معظم الناس. أما المفردة فتكون سيئة الملاءمة في حالتين،هما: إذا كانت صعوبتها غير مستقرة بالنسبة لباقي المفردات عبر المستويات المختلفة للقدرة، وإذا كانت المفردة تقيس صفة مختلفة عما تقيسه باقي مفردات الاختبار(أمينة كاظم، ١٩٩٦أ، ص ص٣٣٦ – ٣٣٧).

ووفقا لإحصاء **ZSTD** يصنف الأفراد والمفردات على أنها سيئة الملاءمة (Misfit) عندما ترتفع القيم الموجبة لإحصاءي الملاءمة التقاربي و التباعدي لها. كما تصنف على أنها قد تجاوزت حدود الملاءمة (Overfit) عندما تنخفض القيم السالبة لإحصاءي الملاءمة لها (Wright & Stone , 1979 , p127). ومن ثم لا تحقق هذه الأفراد والمفردات فروض النموذج ،وينبغي حذفها (أمينة كاظم، ١٩٩٤ ، ص ١٢٨).

ولا توجد قيمة مطلقة على تدريج الملاءمة يمكن اعتبارها حدا فاصلا بين الملاءمة وعدم الملاءمة، بل يمكن تحديد هذه القيمة وفقا لموقف القياس. ويمكن اتخاذ القيمة (± ٢) كحدود للملاءمة (Wright & Linacre , 1985 , p4-13) وهى الحدود المستخدمة لملاءمة الأفراد في الدراسة الحالية. وقد استخدمت القيمة (±٢,٥) كحدود لملاءمة المفردات في العديد من الدراسات (مثل: أمينة كاظم ، ١٩٩٤ ، ص١٩٢). وهى الحدود المستخدمة لملاءمة المفردات في الدراسة الحالية.

دالة معلومات المفردة ودالة معلومات الاختبار (في نموذج راش):

تختص دالة معلومات المفردة Item Information Function (IIF) بتحديد كمية المعلومات التي تقدمها المفردة عن قدرة الأفراد (أو السمة المقاسة لديهم) عند المستويات المختلفة من القدرة (أو السمة)، وتقاس بأقصى ارتفاع للمنحني الممثل لدالة المعلومات لتلك المفردة (Hambleton, et al., 1991, p79). وتكون دالة المعلومات أكبر ما يمكن عند مواءمة المفردة مع مستوى الفرد (كروكر

والجينا، ٢٠٠٩، ص ٤٨٦ ؛ Reckase M.D. ,2009 , P 49). **ومن استخدامات دوال معلومات المفردات:** تحديد المفردات المتماثلة في دوال معلوماتها ومن ثم مراعاة ذلك في تكوين صور متكافئة من الاختبار. و تحديد أي المفردات الاختبارية تقيس القدرة المراد قياسها بدرجة أفضل عند مستويات محددة لهذه القدرة. كما يمكن اتخاذها كأساس لانتقاء المفردات الجيدة، مع الأخذ في الاعتبار محتوى المفردة (Toland , 2014 , p144).

وانطلاقا من افتراض استقلالية الموضع فإن المعلومات التي تقدمها المفردة عن القدرة أو السمة موضوع القياس تكون مستقلة عن المعلومات التي تقدمها باقي مفردات الاختبار عن هذه السمة، ولذا فإن **دالة معلومات الاختبار** Test Information Function (TIF) تساوى مجموع دوال معلومات المفردات في هذا الاختبار. ومن ثم تزيد دالة معلومات الاختبار كلما زاد عدد مفرداته (Reckase M.D. ,2009, PP49,58 ؛ كروكر والجينا، ٢٠٠٩، ص ٤٨٧ ؛ Toland , 2014 , pp145).

وحيث أن دالة معلومات الاختبار تقدم معلومات مفيدة عن المعلومات التي يوفرها الاختبار ككل عن القدرة كدالة لقيم القدرة، فإن **من أوجه الاستفادة من دالة معلومات الاختبار:** أنها توظف في تحديد الفجوات على متصل المتغير موضوع القياس. كما تستخدم كمقياس للخطأ المعياري المتوقع لتقدير القدرة (**SEE**)، أو مقدار عدم الثقة في تقدير القدرة عند نقطة ما على متصل القدرة ويساوي مقلوب الجذر التربيعي لمعلومات الاختبار(**SEE=1/SQR(I(θ)**) (Toland , 2014 , p145).

الخطأ المعياري للقياس Standard Error Of Measurement:

يصف خطأ القياس المدى الذي تتباين فيه تقديرات القدرة المتوقعة حول القيم الحقيقية لها، ويتم التعبير عنه بنفس الوحدات التي تقاس بها معالم النموذج. وتباين تلك التقديرات لاختبار ما عند مستوى محدد للقدرة هو مقلوب دالة معلومات هذا الاختبار عند هذا المستوى للقدرة (**1/(I(θ))**). ولذا فإن هذا التباين دالة للتغير في القدرة بحيث يتفرد كل تقدير للقدرة بقيمة لتباين التقديرات المتوقعة. والجذر التربيعي لتباين التقديرات المتوقعة يطلق عليه الخطأ المعياري(Suen , 1990, P98 ؛ Toland , 2014, pp143-144).

ويقترن انخفاض قيمة الخطأ المعياري ومن ثم زيادة دقة القياس مع زيادة عدد مفردات الاختبار وارتفاع قوى تمييز مفرداته وانخفاض فرصة تخمين الإجابة الصواب، بالإضافة إلى مواءمة صعوبة الاختبار مع قدرات المفحوصين (**Hambleton et al., 1991,P93** ؛

رينولدز وليفنجستون، ٢٠١٣ ، ص ص ٢٠١-٢٠٢).

مما سبق يتبين أن تحقيق شروط القياس الموضوعي باستخدام نموذج راش يتطلب:

- إعداد و صياغة مفردات الاختبار بحيث يتوفر فيها كل من الدقة وفقا لقواعد وشروط ومعايير جودة المفردة، و التجانس (أي أن جميع المفردات تتطلب نفس التوليفة من العمليات لحلها). وقد أكد (Damsgard, San, Fors, Anke & Roe, 2007) على أن التحليل الإحصائي وفقا لنموذج راش يعد الخطوة الثانية لتنقية الاختبار وانتقاء المفردات الجيدة وتكوين اختبار جيد، وذلك بعد الخطوة الأولى المتمثلة في كتابة مفردات جيدة.
- مواءمة صعوبة الاختبار لمستوى قدرات الأفراد، حيث ينبغي أن يضم مفردات متنوعة الصعوبة بما يتسق مع مستويات القدرة بحيث يقدم المعلومات المطلوبة عن قدرات الممتحنين.
- استخدام الأساليب المناسبة لاستبعاد الأفراد و المفردات غير الصادقة في تدرجها على متصل المتغير موضوع القياس.
- ضبط زمن الاختبار، وخصائص الموقف الاختباري بما يساعد على تحقيق صدق الاستجابة.

تطبيقات نظرية الاستجابة للمفردة في التربية وعلم النفس:

نظرا لتغلب نماذج الاستجابة للمفردة على كثير من مشكلات القياس السلوكي، فقد انتشرت تطبيقاتها في القياس النفسي والتربوي. وقد ساعد على ذلك انتشار برامج الحاسوب المناسبة، بالإضافة إلى تداول المفاهيم المتعلقة بها والألفة بها في وسط المهتمين بالقياس النفسي و التربوي.

فقد استخدمت نماذج نظرية الاستجابة للمفردة في المؤسسات التربوية والنفسية الدولية والمحلية لتدريج المقاييس وبنوك الأسئلة، مثل: اختبار الكفاءة في اللغة الانجليزية TOEFL، واختبارات التحصيل المستخدمة في الدراسات الدولية مثل TIMSS، ومقاييس القبول بالجامعات الأمريكية مثل SAT، والمقاييس الخاصة بالعديد من مشروعات المركز القومي للامتحانات مثل بطارية المقاييس القومية للذكاء، وبنوك الأسئلة في مشروع مركز بنوك الأسئلة للتقويم الموضوعي بكلية البنات جامعة عين شمس (أمينة كاظم وآخرون، ٢٠٠٦؛ سيد خير الله وآخرون، ٢٠١٥).

كما استخدم الباحثون هذه النماذج في دراسات متنوعة لتدريج أدوات قياس المتغيرات المعرفية مثل التحصيل الدراسي (أحلام الباز، رضا حجازي، ٢٠٠٢ ؛ إكرام حمزة ، ٢٠٠٤؛ نضال الشريفين، ٢٠٠٩ ؛ Bhakta, el. ,2005 ؛ ديانا حماد، ٢٠١٠؛ هالة مندور، محمد سيد أحمد، نسرين عاشور، ٢٠١٣)، الفهم القرائي باللغة الانجليزية (وليد مسعود، ٢١٠)، الاستيعاب والسرعة في القراءة (باسمة الشايجي، ٢٠٠٧)، والقدرات العقلية(اعتدال غازي ، ٢٠٠٣ ؛ علي

ذكري، ٢٠١١)، والتصور البصري المكاني للمجسمات (شادية عبد العزيز،٢٠٠٢)، والاستدلال الاستقرائي (منى ربيع، ٢٠٠٨)، الاستدلال اللغوي(مصطفى قاسم، ٢٠١٥)، مرونة الغلق (جمعة عبد الحميد، ٢٠١٣)، والذكاء (منى ربيع، ٢٠٠٠ ؛ وليد مسعود ، ٢٠٠٤)، وغيرها.

وقد امتدت تطبيقات النظرية لتشمل تدريج مقاييس الجوانب غير المعرفية، مثل: الاتجاهات (ابتسام أبو خليفة، ٢٠٠٩)، ومداخل الدراسة (حصة عبد الرحمن، أنور رياض، محمد إبراهيم، ٢٠٠٩)، والاكتئاب (سليمان صالح، ٢٠١٣)، ودافعية الإنجاز (أسماء عرفان، ٢٠١٠)، والقلق (رانيا ماهر، ٢٠٠٥)، وجودة الحياة الصحية (Hayas, Quinta, Padierna, Bilbao & Munoz, 2010)، وأهداف الدراسة (Muis, Winne & Edwards, 2009)، ورهاب الحركة لمرضى العمود الفقري (Damsgard et. Al., 2007)، والذكاء الوجداني (Carrasco, MacCAN, Geiger, Seybert & Roberts, 2015)، وغيرها.

وقد تباينت هذه الدراسات من حيث أهدافها وطول الاختبار وحجم عينة التدريج ودرجة ملاءمة البيانات لأسس القياس الموضوعي كما تتمثل في عدد الأفراد المفردات المحذوفة ... والنموذج المستخدم في التدريج وإن كان أغلبها قد استخدم نموذج راش.

ثالثا: التحليل البياني لمفردات الاختبار Graphical Item Analysis (GIA):

قدم فان باتنبرج وجاكوب لاروس (١٩٩٧) أسلوب التحليل البياني للمفردات الاختيارية من نوع الاختيار من متعدد، بهدف تحقيق المزيد من دقة القياس النفسي والتربوي باستخدام نماذج نظرية الاستجابة للمفردة. ويتم إجراء التحليل البياني للمفردات باستخدام ملف أوامر SYNTAX يتم إعداده باستخدام البرنامج الإحصائي SPSS.

ويوفر التحليل البياني للمفردة (GIA) تمثيلا بيانيا للعلاقة بين الدرجة الكلية على الاختبار ونسبة الاستجابة لكل بديل من البدائل المختلفة للمفردة. ويمكن من خلال هذا التمثيل البياني استخلاص معلومات أساسية سهلة التفسير عن خصائص المفردة، مثل: الصعوبة والتمييز ومعدل التخمين ووصف المشتتات أو البدائل الخطأ. ومن ثم يعد (GIA) طريقة بصرية مبسطة لتقييم جودة المفردة. ومن خلال هذا العرض البصري يمكن تحقيق فهم أفضل لنظرية الاستجابة للمفردة، وخاصة بالنسبة لمن ليس لديه خبرة سابقة في مجال القياس النفسي والتربوي (Yurdugul &Batenburg, 2006, P209 ; Batenburg & Laros, 2002, P319).

ولم تكن طريقة التحليل البياني للمفردات التي قدمها باتنبرج ولاروس هي المحاولة الأولى

أو الأخيرة لاستخدام الرسوم البيانية بهدف تيسير حساب وفهم خصائص المفردات والاختبارات، فقد استخدم لودلو (١٩٨٦) التحليل البياني للبواقي المعيارية للكشف عن ملاءمة البيانات لنموذج راش(Ludlow , 1986 , P217). كما قام فريق الإحصاء في مركز ETS بتطوير مدخل للتحليل البياني لمفردات الاختبار بما يساعد على القراءة البصرية السريعة للخصائص الكلاسيكية لمفردات الاختبار مثل الصعوبة وقوة التمييز وجاذبية المشتتات(Livingston, Dorans, 2004 , P1).

وفي هذا الإطار أيضا شرح (Izard, 2005, PP35-38) طريقة بيانية لفحص المفردة وتشخيصها من خلال نمط الاستجابة للبديل الصواب عند ثلاث مستويات للقدرة.

كما اهتم بعض الباحثين بتطوير برامج حاسوب لإنشاء التمثيل البياني للمفردات، ومن أمثلة هذه البرامج برنامج Vista - CITA الذي يساعد على تقديم عرض بصري سهل التناول والتفسير للخصائص الكلاسيكية لكل من المفردة والاختبار. مثل الاتساق الداخلي وخطأ القياس وأحادية البعد وصعوبة المفردة وقوة تمييزها (Ledesma, Molina, 2009, PP1161-1162).

الأساس العلمي للتحليل البياني للمفردات:

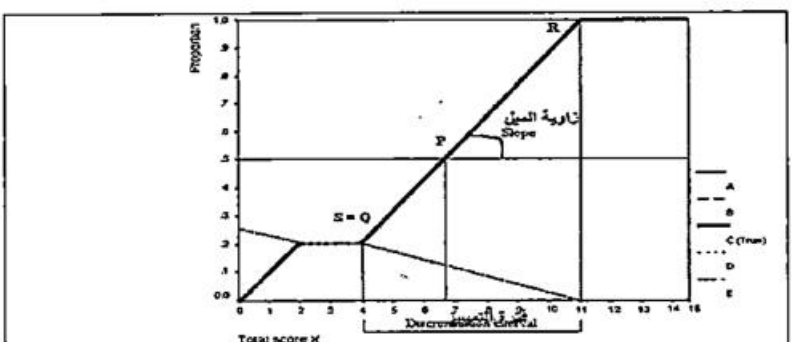
عند تقدير خصائص مفردات الاختيار من متعدد – سواء باستخدام الأساليب التقليدية أو باستخدام نماذج نظرية الاستجابة للمفردة- لا يتم استخدام المعلومات الخاصة بالاستجابات على البدائل الخطأ للمفردة، حيث تكود جميعها بالدرجة (صفر)، مما يعد فقدا للمعلومات، ويساعد التحليل البياني للمفردات في التغلب على هذه المشكلة من خلال دراسة كل من المشتتات والبديل الصواب. وقد أشار (Haladyna, 2004, P 272) إلى أن دراسة المشتتات قد حازت القليل من الاهتمام في الماضي رغم أهميتها.

ومن المعلوم أن ارتفاع قوة تمييز المفردة يعد من مؤشرات جودة المفردة من نوع الاختيار من متعدد. ويقاس تمييز المفردة عادة عن طريق معامل الارتباط بين الدرجة على هذه المفردة والدرجة الكلية على الاختبار. ويجب أن يكون معامل الارتباط موجبا، مشيرا إلى أنه كلما ارتفعت الدرجة الكلية على الاختبار(التحصيلي) دل ذلك على ارتفاع مستوى التحصيل الدراسي ومن ثم يزيد احتمال الإجابة الصواب على المفردة.

ويمكن حساب معاملات ارتباط مشابهة لكل مشتت (لكل بديل من البدائل الأخرى للإجابة على المفردة)، وهذه يجب أن تكون سالبة ، مما يعنى أنه كلما ارتفعت الدرجة الكلية على الاختبار كلما قل معدل اختيار هذا المشتت(Izard, 2005, P56).

ولذا ، لكي تكون المفردة جيدة وفقا لمؤشر التحليل البياني فإن نسبة الطلاب الـذين يختـارون الإجابة الصواب يجب أن يزيد بزيادة الدرجة الكلية، كما أن نسبة الطلاب الـذين يختـارون كـل مشتت يجب أن يتناقص مع زيادة الدرجة الكلية. بالإضافة إلى ضرورة أن تتوزع اختيارات العينة على البدائل المختلفة فلا يترك أي بديل دون أن يختاره أحد أفراد العينة.

ويمكن قراءة ذلك بسهولة ويسر من خلال الخطوط التي تمثل العلاقة بين الدرجة الكلية على المحور الأفقي، ونسبة اختيار كل بديل من بدائل الإجابة على المفردة على المحور الرأسي. ويمكن أيضا من خلال فحص هذه الخطوط البيانية فحص الكثير من الخصائص المفردة (Yurdugul & Batenburg, 2006, P209) . ويتضح ذلك من الشكل التالي:



شكل (١)

نموذج نظري لتحليل البياني لمفردة افتراضية من نوع الاختيار من متعدد

قراءة وتفسير التحليل البياني للمفردات:

في شكل (١): يمثل **الخط السميك الأسود البديل الصواب للإجابة** من بين البدائل الخمسة على المفردة وهو البديل (C) في المثال المذكور. أما الخط المتقطع فيمثل المشتتات حيـث تطابقـت الخطوط التي تمثل المشتتات الأربعة عند جميع مستويات الدرجة الكلية (مـن صـفر حتـى ١٥) .(Batenburg & Laros,2002, P321)

وتلخص الباحثة المعلومات المستخلصة من هذا التمثيل البياني للمفردة كما يلي:

- **فترة التمييز للمفردة:** وهى الفترة التي يزيد فيها نسبة الأفراد الـذين اختـاروا الإجابـة الصواب عن نسبة الأفراد الذين اختاروا أي بديل آخر للإجابة ، وكذا تكـون العلاقـة طردية بين نسبة الأفراد الذين أجابوا صوابا والدرجة الكلية على الاختيار. وهي الفتـرة المحصورة بين النقطة S والنقطة R أي بين الدرجات الكلية (٤ : ١١) في الشكل السابق. وترتبط فترة التمييز بكل من صعوبة المفردة وقوة تمييزها.

- **قوة التمييز للمفردة:** وتقدر بميل الخط الأسود السميك – الذي يمثل العلاقة بين الدرجة الكلية واحتمال الإجابة الصواب على المفردة – وذلك عند النقطة (P) التي تقابل نسبة احتمال الإجابة الصواب (٥٠ %)على المحور الرأسي. فإذا كانت زاوية الميل تساوى صفر درجة ، تكون نسبة احتمال الإجابة الصواب ثابتة ولا ترتبط بالدرجة الكلية على الاختبار ، أما إذا كانت زاوية الميل تساوى (٩٠) درجة فان الخط الذي يمثل الاستجابة الصواب يشبه نموذج جتمان حيث تكون قوة التمييز في حدها الأقصى وفترة التمييز أقل ما يمكن (تتقلص في نقطة تمثل درجة قطع محددة).

وفي الشكل رقم (١) يلاحظ أن قوة تمييز هذه المفردة الافتراضية يساوي ميل الزاوية (٦٠) درجة تقريبا، كما أن المفردة لا تميز بين الأفراد عند مستوى الدرجات المنخفضة جدا (٢- ٤) والمرتفعة جدا (١١- ١٥)، حيث كان خط البديل الصواب أفقيا.

- **صعوبة المفردة:** وتقدر بالدرجة الكلية (على المحور الأفقي) المقابلة للنقطة (P). وهذه الدرجة تساوي (٦,٨) تقريبا في الشكل السابق.
- **معدل تخمين الإجابة الصواب على المفردة:** ويقدر بنسبة الذين اختاروا الإجابة الصواب من ذوى الحد الأدنى للدرجة، أو نسبة الذين اختاروا البديل الصواب للإجابة قبل بداية فترة التمييز. ويفترض أن تكون هذه النسبة ضئيلة.
- **فاعلية المشتتات:** يمكن من خلال التمثيل البياني المقارنة بين جاذبية البدائل المختلفة عند مختلف مستويات القدرة، وتحديد ما إذا كان هناك مشتت مستبعد لم يتم اختياره من قبل أحد أفراد العينة، مع ملاحظة ميل كل خط من الخطوط التي تمثل العلاقة بين الدرجة الكلية ونسبة اختيار المشتت، وعدد نقاط تقاطع خط البديل الصواب مع المشتتات بعد بداية فترة التمييز.

وفي الشكل السابق رقم (١) تتساوى المشتتات الأربعة في احتمال اختيارها من قبل أفراد العينة في جميع مستويات الدرجة الكلية. كما أن العلاقة عكسية بين الدرجة الكلية واحتمال اختيار كل مشتت فيما عدا الفترة (٢-٤) حيث تتساوى نسب الذين اختاروا البديل الصواب (٢٠ %) مع نسب من اختاروا كل بديل من المشتتات الأربعة الأخرى وكانت جميع الخطوط أفقية، ويلاحظ أيضا أنه لا يوجد أي مشتت مستبعد، كما أن المشتتات لا تتقاطع مع خط البديل الصواب بعد بداية فترة التمييز.

- **تقييم جودة المفردة بوجه عام:** من خلال ما سبق يمكن تقييم جودة المفردة. وتعتبر المفردة الافتراضية الممثلة بالشكل رقم (١) مفردة جيدة بوجه عام.

ويرى (Bhakta, el. ,2005, P10) أن قرار حذف المفردة أو الاحتفاظ بها بناء على التحليل البصري للاستجابة للمفردة يتحدد بأهمية الهدف الذي تقيسه المفردة.

وجدير بالذكر أن التحليل البياني للمفردات الواقعية ليس متسقا تماما كما في المثال النظري السابق، حيث أن الخطوط التي تمثل العلاقة بين نسب اختيار كل بديل من بدائل الإجابة والدرجة الكلية على الاختبار تكون اقل انسيابية. ولذا يتم إجراء تحسين Smoothing للرسوم البيانية بحيث تصبح الخطوط أكثر انسيابية وبالتالي يسهل تفسيرها كما يتضح من نتائج الدراسة الحالية.

انتقاء المفردات وفقا لنتائج التحليل البياني:

يتم حذف المفردات غير الملائمة بيانيا في ضوء المحكات الأربعة التي قدمها باتنبرج ولاروس والتي سبق عرضها في التعريف الإجرائي للمتغيرات. ووفقا لهذه المحكات، إذا ظهر بوضوح انخفاض جودة المفردة يتم استبعادها، ثم يعاد التحليل البياني بحيث لا تدخل المفردات السيئة في حساب الدرجة الكلية على الاختبار ومن ثم فقد تتحسن خصائص المفردات التي كان هناك شك في انخفاض جودتها. وبذلك تستبقى المفردات الصالحة للقياس.

ويعتقد باتنبرج ولاروس أن دقة تقدير تحصيل التلاميذ سوف تتحسن بعد حذف المفردات منخفضة الجودة. كما يريان ضرورة دراسة نتائج التحليل البياني في علاقته بمحتوى المفردات والمحتوى الدراسي حتى يكون من الممكن تعديل المفردة منخفضة الجودة، وإذا تم تعديل محتوى المفردة يكون من الضروري إجراء تحليل بياني آخر (Batenburg & Laros,2002, PP330-331).

أهمية التحليل البياني للمفردات واستخداماته:

يستخدم التمثيل البياني لتحقيق المزيد من الفهم للمخرجات الرقمية الصماء التي يوفرها التحليل الإحصائي تبعا لنماذج الاستجابة للمفردة، حيث يمكن استخدام التمثيل البياني لتحقيق أهداف عديدة، منها: الكشف عن الخصائص الهامة فى البيانات، ومقارنة أنماط الاستجابة المتوقعة بالملاحظة، والتحقق من افتراضات النموذج(Cheu & Chen, 2006, P1). وتعطي طريقة GIA معلومات عن جودة القياس، تجعل من الممكن تقرير ما إذا كانت المفردة تمثل مشكلة في الاختبار ويجب حذفها. **وبالتالي يمكن توظيفها بهدف:** تحديد المفردات التي ينبغي حذفها بناء على نتائج التحليل الإحصائي فيما بعد، وتحسين جودة الاختبارات وبنوك الأسئلة (Batenburg & Laros,2002, P319).

كما يمكن استخدامها في كشف تحيز المفردة، وفي الكشف عن الأنماط غير المتوقعة من الاستجابات على المفردات، مثل اختيار مشتت ما (نفس المشتت) بنسب أعلى من اختيار البديل الصواب عبر بعض أو مختلف مستويات القدرة(Bhakta, el. ,2005, PP 6-9).

ويلاحظ أن هذه الأنماط يصعب اكتشافها بدون استخدم إحدى طرق التحليل البياني للمفردات مما يبين جدوى استخدام هذه الطرق لفحص المفردات وانتقائها، وهذا ما تهتم به الدراسة الحالية.

وقد استخدمت **دراسات عديدة** هذه الطريقة في التحليل البياني لمفردات الاختيار من متعدد، **حيث:**

- اهتمت بعض هذه الدراسات بتوضيح كيفية استخدام طريقة التحليل البياني في استكشاف البيانات وحساب بعض الخصائص الإحصائية للمفردة، ومن هذه الدراسات: (Batenburg & Laros,2002) التي شرحت كيفية استخدام التحليل البياني للمفردات وتفسيره مع التطبيق على مثال نظري وآخر واقعي، مع تقديم المحكات التي بناء عليها يتم استبعاد المفردات. ومنها أيضا دراسة (Yurdugul & Batenburg, 2006)التي وضحت إجراء التحليل البياني وكيفية حساب صعوبة المفردة باستخدام التحليل البياني للمفردات.
- اهتمت بعض الدراسات باستخدام التحليل البياني للمفردات بهدف دعم اتخاذ القرار بحذف المفردات غير الملائمة بناء على المحكات الإحصائية للملاءمة، ومن هذه الدراسات (هالة مندور وآخرون، ٢٠١٣).
- اهتمت أغلب الدراسات باستخدام التحليل البياني للمفردات بهدف تفسير وتشخيص المفردات المرفوضة بناء على المحكات الإحصائية للملاءمة، ودعم اتخاذ القرار بحذف المفردات غير الملائمة إحصائيا، ومن هذه الدراسات: (إكرام حمزة، ٢٠٠٢ ؛ Bhakta, el. ,2005 ؛ أمينة كاظم وآخرون، ٢٠٠٦ ؛ باسمة الشايجي، ٢٠٠٧ ؛ منى ربيع، ٢٠٠٨ ؛ وليد مسعود، ٢٠١٠) حيث أشارت هذه الدراسات إلى أن أغلب المفردات التي رفضها النموذج بناء على المحكات الإحصائية تبين وجود مشكلات في بدائل الإجابة الخاصة بها من خلال التحليل البياني. وفي دراسة (باسمة الشايجي، ٢٠٠٧) أضافت الباحثة أن العديد من المفردات الملائمة للنموذج كانت جيدة في التحليل البياني.
- اهتمت دراسات قليلة باستخدام التحليل البياني للمفردات بهدف تشخيص عيوب المفردات وتنقيتها قبل التحليل الإحصائي والتدريج تبعا لنظرية الاستجابة للمفردة، ومن هذه الدراسات (أحلام الباز ورضا حجازي، ٢٠٠٢ ؛ مصطفى قاسم، ٢٠١٥) حيث قام الباحثون بحذف المفردات التي تبين عدم ملاءمتها بناء على محكات التحليل البياني الأولي، ولم تدخل البيانات

الخاصة بها في إنشاء التدريج، وبالتالي لم يهتم الباحثون بفحص إحصاءات ملاءمة هذه المفردات للنموذج.

وفي دراسة (أمينة كاظم وآخرون، ٢٠٠٦) استخدمت نتائج التحليل البياني الأولي للمفردات الملائمة لاستكمال وصف المفردات في التدريج النهائي لبنك الأسئلة. وفي دراسة (مصطفى قاسم، ٢٠١٥) استخدم الباحث نتائج التحليل البياني الأولي في تفسير المفردات المرفوضة تبعا لإحصاءات الملاءمة للنموذج وذلك بدون إعادة التحليل البياني. في حين من المحتمل أن تختلف معالم التحليل البياني بعد استبعاد المفردات غير الملائمة لأسس القياس سواء تم تحديدها تبعا للمحكات الإحصائية أو البيانية.

فروض الدراسة:

١- تتسق محكات التحليل البياني مع المحكات الإحصائية للملاءمة لنموذج راش في تقييم جودة المفردة الاختبارية.

٢- لا يؤثر حذف المفردات غير الملائمة وفقا لمحكات التحليل البياني على إحصاءات المفردات (نسبة المفردات المحذوفة - إحصاءات الملاءمة - تدرج المفردات - دقة التدرج - دالة المعلومات) في **التدريج النهائي** للاختبار التحصيلي تبعا لنموذج راش.

٣- لا يؤثر حذف المفردات غير الملائمة وفقا لمحكات التحليل البياني على إحصاءات الأفراد (نسبة الأفراد المحذوفين - إحصاءات الملاءمة - تدرج الأفراد - دقة التدرج) في **التدريج النهائي** للاختبار التحصيلي تبعا لنموذج راش.

٤- تتحسن نتائج التحليل البياني للمفردات بعد حذف الأفراد والمفردات غير الملائمة لأسس القياس الموضوعي وفقا لكل من المحكات البيانية والإحصائية.

منهج الدراسة وإجراءاتها

منهج الدراسة:

استخدم المنهج الوصفي الارتباطي والمقارن لاختبار فروض الدراسة السيكومترية الحالية، حيث اتبعت الخطوات التالية:

١- إعداد الاختبار التحصيلى النهائي لمقرر نظريات التعلم، وتطبيقه في الموعد الزمني المقرر للامتحان النهائي للمادة.

٢- إدخال البيانات الخاصة بأسئلة الاختيار من متعدد آليا باستخدام جهاز الماسح الضوئي، حيث يتم تدقيق ورقة الإجابة الالكترونية، ثم إدخال البيانات، ثم التحقق من دقة الإدخال.

وقد تمت الخطوتين السابقتين في إطار خطة عمل مشروع مركز بنوك الأسئلة للتقويم الموضوعي بكلية البنات جامعة عين شمس(أحد مشروعات تطوير كليات التربية). وقد شاركت الباحثة في ذلك باعتبارها عضو بالمركز.

٣- إجراء التحليل البياني الأولي لمفردات إحدى نماذج الاختبار، وحذف المفردات التي لا تحقق شروط دقة القياس وفقا لمحكات التحليل البياني.

٤- التحليل الإحصائي للمفردات لتدريجها وفقا لنموذج راش، وذلك في حالتين: بدون حذف المفردات غير الملائمة وفقا لمحكات التحليل البياني، وبعد بحذف هذه المفردات.

٥- دراسة الاتساق بين نتائج التحليل البياني والإحصائي لاختبار الفرض الأول.

٦- المقارنة بين نتائج التحليل الإحصائي في الحالتين لاختبار الفرضين الثاني والثالث.

٧- إعادة التحليل البياني لمفردات الاختبار في صورته النهائية -بعد حذف الحالات غير الملائمة وفقا لكل من المحكات البيانية والإحصائية- ومقارنته بنتائج التحليل البياني الأولي، وذلك لاختبار الفرض الرابع.

٨- استخلاص وعرض النتائج وفقا لفروض الدراسة، وتفسيرها ومناقشتها.

عينة الدراسة:

تتمثل في مجموعة طالبات الفرق الرابعة التربوية بكلية البنات للآداب والعلوم والتربية في الفصل الدراسي الثاني بالعام الجامعي ٢٠٠٦-٢٠٠٧ اللاتي استجبن على النموذج الاختباري المستخدم في الدراسة، وبلغ عددها (١٩٩) طالبة، منها (٣٦) طالبة بالشعبة العلمية التربوية، و(١٦٣) طالبة بالشعبة الأدبية التربوية.

أداة الدراسة:

تتمثل أداة القياس في الجزء الخاص بأسئلة الاختيار من متعدد وذلك في أحد نماذج الاختبار التحصيلي النهائي في مقرر نظريات التعلم. وقد شاركت الباحثة في إعداده ضمن فريق أعضاء هيئة التدريس المعني بتدريس المقرر، وذلك وفقا لخطة عمل مركز بنوك الأسئلة للتقويم الموضوعي، حيث قام المركز بعقد عدة ورش عمل لتدريب المشاركين على تحليل محتوى المقرر الدراسي، وصياغة الأهداف الإجرائية وتحديد المستوى المعرفي لها، وحساب نسبة الأهداف في كل مستوى معرفي لكل موضوع من موضوعات المقرر ومن ثم إعداد جدول مواصفات الورقة الامتحانية، وصياغة المفردات الاختبارية الموضوعية والمقالية المطورة،وتكوين النماذج الاختبارية، وصياغة تعليمات الاختبار.

ووفقا لما سبق، تم إعداد الاختبار التحصيلي الحالي كما يلي:

١. إعداد جدول مبسط لمواصفات الورقة الامتحانية في ضوء محتوى وأهداف المقرر ومستويات الأهداف المعرفية التي اقتصرت على ثلاثة مستويات هي : التذكر والفهم والتطبيق كما يوضح الجدول التالي:

جدول (١)

جدول مواصفات الورقة الامتحانية

الموضوع		المستوى المعرفي			المجموع
		التذكر	الفهم	التطبيق	
مفاهيم أساسية		% ٨,٣	% ٧,٤	% ٣,١	% ١٨,٩
نماذج للنظريات السلوكية	المحاولة والخطأ (ثورندايك)	% ٣,٤	% ٢,٢	% ٢,٢	% ٧,٩
	الترابط (جثري)	% ٥,٨	% ٤,٥	% ٣,٤	% ١٣,٧
	التعزيز (هل)	% ٩,٠	% ٥,٦	% ٣,٨	% ١٨,٤
نماذج للنظريات المعرفية	الجشتالت (فرتهيمر ..)	% ٦,٥	% ٤,٥	% ١,٦	% ١٢,٦
	المجال (ليفين)	% ٦,٥	% ٤,٥	% ١,٦	% ١٢,٦
	البنائية النمائية (بياجيه)	% ٨,١	% ٥,٦	% ٢,٢	% ١٦,٠
المجموع		% ٤٧,٦	% ٣٤,٤	% ١٨,٠	% ١٠٠

٢. كتابة مفردات الاختبار بحيث تغطي محتوى المقرر ومستويات الأهداف المعرفية الثلاثة، وذلك مع مراعاة معايير جودة المفردة.

٣. تكوين عدة نماذج اختبارية بما يتناسب مع الزمن الرسمي المحدد للاختبار، وكتابة تعليمات الإجابة، وإعداد مفاتيح التصحيح.

٤. مراجعة وتحكيم مفردات وتعليمات الاختبار من قبل فريق تدريس المقرر وأحد الأساتذة المتخصصين بقسم علم النفس بالكلية.

وفي ضوء ذلك تم إعداد ثلاثة نماذج للاختبار يتكون كل نموذج من جزأين، الأول منها يضم (٣٠) مفردة من نوع الاختيار من متعدد، ويضم الجزء الثاني أنواعا أخري من الأسئلة. وقد اقتصرت أداة الدراسة الحالية على **الجزء الأول** الذي يضم أسئلة الاختيار من متعدد لأحد هذه النماذج.

التحقق من صدق وثبات الاختبار:

تم التحقق من **صدق القياس** بالطرق التالية:

صدق المحتوى: حيث تم إعداد الاختبار ومراجعته وتحكيمه على النحو المبين فيما سبق. وقد روعي في إعداد الاختبار أن يعرف الاختبار متغيرا واحدا، وألا تعتمد الإجابة عن المفردة على إجابة المفردات الأخرى. مما يدعم صدق هذا الاختبار.

صدق التدرج: إن تدرج مفردات الاختبار الحالي معا على ميزان تدرج واحد مشترك وفقا لنموذج راش – بعد حذف الحالات غير الملائمة من الأفراد و المفردات – يعنى تحقق شروط النموذج، ومنها شرط أحادية البعد. ويعنى ذلك تحقق صدق تدرج المفردات في قياس المتغير موضوع القياس، وكذا صدق تدرج قدرات الأفراد على متصل المتغير والذي يقوم على صدق استجاباتهم على المفردات (Loyd , 1988 , P142 ؛ Henning , 1984 ,P129).

الصدق العاملي: بالإضافة لما سبق، يوفر برنامج Winsteps إحصاء للتحقق من أحادية البعد كبديل للتحليل العاملي، حيث تتضمن مخرجات البرنامج نسبة التباين المفسر باستخدام مفردات الاختبار، وقد أسفرت هذه النتائج عن تأكيد تحقيق الاختبار لشرط أحادية البعد حيث تجاوزت نسبة التباين المفسر الحد الفاصل (٥٠ %)، كما قلت نسبة العامل الثاني عن (٥ %).

وبتحقق كل من صدق المحتوى وصدق التدرج وأحادية القياس يتحقق صدق مفردات الاختبار فى تعريف متغير واحد هو متغير التحصيل الدراسي لمقرر نظريات التعلم.

كما اعتمد حساب ثبات الاختبار على كل من:

معامل الثبات الحدي أو الهامشي: يستخدم هذا المعامل في تحديد ثبات الاختبارات المدرجة باستخدام نظرية الاستجابة للمفردة، ويعتمد في حسابه على الخطأ المعياري لتقدير بارامترات الصعوبة أو القدرة (١- مربع متوسط الخطأ المعيارى). ويتم تفسيره بنفس الطريقة التي يفسر بها معامل الثبات التقليدي. (Toland , 2014 , P145). وقد بلغت قيمة معامل الثبات الهامشي لتقديرات الصعوبة (٠,٩٦) ، كما بلغت قيمة معامل ثبات تقديرات القدرة (٠,٧٦) وهي قيم تدل على أن الاختبار يحقق درجة جيدة من الثبات.

قيم الخطأ المعياري لتقديرات كل من صعوبات المفردات وقدرات الأفراد: تراوحت قيم الخطأ المعياري لتقديرات الصعوبة بين (١) و (٢) منف ، كما تراوحت قيم الخطأ المعياري لتقديرات القدرة المقابلة لكل درجة كلية محتملة على الاختبار بين (٢) و (٥) منف – بعد استبعاد الدرجة التامة والصفرية – وجميعها قيم منخفضة نسبيا مما يدل على استقرار تقديرات الصعوبة والقدرة ومن ثم الاطمئنان إلى ثبات القياس.

معامل الثبات الذي يوفره برنامج Winsteps، وهذا المعامل مكافئ لمعامل كيودر ريتشاردسون- ٢٠ (Linacre , 2006 ,P193). وقد بلغت قيمتي معامل ثبات تقديرات كل من الصعوبة والقدرة (٠,٩٨) و(٠,٥٥) وهي قيم تدل على أن الاختبار يحقق درجة مرتفعة من الثبات لتقديرات صعوبة المفردات، إلا أن هذا المؤشر يعكس درجة منخفضة نسبيا من الثبات

لتقديرات القدرة.

ثبات التدرج: إن تدرج مفردات الاختبار الحالي معاً على ميزان تدرج واحد مشترك وفقا لنموذج راش – بعد حذف الحالات غير الملائمة من الأفراد و المفردات – يعنى تحقق شروط النموذج، ومنها شرط استقلالية القياس. ويعنى ذلك ثبات تقديرات كل من الصعوبة والقدرة وعدم تأثرها باختلاف أداة القياس، أو باختلاف عينة الأفراد (أمينة كاظم ،٢٠٠٠ ، ص٣٣١).

أساليب المعالجة الإحصائية والبيانية والكيفية:

- المقاييس الإحصائية الوصفية، ومعامل الثبات الهامشي، وحجم التأثير.
- التحليل الإحصائي تبعا لنموذج راش باستخدام برنامج الحاسب الآلي Winsteps لحساب إحصاءات كل من الفرد والمفردة (تقدير الصعوبة أو القدرة ، والخطأ المعياري لها، وإحصاءات الملاءمة، والحصول على دالة معلومات الاختبار) ، وكذا لحساب معامل ثبات الاختبار ومؤشر أحادية البعد .
- التحليل البياني للمفردات باستخدام ملف الأوامرSyntax الذي تم إعداده باستخدام برنامج SPSS.
- التحليل والوصف الكيفي لنتائج التحليل البياني في ضوء محكات التحليل البياني، ومعامل فاي لاختبار الفرض الأول.
- ولاختبار الفرضين الثاني والثالث تم حساب :
  - دلالة الفرق بين النسب المئوية لعدد المفردات المحذوفة، وكذا النسب المئوية لعدد الأفراد المحذوفين، وذلك في حالتي التدريج.
  - معاملات الارتباط بين تقديرات الصعوبة، وكذا بين تقديرات القدرة، في حالتي التدريج.
  - اختبار (ت) للعينتين المرتبطتين والمستقلتين، واختبار ولكوكسن واختبار مان ويتني، لاختبار دلالة الفروق بين قيم إحصاءات الملاءمة، وكذا قيم الخطأ المعياري.
- التحليل والوصف الكيفي لنتائج التحليل البياني، واختبار مكنمار لاختبار الفرض الرابع.

نتائج الدراسة ومناقشتها:

أولا : نتائج اختبار الفرض الأول ومناقشتها:

ينص الفرض الأول على: تتسق محكات التحليل البياني مع المحكات الإحصائية للملاءمة لنموذج راش في تقييم جودة المفردة الاختبارية. ولاختبار هذا الفرض تم إجراء التحليل البياني لمفردات اختبار نظريات التعلم بناء على استجابات عينة الدراسة (١٩٩ فردا) على نموذج الاختبار (٣٠ مفردة). ثم تحديد المفردات الملائمة وغير الملائمة وفقا لمحكات التحليل البياني.

ولتحري الدقة في اتخاذ القرار المناسب بشأن المفردات غير الملائمة بيانيا، تم حذف أسوأ تلك المفردات، ثم أعيد التحليل البياني، ثم اعتبرت المفردات التي لم يطرأ عليها تحسن ملحوظ، مفردات غير ملائمة وفقا لمؤشر التحليل البياني. وقد أسفر هذا التحليل عن حذف ثلاث مفردات غير ملائمة وفقا لمحكات التحليل البياني، وهي المفردات (٨ ، ١٥ ، ١٦).

كما استخدم برنامج **WINSTEPS** لإجراء التحليل الإحصائي للبيانات الكلية ((٣٠) مفردة، و(١٩٩) فردا) وتحديد المفردات الملائمة وغير الملائمة لنموذج راش وفقا لإحصاءي الملاءمة التقاربي والتباعدي للمؤشر (ZSTD)، حيث اعتبر تجاوز المدى (±٢,٥) محكا لعدم الملاءمة. وقد أسفر التحليل الإحصائي عن حذف مفردة واحدة غير ملائمة إحصائيا، وهي المفردة رقم (٨).

وبذلك تتفق نتائج التحليل البياني والإحصائي على تقرير ملاءمة (٢٧) مفردة وعدم ملاءمة مفردة واحدة ، في حين صنفت مفردتان غير ملائمة وفقا لمحكات التحليل البياني فقط ولم تصنف أي مفردة غير ملائمة وفقا للمحكات الإحصائية فقط. وباستخدام معامل فاي لاختبار دلالة الاتساق بين المؤشرين كانت النتائج كما بالجدول التالي:

جدول (٢)

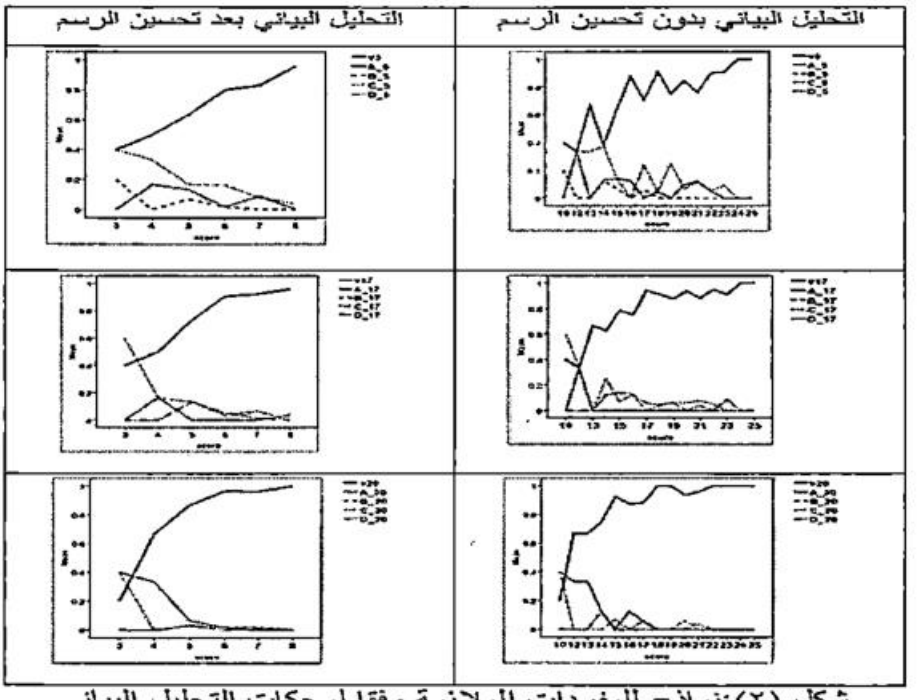
حجم ودلالة الاتساق بين محكات الملاءمة الإحصائية والبيانية في تقييم ملاءمة مفردات الاختبار

بيانيا / إحصائيا	غير ملائمة	ملائمة	المجموع	معامل ارتباط فاي	مستوى الدلالة
ملائمة	٢	٢٧	٢٩	٠,٥٦	٠,٠٠٣
غير ملائمة	١	-	١		
المجموع	٣	٢٧	٣٠		

من الجدول السابق يتضح اتساق نتائج التحليل البياني والتحليل الإحصائي في تقييم جودة المفردة الاختبارية حيث أن معامل الارتباط متوسط ودال إحصائيا عند مستوى دلالة أقل من (٠,٠١). وقد تأكدت هذه النتائج باستخدام معامل الاتفاق (كابا) حيث كان الاتساق متوسطا (٠,٤٧٤**) ودالا إحصائيا. وبذلك تتحقق صحة الفرض الأول من فروض الدراسة.

ولتوضيح محكات حذف المفردات وفقا لنتائج التحليل البياني تعرض الدراسة فيما يلي نماذج للتحليل البياني للمفردات.

١- نماذج لنتائج التحليل البياني للمفردات الملائمة بيانيا:



شكل (٢):نماذج للمفردات الملائمة وفقا لمحكات التحليل البياني

يوضح شكل (٢) نماذج لنتائج التحليل البياني للمفردات الملائمة بيانيا (٥ ،١٧، ٢٠). ويتضح من الشكل ما يلي:

- ان نتائج التحليل البياني للمفردات على الجانب الأيمن لا تكون انسيابية تماما مثل النموذج النظري السابق الإشارة إليه (شكل (١))، ولذا يتم معالجة الرسم ليصبح أكثر انسيابية كما يبدو في الجانب الأيسر (بعد تحسين الرسم).
- في حالة المفردة رقم (٥) يتزايد معدل اختيار البديل الصواب (د) كلما زادت الدرجة الكلية على الاختبار عبر جميع مستويات القدرة، بينما يتناقص بوجه عام معدل اختيار المشتتات (أ،ب،ج)، والمفردة متوسطة الصعوبة تقريبا، وتميز بشكل جيد بين مختلف مستويات القدرة حيث أن زاوية ميل خط البديل الصواب تقترب من (٤٥) درجة، ولا يوجد مشتت متروك تماما. وينطبق مثل هذا الوصف بدرجة كبيرة على المفردتين رقم (١٧ ، ٢٠).

وقد أكدت المحكات الإحصائية ملاءمة المفردات الثلاث حيث بلغت قيم إحصاءي الملاءمة (ZSTD) التقاربية والتباعدية في التحليل الأولي للاختبار (-٠,٠٩ ، -٠,٢٧) و (-٠,٢٤ ، -٠,٣٧) و (-٠,٤٥ ، -١,١٥) على الترتيب. إلا أن المفردة رقم (٢٠) تعتبر تجاوزت الحدود

الدنيا (٠,٨) للملاءمة التباعدية وفقا لمؤشر (MNSQ)، حيث بلغت قيم إحصاءي الملاءمة التقاربية والتباعدية للمفردات الثلاث وفقا لهذا المؤشر(٠,٩٩، ٠,٩٦) و (٠,٩٦، ٠,٩١) و(٠,٨٩، ٠,٦٨) على الترتيب.

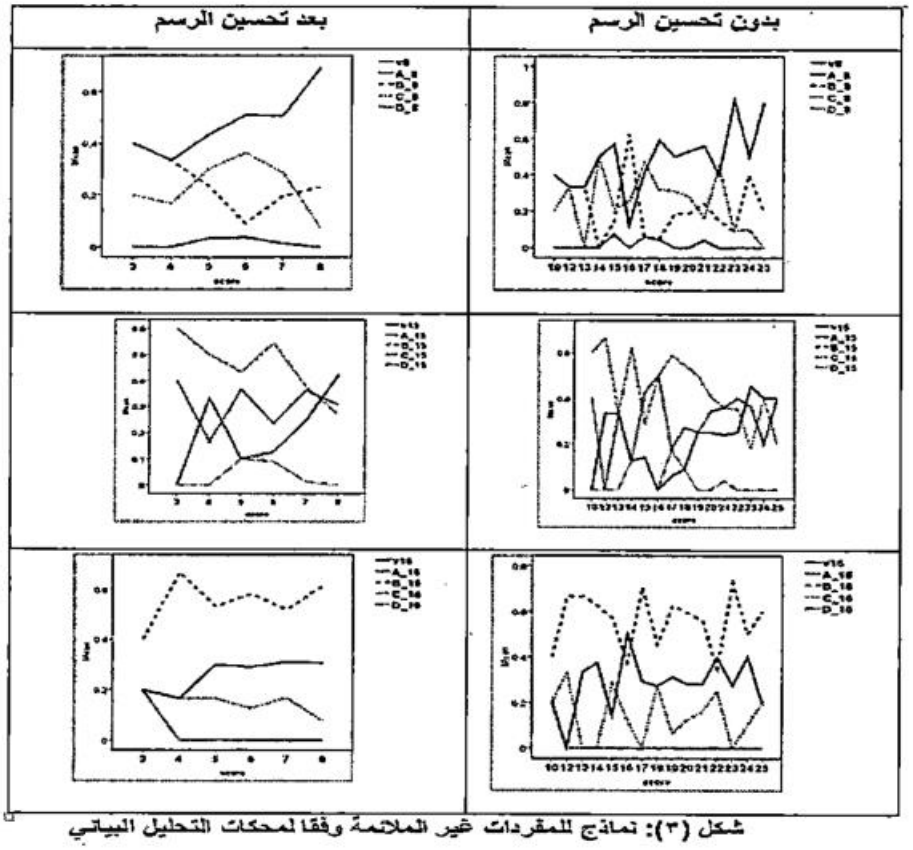
٢- نتائج التحليل البياني للمفردات الثلاث غير الملائمة بيانيا:

يوضح شكل (٣) نتائج التحليل البياني للمفردات غير الملائمة بيانيا (٨ ،١٥، ١٦). ويتضح من الشكل ما يلي:

- تعد كل من المفردتين (١٥ و ١٦) نموذجا للمفردة غير الملائمة وفقا لمحكات التحليل البياني، ومن ثم تم حذفها من الاختبار نظرا لتحقق المحكات الأربعة لضعف المفردة بدرجة كبيرة، ففي المفردة رقم (١٦):

- تتعدد النقاط التي لا يتحقق فيها شرط زيادة نسبة اختيار البديل الصواب (د) مع زيادة الدرجة الكلية على الاختيار .
- تعدد النقاط التي لا يتحقق فيها شرط تناقص نسبة اختيار المشتتين (ب) و(ج) مع زيادة الدرجة الكلية على الاختيار.ويلاحظ أن معدل اختيار المشتت (أ) ثابت تماما عبر أغلب مستويات القدرة.
- يلاحظ أن المشتت (ب) أقوى جذبا من البديل الصواب عبر مختلف مستويات القدرة.
- تعدد التقاطعات بين خط البديل الصواب وخطوط المشتتات، وتكاد لا توجد فترة تمييز للمفردة.
- ويلاحظ بوضوح انخفاض ميل خط البديل الصواب، بل ويلاحظ أيضا انخفاض ميل خطوط جميع المشتتات، مما يدعم ضعف تمييز المفردة عبر مختلف مستويات القدرة.

كما أن المؤشرات الإحصائية للملاءمة لها تبتعد بدرجة كبيرة عن الصفر إلا أنها لم تتجاوز الحدود الموجبة للملاءمة الإحصائية (٢,٥) المأخوذ بها في الدراسة وفي كثير من الدراسات السابقة، حيث بلغت قيم إحصاءي ملاءمتها التقاربية والتباعدية وفقا لإحصاء (ZSTD)(١,٩١ ، ٢,١٥) على الترتيب. إلا أن هذه المفردة تعتبر تجاوزت الحد الأعلى (١,٢) للملاءمة التباعدية وفقا لمؤشر (MNSQ) ، حيث بلغت قيم إحصاءي الملاءمة التقاربية والتباعدية لها وفقا لهذا المؤشر (١,١٣ ، ١,٢٢).



شكل (٣): نماذج للمفردات غير الملائمة وفقا لمحكات التحليل البياني

- وينطبق الوصف السابق بدرجة كبيرة على المفردة رقم (١٥) مع اختلاف درجة ميل خط البديل الصواب والمشتتات.

ورغم تعدد محكات التحليل البياني التي تبين ضعف جودة هذه المفردة، إلا أن المؤشرات الإحصائية للملاءمة لم تؤيد ذلك، حيث بلغت قيم إحصاءي ملاءمتها التقاربية والتباعدية وفقا لإحصاء (ZSTD) (٠,٢ ، ٠,٥١) على الترتيب، كما بلغت قيم إحصاءي الملاءمة التقاربية والتباعدية وفقا لإحصاء (MNSQ) (١,٠٢ ، ١,٠٧) على الترتيب.

- أما المفردة رقم (٨) فبالرغم من أن قوة جذب البديل الصواب مقارنة بالمشتتات أقوى من المفردتين السابقتين مما يجعلها أفضل نسبيا، إلا أن هناك الكثير من محكات ضعف المفردة السابق الإشارة إليها، حيث يلاحظ تعدد نقاط تقاطع خط البديل الصواب مع خطوط المشتتات، كما يلاحظ بوضوح في الرسم بعد التحسين تزايد معدل اختيار المشتت (ج) مع

زيادة الدرجة الكلية عند المستويات الأقل من المتوسط والمتوسطة من القدرة، وما أن يبدأ هذا المعدل في الانخفاض حتى يبدأ تزايد معدل اختيار المشتت (ب) بالتوازي مع البديل الصواب عند المستويات المرتفعة من القدرة .

وقد تأكدت عدم ملاءمة هذه المفردة وفقا للمؤشرات الإحصائية للملاءمة، حيث بلغت قيم إحصاءي ملاءمتها التقاربية والتباعدية وفقا لمؤشر (ZSTD) (٣,٠٤ ، ٣,٣١) على الترتيب، إلا أن هذه المفردة لم تتجاوز الحدود المثلى للملاءمة وفقا لمؤشر (MNSQ) وان كانت اقتربت من الحدود العليا للملاءمة التباعدية حيث بلغت قيمتي إحصاء الملاءمة التقاربي والتباعدي وفقا لهذا المؤشر (١,١٣ ، ١,١٨) على الترتيب.

وتبين نتائج التحليل الإحصائي أن صعوبة المفردة رقم (٨) (٥٧,٣ منف) أعلى من المتوسط قليلا، ويلاحظ من التحليل البياني لها أن أهم المشكلات في خط البديل الصواب (والمشتتات) كانت في الدرجات المتوسطة والأعلى من المتوسط والمتطرفة، أي أن الأفراد الذين يقترب مستوى قدرتهم كثيرا من مستوى صعوبة المفردة (١٩ درجة = ٥٧,٥ منف) لا يستجيبون عليها بما يتناسب والفرق بين القدرة والصعوبة، وكذلك الحال بالنسبة لذوي القدرة المنخفضة عن صعوبة المفردة (١٠درجات = ٤٠,٥ منف) أو المرتفعة (٢٤ درجة = ٦٨,٥ منف) مما يؤدي إلى ارتفاع قيمة إحصاءي الملاءمة التقاربي والتباعدي لها.

يلاحظ مما سبق أن المفردة الوحيدة التي رفضها النموذج تبين أيضا عدم ملاءمتها بيانيا، ويتسق ذلك مع ما لاحظه الباحثون في عدد من الدراسات السابقة مثل (إكرام حمزة، ٢٠٠٢ ؛ Bhakta, el. ,2005 ؛ منى ربيع، ٢٠٠٨ ؛ وليد مسعود، ٢٠١٠) حيث لاحظ الباحثون أثناء مناقشتهم للمفردات المحذوفة بناء على المحكات الإحصائية للملاءمة أن التحليل البياني لهذه المفردات يبين أيضا أنها مفردات غير جيدة. كما يلاحظ ارتفاع نسبة المفردات (٩٠%) التي اتفقت المحكات الإحصائية والبيانية على ملاءمتها، ويتفق ذلك مع ما ذهبت إليه (باسمة الشايجي، ٢٠٠٧) حيث ذكرت أن العديد من المفردات الملائمة إحصائيا كانت نتائج التحليل البياني الخاصة بها جيدة.

وحيث أن المحكات الإحصائية للملاءمة تعتمد على دراسة الاستجابات الصواب في مقابل الاستجابات الخطأ (التي ترمز جميعها بالدرجة (صفر)) مع التركيز على استجابات الأفراد الذين تقترب أو تبتعد قدرتهم عن صعوبة المفردة على متصل التدريج المشترك لكل من الصعوبة والقدرة، ولا تهتم بدراسة الاستجابة على كل مشتت على حده ، فإن بعض المفردات قد تكون

التكامل بين التحليل الإحصائي والبياني في انتقاء وتدرج مفردات الاختيارمن متعدد ملائما إحصائيا وغير ملائمة بيانيا أو العكس، وإن كانت نسبة الاتفاق أكبر من نسبة الاختلاف بين المؤشرين البياني والإحصائي في تقدير ملاءمة مفردات الاختبار.

وتتسق هذه النتائج مع ما ذهب إليه (Izard, 2005, p65) من أن الأساليب التقليدية تتفق في أغلب الأحوال مع الأساليب الحديثة في تقييم جودة المفردات، إلا أن ذلك لا ينفي وجود الاختلاف بينها. كما أن الدراسات التي اهتمت بمقارنة المدخل التقليدي بنماذج الاستجابة للمفردة في تقرير ملاءمة المفردات (مثل:أحمد عودة، ١٩٩٢ ؛ Kirishnan, 2013, P10) أسفرت عن وجود اتفاق غير تام بين المدخلين.

ويلاحظ أن المفردة رقم (١٦) قد اقتربت من الحدود الموجبة لعدم الملاءمة الإحصائية التقاربية والتباعدية وفقا لمؤشر (ZSTD) وإن لم تتجاوزها، كما تخطت الحدود العليا للملاءمة التباعدية وفقا لمؤشر (MNSQ) في حين ابتعدت المفردتين (٨) و (١٥) عن القيمة المثلى (١) للملاءمة وفقا لهذا المؤشر وان لم تتجاوز القيمة الفاصلة لها.بالإضافة إلى أن المفردة (٨) قد تجاوزت الحدود الموجبة للملاءمة الإحصائية بنوعيها لمؤشر (ZSTD). وهذه النتيجة قد تفيد بأن استخدام مؤشري الملاءمة معا قد يعطي نتائج أكثر اتساقا مع نتائج التحليل البياني للمفردات. كما أن هذه النتائج تطرح السؤال التالي: أي من مؤشري الملاءمة الإحصائية السابقين أكثر اتساقا مع محكات التحليل البياني في تقرير جودة المفردة الاختبارية؟

والخلاصة أن المفردات التي تصنف على أنها ملائمة إحصائيا وغير ملائمة بيانيا تشير نتائج التحليل الإحصائي أيضا إلى وجود مشكلات بها وان لم تصل إلى حد عدم الملاءمة، حيث تكون المحكات الإحصائية أكثر حساسية للمشكلات التي تتعلق بالبديل الصواب للإجابة ونسق الاستجابات الصواب، في حين يكون التحليل البياني أكثر حساسية لمشكلات أخرى في المفردات الاختبارية حيث يأخذ في الاعتبار أيضا معدلات اختيار كل مشتت والعلاقة بين خطوط المشتتات وخط الإجابة الصواب وميل كل منها.

ويؤيد ذلك اتساق وتكامل نتائج التحليل البياني مع نتائج التحليل الإحصائي. هذا بالإضافة إلى توفر سهولة قراءة الرسوم البيانية وتفسيرها من قبل الباحثين غير المتخصصين في القياس الأمر الذي يدعم جدوى استخدامه عند تدريج المقاييس النفسية والتربوية.

ثانيا : نتائج اختبار الفرض الثاني ومناقشتها:

ينص الفرض الثاني على : لا يؤثر حذف المفردات غير الملائمة وفقا لمحكات التحليل

البياني على إحصاءات المفردات (نسبة المفردات المحذوفة – وإحصاءات الملاءمة – تدرج المفردات – دقة التدرج– دالة معلومات الإختبار) في التدريج النهائي للاختبار تبعا لنموذج راش.

لاختبار هذا الفرض تم إجراء التحليل البياني لمفردات اختبار نظريات التعلم وتحديد المفردات الثلاث غير الملائمة وفقا لمحكات التحليل البياني كما سبق الإشارة في نتائج الفرض الأول. كما تم إجراء التحليل الإحصائي لهذه البيانات تبعا لنموذج راش باستخدام برنامج (WINSTEPS). لتحديد الحالات غير الملائمة وفقا للمحكات الإحصائية وتدريج الاختبار، وذلك في حالتي التدريج. وفيما يلي عرض لملخص نتائج التدريج، يليه نتائج المقارنة بين إحصاءات الاختبار في الحالتين، ومناقشتها:

١- ملخص نتائج المفردات في حالتي التدريج، ومناقشتها:

لتدريج الاختبار تبعا لنموذج راش، يجرى التحليل الإحصائي للبيانات على ثلاث مراحل بهدف حذف الأفراد غير الملائمين ثم المفردات غير الملائمة، ثم الحصول على التدريج النهائي للاختبار. ويلخص الجدول(٣) نتائج المفردات في الحالة الأولى للتدريج(الإحصائي) الخاصة بالاختبار الكلي بدون حذف المفردات غير الملائمة بيانيا، وكذا نتائج المفردات في الحالة الثانية للتدريج (البياني/الإحصائي) الخاصة بالاختبار بعد حذف تلك المفردات:

جدول (٣)

ملخص نتائج المفردات في حالتي تدريج الاختبار

حالة التدريج	مراحل التدريج	عدد المفردات	تقديرات الصعوبة بالمئات		الخطأ المعياري لها بالمئات		معامل الثبات
			المتوسط	الانحراف المعياري	المتوسط	الانحراف المعياري	
الإحصائي (غير الإحصائي)	الأولى	٣٠	٥٠	٦,٦٠	٠,٩٥	٠,٢٥	٠,٩٨
	النهائي	٢٩	٥٠	٧,١٠	١,٠٥	٠,٣٠	٠,٩٨
	الأولى	٢٧	٥٠	٦,٤٥	٠,٩٥	٠,٢٥	٠,٩٨
	النهائي	٢٧	٥٠	٦,٧٥	١,٠٠	٠,٣٠	٠,٩٨

يتبين من الجدول السابق ما يلي :

- تم حذف مفردة واحدة لا تصلح للتدريج وفقا للمحكات الإحصائية للملاءمة، وتعتبر نسبة المفردات المحذوفة (٠,٠٣٣ %) صغيرة جدا مقارنة بنتائج بعض الدراسات السابقة التي استخدمت نموذج راش في تدريج الاختبارات المعرفية ، مثل (أمين نور الدين، ١٩٩٥ ؛ شادية عبد العزيز، ١٩٩٧ ؛ منى ربيع ، ٢٠٠٠ ؛ شادية عبد العزيز، ٢٠٠٢ ؛ أحلام الباز ورضا حجازي، ٢٠٠٢ ؛ Bhakta, el. ,2005 ؛ نضال الشريفين، ٢٠٠٩ ؛ علي

زكري، ٢٠١١). فقد تراوحت نسبة عدد المفردات المحذوفة في تلك الدراسات بين (٣,٨ %)و (٣٢,٤٣ %)، عدا دراسة نضال الشريفين الذي درج سبعة نماذج اختبارية وتراوحت نسب المفردات المحذوفة منها بين (صفر%) و (١١%) مما يؤكد ملاءمة الاختبار المستخدم في الدراسة الحالية لأسس القياس كما تتمثل في نموذج راش.

- وفي حالة الأخذ في الاعتبار نتائج التحليل البياني قبل التدريج ، لم يسفر التحليل الإحصائي عن وجود مفردات إضافية غير صالحة للتدريج مقارنة بنتائج التدريج الكلي.
- بلغت قيمة متوسط تقديرات صعوبة المفردات في جميع مراحل التحليل في الحالتين (٥٠) منف أي (صفر) لوجيت وهو صفر اعتباري حيث يعتبر البرنامج أن متوسط صعوبات المفردات هو صفر التدريج، وحيث أنه لم تتم معادلة التدريج فإن المتوسط الحقيقي لهذه التقديرات قد يكون مختلفا.
- يلاحظ انخفاض قيم الخطأ المعياري لتقديرات الصعوبة في حالتي التدريج مما يشير إلى دقة وثبات تقديرات الصعوبة، ويؤكد ذلك ارتفاع قيم معاملات الثبات (٠,٩٨) التي لم تختلف في جميع مراحل التحليل وحالاته.

٢- نتائج مقارنة نسبة المفردات المحذوفة وفقا للمحكات الإحصائية في حالتي التدريج، ومناقشتها:

يوضح الجدول التالي المفردات غير الملائمة لأسس القياس الموضوعي وفقا للمحكات الإحصائية للملاءمة، وذلك في حالتي التدريج. وقد سبق الإشارة إلى أن استبعاد الأفراد غير الصالحين للتدريج يسبق تحديد المفردات غير الملائمة إحصائيا عند تدريج الاختبار تبعا لنموذج راش، ولذا قد يلاحظ اختلاف طفيف في قيم إحصاءات الملاءمة المتضمنة بالجدول التالي مقارنة بتلك الواردة في نتائج الفرض الأول:

جدول (٤)

المفردة التي تجاوزت الحدود (±٢,٥) على تدريج الملاءمة في حالتي التدريج

رقم المفردة	حالة التدريج	الصعوبة بالمنقف	الخطأ المعياري	إحصاء الملاءمة	
				التقاربي	التباعدي
٨	الإحصائي	٥٤٠	٠,٨٠	٣,٠٠	٣,٤٢
	البياني| الإحصائي	--	--	--	--

يتبين من الجدول السابق أن المفردة تعتبر غير صالحة للتدريج في حالة تدريج الاختبار الكلي، وأنها قد تجاوزت الحدود الموجبة للملاءمة مما قد يعني سوء صياغتها ، أو أنها تعرف

متغيرا آخر يختلف عن المتغير الذي تعرفه باقي المفردات (Wright & Linacre , 1985 , pp4-15 : 4-20). ولذا فإن المفردة لا تحقق أسس القياس الدقيق ومن ثم فقد حذفت من التدريج.

وبمراجعة نص المفردات وجد أن المفردة صياغتها جيدة ومباشرة وتقيس المستوى المعرفي التذكر، وبمراجعة محتوى المقرر الذي تم تدريسه وجد أن بعض المجموعات لم تدرسن مع أستاذ المادة الجزء من المحتوى التي تنتمي إليه المفردة، حيث يقوم عدة أساتذة بتدريس المقرر. ويرجح أن ذلك قد أدى إلى اختلاف صعوبة المفردة لدى المجموعات المختلفة للطالبات مما أدى إلى ارتفاع قيم إحصاءي ملاءمتها.

ويلاحظ أن هذه المفردة قد تبين عدم ملاءمتهما وفقا لنتائج التحليل البياني، ولذا فقد حذفت ضمن المفردات الثلاث غير الملائمة بيانيا قبل التحليل الإحصائي، وبإجراء التحليل الإحصائي لتدريج باقي المفردات (٢٧ مفردة) وفقا لنموذج راش لم تسفر النتائج عن مفردات أخرى غير ملائمة وفقا للمحكات الإحصائية . وتختلف هذه النتيجة عن نتائج (أحلام الباز ورضا حجازي، ٢٠٠٢ ؛ مصطفى قاسم، ٢٠١٥) التي قامت باستبعاد المفردات غير الملائمة بيانيا قبل التحليل الإحصائي وتدريج الاختبار، حيث وجد الباحثون مفردات أخرى غير ملائمة وفقا للمحكات الإحصائية للملاءمة.

وبذلك يمكن القول بأن: حذف المفردات السيئة وفقا لمؤشر التحليل البياني لا يؤدي إلى اختلاف هوية ونسبة المفردات غير الملائمة وفقا لمحكات الملاءمة الإحصائية عند تدريج باقي مفردات المقياس وفقا لنموذج راش.

٣- نتائج مقارنة إحصاءات ملاءمة المفردات في حالتي التدريج النهائي، ومناقشتها:

يوفر برنامج (Winsteps) مؤشرين لكل من إحصاءي الملاءمة التقاربي والتباعدي هما (ZSTD) و (MNSQ)، وتتضمن مخرجات البرنامج قيم إحصاءات الملاءمة بنوعيها التقاربي والتباعدي لكل مفردة ولكل فرد وكذا متوسط هذه القيم للاختبار ولعينة التدريج ككل.

وقد تضمنت عملية التدريج حذف المفردات غير الملائمة لأسس القياس وفقا لمؤشر (ZSTD) واستبقيت فقط المفردات الصالحة للتدريج، وهذه المفردات وإن لم تتجاوز الدرجة الفاصلة للملاءمة(±٢,٥) إلا أنها تتباين من حيث اقترابها من درجة الملاءمة التامة (الصفر). وبالمثل فإن هذه المفردات تتفاوت من حيث درجة اقترابها من القيمة المثلى (واحد صحيح)

= التكامل بين التحليل الإحصائي والبياني في انتقاء وتدرج مفردات الاختيارمن متعدد= لمؤشر(MNSQ) الذي يعبر عن حجم الملاءمة للنموذج.

وتهتم الدراسة باختبار ما إذا كان البدء بالتحليل البياني وحذف المفردات غير الملائمة بيانيا ثم تدريج الاختبار يؤثر في درجة ملاءمة الاختبار وملاءمة مفرداته في التدريج النهائي، وذلك كما يتضح مما يلي:

أ- نتائج مقارنة ملاءمة الاختبار في تدريجه النهائي، ومناقشتها:

الجدول التالي يتضمن قيم المتوسط والوسيط والانحراف المعياري لإحصاءات ملاءمة المفردات المتضمنة في مخرجات البرنامج في التدريج النهائي في حالتي التدريج:

جدول (٥)

المقارنة بين إحصاءات الملاءمة في التدريج النهائي للاختبار في حالتي التدريج

إحصاء الملاءمة		حالة التدريج	عدد المفردات	المتوسط	الوسيط	الانحراف المعياري
ZSTD	التقاربي	الإحصائي	٣٩	٠,١٠	٠,٠٠	٠,٦٠
		البياني/ الإحصائي	٣٧	٠,١٠	٠,٠٨	٠,٨٠
	التباعدي	الإحصائي	٣٩	٠,٠٠	-٠,٢١	٠,٨٠
		البياني/ الإحصائي	٣٧	٠,٠٠	-٠,٣١	١,٠٠
MNSQ	التقاربي	الإحصائي	٣٩	١,٠٠	٠,٩٩	٠,٠٥
		البياني/ الإحصائي	٣٧	١,٠٠	٠,٩٩	٠,٠٦
	التباعدي	الإحصائي	٣٩	٠,٩٦	٠,٩٦	٠,١١
		البياني/ الإحصائي	٣٧	٠,٩٦	٠,٩٧	٠,١٤

يبين الجدول السابق أن الاختبار في تدريجه النهائي يحقق درجة مرتفعة من الملاءمة لنموذج راش سواء اعتمد التدريج على نتائج التحليل الإحصائي فقط أو على نتائج كل من التحليل البياني والإحصائي، حيث تقترب الملاءمة في الحالتين من الملاءمة التامة.

وقد كان الفرق بين ملاءمة الاختبار في الحالتين صفريا حيث تتساوى قيمتي متوسط إحصاء الملاءمة التقاربي تماما في حالتي التدريج، كما تتساوى قيمتي متوسط إحصاء الملاءمة التباعدي تماما في حالتي التدريج، ومن ثم لا تكون هناك حاجة لاستخدام أحد الأساليب الإحصائية الاستدلالية لاختبار دلالة الفروق بينها.

وحيث أن القيم السالبة للملاءمة بنوعيها لمؤشر ZSTD كلما ابتعدت عن الصفر كلما كانت المفردة أقل ملاءمة ، وعلى ذلك فإن القيمة (- ٢) مثلا ليست أقل من القيمة (-١) من حيث مدلولها على متصل الملاءمة ، لذا فقد رأت الباحثة أن تقوم بمقارنة القيم المطلقة لإحصاءات الملاءمة في التدريج النهائي في حالتي التدريج، فأكدت النتائج عدم وجود فروق دالة إحصائيا بين

إحصاءات ملاءمة الاختبار في تدريجه النهائي في حالتي التدريج، حيث بلغ مستوى دلالة الفرق (٠,٦٣) و(٠,٤٣) بالنسبة لإحصاءي الملاءمة التقاربي والتباعدي على الترتيب.

وبالمثل تمت مقارنة انحراف قيم مؤشر **MNSQ** عن القيمة المثلى (الواحد الصحيح)، وكذا القيم المطلقة لهذا الانحراف، وذلك لكل من إحصاءي الملاءمة التقاربي والتباعدي. وقد أكدت النتائج عدم وجود فروق دالة إحصائيا بين إحصاءات ملاءمة الاختبار في تدريجه النهائي في حالتي التدريج، حيث تراوحت قيم مستوى الدلالة بين (٠,٤٠) و (٠,٩١). وقد يرجع ذلك إلى ارتفاع درجة ملاءمة الاختبار لأسس القياس الموضوعي قبل استبعاد المفردات غير الملائمة سواء وفقا للمحكات البيانية أو الإحصائية.

ب- نتائج مقارنة القيم المتناظرة لإحصاءات ملاءمة المفردات المشتركة في التدريج النهائي للاختبار في حالتي التدريج، ومناقشتها:

لإجراء هذه المقارنة، تم استخدام اختبار (ت) حالة العينتين المرتبطتين، وكانت النتائج كما يبين الجدول التالي:

جدول (٦)

مقارنة إحصاءات ملاءمة المفردات المشتركة في حالتي التدريج النهائي

إحصاء الملاءمة		متوسط الفروق	الانحراف المعياري	ت	د.ح	مستوى الدلالة	حجم الفرق	معامل الارتباط	مستوى الدلالة
ZSTD	التقاربي	٠,٠٤	٠,٢٩	٠,٦٦	٢٦	٠,٥١	صغير	٠,٩٦	٠,٠٠
	التباعدي	٠,٠٧	٠,٢٨	١,٣	٢٦	٠,٢١	متوسط	٠,٩٧	٠,٠٠
MNSQ	التقاربي	٠,٠٠٢	٠,٠٢٥	٠,٤٦	٢٦	٠,٦٥	صغري	٠,٩١	٠,٠٠
	التباعدي	٠,٠٠٢	٠,٠٤٢	٠,٢٣	٢٦	٠,٨٢	صغري	٠,٩٧	٠,٠٠

يبين الجدول السابق أن القيم المتناظرة لإحصاءات الملاءمة بنوعيها والخاصة بالمفردات المشتركة في التدريج النهائي لم تختلف في حالتي التدريج: التحليل الإحصائي فقط و التحليل البياني ثم الإحصائي. كما أن توزيع المفردات على متصل الملاءمة بنوعيها لم يختلف في حالتي التدريج، حيث كانت قيم معامل الارتباط مرتفعة ودالة إحصائيا عند مستوى دلالة أقل من (٠,٠١).

مما سبق يتبين عدم اختلاف درجة ملاءمة الاختبار في تدريجه النهائي بالإضافة إلى عدم اختلاف توزيع المفردات على متصل الملاءمة في حالة اعتماد التدريج على نتائج التحليل الإحصائي فقط وحالة اعتماده على نتائج كل من التحليل البياني والإحصائي. وبذلك يمكن القول

بأن : حذف المفردات غير الملائمة وفقا لمحكات التحليل البياني لا يؤثر على إحصاءات ملاءمة المقياس في تدريجه النهائي وفقا لنموذج راش.

وقد يرجع ذلك إلى أن الاختبار في تدريجه النهائي يحقق درجة مرتفعة من الملاءمة لأسس القياس الموضوعي في الحالتين. فقد تم إنشاء التدريج النهائي في الحالة الأولى باستخدام المفردات والأفراد الصالحة للتدريج بعد استبعاد غير الملائم منها وفقا للمحكات الإحصائية. كما أن التدريج في الحالة الثانية وان كان قد بدأ بانتقاء المفردات وفقا لمحكات التحليل البياني إلا أنه أيضا اعتمد إضافة إلى ذلك على المحكات الإحصائية لانتقاء المفردات والأفراد الصالحة للتدريج. وقد كان الأفراد المستبعدين لعدم صلاحيتهم لإنشاء التدريج مختلفا في الحالتين – كما يتبين من نتائج الفرض الثالث – بحيث كانت مصفوفة الاستجابات النهائية تحقق شروط النموذج بما أدى إلى تكافؤ خصائص المفردات في التدريج النهائي في حالتي التدريج.

٤- نتائج مقارنة تدرج المفردات على متصل الصعوبة في حالتي التدريج، ومناقشتها:

نظرا لاختلاف صفر تدريج الصعوبة فقد استخدم المدى ومعامل الاختلاف ومعامل الارتباط بين تقديرات الصعوبة في حالتي التدريج وذلك لدراسة أثر حذف المفردات غير الملائمة وفقا لمحكات التحليل البياني على تدرج المفردات على متصل الصعوبة. ومن ثم اقتصرت المقارنة على المفردات المشتركة فقط في حالتي التدريج النهائي والبالغ عددها (٢٧) مفردة، ولذا يلاحظ اختلاف بعض هذه الإحصاءات نسبيا في الجدول التالي مقارنة بتلك الواردة بملخص نتائج المفردات:

جدول (٧)

المقارنة بين تدرج المفردات المشتركة في حالتي التدريج النهائي (ن=٢٧ مفردة)

حالة التدريج	تقديرات الصعوبة بالمئة			معامل الاختلاف	معامل الارتباط	مستوى الدلالة
	المدى	المتوسط	الانحراف المعياري			
الإحصائي	٢٥	٤٩,١٩	٦,٨٠	٠,١٤	٠,٩٩٨	٠,٠٠٠
البياني/ الإحصائي	٢٦	٥٠	٦,٧٥	٠,١٤		

من الجدول السابق يلاحظ ما يلي:

- لا يختلف مدى تقديرات الصعوبة في التدريج النهائي لمجموعة المفردات المشتركة (٢٧ مفردة) في حالتي التدريج مع الحذف وبدون الحذف وفقا لنتائج التحليل البياني، حيث أن الفرق بين قيمتي المدى لا يتجاوز متوسط الخطأ المعياري لتقديرات صعوبة تلك المفردات في إحدى

حالتي التدريج.

- لم يتغير تشتت تقديرات الصعوبة في حالتي التدريج مع الحذف وبدون الحذف وفقا لنتائج التحليل البياني، حيث تتساوي قيم معاملي الاختلاف مما يدل على عدم اختلاف تباين تلك التقديرات في الحالتين.
- يوجد ارتباط طردي قوي – يكاد يكون تام – دال إحصائيا عند مستوى دلالة أقل من (٠,٠١) بين قيم تقديرات الصعوبة قبل وبعد حذف المفردات غير الملائمة وفقا لنتائج التحليل البياني، مما يعني اتساق توزيع المفردات على متصل الصعوبة في الحالتين.

ومن ثم يمكن القول بأن : حذف المفردات غير الملائمة وفقا لمحكات التحليل البياني لا يؤثر على تدرج المفردات على المتغير موضوع القياس عند تدريج الاختبار وفقا لنموذج راش. وتتسق هذه النتائج ما يحققه نموذج راش من استقلالية القياس.

٥- نتائج مقارنة دقة تدرج المفردات في حالتي التدريج، ومناقشتها:

اعتمدت الدراسة في تقدير دقة تدرج المفردات على كل من معامل ثبات كيودر ريتشاردسون-٢٠ الذي يحسبه برنامج Winsteps، بالإضافة إلى الخطأ المعياري لتقديرات الصعوبة.

أ- فيما يتعلق بمعامل الثبات فقد سبق الإشارة إلى أن معامل ثبات تقديرات الصعوبة لم يختلف في جميع مراحل التحليل وحالات التدريج، حيث بلغت قيمته (٠,٩٨)، وتدل هذه القيمة على ارتفاع ثبات تقديرات الصعوبة في الحالتين. وتتفق هذه النتيجة مع نتائج الدراسات السابقة (مثل: شادية عبد العزيز، ١٩٩٧؛ إكرام حمزة، ٢٠٠٢ ؛ Bhakta, el. ,2005 ؛ باسمة الشايجي، ٢٠٠٧) حيث حصلت تلك الدراسات على قيم مرتفعة لمعاملات ثبات تقديرات الصعوبة عند تدريج المقاييس المعرفية باستخدام نموذج راش.

ب- أما فيما يتعلق بالخطأ المعياري فقد تمت المقارنة بين قيم الخطأ المعياري لتقديرات صعوبة مفردات الاختبار ككل وكذا بين القيم المتناظرة للخطأ المعياري لتقديرات صعوبة المفردات المشتركة في حالتي التدريج النهائي، وذلك كما يلي:

- نتائج المقارنة بين قيم الخطأ المعياري لتقديرات صعوبة مفردات الاختبار في حالتي التدريج النهائي:

أجريت هذه المقارنة باستخدام اختبار مان ويتني – نظرا لالتواء التوزيع- وكانت النتائج كما يلي:

جدول (٨)

المقارنة بين قيم الخطأ المعياري لتقديرات الصعوبة في حالتي التدريج النهائي

حالة التدريج	ن	الخطأ المعياري بالمنف			مجموع الرتب	z	مستوى الدلالة	حجم التأثير
		المدى	التوسيط	الانحراف الربيعي				
الإحصائي	٢٩	١,٣	٠,٩٠	٠,١٥	٨٤٥,٥	٠,٣٢	٠,٧٥	٠,٠٥ (صغير)
البياني/ الإحصائي	٢٧	١,٥	٠,٩٠	٠,١٥	٧٥٠,٥			

يتضح من الجدول السابق ما يلي:

- لا يوجد فرق بين قيم الخطأ المعياري لتقديرات الصعوبة لمفردات الاختبار ككل في تدريجه النهائي في حالتي التدريج، حيث تتساوى تماما قيمت الوسيط والانحراف الربيعي، وقد أكد اختبار مان ويتني ذلك حيث بلغ مستوى الدلالة (٠,٧٥).
- يلاحظ انخفاض قيم الخطأ المعياري في حالتي التدريج حيث لا يتجاوز متوسطها منف واحد تقريبا، وقد تراوحت قيم الخطأ المعياري بين (٠,٨٠) و (٢,٣٠) منف، وبالنظر إلى أن تدريج المنف هو تدريج مئوي متوسطه يساوي (٥٠) وانحرافه المعياري يساوي (٥) فإن قيم الخطأ المعياري تؤكد ارتفاع ثبات تقديرات صعوبة مفردات الاختبار في حالتي التدريج. وقد سبق الإشارة إلى ارتفاع قيم معامل الثبات الهامشي للاختبار الذي يعتمد على الخطأ المعياري.

• نتائج المقارنة بين قيم الخطأ المعياري لتقديرات صعوبة المفردات المشتركة في حالتي التدريج النهائي:

لإجراء هذه المقارنة، تم استخدام اختبار ولكوكسن ومعامل ارتباط سبيرمان نظرا لالتواء التوزيع، وكانت النتائج كما بالجدول التالي:

جدول (٩)

المقارنة بين قيم الخطأ المعياري للمفردات المشتركة في حالتي التدريج النهائي (ن= 27 مفردة)

مجموع الرتب الموجبة(*)	مجموع الرتب السالبة	z	مستوى الدلالة	حجم التأثير	معامل الارتباط	مستوى الدلالة
١٦٢	٢١٦	٠,٦٥	٠,٥٢	٠,١٤ (ضعيف)	٠,٩٩	٠,٠٠

(*)الخطأ المعياري في الحالة الثانية للتدريج أكبر من الخطأ المعياري في الحالة الأولى.

يتضح من الجدول السابق ما يلي:

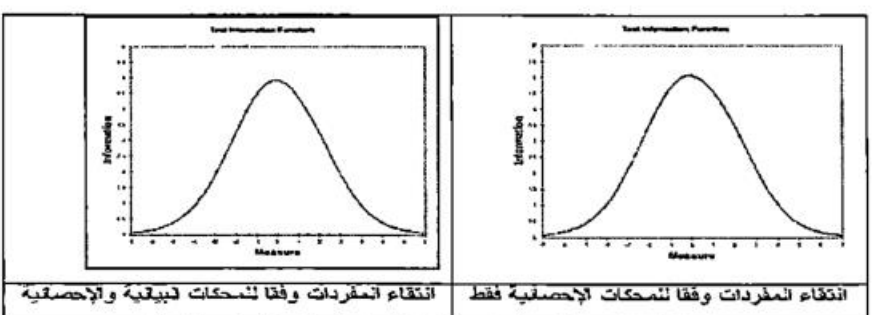
- لا تختلف قيم الخطأ المعياري لتقديرات صعوبة المفردات المشتركة في التدريج النهائي في حالتي التدريج – بحذف وبدون حذف المفردات غير الملائمة وفقا لنتائج التحليل البياني – حيث بلغ مستوى دلالة الفرق (٠,٥٢) باستخدام اختبار ولكوكسون.
- يوجد ارتباط طردي قوي – يكاد يكون تام – دال إحصائيا عند مستوى دلالة أقل من (٠,٠١) بين قيم الخطأ المعياري لتقديرات الصعوبة قبل وبعد حذف المفردات غير الملائمة وفقا لنتائج التحليل البياني، مما يعني اتساق توزيع قيم الخطأ المعياري للمفردات المتناظرة في حالتي التدريج.

كما وجد أن تشتت قيم الخطأ المعياري للمفردات المشتركة متساو في حالتي التدريج، حيث بلغت قيمة الانحراف الربيعي (٠,١٥) منف في كل من الحالتين مع تساوي قيمتي الوسيط (٠,٩ منف). وتؤكد هذه النتائج عدم تغير ثبات تقديرات الصعوبة عند حذف المفردات غير الملائمة بيانيا قبل التدريج الإحصائي. وقد يرجع ذلك إلى ارتفاع ملاءمة الاختبار لأسس القياس الموضوعي في حالتي التدريج كما سبقت الإشارة.

ومن ثم يمكن القول بأن : حذف المفردات غير الملائمة وفقا لمحكات التحليل البياني لا يؤثر على دقة التدرج النهائي للمفردات على متصل المتغير موضوع القياس عند تدريج الاختبار وفقا لنموذج راش.

٦- نتائج مقارنة دالة معلومات الاختبار:

يتيح برنامج Winsteps الحصول على دوال معلومات المفردات والاختبار، والشكل التالي يوضح دالة معلومات الاختبار في التدريج النهائي في حالتي التدريج:



انتقاء المفردات وفقا للمحكات الإحصائية فقط | انتقاء المفردات وفقا للمحكات البيانية والإحصائية

شكل(٤): دالة معلومات الاختبار في حالتي التدريج النهائي

يتضح من الشكل السابق أن دالة معلومات الاختبار جيدة في حالتي التدريج، حيث بلغت قيمتها (٤,١٤) و(٣,٧٠) لوجيت في حالتي التدريج على الترتيب، وذلك مع ملاحظة أنها تتأثر بقلة عدد مفردات الاختبار، حيث انخفض عدد مفردات الاختبار في الحالة الثانية إلى (٢٧) مفردة مقارنة بـ(٢٩) مفردة في الحالة الأولى للتدريج، وكما سبق الإشارة فإن دالة معلومات الاختبار ترتفع بزيادة عدد مفرداته (Toland , 2014 , pp145).

ويلاحظ أنه لم يتم توحيد صفر التدريج في الحالتين، وبالرجوع لأعلى تقدير للقدرة المحتمل على الاختبار – بعد استبعاد الدرجات التامة- نجد أنه يساوي (٤,١٦) و(٤,٠١) لوجيت في حالتي التدريج على الترتيب. وتعد دالة المعلومات في الحالتين مؤشرا لارتفاع دقة تقدير القدرة ولا سيما المستويات المتوسطة من القدرة، حيث يبلغ المنحنى أقصى ارتفاع له عند تقدير القدرة صفر لوجيت في حالتي التدريج.

وعلى ذلك يمكن القول بأن: حذف المفردات غير الملائمة وفقا لمحكات التحليل البياني لا يؤثر على خصائص الاختبار في تدريجه النهائي وفقا لنموذج راش، سواء من حيث ملاءمة الاختبار ككل ومفرداته وتوزيعها على متصل الملاءمة، أو دقة تدرج المفردات على متصل المتغير موضوع القياس أو دالة معلومات الاختبار. وبذلك تتحقق صحة الفرض الثاني من فروض الدراسة.

وبالرغم من أن باتنبرج ولاروس يعتقدان أن خصائص المفردات سوف **تتحسن** بعد حذف المفردات غير الملائمة وفقا لمحكات التحليل البياني. (Batenburg & Laros,2002, P331). إلا أن اهتمام الدراسة الحالية يتركز في خصائص **التدريج النهائي** للاختبار باستخدام نموذج راش – أي بعد الاستفادة من المحكات الإحصائية للملاءمة في انتقاء البيانات والمفردات الصالحة لإنشاء التدريج – ولم تهتم الدراسة بخصائص المفردات (والأفراد) بعد الاستفادة من التحليل البياني على حده. فالدراسة الحالية تهتم بالتحليل البياني كإجراء مساند لعملية التدريج باستخدام نماذج نظرية الاستجابة للمفردة.

ثالثا : نتائج اختبار الفرض الثالث، ومناقشتها:

ينص الفرض الثالث على: **لا يؤثر حذف المفردات غير الملائمة وفقا لمحكات التحليل البياني على إحصاءات الأفراد (نسبة الأفراد المحذوفين – إحصاءات الملاءمة – تدرج الأفراد – دقة التدرج) في التدريج النهائي للاختبار تبعا لنموذج راش.** ولاختبار هذا الفرض تم إتباع نفس إجراءات اختبار الفرض الثاني من فروض الدراسة، وذلك كما يلي:

١- ملخص نتائج الأفراد في حالتي التدريج، ومناقشتها:

يلخص الجدول التالي نتائج الأفراد في الحالة الأولى للتدريج بدون حذف المفردات غير الملائمة بيانيا، وكذا نتائج الأفراد في الحالة الثانية للتدريج بعد حذف تلك المفردات:

جدول (١٠)
ملخص نتائج الأفراد في حالتي تدريج الاختبار

حالة التدريج	مراحل التدريج	عدد الأفراد	تقديرات القدرة بالمئتف		الخطأ المعياري لها بالمئتف		معامل الثبات
			المتوسط	الانحراف المعياري	المتوسط	الانحراف المعياري	
الأصلي	الأولى	١٩٩	٥٣,٩٠	٣,٤٥	٢,٣٥	٠,٢٠	٠,٥٢
	النهائي	١٨٢	٥٤,٣٥	٣,٧٠	٢,٤٥	٠,٢٠	٠,٥٥
المعدل الأصلي	الأولى	١٩٩	٥٤,٨٥	٣,٩٠	٢,٥٥	٠,٣٠	٠,٥٦
	النهائي	١٩١	٥٥,٠٥	٣,٩٥	٢,٦٠	٠,٣٥	٠,٥٧

يتبين من الجدول السابق ما يلي :

- تم حذف (١٧) فردا غير ملائمين لأسس القياس، وفي حالة الأخذ في الاعتبار نتائج التحليل البياني قبل تدريج الاختبار تم حذف (٨) أفراد فقط. وبذلك لم تتجاوز نسبة الأفراد المحذوفين (٨,٥٤ %) انخفضت إلى (٤,٠٢ %) في الحالة الثانية للتدريج، وهي نسبة صغيرة نسبيا مقارنة بنتائج بعض الدراسات السابقة التي استخدمت نموذج راش في تدريج الاختبارات المعرفية ، مثل (أمين نور الدين، ١٩٩٥ ؛ شادية عبد العزيز، ١٩٩٧؛ أحلام الباز ورضا حجازي، ٢٠٠٢ ؛ Bhakta, et.al. ,2005 ؛ نضال الشريفين، ٢٠٠٩ ؛ علي زكري، ٢٠١١ ؛ هالة يوسف وآخرون، ٢٠١٣). فقد تراوحت نسبة الأفراد المحذوفين في هذه الدراسات بين (١,٤ %) إلى (٢٣,٥ %) تقريبا، مما يرجح ملاءمة بيانات الدراسة الحالية لأسس القياس الموضوعي كما تتمثل في نموذج راش.
- تقترب قيم متوسط تقديرات القدرة مع القيم المناظرة لها لمتوسط تقديرات الصعوبة المدرجة بجدول (٣)، حيث أن الفرق بين كل متوسطين متناظرين لا يتجاوز متوسط الخطأ المعياري لها، وذلك في مختلف مراحل وحالات التدريج. ويدل ذلك على أن صعوبة الاختبار موائمة بدرجة كبيرة لمستوى المختبرين، أي أن أداة الدراسة مناسبة لمستوى العينة. ولذا ارتفعت دالة معلومات الاختبار في حالتي التدريج كما سبق الإشارة.
- تتقارب قيم متوسط تقديرات القدرة في الحالتين، وحيث أنه لم يتم معادلة التدريج فإن المتوسط الحقيقي لهذه التقديرات قد يكون أو لا يكون مختلفا.
- تعتبر قيم الخطأ المعياري منخفضة في جميع مراحل وحالات التدريج، مما يشير إلى دقة

وثبات تقديرات القدرة. وقد يؤيد ذلك أن متوسط قيم الخطأ المعياري أقل من الانحراف المعياري لتقديرات القدرة.

- يلاحظ ارتفاع قيم الخطأ المعياري وانخفاض معاملات الثبات لتقديرات القدرة نسبيا مقارنة بقيم الخطأ المعياري والثبات للمفردات (جدول (٣))، وتتسق هذه النتيجة مع نتائج الدراسات التطبيقية السابقة في مجال تدريج الاختبارات المعرفية وغير المعرفية باستخدام نموذج راش (مثل: منى ربيع ، ٢٠٠٠ ؛ شادية عبد العزيز، ٢٠٠٢ ؛ أحلام الباز ورضا حجازي، ٢٠٠٢؛ أمينة كاظم وآخرون، ٢٠٠٦ ؛ باسمة الشايجي، ٢٠٠٧ ؛ Muis, et.al. ,2009 ؛ نضال الشريفين، ٢٠٠٩ ؛ حصة عبد الرحمن وآخرون، ٢٠٠٩ ؛ علي زكري، ٢٠١١) حيث يلاحظ ارتفاع قيم معاملات ثبات تقديرات المفردات مقارنة بمعاملات ثبات تقديرات الأفراد مع تباين أحجام العينات وأطوال الاختبارات المستخدمة في تلك الدراسات.

٢- نتائج مقارنة نسبة الأفراد غير الملائمين وفقا للمحكات الإحصائية في حالتي التدريج، ومناقشتها:

تم تصنيف أفراد عينة الدراسة (١٩٩ فردا) إلى ملائمين وغير ملائمين (تجاوزوا الحدود (±٢) من تدريج الملاءمة التقاربية و(أو) التباعدية) وفقا للمحكات الإحصائية في حالتي تدريج الاختبار. فوجد أن (١١) فردا من الأفراد غير الملائمين في الحالة الأولى للتدريج قد تجاوزوا الحدود الموجبة للملاءمة وتجاوز الستة الباقون الحدود السالبة للملاءمة(جدول (١١)). ونظرا لأن تجاوز الحدود الموجبة للملاءمة يعني أن نمط استجابات الفرد يختلف عما يتوقع منه نتيجة للتسرع أو الإهمال أو التخمين ، وتجاوز الحدود السالبة للملاءمة يعني أن نمط استجابات الفرد متسق بدرجة غير واقعية نتيجة للبطء أو الحرص الشديد أو الغش (Wright & Stone , 1979 , pp170-171)، فإن استجابات هؤلاء الأفراد لا تحقق أسس القياس الدقيق ومن ثم فقد حذفوا من عينة التدريج.

أما في الحالة الثانية للتدريج، فقد تجاوز ستة أفراد الحدود الموجبة للملاءمة وتجاوز فردان الحدود السالبة للملاءمة، وتم حذفهم من عينة هذا التدريج. وللمقارنة بين نسبة الأفراد الملائمين وغير الملائمين إحصائيا في الحالتين تم استخدام اختبار النسبة الحرجة لدلالة الفرق بين نسبتين مئويتين مرتبطتين كما يوضح جدول (١١) التالي:

جدول (١١)

المقارنة بين الأفراد الذين تجاوزوا الحدود (±٢) من تدريج الملاءمة في حالتي التدريج

حالة التدريج	رقم الفرد	إحصاء الملاءمة ZSTD		تابع رقم الفرد	إحصاء الملاءمة ZSTD		النسبة المئوية	Z
		التقاربي	التباعدي		التقاربي	التباعدي		
الأصلي	١٢٧	-٢,٨٢	-٢,١٤	٨٥	٢,٠٨	٢,٢٧	%٨,٥٤	٢,٧١
	١٩١	-٢,٦٠	-١,٨٩	٤٠	٢,١٧	٢,٣٨		
	١٣٨	-٢,٠٨	-١,٠٠	٥٣	٢,٢٦	٢,١٠		
	١٩٩	-٢,٠٤	-١,٧١	١٥٣	٢,٢٦	٢,١٠		
	٩	-٢,٠٣	-١,٧٦	٦٥	٢,٣٧	٢,٠٦		
	٨	-٢,٠٣	-١,٧٦	١٥١	٢,٤٨	٢,٣٦		
	٤٣	١,٠٧	٢,٩٣	١٢٤	٢,٨٢	٣,٣٨		
	٤٧	١,٢٧	٢,٥٤	٥	٢,٨٦	٢,٥٤		
	١٤٥	١,٩٣	٢,٠٩	-	-	-		
البديل الإحصائي	١٢٧	-٢,٣٢	-١,٦٩	٩١	١,٤٦	٢,٠٩	%٤,٠٢	
	١٩١	-٢,٣٢	-١,٦٩	١٢٤	٢,١٥	٢,٨٣		
	٤٣	١,٤١	٢,٧٥	٦٥	٢,٢٠	٢,٠٣		
	٤٧	١,٠٩	٢,٣٢	٥	٢,٣٢	٢,٢٩		

يتبين من الجدول السابق اختلاف نسبة الأفراد الذين تجاوزوا الحد الفاصل للملاءمة في حالتي التدريج، حيث بلغت قيمة النسبة الحرجة (٢,٧١) وهي قيمة دالة إحصائيا عند مستوى الدلالة أقل من (٠,٠١). أي أن نسبة الأفراد الذين تجاوزوا الحد الفاصل للملاءمة قد انخفضت انخفاضا دالا إحصائيا في الحالة الثانية للتدريج.

وهذه النتيجة منطقية لأن تحديد واستبعاد الأفراد غير الملائمين إحصائيا يتم في المرحلة الأولى للتدريج أي قبل استبعاد المفردات غير الملائمة إحصائيا، وبالتالي فقد ارتفعت نسبة الاستجابات غير الملائمة في الحالة الأولى للتدريج نتيجة وجود المفردات غير الملائمة سواء إحصائيا أو بيانيا، في حين انخفضت هذه النسبة في الحالة الثانية نتيجة لاستبعاد تلك المفردات بناء على محكات التحليل البياني قبل البدء في مراحل التحليل الإحصائي بهدف تدريج الاختبار.

وبمراجعة أرقام الأفراد غير الملائمين في الجدول السابق يلاحظ أن حذف المفردات غير الملائمة بيانيا قبل التحليل الإحصائي أدى إلى تحسن ملاءمة عشرة أفراد، في حين ساءت الملاءمة التباعدية فقط لفرد واحد، وظل سبعة أفراد غير ملائمين في الحالتين. وتؤكد نتائج اختبار مكنمار المتضمنة بالجدول التالي دلالة هذا التغير في تصنيف الأفراد إلى ملائمين وغير ملائمين:

جدول (١٢)

اختبار دلالة التغير في تصنيف الأفراد إلى ملائمين وغير ملائمين في حالتي التدريج

		بعد الحذف		المجموع	مستوى دلالة التغير
		غير ملائم	ملائم		
القبلي	ملائم	١	١٨١	١٨٢	٠,٠١٢
	غير ملائم	٧	١٠	١٧	
المجموع		٨	١٩١	١٩٩	

ويعني ذلك أنه عند تدريج الاختبار وفقا لنموذج راش قد يكون من المفيد إجراء التحليل البياني وحذف المفردات غير الملائمة وفقا لمحكات التحليل البياني قبل إجراء التحليل الإحصائي، حيث يؤدي ذلك إلى تقليل عدد الأفراد المحذوفين من عينة التدريج وفقا للمحكات الإحصائية للملاءمة وبالتالي زيادة حجم عينة التدريج النهائي. ويفيد هذا الإجراء عندما يكون الحفاظ على العينة أمرا ضروريا، كما في حالة إعلان نتائج الأفراد بعد التدريج واتخاذ قرارات بشأنهم.

وبذلك يمكن القول بأن : حذف المفردات غير الملائمة وفقا لمحكات التحليل البياني يؤدي إلى انخفاض نسبة الأفراد غير الملائمين وفقا لمؤشرات الملاءمة الإحصائية، وذلك عند تدريج باقي مفردات المقياس تبعا لنموذج راش.

٣- نتائج مقارنة إحصاءات ملاءمة الأفراد في التدريج النهائي في حالتي التدريج، ومناقشتها:

اتبعت الدراسة نفس الإجراءات المناظرة الخاصة بمقارنة ملاءمة المفردات والسابق بيانها في عرض نتائج الفرض الثاني، وذلك كما يلي:

أ- نتائج المقارنة بين إحصاءات ملاءمة عينة التدريج النهائي في حالتي التدريج، ومناقشتها:

يوضح الجدول التالي نتائج هذه المقارنة:

جدول (١٣)

المقارنة بين إحصاءات ملاءمة عينة التدريج النهائي للاختبار في حالتي التدريج

إحصاء الملاءمة		حالة التدريج	عدد الأفراد	المتوسط	الوسيط	الانحراف المعياري
ZSTD	التقاربي	الإحصائي	١٨٢	٠,٠٠	٠,٠٦	٠,٩٠
		البياني/ الإحصائي	١٩١	٠,٠٠	٠,١١	٠,٩٠
	التباعدي	الإحصائي	١٨٢	٠,٠٠	-٠,١٤	٠,٨٠
		البياني/ الإحصائي	١٩١	٠,٠٠	-٠,٠٤	٠,٨٠
MNSQ	التقاربي	الإحصائي	١٨٢	١,٠٠	١,٠٠	٠,٢٠
		البياني/ الإحصائي	١٩١	١,٠٠	١,٠٠	٠,٢١
	التباعدي	الإحصائي	١٨٢	٠,٩٦	٠,٩٦	٠,٣٦
		البياني/ الإحصائي	١٩١	٠,٩٦	٠,٩٦	٠,٣٨

يبين الجدول السابق أن عينة التدريج النهائي للاختبار تحقق درجة مرتفعة من الملاءمة لأسس القياس الموضوعي سواء اعتمد التدريج على نتائج التحليل الإحصائي فقط أو على نتائج كل من التحليل البياني والإحصائي، حيث كان متوسط إحصاءات ملاءمة عينة التدريج مساويا للقيمة المثلى للملاءمة التقاربية والتباعدية سواء لمؤشر ZSTD (صفر) أو لمؤشر MNSQ (واحد صحيح).

وقد كان الفرق بين متوسطي الملاءمة بأنواعها في الحالتين صفريا حيث تتساوى قيمتي متوسط إحصاء الملاءمة لكل مؤشر تماما في حالتي التدريج، ومن ثم لا تكون هناك حاجة لاستخدام اختبار تاء لاختبار دلالة الفروق بينها. وذلك باستثناء الفروق في إحصاء الملاءمة التباعدي لمؤشر ZSTD حيث كان توزيعه ملتويا وكانت قيمتي الوسيط مختلفتين، وباستخدام اختبار مان ويتني تبين عدم وجود فرق دال إحصائيا في ملاءمة الأفراد في الحالتين، حيث بلغ مستوى الدلالة (٠,٦٥٣).

كما تم اختبار دلالة الفرق بين القيم المطلقة لكل من إحصاءي الملاءمة التقاربي والتباعدي وفق مؤشر ZSTD، فأكدت النتائج عدم وجود فروق دالة إحصائيا بين إحصاءات ملاءمة عينة التدريج النهائي للاختبار في حالتي التدريج، حيث بلغ مستوى دلالة الفرق (٠,٥٥) و(٠,٥٦) بالنسبة لإحصاءي الملاءمة التقاربي والتباعدي على الترتيب.

وبالمثل تمت مقارنة انحراف قيم مؤشر MNSQ عن القيمة المثلى (الواحد الصحيح)، وكذا القيم المطلقة لهذا الانحراف، وذلك لكل من إحصاءي الملاءمة التقاربي والتباعدي. وقد أكدت النتائج عدم وجود فروق دالة إحصائيا بين إحصاءات ملاءمة الاختبار في تدريجه النهائي في حالتي التدريج، حيث تراوحت قيم مستوى الدلالة بين (٠,٣٤) و (٠,٩٣).

ب-نتائج المقارنة بين إحصاءات ملاءمة الأفراد المشتركين في عينة التدريج النهائي للاختبار في حالتي التدريج، ومناقشتها:

تم إجراء المقارنة بين القيم المطلقة لإحصاءات ملاءمة الأفراد المشتركين (١٨١ فردا) في عينة التدريج النهائي للاختبار في حالتي التدريج، وذلك باستخدام اختبار ولكوكسن نظرا لالتـواء التوزيع. وكانت النتائج كما بالجدول التالي:

جدول (١٤)

المقارنة بين إحصاءات ملاءمة الأفراد المشتركين في حالتي التدريج النهائي للاختبار

إحصاء الملاءمة		الرتب الموجبة *		الرتب السالبة		ذ	مستوى الدلالة	حجم الفرق	معامل الارتباط
		المجموع	العدد	المجموع	العدد				
ZSTD	التقاربي	٧٤٣٤,٥	٨٥	٨٨٥٥,٥	٩٥	١,٠٢	٠,٣١	-٠,٠٩ (صغير)	٠,٨٨
	التباعدي	٦١٥٠	٧٥	١٠٣٢١	١٠٦	٢,٩٦	٠,٠٠٣	-٠,٢٥ (ضعيف)	٠,٨٩
MNSQ	التباعدي	٨٢٣٠	١٠٧	٧٥٢٣	٧٠	٠,٥٢	٠,٦٠	-٠,٠١ (صغير)	٠,٩٣

(*) إحصاء ملاءمة الفرد في الحالة الثانية للتحليل أكبر من إحصاء ملاءمته في الحالة الأولى

يتضح من الجدول السابق ما يلي:

- انخفضت قيم إحصاءات الملاءمة التباعدية لمؤشر **ZSTD** والخاصة بالأفراد المشتركين في التدريج النهائي انخفاضا دالا إحصائيا في الحالة الثانية للتدريج، مما يعني أن استبعاد المفردات غير الملائمة وفقا لمحكات التحليل البياني يؤدي إلى تحسن إحصاءات الملاءمة التباعدية للأفراد، وإن كان حجم هذا التحسن ضعيفا.
- لم تختلف إحصاءات الملاءمة التقاربية لهم وفق مؤشرZSTD. كما لم تختلف إحصاءات الملاءمة التباعدية لهؤلاء الأفراد في حالتي التدريج وفقا لمؤشر **MNSQ**.
- يوجد ارتباط قوي دال إحصائيا بين قيم إحصاءات الملاءمة التقاربية وكذا التباعدية في حالتي التدريج النهائي، مما يدل على اتساق توزيع الأفراد على تدريج الملاءمة في حالتي التدريج.

أما إحصاء الملاءمة التقاربي لمؤشر **MNSQ** للأفراد المشتركين في التدريج النهائي في حالتي التدريج، فقد أجريت المقارنة باستخدام اختبار تاء. وبينت النتائج عدم اختلاف إحصاءات الملاءمة في حالتي التدريج حيث بلغ مستوى الدلالة (٠,٧٤). كما بلغ معامل الارتباط بينها (٠,٨٨) وكان دالا إحصائيا عند مستوى دلالة أقل من (٠,٠١)، مما يؤكد اتساق توزيع الأفراد

على متصل الملاءمة في حالتي التدريج.

وبذلك يمكن القول بأن : حذف المفردات غير الملائمة وفقا لمحكات التحليل البياتي يؤدي إلى انخفاض نسبة الأفراد غير الملائمين لأسس القياس الموضوعي، وتحسن الملاءمة التباعدية للأفراد وفق مؤشر ZSTD فقط، ولا يؤثر في قيم إحصاءات الملاءمة الأخرى للأفراد الملائمين وتوزيعهم على تدريج الملاءمة، وذلك في التدريج النهائي وفقا لنموذج راش.

٤- نتائج مقارنة تدرج الأفراد على متصل القدرة في حالتي التدريج، ومناقشتها:

نظرا لاختلاف صفر تدريج القدرة في حالتي التدريج فقد استخدم المدى ومعامل الاختلاف ومعامل الارتباط لدراسة أثر حذف المفردات غير الملائمة وفقا لمعايير التحليل البياتي على تدرج الأفراد على متصل القدرة، وقد اقتصرت المقارنة على الأفراد المشتركين، وذلك كما يلي:

جدول (١٥)

المقارنة بين تدرج الأفراد المشتركين في حالتي التدريج النهائي (ن= ١٨١ فردا)

حالة التدريج	تقديرات القدرة بالمئف			معامل الاختلاف	معامل الارتباط	مستوى الدلالة
	المتوسط	المدى	الانحراف المعياري			
الإحصائي	٥٤,٣٨	٢٣,٥٥	٣,٧٠	٠,٠٧	٠,٩٨	٠,٠٠
البياني/ الإحصائي	٥٥,٢٣	٢٨,٤٠	٣,٩٢	٠,٠٧		

من الجدول السابق يلاحظ ما يلي:

- يتكافأ إحصائيا مدى تقديرات القدرة لمجموعة الأفراد المشتركين (١٨١ فردا) في التدريج النهائي في حالتي التدريج، حيث أن قيمة الاتساع الملاحظ في الحالة الثانية للتدريج لا تتجاوز مجموع متوسطي الخطأ المعياري لتقديرات القدرة في حالتي التدريج.
- لا يختلف تشتت تقديرات القدرة في حالتي التدريج، حيث تتساوي قيم معاملي الاختلاف مما يدل على عدم اختلاف تباين تلك التقديرات في الحالتين.
- يوجد ارتباط طردي قوي دال إحصائيا عند مستوى دلالة أقل من (٠,٠١) بين قيم تقديرات القدرة في حالتي التدريج النهائي. مما يعني اتساق توزيع الأفراد على متصل القدرة في الحالتين. ويتسق ذلك مع ما يحققه نموذج راش من استقلالية في القياس.

يتبين مما سبق أن : حذف المفردات غير الملائمة وفقا لمحكات التحليل البياني لا تؤثر على تدرج الأفراد على المتغير موضوع القياس عند تدريج الاختبار وفقا لنموذج راش.

٥- نتائج مقارنة دقة تدرج الأفراد على متصل القدرة في حالتي التدريج، ومناقشتها:

اعتمدت الدراسة في تقدير دقة تدرج الأفراد على كل من معامل ثبات كيودر ريتشاردسون-٢٠، والخطأ المعياري لتقديرات القدرة.

أ- فيما يتعلق بمعامل الثبات فقد سبق الإشارة إلى التحسن النسبي في ثبات تقديرات القدرة بعد حذف المفردات غير الملائمة بيانيا قبل إجراء التحليل الإحصائي، حيث ارتفعت قيمة معامل الثبات في التدريج النهائي من (٠,٥٥) إلى (٠,٥٧) أي إلى (٠,٥٩) بعد تعديله باستخدام معادلة سبيرمان براون (Izard, 2005, P57) بافتراض تساوي طولي الاختبار النهائي في حالتي التدريج (٢٩ مفردة).

وقد تمت مقارنة معاملي الثبات باستخدام إحصاء W ، ويتوزع هذا الإحصاء مثل النسبة الفائية (Feldt, 1987, P96). وقد بينت المقارنة أن التحسن في معامل الثبات كان دالا إحصائيا عند مستوى الدلالة (٠,٠٥). ورغم هذا التحسن إلا أن ثبات تقديرات القدرة يعتبر متوسط. ووفقا لدليل البرنامج المستخدم في التحليل الإحصائي (Linacre, 2006, P320) قد يرجع ذلك لانخفاض عدد مفردات الاختبار كما سبق الإشارة، كما قد يرجع لتجانس عينة التدريج أي ضيق مدى تقديرات القدرة الذي لم يتجاوز (٢٨) منف تقريبا على تدريج منوي في حالتي التدريج.

ب-أما فيما يتعلق بالخطأ المعياري فقد تمت المقارنة بين قيم الخطأ المعياري لتقديرات قدرة عينة التدريج النهائي للاختبار، وكذا الأفراد المشتركين في حالتي التدريج النهائي، وذلك كما يلي:

- نتائج المقارنة بين قيم الخطأ المعياري لتقديرات قدرة عينة التدريج النهائي للاختبار في حالتي التدريج:

لإجراء هذه المقارنة تم استخدام اختبار مان ويتني نظرا لالتواء التوزيع، وكانت النتائج كما بالجدول التالي:

جدول (١٦)

المقارنة بين قيم الخطأ المعياري لتقديرات قدرة عينة التدريج النهائي للاختبار في حالتي التدريج

حالة التدريج	ن	التقديرات بالمنف			مجموع الرتب	Z	مستوى الدلالة	حجم التأثير
		المدى	الوسيط	الانحراف الربيعي				
الإحصائي	١٨٢	١,١٠	٢,٣٠	٠,٢٠	٢٧٨٩٤,٥	٥,٩٤	٠,٠٠	١ (كبير)
البياني/ الإحصائي	١٩١	٣,٠٠	٢,٤٥	٠,٣٥	٤١٨٥٦,٥			

يتضح من الجدول السابق ما يلي:

- تعتبر قيم الخطأ المعياري لتقديرات القدرة منخفضة بوجه عام في حالتي التدريج حيث لا يتجاوز متوسطها (٢,٥) منف تقريبا، ويبلغ أقصى مدى (٣) منف. وبالنظر إلى أن تدريج المنف هو تدريج منوي متوسطه يساوي (٥٠) وانحرافه المعياري يساوي (٥) فإن قيم الخطأ المعياري في مجملها تؤكد دقة تقديرات القدرة لأفراد عينة التدريج النهائي للاختبار في حالتي التدريج. وقد سبق الإشارة إلى ارتفاع قيم معامل الثبات الهامشي للاختبار الذي يعتمد على الخطأ المعياري.
- ارتفعت قيم الخطأ المعياري لتقديرات قدرة عينة التدريج النهائي للاختبار في الحالة الثانية للتدريج مقارنة بالحالة الأولى للتدريج – أي بعد حذف المفردات غير الملائمة وفقا لمحكات التحليل البياني. وكان هذا الارتفاع كبيرا ودالا إحصائيا عند مستوى دلالة أقل من (٠,٠١).

• نتائج المقارنة بين قيم الخطأ المعياري لتقديرات قدرة الأفراد المشتركين في حالتي التدريج النهائي:

أجريت هذه المقارنة لقيم الخطأ المعياري لتقديرات القدرة للأفراد المشتركين (١٨١ فردا)، وذلك باستخدام اختبار ولكوكسن نظرا لالتواء التوزيع، وكانت النتائج كما بالجدول التالي:

جدول رقم (١٧)

المقارنة بين قيم الخطأ المعياري لتقديرات قدرة الأفراد المشتركين في حالتي التدريج النهائي

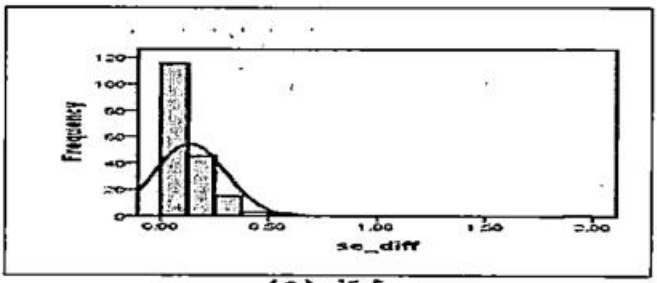
الرتب الموجبة(*)		الرتب السالبة		Z	مستوى الدلالة	حجم التأثير	معامل الارتباط	مستوى الدلالة
العدد	المجموع	العدد	المجموع					
١٦٦	١٣٨٦١	-	صفر	١١,٢٦	٠,٠٠	١,٠٠ (قوي)	٠,٩٢	٠,٠٠

(*) الخطأ المعياري في الحالة الثانية أكبر من الحالة الأولى للتدريج.

يتضح من الجدول السابق ما يلي:

• تؤكد النتائج ارتفاع قيم الخطأ المعياري لتقديرات القدرة في الحالة الثانية للتدريج.

وبمراجعة الفروق بين قيم الخطأ المعياري لتقديرات قدرة الأفراد في حالتي التدريج، وجد أن هذه الفروق كانت صغيرة، حيث تراوحت بين (صفر) و (٠,٥٥) منف، فيما عدا حالة واحدة بلغ الفرق بين الخطأ المعياري لتقدير قدرتها (١,٩٥) منف، وبلغ وسيط هذه الفروق (٠,١) منف وكان توزيعها ملتويا التواء موجبا حيث كانت أغلب الفروق منخفضة جدا كما يتضح من الشكل التالي:



شكل (٥)

توزيع الفروق بين الخطأ المعياري لتقديرات القدرة في حالتي التدريج

وبالرغم من أن الفروق صغيرة إلا أنها كانت إما صفرية أو لصالح التحليل الإحصائي فقط، فقد ارتفعت قيم الخطأ المعياري لتقديرات قدرة (١٦٦) فردا، ولم تتغير قيم الخطأ المعياري لتقديرات قدرة (١٥) فردا، مما أدى إلى ظهور الفروق الدالة إحصائيا، حيث أثر اتجاه الفرق بغض النظر عن قيمته في نتائج اختبار ولكوكسن وحجم التأثير الذي يعتمد على رتب الفروق.

وقد يرجع السبب في ارتفاع قيم الخطأ المعياري لتقديرات قدرة الأفراد (سواء عينة التدريج أو الأفراد المشتركين فقط) في التدريج النهائي للاختبار إلى انخفاض عدد المفردات التي أجاب عنها كل فرد بعد حذف المفردات غير الملائمة بيانيا. ويلاحظ أنه لنفس هذا السبب ارتفعت القيمة الملاحظة لمتوسط الخطأ المعياري لتقدير القدرة في التدريج النهائي مقارنة بالتدريج الأولي وذلك في كل من حالتي التدريج(جدول (١٠)).

- هناك ارتباط قوي ودال إحصائيا عند مستوى دلالة أقل من (٠,٠١) بين القيم المتناظرة للخطأ المعياري لتقديرات قدرة الأفراد المشتركين في حالتي التدريج. مما يؤكد النتائج السابقة بعدم اختلاف ترتيب تدرج الأفراد على متصل متغير التحصيل الدراسي لمقرر نظريات التعلم في حالتي التدريج.

وبذلك يتضح أن إجراء التحليل البياني وحذف المفردات غير الملائمة وفقا لمحكات التحليل البياني قبل إجراء التحليل الإحصائي يؤدي إلى ارتفاع قيمة معامل ثبات تقديرات قدرة الأفراد كما يقدر بمعامل ثبات كيودر- ريتشاردسون ومن ثم ارتفاع دقة تقدير القدرة، إلا أن نفس الإجراء يؤدي إلى انخفاض عدد مفردات الاختبار مما يؤدي إلى ارتفاع قيم الخطأ المعياري لتقديرات القدرة، ومن ثم انخفاض دقة تقدير القدرة.

وقد يفسر هذا التناقض بأن معامل ثبات كيودر ريتشاردسون يعتمد على التجانس الداخلي لاستجابات الأفراد على مفردات الاختبار. ومن المرجح أن يزداد هذا التجانس مع حذف المفردات

غير الملائمة بيانيا وما يتعلق بها من استجابات لا تتسق مع الاستجابات على باقي المفردات. أما الخطأ المعياري فهو دالة لعدد مفردات الاختبار ولذا فقد ارتفعت قيمته لغالبية الأفراد نتيجة لانخفاض عدد المفردات. ولنفس هذا السبب فقد لوحظ أن معامل كيودر ريتشاردسون أيضا رغم تحسنه إلا أن قيمته لا تزال متوسطة.

ممّا سبق يتبين أن: حذف المفردات غير الملائمة وفقا لمحكات التحليل البياني يؤدي إلى انخفاض نسبة الأفراد غير الملائمين لأسس القياس الموضوعي وفقا لمحكات الملاءمة الإحصائية، وارتفاع معامل ثبات تقديرات قدرة عينة التدريج النهائي، وارتفاع قيم الخطأ المعياري لهذه التقديرات، وارتفاع الملاءمة التباعدية للأفراد وفق مؤشر ZSTD في حين لا يؤثر على قيم وتوزيع إحصاءات الملاءمة الأخرى للأفراد الملائمين في التدريج النهائي وفقا لنموذج راش، كما لا يؤثر على تدرج الأفراد على المتغير موضوع القياس.

وبذلك تتحقق جزئيا صحة الفرض الثالث من فروض الدراسة.

رابعا: نتائج اختبار الفرض الرابع ومناقشتها:

ينص الفرض الرابع على : تتحسن نتائج التحليل البياني للمفردات بعد حذف الأفراد والمفردات غير الملائمة لأسس القياس الموضوعي وفقا لكل من المحكات البيانية والإحصائية.

ولاختبار هذا الفرض أجري التحليل البياني للمفردات مرتين، هما:

التحليل البياني الأولي باستخدام البيانات الكلية، أي التحليل البياني بناء على استجابات عينة الدراسة (١٩٩ فردا) على الاختبار الكلي (٣٠ مفردة). وبناء على نتائج هذا التحليل تم تحديد المفردات غير الملائمة وفقا لمحكات التحليل البياني كما سبق الإشارة.

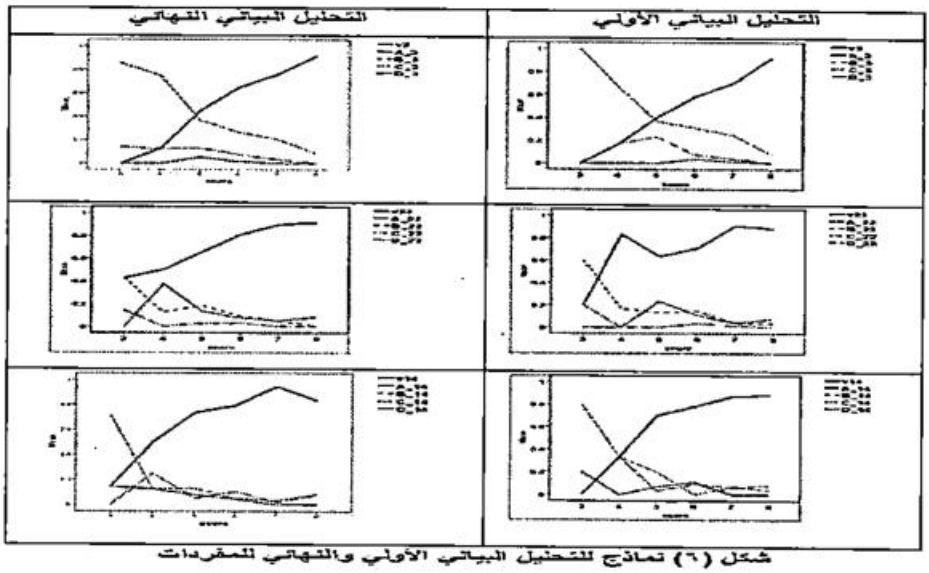
التحليل البياني النهائي: حيث أعيد التحليل البياني بعد حذف المفردات غير الملائمة وفقا لمحكات التحليل البياني، بالإضافة إلى حذف الحالات (الأفراد والمفردات) غير الملائمة لأسس القياس الموضوعي وفق المحكات الإحصائية للملاءمة لنموذج راش. وبذلك تم التحليل البياني باستخدام استجابات الأفراد الملائمين (١٩١) على المفردات الملائمة (٢٧ مفردة).

ثم أجريت المقارنة بين نتائج التحليل البياني الأولي والنهائي للمفردات المشتركة (٢٧ مفردة). وقد تمت المقارنة في ضوء المحكات التي وضعها باتنبرج ولاروس (٢٠٠٢)، حيث تم تصنيف المفردات إلى:

- مفردات لم تتغير معالم التحليل البياني لها، مثل المفردتين رقمي (٥) و(٢٠) السابق عرض التحليل البياني لكل منها في نتائج الفرض الأول (شكل (٢)).
- مفردات تغيرت معالم التحليل البياني لها إلى الأفضل (أكثر ملاءمة)، مثل المفردتين رقمي

(١٣ ، ٢٢) المبينة بشكل (٦).

- مفردات تغيرت معالم التحليل البياني لها إلى الأقل ملاءمة، مثل المفردة رقم (١٤) المبينة بالشكل التالي.



شكل (٦) نماذج للتحليل البياني الأولي والنهائي للمفردات

يلاحظ من الشكل السابق أن المفردات بوجه عام جيدة، ولكن تغيرت بعض ملامح التحليل البياني لها بحيث تصبح خطوط البدائل والمشتتات أكثر انسيابية مثل المفردة رقم (٣) أو أكثر جودة وملاءمة مثل المفردة رقم (٢٢)، أو أقل جودة مثل المفردة رقم (١٤). ولاختبار دلالة هذا التغير تم استخدام اختبار مكنمار، كما يوضح الجدول التالي :

جدول رقم (١٨)

المقارنة بين نتائج التحليل البياني الأولي والنهائي للمفردات المشتركة (ن = ٢٧ مفردة)

التحليل البياني		النهائي		المجموع	اختبار مكنمار	
		أكثر ملاءمة	أقل ملاءمة		الإحصاء	مستوى الدلالة
الأولي	أقل ملاءمة	١٥	-	١٥	٦,٣٧	٠,٠١٩
	أكثر ملاءمة	٨	٤	١٢		
المجموع		٢٣	٤	٢٧		

يتضح من الجدول السابق تحسن ملاءمة (١٥) مفردة في التحليل النهائي وفقا لمحكات التحليل البياني في حين أصبحت أربع مفردات أقل ملاءمة ولم تتغير ملاءمة ثمان مفردات مقارنة بالتحليل البياني الأولي. وقد كان هذا التغير في معالم التحليل البياني للمفردات دالا إحصائيا عند مستوى

دلالة أقل من (٠,٠٥)، مما يفيد تحسن معالم التحليل البياني للمفردات في الصورة النهائية للاختبار وذلك بعد استبعاد المفردات غير الملائمة وفقا للمحكات البيانية والإحصائية وكذا الأفراد غير الملائمين لأسس القياس الموضوعي وفقا للمحكات الإحصائية للملاءمة.

وتتسق هذه النتيجة مع ما ذكره بانتبرج ولاروس (٢٠٠٢) من أن حذف أسوأ المفردات وفقا لمعايير التحليل البياني يمكن أن يؤدي إلى تحسن نتائج التحليل البياني لباقي مفردات الاختبار.

كما أن هذه النتيجة تتفق مع المنطق، حيث أن الفرد غير الملائم إحصائيا والذي يتم تحديده باستخدام إحصاءي الملاءمة التقاربي والتباعدي تختلف الاستجابات الخاصة به عن ما هو متوقع لها – أو عن استجابات غالبية الأفراد المكافئين له في القدرة – إما على المفردات التي تقترب صعوبتها من قدرة هذا الفرد على متصل السمة موضوع القياس فتؤدي إلى تجاوزه لحدود الملاءمة التقاربية، أو على المفردات التي تبتعد صعوبتها عن قدرته على متصل هذه السمة فتؤدي إلى تجاوز حدود الملاءمة التباعدية (Linacre, 2006, P213-214). ومن ثم تسهم استجابات مثل هذا الفرد في الإخلال – بدرجة ما – ببعض معايير جودة التحليل البياني للمفردات، وبالتالي يؤدي استبعاد الأفراد غير الملائمين إحصائيا إلى تحسن نتائج التحليل البياني لهذه المفردات.

وبالمثل فإن المفردة غير الملائمة تختلف الاستجابات عليها عما هو متوقع لها إما من قبل الأفراد الذين تقترب قدرتهم من صعوبة هذه المفردة على متصل السمة موضوع القياس، فتؤدي إلى تجاوز حدود الملاءمة التقاربية و (أو) من قبل الأفراد التي تبتعد قدرتهم عن صعوبتها على متصل هذه السمة فتؤدي إلى تجاوز حدود الملاءمة التباعدية (Linacre, 2006, P210-211). ومن ثم تسهم الاستجابات على مثل هذه المفردة في الإخلال – بدرجة ما – ببعض معايير جودة التحليل البياني لباقي المفردات، وبالتالي يؤدي استبعاد المفردات غير الملائمة وعدم دخولها في حساب الدرجة الكلية للأفراد إلى تحسن نتائج التحليل البياني لباقي المفردات، ولا سيما أن طول الاختبار الحالي ليس كبيرا ولذا أدى حذف المفردات الثلاث غير الملائمة إلى حدوث تأثير ملموس في الدرجة الكلية للفرد.

مما سبق يتبين أن: حذف المفردات غير الملائمة لأسس القياس وفقا لكل من المحكات البيانية والإحصائية وكذا حذف الأفراد غير الملائمين وفقا للمحكات الإحصائية يؤدي إلى تحسن نتائج التحليل البياني لباقي مفردات الاختبار في صورته النهائية.

وبذلك تتحقق صحة الفرض الرابع من فروض الدراسة.

خلاصة وتوصيات ومقترحات:

- أظهرت النتائج وجود تكامل واتساق بين المحكات الإحصائية والبيانية في انتقاء المفردات الملائمة لأسس القياس، ولذا توصي الدراسة باستخدام كل من التحليل البياني والإحصائي عند تدريج الاختبارات والمقاييس وبنوك الأسئلة، حيث يساعد ذلك على المزيد من الاستبصار بخصائص المفردات ويتيح إمكانية تعديلها على أساس علمي دقيق.
- بينت النتائج انخفاضا دالا إحصائيا في نسبة الأفراد غير الملائمين لأسس القياس الموضوعي، وارتفاع الملاءمة التباعدية للأفراد وفق مؤشر ZSTD، بالإضافة إلى ارتفاع معامل ثبات تقدير القدرة، وذلك في حالة البدء بحذف المفردات غير الملائمة بيانيا قبل إجراء التحليل الإحصائي. ولذا توصي الدراسة الباحثين باستخدام التحليل البياني لتنقية الاختبار قبل التدريج عندما يكون من الضروري الحفاظ على عينة التدريج ولاسيما عندما يكون الاختبار طويلا.
- أكدت الدراسة على تحقيق الاختبار في صورته النهائية لاستقلالية القياس، حيث لم يختلف تدرج كل من الأفراد والمفردات على متصل التحصيل الدراسي في حالتي التدريج. ولذا توصي الدراسة باستخدام نماذج الاستجابة للمفردة في تدريج المقاييس والاختبارات وبنوك الأسئلة نظرا لما تحققه من استقلالية وموضوعية في القياس.
- أظهرت النتائج ارتفاع قيم الخطأ المعياري نسبيا في حالة البدء بحذف المفردات غير الملائمة بيانيا قبل التحليل الإحصائي، كما أظهرت أن معامل ثبات كيودر ريتشاردسون لتقدير القدرة يعتبر متوسطا وأقل من معامل ثبات تقديرات الصعوبة، وقد رجحت الباحثة أن ذلك قد يعزى إلى قصر الاختبار أو ضيق مدى القدرة (تجانس عينة التدريج). ولذا توصي الدراسة بمحاولة تضمين عينة التدريج مستويات متباينة من السمة موضوع القياس – كلما أمكن ذلك – مع زيادة طول الاختبار، وذلك عند استخدام كل من التحليل البياني والإحصائي لانتقاء مفردات التدريج النهائي للاختبار وما قد يترتب عليه من زيادة المفردات المرفوضة.
- أسفرت النتائج عن تحسن معالم التحليل البياني للمفردات في الصورة النهائية للاختبار بعد حذف الحالات غير الملائمة من الأفراد والمفردات بناء على كل من المحكات البيانية والإحصائية. ولذا توصي الدراسة بإعادة إجراء التحليل البياني للمفردات بعد حذف أي أفراد أو مفردات من مصفوفة البيانات، وذلك قبل استخدامه لغرض ما مثل تفسير نتائج

التحليل الإحصائي، أو استكمال توصيف المفردات النهائية.

- وفي ضوء النتائج الحالية تقترح الدراسة إجراء بحوث أخرى للإجابة على الأسئلة التالية:

- هل يختلف الاتساق بين التحليل البياني والإحصائي في تقرير جودة المفردات باختلاف المحكات الإحصائية للملاءمة؟
- هل يختلف أثر استبعاد المفردات غير الملائمة بيانيا قبل التحليل الإحصائي على خصائص الأفراد والمفردات في التدريج النهائي تبعا لنموذج راش باختلاف كل من: حجم العينة وطول الاختبار- درجة ملاءمة الاختبار للنموذج - النموذج المستخدم لتدريج الاختبار؟
- هل يؤثر طول الاختبار على الخطأ المعياري لتقدير قدرة الأفراد في التدريج النهائي بعد استبعاد المفردات غير الملائمة وفقا للمحكات البيانية والإحصائية معا؟
- هل يختلف دلالة وحجم التحسن في معالم التحليل البياني النهائي للاختبار باختلاف طول الاختبار وحجم عينة التدريج؟

المراجع

أحمد عودة (١٩٩٢). مدى التوافق بين نموذج راش والمؤشرات التقليدية فى اختيار فقرات مقياس اتجاه سباعي التدريج . *مجلة كلية التربية*: كلية التربية جامعة الإمارات العربية المتحدة. العدد الثامن، السنة السابعة : ١٥٣-١٧٨.

أحلام الباز ورضا السيد حجازي (٢٠٠٢). تقويم تحصيل تلاميذ مرحلة التعليم الأساسي في العلوم باستخدام نظرية السمات الكامنة. *الجمعية المصرية للتربية العلمية*. المؤتمر العلمي السادس. التربية العلمية وثقافة المجتمع. يوليو. المجلد الثاني. ٦١١-٦٤٨.

ابتسام توفيق أبو خليفة (٢٠٠٩). توظيف نموذج راش في انتقاء فقرات مقياس تقدير لتقييم اتجاهات طلبة كلية العلوم التربوية نحو تخصص معلم الصف. *مجلة العلوم التربوية والنفسية*. ١٠(٤): ١٤٥-١٦٤.

أسماء عبد المنعم عرفان (٢٠١٠). *دراسة سيكومترية حول قياس دافعية الإنجاز باستخدام نظرية الاستجابة المفردة*. رسالة ماجستير غير منشورة. جامعة عين شمس: كلية البنات للآداب والعلوم والتربية.

إكرام حمزة السيد حبشي (٢٠٠٤). *استخدام اختبارات متعادلة القياس في دراسة اثر التغذية المرتدة على النواتج المعرفية للتعلم في مادة الكيمياء لدى طلاب المرحلة الثانوية العامة*. رسالة

ماجستير غير منشورة. جامعة عين شمس: كلية البنات للآداب والعلوم والتربية.

اعتدال غازي عبد الله (٢٠٠٣). *استخدام نموذج راش في تدريج مقياس للقدرة العقلية لدراسة بعض العوامل المؤثرة على دقة القياس*. رسالة دكتوراه غير منشورة. جامعة عين شمس: كلية البنات للآداب والعلوم والتربية.

أمين محمد صبري نور الدين (١٩٩٥). *بعض الخصائص السيكومترية لمقياس ستانفورد-بينيه المعدل لدى عينة من الأطفال في مرحلة ما قبل المدرسة*. رسالة ماجستير غير منشورة. جامعة عين شمس: كلية التربية.

أمينة محمد كاظم (١٩٩٤). في محمد عماد الدين إسماعيل وناهد رمزي وليلي كرم وأمينة كاظم وهدي ناشف، *معايير نمو طفل ما قبل المدرسة*، المجلد الثاني، الدراسة النفسية. القاهرة: رئاسة مجلس الوزراء، المجلس القومي للطفولة والأمومة.

أمينة محمد كاظم (١٩٩٦أ). دراسة نظرية نقدية حول القياس الموضوعي للسلوك- نموذج راش. في: أنور محمد الشرقاوي، سليمان الخضري الشيخ، أمينة محمد كاظم، نادية محمد عبد السلام، *اتجاهات معاصرة في القياس والتقويم النفسي والتربوي*. القاهرة: مكتبة الأنجلو المصرية،٢٨١-٤٣٠.

أمينة محمد كاظم (١٩٩٦ب). مستوى العينة وتدرج بنك الأسئلة باستخدام نموذج (راش) (دراسة تجريبية) ، فى : أنور الشرقاوى ، وسليمان الخضرى ، وأمينة كاظم ، ونادية عبد السلام. *اتجاهات معاصرة في القياس والتقويم النفسى والتربوى* . القاهرة: مكتبة الأنجلو المصرية، ٥٤٧-٥٨٣ .

أمينة محمد كاظم (٢٠٠٠). اتجاهات معاصرة فى بناء بنوك الأسئلة ، في: الأسس التربوية لإعداد المعلم الجامعى . القاهرة: جامعة عين شمس.(ط٣)، ٣٢١-٣٤٢..

أمينة محمد كاظم، شادية عبد العزيز منتصر، سوسن عبد الرحمن (٢٠٠٦). التقرير العلمي عن بناء بنوك الأسئلة وتحقيق الموضوعية والعدالة، وجدوى ميكنة الامتحانات. مشروع تطوير كليات التربية: مركز بنوك الأسئلة للتقويم الموضوعي بكلية البنات للآداب والعلوم والتربية جامعة عين شمس.

أمينة محمد كاظم، شادية عبد العزيز منتصر، رباب عبد المنعم سيف (٢٠١٥). معايير الورقة الامتحانية. ورقة تدريبية غير منشورة. جامعة عين شمس: كلية البنات للآداب

والعلوم والتربية.

باسمة فهد عبد الرحمن الشايجي (٢٠٠٨). *تقويم برنامج لتنمية الاستيعاب والسرعة في القراءة وعلاقته بأنماط الشخصية وفق دليل مايرز وبريجز*. دراسة دكتوراه غير منشورة. جامعة عين شمس: كلية البنات للآداب والعلوم والتربية.

جابر عبد الحميد جابر (١٩٨٣). التقويم التربوي والقياس النفسي. القاهرة: دار النهضة العربية.

حصة عبد الرحمن فخرو، أنور رياض عبد الرحيم، محمد إبراهيم محمد (٢٠٠٩). فاعلية استخدام سلم التقدير الجزئي في تحليل مفردات مقياس مداخل الدراسة لدى طلبة جامعتي قطر والمنيا. *مجلة البحث في التربية وعلم النفس*-كلية التربية جامعة المنيا. ٢٢ (٢)، ١٣٥-١٧٥.

ديانا فهمي علي حماد (٢٠١٠). علاقة الحكمة الاختبارية بالأداء الناتج من اختبار تحصيلي ذي اختيار من متعدد مبني وفق نموذج راش لدى طالبات كلية التربية للبنات الأقسام الأدبية بجامعة أم القرى. رسالة دكتوراه غير منشورة. جامعة أم القرى: كلية التربية.

رانيا ماهر وهدي (٢٠٠٥). دراسة سيكومترية مقارنة بين نموذج راش والقياس التقليدي حول دقة التنبؤ بحالة القلق من سمة القلق. رسالة ماجستير غير منشورة. جامعة عين شمس: كلية البنات للآداب والعلوم والتربية.

رجاء محمود أبو علام (٢٠٠٥). تقويم التعلم. عمان: دار المسيرة.

رينولدز، س. ر. وليفنجستون، ر. ب.(٢٠١٣). إتقان القياس النفسي الحديث، النظريات والطرق. ترجمة صلاح الدين محمود علام. عمان: دار الفكر.

سامي محمد ملحم (٢٠١٥). القياس والتقويم في التربية وعلم النفس. ط ٧. عمان: دار المسيرة للنشر والتوزيع والطباعة.

سيد محمد خير الله، أمينة محمد كاظم ، محمد رجب فضل الله، مجدي محمد أمين،أحلام الباز حسن، هشام حبيب الحسيني، سعد محمد حنفي،... ميرفت العيدروس أبو العينين(٢٠١٥) . بطارية المقاييس القومية للذكاء، المقياس الأول : مقياس الذكاء المعرفي للمرحلة العمرية (١١-١٨) سنة . المركز القومي للامتحانات والتقويم

التربوي.

شادية عبد العزيز منتصر (١٩٩٧). *استخدام نموذج راش في بناء بنك للأسئلة لمقرر علم النفس التعليمي وتحديد الدرجات الفاصلة المقابلة للتقديرات الجامعية*. رسالة ماجستير. غير منشورة. جامعة عين شمس: كلية البنات للآداب والعلوم والتربية.

شادية عبد العزيز منتصر (٢٠٠٢). *التصور البصري المجسم لدي عينة من طالبات الجامعة: قياسه وتنميته*. رسالة دكتوراه. غير منشورة. جامعة عين شمس: كلية البنات للآداب والعلوم والتربية.

صلاح الدين محمود علام (٢٠٠٥). *نظرية الاستجابة للمفردة أحادية البعد ومتعددة الأبعاد وتطبيقاتها في القياس النفسي والتربوي*. القاهرة: دار الفكر العربى.

صلاح الدين محمود علام (٢٠٠٧). التقويم التربوى البديل، أسسه النظرية والمنهجية وتطبيقاته الميدانية. القاهرة: دار الفكر العربى.

صلاح الدين محمود علام (٢٠١٥). القياس والتقويم التربوي والنفسي، أساسياته وتطبيقاته وتوجهاته المعاصرة(ط٦). القاهرة: دار الفكر العربى.

علي محمد زكري (٢٠١١). تطوير وتدريج اختبار أوتيس-لينون للقدرة العقلية المستوى المتوسط الصورة (ل) باستخدام نموذج راش. *مجلة كلية التربية جامعة المنصورة*. ٢ (٧٥)، ٦٤-١١٩.

كروكر، ل. والجينا، ج. (٢٠٠٩). مدخل إلى نظرية القياس التقليدية والمعاصرة. ترجمة زينات يوسف دعنا. عمان: دار الفكر.

ماهر اسماعيل صبري ومحب محمود الرافعي(٢٠٠٣). التقويم التربوي، أسسه وإجراءاته(ط٣). الرياض: مكتبة الرشد.

محمد منصور الشافعي (١٩٩٦). *أثر طرق معادلة درجات الاختبار وضوابط اختيار العينة علي تدريج بنود بنك الأسئلة باستخدام نموذج راش*. رسالة دكتوراه غير منشورة. جامعة المنصورة: كلية التربية.

منى ربيع الطنطاوى (٢٠٠٠) . *دراسة سيكومترية لتطوير اختبار المصفوفات المتتابعة لرافن باستخدام نموذج (راش)* . رسالة ماجستير غير منشورة . جامعة عين شمس : كلية

البنات للآداب والعلوم والتربية.

منى ربيع الطنطاوى متيرد (٢٠٠٨). *استخدام نظرية الاستجابة للمفردة في تقويم فاعلية برنامج لتنمية الاستدلال الاستقرائي لدى عينة من تلاميذ المرحلة الابتدائية*. رسالة دكتوراه غير منشورة. جامعة عين شمس: كلية البنات للآداب والعلوم والتربية.

مصطفى أحمد قاسم (٢٠١٥). قياس الاستدلال بأسلوب الحوسبة باستخدام نظرية الاستجابة للمفردة، دراسة سيكومترية. رسالة ماجستير غير منشورة. جامعة عين شمس: كلية البنات للآداب والعلوم والتربية.

نضال كمال الشريفين (٢٠٠٩). معادلة درجات نماذج مختلفة من اختبار الكفاءة اللغوية في اللغة الانجليزية لدى طلبة جامعة اليرموك. مجلة جامعة أم القرى للعلوم التربوية والنفسية، ١ (٢)، ١١-٦٢.

هالة يوسف مندور، محمد محمد فتح الله، نسرين محمد عاشور (٢٠١٣). بناء وتدريج بنك أسئلة لمقرر الهوكي باستخدام نظرية الاستجابة للمفردة IRT في ضوء معايير الجودة. المجلة المصرية للدراسات النفسية. ٢٣ (٨٠). ٢٤٧-٣٩٥.

وليد أحمد مسعود (٢٠١٢). *دراسة اثر القياس الديناميي علي فاعلية برنامج لتنمية الفهم القرائي باستخدام نظرية الاستجابة للمفردة*. رسالة دكتوراه غير منشورة. جامعة عين شمس: كلية البنات للآداب والعلوم والتربية.

Bahakta, B. , Tennat, A. ; Horton, M. Lawton, G. & Andrich, D. (2005). Using Item Response Theory to Explore the Psychometric Properties of Extended Matching Questions Examination in Undergraduate Medical Education. BMC Education:5-9. Retrieved on December , 2016 from www.biomed central.com/1472-6920/59.

Batenburg, T. Laros, J.(2002). Graphical analysis of test items. *Educational Research and Evaluation. 8*(2), 319-333.

Carrasco,C. , MacCan, C. , Geiger, M. , Seybert, J. & Roberts , C. (2015). Development of A Forced-Choice Measure of Typical Performance Emotional Intelligence. *Journal of Psychological Assessment. 30*(1), 83-97.

Cheu, C. & Chen, C. (2006). Turning Output of Item Response Theory Data Analysis Into Graphs With R. National Chung Cheng

University. Retrieved on March , 2016 from www.r-project.org/conferences/user-2006/slides/sheu+shen.pdf.

Damsgård,E. , San, C. , Fors,T. , Anke, A. & Rose , C. (2007). The TAMPA Scale of Kinesiophobia: A Rasch Analysis of Its Properties in Subjects With Low Back And More Widespread Pain. J Rehabil Med. 39, 672–678.

Elliott, c. (1983). British Ability Scales, Manual 1; introductory handbook. windsor , England: National Foundation for Educational Research.

Embretson, S.E. & Reise, S. P. (2000). Items Response Theory for Psychologists. London, Lawrence Erllaum Associates.

Feldit, L. ,Woodruff, D. & Salih, F. (1987). Statistical Inference for Coefficient Alpha. *Applied Psychological Measurement. 11*(1), 93-103.

Haladyana, T. (2004). Developing and Validating Multiple-Choice Test Items. 4th ed. New York: Rout Ledge. Retrieved on June, 2016 from https://books.google.com.eg.

Hambleton, R.K. & Swaminthan, H. (1985). Item Response Theory "Principles and Applications". Norwell, Assinippi, U.S.A.: Kluwer Academic Publishers.

Hambleton, R.K. & Swaminthan, H. & Rogers, H. J. (1991). Fundamentals of Item Response Theory. Teler Road, Newbury Park, California, U.S.A.: Sage Publications.

Hayas, C. , Quintana1, J. , Padierna, J. , Bilbao, A. & Munoz, P.(2010). Use of Rasch Methodology to Develop A Short Version of The Health Related Quality of Life for Eating Disorders Questionnaire: A Prospective Study. Health and Quality of Life Outcomes. 8-29. Retrieved on April , 2016 from http://www.hqlo.com/content/8/1/29.

Henning , G. (1984). advantages of latent trait measurement in language testing . Language Testing . (1) : 123 – 33.

Isaac, S. & Michael. W. (1982). A handbook in research and evaluation for Educational and the behavioral sciences. 2nd ed. , Sandiego: Edits Publishers.

Izard, j. (2005). Trial Testing And Item Analysis in Test Construction. In Ross,k. (ed.): Quantitative Research Methods in

Educational Planning. UNESCO International Institute for Educational Planning(IIEP). Retrieved on November , 2016 from www.unesco.org/iiep.

Kirishnan, V. (2013). The Early Child Development Instrument (EDI): An Item Analysis Using Classical Test Theory(CTT) on Alberta's Data. ECMAP: Early Child Development Mapping Project Alperta. Retrieved on December , 2016 from www.cup.ualberta.ca/wp-context/uploods/2013/04/itemanalysisCTTcup website_10april2013.pdf.

Ledesma, R. & Molina, J. (2009). Classical Item and Test Analysis with Graphics: The Vista-CITA Program. *Behavioral Research Methods. 41*(4), 1161-1168.

Linacre, J.M. (2006) *A User's Guide to WINSTEPS & MINISTEPS: Rasch-Model Computer Programs.* Chicago: MESA.

Livingston, S. & Dorans, N. (2004). A Graphical Approach To Item Analysis. ETS Research Report. Retrieved on January, 2017 from www.ets.org/research/policy_research_reports.

Loyd, B. (1988). Implications of Item Response Theory for the Measurement Practitioner, *Applied Measurement in Education*, 1(2),135-43.

Ludlow, L. (1986). Graphical Analysis of Item Response Theory Residuals. *Applied psychological Meäsurement. 10*(3), 217-229.

Muís, K. , Winne, P. , Ordene V. & Edwards, O.(2009). Modern Psychometrics for Assessing Achievement, goal orientation: A Rasch analysis. *British Journal of Educational Psychology* 79, 547-576.

Recckase, D. (2009). Multidimensional Item Response Theory. New York: Springer.

Suen, H. (1990). *Principles of Test Theories*, New Jersey, U.S.A.: Laurence Erlbaum Associates Publishers.

Tay,L. , Meade, a. & Cao, M. (2015). An Overview and Practical Guide to IRT Measurement Equivalence .*Organizational Research Methods. 8*(1):3-46.

Toland, M. (2014). Practical Guide to Conducting an Item Response Theory Analysis. *Journal of Early Adolescence. 34*(1), 120-151.

Wright, B. & Linacre, J. (1985). Microscale Manual. Ver.2 . Mediax Interactive Technologies .

Wright, B. & Masters, G. (1982). Rating Scale Analysis: Rasch Measurement. Chicago: META Press .

Wright, B. & Stone, M.(1979). Best Test Design: A Hand Book for Rasch Measurement . Chicago: META Press .

Yurdgul, H. & Batenburg, T. (2006). Item Difficulty From Graphical Item Analysis. Eurasian *Journal of Educational Research*. 24, 209-218.

Integration Between the Statistical and Graphical Analysis in Selecting and Calibrating Multiple Choice Items using Item Response Theory

Dr. Shadia Abd-Elaziz Mohtadi Montasir

Psychology department, Women college, Ain Shams university.

Abstract:

This study aimed at examining the consistency between graphical item analysis (GIA) criteria and Rasch model fitness statistics in evaluating the test item quality, and identifying the effect of using both graphical and statistical criteria in selecting the fit items upon the final scale characteristics and the final GIA outputs. A (30) MCQ test was administered on (199) college female students as a part of the final learning theories exam. Data was analyzed using Winsteps, syntax, and SPSS computer programs. Results: The GIA criteria was consistent with and complement the fitness statistics in evaluating the item quality. Excluding the graphical misfit items before calibrating the test according Rasch model had no effect on the item statistics in the final scale, however it reduced the ratio of misfit persons and persons ZSTD outfit values, and raised the reliability coefficient and the ability estimates standard error values. Excluding the graphical and statistical misfit items and the statistical misfit persons enhanced the final GIA outputs.