# معادلة درجات الاختبارات متعددة الاستجابة في تصميم الفقرات المشتركة باستخدام طرق النظرية التقليدية في القياس (دراسة تجريبية)

*يوسف عبد العاطي المحروق**

## ملخص

هدفت هذه الدراسة إلى معادلة درجات الاختبارات متعددة الحدود في تصميم الفقرات المشتركة باستخدام طرق النظرية التقليدية في القياس، من خلال دراسة المتغيرات:حجم العينة، طول الاختبار، مستويات الصعوبة، تم توليد بيانات تجريبية باستخدام برمجية (Wingen2) الذي يعتمد على استخدام نظرية استجابة الفقرة لتوليد استجابات المفحوصين على فقرات الاختبار، أشارت النتائج إلى أن حجم العينات الكبير يقلل من الخطأ المعياري للمعادلة، ويقلل من البواقي المعيارية.كذلك بينت النتائج أن اختلاف النماذج في صعوبتها يزيد من الخطأ المعياري والبواقي المعيارية. كذلك أظهرت النتائج أن طول الاختبار يوثر في الخطأ المعياري؛ فالاختبار الطويل يعطي قيماً كبيرة للخطأ المعياري، والبواقي المعيارية.

**الكلمات الدالة:** معادلة الاختبارات، تصميم الفقرات المشتركة، الاختبارات متعددة الحدود.

**المقدمة**

اهتم علماء النفس ومنذ ظهور حركة القياس النفسى بتحقيق صدق وثبات الاختبارات والمقاييس النفسية، سعياً منهم لتحقيق أعلى درجة من الموضوعية فى هذه الأدوات، عند استخدامها فى عملية القياس، لذلك فقد سيطرت نظرية القياس النفسي التربوي الكلاسيكية على الفكر التربوي لفترة طويلة من الوقت، وظل العاملون بالقياس النفسي يستخدمون مبادئ وأسس هذه النظرية في بناء الاختبارات والمقاييس بأشكالها المختلفة وتفسير الدرجات المتحققة عليها (Hambelton &linden, 1982)، ووفقاً لهذه النظرية يمكن التعبير عن قدرة الفرد من خلال الدرجة الحقيقية والتى تتضح من خلال أدائه على الاختبار، وبناءً عليه فإنه سيتغير وضع قدرة الفرد حسب تغير مستوى الاختبار، فالاختبار وفقراته تتغير خصائصها بتغير خصائص الأفراد، كما أن خصائص الأفراد تتغير بتغير خصائص الاختبار من حيث السهولة والصعوبة.

لذلك تستخدم علامة الاختبار غالباً كمصدر للمعلومات من أجل اتخاذ قرارات مهمة؛ فبعض هذه القرارات يكون على مستوى الفرد مثل عندما يقرر الطالب أي الكليات سوف يلتحق بها وما المبحث الذي سوف يدرسه، وبعض القرارات الأخرى يكون على المستوى المؤسسي، كأن تقرر مؤسسة ما علامة القبول المطلوبة لتأهيل الأفراد لمهنة ما أو لقبول الطلاب في الكلية أو الالتحاق بوظيفة، وبعضها الآخر يكون على مستوى الدولة ورسم سياسة عامة مثل عندما يتم تقييم المتغيرات في النظام التعليمي من أجل تحسينه (مدانات، 2008).

إن اتخاذ مثل هذه القرارات يتطلب أن تكون العلامات المستوحاة من نتائج الاختبار دقيقة، وهذا يتطلب إدارة الامتحانات بشكل صحيح، خاصة وأن بعض الاختبارات يتم إجراؤها في أوقات مختلفة وبصورة مختلفة، فمثلاً قد تجرى اختبارات القبول 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 أحد المفحوصين قد يتعرض لنفس أسئلة الاختبار، في هذه الحالات قد يصبح الاختبار مقياساً لمدى تذكر الإجابات، وليس مقياساً للمفهوم المفترض قياسه، كما أن النماذج المقدمة في أوقات مختلفة قد تختلف في مستوى صعوبتها، وتعالج هذه المشكلات بإجراء إحصائي لتعديل علامات الاختبارات كي يكون بالإمكان استخدام العلامات على النماذج بشكل متبادل حيث يتم تعديل الفروق في صعوبة الصور المقدمة، والتي يتم بناؤها على اعتبار أنها تحتوي على نفس المحتوى ونفس مستوى الصعوبة، ويطلق على هذه العملية معادلة الاختبارات.

* إدارة المناهج، وزارة التربية والتعليم، مملكة البحرين. تاريخ استلام البحث 2014/11/09، وتاريخ قبوله 2014/12/22.


2017© عمادة البحث العلمي وضمان الجودة/ الجامعة الأردنية. جميع الحقوق محفوظة.

إن نظرية القياس التقليدية بدأت من فكرة أن درجة الفرد على الاختبار تنتج من تأثير مجموعتين من العوامل هما: عوامل تؤدى إلى الاتساق، وهى عوامل مرغوبة وتتكون من صفات مستقرة لدى الفرد يتم قياسها فى الاختبار، والثانية عوامل تؤدى إلى عدم الاتساق أو الاختلاف وتتضمن مجموعة من العوامل غير مرغوب فيها تؤثر فى درجة الفرد على الاختبار وتسمى بعوامل الخطأ. وبغرض تفسير الدرجة من حيث صحتها من عدمه تضع هذه النظرية مجموعة من الافتراضات حددها (علام، 2000) في:

– الدرجة الحقيقية للفرد يفترض أن تكون درجة أو قيمة ثابتة ذلك أنها تمثل قدرة الفرد المقاسة.

– الدرجة التى يحصل عليها الفرد ليس من الضرورى أن تمثل درجته الحقيقية ولذا فالدرجة التى يحصل عليها قابلة للتغيير حسب الظروف الاختبارية.

– الدرجة التى يحصل عليها الفرد هى نتاج الدرجات الحقيقية ودرجات الخطأ.

– الدرجة الخاطئة ليست محددة وثابتة فى كل المواقف والظروف بل تتغير بتغير هذه المواقف والظروف.

– عدم وجود علاقة دالة بين الدرجات التى يحققها الأفراد وبين مقدار الخطأ في القياس.

– عدم وجود ارتباط بين الدرجات الخاطئة فى الاختبارات المختلفة وهذا يرجع إلى الاختلاف فى طبيعة الاختبارات.

– الدرجات الخاطئة قد تكون منتظمة وقد لا تكون منتظمة، أي أنها قد تتكرر بنفس الصورة وبنفس المستوى فى كل الحالات التى يتم بها تطبيق الاختبار.

– أن الدرجات التى يحصل عليها الفرد فى مفردات الاختبار يمكن جمعها كما لو كانت تمثل ميزاناً خطياً **Linear Scale**، وأن المفردات المتعلقة بالمتغير المراد قياسه تحمل المعنى نفسه لدى جميع المختبرين.

لقد وظفت النظرية الكلاسيكية في القياس في بناء 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، وهناك حالات يتعرض فيها المفحوصون لاختبارين صمما لقياس نفس المفهوم النفسي أو المحتوى، ولكن قلما نحصل على نموذجين متكافئين تماماً في المستوى وفي مدى الصعوبة لنفس الاختبار، لذا يصبح من الضروري أن نعادل هذه النماذج؛ بمعنى أن نقوم بتحويل العلامات في أحد النماذج إلى ما يناظرها أو ما يكافئها في النموذج الآخر، وتتطلب معظم البرامج الاختبارية أن يكون لدينا نماذج متكافئة من نفس الاختبار، ولإجراء المعادلة بين نموذجين (صورتين) لنفس الاختبار فإننا نطبق النموذجين على عينة من الأفراد – حسب التصميم المناسب الذي تم اختياره– ثم نجري التحليلات الإحصائية المناسبة التي تمكننا من تحويل العلامات من نموذج إلى آخر عندما يطبق على أفراد أحد النموذجين (Angoff.1971).

وبعد مجال معادلة الاختبارات من المجالات التطبيقية الهامة في القياس التربوي؛ ففي الكثير من البرامج الاختبارية تنشأ الحاجة إلى استخدام عدة صور من الاختبار الواحد لضمان السرية، كما أن الصور المتعددة ضرورية في اختبارات القبول وفي التقويم البنائي للطالب حيث يلزم تطبيق اختبارات دورية ومتعددة اثناء الفصل الدراسي ومقارنة العلامات التي يحصل عليها عبر الزمن (kolen, Brenann, 2004)، لذلك تقوم العديد من المؤسسات التربوية التي تعنى في قياس تحصيل الطلاب بإعداد نماذج متعددة للاختبارات؛ ضمانا لامن هذه الاختبارات، وذلك من خلال تضمين فقرات مختلفة في مواقع مختلفة في هذه الاختبارات لتصبح هذه الاختبارات تمتاز بالعدالة لدرجة مقبولة، بحيث أن العلامة على أي صورة يكون لها نفس الدلالة ونفس القيمة لقياسية لو تحققت نفسها على صورة أخرى، ويقال عندئذ أن هناك تكافؤا تاماً بين صور الاختبار الواحد، لكن نادراً ما يتحقق مثل هذا التكافؤ التام بين صور أعدت عن نفس الاختبار، وعندها يصبح من الضروري معادلة هذه الصور بتحويل نظام الوحدات في إحدى الصور إلى نظام الوحدات في الصورة الأخرى حتى تكون العلامات المشتقة من الصورتين بعد عملية التحويل متكافئة تماماً (مدانات، 2008).

وعندما يتم استخدام وتطبيق الاختبار بشكل متكرر، فإن فقرات الاختبار يمكن أن تصبح فقرات سهلة ومعروفة للمفحوصين المستقبليين؛ لذلك يتم إعداد نماذج متعددة للاختبارات في العديد من برامج الاختبارات لمنع كشف الاختبار. وعلى الرغم من وجود النماذج المتعددة للاختبارات والتي تم بناؤها بالاعتماد على نفس الخصائص (نفس المحتوى، ونفس مستوى الصعوبة)، إلا أن نماذج الاختبار لا تكون متكافئة بالضبط؛ ولهذا السبب فإن بعض المفحوصين الذين يأخذون الاختبار الأسهل سيكون لهم أفضلية على أولئك الذين يأخذون الاختبار الأصعب. لذلك إذا تمت المعادلة بشكل مثالي، فان المفحوصين سوف يحصلون على نفس الدرجات بغض النظر عن أي اختبار يتم التقدم له (Kolen, Brenann,2004) فالمعادلة هي إجراء إحصائي يتم فيه تعديل الاختلاف في مستوى الصعوبة من نموذج إلى آخر، ونتيجة

لهذا التعريف ظهر جدل واسع بين علماء القياس حول ما هو نوع الإجراء الإحصائي المميز أو الظروف التي ستكون مقنعة إذا تم إجراء المعادلة؟

تعد المعادلة إحدى الإجراءات المستخدمة لتحقيق عدالة الاختبار.فعندما يتم تطبيق اختبارات متعاقبة ومتسلسلة، فإن المعادلة تصبح ضرورية لوضع الدرجات في الاختبارات المختلفة على نفس المتصل.

عملياً، يوجد نوعان من أنواع المعادلة للاختبارات وهما: المعادلة الأفقية Vertical equating والمعادلة الرأسيةHorizontal equating. ففي المعادلة الأفقية تكون الصور المختلفة للاختبارات متقاربة في مستوى صعوبتها، والهدف من إجراء عملية المعادلة هو تعديل الفروق الناتجة عن الاختلاف في مستوى الصعوبة، بينما في المعادلة الرأسية تكون صور الاختبارات مختلفة في مستوى صعوبتها، كذلك تكون عينات الأفراد التي تطبق عليها هذه الصور مختلفة في مستوى قدرتها، ويعتبر لورد (Lord) صور الاختبار متكافئة عندما لا يشكل تطبيق أي اختبار منها أي فارق في أداء المفحوص، حيث إن الاختبارين (X، y) لا يمكن أن يكونا متكافئين إلا إذا كانت القدرة المستنتجة للفرد هي ذاتها، سواء تقدم الفرد للاختيار X أم للاختبار y (Lord, 1980, P171).

**مشكلة الدراسة وأسئلتها:**

تعتمد موضوعية وصدق نتائج الاختبارات على دقة الأساليب التى استخدمت فى بنائها واختيار فقراتها وتفسير نتائجها، وكذلك فى وصفها للقدرة التى يقيسها الاختبار، لذلك فإن الأمر يتطلب ضرورة استخدام اساليب حديثة فى القياس والتى أثبتت البحوث التجريبية أنها تحقق الدقة والموضوعية المنشودة فى القياس التربوي والنفسي، وهذا الامر دفع الباحث لاختيار النظرية الكلاسيكية في القياس للكشف عن مدى دقتها في معادلة درجات الاختبارات متعددة الاستجابة. وبعبارة أخرى فان الغرض الأساسي من الدراسة هو التحقق من دقة معادلة درجات الاختبارات ذات الجذع المشترك للمجموعات المتكافئة باستخدام النظرية الكلاسيكية في القياس. وبالتحديد فان هذه الدراسة حاولت الإجابة عن السؤالين الآتيين:

- ما أثر طول الاختبار، حجم العينة، وتشابه مستويات الصعوبة في الخطأ المعياري للمعادلة عند النقاط المختلفة على سلم الدرجات ؟
- ما أثر طول الاختبار، حجم العينة، وتشابه مستويات الصعوبة في البواقي المعيارية للمعادلة عند النقاط المختلفة على سلم الدرجات ؟

واختارت هذه الدراسة الطريقة المئينية باعتبارها أكثر الطرق شيوعاً في المعادلة؛ وذلك للكشف عن دقة معادلة الدرجات في الاختبارات مستخدمين تصميم المجموعات العشوائية ذات الفقرات المشتركة.

**أهمية الدراسة:**

تكتسب الدراسة أهميتها من خلال استخدامها الطريقة المئينية للكشف عن دقة معادلة درجات الاختبارات تحت ظروف: البيانات متعددة الاستجابة، الفقرات المشتركة، حجم العينة، طول الاختبار، وتجانس مستوى الصعوبة. ويمكن تلخيص أهمية الدراسة في الجوانب الآتية:

الأهمية النظرية

- الإسهام في إلقاء الضوء على فاعلية الطريقة المئينية في معادلة درجات الاختبارات متعددة الحدود والمستخدمة في هذه الدراسة، بحيث تسهم في إعطاء صورة واضحة للمتخصصين في مجال بناء الاختبارات لاختيار طريقة المعادلة التي تتواءم مع طبيعة الاختبار الذي يتم بناؤه للحصول على نتائج تمتاز بالشمولية والدقة.
- التحقق من النتائج المحتملة بالإجابة عن الأسئلة التي طرحتها الدراسة، أي كيف يمكن استخدام الطرق المختلفة لمعادلة درجات الاختبارات؟

الأهمية العملية

- يمكن أن تسهم هذه الدراسة في توضيح طرق جديدة لمصممي ومحللي الاختبارات في وزارات التربية والتعليم والجامعات العربية المهتمين بموضوع إيجاد صور متكافئة للاختبارات.
- تعريف المختصين ومحللي الاختبارات على بعض البرامج الحاسوبية التي تساعدهم في معادلة درجات الصور المختلفة للاختبارات من خلال توفير معلومات وإرشادات للمهتمين في بناء الاختبارات حول كيفية توظيف البرامج الحاسوبية المستخدمة في معادلة الدرجات وتطبيقها بكل يسر وسهولة وخصوصا برمجية (Equating Receips).

التعريفات الإجرائية:

معادلة درجات الاختبارات: هو إجراء إحصائي يتم فيه تحويل سلم الدرجات على أحد الاختبارات إلى سلم الدرجات على الاختبار الآخر، بحيث يمكن معرفة درجة الفرد على أحد الاختبارات إذا علمنا درجته على الاختبار الآخر.

دقة المعادلة: هو أسلوب إحصائي يستخدم للتأكد من مدى فاعلية المعادلة باستخدام اختبار الجذع المشترك.

البيانات متعددة الاستجابة: وهي فقرات اختباريه تكون

الاستجابة عليها متعددة (1-2-3-4-5) حيث يكلف المفحوص بتحديد الاستجابة التي يراها مناسبة.

اختبار الجذع المشترك: هو اختبار مكون من مجموعة من الفقرات المشتركة بين اختبارات الصفوف المختلفة.

**الدراسات السابقة:**

للوقوف على فاعلية النظرية الكلاسيكية في معادلة درجات الاختبارات متعددة الحدود، فقد أجريت العديد من الدراسات، ومنها دراسة لي وآخرون (Lee, Kolen, Frezby, 2001) التي هدفت من خلالها إلى المقارنة بين معادلة اختبارين أحداهما ثنائي التدريج والآخر متعدد التدريج وفق نماذج النظرية الحديثة والنظرية الكلاسيكية في المعادلة، حيث تم استخدام طرق المعادلة التقليدية الآتية لمعادلة الاختبار الثنائي التدريج وهي (الأوساط الخطية والمئينية)، بينما تم استخدام طرق المعادلة وفق النظرية الحديثة وهي طريقتي المعادلة التاليتين((GRM) و(NM)) في معادلة الاختبار المتعدد التدريج وتم استخدام محك الجذر ألتربيعي للأوساط غير الموزونة للحكم على فاعلية الطريقة المستخدمة في المعادلة. أظهرت النتائج تشابهاً بين طرق معادلة العلامات الحقيقية والملاحظة في الاختبار المتعدد التدريج وفق نماذج النظرية الحديثة، والطرق الكلاسيكية المستخدمة في المعادلة (الأوساط الخطية والمئينية) في الاختيار ثنائي التدريج للعلامات الحقيقية.

أما هاو(Hue, 2004) فقد أجرى دراسة إلى معرفة أثر اختلاف طرق اختيار الفقرات في اختبار الجذع المشترك على دقة معادلة الاختبار التكيفي. حيث أجريت الدراسة على طلبة جامعة أوهايو الأمريكية من ذوي الثقافات المختلفة كالصينية والكورية والاسبانية وهدفت إلى التعرف إلى أثر اختلاف طرق اختيار فقرات الجذع المشترك على دقة عملية معادلة الدرجات وذلك من خلال مقارنة القسم اللفظي للاختبار بالقسم غير اللفظي واستخدم الباحث طريقتين لاختيار فقرات الجذع المشترك هما طريقة المعالم للفقرة (الصعوبة للتميز) وطريقة دالة الفقرة التفاضلية.وهدفت أيضاً إلى التعرف فيما إذا كانت إحدى الطريقتين أفضل من غيرها في تحديد دقة المعادلة.دلت النتائج أن اختيار فقرات الجذع المشترك باستخدام طريقة معالم الفقرات كانت أفضل من حيث الدقة.

وفي دراسة (الصمادي، 2007) التي هدفت إلى التعرف إلى دقة المعادلة المئينية العمودية من خلال استخدام تصميم الجذع المشترك للمجموعات غير المتكافئة وذلك من خلال الخطأ المعياري للمعادلة المئينية العمودية. وقد استخدم الباحث طريقة المعادلة العمودية المنبثقة عن النظرية التقليدية في القياس، وخلصت الدراسة إلى أن دقة المعادلة المئينية العمودية كانت أعلى ما يمكن عندما كان عدد فقرات اختبار الجذع المشترك أكثر من 20 فقرة وممثلة للمحتوى والصعوبة والتمييز وأن توفر خاصية التمييز في الفقرات لاختبار الجذع المشترك تعطي دقة أكبر مما لو توفرت خاصية الصعوبة في الفقرات.

أما فيكتور (Victor, 2007) فقد أجرى دراسة هدفت إلى فحص دقة المعادلة باستخدام عينات صغيرة الحجم وباستخدام تصميم المجموعات العشوائية عندما تختلف الاختبارات في متوسط الصعوبة وبمستويات مختلفة من الصعوبة، كذلك كان الهدف الثاني من الدراسة هو التحقق من اثر خطا المعادلة في تصنيف الأخطاء عندما تقدم النتائج تبعاً للأداء المعياري. تم تحديد دقة المعادلة من خلال فحص مقدار الإحصائيات الثلاث الآتية: البواقي المعيارية والخطأ العشوائي والتحيز. تم تطبيق تسع طرق للمعادلة على بيانات مولدة وبمستويات مختلفة من الصعوبة وبستة أحجام من العينات (25-50-75-100-150-200). أشارت النتائج المتعلقة بطول الاختبار وتصميم المعادلة المستخدم في هذه الدراسة إلى أن دقة المعادلة في العينة صغيرة الحجم يعتمد على: اختلافات الصعوبة بين نماذج الاختبار، مدى الدرجات التي تقيم طريقة المعادلة وفقا له، حجم العينة، وطريقة المعادلة التي تطبق، كذلك أشارت النتائج إلى أن الاختبارات المتوسطة وذات الصعوبة الكبيرة ومن خلال التركيز على خطا العينة لتحديد فيما إذا كان سيتم تطبيق المعادلة على أحجام عينات صغيرة هو أمر غير مفيد، في حين أن العينات الصغيرة أنتجت أخطاء في العينات بشكل كبير.

كما أجرت امائدا (Amanda, 2008) دراسة هدفت من خلالها إلى مقارنة طرق المعادلة في النظرية الكلاسيكية ونظرية استجابة الفقرة باستخدام تقييم عدد الدرجات الصحيحة أو إحراز الصيغة. حيث تم تطبيق سبع طرق مختلفة لمعادلة هذين النوعين من الدرجات وهي: ثلاث طرق كلاسيكية (طريقة توكر الخطية،الطريقة غير الممهدة،وطريقة تشين المئينية)، وأربع طرق من طرق معادلة الدرجات في نظرية استجابة الفقرة (النموذج الأحادي،النموذج الثنائي، النموذج الثلاثي، ونماذج الاختيار من متعدد).تم مقارنة الطرق السبعة المستخدمة في هذه الدراسة باستخدام بيانات حقيقية والتي تم جمعها تحت تعليمات SAT وبيانات مولدة، فالبيانات الحقيقية استخدمت فيها بيانات إحراز الصيغة، بعد ذلك تم استخدام نماذج الاختيار من متعدد حتى يكمل الاستجابات المحذوفة للبيانات المولدة ليتبع تعليمات إحراز الدرجات الصحيحة. في البيانات المولدة تم تطبيق نموذج الاختيار من متعدد لتوليد البيانات

بحيث تمثل كلا التطبيقين. أشارت النتائج إلى أن الطريقة التي أنتجت أقل قيمة للتحيز في البيانات الحقيقية والمولدة هي طريقة توكر الخطية.

**التعقيب على الدراسات السابقة:**

من مجمل الدراسات السابقة يمكن القول إن هذه الدراسات تناولت معادلة الدرجات من جوانب متعددة، فمنها من تناول معادلة الدرجات باستخدام الاختبارات المشتركة الداخلية أو الخارجية مع التصاميم المختلفة لجمع المعلومات، وبعضها حاول التحقق من دقة معادلة طريقة من بين طرق المعادلة المختلفة.

وتطرقت بعض الدراسات إلى الحديث عن العوامل التي تؤثر في دقة معادلة الدرجات من مثل تلك المتعلقة بصعوبة الفقرات،أو حجم العينات،أو طبيعة البيانات المستخدمة،كما عالجت بعض الدراسات اثر زيادة كل من عدد فقرات الاختبار وحجم العينة على دقة معادلة الدرجات. وقد لاحظ الباحث أن بعض هذه الدراسات عانت من بعض جوانب القصور على الرغم من أن بعض هذه الجوانب تم الإشارة إليها كمحددات في هذه الدراسة مثل اقتصار عينتها على فئة معينة من الفئات المختلفة، إضافة إلى اقتصار بعض الدراسات على طريقة واحدة أو اثنتين من طرق معادلة درجات الاختبار المختلفة. وقد لاحظ الباحث أيضا أن الدراسات العربية رغم محدودية عددها لم تشمل طريقتي بيتا4 وكيرنيل،حيث ركزت معظمها على معادلة الدرجات باستخدام طرق نظرية استجابة الفقرة. ولعل هذه الدراسة تتميز عن الدراسات السابقة في كونها تناولت البيانات متعددة الحدود ومعادلتها بحسب النظرية الكلاسيكية في معادلة الدرجات والتي لم تبحث سابقا في الوطن العربي– في حدود علم الباحث– والتي لم تدرس من قبل. كذلك تتميز هذه الدراسة بتناولها لمتغيرات عديدة مثل حجم العينة، صعوبة الفقرات، وعدد فقرات الاختبار للوقف على أثرها في دقة معادلة الدرجات متعددة الحدود.

**الطريقة والإجراءات:**

هدفت هذه الدراسة إلى معادلة الدرجات الاختبارات متعددة الاستجابة في تصميم الفقرات المشتركة باستخدام طرق النظرية التقليدية في القياس، وفيما يلي وصفاً للمنهجية المتبعة في هذه الدراسة.

**أولاً:** تصميم الدراسة ومتغيراتها:

تكونت متغيرات الدراسة المستقلة من:1– طول الاختبار2– حجم العينة3– مستويات الصعوبة.

طول الاختبار: وقصد به عدد الفقرات في الاختبار الواحد بما فيها فقرات الجذع المشترك، وتكون من مستويين:1– 30 فقرة. 2– 60 فقرة. حجم العينة: ويقصد به عدد الأفراد في العينة الواحدة، وتكونت من ثلاثة مستويات:1– 200. 2– 600. 3– 1000. مستويات الصعوبة، وله مستويان:1– التشابه في متوسط معدل الصعوبة للاختبارات 2– الاختلاف في متوسط معدل الصعوبة.

وفي ضوء المتغيرات الثلاثة بمستوياتها،كان تصميم الدراسة (2 X 2 X 3) وبعدد خلايا (12) خلية، والجدول (1) يبين التصميم وخلاياه:

**الجدول (1)**

خلايا التصميم متضمناً المتغيرات المستقلة في الدراسة وعدد مرات التكرار في كل خلية

| طول الاختبار | طريقة المعادلة | التشابه في الصعوبة | | | الاختلاف في الصعوبة | | |
|---|---|---|---|---|---|---|---|
| | | SEE | | | RMSE | | |
| | | 200 | 600 | 1000 | 200 | 600 | 1000 |
| 30 | المئينات | √ | √ | √ | √ | √ | √ |
| 60 | المئينات | √ | √ | √ | √ | √ | √ |

**ثانياً:** المتغيرات التابعة:

وهو معيار الحكم على دقة معادلة درجات الصور المختلفة للاختبار وهو الخطأ المعياري للمعادلة، والجذر التربيعي لمتوسط مجموع مربعات الانحرافات الخطأ.

ويحسب الخطأ المعياري للمعادلة بالمعادلة الآتية:

$$SEE = \sqrt{\left[\frac{1}{n-1}\sum_{1}^{n}(\hat{e}_x(y_k) - \bar{\hat{e}}_x(y_k)\right]^{z'}}$$

حيث: $n$= عدد مرات توليد البيانات. $y_k$ = تمثل الدرجات على النموذج $y$.

$e_x(y_k)$= هي الدرجة المعادلة على النموذج $x$ من خلال النموذج $y_k$.

$\bar{\hat{e}}_x(y_k)$ = هي متوسط درجات المعادلة للدرجة $y_k$ عبر مرات التكرارات.

الجذر التربيعي لمتوسط مجموع مربعات الانحرافات الخطأ Root Mean Squared Error (RMSE)

ويمكن حساب جذر متوسط مربع الأخطاء المعيارية(RMSE ) من خلال المعادلة الآتية:

$$RMSE = \left[ \frac{1}{n_s} \sum \frac{1}{n_e} \sum_j (y_j - E_j)^2 \right]^{\frac{1}{2}}$$

حيث أن: $K$= عدد المفحوصين. $j$ = رمز المفحوص. $n_s$ = عدد العينات والمتمثلة في عدد مرات توليد البيانات. $n_e$= عدد المفحوصين في العينة الواحدة. $y_j$ = الدرجة على الصورة الأولى للمفحوص j في ضوء درجته على الصورة الثانية. $E_j$= الدرجة الحقيقية المتوقعة للمفحوص على الصورة الأولى.

توليد بيانات الاستجابة على الفقرات ومعادلة الدرجات

تم توليد البيانات ومعادلة الدرجات وتقييم دقة المعادلة وفق الخطوات الآتية:

**1**- توليد قدرات الأفراد: ولدت θ والممثلة لقدرات أفراد كل عينة باستخدام برمجية (WINGEN2)، حيث ولدت θ عندما يكون التوزيع توزيعا سوياً بمتوسط صفر وانحراف معياري واحد.

2- تقدير معالم الفقرات: تم تقدير معالم الفقرات باستخدام نموذج الاستجابات المتعددة، أما معالم الفقرات التي تم تمثيلها فهي (الصعوبة والتمييز)، ولتحقيق هذا الغرض استخدمت برمجية WINGEN2.

.3- اختيار الفقرات: تم تحديد فقرات كل صورة من صور الاختبار، كما حددت فقرات الجذع المشترك وبنسبة الثلث من عدد فقرات الاختبار الكلي.

4- توليد استجابات المفحوصين: ولدت استجابات كل مفحوص من إفراد العينة ( 200، 600، 1000) وذلك وفق مستويات كل متغير من المتغيرات المتمثلة في التصميم.حيث استخدمت قدرة المفحوص، كما تم توليدها سابقا ومعالم كل فقرة في نموذج الاستجابات المتعدد لتحديد احتمالية الإجابة عن الفقرة واستخدمت نفس البرمجية لتوليد الاستجابات.وبنفس الإجراءات تم توليد استجابات للمفحوصين على الصورة الثانية، وبالتالي تكون لكل مفحوص من إفراد العينة (200، 600، 1000) درجة كلية على الصورة الثانية من الاختبار ودرجة على فقرات الجذع المشترك.

5- تكرار الخطوة السابقة بتوليد الاستجابات للمفحوصين باستخدام معالم الفقرات ومعالم قدرات المفحوصين 200 مرة في كل خلية من خلايا التصميم

كررت الخطوة السابقة وذلك باستخدام نفس البرمجية حيث كان يتم في كل مرة توليد قدرات الأفراد على التوزيع المطلوب، وباستخدام نفس معالم الفقرات لتحديد استجابات المفحوصين وفق النموذج المتعدد الاستجابات.

6- تخزين استجابات المفحوص في كل عينة وفي كل مرة تكرر فيها للصورة الأولى والثانية في الاختبار.

**7**- تم تحويل الملفات إلى ملفات نصية لصورتي الاختبار الأولى والثانية ؛ وذلك من اجل التعامل مع البيانات، وتم إجراء المعادلة وفق الخطوات الآتية:

* تم حساب درجة كل مفحوص على الصورة الأولى من الاختبار ككل، ودرجته على فقرات الجذع المشترك.

* تم حساب درجة كل مفحوص على الصورة الثانية من الاختبار ككل، ودرجته على فقرات الجذع المشترك.

* تم حساب المتوسط الحسابي والانحراف المعياري لأفراد العينة، وذلك للدرجة الكلية وللدرجة على فقرات الجذع المشترك في الصورة الأولى، وكذلك الأمر في الصورة الثانية، ثم تم إجراء المعادلة المئينية المبنية على تصميم الاختبار المشترك ويمكن إجراؤها وفقا للخطوات الآتية:

1-إعداد توزيع العلامات لاختبار الجذع المشترك (u) للمجموعة الكلية والتي تشمل مجموعة الأفراد الذين أجابوا على الصورتين X, Y.

2-إيجاد العلامات على الاختبار (u) في توزيع المجموعة الكلية التي تناظر مئينات معينة مثل (p5,p10,p15…p95 ).

3-إعداد التوزيعات المنفصلة للاختبار المشترك (u) لكل مجموعة جزئية أجابت عن إحدى صورتي الاختبار المراد معادلتهما.

4-يتم تحديد الرتب المئينية المناظرة للعلامات التي تم تعريفها في الخطوة رقم (3) وذلك لكل مجموعة جزئية.

5-إيجاد العلامة X أو Y المناظرة للرتبة المئينية التي حسبت في الخطوة (4).

6-تعد العلامتان X, Y متكافئتين حين تناظر كل منهما الرتبة المئينية ذاتها على اختبار الجذع المشترك والتي حددت من المجموعة الكلية التي أعطيت الاختبارين.

7-رسم منحنى ممهد يمثل النقاط التي عينت في الخطوة رقم (6) والذي يستفاد منه في تحديد التكافؤ بين صورتي الاختبار Y,X.

**8**- تم إنشاء ملف يحتوي على قيم (X) وما يقابلها من قيم (Ŷ) لكل مفحوص من المفحوصين، حيث يتم وفق هذا الملف حساب معياري تقدير دقة المعادلة والبواقي المعيارية.

**9**-حساب SEE, RMSE

**نتائج الدراسة:**

هدفت هذه الدراسة إلى معادلة درجات الاختبارات

متعددة الحدود في تصميم الفقرات المشتركة باستخدام طرق النظرية التقليدية في القياس، وذلك في ظل ظروف تجريبية مختلفة وفيما ياتي عرضا لنتائج الدراسة وذلك بتناول كل متغير مستقل على حده، ثم تناول المتغيرات المستقلة مجتمعة:

**أولاً:** النتائج المتعلقة بنماذج الاختبار الذي يتألف من 30 فقرة:

ا- التشابة في مستويات صعوبة الفقرات:

**1- الخطأ المعياري للمعادلة:** يوضح الجدول (2) قيم الخطأ المعياري للمعادلة لطريقة المئينات في معادلة الدرجات للاختبار الذي يتألف من 30 فقرة والمتشابهة في الصعوبة:

**الجدول (2)**

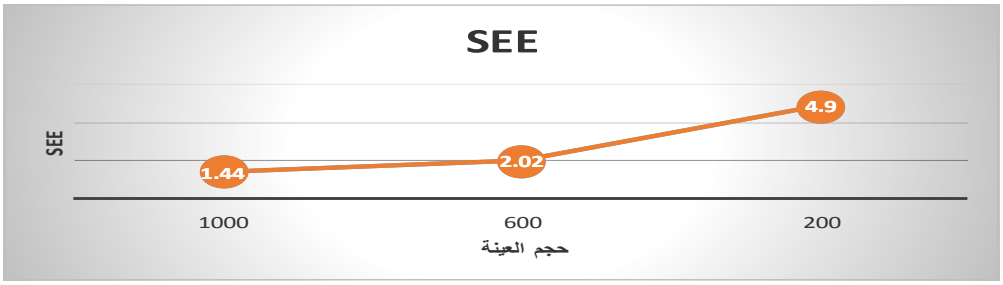
قيم الخطأ المعياري للمعادلة لطريقة المئينات في معادلة الدرجات للاختبار الذي يتألف من 30 فقرة والمتشابه في الصعوبة

| حجم العينة | 200 | 600 | 1000 |
|---|---|---|---|
| SEE | 4.90 | 2.02 | 1.44 |

**الجدول (3)**

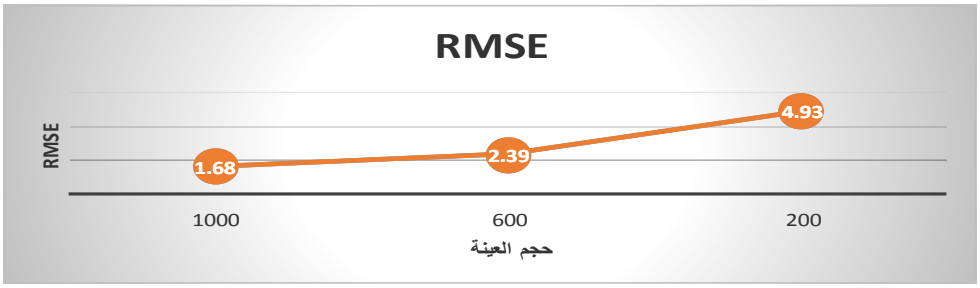
قيم (RMSE) لطريقة المئينات في معادلة الاختبارات للاختبار الذي يتألف من 30 فقرة والمتشابه في الصعوبة

| حجم العينة | 200 | 600 | 1000 |
|---|---|---|---|
| RMSE | 4.93 | 2.39 | 1.68 |




**الشكل (1)**

الخطأ المعياري للمعادلة لطريقة المئينات في معادلة درجات الاختبار الذي يتالف من (30) فقرة والمتشابه في الصعوبة




**الشكل (2)**

قيم (RMSE) لطريقة المئينات في معادلة الدرجات على سلم الدرجات وعبر اختلاف احجام العينات للاختبار الذي يتألف من 30 فقرة والمتشابه في الصعوبة

يظهر الجدول (2) أن قيم الخطأ المعياري للمعادلة لطريقة المئينات في معادلة الدرجات للاختبار الذي يتألف من 30 فقرة والمتشابه في الصعوبة قد تراوحت بين (1.44–4.90)، حيث بلغت أقل قيمة للخطأ المعياري (1.44)، في حين بلغت أعلى قيمة (4.90)، كذلك يظهر الجدول ان قيم الخطأ المعياري للمعادلة تنخفض مع ازدياد حجم العينة، فعندما تكونت العينة من (1000) مفحوص، بلغت قيمة الخطأ المعياري (1.44)، وعندما بلغ حجم العينة (600) مفحوص، بلغت قيمة الخطأ المعياري (2.02)، وعندما بلغ حجم العينة (200) مفحوص، بلغت قيمة الخطأ المعياري (4.90). ويوضح الشكل رقم (1) قيم الخطأ المعياري للمعادلة لطريقة المئينات على سلم الدرجات وعبر اختلاف حجم العينات.

يوضح الشكل (1) أن قيمة الخطأ المعياري للمعادلة تنخفض مع ازدياد حجم العينة، وأن المنحنيات تميل إلى الانخفاض وتقترب من الصفر عندما ترتفع حجم العينة، وهذا يشير إلى أن ارتفاع حجم العينة يقلل من قيمة الخطأ المعياري للمعادلة (SEE).

**2– (RMSE):** يوضح الجدول (3) قيم جذر متوسط مربعات الفروق (RMSE) لطريقة المئينات في معادلة درجات الاختبار الذي يتألف من (30) فقرة والمتشابه في الصعوبة.

يظهر الجدول (3) أن قيم (RMSE) لطريقة المئينات في معادلة الدرجات للاختبار الذي يتألف من 30 فقرة والمتشابه في الصعوبة قد تراوحت بين (1.68 –4.93)، حيث بلغت أقل قيمة ل (RMSE) (1.68)، في حين بلغت أعلى قيمة (4.39)، كذلك يظهر الجدول أن قيم (RMSE) للمعادلة تنخفض مع ازدياد حجم العينة، فعندما تألفت العينة من (1000) مفحوص، بلغت قيمة (RMSE) (1.68)، وعندما بلغ حجم العينة (600) مفحوص، بلغت قيمة (RMSE) (2.39)، وعندما بلغ حجم العينة (200) مفحوص، بلغت قيمة (RMSE) (4.93). ويوضح الشكل (2) قيم (RMSE) لطريقة المئينات في معادلة الدرجات على سلم الدرجات وعبر اختلاف احجام العينات للاختبار الذي يتألف من 30 فقرة والمتشابه في الصعوبة.

يوضح الشكل (2) ان طريقة المئينات وعبر الاحجام المختلفة للعينات كانت قيمة (RMSE) فيها أقل قيمة عندما كانت حجم العينة كبيرة (1000 مفحوص)، في حين كانت أعلى قيمة عند استخدام حجم عينة صغير (200 مفحوص)، ويلاحظ من الشكل (2) ان قيمة (RMSE) تنخفض مع ازدياد حجم العينة.

**ب**– الاختلاف في مستويات الصعوبة

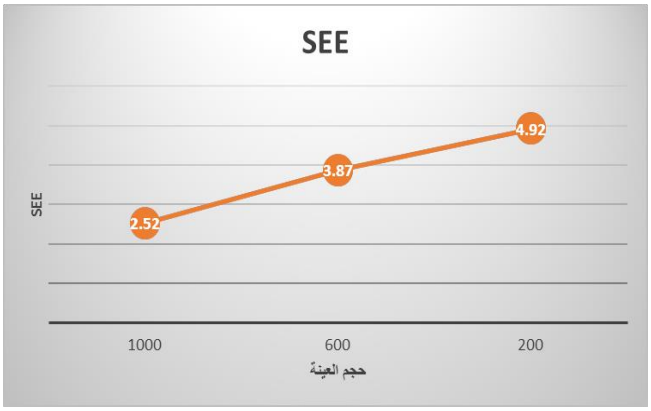
1–الخطأ المعياري للمعادلة: يوضح الجدول (4) قيم الخطأ المعياري الخطأ المعياري للمعادلة لطريقة المئينات في معادلة الدرجات للاختبار الذي يتألف من 30 فقرة والمختلف في الصعوبة:

**الجدول (4)**

قيم الخطأ المعياري للمعادلة لطريقة المئينات في معادلة الدرجات للاختبار الذي يتألف من 30 فقرة والمختلف في الصعوبة

| حجم العينة | 200 | 600 | 1000 |
|---|---|---|---|
| SEE | 4.92 | 3.87 | 2.52 |

يظهر الجدول (4) أن قيم الخطأ المعياري للمعادلة لطريقة المئينات في معادلة الدرجات للاختبار الذي يتألف من 30 فقرة والمختلف في الصعوبة قد تراوحت بين (2.52–4.92)، حيث بلغت أقل قيمة للخطأ المعياري (2.52)، في حين بلغت أعلى قيمة (4.92)، كذلك يظهر الجدول أن قيمة الخطأ المعياري للمعادلة تنخفض مع ازدياد حجم العينة، فعندما تكونت العينة من (1000) مفحوص، بلغت قيمة الخطأ المعياري (2.52)، وعندما بلغ حجم العينة (600) مفحوص، بلغت قيمة الخطأ المعياري (3.87)، وعندما بلغ حجم العينة (200) مفحوص، بلغت قيمة الخطأ المعياري (4.92). ويوضح الشكل (3) قيم الخطأ المعياري للمعادلة لطريقة المئينات على سلم الدرجات وعبر اختلاف حجم العينات.




**الشكل (3)**

الخطأ المعياري للمعادلة لطريقة المئينات في معادلة الدرجات على سلم الدرجات وعبر اختلاف احجام العينات للاختبار الذي يتالف من (30) فقرة والمختلف في الصعوبة

يوضح الشكل (3) أن قيمة الخطأ المعياري للمعادلة تنخفض مع ازدياد حجم العينة، وأن المنحنيات تميل إلى

الانخفاض وتقترب من الصفر عندما ترتفع حجم العينة، وهذا يشير إلى أن ارتفاع حجم العينة يقلل من قيمة الخطأ المعياري للمعادلة (SEE).

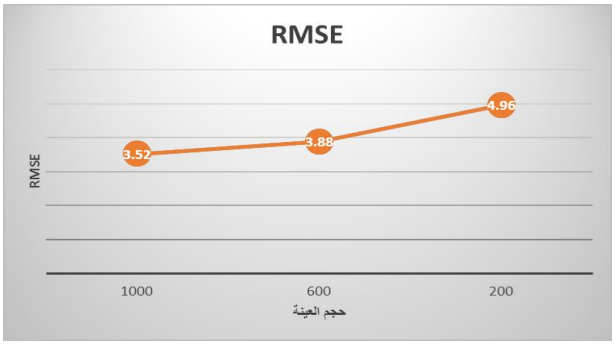
2- (RMSE): يوضح الجدول (5) قيم جذر متوسط مربعات الفروق (RMSE) لطريقة المئينات في معادلة الدرجات للاختبار الذي يتألف من 30 فقرة والمختلف في الصعوبة:

**الجدول (5)**

قيم(RMSE) لطريقة المئينات في معادلة الاختبارات للاختبار الذي يتألف من 30 فقرة والمختلف في الصعوبة

| حجم العينة | 200 | 600 | 1000 |
|---|---|---|---|
| RMSE | 4.96 | 3.88 | 3.52 |

يظهر الجدول (5) أن قيم (RMSE) لطريقة المئينات في معادلة الدرجات للاختبار الذي يتألف من 30 فقرة والمختلف في الصعوبة قد تراوحت بين (3.52 -4.96)، حيث بلغت أقل قيمة ل (RMSE) (3.52)، في حين بلغت أعلى قيمة (4.96)، كذلك يظهر الجدول أن قيم (RMSE) للمعادلة تنخفض مع ازدياد حجم العينة، فعندما تألفت العينة من (1000) مفحوص، بلغت قيمة (RMSE) (3.52)، وعندما بلغ حجم العينة (600) مفحوص، بلغت قيمة(RMSE) (3.88)، وعندما بلغ حجم العينة (200) مفحوص، بلغت قيمة (RMSE) (4.96). ويوضح الشكل (4) قيم (RMSE) لطريقة المئينات في معادلة الدرجات على سلم الدرجات وعبر اختلاف احجام العينات للاختبار الذي يتألف من 30 فقرة والمختلف في الصعوبة




**الشكل (4)**

قيم (RMSE) لطريقة المئينات في معادلة الدرجات على سلم الدرجات وعبر اختلاف احجام العينات للاختبار الذي يتألف من 30 فقرة والمختلف في الصعوبة

يوضح الشكل (4) ان قيم RMSE للمعادلة تنخفض مع ازدياد حجم العينة، وأن المنحنيات تميل إلى الانخفاض وتقترب من الصفر عندما ترتفع حجم العينة، وهذا يشير إلى أن ارتفاع حجم العينة يقلل من قيمة الخطأ المعياري للمعادلة (SEE).

**ثانياً:** النتائج المتعلقة بنماذج الاختبار الذي يتألف من 60 فقرة:

أ - التشابه في مستويات صعوبة النماذج:

1- الخطأ المعياري للمعادلة (SEE): يوضح الجدول (6) قيم الخطأ المعياري الخطأ المعياري للمعادلة لطريقة المئينات في معادلة الدرجات للاختبار الذي يتألف من 60 فقرة والمتشابه في الصعوبة:

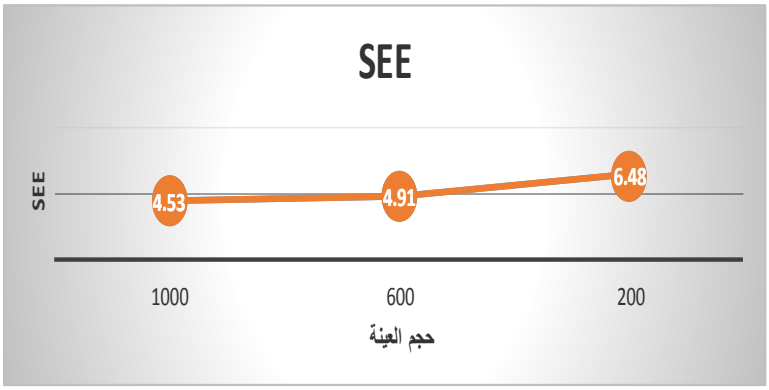
**الجدول (6)**

قيم الخطأ المعياري للمعادلة لطريقة المئينات في معادلة الدرجات للاختبار الذي يتألف من 60 فقرة والمتشابه في الصعوبة

| حجم العينة | 200 | 600 | 1000 |
|---|---|---|---|
| SEE | 6.48 | 4.91 | 4.53 |

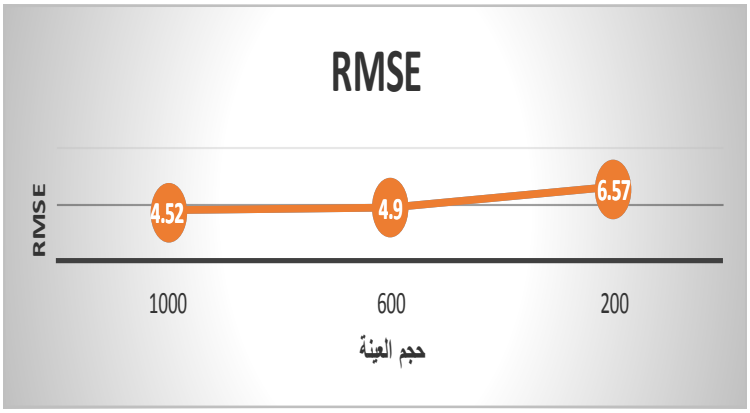
يظهر الجدول (6) أن قيم الخطأ المعياري للمعادلة لطريقة المئينات في معادلة الدرجات للاختبار الذي يتألف من 60 فقرة والمتشابه في الصعوبة قد تراوحت بين (4.53-6.48)، حيث بلغت أقل قيمة للخطأ المعياري (4.53)، في حين بلغت أعلى قيمة (6.48)، كذلك يظهر الجدول ان قيمة الخطأ المعياري للمعادلة تنخفض مع ازدياد حجم العينة، فعندما تكونت العينة من (1000) مفحوص، بلغت قيمة الخطأ المعياري (4.53)، وعندما بلغ حجم العينة (600) مفحوص، بلغت قيمة الخطأ المعياري (4.91)، وعندما بلغ حجم العينة (200) مفحوص، بلغت قيمة الخطأ المعياري (6.48). ويوضح الشكل رقم (5) قيم الخطأ المعياري للمعادلة لطريقة المئينات في معادلة الدرجات للاختبار الذي يتألف من 60 فقرة والمتشابه في الصعوبة.

يوضح الشكل (5) أن قيمة الخطأ المعياري للمعادلة تنخفض مع ازدياد حجم العينة، وأن المنحنيات تميل إلى الانخفاض وتقترب من الصفر عندما ترتفع حجم العينة، وهذا يشير إلى أن ارتفاع حجم العينة يقلل من قيمة الخطأ المعياري للمعادلة (SEE).




**الشكل (5)**

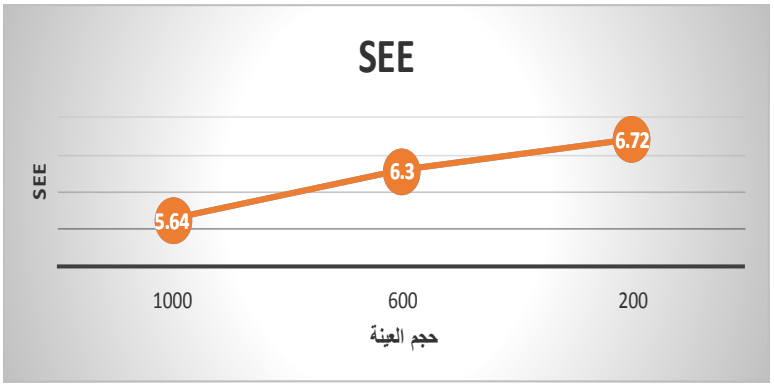
الخطأ المعياري للمعادلة لطريقة المئينات في معادلة الدرجات للاختبار الذي يتألف من 60 فقرة والمتشابه في الصعوبة




**الشكل (6)**

قيم RMSE للمعادلة لطريقة المئينات في معادلة الدرجات للاختبار الذي يتألف من 60 فقرة والمتشابه في الصعوبة

**2**




**الشكل (7)**

الخطأ المعياري للمعادلة لطريقة المئينات في معادلة الدرجات للاختبار الذي يتألف من 60 فقرة والمختلف في الصعوبة

**الجدول (7)**

قيم(RMSE) لطريقة المئينات في معادلة الاختبارات للاختبار الذي يتألف من 60 فقرة والمتشابهة في الصعوبة

| حجم العينة | 200 | 600 | 1000 |
|---|---|---|---|
| RMSE | 6.57 | 4.90 | 4.52 |

يظهر الجدول (7) أن قيم (RMSE) لطريقة المئينات في معادلة الدرجات للاختبار الذي يتألف من 60 فقرة والمتشابهة في الصعوبة قد تراوحت بين (4.52 –6.57)، حيث بلغت أقل قيمة ل (RMSE)(4.52)، في حين بلغت أعلى قيمة (6.57)، كذلك يظهر الجدول أن قيم (RMSE) للمعادلة تنخفض مع ازدياد حجم العينة، فعندما تألفت العينة من (1000) مفحوص، بلغت قيمة (RMSE) (4.52)، وعندما بلغ حجم العينة (600) مفحوص، بلغت قيمة (RMSE) (4.90)، وعندما بلغ حجم العينة (200) مفحوص، بلغت قيمة (RMSE)(6.57). ويوضح الشكل (6) قيم (RMSE) لطريقة المئينات في معادلة الدرجات على سلم الدرجات وعبر اختلاف احجام العينات للاختبار الذي يتألف من 60 فقرة.

يوضح الشكل (6) أن طريقة المئينات وعبر الأحجام المختلفة للعينات كانت قيمة (RMSE) فيها أقل قيمة عندما كانت حجم العينة كبيرة (1000 مفحوص)، في حين كانت أعلى قيمة عند استخدام حجم عينة صغير (200 مفحوص)، ويلاحظ من الشكل (6) أن قيمة (RMSE) تنخفض مع ازدياد حجم العينة.

**ب**– الاختلاف في مستويات الصعوبة

**1**–الخطأ المعياري للمعادلة (SEE): يوضح الجدول (8) قيم الخطأ المعياري الخطأ المعياري للمعادلة لطريقة المئينات في معادلة الدرجات للاختبار الذي يتألف من 60 فقرة والمختلف في الصعوبة.

–(RMSE): يوضح الجدول (7) قيم جذر متوسط مربعات الفروق (RMSE) لطريقة المئينات في معادلة الدرجات للاختبار الذي يتألف من 60 فقرة والمتشابه في الصعوبة:

**الجدول (8)**

قيم الخطأ المعياري للمعادلة لطريقة المئينات في معادلة الدرجات للاختبار الذي يتألف من 60 فقرة والمختلف في الصعوبة

| حجم العينة | 200 | 600 | 1000 |
|---|---|---|---|
| SEE | 6.72 | 6.30 | 5.64 |

يظهر الجدول (8) أن قيم الخطأ المعياري للمعادلة لطريقة المئينات في معادلة الدرجات للاختبار الذي يتألف من 60 فقرة والمختلف في الصعوبة قد تراوحت بين (5.64–6.72)، حيث بلغت أقل قيمة للخطأ المعياري (5.64)، في حين بلغت أعلى قيمة (6.72)، كذلك يظهر الجدول أن قيمة الخطأ المعياري للمعادلة تنخفض مع ازدياد حجم العينة، فعندما تكونت العينة من (1000) مفحوص، بلغت قيمة الخطأ المعياري (5.64)، وعندما بلغ حجم العينة (600) مفحوص، بلغت قيمة الخطأ المعياري (6.30)، وعندما بلغ حجم العينة (200) مفحوص، بلغت قيمة الخطأ المعياري (6.72). ويوضح الشكل (7) قيم الخطأ المعياري للمعادلة لطريقة المئينات على سلم الدرجات وعبر اختلاف حجم العينات.

يوضح الشكل (7) أن قيمة الخطأ المعياري للمعادلة تنخفض مع ازدياد حجم العينة، وأن المنحنيات تميل إلى الانخفاض وتقترب من الصفر عندما ترتفع حجم العينة، وهذا يشير إلى أن ارتفاع حجم العينة يقلل من قيمة الخطأ المعياري للمعادلة (SEE).

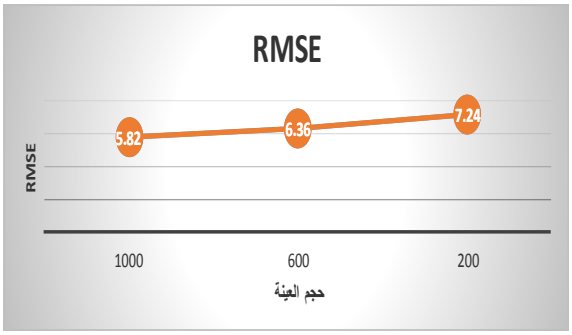
1–(RMSE): يوضح الجدول (9) قيم(RMSE) لطريقة المئينات في معادلة الاختبارات للاختبار الذي يتألف من 60 فقرة والمختلف في الصعوبة:

**الجدول (9)**

قيم(RMSE) لطريقة المئينات في معادلة الاختبارات للاختبار الذي يتألف من 60 فقرة والمختلف في الصعوبة

| حجم العينة | 200 | 600 | 1000 |
|---|---|---|---|
| RMSE | 7.24 | 6.36 | 5.82 |

يظهر الجدول (9) أن قيم(RMSE) لطريقة المئينات في معادلة الدرجات للاختبار الذي يتألف من 60 فقرة والمتختلف في الصعوبة قد تراوحت بين (5.82–7.24)، حيث بلغت أقل قيمة ل (RMSE) (5.82)، في حين بلغت أعلى قيمة (7.24)، كذلك يظهر الجدول أن قيم (RMSE) للمعادلة تنخفض مع ازدياد حجم العينة، فعندما تألفت العينة من (1000) مفحوص، بلغت قيمة (RMSE) (5.82)، وعندما بلغ حجم العينة (600) مفحوص، بلغت قيمة (RMSE) (6.36)، وعندما بلغ حجم العينة (200) مفحوص، بلغت قيمة (RMSE) (7.24). يوضح الشكل (8) قيم (RMSE) لطريقة المئينات في معادلة الدرجات على سلم الدرجات وعبر اختلاف احجام العينات للاختبار الذي يتألف من 60 فقرة.




**الشكل (8)**
قيم RMSE للمعادلة لطريقة المئينات في معادلة الدرجات للاختبار الذي يتألف من 60 فقرة والمختلف في الصعوبة

يوضح الشكل (8) ان طريقة المئينات وعبر الاحجام المختلفة للعينات كانت قيمة (RMSE) فيها أقل قيمة عندما كانت حجم العينة كبيرة (1000 مفحوص)، في حين كانت أعلى قيمة عند استخدام حجم عينة صغير (200 مفحوص)، ويلاحظ من الشكل (8) أن قيمة (RMSE) تنخفض مع ازدياد حجم العينة.

**مناقشة النتائج والتوصيات**

**أولاً:** مناقشة النتائج المتعلقة بنماذج الاختبار الذي يتألف من 30 فقرة:

ا- التشابة في مستويات صعوبة الفقرات:

**1**-الخطأ المعياري للمعادلة: أظهرت النتائج أن قيم الخطأ المعياري للمعادلة لطريقة المئينات في معادلة الدرجات للاختبار الذي يتألف من 30 فقرة والمتشابه في الصعوبة قد تراوحت بين (1.44 -4.90)، حيث بلغت أقل قيمة للخطأ المعياري (1.44)، في حين بلغت أعلى قيمة (4.90)، كذلك يظهر الجدول أن قيم الخطأ المعياري للمعادلة تنخفض مع ازدياد حجم العينة، فعندما تكونت العينة من (1000) مفحوص، بلغت قيمة الخطأ المعياري (1.44)، وعندما بلغ حجم العينة (600) مفحوص، بلغت قيمة الخطأ المعياري (2.02)، وعندما بلغ حجم العينة (200) مفحوص، بلغت قيمة الخطأ المعياري (4.90).

**2**-(RMSE): أظهرت النتائج أن قيم (RMSE) لطريقة المئينات في معادلة الدرجات للاختبار الذي يتألف من 30 فقرة والمتشابه في الصعوبة قد تراوحت بين (1.68 -4.93)، حيث بلغت أقل قيمة ل (RMSE)(1.68)، في حين بلغت أعلى قيمة (4.39)، كذلك يظهر الجدول أن قيم (RMSE) للمعادلة تنخفض مع ازدياد حجم العينة، فعندما تألفت العينة من (1000) مفحوص، بلغت قيمة (RMSE) (1.68)، وعندما بلغ حجم العينة (600) مفحوص، بلغت قيمة (RMSE) (2.39)، وعندما بلغ حجم العينة (200) مفحوص، بلغت قيمة (RMSE) (4.93).

**ب**- الاختلاف في مستويات الصعوبة

1-الخطا المعياري للمعادلة:أظهرت النتائج أن قيم الخطأ المعياري للمعادلة لطريقة المئينات في معادلة الدرجات للاختبار الذي يتألف من 30 فقرة والمختلف في الصعوبة قد تراوحت بين (2.52-4.92)، حيث بلغت أقل قيمة للخطأ المعياري (2.52)، في حين بلغت أعلى قيمة (4.92)، كذلك يظهر الجدول أن قيمة الخطأ المعياري للمعادلة تنخفض مع ازدياد حجم العينة، فعندما تكونت العينة من (1000) مفحوص، بلغت قيمة الخطأ المعياري (2.52)، وعندما بلغ حجم العينة (600) مفحوص، بلغت قيمة الخطأ المعياري (3.87)، وعندما بلغ حجم العينة (200) مفحوص، بلغت قيمة الخطأ المعياري (4.92).

**2**-(RMSE): أظهرت النتائج أن قيم (RMSE) لطريقة المئينات في معادلة الدرجات للاختبار الذي يتألف من 30 فقرة والمختلف في الصعوبة قد تراوحت بين (3.52 -4.96)، حيث بلغت أقل قيمة ل (RMSE) (3.52)، في حين بلغت أعلى قيمة (4.96)، كذلك يظهر الجدول أن قيم (RMSE) للمعادلة تنخفض مع ازدياد حجم العينة، فعندما تألفت العينة من (1000) مفحوص، بلغت قيمة (RMSE)(3.52)، وعندما بلغ حجم العينة (600) مفحوص، بلغت قيمة (RMSE) (3.88)، وعندما بلغ حجم العينة (200) مفحوص، بلغت قيمة (RMSE) (4.96).

**ثانياً:** مناقشة النتائج المتعلقة بنماذج الاختبار الذي يتألف من 60 فقرة:

أ - التشابه في مستويات صعوبة النماذج:

**1**-الخطا المعياري للمعادلة (SEE): أظهرت النتائج أن قيم الخطأ المعياري للمعادلة لطريقة المئينات في معادلة الدرجات للاختبار الذي يتألف من 60 فقرة والمتشابه في الصعوبة قد تراوحت بين (4.53 -6.48)، حيث بلغت أقل قيمة للخطأ المعياري (4.53)، في حين بلغت أعلى قيمة (6.48)، كذلك يظهر الجدول ان قيمة الخطأ المعياري للمعادلة تنخفض مع ازدياد حجم العينة، فعندما تكونت العينة من (1000) مفحوص، بلغت قيمة الخطأ المعياري (4.53)، وعندما بلغ حجم العينة (600) مفحوص، بلغت قيمة الخطأ المعياري (4.91)، وعندما بلغ حجم العينة (200) مفحوص، بلغت قيمة الخطأ المعياري (6.48).

**2**-(RMSE): أظهرت النتائج أن قيم (RMSE) لطريقة المئينات في معادلة الدرجات للاختبار الذي يتألف من 60 فقرة والمتشابهة في الصعوبة قد تراوحت بين (4.52-6.57)، حيث بلغت أقل قيمة ل (RMSE) (4.52)، في حين بلغت أعلى قيمة (6.57)، كذلك يظهر الجدول أن قيم (RMSE) للمعادلة تنخفض مع ازدياد حجم العينة، فعندما تألفت العينة من (1000) مفحوص، بلغت قيمة (RMSE) (4.52)، وعندما بلغ حجم العينة (600) مفحوص، بلغت قيمة (RMSE) (4.90)، وعندما بلغ حجم العينة (200) مفحوص، بلغت قيمة (RMSE) (6.57).

**ب**- الاختلاف في مستويات الصعوبة.

**1**-الخطأ المعياري للمعادلة(SEE): أظهرت النتائج أن قيم الخطأ المعياري للمعادلة لطريقة المئينات في معادلة الدرجات للاختبار الذي يتألف من 60 فقرة والمختلف في الصعوبة قد تراوحت بين (5.64-6.72)، حيث بلغت أقل قيمة للخطأ المعياري (5.64)، في حين بلغت أعلى قيمة (6.72)، كذلك يظهر الجدول ان قيمة الخطأ المعياري للمعادلة تنخفض مع ازدياد حجم العينة، فعندما تكونت العينة من (1000) مفحوص، بلغت قيمة الخطأ المعياري (5.64)، وعندما بلغ حجم العينة (600) مفحوص، بلغت قيمة الخطأ المعياري (6.30)، وعندما بلغ حجم العينة (200) مفحوص، بلغت قيمة الخطأ المعياري (6.72).

**2**-(RMSE): أظهرت النتائج أن قيم(RMSE) لطريقة المئينات في معادلة الدرجات للاختبار الذي يتألف من 60 فقرة والمتختلف في الصعوبة قد تراوحت بين (5.82 -7.24)، حيث بلغت أقل قيمة ل (RMSE) (5.82)، في حين بلغت أعلى قيمة (7.24)، كذلك يظهر الجدول أن قيم (RMSE) للمعادلة تنخفض مع ازدياد حجم العينة، فعندما تألفت العينة من (1000) مفحوص، بلغت قيمة (RMSE) (5.82)، وعندما بلغ حجم العينة (600) مفحوص، بلغت قيمة (RMSE) (6.36)، وعندما بلغ حجم العينة (200) مفحوص، بلغت قيمة (RMSE) (7.24).

تحت ظرف اختلاف حجم العينة، فإن النتائج أظهرت تقارب نتائج الخطأ المعياري للمعادلة (SEE) مع نتائج (RMSE) عندما ترتفع حجم العينة. حيث أظهرت النتائج أن قيم الخطأ المعياري وقيم (RMSE) تتأثر وبدرجة كبيرة بحجم العينة، فالعينات الكبيرة تنتج قيما صغيرة للخطأ المعياري وقيم (RMSE)، والعينات الصغيرة تنتج قيماً كبيرة للخطأ المعياري وقيم (RMSE)، كما أظهرت النتائج أن قيم الخطأ المعياري للمعادلة وقيم (RMSE) تقل مع ازدياد حجم العينات.

يتضح من النتائج السابقة أن قيمة الخطأ المعياري تتناقص تدريجياً بتزايد القدرة حتى تصل إلى أقصى قيمة ممكنة عندما تكون قيمة القدرة صفراً لوجيتاً، وهي تقابل متوسط الصعوبة للمفردات صفر لوجيت، إذ يكون الخطأ المعياري صغيرا كلما اقترب مستوى القدرة للأفراد من مستوى الصعوبة للمفردات، وبالتالي تزداد كمية المعلومات بنقصان الخطأ المعياري؛ لذلك فإن قيمة الخطأ المعياري يتناسب عكسياً مع حجم العينة، حيث إنه يزداد بزيادة حجم العينة، وتقل قميته بنقصان حجم العينة، وهذا أمر طبيعي؛ إذ أن كمية المعلومات عند أي مستوى من مستويات القدرة تتناسب عكسياً مع الخطأ المعياري، فعند زيادة حجم العينة يقترب متوسط معلمة التخمين من الصفر وهذا يؤدي إلى التقليل من قيمة الخطأ المعياري، حيث إن قيمة الخطأ المعياري تقل كلما قلت قمية التخمين، وقد يكون سبب ذلك أنه عند توليد البيانات باستخدام النموذج الثلاثي يتم أخذ معلمة التخمين بعين الاعتبار، مما يقلل من أثر التخمين وبالتالي التقليل من قيمة الخطأ المعياري والبواقي المعيارية، وهذا ما تؤيده نتائج دراسة (Victor, 2007) التي أشارت إلى أن العينات الصغيرة تؤدي إلى ارتفاع قيمة الخطأ المعياري، وأن اختلاف معامل الصعوبة بين نماذج الاختبار يؤدي إلى ارتفاع قيم الخطأ المعياري.

**ثالثاً:** مقارنة نتائج الخطأ المعياري للمعادلة وقيم (RMSE) من خلال طول الاختبار:

لقد أظهرت النتائج أن طول الاختبار يوثر في الخطأ المعياري للمعادلة، فالاختبار الطويل يعطي قيما كبيرة للخطأ المعياري للمعادلة، في حين أن تقليل طول الاختبار من (60) إلى (30) يؤدي إلى تخفيض كبير للخطأ المعياري للمعادلة. فعلى سبيل المثال عندما كان الخطأ المعياري (4.53) للاختبار الذي يتكون من (60) فقرة وعند حجم عينة (1000) مفحوص، فان هذا الخطأ المعياري انخفض إلى (1.44) للاختبار الذي يتكون من (30) فقرة وبنفس حجم العينة. أي ان الاختلاف في قيم الخطأ المعياري للمعادلة يصبح أقل عندما ينخفض طول الاختبار. كذلك أظهرت النتائج ان طول الاختبار يوثر في قيم (RMSE)، فالاختبار الطويل يعطي قيما كبيرة ل (RMSE)، فعند تقليل طول الاختبار من (60) إلى (30) فقرة، أدى هذا إلى انخفاض كبير لقيم (RMSE) مقارنة مع الانخفاض عندما.

يمكن تفسير هذه النتيجة في عاملين رئيسين:

أ-أن قيمRMSE، SEE المستخدمة هنا ليست قيما معيارية (Standardized)، لأن فقرات الاختبارات لم تقسم باستخدام تطابق الانحراف المعياري للعلامات المشاهدة؛ ولأن الانحراف

المعياري للدرجات الصحيحة يرتفع مع زيادة طول الاختبار، فانه يتوقع أن القيم المعيارية لقيم SEE،RMSE سوف تظهر ميلاً قليلاً إلى الارتفاع مع ازدياد طول الاختبار.

ب–أن الاختبار الطويل يعني احتمالية عالية عند نقاط الدرجة، وبالتالي مفحوصين قليلين لكل درجة للعينات المركبة، وبالتالي فإن العدد القليل من المفحوصين عند كل درجة يقلل من دقة المعادلة عند كل النقاط. وهذه النتيجة لا تتفق مع نتائج دراسة (الصمادي، 2007) والتي أشارت إلى ان طول الاختبار يؤدي إلى نقصان قيمة الخطأ المعياري.

**خامساً:** مقارنة نتائج معادلة درجات الاختبار من خلال حجم العينة وصعوبة الفقرات

**1**–الخطأ المعياري للمعادلة: تحت الظروف المختلفة لحجم العينة والتشابه في صعوبة النماذج، أظهرت النتائج أن قيم الخطأ المعياري للاختبار الذي يتكون من (60) فقرة تنخفض عندما تتشابه النماذج في مستوى صعوبتها، فقد بلغت قيمة الخطأ المعياري للعينة التي تكونت من (1000) مفحوص والمتشابهة في مستوى الصعوبة (4.53)، في حين فقد بلغت قيمة الخطأ المعياري لنفس العينة التي تكونت من (1000) مفحوص والمختلفة في مستوى الصعوبة (5.64)، كما بلغت قيمة الخطأ المعياري للعينة التي تكونت من (600) مفحوص والمتشابهة في مستوى الصعوبة (4.91)، في حين بلغت قيمة الخطأ المعياري لنفس العينة التي تكونت من (600) مفحوص والمختلفة في مستوى الصعوبة (6.30)، وكذلك بلغت قيمة الخطأ المعياري للعينة التي تكونت من (200) مفحوص والمتشابهة في مستوى الصعوبة (6.48)، في حين بلغت قيمة الخطأ المعياري لنفس العينة التي تكونت من (200) مفحوص والمختلفة في مستوى الصعوبة (7.24)، وهذا يعطي دلالة واضحة بان النماذج الممتشابهة في صعوبتها تميل قيم الخطأ المعياري إلى الانخفاض عندما تتشابه مستويات الصعوبة فيها، والنماذج المختلفة في صعوبتها تميل قيم الخطأ المعياري إلى الارتفاع عندما تتشابه مستويات الصعوبة فيها.

**2**–(RMSE): تحت الظروف المختلفة لحجم العينة والتشابه في صعوبة النماذج، أظهرت النتائج أن طريقة المعادلة باستخدام (IRT) أظهرت قيما منخفضة لقيم (RMSE) وعبر الاحجام المختلفة للعينات. كما أظهرت النتائج ان قيم (RMSE) للاختبار الذي يتكون من (60) فقرة تنخفض عندما تختلف النماذج في مستوى صعوبتها، فقد بلغت قيمة (RMSE) للعينة التي تكونت من (1000) مفحوص والمتشابهة في مستوى الصعوبة (4.52)، في حين فقد بلغت قيمة (RMSE) لنفس العينة التي تكونت من (1000) مفحوص والمختلفة في مستوى الصعوبة (5.82)، كما بلغت قيمة (RMSE) للعينة التي تكونت من (600) مفحوص والمتشابهة في مستوى الصعوبة (4.90)، في حين بلغت قيمة (RMSE) لنفس العينة التي تكونت من (600) مفحوص والمختلفة في مستوى الصعوبة (6.36)، وكذلك بلغت قيمة (RMSE) للعينة التي تكونت من (200) مفحوص والمتشابهة في مستوى الصعوبة (6.57)، في حين بلغت قيمة (RMSE) لنفس العينة التي تكونت من (200) مفحوص والمختلفة في مستوى الصعوبة (7.24)، وهذا يعطي دلالة واضحة بان النماذج المتشابهة في صعوبتها تميل قيم (RMSE) إلى الانخفاض عندما تتشابه مستويات الصعوبة فيها، والنماذج المختلفة في صعوبتها تميل قيم (RMSE) إلى الارتفاع عندما تختلف مستويات الصعوبة فيها.

من المعلوم أن الدقة في تقدير معلمة الصعوبة والقدرة تزداد عندما يكون مدى القدرة للمفحوص متوافقا مع مدى صعوبة الفقرات، وكذلك تزداد الدقة في تقدير معلمة التخمين عند استخدام عينة من ذوي القدرة المتدنية في معايرة الفقرات، كما أن تقديرات معالم الفقرة تتسم بالاستقرار النسبي عند معايرتها ضمن المقياس ككل أو بصورة مستقلة باستخدام العينة نفسها من الأفراد، وأن الدقة في تقدير معالم الفقرة تزداد بزيادة حجم عينة المفحوصين، كما أن الخطأ المعياري في تقدير معلمة صعوبة الفقرة يتناقص بزيادة عدد أفراد العينة، وهذا ما تؤيده نتائج دراسة (Victor,2007) التي أشارت إلى ان العينات الصغيرة تؤدي إلى ارتفاع قيمة الخطأ المعياري، وان اختلاف معامل الصعوبة بين نماذج الاختبار يؤدي إلى ارتفاع قيم الخطأ المعياري.

**الاستنتاجات**

خلصت الدراسة إلى الاستنتاجات الآتية:

1– أظهرت نتائج هذه الدراسة أن أحجام العينات الكبيرة يوثر على معايير دقة المعادلة، فحجم العينات الكبير يؤدي إلى تقليل الخطأ المعياري SEE وتقليل قيم RMSE.

2– بصورة عامة فإن الإختلاف في الخطأ المعياري للمعادلة وقيم RMSE يصبح أقل عندما ترتفع حجم العينة.

3– كذلك أشارت النتائج إلى أن اختلاف النماذج في صعوبتها يزيد من الخطأ المعياري والبواقي المعيارية.

4– كذلك أظهرت النتائج ان طول الاختبار يوثر في الخطأ المعياري، فالاختبار الطويل يعطي قيماً كبيرة للخطأ المعياري، والبواقي المعيارية.

**التوصيات:**

1- يوصي الباحث بتطبيق الطريقتين على بيانات حقيقية متعددة الاستجابة؛ وذلك للوقوف على مقدار الإختلاف في المعادلة بين البيانات الحقيقية والبيانات التجريبية.

2- استخدام طرق أخرى للمعادلة للوقوف على مدى دقتها في معادلة الدرجات.

استخدام تصميم آخر غير تصميم الفقرات المشتركة للوقوف على مقدار الإختلاف في دقة المعادلة لنفس الطرق المستخدمة في هذه الدراسة مثل: تصميم المجموعة المتكافئة (Equivalent Group)، أو تصميم المجموعة المفردة (Single Group Design) أو تصميم المجموعات العشوائية المتوازنة (Counter-Balanced Design).

## المصادر والمراجع


الصمادي، إ (2007). أثر طريقة اختيار الفقرات في اختبار الجذع المشترك على دقة معادلة اختبار متعدد المستوى في الرياضيات للمرحلة الأساسية في الأردن، رسالة دكتوراه غير منشوره،جامعة عمان العربية، عمان، الأردن.

علام، ص (2000). القياس والتقويم التربوى والنفسى – أساسياته وتطبيقاته وتوجهاته المعاصرة، القاهرة، دار الفكر العربى.

المدانات، ر. (2008). أثر طريقة المعادلة باستخدام جذع مشترك وعدد فقراته وحجم العينة على القيم المعادلة والخطأ في المعادلة بين صورتي اختبار في الفيزياء، رسالة دكتوراه غير منشورة، جامعة عمان العربية،عمان، الأردن.

Amanda, A. (2008). A comparison of classical test theory and item response theory methods for equating Number-right scored to formula scored assessments. Unpublished DoctoralDissertation, University of Kansas, USA. Angoff, W.H. (1971). Scales, norms, and equivalent scores. InR L.Thorndike (Ed), Educational measurement (2nded.) (pp.508-600).Washington: American Council on Education.

Hambleton, (1982). Advances in Item Response Theory and Applications: An Introduction. Applied Psychological Measurement; 6, 373-378.

Hue, G. (2004).The effect of different anchor tests on the accuracy of test equating for test adaptation. Unpublished Doctoral Dissertation, Ohio University, USA.

Kolen, M. and Brennan, R. L. (2004). Test equating, scaling, and linking and practices, (2nded). New York: Springer- Verlag.

Lee, G.Hill, Kolen, M. J., Frisbee, D. & Ankenmann, R. (2001). Comparison of dichotomous and polytomous item Response models in equating scores from tests composed of test lets. Applied Psychological Measurement, 25, (4): 357-372

Lord, D. M. (1980). Applications of item Response Theory to Practical Testing Problems, Hillsdale. N.J: Erlbaum.

Victor K. (2007). Equating accuracy using small samples in the random groups design. Unpublished Doctoral Dissertation, Ohio University, USA.

# Equating Polynomial Scores under Anchor Items Design Using Classical Test Theory Methods (CTT) "Experimental Study"

***Yousef A. Al-Mahrouk****

## ABSTRACT

This study explored the effects of sample size, test length, and difficulty level on final equating results of equipercentile method when using the nonequivalent groups anchor test (NEAT) design. This study used Simulation data. The equating method was evaluated using an equating criterion (SEE, RMSE). Standard error of equating differences between the criterion scores and equated scores, and root mean square error of the differences (RMSE) were used as a measure to compare the method to the criterion equating, which in this study is an equivalent groups (EG) equipercentile equating function. The results indicated that the large sample size reduces the standard error of the equating and reduces residuals, also results indicated that difficulty of tests increases the standard error and standard residuals. The results also showed that the test length has affected on the standard error. The long test gives large standard error, and the residuals standard.

**Keywords:** Equating Test, NEAT Design, Polynomial Scores, Equating Polynomial Scores under Anchor Items Design Using Classical Test Theory Methods (CTT).

* Directorate of Curricula, Ministry of Education, Kingdom of Bahrain. Received on 09/11/2014 and Accepted for Publication on 22/12/2014.