DOI: http://dx.doi.org/10.24200/jeps.vol15iss2pp277-293

معادلة درجات امتحان دبلوم التعليم العام لمادة الفيزياء في سلطنة عمان باستخدام نظرية الاستجابة للمفردة

نبيل بن سيف الحبسي* ويوسف عبدالقادر أبوشندي وحميراء بنت سليمان السليمانية
وزارة التربية والتعليم، سلطنة عمان جامعة السلطان قابوس، سلطنة عمان

استُلم بتاريخ: ٢٠٢٠/١٠/٥ قُبل بتاريخ: ٢٠٢٠/١٢/١

ملخص: هدفت الدراسة معادلة درجات امتحان دبلوم التعليم العام لمادة الفيزياء في سلطنة عمان باستخدام نظرية الاستجابة للمفردة، اشتملت عينة الدراسة جميع الطلبة الذين اجتازوا امتحان مادة الفيزياء للعامين الدراسيين (٢٠١٤/٢٠١٥م) و(٢٠١٥/٢٠١٦م) والبالغ عددهم ١٨٩٦٦ و١٧١٧٥ على التوالي، وتمثلت أداة الدراسة في صورتين، حيث احتوت الصورة الأولى على ٤٠ مفردة، و٣٦ مفردة للصورة الثانية. وبعد ما تم التحقق من افتراضات نظرية الاستجابة للمفردة في البيانات تم تقدير معالم المفردات وقدرات المفحوصين حسب النموذج اللوجستي ثنائي المعلم، ومن ثم استخدمت طريقة المعادلة بالدرجات الحقيقية، حيث أظهرت نتائج المعادلة تقارب المتوسطات الحسابية للدرجات الحقيقية للصورة المرجعية والدرجات المعادلة للصورة الجديدة، وتم التحقق من دقة المعادلة بطريقة (Bias).

كلمات مفتاحية: نظرية الاستجابة للمفردة، امتحان الدبلوم العام، الفيزياء، سلطنة عمان، معادلة الدرجات.

Equating General Education Diploma Results of Physics Examination in the Sultanate of Oman Using Item Response Theory

Nabil S. Al Habsi,* Yusef A. Abu Shindi, & Humaira S. Al Sulaimani
Ministry of Education, Sultanate of Oman Sultan Qaboos University, Sultanate of Oman

Abstrac: The study aimed to equate General Education Diploma results of physics exam in the Sultanate of Oman using Item Response Theory (IRT) and setting them in a common scale. The study sample included all of the General Education Diploma students both male and female who passed the physics exam in the academic years 2014/2015 and 2015/2016 with atotal sample of (17175 & 18966) respectively. Two different copies of physics exam were used for the purpose of the study. The first one consisted of 40 items and the other was 36 items. The data was analyzed by Multilog 7.03 program, to estimate the parameters of the items and the abilities of examinees and to check the IRT hypotheses. Moreover, the method of True Score Equation was used. The Results showed that there are no statistically significant differences between the means of the true scores of the criteria exam and the results, which were equated to the new exam. The equation accuracy was checked by Equating Bias method

Keywords: Item response theory, general education exam, physics, equating test scores, Sultanate of Oman.

*nabil121@moe.om

تعدّ الامتحانات من أدوات القياس التربوية متعددة الأهداف، فمن خلالها يستند المختصون إلى مؤشرات هامة تساعدهم في عملية التقويم من أجل قياس مستوى التحصيل والوقوف على نواحي التقدم في التعلم، وتستخدم في اتخاذ القرارات فمن خلال نتائجها يتم الانتقاء والقبول والتوجيه، وكلما زادت الحاجة للاستفادة منها زادت الحاجة لأن تكون المعلومات التي نحصل عليها من الامتحانات أكثر دقة وذات صلة بالغرض الذي أعدت من أجله، وبالتالي لابد من بناء امتحان يقيس مستوى التحصيل (عبدالهادي، ٢٠٠١).

وتقدم في غالب الأحيان نماذج ونسخ مختلفة من الامتحانات سواء في السنوات المختلفة أو في نفس الاختبار بغرض ضبط الانتحال والغش، ويخشى من اختلاف مستويات هذه النسخ عبرالتطبيقات المختلفة، لذلك اهتمت جمعية علم النفس الأمريكية والرابطة الأمريكية للبحث التربوي، بمعادلة درجات الامتحانات لزيادة عددها وتنوعها ولاستمراريتها، ولتحقيق العدالة (Kolen & Brennan, 2004)، ويمكن تعريف معادلة صورتي الامتحان بأنها تحويل نظام وحدات القياس الخاص بإحدى الصورتين إلى نظام قياس خاص بالصورة الأولى بحيث تصبح القياسات المستمدة من درجات كلٍ من الصورتين متعادلة بعد إجراء هذا التحويل (الشمري، ٢٠١٥).

وإضافة إلى تحقيق العدالة تفيد معادلة الامتحانات في عمليات تطوير بنوك الأسئلة، التي تحتاج إلى إضافة أو حذف للمفردات، وبالتالي تحتاج إلى تدريج المفردات المضافة على نفس مقياس المفردات الأصلية، أو التوصل إلى تدريج واحد للمفردات الأصلية والجديدة، حتى يتم التوصل في النهاية إلى مقياس واحد، وتتحدد صحة المعادلة بناءً على مجموعة من العناصر التي يتم اختيارها بعناية لكي تكون عملية المعادلة ذات موثوقية عالية، منها اختيار التصميم المناسب، واختيار الطريقة المناسبة، وكذلك اختيار النموذج الرياضي المناسب، فاختيار النموذج الرياضي المناسب يقود إلى مطابقة جيدة للمفردات، حيث أن لكل نموذج خصائصه المحددة، فقد يتبين لاحقاً أن مفردات الامتحان التي قام بإعدادها لا تناسب النموذج (الشواورة، ٢٠١٣).

وينتج من معادلة الامتحانات علاقة اقتران بين درجات نماذج الامتحان، بحيث تقترن كل درجة من درجات أحد النماذج بما يعادلها من درجات النموذج الآخر، وتقوم عملية معادلة نماذج الامتحان على مبدأ تعديل الاختلافات في مستوى الصعوبة لنماذج الامتحان لكي تكون متساوية في الصعوبة ومن ثم تستخدم بطريقة تبادلية (المحرزي، ٢٠١٤).

وتُعد نظرية الاستجابة للمفردة من النظريات السيكومترية المعاصرة التي ظهرت في الثلث الأخير من القرن الماضي، والتي ساعدت في تقديم طرق حديثة في الجوانب التربوية، وخاصة في معادلة الامتحانات (Reev, 2002). حيث كان لها دور رئيسي في مجال معادلة الامتحانات (IRT Equating) لما لها من إمكانية في وضع عدة امتحانات وعدة مجموعات من المفحوصين على تدريج مشترك (Common Scale) (الكوفحي، ٢٠١٣).

وتمتلك النظرية أساس قوي ونماذج رياضية تعمل على تقدير احتمالية نجاح الأفراد على مستوى المفردة، ولا تجرى التحليلات للنماذج لمستوى درجات الامتحان فقط، ولكن على مستوى المفردة أيضا، وتركز على معلومات مستوى المفردة (Fan, 1998).

معادلة درجات الامتحانات باستخدام نظرية الاستجابة للمفردة

يشير سون (Suen, 1990) إلى أن المعادلة تستخدم حين تتوفر أكثر من صورة لامتحان ما، وللحصول على تفسيرات ملائمة لدرجات صور الامتحان فإنه لابد من النظر في مسألة معادلة درجات نماذج الامتحان بالاعتماد إلى نموذج مرجعي. ويعد مجال معادلة الامتحانات Test Equating من المجالات التطبيقية المهمة في القياس النفسي والتربوي، حيث أن البرامج

الامتحانية تتطلب توافر صيغ متعددة من الامتحانات لضمان السرية، لذلك ينبغي إجراء تعادل للصور المختلفة لتحويل الدرجات الخام، بحيث تصبح الدرجات المحولة لكل صور منها متعادلة (علام، ٢٠٠٥).

ويرى أنجوف (Angoff, 1986) أن درجة المفحوص في الامتحان تعتمد على عاملين رئيسين: قدرة المفحوص، وصعوبة الامتحان، ويمكن إزالة أثر الصعوبة في نماذج الاختبار المختلفة من خلال المعادلة للنماذج ووضعها على مقياس معياري موحد، حيث أن المعادلة هنا تساعد على مقارنة الدرجات المأخوذة من عدة نماذج دون التخوف من اختلاف مستوى صعوبة النماذج.

وتقسم المعادلة لنوعين وهما: المعادلة الأفقية Horizontal quating والمعادلة الرأسية Vertical Equating. حيث تكون صور الامتحانات متقاربة الصعوبة في المعادلة الأفقية، بينما تختلف في المعادلة الرأسية (Skaggs & Lissitz, 1986). وفي التعادل الأفقي تكون الصيغ المختلفة متماثلة في المحتوى، ومتقاربة في مستوى صعوبة مفرداتها، وتقيس السمة نفسها في مجتمع معين من الأفراد (علام، ٢٠٠٥). والهدف من إجراء التعادل الأفقي هو تعديل الفروق الناتجة عن اختلاف مستوى الصعوبة، إذا كانت الصور المختلفة متقاربة في صعوبتها، ولذا فإن التعادل الأفقي يكون ملائما عندما يكون لدينا صور مختلفة من الامتحان نفسه، ولكنها ليست متطابقة في المفردات (الدوسري، ٢٠٠١)، ومثال عليها تكافؤ صيغ الامتحان عبر السنوات.

وأما في التعادل الرأسي تكون الصيغ الامتحانية مختلفة في مستوى صعوبتها ولكن تقيس نفس المحتوى أو المهارات، ولكن عينات الأفراد التي تطبق عليها هذه الصيغ مختلفة في مستوى قدرتها (علام، ٢٠٠٥؛ والمدانات، ٢٠١٢). كما وتهدف المعادلة الرأسية إلى وضع تدريج جميع نماذج الامتحان ذات الصعوبة المختلفة على ميزان موحد، بحيث يسهل مقارنة درجات الطلبة عبر المستويات التعليمية المختلفة (المحرزي، ٢٠١٤).

متطلبات إجراء معادلة الامتحانات

تهدف معادلة الامتحانات لايجاد علاقة بين صور الامتحان، بحيث يمكن استخدام أي صورة في تقييم الأفراد واتخاذ قرارات وإصدار أحكام حول درجات الصور، ولكن هنالك متطلبات لإجراء المعادلة والتي اتفق عليها بعض الباحثين (علام، ٢٠٠٥؛ Hambleton & Swaminathan, ; Lord, 1980 Kolen & Brennan, 2004; 1985)، والتي نتطرق لها دون التعرض للمعادلات الرياضية المرتبطة بها وهي:

- أن تقيس الامتحانات المراد معادلتها نفس السمة (القدرة)، فمثلا في مادة الفيزياء يجب أن تشتمل الامتحانات على مفردات تتعلق بفروع مادة الفيزياء، وأن يوجد قاسم مشترك بينها، وكذلك يشتمل كل امتحان على مزيج من نفس المحتوى.
- العدالة أو المساواة (Equity)، وهو أن تكون خصائص التوزيع التكراري المشروط للدرجات عند مستوى معين من مستويات القدرة (θ) للامتحان الأول بعد تحويل الدرجات نفسها للتوزيع التكراري المشروط للدرجات عند المستوى نفسه من القدرة (θ) للامتحان الثاني.
- اللاتباين (Invariance) في مجتمع الدراسة، أي أن الدالة التي تستخدم في تحويل درجات امتحان ما، يجب ان يبقى كما هو بصرف النظر عن عينة المفحوصين، وبمعنى آخر استقلالية دالة التحويل عن البيانات المستمدة من مجتمع معين.
- تحويل الدرجات من صورة إلى اخرى قابلة للانعكاس، أي أن الدالة التي تحول بها الدرجات من الصورة الاولى للصورة الثانية من الامتحان، والدالة التي تحول بها الدرجات من الصورة الثانية للصورة الأولى من الامتحان تعطيان نفس النتائج وهذا ما يسمى التماثل (Symmetry).

- أن تتمتع الامتحانات بثبات عالي (Reliability).

الشروط المذكورة سابقاً تستخدم كأساس لضبط إجراءات عملية معادلة الامتحانات، حيث لا يوجد هنالك بنية نظرية لعملية المعادلة، وقد لا تحقق كلها وقد تكون غير عملية، ولكن لابد العمل على تحقيق هذه الشروط بقدر الإمكان (السيد، ٢٠١٢).

اهتمت الدراسات السابقة بطرق المعادلة ومعرفة أفضل طرقها، مع وجود برمجيات حاسوبية تتعامل مع الاستجابات سواء كانت ثنائية أو متدرجة، ومنها دراسة أينل وأركان (İnal & Arikan, 2017) للتحقق من حالة الثبات في المجموعة طبقاً لـ تاكر وليفين ومعادلة الدرجات بطريقة المعادلة الخطية. حيث تم استخدام الكتيبين الرابع والسادس من الاختبار الفرعي للرياضيات (PIAS). وقد تم معادلة الكتيبات وفقاً لمتغيري المجموعة والنوع الفرعي، ثم تم حساب ثبات المجموعة لكل حالة وحساب قيم Weighted Mean Squares Error (WMSE) وتم تطبيق مؤشرات الثبات في مجموعتي Root Expected Mean Square Difference (REMSD) و Root Mean Square Difference (RMSD). وأشارت نتائج الدراسة إلى أنه عندما تمت مقارنة قيم (WMSE) التي تم الحصول عليها وفقاً لطرق المعادلة فقد لاحظ تاكر طرق معادلة الدرجات فيما يتعلق بالمجموعة الكاملة والمجموعات الفرعية للنوع أنتجت أدنى مستوى من الخطأ، وعندما تم فحص قيم (REMSD) و(RMSD) وفقاً للمجموعات الفرعية للنوع بواسطة طرق التعادل الخطية، فقد وجد أن قيمة الثبات للمجموعة أصغر من قيمة المعيار في طريقة المعادلة لتاكر، بينما كانت أكبر من القيمة المعيارية لطريقة معادلة ليفين.

حيث أجرت الشمري (٢٠١٥) دراسة لمعادلة صور من امتحان القدرات المعرفية لدى طلبة الثانوية العامة بالمملكة العربية السعودية، وتألف كل نموذج من ٩٦ مفردة ثنائية التدرج، واستخدمت المعادلة الأفقية وفق نموذج راش من خلال عمل مقارنة مع النظرية الكلاسيكية بالطريقتين الخطية والمئينية حيث بلغت معاملات الصدق التقاطعي أقل مايمكن للنموذج أحادي المعلمة مقارنة بالطريقتين، ومن ناحية الخطأ المعياري للمعادلة كان النموذج أحادي المعلمة أكبر فاعلية من طريقتي المعادلة الخطية والمئينية.

وأجرى عليمات (٢٠١٣) دراسة لمعادلة درجات ثلاث صور مختارة من الامتحان الوطني للغة الانجليزية في الأردن وفق نظرية الاستجابة للمفردة والنظرية الكلاسيكية، مستخدماً المعادلة الخطية والرتب المئينية، وتوصلت النتائج إلى أن طرق المعادلة للمفردات المطابقة لنماذج نظرية الاستجابة للمفردة باستخدام طريقتي المعادلة التي ترتكز عليها النظرية الكلاسيكية (الخطية والمئينية) كانت أكثر دقة وفاعلية في النموذج أحادي المعلمة، تليها معادلة الصور المكونة للمفردات المطابقة للنموذج الثنائي ثم الصور المكونة من المفردات المطابقة للنموذج الثلاثي المعلمة، وعند استخدام محك الصدق التقاطعي للمقارنة بين فاعلية طرق المعادلة وفق النظريتين تبين أن المعادلة وفق النموذج أحادي المعلمة كانت الأكثر فاعلية تلتها المعادلة وفق النموذج الثنائي المعلمة فالمعادلة وفق النموذج ثلاثي المعلمة ثم المعادلة المئينية والمعادلة الخطية.

وأجرى كيم وهانسون (Kim & Hanson, 2002) دراسة لمعادلة نموذجين من امتحانات الرياضيات (ACT) من نوع اختيار من متعدد، حيث استخدمت طريقة منحنى خصائص المفردة بالاعتماد على تصميم المجموعات العشوائية المتكافئة، وتم تطبيق النموذج الأول عام ١٩٩٧ على ٢١٦٣ فرداً، وطبق النموذج الثاني عام ١٩٩٨ على ٢٩٠٩ فرداً، وأظهرت نتائج الدراسة فاعلية تصميم المجموعات العشوائية المكتافئة.

كما أجرى يوآن ورفاقه (Yuan, et al., 1999) دراسة لمعادلة معاملات الامتحان وفق نظرية الاستجابة للمفردة للامتحانات ذات المفردات الثنائية والمتعددة التدريج حيث استخدم تصميم المفردات المشتركة والتي تحتوي امتحاناتها مجموعة من المفردات المشتركة وطبقت على

مجموعتين من المفحوصين. وأشارت النتائج إلى اقتراب قيمة المؤشرات الإحصائية للتحيز في معادلة المعاملات وعبر كل الظروف من الصفر.

مشكلة الدراسة

هنالك مشكلات يواجهها المختصون في تصميم ووضع الامتحانات وعدالتها للفئات المستهدفة، وخاصة عند اختلاف مواعيد تطبيق الامتحانات من وقت لآخر، والذي بدوره يعمل على استنزاف الكثير من الجهد والوقت لإعدادها. ولذلك فلابد أن تتصف هذه الامتحانات بالعدالة والصدق والموضوعية في قياس سمة التحصيل. حيث أشار أنجوف (Angoff, 1986) إلى أن تشابه نماذج الامتحان في المحتوى المقاس من خلال جدول المواصفات لا يضمن تساوي مستوى الصعوبة لها، وعلى الرغم من ذلك فإن معدي الامتحانات يحاولون قدر الإمكان بناء نسخ امتحانية متقاربة، إلا أنه مازال هنالك اختلاف في مستوى صعوبتها وفي ظروف تطبيقها وتصحيحها، وربما تكون الامتحانات المقدمة للطلبة غير متعادلة من عام لآخر. وقد أشار الشريفين (٢٠٠٩) إلى أهمية إجراء معادلة لنماذج الامتحان المختلفة، حيث أن عدم تعادل درجات النماذج قد يؤدي إلى اتخاذ قرارات خاطئة للحكم على مستويات الطلبة، وبالتالي فإنه بات من الضروري جداً إجراء المعادلة لدرجات الامتحانات بواسطة نظرية الاستجابة للمفردة لما يترتب عليها من قرارات متعلقة بمستوى تحصيل الطلبة بشكل عام، وكذلك في تحديد نسب القبول للالتحاق بمؤسسات التعليم العالي، وبالتالي مساعدة المختصين في الحصول على تفسيرات صحيحة لدرجات تطبيق عدة صور لامتحان يقيس نفس السمة.

وقد صممت امتحانات دبلوم التعليم العام في سلطنة عمان كونها أحد أدوات التقويم الختامية، والتي تنفذ في نهاية الفصل الدراسي للتعرف على مستويات الطلبة ولمعرفة مدى امتلاكهم للمعارف والمهارات المتعلقة بما درسه الطالب في المادة الدراسية، والتي يتم إعدادها من قبل فريق فني، وفي كل عام تختلف الصورة الجديدة للامتحان في مفرداتها عن مفردات الصورة السابقة، ولكن لا يوجد دليل على أن هذه الصور من الامتحانات متعادلة، ولا يمكن عمل مقارنة علمية بين مستويات الطلبة عبر الصور المختلفة للامتحانات، حيث أشار تشولو وسيرسي (Chulu & Sireci, 2011) إلى أن عملية معادلة نماذج الامتحان ضرورية لضمان تساوي مستوى الصعوبة للنماذج واستخدامها بشكل تبادلي. وعليه فإن هذه الدراسة تحاول إجراء معادلة لدرجات صورتين تقيسان مخرجات مادة الفيزياء وبنفس الأوزان لمجالات امتحان دبلوم التعليم العام في سلطنة عمان باستخدام نظرية الاستجابة للمفردة، وعلى وجه الخصوص تقصي معادلة صور امتحانات الفيزياء للعامين الدراسيين (٢٠١٤/٢٠١٥) و(٢٠١٥ /٢٠١٦)، ووضعها في تدريج مشترك. وستحاول الدراسة الحالية الإجابة عن الأسئلة الآتية:

- **ما خصائص توزيع الدرجات الخام وتقدير معالم المفردات، لصورتي امتحان الفيزياء لطلبة دبلوم التعليم العام بسلطنة عمان؟**
- **ما قيم الدرجات المتعادلة باستخدام طريقة المعادلة بالدرجات الحقيقية في صورتي امتحان مادة الفيزياء لدى طلبة دبلوم التعليم العام بسلطنة عمان باستخدام النموذج الثلاثي والنموذج المتدرج؟**

أهمية الدراسة

نظراً لأهمية امتحانات دبلوم التعليم العام لمادة الفيزياء، كونها تساعد في الكشف عن المستويات التحصيلية للطلبة وما يترتب عليها قبول الطلبة في مؤسسات التعليم العالي، جاءت هذه الدراسة - وهي من الدراسات الحديثة في سلطنة عمان- لتسليط الضوء على ضرورة إجراء معادلة لامتحانات الفيزياء في دبلوم التعليم العام، وكذلك تسلط الضوء على توظيف نظرية الاستجابة للمفردة في إجراء المعادلة مما قد يوفر معلومات عن طريقة حديثة للمعادلة تعين الباحثين وطلبة الدراسات العليا.

وتعتبر هذه الدراسة مهمة ايضاً لتوفيرها معلومات لوزارة التربية والتعليم في سلطنة

عمان، حيث تساعد في تعريف المختصين ببناء الامتحانات والمقاييس بموضوع معادلة الامتحانات باستخدام نظرية الاستجابة للمفردة وطرقها للحصول على مؤشرات هامة من خلال الدرجات المعادلة لصور الامتحان، وتفسير درجاتها واتخاذ قرارات حول الرسوب والنجاح. وتفتح هذه الدراسة الباب لإجراء معادلات لنماذج الاختبارات المختلفة في سلطنة عمان، سواء كان لاختبارات دبلوم التعليم العام أو لاختبارات المراحل الدراسية المختلفة.

محددات الدراسة

- ✓ استخدام النموذج اللوجستي ثلاثي المعلم لمعايرة المفردات الموضوعية.
- ✓ استخدام النموذج المتدرج لمعايرة المفردات المقالية.
- ✓ استخدام تصميم المجموعات العشوائية (المتكافئة) لجمع البيانات في المعادلة.
- ✓ نتائج طلبة دبلوم التعليم العام لمادة الفيزياء لعامي (٢٠١٤/٢٠١٥) و(٢٠١٥/٢٠١٦) بسلطنة عمان.
- ✓ تعتمد دقة النتائج على دقة تقدير البرمجيات المستخدمة في الدراسة الحالية.

مصطلحات الدراسة

امتحان دبلوم التعليم العام لمادة الفيزياء: أداة تقويم ختامية تطبق بنهاية الفصل الدراسي، ويتطلب لإعداده مواصفات تفصيلية بحيث تأخذ في الاعتبار تحديد مخرجات المنهج وتحديد مستويات التعلم مقرونة بالأوزان النسبية لكل مستوى من مستويات الأسئلة وعدد الجزئيات والمفردات، ويُعد متطلب قبول للدخول في مؤسسات التعليم العالي وما في مستواها (وزارة التربية والتعليم، ٢٠١٥).

المفردة ثنائية الاستجابة: المفردات التي تنقسم فيها الإجابة إلى طريقة يمنح فيها الطالب علامة ١ للاستجابة بطريقة صحيحة، والعلامة صفر عند استجابته بطريقة خاطئة (الشريفين، ٢٠٠٣).

المفردة متعددة الاستجابة (Polytomous Response Item): التي يتم الإجابة عنها في عدة خطوات، بحيث توضع درجات على الخطوات التي يتمها الطالب بشكل صحيح، وعادة ما تكون هذه المفردة في الأسئلة المقالية (الشريفين، ٢٠٠٣).

معلمة صعوبة المفردة: النقطة على متصل القدرة (θ)، والتي تكون عندها احتمالية استجابة الفرد (المفحوص) صحيحة عند تلك المفردة مساوية لـ (1+ci/2) ويرمز لها بالرمز (bi)، وتتراوح قيمة معلمة الصعوبة عملياً -٣ و+٣ (Baker, 2001).

معلمة تمييز المفردة: مدى قدرة المفردة على التمييز بين المفحوصين ذوي القدرات التي تقع دون موقع المفردة على متصل القدرة، وأولئك التي تقع قدراتهم فوق موقع المفردة، ونظرياً تقع على الفترة (-∞،+∞) وعمليا على الفترة ٠، ٢ (Hambleton & Swaminathan,1985). وعملياً تعرف بأنها ميل المماس لمنحنى خاصية المفردة عند النقطة التي تحدد صعوبتها (Baker, 2001).

معادلة درجات الامتحانات Equating Test Scores: عملية إحصائية يتم فيها وضع درجات الامتحانات المختلفة والتي تقيس نفس السمة على مقياس واحد مشترك بحيث يحصل المفحوص على نفس التقدير للسمة التي يقيسها الامتحان بغض النظر عن اختلاف صوره (بركات، ٢٠١٠). ويعرفها هولند وروبن (Holland & Rubin, 1982) بأنها إجراء تعديلات عددية على درجات صور الامتحانات للتعويض عنها عند اختلاف الصعوبة من صورة لأخرى.

وفي هذه الدراسة يقصد بمعادلة درجات امتحان مادة الفيزياء تحويل درجات صورة امتحان مادة الفيزياء (٢٠١٥/٢٠١٦) إلى نموذج مرجعي، من خلال حساب الدرجات الخام لمجموعة الإجابات الصحيحة لمفردات أحد صورتي الامتحان وتحويل درجات أحد النماذج (الصور) إلى النموذج المرجعي الذي اعتمده الباحثين، وهو صورة امتحان العام الدراسي (٢٠١٤/٢٠١٥).

المنهجية والاجراءات

بيانات الدراسة

تكوّن مجتمع الدراسة من طلبة دبلوم التعليم العام لجميع محافظات سلطنة عمان بالمدارس الحكومية الذين خضعوا لامتحان مادة الفيزياء في العام الدراسي ٢٠١٤/٢٠١٥ للصورة الأولى وعددهم ١٨٩٦٦(٩٧١٠ ذكور، ٩٢٥٦ إناث)، والعام الدراسي ٢٠١٥/٢٠١٦ للصورة الثانية وعددهم١٧١٧٥(٨٣٦٠ ذكور، ٨٨١٥ إناث) (المديرية العامة للتقويم التربوي، ٢٠١٥).

أداة الدراسة

تكونت من امتحانين لدبلوم التعليم العام لطلبة الصف الثاني عشر لمادة الفيزياء للفصل الدراسي الأول للعامين الدراسيين (٢٠١٤/٢٠١٥) و(٢٠١٥/٢٠١٦)، وتألفت الصورة الأولى من ٤٠ مفردة، والصورة الثانية من ٣٦ مفردة، وعدد المفردات الموضوعية ١٤ لكل صورة من نوع الاختيار من متعدد بأربعة بدائل وتوضع الدرجات ٠ أو ٢، وبالنسبة للمفردات المقالية وتوضع لها درجات متدرجة (متعددة)، وبلغت ٢٦ مفردة لعام ٢٠١٤/٢٠١٥، و٢٢ مفردة لعام ٢٠١٥/٢٠١٦، ويتم إعدادها وفق وثيقة تقويم تعلم الطلبة في مواد العلوم بشكل مركزي من قبل فريق فني، وتتراوح الدرجات الخام لامتحان دبلوم التعليم العام لمادة الفيزياء بين (صفر) إلى ٧٠، ويوضح مواصفاتها جدول ١ (وزارة التربية والتعليم، ٢٠١٥).

تصميم الدراسة

اعتمدت الدراسة تصميم المجموعات المتكافئة Equivalent-Group Design، من خلال استخدام بيانات صورتي امتحان مادة الفيزياء لعامين دراسيين مختلفين والتي تقيس نفس السمة، حيث طُبقت صور الامتحان على مجموعتين من المفحوصين. مع مراعاة تماثل قدرات المجموعتين بأقصى درجة ممكنة كي لا يدخل عامل التحيز في عملية المعادلة، لكن هذا الأمر يصعب تحقيقه بشكل عام، ولكن يمكن تحقيقه نسبياً عندما تكون حجوم عينات المفحوصين كبيرة (Livingston, 2004).

الإجراءات والمعالجة الإحصائية: للإجابة عن أسئلة الدراسة تم القيام بالإجراءات والتحليلات الإحصائية وتتمثل في الآتي:

أولاً: تهيئة بيانات صورتي الامتحان

- الحصول على بيانات صورتي امتحان الفيزياء للأعوام الدراسية ٢٠١٤/٢٠١٥م (الصورة الأولى) و٢٠١٥/٢٠١٦م (الصورة الثانية).

تعديل على صورة امتحان العام الدراسي ٢٠١٤/٢٠١٥ من حيث عدد المفردات الجزئية، حيث كانت المفردات ٤٠ مفردة، ولغرض المعايرة تم دمج درجات ٤ مفردات مقالية (تقع في مستوى التطبيق)؛ ليصبح عدد المفردات الكلي ٣٦ للصورة الأولى، وقد روعي عند دمج الدرجات للمفردات الأربع مستوى التعلم الذي تقسيه، ليصبح عدد المفردات في كلا الامتحانيين متساوٍ ٣٦ مفردة لكل امتحان، مع بقاء تساوي الدرجة الكلية للامتحانين ٧٠ درجة لكل امتحان.

جدول١

توزيع الموضوعات ونسب مستويات الأهداف لامتحان مادة الفيزياء حسب وثيقة تقويم تعلم الطلبة لمواد العلوم

الأهداف / المحتوى	معرفة (٣٠)	تطبيق (٥٠)	استدلال (٢٠)*	المجموع
الكهرباء ١٦%	٤	٥	٢	١١
الحث الكهرومغناطيسي ٣٤%	٧	١٢	٥	٢٤
الموجات الميكانيكية ١٦%	٣	٦	٢	١١
الصوت ٣٤%	٧	١٢	٥	٢٤
المجموع	٢١	٣٥	١٤	٧٠

*قدرات عليا حسب تصنيف وزارة التربية والتعليم

ثانياً: إجراءات معايرة صورتي الامتحان

- تهيئة درجات الصورتين لتتوافق مع برنامج (Multilog7.03)، من خلال إعادة ترميز درجات المفردات بضرب الدرجات الخام بـالرقم ٢، لغرض التخلص من الأنصاف في الدرجات وعليه توحدت المسافات بين درجات المفردات، وأصبحت الدرجة الكلية لكل امتحان ١٤٠ درجة.
- معايرة درجات (بيانات) كل صورة من امتحانات الفيزياء باستخدام برنامج (Multilog7.03) لتقدير معالم المفردات (الصعوبة-التمييز-التخمين)، والتأكد من ثبات الصور حسب مخرجات البرنامج، حيث أشارت النتائج إلى ثبات الصورتين، حسب معايرة مفردات الامتحانات بواسطة البرنامج. وبلغ معامل ثبات صورة العام الدراسي ٢٠١٤/٢٠١٥م (٠,٩٢)، وثبات صورة العام الدراسي ٢٠١٥/٢٠١٦م (٠,٩٢) وهذه معاملات ثبات مرتفعة تدل على ثبات استجابات المفحوصين لصورتي الامتحان، وهذا يتفق مع أحد شروط المعادلة أن تتميز صور الامتحان المراد معادلتها بالثبات (عليمات، ٢٠١٣).
- تقدير قدرات الأفراد في صورتي الامتحان وذلك باستخدام برنامج (Multilog7.03).
- اختيار التصميم المناسب لجمع البيانات لمعادلة درجات الصورتين، وفي هذه الدراسة استخدم تصميم المجموعات المتكافئة.
- التحقق من الافتراضات الرئيسة التي تقوم عليها نظرية الاستجابة للمفردة، والتحقق من مطابقة المفردات والأفراد المطابقين لنموذج استجابة المفردة، من خلال استخراج البواقي المعيارية لكل نموذج على حدة، ومقارنة المؤشرات بقيمة مربع كاي عند مستوى دلالة ٠,٠٥ وبدرجة حرية ١، وذلك من خلال عمل الإجراءات الآتية:

فحص افتراض أحادية البُعد لصورتي الامتحان: وتفترض نماذج نظرية الاستجابة للمفردة أحادية البعد أن مفردات الاختبار تقيس سمة واحدة، وهي التي هدفت لقياسها، ويسمى هذا الافتراض بأحادية البُعد Unidimensionality، فإذا كان العامل الأول يفسر تقريباً ٢٠% من التباين الكلي في الدرجات فإنه يدل على أحادية السمة، كما أنه إذا كان ناتج قسمة قيمة الجذر الكامن للعامل الأول على قيمة الجذر الكامن للعامل الثاني أكبر ٢ فإن ذلك يعني أن الامتحان يقيس بعد واحد (Reckase, 1985; Önder, 2007). وقد استخدمت الدراسة عدة طرق لذلك:

استخدام أسلوب التحليل العاملي (Factor Analysis): حيث استخدم التحليل العاملي للمكونات الرئيسية (Principal Component Analysis)، وتم حساب قيم الجذر الكامن (Eigen Value) ونسبة التباين المفسر (Explained Variance) لكل صورة كم هو مبين في جدول ٢

يتبيّن من جدول٢ وجود عامل واحد يفسر وبفارق كبير عن العوامل الأخرى التباين بين المفحوصين في صورتي الامتحان، وكانت قيمة الجذر الكامن للصورة الأولى من الامتحان للعامل الأول ٩,٢٣ قد فسر ما نسبته ٢٣,٠٧% من التباين الكلي، بينما كانت قيمة الجذر الكامن للعامل الثاني ٢,٠٠ وقد فسر ما نسبته ٥,٠٠% من التباين الكلي، ويلاحظ هنا أن نسبة التباين المفسّرة بالعامل الأول عالية وهذا مؤشر على أحادية البعد. وكذلك الحال بالنسبة للصورة الثانية، حيث بلغت قيمة الجذر الكامن للصورة الثانية للعامل الأول ٨,٩١ وقد فسر ما نسبته ٢٤,٧٦% من التباين الكلي لامتحان الصورة الثانية مما يرجح وجود عامل كامن سائد يفسر النسبة الكبرى من التباين.

كذلك تم استخدام تحليل البواقي Residual Analysis: من خلال برنامج NOHARM بغرض

التحقق من أحادية البعد للبيانات بإجراء تحليل البواقي Residual Analysis، حيث يتم فحص الموائمة بين عدد الأبعاد التي يتم تحديدها والبيانات من خلال حساب مصفوفة البواقي Residual Matrix لإيجاد قيمة الجذر التربيعي لمتوسطات مربعات البواقي.

وبين ميلر (Miller, 1991) أنه إذا كانت قيمة مؤشرRMSR قريبة من الصفر، فهذا دليل على تحقق أحادية البُعد للبيانات، وفي الوقت نفسه إذا كانت قيمة مؤشر تاناكا Tanaka Index قريبة من الواحد يُعد دليلاً آخر على أحادية البعد، وقد تحقق ذلك في صورتي الاختبار؛ وكان للصورة الأولى (RMSR= 0.01,TANAKA index=0.92)، وللصورة الثانية (RMSR= 0.02,TANAKA index=0.91).

التحقق من افتراض الاستقلال الموضعي: استخدم مؤشر (Q3) ويسمى مؤشر ين (Yen, 1984) لقياس الارتباط الموضعي بين مفردات الامتحان، وهو مؤشر يعبر عن العلاقة بين البواقي لزوج من المفردات بعد ضبط السمة المقاسة، ويتبع توزيع المؤشر توزيعاً عينياً بمتوسط حسابي (-1/ n-1) وانحراف معياري قدره (1/ n-3)، حيث (n) عدد المفردات، حيث استخدمت المعادلة التالية (البواقي= درجة الطالب – الدرجة المتوقعة) لجميع المفردات في برنامج الأكسل بشكل منفصل لكل صورة، ومن ثم حساب معاملات الارتباط للبواقي بمؤشر(Q3) لكل زوج من مفردات كل صورة بواسطة برنامج (SPSS)، وجاءت النتائج كالتالي -٠,٠٠١ للصورة الاولى و-٠،٠٠٢ للصورة الثانية، كما أن معاملات الارتباط للبواقي كانت أقل من ٠,٠٥، مما يدل على تحقق افتراض الاستقلال الموضعي (Greer, 2004).

التحقق من مطابقة الاستجابات على مفردات صورتي امتحان مادة الفيزياء لنماذج نظرية الاستجابة للمفردة: استخدم برنامج (Multilog7.03) المصمم لتحليل البيانات وفقاً لنماذج نظرية الاستجابة للمفردة وبرنامج الأكسل، وتحليل مفردات كل صورة ووضعها على مقياس مشترك للتحقق من مدى ملائمة النموذج الثلاثي والنموذج المتدرج للبيانات Goodness of fit. فقد أشار علام (٢٠٠٥) أن تقدير معالم أي نموذج من نماذج نظرية الاستجابة للمفردة استناداً إلى البيانات الامتحانية لا يعني أن هذا النموذج مطابق للبيانات، وبالتي لايمكن الاستفادة من النموذج في حالة عدم مطابقة البيانات. ولمعرفة مدى ملائمة بيانات الامتحان مع النماذج المستخدمة في الدراسة تم استخراج متوسط مؤشر البواقي المعيارية باستخدام دالة مربع كاي (CHISQ.DIST.RT) وبدرجات حرية مقدراها ١، كمؤشر لحسن مطابقة المفردات أو الأفراد، وتوصلت النتائج إلى مطابقة البيانات مع توقعات النماذج، وكانت قيمة مؤشر الصورتين غير دالة مقارنة مع قيمة مربع كاي، حيث كانت نتائج الصورة الأولى (sig=0.05، والصورة الثانية ($\chi2=0.36$, sig=0.05). ($\chi2=0.37$, .

جدول٢

الجذور الكامنة ونسبة التباين المفسر للعوامل المستخلصة في العينة الرئيسية لصورتي الامتحان

العامل	الصورة الأولى			الصورة الثانية		
	الجذر الكامن	نسبة التباين المفسر %	التباين المفسر التراكمي %	الجذر الكامن	نسبة التباين المفسر%	التباين المفسر التراكمي %
١	٩,٢٣	٢٣,٠٧	٢٣,٠٧	٨,٩١	٢٤,٧٦	٢٤,٧٦
٢	٢,٠٠	٥,٠٠	٢٨,٠٧	١,٩٤	٥,٣٩	٣٠,١٥
٣	١,٦١	٤,٠٤	٣٢,١٠	١,٢٤	٣,٤٤	٣٣,٥٩
٤	١,٣٩	٣,٤٧	٣٥,٥٨	١,٠٩	٣,٠٣	٣٦,٦١
٥	١,٢٤	٣,١١	٣٨,٦٨	١,٠١	٢,٨١	٣٩,٤٣
٦	١,١٧	٢,٩٢	٤١,٦٠			
٧	١,٠٧	٢,٦٦	٤٤,٢٦			

كما تم تحليل استجابات الطلبة الذين أجابوا عن مفردات الامتحان في كلا الصورتين، باستخدام مؤشر البواقي المعيارية للكشف عن مطابقة استجابات المفحوصين للنموذج الثلاثي والمتدرج. وبيّنت النتائج أن هناك ١٠ طلبة في الصورة الأولى و٤٢ في الثانية لم تكن استجاباتهم مطابقة لتوقعات النموذج (أي كانت هناك دلالة إحصائية). ويعتبر الأفراد غير مطابقين للنموذج، لأن استجاباتهم كانت بعيدة عن توقعات النموذج، فمثلا قد يجيب الطالب عن مفردة ما إجابة خاطئة على الرغم من صعوبتها دون مستوى قدرته، أو يجيب على مفردة ما إجابة صحيحة على الرغم من مستوى صعوبتها أعلى من مستوى قدرته (الشريفين، ٢٠٠٩).

نتائج الدراسة

للإجابة عن السؤال البحثي الأول: ما خصائص توزيع الدرجات الخام وتقدير معالم المفردات، لصورتي امتحان الفيزياء لطلبة دبلوم التعليم العام بسلطنة عمان؟ تم إجراء تحليلات على صورتي الامتحان قبل البدء في عملية تقدير معالم المفردات والقدرات للأفراد لصورتي الامتحان بواسطة نظرية الاستجابة للمفردة، وتم التحقق من ثبات الصورتين، وتحقق أحد شروط معادلة صور الامتحان. وهذا يتفق مع ما ذكره لورد (Lord, 1980) حيث أشار إلى أن صور الامتحان المراد معادلتها يجب أن تتصف بالثبات. كما تم استخراج الإحصاءات الوصفية للدرجات الخام والتي يلخصها جدول ٣.

يوضح جدول ٣ الإحصاءات الوصفية للدرجات الخام لصورتي الامتحان قبل إجراء عملية المعادلة، ويلاحظ اختلاف في المتوسط الحسابي لكلا الصورتين، كما أن قيم الالتواء والتفلطح للدرجات الخام كانت ضمن نطاق التوزيع الاعتدالي، كما تم استخدام اختبار (ت) للعينات المستقلة للكشف عن الفروق بين المتوسطات الحسابية لدرجات صورتي الامتحان، ويلاحظ وجود فروق ذو دلالة إحصائية بين المتوسطات الحسابية للدرجات الخام، وهذا يؤكد ضرورة إجراء المعادلة. وتتفق هذه النتيجة مع دراسة الشمري (٢٠١٥) وعليمات (٢٠١٣) والحواري (٢٠٠٨) الذين أشاروا إلى أن اختلاف المتوسطات الحسابية يدعو لضرورة إجراء المعادلة.

كما تم إجراء تقدير معالم المفردات لصورتي الامتحان باستخدام برنامج (Multilog7.03) لمعرفة الخصائص السيكومترية، حيث استخرجت الإحصاءات الوصفية لمعالم مفردات، وتوصلت النتائج إلى أن قيم معاملات التمييز للصورة الأولى تراوحت بين ٠,٥١ ، ٢,٧١ وبمتوسط حسابي ١,٦٨ ، بينما نجد أن قيم معاملات التمييز للصورة الثانية تتراوح بين ٠,٣٢ ، ٢,٢٦ وبمتوسط حسابي ١,٥٠ ، وحسب ما أشار هامبلتون وسوامينثان (Hambleton & Swaminathan, 1985) أن القيمة العملية للتمييز تقع على الفترة ٠ و ٢، وعليه فإن هنالك اختلاف في متوسطات معاملات التمييز لصورتي الامتحان.

وكذلك بينت النتائج أن قيم معاملات الصعوبة للصورة الأولى قد تراوحت بين -١,٠٩، ٢,٩٢ وبمتوسط حسابي ٠,٨٣، بينما نجد أن قيم معاملات الصعوبة للصورة الثانية للامتحان تتراوح بين -١,٣٧، ٢,٥٢ وبمتوسط حسابي٠,٤٤، وتتراوح قيمة معلمة الصعوبة عمليا -٣ و+٣ (Baker, 2001). ونتيجة هذا الاختلاف في متوسطات معاملات الصعوبة فإن ذلك يُعد مؤشراً على عدم تكافؤ الصورتين، ونجد أن قيم معاملات التخمين

جدول٣

الإحصاءات الوصفية للدرجات الخام لصورتي امتحان مادة الفيزياء قبل إجراء عملية المعادلة

صورة الامتحان	حجم العينة	المتوسط الحسابي	الانحراف المعياري	التفلطح	الالتواء	قيمة "ت" المحسوبة	القيمة الاحتمالية
الأولى	١٨٩٦٦	٢٧,٩٨	١٢,٧٤	٠,٦٥	-٠,٠٩	-٢٨,٣٢	<٠,٠٠٠
الثانية	١٧١٧٥	٣٢,٠١	١٤,٣٣	٠,٥٣	-٠,٤٩		

للصورة الأولى قد تراوحت بين ٠٢٢، ٠,٦ وبمتوسط حسابي ٠,٣٨، بينما نجد أن قيم معاملات التخمين للصورة الثانية تتراوح بين ٠,٠٧، ٠,٧٣ وبمتوسط حسابي٠,٤١. فإن ذلك يُعد مؤشراً على عدم تكافؤ صورتي الامتحان.

وبشكلٍ عام تشير النتائج إلى أن الصورتين تتمتع بخصائص سيكومترية مقبولة من حيث الثبات والصعوبة والتمييز، إلا أن اختلاف متوسطات معاملات الصعوبة، وكذلك خلال نتائج اختبار (ت) للعينات المستقلة التي أُجريت على الدرجات الخام تشير إلى ضرورة إجراء عملية معادلة لدرجات صورتي امتحان مادة الفيزياء لطلبة دبلوم التعليم العام.

وعلى الرغم من سعي مطوري الاختبارات في وزارة التربية والتعليم إلى تحقيق التكافؤ بينها من خلال الاعتماد على جدول مواصفات موحد، ألا أن تحقيق هذا التكافؤ بشكل فعلي يعتبر أمر في غاية الصعوبة وهذا ما دللت عليه المؤشرات الإحصائية سابقة الذكر، وهذا أيضاً يوافق ما ذكره الشريفين (٢٠٠٩) و(Baker, 2001) حيث أشاروا إلى أنه من الصعب الحصول على امتحانات متكافئة في الخصائص السيكومترية، فقد تكون هنالك عوامل مؤثرة، مثل تفاعل بعض المفردات ومستويات المفحوصين وغيرها. ولذا فإن عملية معادلة الدرجات ضرورية لضمان تساوي مستوى الصعوبة، ولاستخدامها بشكل تبادلي حسب ما أشار إليها تشولو وسيرسي (Chulu & Sireci, 2011).

للإجابة عن السؤال البحثي الثاني: ما قيم الدرجات المتعادلة باستخدام طريقة المعادلة بالدرجات الحقيقية في صورتي امتحان مادة الفيزياء لدى طلبة دبلوم التعليم العام بسلطنة عمان باستخدام النموذج الثلاثي والنموذج المتدرج؟ تم نقل نتائج المعايرة لقيم معاملات الصعوبة والتمييز والتخمين ومعالم القدرات للمفحوصين إلى البرنامج الإحصائي Polyequate(v0.5)؛ لإيجاد قيم الدرجات المتعادلة باستخدام طريقة المعادلة بالدرجات الحقيقية، باعتبار الصورة الأولى للعام الدراسي (٢٠١٤/٢٠١٥) كنموذج مرجعي يرجع إليه تفسير درجات الصورة الثانية، ويوضح جدول٤ الدرجات المعادلة للصورة الجديدة بالنسبة إلى الصورة الأولى (المرجعية).

يلاحظ من خلال جدول ٤ أن الدرجات الخام للصورة الجديدة (الثانية) من الامتحان تراوحت بين صفر و ٧٠ (الوسط الحسابي= ٢٨,٤، الانحراف المعياري=١٣,٢)، وتضمنت أنصاف الدرجة كون أن درجات المفردات كانت تحمل أعداداً صحيحة وأنصاف الدرجات، كما يلاحظ أن الدرجات المعادلة للصورة المرجعية متقاربة جداً من الدرجات الخام للصورة الجديدة عند الدرجات الأقل من ٢١,٥٠(الوسط الحسابي=٢٧,٥، والانحراف المعياري= ١٣,٦) وبينت النتائج تقارب الإحصاءات الوصفية للصورة المرجعية والصورة الجديدة (الثانية) وهذا ما أكده Kolen & (Brennan, 2004)، حيث أن الدرجات المعادلة في الصورة الجديدة بعد تحويلها تتقارب مع توزيع الدرجات في الصورة المرجعية. بينما نلاحظ وجود اختلاف بين قيم الدرجات المعادلة للصورة المرجعية والدرجات الخام للصورة الجديدة يكون واضحاً عند الدرجات الأعلى من الدرجة ٣٢,٥٠ حتى الدرجة المعادلة ٦٥، ثم يلاحظ تقارب الدرجات بعد ذلك.

ويوضح الشكل ١ الدرجات المعادلة للصورة الجديدة (الثانية) من امتحان مادة الفيزياء بالنسبة إلى الصورة الأولى (المرجعية)، حيث يلاحظ تقارب الدرجات كما تم الإشارة له مسبق. كما أن نتائج الدراسة الحالية تتفق مع المبدأ التي تقوم عليه المعادلة من حيث أنه تعمل على ضبط الاختلاف بين مستوى صعوبة صور الامتحان، بحيث إذا كان هنالك اختلاف في الصعوبة بين الصور فإن عملية المعادلة تعمل على تعديل لدرجات الصورة الجديدة ليتوافق مع تدريج الصورة المرجعية، وهذا يتفق مع ما ذكره المحرزي (٢٠١٥) وبركات (٢٠١٠)، حيث ينتج من المعادلة تساوي المتوسطات الحسابية والانحرافات المعيارية للدرجات المعادلة للنموذج الجديد قبل المعادلة مع درجات النموذج المرجعي، أي تقارب مستوى صعوبة صورتي الامتحان.

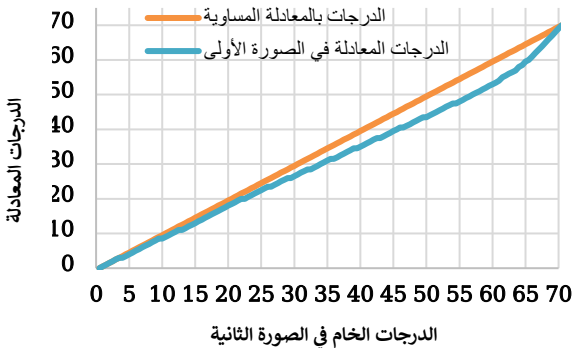
جدول ٤

الدرجات المعادلة للصورة الجديدة من امتحان مادة الفيزياء بالنسبة إلى الصورة الأولى (المرجعية)

الدرجة الخام في الصورة الثانية	الدرجة المعادلة في الصورة الأولى (المرجعية)	الدرجة الخام في الصورة الثانية	الدرجة المعادلة في الصورة الأولى (المرجعية)	الدرجة الخام في الصورة الثانية	الدرجة المعادلة في الصورة الأولى (المرجعية)	الدرجة الخام في الصورة الثانية	الدرجة المعادلة في الصورة الأولى (المرجعية)
٠	٠	١٨	١٦,٥	٣٦	٣٢	٥٤	٤٧,٥
٠,٥	٠,٥	١٨,٥	١٧	٣٦,٥	٣٢,٥	٥٤,٥	٤٨
١	١	١٩	١٧,٥	٣٧	٣٣	٥٥	٤٨,٥
١,٥	١,٥	١٩,٥	١٨	٣٧,٥	٣٣,٥	٥٥,٥	٤٩
٢	٢	٢٠	١٨,٥	٣٨	٣٤	٥٦	٤٩,٥
٢,٥	٢,٥	٢٠,٥	١٩	٣٨,٥	٣٤,٥	٥٦,٥	٥٠
٣	٣	٢١	١٩,٥	٣٩	٣٤,٥	٥٧	٥٠,٥
٣,٥	٣	٢١,٥	٢٠	٣٩,٥	٣٥	٥٧,٥	٥١
٤	٣,٥	٢٢	٢٠	٤٠	٣٥,٥	٥٨	٥١,٥
٤,٥	٤	٢٢,٥	٢٠,٥	٤٠,٥	٣٦	٥٨,٥	٥٢
٥	٤,٥	٢٣	٢١	٤١	٣٦,٥	٥٩	٥٢,٥
٥,٥	٥	٢٣,٥	٢١,٥	٤١,٥	٣٧	٥٩,٥	٥٣
٦	٥,٥	٢٤	٢٢	٤٢	٣٧,٥	٦٠	٥٣,٥
٦,٥	٦	٢٤,٥	٢٢,٥	٤٢,٥	٣٧,٥	٦٠,٥	٥٤
٧	٦,٥	٢٥	٢٣	٤٣	٣٨	٦١	٥٥
٧,٥	٧	٢٥,٥	٢٣,٥	٤٣,٥	٣٨,٥	٦١,٥	٥٥,٥
٨	٧,٥	٢٦	٢٣,٥	٤٤	٣٩	٦٢	٥٦
٨,٥	٨	٢٦,٥	٢٤	٤٤,٥	٣٩,٥	٦٢,٥	٥٦,٥
٩	٨,٥	٢٧	٢٤,٥	٤٥	٤٠	٦٣	٥٧
٩,٥	٨,٥	٢٧,٥	٢٥	٤٥,٥	٤٠,٥	٦٣,٥	٥٨
١٠	٩	٢٨	٢٥,٥	٤٦	٤٠,٥	٦٤	٥٨,٥
١٠,٥	٩,٥	٢٨,٥	٢٦	٤٦,٥	٤١	٦٤,٥	٥٩,٥
١١	١٠	٢٩	٢٦	٤٧	٤١,٥	٦٥	٦٠
١١,٥	١٠,٥	٢٩,٥	٢٦,٥	٤٧,٥	٤٢	٦٥,٥	٦١
١٢	١١	٣٠	٢٧	٤٨	٤٢,٥	٦٦	٦٢
١٢,٥	١١	٣٠,٥	٢٧,٥	٤٨,٥	٤٣	٦٦,٥	٦٣
١٣	١١,٥	٣١	٢٨	٤٩	٤٣,٥	٦٧	٦٤
١٣,٥	١٢	٣١,٥	٢٨,٥	٤٩,٥	٤٣,٥	٦٧,٥	٦٥
١٤	١٢,٥	٣٢	٢٨,٥	٥٠	٤٤	٦٨	٦٦
١٤,٥	١٣	٣٢,٥	٢٩	٥٠,٥	٤٤,٥	٦٨,٥	٦٧
١٥	١٣,٥	٣٣	٢٩,٥	٥١	٤٥	٦٩	٦٨
١٥,٥	١٤	٣٣,٥	٣٠	٥١,٥	٤٥,٥	٦٩,٥	٦٩
١٦	١٤,٥	٣٤	٣٠,٥	٥٢	٤٦	٧٠	٧٠
١٦,٥	١٥	٣٤,٥	٣١	٥٢,٥	٤٦,٥		
١٧	١٥,٥	٣٥	٣١,٥	٥٣	٤٧		
١٧,٥	١٦	٣٥,٥	٣١,٥	٥٣,٥	٤٧,٥		

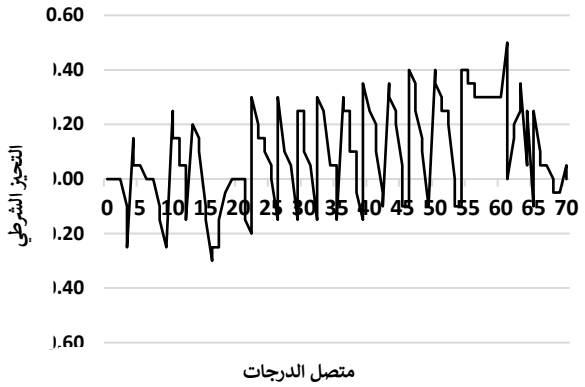
شكل ١

الدرجات المعادلة للصورة الجديدة (الثانية) من امتحان مادة الفيزياء بالنسبة إلى الصورة الأولى (المرجعية)



شكل ٢

التحيز الشرطي للمعادلة عبر متصل الدرجات



التحقق من دقة المعادلة

استخدمت طريقة التحيز الشرطي للمعادلة Bias للتحقق من دقة المعادلة، حيث تم اختيار ١٠ عينات عشوائية من الصورة الأولى و١٠ عينات عشوائية من الصورة الثانية بحيث أصبح عدد العينات الكلية ٢٠ عينة لصورتي الامتحان. وتم حساب التحيز الشرطي للمعادلة والذي مثل الفرق بين الدرجة المحسوبة المكافئة من التدريج المرجعي ومتوسط الدرجات المكافئة ويبين شكل ٢ التحيز الشرطي للمعادلة عبر متصل الدرجات. وتمثل النقاط في الشكل التحيز الشرطي للمعادلة Bias، حيث يلاحظ اقتراب المنحنى من الصفر تارة وابتعاده تارة أخرى على متصل الدرجات، وبالتالي فذلك يعدّ ذلك مؤشراً على درجة استقرار الدرجات المتعادلة، وهذا يتفق مع دراسة يوآن وآخرون (Yuan etal., 1999)، التي أشارت إلى اقتراب مؤشرات التحيز في المعادلة إلى الصفر.

وبشكلٍ عام يلاحظ أن قيمة متوسط التحيز الشرطي صغيرة جدا حيث بلغت ٠,٠٩، كما بلغت أعلى قيمة ٠,٥ وأقل قيمة -٠,٣، وتستخدم هذه الطريقة في الحكم على دقة عملية التعادل بالاعتماد على اقترابها من الصفر أو ابتعادها عنه، ويتم ذلك من خلال عرض رسومبيانية توضيحية للأخطاء على طول متصل الدرجة (Hou, 2007).

التوصيات والمقترحات

هدفت الدراسة الحالية إلى معادلة درجات صورتين من امتحان دبلوم التعليم العام لمادة الفيزياء بسلطنة عمان من خلال نظرية الاستجابة للمفردة، وكذلك وضع درجات المفحوصين على تدرج مشترك بحيث يسهل المقارنة بينهم، حيث بيّنت النتائج تقارب نتائج المعادلة الكلية لصورتي الامتحان.

أولاً: التوصيات

- ومن خلال النتائج التي توصلت إليها الدراسة، يمكن الخروج بالتوصيات الآتية:
- الاستفادة من إجراءات الدراسة الحالية وتطبيقها على نماذج لبعض امتحانات شهادة دبلوم التعليم العام؛ نظراً لأهمية هذه الامتحانات وأهمية القرارات المبنية عليها.
- استخدام نتائج معادلة امتحان دبلوم التعليم العام لمادة الفيزياء بالدرجات الحقيقية للمقارنة بين الطلبة باختلاف صور الامتحان حسب العام الدراسي.
- الاهتمام بموضوع معادلة درجات امتحانات دبلوم التعليم العام نظراً لأهميتها، ولما ينتج عنها من قرارات في تحديد درجات بعض المواد، كشرط

للالتحاق ببعض التخصصات في الجامعات.

- استخدام الدرجات المعادلة لاتخاذ قرارات نجاح الطلبة ورسوبهم.

ثانياً: المقترحات

في ضوء النتائج التي توصلت إليها الدراسة، يمكن الخروج بالمقترحات الآتية:

- إجراء دراسات على البيانات المتوفرة لدى وزارة التربية والتعليم، والعمل على نشر مبادئ معادلة صور الامتحان؛ لما لها من أهمية لدى الأعضاء الفنيين بمركز القياس والتقويم التربوي.
- دراسة نتائج معادلة الدرجات باختلاف الجنس (النوع) بواسطة صور من امتحانات بعض المواد الدراسية.
- عمل بنك أسئلة من مفردات صور الامتحان لمادة الفيزياء، والتي تتميز معالم مفرداتها (الصعوبة والتخمين والتمييز) بقيم مقبولة وفق نظرية الاستجابة للمفردة.

المراجع
References

بركات، مايا. (٢٠١٠). *أثر تصميمات المعادلة ومتوسط صعوبة الامتحانات وتوزيع القدرة على معادلة الامتحانات متعددة الأبعاد باستخدام نظرية الاستجابة للمفردة* (رسالة دكتوارة غير منشورة). جامعة القاهرة.

الحواري، أروى. (٢٠٠٨). *الخصائص السيكومترية لصور مختارة من امتحانات الرخصة الدولية لقيادة الحاسوب في الاردن ومعادلة درجاتها* (رسالة ماجستير غير منشورة). جامعة اليرموك.

الدوسري، راشد. (٢٠٠١). معادلة الامتحانات مفهومها، وطرقها، ومشكلات تطبيقها. *مجلة العلوم النفسية والتربوية، جامعة البحرين*، ٢ (٤)، ١٠٦-١٤١.

السيد، إيهاب. (٢٠١٢). تأثر بعض المتغيرات السيكومترية على دقة معادلة الامتحانات باستخدام تصميم المفردات المشتركة (دراسة محاكاة). *مجلة القراءة والمعرفة – مصر*، ١٣٠، ١١٤-١٤٢.

الشريفين، نضال. (٢٠٠٣). *مدى تحقق المعايير الفاعلية في معادلة امتحانين أحدهما ثنائي التدريج والأخر متعدد التدريج وفق نماذج النظرية الكلاسيكية والنظرية الحديثة في القياس والتقويم* (رسالة دكتوارة غير منشورة). جامعة عمّان العربية.

الشريفين، نضال. (٢٠٠٩). معادلة درجات نماذج مختلفة من امتحان الكفاءة اللغوية في اللغة الإنجليزية لدى طلبة جامعة اليرموك. *مجلة جامعة أم القرى للعلوم التربوية والنفسية*، ١ (٢)، ١١-٦٢.

الشمري، مها. (٢٠١٥). *معادلة درجات صور مختلفة من امتحانات القدرات المعرفية لدى طلبة الثانوية العامة في المملكة العربية السعودية* (رسالة ماجستير غير منشورة). جامعة اليرموك.

الشواورة، ياسين. (٢٠١٣). *دراسة مقارنة بين نموذج التقدير الجزئي والإستجابة في معادلة درجات الامتحانات* (رسالة دكتوارة غير منشورة). جامعة القاهرة.

عبد الهادي، نبيل. (٢٠٠١). *القياس والتقويم التربوي، واستخدامه في مجال التدريس الصفي(ط٢)*. دار وائل للنشر.

علام، صلاح. (٢٠٠٥). *نماذج الاستجابة للمفردة الامتحانية أحادية البعد ومتعددة الأبعاد وتطبيقاتها في القياس النفسي والتربوي*. دار الفكر العربي.

عليمات، مهدي. (٢٠١٣). *معادلة صور مختارة من الامتحان الوطني للغة الإنجليزية في الاردن وفق نظرية الامتحان ونظرية المفردة*، (رسالة ماجستير غير منشورة). جامعة اليرموك، الأردن.

الكوفحي، موسى. (٢٠١٣). *أثر طريقة تصفية المموهات في مفردات الاختيار من متعدد على معادلة الامتحان* (رسالة دكتوارة غير منشورة). الجامعة الأردنية.

المحرزي، راشد. (٢٠١٤). المعادلة بين نتائج طرق المعادلة الكلاسيكية لدرجات نماذج امتحان القدرات العامة باستخدام تصميم المجموعات المتكافئة. *مجلة رسالة الخليج، ١٣٤* (٣٥)، ١٥- ٤٢.

المحرزي، راشد (٢٠١٥). المفاضلة بين الدرجات المكافئة لنماذج اختبار القدرات العامة باستخدام طرق المعادلة الكلاسيكية في تصميم المفردات المشتركة بجماعات غير متكافئة. *مجلة العلوم التربوية والنفسية - البحرين*، ١٦(٣)، ٣٩٤- ٤٢٩.

المدانات، رائد. (٢٠١٢). مقارنة فاعلية طريقتى معادلة العلامات الحقيقية والمشاهدة فى معادلة الامتحانات باستخدام جذع مشترك ومجموعات غير متكافئة. *مجلة العلوم النفسية والتربوية- البحرين*، ١٣(٢)، ٣٦٥- ٣٩٤.

المديرية العامة للتقويم التربوي. (٢٠١٥). *احصائيات المديرية العامة للتقويم التربوي*. سلطنة عمان.

وزارة التربية والتعليم. (٢٠١٥). *وثيقة تعلم مواد العلوم. المديرية العامة للتقويم التربوي*. سلطنة عمان.

Abdel Hadi, N. (2001). *Educational measurement and evaluation, and its use in the field of classroom teaching (2nd ed)*. Wael Publishing House.

Al Koufhi, M. (2013).*The effect of the method of filtering the unknown in the multiple-choice vocabulary on the exam equation* [Unpublished doctoral dissertation].University of Jordan. (In Arabic)

Al Sayed, I. (2012). The influence of some psychometric variables on the accuracy of the examination equation using the common vocabulary design (simulation study). *Reading and Knowledge Journal, 130*, 114-142. (In Arabic)

Al Shamare, M. (2015). *Equating Scores for Different Forms of Ability Test among Secondary School Students in Saudi Arabia* [Unpublished master's thesis]. Yarmouk University. (In Arabic)

Al Sharifain, N. (2009). Equating Scors for Different forms of the English Language Proficiency Exam for Yarmouk University students *and Umm Al-Qura University .Journal of Educational and Psychological Sciences, 1*(2), 11-62. (In Arabic)

Al Sharifain, N. (2003). *Verification of Criteria of Efficiency in Equating a Dichotomous and a Polytomous Achievement Test Using Classical and Modern Test Theory Models* [Unpublished doctoral dissertation]. University of Jordan. (In Arabic)

Al Shawoura, Y. (2013). *A comparative study between the partial assessment model and the response in the equivalence of exam scores* [Unpublished doctoral dissertation]. Cairo University. (In Arabic)

Alimat, M. (2013). *Equivalence of selected images from the national examination of the English language in Jordan according to the examination theory and the singular theory* [Unpublished master's thesis]. Yarmouk University. (In Arabic)

Allam, S. (2005). *Models of response to the exam vocabulary one-dimensional and multi-dimensional and their applications in psychological and educational measurement*. Arab Thought House. (In Arabic)

Almehrizi. R. (2014). Equation between the results of the classic equivalency methods

for the scores of the General Aptitude Examination Models using the equivalent groups design. *Gulf Message Magazine, 134* (35), 15-42. (In Arabic)

Almehrizi. R. (2015). Differentiation among Equated Scores for General Abilities Test Forms Using Classical Equating Methods with Common-Item Non- Equivalent Groups Design. *Journal of Educational and Psychological Sciences - Bahrain, 16* (3), 394-429. (In Arabic)

Angoff, W. H. (1986). Some contributions of the College Board SAT to psychometric theory and practice. *Educational Measurement: Issues and Practice*, *5*(3), 7-11.

Arwa, A. (2008). *Sychometric Properties for Selected Forms of International Computer Driving License Tests in Jordan, and Equating Their Scores* [Unpublished master's thesis]. Yarmouk University. (In Arabic) (In Arabic)

Baker, F. B. (2001). *The basics of item response theory*. ERIC. http://ericae. Net/irt/baker.

Barakat, M. (2010). *The effect of equation designs, average difficulty of exams, and the distribution of ability to equate multidimensional exams using the singular response theory* [Unpublished doctoral dissertation].Cairo University. (In Arabic)

Chulu, B. W., & Sireci, S. G. (2011). Importance of Equating High Stakes Educational Measurements. *International Journal of Testing, 11*, 38-52.

Fan, X. (1998). Item Response Theory and Classical Test Theory: An Empirical Comparison of their Item/Person Statistics. *Educational and Psychological Measurement, 58*(3), 357-382.

General Directorate of Educational Assessment (2015). *Statistics of the General Directorate of educational evaluation*. Sultanate of Oman.

Greer, T. G. (2004). *Detection of differential item functioning (DIF) on the SATV: A comparison of four methods: Mantel- Haenszel, logistic regression, simultaneous item bias and likelihood ratio test* [Unpublished Doctoral Dissertation]. Houston University.

Hambleton, R. K., & Swaminathan, H. (1985). *Item response theory: Principles and Applications*. Kluwer Nijhoff.

Holland, P., & Rubin, D. (1982). *Test equating*. Academic press.

Hou, J. (2007). *Effectiveness of the Hybrid Levine Equipercentile and Modified Frequency Estimation Equating Methods under the Common-Item Nonequivalent Groups Design*. [Unpublished Doctoral Dissertation]. Iowa University.

Inal, H., Arikan A., (2017). An Investigation of Group Invariance in Test Equating According to Gender. *Journal of Measurement and Evaluation in Education and Psychology, 8*(1) 128-145.

Kim, S. & Hanson, B. (2002). Test equating under the multiple-choice model. *Applied Psychological Measurement*, *26*(3), 255 – 270.

Kolen, M., J. & Brennan, R.L. (2004). *Test equating, scaling, and linking: Methods and Practice.* Springer.

Livingston, S. A. (2004). *Equating Test Scores (without IRT)*. Educational testing service.

Lord, F. M. (1980). *Application of Item Response Theory to Practical Testing Problems*. N. J. Erlbaum.

Miller, T. R. (1991). *Empirical Estimation of Standard Errors of Compensatory MI Model Parameters Obtained from the NOHARM estimation program*. (ACT Research Report No. onr91-2). ACT Inc.

Ministry of Education (2015). *Science Materials Learning Document. General Directorate of Educational Evaluation*. Sultanate of Oman.

Önder, I. (2007). An Investigation of goodness of model data fit model veri uyumunun araştirilmasi. *Hacettepe Universitesi Eğitim Fakültesi Dergisi, (32)*, 210-220.

Raed, M. (2012).Comparing the Efficiency of True and Observed Score Equating Methods in Equating Tests Utilizing the Design with an Anchor Test and Nonrandom Groups. *Journal of Educational and Psychological Sciences, 13* (2), 365-394.

Rashid, H. (2001). Technical and Practical Issues on Test Equating. *Journal of Psychological and Educational Sciences, University of Bahrain*, *2*(4), 106-141.

Reckase, M. D. (1985). The difficulty of test items that measure more than one ability. *Applied Psychological Measurement*, *9*, 401-412

Reeve, B. (2002). *An Introduction to modern measurement theory*. Division of cancer control and population sciences, National Cancer Institute.

Skaggs, G., & Lissitz, R. (1986). IRT test equating: Relevant Issues and a Review of Recent Research. *Review of Educational Research, 56*(4), 495-529.

Suen, H. (1990). *Principles of test theories*. Lawrence Erlbaum.

Yen, W. M. (1984). Effect of local item dependence on the Fit and Equating Performance of the Three- Parameter logistic Model. *Applied Psycholigical Measurement*, *8*, 125-145.

Yuan, L. H. Lissitz, R. W., & Yang, Y. N. (1999, April 19-23). *Estimating IRT Equating Coefficients for Tests with Polytomously and Dichotomously Scored Items* [Conference session]. National Council on Measurement in Education, Montreal, Quebec, Canada.