Journal of Educational and Psychological Sciences
Volume (6), Issue (52): 30 Nov 2022
P: 30 - 54
ISSN: 2522-3399
مجلة العلوم التربوية والنفسية
المجلد (6)، العدد (52): 30 نوفمبر 2022م
ص: 30 - 54

Comparing the effectiveness of a number of statistical methods to equalize the test scores using the common anchor design in the light of some variables

Wadea Areesh Raedi
College of Education || Umm Al-Qura University || KSA

Abstract: The present study aimed to compare the effectiveness of three statistical methods to equalize the test scores which is Levin way for watching signs and Ticker linear way and Unsmoothed Unequipercetile Equating under different experimental conditions whence adjusting the power distribution. The sub-sample has reache ; the small (500), medium (1000), and large (1500), were also conducted at the level of the distribution of natural power and regular.

The study found the following results:

The standard error rates vary depending on the method of equivalent grades. Accordingly, for ways of equation scores vary according to the bilateral interaction between the equation way and capacity distribution. The scores equation vary in the bilateral interaction between the way the equation, and the sample size. Accordingly, the ways equation vary in the triple interaction between the way the equation and distribution capacity and the size of the sample.

Accordingly, the researcher recommends the following:

1. Not shorten in the equation of test scores on the specific way; because of the different ways the effectiveness of the equation depending on the nature of the test models, where there is no way is the best in all situations.

2. Training interested in conducting tests on the use of software that facilitates the process of making the equation between the scores of experimental models.

Keywords: Equation scores, Effectiveness, Statistical methods, Experimental models, Anchor test Design.

مقارنة فاعلية عدد من الطرق الإحصائية لمعادلة درجات الاختبارات باستخدام تصميم الجذع المشترك في ضوء بعض المتغيرات

وديع عريش رعيدي
كلية التربية || جامعة أم القرى || المملكة العربية السعودية

المستخلص: هدفت الدراسة الحالية إلى مقارنة فاعلية ثلاث طرق إحصائية لمعادلة درجات الاختبارات -طريقة ليفين للعلامات المشاهدة، وطريقة تكر الخطية، وطريقة المئينيات غير المهدة- تحت ظروف تجريبية مختلفة من حيث طريقة المعادلة وحجم العينة وتوزيع القدرة.

وقد بلغت العينة الفرعية، صغيرة (500)، متوسطة (1000)، وكبيرة (1500)، كما أجريت على مستوى توزيعي القدرة الطبيعي والمنتظم.

وتوصلت الدراسة للنتائج التالية:

تختلف معدلات الخطأ المعياري باختلاف طريقة معادلة الدرجات، وتبعا تختلف طرق معادلة الدرجات باختلاف التفاعل الثنائي بين طريقة المعادلة وتوزيع القدرة، وتختلف طرق معادلة الدرجات في التفاعل الثنائي بين طريقة المعادلة وحجم العينة، وتبعًا تختلف طرق معادلة التفاعل الثلاثي بين طريقة المعادلة وتوزيع القدرة وحجم العينة وعليه فإن الباحث يوصي بما يلي:

1. عدم الاقتصار في معادلة درجات الاختبارات على طريقة محددة؛ لعدم وجود طريقة هي الأفضل في كل الحالات.

2. تدريب المهتمين على استخدام البرمجيات التي تسهل إجراء المعادلة بين درجات النماذج الاختبارية.

الكلمات المفتاحية: معادلة الدرجات، فاعلية، الطرق الإحصائية، النماذج الاختبارية، تصميم الجذع المشترك.

المقدمة.

في العديد من المجالات التطبيقية في القياس النفسي والتربوي يلجأ المتخصصون والباحثون إلى استخدام أدوات (اختبارات ومقاييس مختلفة) لقياس المفحوصين، يفترض في هذه الأدوات أنها تقيس السمة نفسها، فمثلاً في اختبارات القدرات العامة الذي يطبق على طلاب المرحلة الثانوية بالمملكة العربية السعودية العديد من الصيغ التي تبنى بحيث يتم تطبيقها عدة مرات خلال العام، فإذا كان توزيع الدرجات لهذه الصيغ ليس نفسه؛ فمن الضروري تحديد الدرجات المتكافئة من الصيغ الاختبارية المختلفة، تحقيقًا لمبدأ العدالة بين الطلاب المختبرين خلال العام، ويطلق على عملية تحديد الدرجات المتكافئة للصيغ المختلفة في المثال السابق اسم المعادلة الأفقية "Horizontal equating"، بينما لو تم تحديد الدرجات المتكافئة لصيغ بطارية اختبارات تحصيلية مثلاً تقيس مستويات عدة فإنه يطلق عليها اسم المعادلة العمودية"Vertical equating".

ويمكن اعتبار عملية المعادلة تأسيس أو اشتقاق درجات متكافئة من اختبارات مختلفة، فإذا عودلت درجات اختبارين (X, Y) وتقدم مفحوص لاختبار (x) فإنه يمكن تحديد الدرجة المكافئة لها على الاختبار (y). يمكن كتابة الدالة التي تربط بين درجات (x) والدرجات المكافئة لها على الاختبار (y) على النحو التالي $f(x) = y^*$ (كروكر والجينا، 2009).

وتختلف إجراءات جمع البيانات اللازمة لعملية معادلة درجات الاختبارات تبعاً للتصميم الذي يتبعه الباحث، ومن أشهرها تصميم المجموعة الواحدة، وتصميم المجموعات العشوائية المتكافئة، وتصميم المجموعات غير المتكافئة ذات المفردات المشتركة أو اختبار الجذع المشترك، وتصميم الأخير هو الذي اتبعه الباحث في جمعه لبيانات المعادلة نظراً لكونه يسمح بالتخلص من الفروق العشوائية في القدرة أو السمة المراد قياسها بين أفراد المجموعتين (بالخيور، 2014).

ويوجد العديد من أساليب وطرق إجراء المعادلة التي تتباين في دقتها وفاعليتها، والتي تختلف باختلاف نظرية القياس التي تنبثق منها، وقد اهتمت الكثير من الأبحاث بالمقارنة بين طرق وأساليب المعادلة من أجل توفير أساس علمي متين ينطلق منه المشتغلين والمهتمين بالاختبارات في إجرائهم للمعادلة، كما نجد ذلك في دراسة (بالخيور، 2014) و(الزعابي، المحرزي، وحسن، 2019)، ويأتي هذا الدراسة ضمن الجهود المبذولة المهتمة بهذه المقارنات بين الطرق.

مشكلة الدراسة وتساؤلاتها:

مع التطور الذي يشهده علم القياس والتقويم التربوي، ونشوء نظريات لتفسير المشكلات التي تواجه هذا العلم، والتي من ضمنها حاجة المتخصصين والباحثين وكذلك التربويين لمعادلة درجات نماذج مختلفة لاختبار ما؛ ليضمنوا عدالة نتائج الاختبار لجميع المفحوصين بغض النظر عن النموذج الذي أُدِيَ الاختبار من خلاله.

فقد بنيت عدة طرق لمعادلة درجات الاختبارات، اعتمدت بعضها النظرية التقليدية في القياس، واعتمد البعض الآخر منها النظرية الحديثة، إلا أن الباحثين والمتخصصين وجدوا تبايناً بين فاعلية هذه الطرق من حيث دقة نتائجها. حيث يرى شولو وسيرس (2011) أنه قد تتشابه أو تختلف نتائج عملية المعادلة باستخدام الطرق الإحصائية المختلفة (أورد في: المحرزي، 2015). وبالرغم من جهود المبذولة من الباحثين في مقارنة فاعلية طرق معادلة الدرجات

(Von وKong، 2005)، (الوليلي، 2005)، (حمادنة، 2012)، (الزعابي، المحرزي، وحسن، 2019)، إلا أن هذه المقارنة تتطلب مزيداً من الدراسة بسبب اختلافها باختلاف المتغيرات التي تركز المقارنة عليها. وبناءً على ما سبق فقد رأى الباحث أن يقدم بحثه في المقارنة بين عدد من الطرق المختلفة للمعادلة والتي اعتمدت على النظرية التقليدية في القياس، مستخدماً أسلوب المحاكاة في توليد البيانات، ووفقاً لمتغيرَي حجم العينة وتوزيع القدرة.

وقد تحددت مشكلة الدراسة في التساؤلات التالية:

1- هل تختلف معدلات الخطأ المعياري للمعادلة باختلاف الطرق المختلفة للمعادلة؟
2- هل تختلف معدلات الخطأ المعياري للمعادلة باختلاف التفاعل الثنائي بين طريقة المعادلة وتوزيع القدرة؟
3- هل تختلف معدلات الخطأ المعياري للمعادلة باختلاف التفاعل الثنائي بين طريقة المعادلة وحجم العينة؟
4- هل تختلف معدلات الخطأ المعياري للمعادلة باختلاف التفاعل الثلاثي بين طريقة المعادلة وتوزيع القدرة وحجم العينة؟

فروض الدراسة:

لدراسة التساؤلات التي أثيرت عن مشكلة الدراسة تم طرح الفروض التالية:

1- لا تختلف معدلات الخطأ المعياري للمعادلة باختلاف طرق المعادلة.
2- لا تختلف معدلات الخطأ المعياري للمعادلة باختلاف التفاعل الثنائي بين الطريقة وتوزيع القدرة.
3- لا تختلف معدلات الخطأ المعياري للمعادلة باختلاف التفاعل الثنائي بين الطريقة وحجم العينة.
4- لا تختلف معدلات الخطأ المعياري للمعادلة باختلاف التفاعل الثلاثي بين الطريقة وحجم العينة وتوزيع القدرة.

أهداف الدراسة

تهدف الدراسة إلى مقارنة عدد من الطرق الإحصائية لمعادلة درجات الاختبارات والتي تتبع النظرية التقليدية في القياس، وهي طريقة ليفين للعلامات المشاهدة، وطريقة تكر الخطية، وطريقة المئينيات غير المهدة، وباستخدام تصميم الجذع المشترك في ضوء متغيرَي توزيع القدرة وحجم العينة.

ويمكن أن تصاغ أهداف الدراسة بشكل تفصيلي كما يلي:

1- التعرف على أثر اختلاف معدلات الخطأ المعياري للمعادلة باختلاف الطرق المختلفة للمعادلة؟
2- التعرف على أثر اختلاف معدلات الخطأ المعياري للمعادلة نتيجة التفاعل الثنائي بين طريقة المعادلة وتوزيع القدرة؟
3- التعرف على أثر اختلاف معدلات الخطأ المعياري للمعادلة نتيجة التفاعل الثنائي بين طريقة المعادلة وحجم العينة؟
4- التعرف على أثر اختلاف معدلات الخطأ المعياري للمعادلة نتيجة التفاعل الثلاثي بين طريقة المعادلة وتوزيع القدرة وحجم العينة؟

أهمية الدراسة:

للدراسة الحالية أهمية كبيرة من الناحيتين النظرية والتطبيقية، وكما يلي:

- الأهمية النظرية للدراسة؛ وتتمثل في أنها:

- تقدم إطارًا نظريًّا مفصلًا تتضح من خلاله ماهية المعادلة، والطرق المختلفة لها، والمقارنة بينها من حيث فاعليتها ودقتها في تحقيق هدف المعادلة مما يثري المكتبة العربية، حيث يلاحظ فيها ندرة الدراسات والمراجع التي تناولت موضوع المعادلة.
- وكذلك إضافة أطر ذات كفاءة ودقة في إجراء المعادلة بين درجات الصور المختلفة للاختبار.

- أما الأهمية التطبيقية فتتمثل في عدة نقاط كما يلي:

- تقديم عدد من طرق المعادلة، وتوضيح فاعليتها وكيفية الاستفادة منها وفق الظروف التجريبية التي يفضل استخدامها معها.
- تقديم دليلًا موجهًا للباحثين والمهتمين يساعد في اختيار الطريقة الأكثر فاعلية في تحقيق دقة وإحكام لنتائج الاختبارات مهما اختلفت صورها.
- تقدير القدرة الحقيقية لكل مفحوص، مع المحافظة على ثبات هذا التقدير مع استخدام الصور الاختبارية المتعددة، والتي قد تختلف فيها صعوبة المفردات المكونة لها.
- تحسين قدرة الاختبارات على تحقيق مبدأ العدل في قياس سمة ما لدى مجموعة من الطلبة، تعرضوا لصور مختلفة من اختبار معين.

حدود الدراسة:

- **النظرية العلمية**: اقتصار الدراسة في مقارنتها على بعض الطرق التي تتبع النظرية الكلاسيكية في القياس.
- **المتغيرات المستقلة في الدراسة**: وتتحدد بـ (طريقة المعادلة، حجم العينة، نوع توزيع القدرة).
- **مصدر البيانات**: تعتمد الدراسة على بيانات يتم توليدها بواسطة برامج حاسوبية.
- **محك الدراسة**: اعتماد معدلات الخطأ المعياري معيارًا للمقارنة بين دقة الطرق المختلفة للمعادلة.

مصطلحات الدراسة:

- **معادلة الاختبارات Test Equating:** "إجراء احصائي لإيجاد علاقة بين درجات صورتين أو أكثر للاختبار الواحد، وهذا الإجراء الإحصائي من شأنه وضع صورتين أو أكثر على مقياس مشترك" (بالخيور، 2014، ص: 55-56).
- **الطرق الإحصائية لمعادلة درجات الاختبارات Methods Equating**: أساليب إجراء معادلة الاختبارات التي تستند إلى النظرية الكلاسيكية في القياس، وهي طريقة ليفين للعلامات المشاهدة، وطريقة تكر الخطية، وطريقة المئينيات غير المهدة.
- **تصميم الجذع المشترك Anchor-Test Design:** "تصميم لجمع البيانات من أجل معادلة درجات الاختبارات، وفيه تقدم الاختبارات المختلفة إلى مجموعات مختلفة من المفحوصين، بالإضافة لاختبار مشترك يتقدم إليه جميع المفحوصين، ولا يشترط في هذا التصميم توزيع الأفراد عشوائيًّا على المجموعات" (كروكر والجينا، 2009، ص: 604).
- **طريقة الوسط الحسابي Mean Equating:** "هي طريقة يتم من خلالها إجراء تحويل رياضي يؤدي إلى تساوي متوسطي الدرجات في صورتي الاختبار". (الدوسري، 2001، ص: 137)
- **طريقة المئينيات Equipercntile Equating:** "تعتبر درجات المفحوصين متعادلة إذا تساوت رتبها المئينية بالنسبة لمجموعة معينة من الأفراد". (كوفحي وزواهرة، 2017، ص: 349)

- **الطريقة الخطية :Linear Equating** "تقوم هذه الطريقة على تحويل الدرجات رياضياً، بحيث تتساوى المتوسطات الحسابية والانحرافات المعيارية لدرجات مفردات صورتي الاختبار المراد معادلتهما". (بالخيور، 2014، ص: 99)
- **معايير الفاعلية Criteria of Efficiency:** وهي المعايير المستخدمة للتعبير عن دقة نتائج معادلة درجات الاختبارات، مثل الخطأ المعياري للمعادلة، ومعامل الصدق التقاطعي.
- **الخطأ المعياري للمعادلة Standard Error Of Equating:** "يمثل الانحراف المعياري للدرجات المعادلة من إحدى صورتي الاختبار للأخرى من تحويل مبني على عينات اختيرت بشكل مستقل" (كوفحي وزواهرة، 2017، ص: 349)

2- الإطار النظري والدراسات السابقة.

أولاً- الإطار النظري:

يتناول الباحث في عرضه للإطار النظري مفهوم المعادلة، وأنواعها، والأساليب التي تتبع لإجراء المعادلة، والشروط التي يُتحقق منها قبل إجراء المعادلة، والتصاميم المختلفة لجمع البيانات من أجل إجراء المعادلة، وعرض لطرق معادلة الاختبارات في النظرية التقليدية، والمعايير التي تُقيّم دقة المعادلة على أساسها، وأخيراً عرض الدراسات السابقة ذات الصلة بالموضوع.

مفهوم المعادلة Equating:

يعرف كل من كولن وبرنان(Brennan & Kolen, 1987) المعادلة بأنها العملية الإحصائية التي تستخدم لتعديل الدرجات في النماذج المختلفة للاختبار، بحيث يمكن استخدامها بالتبادل، فهي تعمل على ضبط الاختلافات في الصعوبة بين النماذج التي يتم بناؤها، لتكون متماثلة في الصعوبة والمحتوى.

وتهدف المعادلة إلى تكوين علاقة بين درجات صورتين للاختبار بحيث يمكن استخدام درجات كل صورة من صور الاختبار كما لو أنها جاءت من نفس الاختبار (Davier, 2011).

كما يمكن اعتبار المعادلة عملية اشتقاق درجات متكافئة من اختبارات مختلفة، فإذا عودلت درجات اختبارينX وY وتقدم مفحوص ما لاختبار X فإنه من الممكن تحديد درجته على الاختبار Y، بشرط أن يقيس الاختبار نفس السمة بدرجات ثبات متساوية لها نفس الرتب المئينية (كروكر والجينا، 2009).

ويرى علام (2005 , ص214) "أن المقصود بتكافؤ صيغتي اختبار معين، تحويل نظام وحدات القياس الخاص بإحدى الصيغتين إلى نظام وحدات القياس الخاص بالصيغة الأخرى، بحيث تصبح القياسات المستمدة من درجات كل من الصيغتين متكافئة بعد إجراء هذا التحويل".

وبعد عرض التعاريف السابقة فإنه يمكن تعريف المعادلة بأنها: إجراء إحصائي لتعديل درجات الصور المختلفة للاختبار، ووضعها ضمن تدريج مشترك Common Scale بحيث تصبح متكافئة، ويمكن استخدامها بالتبادل بعد تحقق شروط المعادلة، ويستفاد منها في تفسير ملائم ذي معنى للمقارنة بين الأفراد، أو بين درجات المفحوص نفسه.

أنواع المعادلة:

حدد علماء القياس والتقويم نوعين من المعادلة يعتمد تحديد كل منهما على حسب مستوى الصعوبة، ومستوى توزيع القدرة لدى المفحوصين، والهدف من إجراء الاختبار، وهذان النوعان هما:

1- المعادلة الأفقية Horizontal Equating: ويرى كل من Hambleton & Swaminathan (1985) أن المعادلة الأفقية تستخدم عندما يكون لدينا اختباران متكافئان في مستوى الصعوبة، وتكون توزيعات السمة المراد قياسها لدى المفحوصين متشابهة، حيث يتم إعداد صور متكافئة من الاختبار نفسه ذات عبارات مختلفة. (أورد في: الدوسري،2001).

ويُحتاج إلى ذلك عندما نريد المحافظة على سرية الاختبار، أو تيسير إعادة الاختبار لمن لم يجتازه.

2- المعادلة العمودية Vertical equating: ويرى كل من Hambleton & Swaminathan(1985) أنها تستخدم عندما يكون لدينا اختباران متفاوتان في مستوى الصعوبة، وتوزيعات السمة المراد قياسها متباينة بين المفحوصين الذين طبق عليهم الاختبار، حيث يصمم اختباران متفاوتان في مستوى الصعوبة، ولتوزيعات مختلفة من السمة المراد قياسها، وينطبق ذلك على بطاريات الاختبارات، وكذلك اختبارات الصفوف المختلفة في مستواها التعليمي (الرابع، الخامس، السادس...). (أورد في: الدوسري،2001).

فهي إذن ملائمة عندما نريد بناء تدريج Scale يقارن بين مستويات المفحوصين في سمة معينة.

شروط معادلة الاختبارات

تم اقتراح العديد من الخصائص الضرورية للمعادلة في أدبيات القياس والتقويم، فبعض هذه الخصائص ركزت على درجات الأفراد؛ حيث أن المفحوص اختبر إحدى صور الاختبار ومن حقه الحصول على نتيجة اختباره بغض النظر عن كيفية استحقاقه لها، وبعض الخصائص ركزت على توزيع الدرجات لمجموعة المفحوصين؛ وأنه يجب أن يحصل المفحوص بنفس النسبة على الدرجة المذكورة له أو أقل ((Kolen & Brennan, 2004.

ويرى الباحث أن الباحثين ينظرون للخصائص على أنها شروط يجب التأكد منها في بناء المعادلة.

وهذه الشروط هي:

1- قياس نفس السمة Same Construct:

فالاختباران يجب أن يقيسا نفس السمة، أو القدرة، أو المهارة، ولا يشترط أن يشتمل الاختباران على مفردات أحادية البعد، ولكن يجب أن يقيسا البعد نفسه؛ بمعنى أن يشتمل الاختباران على مفردات تتناول نفس المزيج من المحتوى، وهو ما يعرف بتوازي المحتوى Content Parallelism، أو توازنه Content Balancing.

2- التعادل Equity:

عندما يكون لدينا اختباران Y،X يقيسان سمة أو قدرة أحادية البعد (θ). فإن شرط التعادل يعني أن التوزيع التكراري المشروط للدرجة (x)، أي أن $f_{x|\theta}$ هو نفس التوزيع التكراري المشروط $f_{x(y)|\theta}$، حيث x(y) دالة تناظر أحادي تستخدم في تحويل درجات الاختبار (y) إلى ميزان Scale الاختبار (x).

3- التماثل Symmetry:

تتطلب هذه الخاصية أن تكون التحويلات باستخدام المعادلة قابلة للانعكاس، بحيث لو تم تحويل الدرجات من النموذج X إلى ميزان الدرجات في النموذج Y، ثم تم تحويل الدرجات من النموذج Y إلى ميزان الدرجات في النموذج X فإن ناتج المعادلة يبقى كما هو. (Kolen & Brennan, 2004).

4- استقلالية دالة التحويل عن البيانات المستمدة من مجتمع معين Invariance:

وهذا يعني أن دالة التحويل المستخدمة في تحويل درجات اختبار ما إلى ميزان اختبار آخر لا تتغير بتغير مجتمع الأفراد المفحوصين، وذلك لأن عملية التحويل Transformation تعد علاقة تناظر أحادي بين الدرجات.

تصاميم جمع البيانات لإجراء المعادلة:

توجد أربعة من تصاميم الربط بين الاختبارات لإجراء المعادلة والمقارنة بين المفحوصين فيما بعد، والتي تسمح بمقايسة معالم المفردات (أو تقديراتها) وهي كما يلي(Hambleton, Swaminathan, & Rogers, 1991):

1- تصميم المجموعة الواحدة Single-Group Design:

ويتميز هذا التصميم بيسره وسهولته في التطبيق حيث يتم تطبيق صورتي الاختبار المراد ربطهما على نفس مجموعة المختبرين، ولكن يعاب عليه أنه غير عملي في التطبيق حيث يتطلب وقتاً زمنيًّا طويلاً، إضافة إلى ذلك فإن أثر الممارسة والتدريب والتعب قد تؤثر على تقدير المعالم إذا تم تطبيق الاختبارين متتالين في نفس الوقت.

2- تصميم المجموعات المتكافئة Equivalent-Groups Design:

وفيه يتم تطبيق صورتي الاختبار على مجموعتين متكافئتين – تم تعيينها عشوائيًّا -ولكنها ليستا متماثلتين في القدرة، ويعتبر هذا التصميم عمليًّا أكثر ويمكن من خلاله تجنب تأثيرات التعب والممارسة.

3- تصميم المجموعات غير المتكافئة ذات الجذع المشترك Design Anchor-Test:

لتطبيق هذا التصميم يقسم المفحوصون إلى مجموعتين مختلفتين، ثم تطبق الصورة الأولى للاختبار على المجموعة الأولى وتطبق الصورة الثانية للاختبار على المجموعة الثانية، وتشتمل صورتا الاختبار على مفردات مشتركة يمكن أن تكون داخلية – ضمن عبارات الاختبار- أو خارجية – في اختبار مستقل- ويعتبر هذا التصميم عمليًّا إذا تم اختيار مفردات الجذع المشترك بدقة، ويمكن من خلاله تجنب عيوب استخدام تصميمات المجموعة الواحدة والمجموعات المتكافئة.

4- تصميم مجموعة الأفراد المشتركين Common-Person Design:

يتم تطبيق صورتي الاختبار على مجموعتين مختلفتين بحيث تختبر كل مجموعة إحدى صور الاختبار، مع وجود مجموعة مشتركة من المختبرين تطبق عليهم الصورتان معاً، ونظراً لأن زمن الاختبار سيكون طويلاً بالنسبة للمجموعة المشتركة فإن هذا التصميم له نفس عيوب تصميم المجموعة الواحدة.

طرق معادلة الاختبارات:

يصنف الباحثون والمختصون في مجال القياس والتقويم التربوي طرق معادلة درجات الاختبارات إلى (Kolen & Brennan, 2004):

أ- **طرق تتبع النظرية الكلاسيكية في القياس (CTT،Classical Tests Theory)** وتتضمن الطرق التالية:

- معادلة المتوسط الحسابي (Mean Equating).
- طرق المعادلة الخطية ومنها: طريقة تكر الخطية (Tucker Liner Method)، طريقة ليفين للعلامات المشاهدة (Levine Observed Score Method)، وطريقة براون-هولاند الخطية (Braun-Holland Linear Method).
- طريقة المعادلة المئينية (Equipercentile Equating).

ب- **طرق تتبع النظرية الحديثة في القياس (IRT ،Item Response Theory).**

وتتضمن الطرق التالية:

- طريقة معادلة العلامات الحقيقية (IRT true score equating)

- طريقة معادلة العلامات المشاهدة (IRT observed score equating)

تقييم دقة المعادلة

يوجد العديد من المحكات للاستدلال على دقة نتائج معادلة علامات إحدى صورتي اختبار على الصورة الأخرى ومنها:

الخطأ المعياري للمعادلة Standard Error Of Equating

يرى ثورنديك أن الخطأ المعياري للمعادلة يمثل الانحراف المعياري للعلامات المعادلة بالنسبة لعينة من المفحوصين خلال عملية المعادلة. وتختلف صيغ إيجاد قيم الخطأ المعياري باختلاف تصميم جمع البيانات المستخدم.

الصدق التقاطعي Cross Validation

يمثل محك الصدق التقاطعي متوسط مربعات الانحرافات، وهو يشير إلى مدى استقرار النتائج المشتقة من المعادلة

ثانياً- الدراسات السابقة:

بالرجوع إلى الدراسات التي تناولت موضوع معادلة درجات الاختبارات Equating نجد أنها اختلفت فيما بينها من حيث الهدف من الدراسة، فبعضها كان الهدف منها هو بحث فعالية الطرق الإحصائية المختلفة للمعادلة، وذلك من خلال مقارنة نتائجها ومدى دقتها، والبعض الآخر كان يهدف إلى دراسة أثر بعض المتغيرات على نتائج المعادلة، وعليه سيتم عرض الدراسات لهذين الهدفين لعلاقتها بموضوع الدراسة.

أ- دراسات اهتمت بمقارنة الطرق الإحصائية المختلفة لمعادلة درجات الاختبارات.

- **يرى شولو وسيرس (2011)** أنه قد تتشابه أو تختلف نتائج عملية المعادلة باستخدام الطرق الإحصائية المختلفة (أورد في: المحرزي، 2015).
- **أكد يانغ (Yang W. L., 1997)** في دراسة والتي هدفت إلى التعرف على أثر استخدام ثلاث طرق للمعادلة، وهي طريقة توكر الخطية وطريقتان تعتمدان على نظرية الاستجابة للمفردة، باستخدام تصميم الجذع المشترك الممثل للمحتوى على دقة المعادلة، حيث قام بتحليل البيانات الناتجة عن شكلين من أشكال أحد الاختبارات المهنية، وقد بينت النتائج أن دقة المعادلة باستخدام الطريقتين المعتمدتين على نظرية الاستجابة للمفردة كانت أكثر دقة من طريقة المعادلة الخطية، وأن دقة المعادلة تعتمد على مدى تمثيل عبارات اختبار الجذع المشترك للمحتوى بغض النظر عن طريقة المعادلة المستخدمة.
- **وفي دراسة أبولبدة (1993)** والتي هدفت إلى بناء مقياس متعدد المستويات للأداء العقلي للأطفال الأردنيين من سن (6-12) والذي يقيس ثلاث قدرات في بناء القدرة العقلية، حيث تكونت عينة الدراسة من (1800) طالباً وطالبة، واعتمدت الدراسة تصميم الجذع المشترك والمجموعات غير المتكافئة لجمع البيانات، واستخدم الباحث طريقتي المعادلة الخطية والمئينية، أظهرت النتائج أن المعادلة الخطية أعطت خطأ معيارياً أقل من المعادلة المئينية، وبالتالي كانت أكثر فاعلية، وأن الطريقتين أعطتا نتائج متقاربة عند الوسط الحسابي للعلامات، بينما كان هناك فرق كبير في العلامات المتطرفة.

- **أما في دراسة أيوب (1993)** عندما قارن بين أربع طرق لمعادلة الاختبارات وهي الطريقة الخطية، والطريقة المئينية المنبثقة من النظرية الكلاسيكية، وطريقة نماذج أحادي المعلمة، وطريقة ثنائي المعلمة المنبثقة من النظرية الحديثة، ولتحقيق أهداف الدراسة تم بناء ثلاثة اختبارات كل اختبار بصورتين، وكانت العينة الكلية مكونة من عينتين مستقلتين الأولى بها (1390) طالباً وطالبة والثانية بها (1412) طالباً وطالبة، حيث استخدم الباحث تصميم المجموعات العشوائية والتصميم القائم على اختبار الجذع المشترك وكذلك تصميم المجموعات غير المتكافئة لأغراض جمع البيانات، فأشارت النتائج فيما يخص المعادلة الأفقية إلى أن نماذج النظرية الحديثة في القياس كانت أكثر فاعلية من طريقتي النظرية الكلاسيكية.
- **وفي دراسة الشريفين (2003)** التي هدفت إلى الكشف عن مدى تحقق معايير الفاعلية في معادلة اختبارين أحدهما ثنائي التدريج والآخر متعدد التدريج وفق نماذج النظرية الكلاسيكية والنظرية الحديثة في القياس، تكونت عينة الدراسة من (1003) من الطلاب والطالبات، استخدم الباحث معيارين للحكم على فاعلية طرق المعادلة المستخدمة وهما: معيار الصدق التقاطعي ومعيار الخطأ المعياري للمعادلة، أظهرت النتائج أن طريقة النموذج أحادي المعلمة كان أكثر فاعلية من طريقتي المعادلة الخطية والمئينية وفق معيار الصدق التقاطعي، أما وفق معيار الخطأ المعياري فقد كانت المعادلة الخطية الأكثر فاعلية وأنتجت أقل خطأ معياري.
- **وفي دراسة للطراونة (2004)** عن بناء اختبار للرياضيات في أربعة مستويات للصفوف الأساسية (3-6) وبعبارات متعددة التدريج، تكونت عينة الدراسة من (1279) طالباً وطالبة موزعين في أربعة صفوف، واستخدمت الدراسة ثلاث طرق للمعادلة لاشتقاق المقياس المتعدد المستويات وهي: المعادلة الخطية، والمعادلة المئينية، والنموذج أحادي المعلمة، تبعاً لتصميم اختبار الجذع المشترك للمجموعات غير المتكافئة، أظهرت النتائج أن المعادلة الخطية كانت أكثر فاعلية من المعادلة المئينية وفق محك الخطأ المعياري، ولم يدخل النموذج الأحادي المعلمة في المقارنة لعدم توفر برنامج مناسب، أما وفق محك الصدق التقاطعي للمعادلة فكان النموذج أحادي المعلمة أكثر فاعلية، تليه المعادلة الخطية، ثم المعادلة المئينية.

ب- دراسات اهتمت بدراسة أثر بعض المتغيرات على نتائج المعادلة.

- سيتم التركيز على المتغيرات المتعلقة بتصميم الجذع المشترك، والتي لها علاقة مباشرة بالدراسة مثل عدد عبارات الجذع المشترك، وتمثيل عبارات الجذع المشترك للمحتوى، وتجانس عبارات الجذع المشترك وغيرها، مع تطرق يسير لمتغير حجم العينة.
- فعند عرض الدراسات التي تناولت متغير تمثيل عبارات الجذع المشترك للمحتوى نجد **أن يانغ (Yang W. L., 1997)** أشار في دراسته على أن دقة المعادلة تعتمد على مدى تمثيل عبارات الجذع المشترك للمحتوى، بغض النظر عن طريقة المعادلة المستخدمة، وكذلك كلٌ من كولن وبرنان **(Brennan & Kolen, 1987)** أكدا على أهمية تمثيل عبارات الجذع المشترك للاختبار الكلي، كما أكد على ذلك يانغ في دراسته (yang, 2000).
- أهمية تجانس المحتوى في الجذع المشترك نجد أن يانغ **(yang، 2000)** أشار في دراسته التي هدفت إلى الكشف عن أثر تمثيل تجانس محتوى اختبار الجذع المشترك وطريقة المعادلة على دقة المعادلة، والتي استخدم بها بيانات أحد الاختبارات الطبية، وكانت العينة مكونة من (1091) طالباً وطالبة تقدموا للصورة الأولى للاختبار، و(1149) طالباً وطالبة تقدموا للصورة الثانية، أظهرت النتائج أن جميع طرق المعادلة المستخدمة كانت ذات دقة متوسطة، وكانت أكثر دقة عندما كان محتوى اختبار الجذع المشترك أكثر تجانساً.
- وكذلك **أشار حمادنة (2012)** على أن تجانس المحتوى في اختبار الجذع المشترك أسهم في دقة المعادلة المئينية.

- **فون وكونغ (Von & Kong, 2005)** في دراستهم التي هدفت إلى استقصاء الخطأ المعياري للمعادلة والدلالة الإحصائية للفروق في الخطأ، والتي استخدم بها الباحثان تصميم المجموعات غير العشوائية القائم على الجذع المشترك وطرق المعادلة الخطية وعينات ذات أحجام مختلفة، خلصت الدراسة إلى أنه إذا كان حجم العينة (10، 000) فرد فإنه يمكن تقدير المعالم بدقة كبيرة، وبالتالي ستكون الفروق قليلة، وكلما صغر حجم العينة يصبح الفرق أكبر وخاصة إذا كانت هناك قيم متطرفة.
- **وكذلك أشارت دراسة فيتزباتريك ووندي (Fitzpatrick & Wendy, 2001)** في نتائجها إلى أن دقة التقدير تكون مقبولة لمعلم العبارة عند استخدام عينة ذات حجم 200 طالب، وأكبر دقة نحصل عليها باستخدام عينة عدد أفرادها 1000 طالب.

تعليق على الدراسات السابقة:

وبناءً على ما سبق تم التوصل إلى الاستنتاجات التالية:

1- يوجد اختلاف في نتائج المقارنة بين طرق المعادلات، وقد يعود ذلك للمشكلات التي تواجه عملية المعادلة وهي أخطاء المعادلة، أو الخطأ المعياري في المعادلة، أو عدم التناظر في شروط التصميم، أو البيانات الحقيقية وشروط المحتوى، كما أشار إليها كل من كولن وبرنان في دراستهم (Brennan & Kolen, 1987).

2- أن أهم المتغيرات التي تأثر في دقة المعادلة هو تمثيل عبارات الجذع المشترك للمحتوى، وكذلك زيادة عدد العبارات وتجانسها في الجذع المشترك.

3- نجد في الدراسات التي استخدمت معيار الخطأ المعياري للمقارنة بين الطريقة الخطية والطريقة المئينية، أن الطريقة الخطية تكون أقل خطأً معياريًّا.

4- عدم دراسة متغير توزيع القدرة وأثره على دقة المعادلة حسب علم الباحث.

ونتيجة لذلك فإن الدراسة الحالية سوف تقوم بدراسة مقارنة بين فاعلية الطرق الثلاث للنظرية التقليدية (الكلاسيكية)، وهي الطريقة المئينية والطريقة الخطية وطريقة المتوسط، وذلك من خلال المقارنة بينها حسب معيار الخطأ المعياري، واستخدام أسلوب المحاكاة لتوليد البيانات وفق شروط محددة، في ضوء اختلاف حجم العينة وتوزيع القدرة، حيث تم صياغة الفروض الصفرية التالية.

3- منهجية الدراسة وإجراءاتها.

منهج الدراسة:

استخدم الباحث المنهج التجريبي للإجابة على تساؤلات بحثه، والمتعلقة بدراسة أثر كل من طريقة معادلة الدرجات، وحجم العينة، وتوزيع القدرة، على تقدير معدلات الخطأ المعياري للمعادلة.

تصميم الدراسة:

تهتم الدراسة بتقدير معدلات الخطأ المعياري، وذلك عند ضبط فوارق حجم العينة، وتوزيع القدرة، لتقويم فاعلية الطرق الإحصائية الثلاث المستخدمة لمعادلة الدرجات، وباستخدام التصميم التجريبي للقياسات المتكررة لثلاثة عوامل تجريبية. وفي هذا التصميم المتغير التابع هو تقدير الخطأ المعياري للمعادلة، والمتغيرات المستقلة هي طرق معادلة الدرجات (ثلاث طرق) والذي يمثل متغير القياسات المتكررة، وحجم العينة (ثلاثة أحجام)، وتوزيع القدرة لمجموعتين حسب نوع التوزيع المنتظم أو المعتدل، وحيث إن التصميم المستخدم من النوع (Two between

One within fector) لذا ينتج تصميم عاملي (3×3×2) وللتحكم في المتغيرات المستقلة بمستوياتها المختلفة وضبطها والحصول على البيانات اللازمة تم استخدام دراسة المحاكاة Simulation Study.

وقد استخدم الباحث أساليب المونت كارلو A monte Carlo Procedures لأنها عملية ومنطقية مقارنة بالطرق البديلة التي تستخدم بيانات حقيقية، حيث يمكن التحكم في خصائص العبارات من حيث الصعوبة، والتمييز، والتخمين، إضافة إلى ذلك فإنها تمكن الباحث من التحكم بأحجام العينات ونوع توزيع القدرة.

وفي هذه الدراسة تم التحكم في ثلاثة عوامل أشارت الدراسات السابقة أنها تؤثر في معدلات الخطأ المعياري، وهذه العوامل تمثل المتغيرات المستقلة في الدراسة وهي كما يلي:

أولاً: طرق معادلة الدرجات، حيث تم اختيار ثلاث طرق للمعادلة تتبع النظرية التقليدية (تكر الخطية، ليفين للعلامات المشاهدة، والمعادلة المئينية غير المهدة)، وفق تصميم المجموعات غير المتكافئة مع عبارات مشتركة.

ثانياً: طريقة توزيع قدرات المفحوصين (θ)، وله مستويان:

- توزيع طبيعي Normal: وفيه تتوزع قدرات المفحوصين توزيعاً اعتداليًا، بمتوسط حسابي صفر وانحراف معياري 1.
- توزيع منتظم:Uniform والذي تتوزع فيه قدرات المفحوصين توزيعاً منتظماً، بحيث تتراوح القدرات بين$\theta = 3$ و$\theta = -3$.

ثالثاً: أحجام العينات، حيث تم اختيار ثلاثة أحجام للعينات وهي 500، 1000، 1500.

وفي ضوء المتغيرات الثلاثة بمستوياتها المختلفة كان تصميم الدراسة ($3 \times 2 \times 3$) وبعدد 18 خلية، كما هو موضح في جدول (1).

جدول (1) خلايا التصميم متضمنة المتغيرات المستقلة في الدراسة وعدد مرات التكرار

التسلسل	الرمز	حجم العينة	توزيع القدرة	عدد العبارات	عدد العينات للطريقة		
					ليفين للمشاهدات المتكررة	تكر الخطية	المئينية غير المهدة
1	AAA	500	طبيعي	116	100	100	100
2	AAB	500	منتظم	116	100	100	100
3	AAC	1000	طبيعي	116	100	100	100
4	ABA	1000	منتظم	116	100	100	100
5	ABB	1500	طبيعي	116	100	100	100
6	ABC	1500	منتظم	116	100	100	100

أما بالنسبة للمتغيرات التابعة فهي معيار الحكم على دقة معادلة درجات الصور المختلفة للاختبار، وهي معدلات الخطأ المعياري للمعادلة.

توليد بيانات الاستجابة على العبارات ومعادلة الدرجات:

الخطوات التي يتم توليد البيانات من خلالها هي:

1- تقدير معالم العبارات:

تم الحصول على معالم عبارات نموذجين من نماذج اختبار القدرات العامة والذي يقدمه مركز القياس الوطني بالمملكة العربية السعودية، وهي النموذج (1030) والنموذج (1031)، وعدد العبارات لكل نموذج (120)،

وينقسم الاختبار إلى جزأين بحسب نوع القدرة المقاسة في الاختبار، وهي القدرة اللغوية وتمثلها (66) عبارة من الاختبار، والقدرة الرياضية وتمثلها (50) عبارة من الاختبار عبارة، وقد تم حذف (4) عبارات من كل واحد منهما بسبب أنها غير جيدة إحصائيًّا، ويضح الجدول (2) أرقام العبارات المستبعدة موزعة حسب نموذج الاختبار ونوع القدرة.

جدول (2) أرقام العبارات المستبعدة من نموذجي الاختبار

نموذج الاختبار	أرقام العبارات المستبعدة	
	1030	1031
القدرة اللغوية	25، 49	56، 60
القدرة الرياضية	5، 39	30، 39

فأصبح عدد العبارات (116) (بالخيور، 2014).

2- اختيار العبارات المشتركة:

تم وضع عبارات مشتركة ضمن أسئلة النموذجين، وذلك للربط بينها بحيث تكون المعادلة وفق تصميم المجموعات غير المتكافئة مع عبارات مشتركة، ويوضح الجدول (3) أرقام العبارات المشتركة لكل جزء من أجزاء الاختبار:

جدول (3) أرقام المفردات المشتركة بين نموذجي الاختبار

نموذج الاختبار	أرقام المفردات المشتركة	
نوع القدرة	النموذج (1030)	النموذج (1031)
الجزء الأول القدرة اللغوية	29	30
	4	31
	34	36
	40	13
	41	65
	52	54
	64	14
الجزء الثاني القدرة الرياضية	70	96
	72	114
	82	73
	83	105
	97	71
	101	94
	110	95
مجموع المفردات المشتركة	14	14

3- توليد قدرات الأفراد:

تم توليد θ والممثلة لقدرات الأفراد باستخدام برمجية wingen2 وفقاً للتوزيع المعتدل بمتوسط صفر وانحراف معياري واحد، وبالمثل للتوزيع المنتظم بمدى ± 3.

4- توليد استجابات المفحوصين:

تم توليد استجابات كل مفحوص من أفراد العينة وفقاً لمستويات توزيع القدرة وحجم العينة، حيث تستخدم قدرة كل مفحوص على الصورة الأولى، ومعالم كل عبارة في النموذج اللوجستي ثنائي المعالم لتحديد احتمال الإجابة على العبارة 1 أو صفر، والذي يقوم بتوليد الاستجابات برمجية wingen3، وبالمثل يتم توليد الاستجابات للمفحوصين على الصورة الثانية، وبالتالي يكون لكل مفحوص على الصورتين الأولى والثانية درجة كلية ودرجة على عبارات الجذع المشترك.

5- تكرار توليد الاستجابات للمفحوصين:

يقوم الباحث بتكرار توليد الاستجابات بنفس البرمجية، باستخدام معالم العبارات ومعالم قدرات المفحوصين 100 مرة في كل خلية من خلايا التصميم.

6- تخزين استجابات المفحوصين في كل عينة وفي كل مرة تكرار لكلا الصورتين.

7- تحويل الملفات إلى برنامج SPSS لصورتي الاختبار الأولى والثانية، وذلك لأجل التعامل مع البيانات وإجراء المعادلة، حيث يتم حساب الدرجة الكلية والدرجة في عبارات الجذع المشترك لكل مفحوص في كل عينة على الصورة الأولى وبالمثل للصورة الثانية.

8- يتم إجراء المعادلة وفقاً لثلاث طرق من طرق النظرية التقليدية وهي معادلة تكر الخطية، ومعادلة ليفين للعلامات المشاهدة، والمعادلة المئينية غير المهدة باستخدام برنامج (CIPE_wg_v20)

9- حساب معدلات الخطأ المعياري لكل طريقة من طرق المعادلة الثلاث.

الأسلوب الإحصائي المستخدم:

ونظراً لأن تصميم الدراسة من النوع المختلط، وهو عبارة عن Three fector Experiment with Repeated Measures، لذا استخدم الباحث أسلوب تحليل التباين للقياسات المتكررة، وذلك لتحليل بيانات الدراسة والإجابة على تساؤلاتها

4- نتائج الدراسة ومناقشتها.

أولاً: عرض نتائج التحليل الإحصائي لبيانات الدراسة:

حيث تمت معادلتها وحساب معدلات الخطأ المعياري لكل طريقة من طرق المعادلة، ومن ثم تم تحليلها باستخدام أسلوب تحليل التباين المختلط للقياسات المتكررة للإجابة على تساؤلات الدراسة والتحقق من فرضياته، وقد تم استخدام أسلوب تحليل التباين المختلط وفقاً للخطوات التالية:

1. تم إجراء تحليل التباين للقياسات المتكررة لدراسة الخطأ المعياري لطرق المعادلة:

ولإجراء ذلك استخدم التحليل متعدد المتغيرات المتمثل في اختبار روي (Roy's) وكانت النتائج كما هو موضح في الجدول (4):

جدول (4) نتائج تحليل التباين للخطأ المعياري للمعادلة باستخدام التحليل متعدد المتغيرات من خلال اختبار روي

الآثار	قيمة اختبار روي	قيمة اختبار ف	درجات حرية الفرضيات	درجات حرية الخطأ	مستوى الدلالة	مربع إيتا الجزئي
الطريقة	4.683	1627.293	2	695	صفر	0.824
تفاعل الطريقة مع القدرة	1.286	446.872	2	695	صفر	0.563
تفاعل الطريقة مع حجم العينة	0.322	112.011	2	696	صفر	0.243
تفاعل الطريقة وتوزيع القدرة وحجم العينة	0.103	35.871	2	696	صفر	0.093

2. استخدام النتائج الواردة في الجدول السابق في الإجابة عن تساؤلات الدراسة والتحقق من فروضه كالتالي:

أ- للتحقق من صحة الفرض الأول "لا تختلف معدلات الخطأ المعياري للمعادلة باختلاف طرق المعادلة".

ولقبول هذا الفرض أو رفضه تم استخدام نتائج التحليل الواردة في الجدول (4) والخاصة بالطريقة، حيث يلاحظ أن قيمة اختبار روي بلغت (4.683) وهي قيمة دالة إحصائيًا عند مستوى دلالة أقل من (0.001) مما يعني وجود فروق ذات دلالة إحصائية، وللتأكد من الدلالة العملية تم إيجاد حجم الأثر باستخدام قيمة مربع إيتا الجزئي والتي بلغت (0.824)، وهي قيمة تدل على أن حجم الأثر كبير.

ومما سبق يتأكد للباحث اختلاف معدلات الخطأ المعياري لطرق المعادلة الثلاثة مما يعني رفض الفرض الصفري، وقبول الفرض البديل الذي ينص على أنه " تختلف معدلات الخطأ المعياري للمعادلة باختلاف طرق المعادلة".

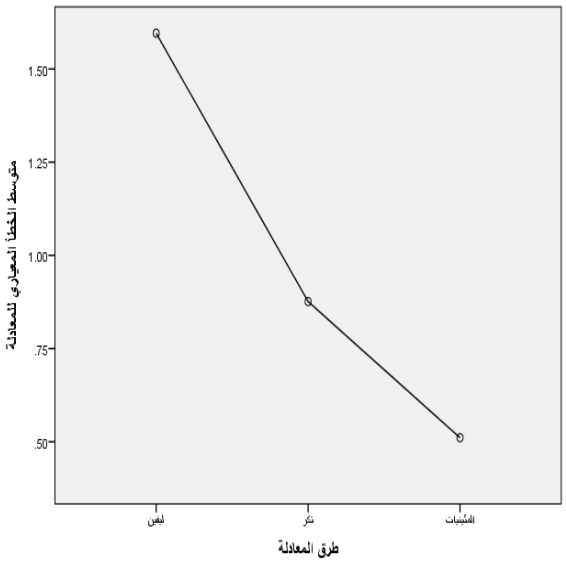
وللمقارنة بين معدلات الخطأ المعياري للطرق الثلاث للمعادلة تم إجراء المقارنات باستخدام اختبار بونفيروني كما في جدول (5).

جدول (5) نتائج المقارنات بين معدلات الخطأ المعياري لطرق المعادلة الثلاث باستخدام اختبار بونفيروني

الطريقة (I)	المتوسط	الطريقة (J)	الفرق بين المتوسطات (I-J)	مستوى الدلالة
ليفين للعلامات المشاهدة	1.596	تكر الخطية	0.720	صفر
		المئينيات غير المهدة	1.085	صفر
تكر الخطية	0.876	ليفين للعلامات المشاهدة	0.720-	صفر
		المئينيات غير المهدة	0.365	صفر
المئينيات غير المهدة	0.511	ليفين للعلامات المشاهدة	1.085-	صفر
		تكر الخطية	0.365-	صفر

يتضح من النتائج في الجدول (5) وجود فروق دالة إحصائيًا بين متوسطات الخطأ المعياري لجميع الطرق، وبالتالي نجد أن الترتيب التصاعدي لطرق المعادلة وفقاً لمتوسطات الخطأ المعياري يكون كالتالي: طريقة المئينيات غير المهدة، ثم طريقة تكر الخطية، ثم طريقة ليفين للعلامات المشاهدة.

ويؤكد النتائج السابقة الرسم البياني الموضح بالشكل (4)



الشكل (4) الرسم البياني لمتوسطات الخطأ المعياري لطرق المعادلة الثلاثة

ب- التحقق من صحة الفرض الثاني:

والذي ينص على أنه "لا تختلف معدلات الخطأ المعياري للمعادلة باختلاف التفاعل الثنائي بين الطريقة وتوزيع القدرة" ولأجل قبوله أو رفضه تم استخدام نتائج التحليل الواردة في الجدول (4) الخاصة بتفاعل الطريقة وتوزيع القدرة حيث يلاحظ أن قيمة اختبار روي قد بلغت (1.286) وهي قيمة دالة إحصائيًا عند مستوى دلالة أقل من (0.001)، مما يعني وجود فروق ذات دلالة إحصائية، وللتأكد من وجود دلالة عملية تم إيجاد قيمة مربع إيتا الجزئي، والتي بلغت (0.563) وهي تدل على أن حجم الأثر كبير.

والنتائج السابقة تؤكد أن معدلات الخطأ المعياري للمعادلة يختلف باختلاف التفاعل الثنائي بين الطريقة وتوزيع القدرة للمجموعتين المرجعية والمستهدفة، مما يعني رفض الفرض الصفري وقبول الفرض البديل والذي ينص على أنه " تختلف معدلات الخطأ المعياري للمعادلة باختلاف التفاعل الثنائي بين الطريقة وتوزيع القدرة".

وللتعرف على طبيعة التفاعل وأثره على معدلات الخطأ المعياري، وللمساعدة في تفسير النتائج تم إيجاد معدلات الخطأ المعياري لكل طريقة وفقاً لتوزيع القدرة، كما هو موضح في جدول (6).

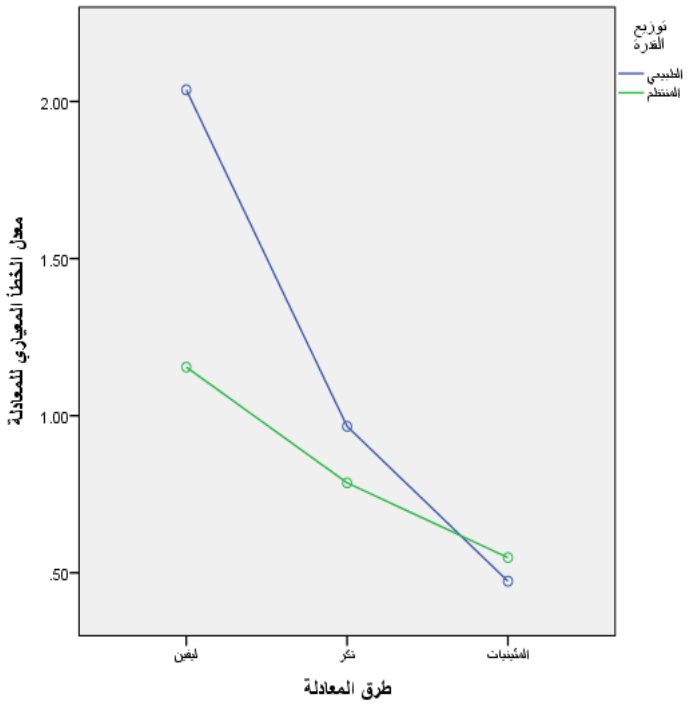
جدول (6) نتائج المقارنات بين معدلات الخطأ المعياري لطرق المعادلة وتوزيع القدرة باستخدام اختبار بونفيروني

توزيع القدرة	طرق معادلة الدرجات		
	ليفين للعلامات المشاهدة	تكر الخطية	المئينيات غير المهدة
الطبيعي	2.0371	0.9661	0.4733
المنتظم	1.1548	0.7863	0.5485

يلاحظ من الجدول (6) أن معدلات الخطأ المعياري للمعادلة بطريقتي ليفين للعلامات المشاهدة وتكر الخطية تقل مع توزيع القدرة المنتظم، بينما تزداد مع توزيع القدرة الطبيعي وعلى العكس نجد في طريقة المئينيات غير المهدة أن معدل الخطأ يقل مع التوزيع الطبيعي للقدرة ويزداد مع التوزيع المنتظم.

كما يلاحظ أنه يمكن ترتيب معدلات الخطأ المعياري للطرق الثلاث تصاعديًا - من أقل معدل خطأ إلى أعلى معدل خطأ - وفق كلا التوزيعين الطبيعي والمنتظم كالآتي: أولاً طريقة المئينيات غير المهدة ثم طريقة تكر الخطية ثم طريقة ليفين للعلامات المشاهدة.

ويؤكد النتيجة السابقة الرسم البياني لتفاعل معدلات الخطأ المعياري للمعادلة بطرقها مع توزيع القدرة والموضح بالشكل (5)



الشكل (5) الرسم البياني لتفاعل طريقة المعادلة وتوزيع القدرة

ج- التحقق من صحة الفرض الثالث: "لا تختلف معدلات الخطأ المعياري للمعادلة باختلاف التفاعل الثنائي بين الطريقة وحجم العينة".

ولغرض قبوله أو رفضه تم استخدام نتائج التحليل الواردة في جدول (4)، والخاصة بتفاعل الطريقة وحجم العينة، حيث يلاحظ أن قيمة اختبار روي قد بلغت (0.322) وهي قيمة دالة إحصائيًا عند مستوى دلالة أقل من (0.001)، مما يعني وجود فروق ذات دلالة إحصائية.

وللتأكد من وجود دلالة عملية تم إيجاد قيمة مربع إيتا الجزئي والتي بلغت (0.243)، والتي تدل على أن حجم الأثر كبير وأن الدلالة عملية.

والنتائج السابقة تؤكد أن معدلات الخطأ المعياري للمعادلة تختلف باختلاف التفاعل الثنائي بين طريقة المعادلة وحجم العينة، مما يعني رفض الفرض الصفري وقبول الفرض البديل.

وللتعرف على طبيعة التفاعل وأثره على معدلات الخطأ المعياري، تم إيجاد معدلات الخطأ المعياري وفقاً لحجم العينة كما في الجدول (7).

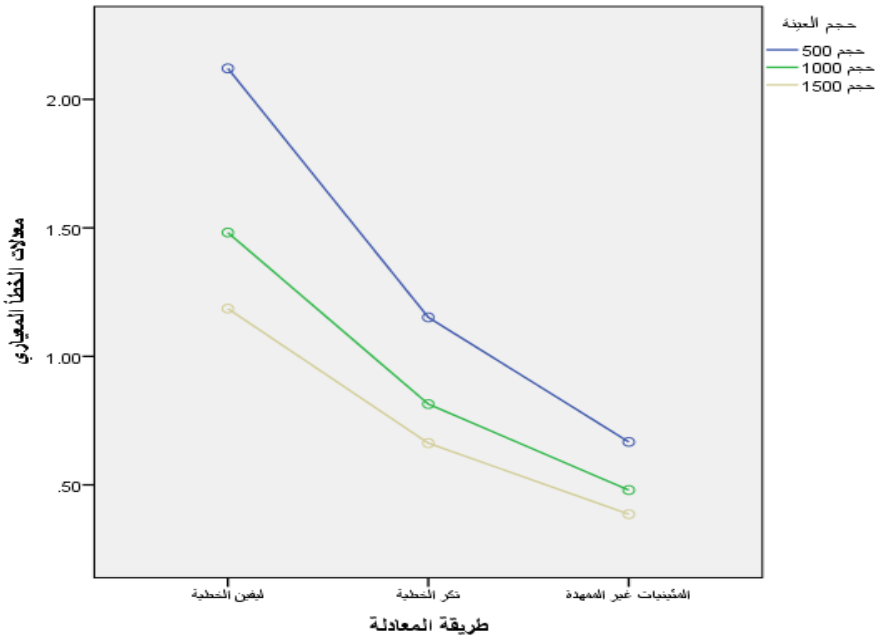
جدول (7) مقارنة معدلات الخطأ المعياري لتفاعل طرق المعادلة وحجم العينة

حجم العينة	طرق معادلة الدرجات		
	ليفين للعلامات المشاهدة	تكر الخطية	المئينيات غير المشاهدة
500	2.1204	1.1519	0.6670
1000	1.4817	0.8144	0.4798
1500	1.1858	0.6624	0.3859

وبالنظر والتدقيق في الجدول (7) نجد أن معدلات الخطأ المعياري لطرق المعادلة الثلاث تتناسب عكسيًّا مع حجم العينة، فكلما زاد حجم العينة قل الخطأ المعياري للمعادلة.

كما يلاحظ أنه يمكن ترتيب معدلات الخطأ المعياري للطرق الثلاث تصاعدياً- من أقل معدل خطأ إلى أعلى معدل خطأ - وفقاً للأحجام الثلاثة (500-1000-1500) كالآتي في طريقة المئينيات غير المهدة، ثم طريقة تكر الخطية، ثم طريقة ليفين للعلامات المشاهدة.

والنتيجة السابقة يوضحها بشكل أكبر الرسم البياني للتفاعل بين طريقة المعادلة وحجم العينة والموضح بالشكل (6).



الشكل (6) الرسم البياني للتفاعل بين طريقة المعادلة وحجم العينة

د- التحقق من صحة الفرض الرابع: "لا تختلف معدلات الخطأ المعياري للمعادلة باختلاف التفاعل الثلاثي بين الطريقة وتوزيع القدرة وحجم العينة".

ولغرض قبوله أو رفضه تم استخدام نتائج التحليل الواردة في جدول (4)، والخاصة بتفاعل الطريقة وتوزيع القدرة وحجم العينة، حيث يلاحظ أن قيمة اختبار روي قد بلغت (0.103) وهي قيمة دالة إحصائيًّا عند مستوى دلالة أقل من (0.001) مما يعني وجود فروق ذات دلالة إحصائية.

وللتأكد من وجود دلالة عملية تم إيجاد قيمة مربع إيتا الجزئي والتي بلغت (0.093)، والتي تدل على أن حجم الأثر متوسط وأن الفروق ذات دلالة عملية أيضاً.

والنتائج السابقة تؤكد أن معدلات الخطأ المعياري للمعادلة تختلف باختلاف التفاعل الثلاثي بين طريقة المعادلة وتوزيع القدرة وحجم العينة، مما يعني رفض الفرض الصفري وقبول الفرض البديل.

وللتعرف على طبيعة التفاعل وأثره على معدلات الخطأ المعياري ومساعدة في تفسير النتائج تم إيجاد معدلات الخطأ المعياري للطرق الثلاث وفقاً لحجم العينة، ولتوزيع القدرة، كما في الجدول (8).

جدول (8) مقارنة معدلات الخطأ المعياري للتفاعل الثلاثي بين طرق المعادلة وتوزيع القدرة وحجم العينة

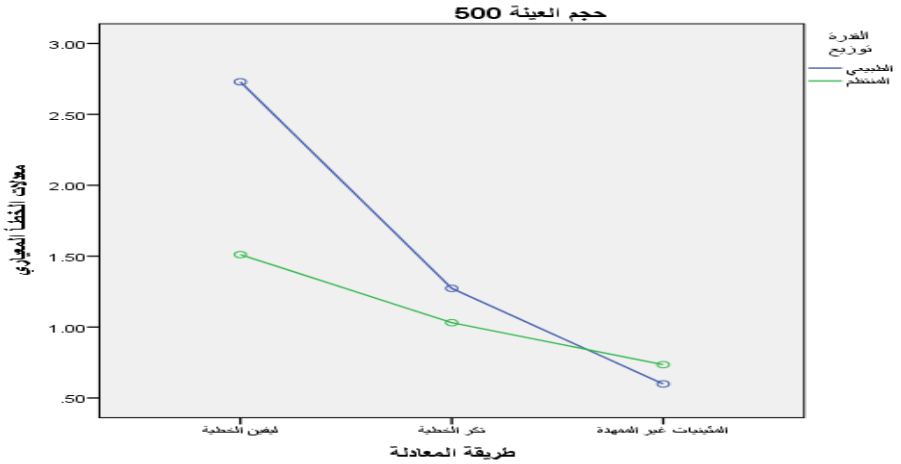
الطريقة	توزيع القدرة	حجم العينة		
		500	1000	1500
ليفين للعلامات المشاهدة	التوزيع الطبيعي	2.7295	1.8886	1.4932
	التوزيع المنتظم	1.5113	1.0747	0.8784
تكر الخطية	التوزيع الطبيعي	1.2728	0.8973	0.7283
	التوزيع المنتظم	1.0309	0.7314	0.5964
المئينيات غير المهدة	التوزيع الطبيعي	0.5990	0.4563	0.3648
	التوزيع المنتظم	0.7351	0.5033	0.4070

بالنظر إلى الجدول يمكن استخلاص النتائج التالية:

1. **في حالة حجم العينة 500:**

يلاحظ أن معدلات الخطأ المعياري لطريقتي المعادلة الخطية تكر وليفين تقل مع التوزيع المنتظم بينما تزداد مع التوزيع الطبيعي، وعلى العكس في طريقة المئينيات غير المهدة نجد أن معدلات الخطأ المعياري تزداد مع التوزيع المنتظم، وتقل مع التوزيع الطبيعي، كما أنها تحقق أقل خطأ مع التوزيع الطبيعي للقدرة إذا ما قورنت بالحالات الأخرى للمعادلة.

ويوضح النتائج السابقة الرسم البياني للتفاعل الثلاثي بين طريقة المعادلة وتوزيع القدرة وحجم العينة 500والموضح في الشكل (7).

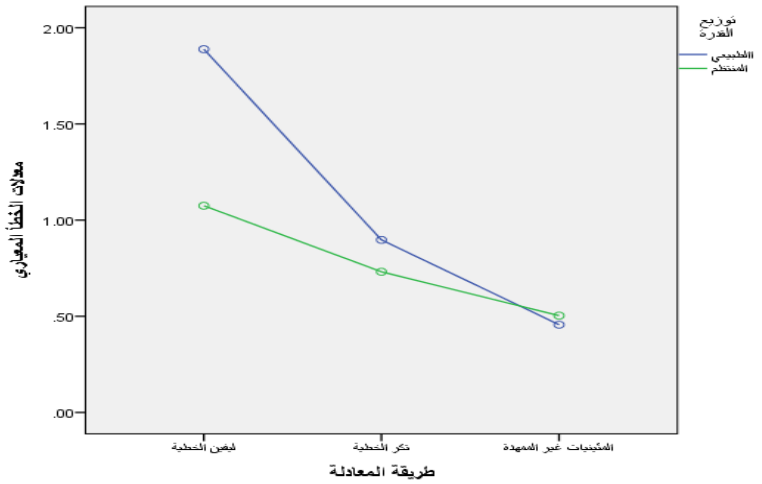


الشكل (7) الرسم البياني للتفاعل الثلاثي بين طريقة المعادلة وتوزيع القدرة وحجم العينة 500

2. **في حالة حجم العينة 1000:**

يلاحظ أن معدلات الخطأ المعياري لطريقتي المعادلة الخطية تكر وليفين تقل مع التوزيع المنتظم بينما تزداد مع التوزيع الطبيعي، وعلى العكس في طريقة المئينيات غير المهدة نجد أن معدلات الخطأ المعياري تزداد مع التوزيع المنتظم وتقل مع التوزيع الطبيعي، كما أنها تحقق أقل خطأ مع التوزيع الطبيعي للقدرة إذا ما قورنت بالحالات الأخرى للمعادلة.

ويوضح النتائج السابقة الرسم البياني للتفاعل الثلاثي بين طريقة المعادلة وتوزيع القدرة وحجم العينة 1000 والموضح في الشكل (8).

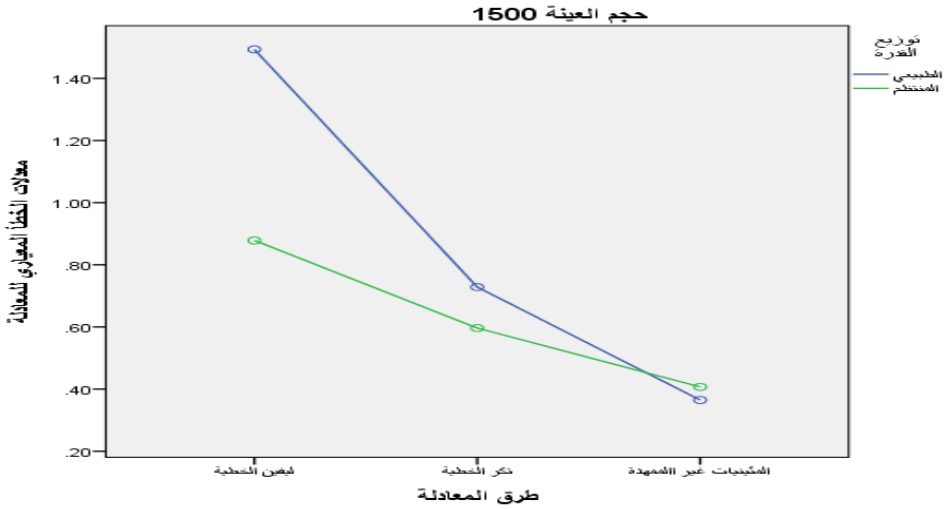


الشكل (8) الرسم البياني للتفاعل الثلاثي بين طريقة المعادلة وتوزيع القدرة وحجم العينة 1000

3. في حالة حجم العينة 1500:

يلاحظ أن معدلات الخطأ المعياري لطريقتي المعادلة الخطية تكر وليفين تقل مع التوزيع المنتظم بينما تزداد مع التوزيع الطبيعي، وعلى العكس في طريقة المئينيات غير المهدة نجد أن معدلات الخطأ المعياري تزداد مع التوزيع المنتظم وتقل مع التوزيع الطبيعي، كما أنها تحقق أقل خطأ مع التوزيع الطبيعي للقدرة إذا ما قورنت بالحالات الأخرى للمعادلة.

ويوضح النتائج السابقة الرسم البياني للتفاعل الثلاثي بين طريقة المعادلة وتوزيع القدرة وحجم العينة 1500 والموضح في الشكل (9).



الشكل (9) الرسم البياني للتفاعل الثلاثي بين طريقة المعادلة وتوزيع القدرة وحجم العينة 1500

ومما سبق يمكن استخلاص نتائج التفاعل الثلاثي لطرق المعادلة مع نوع التوزيع وحجم العينة كما يلي:

1. أن معدلات الخطأ المعياري للطرق الثلاث تقل كلما زاد حجم العينة مع كلا التوزيعين الطبيعي والمنتظم.

2. أنه باختلاف حجم العينة فإن معدلات الخطأ المعياري لطريقتي تكر الخطية وليفين للعلامات المشاهدة تقل مع التوزيع المنتظم وتزداد مع التوزيع الطبيعي للقدرة، بينما نجد أن معدلات الخطأ لطريقة المئينيات غير المهدة تقل مع التوزيع الطبيعي وتزداد مع التوزيع المنتظم.
3. تحقق طريقة المئينيات غير المهدة أقل خطأ معياري إذا كان توزيع القدرة طبيعيًّاً مع اختلاف حجم العينة (500-1000-1500) مقارنة بطريقتي تكر الخطية وليفين للعلامات المشاهدة.
4. يمكن ترتيب معدلات الخطأ المعياري ترتيباً تصاعديًّا - من أقل خطأ إلى أعلى خطأ – كالتالي في طريقة المئينيات غير المهدة مع التوزيع الطبيعي ثم طريقة المئينيات غير المهدة مع التوزيع المنتظم، ثم طريقة تكر الخطية مع التوزيع المنتظم، ثم طريقة تكر الخطية مع التوزيع الطبيعي، ثم طريقة ليفين للعلامات المشاهدة مع التوزيع المنتظم، ثم طريقة ليفين للعلامات المشاهدة مع التوزيع الطبيعي، مع اختلاف حجم العينة (500-1000-1500).

مناقشة وتفسير نتائج التحليل الإحصائي لبيانات الدراسة:

حاول الباحث في دراسته الحالية التحقق من فاعلية ثلاث طرق إحصائية لمعادلة درجات الاختبارات وذلك من خلال مقارنة معدلات الخطأ المعياري لها.

حيث تم تصميم دراسة محاكاة من خلال ضبط أهم المتغيرات التي يعتقد أن لها تأثيراً على فاعلية طرق المعادلة، وهي نوع توزيع القدرة، وأحجام العينة، وقد تم تحليل البيانات المتحصلة من تصميم المحاكاة وعرضها بغرض الإجابة على تساؤلات الدراسة.

وفيما يلي سيتم مناقشة النتائج في ضوء ما أسفر عنه التحليل الإحصائي لبيانات الدراسة، كما سيتم ربطها بما تناولته أدبيات الدراسة في مجالي التربية وعلم النفس، ومقارنة النتائج بنتائج الأبحاث السابقة، مع التعقيب عليها وتفسيرها، وذلك على مستوى جميع الطرق بصفة عامة أولاً، ثم على المستوى التفصيلي لكل طريقة من الطرق.

1- مناقشة وتفسير النتائج الرئيسية بصفة عامة على مستوى جميع الطرق:

تم عرض نتائج التحليل الإحصائي لدراسة الخطأ المعياري لطرق المعادلة بشكل مفصل، وقد أكدت النتائج على ما يلي:

1. اختلاف دقة فاعلية معادلة درجات الاختبارات باختلاف طرق المعادلة المختلفة، من خلال اختلاف معدلات الخطأ المعياري، إذ كانت طريقة المئينيات غير المهدة أفضل الطرق في معادلة الدرجات؛ حيث كان معدل الخطأ المعياري لها أقل من طريقتي المعادلة الخطية.
2. تختلف معدلات الخطأ المعياري باختلاف نوع توزيع القدرة، فنجد أن طريقتي المعادلة الخطية تكر وليفين ينتج عنها معدل خطأ معياري أقل عندما يكون توزيع الدرجات منتظماً بينما يكبر معدل الخطأ المعياري إذا كان توزيع الدرجات طبيعيًّا، وعلى العكس تماماً عندما تعادل الدرجات بطريقة المئينيات غير المهدة نجد أن معدل الخطأ المعياري يقل مع توزيع الدرجات الطبيعي ويزداد مع التوزيع المنتظم.

2- تختلف معدلات الخطأ المعياري لجميع طرق المعادلة باختلاف حجم العينة، حيث يوجد تناسب طردي بين دقة طرق المعادلة وزيادة حجم العينة.

3- مناقشة وتفسير نتائج الدراسة على المستوى التفصيلي لكل طريقة من الطرق

أكدت نتائج الدراسة أن معدلات الخطأ المعياري تختلف باختلاف التفاعل الثلاثي بين طريقة المعادلة وتوزيع القدرة وحجم العينة، لذا سيتم عرض نتائج معدلات الخطأ المعياري للطرق الثلاث وفقاً لتوزيع القدرة وحجم العينة كالتالي:

أولاً: طريقة ليفين الخطية Levine's Linear Method

1- تتناقص معدلات الخطأ المعياري لطريقة ليفين الخطية بشكل مطرد عند زيادة حجم العينة مع كلا التوزيعين الطبيعي والمنتظم.
2- يلاحظ أن معادلة الدرجات ذات التوزيع المنتظم بطريقة ليفين الخطية ينتج عنها معدل خطأ معياري أقل من معدل خطئها المعياري مع الدرجات ذات التوزيع الطبيعي، مع جميع أحجام العينة بالدراسة.
3- كما يلاحظ أن الفرق كبير بين معدلات الخطأ المعياري لمعادلة الدرجات ذات التوزيع المنتظم بطريقة ليفين الخطية، ومعادلة الدرجات ذات التوزيع الطبيعي بنفس الطريقة مع جميع أحجام العينة.

ثانياً: طريقة تكر الخطية Tucker's Linear Method

1- تتناقص معدلات الخطأ المعياري لطريقة تكر الخطية بشكل مطرد عند زيادة حجم العينة مع كلا التوزيعين الطبيعي والمنتظم.
2- يلاحظ أن معادلة الدرجات ذات التوزيع المنتظم بطريقة تكر الخطية ينتج عنها معدل خطأ معياري أقل من معدل خطئها المعياري مع الدرجات ذات التوزيع الطبيعي، مع جميع أحجام العينة بالدراسة.
3- يلاحظ أن الفرق يسير بين معدلات الخطأ المعياري لمعادلة الدرجات ذات التوزيع المنتظم بطريقة تكر الخطية ومعادلة الدرجات ذات التوزيع الطبيعي بنفس الطريقة، مع جميع أحجام العينة إذا ما قورن بذات الفرق مع طريقة ليفين الخطية.

ثالثاً: المعادلة المئينية غير الممهدة Unsmoothed Unequipercentile Equating

1- تتناقص معدلات الخطأ المعياري لطريقة المئينيات غير الممهدة بشكل مطرد عند زيادة حجم العينة مع كلا التوزيعين الطبيعي والمنتظم.
2- يلاحظ أن معادلة الدرجات ذات التوزيع الطبيعي بطريقة المئينيات غير الممهدة ينتج عنها معدل خطأ معياري أقل من معدل خطئها المعياري، مع الدرجات ذات التوزيع المنتظم مع جميع أحجام العينة بالدراسة.
3- يلاحظ أن الفرق ضئيل جدًّا بين معدلات الخطأ المعياري لمعادلة الدرجات ذات التوزيع الطبيعي بطريقة المئينيات غير الممهدة، ومعادلة الدرجات ذات التوزيع المنتظم بنفس الطريقة، مع جميع أحجام العينة إذا ما قورنت بذات الفرق مع طريقة ليفين الخطي

خلاصة الدراسة:

ومن خلال دراسة النتائج والمناقشات والتفسيرات التي تم التوصل إليها، يمكن تلخيص أهم الاستنتاجات للبحث، على النحو الآتي:

1- تختلف معدلات الخطأ المعياري باختلاف طريقة معادلة الدرجات، إذ تعتبر طريقة المئينيات غير الممهدة أفضل طريقة من حيث صغر معدل الخطأ المعياري، تليها طريقة تكر الخطية ثم أخيراً طريقة ليفين الخطية، والنتائج السابقة كانت في حالة المقارنة بين الطرق الثلاث عبر جميع حالات فرق توزيع القدرة ولجميع أحجام العينة.

2- تختلف معدلات الخطأ المعياري لطرق معادلة الدرجات باختلاف التفاعل الثنائي بين طريقة المعادلة وتوزيع القدرة، حيث يزداد الخطأ المعياري لطريقتي المعادلة الخطية تكر وليفين في التوزيع الطبيعي ويقل في التوزيع المنتظم، بينما نجد في طريقة المئينيات غير المهدة يزداد الخطأ المعياري في التوزيع المنتظم ويقل في التوزيع الطبيعي.

3- تختلف معدلات الخطأ المعياري لطرق معادلة الدرجات في التفاعل الثنائي بين طريقة المعادلة وحجم العينة، ويلاحظ أن معدلات الخطأ المعياري تتناسب عكسيًا مع حجم العينة لطرق المعادلة الثلاث؛ فكلما زاد حجم العينة يقل معدل الخطأ المعياري.

4- تختلف معدلات الخطأ المعياري لطرق المعادلة بالتفاعل الثلاثي بين طريقة المعادلة وتوزيع القدرة وحجم العينة حيث نجد أن معدلات الخطأ المعياري لطريقتي المعادلة الخطية تكر وليفين تقل مع التوزيع المنتظم بينما تزداد مع التوزيع الطبيعي، وعلى العكس في طريقة المئينيات غير المهدة نجد أن معدلات الخطأ المعياري تزداد مع التوزيع المنتظم وتقل مع التوزيع الطبيعي، وذلك مع جميع أحجام العينة المختلفة.

التوصيات والمقترحات.

اعتماداً على ما توصل إليه البحث من نتائج يوصي الباحث ويقترح بما يلي:

1- عدم الاقتصار في معادلة درجات الاختبارات على طريقة محددة؛ نظراً لاختلاف فاعلية طرق المعادلة حسب طبيعة النماذج الاختبارية، لعدم وجود طريقة هي الأفضل في كل الحالات.

2- أهمية نشر ثقافة معادلة الدرجات بين المهتمين بإجراء الاختبارات الجمعية، حتى يتم تحقيق العدالة وتكافؤ الفرص بين من تطبق عليهم النماذج الاختبارية، ونضمن الحصول على تفسيرات ملائمة للدرجات، كما نضمن عدم التشكيك في موثوقية نتائج الاختبارات.

3- تدريب المهتمين بإجراء الاختبارات على استخدام البرمجيات التي تسهل عملية إجراء المعادلة بين درجات النماذج الاختبارية.

4- الاهتمام بإنشاء مراكز القياس والتقويم في الجامعات والمؤسسات التعليمية، والتي يوكل إليها الإشراف على بناء الاختبارات بالطرق الصحيحة، وآلية استخراج النتائج، وكيفية إجراء المعادلة للنماذج المختلفة.

5- كما يقترح الباحث إجراء الدراسات التالية:

1. إعادة دراسة المحاكاة الحالية باستخدام نفس العبارات، واستخدام معيار الصدق التقاطعي للمقارنة بين فاعلية طرق المعادلة الثلاث، ومقارنة النتائج بنتائج الدراسة الحالية.
2. إعادة دراسة المحاكاة الحالية باستخدام نفس العبارات، واستخدام معيار مربع الجذر التربيعي، للمقارنة بين فاعلية طرق المعادلة الثلاث، ومقارنة النتائج بنتائج الدراسة الحالية.
3. إعادة الدراسة بتطبيق طرق أخرى لم تستخدم في الدراسة الحالي مثل طريقة المتوسط الحسابي (Mean Equating) وطريقة بروان - هولاند الخطية (Braun-Holland Linear Method).
4. دراسة فاعلية طرق معادلة درجات الاختبارات إذا كان نوع المعادلة المستخدم هو المعادلة الرأسية/ العمودية.

قائمة المراجع

أولاً- المراجع بالعربية:

- أمل خميس الزعابي، راشد سيف المحرزي، وعبد الحميد حسن. (27 3، 2019). معادلة كتيبات اختبار الدراسة الدولية لقياس مهارات القراءة (PIRLS2011) بسلطنة عمان باستخدام نظرية الاستجابة للمفردة.
- إياد محمد حمادنة. (18 7، 2012). أثر تجانس المحتوى وطول الجذع المشترك على دقة معادلة اختبار ثنائي التدريج ومتعدد المستوى في مبحث رياضيات. المنارة، الصفحات 9-43.
- حسين محمد عبد القادر أيوب. (1993). المقارنة بين أربع طرق للمعادلة عندما يكون التصميم من مجموعة متكافئة وغير متكافئة. أطروحة دكتوراة، غير منشورة، الجامعة الأردنية. عمان، الأردن.
- خطاب محمد أبو لبدة. (1993). بناء اختبار متعدد المستويات للأداء العقلي للأطفال الأردنيين من سن (12-6). عمان، الأردن: الجامعة الأردنية.
- راشد حماد الدوسري. (ديسمبر، 2001). معادلة الاختبارات مفهومها وطرقها ومشكلات تطبيقها. العلوم التربوية والنفسية، الصفحات 117-141.
- راشد سيف المحرزي. (2015). المقارنة بين نتائج طرق المعادلة الكلاسيكية لدرجات نماذج اختبار القدرات العامة باستخدام تصميم الجماعات المتكافئة. رسالة الخليج العربي.
- رائد فايز المدانات. (يونيو، 2012). مقارنة فاعلية طريقتي معادلة العلامات الحقيقية والمشاهدة في معادلة الاختبارات باستخدام جذع مشترك ومجموعات متكافئة. العلوم التربوية والنفسية، الصفحات 365-394.
- شفاء عبد الله بالخيور. (2014). فاعلية طرق معادلة نماذج اختبار القدرات العامة بالمركز الوطني للقياس والتقويم وفق نظريتي القياس التقليدية والحديثة في ضوء بعض المتغيرات. مكة المكرمة: جامعة أم القرى.
- صبري حسن الطراونة. (2004). تطوير اختبار رياضيات متعدد المستويات للصفوف الأساسية (6-3) بعبارات متعددة التدريج. أطروحة دكتوراة غير منشورة، جامعة عمان العربية. عمان، الأردن.
- صلاح الدين محمود علام. (2005). نماذج الاستجابة للمفردة الاختبارية. القاهرة: دار الفكر العربي.
- ليندا كروكر، وجيمس الجينا. (2009). مدخل إلى نظرية القياس التقليدية والمعاصرة. (زينات يوسف دعنا، المحرر) عمّان: دار الفكر.
- المدانات، ر. ف .(2008) .أثر طريقة المعادلة باستخدام جذع مشترك وعدد عباراته وحجم العينة على القيم المعادلة والخطأ في المعادلة بين صورتي اختبار في الفيزياء .عمان: جامعة عمان العربية للدراسات العليا.
- موسى محمد كوفحي، وريما فايز زواهرة. (2017). أثر طريقة تصفية المموهات في عبارات الاختيار من متعدد على معادلة الاختبار. صفحة 349.
- نضال كمال الشريفين. (2003). مدى تحقق معايير الفاعلية في معادلة اختبارين أحدهما ثنائي التدريج والآخر متعدد التدريج وفق نماذج النظرية الكلاسيكية والنظرية الحديثة. أطروحة دكتوراة، جامعة عمان العربية للدراسات العليا. عمان، الأردن.
- نضال كمال الشريفين. (يوليو، 2009). معادلة درجات نماذج مختلفة من اختبار الكفاءة اللغوية في اللغة الإنجليزية لدى طلبة جامعة اليرموك. جامعة أم القرى للعلوم التربوية والنفسية، الصفحات 13-62.
- الوليلي. (أكتوبر، 2005). تكافؤ درجات الاختبارات في ضوء نظريتي القياس الكلاسيكية والحديثة. جامعة بنها كلية التربية، الصفحات 99-149.

ثانياً- المراجع بالإنجليزية:

- Alina A von Davier .(2011) .Statistical Models for Test Equating, Scaling, and Linking .New York: Springer.
- Brennan, R. L., & Kolen, M. J. (1987). Some practical issues in equating. Applied Pscychological Measurement, ll.
- Chulu, B. W., & Sireci, S. G. (2011). Importance of equating high stakes educational measurements. Internaltional Journal of Testing, pp. 11, 38-52.
- D J Harris, وM J Kolen .(1990) .Effect of examinee group on equating relationships .Applied Psychological Measurement, .43-35 ،10
- D Von, وN Kong .(2005) .A unified approach to linear equating for "the Non-equivalent group design. Journal of Educational and Behavioral Statistics, .313-342 الصفحات
- Davier, A. A. (2011). Statistical Models for Test Equating, Scaling, and Linking. New York: Springer.
- Fitzpatrick, H., & Wendy, A. (2001). The Effect of test lengh and sample size on the reliability and equating of tests composed of constructed response. items, Monterey, California. CTB/McGraw-Hill.
- G L Marco, N S Petersen, وE E Stewart .(1989) .A test of the adequacy of curvilinear score equating method .Journal of Educational Measurement, (1)12, pp19-32.
- Gialluca, K. A., Crichton, L. I., Yale, C. D., & Ree, M. J. (1983). Methods for Equating Mental Tests. Assessment Systems Corporation.
- Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). Fundamentals of Item Resp0nse Theory. California, Newbury Park: Sage Publications, Inc.
- Kolen, M. J. (1981). Comparison of traditional and item response theory method for equating tests. Journal of Educational Measurement, pp. 18, 1-11.
- Kolen, M. j., & Brennan, R. L. (2004). Test Equating, Scaling, and Linking. New York: Springer.
- Kolen, M. J., & Whitney, D. R. (1982). Comparison of four procedures for equating the Tests of General Educational Development. Journal of Educational Measurement, pp. 19, 279-293.
- R Steven .(1990) .A Comparison of classical and item response theory equating methods using a composite score of direct writing. Dia, Dec.
- Skaggs, S. G., & Lissitz, R. W. (1986). IRT equating: Relevant issues and areview of recent research. Review of Educational Research, pp. 56, 495-529.
- von Davier, A. A., Holland, P. W., & Thayer, D. T. (2004). The Kernel Method of Test Equating. New York: Springer.
- W l Yang, وR T Houang .(1996) .The effect of anchor length and equating method on the accuaracy of test equating: Comparisons of linear and IRT-based equating using an anchor item desing .Paper presented the annual meeting of the American Educational Research Association .New York.

- W yang" .(2000) .The effect of content homogeneity and equating method on the accuracy of common-item test equating ."Paper presented at the annual meeting of American, Educational Research Association .Chicago.
- Yang, W. L. (1997). The Effect of content Mix and Equating Method on the Accuracy of Test Equating Using Anchor - Item Design. Annual Meeting of American Educational Research Association. Chicago, IL.