365

المجلد 13 العدد 2 يونيو 2012

مقارنة فاعلية طريقتي معادلة العلامات الحقيقية والمشاهدة في معادلة الاختبارات باستخدام جذع مشترك ومجموعات غير متكافئة

د. رائد فايز المدانات
وزارة التربية والتعليم
المملكة الأردنية الهاشمية

366
المجلد 13 العدد 2 يونيو 2012

مقارنة فاعلية طريقتي معادلة العلامات الحقيقية والمشاهدة في معادلة الاختبارات باستخدام جذع مشترك ومجموعات غير متكافئة

د. رائد فايز المدانات
وزارة التربية والتعليم
المملكة الأردنية الهاشمية

الملخص

هدفت هذه الدراسة إلى مقارنة فاعلية طريقة معادلة العلامات الحقيقية وطريقة معادلة العلامات المشاهدة في معادلة الاختبارات عند استخدام جذع مشترك وتصميم المجموعات غير المتكافئة، ولتحقيق هذا الهدف تم بناء صورتين متكافئتين لاختبار في الفيزياء عدد فقرات كل منهما (٢٠) فقرة، بالإضافة إلى (١٠) فقرات استخدمت كاختبارٍ جذع مشترك. تكونت عينة الدراسة من مجموعتين من الطلبة تم اختيارهما عشوائياً إذ تقدمت المجموعة الأولى للصورة الأولى من الاختبار في الفصل الدراسي الأول، وتقدمت المجموعة الثانية للصورة الثانية من الاختبار في الفصل الدراسي الثاني وعدت مع المجموعة الأولى مجموعتين غير متكافئتين.

تم استخدام طريقتين للمعادلة تتبعان النظرية الحديثة في القياس وهما: طريقة معادلة العلامات الحقيقية وطريقة معادلة العلامات المشاهدة واربعة أحجام من فقرات اختبار الجذع المشترك. أظهرت نتائج الدراسة وجود فروق دالة إحصائيا في متوسطات القيم المعادلة بطريقة المعادلة بالعلامات الحقيقية وطريقة المعادلة بالعلامات المشاهدة عندما كانت عدد فقرات الجذع المشترك (٤، ٧، ١٠) لصالح طريقة معادلة العلامات المشاهدة، ولكن الفروق لم تكن ذات دلالة عندما كان الجذع المشترك فقرة واحدة. كما أشارت النتائج إلى عدم وجود أثر لعدد فقرات الجذع المشترك باستخدام أي من الطريقتين.

الكلمات المفتاحية: طريقة معادلة العلامات الحقيقية، طريقة معادلة العلامات المشاهدة، معادلة الاختبارات، جذع مشترك.

* تاريخ تسلم البحث: ٢٠١٠/٣/٣م * تاريخ قبوله للنشر: ٢٠١٠/١١/٣م

Comparing the Efficiency of True and Observed Score Equating Methods in Equating Tests Utilizing the Design with an Anchor Test and Nonrandom Groups

Dr. Raed F. AL- Madanat
The Ministry of Education
Hashemite Kingdom of Jordan

Abstract

The aim of the study was to compare the efficiency of true score equating method and observed score equating method in equating tests Utilizing the design that is based on the anchor test and Nonrandom groups. For achieving that aim, two equivalent forms of a physics test were constructed with (20) items for each one of them, in addition to a group of (10) items used as an anchor test.

The population of the study consisted of the secondary stage students/ level three in 2008/2009, while the sample of the study was selected from directories of education in the governorate of karak where the sample of the study consisted of two groups: The first group was given the first form of the test in addition to an anchor test at the first semester. The second group was given the second form of the test in the second semester, and these two groups were considered as two non- equivalent groups.

Two equating methods that follow the IRT theory were used, the true score equating method and the observed score equating method and four numbers of items of the anchor test(1, 4, 7, 10).

The results and according to the performance of the non-equivalent groups revealed that there was a statistical significance in the means of the equated scores using the true score equating method and the observed score equating method when the number of items of the anchor test were (1, 4, 7) in favor of the observed score equating method, but the differences had no statistical significance when the anchor test had (1) item.

The results also revealed that there was no impact of the number of items of the anchor test by using any of the two methods.

Key words: true score equating method, observed score equating method, anchor items, test equating, Nonrandom Groups.

المجلد 13 العدد 2 يونيو 2012

مقارنة فاعلية طريقتي معادلة العلامات الحقيقية والمشاهدة في معادلة الاختبارات باستخدام جذع مشترك ومجموعات غير متكافئة

د. رائد فايز المدانات
وزارة التربية والتعليم
المملكة الأردنية الهاشمية

المقدمة

يُعد مجال معادلة الاختبارات من المجالات التطبيقية المهمة في القياس النفسي والتربوي، فقد بدأَ الاهتمامُ بمعادلة الاختبارات من قبل الكثير من المؤسسات التربوية مثل الرابطة الأمريكية للبحث التربوي، وجمعية علم النفس الأمريكية (Kolen & Brennan, 2004). ويعود سبب هذا الاهتمام إلى تعدد البرامج الاختبارية وتنوعها إذ تستخدمُ صورٌ متعددة من الاختبارات وتتطلبُ أن يكون لدينا نماذج متكافئة من نفس الاختبار. وتُستخدمُ نتائج الاختبارات في المجالات التربوية والنفسية من أجل تزويد صاحب القرار أو المعلم بمعلومات يمكن استخدامها في اتخاذ قرارات تتعلق بالتعيين أو الترقية أو إعطاء الطالب تقديرات تتعلق بتحصيله الأكاديمي. ولإجراء المعادلة بين نموذجين (صورتين) لنفس الاختبار فإننا نختار التصميم المناسب أولا ثم نطبق النموذجين على عينة من الأفراد، وبعد ذلك نجري التحليلات الإحصائية المناسبة التي تمكننا من تحويل العلامات من نموذج إلى آخر (Angoff, 1971). بحيث إن العلامة على أي من الصور يكون لها نفس الدلالة ونفس القيمة القياسية لو تحققت نفسها على صورة أخرى، ويقال عندئذ إنّ هناك تكافؤاً تاماً بين صور الاختبار الواحد.

هناك العديد من طرق المعادلة التي تتبع النظرية الكلاسيكية وأخرى تتبع نظرية الاستجابة للفقرة، والطرق الأكثر استخداماً في النظرية الكلاسيكية: طريقة المعادلة الخطية ومنها طريقة تكر، وطريقة ليفين، وطريقة براون–هولند وطريقة المعادلة المئينية. والطرق الأكثر استخداما في النظرية الحديثة: طريقة معادلة العلامات الحقيقية، وطريقة معادلة العلامات المشاهدة (Kolen & Brennan, 2004).

بسبب أهمية القرارات المترتبة على نتائج تطبيق الاختبارات، وحتى نتجنب شيوعها عند تكرار تطبيقها، وللحد من تكرار تطبيقها على نفس المفحوصين يتم اللجوء إلى بناء عدة صور متكافئة للاختبار، واستخدام أساليب إحصائية لمعادلة هذه الصور (Kolen &

Whitney, 1982)، وتظهر الحاجة للمعادلة في الوضعين الآتيين:

١- المعادلة الأفقية (Horizontal Equating): وتستخدم بوضع يكون لدينا فيه اختباران متكافئان في مستوى الصعوبة، وتكون توزيعاتُ السمةِ المرادِ قياسها متشابهةً بين المفحوصين الذين يطبق عليهم الاختباران، ويكون الهدف من المعادلة هو تعديل الفروق الناتجة عن الاختلاف في مستويات الصعوبة بين صورتي الاختبار.

٢- المعادلة العمودية (Vertical Equating): وتستخدم عندما يكون لاختبارين أو أكثر مستويات مختلفة من الصعوبة، وتوزيعات القدرة للمفحوصين مختلفة، وينطبق ذلك على ما يُعرفُ بالاختبارات متعددة المستويات (الدوسري، ٢٠٠٤؛ Hambleton & Swaminathan, 1985).

مفهوم المعادلة وشروطها

يعرف كروكر والجاينا (Crocker & Algina, 1986) المعادلة بأنّها عملية الحصول على درجات متكافئة لأداتين تقيسان السمة نفسها. ويعرفها كولن (Kolen, 1981) بأنها عملية تحويل نظام وحدات اختبار في صورة منه إلى ما يقابلها في صورة أخرى لكي تكون علامات صورتي الاختبار متكافئةً بعدَ عمليةِ التحويلِ. كما يمكن تعريفها بأنها إجراء إحصائي يؤسس علاقة بين علامات اختبارين أو أكثر ويضع هذه العلامات على مقياس مشترك (المدانات، ٢٠٠٨)، ولإجراء عملية المعادلة يجب توافر عدة شروط وهي (Angoff, 1971; Kolen & Brennan, 2004; Lord, 1980):

١. أنْ تقيسَ الاختبارات نفس السمة الكامنة (Latent Trait) أو القدرة .

٢. العدالة (المساواة) (Equity) ويعني ذلك أنْ يكونَ التوزيعُ التكراري المشروط للدرجات عند مستوى معين من مستويات القدرة (θ) للاختبار (Y) بعد تحويل الدرجات هو نفس التوزيع التكراري المشروط للدرجات عند مستوى معين من مستويات القدرة (θ) للاختبار (X) (Angoff, 1971).

$$F(x|\theta) = F(y|\theta)$$

بمعنى أنّ العلامات على الاختبار (x) والاختبار (y) يجب أن تكون قابلة للتبادل بعد إجراء عملية المعادلة.

٣. اللاتباين (اللاتغير) في مجتمع الدراسة (Population invariance) أيّ إنّ تحويل الدرجات يجب أن يبقى كما هو بصرف النظر عن مجموعة المفحوصين إذ تم اشتقاقه من نتائجها في الاختبار.

٤. التماثل (Symmetry) ويعني أن تحويل الدرجات من صورة إلى أخرى يجب أن يكون قابلاً للانعكاس (Invertible). أي أن الرسم البياني لتوزيع الدرجات من الصورة (X) للاختبار إلى الصورة (Y) يجبُ أنْ يكونَ نفسُ الانعكاس من الصورة (Y) إلى الصورة (X) من نفس الاختبار (Kolen & Brennan, 2004).

٥. الثبات: يجب أن تكونَ الاختباراتُ متساوية في مستوى الثبات حتى يمكن معادلتها (Lord, 1980).

٦. يجب أنْ تكونَ الاختباراتُ المراد معادلتها متكافئة (Parallel) بمعنى أنَّ المتوسط الحسابي والانحراف المعياري والتباين للأداء على الصورة الأولى للاختبار ومستوى الصعوبة ومستوى التمييز لفقرات الصورة الأولى من الاختبار تماثل نظيراتها في الصورة الثانية من نفس الاختبار (Crocker &Algina,1986).

تصميمات جمع البيانات في معادلة الاختبارات

يوجد العديد من التصاميم والطرق التي طورت لجمع بيانات ميدانية من أجل معادلة علامات الاختبار، ومن أكثر التصاميم شيوعا ما يلي (المدانات، ٢٠٠٨):

أولاً: تصميم المجموعة الواحدة:

يتم في هذه الطريقة اختيار عدد كبير من المفحوصين من مجتمع متجانس. إذ تطبق صورتا الاختبار المراد معادلتهما إحداهما تلو الأخرى على نفس مجموعة المفحوصين.

ثانيا: تصميم المجموعات المتكافئة:

يُعطى كلا الاختبارين المراد معادلتهما لمجموعتين عشوائيتين من المفحوصين، بحيث تأخذ كل مجموعة صورة واحدة.

ثالثاً: تصميم المجموعات العشوائية المتوازنة:

يتم تقسيم مجموعة المفحوصين إلى مجموعتين متساويتين بشكل عشوائي، ثم تُطبَّقُ عليهم صورتا الاختبار بالتناوب.

رابعاً: التصميم القائم على اختبار مشترك ومجموعات عشوائية:

يتم تقسيم مجموعة المفحوصين إلى مجموعتين متساويتين بشكل عشوائي، وتُطبَّقُ الصورة الأولى من الاختبار على المجموعة الأولى، وفي الوقت نفسه تطبق الصورة الثانية من الاختبار على المجموعة الثانية ثم يتم تطبيق اختبار مشترك على المجموعتين كلتيهما في الوقت نفسه.

خامساً: التصميم القائم على اختبار مشترك ومجموعات غير عشوائية:

يتم تقسيم مجموعة المفحوصين إلى مجموعتين بشكل غير عشوائي ثم تُطبَّقُ الصورة الأولى من الاختبار على المجموعة الأولى، وتُطبَّقُ الصورةُ الثانيةُ من الاختبار على المجموعة الثانية ثم يتم تطبيق اختبار مشترك على المجموعتين في الوقت نفسه، ولضمان عدم التكافؤ يُفضَّلُ تطبيق الصورتين في أوقات مختلفة قد يفصل بينها أسابيع أو شهور.

وقد يكون الاختبار المشترك داخلياً أو خارجياً: فالاختبار المشترك الداخلي يتكون من مجموعة من الفقرات يتم تضمينها في صورتي الاختبارين. أمّا الاختبار المشترك الخارجي (External Anchor) فهو اختبار منفصل يقدم لكلا المجموعتين خارج الوقت الذي تقدم فيه صورتا الاختبار. وهناك عدة معايير لاختيار الفقرات الجذعية (المشتركة) وهي (Kolen & Brennan, 2004):

١- أن تمثل المحتوى بشكل جيد.

٢- أن يكون عددها عشرين فقرة أو (٢٠٪) من عدد الفقرات الكلي في كل من صورتي الاختبار أيهما أكبر.

٣- أنْ يتم تحديد معاملات الصعوبة والتمييز للفقرات وانتقاء أفضلها.

٤- إيجاد الفقرات المتحيزة وحذفها.

٥- أنْ تتمتع بدرجة ثبات مرتفعةٍ بحيث تُقدم البيانات التي يمكن استخدامها بفعالية لإجراء التعديلات المناسبة على الفروق بين المجموعات.

٦- ينصح بوضع فقرات الاختبار المشترك ضمن صور الاختبار؛ لتجنُّب أخذ هذه الفقراتِ في الجزء النهائي من الاختبار حيثُ يتوقع ظهور آثار السرعة (Angoff, 1971).

طرق معادلة الاختبارات

يصنف المختصون في القياس والتقويم التربوي طرق معادلة الاختبارات إلى (Thorndike, 1982; Hambleton & Swaminathan, 1985; Kolen & Brennan, 2004).

أ) طرق تتبع النظرية الكلاسيكية في القياس وتتضمن الطرق الآتية:-

• طريقة معادلة المتوسط الحسابي

• طرق المعادلة الخطية ومنها: طريقة تكرالخطية ، طريقة ليفين للعلامات المشاهدة ، وطريقة براون-هولند الخطية

• طرق المعادلة المئينية ومنها: المعادلة المئينية ذات التكرارات الممهدة ، والمعادلة المئينية ذات التكرارات غير الممهدة.

ب) طرق تتبع النظرية الحديثة في القياس (IRT) وتتضمن الطرق الآتية:

- طريقة معادلة العلامات الحقيقية.
- طريقة معادلة العلامات المشاهدة .
- المعادلة باستخدام النموذج أحادي المعلم (وثنائي المعلم وثلاثي المعلم).

طرق معادلة الاختبارات استناداً الى نظرية الاستجابة للفقرة (IRT):

الخطوة الأولى في المعادلة باستخدام النظرية الحديثة أو نظرية الاستجابة للفقرة هي تحديد فيما إذا كانت الاختبارات تحوي فقرات سبق إدراجها في نفس العينة، وهنا لا توجد ضرورة لإجراء عملية المعادلة، أما إذا لم تكن الاختبارات معادلة فنتبع الخطوات الآتية (Hambleton & Swaminathan, 1985):

١- اختيار التصميم المناسب لجمع البيانات من أجل المعادلة بالاعتماد على طبيعة الاختبارات المراد معادلتها وخصائص مجموعة المفحوصين ،والتصاميم التي تستخدم هنا هي ذاتها التي تستخدم في النظرية التقليدية التي سبق عرضها.

٢- تحديد نموذج الاستجابة المناسب للفقرة الذي يطابق بيانات الفقرة باستخدام مقاييس جودة المطابقة لتقييم النموذج.

٣- بناء مقياس عام) تدريج مشترك (يربط بين القدرة المراد قياسها) السمة (ومعلم صعوبة الفقرة .وحيث إن وحدة القياس ونقطة الأصل للقدرة والصعوبة غير محددتين فيتم تعينهما بناءً على قدرات المفحوصين الذين تم استخدامهم في معايرة الفقرات ،فيتم جعل المتوسط الحسابي لـ(θ) صفراً والانحراف المعياري يساوي واحداً صحيحاً. فعند استخدام النموذج أحادي أو ثنائي أو ثلاثي المعلمة فإنه يتم إجراء التحويل الخطي على كل من (ai , bi, θ) وفقاً للمعادلات التالية:

$$\theta' = A\theta + B$$

$$b' = Ab_i + B$$

$$a_i' = \frac{1}{A} a_i$$

حيث

A : ميل خط التحويل الخطي

B : المقطع الصادي للتحويل الخطي

وبذلك فإن

Pi(θ)=Pi(θ)

ويتم ذلك لأي تحويل خطي للمعالم.

ويقوم برنامج الحاسوب المستخدم (BILOG-MG) بوضع المعالم على نفس التدريج إذا تم تحليل بيانات المجوعتين الأولى والثانية معا، وهذا ما تم في هذه الدراسة لوضع معلمات القدرة والصعوبة على نفس التدريج. حيث تم الحصول على نقطة أصل ووحدة قياس للسمة ولمستوى الصعوبة. وكان متوسط درجات القدرة هو الصفر وانحرافها المعياري هو الواحد الصحيح. ومن البرامج الحاسوبية التي يمكن أن تضع معلمات القدرة والصعوبة على نفس التدريج برنامج (LOGIST) وبرنامج (ST).

٤- اختيار التدريج المناسب لتسجيل علامات الاختبار؛ أي بمعنى هل تكتب العلامات كعلامات خام (Scores Raw) أم على صورة علامات قدرة (Score Ability) أم على صورة درجات حقيقية (Scores True).

- إذا تم تسجيل العلامات بدلالة القدرة (θ) فإن الإجراء ينتهي.
- إذا تم تسجيل العلامات بدلالة تقديرات العلامة الحقيقية، فيجب تقدير العلامات الحقيقية على الاختبارات عند مستويات مختلفة من القدرة، ثم توضع بجداولَ أو رسوماتٍ لتنفيذ عملية المعادلة.
- إذا أردنا إجراء معادلة العلامات المشاهدة فيجب:

- إنشاء توزيعات نظرية مشروطة للعلامات المشاهدة المناظرة لقدرات أفراد عينة مختارة من المفحوصين.
- إنشاء توزيع نظري هامشي للعلامات المشاهدة.-
- تنفيذ المعادلة المئينية لمعادلة علامات الاختبار.

الطريقة الأولى: معادلة العلامات الحقيقية (True Score Equating)

تتضمن معادلة العلامات الحقيقية تحديد علامات حقيقية متكافئة على كل من صورتي الاختبار، ثم استخدام الدالة الرياضية التي تربط بين العلامات الحقيقية المتكافئة لمعادلة العلامات المشاهدة. وفي نظرية الاستجابة للفقرة تصاغ الدالة الرياضية التي تربط بين تقديرات السمة الكامنة والعلامات الحقيقية على الصورة

$$\xi = \sum_{i=1}^{n} P_i(\theta) \ldots\ldots (1)$$

إذ تشير المعادلة إلى أنه يمكن تحديد العلامة الحقيقية للمفحوص ذي القدرة (θ) على الاختبار (X) من خلال حساب احتمال إجابته إجابة صحيحة ($P_i(\theta)$ على جميع الفقرات المكونة للاختبار ثم إجراء عملية الجمع. ويتم حساب ($P_i(\theta)$ عند قيمة محددة لـ (θ) في نموذج أحادي المعلم (نموذج راش) من خلال هذه العلاقة (Crocker & Algina, 1986):

$$P_i(\theta)=\frac{e^{(\theta-b_i)}}{1+e^{(\theta-b_i)}}\ldots\ldots\ldots\ldots\ldots(2)$$

وعلى افتراض أن صورتي الاختبار تقيسان الخصائص نفسها (سمة كامنة، وقدرة، ومهارة)، وان معالم الفقرات لكلتا الصورتين قد تم وضعهما على نفس التدريج. فإن القيم (xζy & ζ) المناظرة لقيمة محددة لـ (θ) تمثل مستويات متماثلة من القدرة وقيماً تمت معادلتها. وتتم عملية المعادلة للعلامات الحقيقية على النحو الآتي:

لتكن العلامة الحقيقية للطالب على صورة الاختبار (X) هي ξ_x وعلامته الحقيقية على صورة الاختبار (Y) هي ξ_y. وترتبط العلامات الحقيقية بعلامات القدرة بالصورة الرياضية الآتية:

$$\xi_x=\sum_{i=1}^{n}P_i(\theta_x)\ldots\ldots\ldots(3)$$

حيث (xθ) تعبر عن مستوى قدرة المفحوص على صورة الاختبار (X)

و ξ_x تعبر عن العلامة الحقيقية للمفحوص على صورة الاختبار (X)، كذلك

$$\xi_y=\sum_{j=1}^{m}P_j(\theta_y)=\sum_{j=1}^{m}P_j(\alpha\theta_x+\beta)\ldots\ldots\ldots(4)$$

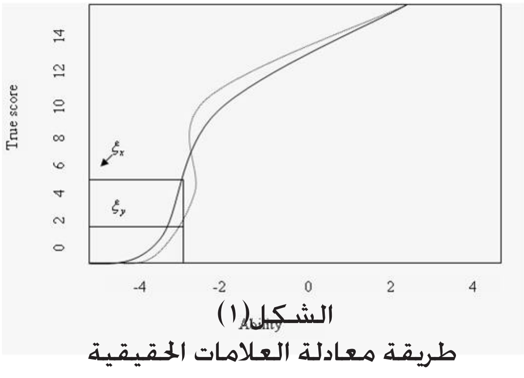
حيث (yθ) تعبر عن مستوى قدرة المفحوص على صورة الاختبار (Y)

و ξ_y تعبر عن العلامة الحقيقية للمفحوص على صورة الاختبار (Y)

وترتبط (y θ& xθ) بالمعادلة الرياضية الآتية :

$$\theta_y=\alpha\theta_x+\beta$$

وعند رسم منحنى العلاقة بين (y θ& xθ)، يمكن تحديد القيمتين (yξ& xξ) عند قيم محددة لـ (xθ) واللتين تمثلان قيماً تمت معادلتها، كما في الشكل التالي:



الشكل(1)
طريقة معادلة العلامات الحقيقية

فمثلا عند قيمة (xθ= -٣) فإن العلامة الحقيقية للمفحوص على صورة الاختبار (X) تساوي (٥) في حين العلامة الحقيقية لنفس المفحوص على صورة الاختبار (Y) تساوي (٢).

ولتحديد القيم (y θ& xθ) يمكن استخدام عدة طرق (Kolen & Bernnan, 2004)

١- طرق الانحدار (Regression Methods)

٢- طريقة المتوسط الحسابي والانحراف المعياري (Mean and Sigma) .

٣- طريقة روباست للمتوسط الحسابي والانحراف المعياري (Robust Mean and Sigma).

٤- طريقة خصائص المنحنى (Characteristic Curve Methods)

ويشير هامبلتون وسوامنثان (Hambleton & Swaminathan, 1985) إلى أن طريقة خصائص المنحنى لتحديد ثوابت المعادلة هي أكثر الطرق مناسبة للاستخدام في معادلة العلامات الحقيقية وتتم هذه الطريقة كالآتي:

لتكن العلامة الحقيقية (ξ) للمفحوص ذي القدرة (θ_a) على الاختبار (X) على الشكل التالي:

$$\xi_{xa} = \sum_{i=1}^{n} P\,(\theta_a; a_{xi}, b_{xi}, c_{xi}) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (5)$$

والعلامة الحقيقية (ξ) للمفحوص ذي القدرة (θ_a) على الاختبار (Y) على الشكل التالي

$$\xi_{ya} = \sum_{i=1}^{n} P\,(\theta_a; a_{yi}, b_{yi}, c_{yi}) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (6)$$

حيث

$$b_{yi} = \alpha b_{xi} + \beta \ldots\ldots\ldots\ldots\ldots (7)$$

$$a_{yi} = \frac{a_{xi}}{\alpha}$$

$$c_{yi} = c_{xi}$$

حيث تمثل

cyi : بارامير التخمين للفقرة (i).

ayi : بارامير التمييز للفقرة (i).

byi : بارامير الصعوبة للفقرة (i).

D : قيمة ثابتة وتساوي (١,٧).

ويجب اختيار قيم الثابتين (β) و(α) بحيث يكون الفرق بين العلامات الحقيقية (xaζ) و(yaζ) أقل ما يمكن. وحسب ما يشير له ستوكنج ولورد (Stoking and Lord, 1983) فإن

المحك المناسب لاختيار قيمة الثابتين (β) و(α) هو

$$F = \frac{1}{N}\sum_{a=1}^{N}(\xi_{xa} - \xi_{ya})^2$$

حيث (N) تشير إلى عدد المفحوصين. والدالة (F) هي اقتران للثوابت (β) و(α) ونحصل على القيمة الدنيا لـِ (F) عندما تكون المشتقة الأولى للدالة بالنسبة (α) وبالنسبة (β) تساوي صفراً.

$$\frac{\partial F}{\partial \alpha} = \frac{\partial F}{\partial \beta} = 0$$

معادلة العلامات المشاهدة: (Observed- Score Equating)

من المشكلات التي تواجه معادلة الاختبارات بطريقة العلامات الحقيقية عدم إمكان معادلة علامة مفحوص علامته الخام أدنى من مستوى الصدفة (درجة التخمين). ففي العلامات الخام تكون ادني علامة هي الصفر، في حين تكون أدنى علاقة في العلاقات الحقيقية هي

$$\xi = \sum_{i=1}^{n} C_i$$

تقوم معادلة الاختبار بطريقة العلامات المشاهدة على فكرة التنبؤ بالتوزيع النظري للعلامات الخام (Theoretical Observed Score Distribution) للاختبار عن طريق بناء التوزيع التكراري الذي تمثله دالة توزيع العلامات الخام $f(r|\theta)$ لمفحوص قدرته (θ). وتزودنا نظرية الاستجابة للفقرة بطرق للتنبؤ بالتوزيع نظريا عند تقديم اختبار ما، وعند استخراج التوزيعات النظرية للعلامات الخام المشاهدة للاختبارين (y & x) يمكن إجراء المعادلة المئينية.

ويمكن استخراج التوزيع النظري للعلامات المشاهدة $f(r|\theta)$ على اختبار من خلال المتطابقة الآتية:

$$\sum_{r=0}^{n} f(r/\theta)t^r = \prod_{i=1}^{n}\left[Q_i(\theta) + P_i(\theta)\right] \ldots\ldots\ldots\ldots\ldots\ldots (8)$$

حيث

r : العلامة الخام المشاهدة.

$f(r|\theta)$: التوزيع النظري للعلامات المشاهدة .

t : معامل

Qi (θ) : احتمال الإجابة الخطأ.

Pi(θ) : احتمال الإجابة الصحيحة.

وحيث إن معلم القدرة (θa) ومعالم الفقرات ستكون غير معروفة فإنه يتم تعويض

تقديرات معلم القدرة وتقديرات معالم الفقرة لاستخراج تقدير لدالة استجابة الفقرة (Pi) وتقدير التوزيع التكراري الهامشي . f(r). وعند استخراج التوزيع التكراري للعلامات المشاهدة يتم تنفيذ المعادلة بالشكل الآتي:

١- وضع معالم القدرة و الفقرة على تدريج مشترك لجميع مجموعات الطلبة وصور الاختبار.

٢- استخراج التوزيع التكراري المشروط $f_x(r/\theta)$ من خلال المعادلة $\xi_x = \sum_{i=1}^{n} P_i(\theta_x)$ على صورة الاختبار (X) لكل مفحوص في مجموعة المفحوصين باستخدام تقديرات معالم القدرة والفقرة.

٣- استخراج التوزيع التكراري الهامشي $f_x(r)$ Distribution Frequency Marginal من خلال المعادلة

$$\xi_y = \sum_{j=1}^{m} P_j(\theta_y) = \sum_{j=1}^{m} P_j(\alpha\theta_x + \beta) \ldots \ldots \ldots (9)$$

٤-إعادة الخطوتين الثانية والثالثة للمفحوصين الذين يتقدمون لصورة الاختبار (Y).

٥- إجراء المعادلة باستخدام طريقة الرتب المئينية المتساوية بين العلامات الخام للاختبار الأول وللاختبار الثاني.

تعددت الدراسات التي تناولت معادلة الاختبارات، فقد قام كولن ووتني (Whitney & Kolen، ١٩٨٢) بدراسة هدفت إلى مقارنة مدى ملاءَمة أربع طرق للمعادلة الأفقية لبطارية اختبار التطوير التربوي العام (GED) Development Education General Of Test)) وطبقت على (١٢٠٠) طالب وطالبة وهذه الطرق هي: الطريقة الخطية، والطريقة المئينية، والنموذج أحادي المعلم، والنموذج ثلاثي المعلم، واستخدم محك الصدق التقاطعي كمعيار لتحديد الملاءمة النسبية. وقد أظهرت الدراسة عدم وجود طريقة مثلى من بين الطرق الأربع على الرغم من تفوق الطرق المعتمدة على نماذج النظرية الحديثة على الطريقة الخطية والمئينية في المعادلة.

كما هدفت دراسة جيالوكا وكريتشون وفالي (Gialluca, Crichton & Vale, 1984) إلى تقصي مدى فعالية طرق المعادلة للاختبارات العقلية، واستخدم في الدراسة بيانات واقعية وبيانات مولدة باستخدام الحاسوب (Simulated) من اختبارات القوة الجوية الأمريكية للمقارنة بين طرق المعادلة المختلفة ووصف أي الاختبارات يعمل بشكل أفضل مع طريقة المعادلة، وقد اختيرت طريقة المعادلة الخطية والمئينية وطرق نظرية الاستجابة للفقرة، بالإضافة إلى طريقة العلامات الحقيقية. وقد أشارت النتائج إلى أن الاختبارات المتكافئة

تكون معادلتها أفضل باستخدام الطرق الخطية والمئينية، وفي المقابل تكون المعادلة أفضل للاختبارات غير المتكافئة عند استخدام الطرق المعتمدة على نظرية الاستجابة للفقرة وبشكل محدد طريقة معادلة العلامات الحقيقية. كما أشارت النتائج إلى أن فائدة قليلة يمكن جنيها عند زيادة حجم العينة من (١٠٠٠) إلى (٢٤٠٠). كما أنّ دقة المعادلة لم تتأثر عند مضاعفة طول اختبار الجذع المشترك أو عند تقديمه سهلاً كان أم صعباً.

وأجرى يانغ وهوبانغ (Yang & Houang, 1996) دراسة هدفت إلى معرفة أثر طول الاختبار المشترك في دقة معادلة الاختبار باستخدام طريقة تكر الخطية (Tucker) وطريقتين تعتمدان على نظرية استجابة الفقرة. وكان الهدف معرفة فيما إذا كانت الدقة ستزداد عند استخدام فقرات جذعية أكثر. وما إذا كان أثر الجذع المشترك يعتمد على طريقة المعادلة المستخدمة، وقد جمعت البيانات من صورتين لاختبار الحد الأدنى للكفاءة الذي كان يشتمل على ١٩٧ و٢٠٣ فقرة على التوالي وتم إعداد ثلاثة أزواج من الصور المصغرة بطريقة التعيين العشوائي للفقرات، ثم تمت معادلة الأزواج بشكل منفصل. أشارت النتائج التي تم الحصول عليها من خلال طرق المعادلة الثلاث إلى أن هناك دقة بدرجة متوسطة بغض النظر عن طريقة المعادلة المستخدمة، إلا أن النتائج تميل لأن تكون أكثر دقة عند زيادة الفقرات الجذعية.

وأجرى تيانكي (Tianqi, 1997) دارسة هدفت إلى فحص التشابه والاختلاف بين إجراءات المعادلة لنماذج نظرية الاستجابة للفقرة والتشابه بين إجراءات المعادلة المئينية وطريقتين تنبثقان من نظرية الاستجابة للفقرة (طريقة معادلة العلامات الحقيقية وطريقة معادلة العلامات المشاهدة). أظهرت نتائج الدراسة أن معادلة العلامات الحقيقية كانت الأكثر استقراراً، وأن معادلة العلامات الملاحظة أكثر استقراراً من نتائج المعادلة بالطريقة المئينية، كما أظهرت النتائج أنه كلما زاد الفرق في الصعوبة بين صورتي الاختبار زاد الاختلاف بين طرق المعادلة.

ومن الدراسات العربية في هذا المجال دراسة أيوب (١٩٩٤) وهدفت إلى المقارنة بين أربع طرق لمعادلة الاختبارات وهي الطريقة الخطية والطريقة المئينية المنبثقتين عن النظرية الكلاسيكية، وطريقتين منبثقتين من النظرية الحديثة وهما طريقة نماذج أحادي المعلمة وطريقة ثنائي المعلمة. ولتحقيق أهداف الدراسة تم بناء ثلاثة اختبارات كل اختبار بصورتين، وتكونت العينة الكلية للدراسة من عينتين مستقلتين الأولى من (١٣٩٠) طالبا وطالبة والثانية من (١٤١٢) طالبا وطالبة. وللتحقق من فاعلية طرق المعادلة تم استخدام معامل الصدق التقاطعي. أشارت الدراسة في جزء من نتائجها التي تختص المعادلة الأفقية إلى أن

نماذج النظرية الحديثة في القياس كانت أكثر فاعلية من طريقتي المعادلة الخطية والمئينية، حيث كانت معاملات الصدق التقاطعي أقل ما يمكن لنموذج ثنائي المعلمة وتزداد قليلاً في نموذج أحادي المعلمة ثم تزداد بصورة واضحة في المعادلة المئينية وتصبح أكبر ما يمكن في الطريقة الخطية.

وأجرى الشريفين (٢٠٠٣) دراسة 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. ولتحقيق ذلك تم بناء اختبارين تحصيليين في الفيزياء أحدهما ثنائي التدريج والآخر متعدد التدريج، تكونت عينة الدراسة من (١٠٠٣) طالب وطالبة. أظهرت نتائج الدراسة وحسب معيار الصدق التقاطعي أن طريقة النموذج أحادي المعلمة كان الأكثر فاعلية من طريقتي المعادلة الخطية والمئينية. أما وفق معيار الخطأ المعياري للمعادلة فقد كانت المعادلة الخطية هي الأكثر فاعلية وأنتجت أقل خطأ معياري للمعادلة.

كما أجرى المدانات (٢٠٠٨) دراسة هدفت إلى تقصي أثر طريقة المعادلة باستخدام جذع مشترك وعدد فقراته وحجم العينة في القيم المعادلة والخطأ في المعادلة بين صورتي اختبار في الفيزياء، وتكونت عينة الدراسة من ثلاث مجموعات من طلبة الثانوية العامة. أظهرت الدراسة في جزء من نتائجها وجود فروق ذات دلالة إحصائية عند مستوى الدلالة ((α= ٠,٠٥ في القيم المعادلة تعزى إلى طريقة المعادلة ولصالح طريقة معادلة العلامات الحقيقية، وطريقة معادلة العلامات المشاهدة .وعدم وجود فروق ذات دلالة إحصائية في متوسطات القيم المعادلة تعزى إلى عدد فقرات الجذع المشترك وعلى جميع طرق المعادلة المستخدمة باستثناء طريقة معادلة العلامات الحقيقية وطريقة معادلة العلامات المشاهدة .وعدم وجود فروق ذات دلالة إحصائية بين القيم المعادلة الناتجة من جميع طرق المعادلة المستخدمة تعزى إلى أحجام العينات المستخدمة.

من استعراض الدراسات السابقة يمكن استخلاص ما يأتي:

١- في ما يخص طرق المعادلة المنبثقة من النظرية الحديثة أشارت بعض الدراسات إلى تفوق طرق المعادلة التي تتبع النظرية الحديثة عند استخدام معيار الصدق التقاطعي مثل دراسة أيوب (١٩٩٤) والشريفين (٢٠٠٣). أما عند استخدام معيار الخطأ المعياري للمعادلة فقد تفوقت الطريقة الخطية ومن الدراسات التي أشارت إلى ذلك دراسة الشريفين (٢٠٠٣).

٢- وفيما يخص عدد فقرات الجذع المشترك أشارت دراسةُ يانغ وهويانغ (Yang & Houang,) 1996) إلى أن النتائج تميل لأن تكون أكثر دقة عند زيادة الفقرات الجذعية. وعلى النقيض

من ذلك فقد توصلت دراسات أخرى إلى عدم تأثير عدد فقرات اختبار الجذع المشترك في دقة المعادلة مثل دراسة جيالوا وكريتشون وفالي (Gialluca, Crichton and Vale, 1984).

٣- حول فعالية طرق المعادلة أشارت دراسة كولن ووتني (Kolen & Whitney, 1982) ودراسة يانغ وهويانغ (Yang & Houang, 1996) إلى عدم وجود طريقة مثلى للمعادلة فيما أشارت دراسة المدانات (٢٠٠٨) إلى أفضلية الطرق المعتمدة على نماذج النظرية الحديثة في القياس.

ونظراً لعدم وجود اتفاق بين الدراسات السابقة، فإن إجراء المزيد من الدراسات وعلى عينات مختلفة وأحجام مختلفة لعدد فقرات الجذع المشترك يمكن أن يضيف بينات لها دلالاتها سعياً وراء الوقوف عند طبيعة هذه العلاقة، وعليه فإن هذه الدراسة تعد إسهاماً في هذا المجال.

مشكلة الدراسة

هدفت هذه الدراسة إلى مقارنة فاعلية طريقة معادلة العلامات الحقيقية وطريقة معادلة العلامات المشاهدة في معادلة الاختبارات عند استخدام جذع مشترك بين صورتي اختبار في الفيزياء للمرحلة الثانوية عند استخدام التصميم القائم على اختبار مشترك ومجموعات غير عشوائية.

فرضيات الدراسة

تحاول هذه الدراسة اختبار الفرضيات الإحصائية الآتية:

١. لا توجد فروق ذات دلالة إحصائية عند مستوى الدلالة (α= ٠,٠٥) لطريقة المعادلة (طريقة معادلة العلامات المشاهدة وطريقة معادلة العلامات الحقيقية) في قيم العلامات الخام المعادلة بين صورتين لاختبار في الفيزياء.

٢. لا توجد فروق ذات دلالة إحصائية عند مستوى الدلالة (α= ٠,٠٥) لعدد فقرات الجذع المشترك في قيم العلامات الخام المعادلة بين صورتين لاختبار في الفيزياء عند استخدام طريقة معادلة العلامات المشاهدة وطريقة معادلة العلامات الحقيقية.

أهمية الدراسة

يقوم معدو ومطورو الاختبارات عادة بإجراء اختبارات يبنى عليها الكثير من القرارات دون الاهتمام بموضوع معادلة علامات الاختبارات على أهميته، ومما يزيد من خطورة المشكلة أن

نتائج الطلبة قد يترتب عليها قرارات تتعلق بالانتقال إلى مراحل دراسية أعلى أو الالتحاق بالجامعات أو بسوق العمل، أو الحصول على رتبة أعلى خاصة وأن بعض الاختبارات يتم إجراؤها في أوقات مختلفة وبصور مختلفة؛ فمثلا تجرى اختبارات الرخصة الدولية لقيادة الحاسوب (ICDL) على مدى سنوات عديدة فإذا أعطيت نفس الأسئلة في كل عام فقد يحصل المفحوصون على بعض نماذج الأسئلة من المفحوصين الذين تقدموا للاختبار في وقت سابق. أو أن احد المفحوصين قد يتعرض لنفس أسئلة الاختبار، في هذه الحالات قد يصبح الاختبار مقياسا لمدى تذكر الإجابات، وليس مقياسا للمفهوم المفترض قياسه. كما أن النماذج المقدمة في أوقات مختلفة قد تختلف في مستوى الصعوبة. وتعالج هذه المشكلات بإجراء إحصائي لتعديل علامات الاختبارات كي يكون بالإمكان استخدام العلامات على النماذج بشكل متبادل إذ يتم تعديل الفروق في صعوبة الصور المقدمة، والتي يتم بناؤها على اعتبار أنها تحوي المحتوى ذاته ونفس مستوى الصعوبة، ويطلق على هذه العملية معادلة الاختبارات.

محددات الدراسة

١. محددات في العينة: تقتصر العينة على طلبة المرحلة الثانوية- الفرع العلمي.

٢. محددات في أداة البحث: تقتصر هذه الدراسة على اختبار في مجال الفيزياء فقراته من نوع الاختيار من متعدد يتم إعداده لأغراض هذه الدراسة.

مصطلحات الدراسة

١. **معادلة الاختبارات (Test Equating):** هي تحويل نظام وحدات القياس الخاص بإحدى صورتي الاختبار إلى نظام وحدات القياس الخاص بالصورة الأخرى بحيث تصبح درجات كل من الصورتين متكافئة في قياس مستوى القدرة لنفس الأفراد (المدانات، ٢٠٠٨).

٢. **القيم المعادلة (Equated Values):** هي القيم الناتجة على إحدى صورتي الاختبار المناظرة والمكافئة لقيم معينة في الصورة الأخرى (المدانات، ٢٠٠٨).

٣. **الجذع المشترك (Anchor Items):** ويمثل الفقرات المشتركة بين صورتي الاختبار عند تجريبهما لأغراض معادلتهما (المدانات، ٢٠٠٨).

٤. **الفقرة ثنائية التدريج (Dichotomous Items):** هي الفقرة التي تنقسم فيها طريقة الاستجابة إلى طريقتين حيث يمنح الفرد عليها العلامة (١) عند استجابته عليها استجابة صحيحة والعلامة (صفرا) عند استجابته عليها استجابة خاطئة (الشريفين، ٢٠٠٣).

منهجية الدراسة وإجراءاتها

المجتمع والعينة

تكوَّن مجتمع الدراسة من طلاب وطالبات المرحلة الثانوية في المدارس الحكومية والخاصة التابعة لمديريات التربية والتعليم في محافظة الكرك وهي: قصبة الكرك، ولواء المزار الجنوبي، ولواء القصر، ولواء الأغوار الجنوبية، وبلغ عدد أفراد مجتمع الدراسة (١٢٣٢) طالباً وطالبة في العام الدراسي (٢٠٠٨/٢٠٠٩) حسب التقارير الإحصائية لمديريات التربية والتعليم في المناطق المذكورة وكما يتضح في الجدول رقم (١).

الجدول رقم (١)
توزيع أفراد مجتمع الدراسة حسب المديرية

المديرية	عدد الطلبة
الكرك	٤٩٢
المزار	٣٧٦
القصر	٢٨٣
الأغوار الجنوبية	٨١
المجموع	١٢٣٢

أما عينة الدراسة فتكونت من مجموعتين من طلبة الثانوية العامة /الفرع العلمي إذ اختيرت المجموعة الأولى في الفصل الدراسي الأول وتكونت من (١٠٢) طالب، كما تم اختيار المجموعة الثانية في الفصل الدراسي الثاني وتكونت من (٩٧) طالباً وذلك حتى تنطبق عليها مع المجموعة الأولى مواصفات تصميم المجموعات غير العشوائية، بالإضافة إلى ذلك تم اختيار عينة التجريب الأولي للتحقق من خصائص الفقرات وتكونت من (٣٧) طالباً، ويوضح الجدول رقم (٢) توزع أفراد العينة على المديريات والفصل الدراسي.

الجدول رقم (٢)
توزيع أفراد عينة الدراسة حسب المديرية والفصل الدراسي

المديرية	الفصل الدراسي		المجموع
	الأول	الثاني	
الكرك	٣٣	٣١	٦٤
المزار	٢٥	٢٥	٥٠
القصر	٢٧	٢٥	٥٢
الأغوار الجنوبية	١٧	١٦	٣٣
المجموع	١٠٢	٩٧	١٩٩

أداة الدراسة

استخدمَ في هذه الدراسة صورتان متكافئتان لاختبار في مبحث الفيزياء في موضوع الكهرباء الساكنة تمثّل كل منهما بعشرين فقرة من نوع الاختيار من متعدد، واختبار جذع مشترك (Anchor Test) تمثّل بعشرفقرات من نوع الاختيار من متعدد، وقد تم بناء فقرات المقياس في شكله النهائي وفق الإجراءات الآتية (عودة، ٢٠٠١؛ الكيلاني وعدس، ١٩٩٣):-

١. تحديد الغرض من الاختبار وهو قياس تحصيل طلبة المرحلة الثانوية في مبحث الفيزياء في موضوع الكهرباء الساكنة.

٢. تحليل محتوى المادة العلمية ثم صياغة الأهداف التدريسية وعرضها على مشرفين تربويين لمبحث الفيزياء، ومعلمي مدارس ذوي خبرة في تدريس مبحث الفيزياء.

٥. إعداد جدول مواصفات يربط عناصر المحتوى ومستويات الهدف كسلوك عقلي معرفي.

٦. تكوين تجمع فقرات اختيار من متعدد وتنقيحها أكثر من مرة، ووضع الإجابة النموذجية بالتعاون مع مدرسين لمبحث الفيزياء. وبعد ذلك عُرضتُ هذه الفقرات على محكمين متخصصين في مجال القياس والتقويم ومشرفين تربويين لتحكيمها بشكل أولي من حيث السلامة اللغوية، ومدى مناسبتها للمقياس، وسلامتها من الناحية العلمية والتطبيقية وقد تم الأخذ بملاحظاتهم وأدخلت التعديلات المناسبة.

٧. وبعد اختيار الشكل الأولي للاختبار وصياغة صورتيه وتعديل عباراتهما وفقاً لما جاء في آراء المحكمين، تمت طباعته وتطبيقه على عينة استطلاعية تألّفت من (٣٧) طالباً من طلبة المرحلة الثانوية، بغرض التحقُّق من وضوح التعليمات ووضوح الصياغة اللغوية والتأكد من فعالية البدائل.

٨. بعد تصحيح إجابات الطلبة على الاختبار التجريبي، وتحليل نتائج تطبيق الفقرات في مرحلة التجريب الأولي باستخدام برنامج الحزمة الإحصائية للدراسات الاجتماعية (SPSS) واستخراج معاملات صعوبة الفقرات وتمييزها واستخراج معاملات الثبات للاختبار ككل، أجريت التعديلات على الصورة الأولية وفق المعايير الآتية:

أ-حذف الفقرات التي أجاب عنها جميع الطلبة أو كانت نسبة الإجابة عنها عالية جداً كما حذفت الفقرات التي لم يجب عنها أحد أو كانت نسبة الإجابة منخفضة جداً.

ب-حذف الفقرات التي كان معامل تمييزها بين (صفر - ٠,٢٠) كما تم حذف الفقرات التي كان معامل تمييزها سالباً.

ج- تمَّ تعديل صياغة بعض الفقرات التي كان معامل تمييزها من (٠,٢٠ ـ ٠,٣٠).

وبهذه المعايير تم حذف عدد من الفقرات، وبذلك تكون الاختبار من صورتين (Y & X) كل منهما تتكون من (٢٠) فقرة، بالإضافة إلى (١٠) فقرات مشتركة بين الصورتين (جذع مشترك) إذ أشارت بعض الدراسات إلى أن العدد القليل من الفقرات المشتركة التي تشكل الجذع المشترك تنجز غالبا نفس ما ينجزه العدد الأكبر من الفقرات (Raju, Edwards & Osberg, 1983) وقد جرى إدخال فقرات اختبار الجذع المشترك في كل من صورتي الاختبار بترتيب محدد (٣،٦،٩،١٢،١٥،١٨،٢١،٢٤،٢٧،٣٠).

معاملات الصعوبة والتمييز

تم إيجاد معاملات الصعوبة والتمييز لفقرات الصورة الأولى والثانية للاختبار وذلك بإيجاد نسبة الإجابة الصحيحة للفقرة للتعبير عن صعوبتها، ومعامل الارتباط الثنائي النقطي (rPbis) بين علامة الفقرة والعلامة الكلية على الاختبار للتعبير عن تمييز الفقرة. ويبين الجدول رقم (٣) قيم الصعوبة ومعاملات التمييز لفقرات الاختبار لعينة الدراسة الاستطلاعية حسب أداء كل مجموعة.

الجدول رقم (٣)
معاملات الصعوبة والتمييز لفقرات الصورة الأولى والثانية من الاختبار

الصورة الأولى (X)				الصورة الثانية (Y)			
رقم الفقرة	رمز الفقرة	الصعوبة	التمييز	رقم الفقرة	رمز الفقرة	الصعوبة	التمييز
١	X_1	.٦٤	.٤٥	١	Y_1	.٦٢	.٤٧
٢	X_2	.٦٥	.٢٠	٢	Y_2	.٥٥	.٥٥
٣×	X_3	.٥٤	.٢٨	٣×	Y_3	.٥٥	.٤٥
٤	X_4	.٧٠	.٤٢	٤	Y_4	.٥٨	.٣٨
٥	X_5	.٦٦	.٢٧	٥	Y_5	.٥٣	.٤٠
٦×	X_6	.٦٩	.٣٦	٦×	Y_6	.٦٦	.٣٧
٧	X_7	.٥٣	.٤٧	٧	Y_7	.٤٨	.٥٦
٨	X_8	.٧٢	.٤٠	٨	Y_8	.٦٧	.٣٠
٩×	X_9	.٧١	.٣٦	٩×	Y_9	.٤٧	.٤٨
١٠	X_{10}	.٥٨	.٣٣	١٠	Y_{10}	.٥٦	.٤٩
١١	X_{11}	.٧١	.٤٣	١١	Y_{11}	.٨٠	.٣٢
١٢×	X_{12}	.٢٦	.٤٤	١٢×	Y_{12}	.٥٩	.٢٤
١٣	X_{13}	.٥٩	.٤٤	١٣	Y_{13}	.٣٣	.٣٢
١٤	X_{14}	.٥٠	.٣٥	١٤	Y_{14}	.٤٩	.٣٨
١٥×	X_{15}	.٦٠	.٥٦	١٥×	Y_{15}	.٦٥	.٣٦
١٦	X_{16}	.٦٤	.٤٤	١٦	Y_{16}	.٥٤	.٤٠

تابع الجدول رقم (٣)

(X) الصورة الأولى				(Y) الصورة الثانية			
رقم الفقرة	رمز الفقرة	الصعوبة	التمييز	رقم الفقرة	رمز الفقرة	الصعوبة	التمييز
١٧	X١٧	.٦٠	.٤٨	١٧	Y١٧	.٧٤	.٣٣
١٨×	X١٨	.٥٥	.٥٤	١٨×	Y١٨	.٥٤	.٣٢
١٩	X١٩	.٥٤	.٣٠	١٩	Y١٩	.٦٨	.٤٣
٢٠	X٢٠	.٦٠	.٤٣	٢٠	Y٢٠	.٥٤	.٥٣
٢١×	X٢١	.٥٣	.٣٤	٢١×	Y٢١	.٦٧	.٤٨
٢٢	X٢٢	.٥٥	.٣٢	٢٢	Y٢٢	.٦٤	.٤٠
٢٣	X٢٣	.٥٤	.٣٨	٢٣	Y٢٣	.٧٣	.٣٢
٢٤×	X٢٤	.١٥	.٤٢	٢٤×	Y٢٤	.٤٤	.٣٨
٢٥	X٢٥	.٥١	.٣١	٢٥	Y٢٥	.٤٢	.٣٥
٢٦	X٢٦	.٤٣	.٣١	٢٦	Y٢٦	.٥٢	.٤١
٢٧×	X٢٧	.٥٩	.٣٩	٢٧×	Y٢٧	.٧٢	.٢٨
٢٨	X٢٨	.٥٣	.٤٢	٢٨	Y٢٨	.٥٩	.٣٣
٢٩	X٢٩	.٥٢	.٣٦	٢٩	Y٢٩	.٤١	.٥٧
٣٠×	X٣٠	.٤٠	.٥٣	٣٠×	Y٣٠	.٤٣	.٣٢

ملحوظة: الفقرات المميزة بالإشارة (×) هي فقرات جذع مشترك.

وحول التأكد من انطباق شروط المعادلة فإن صورتي الاختبار تحققان شروط إجراء المعادلة، فالاختباران يقيسان المهارات عينها وهما متقاربان في مستوى الثبات، وكذلك فإنّ المتوسط الحسابي والانحراف المعياري والتباين للأداء على الصورة الأولى للاختبار ومستوى الصعوبة ومستوى التمييز تماثل نظيراتها في الصورة الثانية حسبما أشار (Crocker & Algina, 1986; Lord, 1980)، ويبين الجدول رقم (٤) ذلك.

الجدول رقم (٤)

متوسطات معاملات الصعوبة والتمييز ومعاملات الثبات ومتوسط الأداء على كل من صورتي الاختبار واختبار الجذع المشترك

صورة الاختبار	عدد الفقرات	متوسط الصعوبة	متوسط التباين	متوسط مؤشرات التمييز	معامل الثبات	متوسط الأداء	الانحراف المعياري
الصورة الأولى (X)	٣٠	.٥٨	٠,٠١	٠,٣٩	.٨٥٠	١٧,٥٦	٦,٢٩
الصورة الثانية (Y)	٣٠	.٥٧	٠,٠١	٠,٣٩	.٨٤٠	١٧,٢٩	٦,١٢

يظهر الجدول السابق أن متوسط (الصعوبة والتباين ومؤشرات التمييز) لفقرات الصورة الأولى والثانية متقاربة بشكل كبير هذا بالإضافة إلى معامل الثبات ومتوسط أداء الطلبة على صورتي الاختبار.

صدق الاختبار

يعبر الصدق عن مدى تحقق الغرض الذي أعد لأجله الاختبار، ولما كان هدف الاختبار قياس تحصيل الطلبة فقد كان من الضرورة التحقق من أن الاختبار يقيس ما أعدّ لقياسه، وقد تم التحقق من ذلك بالطرق الآتية:

١. **صدق المحتوى** (Content Validity) حيث عرض الاختبار في كل مرحلة من مراحل إعداده كما ذكر سابقا في بند أداة الدراسة على مجموعة من المحكمين من ذوي الاختصاص في مجالات أساليب التدريس، والقياس والتقويم، وعدد من معلمي الفيزياء من ذوي الخبرة هذا بالإضافة إلى إعداد جدول المواصفات وتحليل محتوى المادة العلمية.

٢. **الصدق العاملي**: استخدم الصدق العاملي كمؤشر لصدق المقياس إذ تم إجراء التحليل بطريقة المكونات الرئيسة (Principle Component Analysis) لاستخلاص العوامل المسؤولة عن الأداء في كل صورة من صور الاختبار، وكون النظرية الحديثة تقوم على عدة افتراضات أهمها أحادية البعد فقد تم فحص البيانات للتأكد من تحقق هذا الشرط. ويظهر الجدول رقم (٥) نتائج التحليل العاملي من الدرجة الأولى لفقرات الصورة الأولى (X) وقيم الجذر الكامن ونسبة التباين المفسر، والنسبة التراكمية للتباين المفسر ونتائج التحليل العاملي من الدرجة الأولى لفقرات الصورة الثانية (Y) وقيم الجذر الكامن ونسبة التباين المفسر، والنسبة التراكمية للتباين المفسر.

الجدول رقم (٥)
نتائج التحليل العاملي من الدرجة الأولى لفقرات الصورة الأولى والثانية

الصورة الأولى (X)			
العامل	الجذر الكامن	نسبة التباين المفسر	التباين المفسر التراكمي
١	٦,٠١٩	٢٠,٠٦٢	٢٠,٠٦٢
٢	١,٨٢٩	٦,٠٩٨	٢٦,١٦٠
٣	١,٧٢٩	٥,٧٦٤	٣١,٩٢٤
٤	١,٥٦٩	٥,٢٣١	٣٧,١٥٥
٥	١,٥٣٢	٥,١٠٥	٤٢,٢٦٠
٦	١,٤٠٤	٤,٦٨٠	٤٦,٩٤٠
٧	١,٣١٥	٤,٣٨٥	٥١,٣٢٤
٨	١,١٩٩	٣,٩٩٨	٥٥,٣٢٢
٩	١,١٨٩	٣,٩٦٢	٥٩,٢٨٤
١٠	١,٠٧٥	٣,٥٨٢	٦٢,٨٦٦
١١	١,٠٠٥	٣,٣٥١	٦٦,٢١٦

تابع الجدول رقم (٥)

الصورة الأولى (X)			
العامل	الجذر الكامن	نسبة التباين المفسر	التباين المفسر التراكمي
الصورة الثانية (Y)			
١	٥,٨٢٦	١٩,٤٢١	١٩,٤٢١
٢	٢,٣٥٠	٧,٨٣٣	٢٧,٢٥٤
٣	٢,٠١٢	٦,٧٠٧	٣٣,٩٦١
٤	١,٧٨٠	٥,٩٣٢	٣٩,٨٩٣
٥	١,٦١٢	٥,٣٧٥	٤٥,٢٦٨
٦	١,٣٤٦	٤,٤٨٨	٤٩,٧٥٦
٧	١,٢٢٤	٤,٠٧٩	٥٣,٨٣٥
٨	١,١٩٢	٣,٩٧٣	٥٧,٨٠٨
٩	١,١٦١	٣,٨٧١	٦١,٦٨٠

بينت نتائج التحليل العاملي من الدرجة الأولى (First order) لفقرات الصورة الأولى- الجدول رقم (٥)- تشبع الفقرات على (١١) عاملا جذورها الكامنة واحد فأكثر، وفسرت ما مجموعه (٦٦,٢١٦٪) من تباين الأداء على الاختبار، كما يظهر الجدول تشبع فقرات الصورة الثانية من الاختبار على (٩) عوامل كانت قيمة جذورها الكامنة واحدا فأكثر وفسرت ما مجموعه (٦١,٦٨٠٪) من تباين الأداء على الاختبار.

كما تشير نتائج التحليل للصورتين إلى أن قيمة الجذر الكامن للعامل الأول مرتفعة ولبقية العوامل قليلة ومتقاربة مما يرجح وجود عامل سائد يفسر النسبة الكبرى من التباين ويمكن أن يستدل منه على أحادية البعد لأغراض تقدير المعالم؛ أي أن هناك تماثلاً نسبياً شبه استقرار في نسب التباين المفسرة لجميع العوامل باستثناء العامل الأول، وهذا يرجح تحقق أحادية البعد في بيانات هذا الاختبار (Warm, 1978).

ثبات الاختبار: جرى التأكد من توافر دلالات ثبات صورتي الاختبار ممثلاً بتقدير كرونباخ ألفا وهي كما ظهرت في الجدول رقم (٤).

الأساليب الإحصائية

استخدم البرنامج الإحصائي (SPSS) في حساب صعوبة الفقرات بالطريقة الكلاسيكية وحساب قيم الثبات ممثلة بمعاملات كرونباخ الفا وإجراء التحليل العاملي، واستخدمت البرمجية الإحصائية (PIE) لإيجاد القيم المعادلة الناتجة من طرق المعادلة التي تنبثق من النظرية الحديثة في القياس باستخدام النموذج ثلاثي البارامتر (Three-Parameter Logistic

Model) بافتراض أنّ قيمة الثابت (D=1). كما استخدمت البرمجية الإحصائية (-BILOG MG) لإيجاد معالم الفقرات (الصعوبة، والتمييز، والتخمين) ومعلم قدرة الأفراد.

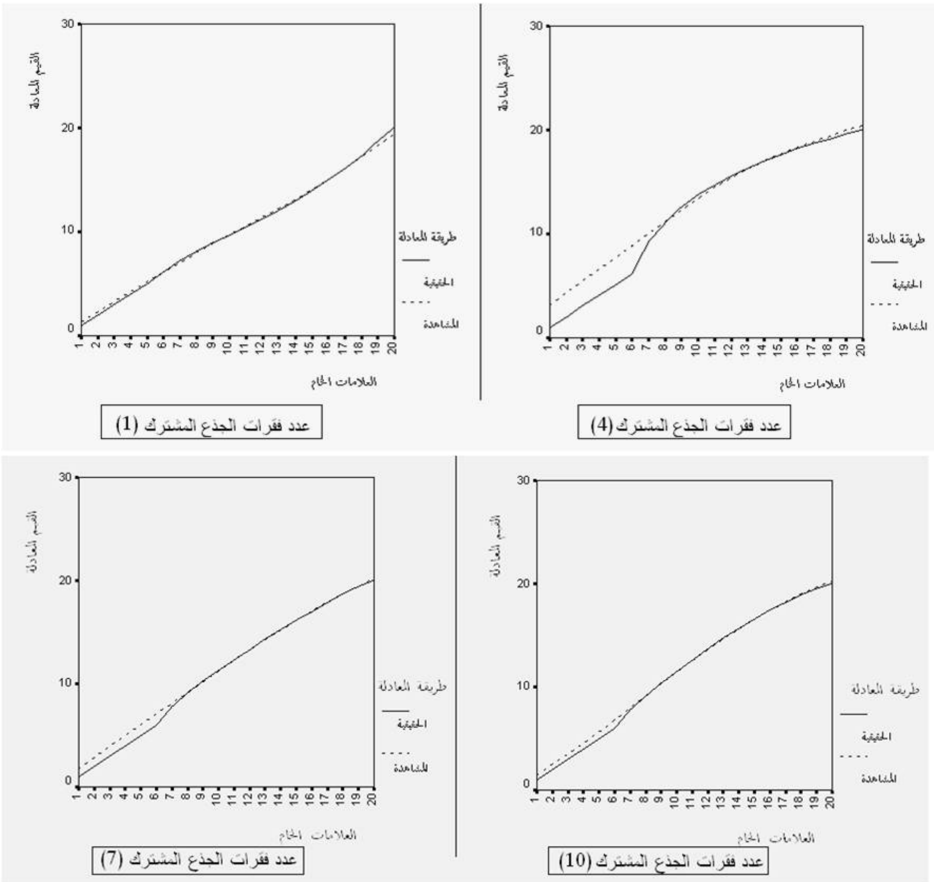
عرض النتائج ومناقشتها:

التحقق من صحة الفرضية الأولى

من أجل التحقق من صحة الفرضية الأولى التي تنص على أنه "لا توجد فروق ذات دلالة إحصائية عند مستوى الدلالة (α= ٠,٠٥) لطريقة المعادلة في قيم العلامات الخام المعادلة بين صورتين لاختبار في الفيزياء" فقد استخدمت البرمجية الإحصائية (PIE) لإيجاد القيم المعادلة الناتجة وفقاً لطريقة معادلة العلامات الحقيقية ومعادلة العلامات المشاهدة عند أعداد متغيرة من فقرات الجذع المشترك، وبعد ذلك استخدم الاختبار التائي لإيجاد دلالة الفروق ويظهر الجدول رقم (٦) ذلك، كما يظهر الشكل (٢) التمثيل البياني لهذه القيم.

الجدول رقم (٦)
قيم العلامات الخام المعادلة بين صورتي الاختبار وفقاً لطريقة معادلة العلامات المشاهدة وطريقة معادلة العلامات الحقيقية

العلامة الخام	الحقيقية الجذع١	المشاهدة الجذع١	الحقيقية الجذع٤	المشاهدة الجذع٤	الحقيقية الجذع٧	المشاهدة الجذع٧	الحقيقية الجذع١٠	المشاهدة الجذع١٠
١	١,٠١	١,٠٣	١,٣٩	١,٥٠	١,٠٠	١,٧٨	١,٠٠	٣,١٠
٢	٢,٠٣	٢,٠٥	٢,٣٥	٢,٥١	٢,٠١	٢,٨٨	٢,٠٠	٤,٣٦
٣	٣,٠٤	٣,٠٧	٣,٣٠	٣,٥٣	٣,٠١	٣,٩٥	٣,٠١	٥,٥٢
٤	٤,٠٥	٤,٠٦	٤,٢٦	٤,٥٧	٤,٠١	٥,٠٠	٤,٠٢	٦,٦١
٥	٥,٠٦	٥,١٢	٥,٢١	٥,٦٤	٥,٠٢	٦,٠٥	٥,٠٢	٧,٧٢
٦	٦,٠٩	٦,١٤	٦,١٥	٦,٧٦	٦,٠٢	٧,٠٩	٦,٠٣	٨,٨٦
٧	٧,٢٤	٩,٢٢	٧,٠٩	٧,٩٣	٧,٨٦	٨,١٢	٧,٨٧	١٠,٠١
٨	٨,١٥	١١,٠٧	٨,٠٠	٩,١٣	٩,١٧	٩,١٥	٩,١٥	١١,١٧
٩	٨,٩٤	١٢,٥٥	٨,٨٩	١٠,٣٠	١٠,٣٦	١٠,١٩	١٠,٢٧	١٢,٣٢
١٠	٩,٧٠	١٣,٧٥	٩,٧٥	١١,٤٤	١١,٥٢	١١,٢٣	١١,٣٣	١٣,٤٤
١١	١٠,٤٦	١٤,٧٤	١٠,٥٧	١٢,٥٥	١٢,٦٤	١٢,٢٨	١٢,٣٥	١٤,٤٩
١٢	١١,٢٥	١٥,٥٨	١١,٣٩	١٣,٦٥	١٣,٧١	١٣,٣١	١٣,٣٤	١٥,٤٤
١٣	١٢,٠٩	١٦,٣٣	١٢,٢٣	١٤,٧٠	١٤,٧٣	١٤,٢٨	١٤,٣٠	١٦,٢٧
١٤	١٢,٩٩	١٧,٠٠	١٣,١٢	١٥,٦٧	١٥,٦٩	١٥,٢٣	١٥,٢٤	١٧,٠٢
١٥	١٣,٩٦	١٧,٦١	١٤,٠٦	١٦,٥٨	١٦,٥٩	١٦,١٥	١٦,١٥	١٧,٧٠
١٦	١٤,٩٩	١٨,١٨	١٥,٠٥	١٧,٤٤	١٧,٤٤	١٧,٠٤	١٧,٠٣	١٨,٣٢
١٧	١٦,١٢	١٨,٧٠	١٦,١٠	١٨,٢٧	١٨,٢٤	١٧,٩٠	١٧,٨٧	١٨,٩١
١٨	١٧,٣٥	١٩,١٧	١٧,٢٠	١٩,٠٣	١٨,٩٦	١٨,٧٣	١٨,٦٧	١٩,٤٣
١٩	١٨,٧٢	١٩,٦١	١٨,٣٣	١٩,٧١	١٩,٥٧	١٩,٤٨	١٩,٤	٢٠,٠٣
٢٠	٢٠,٠٠	٢٠,٠٠	١٩,٤٥	٢٠,٢٧	٢٠,٠٠	٢٠,٢٦	٢٠,٠٠	٢٠,٣٩



الشكل (٢)

التمثيل البياني لقيم العلامات الخام المعادلة بين صورتي الاختبار وفقاً لطريقة معادلة العلامات الحقيقية وطريقة معادلة العلامات المشاهدة

يلاحظ من الجدول رقم (٦) والشكل رقم (٢) وجود عدم تماثل وفروق في القيم المعادلة لدى استخدام كل من طريقتي المعادلة المستخدمة واختلاف هذه القيم عند استخدام أطوال مختلفة للجذع المشترك، وبشكل محدد عند العلامات التي هي أقل من العلامة الخام (٩). وللحكم على معنوية الفروق بين متوسطات القيم المعادلة الناتجة من طريقتي المعادلة المستخدمة وعند عدد متغير من فقرات الجذع المشترك تم استخدام الاختبار التائي للعينات المرتبطة (T-test Samples-paired) إذ أكد لي وتام وتومبكنز (Li, Tam & Tompkins, 2004) على ضرورة استخدام الاختبار التائي للعينات المرتبطة بسبب أن أخذ القياسات في مرتي التطبيق على نفس عينة الأسئلة. ويوضح الجدول رقم (٧) متوسطات القيم المعادلة

ودلالة الفروق بين متوسطات قيم المعادلة .

الجدول رقم (٧)
متوسطات القيم المعادلة ودلالة الفروق بينها

عدد فقرات الجذع المشترك	الطريقة	المتوسط الحسابي	الانحراف المعياري	قيمة ت المحسوبة	الدلالة الإحصائية
١	معادلة العلامات الحقيقية	١٠,١٦	٥,٦٣	-.٦٥٤	٠,٥٢١
	معادلة العلامات المشاهدة	١٠,١٩	٥,٤٦		
٤	معادلة العلامات الحقيقية	١٢,٢٥	٦,٥١	-٣,١٧٣	٠,٠٠٥
	معادلة العلامات المشاهدة	١٣,٠٥	٥,٥٨		
٧	معادلة العلامات الحقيقية	١١,٢٠	٦,١٧	-٣,٠٤١	٠,٠٠٧
	معادلة العلامات المشاهدة	١١,٥٠	٥,٨٣		
١٠	معادلة العلامات الحقيقية	١١,٣٧	٦,٢٩	-٢,٩٠٥	٠,٠٠٩
	معادلة العلامات المشاهدة	١١,٥٥	٦,١١		

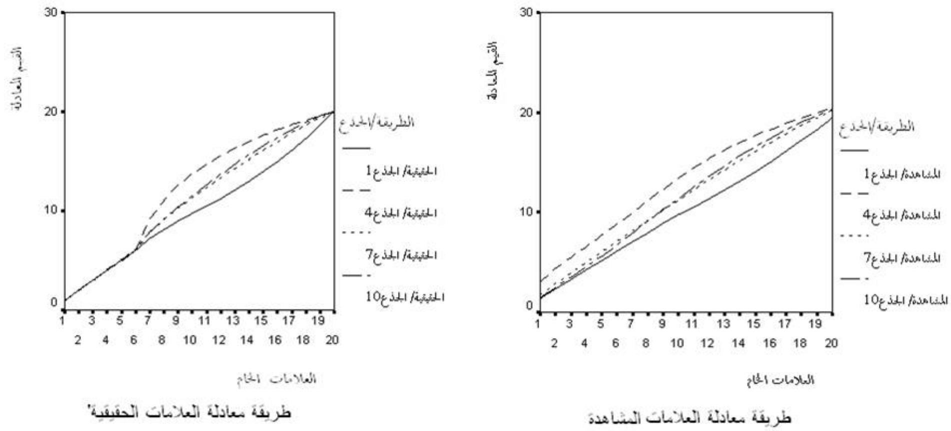
وبشكل تفصيلي أظهرت نتائج الاختبار التائي لدلالة الفروق بين متوسطات قيم العلامات المعادلة الموضحة في الجدول رقم (٧) أن هناك فروقا دالة إحصائيا عند مستوى الدلالة (α= ٠,٠٥) في متوسطات القيم المعادلة بطريقة المعادلة بالعلامات الحقيقية وطريقة المعادلة بالعلامات المشاهدة عندما كانت عدد فقرات الجذع المشترك (٤، ٧، ١٠) لصالح طريقة معادلة العلامات المشاهدة. ولكن الفروق لم تكن ذات دلالة عندما كان الجذع المشترك فقرة واحدة. وتتفق هذه النتيجة مع دراسة (Thorndike, 1982) والشريفين (٢٠٠٣)، وأيوب (١٩٩٤)، وأبي لبدة (١٩٩٣)، وبترسون وكوك وستوكنغ (Peterson, Cook and Stocking, 1983).

التحقق من صحة الفرضية الثانية

ومن أجل التحقق من صحة الفرضية الثانية التي تنص على أنه: "لا توجد فروق ذات دلالة إحصائية عند مستوى الدلالة (α= ٠,٠٥) لعدد فقرات الجذع المشترك في قيم العلامات الخام المعادلة بين صورتين لاختبار في الفيزياء عند استخدام طريقة معادلة العلامات المشاهدة وطريقة معادلة العلامات الحقيقية" فقد استخدمت البرمجية الإحصائية (PIE) لإيجاد القيم المعادلة الناتجة عند أعداد متغيرة من فقرات الجذع المشترك، ولفحص أثر عدد فقرات الجذع المشترك في القيم المعادلة استخدم تحليل التباين الأحادي لمعرفة ما إذا كانت هناك فروق ذات دلالة إحصائية بين متوسطات القيم المعادلة الناتجة من طرق المعادلة المستخدمة، ويبين الجدول رقم (٨) نتائج هذا التحليل.

الجدول رقم (٨)
نتائج تحليل التباين الأحادي لقيم العلامات الخام المعادلة وفقاً لطريقتي المعادلة عند أعداد مختلفة لفقرات الجذع المشترك

طريقة المعادلة	مصدر التباين	مجموع المربعات	درجة الحرية	متوسط المربعات	قيمة ف المحسوبة	الدلالة الإحصائية
معادلة العلامات الحقيقية	بين المجموعات	٤٤,٠٨٠	٣	١٤,٦٩٣	٠,٣٨٧	٠,٧٦٣
	داخل المجموعات	٢٨٨٦,٣٧١	٧٦	٣٧,٩٧٩		
	المجموع	٢٩٣٠,٤٥١	٧٩			
معادلة العلامات المشاهدة	بين المجموعات	٨٢,٠٨٠	٣	٢٧,٣٦٠	٠,٨٢٦	٠,٤٨٤
	داخل المجموعات	٢٥١٨,٥٠٥	٧٦	٣٣,١٣٨		
	المجموع	٢٦٠٠,٥٨٥	٧٩			



الشكل (٣)
التمثيل البياني لقيم العلامات الخام المعادلة بين صورتي الاختبار وفقاً لطريقة معادلة العلامات الحقيقية وطريقة معادلة العلامات المشاهدة عند أعداد مختلفة للجذع المشترك

يتضح من البيانات الواردة في الجدول رقم (٨) وفيما يتعلق بطريقة معادلات العلامات الحقيقية أن قيمة (ف) كانت (٠,٣٨٧) وهي غير دالة إحصائيا عند (α= ٠,٠٥)؛ أي أن الفروق بين علامات الطلبة المعادلة باستخدام أعداد مختلفة لفقرات الجذع المشترك هي فروق غير جوهرية، وليس هناك أثر لعدد فقرات الجذع المشترك. أما فيما يتعلق بطريقة معادلات العلامات المشاهدة فقد كانت قيمة (ف) (٠,٨٢٦) وهي غير دالة إحصائيا عند (α= ٠,٠٥)، وتتفق هذه النتائج مع دراسات أخرى أشارت إلى عدم تأثير عدد فقرات اختبار الجذع المشترك في دقة المعادلة أي أن العدد القليل من الفقرات المشتركة التي تشكل الجذع المشترك تنجز غالبا

نفس ما ينجزه العدد الأكبر من الفقرات مثل دراسة جيالوا وكريتشون وفالي (Gialluca, Crichton & Vale, 1984)، وعلى النقيض من ذلك ما توصلت إليه دراسات أخرى مثل دراسة يانغ وهويانغ (Yang & Houang, 1996)، ودراسة فيتزباتريك ووندي (Fitzpatrick & Wendy, 2001) ودراسة المدانات (٢٠٠٨) إذ أشارت هذه الدراسات إلى أن نتائج المعادلة تميل لأن تكون أكثر دقة عند زيادة عدد الفقرات الجذعية، كما تناقضت هذه الدراسة مع الشروط العامة لاختيار الفقرات الجذعية (Kolen & Brennan, 2004) التي تشير إلى أن عددها يجب أن يكون (٢٠٪) من الاختبار أو (٢٠) فقرة أيهما أكبر.

الاستنتاج والتوصيات

بالاعتماد على ما توصلت إليه الدراسة فإنه يوصى باستخدام طريقة العلامات المشاهدة عند إجراء المعادلة بين اختبارين تم تقديمهما إلى مجموعات غير متكافئة من الطلبة باستخدام التصميم القائم على الجذع المشترك دون أن يكون لعدد فقرات الجذع المشترك أهمية كبيرة ما دامت هذه الفقرات تمثل المحتوى جيداً، وذات خصائص سيكلومترية مناسبة.

المراجع

أبو لبدة، خطاب محمد أحمد (١٩٩٣). **بناء اختبار متعدد المستويات للأداء العقلي للأطفال الأردنيين من سن (٦-١٢)**. رسالة دكتوراه غير منشورة، الجامعة الأردنية: عمان، الأردن.

أيوب، حسين محمد عبد القادر (١٩٩٤). **المقارنة بين أربع طرق للمعادلة عندما يكون التصميم من مجموعات متكافئة وغير متكافئة**. رسالة دكتوراه غير منشورة، الجامعة الأردنية: عمان.

الدوسري، راشد حماد (٢٠٠٤). **القياس والتقويم التربوي الحديث، مبادئ وتطبيقات وقضايا معاصرة**، (ط١). عمان: دار الفكر للنشر والتوزيع.

الشريفين، نضال أحمد (٢٠٠٣) **مدى تحقق معايير الفاعلية في معادلة اختبارين احدهما ثنائي التدريج والأخر متعدد التدريج وفق نماذج النظرية الكلاسيكية والنظرية الحديثة في القياس**، رسالة دكتوراه غير منشورة، جامعة عمان العربية: عمان، الأردن.

عودة، أحمد (٢٠٠١). **القياس والتقويم في العملية التدريسيه**، (ط٥). اربد: دار الأمل للنشر والتوزيع.

الكيلاني، عبدالله زيد وعدس، عبد الرحمن (١٩٩٣). **القياس والتقويم في التعلم والتعليم**. القدس: منشورات جامعة القدس المفتوحة.

المدانات، رائد فايز.(٢٠٠٨). **أثر طريقة المعادلة باستخدام جذع مشترك وعدد فقراته على القيم المعادلة والخطأ في القيم المعادلة بين اختبارين في الفيزياء**. رسالة دكتوراه غير منشورة. جامعة عمان العربية: عمان، الأردن.

Angoff, W. H. (1971). **Scales, norms, and equivalent scores.** In R. L. Thorndike (Ed.), Educational measurement (2nd ed., pp. 508-600). Washington, DC: American Council on Education.

Crocker, L. & Algina, J. (1986). **Introduction to classical and modern test theory.** New York: Holt, Rinehart and Winston Inc.

Fitzpatrick, A. & Wendy, M. (2001). **The effect of test length and sample size on the reliability and equating of tests composed of constructed-response items,** Monterey, California. CTB/McGraw-Hill.

Gialluca, K., Crichton, L. & Yale, C. (1984) Methods for equating mental tests. Assessment Systems Corporation. 233 university Avenue, Suite 310. St. Paul, Minnesota 55114, **(Eric Document Reproduction Service No: ED251512)**.

Hambleton, R. K & Swaminathan, H. (1985). **Item response theory: principles and applications.** Boston: Kluwer.

Kolen, M . J . (1981) . Comparison of traditional and item response theory methods for equating tests . **Journal of Educational Measurement, 18,** 1-11 .

Kolen, M. J., & Brennan, R. L. (2004). **Test equating, scaling, and linking: Methods and practices** (2nd ed.). 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, 9**(4), 279-293.

Lord, F. M.(1980). **Applications of item response theory to practical testing problems.** Hillsdale, NJ: Lawrence Erlbaum Associates.

Li, Y. H. Tam, H. P. & Tompkins, L. J.(2004) A comparison of using the fixed common-precalibrated parameter method and the matched characteristic curve method for linking multiple-test items, **International Journal Of Testing, 4**(3), 267-293

Peterson, N. S., Cook, L. L. & Stocking, M. L. (1983). IRT versus conventional equating methods: A comparative study of scale stability. **Journal of Educational Statistics, 8**(2), 137-156 .

Raju, N. S., Edwards, J. E., & Osberg, D. W.(1983). **The effect of anchor test size in vertical equating with Rasch and three parameter models.** Paper presented at the annual meeting of the national council on measurement in education, Montreal.

Stocking, M. L. & Lord, F. M. (1983). Developing a common metric in item response theory. **Applied Psychological Measurement, 7**(2), pp 201-210.

Thorndike, R. L. (1982). **Educational measurement: Theory and practice.** In D. Spearritt (Ed.), The improvement of measurement in education and psychology: Contributions of latent trait theory (pp. 3-13). Princeton, NJ: ERIC Clearinghouse of Tests, Measurements, and Evaluations.(ED 222 545).

Tianqi, H. (1977). A comparison among IRT true and observed score equating and traditional equipercentile equating. **Applied Measurement Education, 10,** 105–121.

Yang, W. l. & Houang, R. T.(1996). **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.** Paper presented at the annual meeting of the American Educational Research Association. (New York, NY, April 8-12, 1996).

Warm, Thomas A.(1978). **A prime of item response theory. U.S.** Coast Guard Institutem, Oklahoma.