الخصائص السيكومترية لاختبار محكي المرجع في القياس والتقويم التربوي وفق النظرية الحديثة في القياس التربوي والنفسي

د. نضال كمال الشريفين
قسم علم النفس التربوي
جامعة اليرموك- الأردن

الخصائص السيكومترية لاختبار محكي المرجع في القياس والتقويم التربوي وفق النظرية الحديثة في القياس التربوي والنفسي

د. نضال كمال الشريفين
قسم علم النفس التربوي
جامعة اليرموك- الأردن

الملخص

هدفت هذه الدراسة إلى تقدير الخصائص السيكومترية لاختبار محكي المرجع في القياس والتقويم التربوي وفق النظرية الحديثة في القياس. ولتحقيق هدف الدراسة تم بناء اختبار تحصيلي في القياس والتقويم التربوي مؤلف من ٥٠ فقرة من نوع الاختيار من أربعة بدائل. طبق على أفراد عينة الدراسة البالغ عددهم (٢٢٢) طالبا وطالبة. وأشارت النتائج إلى: مطابقة الاستجابات عن (٣٧) فقرة من فقرات الاختبار لافتراضات نموذج راش ،وحذف (١٣). لم تطابق النموذج، حيث كانت جميع فقرات الاختبار بصورته النهائية ضمن حدود المطابقة بالنسبة لمؤشري متوسطات المربعات الداخلية والخارجية، وبلغ معامل الثبات للأفراد (٠,٩٠) ومعامل الثبات للاختبار (٠,٩٥)، وتم التأكد من تمتع الاختبار بمظاهر متعددة من الصدق. كما بينت النتائج أن قيمة دالة المعلومات كانت أقصى ما يمكن عندما كانت ($\theta=b$) وذلك كما يتوقع من نموذج راش. وأن الاختبار يقدم أكبر كمية من المعلومات عند مستويات القدرة المتوسطة، ويقدم أقل كمية من المعلومات عند مستويات القدرة العالية والمتدنية.

* تاريخ تسلم البحث ٢٠٠٥/٩/١٢ * تاريخ قبوله للنشر ٢٠٠٦/٢/٧

The Psychometric Characteristics of a Criterion-Referenced Test in Measurement and Educational Evaluation Estimated According to Modern Theory in Measurement

Dr. Nedal Kamal Alshiraifin
Dept of Educational Psychology
Al-Yarmouk University- Jordon

Abstract

This study aimed at estimating the psychometric characteristics of a criterion- referenced test in measurement and educational evaluation according to the Modern Theory of Measurement. For achieving the objective of the study, an achievement test in measurement and educational evaluation was constructed. The test in its final form consisted of (50) multiple choice items. The test was administered on a sample of (222) male and female students. The results revealed that students' responses for (37) items of the test fit Rasch Model assumptions, whereas (13) items didn't fit the Model. All items in the final form of the test were located in the range of infit / outfit statistic. The coefficients of reliability obtained for persons and items were (0.90) and (0.95) respectively. The test validity was ensured through several approaches. The results also showed that information function for the test was maximum at (θ= b) and it was consistent with Rasch Model assumptions. The test presents the greatest amount of information at medium ability level and presents the least amount of information at both the highest and lowest ability levels.

الخصائص السيكومترية لاختبار محكي المرجع في القياس والتقويم التربوي وفق النظرية الحديثة في القياس التربوي والنفسي

د. نضال كمال محمد الشريفين
قسم علم النفس التربوي
جامعة اليرموك- الأردن

مقدمة الدراسة

يهدف التعليم الجامعي إلى بناء المعارف والمهارات والقيم للوصول بالمتعلم إلى مستوى الإتقان لتلك المهارات والمعارف (علام ، ٢٠٠١). ولتحقيق ذلك فقد حرصت كليات التربية في الجامعات على تنمية مهارات طلبتها، ورفع مستوياتهم المهنية، وإعدادهم للحياة، لا سيما أن خريجي تلك الكليات هم معلمو المستقبل والذين لا ينحصر دورهم في عملية التعليم على تحديد الأهداف المرجو تحقيقها في سلوك المتعلمين واختيار الأساليب والأنشطة التي تيسر تعليمهم ، وإنما ينبغي أيضا تقويمهم لمدى تحقيق تلك الأهداف التي ينعكس أثرها في التغيرات التي تحدث في سلوكهم نتيجة لعملية التعليم. وهذا يتطلب تنمية كفايات طلبة الجامعة في كليات التربية وغيرها في مجال القياس والتقويم. وأن تقدم لهم هذه الكليات برامج كافية ومناسبة لتحقيق ذلك. كما يتطلب توافر اختبارات لقياس تمكن طلبة كليات التربية من أساسيات القياس والتقويم.

وللتحقق من امتلاك الطلبة لهذه الكفايات لا بد من استخدام أدوات قياس مناسبة. حيث تشكل أدوات القياس والتقويم عنصرا أساسيا في الدراسات السيكومترية، إذ لاتكاد تخلو دراسة واحدة من أداة لقياس المتغيرات التي تعالجها، أو السمات التي تحاول قياسها وذلك من أجل تحديد مستوى امتلاك الطلبة لهذه الكفايات والمهارات. لذلك لا بد وأن يكون المعلم متمكنا من أساسيات القياس والتقويم التربوي والنفسي، وقادرا على تصميم وبناء الاختبارات التحصيلية الصفية، وغيرها من أدوات القياس، وأن يتمكن من تحليل وتفسير نتائج الاختبارات من أجل اتخاذ القرارات المناسبة المتعلقة بالمتعلمين، لا سيما وأن الاختبارات التحصيلية هي أكثر أدوات ووسائل القياس استخداما في المدرسة (الجامعة العربية المفتوحة، ٢٠٠٥) حيث تستخدم في التقويم التربوي لأغراض متعددة: في تقويم تحصيل الطلبة وتشخيص جوانب القصور في تعلم الطلبة وكذلك في تنبؤ تحصيل الطالب اللاحق من معرفة تحصيلة السابق ...الخ (الكيلاني والشريفين، ٢٠٠٥) .

والواقع الذي يجب أن لا نغفله، هو أن التعليم لم يعد مقتصرا على مجرد التمييز بين الطلبة في القدرة أو المستوى، وإنما يجب أن يركز على اكتسابهم مهارات معينة، وتحقيق أهداف

محددة ، بل وإتقان تلك المهارات وتلك المعلومات، وهذا ما تهدف إليه الاختبارات محكية المرجع والتي تهدف إلى مقارنة أداء الفرد بمستوى أداء مطلق ومحدد في مجال من السلوك أو بمستوى كفاية معين، دون الحاجة إلى مقارنة أداء الفرد بأداء الآخرين، كما هو الحال في الاختبارات معيارية المرجع. وبهذا فإن الاختبار المحكي المرجع يعتمد في جوهره على تقدير مستوى أداء الفرد بالنسبة لمحك أو مستوى أداء أو درجة قطع (Cut Score) (Popham, 1978). فعندما يود المعلم تحقيق مستوى إتقان معين لدى طلبته فينبغي عليه أن يجيب كل منهم إجابة صحيحة عن نسبة مئوية من فقرات الاختبار، يتم تحديدها مسبقا وتسمى بدرجة القطع؛ ودرجة القطع هي الدرجة التي ينبغي أن يحصل عليها الفرد في المجال الشامل (Universe Specifications) لفقرات الاختبار، لكي يكون متقنا لمحتوى أو مهارة معينة. وأحيانا تشير درجة القطع إلى الحد الأدنى للأداء المقبول، لكي يتمكن الفرد من أداء مهام تالية (Guskey, 1988).

ونظرا لأهمية مستويات الأداء في اتخاذ القرارات التعليمية المتعلقة بالأفراد، فقد اهتم علماء القياس التربوي باقتراح العديد من الطرق، التي يمكن الاسترشاد بها في تحديد هذه المستويات، حيث يزيد عددها في الوقت الحاضر على (٣٥) طريقة (علام، ٢٠٠١). وقد اعتادت مؤسسات التعليم في الأردن تحديد مستويات الأداء في الاختبارات الصفية بـــ (٥٠) من الدرجة الكلية في معظم الاختبارات. فإذا حصل الفرد على هذه النسبة أو أعلى منها عد ناجحا أو اجتاز الاختبار، أما إذا حصل على أقل من هذه النسبة عد راسبا، بالإضافة إلى عدم استناد هذه المستويات إلى أساليب منهجية منظمة مما يجعل قرارات التصنيف مشوبة بقدر كبير من الخطأ. فالحد الأدنى لدرجة النجاح أو الاجتياز الذي تحدده النسبة، تكون موحدة في معظم المواد الدراسية، ولجميع الأفراد المفحوصين، بالرغم من اختلاف صعوبة فقرات هذه الاختبارات. كما أن نسبة (٥٠٪) تمثل مستويات من المعرفة تتباين تباينا كبيرا من اختبار إلى آخر، كما أن المجالات السلوكية التي تقيسها هذه الاختبارات والتي تنسب إليها هذه الدرجات، لا تكون محددة، مما يجعل تفسير معنى ودلالة هذه النسب صعبا في ضوء المعارف والمهارات التي يقيسها الاختبار.

وقد جاءت هذه الدراسة لبيان كيفية بناء اختبار محكي المرجع في أساسيات القياس والتقويم، من خلال انتقاء فقرات لها خصائص سيكومترية محدده، حسب الهدف من الاختبار، سواء أكان لاختيار أفراد بقدرات عالية، أم لتصنيف الأفراد إلى متقنين وغير متقنين، وكذلك لتقدير علامات الأفراد على مجال محدد، من خلال، انتقاء فقرات ممثلة لمجال سلوكي محدد، وبخصائص سيكومترية محددة.

وقد اعتمدت هذه الدراسة في تقييم الخصائص السيكومترية لاختباره وفقراته على النظرية الحديثة في القياس، أو ما يعرف بنظرية السمات الكامنة، أو نظرية استجابة الفقرة باعتبارها تشكل إطارا للتوجه الحالي والمستقبل المنظور في اختيار الفـــقـــرات

(Anastasi , 1982) وكذلك على اعتبار أنها نظرية بديلة للنظرية الكلاسيكية والتي تعرضت لمجموعة من الانتقادات وفق ما أشار هامبلتون وسوامنيثان (Hambleton & Swaminathan, 1985)، بالإضافة لمعاناتها عيوب وجوانب ضعف متعددة (Stocking, 1999). ومن أهم تلك الجوانب أن الخصائص السيكومترية للاختبار تحدد بالنسبة لمجموعة من المفحوصين، كما أن خصائص المفحوصين، تحدد بالنسبة إلى اختبار معين، وليس بشكل مطلق. وكذلك تفترض أن الخطأ المعياري في القياس متساو لكل المفحوصين وهذا يفتقر إلى الدقة ؛ زيادة على أنها لا تبين مدى تحصيل المفحوص على مستوى الفقرة، وإنما على الاختبار ككل (Swaminthan & Rogers, 1999 Hambelton,). ولذلك ظهرت النظرية الحديثة في القياس على أنها نظرية بديلة عن النظرية الكلاسيكية، والتي يفترض فيها تلافي عيوب النظرية الكلاسيكية، كما أنها تعد ثمرة محاولات المهتمين في القياس النفسي والتربوي لتطوير مقاييس أكثر دقة في قياسها للسمات النفسية والتربوية، وكمحاولة لحل المشكلات التي تعانيها النظرية الكلاسيكية (عودة، ١٩٩٢). كما يقال إن هذه النظرية تمكن المختص من الإجابة عن أي تساؤل يثار حول الفقرة أو الاختبار أو المفحوص (الشريفين ،٢٠٠٣). كما أنها تعالج كثيرا من قضايا القياس بشكل أكثر فاعلية من النظرية الكلاسيكية . حيث تفترض هذه النظرية إمكان التنبؤ بأداء الأفراد، وتفسير أدائهم في الاختبار النفسي أو التربوي في ضوء خاصية أو خصائص مميزة لهذا الأداء تسمى السمات (Traits)، وتحاول هذه النظرية تقدير درجات الأفراد على هذه السمات .

وقد انبثقت عن هذه النظرية مجموعة من النماذج والتي تعرف بنماذج السمات الكامنة (Latent Trait Models) وتهدف جميعها إلى تحديد علاقة بين أداء الفرد على الاختبار– وهو ما يمكن ملاحظته ملاحظة مباشرة– وبين السمات أو القدرات التي تكمن وراء هذا الأداء وتفسره. ولقد بين هامبلتون وسوامنيثان (Hambleton & Swaminathan, 1985, p11) ثلاث مزايا رئيسة لنظرية السمات الكامنة أو النظرية الحديثة في القياس هي:

أولا: بافتراض وجود مجموعة كبيرة نسبياً من فقرات الاختبار التي تقيس السمة نفسها، يكون تقدير قدرة الفرد مستقلا عن عينة الفقرات التي تطبق عليه. أي أن تقدير قدرات الأفراد متحررة من الفقرات

ثانيا : بافتراض وجود مجتمع كبير من الأفراد، يكون تقديرات الخصائص السيكومترية للفقرات (مثل معاملات الصعوبة والتمييز) مستقلة عن عينة الأفراد التي استخدمت في تقدير هذه الخصائص، أي لا تتأثر معالم الفقرات بالعينة المختارة من المجتمع.

ثالثا: يمكننا الحصول على إحصائي مثل (الخطأ المعياري في التقدير) لتقدير درجة الدقة في قياس القدرة لكل فرد. وربما يختلف هذا الإحصائي من فرد إلى آخر. وتقوم هذه النظرية على افتراضات أساسية (Swaminathan, 1985; Hambleton & Jonse, 1993

Hambleton &) هي:

- افتراض أحادية البعد :(Unidimensionality) أي أن هناك قدرة واحدة تفسر أداء الفرد في الاختبار، ولذلك تسمى بالنماذج أحادية البعد، أما النماذج التي تفترض وجود أكثر من قدرة واحدة تكمن وراء هذا الأداء فإنها تسمى متعددة الأبعاد.
- افتراض الاستقلال الموضعي : (Local Independence) يقصد بهذا الافتراض أن تكون استجابات الفرد عن الفقرات المختلفة في الاختبار مستقلة استقلالا إحصائيا (Statistically Independent) عند مستوى معين من القدرة.
- افتراض منحنى خصائص الفقرة (Item Characterstic Curve) ويمكن وصف العلاقة بين تحصيل المفحوص على الفقرة وقدرته المقيسة بالاختبار، من خلال اقتران تراكمي صاعد يعرف باسم منحنى خصائص الفقرة، أو دالّة الاستجابة للفقرة. حيث يوفر هذا المنحنى احتمالات إجابة المفحوصين عن الفقرة في مستويات القدرة المختلفة إجابة صحيحة. وكون المنحنى تراكميا صاعدا، يشير بوضوح إلى أن احتمال إجابة الفقرة إجابة صحيحة يزداد بازدياد قدرة المفحوص. وفي العادة توصف هذه المنحنيات في نماذج الاختبارات المصممة لقياس سمة واحدة (أحادية البعد) بدلالة مَعْلَم واحدة (صعوبة الفقرة)، أو معلمين (صعوبة الفقرة وتمييزها)، أو ثلاثة معالم (الصعوبة والتمييز والتخمين) أو أكثر حيث يرجع الاختلاف الرئيس بين نماذج السمات المستخدمة إلى اختلاف شكل منحنى خصائص الفقرة.
- السرعة في الأداء (Speededness) حيث تفترض نماذج النظرية الحديثة أن عامل السرعة لا يقوم بدور في الإجابة عن فقرات الاختبار بمعنى أن إخفاق الأفراد في الإجابة عن فقرات الاختبار يرجع إلى انخفاض قدرتهم وليس إلى تأثير عامل السرعة في إجاباتهم.

واستخدم في تحليل استجابات أفراد عينة الدراسة نموذج المعلم الواحد، أو ما يسمى بنموذج راش (Rasch Model)، والذي يعد أهم نماذج السمات الكامنة، باعتباره يحقق خاصية الموضوعية في القياس النفسي والتربوي (Wright & Stone, 1979) وحتى تتوافر متطلبات الموضوعية في القياس يجب أن تستوفى فروض النموذج وهي: أحادية البعد، واستقلال القياس، والذي يعني استقلال تقدير معلم الصعوبة للفقرة عن تقديرات قدرات الأفراد (Person Free)، واستقلال تقدير معلم القدرة للفرد عن عينة الفقرات، وكذلك توازي منحنيات الخصائص للفقرات. فعندما تكون الفقرات متطابقة للنموذج، يكون هناك شكل أو انحناء عام لمنحنيات الخصائص للفقرات، أي أن هذه المنحنيات متوازية وعندئذ تكون لها القدرة نفسها على التمييز بين الأفراد على متصل السمة. كما أن نموذج راش يتعامل مع الفقرات ذات الاستجابات الثنائية، ليؤلف مجموع العلامات التي يحصل عليها الفرد على هذه الفقرات إحصاءا كافيا، لتقدير قدرة الفرد. كما أن مجموع الأفراد الذين أجابوا إجابة صحيحة عن الفقرة يؤلف إحصاءً كافيا لتقدير درجة صعوبتها. علاوة على ذلك، فإن نموذج راش يمكن التعامل معه بسهولة بالمقارنة مع النماذج الأخرى.

ومن المميزات التي أعطت هذا النموذج هذه الأهمية وفق ما أشار إليه ماسترز (1982 ,Masters) أنه عندما تتطابق البيانات مع هذا النموذج، فإن معالم الفقرات التي تتمثل في درجة صعوبتها يمكن تقديرها بشكل مستقل عن العينة، كما أن قدرات الأفراد يمكن تقديرها بشكل مستقل عن درجة صعوبة الفقرات.

ويعبر عنه بالعلاقة التالية التي تقيس احتمال إجابة الفرد ذي القدرة θ عن الفقرة i. (i=1,2,.....,n)

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

حيث $P_i(\theta)$: تمثل احتمال الإجابة الصحيحة لفرد قدرته؟ عن الفقرة i حيث e=2.718 وهو ثابت رياضي وأن D=1.7 وهو عامل تدريج (Scaling Factor) يجعل القدرة المستنتجة من استعمال هذا النموذج مساوية للقدرة المستنتجة من استعمال النموذج الذي يعتمد على المنحنى الطبيعي التراكمي (Swaminathan 1985, p37 Hambleton,&) والرمز :b_i معلم الصعوبة للفقرة i.

مشكلة الدراسة وأهميتها

تهدف هذه الدراسة إلى تقدير الخصائص السيكومترية لاختبار محكي المرجع في القياس والتقويم، وفق النظرية الحديثة في القياس التربوي والنفسي، وانبثقت مشكلة هذه الدراسة من عدم توافر أداة قياس تتمتع بخصائص سيكومترية مقبولة، لقياس تحصيل الطلبة في مساق القياس والتقويم، وبناء الاختبارات في الجامعات، حيث تبدو الحاجة ملحة لتوافر أداة قياس موضوعية، يمكن من خلالها التحقق من مدى تمكن أو إتقان الطلبة لموضوعات القياس والتقويم. وذلك نظراً لأهمية إتقان مثل هذه المهارات والمعارف في قياس تحصيل الطلبة في الجامعات وغيرها من مؤسسات التعليم. كما أنه يمكن أن يكون هذا الاختبار أداة مناسبة لتشخيص ومعرفة الأهداف التي تحققت، والأهداف التي لم تتحقق، واقتراح الأساليب التعليمية والتدريبية في ضوء نتائج عملية التشخيص. كما تتمثل أهمية هذه الدراسة في بيان آلية بناء اختبار محكي المرجع في أحد المساقات التي تدرس بكلية العلوم التربوية في جامعة آل البيت، والذي يدرس أيضا في كليات التربية في الجامعات الأردنية الأخرى، بحيث يتمتع بدرجة معقولة من الصدق والثبات، لاستخدامه لاحقا، وكذلك للتحقق من التقدم الكبير، الذي أحرزته النظرية الحديثة في تطوير مقاييس أكثر دقة في قياسها للسمات النفسية والتربوية، في التحرر منأاثر قدرة الفرد على معالم الفقرات ومن أثر معالم الفقرات على قدرة الفرد الذي يجيب عنها، وعلى وجه التحديد فإن هذه الدراسة ستحاول بناء اختبار محكي المرجع في القياس والتقويم، يكون صالحا للتطبيق في الجامعات

الأردنية وذلك بتوفير مؤشرات عن دلالات ثبات الاختبار، ودلالات صدق الاختبار، ومعايير الأداء.

تساؤلات الدراسة

لتحقيق الأهداف ستحاول هذه الدراسة الإجابة عن التساؤلات التالية :

١. ما درجة مطابقة الاستجابات عن فقرات الاختبار مع نموذج راش؟

٢. ما الخصائص السيكومترية للاختبار (الصدق والثبات)؟

٣. ما كمية المعلومات التي يقدمها الاختبار عند مستويات القدرة المختلفة؟

تعريف المصطلحات

- **نموذج راش** :(Rash Mdel) أحد نماذج النظرية الحديثة الذي يفترض تساوي جميع فقرات الاختبار في التمييز بين مستويات القدرة المقاسة وأن الفرد لا يلجأ إلى التخمين العشوائي في إجابته عن فقرات الاختبار (Wrtight & Stone, 1979) (c =0).
- **الخصائص السيكومترية**: يقصد بالخصائص السيكومترية وفق النظرية الحديثة: معالم الصعوبة والتمييز والتخمين للفقرة، والصدق والثبات. وفي هذه الدراسة اقتصرت الخصائص السيكومترية على معالم الصعوبة، والصدق والثبات (علام، ٢٠٠١).
- **الاختبار المحكي المرجع**: اختبار يستخدم لتحديد مستوى أداء الفرد بالنسبة لمجال سلوكي محدد تحديداً ومعرف تعريفاً جيداً (علام، ٢٠٠١).

منهجية الدراسة وإجراءاتها:

مجتمع الدراسة

مجتمع الدراسة يتكون من طلبة كليات التربية في الجامعات الأردنية ؛ ونظرا لتعذر تطبيق اختبار موحد على طلبة الجامعات ، فقد تم اختيار العينة بالطريقة المتيسرة (المتاحة).

عينة الدراسة

عينة الدراسة تكونت عينة الدراسة من الطلاب والطالبات المسجلين في مساق القياس والتقويم التربوي في جامعة آل البيت للفصل الدراسي الأول ٢٠٠٤ / ٢٠٠٥ والبالغ عددهم (٢٢٢) طالبا وطالبة موزعين على (٤) شعب.

أداة الدراسة

تم بناء الاختبار المحكي المرجع في موضوعات القياس والتقويم وفق الخطوات الآتية:

١. تم تحديد الغرض من الاختبار، وهو قياس تحصيل الطلبة في موضوعات القياس والتقويم.

٢. تم تحديد محتوى الاختبار والأهداف السلوكية المتعلقة به، حيث استعان الباحث ببعض الدراسات السابقة (علام، ٢٠٠١؛ علام، ١٩٨٥؛ Popham, 1980) والتي أجريت في هذا المجال؛ لتحديد موضوعات المساق، وكذلك من خبرته التدريسية في تدريس مساقات القياس والتقويم، وببعض المراجع التي تدرس في بعض الجامعات كأساسيات في القياس والتقويم، واستخلص منها خمسة موضوعات يفترض أن تتحقق عند المتعلم في مستوى الإتقان التام واشتملت على الموضوعات الآتية: مفاهيم أساسية في القياس والتقويم، الأهداف التدريسية، بناء الاختبارات التحصيلية، خصائص الاختبار الجيد، وتحليل نتائج الاختبارات.

٣. تم صياغة الأهداف السلوكية التي تضمنتها هذه الموضوعات والتي يفترض أن تتحقق عند المتعلمين في مستوى الإتقان.

٤. تم إعداد جدول مواصفات للاختبار جرى فيه ربط مستويات الأهداف بمحتوى المادة الدراسية موضوع الاختبار .

٥. قام الباحث بصياغة (٦٠) فقرة من نوع الاختيار من متعدد، بأربعة بدائل أحدها يمثل الإجابة الصحيحة؛ لقياس كل هدف من الأهداف السلوكية، وقد تطلب قياس بعضها أكثر من فقرة واحدة، وقد روعيت الأسس الفنية في كتابة هذا النوع من الفقرات، ومطابقتها للهدف السلوكي الذي تقيسه من حيث المحتوى والمستوى المعرفي، كما روعي ألا تعتمد إجابة إحدى الفقرات على إجابة الفقرات الأخرى، وهو ما يتطلبه نموذج راش كأحد نماذج النظرية الحديثة في القياس التربوي والنفسي .

٦. للتأكد من صدق المحتوى، تم عرض فقرات الاختبار على خمسة محكمين من أعضاء هيئة التدريس في الجامعات والذين يُدّرسون مساق مبادئ أساسية في القياس والتقويم، وذلك لأخذ وجهات نظرهم حول الفقرات والاختبار ككل واقتراح ما يرونه من تعديل مناسب، وطلب من كل منهم أن يبدي رأيه في: درجة الاتفاق بين الأهداف السلوكية والهدف الأساسي للاختبار وتمثيلها لها، ومدى قياس الفقرات للهدف السلوكي المتعلق بها، ومدى اتساق الاختبار ككل مع الهدف الأساسي منه. وبناء على ملاحظاتهم تم حذف بعض الفقرات التي أجمع المحكمون أن هناك فقرات تؤدي غرضها، وبلغ عددها ٦ فقرات. وبعد ذلك قام الباحث بتعديل الفقرات، ومراجعتها وبلغ عددها في صورتها الأولية (٥٤) فقرة أجمع المحكمون على أنها عينة كافية وممثلة للمجال السلوكي الذي يقيسه الاختبار.

٧. التجريب الأولي للاختبار: تم تجريب الصورة الأولية للاختبار على إحدى شعب القياس والتقويم في نهاية الفصل الصيفي للعام الدراسي (٢٠٠٣ / ٢٠٠٤) بغرض التعرف على مستوى صعوبة الفقرات، وتحديد الزمن المناسب لتطبيق الاختبار، والتأكد من وضوح الصياغة اللغوية للفقرات، والكشف عن الصعوبات المتوقعة أن تواجه الطلبة، وكذلك

إيجاد معامل الثبات، و لم يكشف التجريب الأولي عن أية ملاحظات ذات أهمية فيما يتعلق بوضوح الفقرات و لم تواجه الطلبة أية صعوبات.

المعالجة الإحصائية

وقد أجريت عملية التحليل لبيانات التجريب الأولي باستخدام برنامج التحليل الإحصائي (SPSS) فقد تم حساب معامل الصعوبة لكل فقرة، وذلك بإيجاد نسبة الطلبة الذين أجابوا عن الفقرة إجابة صحيحة من بين المفحوصين الذين حاولوا الإجابة عن هذه الفقرات، كما تم حساب معاملات التمييز للفقرات، والتي تمثل معاملات الارتباط الثنائية بين الفقرات والاختبار المصححة والجدول (١) يوضح معاملات الصعوبة ومعاملات التمييز لكل فقرة من فقرات الاختبار للعينة الاستطلاعية.

الجدول رقم (١)

معاملات الصعوبة ومعاملات التمييز لكل فقرة من فقرات الاختبار على العينة الاستطلاعية

رقم الفقرة	معامل الصعوبة	معامل التمييز	رقم الفقرة	معامل الصعوبة	معامل التمييز	رقم الفقرة	معامل الصعوبة	معامل التمييز
١	٠,٣٧	٠,٤٧	١٩	٠,٣٦	٠,٦٤	٣٧	٠,٤٧	٠,٣٩
٢	٠,٥٦	٠,٥٢	٢٠	٠,٣٢	٠,٤٥	٣٨	٠,٣٦	٠,٢٤
٣	٠,٤٢	٠,٦٧	٢١	٠,٥٢	٠,٣٩	٣٩	٠,٣٩	٠,٥٢
٤	٠,٦١	٠,٧٢	٢٢	٠,٤٨	٠,٤٠	٤٠	٠,٥٦	٠,٦١
٥	٠,٥٧	٠,٥١	٢٣	٠,٥٤	٠,٤١	٤١	٠,٤٤	٠,٥٧
٦	٠,٥٣	٠,٥٨	٢٤	٠,٥٨	٠,٣٧	٤٢	٠,٦١	٠,٥٢
٧	٠,٥٢	٠,٤٣	٢٥	٠,٦١	٠,٦٢	٤٣	٠,٥٢	٠,٧٦
٨	٠,٥٠	٠,٤٤	٢٦	٠,٣٧	٠,٥١	٤٤	٠,٦٠	٠,٨١
٩	٠,٤٤	٠,٥٧	٢٧	٠,٤١	٠,٥٧	٤٥	٠,٥٤	٠,٦٢
١٠	٠,٤٨	٠,٥٦	٢٨	٠,٥٦	٠,٦٣	٤٦	٠,٤٤	٠,٧٢
١١	٠,٤٩	٠,٦٢	٢٩	٠,٥٤	٠,٤٤	٤٧	٠,٥١	٠,٧٣
١٢	٠,٥٩	٠,٧٢	٣٠	٠,٦٠	٠,٥٧	٤٨	٠,١٦	٠,١٩
١٣	٠,٦١	٠,٨٠	٣١	٠,٧٢	٠,٧٢	٤٩	٠,٥٧	٠,٧٤
١٤	٠,٧٢	٠,٧١	٣٢	٠,٧٤	٠,٤١	٥٠	٠,٦٢	٠,٧٥
١٥	٠,٢١	٠,١٨	٣٣	٠,٦٢	٠,٤٨	٥١	٠,٧٧	٠,٧٠
١٦	٠,٦١	٠,٨٤	٣٤	٠,٥١	٠,٤٦	٥٢	٠,٧١	٠,٥١
١٧	٠,٧١	٠,٦١	٣٥	٠,٥٥	٠,٤٧	٥٣	٠,٨٦	٠,٣٢
١٨	٠,٥١	٠,٦٣	٣٦	٠,٤٦	٠,٤٤	٥٤	٠,٦٦	٠,٣٦

يتضح من الجدول (١) أن معاملات الصعوبة في الاختبار تراوحت بين (٠,١٦– ٠,٨٦) بوسط قدره (٠,٥٣) أمّا معاملات التمييز فقد تراوحت بين (٠,١٨ –٠,٨١) بوسط قدره (٠,٥٤) وتم حذف الفقرات التالية (١٥، ٣٨ ، ٤٨) نظرا لتدني معاملات التمييز لها، فقد كانت (٠,١٨، ٠,٢٤، ٠,١٩) على التوالي. أما معاملات الصعوبة لهذه الفقرات فقد كانت على التوالي، (٠,٢١، ٠,٣٦، ٠,١٦) ويلاحظ أنها كانت جميعها مقبولة باستثناء الفقرة (٤٨) حيث بلغت قيمة معامل الصعوبة لها (٠,١٦) وهي قيمة متدنية. وبعد حذف هذه الفقرات أصبح عدد الفقرات بصورتها النهائية (٥١) فقرة تم حذف الفقرة (٥٣) حيث معامل الصعوبة لها (٠,٨٦) ليصبح عدد الفقرات للاختبار بصورته النهائية (٥٠) فقرة، وتم حساب قيمة معامل ثبات الاتساق الداخلي على الفقرات المتبقية باستخدام معادلة كودر–ريتشاردسون رقم (٢٠) (KR-20) وكانت قيمة معامل الثبات للاختبار (٠,٩٠) وهي قيمة مقبولة .

٨. إجراءات جمع البيانات: بعد أن قام الباحث ببناء الاختبار بصورته النهائية انتظر حتى نهاية الفصل الدراسي الأول من العام الجامعي (٢٠٠٤ /٢٠٠٥) وتم تطبيق الاختبار على أربع شعب دراسية من الشعب التي قام بتدريسها . ورغم وضوح تعليمات الاختبار فقد تمت الإجابة عن استفسارات الطلبة جميعها . وبعد أن قام الباحث بجمع استجابات الطلبة عن الاختبار، تمت مراجعة أوراق الطلبة جميعها للكشف عن الاستجابات ذات النمطية الواحدة . وبناء على ذلك لم يستثن الباحث أية ورقة من التحليل.

عرض النتائج ومناقشتها

أولاً: عرض النتائج المتعلقة بدرجة مطابقة الاستجابات عن فقرات الاختبار مع نموذج راش:

لمعرفة درجة مطابقة الاستجابات عن فقرات الاختبار مع نموذج راش، تم تصحيح فقرات الاختبار حيث أعطيت الاستجابة الصحيحة العلامة (١) والخاطئة العلامة (صفر). بعد ذلك قام الباحث بإدخال البيانات في ذاكرة الحاسوب، وأخضعت البيانات للتحليل باستخدام برنامج (BIGSTEPS) المصمم لتحليل البيانات وفقاً لنموذج راش (Linacre & Wright, 1993).

وفيما يلي خطوات تحليل استجابات أفراد عينة الدراسة البالغة (٢٢٢) فردا عن فقرات الاختبار(٥٠) فقرة، وهو عدد مناسب وفق ما أشار إليه فان (Fan, 1998) من أن الحد الأدنى لعدد الأفراد وفق نموذج المعلم الواحد (نموذج راش) يجب أن لا يقل عن ٢٠٠ فرد، بينما نموذج الثلاثة معالم يحتاج ما لا يقل عن ١٠٠٠ فرد للحصول على نتائج دقيقة:

- البرنامج يحذف استجابات الأفراد الحاصلين على العلامة الكاملة ٥٠ أو الحاصلين على العلامة صفر، وكذلك الفقرات التي يجيب عنها أفراد العينة جميعهم إجابة صحيحة أو تلك التي يخفق فيها الأفراد جميعهم. وكذلك الأفراد الذين تزداد استجاباتهم المحذوفة أو

المتروكة كلما اتجهنا نحو نهاية الاختبار. أو نحو بداية الاختبار.

- ليس هناك فرد حصل على العلامة الكاملة ٥٠ وليس هناك فرد واحد حصل على العلامة صفر، وعلى هذا فهناك (٢٢٢) فرداً دخلت استجابتهم في التحليل بوساطة البرنامج. حيث لم يستثن البرنامج أي فرد من التحليل.
- ليس هناك أي فقرة أجاب عنها أفراد عينة الدراسة جميعهم إجابة خاطئة، أو أجاب عنها أفراد العينة جميعهم إجابة صحيحة، وعلى هذا فإن البرنامج لم يستثن أية فقرة من التحليل، وبالتالي فإن الفقرات جميعها وعددها (٥٠) فقرة داخلة في التحليل.
- قبل البدء بتحليل الاستجابات عن الفقرات لا بد من حذف الأفراد غير المطابقين للنموذج، ولمعرفة مؤشرات المطابقة الخاصة بالأفراد، تم تقدير قدرة كل فرد بالإضافة إلى الخطأ المعياري في قياس هذه القدرة، وكذلك قيم إحصائي المطابقة الكلي وله مؤشران: مؤشر المطابقة الداخلية (التقاربي) (ZSTD) ويعبر عنه كذلك بإحصائي متوسط المربعات التقاربي (MNSQ)، وهو مؤشر إحصائي للسلوكات غير المتوقعة التي تؤثر في استجابات الأفراد عن الفقرات التي تكون قريبة من مستوى قدراتهم. وكذلك مؤشر المطابقة الخارجية (التباعدي) (ZSTD) ويعبر عنه كذلك بإحصائي متوسط المربعات التباعدي وهو مؤشر إحصائي يعد بديلاً عن السابق، وله صفات مقاربة ومشابهة، ولكنه أكثر حساسية للسلوكات غير المتوقعة من الأفراد عن الفقرات التي تبتعد عن مستوى قدرة الأفراد. وذلك لكل تقدير من تقديرات القدرة، والجدول (٢) يبين المتوسط الحسابي والانحراف المعياري لكل تقدير من تقديرات القدرة، والخطأ المعياري في قياس هذه القدرة، وقيم إحصائيات المطابقة الداخلية والخارجية ومتوسطات المربعات للمطابقة الداخلية والخارجية.

الجدول رقم (٢)

المتوسط الحسابي والانحراف المعياري لكل من تقديرات قدرات الأفراد والخطأ المعياري في قياس القدرة وإحصاءات المطابقة الداخلية والخارجية

	القدرة	الخطأ المعياري	INFIT إحصائي المطابقة الداخلية		OUTFIT إحصائي المطابقة الخارجية	
			متوسط المربعات MNSQ	قيمة الإحصائي ZSTD	متوسط المربعات MNSQ	قيمة الإحصائي ZSTD
المتوسط الحسابي	٠,١٧-	٠,٣١	١	٠,١-	١,٠١	٠,١-
الانحراف المعياري	٠,٦٧	٠,٠٢	٠,١٦	١,٣	٠,٢٣	١,٤

يتضح من الجدول (٢) اقتراب المتوسط الحسابي لمتوسطات المربعات الموزونة من الواحد وهو الوضع المثالي كما يتوقعه النموذج، كما يلاحظ أن متوسط قيم إحصائي المطابقة الداخلية قد بلغ (−٠,١) والانحراف المعياري لها (١,٣) وهي أيضاً تقترب من القيم المثالية التي يفترضها النموذج وهي (صفر، ١) على الترتيب، وكذلك يلاحظ أن متوسط قيم إحصائي المطابقة الخارجية (التباعدية) قد بلغ (−٠,١) والانحراف المعياري لها يساوي (١,٤)، وهي أيضاً تقترب من القيم المثالية التي يفترضها النموذج وهي (صفر، ١) على الترتيب.

ولدى تفحص قيم إحصائي المطابقة الخارجية للأفراد، أو ما يسمى إحصائي المطابقة الكلية للأفراد، والذي يشير إلى مطابقة قدرة الفرد مع قدرات مجموعة الأفراد التي ينتمي إليها في قياس السمة التي يقيسها الاختبار. ولتعرف كيفية حساب هذا المعامل يمكن الرجوع إلى رايت وماسترز (Wright & Masters, 1982) وإذا كانت قيمة هذا الإحصائي تزيد عن (+٢) تعدّ قدرة الفرد غير متطابقة مع قدرات مجموعة الأفراد (Julian, 1988; Alastair & Hutchinson, 1987). وبناء على ذلك تبين وجود (١٩) فرداً تبتعد استجاباتهم الملحوظة عن الاستجابات المتوقعة تبعاً لقدراتهم، بمعنى أن قيم إحصائي المطابقة الخارجية المقابلة لدرجاتهم تزيد على (+٢) أو قيم متوسطات المربعات المناظرة لهذه الدرجات تزيد على (١) وهي القيم التي يتوقعها النموذج (Wright & Stone, 1979). ويعد هؤلاء الأفراد غير مطابقين للنموذج، لأن استجاباتهم الملحوظة تبتعد عن توقعات النموذج، كأن يجيب الفرد عن فقرة ما إجابة خاطئة على الرغم من أن صعوبة الفقرة دون مستوى قدرته، أو أن يجيب الفرد عن فقرة ما إجابة صحيحة، على الرغم من أن مستوى صعوبة الفقرة تفوق قدرته.

كذلك فإن الجدول (٣) يحوي رسما تخطيطيا معبراً عن مؤشرات المطابقة لكل فرد من الأفراد الذين تعد استجاباتهم من أكثر الاستجابات غير المطابقة للنموذج وكذلك مع التوضيح بالرسم حدود المطابقة وموقع كل فرد من هذه الحدود، وذلك بالنسبة لمؤشري المطابقة الداخلية والخارجية.

الجدول رقم (٣)

رسم تخطيطي معبر عن مؤشرات المطابقة لكل فرد مع توضيح حدود المطابقة وموقع كل فرد من هذه الحدود وذلك لمؤشري المطابقة الداخلية والخارجية وموقع قدرة الفرد

PERSON FIT GRAPH: MISFIT ORDER

ENTRY NUMBER	MEASURE - +	INFIT MEAN-SQUARE 0 0.7 1 1.3 2	OUTFIT MEAN-SQUARE 0 0.7 1 1.3 2	PER
201	*	: . *	: . : *	201
158	*	: . :*	: . : *	158
114	*	: . :*	: . : *	114
12	*	: . *	: . : *	012
118	*	: . *	: . : *	118
173	*	: . :*	: . : *	173
98	*	: . :*	: . : *	098
163	*	: *. :	: . : *	163
119	*	: . *	: . : *	119
165	*	: . *	: . : *	165
55	*	: . *	: . : *	055
2	*	: . *	: . :*	002
96	*	: . *	: . :*	096
200	*	: .* :	: . :*	200
109	*	: . *:	: . :*	109
47	*	: . *:	: . :*	047
116	*	: . *:	: . :*	116
110	*	: . *:	: . :*	110
199	*	: . *:	: . :*	199

ويلاحظ من الجدول (٣) أن الأفراد المستبعدين جميعهم، كانت متوسطات المربعات للمطابقة التباعدية المقابلة لقدراتهم خارج حدود المطابقة، وهي (٠,٧–١,٣). وهذا يعني أن استجاباتهم الملحوظة كانت تبتعد عن القيم المتوقعة، بينما متوسط المربعات للمطابقة الداخلية كان معظمها ضمن المدى المتوقع، وهذا مؤشر إلى أن هؤلاء الأفراد كانت استجاباتهم الملحوظة عن الفقرات الملحوظة من مستوى قدرتهم مقبولة وضمن حدود المطابقة، ولكن استجاباتهم عن الفقرات البعيدة عن مستوى قدرتهم غير متوقعة، وهذا هو السبب وراء عدم مطابقتهم للنموذج.

وكذلك الجدول (٤) يوضح أكثر الاستجابات غير المطابقة من الأفراد المستبعدين وأماكنها بملف البيانات.

الجدول رقم (٤)

أكثر الاستجابات غير الملائمة من قبل الأفراد المستبعدين

```
MOST MISFITTING RESPONSE STRINGS

PERSON    OUTMNSQ  | ITEM
| 111 211 441 43 42321124254 3122 4343334
|98696134153730472622517450653209488979072
high-----------------------------------------
201 201    1.68 A|.........................11......1111.
158 158    1.68 B|...................1111..1....1..1111...
114 114    1.67 C|00.................................111
12 012    1.65 D|.....................1.1..1.......1.1
118 118    1.58 E|00.....................................1.
173 173    1.56 F|...............................1....1.1.
98 098    1.55 G|00.....................................1.
163 163    1.55 H|.............1...........11.......1...1
119 119    1.54 I|0.0.0..................................
165 165    1.51 J|0.............................111.11.1.
55 055    1.50 K|0.............................1.1...1.
2 002    1.49 L|0.............................11...1
96 096    1.47 M|.00.....................................1.
200 200    1.47 N|.................11..1.....1....1...1
109 109    1.43 O|.........................11.....1..1.
47 047    1.42 P|.......................1.11..11..1..
116 116    1.41 Q|.....................................11.
110 110    1.41 R|...................................11.11.
199 199    1.40 S|0.................................11.11.
|-------------------------------------low-
|91116211144134374232112425453122443433342
| 869 134 537 04 26225174506 3209 8897907
```

فمثلاً الفرد رقم (٢٠١) أجاب عن فقرات مستوى صعوبتها أعلى من مستوى قدرته، فقد أجاب إجابة صحيحة عن الفقرات (٣٣، ١٢، ٣٧، ٣٩، ٣٠، ٤٧) رغم أن مستوى صعوبة هذه الفقرات يفوق قدرته، وكذلك فإن الفرد رقم (١١٩) أخطأ في الإجابة عن الفقرات (٩، ١٦، ٦) رغم أن مستوى صعوبتها دون مستوى قدرته.

بعد استبعاد الأفراد الذين لم تتطابق استجاباتهم مع توقعات النموذج، وإعادة التحليل لاختبار مدى مطابقة الفقرات للنموذج، تم تقدير معلم الصعوبة لكل فقرة، بالإضافة إلى الخطأ المعياري في قياس هذا المعلم، وقيم إحصائي المطابقة الداخلية (التقاربية) للفقرات أو ما يسمى بإحصائي المطابقة الكلي للفقرات (ZSTD) ومتوسط المربعات للفقرات ، وكذلك قيم إحصائي المطابقة الخارجية (التباعدية) للفقرات أو ما يسمى إحصائي المطابقة الكلي للفقرات وإحصائي متوسط المربعات لكل معلم من معالم الصعوبة.

والجدول (٥) يبين المتوسط الحسابي والانحراف المعياري لكل من تقديرات معالم الصعوبة للفقرات، والخطأ المعياري في قياس هذه الصعوبة، وكذلك قيم إحصائيات المطابقة الداخلية والخارجية لهذه المعالم.

الجدول رقم(٥)

المتوسط الحسابي والانحراف المعياري لكل من معالم الصعوبة للفقرات والخطأ المعياري في القياس وإحصائيات المطابقة الداخلية والخارجية لهذه المعالم

		الخطأ	إحصائي المطابقة الداخلية		إحصائي المطابقة الخارجية	
	الصعوبة	المعياري	متوسط المربعات MNSQ	قيمة الإحصائي ZSTD	متوسط المربعات MNSQ	قيمة الإحصائي ZSTD
المتوسط الحسابي	صفر	٠,١٦	١	٠,٢-	١,٠١	٠,٢-
الانحراف المعياري	٠,٩٢	٠,٠٢	٠,٠٩	١,٦	٠,١٣	١,٥

يتضح من الجدول (٥) اقتراب المتوسط لمتوسطات المربعات الموزونة من الواحد، وهو الوضع المثالي كما يتوقعه النموذج، كما يلاحظ أن متوسط قيم إحصائي المطابقة الداخلية قد بلغ (-٠,٢) والانحراف المعياري (١,٦) وهي بعيدة نوعاً ما عن القيم المثالية التي يفترضها النموذج وهي (صفر، ١) على الترتيب، وكذلك يلاحظ أن متوسط قيم إحصائي المطابقة الخارجية قد بلغ (-٠,٢) والانحراف المعياري لها (١,٥) وهي أيضاً بعيدة نوعاً ما عن القيم المثالية التي يفترضها النموذج وهي (صفر، ١) على الترتيب.

وفي ضوء محكات المطابقة للفقرات المستخدمة، تبين وجود ثلاث عشرة فقرة تعدت قيم متوسط المربعات الموزونة لها الواحد الصحيح، وهذا مؤشر إلى أنها فقرات غير مطابقة وبعيدة عن توقعات النموذج وأنها قيم مضطربة وتشوش البيانات (Linacre & Wright, 1993). بينما تبين وجود أربع فقرات فقط مرفوضة التمييز، حيث قيم معامل الارتباط الثنائي (r_{pbis}) لها كانت سالبة، ومعامل الارتباط الثنائي الخاص وفق برنامج BIGSTEPS يتم حسابه بين العلامات الملاحظة عن الفقرة أو الفرد والعلامات الكلية للفرد، بعد حذف الفقرة أو العلامات الكلية للفقرة بعد حذف الفرد وتحذف القيم المحسوبة المتطرفة للعلامات، والقيم السالبة لمعاملات تشير إلى مطابقة سيئة أو تقدير بالاتجاه المعاكس، والحروف الأبجدية الموجودة عند المعامل مؤشر على حسن المطابقة للفقرات وفق مؤشر التمييز.

والجدول (٦) يبين قيم إحصاءات المطابقة الداخلية والخارجية ومتوسط المربعات الموزونة وقيم معاملات التمييز (r_{pbis}).

الجدول رقم (٦)

أرقام الفقرات وقيم إحصاءات المطابقة الداخلية والخارجية لها، ومتوسطات المربعات لها وقيم معاملات التمييز (rpbis)

ITEMS STATISTICS: MISFIT ORDER

ENTRY NUMBER	RAW SCORE	COUNT	MEASURE	ERROR	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PTBIS CORR.	ITEMS
47	23	203	2.12	0.23	1.14	0.8	1.36	1.5	-0.05	47=Q47
5	70	203	0.59	0.16	1.27	4.2	1.36	4.4	-0.16	5 =Q5
39	53	203	1.02	0.17	1.21	2.5	1.33	2.9	-0.10	39=Q39
30	40	203	1.42	0.18	1.14	1.3	1.23	1.6	-0.01	30=Q30
49	54	203	0.99	0.17	1.14	1.6	1.23	2.1	E0.01	49=Q49
37	50	203	1.11	0.17	1.13	1.5	1.20	1.7	F0.02	37=Q37
2	30	203	1.79	0.20	1.06	0.5	1.15	0.8	G0.08	2 =Q2
29	64	203	0.73	0.16	1.02	0.3	1.10	1.2	J0.18	29=Q29
25	75	203	0.47	0.15	1.07	1.2	1.10	1.5	K0.13	25=Q25
50	77	203	0.42	0.15	1.08	1.5	1.09	1.4	M0.11	50=Q50
12	74	203	0.49	0.15	1.05	-0.2	1.00	0.0	W0.25	12=Q12
18	178	203	-2.23	0.22	1.00	0.0	1.01	-0.4	X0.15	18=Q18
19	148	203	-1.20	0.16	1.01	-0.2	1.13	-0.1	y0.22	19=Q19
44	74	203	0.49	0.15	0.99	1.4	0.98	1.3	N0.12	44=Q44
4	62	203	0.78	0.16	0.97	0.8	0.98	0.7	O0.16	4 =Q4
1	140	203	-1.00	0.16	0.95	-0.1	0.97	0.6	P0.20	1 =Q1
10	99	203	-0.07	0.15	0.95	1.2	0.96	0.8	Q0.17	10=Q10
15	85	203	0.24	0.15	0.96	0.9	0.96	0.6	R0.18	15=Q15
6	149	203	-1.23	0.16	0.98	0.3	0.99	-0.1	S0.18	6 =Q6
8	107	203	-0.24	0.15	0.97	0.5	0.98	0.2	T0.21	8 =Q8
22	91	203	0.11	0.15	0.96	0.2	0.96	0.3	U0.23	22=Q22
3	113	203	-0.37	0.15	0.95	0.4	0.96	0.2	V0.21	3 =Q3
46	76	203	0.44	0.15	0.98	1.8	0.97	1.6	H0.09	46=Q46
17	116	203	-0.44	0.15	0.97	1.9	0.95	1.7	L0.10	17=Q17
11	82	203	0.31	0.15	0.99	1.6	0.94	1.7	I0.12	11=Q11
35	95	203	0.02	0.15	0.98	-0.1	0.99	-0.2	Y0.25	35=Q35
9	188	203	-2.82	0.27	0.99	0.0	0.87	-0.5	x0.15	9 =Q9
48	62	203	0.78	0.16	0.99	-0.2	0.93	-0.8	w0.28	48=Q48
38	57	203	0.91	0.16	0.98	-0.2	0.97	-0.3	v0.26	38=Q38
7	103	203	-0.16	0.15	0.98	-0.5	0.97	-0.5	u0.27	7 =Q7
21	143	203	-1.07	0.16	0.98	-0.4	0.94	-0.7	t0.25	21=Q21
23	98	203	-0.05	0.15	0.96	-0.9	0.95	-1.0	s0.30	23=Q23
16	162	203	-1.61	0.18	0.96	-0.4	0.86	-1.0	r0.26	16=Q16
14	145	203	-1.12	0.16	0.95	-0.7	0.91	-0.9	q0.28	14=Q14
27	83	203	0.28	0.15	0.95	-1.0	0.93	-1.1	p0.31	27=Q27
36	99	203	-0.07	0.15	0.95	-1.2	0.94	-1.1	o0.32	36=Q36
43	117	203	-0.46	0.15	0.95	-1.2	0.93	-1.2	n0.31	43=Q43
41	112	203	-0.35	0.15	0.95	-1.3	0.93	-1.4	m0.32	41=Q41
42	102	203	-0.13	0.15	0.94	-1.3	0.93	-1.4	l0.33	42=Q42
24	92	203	0.08	0.15	0.91	-2.0	0.90	-2.0	k0.37	24=Q24
33	69	203	0.61	0.16	0.91	-1.5	0.90	-1.4	j0.37	33=Q33
13	142	203	-1.05	0.16	0.91	-1.4	0.86	-1.6	i0.36	13=Q13
32	96	203	0.00	0.15	0.89	-2.7	0.87	-2.6	h0.41	32=Q32
31	95	203	0.02	0.15	0.89	-2.7	0.86	-2.8	g0.42	31=Q31
45	128	203	-0.71	0.15	0.89	-2.3	0.86	-2.0	f0.40	45=Q45
20	65	203	0.71	0.16	0.87	-2.1	0.88	-1.5	e0.42	20=Q20
28	92	203	0.08	0.15	0.88	-2.7	0.86	-2.8	d0.42	28=Q28
26	102	203	-0.13	0.15	0.88	-3.1	0.87	-2.7	c0.43	26=Q26
34	107	203	-0.24	0.15	0.86	-3.4	0.87	-2.7	b0.44	34=Q34
40	107	203	-0.24	0.15	0.86	-3.6	0.84	-3.3	a0.45	40=Q40
MEAN	96.	203.	0.00	0.16	1.00	-0.2	1.01	-0.2		
S.D.	36.	0.	0.92	0.02	0.09	1.6	0.13	1.6		

الجدول (٧) يبين الفقرات ذات أكثر الاستجابات غير المطابقة من قبل الأفراد.

الجدول رقم (٧)

الفقرات ذات أكثر الاستجابات غير المطابقة من قبل الأفراد وأماكنها بملف البيانات

```
MOST MISFITTING RESPONSE STRINGS
  ITEM         OUTMNSQ  |PERSON
                        |11 11111 1 1    11   1  111     11111   1  11   11
                        |972918337352769 099962655016631 8665463345451 6174
                        |038897426935931737859048968864231710108350464514 78
            high--------------------------------------------------
    47 47=Q47     1.36 A|........1...111.1.111....11..111....1....1....1...
     5 5 =Q5      1.36 B|.....................................1....1.11.11
    39 39=Q39     1.33 C|.........................11.1.....1.1...11.....1
    30 30=Q30     1.23 D|..............1....11.1..1....111...1.........1..
    49 49=Q49     1.23 E|.......................1.1....1..1.1..11...1...1
    37 37=Q37     1.20 F|.........................1..1........1.1.1.....
     2 2 =Q2      1.15 G|...................1....11..1..111.11..........
    29 29=Q29     1.10 J|.....................................1.1..1...111
    25 25=Q25     1.10 K|.....................................1.......1..1
    50 50=Q50     1.09 M|.....................................11.11.1....
    12 12=Q12     1.00 W|.....................................1.1.11......
    18 18=Q18     0.93 X|........0..000000..........0......0...............
    19 19=Q19     0.99 y|...0..0.0...0.....................................
                        |-----------------------------------------------low-
                        |1121111171517697119961651116631311111 6331541156111
                        |978918336332931 098592485018642 8665408340451 1474
                        |03 89742 9 5    37   0  968     17101   5  64   78
```

يلاحظ من الجدول (٧) أن الفقرة (٤٧) والبالغة صعوبتها (٢,١٢) لوجيت وهي فقرة صعبة جدا، وقد تمكن عدد كبير من الأفراد من ذوي القدرات المتدنية أن يجيب عنها إجابة صحيحة، فمثلاً الفرد (٧٦) مستوى قدرته أدنى من صعوبة الفقرة ومع ذلك أجاب عنها إجابة صحيحة وكذلك الأفراد ذوو الأرقام (٧٩، ٩١,٦٣،). وكذلك فإن الفقرة (١٨) فقرة سهلة جداً إذ بلغت قيمة معامل الصعوبة لها (-٢,٢٣) و لم يتمكن بعض الأفراد ذوي القدرات العليا من الإجابة عنها إجابة صحيحة، فمثلاً لم يتمكن الأفراد ذوو الأرقام (٧٦، ١٢٥، ٧٩، ٦٣، ...) من الإجابة عنها إجابة صحيحة رغم أن قدراتهم كانت أعلى من معامل الصعوبة للفقرة .

للتأكد من جودة مطابقة البيانات الناتجة عن استجابات أفراد العينة على الاختبار، والتحقق من موضوعية تفسير نتائج الاختبار بصورته النهائية (٣٧) فقرة، أي بعد حذف الأفراد غير المطابقين والفقرات غير المطابقة والبالغة ثلاث عشرة فقرة لم تتطابق بياناتها مع توقعات النموذج، وكذلك التحقق من الافتراضات الخاصة بنموذج راش حيث يعد ذلك تحقيقا لمدى صدق النموذج في تحقيقه لموضوعية القياس، أعيد التحليل للمرة الثالثة للحصول على تقديرات نهائية متحررة لكل من صعوبة الفقرات وقدرات الأفراد ويلخص الجدول (٨) نتائج التحليل للقيم المتحررة لقدرة الأفراد .

الجدول رقم (٨)

نتائج التحليل للقيم المتحررة لقدرة الأفراد

	القدرة	الخطأ المعياري	المدى	الجذر التربيعي لمتوسطات مربعات الأخطاء المعيارية		الانحراف المعياري المعدل من الأخطاء		مؤشر الفصل (مؤشر التباعد)		معامل الثبات لتباعد الأفراد	
				الفعلي	المتوقع	الفعلي	المتوقع	الفعلي	المتوقع	الفعلي	المتوقع
المتوسط الحسابي	٠,١٠	٠,٣٧	-٢,٠٦ ٢,٦٢	٠,٣٨	٠,٣٧	٠,٧٦	٠,٧٦	٣,٠١	٣,٠٧	٠,٩٠	٠,٩٠
الانحراف المعياري	٠,٨٥	٠,٠٤									
أقصى علامة خام	٣٤										
أدنى علامة خام	٥										

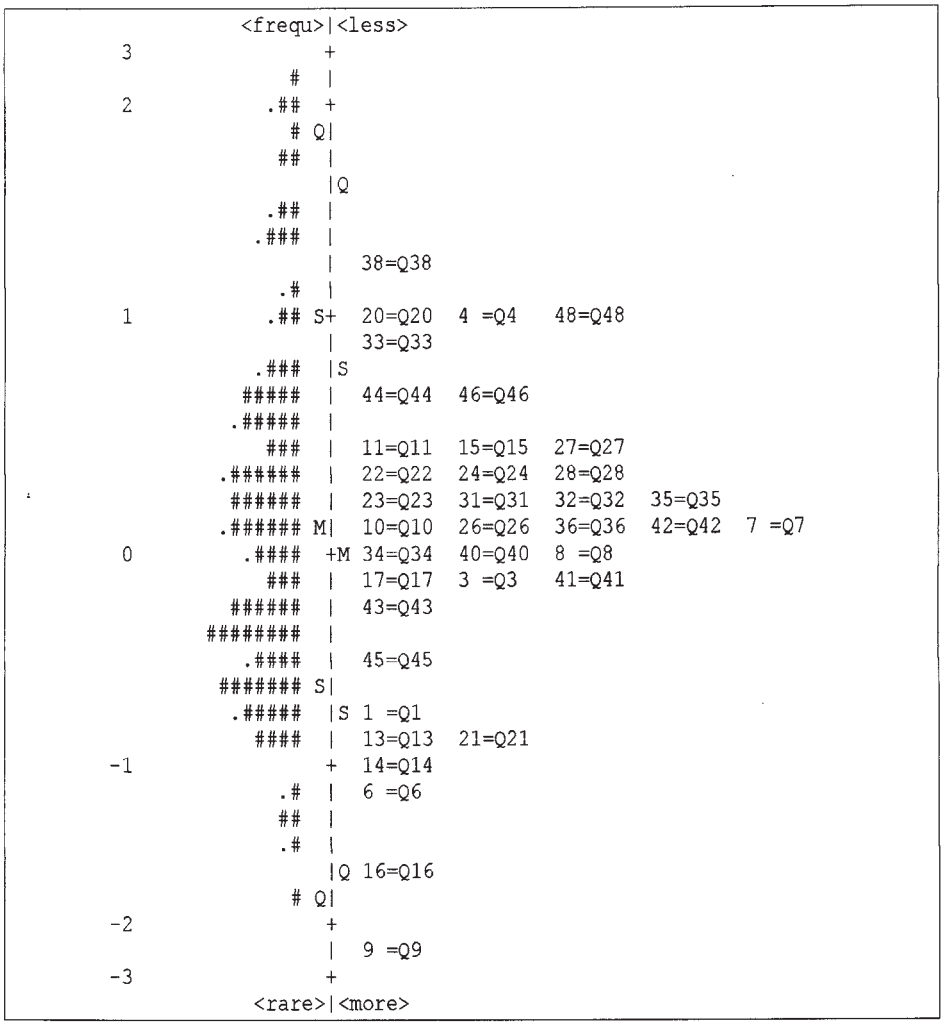
ويلاحظ من الجدول (٨) أن التقديرات النهائية المتحررة لقدرات الأفراد قد تراوحت بين العلامة الكلية (٥) كحد أدنى والعلامة الكلية (٣٤) كحد أقصى. وقد بلغ متوسط توزيع القدرة (٠,١٠) وحدة لوجيت، والانحراف المعياري (٠,٨٥) وحدة لوجيت أي بمدى يتراوح بين (-٢,٠٦) لوجيت للمستوى المنخفض من القدرة و (٢,٦٢) لوجيت للمستوى المرتفع من القدرة؛ وبلغ متوسط الأخطاء المعيارية لتقديرات قدرة الأفراد (٠,٣٧) وهي قيمة متدنية، الأمر الذي يشير إلى دقة تحديد مواقع الأفراد على متصل السمة. وأما الجدول (٩) فيلخص نتائج التحليل للقيم المتحررة لصعوبة الفقرات.

الجدول رقم (٩)

نتائج التحليل المتحررة لصعوبة الفقرات

	القدرة	الخطأ المعياري	المدى	الجذر ألتربيعي لمتوسطات مربعات الأخطاء المعيارية		الانحراف المعياري المعدل من الأخطاء		مؤشر الفصل (مؤشر التباعد)		معامل الثبات لتباعد الأفراد	
				الفعلي	المتوقع	الفعلي	المتوقع	الفعلي	المتوقع	الفعلي	المتوقع
المتوسط الحسابي	صفر	٠,١٦	-٢,٧ ١,١٧	٠,١٦	٠,١٦	٠,٧٥	٠,٧٥	٤,٥٧	٤,٦٦	٠,٩٥	٠,٩٦
الانحراف المعياري	٠,٨٧	٠,٠٢									

ويلاحظ من الجدول (٩) أن قيم التقديرات المتحررة لصعوبة الفقرات قد تراوحت ما بين −٢,٧ إلى ١,١٧ وحدة لوجيت، وتتوزع بمتوسط قدره صفر لوجيت، وانحراف معياري مقداره (٠,٨٧) لوجيت، وقد بلغ متوسط الأخطاء المعيارية للفقرات (٠,١٦) وهي قيمة متدنية، الأمر الذي يشير إلى دقة تقديرات الصعوبة للفقرات،ولدى مقارنة محوري تدريج الفقرات والقدرة −كما في الشكل (١)− يلاحظ تطابقهما إلى حد كبير، بمعنى أن هناك نوعاً من الاتساق في تدرج صعوبة الفقرات وأن الاختبار يقيس مدىً عريضاً من القدرة، ويقيس متغيراً أحادي البعد، الأمر الذي يعكس صدق البناء للاختبار.



شكل رقم (١)

خريطة لتدريج الفقرات والأفراد

للتأكد من جودة مطابقة البيانات الناتجة عن استجابات أفراد العينة على الاختبار الثنائي التدريج، والتحقق من موضوعية تفسير النتائج للاختبار بصورته النهائية (٣٧) فقرة، أي بعد حذف الأفراد غير المطابقين والفقرات غير المطابقة، وكذلك التحقق من الافتراضات الخاصة بنموذج راش، حيث يعدّ ذلك تحقيقاً لمدى صدق النموذج في تحقيقه لموضوعية القياس،ولقد أشارت النتائج إلى تحقق شروط الموضوعية في الاختبار ثنائي التدريج، والتي يمكن تلخيصها بما يلي:

- إن متوسط تقدير قدرات الأفراد قد ازداد بمقدار (٠,٠٨) لوجيت بعد حذف الأفراد غير المطابقين، وبمقدار (٠,٢٧) بعد حذف الأفراد غير المطابقين والفقرات غير المطابقة للنموذج معاً. وقد بلغ متوسط قدرات الأفراد (٠,١٠) لوجيت، أي أن مستوى الاختبار يصبح أكثر ملاءمة لقدرات الأفراد، نظراً لاقتراب قيمة متوسط القدرة للأفراد من قيمة متوسط الصعوبة للفقرات، والبالغ صفرا لوجيت. وهو يتسق مع ما هو متوقع من تزايد احتمال الإجابة الصحيحة عن الفقرات، بتزايد مستوى قدرات الأفراد، وهذا يعني اتساق معطيات الفقرات مع توقعات النموذج، أي حسن مطابقة الفقرات للنموذج، وهذا ما يتبين من إحصاءات المطابقة المحسوبة حيث اقتربت من الوضع المثالي، وهو أن يكون متوسط المربعات للقدرات يساوي واحدا، والانحراف المعياري لهذه الأخطاء يساوي صفرا
- كما أشارت النتائج إلى زيادة بلغت (٠,٠٦) لوجيت في متوسط الخطأ المعياري في القياس لهذه القدرات، وربما يعود السبب في ذلك إلى نقصان عدد أفراد العينة، وكذلك نقصان عدد فقرات الاختبار بمقدار (٢٦٪) ، لتصبح (٣٧) فقرة بدلاً من (٥٠) فقرة. وقد بلغ متوسط الأخطاء المعيارية لتقديرات قدرة الأفراد (٠,٣٧) وهي قيمة متدنية، الأمر الذي يشير إلى دقة تحديد مواقع الأفراد على متصل السمة. حيث يكون تقدير قدرات الأفراد أكثر دقة كلما كانت قيمة الخطأ المعياري في قياسها أقل، ويعدّ ذلك مؤشراً إلى نقصان احتمال الوصول إلى الإجابة الصحيحة عن طريق التخمين، ليتحقق افتراض مهم من افتراضات النموذج، وهو اقتراب فرصة التخمين من الصفر.
- بيّنت النتائج أن متوسط قيم إحصائي المطابقة الخارجية (التباعدي) أو ما يسمى إحصائي المطابقة الكلية للفقرات قد بلغ (– ٠,١) والانحراف المعياري لها (١,١) وهي أيضاً تقترب من القيم المثالية التي يفترضها النموذج وهي (صفر، ١) على الترتيب. وهذا المؤشر يختص باختبار مدى مطابقة الفقرة للنموذج، وذلك بوجه عام من فرد لآخر. فإذا ما حدث اتساق بين الاستجابات الملاحظة للأفراد على الفقرة، واحتمال نجاحهم عليها، كان معنى ذلك أن هناك اتساقاً بين الاستجابات الملاحظة للأفراد عن هذه الفقرة وعلاماتهم الكلية على الاختبار؛ أي استجاباتهم عن باقي فقرات الاختبار، وهذا يدل على الاتفاق بين السمة التي تعبر عنها هذه الفقرة والسمة التي تعبر عنها باقي الفقرات، وذلك عبر العينة كلها. ومعنى هذا مطابقة الفقرة بوجه عام لمتطلبات النموذج واتفاقها في تعريف السمة مع

ذلك الذي تعرفه باقي الفقرات، أي تحقق افتراض أحادية البعد وهو أحد الافتراضات الرئيسة في نموذج راش.

● كما أشارت النتائج إلى أن قيمة إحصائي المطابقة الخارجية (التباعدي) للأفراد أو ما يسمى إحصائي المطابقة الكلية قد بلغ (−٠,١) والانحراف المعياري له (٠,٩) لوجيت. وهي تقترب جداً من القيم المثالية التي يفترضها النموذج (صفر، ١) على الترتيب. وهذا يشير إلى عدم اختلاف الصعوبة النسبية للفقرات عند معظم الأفراد عبر مستويات القدرة. وهذا يعني تحرر تقديرات كل من صعوبة الفقرة وقدرة الفرد من قدرة الأفراد الذين يؤدون الاختبار، أي تحرر القياس من توزيع أداء العينة وهو أحد الافتراضات المهمة في نموذج راش.

● وأشارت النتائج إلى أن متوسط مربعات المطابقة الداخلية بين الفقرات وهو صورة لإحصائي المطابقة بين المجموعات ، قد بلغت قيمته (١) بانحراف معياري (٠,٠٨) وهي قيم قريبة جداً من الوضع المثالي (١، صفر) على الترتيب، وهذا الإحصائي يشير إلى مدى استقرار مستوى الصعوبة النسبي للفقرات، عبر مستويات القدرة المختلفة، أي اختبار مدى ثبات تدرج صعوبة الفقرات عند كل مستوى من مستويات القدرة، ويستخدم هذا الإحصائي للمطابقة بين المجموعات، وذلك لفحص مدى الاتفاق بين المنحنى الملاحظ لخصائص الفقرة، الذي يوضح نسبة الاستجابات الصحيحة الملاحظة عن الفقرة للأفراد عبر مستويات القدرة المختلفة، والمنحنى المتوقع من النموذج والذي يوفر احتمالات الإجابة الصحيحة عن هذه الفقرات عند المستويات المختلفة للقدرة (Mead & Bell, 1980 ,Wright). وهذا يدل على اقتراب المنحنى الملاحظ من المنحنى المتوقع (أفضل منحنى له يطابق النموذج)، وعندئذ يتوافر لهذا المنحنى الملاحظ ما يتوافر للمنحنى المتوقع من النموذج من استقلال لصعوبة الفقرة عن العينة ومن ثم استقرار لهذه الصعوبة عبر مستويات القدرة المختلفة، أي يتوافر افتراض مهم من افتراضات نموذج راش وهو استقلال صعوبة الفقرات عن عينة الأفراد. وعندئذ تكون الفقرات مطابقة للنموذج.

● وأشارت النتائج إلى أن القيم التقديرية المتحررة لصعوبة الفقرات قد تراوحت ما بين (−٢,٧ إلى ١,١٧) وحدة لوجيت، وتتوزع بمتوسط قدره صفر لوجيت، وانحراف معياري مقداره (٠,٨٧) لوجيت، وقد بلغ متوسط الأخطاء المعيارية لصعوبة الفقرات (٠,١٦) وهي قيمة متدنية، الأمر الذي يشير إلى دقة تقديرات صعوبة الفقرات. وعند مقارنة محوري تدريج صعوبة الفقرات وقدرة الأفراد كما هو واضح بما يسمى خريطة الفقرات− الأفراد، أشارت النتائج إلى أن المدى الذي تتشتت فيه صعوبات الفقرات يتراوح ما بين (−٢,٧) لوجيت و (١,١٧) لوجيت وبخطأ معياري تراوح بين (٠,١٥ و ٠,٢٧) على الترتيب ويبدو من الشكل (١) انتظام تدرج هذه الفقرات بوحدة تدريج مقدارها (٠,٢) لوجيت على متصل السمة، وهذا يعني تعريف الفقرات للسمة موضوع القياس

على مدى هذا المتصل، وعلى الرغم من انتظام التدرج ضمن هذا المدى، فقد تبدو بعض الفراغات الضيقة عند بعض المستويات الأخرى للسمة. وبذلك يتحقق افتراض من افتراضات نموذج راش –مرة أخرى– وهو أن تعرف فقرات الاختبار سمة واحدة وهي في حالتنا هذه "تحصيل طلبة الجامعة في القياس والتقويم" ويعدّ تحقيق هذه الفرضية تحقيقاً لصدق الاختبار.

ويختص إحصائي المطابقة بين المجموعات أيضاً باختبار ما إذا كانت منحنيات خصائص الفقرات المطابقة للنموذج ذات ميل أو انحناء متشابه، فعندما تكون الفقرات مستقلة فعلاً عن العينة، فإن قدرة الفقرات على التمييز تكون متساوية، ومن ثم تكون المنحنيات لخصائص الفقرات متوازية، أي ذات انحناء متماثل. وعندئذ يكون متوسط التوزيع الملاحظ لقيم إحصائي المطابقة بين المجموعات (إحصائي المطابقة الداخلية) يقترب من الصفر وانحرافه المعياري يقترب من(واحد) وذلك للفقرات المطابقة، وقد وجد أن قيمة متوسط إحصائي المطابقة الداخلية للفقرات صفر، والانحراف المعياري لها يساوي (٠,٨٧)، وهي قيم قريبة من الوضع المثالي، وبالتالي فإن ذلك مؤشر إلى توازي المنحنيات لخصائص الفقرات، وهذا يعني أن الزيادة المتساوية في مستوى القدرة، يقترن بزيادة متساوية تقريباً في احتمال الإجابة الصحيحة عن أي من هذه الفقرات، وبعبارة أخرى إن القدرة التمييزية للفقرات بين مستويات القدرة للأفراد (ومعنى ذلك استقرار لصعوبات الفقرات عبر مستويات القدرة المختلفة). وهذا يؤكد افتراض استقلال صعوبات الفقرات عند قدرات العينة مما يعني مطابقة الفقرات للنموذج.

وما يؤكد هذه النتيجة التقارب الواضح في قيم معاملات الارتباط الثنائية (r_{pbis}) ،والتي سبق الإشارة إليها، على أنها تعكس تقارباً أيضاً في معاملات التمييز للفقرات، وبالتالي هذا مؤشر إلى تحقق افتراض مهم من افتراضات النموذج، وهو أن تكون للفقرات قدرة تمييزية متكافئة تقريباً. حيث أشار هامبلتون وسوامنيثان (Swaminathan, 1985 & Hambelton) إلى أنه لكي يتحقق افتراض تكافؤ مؤشرات التمييز ومطابقتها، بالتالي للنموذج يجب أن تكون قيمها واقعة ضمن حدود المدى (متوسط معاملات التمييز ± ٠,١٥) وحيث إن متوسط معاملات التمييز(٠,٥٤) فإن حدود المطابقة للفقرات المطابقة تتراوح ما بين(٠,٣٩ و ٠,٦٩) وبالتالي فإن قيم معاملات الارتباط الثنائية جميعها كانت ضمن هذا المدى، وكذلك كانت قيمة الانحراف المعياري لهذه المعاملات صغيرة، وهو مؤشر على تحقق هذا الافتراض. وهذا ما أكده البرنامج ا لذي أشار إلى تطابق الفقرات جميعها من حيث القدرة التمييزية حيث إن معامل التمييز المطابق يظهر بجانبه حرف وجميعها كانت هناك حروف بجانبها، وكل هذه المؤشرات تدل على تكافؤ معاملات التمييز للفقرات وبالتالي تحقق افتراضاً مهماً وهو تكافؤ القدرة التمييزية للفقرات.

أما فيما يتعلق بافتراض انخفاض مؤشر التخمين فلا توجد طريقة مباشرة لتحديد ما إذا

كان الفرد أجاب عن الفقرة إجابة صحيحة بالتخمين أم لا. ولكن تؤخذ بعين الاعتبار حالتان أولاً: رسم خطوط الانحدار غير الخطي لعلامات فقرات الاختبار. وثانياً: الأداء المنخفض على فقرات الاختبار الصعبة. وعند رسم خطوط الانحدار غير الخطي لوحظ أن الخط التقاربي الأسفل يقترب من الصفر في الفقرات جميعها، وهذا يعني أن احتمال الإجابة الصحيحة للفقرة يقترب كثيراً من الصفر عندما تكون قدرة الفرد منخفضة، وهذا يحدث عندما لا يلجأ الفرد إلى التخمين، وهذا مؤشر على حسن مطابقة الفقرات للنموذج. والمؤشر الآخر هو أن أداء الأفراد ذوي القدرات المتدنية على الفقرات الصعبة كان ضعيفاً جداً. ويدل هذان المؤشران على مدى تطابق فقرات الاختبار مع نموذج راش.

والنظرية الحديثة في القياس النفسي والتربوي. تفترض أن عامل السرعة لا يلعب دوراً في الإجابة عن فقرات الاختبار، بمعنى أن إخفاق الفرد في الإجابة عن فقرة يرجع إلى انخفاض قدرته، وليس إلى عامل السرعة على إجابته. ولكن إذا كان عدد الأفراد الذين لم يتمكنوا من الإجابة عن فقرات الاختبار جميعها كبيراً، عندها يمكن القول إن عامل السرعة قد قام بدور في الإجابة، وهذا ما لم يتحقق في هذا الاختبار، حيث إن عددا محدودا جداً من الطلبة لم يجيبوا عن الأسئلة الأخيرة أو تركوا الإجابة عنها كما تبين عندما حذف البرنامج فردا واحدا فقط في بداية التحليل: لذا لا يمكن القول إن عامل السرعة قد قد قام بدور في الإجابة.

ثانياً: عرض النتائج المتعلقة بالخصائص السيكومترية للاختبار (الصدق والثبات) ومناقشتها:

بعد التحقق من جودة مطابقة فقرات الاختبار لنموذج راش، واستبعاد الفقرات التي لم تحقق هذه المطابقة، وتحديد القيم التقديرية لكل من صعوبة الفقرات وقدرات الأفراد، ينبغي الحصول على نوعين من المعاملات، يطلق عليهما الثبات المتعلق بالأفراد ، والثبات المتعلق بفقرات الاختبار. فمفهوم الثبات في إطار نموذج راش، يشير إلى مدى الدقة في تقدير موقع كل من الفقرات والأفراد على متصل السمة، الذي نهدف إلى قياسه. ويمكن تحديد مدى دقة الفقرات في تعريف هذا المتصل، بإيجاد النسبة بين الانحراف المعياري للقيم التقديرية المتحررة لصعوبة الفقرات، ومتوسط الخطأ المعياري لهذه القيم. ويطلق على هذه النسبة معامل الفصل بين الفقرات (G_i) "Item Separation Index" الذي يجب أن تكون قيمته أكبر ما يمكن. فإذا لم تزد هذه القيمة على (٢) يصعب قياس المتغير بوساطة هذه الفقرات. وقد وجد الباحث أن قيمة معامل الفصل بين الفقرات في الاختبار بصورته النهائية (٣,٧٧). ونظراً لأن هذه القيمة تزيد عن (٢) فإن فقرات الاختبار تعد كافية لتعريف متصل السمة الذي تقيسه. وكذلك وجد أن قيمة معامل الفصل بين الأفراد (G_p) (Person Separation Index)(3.01)، وهي ترمز إلى النسبة بين الانحرافات المعيارية

للقيم التقديرية المتحررة لقدرات الأفراد، ومتوسط الخطأ المعياري لهذه القيم وهذه القيمة أيضاً تزيد على (٢)، لذا فإن عينة الأفراد تعد أيضاً كافية في الفصل بين الفقرات، وكذلك يمكن الحصول على معامل الثبات لكل من الفقرات والأفراد باستخدام الصيغة الرياضية التالية: ($R=G_p^2/ 1+G_p^2$) حيث (G) ترمز إلى معامل الفصل (Masters, 1981 & Wright).

وقد وجد الباحث أن قيمة معامل الثبات المتعلق بكل من الأفراد والفقرات (الاختبار) باستخدام الصيغة الرياضية السابقة (٠,٩٠، ٠,٩٥) وعلى الترتيب، وهما قيمتان مرتفعتان، وتدل الأولى على كفاية عينة الأفراد في الفصل بين الفقرات، وبالتالي في تعريف متصل السمة الذي تقيسه هذه الفقرات، وتدل الثانية على كفاية عينة الفقرات في الفصل بين الأفراد، وبالتالي في التمييز بين مستويات التحصيل لدى هؤلاء الأفراد.

ويلاحظ أن قيمة معامل الثبات المتعلق بالأفراد كانت أقل، ولكنها قيمة عالية، وكانت قادرة على إظهار ما مقداره خمس طبقات إحصائية متمايزة تقريبا للأفراد، أي خمسة مستويات للتحصيل لدى هؤلاء الأفراد وفق المعادلة ($H_p=(4G_p + 1) / 3$) حيث : H_p عدد الطبقات الإحصائية للأفراد ، :G_p معامل الفصل بين الأفراد في حين أن معامل الثبات المتعلق بالفقرات كانت قيمته كبيرة جداً، ودلت على أن الفقرات كانت قادرة على إظهار ست طبقات إحصائية متمايزة تقريبا لصعوبة الفقرات لدى أفراد عينة الدراسة، أي إظهار الفروق الفردية بين الأفراد بشكل كبير.

ومما يجدر ملاحظته أن قيم معامل الثبات هذه تقابل معامل الثبات بطريقة (كودر– ريتشاردسون– ٢٠) (KR-20) في النظرية الكلاسيكية، وبخاصة أن قيمة معامل كودر ريتشاردسون تمثل الحد الأدنى لمعامل الثبات. (Julian, 1988).

أما فيما يتعلق بصدق الاختبار (مجتمع الفقرات) فقد تم التحقق منه بطريقتين:

الطريقة الأولى: اعتمدت على الإجراءات التي اتبعت في تحديد مجال التحصيل في موضوعات القياس والتقويم، وتحليل محتوى كتابة الأهداف السلوكية، وكتابة فقرات الاختبار، والتي تمثل الأهداف السلوكية كنتاج تعلمي يراد تحقيقه، والوصول به إلى مستوى الإتقان، والتأكد من صحة هذه الإجراءات، من خلال الأحكام والتقديرات والتوافق الذي برز من خلال المحكمين من المختصين في تدريس القياس والتقويم في الجامعات عن مدى ملاءمة نتائج التحليل لمواصفات مجال التحصيل، والتوافق بين مضمون الفقرات، والأهداف التي تحققها تلك الفقرات .

الطريقة الثانية: الصدق المعياري: أشار لن وجرنلند (Linn & Gronlund, 2000) أن تحديد معيار مناسب يعد أهم وأصعب خطوة في فحص الصدق المعياري، إذ لم يسبق لأفراد عينة الدراسة أن تقدموا لاختبار في موضوعات القياس والتقويم، وبالتالي فإن المهمة في تحديد المعيار تبدو أصعب. لذا استخدم الباحث التحصيل الأكاديمي الجامعي مقدرا بالمعدل

التراكمي في الجامعة كمعيار خارجي . وبلا شك فهذا الأمر ليس دقيقا تماما، لأن المعدل التراكمي يشتمل على مواد ليست ذات علاقة بالتحصيل في القياس والتقويم، وتم حساب معامل الارتباط بين نتائج الطلبة على اختبار القياس والتقويم والمعدل التراكمي، فكانت قيمة (٠,٥١) فقط وهي قيمة متدنية وتفسر ما نسبته (٠,٢٥) من تحصيل الطلبة في القياس والتقويم، من خلال العلاقة الخطية بالمعدل التراكمي، وهذا يدل على ضعف هذه العلاقة، وتدل على عدم وجود معيار خارجي مناسب، لمقارنة التحصيل في اختبار القياس والتقويم.

ويمكن الاحتفاظ بنتائج الاختبار إلى أن يصبح هؤلاء معلمين في الميدان، ويتم تقدير درجة ممارستهم تطبيق مبادئ القياس والتقويم، وبناء الاختبارات وعندئذ يتم حساب معامل لارتباط بين الاختبار والمحك.

وبعد أن تم بناء الاختبار عبر هذه المراحل من استبعاد الفقرات غير المطابقة لنموذج راش، واستبعاد الأفراد غير المطابقين. وبعد تحديد الخصائص السيكومترية للاختبار وفقراته، فإن هذا الاختبار المؤلف من (٣٧) فقرة يمثل المجال بشكل متوازن مع الأهداف، وبالتالي يمكن تكوين اختبارات فرعية من هذا الاختبار؛ وتحديد علامة المجال من هذه الاختبارات. كما يمكن تشكيل اختبارات فرعية من هذا الاختبار، حسب درجة الصعوبة أو اختيار فقرات ذات تمييز مرتفع تصف الأفراد عند نقطة قطع محددة لأغراض محددة .

ثالثاً: عرض النتائج المتعلقة بدالة المعلومات للاختبار ومناقشتها:

لإيجاد دالة المعلومات للاختبار الكلي، فقد استخدم الباحث البرنامج الإحصائي (MINSTEPS) وكون هذا البرنامج لا يحلل استجابات أفراد يزيد عددهم على ٢٠٠ فرد وفقرات لا يزيد عددها على ٣٠ فقرة، لذا قام الباحث باختيار ٣٠ فقرة من فقرات الاختبار تتوزع على متصل السمة، وذلك بالاستعانة بخريطة الفقرات بحيث تغطي المدى الذي تتوزع عليه الفقرات حسب درجة صعوبتها والممتد من (−٢,٧ −١,١٧) لوجيت. أما الأفراد فكان عددهم مناسبا ثم أدخلت البيانات في ذاكرة الحاسوب، وتم تقدير دالة المعلومات للاختبار عند كل مستوى من مستويات القدرة المحددة، والجدول (١٠) يبين تقديرات الأفراد بوحدة اللوجيت وقيم دالة المعلومات عند مستويات مختارة للقدرة.

ويلاحظ من الجدول (١٠) أن قيم دالة المعلومات التي يقدمها الاختبار ثنائي التدريج، تكون أكبر ما يمكن عند مستوى القدرة صفر لوجيت، بمعنى أن الاختبار يعطي معلومات أكثر فاعلية عند الأفراد ذوي القدرة المتوسطة، بينما تكون قيم دالة المعلومات التي يقدمها الاختبار أقل ما يمكن عن مستويات القدرة العالية والمتدنية، بمعنى أن الاختبار يعطي معلومات قليلة عند الأفراد ذوي القدرات العالية والقدرات المتدنية.

الجدول رقم (١٠)

تقديرات القدرة لأفراد العينة وقيم دالة المعلومات التي يقدمها الاختبار عند مستويات مختارة من القدرة للاختبار ثنائي التدريج

القدرة	دالة المعلومات	القدرة	دالة المعلومات	القدرة	دالة المعلومات
-٦	٠,٠٧	-١,٤٤	٢,٨١	٣,١٢	١,٠٤
-٥,٧٦	٠,٠٩	-١,٢	٣,٠٨	٣,٣٦	٠,٨٥
-٥,٥٢	٠,١١	-٠,٩٦	٣,٣٣	٣,٦	٠,٦٩
-٥,٢٨	٠,١٤	-٠,٧٢	٣,٥٤	٣,٨٤	٠,٥٧
-٥,٠٤	٠,١٨	-٠,٤٨	٣,٧٢	٤,٠٨	٠,٤٦
-٤,٨	٠,٢٣	-٠,٢٤	٣,٨٤	٤,٣٢	٠,٣٦
-٤,٥٦	٠,٣٠	٠	٣,٩٢	٤,٥٦	٠,٢٩
-٤,٣٢	٠,٣٧	٠,٢٤	٣,٩١	٤,٨	٠,٢٤
-٤,٠٨	٠,٤٥	٠,٤٨	٣,٨٥	٥,٠٤	٠,١٩
-٣,٨٤	٠,٥٨	٠,٧٢	٣,٧٠	٥,٢٨	٠,١٥
-٣,٦	٠,٧٠	٠,٩٦	٣,٥١	٥,٥٢	٠,١٢
-٣,٣٦	٠,٨٦	١,٢	٣,٢٩	٥,٧٦	٠,١٠
-٣,١٢	١,٠٤	١,٤٤	٣,٠١	٦	٠,٠٨
-٢,٨٨	١,٢٥	١,٦٨	٢,٧٦		
-٢,٦٤	١,٤٨	١,٩٢	٢,٤١		
-٢,٤	١,٧٢	٢,١٦	٢,٠٥		
-٢,١٦	٢,٠٠	٢,٤	١,٧٥		
-١,٩٢	٢,٢٦	٢,٦	١,٤٨		
-١,٦٨	٢,٥٤	٢,٨٨	١,٢٥		

وأشارت النتائج إلى أن قيمة دالة المعلومات للاختبار تتزايد تدريجيا بتزايد القدرة (θ) حتى تصل إلى أقصى قيمة ممكنة لها عندما تكون قيمة القدرة صفرا لوجيت، وهي تقابل متوسط الصعوبة للفقرات صفر لوجيت، وهذا ما يتطابق مع ما هو متوقع من نموذج راش، حيث إن دالة المعلومات تكون أقصى ما يمكن عندما تكون قيمة (θ=b) وذلك لكل من النموذجين الأحادي المعلم والثنائي المعلم. وهذا ما تؤكده قيمة الخطا المعياري للتقدير التي كانت أقل ما يمكن عند متوسط الصعوبة لفقرات الاختبار، إذ بلغت قيمة (٠,١٦) .

إذ يكون الخطأ المعياري صغيرا كلما اقترب مستوى القدرة للأفراد من مستوى الصعوبة للفقرات. وبالتالي تزداد كمية المعلومات بنقصان الخطأ المعياري حسب العلاقة

$$I(\theta) = \frac{1}{\sqrt{S.E(\theta)}}$$

حيث إن(θ)I: كمية المعلومات عند مستوى القدرة (θ)و (θ)S.E: الخطأ المعياري في تقدير القدرة عن مستوى القدرة (θ)، بالإضافة إلى ذلك فإن ميل المماس عند نقطة الانعطاف لمنحنى خصائص الاختبار المتعدد التدريج كان أكبر من ميل المماس للاختبار الثنائي التدريج حيث أن كمية المعلومات تتناسب طردياً مع ميل المماس لمنحنى خصائص الاختبار إذ كلما زاد الميل زادت كمية المعلومات وذلك حسب العلاقة التالية:

$$I(\theta) = \frac{[P^1(\theta)]^2}{[P(\theta)][Q(\theta)]}$$

حيث I(θ) :كمية المعلومات التي نحصل عليها عند مستوى القدرة (θ).

$[P^1(\theta)]^2$: مربع المشتقة الأولى لدالة الاستجابة عن الاختبار (لمنحنى خصائص الاختبار) عند مستوى القدرة (θ).

$[P(\theta)]$: نسبة الناجحين في الاختبار.

$[Q(\theta)]$: ١-p(θ).

خاتمة

هدفت هذه الدراسة إلى بناء اختبار تحصيلي مقنن في القياس والتقويم التربوي، على المستوى الجامعي، اعتمد في بنائه على الأهداف السلوكية التي تمثل الكفايات التعليمية، التي ترغب في تحقيقها لدى الطلبة (معلمي المستقبل) واستخدم الباحث النظرية الحديثة في القياس كإطار لتحليل فقرات الاختبار، واستجابات الطلبة وتحقيق افتراضات النموذج المستخدم (نموذج راش) كأحد نماذج النظرية الحديثة، وبناء أداة قياس موضوعية متحررة من قدرات الأفراد، ومتحررة من صعوبة الفقرات، وضمان الحصول على فقرات ذات خصائص سيكومترية جيدة. أي أننا نقترب باحتمال كبير من العدالة والموضوعية والدقة، مما يمكننا من تشخيص جوانب الضعف، وبيان المتقنين من غير المتقنين. مما يجعل بناء هذا الاختبار يشكل نواة لبنك أسئلة لمساق القياس والتقويم التربوي في الجامعة، نستطيع من خلاله أن نختار أي عدد من الفقرات بطريقة عشوائية، لتحقيق أهداف عملية القياس الموضوعي. كما هدفت الدراسة إلى توضيح المنهجية التي تم من خلالها بناء وتطوير اختبار في القياس والتقويم التربوي، مما يجعل المجال ممهدا لامكانية تطبيق هذه المنهجية في بناء اختبارات ومقاييس في كفايات تعليمية وتحصيليه أخرى.

المراجع

الجامعة العربية المفتوحة (٢٠٠٥). **القياس والتقويم في وبناء الاختبارات المدرسية** (ط١)، الصفاة، الكويت: الجامعة العربية المفتوحة.

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

علاّم، صلاح الدين (١٩٨٥). تحليل بيانات الاختبارات العقلية باستخدام نموذج راش اللوغاريتمي الاحتمالي: دراسة تجريبية. **المجلة العربية للعلوم الإنسانية**. الكويت، **١٩**(٥) ٦٧-٩٣.

علاّم، صلاح الدين. (٢٠٠١). **الاختبارات التشخيصية مرجعية المحك في المجالات التربوية والنفسية والتدريبية**. القاهرة، دار الفكر العربي.

عودة، أحمد (١٩٩٢). مدى التوافق بين نموذج راش والمؤشرات التقليدية في اختيار فقرات مقياس اتجاه سباعي التدريج. **مجلة كلية التربية**، جامعة الإمارات، (٨)، ١٥٣-١٧٩

الكيلاني، عبد الله زيد، والشريفين، نضال كمال (٢٠٠٥). **مدخل إلى البحث في العلوم التربوية والاجتماعية، أساسياته- مناهجه- تصاميمه- أساليبه الإحصائية** ، (ط١)، عمان: دار المسيرة للنشر والتوزيع والطباعة

Alastair, P., & Hutchinson, G. (1987). Calibrating graded assessment: rasch partial credit analysis of performance in writing. **Language Testing**, 4, 73-92.

Allen, M. J., & Yen, W. M. (1979**). Introduction to measurement theory**. California: Cole Publishing Company applications.

Anastasi , A . (1982) , **Psychological testing** (5th ed). New York: McMillan Publishing Co .

Fan, X. (1998). Item response theory and classical theory: An empirical comparison of their Item/ Person Statistics. **Educational and Psychological Measurement**, **58** (3), 357-381.

Guskey, T. (1988). **Implementing Mastery Learning**. California :Wadsworth Publishing , Co .

Hambleton, K & Swaminathan, H. & rogers , H. (1999**). Fundamentals of item response theory**. Newbury park, CA: Sage Publication.

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

Hambleton, R. K., & Jonse, R. W. (1993). Comparison of classical test theory and Item response theory and their applications to test development. **Educational Measurement, Issues and Practice**, **4**, 38-47.

Julian, E. R., (1988). Using computerized patient simulations to measure the clinical competence of physicians**. Applied Measurement in Education**, **4** (1), 299-318.

Linacre, J. M & Wright, B. D. (1993). **Auser's guide to BIGSTEPS**. (Computer Program). Chicago: MESA Press.

Linn , R . & Gronlund, N. (2000)**. Measurement and assessment in teaching**. columbus, OH: Merrill.

Masters, G. N. (1982). A Rasch model for partial credit scoring. **Psychometrika**, **2**, 149-174.

Poham, W. J. (1980). **Modern educational measurement**. N. J.: Prentice – Hall , Inc.

Popham, w. J. (1978). **Criterion– Referenced measurement**. Englewood Cliffs, NJ : Prentice – Hall, Inc.

Stocking. M. (1999). **Item response theory**. In: G . Masters, & J . Keevs (Eds). Advances in Measurement in Educational Research and Assessment .(pp.322) The Boulevard, UK : Pergamon.

Wright, B, Mead, R & Bell, S. (1980). **Bical calibrating items with rasch model. Research Memorandum**. No. 23, Chicago: Statistical Laboratory. Department of Education, University of Chicago.

Wright, B. D., & Masters, G. N., (1982). **Rating scale analysis. A handbook for research measurement**. Chicago: MESA Press.

Wright, D., & Stone, M. (1979). **Best test design: A handbook for research measurement,** Chicago: MESA Press.