Jerash for Research and Studies Journal مجلة جرش للبحوث والدراسات

Volume 10 | Issue 1 Article 4

2009

The Effectiveness of the Partial Credit Model in Calibrating Multi-Step Paragraphs

Mohammed Al-Zoubi
Jerash University, Jordan, MohammedAl-Zoubi22@yahoo.com

Follow this and additional works at: https://digitalcommons.aaru.edu.jo/jpu

Part of the Arts and Humanities Commons, Business Commons, and the Social and Behavioral Sciences Commons

Recommended Citation

Al-Zoubi, Mohammed (2009) "The Effectiveness of the Partial Credit Model in Calibrating Multi-Step Paragraphs," *Jerash for Research and Studies Journal مجلة جرش للبحوث والدراسات*: Vol. 10 : Iss. 1 , Article 4.
Available at: https://digitalcommons.aaru.edu.jo/jpu/vol10/iss1/4

This Article is brought to you for free and open access by Arab Journals Platform. It has been accepted for inclusion in Jerash for Research and Studies Journal مجلة جرش للبحوث والدراسات by an authorized editor. The journal is hosted on Digital Commons, an Elsevier platform. For more information, please contact rakan@aaru.edu.jo, marah@aaru.edu.jo, u.murad@aaru.edu.jo.

جرش للبحوث والدراسات، المجلد العاشر، العدد الأول ٢٠٠٥م

فاعلية نموذج التقدير الجزئي في معايرة فقرات متعددة الخطوات

محمد محمود البشير الزعبي❖

تاريخ تقديم البحث : ٢٠٠٦/٤/٢٥ تاريخ قبوله للنشر : ٢٠٠٧/١/١٧

Abstract

This study aimed at assessing the efficiency of the partial Credit Model, in Calibration of polytomous items

For the purposes of the study,an achievement test in chemistry for the seconnd secondary class was prepared. The test consisted of 20 polytomous items which differed in the number of steps in each item. The test was administered to a sample of 771 male and female students.

Mor specifically, the follwing findings were of particular importance:

1. Item difficulty increased as the number of the item steps increased.
2. There were no statistically significant differences in estimates of a sample of 11 items selected from the test when calibrated in four different groups of student.
3. There were statistically significant differences in estimated ability between the random, the high ability and the low ability groups, based on a sample of difficulty items and a sample of easy items.
4. The relative efficiency of the test increased in general when the number of the item steps increased
5. The relative efficiency of the difficult test compared to the total test was equal to one in the high ability group. However, the relative efficiency of the easy test compared to the total test increased in the low ability group.

Based on the findings of the study, a number of recommendations were proposed.

الملخص

هدفت هذه الدراسة إلى استقصاء فاعلية نموذج التقدير الجزئي في معايرة قدرة الأفراد ومعايرة صعوبة فقرات متعددة الخطوات. ولتحقيق أهداف الدراسة تم إعداد اختبار تحصيلي في مادة الكيمياء للصف الثاني الثانوي العلمي مؤلفا من ٢٠ سؤالا، تختلف في عدد خطوات الاستجابة. طبقت على عينة من طلبة الصف الثاني الثانوي العلمي بلغ عددهم ٧٧١ طالبا وطالبة.

وتم الحصول على تقدير لقدرات الأفراد، و تقدير لقيم صعوبة خطوات استجابة السؤال، والخطأ المعياري في تقديرها، وقياس المطابقة الخارجية والداخلية لعينة الأفراد والفقرات، وتم التوصل إلي النتائج التالية:

١. يزداد الخطأ المعياري لتقدير درجة صعوبة الأسئلة،بزيادة عدد خطوات الاستجابة

٢. عدم وجود فروق ذات دلالة إحصائية عند مستوى $\alpha \geq 0.01$ بين متوسطات الفروق المعايرة لتقديرات معالم صعوبة الأسئلة المحسوبة في المجموعات الأربع المختلفة.

٣. وجود فرق دال إحصائيا عند مستوى $\alpha \geq 0.01$ في متوسطات الفروق المعايرة بين تقديرات قدرات أفراد المجموعة العشوائية والعليا والمتدنية، المحسوبة في الاختبار الصعب والاختبار السهل.

٤. تزداد الكفاية النسبية للاختبار بشكل عام بزيادة عدد خطوات الاستجابة، وتتساوى الكفاية النسبية للاختبار الصعب مع الاختبار الكلي عند أصحاب القدرات العالية بينما تزداد الكفاية النسبية للاختبار السهل عند أصحاب القدرة المتدنية جدا.

وقد تمت مناقشة هذه النتائج، واستخلصت التوصية باجراء مزيد من الدراسات بما يتعلق باستخدام النموذج التقدير الجزئي في تقدير صعوبة الأسئلة في مقرارات التحصيل المختلفة.

❖ أستاذ مساعد/ كلية العلوم التربوية/ جامعة جرش الأهلية/ الأردن

Published by Arab Journals Platform, 2009

Jerash for Research and Studies Journal ???? ??? ?????? ?????????, Vol. 10 [2009], Iss. 1, Art. 4

مقدمة:

بدأ الاهتمام بمعايرة (Calibration) الفقرات وبنوك الأسئلة (Item Bank) في السنوات الأخيرة من قبل المؤسسات الحكومية، مثل مديريات التربية والتعليم، والقوات المسلحة، وبعض شركات النشر والتوزيع والاختبارات. وباشرت في تطوير فقرات بنوك الأسئلة، وخاصة بعد التقدم الكبير في استخدام برامج الحاسوب (Hambleton & Swaminathan). وبنك الأسئلة عبارة عن تجمع من الفقرات ذات خصائص سيكومترية معلومة، حيث تودع الفقرات في البنك بطريقة تسهل على مطور الاختبار سحب مجموعة منها تستخدم في بناء اختبارات مناسبة لأغراض مختلفة في القياس والتقويم التربوي وتحقق الهدف من الاختبار بدقة، وبأقل جهد. (علام).

هنالك أسلوبان لتحليل الفقرات التي يتم انتقاؤها في بنوك الأسئلة، أحدهما يعتمد على النظرية الكلاسيكية في القياس (Classical Test Theory). والأخر يعتمد على نظرية الاستجابة للفقرة في القياس (Item Response Theory)، وتشير المراجع (Crocker & Algina) في النظرية الكلاسيكية إلى مؤشرات تستخدم في انتقاء فقرات الاختبارات والمقاييس منها معاملات التمييز والصعوبة للفقرات، فتقدير معامل التمييز للفقرة من وجهة النظر الكلاسيكية هو معامل الارتباط بين الأداء على الفقرة والأداء على الاختبار بجميع فقراته. أما معامل صعوبة الأسئلة فيتم تقديره حسب النظرية الكلاسيكية في القياس بحساب نسبة النجاح على الفقرة.

لقد سيطرت النظرية الكلاسيكية لفترة غير بسيطة من الزمن على عملية بناء الاختبارات وتطبيقها وتحليل نتائجها وتفسيرها، وقد أشار باحثون إلى وجود جوانب قصور في النظرية الكلاسيكية، قد لا تمكنها من معالجة بعض القضايا الهامة في القياس والتقويم مثل بناء بنوك الأسئلة، بمفهومها الجديد (Hambleton & Swaminathan)، كماأن إحصائيات تحليل الفقرة بالطرق الكلاسيكية لا تقدم معلومات حول كيفية استجابة الفقرة من قبل الأفراد المختلفين في مستوى القدرة المقيسة بهذه الفقرة، وإحدى الطرق لتحليل فقرات الاختبار، والتي تعطي صورة أكثر شمولية حول استجابة الفقرة من قبل الأفراد المختلفين في مستوى القدرة تعرف بنظرية الاستجابة للفقرة (Item Response Theory IRT)، ويرى الباحثون أن نظرية الاستجابة للفقرة تتضمن افتراضات، ومعالجات يمكن بها التغلب على كثير من جوانب القصور في النظرية الكلاسيكية (Lord).

الافتراضات الأساسية في نظرية الاستجابة للفقرة.

تفترض نظرية الاستجابة للفقرة في القياس أنه يمكن التنبؤ بأداء الأفراد في اختبار نفسي أو تربوي معين من خلال خاصية أو خصائص مميزة للأفراد تسمى السمات (Traits) وبالطبع لا يمكن ملاحظة هذه السمات بشكل مباشر، وإنما يتم تقديرها (أو الاستدلال عليها) من خلال أداء الأفراد على مجموعة من فقرات الاختبار. ولهذا يطلق عليها السمات الكامنة (Latent Trait) أي أن نظرية الاستجابة للفقرة تفترض أن هناك دالة احتمالية تربط بين معلمتين على الأقل أحداهما يتعلق بالفرد والآخرى تتعلق بالفقرة التي يختبر بها (علام).

https://digitalcommons.aaru.edu.jo/jpu/vol10/iss1/4 2

وقد انبثق عن نظرية الاستجابة للفقرة مجموعة من النماذج تعرف باسم نماذج السمات الكامنة (Latent Trait Models). تهدف جميعها إلى تحديد العلاقة بين أداء الأفراد على مجموعة من فقرات الاختبار(وهو ما يمكن ملاحظته مباشرة) وبين السمات أو القدرات التي تكمن وراء هذا الأداء وتفسيره، ويرتكز كل نموذج من هذه النماذج على مجموعة من الافتراضات الأساسية التالية التي يجب أن تتوافر في البيانات المستمدة من الاختبار وهي:

١. افتراض أحادية البعد (Unidimensionality)

تفترض نماذج الاستجابة للفقرة وجود قدرة واحدة تفسر أداء الأفراد على الاختبار، وبالطبع فإن هذا الافتراض لا يمكن تحقيقه بشكل مطلق، بسبب وجود سمات أخرى بجانب السمة المقيسة بفقرات الاختبار (Hambleton & Swaminathan)، ولكي يمكن تفسير النتائج بدقة وصدق لابد من التأكد أن الاختبار يقيس سمة واحدة، فكلما زادت احتمالية تحقيق افتراض أحادية البعد، زادت دقة وصدق تفسير النتائج.

٢. افتراض الاستقلال الموضعي (Local Independence)

إن افتراض الاستقلال الموضعي يعني أن أداء الفرد على فقرة لا يؤثر إيجابا أو سلبا على استجابته على فقرة أخرى من فقرات الاختبار(علام)، أي أن محتوى الفقرة يجب أن لا يقدم دليلا للفرد يساعد في أجابته على الفقرات الأخرى في الاختبار، وعندما يتحقق افتراض الاستقلال الموضعي فإن استجابة الفرد لمختلف الفقرات في الاختبار تكون مستقلة إحصائيا أي أن احتمالية الحصول على نمط معين في الدرجات لفرد ما يساوي حاصل ضرب احتمالات حدوث هذا النمط لكل فقرة من الفقرات التي أجابها.

ويرى هامبلتون و سوامناثان أن افتراض الاستقلال الموضعي مكافئ لافتراض أحادية البعد (Hambleton & Swaminathan).

٣. منحنى خصائص الفقرة (Item Characteristic Curve ICC)

يشير افتراض منحنى خصائص الفقرة على أن العلاقة بين احتمالية الإجابة الصحيحة على الفقرة في الاختبار ومستوى قدرة الفرد المقيسة بفقرات الاختبار علاقة انحدار غير خطي، ولقد تم تطوير نماذج مختلفة في نظرية الاستجابة للفقرة، ويكون الاختلاف بين هذه النماذج في اختيار النموذج الرياضي لمنحنى خصائص الفقرة والذي يتميز بعدد من المعالم تصف خصائص الفقرة.

نموذج التقدير الجزئي:

وضع راش صيغة رياضية لاحتمالية إجابة الفرد ذي قدرة محددة على الفقرة ثنائية التدريج بشكل صحيح من خلال التفاعل بين قدرة الفرد وصعوبة الفقرة، وقد عمل كل من ماسترز (Masters) وساماجيما (Samagima)، واندريش (Andrich) وراش (Rasch) على توسيع نموذج راش للاستجابة الثنائية بحيث يشتمل مستويات من الأداء تقع بين الإجابة التامة على الفقرة والإجابة

الخطأ. فكان هناك نمطان من التفكير في مستويات الأداء, الأول يعالج درجة صعوبة كل مستوى على أنه أكثر من درجة صعوبة المستوى الذي يسبقه، ويسميه صعوبة الفاصل (Threshold)، ويتعامل مع هذا النمط من التفكير كل من نموذج ساماجيما للاستجابة المتدرجة (Graded Response Mod-el) ونموذج مقياس التقدير (rating scale) الذي طوره اندريش، والثاني يعالج كل خطوة من خطوات الاستجابة في الفقرة الواحدة على أنها ثنائية التدريج تأخذ العلامة واحد عندما ينجز الفرد هذه الخطوة أو صفرا عندما يخفق في إنجازها، وفي هذا النمط من التفكير يجب على مطور الاختبار أن يبني الفقرات بشكل منضبط مسبقا وبخطوات متسلسلة بحيث لا يمكن للفرد الوصول إلى مستوى معين من مستويات الإجابة إلا بعد أن يصل إلى المستوى الذي يسبقه مباشر. ويتعامل مع هذا النموذج من التفكير نموذج التقدير الجزئي.

ولقد اقتصرت معظم الدراسات السابقة في معايرة فقرات بنك الأسئلة بمفهومها الحديث باستخدام منهج النظرية الحديثة في القياس على الفقرات التي تكون الاستجابة عليها ثنائية التدريج (Dichotomous)، أي يعطى للطالب علامة واحدة إذا أجاب عن الفقرة بشكل صحيح ويعطى العلامة صفر إذا أجاب عن الفقرة بشكل خاطئ كما هو الحال في الفقرات من نوع الاختيار من متعدد (Choice-Multiple) والفقرات من نوع صح أو خطا (True - False) وقد يعود السبب في ذلك إلى توافر البرامج الحاسوبية في عمليات تصحيح الاستجابة على مثل هذا النوع من الفقرات، فقد قام كل من رايت وستون (Wright & Stone)، ورايت وبيل (Wright & Bell) باستخدام نظرية الاستجابة للفقرة في معايرة فقرات بنك الأسئلة ثنائية التدريج.

أما الدراسات التي تعاملت مع الاستجابات المتعددة فقد كانت أقل بكثير، من الدراسات التي استخدمت فيها الفقرات ثنائية التدريج فقد قام كل من هارس ولان وموسينسون (Harris Laan & Mossenson) في بناء اختبارات كتابية (قصصية) (Narrative Writing Tests) ومعايرتها باستخدام نموذج التقدير الجزئي. وتعد دراسة ماسترز (Masters) من الدراسات التي استخدم فيها منهج نظرية الاستجابة للفقرة في معايرة فقرات بنك الأسئلة والتي تكون الاستجابة عليها محددة مسبقا بعدد من المستويات المتدرجة (Polytomously) والتي يتم معالجتها إحصائيا باستخدام نموذج التقدير الجزئي (Partial Credit Model) في بناء فقرات بنك الأسئلة. وقد قام آدمز (Adams) بدراسة بين فيها فاعلية نموذج التقدير الجزئي في بناء فقرات اختبار تشخيصي، وتحديد المفاهيم الخاطئة عند الطلبة في مادة الرياضيات، والاستراتيجيات التي يستخدمونها في الإجابة على الفقرات والتي لا يمكن أجراؤها باستخدام النظرية التقليدية في القياس. وفي دراسة أجراها دود و كوخ (Dodd & Koch) لتقصي فاعلية نموذج التقدير الجزئي في توزيع دالة المعلومات (Informa-tion Function)، للفقرات التي يتغير فيها عدد و ترتيب خطوات الاستجابة. أما الدراسة التي أجرتها سوزان (Susan) لتقصي فاعلية ثلاثة نماذج في نظرية الاستجابة للفقرة في تقدير قدرات الأفراد ودرجة صعوبة الفقرات وهي النموذج اللوجستي أحادى المعلم (نموذج راش) ونموذج التقدير الجزئي، ونموذج الاستجابة المتدرجة، فقد قامت الباحثة من أجل تحقيق أهداف الدراسة بأعداد

اختبارات محكية المرجع (Criterion Referenced) تتضمن مزيجاً من أشكال الفقرات. من نوع الاختيار من متعدد، وفقرات تم بناءها بحيث تكون الاستجابة عليها متعددة الخطوات و بأربعة مستويات من الأداء، تم معايرة الفقرات بثلاث طرق مختلفة وباستخدام نماذج نظرية الاستجابة للفقرة موضوع الدراسة. وفي الدراسة التي أجراها بارك (park) كان الهدف الرئيسي من الدراسة مقارنة أثر كل من نموذج الاستجابة المتدرجة (Graded Response Model GRM) لساماجيما (Samajima).ونموذج التقدير الجزئي (Partial Credit Model PCM).على دقة تقدير قدرات الأفراد.

مشكلة الدراسة:

تنبثق مشكلة الدراسة من الاهتمام الكبير الذي أصبحت توليه المؤسسات التربوية المختلفة في بناء فقرات الاختبار والتي تتمتع بخصائص سيكومترية (صعوبة الفقرة و التمييز) لا تتأثر بخصائص عينة المفحوصين (القدرة). ويمكن تطبيقها على عينات مختلفة من المفحوصين دون تحيز إلى مجموعة دون غيرها. أو تطبيق عينات مختلفة من الفقرات على نفس عينة المفحوصين دون أن تتأثر خصائص الفقرة بخصائص قدرة المفحوصين(القدرة)، ولذلك تحاول هذه الدراسة التحقق من فاعلية نموذج التقدير الجزئي في معايرة الفقرات التي تكون الاستجابة عليها مدرجة بخطوات منضبطة ومتسلسلة مسبقا في مادة الكيمياء للصف الثاني الثانوي العلمي والتي تنسجم مع مستويات قدرة الأفراد.

وتتحدد عناصر المشكلة من خلال المعايير التي ستعتمد في هذه الدراسة لتقدير فاعلية نموذج التقدير الجزئي. ويمكن صياغة عناصر مشكلة الدراسة بعدد من الأسئلة التي تحاول الدراسة استقصاء إجابات عنها.

١- ما دلالات المطابقة لبيانات الاستجابة لفقرات الاختبار المتعددة التدريج لنموذج التقدير الجزئي؟

٢- ما تقدير قدرات الأفراد، وصعوبة الأسئلة، وصعوبة كل خطوة من خطوات استجابة السؤال الواحد باستخدام نموذج التقدير الجزئي؟

٣- ما قيم الأخطاء المعيارية في تقديرات قدرات الأفراد؟

٤- ما قيم الأخطاء المعيارية في تقدير درجة صعوبة الأسئلة، ودرجة صعوبة كل خطوة من خطوات استجابة السؤال الواحد ؟

٥- ما دلالات ثبات معالم الأسئلة (صعوبة الأسئلة) في عينتين مختلفتين من الأفراد؟

٦- ما دلالات ثبات تقدير القدرة لمجموعة واحدة من الأفراد في عينتين مختلفتين من الأسئلة؟

٧- ما الكفاية النسبية لاختبارات فرعية مأخوذة من بنك الأسئلة عند قيم مختارة للقدرة؟

الطريقة والإجراءات

مجتمع الدراسة :

تألف مجتمع الدراسة من طلبة الصف الثاني الثانوي العلمي في المدارس الحكومية لوزارة التربية

Jerash for Research and Studies Journal ???? ??? ?????? ?????????, Vol. 10 [2009], Iss. 1, Art. 4

والتعليم في مديرية تربية عمان الرابعة (عمل الباحث مدة ثلاث سنوات مشرفا في هذه المديرية لمادة الكيمياء)، وقد بلغ عددهم (٢٢٥٦) طالبا و طالبة منهم (١١٣٨) طالبا في ٢١ مدرسة. و(١١١٨) طالبة في ٣٠ مدرسة، و قد تم اختيار الصف الثاني الثانوي العلمي، لتحقيق الغرض من الاختبار في قياس قدرة الطلبة في مادة الكيمياء، ووفقا لمنهاج الكيمياء المقررة من قبل وزارة التربية والتعليم في الأردن فأن منهاج الصف الثاني الثانوي العلمي يمثل سقف عملية تعليم مادة الكيمياء و التي تشتمل على مهارات تراكمية متصلة يبدأ تعلمها من الصفوف التي تسبق هذا الصف.

عينة الدراسة :

تألفت عينة الدراسة من المدارس الثانوية التي تتضمن الصف الثاني تانوي علمي في مديرية عمان الرابعة من مناطق مختلفة، وقد تم اختيار المدارس من مناطق مختلفة وفق اعتبارات فرضتها طبيعة الدراسة من حيث أهمية الحصول على تباين في استجابات الطلاب عند كل خطوة من خطوات الاستجابة، ولهذا كان الاهتمام موجها لاختيار عينة غير متجانسة من حيث القدرة في مادة الكيمياء، بهدف الحصول على استجابات ذات مواصفات يمكن بوساطتها استخدام نموذج التقدير الجزئي، في الإجابة عن أسئلة الدراسة، وقد بلغ حجم العينة ٧٧١ طالب وطالبة.

أداة الدراسة:

نظرا لان هدف الدراسة بشكل عام تقدير فاعلية نموذج التقدير الجزئي، في معايرة فقرات متعدد الخطوات، فقد تم تصميم اختبار تحصيلي مكون من ٢٠ سؤالا، وقد تم عرض أسئلة الاختبار على عدد من معلمي الكيمياء للصف الثاني الثانوي العلمي ممن يحملون درجة البكالوريوس في مادة الكيمياء، وخبرة لا تقل عن خمس سنوات في تدريس مادة الكيمياء في مدارس مديرية عمان الرابعة، وكان عددهم ١٥ معلما ومعلمة. تم تكليف المعلمين في المديرية المعنية ممن يدرسون مادة الكيمياء بتفحص الأسئلة من حيث أن الاستجابة لها تتم بخطوات متسلسلة ومضبوطة بعدد محدد من مستويات الأداء بحيث لا يمكن للطالب الوصول الى المستويات العليا من الأداء دون الوصول الى المستويات الأدنى. وتمت مناقشة صيغ الأسئلة، وعدد خطوات الاستجابة، والإجابات المتوقعة من الطلبة، والفترة الزمنية اللازمة لتطبيق الامتحان مع المعلمين المذكورين وذلك من خلال ورشة عمل خصصت لهذه الغاية.

التجريب الأولي للاختبار:

تم تجريب الاختبار في مدرستين من مدارس مديرية عمان الرابعة. إحداهما للإناث وعدد الطالبات فيها ٤٠ طالبة، والثانية للذكور وعدد الطلاب فيها ٣٧ طالبا، حيث كان الهدف من التجريب الأولي، دراسة ما يلي

١. مدى مطابقة تدرج الاستجابة المتوقعة، مع استجابة الطلبة الحقيقية من خلال الإجابة على

https://digitalcommons.aaru.edu.jo/jpu/vol10/iss1/4 6

أسئلة الاختبار.

٢. تقدير تقريبي لمستوى صعوبة خطوات الأسئلة من خلال عدد الأفراد الذين وقعوا في كل مستوى من مستويات الاستجابة.

٣. تحديد الزمن المناسب لتطبيق الاختبار.

تم تجريب الاختبار، وتحليل نتائجه، ووجد أن علامات الطلبة الكلية تراوحت بين ٢ و٨٦ من العلامة الكلية ٩٠ والتي تساوي مجموع خطوات استجابة أسئلة الاختبار.

بالاستفادة من التجريب الأولي للاختبار، تم تطوير الاختبار في صورته النهائية، وتحديد خطوات الاستجابة لكل سؤال. عقدت ورشة العمل الثانية قبل يومين من تطبيق الاختبار لمعلمي ومعلمات المادة المشتركين في تصحيح أوراق الإجابة، وكان عددهم ١٨ معلما و معلمة. وتمت مناقشة الأمور الآتية:

١. تعليمات الاختبار والتي يجب على الطالب أن يدركها بعناية فائقة قبل البدء بعملية الاستجابة.

٢. آلية التصحيح إعطاء الطالب علامة واحدة عن كل خطوة ينجزها بنجاح، ولا يوجد علامات كسرية.

٣. إذا دمج الطالب خطوتين أو أكثر من خطوات الإنجاز بخطوة واحدة صحيحة، يعطى للطالب العلامة الكلية لجميع الخطوات التي تم إنجازها.

٤. تم عمل جدول خاص لكل طالب تفرغ فيه علاماته ولكل سؤال على حدة، ثم يتم جمع جميع علامات الأسئلة للحصول على العلامة الخام الكلية للطالب.

وزعت أسئلة الاختبار على مدارس العينة قبل يوم واحد من اجراء تطبيق الاختبار، حيث قام بالمراقبة على الطلبة معلمون لا صلة لهم بمادة الكيمياء، وقد تم قراءة تعليمات الاختبار للطلبة بعناية قبل البدء بالاجابة عن الاسئلة. تم تطبيق الاختبار في جلستين كل جلسة مدتها ساعة ونصف تفصل بينهما ربع ساعة استراحة، وقد تم التأكد من أن جميع الطلبة قد حاولوا الاجابة عن جميع الأسئلة حيث أن معظم الطلبة انتهوا من الاجابة قبل انتهاء مدة الاختبار المقررة، ولمعرفة مدى الدقة في عملية تصحيح أوراق الإجابة، تم إعادة تصحيح ٤٠ ورقة، بعد مضي شهر من تطبيق الاختبار، اختيرت عشوائيا من أوراق الإجابة، وحسب معامل الارتباط بين العلامات في التصحيحين وكان مساويا ٠,٩٨، ويمكن اعتبار ذلك مؤشرا قويا وصادقا على دقة عملية التصحيح.

المعالجات الإحصائية:

اعتمدت البرمجيات الإحصائية (SPSS, BIGSTEP) في تحليل النتائج ودراسة الافتراضات التي يجب أن تتوافر في البيانات حتى يتم استخدام نموذج التقدير الجزئي لماسترز في تقدير قدرات الأفراد ومعالم الفقرات والخطأ المعياري للتقديرات بالإضافة إلى فحص مطابقة تقدير قدرة الأفراد، وتقدير درجة صعوبة الفقرات لقيم توقعات النموذج المستخدم، وحساب قيم دالة المعلومات للاختبار عند قيم مختارة من القدرة، ولتحقيق ذلك تم بناء مصفوفة العلامات لجميع أفراد عينة الدراسة في حاسوب شخصي على برنامج (Excel) وحساب العلامة الخام لكل فرد، وبعدها تم نقل البيانات إلى

Jerash for Research and Studies Journal ???? ??? ?????? ?????????, Vol. 10 [2009], Iss. 1, Art. 4

البرامج الإحصائية لأجراء التحليلات اللازمة لتحقيق أهداف الدراسة، والإجابة عن أسئلتها.

النتائج

أولا:- نتائج استقصاء الافتراضات التي يتطلبها نموذج التقدير الجزئي.

١. افتراض أحادية البعد (Unidimensional)

تم تطبيق الرزمة (SPSS) في أجراء عملية التحليل العاملي (Factor Analysis) للبيانات باستخدام المكونات الرئيسية، ويبين جدول ١ قيم الجذر الكامن ونسبة التباين المفسر للعاملين الأول والثاني، وناتج قسمة نسبة التباين المفسر للعامل الأول على نسبة التباين للعامل الثاني.

جدول ١ قيم الجذر الكامن ونسبة التباين المفسر للعاملين الأول والثاني، وناتج قسمة نسبة التباين المفسر للعامل الأول على نسبة التباين للعامل الثاني.

العامل	الأول	الثاني	ناتج القسمة
الجذر الكامن	٩,١٣٨	١,٣٠٦	٦,٩٩٦
نسبة التباين المفسر	٤٥,٦٩٠%	٦,٥٣٠	

يلاحظ من جدول ١ أن العامل الأول قد فسر ما نسبته ٤٥,٦٩٠٪ من التباين الكلي، والعامل الثاني قد فسر ما نسبته ٦,٥٣٠٪ من التباين الكلي، أي أن مجموع نسبة ما يفسره العامل الأول والثاني معا ٥٢,٢٢٪ من التباين الكلي، وأن نسبة ما يفسره العامل الأول يساوي سبعة أضعاف ما يفسره العامل الثاني، وهذه النسبة العالية تشير إلى أن الاختبار أحادي البعد يقيس سمة واحدة.

٢. افتراض الاستقلال الموضعي (Local Independence)

يشير هاملتون وسوامناثان (Hambleton & Swaminathan) إلى أن افتراض استقلالية الموضع يكافئ افتراض أحادية البعد في الاختبار، وهذا يعني أنه إذا تحقق افتراض أحادية البعد في الاختبار فأن الاختبار يحقق افتراض استقلالية الموضع.

٣. يفترض نموذج التقدير الجزئي تساوي جميع أسئلة الاختبار في القدرة التمييزية بين مستويات القدرة المقيسة، وقد تم تقدير معاملات التمييز لأسئلة الاختبار من خلال حساب معاملات الارتباط بين أداء الأفراد على السؤال الواحد وأداء الأفراد على جميع أسئلة الاختبار، وتشير النتائج في جدول ٢ إلى معاملات الارتباط بين الأداء على السؤال الواحد، والأداء على جميع أسئلة الاختبار.

جدول ٢ متوسط علامة الفقرة "كمؤشر لصعوبة الفقرة" ومعاملات بيرسون بين الأداء على الفقرة، والأداء على جميع الفقرات"كمؤشر على تمييز الفقرة".

https://digitalcommons.aaru.edu.jo/jpu/vol10/iss1/4 8

رقم السؤال	الصعوبة	التمييز	رقم السؤال	الصعوبة	التمييز
١	٠,٣٦	٠,٦٣	١١	٠,٦٧	٠,٥١
٢	٠,٦٤	٠,٣٢	١٢	٠,٤٢	٠,٦٣
٣	٠,٧٣	٠,٦٠	١٣	٠,٤٠	٠,٧١
٤	٠,٥٤	٠,٦٨	١٤	٠,٥٤	٠,٧٢
٥	٠,٥٧	٠,٥٣	١٥	٠,٦٤	٠,٦٤
٦	٠,٥١	٠,٦٣	١٦	٠,٥٧	٠,٦٤
٧	٠,٣٧	٠,٦٣	١٧	٠,٥٢	٠,٥٢
٨	٠,٥٥	٠,٦٠	١٨	٠,٥٥	٠,٦٩
٩	٠,٥٦	٠,٧٥	١٩	٠,٥٥	٠,٧٩
١٠	٠,٦٩	٠,٥٩	٢٠	٠,٥٦	٠,٧٥

تشيرمؤشرات درجة تمييز الفقرة في جدول ٢ أنها ذات قيم عالية ومتقاربة إلى حد ما حيث بلغت أعلى قيمة لمعاملات الارتباط ٠,٧٩ وكانت للسؤال رقم ١٩، وبلغت أدنى قيمة لمعاملات الارتباط ٠,٣٢ وكانت للسؤال رقم ٢. وإذا ما استثنينا السؤال رقم ٢ من أسئلة الاختبار، فإن المدى الذي تقع فيه معاملات التمييز هو ٠,٢٨، وإذا استثنينا خمس أسئلة أخرى يقل المدى لمعاملات التمييز إلى ٠,١٣

ثانيا:- عرض النتائج المتعلقة بأسئلة الدراسة:

النتائج المتعلقة بالسؤال الأول الذي ينص على ما يلي:

ما دلالات المطابقة لبيانات الاستجابة لفقرات الاختبار المتعددة التدريج لنموذج التقدير الجزئي؟

يقوم البرنامج الإحصائي (BIGSTEPS) بحذف العلامات الكاملة والعلامات الصفرية تلقائيا من التحليل ولم يعمل البرنامج على حذف أي سؤال من أسئلة الاختبار حيث لم يكن هناك أي سؤال فشل في الإجابة عنه او اجابة اجابة صحيحة جميع أفراد العينة.

ويبين جدول ٣ قيمة متوسط إحصائي المطابقة الداخلية والخارجية لقيم صعوبة الأسئلة، والانحراف المعياري لتقديرها.

جدول ٣ قيم الوسط الحسابي، والانحراف المعياري لإحصائي المطابقة الداخلية والخارجية لقيم صعوبة الأسئلة.

	INFIT(إحصائي المطابقة الداخلي)	OUTFIT(إحصائي المطابقة الخارجي)
الوسط الحسابي	١,٠٠	١,٠٢
الانحراف المعياري	٠,١٩	٠,٢٧

تبين نتائج جدول ٣ أن قيمة متوسط إحصائي المطابقة الداخلية لقيم صعوبة الأسئلة قد بلغ ١,٠٠

لوجيت وهي مطابقة للقيمة المتوقعة من النموذج، بانحراف معياري قدرة ٠,١٩ لوجيت، وهي قيمة قريبة من القيمة المتوقعة من النموذج، حيث أن توقع قيم هذا الإحصائي لقيمة متوسط إحصائي المطابقة، والانحراف المعياري له (١، ٠) على التوالي، وقيمة متوسط المطابقة الخارجية لدرجة صعوبة الأسئلة في الاختبار بلغت ١,٠٢ لوجيت، بانحراف معياري مقداره ٠,٢٧ لوجيت، وهي قيمة قريبة من القيمة المتوقعة من النموذج، مما يشير إلى اتساق كبير بين الاستجابات الملاحظة للأفراد على أسئلة الاختبار، والاستجابة المتوقعة من النموذج المستخدم، أي أن قيمة توقع النموذج لاحتمالية إجابة الأسئلة الصعبة بشكل صحيح من قبل الأفراد ذوي القدرة المحدودة تقترب كثيرا من القيم المشاهدة لاحتمالية إجابة الأسئلة الصعبة بشكل صحيح المحسوبة من عدد الأفراد الذين لهم القدرة نفسها.

النتائج المتعلقة بالسؤال الثاني الذي ينص على ما يلي:-

ما تقدير قدرات الأفراد، وصعوبة الأسئلة، وصعوبة كل خطوة من خطوات استجابة السؤال الواحد باستخدام نموذج التقدير الجزئي؟

أدخلت العلامات الخام إلى البرنامج الإحصائي (BIGSTEPES) بعد حذف استجابات الأفراد التي لم تطابق النموذج وأجري التحليل الإحصائي مرة ثانية، ويبين جدول ٤ الوسط الحسابي والانحراف المعياري للعلامات الخام التي حصل عليها أفراد العينة، والوسط الحسابي لتقديرات قدرات الأفراد.

جدول ٤ الوسط الحسابي والانحراف المعياري للعلامات الخام التي حصل عليها أفراد العينة، والوسط الحسابي لتقديرات قدرات الأفراد

	INFIT(إحصائي المطابقة الداخلي)	OUTFIT(إحصائي المطابقة الخارجي)
الوسط الحسابي	١,٠٠	١,٠٢
الانحراف المعياري	٠,١٩	٠,٢٧

تشير النتائج في جدول ٤ إلى أن متوسط توزيع تقديرات قدرات الأفراد ٠,١٢ لوجيت، وهي قريبة من الوسط الحسابي للمقياس اللوغارتيمي كما يتوقعها البرنامج الإحصائي وهو صفر لوجيت، وأن الانحراف المعياري لتوزيع التقدير يساوي ٠,٧٥ لوجيت مما يشير إلى أن مستوى الاختبار بشكل عام في متناول قدرات أفراد العينة. ويبين جدول ٥ أرقام الأسئلة، وتقدير درجة صعوبة كل خطوة من خطوات الاستجابة في الاختبار.

رقم السؤال	تقدير درجة صعوبة خطوات الاستجابة δ_i					
	δ١	δ٢	δ٣	δ٤	δ٥	δ٦
١	٠,١٥	٠,٤٠-	٠,٢١	١,٢٧	٠,٨٤	٢,٠٠
٢	٠,٥٨	١,٨٣-	٠,٠٢-			
٣	٠,٦١-	٠,٩٠-	٠,٢٠-	٠,١٤	٠,٢١-	
٤	١,٦٦	٠,٠٥-	٠,٢٩	١,٧٣-		
٥	٠,٠٠	٠,١٤-	٠,٧-			
٦	٠,٥٤	٠,٢٠-	٠,٠٤			
٧	١,٠١	٠,٣٧	٠,٢٠-	١,٠٨	٠,١٦	
٨	٠,٥٤-	١,٣١-	٠,٠٥-	١,٠٥	٠,٥٨	٠,٤٨-
٩	٠,٠٥	٠,٠٢-	٠,٢٨-	٠,٧٢	٠,٦٨-	
١٠	٠,٠٤-	١,٣١-	٠,٠٣-			
١١	٠,٤٩-	٠,٤٢-	٠,٢٠	٠,٩٦-		
١٢	٠,٥٩	٠,٣١-	١,٠٩	٠,٢١-	٠,١٣-	
١٣	٠,١١	٠,٠٢-	٠,١٨	٠,٩٣	١,٤١	
١٤	٠,٦٩	٠,١٨-	٠,٠٢-	٠,٣٢-		
١٥	٠,٧٢-	٠,٤٠-	٠,٠٤			
١٦	٠,٣٩	٠,٠١	٠,٢٧	٠,٠٢-	٠,٧٩-	
١٧	٠,٠٧	٠,٣٦-	٠,٥٠	٠,٢٣-	١,١٠	٠,٥٤-
١٨	٠,٠٣	٠,١٥-	٠,٠٠	٠,٦٠-		
١٩	٠,٣٥	٠,٠٧	٠,٠٤	٠,٥٦	٠,١١	١,١١-
٢٠	٠,٣١	٠,٣٥-	٠,٠٩	٠,١٩-	٠,٢٠	

جدول ٥ أرقام الأسئلة، وتقدير درجة صعوبة كل خطوة من خطوات الاستجابة (δi) في الاختبار.
ملاحظة : الجدول صورة

تشير نتائج جدول ٥ الى أن أصعب خطوة من خطوات استجابة أسئلة الاختبار كانت الخطوة السادسة من السؤال رقم ١ حيث بلغت ٢,٠٠ لوجيت، وأن أسهل خطوة من خطوات استجابة أسئلة الاختبار كانت الخطوة الثانية من السؤال رقم ٢ حيث بلغت -١,٨٣ لوجيت، أي أن مدى الصعوبة في خطوات استجابة أسئلة الاختبار التي يمكن أن نقيس بها قدرات الأفراد بلغ ٣,٨٣ لوجيت وهذا المدى من الصعوبة في خطوات استجابة أسئلة الاختبار يمكننا من استخدام أسئلة الاختبار في تقدير قدرات مدى مقبول من القدرة، حيث تتوافر أسئلة متباينة في صعوبتها يمكن أن تقدم إلى الأفراد كل حسب قدراته.

ولقد تم الحصول من التحليل الإحصائي باستخدام نموذج التقدير الجزئي على معاملات الثبات المتعلقة بالأفراد (Person - Reliability) ومعاملات الثبات المتعلقة بدرجة صعوبة أسئلة الاختبار (Test - Reliability) حيث يشير مفهوم الثبات إلى مدى الدقة في تحديد مواقع كل من قدرة الأفراد، ودرجة صعوبة الأسئلة على متصل السمة، ويتم تحديد مدى الدقة بما يسمى مؤشر الفصل بين الأفراد (Separation Index - Person Gp) ومؤشر الفصل بين الأسئلة (Separation In-dex - Item Gi) والذي يجب أن تكون قيمته أكبر من ٢ فإذا لم تزد قيمته عن ذلك يصعب عندئذ تقدير قدرة الأفراد بوساطة هذه الأسئلة أو تقدير درجة صعوبة الأسئلة من خلال عينة الأفراد (Wright & Linacre).

وقد وجد أن قيمة Gp للأفراد بلغت ٣,٢٥ وهي قيمة تزيد عن ٢، لذا فإن عينة الأفراد كافية لتقدير درجة صعوبة أسئلة الاختبار، وقيمة Gi للفقرات بلغت ٩,٢٥ وهذه القيمة كبيرة تؤكد قدرة وكفاية أسئلة الاختبار في تقدير قدرة الأفراد. وقد وجد أن قيمة معامل الثبات المتعلق بالأفراد باستخدام العلاقة (Wright & Linacre) ($R_p = \frac{G_p^2}{1+G_p^2}$) قد بلغ ٠,٩١ وهذه القيمة تشير إلى كفاية عينة الأفراد في الفصل بين الأسئلة وتعريف متصل السمة الذي تقيسه هذه الأسئلة. وبالتالي في إمكانية تقدير درجة صعوبة الأسئلة من خلال عينة الأفراد. وبلغت قيمة معامل الثبات المتعلق بأسئلة الاختبار ٠,٩٩ وهذه القيمة تشير إلى كفاية عينة الأسئلة في الفصل بين قدرات الأفراد، أي في التمييز بين قدرات الأفراد على متصل السمة.

كما تم استخدام المعادلة ($H_i = (4G_i + 1)/3$) في حساب عدد الطبقات المتمايزة للأفراد وكانت تقريبا تساوي خمس طبقات، أي أن أسئلة الاختبار تميز بين خمس طبقات من قدرات الأفراد مراكزها تبتعد عن بعضها بعضا بمقدار ثلاث من وحدات الأخطاء المعيارية في التقدير، في حين كانت الطبقات المتمايزة لدرجة صعوبة الأسئلة تساوي تقريبا ١٣ طبقة، أي أن عينة الأفراد تميز بين ١٣ طبقة من درجة صعوبة الأسئلة مراكزها تبتعد عن بعضها بعضا بمقدار ثلاث من وحدات الأخطاء المعيارية في التقدير.

النتائج المتعلقة بالسؤال الثالث الذي ينص على ما يلي:-

ما قيم الأخطاء المعيارية في تقدير قدرات الأفراد؟

يبين جدول ٦ متوسط الأخطاء المعيارية والانحراف المعياري لتقدير قدرات الأفراد.

جدول ٦ متوسط الأخطاء المعيارية والانحراف المعياري لتقدير قدرات الأفراد ككل.

	الأخطاء المعيارية في تقدير قدرات الأفراد
الوسط الحسابي	٠,٢
الانحراف المعياري	٠,٠٩

يشير جدول ٦ إلى أن الوسط الحسابي للأخطاء المعيارية في تقدير قدرات الأفراد قد بلغ ٠,٢ لوجيت بانحراف معياري بلغ ٠,٠٩ لوجيت وهي قيمة متدنية الأمر الذي يشير إلى دقة تقدير قدرات الأفراد، حيث أن قيمة الخطأ المعياري في تقدير قدرات الأفراد تقل كلما اقتربت قيمة قدرة الفرد من الصفر. أي أن قيمة الخطأ المعياري في التقدير تقل كلما اقتربت قدرة الأفراد من درجة صعوبة الأسئلة.

٤- النتائج المتعلقة بالسؤال الرابع الذي ينص على مايلي:

ما قيم الأخطاء المعيارية في تقدير درجة صعوبة الأسئلة، وصعوبة كل خطوة من خطوات استجابة السؤال الواحد؟

يبين جدول ٧ رقم السؤال، والخطأ المعياري في تقدير كل خطوة من خطوات الاستجابة (Si) في الاختبار.

رقم السؤال	الخطأ المعياري في التقدير Si					
	S١	S٢	S٣	S٤	S٥	S٦
١	٠,١١	٠,١٠	٠,١٠	٠,١٢	٠,١٤	٠,٢٤
٢	٠,١٢	٠,١١	٠,٠٩			
٣	٠,١٨	٠,١٤	٠,١٢	٠,١٠	٠,١٠	
٤	٠,١١	٠,١١	٠,١١	٠,١٠		
٥	٠,١١	٠,١٠	٠,٠٩			
٦	٠,١٠	٠,١٠	٠,١٠			
٧	٠,١٠	٠,١٠	٠,١١	٠,١٢	٠,١٣	
٨	٠,١٧	٠,١٣	٠,١٠	٠,١٠	٠,١١	٠,١١
٩	٠,١٢	٠,١١	٠,١١	٠,١٠	٠,١٠	
١٠	٠,١١	٠,١١	٠,٠٩			
١١	٠,١٤	٠,١١	٠,١٠	٠,١٠		
١٢	٠,١٠	٠,١٠	٠,١١	٠,١١	٠,١٢	
١٣	٠,١١	٠,١٠	٠,١٠	٠,١١	٠,١٧	
١٤	٠,١١	٠,١١	٠,١٠	٠,١٠		
١٥	٠,١٣	٠,١٠	٠,٠٩			
١٦	٠,١٢	٠,١١	٠,١١	٠,١١	٠,١٠	
١٧	٠,١٢	٠,١١	٠,١١	٠,١١	٠,١١	٠,١١
١٨	٠,١١	٠,١١	٠,١٠	٠,١٠		
١٩	٠,١٢	٠,١٢	٠,١٢	٠,١٢	٠,١١	٠,١١
٢٠	٠,١٢	٠,١١	٠,١١	٠,١٠	٠,١٠	

Jerash for Research and Studies Journal ???? ??? ?????? ?????????, Vol. 10 [2009], Iss. 1, Art. 4

جدول ٧ رقم السؤال، والخطأ المعياري لتقدير كل خطوة من خطوات الاستجابة (Si) في الاختبار.

تبين نتائج جدول ٧ أن قيم الأخطاء المعيارية في تقدير درجة صعوبة خطوات استجابة أسئلة الاختبار متدنية، مما يشير إلى دقة التقدير في صعوبة الأسئلة من خلال عينة المفحوصين.

٥- النتائج المتعلقة بالسؤال الخامس الذي ينص على ما يلي:-

ما دلالات ثبات (Invariance) معالم الأسئلة (صعوبة الأسئلة) في عينتين مختلفتين من الأفراد؟

ولفحص دلالات ثبات صعوبة الأسئلة في عينتين مختلفتين من الأفراد، تم استخراج تقديرات صعوبة الأسئلة والخطأ المعياري في التقدير للاختبار المكون من ١٣ فقرة ذات الأرقام ١، ٤، ٦، ٧، ٨، ٩، ١٢، ١٣، ١٤، ١٦، ١٧، ١٨، ١٩، وذلك بمعايرتها من خلال مجموعة القدرات الكلية ومجموعة القدرات العشوائية، ومجموعة القدرات العليا ومجموعة القدرات المتدنية باستخدام نموذج التقدير الجزئي، وقد تم حذف السؤال رقم ٤ والسؤال رقم ١٨ بسبب عدم مطابقتهما لنموذج التقدير الجزئي عند معايرتهما من خلال مجموعة القدرات المتدنية، ويبين جدول ٨ التالي درجة الفرق المعاير في تقدير قيم الصعوبة في مختلف الأسئلة باستعمال مجموعات مختلفة من الأفراد، واعتمادا على نموذج التقدير الجزئي.

جدول ٨ درجة الفرق المعاير في تقدير فيم الصعوبة على مختلف الأسئلة باستعمال مجموعات مختلفة من الأفراد، واعتمادا على نموذج التقدير الجزئي.

رقم السؤال	المجموعة الكلية والمجموعة العشوائية	المجموعة الكلية والمجموعة العليا	المجموعة الكلية والمجموعة المتدنية	المجموعة العليا والمجموعة المتدنية
١	٠,٦١-	٠,٠٩-	٠,٤٠	٢-
٦	٠,١٥	٠,٨٩	٠,٣٢-	٠,٩٦
٧	٠,٣٣-	٠,٣٨-	٢,٥٦-	٤,٥
٨	٠,١٨-	٢,٦	١,٥٨	٠,١٨
٩	٠,٥	٢,٨	١,٠٦	٠,٨٨
١٢	٣,٥	٠,١٢	١,١٨-	١,٣
١٣	٢,١-	٠,٤٦	١,٧٣	١,٥-
١٤	١,٤	٠,٧٩	٠,٧٣-	١,٢٢
١٦	١,٤	١,٠٨-	٣,٥-	٢,٥٣
١٧	١,٥	١,٣٣	٣,٨	٢,٨٥-
١٩	٣	٣,٩	٠,٥٣-	٢,٦٩

ولفحص الدلالة الإحصائية للفرق في الأوساط الحسابية المعايرة في تقدير فيم الصعوبة على مختلف الأسئلة باستعمال مجموعات مختلفة من الأفراد، واعتمادا على نموذج التقدير الجزئي. فقد

https://digitalcommons.aaru.edu.jo/jpu/vol10/iss1/4 14

تم استخدام تحليل التباين الأحادي (ANOVA)، في حساب قيمة F عند مستوى من الدلالة _ $0.01\geq\alpha$، والجدول ٩ يبين مصدر التباين بين المجموعات، وداخل المجموعات، ومجموع المربعات، ووسط المربعات وقيمة F عند مستوى من الدلالة $0.01\geq\alpha$.

جدول ٩ مصدر التباين بين المجموعات، وداخل المجموعات، ومجموع المربعات، ووسط المربعات، وقيمة F عند مستوى من الدلالة $0.01\geq\alpha$

مصدر التباين	مجموع المربعات	درجات الحرية	وسط المجموعات	F	Sig.
بين المجموعات	6.588	3	2.196	.630	.600
داخل المجموعات	139.513	40	3.488		
المجموع الكلي	146.101	43			

يتضح من جدول ٩ أن فروق الأوساط المعايرة في تقديرات معالم الصعوبة غير دالة إحصائيا في جميع أزواج المقارنات الأربع، وذلك لأن قيمة مستوى الدلالة المحسوب أكبر من ٠,٠١

٦- النتائج المتعلقة بالسؤال السادس الذي ينص على ما يلي:-

ما دلالات ثبات (Invariance) تقدير القدرة لمجموعة واحدة من الأفراد في عينتين مختلفتين من الأسئلة؟

ولفحص مدى الثبات لتقدير قدرات الأفراد من خلال عينتين مختلفتين من الأسئلة تم استخدام الأسئلة الصعبة ذات الأرقام ١، ٦، ٧، ٨، ٩، ١٢، ١٣، ١٤، ١٦، ١٧، ١٩ لتقدير أربع تقديرات لقدرات الأفراد في المجموعة الكلية والمجموعة العشوائية والمجموعة العليا والمجموعة المتدنية. كما تم استخدام الأسئلة السهلة ذات الأرقام ٢، ٣، ٥، ٨، ٩، ١٠، ١١، ١٥، ١٦، ١٧، ١٩، ٢٠ لتقدير أربعة تقديرات لقدرات نفس الأفراد في المجموعة الكلية والمجموعة العشوائية والمجموعة العليا والمجموعة المتدنية باستخدام البرمجية (BIGSTEPS). وجدول ١٠ يبين الفروق المعيارية بين تقديرات قدرات نفس عينة الأفراد في المجموعات الأربع من خلال الاختبار الصعب ثم الاختبار السهل.

جدول ١٠ الفروق المعيارية بين تقديرات قدرات نفس عينة الأفراد في المجموعات الأربع من خلال الاختبار الصعب ثم الاختبار السهل.

المجموعة الكلية	المجموعة العشوائية	المجموعة العليا	المجموعة المتدنية
٠,٦٧	١,٣٦	٠,٧٩	١,٠١
-٠٧٤	١,٦٢	١,٠٣	٠,٧٧
١,٣٣	١,٤٠	١,٠٠	٠,٤٨
١,٣٦	١,٣٦	١,١٢	٠,٧٦
٠,٧٨	١,٥٤	١,١٦	٠,٦٤
٠,٣٣	١,٥١	١,٠٧	٠,٥١
٠,٢٧	١,٢١	٠,٩٠	٠,٦٧
٠,٣٣	١,٣٢	٠,٩٢	٠,٧٦
٠,٥٧	١,١١	٠,٩٢	٠,٦٦
٠,٤٩	١,٣٥	٠,٩٢	٠,٥٣
٠,٤١	١,٤١	١,٠٠	٠,٨٦
٠,٤٦	١,٢٧	١,٠٣	٠,٦٥
٠,٣٨	١,٣٢	١,٠٩	٠,٥٦
٠,٢٣	١,٣٦	١,٠٥	٠,٦٢
٠,٠٠٢	١,٣٩	١,٠٩	٠,٦٢
٠,٠١	١,٣٢	١,٠٠	٠,٥٣
-٠,٠٥	١,٨١	١,١٢	٠,٤٢
٠,٠٤	١,٧٩	٠,٩٧	٠,٤٦
-٠,٥١	١,٢٤	١,٠٠	٠,٤٣
-٠,٠٣	١,٥٥	١,٠٦	٠,٤٤
-٠,٠٤	١,٤٧	٠,٩٧	٠,٥٦
-٠,١٦	١,٥١	٠,٩١	٠,٣٨
-٠,٢٣	٠,١٨	٣,٤٨	٠,٢٥

يبين جدول ١٠ أن أكبر فرق معاير كان في تقدير قدرات أفراد المجموعة العليا المحسوبة من خلال الاختبارين الصعب والسهل حيث بلغ ٣,٤٨ لوجيت، وقد يعود السبب في ذلك إلى أن أسئلة الاختبار السهل دون مستوى قدرات أفراد هذه المجموعة مما ترتب عليه إعطاء تقديرات عالية لقدرة أفراد هذه المجموعة المحسوبة من خلال الاختبار السهل.

ولفحص الدلالة الإحصائية للفرق في الأوساط الحسابية المعايرة لمعالم قدرة الأفراد في كل مجموعة محسوبة من خلال الاختبار الصعب والاختبار السهل، فقد تم استخدام تحليل التباين الأحادي (ANOVA)، في حساب قيمة F عند مستوى من الدلالة $\alpha \geq 0.01$، والجدول ١١ يبين مصدر التباين بين المجموعات، وداخل المجموعات، ومجموع المربعات، ووسط المربعات وقيمة F عند مستوى من الدلالة $\alpha \geq 0.01$

جدول ١١ مصدر التباين بين المجموعات، وداخل المجموعات، ومجموع المربعات، ووسط المربعات وقيمة F عند مستوى من الدلالة $\alpha \geq 0.01$

مصدر التباين	مجموع المربعات	درجات الحرية	وسط المربعات	F	Sig.
بين المجموعات	17.318	3	5.773	35.674	.000
داخل المجموعات	14.240	88	.162		
المجموع الكلي	31.557	91			

يتضح من جدول ١١ أن فروق الأوساط المعايرة في تقديرات قدرات الأفراد في جميع أزواج المقارنات الأربع دالة إحصائيا في جميع أزواج المقارنات الأربع، وذلك لأن قيمة مستوى الدلالة المحسوب أقل من ٠,٠١، أي أنه يوجد فرق لتقدير قدرات الأفراد من خلال عينتين مختلفتين من الأسئلة الصعبة والسهلة.

٧- النتائج المتعلقة بالسؤال السابع الذي ينص على ما يلي:

ما الكفاية النسبية لاختبارات فرعية مأخوذة من بنك الأسئلة عند قيم مختارة للقدرة (β)؟

ولتقصي الكفاية النسبية للاختبار الصعب بالنسبة للاختبار الكلي محسوبة من قسمة قيمة دالة المعلومات للاختبار الصعب على قيمة دالة المعلومات للاختبار الكلي و لمستويات مختلفة من القدرة. وقد تم حساب الكفاية النسبية من العلاقة :

$$R.E = \left(\frac{S_h}{Sa} \right)^2$$

حيث : R E هي الكفاية النسبية للاختبار الصعب بالنسبة للاختبار الكلي.

Sh هو الخطأ المعياري لتقدير قدرة الأفراد باستخدام أسئلة الاختبار الصعب.

Sa هو الخطأ المعياري لتقدير قدرة الأفراد باستخدام أسئلة الاختبار الكلي، ويبين جدول ١٢ الأخطاء المعيارية و الكفاية النسبية للاختبار الصعب بالنسبة للاختبار الكلي في تقدير القدرة عند مستويات مختلفة من القدرة (β).

جدول ١٢ الأخطاء المعيارية والكفاية النسبية لتقدير القدرة من خلال الاختبار الصعب مقارنة بالاختبار الكلي عند مستويات مختلفة من القدرة.

Jerash for Research and Studies Journal ???? ??? ?????? ?????????, Vol. 10 [2009], Iss. 1, Art. 4

مستوى القدرة	sh الخطأ المعياري للاختبار الكلي	Sl الخطأ المعياري للاختبار السهل	Sa الخطأ المعياري للاختبار الصعب	الكفاية النسبية للاختيار السهل	الكفاية النسبية للاختبار الصعب
١,٣٥	٠,٣١	٠,٦٣	٠,٣٤	٠,٢٤	٠,٨٣
٠,١١	٠,٢٨	٠,٤٢	٠,٣٤	٠,٣٨	٠,٥٨
٠,٩٨	٠,٢٤	٠,٦٣	٠,٢٣	٠,١٥	١,٠٩
٠,٩٣	٠,٢٣	٠,٢٤	٠,٢٤	٠,٩٢	٠,٩٢
٠,٨٨	٠,٢٢	١,٣٨	٠,٢٣	٠,٠٣	٠,٩١
٠,٨٤	٠,٢١	٠,٢٢	٠,٢	٠,٩١	١,١
٠,٧٥	٠,٢	٠,٢١	٠,٢	٠,٩١	١,٠٠
٠,٦٨	٠,١٩	٠,٤٢	٠,٢٨	٠,٢	٠,٤٦
٠,٦١	٠,١٨	٠,٢٣	٠,٢٣	٠,٦١	٠,٦١
٠,٦٤	٠,١٨	٠,٣٧	٠,١٩	٠,٢٤	٠,٩
٠,٥٨	٠,١٨	٠,٢٧	٠,٣٢	٠,٤٤	٠,٣٢
٠,٥٢	٠,١٧	٠,٣٧	٠,٢	٠,٢١	٠,٧٢
٠,٤٣	٠,١٦	٠,٢٣	٠,٢١	٠,٤٨	٠,٥٨
٠,٣٨	٠,١٦	٠,٢٤	٠,٣٧	٠,٤٤	٠,١٩
٠,١	٠,١٥	٠,٢١	٠,٢٧	٠,٥١	٠,٣١
-٠,٠٥	٠,١٥	٠,٢٣	٠,٣٣	٠,٤٣	٠,٢١
-٠,١٩	٠,١٥	٠,٤٢	٠,٤٩	٠,١١	٠,٠٩
-٠,٥٢	٠,١٨	٠,٢٧	٠,٩٢	٠,٤٤	٠,٠٤
-٠,٧	٠,٢	٠,٣١	٠,٩٤	٠,٤٢	٠,٠٥
-١,٤١	٠,٢٧	٠,٢١	٠,٣	١,٦٥	٠,٨١
الوسط الحسابي	٠,٢	٠,٣٨	٠,٣٤	٠,٤٩	٠,٥٩

يبين جدول ١٢ أن متوسط قيم الكفاية النسبية للاختبار الصعب بالنسبة للاختبار الكلي يساوي ٠,٥٩ تقريبا، أي أن قيمة دالة المعلومات للاختبار الصعب تساوي ٠,٥٩ قيمة دالة المعلومات للاختبار الكلي وأن الكفاية النسبية للاختبار الصعب بالنسبة للاختبار الكلي عند المستويات العالية من القدرة تقترب من الواحد، أي أن دالة المعلومات التي يعطيها الاختبار الصعب تساوي تقريبا دالة المعلومات للاختبار الكلي وذلك للأفراد ذوي القدرات العالية. أما بالنسبة للأفراد من ذوي القدرة المتوسطة فإن دالة المعلومات تقل تدريجيا مع نقصان القدرة، أي أن دالة المعلومات التي يعطيها الاختبار الكلي أفضل من دالة المعلومات التي يقدمها الاختبار الصعب وذلك عند مستويات متوسطة من القدرة. أما بالنسبة للأفراد من ذوي القدرة المتدنية فإن الكفاية النسبية للاختبار الصعب تقل وتعطي قيمة أصغر من الواحد، أي أن دالة المعلومات التي يقدمها الاختبار الكلي أفضل من دالة المعلومات التي يقدمها الاختبار الصعب وذلك عند مستوى من القدرة متدنية. كما تشير نتائج الجدول السابق أن متوسط قيم الكفاية النسبية للاختبار السهل تساوي ٠,٤٩ قيمة دالة المعلومات للاختبار الكلي، وأن متوسط قيم

https://digitalcommons.aaru.edu.jo/jpu/vol10/iss1/4 18

الكفاية النسبية للاختبار السهل أقل من متوسط قيم الكفاية النسبية للاختبار الصعب أي أن دالة المعلومات التي يقدمها الاختبار الصعب أفضل من دالة المعلومات التي يقدمها الاختبار السهل بشكل عام أما بالنسبة للأفراد من ذوي القدرتين العالية والمتوسطة فإن الاختبار السهل يقدم دالة معلومات قليلة مقارنة مع دالة المعلومات للاختبار الكلي ودالة المعلومات للاختبار الصعب. وعند تفحص دالة المعلومات التي يقدمها الاختبار السهل عند المستويات المتدنية من القدرة نجد أن الكفاية النسبية للاختبار السهل تزداد بشكل ملحوظ.

مناقشة النتائج والتوصيات

عندما تتساوى قدرة الفرد مع درجة صعوبة السؤال فإن احتمال الإجابة عن السؤال بشكل صحيح من قبل هذا الفرد تساوي ٠,٥ حسب نموذج راش، وكلما زادت قدرة الفرد عن درجة صعوبة السؤال فإن احتمالية إجابة السؤال تزداد، أي يصبح السؤال عندها بالنسبة للطالب سهلا، وكلما قلت قدرة الفرد عن درجة صعوبة السؤال فإن احتمالية إجابة السؤال تقل، أي يصبح السؤال صعبا. وتشير النتائج إلى أن متوسط الخطأ المعياري في تقدير قدرة الأفراد قيمة متدنية الأمر الذي يشير إلى دقة تحديد مواقع الأفراد على متصل السمة، أي دقة تحديد مواقع الأفراد في معلم القدرة. وذلك بافتراض أن قدرة الفرد الحقيقية (β) ثابتة لسمة معينة. وعندما يتم تقدير هذه القدرة من خلال اختبار معين ينتج لدينا (β^{δ}). حيث (β^{δ}) هي تقدير ل (β) ويمكن القول أن:

$$\beta^{\delta} = \beta \pm K$$

حيث K عدد يتغير من فرد الى آخر ومن مستوى قدرة الى مستوى قدرة أخرى. ويلاحظ من هذه المعادلة أنه كلما صغرت قيمة العدد الثابت K قربت عملية القياس من تقدير القدرة الحقيقية للأفراد. وإذا اعتبرنا أن :

$$K = SM$$

حيث : M عدد ثابت

S الخطأ المعياري في التقدير.

ينتج أن :

$$\beta^{\delta} = \beta \pm SM$$

أي أن قدرات الأفراد ثابتة ولكن تقديرها يكون أكثر دقة كلما قل الخطأ المعياري في قياسها (أي قيمة S).

وبتفحص الأخطاء المعيارية في تقدير قدرات جميع أفراد العينة وجد أن مدى الأخطاء المعيارية يتراوح بين ٠,١٦ و ٠,٦٨ بوسيط مقداره ٠,٤٢ لوجيت في قياس مدى من القدرة يتراوح بين ٠,٣٣ و ٢,٥٥ بوسيط مقداره ١,٤٤ لوجيت، وكان الخطأ المعياري في قياس مدى القدرة الذي تراوح بين –

Jerash for Research and Studies Journal ???? ??? ?????? ?????????, Vol. 10 [2009], Iss. 1, Art. 4

٠,١٩ لوجيت و ٠,٣١ لوجيت بوسيط مقداره ٠,٠٦ لوجيت ثابتا ومساويا للمقدار ٠,١٥ لوجيت، وكان مدى الأخطاء المعيارية يتراوح بين ٠,١٦ و ١,٣٨ بوسيط مقداره ٠,٧٧ في قياس مدى من القدرة يتراوح بين -٠,٢١ لوجيت إلى -٣,٦٧ لوجيت بوسيط مقداره -١,٩٤ لوجيت.

تشير مثل هذه النتائج إلى أن أسئلة الاختبار تقيس قدرات الأفراد بدقة أكبر كلما اقتربت قيم القدرة من صفر تدريج المقياس المستخدم في قياس القدرة، وهذا يعني أنه كلما اقتربت قيمة معلم صعوبة أسئلة الاختبار من قيمة معلم قدرة الأفراد تزداد الدقة في القياس، أي أن اختيار أسئلة ذات صعوبة مطابقة لقدرات الأفراد تعطي تقديرات أكثر دقة في القياس. أما قيمة متوسط درجة صعوبة الأسئلة تشير إلى أن الاختبار يتمتع بدرجة متوسطة من الصعوبة بالنسبة إلى قدرات الأفراد وأن الخطأ المعياري في تقديرها متدن جدا، ويشير هذا الأمر إلى مدى استقرار الصعوبة النسبية للفقرات عبر مستويات القدرة المختلفة، أي أن ثبات تدرج صعوبة الأسئلة عبر مستويات القدرة عال جدا.

تشير نتائج تحليل الصعوبة للأسئلة إلى أن درجة صعوبة الأسئلة تزداد بزيادة عدد خطوات الاستجابة لها، حيث بلغ متوسط صعوبة الأسئلة التي عدد خطوات الاستجابة فيها ست خطوات ٠,١٦ لوجيت (٠,٧٦، -٠,١٣، ٠,٠٩، ٠,٠٠٣) والتي عدد الخطوات فيها خمس خطوات ٠,١٥ لوجيت (-٠,٣٦، ٠,٤٩، -٠,٠٤، ٠,٢١، ٠,٥٢، -٠,٢٠) والتي عدد خطوات الاستجابة فيها أربع خطوات -٠,١٣ لوجيت (٠,٠٤، -٠,٤٢، ٠,٠٤، -٠,١٨) أما الأسئلة التي عدد خطوات استجابتها ثلاث خطوات فقط فقد بلغ صعوبة خطواتها -٠,٢٨ لوجيت (-٠,٤٢، -٠,٢٨، ٠,١، -٠,٤٦، -٠,٣٦) أي أن الأسئلة ذات عدد الخطوات الأكثر عددا تكون أصعب من الأسئلة ذات الخطوات الأقل عددا. وتشير النتائج إلى أن الخطأ المعياري لتقدير درجة صعوبة الأسئلة تأخذ قيما متقاربة بدرجة كبيرة في جميع الأسئلة وقد بلغ ٠,٠٣ في معظم الأسئلة وهي قيمة متدنية تشير إلى الدقة المحققة في تقدير الصعوبة لهذا النوع من الأسئلة (المتعددة الخطوات).

وكانت فروق الأوساط المعايرة في تقديرات معالم الصعوبة غير دالة إحصائيا في جميع أزواج المقارنات الأربع، وذلك لأن قيمة مستوى الدلالة المحسوب أكبر من ٠,٠١ أي لا يوجد فرق في الاوساط المعايرة في تقدير معالم الصعوبة من خلال المجموعات الاربع أي أن هناك ثباتا (Invariance) في معالم الأسئلة (صعوبة الأسئلة) في عينتين مختلفتين من الأفراد، بينما كان فروق الوسط المعاير في تقديرات معالم قدرة الأفراد في عينتين من الأسئلة دالا احصائيا عند مستوى $\alpha \geq 0.01$، ويمكن تفسير ذلك على أساس أنه عندما لايكون هناك تطابق بين مدى القدرة للأفراد ومدى الصعوبة للفقرات فمن المتوقع أن تظهر فروق ذات دلالة بين متوسطات تقديرات الصعوبة أو بين متوسطات تقديرات القدرة والعكس صحيح.

وتشير نتائج الدراسة إلى إمكانية اختيار مجموعة جزئية من الفقرات ذات مواصفات محددة من بنك الأسئلة وذلك حسب القدرة المقيسة بهذه الفقرات، قد يكون لها فاعلية عالية في تصنيف الأفراد حسب قدراتهم في السمة المقيسة. وفي نهاية هذا الدراسة، واعتمادا على الهدف من هذه الدراسة وهو فاعلية نموذج التقدير الجزئي في معايرة فقرات متعددة خطوات الاستجابة في مادة الكيمياء

https://digitalcommons.aaru.edu.jo/jpu/vol10/iss1/4 20

للصف الثاني الثانوي العلمي. وبناء على النتائج التي تحققت والمشكلات التي وجدت أثناء تطبيق البرمجية (Bigsteps). فإنه من الملائم إعطاء التوصيات التالية :

جدير بالذكر أن الدراسة الحالية لم تتحقق من ثبات معالم الفقرات وثبات معالم الأفراد عندما يكون هناك تطابق بين مدى تقديرات قدرات الأفراد ومدى تقديرات صعوبة الأسئلة، ولعله من المفيد أن يتناول باحثون آخرون مثل هذه الحالة.

كذلك لم يتيسر في هذه الدراسة التحقق من فاعلية نموذج التقدير الجزئي في معايرة فقرات بنك الأسئلة على مواد العلوم الأخرى مثل الفيزياء والأحياء، ولعله من المفيد أن يوصى باحثون آخرون بدراسة هذه الموضوعات.

ولم تتناول الدراسة الحالية فاعلية نموذج التقدير الجزئي في معايرة الأسئلة التي تكون الاستجابة لها بخطوات متدرجة من حيث الصعوبة من الأصعب إلى الأسهل أو من الأسهل إلى الأصعب، ويوصي الباحث بدراسة فاعلية نموذج التقدير الجزئي في مثل هذه الحالات.

المراجع العربية

- علام، صلاح الدين محمود، (١٩٨٦)، تطورات معاصرة في القياس النفسي، كلية الآداب، جامعة الكويت، الكويت، إدارة التأليف والترجمة والنشر.
- علام، صلاح الدين محمود، (٢٠٠٢)، القياس والتقويم التربوي والنفسي، أساسياته وتطبيقاته وتوجيهاته المعاصرة، القاهرة: دار الفكر العربي.

المراجع الأجنبية (Reference)

- Crocker. L. and Algina. J. (1986). Introduction to Classical and Modern Test Theory, University of Florida, Published Simultaneously in Canada.
- Hambleton. R. K, and Swaminathan. H. (1985). Item Response Theory Principles and Applications Boston, Kluwer. Nijhoff Publishing.
- Hambleton. R. K, and Swaminathan. H, and Rogers. H. J, (1991). Fundamentals of Item Response Theory. International Educational and Professional. Publisher Newbury Park. London New Delhi.
- Lord. M. F. (1980). Application of Item Response Theory to Practical Testing Problems, Lawrence Erlbum Associates, Publishers Hillsdut. New Jersey.
- Wright. B. D. and Linacre. J. M. (1993). A Users Guide to Bigsteps, Chicago. Version.
- Masters. N. G. (1984). Constructing an Item Bank Using Partial Credit Scoring. Journal of Educational Measurement. Vol.21. No.(1).
- Rasch, g.(1961). On general laws and the meaning of measurement in psychology. Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability.
- Wright. B. D. and STONE, M.H (1979).Best Test Design. Chicago:

Jerash for Research and Studies Journal ???? ??? ?????? ?????????, Vol. 10 [2009], Iss. 1, Art. 4

MESA press.
- Wright. B. D. and BELL, S.R(1981). Fair and useful testing with item banks, Research Memorandum no 32. Chicago:University of Chicago, Department of Education, MESA Psychometrics Laboratory.
- Harris. J, Laan. S, and Mossenson. L. (1988) Applying Partial Credit Analysis to the Construction of Narrative Writing Tests, Applied Measurement in Education. Vol. 1. No (4).
- Adams. J. R. (1988). Applying the Partial Credit Model to Educational Diagnosis Applied Measurement in Education.Vol. 1. No. (4).
- Andrich, D.(1978).Application of a psychometric rating model to ordered categories which are scored with successive integers. Applied Psychological Measurement.
- Dodd. G. B. and Koch. W. (1987). Effect of Variations in Item Test Information in the Partial Credit Model. Applied Psychological Measurement. Vol. 11. No. (4).
- Susan. N. B. (2001). To Meet or not to Meet Standard, Proficiency Estimation Using Different Polytomous IRT Models, University of Washington. Degree: phD.
- Park, R. (2001). The effective of test length and sample size on the estimates and item parameters, using the partial credit model.

https://digitalcommons.aaru.edu.jo/jpu/vol10/iss1/4