أثر طرق تقدير القدرة للأفراد على دقة معادلة الاختبارات في نظرية الاستجابة للمفردة

The Effect of Estimation of Individuals Ability Methods on the Accuracy of Equating Test in Item Response Theory

إعداد

د/ أحمد سالم سعيد القرني

Dr. Ahmed Salem Saeed Al-Qarni

كلية التربية – قسم علم النفس – جامعة الملك سعود

Doi: 10.21608/jasep.2022.265536

استلام البحث : ٢ / ٥ /٢٠٢٢

قبول النشر: ١٢ / ٥ / ٢٠٢٢

القرني ، أحمد سالم سعيد (٢٠٢٢). أثر طرق تقدير القدرة للأفراد على دقة معادلة الاختبارات في نظرية الاستجابة للمفردة. ***المجلة العربية للعلوم التربوية والنفسية،*** المؤسسة العربية للتربية والعلوم والآداب، مصر، مج (٦)،ع(٣٠) أكتوبر ، ١ – ٢٢.

http://jasep.journals.ekb.eg

أثر طرق تقدير القدرة للأفراد على دقة معادلة الاختبارات في نظرية الاستجابة للمفردة

المستخلص:

هدفت الدراسة إلى التعرف على أثر طرق تقدير القدرة للأفراد على دقة معادلة الاختبارات في نظرية الاستجابة للمفردة، وتكونت عينة الدراسة من (٢٣٧٢) من تلاميذ الصفوف العليا بالمرحلة الابتدائية، وكذلك من (٣٢٧٦) من طلاب المرحلة المتوسطة بمدينة جدة. ولتحقيق هدف الدراسة تم استخدام اختبار أوتيس لينون للقدرات بصورتيه (الابتدائي الثاني والمتوسط) وتتكون كل صورة من عدد (٨٠) مفردة اختبارية تم تطبيقها على عينة استطلاعية بحجم (٢٠٨) طلبًا، وتم اختيار عدد (١٠) مفردات من كل صورة بواقع (٢٠) مفردة تمثل الجذع المشترك. ومن ثم طبقت الأدوات على العينة وتم الحصول على بيانات كانت نسبة القيم المفقودة (٥.٥%) للمرحلة الابتدائية و (٥.٦) للمرحلة المتوسطة. وتم التقدير للقدرة من خلال أربع طرق للتقدير وهي طريقة الأرجحية العظمى (MLE) و التوقع البعدي (EAP) والأرجحية العظمى الموزونة (WML) و الطريقة البايزية(BME) وتم استخدام برنامج Xcalibre في التقدير وكذلك للتحقق من مطابقة البيانات للنموذج. وللكشف عن دقة المعادلة العمودية في نظرية الاستجابة للمفردة باختلاف طرق التقدير تم استخدام محك الخطأ المعياري. وقد أظهرت أن طريقة التقدير بالأرجحية العظمى الموزونة (WML) أفضل النتائج وأقل خطأ معياري من خلال برنامج Xcalibre.

Abstract:

The current study aimed to identify the effect of methods of estimation of individuals ability on accuracy of equating test in Item Response Theory. The study sample consisted of students with higher grades from both primary and intermediate schools in Jeddah. To this end, the Otis Lennon Qualification Test in its two forms (second primary and intermediate) was used. Each form composed of (80) items that was administered to a pilot study consisting of (208) students, and (10) items were selected from each form from a total of (20) items that represent the common items. Then the tools were applied to the sample and data were obtained in which the percentages of missing values were (5.5%) for the primary stage and (5.6) for the intermediate school. The estimation of the ability was based on four methods of estimation as follow: Maximum Likelihood Estimation (MLE) method, Expected A Posterior (EAP), Bayesian Method Estimation (BME) and weighted maximum likelihood (WML).

ISSN: 2537-0464 eISSN : 2537-0472

Xcalibre was used to estimate and verify data conformity to the model. To reveal the accuracy of the vertical equation in item response theory with different treatment methods for estimation methods, the standard error criterion was used. The results showed that the weighted maximum likelihood estimation method (WML) provided the best results and the least standard error in the presence of missing values and through the Xcalibre program.

المقدمة:

ظهر القياس النفسي كأحد المواضيع المهمة و الحديثة التي شهدت تطورًا كبيرًا في بدايات القرن الماضي، فقد انصب اهتمام الباحثين في تلك الفترة في قياس الظواهر النفسية والحصول على بيانات دقيقة عن تلك الظواهر موضوع القياس، وكان لإسهامات الرواد في تلك الحقبة كجالتون(١٩٠٨) والفرد بينيه (١٩١١)وجيمس كاتل(١٩٤٤) ولويس تيرمان(١٩٥٦) الدور الكبير ونقطة انطلاق جوهرية في انتشار الاختبارات على نطاق واسع فيما بعد.

وامتدادًا لهذه الجهود فقد شهدت أعوام التسعينيات الميلادية استمرارية وتوسعًا واهتمامًا متزايدًا بالقياس النفسي والذي كان جليًا في أعوام الثمانينيات. ويتضح هذا النمو في تطوير اختبارات جديدة بعضها يمثل مداخل تتميز بالحداثة (انستازي ويورينا،٢٠١٥) كما شهد القرن العشرون طفرة كبيرة في تطوير وبناء الاختبارات النفسية بمختلف أنواعها ونتيجة لهذه الطفرة ظهرت آلاف الاختبارات في مجالات متعددة (الطريري،٢٠١٤).

وكان من الضروري عند بناء هذه الاختبارات أن يأخذ بعين الاعتبار أن تكون هذه الصور الاختبارية عادلة وتساعد في اتخاذ القرارات التي بني من أجلها الاختبار، وبناء عليه فقد ظهرت بعض المشكلات في إعداد تلك الاختبارات من بينها ما يحصل عندما يتعرض مفحوص للاختبار مرتين وبشكل متتالٍ، فقد يحدث أن يحصل المفحوص على درجة أعلى في الاختبار الثاني. إن إحدى التفسيرات لهذا التحسن هو أن هذه الدرجة تعكس مستوى إنجاز المفحوص ولكن هناك تفسيرًا آخر واحتمال أنه قد تم تضخيم درجة الطالب فقط، لأن الطالب تعرض لعناصر الاختبار مرتين وبالتالي كان مقدار التحسن هو نتيجة لعوامل أخرى لا تعود لمستوى القدرة الحقيقية للمختبر ((Kolen&Brennan,2014.

وعليه فقد ظهرت الحاجة لبناء عدة صور اختبارية بحيث تساهم في معالجة تلك المشكلة، ولكن سرعان ما برزت مشكلة أخرى تكمن في استخدام نماذج اختبار متعددة وفي أزمنة مختلفة، فعلى سبيل المثال: قد يتقدم طالبان بطلب للحصول على نفس المنحة الجامعية التي تعتمد جزئياً على نتائج الاختبارات. وقد يختبر الطالبان الاختبار في تواريخ اختبار مختلفة، ويكتسب الطالب الأول درجة أعلى من الطالب الثاني. إن إحدى التفسيرات المحتملة لتفوق الطالب الأول أنه يمتلك مستوى قدرة أعلى من الطالب الثاني، ومع ذلك إذا أخذ الطالب

ISSN: 2537-0464 eISSN : 2537-0472

الأول نموذج اختبار أسهل من الطالب الثاني عندها سيكون للطالب الأول ميزة غير عادلة على الطالب الثاني وفي هذه الحالة قد يكون الفرق في الدرجات بسبب الاختلافات في صعوبة نماذج الاختبار وليس لمستوى قدرات الطلاب ((Livingston,2014.

وسعيًا لتجنب هذه المشكلة ظهرت الحاجة إلى معادلة الاختبارات. فإذا تمت معادلة نماذج الاختبار بنجاح، فإن الفرق في الدرجات المتعادلة لكل من الطالب الأول والطالب الثاني لا يمكن أن يعزى إلى أن إحدى صور الاختبار كانت أسهل من الصورة الأخرى (Angof & Cowell,1986).

ونظرًا لأهمية معادلة صور الاختبار وتحقيقًا للتكافؤ بين المفحوصين فقد ظهرت عدة صيغ للمعادلة منها ما هو مرتبط بنظرية القياس التقليدية والتي تعتبر من الطرق غير المرغوب فيها وذلك بسبب اخفاقها في تحقيق شروط المعادلة كالعدل والمساواة والتماثل واللاتغاير(Hambleton & Swaminthan,1985). ولذلك فإن المعادلة وفقًا لنظرية الاستجابة للمفردة تحل الكثير من المشاكل التي عجزت عنها نظرية القياس التقليدية بشرط مطابقة البيانات للنموذج. (Kolen,1981).

مشكلة الدراسة:

وفقًا لنظرية الاستجابة للمفردة فإن الهدف الأول لتطبيق الاختبار هو تحديد مكان الشخص على مقياس القدرة (بيكر،٢٠١٠) وتعد هذه الخطوة من الخطوات الأساسية على الرغم من تباين وجهات النظر حول الطرق المثلى في تقدير قدرات الأفراد (الرحيل والدرابسة، ٢٠١٤).

كما أظهرت عدة دراسات سابقة من بينها دراسة جلاس (Glass,2005) ودراسة جاري وفيرمنت (Garre & Vermunt,2006) والشريفين (٢٠١٢) أن هناك تباينًا حول الطرق المثلى في دقة تقدير معالم الفقرات والأفراد، وذلك يعود إلى أن هناك عدة طرق لتقدير معالم الفقرات والأفراد منها ما يعتمد على طريقة الأرجحية العظمى والبعض الآخر يعتمد على الطرق البابيزية(أيالا،٢٠١٧).

وبناء على ما سبق ونظرًا لعدم وجود دراسات سابقة في حدود علم واطلاع الباحث من خلال محركات البحث العالمية google scholar وEric وكذلك المكتبة الرقمية السعودية ودور النشر العربية تناولت طرق تقدير القدرة للأفراد على دقة معادلة الاختبارات، فإن هذه الدراسة تسعى للكشف عن أفضل هذه الطرق على دقة معادلة الاختبارات، وسيتم التركيز على أربع طرق للتقدير وهي:

١- طريقة الأرجحية العظمى (Maximum Likelihood Estimation (MLE:
٢- طريقة التوقع البعدي (Expected a posterior (EAP:
٣- طريقة الأرجحية العظمى الموزونة : Weighted Maximum Likelihood
٤- الطريقة البييزية : (Bayesian Method Estimation (BME

ISSN: 2537-0464 eISSN : 2537-0472

وسيتم التحقق من دقة المعادلة من خلال حساب الخطأ المعياري في المعادلة وهو أحد الطرق المستخدمة في التحقق من دقة المعادلة في التقدير (Peterson& Kolen& Hoover, 1989). وستتم معادلة قيم تقديرات قدرات الأفراد في صورة اختبار اوتيس لينون الابتدائي الثاني وفقًا لقيم تقديرات الأفراد في صورة اختبار اوتيس لينون المتوسط وذلك باستخدام المعادلة المشار إليها في رايت وماسترز (Wright & Masters, 1982). وتفترض هذه المعادلة أن الاختبارين يقيسان القُدرة نفسها، وعند قياس معلم القُدرة، ومعالم الفقرات في الوقت نفسه لكلا الاختبارين تكون هذه المعالم على نفس التدريج. وعند تقدير معلم القُدرة، والصعوبة والتمييز لكل اختبار بشكل منفرد فإنه ينتج علاقة خطية بين تدريج القُدرة للاختبارين. وتتم عملية المعادلة بين الاختبارين بتحديد العلاقة الخطية بين علامات القُدرة المقدرة (Hambleton & Swaminthan & Rogers, 1991).

وتتلخص مشكلة الدراسة في الإجابة عن السؤال الرئيسي التالي:

هل يوجد تأثير لكل من طرق معالجة القيم المفقودة على دقة معادلة الاختبارات وفقًا لنظرية الاستجابة للمفردة؟

ويتفرع من السؤال الرئيسي الأسئلة الفرعية التالية:

١- هل يوجد تأثير لطرق تقدير القدرة على الأخطاء المعيارية في معادلة الاختبار وفقًا لنظرية الاستجابة للمفردة؟

٢- هل توجد فروق بين طرق تقدير القدرة في نظرية الاستجابة للمفردة وفقًا للخطأ المعياري في معادلة الاختبارات؟

أهداف الدراسة:

تهدف الدراسة الحالية إلى التعرف على:

١- أثر طرق تقدير القدرة على الأخطاء المعيارية في معادلة الاختبار وفقًا لنظرية الاستجابة للمفردة.

٢- أثر الفروق بين طرق تقدير القدرة في نظرية الاستجابة للمفردة وفقًا للخطأ المعياري في معادلة الاختبارات.

أهمية الدراسة:

الأهمية النظرية:

تكمن أهمية الدراسة الحالية في أهمية الموضوع الذي تتناوله فالحاجة إلى معادلة الاختبارات في ظل تعدد الصور المختلفة من الاختبارات أصبحت ضرورة ملحة لدى العديد من المؤسسات، كما أن معادلة الاختبارات وفقًا لنظرية الاستجابة للمفردة تعتبر من الموضوعات الحديثة في القياس والتي لم تأخذ حقها من البحث والدراسة خاصة على مستوى الوطن العربي، وستعمد الدراسة الحالية إلى دعم القاعدة النظرية للبحوث والدراسات المتعلقة بتفسير نتائج معادلة الاختبارات وفقًا لنظرية الاستجابة للمفردة. كذلك ستسعى هذه الدراسة إلى توسيع الفهم حول طرق المعادلة وفقًا لنظرية الاستجابة للمفردة. كما ستساهم

ISSN: 2537-0464 eISSN : 2537-0472

الدراسة الحالية في زيادة المعرفة بطريقة الحصول على البيانات من المجموعات غير المتكافئة.

الأهمية التطبيقية:

تكمن الأهمية العملية لهذه الدراسة في إمكانية الاستفادة من نتائج وتوصيات الدراسة الحالية وذلك في الكشف عن أفضل الطرق التي يمكن أن يستخدمها الباحث في التقدير في معادلة الاختبارات وفقًا لنظرية الاستجابة للمفردة. كذلك يمكن الاستفادة من نتائج الدراسة على مستوى المؤسسات التي تقدم عدة صور مختلفة من الاختبارات مثل: اختبارات القبول في الجامعات أو الاختبارات التي تقدم من المراكز المتخصصة في بناء الاختبارات وذلك بهدف بناء اختبارات متكافئة تحقق العدالة بين المتقدمين للاختبار.

حدود الدراسة:

الحدود الموضوعية:

تتحدد الدراسة الحالية بموضوعها "أثر طرق تقدير القدرة على دقة معادلة الاختبارات في نظرية الاستجابة للمفردة". وذلك باستخدام المعادلة العمودية ونموذج تصميم الفقرات المشتركة لمجموعات غير المتكافئة.

الحدود المكانية:

المدارس الحكومية بنين التابعة لإدارة تعليم جدة.

الحدود الزمانية:

يتم تطبيق الدراسة في الفصل الأول من العام الدراسي١٤٤١ / ١٤٤٢هـ.

الحدود البشرية:

طلاب الصفوف العليا من المرحلة الابتدائية (الرابع والخامس والسادس) وطلاب المرحلة المتوسطة (الأول والثاني والثالث) من المدارس الحكومية بنين بمدينة جدة.

مصطلحات الدراسة:

١- معادلة الاختبارات Equating Test:

هي مجموعة من الأساليب المستخدمة لتعديل تقديرات مواقع الأفراد على مقاييس مختلفة وذلك لوضع هذه التقديرات على مقايس واحد (أيالا،٢٠١٧).

التعريف الإجرائي:

هي تعديل تقدير مواقع الأفراد على صورتي اختبار أوتيس – لينون الصورة الابتدائي الثاني بصورة المتوسط لجعل هذه التقديرات على مقياس واحد.

٢- تقدير القدرة Ability Estimate:

عملية تكرارية تستخدم لحساب احتمالية الاستجابة الصحيحة للفرد على مفردات الاختبار للحصول على تقدير أولي لقدرة المفحوص، ومن ثم يتم التعديل على هذا التَّقدير بحيث يصبح التغيير في موقع المفحوص أقل ما يمكن (بيكر، ٢٠١٠).

ISSN: 2537-0464 eISSN : 2537-0472

التعريف الإجرائي:

العملية التي تتم فيها معايرة موقع المفحوص على متصل القدرة من خلال: طرق التَّقدير بالأرجحية العظمى (MLE)، وطريقة التوقع البعدي (EAP)، وطريقة الأرجحية العظمى الموزونة (WML)، والطّريقة البايزية (BME).

٣- دقة المعادلة Equation accuracy:

هي عملية إجراء تعديل إحصائي على درجات صورة واحدة من الاختبار ؛وجعل تلك الدرجات مكافئة بطريقة ما لدرجات صورة أخر من الاختبار ذاته وذلك باستخدام الخطأ المعياري. (Hoover & Kum,1995)

التعريف الإجرائي: الأسلوب الذي سيتم من خلاله الحكم على فاعلية القدرات المعادلة على صورتي اختبار أوتيس لينون الابتدائي المستوى الثاني والمتوسط.

الإطار النظري :

تعتبر نظرية الاستجابة للمفردة نظرية اختبار احتمالية، قائمة على النماذج، تتبع من نمط ردود الممتحنين على مجموعة من عناصر الاختبار (Price،2016). وهذا يحدد العلاقة بين الدرجة المشاهدة التي يحصل عليها الفرد عند إجابته على المفردة والسمة الكامنة التي يمتلكها والتي بناءً عليها يختار الفرد إجابته(Hambleton& Swaminthan,1985)، وعلى الرغم أن هذه العلاقة لا تفسر لماذا يعطي شخص ما استجابة معينة لفقرة اختبارية ، أو كيف يقرر الشخص بماذا يجيب (falmagne,1989) إلا أنه يمكن ربط هذه العلاقة بين السمة الكامنة التي يمتلكها الفرد واستجابته على فقرات الاختبار من خلال منحنى خصائص الفقرة والذي يعتبر اللبنة الأساسية في نظرية الاستجابة للمفردة حيث تعتمد عليه بقية مكونات النظرية (بيكر، ٢٠١٠).

ويمكن التعبير عن هذه العلاقة من خلال عدة نماذج رياضية تنقسم إلى قسمين:

١- ثنائية التدريج وتتمثل في النموذج أحادي المعلم، وثنائي المعلم، وثلاثي المعلم، على النحو الآتي:

أ- النموذج الأحادي المعلم:

قدم هذا النموذج العالم الدنماركي جورج راش (Rasch,1960) وتناول راش تحليل البيانات للاختبارات وذلك من منظور نظرية الاحتمالات. على الرغم من كونه انطلق من إطار مرجعي مختلف إلا أنه في النهاية توصل إلى المنحنى المميز للمفردة والذي يعد نموذجًا لوجستيًا (بيكر،٢٠١٠).

ويعتبر النموذج أحادي المعلم أبسط نماذج نظرية (IRT) وذلك لأنه يصف كل فقرة بدلالة معلم واحد بعد تحليل بيانات الفقرة التي تم الاستجابة عليها من قبل الأفراد، وهذا المعلم هو موقع الفقرة على متصل السمة الذي يمثل المفهوم والذي يمثل معلم الصعوبة للفقرة، ويفترض هذا النموذج أن جميع الفقرات لها قيمة تمييزية واحدة وأن التخمين يكون في حده

ISSN: 2537-0464 eISSN : 2537-0472

الأدنى، أي يفترض أن التخمين يساوي صفرا تقريبا وأن التمييز ثابت لكل فقرات الاختبار وأن صعوبة الفقرة تأخذ قيمة متغيرة. وطبقاً لنظرية الاستجابة للمفردة فإن النموذج الرياضي المعياري للمنحنى المميز للفقرة هو الصيغة التراكمية للدالة اللوجستية التالية (Linder & Hambleton,1997):

$$p(x_j = 1|\theta.a_j.\delta_j) = \frac{e^{a_j(\theta-\delta_j)}}{1+e^{a_j(\theta-\delta_j)}}$$

$p(x_j = 1|\theta.a_j.\delta_j)$ = احتمالية الإجابة على المفردة.

e = العدد الطبيعي وهو قيمة ثابتة = ٢,٧١٨

θ = قدرة الفرد.

δ_j = صعوبة المفردة.

a_j = معامل تمييز المفردة وهو قيمة ثابتة في نموذج راش.

الفرق بين النموذج الأحادي ونموذج راش:

يُعد النموذج الأحادي ونموذج راش متشابهين، أما من الناحية الرياضية فهما متكافئان ولذلك يمكن تحويل القيم من أحد النموذجين إلى الآخر باستخدام أساليب التدريج المناسبة. فهما يتطلبان أن تكون قيم تمييز الفقرة (a) ثابتة مع إمكانية اختلاف مواقع الفقرات. ومع ذلك هناك من يرى أن نموذج راش يمثل وجهة نظر فلسفية مختلفة عن تلك التي يتضمنها النموذج الأحادي فالنموذج الأحادي يسعى لمطابقة البيانات بأفضل ما يمكن وفقا لمتطلبات النموذج، بينما نموذج راش يسعى لبناء المتغير المستهدف، أي أنه يُعد المعيار الذي تبنى على أساسه أداة قياس المتغير. كذلك نجد في نموذج راش القيمة الثابتة لمعلم التمييز تساوي القيمة (١) بينما في النموذج الأحادي المعلم ليس من الضروري أن تكون القيمة الثابتة لمعلم التمييز مساوية للقيمة (١) فمن الممكن أن تأخذ أي قيمة أخرى ولكنها ثابتة لجميع الفقرات (أيالا،٢٠١٧).

ب- النموذج الثنائي المعلم:

يعود هذا النموذج إلى الجهود التي قام بها بيرنبوم (Birnbaum,1968) ويختلف هذا النموذج عن نموذج راش في كونه يسمح للمفردات بأن تختلف في معالم الصعوبة والتمييز (Hambleton & Swaminthan,1985) وعندما تختلف الفقرات في تمييزها نجد أن دوال الاستجابة للفقرة تتقاطع فيما بينها وذلك أن تمييز الفقرة يتناسب طرديًا مع ميل الدالة، فكلما زادت قيمة التمييز يصبح ميل الدالة أكثر انحدارًا؛ وبذلك تزداد قدرة المفردة على التمييز بين الأفراد، وبدورها تزداد القيمة القصوى لمعلومات المفردة عند موقع تلك المفردة مما يؤدي إلى تقليل مقدار خطأ التقدير حيث إن العلاقة بين دالة المعلومات والخطأ المعياري في التقدير هي علاقة عكسية. ويوضح النموذج الآتي الصيغة الرياضية للنموذج ثنائي البارامتر(أيالا،٢٠١٧).

ISSN: 2537-0464 eISSN : 2537-0472

$$p(x_j = 1|\theta.a_j.\delta_j) = \frac{e^{a_j(\theta-\delta_j)}}{1+e^{a_j(\theta-\delta_j)}}$$

$p(x_j = 1|\theta.a_j.\delta_j)$ = احتمالية الإجابة على المفردة.

e = العدد الطبيعي وهو قيمة ثابتة = ٢,٧١٨

θ = قدرة الفرد.

δ_j = صعوبة المفردة.

a_j = معامل تمييز المفردة وهو قيمة متغيرة في النموذج الثنائي.

ج- النموذج ثلاثي المعلم:

تعد ظاهرة التخمين في الاختبارات من الحقائق الثابتة حيث نجد أن بعض المختبرين يجيبون على بعض مفردات الاختبار إجابات صحيحة نتيجةً للتخمين، وبالنظر إلى النموذجين السابقين الأحادي المعلم والثنائي المعلم فنلاحظ أنهما لم يضعا التخمين من ضمن اعتبارهما(بيكر،٢٠١٠). وقدم فكرة هذا النموذج لورد (lord,1980) ويطلق على التخمين في هذا النموذج بالخط التقاربي الأدنى.

ويوضح النموذج الآتي الصيغة الرياضية للنموذج ثنائي البارامتر(أيالا،٢٠١٧).

$$p(x_j = 1|\theta, a_j, \delta_j, x_j) = \delta x_j + (1-\delta x_j)\frac{e^{a_j(\theta-\delta_j)}}{1+e^{a_j(\theta-\delta_j)}}$$

$p(x_j = 1|\theta, a_j, \delta_j, x_j)$= احتمالية الإجابة على المفردة.

e = العدد الطبيعي وهو قيمة ثابتة = ٢,٧١٨

θ = قدرة الفرد.

δ_j = صعوبة المفردة.

δx_j= معلم التخمين للمفردة

a_j = معامل تمييز المفردة وهو قيمة متغيرة في النموذج الثلاثي.

ويتناسب معلم التمييز في النموذج الثلاثي طرديًا مع الميل عند نقطة الانعطاف للمنحنى اللوجستي وتشمل العلاقة بين معلم التمييز وميل الدالة معلم التخمين ويمكن حساب التمييز في النموذج الثلاثي من خلال المعادلة:

(0.25*a(1-x)

ويعبر a عن معلم التمييز بينما يعبر x عن معلم التخمين.

وبذلك تكون الزيادة في معلم التخمين تؤدي إلى انخفاض في معلم التمييز. وبالنظر لدالة المعلومات للنموذجين السابقين نجد أن النهايات العليا لدالة المعلومات عند موقع الفقرة، بينما في النموذج الثلاثي نجد أن النهايات العليا للمعلومات تقع أعلى من موقع الفقرة.

ISSN: 2537-0464 eISSN : 2537-0472

ثانيًا معادلة الاختبارات Test Equation:

يعرف دورانز (Dorans, 1990) معادلة الاختبار بأنها: عملية إجراء تعديل إحصائي على درجات صورة واحدة من الاختبار؛ لجعل تلك الدرجات مكافئة بطريقة ما لدرجات صورة أخرى من نفس الاختبار. ويضيف كولن وبرينان (Kolen & Brennan,2014) بأن معادلة الاختبارات هي: إجراء لإزالة الآثار التي يتركها الفرق بين متوسط مستويات الصعوبة لفقرات الاختبار، وكذلك الفروق في متوسط مستويات التمييز لفقرات الاختبار، على درجات الاختبار، وذلك بين صورتين من نفس الاختبار لجعلهما متكافئتين، كما يقدم ليفنجستون (Livingston,2014) تعريفًا بسيطًا وعامًا للمعادلة، حيث تعتبر الدرجة على صورة جديدة والدرجة على صورة أخرى متكافئتين في مجموعة من الأفراد طبقت عليهم الصورة إذا كانوا يمثلون "الموقع النسبي نفسه في المجموعة. بينما يعرف كروكر والجاينا(٢٠١٧) معادلة الاختبارات بأنها : يمكن اعتبار درجتين إحداهما على الأداة x والأخرى على الأداة y متكافئتان إذا كانت كلا الأداتين X و Y تقيسان نفس السمة، وثباتهما متساوٍ، والرتب المئينية المناظرة للدرجات متساوية.

أنواع المعادلة لدرجات الاختبارات:

يمكن تصنف المعادلة إلى نوعين:

١- **المعادلة الأفقية Horizontal Equating:** وهي معادلة نماذج الاختبارات المصممة لقياس نفس السمة على نفس المستوى من القدرة ولنفس المجتمع(Slind & Linn,1977) . وتهدف المعادلة هنا إلى تعديل الفروقات الناتجة عن اختلاف مستويات صعوبة الفقرات بين الصور الاختبارية، والتي يفترض أن تقيس مستوى سلوك واحد أو مستوى محددًا من القدرة؛ وبذلك تكون هذه الصورة الاختبارية موجهة لمجتمع محدد من الأفراد يمثل توزيع تلك القدرة(Skaggs & Lissitz, 1986) .

٢- **المعادلة العمودية Vertical Equating:** وهي عبارة عن تحويل علامات على نماذج للاختبار، صممت لمجتمعات عند مستويات قدرة مختلفة(Slind & Linn,1977) . ويتطلب هذا التصميم بناء اختبارين أو أكثر تقيس السمة نفسها، ومصممة بحيث تكون مختلفة في الصعوبة(كروكر،٢٠١٧).

شروط معادلة الاختبارات:

ذكر هابلتون وسوامنيثان (Hambleton & Swaminthan,1985) عددًا من الشروط لمعادلة الاختبارات يجب التحقق منها قبل البدء في عملية المعادلة:

١- قياس نفس السمة (القدرة): يجب أن تقيس الاختبارات المراد معادلتها البعد نفسه. كمثال النقيض، يجب عدم معادلة الدرجات على اختبار في الجبر مع الدرجات على اختبار في تاريخ الفن.

٢- اللاتغاير في مجتمع الدراسة (Population Invariance): أي أن التحويل يجب أن يكون مستقلا عن مجموعات الأفراد التي استخدمت في تطوير عملية التحويل.

ISSN: 2537-0464 eISSN : 2537-0472

٣- التماثل (Symmetry): ويعني أن تحويل الدرجات من صورة إلى أخرى في الاختبار يجب أن يكون قابلًا للانعكاس.Invertible أي أن الرسم البياني لتوزيع الدرجات من الصورة (X) للاختبار إلى الصورة (Y)، يجب أن يكون هو نفس الانعكاس من الصورة (Y) للاختبار إلى الصورة (X) من نفس الاختبار.

٤- الثبات: ويعني أن الاختبارات التي تتمتع بثبات كامل هي التي يمكن معادلتها.

٥- العدالة (Equity): وهو أن يكون التوزيع التكراري المشروط للدرجات عند مستوى معين من مستويات القدرة (θ) للاختبار(Y) بعد تحويل الدرجات هو نفس التوزيع التكراري المشروط للدرجات عند مستوى معين من مستويات القدرة (θ) للاختبار(X).

تصاميم معادلة الاختبارات:

وهي الطرق المتبعة في جمع البيانات وهناك ثلاثة تصاميم شائعة تستخدم في المعادلة (أيالا،٢٠١٧) (Kolen&Brennan,2014) وهي:

١- تصميم المجموعة العشوائية:

في هذا التصميم يتم توزيع صورتي الاختبار بشكل عشوائي على مجموعة واحدة من المختبرين. ولضبط عملية التكافؤ يتم تقسم المجموعة الواحدة إلى مجموعتين فرعيتين يأخذ المختبر في المجموع الفرعية (١) الصورة (x) متبوعة بالصورة (y) ويليه يأخذ المختبر في المجموعة الفرعية الثانية الصورة (y) متبوعة بالصورة (x) وهكذا يستمر التطبيق بصورة لولبية.

٢- تصميم المجموعة الواحدة:

في هذا التصميم يتم توزيع المجموعة الواحدة بشكل عشوائي وتطبق على كل مجموعة صورة واحدة من صور الاختبار ويستخدم في التطبيق الأسلوب السابق اللولبي؛ بحيث يأخذ المختبر في المجموعة الفرعية (١) الصورة (x) ويأخذ المختبر في المجموعة الفرعية (٢) الصورة (Y).

ويتميز هذا الأسلوب في أنه يتطلب نصف وقت التطبيق في المجموعة العشوائية؛ لأن كل فرد يتلقى صورة واحدة فقط. لكن أيضًا تكون المعلومات المشتركة في إجراء المعادلة أقل مما هي في المجموعة الواحدة المتكافئة؛ ولذلك فهذا التصميم نادرًا ما يستخدم في التطبيقات العملية.

ويشترط في التصميمين السابقين التطبيق في الوقت نفسه.

٣- تصميم الفقرات المشتركة في مجموعات غير متكافئة:

ويتطلب في هذا التصميم استخدام عينتين من الأفراد، وتطبق على كل عينة إحدى الصورتين؛ ويطلق على هذا التصميم، أيضًا، "تصميم المجموعات غير المتكافئة مع اختبار جذعي مشترك، ويتم الحصول على المعلومات التي تستخدم في المعادلة من خلال الجذع المشترك ويوجد نوعان من تصميم الفقرات المشتركة في مجموعات غير متكافئة:

ISSN: 2537-0464 eISSN : 2537-0472

١- فقرات مشتركة داخلية: تحتوي على مجموعات جزئية من الفقرات يتم تضمينها في صورتي الاختبارين المراد معادلتهما.
- يعتبر الأداء عليها جزءًا من درجة المشاهدة.
- تتوزع هذه الفقرات في الأداة في المواقع نفسها في الصورتين.
٢- فقرات مشتركة خارجية: اختبار ذو فقرات مشتركة تطبق في وقتين مختلفين على المجموعتين.
-لا يعتبر الأداء عليها جزءا من الدرجة المشاهدة.
- تعطي هذه المفردات عادة بعد المفردات غير المشتركة.

الدراسات السابقة:

أجرى الرحيل (٢٠١٣) بدراسة هدفت للكشف عن أثر وجود أداء تفاضلي للجنس في فقرات الجذع المشترك على دقة معادلة الاختبارات العمودية لاختبار أوتيس لينون للقدرة العقلية بمستوييه المتوسطين وذلك باستخدام بيانات مولدة، وتكونت عينة الدراسة من (١٠٠١) طالب وطالبة تم تصنيفهم إلى (٥٠٠) من ذوي القدرة المتوسطة و (٥٠١) من ذوي القدرة المرتفعة، وتم استخدام تصميم الفقرات المشتركة والمجموعات غير المتكافئة لجمع البيانات وتضمن كل اختبار فقرات مشتركة عددها (٢٥) فقرة ، وقد اعتمدت الدراسة على محك الصدق التقاطعي فقط كمؤشر لدقة التقدير، وأظهرت النتائج أن إجراء المعادلة في ظل حذف فقرات الجذع المشترك ذات الأداء التفاضلي زاد من دقة المعادلة العمودية بالاستناد إلى قيم معاملات الصدق التقاطعي.

وهدفت دراسة للرحيل والدرابسة (٢٠١٤) للتعرف على أثر طريقة تقدير القُدرة وطريقة التعامل مع القيم المفقودة على دقة تقدير معالم الفقرات والأفراد. ولتحقيق ذلك تم استخدام بيانات مولدة بحجم (١٥٠٠) مفحوص على اختبار مكون من (٨٠) فقرة ثنائية الاستجابة ومطابقة للنموذج الثلاثي المعلمة. وتم الحصول على بيانات مفقودة بنسبة (٥%) تم معالجتها بثلاث طرق وهي: (التعويض المتعدد – نظرية الاستجابة للمفردة – تعظيم التوقعات) وللكشف عن دقة التقدير تم استخدام أسلوب تحليل التباين للقياسات المتكررة وأظهرت النتائج وجود اختلاف في تقدير معلم التمييز يعزى لاختلاف طريقة التقدير ولصالح طريقة التقدير بالأرجحية العظمى، كما وجد اختلاف في تقدير معلم التمييز لصالح طريقة التعويض المتعدد (MI) كذلك اختلاف في دقة تقدير التمييز يعزى للتفاعل بين طريقتي التقدير والمعالجة، بينما لم يكن هناك اختلاف في دقة تقدير معلم الصعوبة يعزى لاختلاف طريقة المعالجة أو التفاعل بين طريقة المعالجة وطريقة التقدير للفقرات والأفراد، باختلاف طرق تعويض القيم المفقودة وطرق تقدير القدرة، أما في دقة تقدير قدرات الأفراد فقد كانت هناك فروق معنوية تعزى لطريقة معالجة القيم المفقودة ولصالح طريقة تعظيم التوقعات(EM)، مع وجود اختلاف في طريقة تقدير الأفراد تعزى لطريقة التقدير ولصالح

ISSN: 2537-0464 eISSN : 2537-0472

طريقة بايز للتوقع البعدي (EAP). كذلك وجد اختلاف في دقة تقدير قدرات الأفراد يعزى للفاعل بين متغيري طرق المعالجة للقيم المفقودة وطريق التقدير.

وفي دراسة أنال وأنيل (Inal & Anil,2018) التي هدفت للكشف عن أثر الأداء التفاضلي في فقرات الجذع المشترك على معادلة الاختبارات في ظل أحجام عينة مختلفة وفق نظرة الاستجابة للمفردة، وتمت الدراسة على بيانات مولدة وفقًا لتصميم المجموعات المتكافئة، وقد اعتمدت الدراسة على مؤشر (REMSD) للكشف عن دقة التقدير في المعادلة كما تم استخدام تحليل التباين للكشف عن الدلالة المعنوية بين مجموعات المعالجة. وقد أظهرت النتائج أن هناك تأثيرا للأداء التفاضلي وبشكل كبير في فقرات الجذع المشترك على دقة معادلة الاختبارات وفقًا لمؤشر (REMSD) في حين لم يظهر تحليل التباين وجود دلالة معنوية وتأثير لحجم العينة.

وقام المحروق (٢٠١٩) بدراسة هدفت إلى الكشف عن فاعلية طريقة العلامات المشاهدة وكيرنيل في معادلة درجات الاختبارات، من خلال دراسة المتغيرات: حجم العينة، والاستماع - قراءة صوتية لكلمات القاموس وعرض القاموس المفصل - طول الاختبار، وقد تم توليد بيانات تجريبية وفقًا لنظرية الاستجابة للفقرة وتم استخدام مؤشر الخطأ المعياري ومؤشر الجذر التربيعي لمتوسط مربعات البواقي (RMSE)، وأظهرت النتائج إلى أن حجم العينات الكبير يقلل من الخطأ المعياري للمعادلة، ويقلل من البواقي المعيارية. كما أن طول الاختبار يوثر في الخطأ المعياري؛ فالاختبار الطويل يعطي قيما كبيرة للخطأ المعياري والبواقي المعيارية. بينما كانت طريقة المعادلة باستخدام الدرجات الملاحظة أكثر دقة من طريقة كيرنيل.

كما أجرى السرحان (٢٠١٨) دراسة هدفت للمقارنة بين طريقتي التقدير بالأرجحية العظمى والقيمة العظمى للتوزيع البعدي في دقة تقدير القدرة وفقا لنظرية الاستجابة للفقرة، واستخدمت الدراسة اختبار تفسير الرموز الصادر من (مركز التعليم السريع Learning Express ,LLC,2001) والذي يضم (3) اختبارات فرعية تقيس كلًا من التشابهات الرمزية (٣٠) فقرة، فرز وتصنيف الأشكال (٣٠) فقرة، التفكير في سلسلة (٣٠) فقرة، وبعد التحقق من مطابقة الفقرات للنموذج تم استبعاد بعض الفقرات غير المطابقة. وتكون الاختبار في صورته النهائية من (٧٣) فقرة، وطبق الاختبار على عينة عشوائية طبقية مكونة من (١٠٠٠) طالب وطالبة. وأظهرت النتائج عن وجود فروق في تقدير القدرة بين الطريقتين ولصالح طريقة التقدير بالقيمة العظمى للتوزيع البعدي من خلال انخفاض قيم الخطأ المعياري لها.

تعقيب على الدراسات السابقة:

من خلال مراجعة الدراسات والأدبيات السابقة وعلى حد علم واطلاع الباحث فإنه لا توجد دراسة سابقة تناولت أثر طرق تقدير القدرة على دقة معادلة الاختبارات في حين أن دراسة الرحيل والدرابسة (٢٠١٤) تعتبر من أقرب الدراسات للدراسة، إلا أن الدراسة

ISSN: 2537-0464 eISSN : 2537-0472

الحالية وتختلف على مستوى المتغير المستقل فنجد أنها تتناول طرق تقدير القدرة والتي لم تتطرق لها دراسة الشريفين. بينما يظهر الاختلاف عن دراسة الرحيل والدرابسة على مستوى المتغير التابع فهذه الدراسة تدرس طرق تقدير القدرة على معادلة الاختبارات. بينما دراسة الرحيل والدرابسة ركزت على معالم الفقرات والأفراد، كذلك تتميز الدراسة بطريقة الأرجحية العظمى الموزونة وهي طريقة لم تستخدمها دراسة الرحيل والدرابسة (٢٠١٤)، أما باقي الدراسات فنجد أنها قد تناولت كل متغير على حدة.

وتتميز الدراسة الحالية باعتمادها على صورتي اختبار أوتيس لينون المتوسط والابتدائي الثاني. أما من حيث نوع البيانات المستخدمة في الدراسات فتتميز الدراسة الحالية عن باقي الدراسات كون البيانات المستخدمة بيانات حقيقية بينما نجد باقي الدراسات اعتمدت على البيانات المولدة.

وفي ضوء ما تقدم عرضه من دراسات سابقة نجد أن هذه الدراسات لم تحسم الجدل حول الطرق المثلى في طرق تقدير القدرة ولم يتطرق أي لطرق تقدير القدرة على الخطأ المعياري في المعادلة وفقًا لنظرية الاستجابة للمفردة، ولذلك فإنه يؤمل من هذه الدراسة أن تضيف معلومات ومعرفة جديدة حول تلك الطرق.

منهج الدراسة وإجراءاتها:

منهج الدراسة:

في هذه الدراسة سيتم استخدام المنهج الوصفي المسحي لملاءمته لطبيعة الدراسة. ويعتبر المنهج الوصفي أحد مناهج البحث الذي يهدف للإجابة على السؤال الأساس في العلم (ماذا) أي ماهي طبيعية الظاهرة وذلك من خلال تحليل بنيتها وبيان العلاقة بين مكوناتها، أي أنه يهتم بأساس الوحدات أو الشروط أو العلاقات أو الفئات أو الأنساق المتعلقة بالظاهرة، وقد يشمل هذا الآراء حولها والاتجاهات إزاءها والعمليات التي تتضمنها وصولا إلى الآثار التي تحدثها. وقد يمتد المنهج الوصفي للماضي كما في المنهج التاريخي أو إلى الحاضر كما في المنهج الإمبيريقي والذي يسمى حينها بالمسحي (أبو حطب وصادق،٢٠١٠).

مجتمع الدراسة:

تكون مجتمع الدراسة من طلاب الصفوف العليا (الرابع والخامس والسادس) من المرحلة الابتدائية بالمدارس الحكومية (بنين) والبالغ عددهم (٦٠٢٦٥) طالبًا، كذلك من طلاب المرحلة المتوسطة بالمدارس الحكومية (بنين) بمدينة جدة والبالغ عددهم (٥٨٤٠٣) طالبًا (وزارة التعليم:٢٠١٨).

عينة الدراسة:

تم اختيار العينة بطريقة العينة العشوائية الطبقية، حيث تم تحديد عدد المكاتب التعليمية بمدينة جدة والبالغ عددها ٦ مكاتب تتوزع عليها المدارس الابتدائية والمتوسطة، ومن ثم تم اختيار عينة ممثلة للمجتمع طُبقت عليها صورتا الاختبار المراد معادلتهما بنمط المجموعة العشوائية غير المتكافئة.

ISSN: 2537-0464 eISSN : 2537-0472

ويوضح الجدول التالي رقم (١) عدد الطلاب في المجتمع وفقًا لكل مكتب تعليمي للصفوف العليا من المرحلة الابتدائية في المدارس الحكومية بمدينة جدة:

جدول (١) توزيع المجتمع للمرحلة الإبتدائية

تصنيف المدرسة	المكتب	عدد المدارس	الرابع الابتدائي		الخامس الابتدائي		السادس الابتدائي		المجموع	النسبة المئوية	المجموع الكلي
			عدد طلاب	عدد فصول	عدد طلاب	عدد فصول	عدد طلاب	عدد فصول			
حكومي	الصفا	53	3685	127	3777	129	3644	120	١١١٠٦	%١٨	٦٠٢٦٥
	النسيم	45	3228	110	3338	117	3164	110	٩٧٣٠	%١٦	
	جنوب جدة	50	3444	123	3535	123	3406	119	١٠٣٨٥	%١٧	
	شرق جدة	63	4753	165	4686	160	4616	157	١٤٠٥٥	%٢٣	
	شمال جدة	40	2611	83	2716	87	2799	89	٨١٢٦	%١٣	
	وسط جدة	31	2251	74	2299	76	2313	77	٦٨٦٣	%١١	
	المجموع		19972	682	20351	692	19942	672			

ويوضح الجدول التالي جدول (٢) عدد أفراد العينة والبالغ عددهم (٢٣٧٢) بنسبة تمثل (٤%) من مجتمع طلاب الصفوف العليا من المرحلة الابتدائية في التعليم الحكومي بمدينة جدة، وكيفية اختيارهم من كل مكتب ونسبتهم إلى المجتمع:

جدول (٢) عدد أفراد العينة

تصنيف المدرسة	المكتب	عدد المدارس	الأول المتوسط		الثاني المتوسط		الثالث المتوسط		المجموع	النسبة المئوية	المجموع الكلي
			عدد طلاب	عدد فصول	عدد طلاب	عدد فصول	عدد طلاب	عدد فصول			
حكومي	الصفا	40	3332	130	3240	١١١٠٦	3203	117	١١١٠٦	0.16737154	٥٨٤٠٣
	النسيم	32	3238	107	3144	٩٧٣٠	3175	111	٩٧٣٠	0.16363885	
	جنوب جدة	41	3496	131	3126	١٠٣٨٥	3170	118	١٠٣٨٥	0.16766262	
	شرق جدة	49	4598	158	4382	١٤٠٥٥	4152	144	١٤٠٥٥	0.22485146	
	شمال جدة	43	3009	100	2887	٨١٢٦	2838	102	٨١٢٦	0.14954711	
	وسط جدة	30	2622	91	2436	٦٨٦٣	2355	87	٦٨٦٣	0.12692841	
	المجموع	159	20295	394	19215	380	18893	372			

كما يوضح الجدول رقم (٣) التالي عدد الطلاب في المجتمع وفقًا لكل مكتب تعليمي للمرحلة المتوسطة في المدارس الحكومية بمدينة جدة:

جدول (٣) عدد الطلاب في المجتمع

مكتب التعليم	عدد أفراد العينة	النسبة للعينة
الصفا	٣٤١	0.14376054
النسيم	٣٣١	0.139544688
جنوب جدة	٦٢٥	0.263490725
شرق جدة	٥٣٥	0.225548061
شمال جدة	٢٣١	0.097386172
وسط جدة	٣٠٩	0.130269815
المجموع	٢٣٧٢	

ISSN: 2537-0464 eISSN : 2537-0472

ويوضح الجدول رقم (٤) التالي عدد أفراد العينة والبالغ عددهم (٣٢٧٦) بنسبة تمثل (٦%) من مجتمع طلاب المرحلة المتوسطة في التعليم الحكومي بمدينة جدة، وكيفية اختيارهم من كل مكتب ونسبتهم إلى المجتمع:

جدول (٤) توزيع العينة من المجتمع للمرحلة المتوسطة

مكتب التعليم	عدد أفراد العينة	النسبة للعينة
الصفا	٣٢٧	0.14
النسيم	٢٨٦	0.12
جنوب جدة	٣٥٣	0.15
شرق جدة	١٢١٦	0.51
شمال جدة	٢٨١	0.12
وسط جدة	٨١٣	0.34
المجموع	٢٣٧٢	

أدوات الدراسة:

تعتمد هذه الدراسة على اختبار القدرة العقلية للعالمين الأمريكيين أرثر أوتيس و روجر لينون ويعد هذا الاختبار أحد أقدم اختبارات الذكاء الجمعية وأكثرها شمولًا، وتشمل هذه السلسلة ستة مستويات، المستوى التمهيدي الأول خاص باختبار القدرات العقلية لرياض الأطفال، والمستوى التمهيدي الثاني معد لطلاب الصف الأول النصف الأول، والمستوى الابتدائي الأول وهذا الاختبار معد لطلاب النصف الثاني من الصف الأول إلى نهاية النصف الثاني من الصف الثالث الابتدائي، والمستوى الابتدائي الثاني وقد أعد هذا المستوى لطلاب الصف الرابع الابتدائي وحتى الصف السادس الابتدائي، والمستوى المتوسط وهذا المستوى معد لطلاب المرحلة المتوسطة من الصف الأول متوسط وحتى الصف الثالث المتوسط، وأخيرا المستوى المتقدم وهو مخصص لطلاب المرحلة الثانوية من الصف الأول ثانوي وحتى الصف الثالث ثانوي(الزهراني،٢٠١٠).

وسيتم التركيز على صورتي المقياس المستوى الابتدائي (مشاط،٢٠٠٩) والمستوى المتوسط (الرويلي،٢٠٠٨).

الإجابة عن السؤال الأول:

" هل يوجد تأثير لطرق تقدير القدرة على الأخطاء المعيارية في معادلة الاختبار وفقًا لنظرية الاستجابة للمفردة؟"

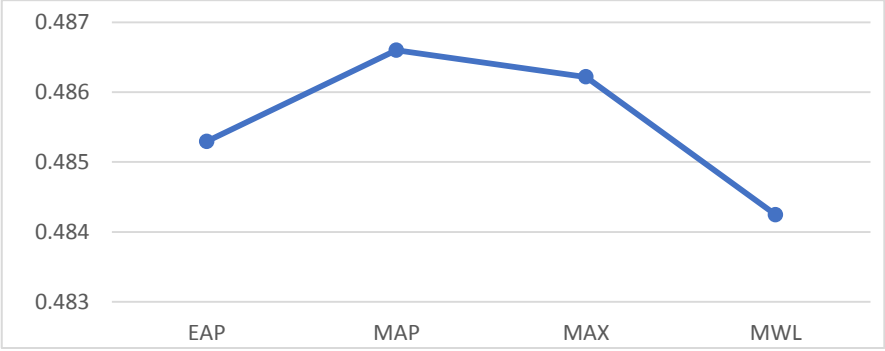
وللإجابة عن هذا السؤال تم القيام بالإجراءات التالية:

للكشف عن التأثير بين طرق التقدير تم الرجوع للبيانات قبل المعالجة وتم تقدير القدرة من خلال برنامج ((Xcalibre وذلك بأربع طرق، وهي طريقة التوقع البعدي (EAP) والتي تعتبر الطريقة الافتراضية في البرنامج ثم الطريقة البييزية (MAP) ثم طريقة الأرجحية العظمى (MLE) وأخيرا تم التقدير بطريقة الأرجحية العظمى الموزونة (WML) وتمت

ISSN: 2537-0464 eISSN : 2537-0472

المقارنة وفقًا للمعالجة الافتراضية للقيم المفقودة في البرنامج وعلى هذا فقد نتج لنا عدد (٨) ملفات بواقع (٤) ملفات لكل صورة اختبارية.

ولحساب الخطأ المعياري في التقدير تم استخدام نفس الإجراءات السابقة في السؤال الأول لحساب الخطأ المعياري في التقدير من خلال معادلة المعالم بين صورتي الاختبار وذلك لكل طريقة من طرق التقدير الأربع ومن ثم حساب احتمالية الاستجابة على الفقرة المعادلة عند كل مستوى من مستويات القدرة وفقًا لمعادلة النموذج الثنائي " المعادلة رقم (٢)" ثم حساب دالة المعلومات عند كل مستوى من مستويات القدرة بناءً على احتمالية الاستجابة للفقرات المعادلة عند كل مستوى من مستويات القدرة ثم حساب مجموع دالة المعلومات ومن ثم حساب الخطأ المعياري، وهذا الإجراء تم مع كل طريقة من طرق التقدير الأربع وفقًا لكل طريقة تعويض عن القيم المفقودة وبالتالي ظهر لنا عدد (٤) ملفات تعبر عن الخطأ المعياري في التقدير عند كل مستوى من مستويات القدرة وفقًا لطريقة التقدير ومن ثم تم أخذ متوسط الخطأ المعياري لكل طريقة تقدير. ويوضح الرسم البياني التالي في الشكل (١٤) والجدول رقم (٢٩) الفروق الظاهرية في متوسطات الخطأ المعياري وفقًا لطريقة تقدير القدرة:



شكل رقم (١) متوسطات الخطأ المعياري وفقاً لطريقة التقدير

جدول (٥) يوضح الخطأ المعياري وفقاً لطرق التقدير

Maximum	Minimum	Std. Deviation	Mean	
0.900211142	0.263089678	0.191878284262	0.48529323843	EAP
0.894737225	0.268212456	0.188418496927	0.48659892501	MAP
0.914609165	0.254425021	0.199805207286	0.48621513711	MAX
0.906434934	0.258580280	0.195048675887	0.48424483069	MWL
0.914609165	0.254425021	0.192932886110	0.48558803281	Total

ISSN: 2537-0464 eISSN : 2537-0472

يتضح من الرسم البياني ومن الجدول السابق جدول متوسطات الخطأ المعياري وفقًا لطرق التقدير بأن هناك فروقًا ظاهرية بين متوسطات الخطأ المعياري في التقدير بين الطرق الأربعة، حيث كانت أقل قيمة للخطأ المعياري لطريقة الأرجحية العظمى الموزونة (WML) بقيمة قدرها (٠.٤٨٤٢٤) ثم تلتها طريقة الأرجحية العظمى (MLE) بقيمة بلغت (٠.٤٨٦٢١) ثم طريقة التوقع البعدي (EAP) بقيمة قدرها (٠.٤٨٥٢٩) ثم جاءت الطريقة البيزية (MAP) بأعلى خطأ معياري يقدر ب (٠.٤٨٦٥٩).

وتشير هذه النتائج إلى أفضلية لطريقة الأرجحية العظمى الموزونة (MWL) على باقي طرق التقدير وفقًا لطريقة التعويض عن القيم المفقودة الافتراضية في البرنامج حيث قدمت لنا أقل خطأ معياري وذلك مقارنة بباقي الطرق، وتختلف هذه النتيجة عن دراسة سرحان (٢٠١٨) والتي تظهر أفضلية لطريقة التقدير البيزية على باقي الطرق وإن كانت هذه الدراسات لم تستخدم طريقة الأرجحية العظمى الموزونة ومع ذلك فطريقة التعويض بالأرجحية العظمى (MLE) أعطت نتائج أفضل من الطرق البايزية، ويمكن تفسير هذا الاختلاف كون الدراسة الحالية ترتكز على دقة تقدير الخطأ المعياري في المعادلة، بينما الدراستان السابقتان تقارن في تقدير الخطأ المعياري في تقدير الأفراد وتقدير المعالم على التوالي. كما تتفق نتائج هذه الدراسة جزئيا مع دراسة الرحيل والدرابسة (٢٠١٤) في أفضلية الأرجحية العظمى.

ثالثًا: الإجابة عن السؤال الثاني:

" هل توجد فروق بين طرق تقدير القدرة في نظرية الاستجابة للمفردة وفقًا للخطأ المعياري في معادلة الاختبارات؟". وللإجابة على هذا السؤال تم القيام بالإجراءات التالية:

أولًا: تم استخدام تحليل التباين الثنائي للكشف عن الفروق إن وجدت من خلال الدلالة الإحصائية

جدول (٦) يوضح تحليل التباين

Tests of Between-Subjects Effects

Dependent Variable: SE

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	.005[a]	11	.000	.012	1.000	.000
Intercept	229.081	1	229.081	5960.529	.000	.861
MISSING	.004	2	.002	.046	.955	.000
ESTMATE	.002	3	.001	.014	.998	.000
MISSING * ESTMATE	.006	6	2.775E-5	.001	1.000	.000
Error	36.896	960	.038			
Total	265.982	972				
Corrected Total	36.901	971				

ISSN: 2537-0464 eISSN : 2537-0472

يتضح من الجدول رقم (٦) السابق جدول تحليل التباين عدم وجود فروق ذات دلالة إحصائية بين كل من طرق التعويض عن القيم المفقودة وطرق تقدير القدرة والتفاعل بينهما، حيث إن قيم sig كانت أكبر من (٠.٠٥). وقد يكمن السبب في عدم وجود فروق معنوية كون الفروق في قيم الخطأ المعياري طفيفة ولذلك نجد كثيرًا من الباحثين يكتفون بمحك الخطأ المعياري في تقدير دقة المعادلة ودراسة المحروق كدراسة الشريفين (٢٠١٧). وهو ما أكدته أيضًا دراسة أنال وأنيل (Inal & Anil,2018) والتي أظهرت عدم وجود فروق جوهرية في قيم الخطأ المعياري لمعادلة الصور الاختبارية.

توصيات الدراسة:

تأتي هذه الدراسة في محاولة لمعرفة أثر طرق تقدير القدرة للأفراد على دقة معادلة الاختبارات في نظرية الاستجابة للمفردة، وذلك وفقًا لأربع طرق تقدير وهي (الأرجحية العظمى (MLE) وطريقة التقدير بالتوقع البعدي (EAP) والتقدير بالأرجحية العظمى (MWE) والطريقة البيزية (BME) وقد أظهرت النتائج أن طريقة الأرجحية العظمى الموزونة (MWL) تعطي أفضل النتائج، وتوصي هذه الدراسة بالأخذ بهذه الطرق عند معادلة الصور الاختبارية مع الأخذ بالاعتبار أن هذه الدراسة اعتمدت في محدداتها على النموذج الثنائي المعلمة، وبناء على النتائج السابقة توصي الدراسة بأهمية الاستفادة من نتائج الدراسة الحالية في المجالات التربوية المختلفة وذلك عند معادلة الصورة الاختبارية ووفقًا لنظرية الاستجابة للمفردة ويمكن الإشارة إلى هذه التوصيات في النقاط التالية:

١- حث المراكز المعنية ببناء الاختبارات على التركيز على نظرية الاستجابة للمفردة في بناء الصور الاختبارية؛ وذلك لما تتميز به هذه النظرية من قوة في تقدير قدرات الأفراد ومعالم الفقرات ومعادلة الصور الاختبارية.

٢- ضرورة معادلة الصور الاختبارية المقدمة من مراكز الاختبارات وذلك بهدف تحقيق العدالة بين المختبرين، فقد تكون الفروق بين المختبرين لا تعود للفروق في قدراتهم وإنما سببه لاختلاف مستويات الصعوبة في الصور الاختبارية.

٣- كذلك من المهم الأخذ بطريقة الأرجحية العظمى الموزونة (MWL) في التقدير وفقًا لنظرية الاستجابة للمفردة IRT لما لها من أفضلية على باقي طرق التقدير وذلك في تقليل الخطأ المعياري وفقًا للنموذج ثنائي المعلمة.

المقترحات:

في ظل الاطلاع والبحث ضمن متغيرات الدراسة الحالية فلا شك أنه قد تم مواجهة عدة مواضيع من شأنها أن تشكل نقاط بحث للعديد من الباحثين، ويمكن اقتراح البحث في أثر العناوين التالية:

١- طرق تقدير القدرة للأفراد على دقة المعادلة الأفقية في نظرية الاستجابة للمفردة.

٢- طرق معالجة القيم المفقودة في الجذع المشترك وتقدير القدرة للأفراد على دقة المعادلة في نظرية الاستجابة للمفردة.

ISSN: 2537-0464 eISSN : 2537-0472

المراجع

المراجع العربية:

أبو حطب، فؤاد وصادق، آمال (٢٠١٠). *مناهج البحث وطرق التحليل الإحصائي*. القاهرة: مكتبة الأنجلو المصرية.

أيالا، أر (٢٠١٧). *النظرية والتطبيق في نظرية الاستجابة للفقرة* (ترجمة عبد الله الكيلاني وإسماعيل البرصان). الرياض: دار جامعة الملك سعود للنشر والتوزيع.

انستازي، أنا ويورينا، سوزان (٢٠١٥). *القياس النفسي* (ترجمة صلاح الدين علام). عمان: دار الفكر.

بيكر، فرانك (٢٠١٠). *أسس نظرية الاستجابة للمفردة* (ترجمة عبدالرحمن الطريري والسيد أبو هاشم و سوسن شلبي). الرياض: دار جامعة الملك سعود للنشر والتوزيع.

الرحيل، راتب (٢٠١٣). أثر وجود أداء تفاضلي في الفقرات المرساوية على دقة المعادلة العمودية لاختبار أوتيس لينون للقدرة العقلية. *المجلة الدولية التربوية المتخصصة*، ٢(٨)، ٧٥٤-٧٧١.

الرحيل، راتب والدرابسة، رياض (٢٠١٤). أثر طريقة تقدير القُدرة وطريقة التعامل مع القيم المفقودة على دقة تقدير معالم الفقرات والأفراد. *المجلة الدولية التربوية المتخصصة*، ٣(٦)، ٢٣-٤٧.

الرويلي، سامي (٢٠٠٨). *دلالات صدق وثبات صورة معدلة للبيئة السعودية لاختبار أوتيس - لينون للقدرة العقلية المستوى الابتدائي الثاني* - الصورة (K)، (رسالة ماجستير غير منشورة)، *جامعة اليرموك*، الأردن.

الزهراني، محمد (٢٠١٠). الخصائص السيكومترية لاختبار (أوتيس – لينون) للقدرة العقلية لدى طلبة المرحلة المتوسطة بمحافظة القنفذة. *مجلة كلية التربية بالزقازيق*، ٦٨، ٢٠١-٢٤٥.

سرحان، محمد(٢٠١٨). المقارنة بين طريقتي الأرجحية العظمى والقيمة العظمى للتوزيع البعدي في تقدير القدرة وفقاً لنظرية الاستجابة للفقرة .(رسالة دكتوراة غير منشورة). *كلية التربية ابن رشد للعلوم الإنسانية*. جامعة بغداد.

الشريفين، نضال كمال (٢٠١٧). أثر طريقة التعويض عن الاستجابات المفقودة على دقة معادلة صور مختلفة لاختبار باستخدام نماذج نظرية الاستجابة للمفردة. *مجلة العلوم التربوية*، ٤٤(٤)، ٢٧٧-٣٠١.

الشريفين، نضال كمال (٢٠١٢). أثر طريقة تقدير معالم الفقرة وقدرات الأفراد على قيم معالم الفقرة والخصائص السيكومترية للاختبار في ضوء تغير حجم العينة. *المجلة التربوية*، ٢٦(١٠٤)، ١٧٧-٢٣٨.

ISSN: 2537-0464 eISSN : 2537-0472

الطريري، عبد الرحمن سليمان (٢٠١٤). *القياس النفسي والتربوي نظريته- أسسه- تطبيقاته* (ط.٢). الرياض: مكتبة الرشد.

كروكر، لندا والجاينا، جيمس (٢٠١٧). *مدخل إلى نظرية القياس التقليدية والمعاصرة* (ترجمة هند الحموري وزينات دعنا). عمان: دار الفكر.

مشاط، سارة (٢٠٠٩). *تقنين اختبار (أوتيس – لينون) للقدرة العقلية المستوى المتوسط* - الصورة (ج) على عينة من طالبات المرحلة المتوسطة بمدينة جدة، (رسالة ماجستير غير منشورة)، *جامعة أم القرى*، مكة المكرمة.

المحروق، يوسف عبدالعاطي (٢٠١٩). فاعلية طريقة العلامات المشاهدة وكيرنيل في معادلة درجات الاختبارات. *عمادة البحث العلمي الجامعة الأردنية* ، ٤٦، ٢٠٤-٢٢٣.

وزارة التعليم (٢٠١٨). *بيانات التعليم العام للعام الدراسي١٤٣٩-١٤٤٠ه*. تم الاسترجاع من موقع

https://departments.moe.gov.sa/PlanningDevelopment/RelatedDepartments/Educationstatisticscenter/OpenData/Pages/GeneralEducation.aspx بتاريخ ٢٠١٩/١٢/٥.

المراجع الأجنبية :

Dorans, N. J. (1990). Equating Methods and Sampling Designs. *Applied Measurement in Education, 3*(1), 3-17. https://doi.org/10.1207/s15324818ame0301_2

Falmagne, K.-C. (1989). A latent trait theory via a stochastic learning theory for a knowledge space. *Psychometrika, 54*(2), 283-303. https://doi.org/10.1007/bf02294521

Galindo Garre, F., & Vermunt, J. K. (2006). Avoiding Boundary Estimates in Latent Class Analysis by Bayesian Posterior Mode Estimation. *Behaviormetrika, 33*(1), 43-59. https://doi.org/10.2333/bhmk.33.43

Hoover, w., & Kum, O.(1995) . Non-Equilibrium Simulations, Journal of Molecular Physics 86, 685- 699.

Hambleton, R., & Swaminathan, H. (1985). 1985: Item response theory: principles and applications. Boston: Kluwer-Nijhoff Publishing.

Hambleton, R. K., Shavelson, R. J., Webb, N. M., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory* (Vol. 2). Sage.

ISSN: 2537-0464 eISSN : 2537-0472

Inal, H., & Anil, D. (2018). Investigation of Group Invariance in Test Equating Under Different Simulation Conditions. *Eurasian Journal of Educational Research*, *18*(78), 1-20. https://doi.org/10.14689/ejer.2018.78.4

Kolen, M. J. (1981). COMPARISON OF TRADITIONAL AND ITEM RESPONSE THEORY METHODS FOR EQUATING TESTS. *Journal of Educational Measurement*, *18*(1), 1-11. https://doi.org/10.1111/j.1745-3984.1981.tb00838.x

Kolen, M. J., & Brennan, R. L. (2014). *Test equating: Methods and practices*. Springer Science & Business Media.

Livingston, S. A. (2014). Equating test scores (without IRT). *Educational testing service*.

Lord, F. (1952). A theory of test scores. *Psychometric monographs*.

Lord, F., Novick, M., & Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. *Statistical Theories on Mental Test Scores, Reading: Mass./Addison-Wesley*.

Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Routledge.

Lord, F. M. (1986). MAXIMUM LIKELIHOOD AND BAYESIAN PARAMETER ESTIMATION IN ITEM RESPONSE THEORY. *Journal of Educational Measurement*, *23*(2), 157-162. https://doi.org/10.1111/j.1745-3984.1986.tb00241.x

Lord, F. M., & Novick, M. R. (2008). *Statistical theories of mental test scores*. IAP.

Masters, G. N. (1982). A rasch model for partial credit scoring. *Psychometrika*, *47*(2), 149-174. https://doi.org/10.1007/bf02296272

Price, L. R. (2016). *Psychometric methods: Theory into practice*. Guilford Publications.

Wright, B. D., & Masters, G. N. (1982). *Rating scale analysis*. MESA press.

ISSN: 2537-0464 **eISSN : 2537-0472**