أثر اختلاف طريقة المعادلة في معادلة صورتي اختبار تحصيلي وفق نظرية الاستجابة للمفردة

إعداد

أ/ أشرف سعد محمد جادالله

أ.د/ محمد سعد محمد علي

أ.د/ إبراهيم سيد أحمد عبد الواحد

قسم علم النفس التعليمي والإحصاء التربوي، كلية التربية

بنين بالقاهرة، جامعة الأزهر، مصر

أثر اختلاف طريقة المعادلة في معادلة صورتي اختبار تحصيلي وفق نظرية الاستجابة للمفردة

[1]أشرف سعد محمد جادالله، محمد سعد محمد علي، إبراهيم سيد أحمد عبد الواحد
قسم علم النفس التعليمي والإحصاء التربوي، كلية التربية بنين بالقاهرة، جامعة الأزهر، مصر
[1]البريد الالكتروني للباحث: ashrafsaad1662020@gmail.com

المستخلص

هدف البحث الحالي إلى التعرف على أثر اختلاف طريقة المعادلة (المتوسط/ المتوسط، المتوسط/ الانحراف المعياري، هيبارا، ستوكينج/ لورد) في معادلة صورتي اختبار تحصيلي وفق نظرية الاستجابة للمفردة، وتمثلت أداتا البحث في صورتي اختبار تحصيلي (X، Y) في مادة العلوم (إعداد الباحث)، وقد بلغ عدد المشاركين في البحث (1032) تلميذاً وتلميذة من بين تلاميذ الصف الثاني الإعدادي بمنطقة القاهرة الأزهرية تم تقسيمهم إلى مجموعتين غير متكافئتين، وقد اعتمد البحث على تصميم المجموعات غير المتكافئة ذات الجذع المشترك في جمع البيانات، وباستخدام البرامج الإحصائية المتمثلة في (SPSS- jMetriK- BILOG-MG- IRT EQ.)، والأساليب الإحصائية المتمثلة في تحليل التباين الأحادي للقياسات المتكررة، واختبار "ت" لعينتين مرتبطتين؛ توصل البحث للنتائج التالية: وجود فروق ذات دلالة إحصائية عند مستوى دلالة (0,01) بين متوسطات القيم الناتجة من عملية المعادلة وفقاً لطرق المعادلة التابعة لنظرية الاستجابة للمفردة حيث أعطت طريقة المتوسط/ الانحراف المعياري أعلى قيم مقارنة بالطرق الأخرى، كما أظهرت نتائج البحث وجود فروق ذات دلالة إحصائية عند مستوى دلالة (0,01) بين متوسطات قيم دقة معادلة صورتي الاختبار وفقاً لطرق المعادلة التابعة لنظرية الاستجابة للمفردة حيث جاءت طريقة المتوسط/ الانحراف المعياري أكثر دقة من الطرق الأخرى.

الكلمات المفتاحية: طريقة المعادلة، معادلة صورتي الاختبار، نظرية الاستجابة للمفردة.

The Effect of the Difference in the Equation Method on Equating Two Forms of Achievement Test According to Item Response Theory

Ashraf Saad Mohammed Gad Allah[1], Mohamed Saad Mohammed Ali, Ibrahim Sayed Ahmed Abdel-Wahed
Department of Educational Psychology and Educational Statistics, Faculty of Education for Boys in Cairo, Al-Azhar University, Egypt
[1]**Corresponding author E-mail**: ashrafsaad1662020@gmail.com

ABSTRACT

The current research aimed at identifying the effect of the difference in the equation method (Mean/Mean, Mean/Standard, Haebara, Stocking/Lord) on equating two forms of achievement test according to the item response theory. The instruments of the research were two forms of achievement test (X&Y) in Science (Prepared by the researcher). The participants were (1032) pupils enrolled in the second year of preparatory stage, Al-Azhar Educational District in Cairo, and they were divided into two non-equivalent groups. To collect data, the research utilized the non-equivalent groups design with anchor test. Using the statistical programs (SPSS- jMetriK- BILOG-MG- IRT EQ.), the statistical techniques (Repeated measures ANOVA, and Paired Samples t-Test). The results revealed that there were statistically significant differences at (0.01) level between the means of the values resulting from the equation method according to the item response theory. The mean /sigma method gave the highest values compared to the other methods. Frthermore, there were statistically significant differences at (0.01) level between the means of the equation accuracy values for the two test forms according to the item response theory. The mean /sigma method is more accurate than the other methods.

Keywords: Equation Method, Two Test Forms Equation, Item Response Theory.

مقدمة:

يواجه العاملون في مجال التربية والتعليم بصفة عامة، والمهتمون بمجال القياس والتقويم بصفة خاصة العديد من الصعوبات المرتبطة بتقويم وقياس تحصيل التلاميذ في المراحل الدراسية المختلفة، والتي على رأسها دقة وكفاءة الاختبارات التحصيلية وتحليل وتفسير نتائجها، والتي تستخدم في الحكم على نجاح التلاميذ ومدى اكتسابهم للمعلومات والمهارات الخاصة بمحتوى دراسي معين، واتخاذ القرارات التعليمية بانتقالهم إلى الصفوف الدراسية الأعلى، أو الحصول على شهادة دراسية تفيد باجتيازهم لهذه المرحلة الدراسية.

لذا نادت العديد من المؤسسات التعليمية بضرورة الاهتمام بجودة الاختبارات التحصيلية لما لها من دور مهم في الحكم على مدى تحقق الأهداف التعليمية المرجوة من تقديم مادة دراسية معينة، والبحث عن أفضل الأساليب التي يمكن استخدامها في إعداد هذه الاختبارات وتحليل مفرداتها والتحقق من صدقها وثباتها.

وظل العاملون في مجال القياس والتقويم التربوي يستخدمون مبادئ وأسس النظرية الكلاسيكية Classical Theory في بناء الاختبارات التحصيلية وتفسير درجاتها فترة طويلة من الوقت، على الرغم من تعدد أوجه القصور التي تعاني منها هذه النظرية منها اختلاف خصائص مفردات الاختبار باختلاف عينات المفحوصين، واختلاف قدرات المفحوصين باختلاف خصائص مفردات الاختبار، كما أن تصميم الاختبارات في ضوء افتراضات النظرية الكلاسيكية لا تزودنا بتقدير دقيق للمفحوصين سواء لذوي القدرة المنخفضة أو القدرة العالية وذلك لكونها تأخذ في اعتبارها عند اختيار مفردات الاختبار أن تتلاءم مع الأفراد متوسطي القدرة.(يوسف المحروق، 2011، 2)

ونتيجة لأوجه القصور التي تعاني منها النظرية الكلاسيكية؛ فقد أدت جهود علماء القياس والتقويم إلى تطوير نظرية معاصرة في القياس النفسي والتربوي عرفت باسم نظرية الاستجابة للمفردة (IRT) Item Response Theory، والتي من خلالها تم التغلب على العديد من أوجه القصور التي كانت تعاني منها النظرية الكلاسيكية، والميزة الأكثر أهمية في هذه النظرية هي استقلال إحصائيات المفردة الاختبارية (الصعوبة، التمييز، التخمين) عن قدرة المفحوصين المستخدمين لإيجاد هذه الإحصائيات، وكذلك استقلال تقدير قدرات المفحوصين عن إحصائيات المفردات المستخدمة في عملية التقدير.(عيسى الحربي، 2009، 3)

وتقوم نظرية الاستجابة للمفردة على مجموعة من الافتراضات التي يجب تحققها في البيانات لكي تؤدي إلى نتائج يمكن الوثوق بها، منها افتراض أحادية البعد والذي يعني أن مفردات الاختبار متجانسة وتقيس قدرة أو سمة واحدة كامنة تفسر أداء المفحوص على الاختبار، والافتراض الثاني هو الاستقلال الموضعي ويقصد به أن احتمال الإجابة الصحيحة للفرد على المفردة الاختبارية يكون مستقلاً عن ناتج إجابته عن أي مفردة أخرى في الاختبار، والافتراض الثالث هو منحنى خاصية المفردة وهو عبارة عن مخطط بياني يربط بين احتمالية الإجابة الصحيحة عن المفردة الاختبارية وقدرة المفحوص، أما الافتراض الرابع فهو التحرر من السرعة أي أن عامل السرعة ليس له تأثير في الاستجابة على مفردات الاختبار. Reise& Waller, (2003, 171 ؛ صلاح الدين علام، 2005، 60-64)

وقد ظهرت في العقدين الأخيرين العديد من التطبيقات العملية لنظرية الاستجابة للمفردة في مجال القياس والتقويم النفسي والتربوي، ومن هذه التطبيقات: بناء الاختبارات التحصيلية وتقنينها، والكشف عن مفردات الاختبار المتحيزة، وبناء بنوك الأسئلة، والاختبارات المواءمة المحوسبة، وتدريج مفردات الاختبار، ومعادلة درجات صور الاختبارات. (Shyu, 2001, 9؛ Embretson& Reise, 2000, 249)

ويُعد مجال معادلة الاختبارات أحد المجالات التطبيقية المهمة في القياس النفسي والتربوي، فالبرامج الاختبارية واسعة النطاق تتطلب توافر صيغ متعددة من الاختبارات المختلفة لضمان سريتها، كما أن الصيغ المتعددة ضرورية في اختبارات القبول كالتي تستخدم في انتقاء الطلاب للالتحاق بكليات معينة، وفي هذه الحالات ينبغي إجراء معادلة للصيغ المختلفة لتحويل الدرجات الخام التي نحصل عليها من صيغتي اختبار معين إلى الصيغة الأخرى بحيث تصبح الدرجات المحولة لكل منهما متعادلة. (صلاح الدين علام، 2005، 213)

فمعظم برامج التقييم واسعة الانتشار تستخدم أكثر من صورة للاختبار الواحد، وبالتالي فإن المعادلة الناجحة أحد الاجراءات المهمة المساعدة في إعطاء تقييم صادق للمقارنة بين الأفراد المختبرين، فهي عملية فنية يتم إجراؤها للحصول على درجات قابلة للمقارنة ذات معنى مكافيء لصور مختلفة من نفس الاختبار مما يسمح باستخدام هذه الدرجات بالتبادل. (Ryan& Brockmann, 2009, 8)

ويشير Dorans, Moses& Eignor (2010,3) إلى أن الهدف من معادلة الاختبار Test Equating هو إجراء ربط بين درجات صورتي اختبار بحيث يمكن استخدام درجات أي صورة من الصورتين كما لو أنها جاءت من نفس الاختبار.

وتوجد عدة طرق لإجراء عملية معادلة الاختبار، منها ما يعتمد على النظرية الكلاسيكية كطريقة المتوسط الحسابي، وطريقة المعادلة الخطية، وطريقة المعادلة المئينية، ومنها ما يعتمد على نظرية الاستجابة للمفردة كطريقة المتوسط/ المتوسط، وطريقة المتوسط/ الانحراف المعياري، وطريقة هيبارا، وطريقة ستوكينج ولورد.(راشد الدوسري، 2001، 118-124؛ Uysal& Kilmen, 2016, 3)

كما توجد عدد من التصميمات التي تستخدم لجمع البيانات اللازمة لإجراء عملية المعادلة من خلال الحصول على الدرجات الخام للأفراد المفحوصين على صورتي الاختبار، فإما أن يُستخدم تصميم المجموعة الواحدة، أو تصميم المجموعات العشوائية، أو تصميم المجموعات المتكافئة ذات الجذع المشترك، أو تصميم المجموعات غير المتكافئة ذات الجذع المشترك.(Gruijter& Kamp, 2008, 202-203)

ولإجراء معادلة الاختبارات؛ فإن هناك مجموعة من الخطوات يجب اتباعها لتنفيذ عملية المعادلة وهي: تحديد الغرض من عملية المعادلة، بناء صور بديلة للاختبار متشابهة في المحتوى والخصائص الإحصائية، اختيار تصميم جمع البيانات المناسب، تنفيذ تصميم جمع البيانات الذي تم تحديده، اختيار واحدة أو أكثر من التعريفات الإجرائية لعملية المعادلة، اختيار واحدة أو أكثر من طرق التقدير الإحصائية المناسبة، تقييم عملية المعادلة للحكم على دقة نتائجها. (Kolen& Brennan, 2014, 7)

ويتم الحكم على دقة المعادلة Equating Accuracy باستخدام مجموعة من المؤشرات أو المعايير للتأكد من كفاءة طريقة المعادلة المستخدمة من أجل الوصول إلى أعلى مستوى دقة لعملية المعادلة، ويستدل عليها من خلال الحصول على أقل قدر ممكن من مقدار الخطأ في معادلة درجات صورتي الاختبار، ومن أهم المعايير المستخدمة للحكم على مستوى دقة عملية المعادلة: معيار الخطأ المعياري للمعادلة، ومعيار الصدق التقاطعي، ومعيار جذر متوسط مربعات الخطأ.(Lee, 2013, 4 ؛ أحمد أبو الحسن، 2014، 456)

مشكلة البحث:

تبرز الحاجة إلى إعداد أكثر من صورة للاختبار الواحد نتيجة إلى أن تكرار تطبيق نفس الاختبار أكثر من مرة سيؤدي في أغلب الأحيان إلى الكشف عن محتوى الاختبار مما يقلل من سرية الاختبارات التي تنادي بها العديد من مؤسسات التعليم اليوم، كما أن هناك بعض البرامج التعليمية التي تتطلب تطبيق أكثر من صورة للاختبار الواحد على مجموعات مختلفة من التلاميذ، بالإضافة إلى انتشار الوسائل التكنولوجية الحديثة التي أصبحت أحد العوامل المساعدة في الكشف عن سرية الاختبارات وسرعة انتشار محتواها بين التلاميذ، ونتيجة لذلك تصبح عملية معادلة صور الاختبار ضرورة عملية تسمح بإجراء المقارنات بين تلميذ تعرض لإحدى صور الاختبار وآخر تعرض لصورة أخرى منه.

وعلى ذلك فإن عملية معادلة درجات صور الاختبار تعتبر متطلباً مهماً لتحقيق العدل والمساواة بين الأفراد المختبرين، فالدرجات المتعادلة (Equated score) يمكن أن تتم مقارنتها بطريقة مباشرة دون أن تكون هناك اختلافات في الدرجة ترجع إلى الأفراد المختبرين الذين تلقوا صورة اختبار سهل أو اختبار أكثر صعوبة. (Langer& swanson, 2010, 510 ؛ Song,2009, 85)

فعندما يتم استخدام الاختبار وتطبيقه بشكل متكرر؛ فإن مفردات الاختبار يمكن أن تصبح سهلة ومعروفة للمفحوصين الذين سوف تُطبق عليهم مستقبلاً، لذلك يتم إعداد نماذج متعددة من الاختبارات في العديد من البرامج التعليمية لمنع كشف محتوى الاختبار، إلا أن أغلب هذه النماذج لا تكون متكافئة في محتواها ومستوى صعوبتها في أغلب الأحيان، ولهذا السبب فإن بعض المفحوصين الذين يأخذون الاختبار السهل سيكون لهم أفضلية على أولئك الذين يأخذون الاختبار الصعب، وعلى ذلك فإنه إذا تمت معادلة درجات نماذج هذه الاختبارات بطريقة علمية صحيحة؛ فإن المفحوصين سوف يحصلون على الدرجات نفسها بغض النظر عن أي اختبار يتم التعرض له.(يوسف المحروق، 2016، 210)

وقد تزايد الاهتمام في الآونة الأخيرة بموضوع معادلة الاختبارات من قبل قطاع كبير من المتخصصين المسئولين عن بناء الاختبارات في الكثير من المؤسسات التربوية: مثل المجلس القومي للقياس في التربية، والرابطة الأمريكية للبحث التربوي، وجمعية علم النفس الأمريكية، نتيجة زيادة عدد برامج الاختبارت وتنوعها التي تستخدم صوراً متعددة من الاختبار الواحد، وإدراك المتخصصين في بناء الاختبارت لأهمية معادلة درجات صور الاختبارات، بالإضافة إلى أن مطوري الاختبارات قد أكدوا على أهمية معادلة الاختبارات لمعالجة الكثير من القضايا الفنية التي يثيرها النقاد في هذه المجال، كما أن قضية العدل في قياس السمة لدى مجموعات متباينة من التلاميذ قد أعطت لعملية معادلة الاختبارات أهمية كبيرة في أوساط المتخصصين في

القياس النفسي والتربوي بشكل عام، ومستخدمي نتائج الاختبارات بشكل خاص. &Kolen (Brennan, 2014, vii)

كما أوصت العديد من الدراسات بضرورة إجراء المزيد من البحوث والدراسات حول العوامل التي قد تؤثر على معادلة صور الاختبار: مثل دراسة كل من (Fitzpatrick& Yen,2001؛ عيسى الحربي، 2009؛ يوسف المحروق، 2011؛ 2015 ,Kabasakal& Kelecioglu؛ يوسف المحروق، 2016؛ Ozdemir,2017؛ أحمد بديوي، 2019) ، كما أوصت هذه الدراسات باستخدام بيانات حقيقية للوقوف على أوجه الاختلاف بين البيانات الحقيقية والبيانات المولدة في نتائج معادلة درجات صورتي الاختبار.

فهناك حاجة ماسة لمعادلة درجات صور الاختبار من أجل تحقيق نزاهة الدرجات التي يمكن الحصول عليها، وعدم الاضطرار إلى تطبيقه أكثر من مرة، كما أن عملية معادلة صور الاختبار تعتبر وسيلة مهمة لضمان العدل في توزيع الأفراد والقضاء على تأثير الممارسة. (Felan, 2002, 4)

لذا برزت الحاجة إلى مزيد من الدراسات والبحوث للوقوف على أنسب طرق المعادلة وأدقها تحت مختلف الظروف والسياقات: حيث أنه يمكن النظر إلى عملية المعادلة كدالة لتأثير مجموعة من العوامل كطريقة المعادلة، وتصميم جمع البيانات، وطول الاختبار، ونسبة المفردات المشتركة بين صورتي الاختبار، وغيرها من العوامل التي قد تؤثر بشكل مستقل على عملية المعادلة بل

وتتفاعل مع بعضها البعض. (إيهاب السيد،2012، 60)

فإذا كان من غير الممكن عملياً أن نبني صوراً متكافئة بصورة تامة للاختبار الواحد؛ وذلك باعتبار الخصائص الإحصائية للاختبار، بالإضافة إلى أن ظروف التطبيق قد تؤثر على أداء الأفراد على صور الاختبار فتتأثر تبعاً لذلك الخصائص الإحصائية للصور، وبالتالي تتأثر درجاتهم التي يحصلون عليها، فإننا بمعادلة درجات صور الاختبار نتمكن بدرجة مقبولة من التخلص من الفروق بين هذه الصور بما يمكننا من مقارنة أداء الأفراد بغض النظر عن الصورة المقدمة إليهم. (آر إيالا،2009/ 2017، 375)

وقد تناقضت نتائج الدراسات والبحوث السابقة التي تناولت تأثير اختلاف طريقة المعادلة المستخدمة في معادلة درجات صورتي الاختبار: حيث توصلت بعض الدراسات إلى وجود فروق ذات دلالة إحصائية ترجع إلى طريقة المعادلة كما في دراسة كل من (يوسف المحروق، 2011؛ رائد المدانات، 2012؛2012 ,Kilmen& Demirtasli؛ Ozdemir,2017 ؛ ديالا العلي ورمضان درويش، 2017؛ أحمد بديوي، 2019)، بينما توصلت دراسات أخرى إلى عدم وجود فروق ذات دلالة إحصائية ترجع إلى طريقة المعادلة كما في دراسة كل من (رائد المدانات، 2010؛ Pang, Madera, Radwan& Zhang,2010 ؛ خالد العنزي، 2014؛ Asiret & sunbul, 2016؛ Arikan& Gelbal, 2018 ؛ Born, Fink, Spoden& Frey,2019).

كما أن الواقع الفعلي وما يشهده الميدان التربوي من انتشار ظاهرة تسرب أسئلة الاختبارات والجدل الحادث كل عام عن ضرورة البحث عن طرق أكثر عملية للحد من هذه الظاهرة يُمثل

متطلباً مهماً لتطبيق أكثر من صورة للاختبار الواحد، وتأتي عملية المعادلة لتمكننا من مقارنة أداء الأفراد على هذه الصورة المختلفة.

وفي ضوء ماسبق؛ نجد أن هناك نتائج متباينة حول أثر طريقة المعادلة في القيم الناتجة من عملية معادلة صورتي الاختبار والحكم على دقتها، بالإضافة إلى الأهمية العملية والتطبيقية لبناء الاختبارات وفق نظرية الاستجابة للمفردة، ومعادلة درجات صورها باستخدام الطرق المستندة إليها.

لذا يسعى البحث الحالي التعرف على أثر اختلاف طريقة المعادلة (المتوسط/ المتوسط، المتوسط/ الانحراف المعياري، هيبارا، ستوكنج ولورد) في معادلة درجات صورتي اختبار تحصيلي وفق نظرية الاستجابة للمفردة.

وبناءً على ما سبق يمكن صياغة مشكلة البحث الحالي في التساؤلات التالية:

1- ما أثر اختلاف طريقة المعادلة (المتوسط/ المتوسط، المتوسط/ الانحراف المعياري، هيبارا، ستوكنج ولورد) في القيم الناتجة من معادلة درجات صورتي اختبار تحصيلي في العلوم وفق نظرية الاستجابة للمفردة.

2- ما أثر اختلاف طريقة المعادلة (المتوسط/ المتوسط، المتوسط/ الانحراف المعياري، هيبارا، ستوكنج ولورد) في دقة معادلة درجات صورتي اختبار تحصيلي في العلوم وفق نظرية الاستجابة للمفردة.

أهدف البحث:

يهدف البحث الحالي إلى التعرف على أثر اختلاف طريقة المعادلة في معادلة صورتي اختبار تحصيلي وفق نظرية الاستجابة للمفردة.

أهمية البحث:

يمكن أن تتضح أهمية البحث الحالي من خلال ما يلي:

1- إلقاء الضوء على نظرية الاستجابة للمفردة وتطبيقاتها في عملية بناء ومعادلة صور الاختبار والتي تعد من أهم التوجهات الحديثة في مجال القياس والتقويم.
2- يسهم البحث الحالي في توجيه مؤسسات التعليم إلى ضرورة الاهتمام بعملية معادلة صور الاختبار، وتوظيف التوجهات الحديثة في القياس النفسي للمحافظة على سرية الاختبارات.
3- يسهم هذا البحث في تقديم معلومات وإرشادات مهمة تساعد المهتمين بمجال القياس والتقويم في انتقاء أفضل الطرق لمعادلة صور الاختبارات.
4- يقدم هذا البحث صورتي اختبار تحصيلي في مادة العلوم وفق نظرية الاستجابة للمفردة يمكن الاستفادة منه في مجال القياس النفسي، وبخاصة مع ندرة الاختبارات التي يتم بناؤها وفق هذه النظرية.

التعريف الإجرائي لمصطلحات البحث:

1- معادلة الاختبار: Test Equating

يعرفها الباحثون إجرائياً بأنها: العملية الإحصائية التي تعمل على تعديل درجات إحدى صورتي اختبار تحصيلي إلى درجات الصورة الأخرى لجعل الدرجات التي يحصل عليها الأفراد من تطبيق الصورتين تمثل المستوى نفسه من القدرة التي يقيسها الاختبار بغض النظر عن الصورة المقدمة لهم، وبالتالي يمكن استخدام صورتي الاختبار بطريقة تبادلية حتى وإن اختلفت الصورتان في مستوى الصعوبة والمحتوى.

2- طرق المعادلة: Equating Methods

إجراءات رياضية تم تطويرها لتحقيق عملية المعادلة بين صورتي الاختبار، والتي يمكن من خلالها حساب ثابتي المعادلة (A& B) المستخدمان في تحويل ميزان درجات إحدي صورتي الاختبار إلى ميزان درجات الصورة الأخرى، وسيتم في هذا البحث استخدام الطرق التابعة لنظرية الاستجابة للمفردة والمتمثلة في طريقة المتوسط/ المتوسط (M/M)، وطريقة المتوسط/ الانحراف

المعياري (M/S)، وطريقة هيبارا (HA)، وطريقة ستوكينج ولورد (SL).

3- الاختبار التحصيلي وفق نظرية الاستجابة للمفردة: Achievement test according to item response theory

مجموعة من المفردات الاختبارية ثنائية الاستجابة (صفر- واحد) من نوع الاختيار من متعدد تقيس أهداف إجرائية خاصة بمحتوى مادة العلوم وحدة (دورية العناصر وخواصها) المقررة على تلاميذ الصف الثاني الإعدادي الفصل الدراسي الأول، يتم إعدادها وفق افتراضات نظرية الاستجابة للمفردة (أحادية البعد، الاستقلال الموضعي، منحنى خاصية المفردة، التحرر من السرعة)، ويتم تحليلها واستخراج بارامتراتها (الصعوبة، التمييز، التخمين) باستخدام أحد نماذج نظرية الاستجابة للمفردة أحادية البعد وهو النموذج اللوغاريتمي ثلاثي البارامتر.

حدود الدراسة:

تتمثل حدود الدراسة الحالية فيما يلي:

1- حدود زمانية: تتمثل في فترة تطبيق صورتي الاختبار التحصيلي خلال الفصل الدراسي الأول من العام الدراسي 2020/ 2021م.

2- حدود مكانية: تتمثل في المعاهد الأزهرية بإدارة (شمال الأزهرية، جنوب الأزهرية، شرق الأزهرية، غرب الأزهرية) التابعة لمنطقة القاهرة الأزهرية.

3- حدود موضوعية: تتمثل في تلاميذ الصف الثاني الإعدادي بمعاهد منطقة القاهرة الأزهرية، وصورتي اختبار تحصيلي في مادة العلوم وحدة (دورية العناصر وخواصها).

4- حدود منهجية: وتتمثل في استخدام النموذج اللوغاريتمي ثلاثي البارامتر، وتصميم تصميم المجموعات غير المتكافئة ذات الجذع المشترك في جمع البيانات، والأساليب الإحصائية المستخدمة في الدراسة، والبرامج الحاسوبية الإحصائية المستخدمة في تحليل بيانات الدراسة.

المفاهيم الأساسية للبحث:

أولاً: نظرية الاستجابة للمفردة Item Response Theory (IRT)

تعد نظرية الاستجابة للمفردة أحد التوجهات السيكومترية المعاصرة التي حاولت التغلب على العديد من أوجه القصور في استخدام النظرية الكلاسيكية لبناء الاختبارات والمقاييس النفسية والتربوية، وتحقيق متطلبات القياس الموضوعي، لذلك حظيت هذه النظرية باهتمام علماء القياس المعاصر والباحثين والمؤسسات التعليمية والتربوية المهتمين ببناء الاختبارات والمقاييس النفسية وتفسير درجاتها.

وتعتمد نظرية الاستجابة للمفردة على افتراضات قوية مقارنة بالنظرية الكلاسيكية مفادها أن هناك استقلالية للمفردات الاختبارية في قياس القدرات المختلفة (استقلال خصائص مفردات الاختبار عن قدرات الأفراد، والعكس)، وأن لكل مفردة اختبارية من مفردات الاختبار خصائصها السيكومترية الخاصة بها، والمُمثلة في منحنى مُميز لها يصف احتمالية الحصول على الإجابة الصحيحة أو الخاطئة بناءً على قدرة الفرد وخصائص المفردات. (Magno, 2009. 2)

وتمثل نظرية الاستجابة للمفردة بمجموعة من النماذج الرياضية التي تحدد العلاقة الاحتمالية بين الاستجابات الملاحظة على الاختبار والسمات الكامنة المسببة لها والتي تقيسها مفردات الاختبار، هذه السمات هي التي تحدد استجابات الأفراد على مفردات الاختبار، وتمثل السمة ببعد كمي متصل يمكن أن يحدد عليه مواقع الأفراد المستجيبين على مفردات هذا الاختبار. (مايا بركات، 2010، 50)

لذا يمكن النظر إلى نظرية الاستجابة للمفردة على أنها عبارة عن مجموعة من الدوال الرياضية الاحتمالية التي تحدد العلاقة بين استجابات الأفراد على أحد مفردات الاختبار والسمات أو القدرات التي يمتلكها الفرد والتي يُفترض أنها السبب وراء هذه الاستجابات، وذلك باستخدام بارامترات هذه المفردات كالصعوبة والتمييز والتخمين، فكلما زادت قدرة الفرد الخاصة بما تقيسه مفردات الاختبار كلما زاد احتمال إجابته الصحيحة على هذه المفردات.

وتستند نظرية الاستجابة للمفردة على مجموعة من الافتراضات التي يجب تحققها في البيانات لكي يتم التوصل إلى نتائج دقيقة يمكن الوثوق بها، تتمثل هذه الافتراضات فيما يلي:

1- أحادية البعد: Unidimensionality

تفترض نماذج نظرية الاستجابة للمفردة وجود قدرة واحدة تفسر أداء الفرد في الاختبار، أي يُعرَف المتغير المقاس بواسطة مجموعة من المفردات الاختبارية تقيس بعداً واحداً على متصل السمة، كما يكون الأفراد ذوي قدرة أحادية البعد هي وحدها التي تحدد مستوى أدائهم على الاختبار. (إسماعيل الوليلي، 2002، 48)

وبالتالي فإن نماذج نظرية الاستجابة للمفردة تعتمد على افتراض أن المفردات الاختبارية تقيس قدرة واحدة على متصل السمة، ويمكن التحقق من ذلك عن طريق التحليل العاملي الاستكشافي على مستوى مفردات الاختبار والذي يكشف عن بنية العوامل الكامنة وراء التغاير الملحوظ بين استجابات الأفراد على مفردات الاختبار، ويمكن اعتبار الاختبار أحادي البعد إذا كانت النسبة بين الجذر الكامن للعامل الأول والجذر الكامن للعامل الثاني أكبر من (2)، أو من خلال كون العامل الأول يفسر نسبة (20%) من نسبة التباين الكلي الذي يفسره جميع العوامل المستخرجة من التحليل العاملي والتي يكون قيمة الجذر الكامن لها أكبر من الواحد الصحيح. (Reeve,2002, 12 ؛ Georgiev, 2008, 10)

2- الاستقلال الموضعي: Local Independence

ويعني احتمال الاستجابة الصحيحة للفرد على أحد مفردات الاختبار يكون مستقلاً عن احتمال استجابته على أي مفردة أخرى عند التحكم في قدرة الفرد ومعالم المفردة، ويمكن التحقق من هذا الافتراض من خلال حساب معاملات الارتباط بين مفردات الاختبار للمجموعة الكلية، وللمجموعة ذات المستويات العليا، وللمجموعة ذات المستويات الدنيا، ثم حساب متوسطات معاملات الارتباط في الحالات الثلاثة، ويتحقق افتراض الاستقلال الموضعي إذا كانت قيم متوسطات معاملات الارتباط بين المفردات بالنسبة للمجموعات الثلاث قريبة من الصفر، وكانت قيمتي متوسطي معاملات الارتباط بين المفردات في المجموعة العليا والمجموعة الدنيا أقل من قيمة متوسط معاملات الارتباط في المجموعة الكلية.(محمد عبدالوهاب، 2010، 55؛ أيمن غانم، 2015، 167)

كما يمكن التحقق أيضاً من افتراض الاستقلال الموضعي باستخدام مؤشر (Q3) التي اقترحته Yen (1984)، وهو عبارة عن معاملات الارتباط بين البواقي لكل زوج من أزواج مفردات الاختبار، ويتحقق الاستقلال الموضعي إذا كانت جميع معاملات الارتباط بين أزواج مفردات الاختبار أقل من نقطة القطع التي اقترحتها "ين" وهي القيمة المطلقة للعدد (0,2)، وكان متوسط قيم مؤشر (Q3) أقل من القيمة المتوقعة للمتوسط الحسابي للتوزيع العيني لهذا المؤشر.(Castaneda, 2017, 19؛ إياد حمادنة، 2011، 62)

3- المنحنى المميز للمفردة: Item Characteristic Curve

يمثل المنحنى المميز للمفردة العمود الفقري لنظرية الاستجابة للمفردة، فكل التكوينات الفرضية الأخرى للنظرية تعتمد على هذه المنحنى، وهو عبارة عن رسم بياني على شكل حرف (S) يصف العلاقة بين احتمال الاستجابة الصحيحة على المفردة وقدرة الفرد، ولكل مفردة من مفردات الاختبار منحنى مميز خاص بها، وتقترب فيه احتمال الاستجابة الصحيحة من الصفر عند المستويات المنخفضة من القدرة، ويزيد احتمال الاستجابة الصحيحة عند المستويات المرتفعة حتى يقترب من الواحد الصحيح.(Baker, 2001, 7)

وتتفق جميع نماذج نظرية الاستجابة للمفردة في وجود هذا المنحنى، والذي يتم إنشاؤه بواسطة النموذج الملائم للبيانات الاختبارية، إلا أنه يختلف من نموذج إلى آخر باختلاف عدد بارامترات (معالم) المفردات التي يتم نمذجتها في صيغة رياضية احتمالية، فمثلاً- كما سيتضح فيما بعد- في النموذج أحادي البارامتر يعتمد شكل المنحنى على بارامتر واحد فقط وهو بارامتر

الصعوبة، أما في النموذج الثنائي فيعتمد شكل المنحنى على بارامترين هما بارامتر الصعوبة وبارامتر التمييز.(Gleason, 2008, 9)

4- التحرر من السرعة

تفترض نظرية الاستجابة للمفردة أن عامل السرعة لا يلعب دوراً مؤثراً في الإجابة على مفردات الاختبار، أي أن إخفاق بعض الأفراد في إجابة بعض مفردات الاختبار يرجع إلى انخفاض قدرتهم وليس إلى تأثير عامل السرعة في إجاباتهم، والاختبار الذي يعتمد على السرعة لا يخل بافتراض أحادية البعد فقط، بل أنه قد يؤثر في تقدير معالم نماذج نظرية الاستجابة للمفردة، وبشكل خاص المفردات التي تقع في نهاية الاختبار، وذلك لأن بعض الأفراد سيتركون بعض المفردات بدون إجابة، مما يؤدي إلى تشويه تقدير قدرات الأفراد الذين تركوا بعض المفردات بدون إجابة.(Dawber, 2004, 18-19؛ هبة ضعضع، 2020، 23)

نماذج نظرية الاستجابة للمفردة:

يُعبر عن نظرية الاستجابة للمفردة بدوال رياضية تستخدم للتنبؤ بأداء الفرد على مفردة باستخدام بارامترات هذه المفردة والقدرات التي تكمن وراء هذا الأداء، وتظهر هذه النماذج بيانياً من خلال دوال رياضية تربط احتمال الاستجابة الصحيحة على مفردات الاختبار بالقدرة التي يمتلكها الفرد، ويتحدد شكل هذه الدوال ببارامترات هذه المفردات. (Stone& Hansen, 2000, 974)

وتنقسم نماذج نظرية الاستجابة للمفردة إلى نماذج أحادية البعد، ونماذج متعددة الأبعاد، وينقسم كل منهما إلى نماذج ثنائية الاستجابة ونماذج متعددة الاستجابة، ويتناول الباحث نماذج نظرية الاستجابة للمفردة أحادية البعد ثنائية الاستجابة- كونها النماذج المستخدمة في البحث الحالي وذلك باعتبار أن مفردات الاختبار تقيس بعداً واحداً والاستجابة عليها ثنائية (إما واحد في حالة الإجابة الصحيحة، وإما صفر في حالة الإجابة الخاطئة)- والمثمثلة في النموذج اللوغاريتمي أحادي البارمتر (نموذج راش)، النموذج اللوغاريتمي ثنائي البارامتر (نموذج لورد) والنموذج اللوغاريتمي ثلاثي البارمتر (نموذج بيرنبوم)، وبرجوع الباحث إلى كل من (Embretson& Reise,2000؛ Baker, 2001؛ إسماعيل الوليلي، 2002؛ صلاح الدين علام، 2005؛ Kim,2007؛ آر إيالا،2017/2009) أمكن تعريف هذه النماذج كما يلي:

- النموذج أحادي البارامتر (نموذج راش): وهو أحد نماذج نظرية الاستجابة للمفردة أحادية البعد ثنائية الاستجابة يكون المتغير التابع فيه هو احتمال أن يجيب فرد (j) إجابة صحيحة على مفردة (i)، أما المتغيرات الكامنة المستقلة فهي درجة قدرة الفرد (θ) وبارامتر صعوبة المفردة (b_i).
- النموذج ثنائي البارامتر (نموذج لورد): وهو أحد نماذج نظرية الاستجابة للمفردة أحادية البعد ثنائية الاستجابة يكون المتغير التابع فيه هو احتمال أن يجيب فرد (j) إجابة صحيحة على مفردة (i)، أما المتغيرات الكامنة المستقلة فهي درجة قدرة الفرد (θ) وبارامتر صعوبة المفردة (b_i) وبارامتر تمييز المفردة (α_i).
- النموذج ثلاثي البارامتر (نموذج بيرنبوم): وهو أحد نماذج نظرية الاستجابة للمفردة أحادية البعد ثنائية الاستجابة يكون المتغير التابع فيه هو احتمال أن يجيب فرد (j) إجابة صحيحة

على مفردة (i)، أما المتغيرات الكامنة المستقلة فهي درجة قدرة الفرد (θ) وبارامتر صعوبة المفردة (b_i) وبارامتر تمييز المفردة (α_i) وبارامتر التخمين (C_i).

ثانياً: معادلة الاختبارات

تستخدم معادلة الاختبارات غالباً في المواقف التي توجد فيها صوراً متعددة من الاختبار الواحد ومقارنة درجات الأفراد الذين أخذوا هذه الصور مع بعضهم البعض، أو إذا أراد أحد الباحثين التغلب على مشكلات تأثير الممارسة على درجات الاختبار، وقد برزت أهمية المعادلة جنباً إلى جنب مع الجوانب العديدة لعملية بناء وتطوير الاختبارات، بالإضافة إلى أن مطوري الاختبارات وجميع المتخصصين في بناء الاختبارات يجب أن يكونوا على دراية كافية بالمفاهيم والإجراءات الإحصائية المرتبطة بعملية معادلة الاختبارات. (Falen, 2002, 3)

مفهوم معادلة الاختبار: Test Equating

تعددت التعريفات التي استهدفت وضع تعريف لمعادلة الاختبار، يعرض منها الباحث ما يلي:

يعرفها صلاح الدين علام (2005، 214) بأنها: عملية تحويل نظام وحدات القياس الخاصة بإحدى صورتي الاختبار إلى نظام وحدات القياس الخاصة بالصورة الأخرى، بحيث تصبح القياسات المستمدة من درجات كل من الصورتين متكافئة بعد إجراء هذا التحويل.

وتعرفها منار طومان (2006، 55) بأنها: العملية الإحصائية التي تُوجِد تكافؤ درجات الاختبارات المختلفة التي أُعدت بنفس المحتوى والمواصفات الإحصائية، وتجعل للمقارنة بينهما معنى، حيث تُعدِل اختلافات الصعوبة بينهما وتضعها على مقياس مشترك فيحصل المفحوص على نفس الدرجة بغض النظر عن سهولة أو صعوبة الاختبار المستخدم.

كما يعرفها Von Davier& Wilson (2007,940) بأنها: تلك العلمية الإحصائية التي يحدث من خلالها تحديد ما تساويه الدرجة التي يحصل عليها الفرد في أحد الاختبارات من درجة يحصل عليها في اختبار آخر (يقيس نفس المتغير) ، وذلك وفقاً لمعيار ثابت (الميزان المشترك الذي تتدرج عليه مفردات كلاً من الاختبارين)، وغالباً ما تستخدم في المواقف التي تتطلب استخدام أكثر من صورة اختبارية.

ويعرفها Chen, Huang& Gregor (2009,3) بأنها: العملية الإحصائية التي تستخدم لتعديل إحدى درجات صورتي الاختبار إلى الصورة الأخرى، مما يسمح باستخدام درجات هاتين الصورتين بالتبادل، أي أن معادلة الاختبار تقوم بتعديل الاختلافات في مستوى الصعوبة بين صورتي الاختبار لتكون متساوية في الصعوبة والمحتوى.

ويضع Livingston (2014,11) تعريفاً عاماً لمعادلة الاختبار ينص على أن الدرجة في الصورة الجديدة للاختبار والدرجة في الصورة المرجعية تكون متعادلة (متكافئة) في المجموعة المتقدمة للاختبار إذا كانوا يمثلون نفس الوضع النسبي في المجموعة.

ويعرفها يوسف المحروق (2019، 211) بأنها: إجراء إحصائي يتم فيه تحويل سلم الدرجات على أحد الاختبارات إلى سلم الدرجات على الاختبار الآخر، بحيث يمكن معرفة درجة الفرد على أحد الاختبارات إذا علمنا درجته على الاختبار الآخر.

وفي ضوء ما سبق؛ يعرف الباحثون معادلة الاختبار بأنها: العملية الإحصائية التي تعمل على تعديل درجات إحدى صورتي اختبار إلى درجات الصورة الأخرى لجعل الدرجات التي يحصل عليها الأفراد من تطبيق الصورتين تمثل المستوى نفسه من القدرة التي يقيسها الاختبار بغض النظر عن الصورة المقدمة لهم، وبالتالي يمكن استخدام صورتي الاختبار بطريقة تبادلية حتى وإن اختلفت الصورتان في مستوى الصعوبة والمحتوى.

تصميمات جمع البيانات في معادلة الاختبارات:

تتمثل تصميمات جمع البيانات اللازمة لمعادلة الاختبارات في *تصميم المجموعة الواحدة Single Group Design* حيث يتم تطبيق صورتي الاختبار المراد معادلة درجاتهما على نفس المجموعة من المفحوصين، *وتصميم المجموعة الواحدة المتوازن Single Group Design with Counterbalancing* حيث يتم تقسيم مجموعة واحدة من المفحوصين إلى مجموعتين متساويتين تقسيماً عشوائياً، ثم تُطبق صورتي الاختبار عليهما بطريقة التوازن العكسي، أي أن إحدى المجموعتين تأخذ الصورة الأولى (X) من الاختبار، وتأخذ المجموعة الأخرى الصورة الثانية (Y) من الاختبار، ثم تأخذ كل مجموعة الصورة التي لم تأخذها في المرحلة الأولى مباشرة، *وتصميم المجموعات العشوائية Random Groups Design* ويتم فيه تقديم صورتي الاختبار إلى مجموعتين مستقلتين عشوائيتين تكونين ممثلتين للمجتمع الأصلي، بحيث تطبق صورة واحدة على كل مجموعة، *وتصميم المجموعات غير المتكافئة ذات الجذع المشترك Anchor Test Non-Equivalent Groups Design*ويعتبر هذا التصميم من أكثر تصميمات جمع البيانات اللازمة لاجراء عملية المعادلة شيوعاً واستخداماً في الدراسات والبحوث. (Felan, 2002, 7؛ Kim, Choi, Lee & Um, 2008, 84)

ففي هذا التصميم (وهو التصميم المستخدم في البحث الحالي) يتم تضمين مجموعة من المفردات المشتركة يطلق عليها مفردات الجذع المشترك بين صورتي الاختبار، بحيث تطبق الصورة الأولى (X) مع اختبار الجذع المشترك (V) على مجموعة من المفحوصين، بينما تطبق الصورة الثانية (Y) مع اختبار الجذع المشترك (V) على مجموعة أخرى من المفحوصين، ويؤدي هذا الاجراء إلى بناء دالة معادلة تضع الدرجات التي يتم الحصول عليها من صورتي الاختبار على مقياس واحد.(Michaelides, 2003, 2-3)

ويجب الإشارة إلى أن تصميم الجذع المشترك له نوعان: فإما أن تُطبق مفردات الجذع المشترك منفصلة في جزء مستقل عن الاختبار فيسمى اختبار خارجي External Anchor وفي هذه الحالة لا تدخل درجات اختبار الجذع المشترك في حساب الدرجة الكلية لصورتي الاختبار، أو تطبق مفردات الجذع المشترك موزعة داخل صورتي الاختبار فيسمى اختبار داخلي Internal Anchor وفي هذه الحالة تدخل درجات اختبار الجذع المشترك في حساب الدرجة الكلية لصورتي الاختبار. (صلاح الدين علام، 2005، 220؛ Kolen& Brennan, 2014, 18)

ويعتبر اختبار الجذع المشترك صورة مصغرة من صورتي الاختبار المراد معادلة درجاتهما، لذا ينبغي أن تتناسب المفردات المشتركة مع صورتي الاختبار من حيث المحتوى والخصائص الإحصائية، ولضمان تحقيق ذلك يجب مراعاة ألا يختلف مواقع مفردات اختبار الجذع المشترك في صورتي الاختبار المراد معادلتهما، وألا تختلف في الصياغة اللغوية أو ترتيب البدائل في صورتي الاختبار، كما يجب أن تكون عدد مفردات اختبار الجذع المشترك كافية لتمثيل

المحتوى المراد قياسه، والقاعدة العامة للحد الأدنى لطول الجذع المشترك هو (20- 25%) من عدد مفردات الاختبار الأصلي.(Ryan& Brockmann, 2009, 40 ؛Felan, 2002, 8)

طرق معادلة الاختبارات:

يصنف العاملون في مجال القياس والتقويم النفسي طرق معادلة الاختبارات إلى فئتين أساسيتين: تتضمن الفئة الأولى *طرق المعادلة وفق النظرية الكلاسيكية (CTT)*، وتتضمن الفئة الثانية *طرق المعادلة وفق نظرية الاستجابة للمفردة (IRT)*، وفيما يلي عرض لطرق المعادلة وفق نظرية الاستجابة للمفردة كونها المستخدمة في البحث الحالي، حيث تُصنف طرق المعادلة وفق نظرية الاستجابة للمفردة بناءً على طرق تقدير ثابتي التحويل إلى:

1- طريقة المتوسط/ المتوسط: Mean/Mean Method (MM)

تمثل هذه الطريقة إحدى الطرق الآنية لمعادلة صور الاختبار وفق نظرية الاستجابة للمفردة التي يتم فيها تقدير ثابتي التحويل من خلال متوسطات بارامتر الصعوبة وبارامتر التمييز للمفردات المشتركة في كل من صورتي الاختبار. (Zhonghua, 2010, 18)

2- طريقة المتوسط/ الانحراف المعياري: Mean/ Sigma Method (MS)

تمثل هذه الطريقة إحدى الطرق الآنية لمعادلة صور الاختبار وفق نظرية الاستجابة للمفردة التي يتم فيها تقدير ثابتي التحويل من خلال المتوسطات والانحرافات المعيارية لبارامتر الصعوبة للمفردات المشتركة في كل من صورتي الاختبار. (Zhonghua, 2010, 19)

3- طريقة هيبارا: Haebara Method (HA)

اقترحت هذه الطريقة من قِبَل هيبارا (Haebara, 1980)، والتي تعتمد على فكرة أن الأفراد المفحوصين الذين لديهم قدرة محددة يمتلكون نفس احتمال الإجابة الصحيحة على المفردة بغض النظر عن الصورة الاختبارية المقدمة إليهم، وبالتالي فإن معادلة صورتي الاختبار تعتمد على تحويل المنحنى المميز لمفردات إحدى الصورتين إلى المنحنى المميز لمفردات الصورة الأخرى. (Michaelides,2006, 24)، وتعتمد طريقة هيبارا في تقدير قيم ثابتي التحويل (A& B) على تقليل الفرق بين مساحتى المنحني المميز لمفردات الصورتين، حيث تستخدم دالة رياضية لحساب الفرق بين المنحنيات المميزة عن طريق حساب مجموع مربع الفرق بين المنحنيات المميزة لكل مفردة عند مستوى قدرة معين للمفحوصين.(Keller& Keller, 2011, 369)

4- طريقة ستوكينج ولورد: Stocking and Lord Method (SL)

اقتُرحت هذه الطريقة من قِبَل ستوكينج ولورد (Stocking and Lord, 1983)، وتعتمد على نفس فكرة هيبارا في تحويل المنحنى المميز لمفردات إحدى الصورتين إلى المنحنى المميز لمفردات الصورة الأخرى، لكنها تعتمد في تقدير قيم ثابتي التحويل (A& B) على حساب الفرق بين المنحنيات المميزة عن طريق حساب مربع مجموع الفرق بين المنحنيات المميزة لكل مفردة عند مستوى قدرة معين للمفحوصين.(Keller& Keller, 2011, 368).

دراسات وبحوث سابقة:

دراسة (رائد المدانات، 2010): هدفت الدراسة إلى مقارنة فاعلية طريقة تكر الخطية وطريقة ليفين في معادلة الاختبارت عند استخدام التصميم القائم على الجذع المشترك ومجموعات غير عشوائية، وتوصلت نتائج الدراسة إلى عدم وجود فروق ذات دلالة إحصائية بين متوسطي القيم المُعادَلة ترجع إلى طريقتي المعادلة، ووجود فروق ذات دلالة إحصائية بين متوسطات القيم المعادلة ترجع إلى عدد مفردات الجذع المشترك (1، 4، 7)، بينما الفروق لم تكن ذات دلالة إحصائية عندما كان عدد مفردات الجذع المشترك (10) مفردات.

دراسة (مايسة أبومسلم، 2010): هدفت الدراسة إلى استخدام بعض طرق معادلة الاختبار في إجراء معادلة بين صورتي اختبار توني للذكاء، وقد أظهرت نتائج الدراسة وجود فروق في القيم المتعادلة بين طرق المعادلة وفق النظرية الكلاسيكية ونظرية الاستجابة للمفردة، كما توصلت الدراسة إلى أن طريقة المتوسط/الانحراف المعياري جاءت أكثر الطرق دقة في معادلة صورتي الاختبار.

دراسة (Pang, Madera, Radwan& Zhang,2010): هدفت الدراسة إلى مقارنة أربع طرق لمعادلة درجات صور الاختبار، وقد توصلت نتائج الدراسة إلى عدم وجود فروق بين متوسطات قيم (RMSE) بين طرق المعادلة باستخدام بارامترات المفردات المشتركة الثابتة، وطريقة ستوكينج ولورد، وطريقة المتوسط /الانحراف المعياري، وتفوق هذه الطرق الثلاث على طريقة المعايرة المتزامنة في دقة معادلة الاختبارات.

دراسة (يوسف المحروق، 2011): هدفت الدراسة إلى مقارنة طريقة كيرنيل والمئينيات ونظرية الاستجابة للمفردة عند استخدام تصميم المفردات المشتركة في دقة معادلة درجات الاختبارات، وتوصلت نتائج الدراسة إلى أن طريقة المعادلة باستخدام الدرجات الملاحظة التي تتبع نظرية الاستجابة للمفردة كانت أكثر دقة في معادلة الاختبار تلها طريقة كيرنيل، وذلك عند مختلف المتغيرات التي تناولتها الدراسة.

دراسة (رائد المدانات،2012): هدفت الدراسة إلى مقارنة فعالية طريقة معادلة الدرجات الحقيقية وطريقة الدرجات الملاحظة في معادلة الاختبارات عند استخدام جذع مشترك وتصميم المجموعات غير المتكافئة، وقد أظهرت نتائج الدراسة وجود فروق ذات دلالة إحصائية في متوسطي القيم المعادلة بطريقة الدرجات الحقيقية وطريقة الدرجات الملاحظة عندما كانت عدد مفردات الجذع المشترك (4، 7 ، 10) لصالح طريقة الدرجات الملاحظة، ولكن الفروق لم تكن ذات دلالة إحصائية عندما كان طول الجذع المشترك مفردة واحدة.

دراسة (Kilmen& Demirtasli, 2012): هدفت الدراسة إلى مقارنة بعض طرق معادلة الاختبار التابعة لنظرية الاستجابة للمفردة، وتوصلت نتائج الدراسة إلى أن طريقة ستوكينج-لورد كانت أكثر دقة في معادلة الاختبار حيث أنها أعطت أقل خطأ لعملية المعادلة، وذلك عند ظروف حجم العينة وتوزيع قدرة المفحوصين، كما أظهرت نتائج الدراسة أن حجم العينة الأكبر يقلل من خطأ المعادلة.

دراسة (خالد العنزي، 2014): هدفت الدراسة إلى الكشف عن أثر طول الجذع المشترك وحجم العينة في دقة معادلة درجات صورتي اختبار في الرياضيات وفق تصميم الجذع

المشترك، وقد أظهرت نتائج الدراسة عدم وجود فروق ذات دلالة إحصائية في متوسطي القيم المعادلة بطريقة الدرجات الحقيقية وطريقة الدرجات الملاحظة، كما توصلت الدراسة إلى أن طريقة معادلة الدرجات الملاحظة أكثر دقة من طريقة الدرجات الحقيقية.

دراسة (Asiret& sunbul, 2016): هدفت الدراسة إلى مقارنة بعض طرق معادلة صور الاختبار باستخدام عينات صغيرة الحجم، وذلك في ضوء المتغيرات التالية: حجم العينة، ومتوسط مستوى صعوبة المفردات، وبارامترات التخمين، وقد توصلت نتائج الدراسة إلى عدم وجود فروق بين متوسطات قيم (RMSE) بين طرق المعادلة المستخدمة عندما كان حجم العينة (50) أو أكثر، كما أعطت طريقة المتوسط والطريقة الدائرية متوسط قيم (RMSE) أقل من الطرق الأخرى عندما كان حجم العينة أقل من (50).

دراسة (Ozdemir, 2017): هدفت الدراسة إلى معادلة درجات اختبار دراسة الاتجاهات العالمية في العلوم والرياضيات (TIMSS)، ومقارنة بعض الطرق الكلاسيكية في معادلة صورتي الاختبار، وتوصلت نتائج الدراسة إلى أن طريقة تيكر، وليفين، وبراون هولاند تعطي قيم (RMSE) أقل من المعادلة المئينية مما يشير إلى تفوق هذه الطرق على المعادلة المئينية في معادلة صورتي الاختبار.

دراسة (ديالا العلي ورمضان درويش، 2017): هدفت الدراسة إلى مقارنة فاعلية طريقتي براون هولاند والمئينيات باستخدام الجذع المشترك بين صورتي اختبار في الرياضيات للصف الرابع الابتدائي، ، وقد توصلت نتائج الدراسة إلى وجود فروق ذات دلالة إحصائية بين متوسطي القيم المعادلة بطريقتي براون هولاند وطريقة المئينيات عندما كان عدد مفردات الجذع المشترك (4، 7، 10)، بينما لم تكن الفروق ذات دلالة إحصائية عندما كان عدد مفردات الجذع المشترك مفردة واحدة.

دراسة (Arikan& Gelbal, 2018): هدفت الدراسة إلى مقارنة طريقة كرنيل ببعض الطرق الكلاسيكية في معادلة الاختبارات، وتمثلت أدوات الدراسة في صورتي اختبار تحصيلي (B, D) في مادة الدراسات الاجتماعية، وقد توصلت نتائج الدراسة إلى عدم وجود فروق بين القيم الناتجة ترجع إلى طريقة المعادلة المستخدمة، كما توصلت الدراسة إلى أنه بالرغم من أن طريقة كرنيل أعطت أقل قيم لخطأ المعادلة إلا أن قيم (RMSE) كانت متساوية تقريباً حيث أن الفروق بينها غير دالة إحصائياً.

دراسة (أحمد بديوي، 2019): هدفت الدراسة إلى الحكم على فاعلية أربع طرق لمعادلة صورتين من اختبار الاستعداد المدرسي في الرياضيات تم تدريجهما وفق نموذج راش، وتوصلت نتائج الدراسة إلى أن طريقتي ستوكينج- لورد، وهيبارا كانت أكثر دقة في معادلة صورتي الاختبار من طريقتي المتوسط/المتوسط، وطريقة المتوسط/الانحراف المعياري وفق مؤشري جذر متوسط مربعات الخطأ(RMSE) والتحيز.

دراسة (أحمد عاشور، 2019): هدفت الدراسة إلى بحث أثر نسبة البيانات المفقودة في الجذع المشترك وطريقة التعامل معها على دقة المعادلة العمودية، وتوصلت نتائج الدراسة إلى تفوق طريقة ستوكينج- لورد على طريقة هيبارا في معادلة صورتي الاختبار.

دراسة (Born, Fink, Spoden& Frey,2019): هدفت الدراسة إلى بحث أثر طريقة المعادلة وصعوبة مفردات الجذع المشترك في دقة تقديرات بارامترات المفردة وجودة المعادلة، وتوصلت نتائج الدراسة إلى أن بزيادة حجم العينة تزداد دقة تقدير بارامترات المفردة، كما توصلت الدراسة إلى عدم وجود فروق في دقة المعادلة ترجع اختلاف طريقة المعادلة المستخدمة، بالرغم من أن طريقتي المعادلة اللحظية أعطت قيم مرتفعة من الخطأ لثوابت التحويل بين صورتي الاختبار.

دراسة (ياسر حسن، 2019): هدفت الدراسة إلى بحث أثر اختلاف طريقة المعادلة (المتوسط/ المتوسط، والمتوسط/ الانحراف المعياري)، وطريقة تقدير الدرجات، وقواعد صياغة مفردات الاختبار على دقة تقدير معالم المفردات وقدرات الأفراد في ضوء النظرية الكلاسيكية والنموذج ثلاثي البارامتر، وقد توصلت نتائج الدراسة إلى أن قيم التحيز وجذر متوسط مربعات الخطأ تقل مع ازدياد حجم العينة وطول الاختبار، وأن طريقة المتوسط/ المتوسط أكثر دقة في معادلة الاختبارات من طريقة المتوسط/ الانحراف المعياري وفق النموذج ثلاثي البارامتر.

دراسة (أمل الزعابي، 2020): هدفت الدراسة إلى المقارنة بين طريقة معادلة الدرجات الحقيقية وطريقة الدرجات المشاهدة في إجراء معادلة كتيبات الاختبارات المكونة من مفردات مختلطة ثنائية الاستجابة ومتعددة الاستجابة، وأظهرت نتائج الدراسة وجود فروق ذات دلالة إحصائية بين متوسطي القيم المعادلة بطريقتي الدرجات الحقيقية والدرجات المشاهدة لصالح المعادلة بالدرجات المشاهدة في جميع الكتيبات ماعدا الكتيب (10) والتي كانت الفروق لصالح المعادلة بالدرجات الحقيقية.

دراسة (Wang, Zhang& You, 2020): هدفت الدراسة إلى مقارنة أثر بعض طرق المعادلة (المعادلة المئينية المتساوية، المعادلة وفق نظرية الاستجابة للمفردة، طريقة كيرنل، المعادلة بطريقة كيرنل وفق نظرية الاستجابة للمفردة)، وقد أظهرت نتائج الدراسة أن طريقة كيرنل وفق نظرية الاستجابة للمفردة أكثر دقة واستقراراً من الطرق الأخرى، كما أظهرت النتائج أيضاً أن المعادلة وفق نظرية الاستجابة للمفردة تعطي أقل أخطاء منهجية وعشوائية مقارنة بالطرق الأخرى، وتنخفض هذه الأخطاء باستخدام اختبار أقصر وحجم عينة أكبر.

وقد استفاد الباحث من عرض الدراسات والبحوث السابقة في صياغة مشكلة الدراسة وتحديد التعريفات الإجرائية للمفاهيم المرتبطة بها، والوقوف على الإجراءات المتبعة في القيام بعملية معادلة صور الاختبارات، وتحديد التصميم المناسب لجمع البيانات، واختيار عينة الدراسة، وتحديد الحجم المناسب للعينة في ضوء عدد أفراد العينات المستخدمة في الدراسات والبحوث السابقة، وتحديد محك الحكم على دقة المعادلة في ضوء المحكات الأكثر استخداماً في الدراسات والبحوث السابقة، وقد استقر الباحث على محك جذر متوسط مربعات الخطأ، حيث أنه الأكثر استخداماً في الدراسات والبحوث السابقة، بالإضافة إلى أنه يجمع بين تقييم الخطأ العشوائي والخطأ المنتظم للمعادلة، وصياغة فروض الدراسة، وتحديد الأساليب الإحصائية المناسبة للتحقق من هذه الفروض، وتحديد البرامج الحاسوبية المستخدمة لتصميم أدوات الدراسة وإجراء عملية معادلة صورتي الاختبار بصورة سليمة وأكثر دقة.

فروض البحث:

يسعى البحث الحالي إلى التحقق من الفروض التالية:

1-لا توجد فروق ذات دلالة إحصائية ترجع إلى أثر طريقة المعادلة (المتوسط/ المتوسط- المتوسط/ الانحراف المعياري- هيبارا- ستوكينج ولورد) في القيم الناتجة من معادلة درجات صورتي اختبار تحصيلي في العلوم وفق نظرية الاستجابة للمفردة.

2-لا توجد فروق ذات دلالة إحصائية ترجع إلى أثر طريقة المعادلة (المتوسط/ المتوسط- المتوسط/ الانحراف المعياري- هيبارا- ستوكينج ولورد) في دقة معادلة درجات صورتي اختبار تحصيلي في العلوم وفق نظرية الاستجابة للمفردة.

إجراءات البحث

أولاً: المنهج وتصميم جمع البيانات

1- المنهج: استخدم الباحث المنهج الوصفي التحليلي لمناسبته لطبيعة البحث الحالي والذي يمكن من خلاله استخدام عمليات التحليل والمقارنة بين طرق معادلة صورتي الاختبار وفق نظرية الاستجابة للمفردة.

2- تصميم جمع البيانات: اعتمد الباحث على تصميم المجموعات غير المتكافئة ذات الجذع المشترك، والذي يتطلب إعداد صورتي اختبار تحصيلي في مادة العلوم كل صورة مكونة من عدد من المفردات الفريدة وعدد من الفقرات المشتركة بين الصورتين تسمى مفردات الجذع المشترك، واختيار مجموعتين غير متكافئتين من الأفراد المشاركين في البحث وتطبيق إحدى الصورتين على المجموعة الأولى بينما تطبيق الصورة الأخرى على المجموعة الثانية.

ثانياً: عينة البحث

وتتمثل في المشاركين الذين طبق عليهم الباحث صورتي الاختبار للتحقق من فروض البحث، والتي تم اختيارهم بطريقة العينة الطبقية العشوائية من مجتمع البحث، والتي تكونت من (1032) تلميذاً وتلميذة من بين تلاميذ الصف الثاني الإعدادي الأزهري بمنطقة القاهرة الأزهرية في العام الدراسي (2021/2020م) الفصل الدراسي الأول، تراوحت أعمارهم الزمنية ما بين (12- 15) عام، بمتوسط (13,57) عام، وانحراف معياري (0,56)، منهم (511) تلميذاً من الذكور، و(521) تلميذة من الإناث، وقد راعي الباحث أن يختار عينة الدراسة بحيث تكون ممثلة لمجتمع الدراسة وفق الإدارة الأزهرية (إدارة شمال الأزهرية، إدارة جنوب الأزهرية، إدارة شرق الأزهرية، إدارة غرب الأزهرية)، والنوع (ذكور، إناث).

وقد تم تقسيم عينة البحث إلى مجموعتين غير متكافئتين طُبقت الصورة الاختبارية الأولى (X) على المجموعة الأولى وعددها (516) تلميذاً وتلميذة من تلاميذ الصف الثاني الإعدادي الأزهري تراوحت أعمارهم الزمنية ما بين (12- 15) عام، بمتوسط حسابي (13,61) عام، وانحراف معياري (0,58)، بينما طُبقت الصورة الاختبارية الثانية (Y) على المجموعة الثانية

وعددها (516) تلميذاً وتلميذة من تلاميذ الصف الثاني الإعدادي الأزهري تراوحت أعمارهم الزمنية ما بين (12- 15) عام، بمتوسط حسابي(13,53) عام، وانحراف معياري (0,55).

ثالثاً: أداتا البحث (إعداد: الباحث)

تمثلت أداتا البحث في صورتي اختبار تحصيلي في مادة العلوم للصف الثاني الإعدادي في وحدة (دورية العناصر وخواصها)، وقد تم إعدادهما وفق الخطوات التالية:

1- تحديد الهدف من صورتي الاختبار: هدفت صورتي الاختبار إلى قياس التحصيل الدراسي

2- في مادة العلوم لتلاميذ الصف الثاني الإعدادي في وحدة (دورية العناصر وخواصها) المقررة عليهم في الفصل الدراسي الأول، وذلك في ضوء بعض مستويات بلوم المعرفية (التذكر، الفهم، التطبيق).

3- الاطلاع على الدراسات والبحوث السابقة التي استخدمت صورتي اختبار تحصيلي لإجراء عملية معادلة درجات صورتي الاختبار وذلك للتعرف على الإجراءات المتبعة في بناء صورتي الاختبار وتوزيع مفردات اختبار الجذع المشترك بين هاتين الصورتين.

4- تحديد محتوى الوحدة التعليمية: قام الباحث باختيار وحدة (دورية العناصر وخواصها) المقررة على تلاميذ الصف الثاني الإعدادي في مادة العلوم في الفصل الدراسي الأول، والتي تشتمل على موضوعات (محاولات تصنيف العناصر، تدرج خواص العناصر في الجدول الدوري الحديث، المجموعات الرئيسية بالجدول الدوري الحديث، الماء وخواصه الفيزيائية والكيميائية).

5- تحليل محتوى الوحدة التعليمية: بعد تحديد موضوعات الوحدة المختارة؛ قام الباحث بتحليل محتوى الوحدة التعليمية إلى مجموعة من الحقائق والمفاهيم والقوانين والتعميمات ، ثم قام بحساب ثبات تحليل المحتوى بالاعتماد على طريقة إعادة التحليل، وذلك عن طريق قيام باحث آخر(1) بتحليل نفس محتوى وحدة (دورية العناصر وخواصها) وحساب نسبة الاتفاق بين نتائج التحليلين باستخدام معادلة هولستي Holsti (رشدي طعيمة،2004، 226)، وبتطبيق المعادلة على القيم الناتجة من التحليل تبين أن معامل ثبات التحليل يساوي (0,974) وهو معامل ثبات مرتفع، مما يشير إلى وجود اتفاق مرتفع بين نتائج تحليل الباحث الحالي وتحليل الباحث الآخر على عناصر التحليل.

6- تحديد وتصنيف الأهداف التعليمية: قام الباحث بتحديد الأهداف التعليمية لموضوعات وحدة (دورية العناصر وخواصها) المقررة على تلاميذ الصف الثاني الإعدادي في مادة العلوم في الفصل الدراسي الأول، مسترشداً بالكتاب المدرسي، ثم صاغها في صورة أهداف إجرائية والتي تفاوت عددها من موضوع لآخر تبعاً لأهمية كل موضوع، وعدد ساعات التدريس اللازمة لتدريسه، وبلغ عددها (84) هدفاً، وقد صنفت هذه الأهداف إلى (34) هدفاً في مستوى التذكر، و(30) هدفاً في مستوى الفهم، و(20) هدفاً في مستوى التطبيق.

[1] * د/ رمضان السيد فرحات علي : مدرس علم النفس التعليمي- كلية التربية- جامعة الأزهر.

7- إعداد جدول المواصفات: قام الباحث بإعداد جدول المواصفات وذلك لتوزيع عدد مفردات كل صورة من صورتي الاختبار على جميع دروس الوحدة وفقاً لعدد الأهداف في كل درس وذلك لتحقيق التوازن في كل صورة من صورتي الاختبار والتأكد من أنه يقيس عينة ممثلة لأهداف التدريس ومحتوى الوحدة الدراسية المراد قياس التحصيل فيها، ويوضح الجدول التالي جدول المواصفات الذي استخرجت منه صورتي الاختبار في ضوء محكات عدد الحصص وعدد الأهداف ومستوياتها في كل درس من دروس الوحدة المختارة، وتوزيع العدد الكلي لمفردات الاختبار (60 مفردة) على دروس الوحدة المختارة:

جدول (1)

جدول مواصفات صورتي الاختبار في ضوء محكات عدد الحصص وعدد الأهداف ومستوياتها

الموضوع	عدد الحصص	عدد أهداف			نسبة الموضوع	عدد مفردات			عدد المفردات
		التذكر	الفهم	التطبيق		التذكر	الفهم	التطبيق	
محاولات تصنيف العناصر	6	10	5	5	%20	5	4	3	12
تدرج خواص العناصر في الجدول الدوري الحديث	9	8	10	8	%30	7	7	4	18
المجموعات الرئيسية بالجدول الدوري الحديث	8	7	9	4	%26,67	6	6	4	16
الماء وخواصه الفيزيائية والكيميائية	7	9	6	3	%23,33	6	5	3	14
المجموع	30	34	30	20		24	22	14	60
مجموع الأهداف			84		%100	عدد المفردات الكلي			
نسبة الأهداف		%40,48	%35,71	%23,81		(60) مفردة			

8- صياغة تعليمات صورتي الاختبار: قام الباحث بصياغة تعليمات صورتي الاختبار في نقاط محددة مبيناً الهدف من الاختبار، وعدد مفرداته، ونوعها، وعدد بدائل الإجابات، وكيفية الإجابة عن المفردة، ومكان تدوين إجابات مفردات صورتي الاختبار.

9- صياغة مفردات الاختبار واختيار مفردات الجذع المشترك: صاغ الباحث (60) مفردة من نوع الاختيار من متعدد، لكل مفردة إجابة صحيحة واحدة من بين أربعة بدائل في ضوء مستوى كل هدف من أهداف الوحدة الدراسية المختارة، منهم (40) مفردة فريدة خاصة بكل صورة من صورتي الاختبار، و(20) مفردة مشتركة تمثل اختبار الجذع المشترك (ما يمثل نسبة 33% من العدد الكلي لمفردات الاختبار)، وقد راعى الباحث عند اختيار مفردات الجذع المشترك أن تقيس هذه المفردات نفس مستويات الأهداف التي تقيسها كل صورة من صورتي الاختبار فلا تركز على مستوى دون الآخر، وأن تكون نسبة هذه المفردات في كل مستوى متقاربة لنسبة مفردات الجذع المشترك للعدد الكلي لمفردات كل صورة اختبارية، كما راعى الباحث أن تحتل هذه المفردات نفس المواضع في الصورتين حيث أنها تتمثل في المفردات التي تشغل الأرقام (3-6-9-12-15-18-21-24-27-30-33-36-39-42-45-48-51-54-57-60) في صورتي الاختبار، وفي ضوء ذلك: تكوَن للباحث صورتي اختبار تحصيلي (X ، Y) بالإضافة إلى اختبار الجذع المشترك (V)، كل صورة تتكون من مفردات فريدة تختلف في صورتي الاختبار عن بعضهما البعض ومفردات مشتركة بين الصورتين.

10- طريقة تصحيح الاختبار: قام الباحث بإعداد مفتاحاً لتصحيح صورتي الاختبار، بحيث تعطي الدرجة (واحد) للإجابة الصحيحة لكل مفردة والتي تتفق مع مفتاح التصحيح، بينما تعطى الدرجة (صفر) للإجابة غير الصحيحة أو المفردة المتروكة، ومن ثم فإن أعلى درجة يمكن أن يحصل عليها التلميذ في أي صورة من صورتي الاختبار تساوي (60) درجة.

11- عرض صورتي الاختبار على مجموعة من المحكمين: عرض الباحث صورتي الاختبار في صورتيهما الأولية على (15) محكماً من الأساتذة المتخصصين في مجال علم النفس التعليمي ومناهج وطرق تدريس العلوم بكليات التربية، وبعض معلمي مادة العلوم في المرحلة الإعدادية، وذلك لإبداء آرائهم ومقترحاتهم حول تمثيل مفردات صورتي الاختبار للهدف الذي وضعت لقياسه، وتمثيل مفردات صورتي الاختبار لمستويات الأهداف المعرفية التي تقيسها، ومناسبة صياغة الأهداف التعليمية وارتباطها بمحتوى الوحدة الدراسية، مع إمكانية الإضافة أو الحذف أو التعديل لما قد يرونه مناسباً، وقد تراوحت نسب اتفاق المحكمين على عناصر التحكيم ما بين (86,67%- 100%)، وجاء متوسط نسب الاتفاق بينهم (95,56 %) وهي نسب اتفاق مرتفعة، مما يشير إلى الثقة في صلاحية صورتي الاختبار ومناسبتهما للتطبيق على عينة البحث، وقد أخذ الباحث ملاحظات ومقترحات السادة المحكمين موضع الاهتمام وقام بإجراء التعديلات التي اقترحوها.

12- التحقق من افتراضات نظرية الاستجابة للمفردة:

أ- التحقق من أحادية البعد:

قام الباحث بالتحقق من افتراض أحادية البعد باستخدام التحليل العاملي الاستكشافي لاستجابات الأفراد على صورتي الاختبار، وذلك بعد حساب مدى ملاءمة حجم العينة أو كفايتها لإجراء التحليل العاملي باستخدام اختبار K.M.O Test (Kaiser- Meyer- Olkin) حيث تتراوح قيمة هذا الاختبار بين الصفر والواحد الصحيح، وقد بلغت قيمة إحصائي اختبار K.M.O في

تحليل الصورة الأولى (0,854)، وللصوة الثانية (0,871) وهما قيمتان أكبر من الحد الأدني الذي اشترطه Kaiser وهو (0,50)، وتقعان في نطاق القيم المرتفعة، وبالتالي فإنه يمكن الحكم بكفاية حجم العينة لإجراء هذا التحليل.

وقد تم إجراء التحليل العاملي لمصفوفة الارتباط بطريقة المكونات الأساسية باستخدام الحزمة الإحصائية (SPSS.v.22)، وقد أخذ الباحث بمحك جيلفورد لمعرفة حد الدلالة الإحصائية للتشبعات الناتجة وهو اعتبار التشبعات التي تصل إلى (0,30) أو أكثر تشبعات دالة، وفي ضوء نتائج التحليل العاملي أمكن استخلاص مجموعة من العوامل الرئيسية لكل صورة الجذر الكامن لكل منهم أكبر من الواحد الصحيح، حيث أسفرت نتائج التحليل العاملي للصورة الأولى (X) عن وجود (21) عاملاً الجذر الكامن لهما أكبر من الواحد الصحيح، وأن النسبة بين الجذر الكامن للعامل الأول والجذر الكامن للعامل الثاني بلغت (3,778)، وهي أعلى من القيمة (2)، مما يشير إلى تحقق أحادية البعد، كما أن العامل الأول يفسر نسبة (24,778) من نسبة التباين الكلي الذي يفسره جميع العوامل المستخرجة من التحليل العاملي وهي نسبة أكبر من (20%)، مما يشير أيضاً إلى تحقق أحادية البعد للصورة الأولى، كما أسفرت نتائج التحليل العاملي للصورة الثانية (Y) عن وجود (18) عاملاً الجذر الكامن لهما أكبر من الواحد الصحيح، وأن النسبة بين الجذر الكامن للعامل الأول والجذر الكامن للعامل الثاني بلغت (4,415)، وهي أعلى من القيمة (2)، مما يشير إلى تحقق أحادية البعد، كما أن العامل الأول يفسر نسبة (28,996) من نسبة التباين الكلي الذي يفسره جميع العوامل المستخرجة من التحليل العاملي وهي نسبة أكبر من (20%)، مما يشير أيضاً إلى تحقق أحادية البعد للصورة الثانية.

ب- التحقق من الاستقلال الموضعي:

قام الباحث بالتحقق من افتراض الاستقلال الموضعي عن طريق حساب معاملات الارتباط بين مفردات الاختبار للمجموعة الكلية، وللمجموعة ذات المستويات العليا، وللمجموعة ذات المستويات الدنيا، ثم حساب متوسطات معاملات الارتباط في الحالات الثلاثة، وقد جاءت قيم متوسطات معاملات الارتباط بين المفردات بالنسبة للمجموعات الثلاث (الكلية، العليا، الدنيا) على الترتيب (0,074- 0,068- 0,062) بالنسبة للصورة الأولى، (0,082- 0,069- 0,065) بالنسبة للصورة الثانية، وجميعها قيم قريبة من الصفر، كما جاءت قيمتي متوسطي معاملات الارتباط بين المفردات في المجموعة العليا والمجموعة الدنيا أقل من قيمة متوسط معاملات الارتباط في المجموعة الكلية بالنسبة لصورتي الاختبار، وهذا يدل على أن مفردات كل صورة لا يوجد من بينها مفردات تتأثر إجابتها بمفردات أخري من مفردات نفس الصورة الاختبارية مما يشير إلى تحقق افتراض الاستقلال الموضعي.

كما قام الباحث بحساب قيم مؤشر (Q3) لفحص معاملات الارتباط بين البواقي لجميع أزواج المفردات وذلك باستخدام برنامج (jMetrik. V.4.1.1)، وقد تراوحت قيم معاملات الارتباط بين البواقي لجميع أزواج مفردات الصورة الأولى بين (-0,0001: -0,1949) وهي قيم أقل القيمة المطلقة للعدد (0,2) ، كما بلغ متوسط قيم مؤشر (Q3) لجميع أزواج المفردات (-0,01034) وهي أقل من القيمة المتوقعة لهذا المؤشر والبالغة (-1/"ن-1"= -0,016)، مما يشير إلى تحقق الاستقلال الموضعي، كما تراوحت قيم معاملات الارتباط بين البواقي لجميع أزواج

مفردات الصورة الثانية بين (-0,0001: -0,1899) وهي قيم أقل من القيمة المطلقة للعدد (0,2) ، كما بلغ متوسط قيم مؤشر (Q3) لجميع أزواج المفردات (-0,01032) وهي أقل من القيمة المتوقعة لهذا المؤشر والبالغة (-1/"ن-1"= -0,0169)، مما يشير إلى تحقق الاستقلال الموضعي.

جـ- التحرر من السرعة:

قام الباحث بالتحقق من افتراض التحرر من السرعة عن طريق حساب النسبة المئوية للتلاميذ الذين تمكنوا من الانتهاء من الإجابة عن جميع مفردات صورتي الاختبار، حيث بلغت النسبة المئوية للتلاميذ الذين تمكنوا من الانتهاء من الإجابة عن جميع مفردات كل من صورتي الاختبار (100%)، مما يشير تحقق افتراض التحرر من السرعة في صورتي الاختبار.

12- الخصائص السيكومترية لصورتي الاختبار وفق نظرية الاستجابة للمفردة:

أ- ملاءمة مفردات صورتي الاختبار للنموذج ثلاثي البارامتر(صدق ملاءمة المفردات):

تم التحقق من ملاءمة مفردات صورتي الاختبار للنموذج ثلاثي البارامتر باستخدام برنامج بايلوج إم جي (BILOG-MG.3.0)، والذي يعتمد على مؤشر مربع كاي (CHI.SQ) والتي تتطلب أن تكون قيمتها غير دالة إحصائياً حتى تكون المفردة ملاءمة للنموذج، وقد أسفرت قيم المطابقة عن عدم ملاءمة (7) مفردات من مفردات الصورة الأولى (X) للنموذج ثلاثي البارامتر، حيث جاءت قيم كا^2 لها دالة إحصائياً وهي المفردات أرقام (1-9-12-17-33-36-47)، كما أسفرت عن عدم ملاءمة (9) مفردات من مفردات الصورة الثانية (Y) للنموذج ثلاثي البارامتر، حيث جاءت قيم كا^2 لها دالة إحصائياً وهي المفردات أرقام (1-9-14-16-17-21-31-33-44)، وقد اتفقت (4) مفردات غير ملائمة في الصورة الأولى مع مفردات الصورة الثانية وهي المفردات أرقام (1-9-17-33) وبالتالي تم حذفهم من الصورتين، كما تم حذف المفردات (12-36-47) غير الملاءمة في الصورة الأولى مع المفردات التي تقيس نفس الأهداف في الصورة الثانية، وبالمثل تم حذف المفردات (14-16-21-31-44) غير الملائمة في الصورة الثانية مع المفردات التي تقيس نفس الأهداف في الصورة الأولى.

وبذلك يكون إجمالي عدد المفردات المحذوفة في الصورتين (12) مفردة وهي المفردات أرقام(1-9-12-14-16-17-21-31-33-36-44-47)، منهم (7) مفردات من المفردات الفريدة في الصورتين، و(5) مفردات من مفردات الجذع المشترك بين الصورتين، لتصبح عدد مفردات كل صورة اختبارية (48) مفردة منهم (33) مفردة فريدة، و(15) مفردة مشتركة بين الصورتين، وقد قام الباحث بإعادة ترقيم مفردات صورتي الاختبار بعد حذف المفردات غير الملاءمة للنموذج ثلاثي البارامتر.

ب- تدريج مفردات صورتي الاختبار وفق النموذج ثلاثي البارمتر (صدق التدريج):

قام الباحث بإعادة التحليل لمفردات صورتي الاختبار بعد حذف المفردات غير الملائمة من صورتي الاختبار وفقاً للنموذج ثلاثي البارامتر لاستخراج تقدير بارامترات المفردات وهي بارامتر الصعوبة، وبارامتر التمييز، وبارامتر التخمين، وقد قام الباحث بتلخيص بارامترات مفردات صورتي الاختبار في الجدول التالي في ضوء مخرجات التحليل الإحصائي لمفردات صورتي الاختبار باستخدام برنامج بايلوج إم جي:

جدول (2)

المتوسطات الحسابية والانحرافات المعيارية لبارامترات مفردات صورتي الاختبار

البارامترات	الصورة الأولى			الصورة الثانية		
	الصعوبة	التمييز	التخمين	الصعوبة	التمييز	التخمين
أقل قيمة	0,023	0,672	0,143	0,048	0,664	0,181
أكبر قيمة	2,425	2,892	0,360	1,885	2,955	0,398
المتوسط الحسابي	1,148	1,337	0,259	0,926	1,229	0,276
الانحراف المعياري	0,506	0,504	0,048	0,405	0,467	0,050

يتضح من جدول (2) أن مفردات صورتي الاختبار تقع معظمها في نطاق المفردات الصعبة، كما تقع معظمها في نطاق المفردات ذات التمييز المتوسط، كما أن مفردات صورتي الاختبار تتأثر بالتخمين إلى حد ما، مما يدل على مناسبة النموذج الثلاثي لهذه المفردات.

ج- ثبات صورتي الاختبار وفق نظرية الاستجابة للمفردة:

- مؤشر الثبات للاختبار Test Reliability Index: حيث يشير مؤشر الثبات إلى ثبات تقديرات قدرة الأفراد وفق السمة المقاسة، وقد قام الباحث بحساب مؤشر الثبات لصورتي الاختبار باستخدام برنامج البايلوج إم جي، وقد بلغ مؤشر الثبات للصورة الأولى (0,908)، وللصورة الثانية (0,913)، وهي مؤشرات ثبات مرتفعة مما يدل على ثبات صورتي الاختبار في تقدير قدرات الأفراد المختبرين بصورتي الاختبار.

- الثبات الإمبريقي للاختبار Test Empirical Reliability: حيث يشير الثبات الامبريقي إلى مدى اقتراب القدرة التي تم تقديرها من خلال نماذج نظرية الاستجابة للمفردة من القدرة الحقيقية للأفراد، وقد قام الباحث بحساب الثبات الإمبريقي لصورتي الاختبار باستخدام برنامج البايلوج إم جي، وقد بلغ مؤشر الثبات للصورة الأولى (0,942)، وللصورة الثانية (0,944)، وهي قيم ثبات مرتفعة مما يدل على ثبات صورتي الاختبار في تقدير قدرات الأفراد المختبرين بصورتي الاختبار.

وتشير البيانات والقيم السابقة التي تم حسابها (صدق الملاءمة، وصدق التدريج، ، مؤشر الثبات، الثبات الامبريقي) إلى تمتع صورتي الاختبار التحصيلي المستخدمان في الدراسة الحالية بقيم صدق وثبات مرتفعة.

وفي ضوء ما سبق تكون للباحث صورتي اختبار تحصيلي تم تدريج مفرداتهما وفق أحد نماذج نظرية الاستجابة للمفردة أحادية البعد وهو النموذج ثلاثي البارامتر بعد حذف المفردات غير الملاءمة له من صورتي الاختبار واستخراج قيم كل من الصعوبة والتمييز والتخمين، كما تم

حساب المؤشرات الدالة على ثباتهما وفق نظرية الاستجابة للمفردة، والجدول التالي يوضح الصورة النهائية لصورتي الاختبار[1] وأرقام المفردات الفريدة والمشتركة في كل منهما:

جدول (3)

الصورة النهائية لصورتي الاختبار التحصيلي

المفردات	أرقام المفردات في صورتي الاختبار	عدد المفردات
الفريدة	25-24-22-21-19-18-16-15-14-13-10-9-8-7-6-4-3-1 47-46-44-43-41-40-38-37-35-33-31-30-28-27-26-	33
المشتركة	48-45-42-39-36-34-32-29-23-20-17-12-11-5-2	15
-	المجموع	48

رابعاً: خطوات تنفيذ البحث

لتنفيذ البحث الحالي اتبع الباحث الخطوات التالية:

1- الاطلاع على الدراسات والبحوث السابقة ذات الصلة بموضوع البحث الحالي، وكتابة الإطار
النظري الخاص بنظرية الاستجابة للمفردة، ومعادلة الصور المختلفة للاختبار والمتغيرات المرتبطة بها.

2- إعداد أداتا البحث والممثلة في صورتي اختبار تحصيلي في مادة العلوم (X، Y) للصف الثاني الإعدادي، واللتان تحتويان على مفردات الجذع المشترك.

3- اختيار عينة البحث بطريقة العينة الطبقية العشوائية من بين تلاميذ الصف الثاني الإعدادي الأزهري بمنطقة القاهرة الأزهرية.

4- تطبيق صورتي الاختبار (X، Y) على أفراد العينة الاستطلاعية والأساسية وذلك لحساب الخصائص السيكومترية الأولية لصورتي الاختبار، والتحقق من افتراضات نظرية الاستجابة للمفردة، والمتمثلة في افتراض أحادية البعد، والاستقلال الموضعي، والتحرر من السرعة باستخدام برنمجي (SPSS) و(jMetriK).

5- التحقق من مطابقة بيانات صورتي الاختبار لأحد نماذج نظرية الاستجابة للمفردة أحادية البعد (النموذج أحادي البارامتر، النموذج ثنائي البارامتر، النموذج ثلاثي البارامتر) باستخدام برنامج (.BILOG-MG) المُعد لهذا الغرض، وقد أوضحت النتائج ملاءمة البيانات للنموذج ثلاثي البارامتر.

[1] * للحصول على ملاحق البحث يمكن الرجوع إلى رسالة الدكتوراه الخاصة بالباحث بعنوان" أثر التفاعل بين نوع الطريقة وحجم العينة وطول الجذع المشترك في معادلة صورتي اختبار تحصيلي وفق نظرية الاستجابة للمفردة" كلية التربية بنين بالقاهرة- جامعة الأزهر.

6- استبعاد المفردات وكذلك استجابات الأفراد التي لا تتطابق استجاباتهم مع النموذج ثلاثي البارامتر، وتدريج مفردات صورتي الاختبار، وقدرات الأفراد وفق النموذج الثلاثي، وحساب كلاً من مؤشرات الثبات والثبات الامبريقي لصورتي الاختبار.

7- إجراء عملية معادلة درجات صورتي الاختبار التحصيلي (X، Y) باستخدام برنامج (IRT EQ.) المُستخدم في معادلة صور الاختبارات وفق نظرية الاستجابة للمفردة.

8- إجراء المعالجات الإحصائية المناسبة في ضوء فروض البحث، وعرض النتائج وتفسيرها في ضوء الإطار النظري والدراسات والبحوث السابقة.

9- وضع مجموعة من التوصيات التي أسفرت عنها نتائج البحث، وتقديم مجموعة من الدراسات والبحوث المستقبلية المقترحة.

خامساً: الأساليب الإحصائية المستخدمة في البحث:

استخدم الباحث الأساليب الإحصائية التالية في التحقق من فروض البحث:

1- تحليل التباين للقياسات المتكررة Repeated Measures ANOVA
2- اختبار ت لعينتين مرتبطتين: Paired Samples t-test

نتائج البحث ومناقشتها

لاختبار فروض البحث يلزم اختيار أحد صورتي الاختبار لتكون الصورة المرجعية (القاعدة)، وتكون الصورة الأخرى هي الصورة الهدف المراد معادلة درجاتها على تدريج الصورة الأولى، وبالرغم من أن اختيار أحد الصورتين كقاعدة والأخرى كهدف لا يؤثر في عملية المعادلة؛ إلا أن الباحث اختار الصورة الثانية (Y) لتكون هي القاعدة، والصورة الأولى (X) هي الهدف المراد معادلة درجاتها على تدريج الصورة الثانية، وذلك كون قيم ثبات الصورة الثانية (مؤشر الثبات، الثبات الامبريقي) أكبر من قيم ثبات الصورة الأولى، وقد قام الباحث بإدخال بارامترات جميع مفردات صورتي الاختبار وفي وجود حجم عينة (516)، وطول جذع مشترك (15 مفردة) إلى برنامج (IRTEQ.) للحصول على قيم الدرجات الخام المتعادلة بين صورتي الاختبار وثابتي المعادلة حسب طرق المعادلة وفق نظرية الاستجابة للمفردة، وقد اعتمد الباحث على محك جذر متوسط مربعات الخطأ للحكم على دقة نتائج طرق المعادلة، حيث تُعد طريقة المعادلة أكثر دقة في حالة الحصول على قيم أقل لهذا المؤشر، واستخدم الباحث برنامج Microsoft) (Excel 2010لاستخراج قيم جذر متوسط مربعات الخطأ، وذلك لادخالها إلى برنامج (SPSS.V.22) لإجراء عمليات التحليل، وفيما يلي نتائج فروض البحث ومناقشتها:

1- نتائج الفرض الأول ومناقشته:

ينص الفرض الأول على أنه " لا توجد فروق ذات دلالة إحصائية ترجع إلى أثر طريقة المعادلة (المتوسط/ المتوسط- المتوسط/ الانحراف المعياري- هيبارا- ستوكينج ولورد) في القيم الناتجة من معادلة درجات صورتي اختبار تحصيلي في العلوم وفق نظرية الاستجابة للمفردة".

وللتحقق من هذا الفرض استخدم الباحث اختبار تحليل التباين الأحادي للقياسات المتكررة للتعرف على الفروق بين متوسطات القيم الناتجة من عملية المعادلة وفقاً لطرق المعادلة التابعة لنظرية الاستجابة للمفردة، والجدول التالي يوضح ذلك:

جدول (4)

قيمة "ف" ودلالتها الإحصائية للفروق بين متوسطات القيم الناتجة من عملية المعادلة

مصدر التباين	مجموع المربعات	درجات الحرية	متوسط المربعات	قيمة "ف"	مستوى الدلالة
بين الدرجات	34741,050	47	739,171		
بين الطرق	31,990	3	10,663	14,514**	0,01 دالة
الخطأ	103,588	141	0,735		
المجموع الكلي	34876,628	191	182,600		

يتضح من جدول (4) أن قيمة "ف" للفروق بين متوسطات القيم الناتجة من عملية المعادلة وفقاً لطرق المعادلة التابعة لنظرية الاستجابة للمفردة (المتوسط/ المتوسط- المتوسط/ الانحراف المعياري- هيبارا- ستوكينج ولورد) بلغت (14,514) وهي قيمة دالة إحصائياً عند مستوى (0,01)، وهذا يعني رفض الفرض الصفري وقبول الفرض البديل أي " توجد فروق ذات دلالة إحصائية بين متوسطات القيم الناتجة من عملية المعادلة وفقاً لطرق المعادلة التابعة لنظرية الاستجابة للمفردة"، ولتحديد اتجاه الفروق بين متوسطات القيم الناتجة من عملية المعادلة وفقاً لطرق المعادلة الأربعة؛ استخدم الباحث اختبار "ت" لعينتين مرتبطتين كما في الجدول التالي:

جدول (5)

قيمة "ت" ودلالتها الإحصائية لمعرفة اتجاه الفروق بين متوسطات القيم الناتجة من عملية المعادلة

م	طريقة المعادلة	العدد	المتوسط الحسابي	الانحراف المعياري	متوسط الفروق	الانحراف المعياري للفروق	قيمة "ت"	مستوى الدلالة
1	م/م	48	24,172	13,993	0,807-	1,680	3,329**	0,01 دالة
	م/ع		24,980	13,067				
2	م/م	48	24,172	13,993	0,380-	0,941	2,799**	0,01 دالة
	هيبارا		24,552	13,418				
3	م/م	48	24,172	13,993	0,277	0,276	6,961**	0,01 دالة
	ستوكينج/لورد		23,895	13,954				
4	م/ع	48	24,980	13,067	0,427	0,756	3,914**	0,01

م	طريقة المعادلة	العدد	المتوسط الحسابي	الانحراف المعياري	متوسط الفروق	الانحراف المعياري للفروق	قيمة "ت"	مستوى الدلالة
	هيبارا		24,552	13,418				دالة
5	م/ع ستوكينج/لورد	48	24,980 23,895	13,067 13,954	1,084	1,819	**4,130	0,01 دالة
6	هيبارا ستوكينج/لورد	48	24,552 23,895	13,418 13,954	0,657	1070	**4,256	0,01 دالة

يتضح من جدول (5) وجود فروق ذات دلالة إحصائية عند مستوى دلالة (0,01) بين متوسطات القيم الناتجة من عملية المعادلة وفقاً لطرق المعادلة التابعة لنظرية الاستجابة للمفردة لصالح طريقة المتوسط/ الانحراف المعياري عند المقارنة بينها وبين باقي الطرق، ووجود فروق ذات دلالة إحصائية عند مستوى دلالة (0,01) بين متوسطات القيم الناتجة من عملية المعادلة وفقاً لطرق المعادلة التابعة لنظرية الاستجابة للمفردة لصالح طريقة هيبارا عند المقارنة بينها وبين طريقتي المتوسط/ المتوسط، وستوكينج ولورد، ووجود فروق ذات دلالة إحصائية عند مستوى دلالة (0,01) بين متوسطات القيم الناتجة من عملية المعادلة وفقاً لطرق المعادلة التابعة لنظرية الاستجابة للمفردة لصالح طريقة المتوسط/ المتوسط عند المقارنة بينها وبين طريقة ستوكينج ولورد، وهذا يشير إلى أن طريقة المتوسط/ الانحراف المعياري تعطي قيماً للمعادلة أكبر من القيم الأخرى تليها طريقة هيبارا، ثم طريقة المتوسط/ المتوسط، ثم طريقة ستوكينج ولورد.

ويفسر الباحث هذه النتيجة في ضوء خصائص طريقة المتوسط/ الانحراف المعياري والتي تعتمد في حسابها لثابتي التحويل على المتوسطات الحسابية والانحرافات المعيارية لبارامتر صعوبة المفردات المشتركة بين صورتي الاختبار فقط، والذي يجعل القيم المتعادلة وفق هذه الطريقة ترتفع عن القيم المتعادلة وفق الطرق الأخرى التي تعتمد في حساب ثابتي التحويل على متوسطات بارامتري الصعوبة والتمييز للمفردات المشتركة بين صورتي الاختبار كما في طريقة المتوسط/ المتوسط، أو تعتمد على تقليل الفرق بين مساحتي المنحني المميز لمفردات الصورتين كما في طريقتي هيبارا وستوكينج ولورد.

كما يفسر الباحث هذه النتيجة في ضوء اختلاف بارامتري الصعوبة والتمييز للمفردات المشتركة بين صورتي الاختبار مما يؤثر في عملية تحويل تدريج درجات الصورة (X) إلى تدريج درجات الصورة (Y) نتيجة اختلاف طرق المعادلة في حساب ثابتي التحويل.

وتتفق هذه النتيجة مع ما توصلت إليه دراسة كل من (رائد المدانات، 2012؛ ديالا العلي ورمضان درويش، 2017؛ أمل الزعابي، 2020) من وجود فروق ذات دلالة إحصائية في القيم

الناتجة من عملية المعادلة ترجع لطرق المعادلة المستخدمة، وتختلف مع ما توصلت إليه دراسة كل من (رائد المدانات، 2012؛ خالد العنزي، 2014؛ Arikan& Gelbal, 2018) من عدم وجود فروق ذات دلالة إحصائية في القيم الناتجة من عملية المعادلة ترجع لطرق المعادلة المستخدمة.

2- نتائج الفرض الثاني ومناقشته:

ينص الفرض الثاني على أنه "لا توجد فروق ذات دلالة إحصائية ترجع إلى أثر طريقة المعادلة (المتوسط/ المتوسط- المتوسط/ الانحراف المعياري- هيبارا- ستوكينج ولورد) في دقة معادلة درجات صورتي اختبار تحصيلي في العلوم وفق نظرية الاستجابة للمفردة".

وللتحقق من هذا الفرض استخدم الباحث اختبار تحليل التباين الأحادي للقياسات المتكررة للتعرف على الفروق بين متوسطات قيم دقة معادلة صورتي الاختبار وفقاً لطرق المعادلة التابعة لنظرية الاستجابة للمفردة، والجدول التالي يوضح ذلك:

جدول (6)

قيمة "ف" ودلالتها الإحصائية للفروق بين متوسطات قيم دقة معادلة صورتي الاختبار

مصدر التباين	مجموع المربعات	درجات الحرية	متوسط المربعات	قيمة "ف"	مستوى الدلالة
بين الأفراد	5771,814	515	11,207		
بين الطرق	489,510	3	163,170	**207,510	0,01
الخطأ	1214,870	1548	0,786		دالة
المجموع الكلي	7476,194	2063	3,624		

يتضح من جدول (6) أن قيمة "ف" للفروق بين متوسطات قيم دقة معادلة صورتي الاختبار وفقاً لطرق المعادلة التابعة لنظرية الاستجابة للمفردة (المتوسط/ المتوسط- المتوسط/ الانحراف المعياري- هيبارا- ستوكينج ولورد) بلغت (207,510) وهي قيمة دالة إحصائياً عند مستوى (0,01)، وهذا يعني رفض الفرض الصفري وقبول الفرض البديل أي " توجد فروق ذات دلالة إحصائية بين متوسطات قيم دقة معادلة صورتي الاختبار وفقاً لطرق المعادلة التابعة لنظرية الاستجابة للمفردة"، ولتحديد اتجاه الفروق بين متوسطات قيم دقة معادلة صورتي الاختبار وفقاً لطرق المعادلة الأربعة؛ استخدم الباحث اختبار "ت" لعينتين مرتبطتين كما في الجدول التالي:

جدول (7)

قيمة "ت" ودلالتها الإحصائية لمعرفة اتجاه الفروق بين متوسطات قيم دقة معادلة صورتي الاختبار

م	طريقة المعادلة	العدد	المتوسط الحسابي	الانحراف المعياري	متوسط الفروق	الانحراف المعياري للفروق	قيمة "ت"	مستوى الدلالة
1	م/م	516	3,621	1,749	0,954	1,691	12,814**	0,01 دالة
	م/ع		2,667	1,888				
2	م/م	516	3,621	1,749	0,477	0,980	11,073**	0,01 دالة
	هيبارا		3,143	1,847				
3	م/م	516	3,621	1,749	0,333-	0,256	29,613**	0,01 دالة
	ستوكينج/لورد		3,954	1,878				
4	م/ع	516	2,667	1,888	0,476-	0,897	12,048**	0,01 دالة
	هيبارا		3,143	1,847				
5	م/ع	516	2,667	1,888	1,187-	1,861	15,715**	0,01 دالة
	ستوكينج/لورد		3,954	1,878				
6	هيبارا	516	3,143	1,847	0,811-	1,129	16,320**	0,01 دالة
	ستوكينج/لورد		3,954	1,878				

يتضح من جدول (7) وجود فروق ذات دلالة إحصائية عند مستوى دلالة (0,01) بين متوسطات قيم دقة معادلة صورتي الاختبار وفقاً لطرق المعادلة التابعة لنظرية الاستجابة للمفردة لصالح طريقة المتوسط/ الانحراف المعياري عند المقارنة بينها وبين باقي الطرق، ووجود فروق ذات دلالة إحصائية عند مستوى دلالة (0,01) بين متوسطات قيم دقة معادلة صورتي الاختبار وفقاً لطرق المعادلة التابعة لنظرية الاستجابة للمفردة لصالح طريقة هيبارا عند المقارنة بينها وبين طريقتي المتوسط/ المتوسط، وستوكينج ولورد، ووجود فروق ذات دلالة إحصائية عند مستوى دلالة (0,01) بين قيم دقة معادلة صورتي الاختبار وفقاً لطرق المعادلة التابعة لنظرية الاستجابة للمفردة لصالح طريقة المتوسط/ المتوسط عند المقارنة بينها وبين طريقة ستوكينج ولورد، وهذا يشير إلى أن طريقة المتوسط/ الانحراف المعياري أكثر دقة من الطرق الأخري تليها طريقة هيبارا، ثم طريقة المتوسط/ المتوسط، ثم طريقة ستوكينج ولورد.

وهذا يعني أن طريقة المتوسط/ الانحراف المعياري أعطت قيماً أقل لجذر متوسط مربعات الخطأ وبالتالي فهي أفضل من الطرق الأخرى من حيث الدقة في معادلة صورتي الاختبار، ويفسر الباحث هذه النتيجة في ضوء تقارب قيمتى الانحراف المعياري لصعوبة المفردات المشتركة بين صورتي الاختبار، والذي أدي إلى تقارب التدريج الأصلي للصورة الهدف من التدريج الذي تم تحويله باستخدام ثابتي التحويل واللذان يعتمدان في حسابهما على الانحراف المعياري لبارامتر صعوبة المفردات المشتركة في حالة استخدام طريقة المتوسط/ الانحراف المعياري.

كما أن اعتماد طريقة المتوسط/ الانحراف المعياري على بارامتر الصعوبة فقط في حساب ثابتي التحويل، والتي تختلف عن طريق المتوسط/ المتوسط التي تعتمد على بارامتري الصعوبة والتمييز وطريقتي هيبارا وستوكينج/لورد اللتان تعتمدان على جميع بارمترات المفردات، قد يؤدي إلى تقارب الدرجات الحقيقية والدرجات المقدرة في حالة طريقة المتوسط/ الانحراف المعياري عن باقي الطرق نتيجة لزيادة عدد البارمترات التي يتم تقديرها في الطرق الأخرى.

وتتفق هذه النتيجة مع ما توصلت إليه دراسة كل من (يوسف المحروق، 2011؛ Kilmen& Demirtasli, 2012؛ Ozdemir,2017 ؛ أحمد بديوي، 2019) من وجود فروق ذات دلالة إحصائية في دقة المعادلة ترجع لطرق المعادلة المستخدمة، كما تتفق هذه النتيجة مع ما توصلت إليه دراسة (مايسة أبومسلم، 2010) من أن طريقة المتوسط/ الانحراف المعياري أكثر دقة من طريقة المتوسط/ المتوسط، وتختلف مع ما توصلت إليه دراسة كل من (Born, Fink, Spoden& Frey,2019؛ Arikan& Gelbal, 2018) من عدم وجود فروق ذات دلالة إحصائية في دقة المعادلة ترجع لطرق المعادلة المستخدمة، كما تختلف هذه النتيجة مع ما توصلت إليه دراسة (ياسر حسن، 2019) من أن طريقة المتوسط/ المتوسط كانت أكثر دقة من طريقة المتوسط/ الانحراف المعياري.

توصيات الدراسة:

في ضوء ما أسفرت عنه نتائج البحث يقدم الباحث التوصيات السيكومترية والتربوية التالية:

1- عند استخدام طرق معادلة صور الاختبار وفق نظرية الاستجابة للمفردة يجب مراعاة أن بعض طرق المعادلة قد تعطي قيم أكبر من الطرق الأخرى، وبالتالي يجب استخدام طريقة واحدة من طرق المعادلة أثناء معادلة درجات أكثر من مجموعة طُبقت عليها هذه الصور الاختبارية، واستخدام الطريقة التي تعطي قيماً أكثر دقة.
2- يمكن الاعتماد على نظرية الاستجابة للمفردة في بناء بنوك الأسئلة واستخراج صور اختبارية متعددة منه، لتطبيقها على التلاميذ الملتحقين بنفس الصف الدراسي ثم معادلة درجاتها بعد التطبيق لتحقيق العدالة والمساواة بين التلاميذ والمحافظة على سرية الاختبار.
3- توجيه نظر القائمين على المؤسسات التعليمية إلى ضرورة الاهتمام باستخدام نظرية الاستجابة للمفردة وتطبيقاتها في مجال بناء الاختبارات التحصيلية ومعادلة درجاتها وبناء بنوك الأسئلة والاختبارات المواءمة المحوسبة، وتوفير البرامج الحاسوبية اللازمة لذلك والتدريب عليها.

دراسات وبحوث مقترحة:

في ضوء نتائج البحث يمكن تقديم المقترحات البحثية التالية:

1- فاعلية بعض نماذج نظرية الاستجابة للمفردة في التنبؤ بالتلاميذ الموهوبين ذوي صعوبات التعلم.

2- أثر اختلاف بارمترات المفردات الاختبارية (الصعوبة، التمييز، التخمين) بين صورتي الاختبار في القيم الناتجة من عملية المعادلة ودقتها.

3- المعادلة الرأسية لاختبار (كاتل) للذكاء باستخدام نظرية الاستجابة للمفردة لدى تلاميذ المرحلة الإعدادية والثانوية.

قائمة المراجع

أولاً: المراجع العربية

أحمد أبو الحسن (2014). فاعلية طرق معادلة صورتي اختبار مبني في ضوء نظرية الذكاء الناجح لستيرنبرج وفقاً لنظرية القياس التقليدية. *المجلة العلمية لكلية رياض الأطفال، جامعة المنصورة*،*1*(1)، 447- 503.

أحمد بديوي (2019، إبريل 27-28). *الفاعلية النسبية لطرق معادلة صورتين من اختبار الاستعداد المدرسي (I SAT) في مادة الرياضيات في ضوء نظرية الاستجابة للمفردة IRT*. المؤتمر الدولي الخامس لكلية التربية بنين بالقاهرة، جامعة الأزهر "التعليم قبل الجامعي الأزهري والعام وتحديات القرن الواحد والعشرين- الواقع والمأمول"،ج2، 695- 821.

أحمد عاشور (2019). *أثر نسبة الفقرات المفقودة في الجذع المشترك وطريقة التعامل معها على دقة المعادلة العمودية*. (رسالة دكتوراه غير منشورة)، كلية التربية، جامعة اليرموك.

آر إيالا (2017). *النظرية والتطبيق في نظرية الاستجابة للفقرة*. (ترجمة: عبدالله الكيلاني وإسماعيل البرصان)، الرياض: دار جامعة الملك سعود للنشر.(نشر العمل الأصلي 2009)

إسماعيل الوليلي (2002). *دراسة سيكومترية مقارنة لبعض نماذج الاستجابة للمفردة في انتقاء مفردات الاختبارات مرجعية المحك*. (رسالة دكتوراه غير منشورة)، كلية التربية، جامعة الأزهر.

أمل الزعابي (2020).معادلة كتيبات اختبار الدراسة الدولية لقياس مهارات القراءة (PIRLS2011) بسلطنة عمان باستخدام نظرية الاستجابة للمفردة. *مجلة العلوم التربوية، كلية التربية جامعة قطر*، (12)، 68- 93.

إياد حمادنة (2011). الكشف عن الارتباط الموضعي بين أزواج فقرات الاختبار باستخدام مؤشر Q3. *مجلة العلوم التربوية والنفسية 12*(1)، 40-68.

أيمن غانم (2015). *بناء بطارية اختبارات لقياس الذكاء الناجح لدى طلاب المرحلة الثانوية في ضوء نظرية الاستجابة للمفردة "دراسة سيكومترية"*. (رسالة دكتوراه غير منشورة)، كلية البنات للآداب والعلوم والتربية، جامعة عين شمس.

إيهاب السيد (2012).*معادلة الاختبارات وعلاقتها ببعض المتغيرات السيكومترية، دراسة محاكاة وتطبيقية*. (رسالة دكتوراه غير منشورة)، كلية التربية، جامعة عين شمس.

خالد العنزي (2014). *أثر طول الجذع وحجم العينة في دقة معادلة درجات صورتي اختبار في الرياضيات وفق تصميم الجذع المشترك*. (رسالة دكتوراه غير منشورة)، معهد الدراسات التربوية، جامعة القاهرة.

ديالا العلي ورمضان درويش (2017). مقارنة فاعلية طريقتي براون–هولد والمئينيات باستخدام الجذع المشترك بين صورتي اختبار تحصيلي في الرياضيات. *مجلة جامعة البعث، 39*(43)، 119– 148.

راشد الدوسري (2001). معادلة الاختبارات مفهومها، وطرقها، ومشكلات تطبيقها. *مجلة العلوم التربوية والنفسية، 2*(4)، 107– 141.

رائد المدانات (2010). مقارنة فاعلية طريقة تكر الخطية وطريقة ليفين في معادلة الاختبارات عند استخدام التصميم القائم على جذع مشترك ومجموعات غير عشوائية. *مجلة مؤتة للبحوث والدراسات، سلسلة العلوم الإنسانية والاجتماعية، 25*(7)، 11– 53.

رائد المدانات (2012). مقارنة فاعلية طريقتي معادلة العلامات الحقيقية والمشاهدة في معادلة الاختبارات باستخدام جذع مشترك ومجموعات غير متكافئة. *مجلة العلوم التربوية والنفسية، 13*(2)، 365– 394.

رشدي طعيمة (2004). *تحليل المحتوى في العلوم الإنسانية*. القاهرة: دار الفكر العربي.

صلاح الدين علام (2005). *نماذج الاستجابة للمفردة الاختبارية أحادية البعد ومتعددة الأبعاد وتطبيقاتها في القياس النفسي والتربوي*. القاهرة: دار الفكر العربي.

عيسى الحربي (2009).*أثر تمثيل الفقرات للمحتوى ونسبة الفقرات المشتركة وطرق توزين الفقرات على دقة معادلة درجات الصور للاختبار عندما يكون التصميم المستخدم هو الجذع المشترك (الفقرات المشتركة)*. (رسالة دكتوراه غير منشورة)، كلية الدراسات العليا، الجامعة الأردنية.

مايا بركات (2010).*أثر تصميمات المعادلة ومتوسط صعوبة الاختبارات وتوزيع القدرة على معادلة درجات الاختبارات متعددة الأبعاد باستخدام نظرية الاستجابة للمفردة*. (رسالة دكتوراه غير منشورة)، معهد الدراسات التربوية، جامعة القاهرة.

مايسة أبومسلم (2010). معادلة صورتي اختبار توني للذكاء غير اللفظي باستخدام طرق مختلفة للمعادلة في ضوء بعض المتغيرات المؤثرة على نتائجها. *المجلة المصرية للدراسات النفسية، 20*(66)، 371– 411).

محمد عبدالوهاب (2010). *استخدام نماذج الاستجابة للمفردة الاختبارية في تدريج مفردات بعض الاختبارات المعرفية*. (رسالة دكتوراه غير منشورة)، كلية التربية، جامعة المنيا.

منار طومان (2006). *فاعلية استخدام نظرية الاستجابة للمفردة (IRT) في عملية معادلة درجات الاختبارات متعددة الأبعاد والمتغيرات المؤثرة فيها*. (رسالة دكتوراه غير منشورة)، معهد الدراسات التربوية، جامعة القاهرة.

هبة ضعضع (2020). *أثر طرائق معالجة القيم المفقودة في تقدير معالم بعض نماذج نظرية الاستجابة للمفردة*. (رسالة دكتوراه غير منشورة)، كلية التربية، جامعة حلب.

ياسر حسن (2019). أثر اختلاف طريقة المعادلة وطرق تقدير الدرجات وقواعد صياغة الفقراتعلى دقة معالم الفقرات وقدرات الأفراد في ضوء القياس الكلاسيكي والنموذج اللوجيستي ثلاثي البارامتر. *مجلة كلية التربية– جامعة أسيوط، 35* (7)، 352-434.

يوسف المحروق (2011). *مقارنة طريقة كيرنيل والمئينيات وطرق نظرية الاستجابة الفقرة عند استخدام تصميم الفقرات المشتركة في دقة معادلة درجات الاختبارات متعددة الحدود*. (رسالة دكتوراه غير منشورة)، كلية الدراسات العليا، الجامعة الأردنية.

يوسف المحروق (2016). أثر حجم العينة وطول الاختبار في دقة معادلة درجات الاختبارات متعددة الحدود باستخدام طريقة كيرنيل. *مجلة العلوم التربوية والنفسية، 17*(3)، 201- 228.

يوسف المحروق (2019). فاعلية العلامات المشاهدة وكيرنيل في معادلة درجات الاختبارات. *مجلة دراسات العلم التربوية، الجامعة الأردنية*، 46، 204-223.

ثانيا: المراجع العربية مترجمة:

Ahmed A. H., (2014). The effectiveness of two-form equation methods A test built in the light of Sternberg's successful intelligence theory according to the traditional measurement theory. *The Scientific Journal of the Faculty of Kindergarten, Mansoura University*, 1(1), 447-503.

Ahmed B., (2019, April 27-28). The relative effectiveness of the methods of equivalence of two forms of the School Readiness Test (SAT I) in mathematics in the light of the IRT response theory. *The Fifth International Conference of the Faculty of Education for Boys in Cairo, Al-Azhar University, "Al-Azhar and General Pre-University Education and the Challenges of the Twenty-first Century - Reality and Expectations",* Volume 2, 695-821.

Ahmed A., (2019). *The effect of the percentage of missing vertebrae in the common trunk and the way to deal with it on the accuracy of the vertical equation*. (Unpublished PhD thesis), College of Education, Yarmouk University.

Ayala R., (2017). *Theory and practice in paragraph response theory.* (Translated by: Abdullah Al-Kilani and Ismail Al-Bursan), Riyadh: King Saud University Press. (Publishing the original work 2009)

Ismael El., (2002). *A comparative psychometric study of some response models for the item in the selection of items for the benchmark reference tests*. (Unpublished PhD thesis), Faculty of Education, Al-Azhar University.

Amal Al., (2020). The equivalence of the booklets of the international study test to measure reading skills (PIRLS2011) in the Sultanate of Oman using the vocabulary response theory. *Journal of Educational Sciences, College of Education, Qatar University*, (12), 68-93.

Iyad H., (2011). Detection of positional correlation between pairs of test items using the Q3 index. *Journal of Educational and Psychological Sciences* 12(1), 40-68.

Ayman G., (2015*). Building a battery of tests to measure the successful intelligence of secondary school students in the light of the theory of response to the single "psychometric study"*. (Unpublished PhD thesis), Girls' College of Arts, Sciences and Education, Ain Shams University.

Ihab El., (2012). *The equation of tests and their relationship to some psychometric variables, a simulation and applied study.* (Unpublished PhD thesis), Faculty of Education, Ain Shams University.

Khaled Al., (2014). *The effect of stem length and sample size on the accuracy of the equation scores for two math test images according to the common stem design*. (Unpublished PhD thesis), Institute of Educational Studies, Cairo University.

Diala Al. & Ramadan D., (2017). Comparison of the effectiveness of the Brown-Hold and percentile methods using the common stem between the two forms of an achievement test in mathematics. *Al-Baath University Journal,* 39(43), 119-148.

Rashid Al., (2001). Equation of tests its concept, methods, and problems of its application. *Journal of Educational and Psychological Sciences*, 2(4), 107-141.

Rayid Al., (2010). Comparison of the effectiveness of the linear Tucker method and Levine's method in equation tests when using a design based on a common trunk and non-random groups. *Mutah Journal for Research and Studies, Humanities and Social Sciences Series*, 25(7), 11-53.

Rayid Al., (2012). Comparing the effectiveness of the two methods of equivalence of real and observed scores in equivalence of tests using common trunk and unequal groups. *Journal of Educational and Psychological Sciences*, 13(2), 365-394.

Rushdi Taima (2004). *Content analysis in the humanities*. Cairo: Arab Thought House.

Salah El. A., (2005). *One-dimensional and multi-dimensional test response models and their applications in psychological and educational measurement*. Cairo: Arab Thought House.

Issa Al., (2009). *The effect of paragraph representation of the content, percentage of common paragraphs, and methods of paragraph weighting on the accuracy of the equation scores for the test images when the design used is the common stem (common items)*. (Unpublished PhD thesis), College of Graduate Studies, University of Jordan.

Maya B., (2010). *The effect of equation designs, average test difficulty, and power distribution on equivalence of the scores of multidimensional tests using singular response theory.* (Unpublished PhD thesis), Institute of Educational Studies, Cairo University.

Maysa A., (2010). Equation of two images of Tony's nonverbal intelligence test using different methods of equation in light of some variables affecting its results. *The Egyptian Journal of Psychological Studies*, 20 (66), 371-411.

Mohamed A., (2010). *The use of test item response models in grading the vocabulary of some cognitive tests*. (Unpublished PhD thesis), Faculty of Education, Minia University.

Manar T., (2006). *The effectiveness of using IRT (Individual Response Theory) in the process of evaluating the scores of multi-dimensional tests and the variables affecting them.* (Unpublished PhD thesis), Institute of Educational Studies, Cairo University.

Heba D., (2020). *The effect of missing values processing methods in estimating parameters of some response theory models*. (Unpublished PhD thesis), College of Education, University of Aleppo.

Yasser H., (2019). The effect of the difference in the equation method, the methods of estimating the grades and the rules for formulating the items on the accuracy of the paragraph's parameters and the abilities of the individuals in the light of the classical measurement and the three-parameter logistic model. *Journal of the Faculty of Education - Assiut University*, 35 (7), 352-434.

Youssef Al., (2011*). Comparison of kernel and percentile method and paragraph response theory methods when using common paragraph design in polynomial test score equivalency accuracy*. (Unpublished PhD thesis), College of Graduate Studies, University of Jordan.

Youssef Al., (2016). The effect of sample size and test length on the accuracy of polynomial test scores equation using Kernel method. *Journal of Educational and Psychological Sciences,* 17(3), 201-228.

Youssef El., (2019). The effectiveness of the observed and kernel marks in the equivalency of test scores. *Journal of Educational Science Studies, University of Jordan*, 46, 204-223.

ثالثاً: المراجع الأجنبية

Arikan, C. & Gelbal, S. (2018). A comparison of traditional and kernel equating methods. *International Journal of Assessment Tools in Education, 5*(3). 417- 427.

Asiret, S. & Sunbul, S. (2016). Investigating test equating methods in small samples through various factors. *Educational Sciences: Theory& Practice, 16*(2), 647- 668.

Baker, F. (2001): *The Basics of Item Response Theory*. (2nd ed.), USA: The ERIC Clearinghouse on Assessment and Evaluation.

Born, S., Fink, A., Spoden, C. & Frey, A. (2019). Evaluating Different equating setups in the continuous item pool calibration for computerized adaptive testing. *Frontiers in Psychology. 10*, 1-14.

Castaneda, R. (2017). *A Model-building approach to assessing Q3 values for local item dependence*. (Unpublished doctoral dissertation), University of California, Merced.

Chen, F., Huang, X.& Gregor, D. (2009). *Equating or linking: Basic concepts and a case study*. Washington: Center for Applied Linguistics(CAL).

Dawber, T. (2004). *Robustness of Lord Formulas for Item Difficulty And Discrimination Conversions Between Classical And Item Response Theory Models*. (Unpublished doctoral dissertation), University Of Alberta.

Dorans, N., Moses, T. & Eignor, D. (2010). *Principles and practices of test score equating. New Jersey*: Educational Testing Service (Research Report: ETS RR-10-29).

Embretson, S.,& Reise, S.(2000). *Item Response theory for Psychologists*. New Jersey: Lawrence Erlbaum Associates.

Felan, G. (2002, February14-16). *Test equating: Mean, linear, Equipercentile, and Item response theory*. Paper presented at the annual meeting of the southwest educational research association. Austin, TX, 1-24.

Fitzpatrick, A. & Yen, W. (2001). The effects of test length and sample size on the reliability and equating of tests composed of constructed- response items. *Applied Measurement in Education, 4*(1), 31-57.

Georgiev, N. (2008). Item analysis of C, D and E series from Raven's standard progressive matrices with item response theory two parameter logistic model. *Europe's Journal of Psychology*, 8, 1-17.

Gleason, J. (2008). An evaluation of mathematics competitions using item response theory. *Notices of The Ams, 55*(1), 8-15.

Gruijter, D. & Kamp, L. (2008). *Statistical test theory for the behavioral sciences*. London: Taylor& Francis Group.

Kabasakal, K. & Kelecioglu, H. (2015). Effect of differential item functioning on test equating. *Educational Sciences: Theory& Practice, 15*(5), 1229- 1246.

Keller, L. & Keller, R. (2011). The long-term sustainability of different item response theory scaling methods. *Educational and Psychological Measurement, 71*(2), 362–379.

Kilmen, S.& Demirtasli, N. (2012). Comparison of test equating methods based on item response theory according to the sample size and ability distribution. *Procedia- Social and Behavioral Sciences*, *46*, 130- 134.

Kim, D., Choi, S., Lee, G. &Um, K. (2008). A comparison of the common-Item and random-groups equating designs using empirical data. *International Journal of Selection and Assessment. 16*(2), 83-92.

Kim, J. (2007). *A comparison of calibration methods and proficiency estimators for creating IRT vertical scales*. (Unpublished doctoral dissertation), The University Iowa.

Kolen, M.& Brennan, R. (2014). *Test equating, Scaling, and Linking.* (3rd ed), New York: Springer.

Langer, M. & swanson, D. (2010) practical considerations in equating progress test. *Medical Teacher*, (32), 509- 512.

Lee, E. (2013). *Equating multidimensional test under a random groups design: A comparison of various equating procedures*. (Unpublished doctoral dissertation), the University Of Lowa.

Livingston, A. (2014). *Equating test scores (without IRT)*. (2rded) Princeton, NJ: Educational Testing Service.

Magno, C. (2009). Demonstrating the Difference between Classical Test Theory and Item Response Theory Using Derived Test Data, *The International Journal of Educational and Psychological Assessment, 1*(1), 1-11.

Michaelides, M. (2003, April 21-25). *Sensitivity of IRT equating to behavior of test equating items*. Paper presented at the Annual Meeting of the American Educational Research Association. Chicago, IL, 1-19.

Michaelides,M. (2006). *Effects of misbehaving common items on aggregate scores and an application of mantel-haenszel statistic in test equating*. University of California, Los Angeles: Center for the Study of Evaluation (CSE).

Ozdemir, B. (2017). Equating TIMSS Mathematic subtests with nonlinear equating methods using NEAT design: Circle- Arc equating approaches. *International Journal of Progressive Education, 13*(2), 116-132.

Pang, X. Madera, E. Radwan, N. & Zhang, S. (2010). *A comparison of four test equating methods*. Ontario: Report prepared for the Education Quality and Accountability Office (EQAO).

Reeve, B. (2002). *An Introduction to Modern Measurement Theory, Outcomes Research Branch*, National Cancer Institute: Applied Research Program.

Reise, S.& Waller, N.(2003). How many IRT parameters does it take to model psychopathology items?. *Psychological Methods, 8*(2), 164- 184.

Ryan, J.& Brockmann, F. (2009). *A Practitioner's Introduction to equating with primers on classical test theory and item response theory*. Washington: Council of Chief State School Officers (CCSSO).

Shyu, C. (2001). *Estimating error indexes in estimating proficiencies and constructing confidence interval in item response Theory*. (Unpublished doctoral dissertation), University of Lowa.

Song, T. (2009). Investigating different item response models in equating the examination for the certificate of proficiency in English. *Spaan Fellow Papers in Second or Foreign Language Assessment, 7*, 85- 98.

Stone, C. & Hansen, M. (2000). The Effect of Errors in Estimating Ability on Goodness-of-Fit Tests For IRT Models. *Educational and Psychological Measurement, 60*(6), 974-991.

Uysal, I. & Kilmen, S. (2016). Comparison of item response theory test equating method for mixed format tests. *International Online Journal of Educational Sciences, 8*(2), 1-11.

Von Davier, A. & Wilson, C. (2007). IRT True-Score test equating (A guide through assumption and applications). *Educational and Psychological Measurement, 67*(6), 940-957.

Wang, S. Zhang, M. & You, S. (2020). A Comparison of ITR observed score Kernel equating and several equating methods. *Frontiers in Psychology, 11*(308), 1-19.

Zhonghua, Z. (2010). *Comparison of different equating methods and an application to link test- based tests*. (Unpublished doctoral dissertation), the Chinese University of Hong Kong.