مجلة جامعة تشرين للبحوث والدراسات العلمية _ سلسلة الآداب والعلوم الإنسانية المجلد (34) العدد (5)2012
Tishreen University Journal for Research and Scientific Studies - Arts and Humanities Series Vol. (34) No. (5) 2013

المقارنة بين الطريقتين الخطية والمئينية لمعادلة درجات صورتي اختبار تحصيلي في مادة الرياضيات للصف الرابع من التعليم الأساسي في مدينة دمشق

الدكتور رمضان محمد درويش*
رانيه جميل رضوان**

(تاريخ الإيداع 8 / 7 / 2012. قبل للنشر في 2 / 10 / 2012)

□ ملخّص □

هدفت الدراسة إلى المقارنة بين الطريقة الخطية والطريقة المئينية لمعادلة درجات صورتي اختبار تحصيلي في مادة الرياضيات للصف الرابع الأساسي، وذلك بالاعتماد على معيار الخطأ المعياري للمعادلة، ولتحقيق ذلك استخدم المنهج الوصفي التحليلي، وأعدت صورتين ذي الاختيار من متعدد وتألفت كل صورة من (24) سؤالاً، حيث جرى التحقق من صدقهما وثباتهما. تكونت العينة من (1150) تلميذاً وتلميذةً من تلاميذ الصف الرابع الأساسي في محافظة دمشق.

أظهرت النتائج أنّ الطريقة الخطية أكثر فاعلية من الطريقة المئينية عند استخدام تصميم المجموعة الواحدة في المعادلة الأفقية، واقترح استخدام الطريقة الخطية لمعادلة درجات صور الاختبار التحصيلي في مادة الرياضيات للصف الرابع.

الكلمات المفتاحية: معادلة الدرجات، الطريقة الخطية، الطريقة المئينية.

* أستاذ مساعد – قسم القياس والتقويم التربوي والنفسي– كلية التربية– جامعة دمشق– دمشق– سورية.
** طالبة دراسات عليا (دكتوراه)– قسم القياس والتقويم التربوي والنفسي– كلية التربية– جامعة دمشق– دمشق–سورية.

مجلة جامعة تشرين للبحوث والدراسات العلمية _ سلسلة الآداب والعلوم الإنسانية المجلد (34) العدد (5)2012
Tishreen University Journal for Research and Scientific Studies - Arts and Humanities Series Vol. (34) No. (5) 2013

A Comparison of Linear and Equipercentile Methods for Equating Two Forms of Achievement Test in Mathematics for the Fourth Elementary Class in Damascus

Dr. Ramadan Darwesh*
Rania Jameel Redwan**

(Received 8 / 7 / 2012. Accepted 2 / 10 / 2012)

□ ABSTRACT □

The purpose of this study is to compare between linear and equipercentile methods for equating two forms of achievement test in mathematics for the fourth elementary class, based on the standard error of equating criterion.

For that, the researcher used the descriptive analytical approach, and prepared two reliable valid forms of achievement test using multiple-choice. Each form consists of (24) items. The research sample consists of (1150) pupils of the fourth elementary class in Damascus.

The results showed that the linear method is more efficient than the equipercentile method when the horizontal equating undertook a single group. The researcher suggested using the linear method for equating two forms of achievement test in mathematics for the fourth elementary class.

Key Words: Test Equating, Linear Method, Equipercentile Method.

*Assistant Professor, Department of Educational and Psychological Assessment and Evaluation, Faculty of Education, Damascus University, Damascus, Syria.
**Postgraduate Student, Department of Educational and Psychological Assessment and Evaluation, Faculty of Education, Damascus University, Damascus, Syria.

مقدمة:

تشكل الاختبارات المدرسية أحد أهم أدوات التقويم التربوي في المدرسة، وهي تزود المطبق بالمعلومات الضرورية اللازمة لاتخاذ القرار، ولا يقتصر مثل هذا القرار على عمليات الترفيع أو الترسيب، ولكنه يتجاوزها إلى مجالات أخرى ذات أهميّة بالغة كتصنيف الطلبة في فئات وفق مستوياتهم التحصيلية، أو تحديد البرنامج التعليمي اللازم لهذه الفئة أو تلك، وبخاصة ما يتصل ببرامج التعليم العلاجي، كما يمكن الإفادة من المعلومات المستمدة من نتائج الاختبارات في اتخاذ قرارات ذات صلة بالنظام الإداري للمؤسسة أو بتطوير البرنامج التعليمي ذاته (الحموري، 1991، 2). ونظراً لأهمية هذه القرارات المأخوذة من الاختبارات ولعدم الإخلال بسرية الاختبار ولتجنب تطبيقه أكثر من مرة على نفس المفحوص فأنه غالباً ما يتم اللجوء إلى بناء عدة صور أو نماذج للاختبار الواحد، حيث إن علامة المفحوص على أي من هذه الصور تكون لها الدلالة نفسها والقيمة القياسية فيما لو تحققت على صورة أخرى للاختبار، ويقال عندئذ أنّ هناك تكافؤاً تاماً بين صور الاختبار الواحد، ولكن بناء مثل هذه الصور قد لا يكون أمراً سهلاً في كثير من الأحايين ، كما أنه من النادر أن يحدث مثل هذا التكافؤ، لذا يصبح من الضروري معادلة الدرجات على هذه الصور للاختبار، ويطلق على هذه العملية **معادلة الاختبار** (***Test Equating***) (الشريفين، 2003، 7).

وتعد معادلة الاختبارات من الموضوعات التي تلقى كثيراً من الاهتمام عند علماء القياس النفسي والتربوي. وهناك العديد من المتغيرات والمواقف الاختبارية التي تؤثر في دقة عملية المعادلة منها: التصميم المستخدم، والعينة، والطرائق المستخدمة للمعادلة، وغيرها، ونظراً لحداثة هذا المجال فإنه يوجد العديد من التصميمات والطرائق للمعادلة التي اقترحها بعض العلماء والمشتغلين بالقياس والتي ما زالت طور البحث (علي، 2004، ص 4).

وقد حظي هذا الجانب باهتمام العديد من الباحثين حيث أجريت العديد من الدراسات التي تقارن بين طرائق المعادلة التقليدية وطرائق المعادلة في النظرية الحديثة للتعرف على أكثرها دقة في معادلة الاختبارات كدراسة (بالخيور 2008) التي أشارت إلى أنّ طرائق المعادلة وفق النظرية الحديثة أكثر دقة منها وفق النظرية التقليدية بوجه عام، وفي دراسة (أيوب 1994) التي توصلت إلى أن نماذج النظرية الحديثة في القياس أكثر فاعلية من طريقتي المعادلة الخطية والمئينية في حالة المعادلة الأفقية بينما تفوقت المعادلة المئينية في حالة المعادلة العمودية، ودراسة (هاريس وكولن)(Harris& Kolen, 1990) التي أشارت إلى عدم تميز أي طريقة من الطرائق الآتية (الخطية، والمئينية، والمعادلة وفق النموذج الثلاثي) في إعطاء نتائج أفضل للمعادلة، وهذا ما توصل إليه كل كولن وويتني في دراستهما (Kolen& Whitney,1982) التي أظهرت بأنه لا توجد طريقة معادلة مثلى من بين الطرائق الآتية (الخطية، المئينية، وباستخدام نموذج راش، وباستخدام النموذج ثلاثي المعلمة)، وبينما وجد كولن (Kolen, 1981) بأن المعادلة باستخدام نموذج راش وباستخدام الطريقة الخطية لم تكونا فاعلتين أما النموذج الثلاثي المعلمة والمعادلة المئينية فقد عملا بشك فاعل.

يلاحظ مما سبق تنوع نتائج هذه الدراسات واختلافها_ بل تناقضها_ بما يوحي بحاجة إلى المزيد من البحث والتقصي، كما لم تجد الباحثة أي دراسة محلية تناولت هذا الجانب في الجمهورية العربية السورية، ما أدى إلى ضرورة دراسة ذلك الجانب بهدف الوصول إلى طريقة المعادلة التي تحقق أفضل دقة لمعادلة الاختبار، من هنا جاءت الدراسة الحالية للمقارنة بين طريقتين من طرائق المعادلة وفق النظرية التقليدية، ويمكن تحديد مشكلة البحث الحالي في الإجابة عن السؤال الرئيس الآتي: " أي الطريقتين أكثر فاعلية الخطية أم المئينية في معادلة درجات صورتي اختبار تحصيلي في مادة الرياضيات للصف الرابع من التعليم الأساسي في مدينة دمشق؟".

ويتفرع عن هذا السؤال الأسئلة الفرعية الآتية:

1- هل صورتي الاختبار التحصيلي ذي الاختيار من متعدد متكافئة؟

2- ما الدرجات المتعادلة لصورتي اختبار تحصيلي ذي الاختيار من متعدد في الرياضيات للصف الرابع وفق كل من الطريقة الخطية، والطريقة المئينية؟

3- أيهما أكثر فاعلية الطريقة الخطية أو الطريقة المئينية في حالة المعادلة الأفقية باستخدام تصميم المجموعة الواحدة؟

أهميّة البحث وأهدافه:

يكتسب البحث أهميته العلمية من محاولته الكشف عن طريقة معادلة درجات صورتي الاختبار التي تحقق أفضل دقة للاختبار عندما يكون التصميم المستخدم هو تصميم المجموعة الواحدة.

كما تكمن أهمية البحث التطبيقية في النقاط الآتية: 1- بناء بنوك للأسئلة، إذ تعد معادلة درجات الاختبارات الأساسية لبناء بنوك الأسئلة. 2- حلّ العديد من الإشكالات التي تحدث أثناء تطبيق الاختبار والتي تحتم وجود عدة صور للاختبار الواحد عندما يجري تطبيقه على عدد من المجموعات التي قد تختلف في أماكن أو مواعيد تواجدها. 3- توضيح خطوات عملية المعادلة باستخدام الطريقتين الخطية، والمئينية، ومناقشتهما وتفسير نتائجهما بما يعد دليلاً يتيح للباحث فرصة الاستفادة عند القيام بمعادلة درجات اختبارات أخرى، وفي مواد مختلفة، ما يسمح بمقارنة التلاميذ في نفس الصف إذا ما طبقت عليهم نماذج مختلفة للاختبار. 4- تقديم صورتين لاختبار تحصيلي في الوحدة (2، و3، و4) من مادة الرياضيات للصف الرابع صالحتين للاستخدام مستقبلاً لغرض تقويم التلاميذ في الصف الرابع في مدارس التعليم الرسمية في مدينة دمشق. 5- البحث الحالي أول دراسة محلية _حسب علم الباحثة_ تسعى لمقارنة طرائق معادلة درجات الاختبار وفق النظرية التقليدية في الجمهورية العربية السورية، وبالتالي فإنّ هذا البحث قد يثير الحافز لدى باحثين مهتمين لإجراء مزيد من الأبحاث حول معادلة الدرجات.

ولذا يهدف البحث الحالي إلى:

1- مقارنة طريقتين من طرائق معادلة درجات صورتي اختبار تحصيلي في الرياضيات ذي الاختيار من متعدد للصف الرابع من التعليم الأساسي وفق النظرية التقليدية وهي (الطريقة الخطية، والطريقة المئينية).

2- التعرف على الطريقة التي تؤدي إلى أفضل دقة معادلة.

3- التعرف على الدرجات المتعادلة لصورتي الاختبار وفق كل من الطريقتين الخطية والمئينية.

طرائق البحث ومواده:

تصميم البحث ومتغيراته: جرى حساب معادلة صورتي الاختبار وفق نمط المعادلة الأفقية بإتباع تصميم المجموعة الواحدة باستخدام طريقتين من طرائق النظرية التقليدية (الخطية، والمئينية) حيث قامت الباحثة بتطبيق الصورة الأولى على تلاميذ الصف الرابع، وبعد فاصل زمني مدته أسبوع قامت بتطبيق الصورة الثانية على نفس المجموعة.

*المتغير المستقل: طريقة معادلة درجات التلاميذ على صورتي الاختبار، وتتمثل بالطريقتين: 1-الخطية. 2- المئينية.

*المتغير التابع: دقة معادلة الاختبار المتمثلة بالخطأ المعياري للمعادلة.

منهجية البحث:

يعتمد البحث الحالي المنهج الوصفي التحليلي ببناء أدوات البحث، مروراً بالإطار النظري، ومناقشة النتائج وتحليلها، فالمنهج الوصفي التحليلي يهتم بوصف الظاهرة وصفاً دقيقاً، ويعبر عنها تعبيراً كيفياً أو كمياً، فالتعبير الكيفي يصف لنا الظاهرة ويوضح خصائصها، أما التعبير الكمي فيعطينا وصفاً رقمياً يوضح مقدار هذه الظاهرة أو حجمها ودرجات ارتباطها بالظواهر المختلفة (ساعاتي، 1999، 71).

مجتمع البحث وعينته:

تكون المجتمع الأصلي للبحث من جميع تلاميذ الصف الرابع الذين يدرسون في المدارس الرسمية في مدينة دمشق في العام الدراسي 2010/ 2011 م، والبالغ عددهم (27270) تلميذاً وتلميذةً.

العينة الاستطلاعية للبحث: تكونت عينة البحث الاستطلاعية من (70) تلميذاً وتلميذةً من تلاميذ الصف الرابع من التعليم الأساسي، جرى اختيارهم من إحدى المدارس الرسمية من مدينة دمشق في خارج عينة البحث الأساسية، وذلك لاستخراج معاملات الصعوبة والتمييز والثبات، وحساب الزمن الملائم للصورتين.

العينة الأساسية للبحث: تكونت عينة البحث الأساسية من (1150) تلميذاً وتلميذةً من تلاميذ الصف الرابع، وشكلت ما نسبته(4.217%) من تلاميذ المجتمع الأصلي للبحث.

حدود البحث: يتحدد البحث وفقاً لما يلي: 1- الحدود الموضوعية: يتحدد البحث الحالي بموضوعه وهو المقارنة بين طريقتين من طرائق المعادلة في النظرية التقليدية وهما: الخطية والمئينية، لمعادلة درجات صورتي اختبار تحصيلي ذي الاختيار من متعدد، ويقتصر البحث على الوحدة (2، و3، و4) من مادة الرياضيات للصف الرابع.

2- الحدود البشرية: يقتصر البحث على تلاميذ الصف الرابع من التعليم الأساسي في المدارس الرسمية في مدينة دمشق. 3- الحدود الزمانية: جرى تطبيق البحث خلال الفصل الدراسي الأول من العام الدراسي 2010 / 2011م.

المصطلحات والتعريفات الإجرائية:

معادلة الاختبار*Test Equating*: وتعني تحويل نظام وحدات القياس الخاص بأحد الاختبارين إلى نظام وحدات القياس الخاص بالاختبار الآخر، بحيث تصبح القياسات المستمدة من درجات كل من الاختبارين متكافئة بعد إجراء هذا التحويل (الشريفين، 2003، 8). التعريف الإجرائي لمعادلة الاختبار: إجراء إحصائي سيجري باستخدام المعادلة الأفقية وفق كِلا الطريقتين (الخطية، والمئينية) لتحويل درجات التلاميذ في الصف الرابع على الصورة الأولى إلى ما يقابلها من درجاتهم على الصورة الثانية حيث يمكن المقارنة بين درجاتهم على الصورتين.

دقة المعادلة *Equating Accuracy*: أسلوب إحصائي يستخدم للتأكد من مدى فاعلية المعادلة، ويستخدم الخطأ المعياري للمعادلة كأحد معايير التأكد من دقة المعادلة، وهو الانحراف المعياري للدرجات المحولة لإحدى الصورتين y المناظرة لدرجة معينة على الصورة الأخرى x تؤخذ من تحويل مبني على عينات مستقلة أخذت من

مجتمعها (الصمادي، 2008، 9). التعريف الإجرائي لدقة المعادلة: التحقق من دقة معادلة درجات التلاميذ على الصورتين الأولى والثانية، باستخدام الخطأ المعياري للمعادلة وفق الطريقتين موضوع الدراسة، وتكون المعادلة أكثر دقة عندما تكون قيمة الخطأ المعياري قليلة، والعكس صحيح.

الاختبار التحصيلي *Achievement Test*: هو أداة تُستخدم لتحديد مستوى كسب الطالب لمعلومات ومهارات في مادة دراسية كان قد تعلمها مسبقاً بصفة رسمية، من خلال إجاباته عن عينة من الأسئلة (الفقرات) التي تمثل محتوى المادة الدراسية (العبسي، 2010، 147). التعريف الإجرائي لصورتي الاختبار التحصيلي: هما اختبارين معياريّ المرجع من نوع الاختيار من متعدد من إعداد الباحثة لتلاميذ الصف الرابع من التعليم الأساسي في مادة الرياضيات للفصل الدراسي الأول لعام 2010/ 2011 م.

الإطار النظري:

مفهوم معادلة الاختبار: أشارت العديد من الدراسات والبحوث إلى عدة تعريفات للمعادلة، ومن بين تلك التعريفات ما أورده الشريفين (2003) بأنّها تحويل نظام وحدات القياس الخاص بأحد الاختبارين إلى نظام وحدات القياس الخاص بالاختبار الآخر، بحيث تصبح القياسات المستمدة من درجات كل من الاختبارين متكافئة بعد إجراء هذا التحويل (الشريفين، 2003، 8)، ويعرفها علام (2002) بأنّها عملية التوصل إلى درجات متناظرة للصيغ المختلفة من الاختبار، ويطلق على عملية التعادل اسماً آخر هو التكافؤ (علام، 2002، 696- 697). وكما عرفها سوناسي (Sunnassee, 2011) بأنّها إجراء إحصائي 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 الاختبار نفسه لجعلهما متكافئتين (Sunnassee, 2011, 16). فالغرض من معادلة الدرجات هو تأسيس الدرجات التي تعد متكافئة على أدوات مختلفة والتي تقيس نفس السمة (عبد الله، 70،2003).

أنواع المعادلة: يوجد نوعان من أنواع المعادلة للاختبارات: المعادلة الأفقية (Horizontal Equating)، والمعادلة العمودية أو الرأسية (Vertical Equating). ففي المعادلة الأفقية تكون الصور المختلفة للاختبار متقاربة في مستوى صعوبتها، والهدف من إجراء المعادلة هو تعديل الفروق الناتجة عن الاختلاف في الصعوبة، والافتراض في مثل هذه الحالة أنّ صور الاختبار الواحد معدة لمنطقة سلوك واحدة أو نفس التوزيع للقدرة أو السمة المقيسة وبالتالي لمجتمع محدد من الأفراد يتمثل فيه توزيع القدرة نفسه (Bernnan & Kolen, 1987, 248). وستستخدم الباحثة المعادلة الأفقية عند معادلة درجات تلاميذ الصف الرابع على الصورتين الأولى والثانية.

وتستخدم المعادلة العمودية لمعادلة درجات اختبارين أو أكثر يعتقد أنهما يقيسان السمة نفسها مصممة بحيث تكون مختلفة في درجة الصعوبة، وكذلك تكون عينات الأفراد التي تطبق عليها هذه الاختبارات مختلفة في مستوى قدراتها (الشريفين، 2003، 8).

تصاميم جمع البيانات لإجراء المعادلة: تستخدم عدة أنواع من التصاميم لجمع البيانات وذلك بغرض إجراء عملية المعادلة، وتتدخل عدة اعتبارات في اختيار التصميم المناسب، وبشكل عام فإنّ أبرز التصاميم المستخدمة لجمع البيانات لأغراض المعادلة هي:تصميم المجموعات العشوائية المتوازنة، تصميم المجموعات غير المتكافئة ذات الجذع المشترك، تصميم المجموعات المتكافئة ذات الجذع المشترك، تصميم المجموعة الواحدة وفي هذا التصميم تعطى صورتا الاختبار لمجموعة واحدة من المفحوصين بحيث تعطى إحدى الصورتين متبوعة بالثانية ويفضل في نفس اليوم

حتى لا تؤثر عوامل دخيلة أخرى مثل الخبرة على استجابات الأفراد، ويجري تدريج فقرات كل صورة من الاختبار على حدة، فيحصل كل فرد على تقديرين مختلفين للقدرة، تنتج الدرجة الأولى عن صورة الاختبار الأولى وتنتج الدرجة الثانية عن أداء المفحوص على صورة الاختبار الثانية، لذا فإنّ أداء المفحوصين الذين تلقوا فقرات صورتي الاختبار يعد هو الأساس في المقارنة ما بين التدريجات المنفصلة الناتجة من الدرجات الكلية على كلتي الصورتين، وفي هذا التصميم لا تختلط الفروق في مستويات الصعوبة للصورتين بالفروق في مستويات قدرات المفحوصين (الحربي، 2009، 21)، وقد استخدمت الباحثة هذا التصميم في البحث الحالي.

طرائق المعادلة: هناك طرائق متعددة لإجراء المعادلة للاختبارات منها ما هو منبثق عن النظرية التقليدية وأخرى منبثقة عن النظرية الحديثة في القياس، فمن الطرائق المستخدمة في معادلة الاختبارات ضمن النظرية التقليدية: معادلة الاختبار بطريقة الوسط الحسابي Mean Equating، ومعادلة الاختبار خطياً Linear Equating، وطريقة الرتب المئينية المتساوية Equipercentile Equating (الدوسري، 2001، 118- 120). وهناك ثلاث طرائق رئيسة لمعادلة الاختبار ضمن النظرية الحديثة، وهي: معادلة الاختبار باستخدام درجات القدرة (السمة) Ability Score Equating، والمعادلة باستخدام الدرجات الملاحظة (الدرجات الخام) Observed Score Equating، والمعادلة باستخدام الدرجات الحقيقية True Score Equating (Yang & Houang, 1996, 21) وقد استخدمت الباحثة في الدراسة الحالية الطريقتين الخطية وطريقة الرتب المئينية المتساوية لمعادلة درجات تلاميذ الصف الرابع على صورتي اختبار تحصيلي في الرياضيات.

الدراسات السابقة:

فيما يلي عرض لبعض الدراسات السابقة ذات الصلة بموضوع البحث التي تمكنت الباحثة من جمعها، ونظراً لعدم توافر دراسات محلية ستقوم الباحثة بعرض الدراسات العربية أولاً، ثم الدراسات الأجنبية، وبترتيب زمني من الأقدم إلى الأحدث:

أولاً- الدراسات العربية:

دراسة حسين أيوب (1994): بعنوان (المقارنة بين أربع طرق للمعادلة عندما يكون التصميم من مجموعات متكافئة وغير متكافئة).

هدفت الدراسة إلى المقارنة بين أربع طرائق لمعادلة الاختبارات هي الطريقة الخطية والطريقة المئينية المنبثقتان عن النظرية التقليدية في القياس ونماذج النظرية الحديثة في القياس أحادي وثنائي المعلمة عندما يكون التصميم (أ) مجموعات متكافئة أو عشوائية، (ب) مجموعات غير متكافئة مع فقرات مشتركة، باستخدام معامل الصدق التقاطعي معياراً لدراسة مدى استقرار النتائج للعلامات المعادلة بين عينتي الدراسة. لتحقيق هدف الدراسة جرى بناء ثلاثة اختبارات بصورتين أ، ب لمادة الرياضيات للصفوف الرابع والخامس والسادس، وقد تكون كل اختبار من 40 فقرة. كما اختيرت عينتان مستقلتان تكونت الأولى من (1390) طالباً وطالبةً، والثانية من (1412) طالباً وطالبةً، وجرى اختيارهم عشوائياً من مجتمع الدراسة الأصلي في مدارس منطقة عمان الكبرى.

وقد أشارت النتائج المتعلقة بالمعادلة الأفقية التي أجريت على المجموعات المتكافئة إلى أنّ نماذج النظرية الحديثة في القياس كانت أكثر فاعلية من طريقتي المعادلة الخطية والمئينية. أما فيما يتعلق بالمعادلة العمودية والتي

أجريت على المجموعات غير المتكافئة فقد أشارت النتائج إلى أنّ الطريقة المئينية كانت أكثر فاعلية من الطرائق الأخرى المستخدمة في الدراسة، يليها نموذج ثنائي المعلمة ، ثم نموذج : أحادي المعلمة، وأخيراً : الطريقة الخطية.

دراسة نضال الشريفين (2003): بعنوان (مدى تحقق معايير الفاعلية في معادلة اختبارين أحدهما ثنائي التدريج والآخر متعدد التدريج وفق نماذج النظرية الكلاسيكية والنظرية الحديثة في القياس).

هدفت الدراسة إلى الكشف عن مدى تحقق معايير الفاعلية في معادلة اختبارين أحدهما ثنائي التدريج والآخر متعدد التدريج وفق طريقتين من طرائق النظرية الكلاسيكية هما: (الخطية، والمئينية)، وطريقتين من طرائق المعادلة في النظرية الحديثة في القياس هما: (نموذج أحادي المعلمة، ونموذج التقدير الجزئي)، وذلك باستخدام معياران للحكم على فاعلية طرائق المعادلة المستخدمة هما: معامل الصدق التقاطعي، والخطأ المعياري للمعادلة. لتحقيق هدف الدراسة قام الباحث ببناء اختبارين تحصيليين في الفيزياء أحدهما ثنائي التدريج مؤلف من 75 فقرة من نوع الاختيار من متعدد، والآخر متعدد التدريج مؤلف من 20 سؤالاً (كل سؤال مكون من أربع خطوات). تكونت عينة الدراسة من (1003) طلاباً وطالبة جرى اختيارها بالطريقة العشوائية العنقودية. ومن أهم ما توصلت إليه الدراسة: كانت نتائج النموذج (أحادي المعلمة) أكثر فاعلية من طريقتي (المعادلة الخطية والمئينية) ، وكانت جميعها ذات فاعلية كبيرة. أما وفق معيار الخطأ المعياري للمعادلة فقد كانت المعادلة الخطية أكثر فاعلية من المعادلة المئينية، ولم يتمكن الباحث من استخدام هذا المحك في النظرية الحديثة نظراً لعدم توافر البرامج الإحصائية المناسبة.

دراسة شفاء بالخيور (2008): بعنوان (فاعلية طرق معادلة نماذج اختبار القدرات العامة بالمركز الوطني للقياس والتقويم وفق نظريتي القياس التقليدية والحديثة في ضوء بعض المتغيرات).

هدفت الدراسة إلى مقارنة طرائق معادلة النماذج الاختبارية الأربعة لاختبار القدرات العامة الذي أعدّه المركز الوطني للقياس والتقويم في التعليم العالي، والذي يطبقه على طلاب الصف الثالث الثانوي كمعيار مكمل لدرجات تحصيلهم الدراسي في اختبار الثانوية العامة لقبولهم بالجامعات ومعظم الكليات، كما تهدف إلى مقارنة طرائق معادلة الاختبار_ بالإضافة إلى مستوى العينة الكلية_ عند مستويات حجم العينة الفرعية صغيرة (100)، متوسطة (500)، كبيرة (1000). ولتحقيق المقارنة بين طرائق المعادلة المختلفة جرى استخدام معيارين هما: جذر متوسط مربعات الخطأ (RMSE) كمؤشر للدقة، والصدق التقاطعي (CV) كمؤشر لاستقرار النتائج، وجرى حساب معادلة نماذج الاختبار وفق نمط المعادلة الأفقية بإتباع تصميم المجموعات المتكافئة أو العشوائية مع وجود مفردات مشتركة، وذلك باستخدام طرائق النظرية التقليدية (CTT): المتوسط، الخطية، توكر الخطية، المئينيات غير الممهدة، المئينيات الممهدة، وطرائق النظرية الحديثة (IRT): المتوسط/المتوسط، المتوسط/سيجما طبقاً للنماذج اللوجستية الثلاثة (أحادي، وثنائي، وثلاثي المعلمة).

جمعت البيانات من خلال نتائج استجابات المجتمع الأصلي للطلاب الذكور من التخصص العلمي على النماذج الاختبارية الأربع، وقد بلغ مجموع عددهم على الترتيب: (43352)، (42366)، (37389)، (34099)، ومن أهم نتائج الدراسة: 1- بالنسبة لمقارنة طرائق المعادلة على مستوى العينة الكلية من حيث الدقة فتشير النتائج إلى أنّ طرائق المعادلة وفق نظرية (IRT) أكثر دقة منها وفق نظرية (CTT) بوجه عام، أما من حيث الاستقرار تعد طريقة المتوسط والمئينيات بنوعيها أكثر استقراراً في نتائجها. 2- بالنسبة لمقارنة طرائق المعادلة على مستوى العينة الفرعية من حيث الدقة فإنّ طريقة المتوسط وفق نظرية (CTT)، وطريقتا المتوسط/المتوسط، والمتوسط/ سيجما لنموذج المعلمة

الواحدة وفق نظرية (IRT) تعد أفضل الطرائق عند المستويات الثلاثة للعينة. أما من حيث الاستقرار فإنّ طرائق المعادلة وفق نظرية (IRT) تتفوق بشكل عام على طرائق المعادلة وفق نظرية (CTT) عند المستويات الثلاثة للعينة.

ثانياً– الدراسات الأجنبية:

دراسة كولن (Kolen, 1981): بعنوان (مقارنة الطرائق وفق النظرية التقليدية ونظرية الاستجابة للمفردة لمعادلة الدرجات).

"Comparison of Traditional and Item Response Theory Methods of Equating Tests".

هدفت الدراسة إلى مقارنة عدداً من نماذج النظرية الحديثة في القياس مع طريقتين من طرائق النظرية التقليدية وهما: (الخطية والمئينية)، في معادلة الاختبارات. لتحقيق هدف الدراسة قام الباحث بمعادلة صورة جديدة مع صورة قديمة لاختبار أيوا للنمو التربوي لمجموعتين (ITED) والمصمم لإجرائه على الطلبة في المدارس الثانوية ويتكون من قدرات لفظية ولغوية والقدرة على التفكير الكمي، وأجري على عينة من طلبة الصفين التاسع والعاشر من ذوي التحصيل المرتفع، وبلغ عدد أفراد العينة (7030) طالباً وطالبةً، واعتمدت الدراسة معيار الصدق التقاطعي للحكم على فاعلية طرائق المعادلة المختلفة. وأشارت نتائج هذه الدراسة إلى أنّ المعادلة باستخدام نموذج راش والمعادلة بالطريقة الخطية لم تكونا فاعلتين، أما النموذج الثلاثي المعلمة والمعادلة المئينية فقد عملا بشكل فاعل، وقد تباينت النتائج في المعادلة الأفقية.

دراسة كولن وويتني (Kolen & Whitney, 1982): بعنوان (مقارنة أربع طرائق لمعادلة اختبارات التطور التربوي العام).

"Comparison of Four Procedures for Equating The Tests of General Educational Development".

هدفت الدراسة إلى مقارنة فاعلية أربع طرائق لمعادلة اختبارات التطور التربوي العام (GED) وهذه الطرائق هي (الخطية، والمئينية، وباستخدام نموذج راش، وباستخدام النموذج ثلاثي المعلمة) وفق المعادلة الأفقية واستخدم معيار الصدق التقاطعي للحكم على فاعلية الطرائق، وقد بلغ عدد أفراد عينة الدراسة (1200) طالباً وطالبةً.

أظهرت نتائج الدراسة أنّه لا توجد طريقة معادلة مثلى من بين الطرائق الأربع، على الرغم مما أبدته نماذج النظرية الحديثة من تفوق على الطرائق الكلاسيكية في المعادلة.

دراسة بترسون، وكوك، وستوكنغ (Peterson, Cook & Stocking, 1983): بعنوان (طرائق المعادلة وفق نظرية الاستجابة للمفردة مقابل الطرائق التقليدية: دراسة مقارنة لمقياس الاستقرار).

"IRT Versus Conventional Equating Methods: A Comparative study of Scale Stability".

هدفت الدراسة إلى مقارنة فاعلية طريقة المعادلة الخطية وطريقة المعادلة وفق النموذج ثلاثي المعلمة، وذلك من حيث ثبات المعادلة للاختبارات واستقرارها باستخدام محك الصدق التقاطعي، وقد استخدم الباحثون اختبار الاستعداد الدراسي (SAT) والذي يقيس قدرات التطور اللغوي والتفكير الرياضي، ويستخدم كمقياس لقبول الطلبة في الجامعات، وأجريت الدراسة على عينة من (2670) طالباً وطالبةً.

أظهرت نتائج الدراسة أنّ طريقة المعادلة الخطية هي الأكثر فاعلية إذا كانت الاختبارات المستخدمة للمعادلة متكافئة (Parallel– Tests)، أما إذا اختلفت الاختبارات من حيث المحتوى أو طول الاختبار فإنّ النموذج ثلاثي المعلمة يعطي ثباتاً واستقراراً أفضل للاختبارات المعادلة.

دراسة ماركو، وبترسن، وستيورات (Marco, Petersen, & Stewart, 1989): بعنوان (التحقق من كفاءة نماذج معادلة الدرجات الخطية).

"A Test of The Adequacy of Curvilinear Score Equating Models".

هدفت الدراسة إلى دراسة فاعلية طرائق المعادلة الخطية، والمئينية، ونموذج راش، والنموذج ثلاثي المعلمة، حيث قام الباحثون بجمع البيانات باستخدام اختبار الاستعداد المدرسي اللفظي (SAT– Verbal) واعتمدت الدراسة التصميم القائم على أساس وجود اختبار جذع مشترك (داخلي وخارجي). ومن أهم نتائج الدراسة: عندما كان الاختبار المشترك منفصلاً عن كل من الاختبارين المراد معادلتهما كانت المعادلة الخطية والمعادلة وفق نماذج النظرية الحديثة في القياس فعّالة في المعادلة الأفقية. أما عندما كان الاختبار المشترك داخلياً فإنّ طريقة المعادلة المئينية فعّالة بدرجة جيدة كما في المعادلة الخطية والمعادلة المستندة إلى النظرية الحديثة في القياس.

دراسة هاريس وكولن (Harris & Kolen, 1990): بعنوان (أثر جماعة المفحوصين في علاقات المعادلة).

"Effect of Examinee Group on Equating Relationships".

هدفت الدراسة إلى معرفة فاعلية طرائق المعادلة باختلاف مستويات القدرة، وقد استخدمت الطرائق التالية: طريقتين من طرائق المعادلة في النظرية التقليدية وهما (الخطية، والمئينية)، والمعادلة وفق النموذج الثلاثي في النظرية الحديثة. استخدم الباحثان لتحقيق الهدف خمس صور لاختبارات المفاهيم الحسابية (Arithmetic Concept Tests "ACD") بحيث كانت متماثلة في درجة ثباتها وخصائص الصعوبة والمحتوى، وقد أشارت نتائج الدراسة إلى عدم تميز أي من طرائق المعادلة المستخدمة في إعطاء نتائج أفضل للمعادلة.

دراسة هيونة (Huynh, 1994): بعنوان (مقارنة الطريقة المئينية وطريقة التقدير الجزئي لمعادلة تقييم الأداء المؤلف من فقرات ذات الاستجابات الحرة).

"A Comparison of Equipercentile and Partial Credit Equating for Performance–based Assessment Composed of Free– response Items".

هدفت الدراسة إلى مقارنة المعادلة المئينية وطريقة المعادلة باستخدام نموذج التقدير الجزئي. واستخدمت لذلك الغرض العلامات الخام المشتقة من اختبار التقييمات المبنية على الأداء على فقرات استجاباتها حرة والتي جمعت حول مختارات قراءة طويلة أو مسائل رياضية متعددة الخطوات/ التدريج، وأخذت البيانات من برنامج تقييم الأداء ي مدرسة مريلاند.

ومن أهم النتائج أنّ نموذج التقدير الجزئي لماسترز قد يكون مفيداً لمعادلة الاختبارات التي تتكون من فقرات سهلة نوعاً ما (أو ليست صعبة جداً) بحيث تتشارك مع المكون الرئيس الأول على الأقل بما نسبته25% من التباين الكلي، وتبقى هذه النتيجة قائمة حتى بوجود مستوى ما من الاعتمادية في الاستجابة عن الفقرات في نطاق كل فئة، كما وجدت تفاوتات حادة بين نموذج التقدير الجزئي والطريقة المئينية في درجة الالتواء لتوزيعات العلامات، ولكن اتجاه

هذه التفاوتات غير متنبأ به، لذا، لا يمكن الاستنتاج أي من الطريقتين تعطيان نتائج متكافئة عندما تكون التوزيعات ملتوية وبشكل واضح.

دراسة يانغ وهوانج (Yang & Houang, 1996): بعنوان (أثر طول الجذع المشترك وطريقة المعادلة في دقة معادلة الاختبار: مقارنة المعادلة بالطريقة الخطية والمعادلة بطرائق نظرية الاستجابة للمفردة باستخدام تصميم الجذع المشترك).

"The Effect of Anchor Length and Equating Method on the Accuracy of Test Equating: Comparisons of Linear and IRT Based Equating Using an Anchor- Item Design".

هدفت الدراسة إلى التحقق من أثر طول اختبار الجذع المشترك في دقة المعادلة، وإلى المقارنة بين طريقة توكر الخطية وطريقتين من طرائق المعادلة باستخدام نظرية الاستجابة للمفردة إحداهما تعتمد تقدير معالم فقرات الجذع المشترك ثم إيجاد التحويلات الخطية لها وذلك باستخدام النموذج ثلاثي المعلم، والأخرى باستخدام معادلة الدرجات الحقيقية. لتحقيق أهداف الدراسة استخدم الباحثان بيانات صورتي اختبار الكفاءة الدنيا (Minimum Competency Examination) وهي من نوع الاختيار من متعدد، وقد اشتملت الصورة الأولى على (197) فقرة، والصورة الثانية على (203) فقرة، وجرى تصميم 3 أزواجا من النماذج القصيرة من خلال الاختيار العشوائي للفقرات من الصورتين السابقتين باستخدام المعاينة العشوائية البسيطة، وجرى معادلة كل نموذجين على حدة، وقد اتخذت الدرجة الكلية على اختبار الجذع المشترك والذي تكون من (145) فقرة كمحك، واتبعت الدراسة المنهج التجريبي. تألفت العينة من (2241) مفحوصا من طلبة الطب الذين أخذوا الاختبار تحت التدريب، حيث بلغ عدد المفحوصين الذين طبقت عليهم الصورة الأولى (1092) مفحوصا ، وعدد المفحوصين الذين طبقت عليهم الصورة الثانية (1149) مفحوصا .

توصلت الدراسة إلى أنّ نتائج المعادلة باستخدام الطريقتين المعتمدتين على نظرية الاستجابة للمفردة بينهما ارتباط تام تقريباً، وذلك بغض النظر عن عدد فقرات الجذع المشترك، كما كان من بين نتائج الدراسة أنّ ارتباط الدرجات التي جرى التوصل إليها باستخدام المعادلة حصلت على معاملات ارتباط مختلفة حسب عدد فقرات الجذع المشترك مع الدرجات الخاصة بالمحك، وقد كان الارتباط مرتفعاً وفقاً لزيادة عدد فقرات الجذع المشترك، وذلك بناءً على النتائج المتوصل إليها باستخدام طرائق المعادلة المعتمدة على نظرية الاستجابة للمفردة، كذلك الأمر مع الطريقة الخطية للمعادلة والمعتمدة على النظرية التقليدية في القياس، إلا أنّ نتائج المعادلة وفقاً للطريقتين المعتمدتين على نظرية الاستجابة للمفردة أعطت نتائج أكثر دقة في معادلة صور الاختبار.

دراسة يانغ (Yang, 1997): بعنوان (أثر المحتوى المركب وطريقة المعادلة في دقة معادلة الاختبار باستخدام تصميم الجذع المشترك).

"The Effects of Content Mix and Equating Method on the Accuracy of Test Equating Using Anchor-Item Design".

هدفت الدراسة إلى التعرف على أثر استخدام ثلاث طرائق لمعادلة الاختبار (طريقة توكر الخطية، و طريقتين من طرائق نظرية الاستجابة للمفردة) باستخدام الجذع المشترك في دقة المعادلة، والتعرف على أثر تمثيل فقرات الجذع المشترك للمحتوى في دقة المعادلة. استخدم الباحث البيانات الناتجة عن تطبيق صورتين لأحد اختبارات الكفاءة المهنية (Professional Competency Examination)، واشتملت الصورة الأولى على (197) فقرة، والصورة الثانية على

(203) فقرة، واشتمل اختبار الجذع المشترك المضمن في الشكلين على (145) فقرة، كما جرى تشكيل أربعة أزواج من اختبارات الجذع المشترك تختلف في مدى تمثيلها للمحتوى. تألفت عينة الدراسة من (2241) مفحوصا، حيث طبقت الصورة الأولى على (1092) مفحوصا، والصورة الثانية على (1149) مفحوصا.

أشارت النتائج إلى أنّ المعادلة باستخدام الطريقتين المعتمدتين على نظرية الاستجابة للمفردة كانت أكثر دقة من المعادلة بطريقة توكر الخطية، وأنّ دقة المعادلة تعتمد على مدى تمثيل فقرات اختبار الجذع المشترك للمحتوى بغض النظر عن طريقة المعادلة المستخدمة.

دراسة يانغ (Yang, 2000): بعنوان (أثر تجانس المحتوى وطريقة المعادلة في دقة المعادلة باستخدام الجذع المشترك).

"The Effects of Content Homogeneity and Equating Method on the Accuracy of Common-Item Test Equating".

هدفت الدراسة إلى التعرف على أثر كل من تجانس المحتوى، وتمثيل فقرات الجذع المشترك للمحتوى، وطريقة المعادلة (طريقة توكر الخطية، والمئينيات، وطريقتين من طرائق نظرية الاستجابة للمفردة) في دقة المعادلة، ولتحقيق أهداف الدراسة قام الباحث باستخدام صورتين من نوع الاختيار من متعدد لأحد اختبارات التخصصات الطبية في الولايات المتحدة الأمريكية، وتتألف الصورة الأولى من (197) فقرة، والصورة الثانية من (203) فقرة، كما جرى تشكيل أربعة أزواج من الاختبارات الفرعية للصورتين، حيث يشترك كل زوج في مجموعة من أسئلة الجذع المشترك. تألفت عينة البحث من (20241) مفحوصا معظمهم من طلبة المدارس الطبية أو متخرجين منها، إذ طبقت الصورة الأولى على (1092) مفحوصا ، بينما طبقت الصورة الثانية على (1149) مفحوصا .

أظهرت نتائج الدراسة بأن طرائق المعادلة الأربع كانت ذات دقة متوسطة، وكانت أكثر دقةً عندما كانت فقرات الجذع المشترك أكثر تمثيلاً للمحتوى أو عندما كان المحتوى الذي يمثله الجذع المشترك يركز على مواضيع محددة. كما بينت النتائج بأنّ المعادلة باستخدام نظرية الاستجابة للمفردة أكثر دقة من طرائق المعادلة التقليدية.

يتضح من عرض الدراسات السابقة بأنّه لا توجد طريقة واحدة من طرائق المعادلة وفق النظرية الحديثة أو التقليدية تتفوق بشكل دائم على طرائق المعادلة الأخرى سواء في المعادلة الأفقية أم العمودية، كما استخدمت الدراسات السابقة عدة معايير للمقارنة بين طرائق المعادلة واعتمدت تصميمات مختلفة للمعادلة فبعضها استخدم تصميمات المجموعات المتكافئة وغير المتكافئة كدراسة أيوب (1994)، وبعضها استخدم تصميم المجموعات المتكافئة كدراسة بالخيور (2008)، كما تنوعت العينات من تلاميذ التعليم الأساسي كدراسة أيوب (1994)، وطلبة الثانوية كدراستي كولن (1981) والشريفين (2003)، وطلبة الجامعات كدراسة يانغ وهوانج (1996) ويانغ (1997) ويانغ (2000)، وتنوعت الأدوات المستخدمة من اختبارات محكية المرجع كدراسة كولن (1981) وكولن وويتني (1982) ويانغ (1997) وبالخيور (2008)، واختبارات معيارية المرجع كدراسة أيوب (1994) ودراسة يانغ (2000) والشريفين (2003).

وتسعى الدراسة الحالية إلى مقارنة طريقتين في حالة المعادلة الأفقية باستخدام صورتي اختبار تحصيلي معياري المرجع ذي الاختيار من متعدد للصف الرابع من التعليم الأساسي، كما استخدمت الباحثة تصميم المجموعة الواحدة وهو ما لم تطرق له الدراسات السابقة، كما اعتمدت معيار الخطأ المعياري للمعادلة كأساس للمقارنة بين الطريقتين.

أدوات البحث:

قامت الباحثة ببناء صورتين لاختبار تحصيلي ذي الاختيار من متعدد في الوحدة (2، و3، و4) من مادة الرياضيات للصف الرابع وتألفت كل صورة من 30 سؤالاً جرى صياغتها على نحو يتوافق مع الأهداف السلوكية، وبثلاثة بدائل لكل سؤال لتلائم المرحلة العمرية للتلاميذ وخصائصهم، وقد تحققت الباحثة من صدق المحتوى للصورتين باستخدام جدول المواصفات لبناء الصورتين، وتحكيم أسئلتهما من عدد من المحكمين. ثم قامت الباحثة بتطبيق الصورتين على أفراد العينة الاستطلاعية للحصول على ملاحظات التلاميذ عن مقدار وضوح الأسئلة والتعليمات وحساب الزمن الضروري لتطبيق كل صورة، كما قامت بالتأكد من فعالية المموهات (المشتتات)، وحساب معاملات الصعوبة والتمييز والثبات بطريقة كرونباخ ألفا لكل سؤال وللاختبار كاملاً، إذ بلغت قيمة كرونباخ ألفا للصورة الأولى (0.70)، وللصورة الثانية (0.71) وهي قيم تشير إلى ثبات جيد. وفي ضوء نتائج التجربة الاستطلاعية حذفت ستة أسئلة، كما عدلت بعضها الآخر، وقامت بإعادة ترتيب الأسئلة وفقاً لمعاملات الصعوبة من الأسهل إلى الأصعب، مع مراعاة الشكل الأنسب للصياغة وموضوع السؤال، فتكونت كل صورة من الصورتين من 24 سؤالاً، وجرى تصحيحهما بالطريقة التقليدية لتصحيح الاختيار من متعدد، كما حدد الزمن الملائم لتطبيق كل صورة والذي بلغ (60) دقيقة.

النتائج والمناقشة:

نتائج السؤال الأول: هل صورتي الاختبار التحصيلي ذي الاختيار من متعدد متكافئة؟

للإجابة عن السؤال الأول للبحث جرى استخدام ثلاثة معايير وضعها ألن وين (Allen & Yen, 1979) من أجل التحقق من تكافؤ الصورتين والأول: تساوي أوساط الدرجات الخام للمفحوصين على الصورتين، والثاني: تساوي التباين للدرجات الخام للمفحوصين على الصورتين، والثالث: تساوي الارتباط لدرجات المفحوصين على الصورتين مع الدرجات على اختبار آخر (الشريفين، 2009، 32). لذا حسبت متوسطات التلاميذ على الصورتين، والانحرافات المعيارية، ومعاملات الارتباط لكل صورة من الصورتين مع اختبار الامتحان النصفي الذي يجري في المدارس في مادة الرياضيات، والجدول (1) يوضح ذلك:

الجدول (1) المتوسطات الحسابية والانحرافات المعيارية للعلامات الخام للصورتين ومعاملات الارتباط

الصورة / الإحصائي	المتوسط الحسابي	الانحراف المعياري	معامل الارتباط
الصورة (Y)	12.91	4.13	0.447
الصورة (X)	13.65	4.79	0.504

للكشف عما إذا كانت هناك فروق ذات دلالة إحصائية عند مستوى الدلالة (0.05) بين المتوسطات الحسابية لدرجات التلاميذ تعزى للصورة المستخدمة استخدم اختبار (t) ستودنت للعينات المرتبطة، والجدول (2) يوضح ذلك:

الجدول (2) نتائج اختبار (t) ستودنت للعينات المرتبطة

الصورة	المتوسط	الانحراف المعياري	الخطأ المعياري للمتوسط	قيمة t	درجة الحرية	القيمة الاحتمالية
الصورة (Y)	12.91	4.13	0.121	6.796	1149	0.000
الصورة (X)	13.65	4.79	0.141			

يتضح من الجدول السابق وجود فروقا ذات دلالة إحصائية عند مستوى الدلالة (0.05) بين متوسط الدرجات الخام على الصورتين لصالح درجاتهم على الصورة (X).

وللتحقق من وجود فروق ذات دلالة إحصائية عند مستوى الدلالة (0.05) بين معاملات الارتباط لدرجات التلاميذ على كل صورة ودرجاتهم على اختبار الرياضيات في الامتحان النصفي استخدم الاختبار الإحصائي Z وفق المعادلة التالية: $Z = \frac{Z_1 - Z_2}{\sqrt{\left(\frac{1}{n_1-3}\right)+\left(\frac{1}{n_2-3}\right)}}$ ، حيث Z_1، Z_2 : قيم فيشر للمعاملين، و n_1، n_2 : عدد أفراد كل من المجموعتين (علام، 2005، 221). وذلك بعد تحويل معاملات الارتباط إلى قيم Z بطريقة فيشر وفق المعادلة التالية: $Z_i = \frac{1}{2}\ell n\left[\frac{(1+r_i)}{(1-r_i)}\right]$ (علام، المرجع السابق، 156).

وبلغت قيمة (Z= 3.55) وهي أكبر من Z الجدولية عند مستوى الدلالة (0.05) وهي (1.96) وبالتالي توجد فروق ذات دلالة إحصائية عند مستوى الدلالة (0.05) بين معاملي ارتباط كل من الصورتين باختبار الرياضيات في الامتحان النصفي.

يتضح مما سبق أن الصورتين لم تحققا معايير التكافؤ جميعها، وعدم التكافؤ ينشأ عنه صعوبة المقارنة بين درجات التلاميذ على الصورتين، ولذا تنشأ الحاجة إلى معادلة درجاتهم على الصورتين.

نتائج السؤال الثاني: ما الدرجات المتعادلة لصورتي اختبار تحصيلي ذي الاختيار من متعدد في الرياضيات للصف الرابع وفق كل من الطريقة الخطية، والطريقة المئينية؟

للإجابة عن السؤال الثاني للبحث جرى حساب الدرجات المتعادلة للصورتين وفق كل طريقة كما يلي:

أولاً– طريقة المعادلة الخطية الأفقية: جرى تحويل الدرجة على الصورة (Y) إلى ما يناظرها على الصورة (X) وفق المعادلة التالية: $y = \left(\frac{S_y}{S_x}\right)x + \left(\overline{Y} - \frac{S_y}{S_x}\overline{X}\right)$

حيث إنّ yهي الدرجة على الصورة الأولى والمعادلة للدرجة x من نتيجة الصورة الثانية، و S_x و $\overline{X}$ هما الانحراف المعياري والمتوسط الحسابي للدرجات على الصورة الأولى، أما S_y و $\overline{Y}$ هما الانحراف المعياري والمتوسط الحسابي للدرجات على الصورة الثانية (الحربي، 2009، 24)، والجدول (3) يوضح الدرجات الخام المتعادلة للصورتين وفق الطريقة الخطية:

الجدول (3) الدرجات الخام المتعادلة للصورتين وفق الطريقة الخطية

الدرجات الخام المتعادلة للصورتين		الدرجات الخام المتعادلة للصورتين	
الصورة (X)	الصورة (Y)	الصورة (X)	الصورة (Y)
0	1.17	13	12.35
1	2.03	14	13.21
2	2.89	15	14.07
3	3.75	16	14.93
4	4.61	17	15.79
5	5.47	18	16.65
6	6.33	19	17.51

7	7.19	20	18.37
8	8.05	21	19.23
9	8.91	22	20.09
10	9.77	23	20.95
11	10.63	24	21.81
12	11.49		

يلاحظ من الجدول السابق أن الدرجات لا تتكافأ في الصورتين، مما قد يؤدي إلى اتخاذ قرارات خاطئة فيما لو طبقت إحدى الصورتين على بعض التلاميذ وطبقت الأخرى على بعضهم الآخر.

ثانياً– طريقة المعادلة المئينية الأفقية: لإيجاد الدرجات الخام المتعادلة للصورتين وفق الطريقة المئينية الأفقية جرى إيجاد الدرجات الخام التي تناظر رتباً مئينية مختارة، حيث حسبت الرتب المئينية المختارة (1، 10، 20،...، 100) وذلك لتوزيعات الدرجات على صورتي الاختبار، حيث تعد الدرجات الخام متكافئة على الصورتين إذا كان لها الرتبة المئينية نفسها، والجدول (4) يبين الدرجات الخام المتعادلة في نماذج الاختبار عند الرتب المئينية المختارة:

الجدول (4) الدرجات الخام المتعادلة للصورتين وفق الطريقة المئينية

الرتبة المئينية	الدرجات الخام المتعادلة للصورتين		الرتبة المئينية	الدرجات الخام المتعادلة للصورتين	
	الصورة (X)	الصورة (Y)		الصورة (X)	الصورة (Y)
1	2	3	60	15	14
10	7	8	70	16.7	15
20	9	9	80	18	17
30	11	11	90	20	18
40	13	12	100	24	24
50	14	13			

يلاحظ من النتائج في الجدول السابق أنها تتفق ونتائج المعادلة التي جرى التوصل إليها باستخدام الطريقة الخطية، إذ إنّ درجات التلاميذ على الصورتين غير متكافئة.

نتائج السؤال الثالث: أيهما أكثر فاعلية الطريقة الخطية أم الطريقة المئينية في حالة المعادلة الأفقية باستخدام تصميم المجموعة الواحدة؟

للتحقق من فاعلية المعادلتين اعتمد الخطأ المعياري للمعادلة كأساس للمقارنة فيما بينهما، ولمعرفة دقة المعادلة التي جرت من خلال الطريقة الخطية حسب الخطأ المعياري للمعادلة وفق المعادلة التالية:

$$S.E_y = \sqrt{\frac{2S_x^2(Z_y+2)}{N_T}}$$

حيث E_y.: الخطأ المعياري لـ y عند الدرجة y ، Z_y : الدرجة المعيارية المناظرة للدرجة y على التوالي، N_T: العدد الكل للأفراد في المجموعة التي أجابت عن صورتي الاختبار X و Y (الشريفين، 2003، 50)، والجدول (5) يوضح الخطأ المعياري للمعادلة الخطية عند درجات معيارية معينة:

الجدول (5) الخطأ المعياري للمعادلة الخطية عند درجات معيارية معينة

الدرجة المعيارية	الخطأ المعياري للمعادلة الخطية S.E
0.0	0.008
0.5±	0.009
1.0±	0.0102
1.5±	0.0110
2.0±	0.0117
2.5±	0.0124
3.0±	0.0131

لمعرفة دقة المعادلة التي جرت من خلال الطريقة المئينية حسب الخطأ المعياري للمعادلة وفق المعادلة التالية:

$$SE_{[e_x(y_0)]} = \sqrt{\sigma_x^2 \left(\frac{Pq}{\emptyset^2}\right)\left(\frac{1}{N_x}+\frac{1}{N_y}\right)}$$

حيث $SE_{[e_x(y_0)]}$: الخطأ المعياري للعلامات المحولة $[e_x(y_0)]$ بالطريقة المئينية، P : نسبة العلامات التي تقل عن قيمة معينة في الصورة Y و $q = (1-p)$، N_x : حجم العينة التي طبق عليها الاختبار x، N_y : حجم العينة التي طبق عليها الاختبار y، $\emptyset$: الاحداثي الصادي لمنحنى التوزيع الطبيعي المعياري عند النقطة التي تكون المساحة تحتها تساوي P (ارتفاع المنحنى)، σ_x^2 : تباين العلامات على الاختبار x (Peterson, Kolen & Hoover, 1989,252)، والجدول (6) يوضح الخطأ المعياري للمعادلة المئينية عند درجات معيارية معينة:

الجدول (6) الخطأ المعياري للمعادلة المئينية عند درجات معيارية معينة

الدرجة المعيارية	الخطأ المعياري للمعادلة المئينية S.E
0.0	0.25
0.5±	0.26
1.0±	0.29
1.5±	038
2.0±	0.55
2.5±	0.89
3.0±	1.614

بالنظر إلى الجدولين (5، و6) يلاحظ أن الخطأ المعياري للمعادلتين الخطية والمئينية يختلف باختلاف الدرجة المعيارية، وهو أقل ما يمكن عند المتوسط الحسابي (Z=0) ويزداد هذا الخطأ كلما ابتعدنا عن المتوسط من أي الجهتين، وهذا يشير إلى أنّ المعادلة تكون أدق ما يمكن عند القيم المتوسطة للدرجات وتتفق هذه النتيجة مع ما توصل إليه الشريفين (2003).

أما عند مقارنة الخطأ المعياري للمعادلة الخطية بالخطأ المعياري للمعادلة المئينية نلاحظ بأن الخطأ المعياري للمعادلة المئينية أكبر من نظيره في المعادلة الخطية ما يؤدي إلى أنّ المعادلة الخطية هي أكثر دقة وتتفق هذه النتيجة مع ما توصلت إليه كل من دراسة ماركو وبترسن وستيورات (1989)، ودراسة الشريفين (2003)، ودراسة بترسون وكوك وستوكنغ (1983) والتي أشارت إلى أنّ طريقة المعادلة الخطية هي الأكثر فاعلية إذا كانت الاختبارات المستخدمة للمعادلة متكافئة، وتتعارض مع ما توصلت إليه كل من دراسة كولن (1981)، ودراسة كولن وويتني (1982). ويمكن رد ذلك إلى كون توزيع الرتب المئينية في مجموعة من البيانات يقترب توزيعها من التوزيع الطبيعي متباعداً في الأطراف وكثيفاً في الوسط، وانتقال معين في الرتبة المئينية يناظره انتقالاً غير متناسب معه في الدرجة الخام، بينما يعتمد التحويل الخطي على الانحراف المعياري كوحدة قياس ويتحقق فيه مقياس الفترات (*Interval Scale*) ويكون الانتقال في الدرجة الخام متناسباً مع الانتقال في الدرجة المعيارية. كما يمكن أن يعزى ذلك إلى أنّ المعادلة الخطية تحليلية وتخلو من أخطاء التمهيد والتي قد تكون خطيرة عند بعض القيم للدرجات التي تكون عندها البيانات قليلة، والتي تكون عادة في ذيلي التوزيع للدرجات، كما أن المعادلة المئينية تعاني من أثري كل من سقف الاختبار وأرضيته، إذ إنه بالتعريف تكون أكبر درجة خام على إحدى الصورتين معادلة لأكبر درجة خام على الصورة الثانية، كما أنّ أصغر درجة في الصورة الأولى تعادل أصغر درجة في الاختبار الثاني وهذا لا يحدث في المعادلة الخطية، مثلاً إذا كانت الصورة (X) أقل صعوبة من الصورة (Y) فإنّ استعمال طريقة المعادلة المئينية يعمل على ضغط الدرجات للأفراد على الصورة (X) حتى يؤدي ذلك إلى حصول الأفراد على درجات متشابهة، لكن استخدام المعادلة الخطية يعمل على المحافظة على الفروق بين الأفراد وهذا يكون من خلال تحويل الدرجات على الصورة (X) إلى درجات على الصورة (Y) خارج مدى درجات الصورة (X).

وقد يكون السبب في هذه النتيجة كما بين كولن (Kolen, 1988) أنّه للحصول على درجة متساوية من الدقة في التحويلين المئيني والخطي فإنّ النسبة بين حجم العينة في المعادلة الخطية إلى حجم العينة في المعادلة المئينية تساوي 1: 1,57 ، وذلك عند قيمة (Z=0) وهذا يعني أنه يلزم 1570 مفحوصا في عينة المعادلة المئينية للحصول على نفس درجة الدقة في المعادلة الخطية عندما يستخدم في العينة 1000 مفحوصا فقط (الطراونة، 2005، 120).

الاستنتاجات والتوصيات:

يهدف البحث إلى المقارنة بين الطريقتين الخطية والمئينية لمعادلة درجات صورتي اختبار تحصيلي ذي الاختيار من متعدد للصف الرابع من التعليم الأساسي، وبناءً على نتائج البحث يمكن التوصل إلى الاستنتاجات والتوصيات الآتية: – وجود فروق ذات دلالة إحصائية بين متوسط درجات التلاميذ على صورتي الاختبار، ولم تحقق الصورتين جميع معايير التكافؤ.

– يختلف الخطأ المعياري للمعادلتين الخطية والمئينية باختلاف الدرجة المعيارية وهو أقل ما يمكن عند المتوسط الحسابي.

– الطريقة الخطية أكثر فاعلية من الطريقة المئينية لمعادلة درجات تلاميذ الصف الرابع على صورتي اختبار تحصيلي ذي الاختيار من متعدد في مادة الرياضيات .

– استخدام الطريقة الخطية عند معادلة درجات صور الاختبارات التحصيلية في حالة المعادلة الأفقية باستخدام تصميم المجموعة الواحدة.

– القيام بمزيد من الدراسات في مجال المعادلة، واستخدام طرائق مختلفة والمقارنة بينها باستخدام معايير مختلفة، وباستخدام تصميمات مختلفة ولصفوف دراسية أخرى.

المراجع:

1. أيوب، حسين. *المقارنة بين أربع طرق للمعادلة عندما يكون التصميم من مجموعات متكافئة وغير متكافئة*. رسالة دكتوراه غير منشورة. الجامعة الأردنية، 1994، 182.
2. بالخيور، شفاء. *فاعلية طرق معادلة نماذج اختبار القدرات العامة بالمركز الوطني للقياس والتقويم وفق نظريتي القياس التقليدية والحديثة في ضوء بعض المتغيرات*. رسالة دكتوراه غير منشورة. جامعة أم القرى، 2008،
3. الحربي، عيسى. *أثر تمثيل الفقرات للمحتوى ونسبة الفقرات المشتركة وطرق توزين الفقرات على دقة معادلة درجات الصور للاختبار عندما يكون التصميم المستخدم هو الجذع المشترك "الفقرات المشتركة"*. رسالة دكتوراه غير منشورة. الجامعة الأردنية، 2009، 103.
4. الحموري، هند. *بناء اختبار رياضيات متعدد المستويات لطلبة الصفوف الأربعة الأولى من المرحلة الأساسية*. رسالة دكتوراه غير منشورة. الجامعة الأردنية، 1991، 156.
5. الدوسري، راشد. معادلة الاختبارات مفهومها، وطرقها، ومشكلات تطبيقها. *مجلة العلوم التربوية والنفسية*. مؤسسة الأيام للصحافة والنشر، البحرين. المجلد الثاني، العدد الرابع، 2001، 107– 141.
6. ساعاتي، أمين. *تبسيط كتابة البحث العلمي من البكالوريوس وثم الماجستير وحتى الدكتوراه*. ط1، المركز السعودي للدراسات الإستراتيجية، مصر الجديدة، 1999، .
7. الشريفين، نضال. *مدى تحقق معايير الفاعلية في معادلة اختبارين أحدهما ثنائي التدريج والآخر متعدد التدريج وفق نماذج النظرية الكلاسيكية والنظرية الحديثة في القياس*. رسالة دكتوراه غير منشورة. جامعة عمان العربية للدراسات العليا، 2003، 259.
8. الصمادي، اسماعيل. *أثر طريقة اختيار الفقرات في اختبار الجذع المشترك على دقة معادلة اختبار متعدد المستوى في الرياضيات للمرحلة الأساسية في الأردن*. رسالة دكتوراه غير منشورة. جامعة عمان العربية للدراسات العليا، 2008، 136.
9. الطراونة، صبري. تطوير اختبار رياضيات متعدد المستويات للصفوف الأساسية (3 – 6) بفقرات متعددة التدريج. رسالة دكتوراه غير منشورة. جامعة عمان العربية للدراسات العليا، 2005، 165.
10. عبد الله، اعتدال. *استخدام نموذج راش في تدريج مقياس للقدرة العقلية لدراسة بعض العوامل المؤثرة على دقة القياس*. رسالة دكتوراه غير منشورة. جامعة عين شمس، 2003، 256.
11. العبسي، محمد. *التقويم الواقعي في العملية التدريسية*. دار المسيرة للنشر والتوزيع، عمان، 2010، 215.
12. علام، صلاح الدين. *القياس والتقويم التربوي والنفسي "أساسياته، وتطبيقاته وتوجهاته المعاصرة"*. دار الفكر العربي، القاهرة، 2002، 823.
13. علام، صلاح الدين. *الأساليب الإحصائية الاستدلالية في تحليل بيانات البحوث النفسية والتربوية والاجتماعية – البارامترية واللابارامترية–*. دار الفكر العربي، القاهرة، 2005، 630.

14. علي، أسامة. *دراسة مقارنة لبعض إجراءات تكافؤ البيانات الاختبارية التعويضية باستخدام نظرية الاستجابة للمفردة متعددة الأبعاد*. رسالة ماجستير غير منشورة. جامعة جنوب الوادي، 2004، 227.

15. HARRIS, D. J. & KOLEN, M. J. *Effect of Examinee Group on Equating Relationships.* Applied Psychological Measurement. Vol.10, 1990, 35- 43.
16. HUYNH, H. *A Comparison of Equipercentile and Partial Credit Equating for Performance- based Assessment Composed of Free- response Items.* Journal of Educational Measurement. Vol.31, N.2, 1994,125- 141.
17. KOLEN, M.J. *Comparison of Traditional and Item Response Theory Methods of Equating Tests.* Journal of Educational Measurement. Vol.18, N.1, 1981, 1- 11.
18. KOLEN, M.J., & WHITNEY, D. R. *Comparison of Four Procedures for Equating The Tests of General Educational Development.* Journal of Educational Measurement. Vol.9, N.4, 1982, 279- 293.
19. MARCO, G. L., PETERSEN, N. S, & STEWART, E. E. *A Test of The Adequacy of Curvilinear Score Equating Models.* In D. J. Weiss (Ed), New Horizons in Testing: Latent Trait Test Theory and Computerized Adaptive Testing. New York, 1989, 147- 177.
20. PETERSON, N. S., COOK,L.L. & STOCKING, M.L. *IRT Versus Conventional Equating Methods: A Comparative study of Scale Stability.* Journal of Educational Statistics. Vol.8, 1983, 137- 156.
21. PETERSON, S., KOLEN, J., & HOOVER, D. *Scaling Norming and Equating.* Educational Measurement. (ED). By Linn, L.: 3rd ed, American Council on Education : Washington, 1989, 241- 262.
22. SUNNASSEE, D. *Conditions Affecting the Accuracy of Classical Equating Methods For Small Samples Under the NEAT Design*: A Simulation Study. A Published Doctoral Dissertation, University of North Carolina, USA, 2011, 230.
23. YANG, L. *The Effects of Content Mix and Equating Method on the Accuracy of Test Equating Using Anchor-Item Design.* A paper presented at the Annual Meeting of the American Educational Research Association, Chicago,1997, March 24- 28, 37.
24. YANG, L. *The Effects of Content Homogeneity and Equating Method on the Accuracy of Common-Item Test Equating***.** A paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, 2000, April 24-28, 27.
25. YANG, L., & HOUANG, T. *The Effect of Anchor Length and Equating Method on the Accuracy of Test Equating: Comparisons of Linear and IRT Based Equating Using an Anchor- Item Design.* A paper presented at the Annual Meeting of the American Educational Research Association, New York, 1996, April 8- 12, 81.