المجلد 16 العدد 3 سبتمبر 2015

المفاضلة بين الدرجات المكافئة لنماذج اختبار القدرات العامة باستخدام طرق المعادَلة الكلاسيكية في تصميم المفردات المشتركة بجماعات غير متكافئة

د. راشد سيف المحرزي
قسم علم النفس
كلية التربية- جامعة السلطان قابوس
mehrzi@squ.edu.om

المفاضلة بين الدرجات المكافئة لنماذج اختبار القدرات العامة باستخدام طرق المعادَلة الكلاسيكية في تصميم المفردات المشتركة بجماعات غير متكافئة

د. راشد سيف المحرزي
قسم علم النفس
كلية التربية- جامعة السلطان قابوس

الملخص

تهدف الدراسة إلى مقارنة الدرجات المكافئة لنماذج اختبار القدرات العامة في تصميم المفردات المشتركة بجماعات غير متكافئة حسب الطرائق الإحصائية للمعادلة وفقا للنظرية الكلاسيكية للاختبار. طُبقت الدراسة من خلال إجراء معادلة لدرجات نموذجين جديدين مع نموذج مرجعي في اختباري القدرة اللفظية والقدرة الكمية من اختبار القدرات العامة. استخدمت في الدراسة خمس طرائق للمعادلة وهي: المعادلات الخطية (Tucker, Levine, Braun-Holland) والمعادلات غير الخطية (المئينية والمئينية المهدة بعديا). وتوصلت الدراسة إلى اختلاف الإحصاءات الوصفية والتوزيعات التكرارية لدرجات نماذج الاختبار ولدرجات المفردات المشتركة. وعند إجراء المعادلة، تظهر النتائج أن الدرجات المكافئة الناتجة من الطرائق المئينية تتشابه مع طرائق المعادلات الخطية عند أغلب درجات الاختبار، ويتمركز الاختلاف بينها عند الدرجات الطرفية (المنخفضة والمرتفعة). كما تظهر النتائج أن الطريقة المئينية أصغر متوسط للأخطاء المعيارية للمعادلة.

الكلمات المفتاحية: معادلة نماذج الاختبار، طرق المعادلة الكلاسيكية، تصميم المفردات المشتركة، اختبار القدرات العامة.

* تاريخ تسلم البحث: ٢٠١٣/١٠/٢١م
* تاريخ قبوله للنشر: ٢٠١٤/٤/٩م

Differentiation among Equated Scores for General Abilities Test Forms Using Classical Equating Methods with Common-Item Non-Equivalent Groups Design

Dr. Rashid S. Almehrizi

Psychology Department-College of Education
Sultan Qaboos University

Abstract

The study aims to differentiate among equated scores for the General Abilities Test forms within common item non-equivalent group design using classical equating methods. Two new forms are equated to a reference form for the Verbal ability test and the Quantitative ability test. The study uses five classical equating methods: Linear equating methods (Tucker, Levine, and Braun-Holland) and Nonlinear equating methods (Equipercentile and Smoother Equipercentile). The results show differences among the descriptive statistics and frequency distributions for test forms. After equating, results show that the equating score from equipercentile methods at the majority of test scores, and different equating scores are obtained at the two extreme test scores (low and high scores). In addition, equipercentile method show smaller standard error of equating.

Keywords: test forms equating, classical equating methods, common item design, general ability test.

المفاضلة بين الدرجات المكافئة لنماذج اختبار القدرات العامة باستخدام طرق المعادَلة الكلاسيكية في تصميم المفردات المشتركة بجماعات غير متكافئة

د. راشد سيف المحرزي
قسم علم النفس
كلية التربية- جامعة السلطان قابوس

المقدمة:

تتميز الاختبارات واسعة النطاق (Large scale tests) بالتطبيقات العديدة –غالبا ما تكون في أيام مختلفة- على عينات كبيرة ومتنوعة من الأفراد، باستخدام أعداد كبيرة من المفردات التي تظهر على صورة نماذج مختلفة من الاختبار. كما تتميز هذه الاختبارات واسعة النطاق غالبا بأهمية وخطورة القرارات التي تبنى على نتائج الطلاب فيها نيتكو وبروكهارت (٢٠١٢). ولذلك تتطلب هذه الاختبارات تطبيقا دقيقا لنوع معين من الإجراءات في صياغة مفرداتها، وبناء نماذج اختباراتها، وتطبيقها وإدارتها، وطريقة تصحيحها وحساب درجاتها المعيارية وتفسير درجاتها، والتأكد من ثبات وصدق نتائجها وخصائصها السيكومترية الأخرى. وتعد إجراءات بناء نماذج متكافئة من هذه الاختبارات وإجراءات معادلة لدرجات هذه النماذج من بين الإجراءات الهامة لهذا النوع من الاختبارات، حيث تعد هذه الإجراءات أحد مجالات البحث والدراسة في قياس وتقويم السمات التربوية والنفسية (Dorans, Moses, & Eignor, 2011).

وتنبع الحاجة الماسة إلى وجود نماذج مختلفة لهذا النوع من الاختبارات وتطبيق إجراءات المعادلة لها استنادا إلى الخطورة المترتبة على الاقتصار على نموذج واحد فقط من الاختبار؛ وذلك بسبب سهولة تسرب مفرداته، والذي يؤدي إلى القضاء على مصداقية درجات الطلبة، وعدم تحقيق العدالة بين الطلبة في الأيام المختلفة لتطبيق الاختبار. كما أن صياغة نماذج مختلفة من الاختبار، وتطبيقها على الطلبة من دون التأكد من تكافؤ هذه النماذج، وعدم تطبيق إجراءات معادلة درجاتها يؤدي إلى عدم تحقيق العدالة بين الطلبة، حين يقدم بعض الطلبة نماذج قليلة الصعوبة؛ بينما يقدم آخرون نماذج شديدة الصعوبة (Linvingston, 1996).

وتتطلب عملية معادلة درجات نماذج الاختبار مرحلتين أساسيتين: (١) تكافؤ الصياغة: وتتطلب بناء نماذج متكافئة من حيث جدول المواصفات والخصائص السيكومترية لمفرداتها، (٢) تكافؤ الدرجات: وتتطلب اختيار تصميم ملائم لتطبيق نماذج الاختبار يساعد في الربط بين درجات نماذج الاختبار، وتطبيق الطريقة الإحصائية الملائمة لمعادلة درجات النماذج وفقا لتصميم المعادلة. وبرغم أهمية مرحلة تكافؤ الصياغة لتحقيق نماذج متكافئة للاختبار، فإنها لا تحقق تكافؤ درجات نماذج الاختبار ولا تغني عنها (Angoff, 1987). فتطبيق نماذج الاختبار (وحتى لو كانت متكافئة من حيث الصياغة) على عينات مختلفة من الأفراد يؤدي -في غالب الأحيان- إلى اختلاف متوسطاتها الحسابية (أو جميع إحصائياتها الحسابية وتوزيعها التكراري)؛ وذلك إما بسبب اختلاف صعوبة نماذج الاختبار، أو بسبب اختلاف مستوى قدرة عينات مقدمي نماذج الاختبار أو اختلاف ظروف تطبيق نماذج الاختبار. ولا يمكن الجزم بالسبب الرئيس وراء هذا الاختلاف بين نتائج نماذج الاختبار، فهناك عدد من العوامل المرتبطة بعملية تطبيق الاختبار ومستويات الطلبة وسلوك المفردات أثناء التطبيق، وتفاعل الطلبة معها تجعل النتائج التي يتحصل عليها من تطبيق نماذج الاختبار غير متطابقة.

وتعرف عملية معادلة نماذج الاختبار Test form Equating بالعملية الإحصائية التي تستخدم لتعديل الدرجات في النماذج المختلفة للاختبار، بما يُمكن لاحقا استخدام درجات تلك النماذج بطريقة تبادلية (Kolen & Brennan, 2004). وبالتالي يمكن القول بأن عملية معادلة نماذج الاختبار ضرورية لضمان تساوي مستوى صعوبة هذه النماذج واستخدامها بشكل تبادلي (Chulu & Sireci, 2011). وتقوم عملية معادلة نماذج الاختبار على مبدأ تعديل الاختلافات في مستوى الصعوبة لنماذج الاختبار التي بنيت لكي تكون متشابهة في الصعوبة والمحتوى ولتستخدم بطريقة تبادلية. وتتطلب هذه العملية تطبيق تصميماً خاصاً يسمى تصميم المعادلة (Equating design) لتطبيق النموذج الجديد للاختبار من أجل معادلة درجاته مع درجات النموذج المرجعي. ويهدف تصميم المعادلة إلى ضبط مستوى القدرة والسمة التي يقيسها الاختبار بين العينات التي يطبق عليها الاختبار؛ ليتسنى بعد ذلك التأكد من التوزيع التكراري والمتوسط الحسابي للنماذج المختلفة. فإذا تبقى اختلاف في المتوسط الحسابي (أو التوزيع التكراري) لنماذج الاختبار، فيرجع ذلك إلى الاختلاف في مستوى صعوبتها، والذي يتطلب أن تُطبق عملية إحصائية على درجات نماذج الاختبار؛ للحصول على الدرجات المكافئة للنموذج الجديد مع النموذج المرجعي.

وتزخر أدبيات مجال معادلة نماذج الاختبار بتصاميم عديدة ومتجددة لجمع البيانات

بغرض إجراء عملية المعادلة، والتي تقوم كلها على مبدأ استبعاد اختلاف مستوى القدرة المقيسة لدى عينات الأفراد التي تقدم النماذج المختلفة للاختبار. ويلخص كل من تشن وليفنجستون وهولاند (Chen, Livingston, & Holland, 2011) وكولن وبرنن (Kolen & Brennan, 2004) ثلاثة تصاميم رئيسة لجمع بيانات معادلة الاختبارات تتمثل في (١) تصميم الجماعة الواحدة (Single Group Design) والذي يتطلب تقديم نماذج الاختبار (نموذجين أو أكثر) لجماعة معينة من الأفراد، بحيث يقدم كل فرد جميع نماذج الاختبار المراد معادلة درجاتها، (٢) تصميم الجماعات المتكافئة (Equivalent Groups Design)، والذي يقوم على تطبيق نماذج الاختبار على جماعات من الأفراد متكافئة في السمة المقاسة في الاختبار تحت نفس ظروف التطبيق، بحيث يقدم الفرد الواحد نموذجا واحدا فقط من الاختبار، (٣) تصميم المفردات المشتركة بجماعات غير متكافئة (Common-Item Non-equivalent Groups Design) أو تصميم الاختبار الرابط لجماعات غير متكافئة (Non-equivalent Anchor Test - NEAT)، والذي يتطلب وضع مفردات مشتركة بين نماذج الاختبار، وتقوم جماعات طلابية غير متكافئة في مستوى قدرتها بتقديم نماذج الاختبار، بحيث تقدم كل جماعة طلابية نموذجا واحدا بدون اشتراط تقديمها في نفس الوقت والظروف.

يتيح تصميم المفردات المشتركة بجماعات غير متكافئة عددا من التسهيلات العملية في تطبيق نماذج الاختبار على الأفراد، فهو يسمح بتقديم نماذج مختلفة من الاختبار في كل يوم من أيام التطبيق، و لا يتطلب من الطالب أن يقدم أكثر من نموذج واحد سواء في نفس التوقيت أو توقيت مختلف. ويساعد هذا التصميم على التقليل من احتمالية تسرب مفردات الاختبار، وتقليل الآثار المترتبة على حدوث نوع معين من التسرب في حالة حصوله. بينما يُدخل هذا التصميم شروطا أكثر في عملية بناء نماذج الاختبار، تتمثل في بناء مفردات مشتركة Common Items أو اختباراً رابطاً Anchor test بين نماذج الاختبار، والتي يتطلب أن يتوفر بها عدد من الشروط، من حيث عددها، وتمثيلها لمحتوى الاختبار ككل، وخصائصها السيكومترية، وترتيبها في الورقة الامتحانية، وغيرها. فيشترط أن يكون عددها ملائما مع العدد الكلي لمفردات الاختبار بحيث تتراوح النسبة المئوية لعددها إلى عدد المفردات الكلية بين ٢٠٪ إلى ٤٠٪ (Chen, Livingston, & Holland, 2011). وبينما لا يوجد اتفاق على عدد المفردات المشتركة نظرا لاختلاف نوعية الاختبارات وأغراضها، فهناك اتفاق على ضرورة تمثيلها مجتمعة لمفردات الاختبار الكلي، بحيث تشكل نسخة مصغرة من الاختبار الأصلي، وأن تتوزع بنفس النسب على مكونات جدول مواصفات الاختبار الكلي، وأن تتكافأ خصائصها

السيكومترية مع الخصائص السيكومترية للاختبار الكلي.

ويمكن دمج المفردات المشتركة مع مفردات الاختبار بطريقتين: تسمى الطريقة الأولى بالمفردات المشتركة الداخلية Internal Common Items، وتسمى الثانية بالمفردات المشتركة الخارجية (External Common Items Raykov, 2010). وتعتبر المفردات المشتركة الداخلية جزءاً من مفردات كل نموذج من نماذج الاختبار، وتدخل في حساب درجة الطالب في الاختبار؛ بينما لا تعتبر المفردات المشتركة الخارجية جزءاً من مفردات نماذج الاختبار، بل زيادة عليها، ولا تدخل في حساب درجة الطالب في الاختبار. فمثلا، إذا كان عدد مفردات الاختبار ٥٠ مفردة، وعدد المفردات المشتركة ١٥ مفردة، فطريقة المفردات المشتركة الداخلية تعني أن ١٥ مفردة من مفردات الاختبار الخمسين هي مفردات مشتركة ومكررة في نماذج الاختبار؛ بينما تعني طريقة المفردات المشتركة الخارجية أن الاختبار مكون من ٥٠ مفردة، وهناك ١٥ مفردة إضافية، يتم تقديمها مع نماذج الاختبار المختلفة. وتتوزع المفردات المشتركة الداخلية بشكل متداخل مع مفردات الاختبار في الورقة الامتحانية، وتحتفظ بترتيبها في الورقة الامتحانية في نماذج الاختبار المختلفة. في المقابل، توضع المفردات المشتركة الخارجية مجتمعة وغالبا في قسم منفصل عن مفردات الاختبار، ولكن لا يتم تمييزه عن أقسام الاختبار الأخرى، بحيث لا توحي للطالب أنها مفردات لا تدخل درجاتها في درجته الفعلية في الاختبار، وبالتالي يختلف أداؤه فيها عن أدائه الحقيقي.

وتعتمد نتائج معادلة درجات نماذج الاختبار في تصميم المفردات المشتركة على أداء الجماعات الطلابية التي قدمت نماذج الاختبار في المفردات المشتركة أو الاختبار الرابط والخصائص السيكومترية لهذه المفردات. ويناقش كل من مروك وسوه وكين وربكي (Mroch, Suh, Kane, & Ripkey, 2009) الشروط الضرورية لتطبيق تصميم المفردات المشتركة. ويؤكدون على صعوبة اعتماد نتائج تطبيق معادلة درجات نماذج الاختبار باستخدام تصميم المفردات المشتركة في حالة اختلاف الخصائص السيكومترية للمفردات المشتركة مثل معاملات الصعوبة والتمييز ومعامل الثبات ومعاملات الارتباط مع مؤشرات أخرى.

ويعاب على المفردات المشتركة (بنوعيها) أنها عرضة للتسرب أكثر من بقية المفردات؛ بسبب تكرارها بين نماذج الاختبار، وبالتالي قد يكون الأداء فيها مختلفا -في حالة تسربها- عن الأداء الفعلي في حالة عدم تسربها. وبينما تتميز المفردات المشتركة الداخلية أنها جزء من الاختبار و لا يضيع جهد الطالب فيها سدى (يقلل عدد المفردات التي يقدمها الفرد ويوفر وقت الطالب وإدارة الاختبار)، إلا أن تسرب بعض المفردات المشتركة الداخلية يؤثر على

نتائج الطلاب أكثر من المفردات المشتركة الخارجية؛ نظرا لاحتساب درجة الطالب بها ضمن درجته الكلية. ويمكن التقليل من فرصة تسرب المفردات المشتركة من خلال استخدام مفردات مشتركة متشابهة بين بعض نماذج الاختبار ومختلفة عن نماذج أخرى، ثم تطبيق ما يعرف بالربط بين المعادلات (Equating Linkage) (Tanguma, 2000).

ونظرا لاختلاف الجماعات الطلابية التي تقدم نماذج الاختبار في تصميم المفردات المشتركة، فإن جميع طرائق المعادلة في هذا التصميم تقوم على مبدأ تعديل الفروق بين أداء هذه الجماعات قبل إجراء معادلة درجات النماذج. ويتطلب القيام بهذا وضع افتراضات إحصائية قوية بسبب أن كل فرد في جماعة معينة يقدم نموذجا واحدا فقط. وتختلف طرائق المعادلة في نوعية وماهية هذه الافتراضات الإحصائية. ولكي تتم عملية معادلة درجات نماذج الاختبار، وضع براون وهولاند (Braun and Holland, 1982) أساسا لطريقة دمج جماعات الأفراد التي تقدم كل منها نموذجا واحدا من خلال افتراض أن جماعة طلابية واحدة قدمت كلا النموذجين المراد معادلة درجاتهما، أطلقا عليها مسمى "المجتمع المصطنع Synthetic Population"، وذلك من خلال جمع درجات جماعات الأفراد في نماذج الاختبار باستخدام أوزان نسبية مختلفة لكل جماعة (الأوزان النسبية تتراوح بين صفر وواحد، ومجموعها يساوي الواحد الصحيح).

يتوفر في تصميم المفردات المشتركة عدد من طرائق المعادلة التي يمكن تقسيمها إلى قسمين رئيسين: الطرائق الخطية (تفترض العلاقة الخطية بين درجات النموذجين)، والطرائق المئينية (تفترض العلاقة غير الخطية بين درجات النموذجين). وتختلف الطرق الإحصائية المتبعة لمعادَلة نماذج الاختبار في تصميم المفردات المشتركة حسب نظرية الاختبار التي تتبناها الجهة المسؤولة عن إعداد الاختبار وتقديمه. فهناك نظريتان أساسيتان في بناء الاختبارات ومعادلتها، هما: نظرية الاستجابة للمفردة والنظرية الكلاسيكية للاختبار (Kolen & Brennan, 2004؛ الدوسري، ٢٠٠١). وتقوم نظرية الاستجابة على المفردة على تفسير العلاقة بين السمة التي يقيسها الاختبار والاستجابة على كل مفردة من مفردات الاختبار من خلال دالة إحصائية متزايدة. وتتعدد النماذج الرياضية التي تتنبأ بهذه الدالة حسب عدد السمات المقاسة ونوعية الخصائص السيكومترية للمفردات. وبالتالي تتطلب طرق المعادَلة في نظرية الاستجابة للمفردة ملائمة البيانات لأحد نماذج الدالة الإحصائية قبل تطبيق طرق المعادَلة فيها. وفي حالة ملاءمة البيانات للنموذج الاحصائي المتبع في نظرية الاستجابة للمفردة، فتقدم هذه النظرية عددا من الطرق الإحصائية المناسبة لمعادلة نماذج

الاختبار. ومن بين هذه الطرق الإحصائية ما يعرف بمعادلة درجات القدرة (Ability Score Equating)، ومعادلة الدرجات الحقيقية (True Score Equating)، ومعادلة الدرجات المشاهدة (Observed Score Equating) (Baker & Al-Karni, 1991; Harris & Kolen, 1988; Lord, 1980).

بينما تتميز طرق المعادَلة في النظرية الكلاسيكية (أو طرق المعادلة الكلاسيكية) بعدم استخدام نماذج رياضية يلزم التأكد من ملاءمة بيانات الاختبار لافتراضاتها كما هو الحال في نماذج نظرية الاستجابة للمفردة. فتعتبر الطرق الإحصائية للمعادَلة في النظرية الكلاسيكية للاختبار أكثر قابلية للتطبيق. وأوضحت الدراسات تقارب نتائج المعادَلة وفقا للنظريتين إلى حد كبير (Wang, Lee, Brennan, & Kolen, 2008). فمثلا أظهرت دراسة جلواكي (Glowacki, 1991)، عند مقارنته بين الطرق الإحصائية في نظرية الاستجابة للمفردة والنظرية الكلاسيكية لمعادَلة اختبارات اللغة والقراءة والرياضيات في اختبار ولاية ألباما لتخريج طلبة المرحلة الثانوية بالولايات المتحدة الأمريكية، تقارب نتائج جميع الطرق المستخدمة في الدراسة. وأوصت الدراسة باستخدام طرق النظرية الكلاسيكية نظرا لسهولتها واستخدامها في بناء هذا الاختبار وعدم اختلاف نتائجها عن طرق نظرية الاستجابة للمفردة.

وتتمثل الطرائق الخطية باستخدام النظرية الكلاسيكية للاختبار في تصميم المفردات المشتركة في طريقة معادلة توكر (Tucker) (Gulliksen, 1950) للدرجات المشاهدة (تتمحور الافتراضات الإحصائية حول الدرجات المشاهدة)، وطريقة معادلة ليفين (Levin, 1955) للدرجات المشاهدة (تتمحور الافتراضات الإحصائية حول الدرجات المشاهدة)، وطريقة معادلة ليفين (Levine, 1955) للدرجات الحقيقية (تتمحور الافتراضات الإحصائية حول الدرجات الحقيقية)، وطريقة معادلة براون-هولاند (Braun & Holland, 1982) للدرجات المشاهدة (تتمحور الافتراضات الإحصائية حول التوزيعات التكرارية). تستخدم طرائق المعادلة الخطية المتوسط الحسابي والانحراف المعياري فقط للدرجات المشاهدة أو الحقيقية (حسب طريقة المعادلة) في النموذجين المراد معادلة درجاتهما في كل من الجماعتين اللتين قدمتا النموذجين والجماعة الافتراضية (المجتمع المصطنع)، وينتج منها تحويلا خطيا لدرجات النموذج الجديد إلى درجات النموذج المرجعي.

وتتمثل الطرائق غير الخطية باستخدام النظرية الكلاسيكية للاختبار في تصميم المفردات المشتركة في طريقة المعادلة المئينية بتقدير التكرارات Frequency Estimation Equipercentile Equating Angoff, 1971 (تتمحور الافتراضات الإحصائية حول التوزيعات التكرارية)، و طريقة المعادلة المئينية السلسلية Chained Equipercentile

Equating Angoff, 1971 (تتمحور الافتراضات الإحصائية حول التوزيعات التكرارية). وتستخدم طرائق المعادلة المئينية التوزيع التكراري لجميع الدرجات المشاهدة، وينتج منها تحويل غير خطي لدرجات النموذج الجديد إلى درجات النموذج المرجعي. وتتوفر طرق إحصائية لتمهيد التحويل غير الخطي الناتج من طرق المعادلة المئينية إما عن طريق تمهيد التوزيعات التكرارية لدرجات نماذج الاختبار (تسمى الطريقة المئينية المهدة قبليا) أو عن طريق تمهيد الدرجات المكافئة المتحصل عليها من المعادلة المئينية (تسمى الطريقة المئينية المهدة بعديا). ويمكن الرجوع إلى كولن وبرنن (Kolen and Brennan, 2004) لمعرفة المزيد حول الافتراضات الإحصائية والمعادلات الحسابية المستخدمة في كل طريقة من طرائق المعادلة المذكورة.

تختلف نتائج المعادلة بين طرائق المعادلة المستخدمة في كل تصميم من تصاميم جمع بيانات المعادلة (Chulu & Sireci, 2011). فهناك عدد من العوامل التي تؤثر على نتائج كل طريقة من طرائق المعادلة، مما يؤدي إلى الاختلاف فيما بينها، ومن بين هذه العوامل مقدار الاختلاف بين نماذج الاختبار في الإحصاءات الوصفية والتوزيعات التكرارية، وعدد أسئلة الاختبار ، وعدد المفردات المشتركة وخصائصها، وتوفر الافتراضات الإحصائية بهاويؤثر عدد المفردات المشتركة بين نماذج الاختبار المراد معادلة درجاتها وخصائصها السيكومترية بصفة خاصة على نتائج المعادلة، ويختلف هذا التأثير من طريقة إلى أخرى من طرائق المعادلة. قام ريكر وفون دافيير (Ricker & Von Davier, 2007) بدراسة للتعرف على تأثير طول المفردات المشتركة (الاختبار الرابط) على نتائج المعادلة في تصميم المفردات المشتركة بجماعات غير متكافئة باستخدام عدد من طرائق المعادلة، تضم فيما بينها الطريقة المئينية بتقدير التكرارات. وكطريقة محكية للمقارنة بين طرائق المعادلة، استخدمت الدراسة الطريقة المئينية في تصميم الجماعات المتكافئة لنماذج اختبار متكافئة، مكونة من استجابات ميدانية لمفردات من اختبار تقدمه مؤسسة خدمات الاختبارات التربوية Educational Testing Services بغرض إصدار تصاريح مهنية (Professional Licensure). أوضحت الدراسة أن مقدار التحيز (الفروق) بين نتائج كل طريقة من طرائق المعادلة وبين الطريقة المحكية المستخدمة في الدراسة تزداد كلما نقص عدد المفردات المشتركة. وكانت الطريقة المئينية بتقدير التكرارات من أقل طرائق المعادلة المستخدمة في الدراسة تأثرا بنقص عدد المفردات المشتركة.

كما توضح الدراسات تأثير حجم عينة الجماعات الطلابية على كل من تحيز طرائق المعادلة وخطأها المعياري في تصميم المفردات المشتركة، فأوضحت دراسة سوناسي (Sunnassee, 2011) ضرورة تجنب استخدام عينات صغيرة عند تطبيق تصميم المفردات المشتركة، وخاصة

عند الظروف التالية: (١) حجم العينة لأي جماعة تقدم أحد نماذج الاختبار أصغر من ٥٠ فردا، (٢) مقدار الفروق بين قدرات الجماعتين أكبر من ١,٠ انحراف معياري، (٣) اختلاف معامل الصعوبة لنماذج الاختبار أكبر من ٠,٢٥ انحراف معياري، (٤) متوسط معاملات التمييز لأي نموذج جديد أصغر من معاملات تمييز مفردات النموذج المرجعي، (٥) عدد مفردات النموذج الجديد صغير (٣٠ فأقل)، (٦) متوسط صغير لمعاملات التمييز لمفردات النموذج المرجعي. وأوضحت الدراسة أن طرائق المعادلة الخطية (توكر وليفين) حصلت على تحيز أقل من الطريقة المئينية بتقدير التكرارات في العينات الصغيرة، بينما حصلت على أخطاء معيارية للمعادلة أكبر من الطريقة المئينية بتقدير التكرارات.

كما قام كولن وبرنن (Kolen & Brennan, 2004) بدراسة لمقارنة طرائق المعادلة في تصميم المفردات المشتركة باستخدام نموذجين لاختبار، يتكون كل منهما من ٣٦ مفردة، من بينها ١٢ مفردة مشتركة (داخلية). تم وضع المفردات المشتركة في النموذجين في نفس المواقع في الورقة الامتحانية. تم تطبيق النموذج الجديد على عينة قوامها ١،٦٥٥ طالبا، والنموذج المرجعي على عينة قوامها ١،٦٣٨ طالبا. طبقت الدراسة خمس طرائق للمعادلة هي: توكر، وليفين (الدرجات المشاهدة) وبراون-هولاند، و الطريقة المئينية بتقدير التكرارات، والطريقة المئينية المهدة بعديا. أظهرت الدراسة العلاقة الخطية للدرجات المكافئة الناتجة من طريقة توكر وليفين وبراون-هولاند، مع تقارب الدرجات المكافئة الناتجة من طريقة توكر براون-هولاند واختلافهما عن طريقة ليفين. واعتبر الباحثان أن الدرجات المكافئة الناتجة من طريقة براون-هولاند أفضل نظرا لتوافق النتائج مع الافتراضات الإحصائية لطريقة براون-هولاند (العلاقة غير الخطية بين درجات المفردات المشتركة في عينتي الدراسة). كما أظهرت الدراسة العلاقة غير الخطية للدرجات المكافئة الناتجة من الطريقة المئينية بتقدير التكرارات واختلافها عن الدرجات المكافئة الناتجة من طريقة براون-هولاند وخاصة عن الدرجات الطرفية (المنخفضة والمرتفعة). كما أظهرت الطريقة المئينية المهدة بعديا علاقة غير خطية ومتقاربة مع الطريقة المئينية مع تعرج أقل لهذه العلاقة.

مشكلة الدراسة:

تحرص مؤسسات إعداد الاختبارات ومن بينها المركز الوطني للقياس والتقويم بالمملكة العربية السعودية على اتباع الخطوات العلمية في بناء نماذج متكافئة للاختبارات واسعة النطاق، وفي تطبيقها وتصحيحها ومعايرتها بالطرائق العلمية، والتي تزيد من تشابه التوزيعات

التكرارية والإحصاءات الوصفية (المتوسطات الحسابية والانحرافات المعيارية) للدرجات الخام والدرجات المعيارية للطلبة في النماذج المختلفة للاختبار. ويعد اختبار القدرات العامة (يضم اختبار القدرة اللفظية واختبار القدرة الكمية) الذي يقدمه المركز الوطني للقياس والتقويم من الاختبارات التي تلاقي اهتماما كبيرا من قبل الطلبة المتقدمين الذي يسعون للحصول على درجات عالية، تضمن لهم مقاعد دراسية بمؤسسات التعليم العالي، وكذلك من قبل مؤسسات التعليم العالي نفسها؛ لضمان انتقاء أفضل المدخلات لبرامجها الدراسية. وبرغم الحرص الشديد من القائمين على اختبار القدرات العامة على تطوير الاختبار وبنائه بأفضل الطرق السيكومترية الملائمة لضمان تحقيق العدالة والمساواة بين الطلبة المتقدمين للاختبار وتحقيق صدق الدرجات التي تستخدمها مؤسسات التعليم العالي، ألا أنه يحرص المسؤولون عن هذا الاختبار على تطوير الطرق المتبعة حاليا في بناء نماذج الاختبار المختلفة لتحقيق مزيداً من تكافؤ نماذج الاختبار.

كما يتبع المركز الوطني للقياس والتقويم سياسة توزيع الطلبة على أيام ونماذج الاختبار بطريقة تضمن عشوائية توزع الطلاب على نماذج الاختبار، مما يزيد من فرص تكافؤ الجماعات الطلابية التي تقدم النماذج المختلفة للاختبار. وبرغم ذلك، يدرك المسؤولون أن هناك عدداً من العوامل المرتبطة بتطبيق نماذج الاختبار على الطلبة في أيام وظروف مختلفة تجعل افتراض تكافؤ نماذج الاختبار غير مسلم به. ومن أمثلة هذه العوامل التباين في مستوى قدرات الطلبة الذين يقدمون نماذج الاختبار، والتباين في ظروف التطبيق، والتباين في استعداد الطلبة بين أيام التطبيق، والتباين في عدد الطلبة الذين قدموا الاختبار لأكثر من مرة سابقا، والتفاعل الذي يحدث بين الطلبة ومفردات كل نموذج من نماذج الاختبار. وقد تؤدي هذه العوامل إلى وجود اختلافات في الإحصاءات الوصفية (المتوسطات الحسابية والتباين) وشكل التوزيعات التكرارية في بيانات نماذج الاختبار.

وحيث إن مبدأ استخدام تصميم المفردات المشتركة في معادلة نماذج الاختبار يقوم على تعديل الاختلافات المحتملة بين أداء الجماعات الطلابية التي تقدم نماذج الاختبار، فإنه من الجدير تجريب هذا التصميم في معادلة نماذج اختبار القدرات العامة، ومقارنة نتائج المعادلة باستخدام الطرائق الإحصائية المتوفرة فيه. ويتوافق هذا التوجه مع التوصية التي قدمها كولن وبرنن (Kolen and Brennan, 2004) حول عملية انتقاء تصميم المعادلة المناسب، حيث أكدا على صعوبة تفضيل تصميماً معيناً على آخر ، وأن السياق المحيط بالاختبار هو الموجه لانتقاء التصميم الملائم (Kolen & Brennan, 2004, p. 298). واقترحا ضرورة

إجراء دراسات ميدانية على معادلة الاختبار باستخدام تصاميم مختلفة وبطرائق معادلة مختلفة؛ للوقوف بشكل عملي على مقدار تشابه أو اختلاف نتائج المعادلة في الطرائق المتعددة في التصاميم المختلفة، ومناقشة ذلك مع الجوانب العملية المرتبطة بسياق الاختبار.

وتختلف مؤسسات إعداد الاختبارات في تصاميم وطرق معادَلة نماذج اختباراتها. فتستخدم اختبارات الكليات الأمريكية ACT (American College Test) مثلا طريقة المعادَلة المئينية في تصميم الجماعات المتكافئة، بينما يستخدم اختبار التقييم الدراسي Scholastic SAT (Assessment Test) طريقة المعادَلة المئينية في تصميم المفردات المشتركة. وتستخدم اختبارات التنمية التربوية العامة GED (General Educational Development) الذي يقدمه المجلس الأمريكي في التربية American Council on Education طريقة المعادَلة المئينية المهدة قبليا في تصميم الجماعات المتكافئة لمعادَلة نماذج الاختبارات في العام الواحد وعبر الأعوام بتقديم النموذج المرجعي المعد في عام ٢٠٠٢م مع النماذج الجديدة (Ezzelle & Setzer, 2009).

تهدف هذه الدراسة إلى التحقق من تكافؤ الجماعات الطلابية التي قدمت نماذج اختباري القدرة اللفظية والقدرة الكمية في اختبار القدرات العامة من حيث تساوي متوسطاتها الحسابية، وتباين الدرجات، وتشابه توزيعاتها التكرارية. وتسعى الدراسة بعد ذلك إلى تطبيق عدداً من طرق المعادلة الكلاسيكية في تصميم المفردات المشتركة على نماذج الاختبار، ومقارنتها في الاختبارين. واقتصرت الدراسة الحالية على تطبيق طرق المعادلة الكلاسيكية؛ لما تتميز بها هذه الطرق من سهولة في تطبيقها، وسهولة تحقق افتراضاتها الإحصائية. كما دعمت الدراسات المذكورة سابقا استخدام طرق المعادلة الكلاسيكية وتحقيقها لنتائج معادلة تقترب من نتائج طرق المعادلة باستخدام نماذج نظرية الاستجابة للمفردة، وفي بعض الدراسات تتفوق طرق المعادلة الكلاسيكية على نظيراتها في نظرية الاستجابة للمفردة (Kolen & Whitney, 1982; Slinde & Linn, 1997). كما يتم حساب الدرجات الخام والدرجات المعيارية في اختبار القدرات العامة باستخدام النظرية الكلاسيكية، مما يتوافق مع توظيف الطرق الإحصائية في النظرية الكلاسيكية لإجراء معادلة نماذج الاختبار. وتتلخص أهداف الدراسة في المحاولة للإجابة عن الأسئلة التالية:

١. هل تختلف الإحصاءات الوصفية لدرجات النماذج الثلاثة لاختباري القدرة اللفظية والقدرة الكمية قبل وبعد المعادلة؟

٢. ما مقدار الفروق في الدرجات المكافئة (الخام والمعيارية) للنموذج (٠٨٢) في اختباري

القدرة اللفظية والقدرة الكمية حسب طرائق المعادلة الكلاسيكية في تصميم المفردات المشتركة؟

٣. هل تتشابه نتائج معادلة درجات النموذج (٠٠٨٣) مع نتائج معادلة درجات النموذج (٠٠٨٢) في اختباري القدرة اللفظية والقدرة الكمية كمؤشر على صدق نتائج المعادلة عبر النماذج؟

٤. ما مقادير الأخطاء المعيارية للمعادلة في طرائق المعادلة الكلاسيكية للنموذجين (٠٠٨٢، ٠٠٨٣) في اختباري القدرة اللفظية والقدرة الكمية؟

أهمية الدراسة:

تكمن أهمية هذه الدراسة في مقارنة عدد من الطرائق الإحصائية لإجراء معادلة نماذج اختبار القدرات العامة في تصميم المفردات المشتركة، مما يساعد في اتخاذ القرار المناسب في تبني الطريقة الإحصائية الأكثر ملاءمة لإجراء معادلة النماذج. كذلك برغم انتشار تصميم المفردات المشتركة في الأدبيات المتخصصة، إلا أنه يندر استخدامه في الاختبارات بالمنطقة العربية على حد علم الباحث. كما تكتسب هذه الدراسة أهميتها من القيمة التي يمثلها اختبار القدرات العامة من حيث عدد المتقدمين وخطورة القرارات المبنية على نتائجه. كما يتوقع أن تسهم هذه الدراسة في نشر ثقافة معادلة نماذج الاختبارات في البيئة العربية.

مصطلحات الدراسة:

معادَلة نماذج الاختبار (Test Forms Equating): تحويل درجات أحد نماذج الاختبار إلى ما يناظرها من درجات النموذج المرجعي للاختبار، بحيث تصبح القياسات المستمدة من درجات كل من النموذجين متكافئة. وفي هذه الدراسة تم حساب الدرجات الخام لنماذج الاختبار (مجموع الإجابات الصحيحة على فقرات نموذج الاختبار)، والدرجات المعيارية المقابلة لها (تحويل خطي للدرجات الخام بحيث يصبح متوسطها الحسابي ٦٥ وانحرافها المعياري ١٠).

نماذج الاختبار المتكافئة (Equivalent Test Forms): نماذج الاختبار التي يتم بناؤها بحيث تحتوي على مفردات مختلفة وفي ذات الوقت تقيس سمة واحدة وتتبع جدول مواصفات موحد، وتتساوى متوسطاتها الحسابية وتباين درجاتها، وتتشابه توزيعاتها التكرارية إذا طبقت على جماعات طلابية متكافئة في السمة المقيسة. وفي هذه الدراسة، تم التحقق من أن نماذج اختبار القدرات العامة متكافئة من حيث المتوسط الحسابي والتباين والالتواء والتفلطح.

طرق المعادَلة الكلاسيكية (Classical Theory Equating Methods): طرق المعادلة التي تستخدمها النظرية الكلاسيكية للاختبار، والتي تتميز ببساطة وسهولة تحقق افتراضاتها، والتي تعتمد على مجموع درجات المفردات للطلاب. وتقدم النظرية الكلاسيكية طرقا إحصائية متعددة لمعادلة نماذج الاختبار في تصميم المفردات المشتركة. وتتمثل طرق المعادلة الكلاسيكية المستخدمة في هذه الدراسة في المعادلات الخطية لتوكر وليفين وبراون-هولاند والمعادلة المئينية بطريقة تقدير التكرارات والمعادلة المئينية المهدة (بعديا).

المفردات المشتركة الداخلية أو الاختبار الرابط (Internal Common Items or Anchor Test): مجموعة من المفردات التي تربط نماذج الاختبار المراد معادلة درجاتها، بحيث تتكرر في نماذج الاختبار ويتوفر بها مجموعة من الشروط من بينها: تمثيلها لجدول مواصفات الاختبار الكلي (اختبار مصغر للاختبار الكلي)، وتشابه خصائصها السيكومترية مع الاختبار الكلي، وملائمة عددها مع طول الاختبار، وتشابه مواقعها في الورقة الامتحانية لنماذج الاختبار، وغيرها.

اختبار القدرات العامة: يهدف اختبار القدرات العامة إلى قياس الاستعداد الدراسي للقبول بمؤسسات التعليم العالي بالمملكة العربية السعودية، ويسعى إلى تقديم معلومات صادقة عن قدرة الطالب على التعامل مع المهارات والعمليات التي تتطلبها البرامج الدراسية بمؤسسات التعليم العالي. ويقيس اختبار القدرات العامة قدرتين أساسيتين للنجاح في التعليم العالي وهي القدرة اللفظية والقدرة الكمية. ويهدف قياس القدرة اللفظية في الاختبار إلى إعطاء معلومات عن مستوى الطالب ومقدار تمكنه من فهم محتوى ومعنى ما يقرؤه وقدرته على استنباط الأفكار من النصوص والربط بينها والوصول إلى تعميمات صحيحة تستند إلى أدلة وشواهد. ويهدف قياس القدرة الكمية في الاختبار إلى تقديم معلومات عن مستوى الطالب ومقدار تمكنه من فهم المفاهيم الرياضية والعمليات الحسابية المختلفة والجبر والهندسة والعلاقات الرياضية والقدرة على الوصول إلى تعميمات رياضية صحيحة تستند إلى أدلة وشواهد.

حدود الدراسة:

١. اقتصرت الدراسة على ثلاثة نماذج لكل من اختبار القدرة اللفظية واختبار القدرة العددية في اختبار القدرات العامة الذي يعده ويقدمه المركز الوطني للقياس والتقويم في التعليم العالي بالمملكة العربية السعودية.

٢. اقتصرت الدراسة على عينة الدراسة التي قدمت تطبيقات النماذج الثلاثة لكل من اختبار القدرة اللفظية واختبار القدرة العددية في العام الأكاديمي ١٤٣٠/١٤٣١هـ.

٣. اقتصرت الدراسة على استخدام مفردات مشتركة داخلية.
٤. اقتصرت الدراسة على طرق المعادلة الإحصائية في النظرية الكلاسيكية للقياس.

الطريقة والإجراءات:

المجتمع والعينة

تم الحصول على بيانات نماذج اختبار القدرات العامة من المركز الوطني للقياس والتقويم بالمملكة العربية السعودية. وتكونت عينة الدراسة من طلبة المرحلة الثانوية (الصفين الثاني الثانوي والثالث الثانوي) من مختلف مناطق المملكة العربية السعودية في الفترة الثانية بنهاية العام الأكاديمي ١٤٣٠/١٤٣١هـ ممن طبقت عليهم نماذج الاختبار المراد معادلتها في أيام تطبيق تصميم المفردات المشتركة، على النحو الذي يبينه الجدول (١).

جدول (١)
توزيع عينات الدراسة على نماذج الاختبار وأيام التطبيق

يوم التطبيق	رقم النموذج	العينة
الخميس	١٠٨١	٢٣٠٠
السبت	٠٠٨٢	٢٣٢٢
الخميس التالي	٠٠٨٣	٢٣٢٦

أدوات الدراسة

تم إعداد نماذج اختبار القدرات العامة من قبل مختصين بالمركز الوطني للقياس والتقويم بالمملكة العربية السعودية. وتم الحصول على بيانات لثلاثة نماذج لكل من اختبار القدرة اللفظية واختبار القدرة الكمية في اختبار القدرات العامة المطبقة سلفا بواسطة المركز الوطني للقياس والتقويم. وتحمل هذه النماذج الأرقام التالية للتمييز بينها، وهي: (١٠٨١)، (٠٠٨١)، (٠٠٨٢). وتتكون جميع نماذج اختبار القدرة اللفظية من ٦٨ مفردة من نوع الاختيار من متعدد (لكل منها أربعة بدائل) موزعة على أربعة مكونات تتمثل في: (١) معاني الكلمات، ويشكل ما وزنه ٢٠٪ تقريبا؛ (٢) إكمال الجمل، ويشكل ما وزنه ٢٣٪ تقريبا؛ (٣) التناظر اللفظي، ويشكل ما وزنه ٢٦٪ تقريبا؛ (٤) استيعاب المقروء، ويشكل ما وزنه ٣١٪ تقريبا. وتتكون جميع نماذج اختبار القدرة الكمية من ٥٢ مفردة من نوع الاختيار من متعدد (لكل منها أربعة بدائل)، تقيس مختلف جوانب القدرة الكمية مثل العمليات الحسابية والجبر والهندسة العامة وهندسة المثلثات وغيرها. حُسب معامل ثبات نماذج الاختبارات، وتراوح معامل ثبات ألفا لكرونباخ لجميع نماذج اختبار القدرة اللفظية والقدرة الكمية، بين ٠,٨٦٨ و ٠,٨٨٧، مما يشير إلى ثبات عالٍ ومتقارب.

وتتراوح الدرجة الخام (عدد الإجابات الصحيحة في أسئلة كل نموذج) لاختبار القدرة اللفظية بين الدرجة صفر والدرجة ٦٨، بينما تتراوح الدرجة الخام لاختبار القدرة الكمية بين الدرجة صفر والدرجة ٥٢. ولتفسير درجات الطلاب معياريا، تحسب درجة معيارية لكل نموذج من خلال حساب الدرجة الزائية (باستخدام المتوسط الحسابي والانحراف المعياري لمجموعة الطلبة الذين طبقوا كل نموذج)، وتحويلها خطيا لتأخذ متوسطا حسابيا قدره ٦٥، وانحرافا معياريا قدره ١٠، بحيث تتراوح بين صفر و١٠٠. وفي حالة تخطي الدرجة المعيارية لهذين الحدين، يتم تثبيتها عن ذلك الحد.

يمثل النموذج (١٠٨١) النموذج المرجعي للمعادلة، نظرا لتطبيقه كأول نموذج في أيام تطبيق نماذج اختبار القدرات العامة في الفترة الثانية بنهاية العام الأكاديمي ١٤٣٠/١٤٣١هـ. ويمثل النموذجان (٠٠٨٢)، (٠٠٨٣) النموذجين الجديدين المطلوب معادلة درجات كل منهما مع درجات النموذج المرجعي (١٠٨١). أُعد النموذج المرجعي (١٠٨١) بالطريقة الاعتيادية التي يتبعها المركز، ليتكون من ٦٨ مفردة في اختبار القدرة اللفظية، و٥٢ مفردة في اختبار القدرة الكمية وموزعة على مكونات الاختبارين بنفس الأوزان المذكورة سابقا. بينما أُعد النموذجان (٠٠٨٢)، (٠٠٨٣) بطريقة مختلفة تتناسب مع تصميم المفردات المشتركة، بحيث يحتوي كل نموذج من النموذجين الجديدين على مجموعتين من المفردات: مفردات خاصة بالنموذج، ومفردات مشتركة مع النموذج المرجعي. ففي اختبار القدرة اللفظية مثلا، تكون كل نموذج من ٤٨ مفردة خاصة بالنموذج، و٢٠ مفردة مشتركة مع النموذج المرجعي (وتشكل ما نسبته ٣٠٪ تقريبا من العدد الإجمالي للمفردات). وأعدت المفردات الخاصة بالنموذج (٤٨ مفردة) لتقيس مكونات كل قدرة بنفس الأوزان في الاختبار المرجعي. أما المفردات المشتركة فتم انتقاؤها من النموذج المرجعي بحيث تمثل اختبارا مصغرا (أي تقيس مكونات القدرة المقاسة بنفس الأوزان، ولها الخصائص السيكومترية ذاتها للاختبار الكلي). وتحقق هذه الطريقة في إعداد النماذج الجديدة تكافؤ الصياغة للنموذج الجديد مع النموذج المرجعي مع وجود مفردات مشتركة تعمل دور الرابط بين النموذجين. وتم تطبيق الطريقة ذاتها في تكوين النماذج الجديدة في اختبار القدرة الكمية، حيث كان عدد المفردات الخاصة بكل نموذج جديد ٣٧ مفردة، وعدد المفردات المشتركة ١٥ مع النموذج المرجعي (وتشكل ما نسبته ٣٠٪ تقريبا من العدد الاجمالي لمفردات القدرة الكمية).

ويجدر التنويه إلى أن المفردات المشتركة في نماذج كل اختبار (القدرة اللفظية والكمية)، والتي تم انتقاؤها من النموذج المرجعي متشابهة بين النماذج الثلاثة. وتحتوي الورقة الامتحانية

لكل نموذج جديد على المفردات الخاصة بالنموذج والمفردات المشتركة بشكل مختلط. ولزيادة تكافؤ النماذج، روعي أن تظهر المفردات المشتركة في الورقة الامتحانية للنموذجين الجديدين في المواقع ذاتها للنموذج المرجعي.

الإجراءات

طُبقت النماذج الثلاثة (١٠٨١، ٠٠٨٢، ٠٠٨٣) لاختبار القدرات العامة في أيام مختلفة، حيث لا يتطلب تصميم المفردات المشتركة تطبيق النماذج في نفس التوقيت. وفي كل يوم من أيام التطبيق، اختيرت عشرة مراكز لتطبيق الاختبار، موزعة على مختلف مناطق المملكة العربية السعودية للحصول على أعلى درجة من تمثيل مجتمع الدراسة. وحيث أن تصميم المفردات المشتركة لا يشترط تكافؤ وتشابه الجماعات الطلابية التي تقدم نماذج الاختبار، لم يراعى كثيرا تكافؤ الجماعات الطلابية في هذه الدراسة أثناء توزيع نماذج الاختبارات على الطلاب في مراكز التطبيق؛ وذلك من أجل زيادة درجة التشابه بين إجراءات الدراسة والتطبيق الميداني الفعلي لنماذج الاختبارات باستخدام تصميم المفردات المشتركة.

استخدمت خمس طرائق إحصائية للمعادلة الكلاسيكية لإجراء معادلة درجات كل من النموذجين الجديدين مع النموذج المرجعي تتوافق مع تصميم المفردات المشتركة بجماعات غير متكافئة. وتم إجراء المعادلة باستخدام برنامج (Common Item Program for Equating CIPE) من إعداد كولن (Kolen, 2004). ومن بين طرائق المعادلة الخمس المستخدمة في الدراسة ثلاث طرائق للمعادلة الخطية، وهي: طريقة توكر للدرجات المشاهدة (Tucker)، وطريقة ليفين للدرجات المشاهدة (Levine) وطريقة براون-هولاند (Braun-Holland)، وطريقتان للمعادلة غير الخطية، وهما: المعادلة المئينية بطريقة التكرارات (Frequency Estimation Equipercentile Equating)، والمعادلة المئينية المهدة بعديا (Postsmoothed Equipercentile Equating). ويضم كتاب كولن وبرنن (Kolen & Brennan,2004) الافتراضات الإحصائية والمعادلات الحسابية لكل طريقة من طرائق المعادلة الخمس.

وفي المعادلة المئينية المهدة بعديا، تم تجريب عدة درجات وطرائق لتمهيد المعادلة المئينية المتوفرة في البرنامج، وأظهرت الدرجة ٠,١ في طريقة «الدالة التكعيبية Cubic Spline" تمهيدا ملائما للمعادلة المئينية في جميع النماذج، مع عدم ابتعاد المعادلة المئينية المهدة بعديا عن المعادلة المئينية الأصلية. وبالتالي اتبعت هذه الدرجة من التمهيد في جميع المعادلات في هذه الدراسة. ويجدر التنويه إلى أنه لا يشترط اتباع نفس درجة التمهيد للمعادلة المئينية بين

النماذج، ولكن يعتمد ذلك على التوزيع التكراري لدرجات نماذج الاختبار، وطبيعة الدرجات المكافئة المتحصل عليها.

بعد الحصول على الدرجات الخام المكافئة لدرجات النموذج الجديد مع النموذج المرجعي من خلال برنامج CIPE، تم تحويلها إلى درجات معيارية مكافئة باستخدام الدرجات المعيارية للنموذج المرجعي. وفي حالة كون الدرجة الخام المكافئة عددا صحيحا، فإن الدرجة المعيارية المكافئة هي الدرجة المعيارية لتلك الدرجة الخام حسب النموذج المرجعي. أما إذا كانت الدرجة الخام المكافئة عددا غير صحيح (بها أرقام عشرية)، لا يتم تقريبها إلى عدد صحيح)، بل يتم استخدام معادلة خاصة للحصول على الدرجة المعيارية المكافئة حسب النموذج المرجعي (انظر Kolen & Brennan, 2004, p.58).

المعالجات الإحصائية:

وللإجابة عن أسئلة الدراسة، أجريت العديد من التحليلات الإحصائية. فتمت الإجابة عن السؤال الأول للدراسة من خلال حساب المتوسطات الحسابية والانحرافات المعيارية والتوزيعات التكرارية للدرجات الخام لجميع المفردات وللمفردات المشتركة في جميع نماذج الاختبارين في الدراسة. وتمت الإجابة عن السؤال الثاني للدراسة من خلال حساب الفروق بين الدرجات المكافئة (الخام والمعيارية) للنموذج (٠٠٨٢) الناتجة من إجراء المعادلة (بطرائق المعادلة الخمس المستخدمة) والدرجات المكافئة الناتجة من افتراض تكافؤ درجات النموذجين (الجديد والمرجعي). وتفترض الطريقة الأخيرة تساوي الدرجات الخام للنموذجين المراد معادلتهما، وأن درجات الطلاب في النموذجين تتبع نفس تدريج الدرجات، ويمكن استخدام الدرجات المعيارية للنموذج المرجعي لتفسير درجات الطلاب في النموذج الجديد بشكل تلقائي. ويسمي كولن وبرنن (Kolen & Brennan, 2004) هذه المعادلة بالمعادلة المتطابقة Identity Equating. وتقدم هذه الطريقة محكا مشتركا بين جميع طرائق المعادلة ومستقلا عن أي منها، ويساعد في معرفة مقدار الإضافة التي تقدمها عملية المعادلة الناتجة من كل طريقة، مع توفير أرضية للمفاضلة بين طرائق المعادلة الخمس المستخدمة في الدراسة، بحيث لا تتحيز لأي طريقة من طرائق المعادلة. كما يساعد حساب الفروق بهذه الطريقة في معرفة مقدار اختلاف الدرجات المكافئة للنموذج الجديد ودرجاته الأصلية. وحُسبت مقادير الفروق للدرجات المعيارية المباشرة لكل نموذج (حسب الطريقة المتبعة حاليا بالمركز، وهي الدرجات المعيارية المحسوبة باستخدام المتوسط الحسابي والانحراف المعياري لكل نموذج)؛ وذلك من أجل المقارنة مع الدرجات المعيارية المكافئة المتحصل عليها من طرائق المعادلات الخمس.

واستخدمت العملية الحسابية التالية لحساب الفرق لكل درجة (خام أو معيارية) من درجات التدريج:

الفرق لكل درجة = الدرجات المكافئة لكل درجة في النموذج الجديد – الدرجة المكافئة من المعادَلة المتطابقة

ولتلخيص فروقات الدرجات المكافئة (الخام أو المعيارية) عبر درجات الاختبار، حُسب المتوسط غير الموزون للفروق (Unweighted Mean of Signed Difference):

المتوسط غير الموزون للفروق = مجـ (الفرق لكل درجة) / (عدد الأسئلة +١) ،

والمتوسط الموزون للفروق (حسب تكرار الدرجة في النموذج الجديد) (Weighted Mean of Signed Difference):

المتوسط الموزون للفروق = مجـ (الفرق لكل درجة X تكرار الدرجة) / (حجم العينة)

حيث تم حساب المجموع عبر جميع درجات تدريج كل اختبار.

وتمت الإجابة عن السؤال الثالث للدراسة من خلال إجراء نفس المعالجات المستخدمة للإجابة عن السؤال الثاني ولكن بالتطبيق على النموذج (٠٠٨٣) بغرض التأكد من صدق النتائج المتحصل عليها عبر نماذج الاختبار. ويجدر الذكر أنه يمكن تبادل أدوار النموذجين (٠٠٨٢) و(٠٠٨٣) من دون أن يكون له أثر على نتائج الدراسة.

وللإجابة عن السؤال الرابع للدراسة، حُسب الخطأ المعياري للمعادلة (Standard Error of Equating) لكل درجة خام مكافئة كمؤشر للخطأ العشوائي في المعادلة، والذي ينتج من اختلاف العينات. وحُسب الخطأ المعياري للمعادلة باستخدام معادلات تقريبية، تم اشتقاقها من قبل باحثين بطريقة Delta للدرجات المكافئة الناتجة من طرائق المعادلة. وتستخدم طريقة (Delta Kendall & Stuart, 1977) لاشتقاق دالة تقريبية للخطأ المعياري لمؤشر إحصائي مُعرف بدالة تربط بين عدد من المؤشرات الإحصائية الأخرى والتي يكون خطؤها المعياري محددا. وقدم براون وهولاند (Braun & Holland, 1982) معادلة حسابية بطريقة دلتا (Delta Kendall & Stuart, 1977) لتقدير الخطأ المعياري للمعادلة الخطية، وقدم لورد (Lord, 1982) معادلة حسابية بطريقة دلتا Delta لتقدير الخطأ المعياري للمعادلة المئينية. وقد حُسب الخطأ المعياري في هذه الدراسة لكل من طريقة توكر وليفين والطريقة المئينية غير المهدة فقط؛ بينما لم يُحسب لطريقة براون–هولاند والطريقة المئينية المهدة بعديا لعدم وجود معادلات تعطي مقدار خطأها المعياري. كذلك لم تُحسب الأخطاء المعيارية للمعادلة للدرجات المعيارية المكافئة لعدم وجود معادلات لها.

بعدما حُسب الخطأ المعياري لكل درجة خام مكافئة، حُسبت المتوسطات الموزونة لهذه الأخطاء المعيارية للمعادلة عبر جميع درجات الاختبار للحكم الكلي على الدرجات المكافئة المتحصل عليها من كل طريقة من طرائق المعادلة المستخدمة باستخدام المعادلة الآتية:
المتوسط الموزون للخطأ= مجـ (مربع الخطأ المعياري للمعادلة لكل درجة X تكرار الدرجة) / (حجم العينة)

نتائج الدراسة:

الإحصاءات الوصفية:

هل تختلف الإحصاءات الوصفية لدرجات النماذج الثلاثة لاختباري القدرة اللفظية والقدرة الكمية قبل وبعد المعادلة؟

١) الإحصاءات الوصفية قبل المعادلة:

يعرض الجدول (٢) الإحصاءات الوصفية للنماذج الثلاثة لاختبار القدرة اللفظية (١٠٨١)، (٠٠٨٢)، (٠٠٨٣). وتظهر النتائج اختلافا في المتوسطات الحسابية لدرجات النموذجين الجديدين عن درجات النموذج المرجعي؛ بينما تتشابه انحرافاتها المعيارية لحد كبير. فكانت المتوسطات الحسابية لدرجات النماذج الثلاثة كالتالي: ٣٣,٤٤٦ للنموذج (١٠٨١)، و٣٥,٤٨٦ للنموذج (٠٠٨٢)، و ٣٤,٤٧١ للنموذج (٠٠٨٣). ويبين اختبار ت لمجموعتين مستقلتين دلالة الفروق بين المتوسطات الحسابية لكل نموذجين (الجديد والمرجعي). كما لا تتشابه الخصائص الأخرى للتوزيعات التكرارية لدرجات الطلبة في النماذج الثلاثة. فعلى سبيل المثال، يميل التوزيع التكراري للنموذجين (٠٠٨٢) و (١٠٨١) للالتواء جهة اليمين مقارنة بالنموذج (٠٠٨٣). وعند تفحص الإحصاءات الوصفية لدرجات المفردات المشتركة (٢٠مفردة) بين النماذج الثلاثة، يلاحظ عدم تساوي خصائص توزيعاتها التكرارية، فمثلا تختلف متوسطاتها الحسابية لحد ما، حيث كانت كالتالي: ٨,٨٢٤ للنموذج (١٠٨١)، ٨,٤٨٦ للنموذج (٠٠٨٢)، ٨,٨٩١ للنموذج (٠٠٨٣). ويبين اختبار ت لمجموعتين مستقلتين أيضا دلالة الفروق بين المتوسطات الحسابية للنموذجين (٠٠٨٢) و(١٠٨١). وحيث إن المفردات المشتركة تعطي أرضية ملائمة لمقارنة أداء الجماعات الطلابية التي قدمت نماذج الاختبار، فإن هذه المتوسطات تشير إلى اختلاف في مستوى القدرة اللفظية لدى هذه الجماعات الطلابية. وتدل الاختلافات في المتوسطات الحسابية للدرجات الكلية ولدرجات المفردات المشتركة في النماذج الثلاثة على أن هناك اختلاف في صعوبة النماذج، زيادة على الاختلاف في مستوى القدرة اللفظية لدى الجماعات الطلابية، مما يدل على الحاجة إلى إجراء معادلة لنماذج القدرة اللفظية.

جدول (٢)
الإحصاءات الوصفية للنماذج الثلاثة في اختبار القدرة اللفظية

اليوم	رقم النموذج	المتوسط الحسابي	الانحراف المعياري	الالتواء	التفلطح
الخميس	١٠٨١	٣٣,٤٤٦	١٠,٩٧٣	٠,١٤٥	٢,٢٠٣
السبت	٠٠٨٢	*٣٥,٤٨٦	١٠,٢٤٣	٠,١٣٧	٢,٣٠٢
الخميس التالي	٠٠٨٣	*٣٤,٤٧١	١٠,٤٤٩	-٠,٠٠٣	٢,١٩٢
المفردات المشتركة في ١٠٨١		٨,٨٢٤	٣,٦٤٣	٠,٢٧٨	٢,٣١٠
المفردات المشتركة في ٠٠٨٢		*٨,٤٨٦	٣,٥٦٤	٠,٢٥٨	٢,٤٧١
المفردات المشتركة في ٠٠٨٣		٨,٨٩١	٣,٥٣٥	٠,٠٩١	٢,٣١٠

*دالة عند α = ٠,٠٥

يعرض الجدول (٣) الإحصاءات الوصفية للنماذج الثلاثة لاختبار القدرة الكمية. وتظهر النتائج اختلافا في المتوسطات الحسابية والانحرافات المعيارية لدرجات النموذجين الجديدين عن درجات النموذج المرجعي. فكانت المتوسطات الحسابية لدرجات النماذج الثلاثة كالتالي: ٢٢،٧٨٠ للنموذج (١٠٨١)، ٢٠،٧٧٠ للنموذج (٠٠٨٢)، ١٨،٤٥٦ للنموذج (٠٠٨٣)، مما يشير إلى أن نماذج الاختبار تختلف فيما بينها في متوسطاتها الحسابية. كما لا تتشابه الخصائص الأخرى للتوزيعات التكرارية لدرجات الطلبة في النماذج الثلاثة. وتظهر الإحصاءات الوصفية لدرجات المفردات المشتركة (١٥مفردة) في النماذج الثلاثة تقارباً في خصائص توزيعاتها التكرارية مقارنة بالدرجات الكلية للنماذج، وتدل الاختلافات في المتوسطات الحسابية للدرجات الكلية ولدرجات المفردات المشتركة في النماذج الثلاثة على الحاجة إلى إجراء معادلة لنماذج القدرة الكمية أيضا.

جدول (٣)
الإحصاءات الوصفية للنماذج الثلاثة في اختبار القدرة الكمية

اليوم	رقم النموذج	المتوسط	الانحراف المعياري	الالتواء	التفلطح
الخميس	١٠٨١	٢٢,٧٨٠	٨,٠٤٧	٠,٧٤١	٣,٣٣٩
السبت	٠٠٨٢	*٢٠,٧٧٠	٧,٤٥٢	١,٠٦١	٤,٠٤٥
الخميس التالي	٠٠٨٣	*١٨,٤٥٦	٦,٠٧٤	١,٠٦٣	٥,٠٢٠
المفردات المشتركة في ١٠٨١		٦,٥٤٨	٢,٤٧٥	٠,٢١٧	٢,٨٩٦
المفردات المشتركة في ٠٠٨٢		*٦,٣٤٨	٢,٤١٠	٠,٣٨٢	٣,٠٤٠
المفردات المشتركة في ٠٠٨٣		٦,٥٥٩	٢,٤٩١	٠,١١٧	٢,٨٤٥

*دالة عند α = ٠,٠٥

٢) الإحصاءات الوصفية بعد المعادلة:

يظهر الجدول (٤) والجدول (٥) الإحصاءات الوصفية للدرجات الخام المكافئة للنموذج (٠٠٨٢) فقط، الناتجة من إجراء المعادلة بالطرائق الخمس في اختبار القدرة اللفظية واختبار القدرة الكمية، على التوالي. ويظهر الجدولان أن المتوسطات الحسابية والانحرافات المعيارية للدرجات الخام المكافئة الناتجة من طرائق المعادلة الخمس متقاربة من بعضها البعض. كذلك تقترب هذه المتوسطات الحسابية والانحرافات المعيارية من مثيلاتها للدرجات الخام للنموذج المرجعي. كما يلاحظ أن قيم الالتواء والتفلطح للدرجات الخام المكافئة للنموذج الجديد الناتجة من طرائق المعادلة الخطية الثلاث لم تتغير مقارنة لما كانت عليه قبل إجراء المعادلة، بينما اختلفت عند استخدام طرائق المعادلة المئينية. وتتفق هذه النتائج مع طبيعة طرائق المعادلة، حيث إن الطرائق الخطية تؤثر على الإحصاءات الوصفية الأولى والثانية (المتوسطات والانحرافات المعيارية) فقط مع عدم إحداث تغييراً في شكل التوزيع التكراري للدرجات؛ بينما تؤثر الطرائق غير الخطية (المئينية غير المهدة والمهدة) على جميع الإحصاءات الوصفية وشكل التوزيع التكراري للدرجات (Chen, Livingston, & Holland, 2011).

جدول (٤)
الإحصاءات الوصفية للدرجات الخام المكافئة في النموذج الجديد (٠٠٨٢) لاختبار القدرة اللفظية

الدرجات		المتوسط الحسابي	الانحراف المعياري	الالتواء	التفلطح
الدرجات المكافئة	الخطية (Tucker)	٣٢,٥٤٧	١٠,٧٨٨	٠,١٣٧	٢,٣٠٢
	الخطية (Levine)	٣٢,٢٩٦	١٠,٦٦٩	٠,١٣٧	٢,٣٠٢
	الخطية (Braun-Holland)	٣٢,٥٨٢	١٠,٧٣٤	٠,١٣٧	٢,٣٠٢
	المئينية	٣٢,٥٨٢	١٠,٧٣٤	٠,١٩٧	٢,٢٦٧
	المئينية المهدة بعديا	٣٢,٥٨٢	١٠,٧٣٩	٠,٢٠٣	٢,٢٨٠
الدرجات قبل المعادلة	النموذج (٠٠٨٢)	٣٥,٤٨٦	١٠,٢٤٣	٠,١٣٧	٢,٣٠٢
	النموذج (١٠٨١)	٣٣,٤٤٦	١٠,٩٧٣	٠,١٤٥	٢,٢٠٣

جدول (٥)
الإحصاءات الوصفية للدرجات الخام المكافئة في النموذج الجديد (٠٠٨٢) لاختبار القدرة الكمية

الدرجات		المتوسط الحسابي	الانحراف المعياري	الالتواء	التفلطح
الدرجات المكافئة	الخطية (Tucker)	٢٢,٢٦١	٧,٩١٢	١,٠٦١	٤,٠٤٥
	الخطية (Levine)	٢١,٩٦٦	٧,٧١٢	١,٠٦١	٤,٠٤٥
	الخطية (Braun-Holland)	٢٢,٢٠٠	٧,٨٦٤	١,٠٦١	٤,٠٤٥

تابع جدول (٥)

الدرجات		المتوسط الحسابي	الانحراف المعياري	الالتواء	التفلطح
الدرجات المكافئة	المئينية	٢٢,٢٠٠	٧,٨٩٤	٠,٨٦٢	٣,٦٣٩
	المئينية المهدة بعديا	٢٢,٢٠٣	٧,٨٧٣	٠,٨٦٧	٣,٦٣٩
الدرجات قبل المعادلة	النموذج (٠٠٨٢)	٢٠,٧٧٠	٧,٤٥٢	١,٠٦١	٤,٠٤٥
	النموذج (١٠٨١)	٢٢,٧٨٠	٨,٠٤٧	٠,٧٤٢	٣,٣٣٩

الفروق في الدرجات المكافئة (الخام والمعيارية):

ما مقدار الفروق في الدرجات المكافئة (الخام والمعيارية) للنموذج (٠٠٨٢) في اختباري القدرة اللفظية والقدرة الكمية حسب طرائق المعادلة الكلاسيكية في تصميم المفردات المشتركة؟

١) **القدرة اللفظية:** يظهر الجدول (٦) المتوسطات غير الموزونة والموزونة للفروق بين الدرجات المكافئة (الخام والمعيارية) للقدرة اللفظية في النموذج (٠٠٨٢) (مع النموذج المرجعي ١٠٨١)، والدرجات المكافئة (الخام والمعيارية) من المعادلة المتطابقة. وتشير المتوسطات غير الموزونة للفروق في جميع طرائق المعادلة إلى أن قيمها سالبة (تقل الدرجات المكافئة (الخام والمعيارية) للنموذج (٠٠٨٢) بمجملها عن الدرجات المكافئة من المعادلة المتطابقة). وتدل هذه النتائج على أن الدرجة التي يحصل عليها الطالب في النموذج (٠٠٨٢) تتحول إلى درجة مكافئة أصغر حسب تدريج النموذج المرجعي لاختبار القدرة اللفظية، أي أن الطالب بمستوى معين من القدرة اللفظية سيحصل على درجة أقل في حالة تقديمه للنموذج المرجعي مقارنة بتقديمه للنموذج (٠٠٨٢).

جدول (٦)
المتوسطات الموزونة وغير الموزونة للفروق بين الدرجات المكافئة (الخام والمعيارية) للنموذج (٠٠٨٢) لاختبار القدرة اللفظية والقدرة الكمية

الدرجة	الطريقة	الدرجة الخام المكافئة *	الدرجة المعيارية المكافئة *
القدرة اللفظية	الخطية (Tucker)	-٣,٠١٨ (٠,٠٤٣-)	-٣,٠١٦ (٠,٠٤٣-)
	الخطية (Levine)	-٣,٢٥١ (٠,٠٤٦-)	-٣,٢٤٣ (٠,٠٤٦-)
	الخطية (Braun-Holland)	-٢,٩٧٥ (٠,٠٤٢-)	-٢,٩٧٣ (٠,٠٤٢-)
	المئينية	-٢,٠٠٣ (٠,٠٤٢-)	-٢,٠٠٣ (٠,٠٤٢-)
	المئينية المهدة بعديا	-١,٩٥٨ (٠,٠٤٢-)	-١,٩٥٧ (٠,٠٤٢-)
	الدرجة المعيارية المباشرة	-	-٢,١٠٨ (٠,٠٤٢-)
القدرة الكمية	Tucker	١,٨١٤ (٠,٠٢٨)	١,٩٩٥ (٠,٠٣٥)
	Levine	١,٣٧٨ (٠,٠٢٣)	١,٥٦٦ (٠,٠٢٨)

تابع جدول (٦)

الدرجة	الطريقة	الدرجة الخام المكافئة *	الدرجة المعيارية المكافئة *
القدرة الكمية	Braun-Holland	١,٧٤٠ (٠,٠٢٧)	١,٩١٩ (٠,٠٣٤)
	المئينية	١,١٧٠ (٠,٠٢٧)	١,٤٥٣ (٠,٠٣٤)
	المئينية المهدة بعديا	١,٢١٣ (٠,٠٢٧)	١,٤٩٧ (٠,٠٣٤)
	الدرجة المعيارية المباشرة	-	٢,٥٧٤ (٠,٠٥١)

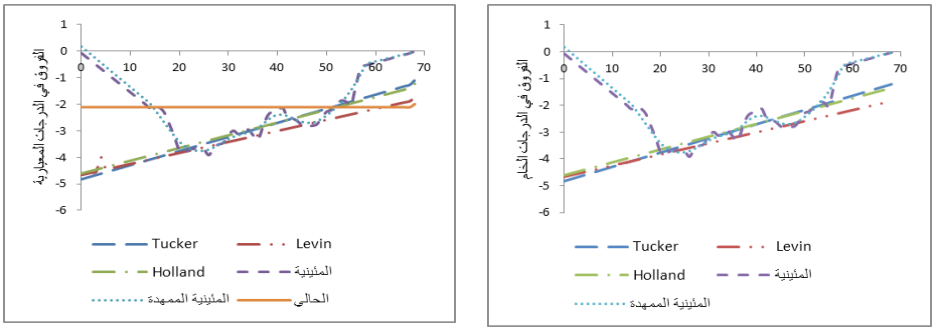
* تمثل الأرقام خارج الأقواس المتوسطات غير الموزونة للفروق، وتمثل الأرقام داخل الأقواس المتوسطات الموزونة للفروق.

كما يظهر في النتائج وجود اختلاف المتوسطات غير الموزونة للفروق في الدرجات المكافئة بين طرائق المعادلة الخمس. وعند وزن الفروق بين الدرجات المكافئة باستخدام تكرارات الدرجات الخام لكل نموذج جديد، تنقص المتوسطات الموزونة للفروق عن المتوسطات غير الموزونة للفروق مع استمرار اتجاه الفروق الدالة على نقصان الدرجات الخام المكافئة لطرائق المعادلة الخمس عن الدرجات الخام المكافئة بالمعادلة المتطابقة. وفيما يخص الدرجات المعيارية المكافئة، فيظهر الجدول (٦) تقارب المتوسطات الموزونة وغير الموزونة للفروق بين الدرجات المعيارية المكافئة، وتلك الناتجة من المعادلة المتطابقة مع نظيراتها في الدرجات الخام المكافئة.

ويعرض الشكل (١) الفروق بين الدرجات المكافئة (الخام والمعيارية) للنموذج (٠٠٨٢) مع النموذج (١٠٨١) والدرجات المكافئة من المعادلة المتطابقة في اختبار القدرة اللفظية عبر متصل الدرجات الخام للنموذج المرجعي. وتشير النتائج المعروضة في الرسم البياني على يمين الشكل (١) أن الفروق لجميع الدرجات الخام المكافئة للنموذج (٠٠٨٢) في طرائق المعادلة الخمس تقل عن درجات المعادلة المتطابقة عبر تدريج الدرجات. كما تُشير هذه الفروق إلى تشابه الطرائق الثلاث للمعادلة الخطية عبر متصل الدرجات، واختلافها عن طريقتي المعادلة المئينية. وتُظهر الفروق في الدرجات الخام المكافئة لطرائق المعادلة الخطية انتظاما خطيا، بحيث إن الفروق في الدرجات الخام المكافئة تقل بالانتقال من الدرجات المنخفضة إلى الدرجات المرتفعة مع بقاء الاتجاه السالب للفروق، بينما تُظهر الفروق في الدرجات المكافئة لطريقتي المعادلة المئينية انتظاما غير خطي ومتعرجا، بحيث تزيد هذه الفروق عند الدرجات الوسطية، بينما تقل هذه الفروق عند الدرجات الطرفية (المنخفضة والمرتفعة).

كما يشير الرسم البياني على يسار الشكل (١) إلى أن الفروق للدرجات المعيارية المكافئة بجميع طرائق المعادلة تشبه تلك المشاهدة مع الدرجات الخام المكافئة. كما يجدر الذكر أن طرائق المعادلة المئينية بنوعيها تعطي درجات مكافئة (خام ومعيارية) متشابهة لحد كبير مع طرائق المعادلة الخطية الثلاث عند أغلب درجات اختبار القدرة اللفظية ما عدا عند الدرجات

الطرفية فقط (المنخفضة والمرتفعة). وفيما يخص الدرجات المعيارية المباشرة (الحالي)، فيظهر الشكل (١) أنها تختلف عن الدرجات المعيارية الناتجة من بقية طرائق المعادلة الخمس. فتشير الفروق إلى اختلافها عن الدرجات الناتجة من المعادلة المتطابقة بمقدار ثابت (ويساوي الفرق بين المتوسط الحسابي للنموذج الجديد والنموذج المرجعي) عبر جميع الدرجات، مما تدل على أن الدرجات المعيارية المباشرة تتشابه مع المعادلة المتطابقة في أن كلا منها تفترض تطابق درجات النموذج الجديد مع النموذج المرجعي مع اختلاف بينهما بمقدار ثابت.



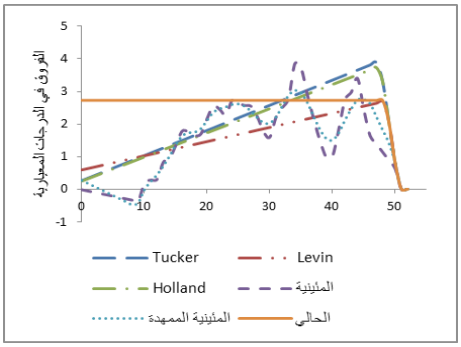
شكل (١)
الفروق بين الدرجات المكافئة (الخام والمعيارية) للنموذج (٠٠٨٢) مع النموذج (١٠٨١) والدرجات المكافئة من المعادلة المتطابقة في اختبار القدرة اللفظية

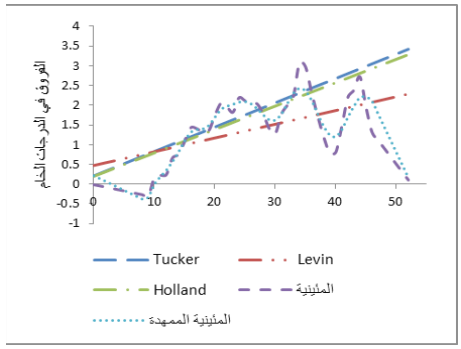
٢) **القدرة الكمية:** يُظهر الجدول (٦) أيضا المتوسطات غير الموزونة والموزونة للفروق بين الدرجات المكافئة (الخام والمعيارية) للقدرة الكمية في النموذج (٠٠٨٢) (مع النموذج المرجعي ١٠٨١) والدرجات المكافئة (الخام والمعيارية) من المعادلة المتطابقة. وتشير المتوسطات غير الموزونة للفروق في جميع طرائق المعادلة إلى أن قيمها موجبة (تزيد الدرجات المكافئة (الخام والمعيارية) للنموذج (٠٠٨٢) بمجملها عن الدرجات المكافئة بالمعادلة المتطابقة). كما تشير النتائج إلى اختلاف المتوسطات غير الموزونة للفروق في الدرجات المكافئة بين طرائق المعادلة الخمس. وعند وزن الفروق بين الدرجات المكافئة باستخدام تكرارات الدرجات الخام لكل نموذج جديد، تنقص المتوسطات الموزونة للفروق عن المتوسطات غير الموزونة للفروق مع استمرار اتجاه الفروق الدالة على زيادة الدرجات الخام المكافئة لطرائق المعادلة الخمس عن الدرجات الخام المكافئة بالمعادلة المتطابقة. وتدل هذه النتائج على أن الدرجة التي يحصل عليها الطالب في النموذج (٠٠٨٢) تتحول إلى درجة مكافئة أكبر حسب تدريج اختبار القدرة الكمية، أي مستوى أعلى من القدرة الكمية. وفيما يخص الدرجات المعيارية المكافئة، فيظهر

الجدول (٦) ارتفاع المتوسطات الموزونة وغير الموزونة للفروق بين الدرجات المعيارية المكافئة وتلك الناتجة من المعادلة المتطابقة مع نظيراتها في الدرجات الخام المكافئة، مما يشير إلى تأثير طرائق المعادلة على الدرجات المعيارية المكافئة.

ويُظهر الشكل (٢) أن الفروق لجميع الدرجات المكافئة (يتشابه النمط في الدرجات الخام والدرجات المعيارية) للقدرة الكمية في النموذج (٠٠٨٢) في طرائق المعادلة الخمس تقل عن درجات المعادلة المتطابقة. كما تشير الفروق للدرجات المكافئة إلى تشابه الطرائق الثلاث للمعادلة الخطية عبر تدريج الدرجات، حيث تظهر انتظاما خطيا، مع بقاء الاتجاه السالب للفروق. كذلك تعطي كل من طريقتي Tucker و Braun-Holland درجات مكافئة شبه متطابقة، بينما تختلف الدرجات المكافئة الناتجة من طريقة Levine، وخاصة عند الدرجات المرتفعة في اختبار القدرة الكمية. وقد يفسر ذلك اختلاف افتراضات طريقة Levine (تتمحور حول الدرجات الحقيقية) عن افتراضات طريقتي Tucker و Braun-Holland (تتمحور حول الدرجات المشاهدة). بينما تظهر الفروق في الدرجات المكافئة لطريقتي المعادلة المئينية انتظاما وتذبذبا غير خطي، بحيث تزيد هذه الفروق عند الدرجات الوسطية، بينما تقل هذه الفروق عند الدرجات الطرفية (المنخفضة والمرتفعة).

ويظهر الشكل (٢) تخطي الدرجات المعيارية المكافئة عتبة الدرجة القصوى (١٠٠) عند الدرجات الخام المرتفعة والتي يتطلب تثبيتها عند تلك الدرجة. وتعد هذه النتيجة كأحد سلبيات المعادلة الخطية، نظرا للانتظام الخطي للدرجات المكافئة. وتؤدي هذه الظاهرة في المعادلة الخطية إلى منح الأفراد ذوي الدرجات المرتفعة في النموذج (٠٠٨٢) الدرجة المعيارية الكاملة برغم تمايزهم في مستوى القدرة الكمية كما تظهرها درجاتهم الخام. كما يجدر بالذكر أن طرائق المعادلة المئينية بنوعيها تعطي درجات مكافئة (الخام والمعيارية) متشابهة لحد كبير مع طريقتي المعادلة الخطية (Tucker وBraun-Holland) عند الدرجات الوسطية في اختبار القدرة الكمية. وفيما يخص الدرجات المعيارية المباشرة (الحالي)، فيظهر الشكل (٢) أنها تختلف عن الدرجات المعيارية الناتجة من بقية طرائق المعادلة الخمس، حيث إن الفروق تشير إلى اختلافها عن الدرجات الناتجة من المعادلة المتطابقة بمقدار ثابت (ويساوي الفرق بين المتوسط الحسابي للنموذج الجديد والنموذج المرجعي) عبر جميع الدرجات. وتدل هذه النتيجة على أن الدرجات المعيارية المباشرة تتشابه مع المعادلة المتطابقة في أن كلا منها تفترض تطابق درجات النموذج الجديد مع النموذج المرجعي مع اختلافهما بمقدار ثابت.





شكل (٢)
الفروق بين الدرجات المكافئة (الخام والمعيارية) للنموذج (٠٠٨٢) مع النموذج المرجعي والدرجات المكافئة من المعادلة المتطابقة في اختبار القدرة الكمية

صدق النتائج عبر النماذج (معادلة درجات النموذج (٠٠٨٣)):

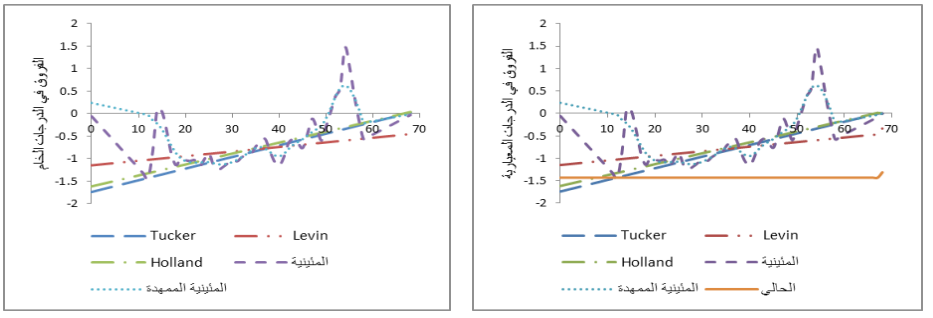
هل تتشابه نتائج معادلة درجات النموذج (٠٠٨٣) مع نتائج معادلة درجات النموذج (٠٠٨٢) في اختباري القدرة اللفظية والقدرة الكمية كمؤشر على صدق نتائج المعادلة عبر النماذج؟

١) القدرة اللفظية: تم إجراء تحليلات مشابهة باستخدام النموذج (٠٠٨٣) كمؤشر على صدق النتائج المتحصل عليها من النموذج (٠٠٨٢). ويظهر الجدول (٧) المتوسطات غير الموزونة والموزونة للفروق بين الدرجات المكافئة (الخام والمعيارية) للنموذج (٠٠٨٣) مع النموذج المرجعي (١٠٨١) والدرجات المكافئة (الخام والمعيارية) من المعادلة المتطابقة في اختبار القدرة اللفظية. وتشير المتوسطات غير الموزونة للفروق في جميع طرائق المعادلة نفس النمط المشاهد في القدرة اللفظية في النموذج السابق (٠٠٨٢)، من حيث إن الدرجات الخام المكافئة بمجملها للنموذج (٠٠٨٣) تقل عن الدرجات الخام المكافئة من المعادلة المتطابقة، ولكن بمقادير أقل مقارنة بالنموذج السابق. وكما تتشابه النتائج في النموذجين من حيث اتجاه ونمط الفروق للدرجات الخام المكافئة للنموذج (٠٠٨٣) عبر متصل الدرجات الخام للقدرة اللفظية في طرائق المعادلة الخمس كما يظهر في الشكل (٣). وقد تختلف نتائج النموذج (٠٠٨٣) عن النموذج (٠٠٨٢) في بعض أجزاء متصل الدرجات الخام، وبالتحديد عند الدرجات المحصورة بين ٥٠ و٥٧، حيث كانت هذه الفروق موجبة (أي أن الدرجات الخام المكافئة بالمعادلة المثينية أكبر من الدرجات المكافئة بالمعادلة المتطابقة).

جدول (٧)
المتوسطات الموزونة وغير الموزونة للفروق بين الدرجات المكافئة (الخام والمعيارية) للنموذج (٠٠٨٢) لاختبار القدرة اللفظية والقدرة الكمية

الدرجة	الطريقة	الدرجة الخام المكافئة*	الدرجة المعيارية المكافئة *
القدرة اللفظية	الخطية (Tucker)	٠,٨٥٧ – (٠,٠١٢ –)	٠,٨٥٨ – (٠,٠١٢ –)
	الخطية (Levine)	٠,٨٠٠ – (٠,٠١٢ –)	٠,٧٩٨ – (٠,٠١٢ –)
	الخطية (Braun-Holland)	٠,٧٨٨ – (٠,٠١١ –)	٠,٧٨٨ – (٠,٠١١ –)
	المئينية	٠,٥٧١ – (٠,٠١١ –)	٠,٥٧٠ – (٠,٠١١ –)
	المئينية المهدة بعديا	٠,٣٨٨ – (٠,٠١١ –)	٠,٣٨٨ – (٠,٠١١ –)
	الدرجة المعيارية المباشرة	–	١,٤٢٨ – (٠,٠٢١ –)
القدرة الكمية	Tucker	٦,٨٤٥ (٠,٠٨٢)	٥,٨٩٨ (٠,١٠٢)
	Levine	٦,٩٢٣ (٠,٠٨٢)	٥,٩٢١ (٠,١٠٢)
	Braun-Holland	٦,٩٠٤ (٠,٠٨٢)	٥,٩١٨ (٠,١٠٢)
	المئينية	٤,٥١٧ (٠,٠٨٣)	٥,٦٠٠ (٠,١٠٣)
	المئينية المهدة بعديا	٤,٥٧٥ (٠,٠٨٣)	٥,٦٦١ (٠,١٠٣)
	الدرجة المعيارية المباشرة	–	٥,٠٤٧ (٠,١٠٣)

* تمثل الأرقام خارج الأقواس المتوسطات غير الموزونة للفروق، وتمثل الأرقام داخل الأقواس المتوسطات الموزونة للفروق.

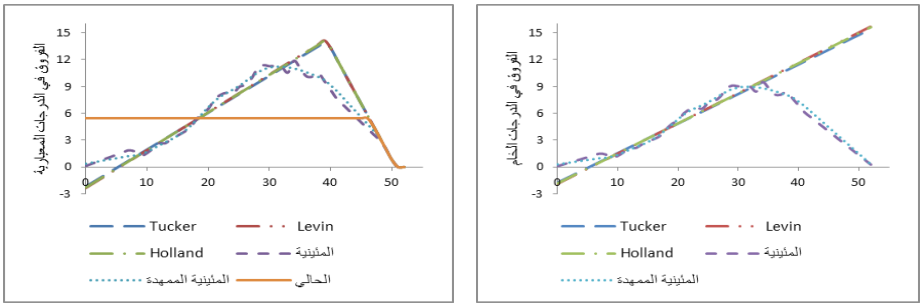


شكل (٣)
الفروق بين الدرجات الخام المكافئة للنموذج (٠٠٨٣) مع النموذج (١٠٨١) والدرجات المكافئة من المعادلة المتطابقة في اختبار القدرة اللفظية

٢) **القدرة الكمية:** يظهر الجدول (٧) ارتفاع المتوسطات الحسابية الموزونة وغير الموزونة للفروق بين الدرجات الخام المكافئة للنموذج (٠٠٨٣) من اختبار القدرة الكمية في جميع طرائق المعادلة الخمس وبين الدرجات المكافئة (الخام والمعيارية) من المعادلة المتطابقة، وتتشابه هذه النتائج مع تلك المتحصل عليها للقدرة الكمية في النموذج السابق (٠٠٨٢) من حيث إن قيم المتوسطات غير الموزونة للفروق في جميع طرائق المعادلة موجبة (تزيد الدرجات المكافئة (الخام والمعيارية) للنموذج (٠٠٨٢) بمجملها عن الدرجات المكافئة بالمعادلة المتطابقة)،

ولكن بمقادير أكبر مقارنة بالنموذج السابق. وكما يظهر في الشكل (٤)، اتفق النموذجان من حيث إن الانتظام الخطي للمعادلات الخطية أدى إلى تخطي الدرجات المعيارية المكافئة للدرجة المعيارية القصوى (١٠٠) لجميع الدرجات الخام فوق الدرجة ٤٠ (من أصل ٥٢ درجة) وتم تثبيتها عند تلك الدرجة. مما يعني عدم القدرة على التمييز بين هؤلاء الطلبة عند استخدام الدرجات المعيارية المكافئة الناتجة من طرائق المعادلة الخطية الثلاث.

وعند مقارنة متوسطات الفروق بين الدرجات المكافئة مع الفروق بين المتوسطات الحسابية للدرجات الخام في النموذجين (٠٠٨٢) و (٠٠٨٣)، نلاحظ أن المتوسطات الموزونة وغير الموزونة للفروق في طرائق المعادلات الخمس تزداد كلما زاد الفرق بين المتوسطات الحسابية للدرجات الخام في النموذجين.



شكل (٤)
الفروق بين الدرجات الخام المكافئة للنموذج (٠٠٨٣) مع النموذج (١٠٨١) والدرجات المكافئة من المعادلة المتطابقة في اختبار القدرة الكمية

الأخطاء المعيارية للمعادلة:

ما مقادير الأخطاء المعيارية للمعادلة في طرائق المعادلة الكلاسيكية للنموذجين (٠٠٨٢، ٠٠٨٣) في اختباري القدرة اللفظية والقدرة الكمية؟

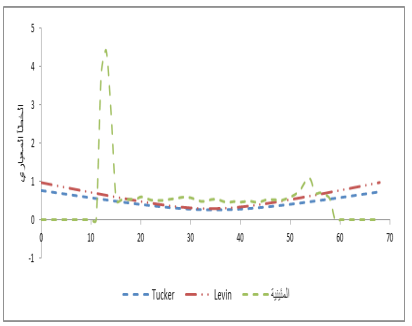
يظهر الجدول (٨) متوسط الأخطاء المعيارية للدرجات الخام المكافئة لنماذج اختبار القدرة اللفظية واختبار القدرة الكمية على التوالي. وتظهر النتائج أن متوسط الأخطاء المعيارية للدرجات الخام المكافئة للمعادلة بالطريقة المثينية أقل من متوسط الأخطاء المعيارية للمعادلة بالطرائق الخطية (Tucker, Levine) للمعادلتين (٠٠٨٢ مع ١٠٨١؛ ٠٠٨٣ مع ١٠٨١) في الاختبارين. كما حصلت الطريقة الخطية (Levine) على أكبر متوسط للأخطاء المعيارية للمعادلة في جميع نماذج الاختبار . وعند استعراض الأخطاء المعيارية للدرجات الخام المكافئة لكل نموذجين تم معادلتهما في اختباري القدرة اللفظية والقدرة الكمية عبر درجات الاختبار (انظر الشكلين ٥، ٦)، تشير النتائج إلى تقارب مقدار الأخطاء المعيارية

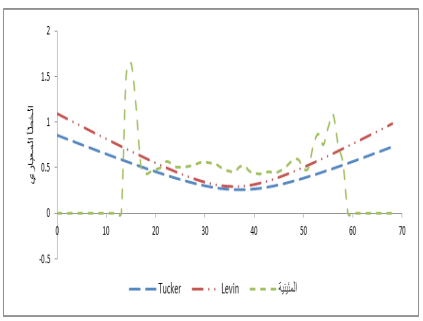
للمعادلة في طرائق المعادلة الثلاث عند الدرجات الوسطية؛ بينما يختلف سلوك الأخطاء المعيارية للمعادلات الثلاث عند الدرجات الطرفية من الجهتين. فيلاحظ أن الأخطاء المعيارية للمعادلات الخطية (Tucker, Levine) تزيد بالانتقال إلى الدرجات الطرفية؛ بينما يلاحظ أن الأخطاء المعيارية للمعادلة المئينية ترتفع بشكل مفاجئ عند درجات معينة (الدرجات ١٢-١٨ والدرجات ٥٢-٥٦)، ثم تنخفض إلى درجة انعدام الأخطاء المعيارية عند الانتقال إلى الدرجات الطرفية. وبرغم هذا السلوك للمعادلة المئينية عند الدرجات الطرفية، فإن متوسط أخطاءها المعيارية أصغر من متوسط الأخطاء المعيارية للمعادلات الخطية. وقد يعود السبب في الارتفاع المفاجئ للأخطاء المعيارية للمعادلة المئينية عند تلك الدرجات إلى قلة عدد الطلبة الذين حصلوا على هذه الدرجات.

جدول (٨)
متوسط الأخطاء المعيارية × للدرجات الخام المكافئة لنماذج اختباري القدرة اللفظية والقدرة الكمية

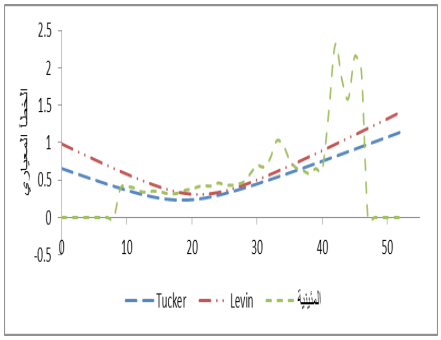
الطريقة	القدرة اللفظية		القدرة الكمية	
	النموذج (٠٠٨٢)	النموذج (٠٠٨٣)	النموذج (٠٠٨٢)	النموذج (٠٠٨٣)
الخطية (Tucker)	٠,٤٨٤	٠,٤٦٠	٠,٥٥٢	٠,٧٨٦
الخطية (Levine)	٠,٦١١	٠,٥٧٨	٠,٧١٠	٠,٩٦٠
الخطية (Braun-Holland)	–	–	–	–
المئينية	٠,٤٠٠	٠,٥١٨	٠,٥٢٤	٠,٤٩٥
المئينية المهدة بعديا	–	–	–	–

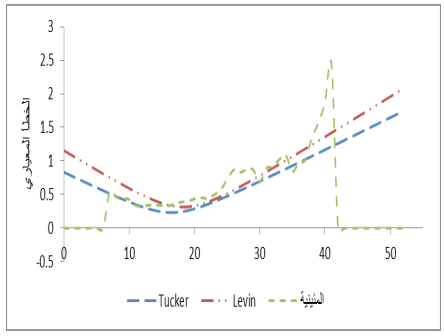
* لا تتوفر طريقة لحساب الأخطاء المعيارية للمعادلة الخطية لطريقة براون-هولاند وللمعادلة المئينية المهدة بعديا.





شكل (٥)
الأخطاء المعيارية للدرجات الخام المكافئة للنموذج (٠٠٨٢) والنموذج (٠٠٨٣) مع (١٠٨١) في اختبار القدرة اللفظية





شكل (٦)
الأخطاء المعيارية للدرجات الخام المكافئة للنموذج (٠٠٨٢)
والنموذج (٠٠٨٣) مع (١٠٨١) في اختبار القدرة الكمية

المناقشة والاستنتاجات:

هدفت الدراسة الحالية إلى مقارنة نتائج الطرائق الإحصائية للمعادلة في تصميم المفردات المشتركة من خلال معادلة درجات النموذجين (٠٠٨٣) و (٠٠٨٢) مع درجات النموذج المرجعي (١٠٨١) في اختباري القدرة اللفظية والقدرة الكمية في اختبار القدرات العامة. وأظهرت النتائج اختلاف هذه النماذج في متوسطاتها الحسابية بمقادير مختلفة، وكذلك الإحصاءات الوصفية الأخرى وتوزيعاتها التكرارية. وكان الاختلاف بين الإحصاءات الوصفية لنماذج اختبار القدرة الكمية أكثر وضوحا من اختبار القدرة اللفظية. ولم يقتصر الاختلاف في الإحصاءات الوصفية على درجات جميع المفردات في الاختبار، بل ظهر أيضا في المفردات المشتركة بين النماذج الثلاثة في كلا الاختبارين. وتتوافق هذه النتائج مع ما ذهب إليه العديد من الباحثين (مثل Angoff, 1987; Kolen & Brennan, 2004)، حيث إنه وبرغم الحرص الشديد من قبل القائمين على اختبار القدرات العامة على أن تكون هذه النماذج متكافئة في جدول مواصفاتها وخصائصها السيكومترية عند بنائها ؛ فإنها لم تعط بعد تطبيقها درجات متكافئة في إحصاءاتها الوصفية بسبب ظروف التطبيق ومستوى الجماعات الطلابية التي تقدم نماذج الاختبارات.

وأسفرت النتائج عن الارتفاع النسبي الملحوظ في المتوسطات غير الموزونة والموزونة للفروق في جميع طرائق المعادلة مقارنة بالدرجات الأصلية للنموذج في حالة عدم إجراء المعادلة لها، وخاصة في النموذج (٠٠٨٣) من اختبار القدرة الكمية، مما يشير إلى وجود اختلاف بين تدريج الاختبار في حالة إجراء المعادلة وبين عدم إجراء المعادلة. وكان اتجاه الفروق سالبا

عندما كان النموذج الجديد أقل صعوبة من النموذج المرجعي، وبالتالي تحولت الدرجات الخام الأصلية للنموذج الجديد إلى درجات مكافئة أصغر منها؛ بينما كان اتجاه الفروق موجبا عندما كان النموذج الجديد أكثر صعوبة من النموذج المرجعي، والذي يؤدي إلى أن تتحول الدرجات الخام الأصلية للنموذج الجديد إلى درجات مكافئة أكبر منها.

كما أشارت النتائج إلى ارتباط مقدار الفروق في الدرجات المكافئة بين النموذجين في جميع طرائق المعادلة مع مقدار الاختلاف في مستوى صعوبة النموذجين، فكلما كان الاختلاف بين مستوى صعوبة النموذج الجديد والنموذج المرجعي كبيرا، كانت المتوسطات غير الموزونة والموزونة للفروق في الدرجات المكافئة أكبر. وتتسق هذه النتيجة مع المبدأ التي تقوم عليه معادلة نماذج الاختبار من حيث إنها تضبط الاختلاف بين مستوى صعوبة نماذج الاختبار، بحيث إذا اختلف النموذجان كثيرا في مستوى صعوبتهما كلما احتاج النموذج الجديد إلى تطبيق عملية تعديل درجاته ليتوافق مع تدريج درجات النموذج المرجعي، وبالتالي تختلف الدرجات المكافئة للنموذج الجديد عن درجاته الأصلية بمقدار أكبر. وظهرت هذه العلاقة في جميع طرائق المعادلة الخطية منها والمئينية.

كما توصلت الدراسة إلى أن طرائق المعادلة المئينية والمعادلة الخطية تتشابه في مقدار الفروق للدرجات المكافئة عند درجات الاختبار الأصلية بالقرب من المتوسط الحسابي لدرجات الطلاب، بينما تختلف فيما بينها عند الدرجات الطرفية (الدرجات المرتفعة والمنخفضة) وتزداد هذه الاختلافات بالاقتراب من حدود درجات الاختبار. وتتفق هذه النتيجة مع ما ذهبت إليه العديد من الدراسات (Kolen & Brennan, 2004; Wang, et. al., 2008) إلى أن الطرائق المئينية أكثر ملائمة عندما يتطلب اتخاذ قرارات حول أداء الطلبة استخدام جميع درجات الاختبار؛ بينما تفي طرائق المعادلة الخطية للغرض عندما يتم اتخاذ القرارات التعليمية لأداء الطلبة باستخدام درجة المتوسط الحسابي وما حولها. وفي ذات الوقت، أظهرت الطريقة المئينية (غير المهدة) تباينا غير خطي ومتذبذبا في الفروق عبر درجات الاختبار وأظهرت فروقا كبيرة في بعض الدرجات الطرفية للاختبار (المنخفضة والمرتفعة). بينما أظهرت الطريقة المئينية المهدة بعديا تباينا أقل تذبذبا وأكثر تمهيدا للمعادلة، مما يزيد من تفضيل الطريقة المئينية المهدة لمعادلة درجات الاختبار، وخاصة تلك الدرجات التي لا يحصل عليها عدد كاف من الطلبة.

كما أظهرت الدراسة عدم ملائمة طريقة Levine في معادلتين من المعادلات الأربع في الدراسة، من حيث اختلاف الدرجات المكافئة المتحصل عليها عن تلك الدرجات المكافئة

المتحصل عليها من طرائق المعادلة الخطية الأخرى. وتدل هذه النتيجة إلى عدم توفر افتراضات طريقة Levine الخاصة بالدرجات الحقيقية بدلا من الدرجات المشاهدة في بيانات نماذج اختباري القدرة اللفظية والقدرة الكمية. لذا لا ينصح باستخدام طريقة Levine لمعادلة درجات اختبارات القدرات العامة.

تشير النتائج السابقة للأخطاء المعيارية لمعادلة الدرجات الخام بالطرائق الخطية والطرائق المئينية في تصميم المفردات المشتركة إلى أن إجراء المعادلة للنموذجين (٠٠٨٢، ٠٠٨٣) مع النموذج المرجعي (١٠٨١) في اختباري القدرة اللفظية والقدرة الكمية لا تخلو من أخطاء معيارية، وتتفاوت فيما بينها في مقدار هذه الأخطاء المعيارية. فالطريقة المئينية أظهرت أخطاء معيارية أصغر نسبيا من الطرائق الخطية، مما يتماشى مع أفضلية الطريقة المئينية في أنها تعطي نتائج أكثر دقة وأقل أخطاء معيارية عند جميع درجات الاختبار وليس حول المتوسط الحسابي فقط. وتشير النتائج إلى تقارب قيم الأخطاء المعيارية للمعادلة لدرجات النموذجين الجديدين في اختباري القدرة اللفظية والقدرة الكمية برغم اختلاف مقادير الفروق بين صعوبة كل نموذج والنموذج المرجعي. ويشير ذلك إلى عدم تأثر الأخطاء المعيارية بمقدار الاختلاف في صعوبة النموذجين المراد معادلتهما .

ويمكن تلخيص نتائج معادلة النماذج الثلاثة لاختباري القدرة اللفظية والقدرة الكمية باستخدام تصميم المفردات المشتركة في أن عملية المعادلة تعطي درجات خام ودرجات معيارية مختلفة عن الدرجات الخام والمعيارية الأصلية للنماذج (من دون معادلة). كما يتضح أن المعادلة بالطريقة المئينية والمئينية المهدة بعديا حققت مؤشرات إحصائية أفضل من بقية طرائق المعادلة، وخاصة عند الدرجات المرتفعة في الاختبارين. وهذه النتائج تدعم استخدام الطريقة المئينية (غير المهدة أو المهدة بعديا) لإجراء عمليات المعادلة لنماذج اختبارات القدرات العامة؛ نظرا للاستخدام الرئيس لهذه الاختبارات كمعايير لقبول الطلبة بأغلب مؤسسات التعليم العالي بالمملكة العربية السعودية. واستخدام الاختبار كمعيار قبول بمؤسسات التعليم العالي يتطلب اتخاذ قرارات حول إمكانية قبول الطلبة الحاصلين على الدرجات العالية في الاختبارين ممن رغبوا في القبول بهذه المؤسسات. لذا تُعد معادلة الدرجات المرتفعة في نماذج اختباري القدرات العامة أكثر حرجا وتأثيرا في القرارات المتخذة على كل من الطالب ومؤسسات التعاليم العالي. وبالتالي تظهر الحاجة ماسة أكثر إلى اختيار طريقة المعادلة التي تعطي درجات معادلة أكثر دقة (تميز بين الطلاب الحاصلين على الدرجات المرتفعة) وأقل أخطاء معيارية. ومن خلال نتائج الدراسة، تتوفر هذه الخصائص المطلوبة في المعادلة المئينية والمئينية المهدة بعديا، حيث إنها تعطي درجات معادِلة داخل إطار تدريج الاختبار، ولا تتطلب عملية تثبيت للدرجة (Truncation) عند الدرجة المعيارية القصوى (١٠٠) مع حصولها على

أخطاء معيارية صغيرة للمعادلة عند طرفي تدريج الاختبار. ولذا توصي الدراسة باستخدام الطريقة المئينية أو الطريقة المئينية المهدة بعديا لمعادلة نماذج اختبارات القدرات العامة باستخدام تصميم المفردات المشتركة. وتقترح الدراسة إجراء المزيد من الدراسات على نماذج أخرى لاختبار القدرات العامة للتأكد من فعالية الطريقة المئينية والمهدة بعديا في معادلة نماذج هذه الاختبارات.

وبرغم مراعاة الدراسة لتمثيل المفردات المشتركة لجدول مواصفات الاختبار الكلي ومراعاة أن يكون عددها متوافق مع النسبة الموصى بها في دراسات تصميم المفردات المشتركة، إلا أن الدراسة اقتصرت على عينة واحدة من المفردات المشتركة وبعدد محدد، والذي قد يؤثر على نتائج الدراسة. فتقترح الدراسة إجراء دراسات أخرى على معادلة نماذج اختباري القدرة اللفظية والقدرة الكمية باستخدام عينات مختلفة من المفردات المشتركة وبأعداد مختلفة، ومقارنة الدرجات المكافئة الناتجة منها، ومعرفة تأثير كل من حجم ونوعية عينة المفردات المشتركة.

اقتصرت الدراسة على استخدام المفردات المشتركة الداخلية (Internal Common Items)، والذي قد يعاب عليه التأثير الناتج من تسرب المفردات المشتركة –في حالة حدوثه– على درجات الطلاب. ويقل تأثير تسرب المفردات المشتركة في حالة استخدام تصميم المفردات المشتركة الخارجية (External Common Items)، حيث أن درجاتها لا تحتسب ضمن الدرجة الفعلية للطالب. لذا تقترح الدراسة أيضا مقارنة نتائج معادلة نماذج الاختبار عند استخدام مفردات مشتركة خارجية. كما تقترح الدراسة إجراء دراسات تجريبية للتحقق من صدق الدرجات المكافئة المتحصل عليها من خلال القدرة التنبؤية لها مع المعدل الجامعي في بعض مؤسسات التعليم العالي. كما تقترح الدراسة إجراء دراسات مستقبلية لمقارنة معادَلة نماذج اختبار القدرات العامة باستخدام الطرق الإحصائية في النظرية الكلاسيكية مع نظرية الاستجابة للمفردة.

المراجع:

الدوسري، راشد حماد (٢٠٠١). معادلة الاختبارات: مفهومها، وطرقها، ومشكلات تطبيقها. مجلة العلوم التربوية والنفسية، جامعة البحرين. ٢، ١٠٥-١٤١.

نيتكو، أنتوني؛ وبروكهارت، سوزان (٢٠١٢). التقييم التربوي للطلبة (علي القرني ؛ راشد المحرزي؛ إبراهيم الدوسري ؛ حسين الخروصي، ترجمة). مكتب التربية العربي لدول الخليج (تم نشر النسخة الأصلية للكتاب في ٢٠٠٩).

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. (Reprinted as W. A. Angoff, Scales, norms, and equivalent scores. Princeton, NJ: Educational Testing Services, 1984).

Angoff, W. H. (1987). Technical and practical issues in equating: A discussion of four papers. *Applied Psychological Measurement, 11,* 291-300.

Baker, F. B. & Al-Karni, A. (1991). A comparison of two procedures for computing IRT equating coefficients. *Journal of Educational Measurement, 28*, 147-162.

Braun, H. I., & Holland, P. W. (1982). Observed-score test equating: A mathematical analysis of some ETS equating procedures. In P. W. Holland & D. B. Rubin (Eds.), *Test equating* (pp. 9-49). New York, NY: Academic.

Chen, H. H., Livingston, S. A., & Holland, P. W. (2011). Generalized equating functions for NEAT designs. In Alina A. Von Davier. *Statistical Models for Test Equating, Scaling, and Linking*. New York: Springer Science.

Chulu, B. W. & Sireci, S. G. (2011). Importance of equating high stakes educational measurements. *International Journal of Testing, 11,* 38-52.

Dorans, N. j., Moses, T. P., & Eignor, D. R. (2011). Equating test scores: Toward best practices. In A. A. Von Davier, *Statistical models for test equating, scaling, and linking* (pp. 21-42). New York: Springer Science.

Ezzelle, C. & Setzer, J. C. (2009). *Technical manual: 2002 series GED tests*. Washington, DC: American Council on Education.

Glowacki, M. L. (1991). *An analysis of test equating models for the Alabama high school graduation examination*. Paper presented at the annual meeting of Mid-South Educational Research Association, Lexington. (ED 340720).

Gulliksen, H. (1950). *Theory of mental tests*. New York: Wiley.

Harris, D. J. & Kolen, M. J. (1990). Effect of examinee group on equating relationships. *Applied Psychological Measurement, 10,* 35-43.

Kendall, M. & Stuart, A. (1977). The advanced theory of statistics (4th ed., Vol. 1). New York: Macmillan.

Kolen, M. J. (1988). An NCME instructional module on traditional equating methodology. *Educational Measurement: Issues and Practice, 7,* 29-36.

Kolen, M. J. (2004b). *Common item program for equating (CIPE) (Windows Console version)*. Iowa City, IA: Center for Advanced Studies in Measurement and Assessment, The University of Iowa. (Available on http:// www. Education.uiowa.edu/casma).

Kolen, M. J. & Brennan, R. L. (2004). *Test equating, scaling, and linking: Methods and practices* (2nd ed.). New York: Springer-Verlag.

Kolen, M. J. & Whitney, D. R. (1982). Comparison of four procedures for equating the Tests of General Educational Development. *Journal of Educational Measurement, 19,* 279-293.

Levine, R. (1955). *Equating the score scales of alternate forms administered to samples of different ability* (Research Bulletin 55-23). Princeton, NJ: Educational Testing Services.

Linvingston, S.A. (1996). Book review of test equating. *Journal of Educational Measurement, 33,* 369-373.

Lord, F. M. (1982). The standard error of equipercentile equating. *Journal of Educational Statistics, 7* , 165-174.

Mroch, A., Suh, Y., Kane, M. T., & Ripkey, D. R. (2009). An evaluation of five linear equating methods for the NEAT design. *Measurement: Interdisciplinary Research and Perspectives, 7,* 174-193.

Raykov, T. (2010). Test Equating under the NEAT Design: A Necessary Condition for Anchor Items. *Measurement: Interdisciplinary Research and Perspectives, 8*, 16-20.

Ricker, K. L. & Von Davier, A. A. (2007). *The Impact of Anchor Test Length on Equating Results in a Nonequivalent Groups Design*. Princeton, NJ: Educational Testing Services.

Slinde, J. A. & Linn, R. L. (1997). An exploration of adequacy of the rasch model for the problem of vertical equating. *Journal of Educational Measurement, 15,* 23-35.

Sunnassee, D. (2011). *Conditions affecting the accuracy of classical equating methods for small samples under the neat design: A simulation study*. Phd Dissertation, University of North Carolina at Greensboro.

Tanguma, J. (2000). *Equating test scores using the linear method: A primer.* Paper presented at the Annual Meeting of Southwest Educational Research Association, Dallas, Texas.

Wang, T., Lee, W., Brennan, R. L., & Kolen, M. J. (2008). A comparison of the frequency estimation and chained equipercentile methods under the common-item non-equivalent groups design. *Applied Psychological Measurement, 32,* 632-651.