المقارنة بين طريقة الاحتمال المقترح للإجابة الصحيحة والطريقة التجريبية لتقدير درجات أسئلة الاختيار من متعدد وفق درجة الارتباط بإحصائيات المفردة والأفراد المقدرة بالنموذج اللوغاريتمي الثلاثي المعلم

إعداد

د/ مشعل مونس نخيلان الرويلي

قسم التربية وعلم النفس – كلية التربية – جامعة الجوف – المملكة العربية السعودية

ملخص الدراسة:

هدفت الدراسة إلى المقارنة بين بين طريقتين لتقدير درجات أسئلة الاختيار من متعدد وهما (طريقة الاحتمال المقترح للإجابة الصحيحة، والطريقة التجريبية) في ضوء ارتباط احصائيات الافراد والمفردات المستمدة من هاتين الطريقتين بالإحصائيات المناظرة المقدرة بالنموذج اللوغاريتمي الثلاثي المعلم، ولتحقيق أهداف الدراسة والإجابة عن أسئلتها قام الباحث بإعداد اختبار من نوع الاختيار من متعدد ذو الأربع بدائل وذلك في وحدتي (المعادلات الخطية، والمتباينات الخطية) من مادة الرياضيات للصف الثالث متوسط، وبعد التحقق من صلاحية أداة الدراسة للغرض الذي أعدت له تم تطبيقها على عينة بلغت في صورها النهائية (٩٠٣) طالبا موزعين على (١٥) مدرسة تم اختيارهم بطريقة العينة العشوائية العنقودية من مجتمع طلاب الصف الثالث متوسط في مدينة عرعر بمنطقة الحدود الشمالية بالمملكة العربية السعودية، وبعد جمع البيانات تم تحليلها باستخدام البرنامج الإحصائي (SPSS-22)، وبرنامج Bilog-MG وقد أسفر التحليل الإحصائي للبيانات عن عدد من النتائج كان من أهمها ما يلي:

١-أظهر التحليل الإحصائي لقدرات الطلاب أن:

أ ـ كانت الطريقة التجريبية هي الأكثر ارتباطا بدرجات القدرات المقدرة بالنموذج اللوغاريتمي الثلاثي المعلم وذلك بمعامل ارتباط (٠,٧٦) ثم طريقة الاحتمال المقترح للإجابة الصحيحة بمعامل ارتباط (٠,٧٤).

ب ـ الطريقة الأكثر قربا من النموذج اللوغاريتمي الثلاثي المعلم في متوسط الفروق المطلقة بين الدرجات التائية هي الطريقة التجريبية بفارق (٥,٧) درجة، ثم طريقة الاحتمال بفارق (٥,٩) درجة.

٢-أظهرت نتائج مقارنات إحصائيات الفقرة (الصعوبة والتمييز) أن:

أ ـ الطريقة الأكثر ارتباطا بالنموذج اللوغاريتمي الثلاثي في تقدير صعوبة الفقرات هي الطريقة التجريبية بمعامل ارتباط (٠,٨٣)، ثم طريقة الاحتمال المقترح للإجابة الصحيحة بمعامل ارتباط (٠,٨١)، وتشير القيمتين إلى ارتباط موجب عالي.

ب ـ الطريقة الأكثر ارتباطا بالنموذج اللوغاريتمي الثلاثي في تقدير تمييز الفقرات هي الطريقة التجريبية ثم طريقة الاحتمال المقترح للإجابة الصحيحة، وقد كانت درجة ارتباط كل منهما ضعيفة جدا وغير دالة إحصائيا(عند مستوى الدلالة ٠,٠٥).

الكلمات المفتاحية : طرق تقدير الدرجات،النظرية الكلاسيكية في القياس، نظرية الاستجابة للمفردة، النموذج اللوغاريتمي الثلاثي.

Abstract:

A **comparison between of method of probability assigned to the correct answer and the experimental method for estimating the scores of the multiple choice questions according to the degree of correlation with the statistics of individual and Three-Parameter Logistic Model (3PL) estimates items.**

The study aimed to compare between two methods of estimating the multiple-choice questions (method of probability assigned to the correct answer, and experimental method) in light of the correlation of the statistics

of individuals and items that are derived from these methods with the corresponding statistics that are estimated by Three-Parameter Logistic Model (3PL). In order to achieve the aims of this study and to answer its questions, the researcher prepared four-multiple choice test at the units (linear equations and linear Inequalities) of Math subject regarding third intermediate level. After getting sure of the study tool propriety to the purpose of the study, it has been applied on a sample of (903 students) distributed on 15 schools and has been chosen by cluster random sample from the population of third intermediate level in Arar city at the north of Kingdom of Saudi Arabia. After collecting data and analyzing them through using SPSS program and Bilog-MG, the statistical analyzing of data revealed the following results:

1-The statistical analysis of the students' ability revealed that:

- **A. The means of the students' scores according to the two methods (probability assigned to the correct answer and the experimental method) were convergent in general. but, the experimental method was the highest in estimating the students' scores then the method of probability assigned to the correct answer.**
- **B. The experimental method was most related method to the abilities scores that are estimated by the three-parameter logarithmic model with a correlation coefficient (0.76) then the method of probability assigned to the correct answer with a correlation coefficient (0.74)**
- **C. The method that was closest to the Three-Parameter Logistic Model in the means of absolute differences between the T-scores, was the experimental method with a difference of (5.7) then the method of probability assigned to the correct answer with a difference of (5.9)**

2- The results of comparisons of item statistics (difficulty and discrimination), showed:

- **a- The method that was most related of the Three-Parameter Logistic Model in estimating the difficulty of items is the experimental method of correlation coefficient (0.83), then the method of probability assigned to the correct answer with a correlation coefficient (0.81). The two values reveal a high positive correlation.**
- **b- The method that was most related with the Three-Parameter Logistic Model in estimating the discrimination of items is the experimental method then the method of probability assigned to the correct answer .and The degree of their correlation was very weak and not statistically significant at (0.50)**

Key words: Methods of Scoring, Classical Test Theory, Item Response Theory, Tree Parameter Logistic Model

مقدمة الدراسة:

يحتل موضوع التقويم الصفي أهمية كبيرة في العملية التدريسية، فهو أحد العناصر الأساسية للمنهج المدرسي، والموجه الأساسي لسير بقية العناصر، كما أن أغلب القرارات والإجراءات التدريسية لابد وأن تتم في ضوء نتائج التقويم، ومن ثم ففاعلية هذه الإجراءات ومدى الاستفادة منها مرتبط بدقة البيانات التي يقدمها

ويعد الاختبار هو الاسلوب التقليدي المستخدم من قبل المعلمين في تقويم أداء الطلاب والكشف عن نواتج التعلم (Chang,Lin&Lin,2007)، والذي يأخذ عدة صيغ من أهمها وأكثرها استخداما اختبارات الاختيار من متعدد (Lau,Lau,Hong&Usop,2011).

وعلى الرغم مما تتميز به اختبارات الاختيار من متعدد إلا أن درجة الاختبار فيها تتحدد جزئيا بواسطة سلوك التخمين لدى الطالب عندما لا يكون متأكدا من إجابة فقرة معينة (Roger,1971,297)، وهذا يؤدي إلى انخفاض صدق الاختبار وذلك لتأثر درجة الطالب بعامل أخر غير السمة المراد قياسها وهو عامل التخمين (Kurz,1999,2)

كذلك فإن فقرات هذه الاختبارات -عند تصحيحها بالطريقة التقليدية -تكون غير قادرة على تقدير معرفة الطالب الجزئية فيما تقيسه الفقرة الاختبارية (Kurz,1999,2)، حيث أن الطالب قد يكون لديه معرفه جزئية عن المادة العلمية التي تقيسها الفقرة إلا أنه إذا فشل في اختيار البديل الصحيح فسوف تكون درجته صفرا متساويا بذلك مع الطالب الذي ليس لديه أية معرفة عن تلك الفقرة.

أن ما سبق يعد من أهم عوامل القصور في الطريقة التقليدية لتقدير الدرجات في فقرات اختبارات الاختيار من متعدد وهو كذلك من أهم المرتكزات التي قامت عليها الطرق البديلة في تقدير درجات تلك الفقرات، ومن بينها طريقتي (الاحتمال المقترح للإجابة الصحيحة والطريقة التجريبية) والتي جاءت بهدف الوصول إلى أفضل تقدير لقدرات الطلاب، عن طريق الحصول على أكبر قدر من المعلومات الكمية، وتقليل خطأ القياس إلى أقل حد (غنيم،١٩٨٦، ٧٤)، وهو ذات الهدف النهائي الذي تهتم به جميع نظريات القياس ومنها نظرية استجابة الفقرة التي جاءت كثورة في مجال القياس التربوي والنفسي حيث قدمت الحلول للعديد من أوجه القصور في القياس الكلاسيكي، كما أصبحت وسيلة أساسية وشائعة في بناء وتطوير الاختبارات حيث أنها تقدم بديلاً عن نظرية القياس التقليدية في تقدير معالم الفرد والفقرة بأقل قدر من الخطأ (عبدالحافظ،١٩٩٩).

وقد انبثق عن نظرية الاستجابة للمفردة مجموعة من النماذج التي تهدف إلى تحديد العلاقة بين أداء الفرد في الاختبار وبين السمات أو القدرات التي تكمن وراء هذا الأداء وتفسره (علام، ١٩٨٦، ١٠٠)،ومن أهم هذه النماذج وأكثرها شيوعا نموذج راش Rasch Model أحادي المعلم، ونموذج بيرنبوم Birnbaum Model ثنائي المعلم، ونموذج لورد Lord Model ثلاثي المعلم (علام، ٢٠٠٢، ٦٩٣).

ويعد النموذج اللوغاريتمي الثلاثي المعلم هو النموذج الكامل والحالة العامة بين النماذج البارامترية الثلاثة السابقة حيث يوصف المنحنى المميز للفقرة وفق هذا النموذج من خلال ثلاثة معالم تشتق رياضيا من البيانات الإمبريقية وهي معلم صعوبة الفقرة، ومعلم التمييز، ومعلم التخمين (اناستازي وأورينا، ٢٠١٥، ٢٤٣)، وهذا يجعله يتميز هذا النموذج عن النموذجين الآخرين بمراعاة عامل التخمين وهو ما يتوقع أن يحصل في كثير من اختبارات الصح والخطأ أو الاختيار من متعدد مما يمكن أن يؤثر على دقة تقدير قدرات الطلاب في هذه الأنواع من الاختبارات

وقد تميزت نظرية الاستجابة للمفردة بقوتها على المستوى التنظيري وقدرتها على إعطاء تقديرات أفضل للمستويات الحقيقية للأفراد على متصل السمة، كما أنها توفر تقديراً للقدرة مستقلاً عن خصائص العينة وبمستوى قياس يحقق مميزات القياس ذي الفئات المتساوية (النفيعي،٢٠١٢، ١٧٨؛ كاظم، ١٩٨٨، ٤١)، وقد اكد تفوق هذه النظرية عددا من الدراسات هدفت لمقارنتها مع المدخل الكلاسيكي مثل : (Hambleton&Jones,1993; Ainol&Noor,2006; dedoyin,2010; Ayhan,2015). إلا أن هذا التفوق رافقه صعوبات كبيرة حدت من استخدام نماذج هذه النظرية كاشتراطها توافر افتراضات قوية يصعب تحقيقها، وتطلبها أعداد كبيرة من الأفراد لتقدير بارامترات الفقرة واستخدامها لعمليات رياضية متقدمة وبرامج حاسوبية خاصة (Hambleton&Jones,1993,43; Baker,2001,1)

لهذا يفضل عدد من الباحثين استخدام المدخل الكلاسيكي في المواقف الاختبارية التقليدية، كالاختبارات الصفية التي يكون فيها الأفراد المستهدفون متجانسين في السمة المراد قياسيها، ويطبق عليهم نفس فقرات الاختبار أو اختبارات متوازية (Lisa, Carolyn & Julie, 2002,530)، حيث أن النظرية الكلاسيكية تتميز بسهولة تحقق افتراضاتها، بالإضافة إلى إمكانية تحليل فقرات الاختبار في ضوئها دون الحاجة إلى تطبيقه على عينة كبيرة من الأفراد أو استخدام عمليات رياضية معقدة. (Fan,1998).

مشكلة الدراسة وأسئلتها : لقد أجريت العديد من الدراسات التي تناولت تقييم الطرق الكلاسيكية البديلة في تقدير درجات اختبارات الاختيار من متعدد كالدراسات التي أجراها كل من (Lau et al,2011; Rippey,1970 ؛ العدل،١٩٨٦؛ سواقد، ١٩٩٨٣؛ سواقد،١٩٩٢؛ عبد الوهاب، ٢٠٠١؛ اللحياني، ٢٠٠٩)، إلا جميع هذه الدراسات كانت محكات الحكم فيها تستند إلى نظرية القياس الكلاسيكي، كما أن النتائج التي خرجت بها جاءت متباينة حول أفضلية أي من هذه الطرق.

وحيث أن نظرية استجابة الفقرة ثورة في عالم القياس التربوي والنفسي وذلك لكونها تراعي عدد من المتغيرات التي أغفلتها نظرية القياس الكلاسيكي مما أثمر عن قياس أكثر دقة وموضوعية في تقدير قدرات الأفراد، وقد أكد هذا التفوق عدد من الدراسات التي قارنت بين هاتين النظريتين كدراسات: Hambleton&Jones,1993; Ainol&Noor,2006; Ayhan,2015 (Adedoyin,2010;؛ عبد الوهاب،٢٠١٣). إلا أن هذا التفوق رافقه صعوبات كبيرة حدت من استخدام نماذج هذه النظرية كاشتراطها توافر افتراضات قوية يصعب تحقيقها، وتطلبها أعداد كبيرة من الأفراد لتقدير بارامترات الفقرة واستخدامها لعمليات رياضية متقدمة وبرامج حاسوبية خاصة (Hambleton&Jones,1993,43; Baker,2001,1).

لهذا جاءت الدراسة الحالية لتكمل البناء البحثي حول هذا الموضوع وذلك من خلال المقارنة بين طريقتين من هذه الطرق البديلة وهي (الطريقة التجريبية، وطريقة الاحتمال المقترح للإجابة الصحيحة) وفق معيار لم تتطرق له الدراسات السابقة وهو علاقة إحصائيات الفقرات والأفراد (صعوبة الفقرات، تمييز الفقرات، وقدرات الأفراد) المستمدة من هاتين الطريقتين بمعالم الفقرات والأفراد المقدرة بالنموذج اللوغاريتمي الثلاثي المعلم والذي يتميز بمراعاة أثر عامل التخمين على تقدير قدرات الأفراد، وهو ما اهتمت به كذلك الطريقتين الكلاسيكيتين التي تناولتهما الدراسة الحالية،ويمكن صياغة مشكلة الدراسة بالسؤال الرئيس التالي :

ما درجة ارتباط إحصائيات الفقرات والأفراد المستمدة من طريقتي (الاحتمال المقترح للإجابة الصحيحة والطريقة التجريبية) بإحصائيات الفقرات والأفراد المستمدة من النموذج اللوغاريتمي الثلاثي المعلم، وينقسم هذا السؤال إلى عدد من الأسئلة الفرعية على النحو التالي:

١- ما درجة ارتباط تقديرات قدرات الطلاب عند استخدام النموذج اللوغاريتمي الثلاثي المعلم بتقديرات درجاتهم عند استخدام كل من طريقة من طرق تقدير الدرجات الكلاسيكية (طريقة الاحتمال المقترح للإجابة الصحيحة، والطريقة التجريبية)؟
٢- ما درجة الاختلاف بين قدرات الطلاب المقدرة وفق النموذج اللوغاريتمي الثلاثي المعلم وقدراتهم المقدرة وفق كل من (طريقة الاحتمال المقترح للإجابة الصحيحة، والطريقة التجريبية)؟
٣- ما درجات ارتباط قيم معاملات صعوبة الفقرات عند استخدام كل من طريقة الاحتمال المقترح للإجابة الصحيحة، والطريقة التجريبية مع قيم معاملات صعوبة الفقرات عند استخدام النموذج اللوغاريتمي الثلاثي المعلم؟
٤- ما درجات ارتباط قيم معاملات تمييز الفقرات عند استخدام كل من طريقة الاحتمال المقترح للإجابة الصحيحة والطريقة التجريبية مع قيم معاملات تمييز الفقرات عند استخدام النموذج اللوغاريتمي الثلاثي المعلم؟

مصطلحات الدراسة الإجرائية:

طرق تقدير الدرجات Methods of Scoring:_ هي القاعدة التي يعطى في ضوئها قيماً كمية تعكس الدرجة المستحقة للطالب في كل فقرة من فقرات الاختبار، وقد اشتملت الدراسة الحالية على ثلاث طرق لتقدير درجات فقرات الاختيار من متعدد وهي: الطريقة التقليدية، والطريقة التجريبية، وطريقة الاحتمال المقترح للإجابة الصحيحة.

طريقة الاحتمال المقترح للإجابة الصحيحة Method of probability assigned to the correct answer : في هذه الطريقة يقوم الطالب بإعطاء نسب مئوية تعبر عن مدى تقديره لصحة كل بديل من بدائل فقرة الاختيار من متعدد، بحيث يكون مجموع هذه النسب مساويا لـ ١٠٠ ٪، ويتم تقدير درجة المفحوص بأخذ النسبة المئوية التي اقترحها للبديل الصحيح لتعبر عن درجته على الفقرة (Rippey,1970).

الطريقة التجريبية Experimental Method: في هذه الطريقة تختلف إجابة الطالب ودرجته على السؤال بحسب مدى ثقته بمعرفة البديل الصحيح، وفيما يلي توضيح لكيفية الإجابة على اختبار الاختيار من متعدد المستخدم في هذه الدراسة (المكون من أربع بدائل)، وكذلك كيفية تقدير الدرجات وفق هذه الطريقة:

- إذا كان الطالب متأكدا من معرفة البديل الصحيح، فإنه يضع أمامه رقم (١) ويكتفي بذلك.
- إذا كان الطالب يشك في صحة بديلين فإنه يضع أمام أحدهما رقم (١)، وأمام الثاني رقم (٢).
- إذا كان الطالب يشك في صحة ثلاث بدائل فإنه يضع أمام أحدهما رقم (١)، وأمام الثاني رقم (٢) وأمام الثالث رقم (٣) .
- لا يحق للطالب اختيار جميع بدائل السؤال الأربعة، وإن فعل تكون درجته صفر.

وعند تقدير درجة الطالب فإن درجته تقل بزيادة عدد البدائل التي يختارها بحيث يحصل الطالب على درجة الفقرة كاملة (ثلاث درجات) إذا اكتفى باختيار بديل واحد وكان هذا البديل هو البديل الصحيح، ويحصل الطالب على درجتين إذا قام باختيار بديلين أحدهما هو البديل الصحيح، ويحصل الطالب على درجة واحدة إذا قام باختيار ثلاث بدائل أحدهم هو البديل الصحيح، فيما يحصل الطالب على درجة (صفر) إذا لم يقع البديل الصحيح ضمن البدائل التي اختارها، أو إذا قام الطالب باختيار جميع بدائل السؤال.

النموذج اللوغاريتمي ثلاثي المعلمة Three Parameter Logistic Model : هو أحد نماذج نظرية الاستجابة للمفردة أحادية البعد ثنائية الاستجابة وهي النماذج التي تستخدم في تدريج فقرات الاختبار الذي يقيس سمة أو قدرة واحدة والتي يتم تقدير درجة الفرد فيها من بين درجتين هما (صفر أو واحد)، ويراعي هذا النموذج إمكانية تفاوت فقرات الاختبار في صعوبتها، وقوتها التمييزية، واحتمالية الإجابة عنها عن طريق التخمين، حيث يفترض أن احتمال وصول الفرد للإجابة(علام،٢٠٠١)

صعوبة الفقرة Item Difficulty:

١- صعوبة الفقرة وفق النظرية الكلاسيكية: في حالة كانت إجابة الفقرة ثنائية (إما صفرا أو واحدا) فتعرف صعوبتها بأنها نسبة الطلاب الذين أجابوا إجابة صحيحة عن الفقرة من بين أفراد العينة (ملحم،٢٠٠٢،٢٣١)، أما في حالة كانت إجابة الفقرة متدرجة (تراعي المعرفة الجزئية) فإن صعوبتها تعرف بأنها نسبة الدرجات التي حصل عليها الطلاب عند إجابتهم لهذه الفقرة (عودة،٢٠٠٢، ٢٩٠).

٢- صعوبة الفقرة وفق النموذج اللوغاريتمي الثلاثي (b): هي نقطة على مقياس القدرة يكون عندها احتمال الإجابة الصحيحة = 2 / (C +1)، حيث C احتمال إجابة الأفراد ذوي المستويات المنخفضة من القدرة إجابة صحيحة عن الفقرة عن طريق التخمين . (Baker,2001,28-30).

تمييز الفقرة Item Discrimination:

١- تمييز الفقرة وفق النظرية الكلاسيكية: هو مؤشر يستخدم للكشف عن مدى مقدرة الفقرة على التمييز بين الطلبة ذوي مستويات القدرة المختلفة (Demars,2010)، وفي الدراسة الحالية تم حساب قيم معاملات تمييز الفقرات كلاسيكيا باستخدام الطريقة الارتباطية والتي يتم فيها إيجاد معامل الارتباط بين الدرجة على الفقرة من جهة والدرجة الكلية من جهة أخرى (النبهان، ٢٠٠٤، ٢٠١).

٢- تمييز الفقرة وفق النموذج اللوغايتمي الثلاثي (a): هو مؤشر إحصائي يعبر عن قدرة الفقرة على التمييز بين المختبرين ذوي المستويات المختلفة من السمة الكامنة أو القدرة، حيث يعكس مدى التغير في احتمالية الاجابة الصحيحة بتغير مستوى السمة الكامنة لدى الفرد، ويسمى أحيانا بالميل(Slope) لأنه يمثل ميل المنحنى المميز للفقرة عند نقطة الانقلاب والتي يكون فيها احتمال إجابة الفرد عن الفقرة إجابة صحيحة يساوي (1+C)/2 (Baker,2001;Georgiev,2008;Demars,2010)

الإطار النظري والدراسات السابقة:

لقد حاول العلماء المشتغلين بالقياس التربوي والنفسي الوصول إلى تقدير دقيق لقدرات المفحوصين وذلك عن طريق الحصول على المزيد من المعلومات الكمية عنهم وتقليل خطأ القياس بقدر الإمكان وكان من نتائج ذلك ابتكار عدة طرق لتقدير الدرجات، وقد حظيت الفقرات ذات الاختيار من متعدد بالنصيب الأكبر من هذه الطرق (غنيم، ١٩٨٦، ٧٤). و طرق تقدير الدرجات هي أساليب وإجراءات تتعلق بتعليمات تطبيق الاختبار وتصحيحه يتم من خلالها تقدير درجة المفحوص على كل فقرة من فقرات الاختبار وفق نظام رقمي محدد يختلف من طريقة لأخرى، ففي حين يقتصر هذا النظام الرقمي على الرقمين (صفر، واحد) لكل فقرة عند استخدام الطريقة التقليدية، فإنه يأخذ قيماً متعددة في طرق أخرى.

وتعد الطريقة التقليدية من أهم طرق تقدير الدرجات في فقرات الاختيار من متعدد ومن أكثرها استخدامها، وفيها يطلب من المفحوص فقط أن يحدد البديل الصحيح من بين بقية البدائل، وعند تصحيح الفقرة تكون درجتها أما واحداً وذلك في حالة كان اختيار البديل صحيحاً أو صفر في حالة كان الاختيار خاطئاً. ألا أنه وعلى الرغم من كثرة استخدام هذه الطريقة في تقدير الدرجات إلا أنه وجه لها الكثير من النقد وذلك لكونها تشجع الطلاب على التخمين كما أنها لا تعطي معلومات عن المعرفة الجزئية للطالب(Lau et al.,2011)، حيث أن الطالب الذي يحصل على الدرجة صفر كثيرا ما يكون لديه معرفة جزئية حول المادة التي تقيسها الفقرة ومن ثم فهو يستحق عليها جزءا من الدرجة يقابل تلك المعرفة (Kansup & Hakstian,1975) وانطلاقا من هذه العيوب للطريقة التقليدية فقد ظهرت العديد من طرق تقدير الدرجات الأخرى والتي يمكن استخدامها في فقرات الاختيار من متعدد .

وقد قام الباحث بتصنيف هذه الطرق في أربعة مجموعات رئيسية وذلك وفقاً لطبيعة كل مجموعة من هذه الطرق، حيث تضم المجموعة الأولى الطرق التي لا تتطلب توزيع تقديرات احتمالية لجميع بدائل الفقرة، والمجموعة الثانية تضم الطرق التي تشترط قيام الطلبة بتوزيع توقعات احتمالية أو نسب ثقة على جميع بدائل الفقرة، والمجموعة الثالثة تضم الطرق اللوغاريتمية، أما المجموعة الرابعة فتقتصر على طرق التي تستخدم المعادلات الثلاثة للتصحيح من أثر التخمين، وفيما يلي عرض لأهم هذه الطرق :

أولاً، طرق لا تتطلب توزيع تقديرات احتمالية لجميع بدائل الفقرة : وتضم هذه المجموعة عدد من الطرق المتباينة فيما بينها سواء في تعليمات الإجابة عليها أو في النظام الرقمي الذي يتم في ضوئه تقدير درجات فقرات الاختبار، إلا أنها تشترك في أن جميع هذه الطرق لا تلزم الطالب بأن يوزع نسبا مئوية على جميع البدائل لتعكس مدى ثقته بصحتها كما في المجموعة الثانية الطرق، ولا تستخدم اللوغاريتمات في تقدير الدرجة كما في المجموعة الثالثة، ولا المعادلات التقليدية للتصحيح من أثر التخمين كما في المجموعة الرابعة، وفيما يلي عرض موجز يوضح أهم هذه الطرق :

١ ـ الطريقة التقليدية في تقدير الدرجات Number right (NR) : وفي هذه الطريقة يقوم الطالب بتفحص جميع بدائل السؤال ثم اختيار أحد هذه البدائل فقط ليكون إجابة السؤال، ويقوم المعلم بمنحه الدرجة واحد في حالة كان الاختيار صحيحا والدرجة صفر في حالة كون الاختيار خاطئاً أو أن كونه ترك السؤال دون إجابة (Lau et al.,2011)، ثم يقوم المعلم بجمع درجات الفقرات التي قام الطالب بالإجابة عليها إجابة صحيحة ويرصدها للطالب لتعبر عن الدرجة الكلية للطالب على الاختبار وذلك دون أي إجراء تعديل عليها. والمزايا الرئيسية لهذه الطريقة هي سهولة تطبيق الاختبار وتصحيحه، ومناسبتها للتطبيق على عينات كبيرة من الأفراد، بالإضافة إلى موضوعيتها في تقدير الدرجات. (Ben-Simon, Budescu, & Nevo, 1997)

إلا انه يعاب عليها عدد من الأمور من أهمها أن الطالب الذي ليست لدية المعرفة اللازمة للإجابة عن الفقرة الاختبارية قد يحصل على درجات إضافية عن طريق التخمين فقط، مما يجعل التخمين عامل عشوائياً يضاف للعوامل التي تؤثر في درجة الاختبار مما يساهم في خفض ثبات درجات الاختبار، وكذلك صدق نتائجه، فمن يقوم بتصحيح الاختبار أو من يحلل نتائجه لن يكون قادراً على التمييز بين الدرجات التي تم الحصول عليها نتيجة توفر المعرفة لدى الطالب وبين تلك الدرجات التي تم الحصول عليها عن طريق التخمين فقط (Lesage, Valcke & Sabbe,2013)

٢ـ طريقة الحذف أو الاستبعاد Elimination Method: وتسمى الاختبارات التي تستخدم هذه الطريقة باختبارات الحذف أو الاستبعاد (Elimination Testing (ET، وأول من اقترحها كل من Coombs, Milholland, and Womer،وذلك عام 1956 ، وفي هذه الطريقة يطلب من الطالب أن يستبعد جميع البدائل التي يثق أنها غير صحيحة، ويبلغ بأنه سيمنح درجة عن كل بديل خاطئ قام الطالب باستبعاده،بشرط أن لا يقوم باستبعاد البديل الصحيح لأن هذا سيترتب عليه الخصم منه بما مقداره الدرجة العظمى للفقرة (والتي تساوي عدد البدائل – ١) كما يحق له أن يترك البدائل التي يشك في صحتها دون استبعاد (Lau et al,2011).،وتتميز هذه الطريقة في أنها تراعي المعرفة الجزئية للطالب وتعبر عنها بشكل واضح، ففي المثال السابق نجد أن الدرجة ٣ تدل على معرفة الطالب الكاملة لما تقيسه الفقرة ، بينما تدل الدرجات (٢، ١) على المعرفة الجزئية، وتدل الدرجة (صفر) على غياب المعرفة، وتدل الدرجات (-١ أو -٢) على سوء فهم جزئي، بينما تدل الدرجة (-٣) على سوء فهم كامل (Bradbard & Green,1986).

٣ ـ طريقة تجمع بين طريقتي الحذف والطريقة التقليدية: وفي هذه الطريقة على الطالب أن يستبعد الإجابات التي يثق أنها خاطئة، كما عليه أن يختار احد البدائل ليكون إجابة للفقرة، حيث تكون الخيارات التي يضعها الطالب هي علامة (√) وذلك أمام البديل الذي يرى أنه صحيح أو علامة (×) أما البديل الذي يكون متأكدا من أنه خاطئ، أو يضع عبارة (غير متأكد) عندما يكون غير متأكد من عدم صحة بديل ما، ويتم تقدير الدرجات في هذه الطريقة حسب ما ذكر Lau et al,2011) على النحو التالي :

- يعطي الطالب درجة واحدةً عن كل بديل خاطئ يتم استبعاده
- يعطى الطالب درجة واحدةً في حالة استطاع اختيار البديل الصحيح
- يخصم من الطالب ثلاث درجات (-٣) في حالة قيامه باستبعاد البديل الصحيح
- تتراوح درجة الطالب بين (-٣ و ٤)، ولا يضاف للطالب أي درجة (يعطي درجة صفر) عند ا اختيار البديل (غير متأكد).

وتختلف هذه الطريقة عن الطريقة السابقة (طريقة ا لحذف) في أنها تشترط على الطالب أن يختار أحد البدائل كإجابة للفقرة، وذلك بعكس الطريقة السابقة والتي كانت تكتفي فقط بالطلب منه أن يستبعد البدائل التي يرى أنها خاطئة ومن ثم فإن الطالب في طريقة الحذف أو الاستبعاد يحق له أن يبقي بدلين أو أكثر دون أن يحدد أيهم هو البديل الصحيح،ويلاحظ أن هذه الطريقة عملت على التقليل من أثر التخمين وذلك من خلال إجرائين :

الأول: إيجاد البديل (غير متأكد) وفي هذا البديل مخرج للطالب أن يختاره دون أن يخمن.

الثاني: في إيجاد العقوبة (حسم ثلاث درجات في حالة تم استبعاد البديل الصحيح) .

٤ ـ الطرق الوزنية: في هذه الطرق يقوم الطالب باختيار البديل الذي يراه صحيحاً فقط وذلك كما في الطريقة التقليدية، إلا أنه يتم تصحيح الاختبار وتقدير درجاته بطريقة مختلفة، وهذه الطرق تشمل (الطريقة الوزنية لجتمان، والطريقة الوزنية لديفز) :

أ ـ الطريقة الوزنية لجتمان: اقترح هذه الطريقة جتمان عام ١٩٤١،وفيها يسمح للطالب أن يجيب باختيار البديل الصحيح كما في الطريقة التقليدية، إلا أن هناك عدد من الإجراءات والعمليات الحسابية يتم تطبيقها بعد تصحيح الاختبار يمكن من خلالها الوصول إلى وزن نسبي لكل بديل من بدائل الاختبار ليتم اعتماده كدرجة وزنية لمن يختار هذا البديل ، ويورد غنيم (١٩٨٦) أهم

الخطوات والإجراءات التي ينبغي اتباعها لتقدير الدرجات وفق هذه الطريقة وذلك على النحو التالي :

- يتم تحديد جميع الأفراد الذين قاموا باختيار بديل معين من بدائل الفقرة، وتعزل أوراق إجاباتهم تمهيدا لإجراء عدد من العمليات اللازمة لحساب الدرجة الوزنية لهذا البديل .
- تحسب الدرجة الكلية لكل من هؤلاء الأفراد على بقية فقرات الاختبار الأخرى مع مراعاة الآتي:

١- عدم إدخال درجة الفقرة التي تحتوى البديل المراد تقدير وزنه فمثلا لو كان الاختبار مكون من عشرة فقرات يتم حساب الدرجة على التسعة فقرات الأخرى وذلك بعد إلغاء درجة الفقرة المتضمنة للبديل المراد تقدير وزنة.

٢- تحسب درجة كل فرد بطريقة التصحيح لأثر التخمين وفقاً للقانون التالي :

الدرجة المصححة = عدد الإجابات الصحيحة – [عدد الإجابات الخاطئة ÷ (عدد البدائل – ١)]

٣- يؤخذ المتوسط الحسابي للدرجات الناتجة من الخطوة السابقة، ويعتبر هذا المتوسط قيمة وزنية مبدئية لذلك البديل

٤- تتبع نفس الخطوات السابقة مع بقية بدائل الفقرة، كما تتبع نفس الخطوات لبقية بدائل جميع الفقرات ليتم احتساب درجة وزنية ابتدائية لكل بديل من بدائل فقرات الاختبار .

٥- يتم حساب التجانس الداخلي لفقرات الاختبار وذلك بحساب معاملات الارتباط بين درجة كل فقرة والدرجة الكلية للاختبار.

٦- يعاد حساب متوسط الحسابي لدرجات الأفراد الذين اختاروا كل بديل من بدائل كل فقرة من فقرات الاختبار بنفس الطريقة السابقة، مما ينتج درجات وزنية جديدة لكل بديل من البدائل

٧- يتم حساب التجانس الداخلي لفقرات الاختبار مرة أخرى، وذلك وفقا للدرجات الوزنية الجديدة .

٨- تتم المقارنة بين قيم التجانس الداخلي في المرة الأولى والثانية، فإذا وجد هناك فروق ذات دلالة بينهما فإنها تكرار حساب الخطوات السابقة والوصول إلى درجات وزنية جديدة .

٩- تستمر عملية أعادة تقدير الدرجات الوزنية لبدائل الاختبار لحين نصل إلى مصفوفتين متتالين لقيم معاملات التجانس الداخلي ليس بينهما فروق ذات دلالة إحصائية .

١٠- يكون أخر تقدير للدرجات الوزنية لكل بديل من بدائل الاختبار هو المعبر عن الدرجة المعتمدة التي يتم إعطائها لمن يختار هذا البديل .

ويلاحظ على هذه الطريقة أنها بوجه عام تقوم على تقدير الأوزان النسبية لبدائل فقرة الاختيار من متعدد على أساس الدرجات الكلية لمن قاموا باختيار هذا البديل وذلك بغض النظر عن صحة هذا البديل من عدمه، فالبديل الذي يختاره الطلبة ذوي المستويات العليا تكون درجته الوزنية أعلى من البديل الذي يختاره الطلبة ذوي المستويات الأقل، كذلك يلاحظ على هذه الطريقة أنها

تسير وفق خطوات ومراحل طويلة نسبياً، كما أنها تتطلب الإلمام ببعض العمليات إحصائية التي قد يصعب على غير المختصين الإلمام بها، وكل هذا مما يحد من انتشار استخدامها.

ب ـ طريقة الدرجة الوزنية للاختيار لديفز Choice Weight Scoring: تعتمد هذه الطريقة على صعوبة وسهولة الأسئلة بحيث لا تكون درجات الأسئلة متساوية، بل تحدد درجة السؤال على حسب صعوبته (اللحياني، ٢٠٠٩)

كما تحدد درجة كل بديل من البدائل بحسب ما يتطلب الوصول إليه من عمليات معرفيه ويلاحظ أنه يمكن أن تأخذ البدائل غير الصحيحة درجات موجبه أو صفر أو درجات سالبة، وهذا يختلف من سؤال لآخر ويتم تقدير الدرجات الوزنية لكل بديل من البدائل بواسطة المحكمين أو واضع الاختبار، أو من خلال دراسة استطلاعية (غنيم، ١٩٨٦).

فمثلاً إذا كانت فقرات الاختبار مكونة من أربعة بدائل، فإن أحدى فقرات هذا الاختبار يمكن أن نعين لبدائلها الدرجات التالية : الدرجة (+٣) للبديل الصحيح، و الدرجات (صفر، -١، -٣) لكل بديل من البدائل الخاطئة، في حين يمكن أن نعين لبدائل فقرة أخرى من نفس الاختبار درجات مختلفة على النحو التالي : الدرجة (+١) للبديل الصحيح، و الدرجات (صفر ، صفر ، -٢) لكل بديل من البدائل الخاطئة. (Sabers &White ,1969, 93)

ويلاحظ على هذه الطريقة أنه على الرغم من كون أساليبها في تقدير أوزان البدائل تتبع نفس المبدأ إلا هناك اختلاف واضح فيما بينها في آلية تقدير هذه الأوزان، مما يمكن أن يعطي نتائج مختلفة عند استخدام أي من هذه الأساليب، حيث يرى الباحث أن كل أسلوب من هذه الأساليب يمكن اعتباره طريقة مستقله.

٥ـ طريقة الدرجات الذاتية Self-Scoring Methods : تسمى الاختبارات التي تستخدم هذه الطريقة باختبارات المحاولة حتى الوصول للإجابة الصحيحة Answer-Until-Correct (AUC) وفي هذه الطريقة يقوم الطالب مواصلة الاختيار من بدائل فقرة الاختيار من متعدد واحداً تلو الآخر حتى يتوصل إلى البديل الصحيح، ولتسهيل هذه العملية غالبا ما يتم وضع طلاء كربوني قابل للمحو أمام كل بديل من البدائل ثم يقوم الطالب بمحو الطلاء الكربوني الذي أمام البديل الذي يراه صحيحاً، وعند المحو ستظهر للطالب علامة تبين له ما إذا كان البديل الذي اختاره صحيحاً أم لا، فإذا لم يكن صحيحا فعليه أن يمحو الطلاء الذي أمام بديل آخر وهكذا إلى أن يصل إلى البديل الصحيح (Rand,1982)، ويتم تصحيح الفقرات في أسلوب الدرجات الذاتية على النحو التالي :

الدرجة = عدد بدائل السؤال – عدد المحاولات للوصول إلى الإجابة الصحيحة، فعلى سبيل المثال إذا كان عدد البدائل في اختبار ما أربعة بدائل وقد توصل الطالب للإجابة الصحيحة بعد محاولتين فإن درجته على الفقرة تساوي: ٤ – ٢ = ٢ (اللحياني، ٢٠٠٩). وعلى الرغم من أن هذه الطريقة تتميز بعدد من المميزات الهامة كمراعاتها للمعرفة الجزئية، وأنها تقدم تغذية راجعة مباشرة، وأنها تساهم في خفض قلق الاختبار لدى الطلاب (Thompson,1976 ;Farland, Barlow, Levi Lancaster& Franks, 2015)، إلا أنه من ا لواضح انها تحتاج إلى تجهيزات خاصة وتستغرق المزيد من الوقت والجهد سواء في أعدادها أو تطبيقها.

٦ـالطريقة التجريبية Experimental Method: في هذه الطريقة يسمح للطالب أن يختار أكثر من بديل في حالة عدم تأكده من صحة الإجابة (Lesage et al,2013)، ويتم ذلك على النحو التالي :

يطلب من الطالب أن يضع رقم (١) أمام البديل الذي يرى أنه البديل الصحيح، فإذا كان متأكدا من صحته يتوقف ولا يضيف أية أرقام أخرى أما إذا كان يشك في صحة بديل أخر فإنه يضع أمامه رقم (٢)، كذلك عليه التوقف إذا لم يكن يشك في صحة أي بديل آخر أو الاستمرار بترقيم البدائل التي يشك في صحتها بشرط أن يبقي ـ على الأقل ـ بديل واحد دون ترقيم، وطريقة تصحيح الاختبار تعتمد على مدى نجاح الطالب في اختيار البديل الصحيح وعلى عدد البدائل التي قام باختيارها، ففي حالة نجاح الطالب في اختيار البديل الصحيح فإن درجته تقل بزيادة عدد البدائل التي قام باختيارها حيث أن الدرجة التي سيحصل عليها عند اختياره لبديل واحد فقط وكان هو البديل الصحيح تساوي (عدد بدائل الفقرة ـ١) ثم تأخذ بالتناقص بحسب عدد البدائل التي تم اختيارها.(غنيم،١٩٨٦؛ اللحياني،٢٠٠٩، ؛James,1975)

فمثلا إذا كان عدد البدائل أربعة بدائل فإن الطالب سيحصل على ثلاث درجات إذا قام باختيار بديل واحد فقط وكان هذا البديل هو البديل الصحيح، فيما سيحصل على درجتين إذا قام باختيار بديلين احدهما صحيح، ودرجة واحدة إذا اختار ثلاث بدائل أحدها صحيح، فيما سيحصل على (صفر) إذا قام باختيار جميع البدائل الأربع، أو إذا كان البديل الصحيح ليس ضمن البدائل التي قام باختيارها.

وتتشابه هذه الطريقة مع طريقة الاختيار الجزئي التي استخدمت في دراستي (سواقد،١٩٨٣؛ سواقد، ١٩٩٢) والتي تمت التوصية باستخدامها في كلا الدراستين، كما تم استخدام هذه الطريقة في عدد من الدراسات الأخرى مثل دراسة (عبدالوهاب، ٢٠٠١) والتي أظهرت نتائجها تفوق هذه الطريقة على الطريقة التقليدية في الرفع من قيمة معامل صدق الاختبار، وفي الدراسة الحالية تم استخدام هذه الطريقة ضمن طرق تقدير الدرجات التي تجرى المقارنة فيما بينها.

٧ ـ طريقة مستويات الثقة لبديل واحد : في هذه الطريقة يقوم الطالب باختيار بديل واحد فقط لكل فقرة من فقرات الاختبار،كما في الطريقة التقليدية، إلا أنه يضع أمام هذا البديل مستوى ثقته بأنه هو البديل الصحيح وذلك في ثلاث مستويات (تخمين، ثقة متوسطة، ثقة عالية)، ويوضح ستانلي (Stanley,(1975 آلية تقدير الدرجات وفق هذه الطريقة على النحو التالي:

إذا كانت إجابة الطالب صحيحة فإنه يتم تقدير درجته على الفقرة على النحو التالي:

- إذا حدد مستوى الثقة (ثقة عالية) فإنه يعطى (ثلاث درجات)
- إذا حدد مستوى الثقة (ثقة متوسطة) فإنه يعطى (درجتين)
- إذا حدد مستوى الثقة (تخمين) فإنه يعطى (درجة واحدة)

أما إذا كانت إجابة الطالب خاطئة فإنه يتم تقدير درجته على النحو التالي :

- إذا حدد مستوى الثقة (ثقة عالية) فإنه يحسم منه (ثلاث درجات)
- إذا حدد مستوى الثقة (ثقة متوسطة) فإنه يحسم منه (درجتين)
- إذا حدد مستوى الثقة (تخمين) فإنه لا يعطى أي درجة (صفر)

والفكرة الرئيسية لهذه الطريقة هي أن الطالب يعطى درجة أكبر بحسب درجة ثقته بالإجابة الصحيحة ،وفي هذا محاولة لتقدير ادق لمستوى المعرفة لديه، وفي المقابل فإنه عندما يعطي ثقة عالية لإجابة غير صحيحة يعاقب بالدرجة الكافية لمنعه من التظاهر غير المبرر بالثقة في الإجابة (Stanley,1975).

ثانياً: طرق تتطلب توزيع توقعات احتمالية أو نسب ثقة على جميع بدائل الفقرة : وتشترك جميع هذه الطرق في أنه يتم فيها الطلب من الطالب أن يعطي درجات تقدير أو نسب مئوية لمدى ثقته بكل بديل من بدائل فقرة الاختيار من متعدد، ومن هذه الطرق ما يلي:

١ـ طريقة الاحتمال المقترح للإجابة الصحيحة Method of Probability assigned to the correct answer: في هذه الطريقة يقوم الطالب بإعطاء نسب مئوية تعبر عن مدى تقديره لصحة كل بديل من البدائل بحيث يكون مجموع هذه النسب مساويا لـ ١٠٠ ٪ ، ويتم تقدير درجة المفحوص بأخذ النسبة المئوية التي اقترحها للبديل الصحيح لتعبر عن درجته على الفقرة، وفي حين ان الطالب يقوم بوضع نسب مئوية أو مستويات ثقة ويوزعها على جميع بدائل الفقرة إلا أنه وعند تصحيح الاختبار فإنه يتم التعامل مع البديل الصحيح فقط (Rippey,1970 ;عبدالوهاب،٢٠٠١; اللحياني،٢٠٠٩).

وتؤكد نتائج الدراسة التي أجراها ريبي (1970) Rippey أنه على الرغم من بساطه هذه الطريقة إلا أنها أظهرت تفوقا على غيرها من الطرق في ثبات درجات الاختبار القائم عليها وذلك عند مقارنتها بخمس طرق أخرى لتقدير درجات فقرات الاختيار من متعدد، كما أظهرت تفوقا على الطريقة التقليدية في قيم معاملات الصدق وفقا لنتائج الدراسة التي أجراها (عبد الوهاب، ٢٠٠١)، وهذه الطريقة هي أحدى طرق تقدير الدرجات التي تضمنتها الدراسة الحالية .

٢ ـ طريقة الاختيار الاستنتاجي Inferred Choice Method : يقوم الطالب في هذه الطريقة بوضع نسب مئوية لكل بديل من البدائل بحسب رأيه حول مدى صحة كل بديل، ثم يتم حساب درجة الفقرة وفق القاعدة التالية :

يعطي الطالب الدرجة (١) إذا كانت النسبة التي اقترحها للبديل الصحيح هي الأكبر بين النسب المقترحة للبدائل الأخرى، ويعطي الدرجة (صفر) إذا أعطى الطالب لأحد البدائل الخاطئة نسبة مئوية أعلى مما أعطاه للبديل الصحيح (Rippey,1970؛عبد الوهاب،٢٠٠١). وبطبيعة الحال يشترط في هذه الطريقة أن لا يعطي الطالب نفس النسبة المئوية لأكثر من بديل، لأن مثل هذا الفعل لا يمكننا من الكشف عن البديل الذي يراه الطالب الأقرب للإجابة الصحيحة، كذلك يجب التأكد من أن مجموع النسب التي وضعها الطالب لجميع البدائل مساويا لـ ١٠٠٪، ويلاحظ على هذه الطريقة أنها وإن كانت تعطينا معلومة حول أي البدائل هو أكثر احتمالا لأن يكون الإجابة الصحيحة من وجهة نظر الطالب، إلا أنها لا تتعامل مع مستويات ثقة الطالب بهذا البديل.

ويرى الباحث أنه على الجانب النظري فإن المعلومة التي تمدنا بها هذه الطريقة لا تختلف كثيراً عن الطريقة التقليدية حيث أننا عند استخدام الطريقة التقليدية كذلك نعرف أن الطالب سيختار البديل الذي يراه أكثر احتمالا لأن يكون هو البديل الصحيح.

٣ ـ الطريقة ا لكروية Spherical Method: في هذه الطريقة يقوم الطالب بتوزيع درجات ثقة أو توقعات احتمالية على جميع بدائل الفقرة، ويكون ذلك في صورة نسب مئوية، ثم يتم حساب درجة الفقرة وفق القانون التالي (اللحياني،٢٠٠٩ ؛ Rippey,1970) :

$$\text{الدرجة الكروية (ك)} = \frac{\text{ع ت}}{\left[\sum_{\text{س}=1}^{\text{ن}} (\text{ع س})^{2}\right]^{\frac{1}{2}}}$$

حيث:

ع ت : التوقع الاحتمالي أو النسبة المئوية التي أعطاها المفحوص للبديل الصحيح

ع س : التوقعات الاحتمالية أو النسب المئوية التي أعطاها المفحوص لكل بدائل الفقرة

ن : عدد بدائل الفقرة .

٤ - الطريقة الإقليدية Euclidean Function Method : تستخدم هذه الطريقة مع الفقرات التي لها إجابة وحيدة وكذلك تلك التي تتطلب توزيع أفضلية الإجابة على البدائل،ولتقدير درجة الفقرة يتم حساب متوسط آراء الجماعة المرجعية (الخبراء) لدرجة صحة (أفضلية) كل بديل من البدائل ويكتب على شكل فترة ثقة أو نسبة مئوية ويرمز له بالرمز م(س) (عبدالوهاب،٢٠٠١ ,Rippey,1970)، ثم يتم استخدام المعادلة التالية :

$$\text{الدرجة الإقليدية} = ١ - \frac{\left[\sum_{\text{س}=1}^{\text{ن}} (\text{ع س} - \text{م س})^{2}\right]^{\frac{1}{2}}}{\sqrt{٢}}$$

حيث:

ع س : التوقعات الاحتمالية التي أعطاها الطالب لكل بديل.

م س :متوسط الآراء الاحتمالية للجماعة المرجعية لكل بديل .

ن : عدد بدائل الفقرة .

٥-طريقة الدالة التربيعية Square function method: كذلك في هذه الطريقة يقوم الطالب بتوزيع درجات ثقة أو توقعات احتمالية على جميع بدائل الفقرة، ويكون ذلك في صورة نسب مئوية، ثم يتم تقدير درجة الطالب على الفقرة وفقاً للقانون التالي: (De Finetti,1965)

$$\text{الدرجة التربيعية} = ٢ \times \text{ع}(\text{ن}) - \left[\sum_{\text{س}=١}^{\text{ن}} (\text{ع}(\text{س}))^{٢} \right]$$

حيث :

ع ن : التوقع الاحتمالي أو النسبة المئوية التي أعطاها المفحوص للبديل الصحيح

ع س : التوقعات الاحتمالية أو النسب المئوية التي أعطاها المفحوص لكل بدائل الفقرة.

ن : عدد بدائل الفقرة .

ثالثاً، الطرق اللوغاريتمية Logarithmic Methods : تشترك هذه الطرق في أنها تتطلب استخدام اللوغاريتمات لتقدير درجة الطالب على الفقرة، ومن أهم هذه الطرق ما يلي:

١ـ الطريقة اللوغاريتمية الأولى : في هذه الطريقة يطلب من المفحوص أن يضع تقديرا لمدى ثقته بصحة كل بديل من البدائل بحيث يكون مجموع النسب التي يضعها ١٠٠ ٪، وعند تصحيح الإجابة يتم الاعتماد على النسبة التي اقترحها الطالب للبديل الصحيح فقط ثم يتم حساب الدرجة اللوغاريتمية كما يلي :

☒ $\frac{٢ + \text{لـو}_{١٠} \text{ع ن}}{٢}$ ، في حالة كانت ع ن ≤ ٠,٠١

☒ صفر ، في حالة كانت ع ن > ٠,٠١

٢ ـ الطريقة اللوغاريتمية الثانية : في هذه الطريقة يجيب الطالب باختيار بديل واحد فقط من بين بدائل الفقرة، ثم يقوم بتقدير مدى ثقته بصحة هذا البديل بإعطاء درجة من عشرة أو مئة أو أي نهاية عظمى مناسبة، ففي حالة كان البديل الذي تم اختياره صحيحا فإنه يتم تقدير درجة الفقرة بحساب لوغاريتم درجة الثقة التي حددها الطالب لهذا البديل، إما إذا كان البديل الذي تم اختياره خاطئا فيتم طرح درجة الثقة التي قدرها من النهاية العظمى للثقة، ثم يتم احتساب اللوغاريتم لناتج حاصل الطرح.

٣ـالطريقة اللوغاريتمية الثالثة : تعتمد هذه الطريقة كذلك على تقدير الطالب لثقته بصحة البديل الذي يقوم باختياره، إلا أنه يتم حساب الدرجة وفق القانون التالي :

أ ـ في حالة كون البديل الذي تم اختياره صحيحاً فإن :

$$\text{درجة الفقرة} = ١ - \text{لـو}_{١٠} (١٠ - \text{ث}١)$$

حيث: ث١ : هي تقدير الطالب لصحة البديل من عشرة.

ب. في حالة كون البديل الذي تم اختياره خاطئا فإن:

درجة الفقرة = ١ -لـو (١٠ -ث٢)
١٠

حيث : ث٢ تساوي (١٠ - درجة ثقة الطالب بصحة البديل من عشرة)

(Kansup & Hakstian,1975)

ويلاحظ أن الطريقة اللوغاريتمية الأولى تتطلب قيام الطلبة بتوزيع درجات الثقة على جميع بدائل الفقرة، أما في الطريقتين الثانية والثالثة فإن الطلبة يقومون بإعطاء درجة ثقة واحده تخص بديل وحيد وهو البديل الذي قاموا باختياره كإجابة للفقرة، كما يجب التنبه أنه في كل الطرق السابقة يجب أن يوضح للطالب طريقة التصحيح ومدى تأثرها بزيادة تقدير درجة الثقة إيجابا في حالة كون الإجابة صحيحة وسلبا في حالة كون الإجابة خاطئة، وذلك للحد من تقدير درجات ثقة لا تعكس الواقع الفعلي لثقة الطالب بالبديل الذي اختاره.

رابعاً، طرق التصحيح من أثر التخمين Correction for Guessing Methods: التصحيح من أثر التخمين هو العملية التي يتم بها تعديل علامة الطالب ن في الاختبار الذي تتهيأ فيه الفرصة للطالب أن يخمن الإجابة (عودة،٢٠٠٢، ٢٢٢). ويلجأ الطالب للتخمين عندما لا تكون لديه المعرفة التامة أو الأكيدة بالإجابة الصحيحة (سواقد، ١٩٩٢). ويعد التخمين هو أحد أهم مصادر الأخطاء العشوائية التي من الممكن أن تؤثر على خصائص الاختبار السيكومترية (الهويدي، ٢٠٠٤)، وتقوم معادلة التصحيح من أثر التخمين على افتراض أن التخمين يتم بطريقة عشوائية، بحيث إذا كان لدينا فقرة اختيار من متعدد تحتوى على عدد (ن) من البدائل، وأحد هذه البدائل هو فقط البديل الصحيح، فإن احتمال أن يتوصل الطالب إلى الإجابة دون أن يكون يعرفها هو (Rand,1981)،وهناك ثلاثة طرق رئيسية يمكن خلالها الحد من التخمين العشوائي لدى الطلبة وهذه الطرق هي:

١-الطريقة الأولى (أسلوب العقاب) : تعد هذه الطريقة هي الطريقة التقليدية والأكثر شهرة واستخداما لتصحيح أثر التخمين، وفيها يمنح الطالب درجة عن كل إجابة صحيحة، وصفر عن كل فقرة تم تركها دون إجابة، و يطرح من درجته ١/ (البدائل – ١) عن كل فقرة قام بالإجابة عنها بشكل خاطئ (Andra& Magnano,2011,236)،والافتراض الذي تقوم عليه هذه الطريقة هو أن المفحوص في حالة عدم معرفته للإجابة الصحيحة للفقرة سيختار عشوائياً من بين جميع بدائل الفقرة، وبناء على ذلك فإن استخدام هذه الطريقة يقتضي أن يخصم جزء من العلامة التي يحصل عليها المفحوص مقابل ذلك الجزء من علامته الذي يتوقع أن يكون قد حصل عليها نتيجة استخدامه التخمين العشوائي، وبما أن مقدار ذلك الجزء غير معروف فإنه يتم تقديره من عدد الإجابات غير الصحيحة، بحيث أن كل عدد من الإجابات غير الصحيحة التي تساوي (عدد بدائل الفقرة -١) يقابله إجابة صحيحة واحده حصل عليها المفحوص باستخدام التخمين العشوائي (سواقد،١٩٩٢).

ووفقاً لهذه الطريقة فإن درجة الطالب بعد تعديلها من أثر التخمين يمكن حسابها وفق القانون التالي:

الدرجة المصححة من أثر التخمين = س − خ / (ل − ١)

حيث : س : عدد الفقرات التي أجاب عليها الطالب إجابة صحيحة، خ : عدد الفقرات التي أجاب عليها الطالب إجابة خاطئة، ل: عدد بدائل فقرة الاختيار من متعدد . (Baradaran,2009)، ويلاحظ أن الطالب إذا أجاب بطريقة عشوائية عن جميع الأسئلة ستكون درجته المتوقعة بعد تطبيق هذه المعادلة صفر (Andra& Magnano,2011)

٢-الطريقة الثانية (أسلوب المكافأة) : الفكرة الرئيسية لهذه الطريقة هي تحفيز الطلاب لكيلا يخمنوا عشوائياً، وذلك عن طريق إضافة بعض الدرجات للطالب الذي لم يخمن، ويتم تقدير الدرجات المضافة في ضوء عدد الفقرات التي تركها الطالب دون إجابة، وذلك وفقاً للمعادلة التالية:

$$\text{الدرجة المعدلة} = \text{س} + \frac{\text{م}}{\text{ل}}$$

حيث : س : عدد الفقرات التي أجاب عليها الطالب إجابة صحيحة، م : عدد الفقرات التي تركها الطالب دون إجابة (المحذوفة)، ل : عدد البدائل (عودة، ٢٠٠٢).

٣-الطريقة الثالثة (أسلوب المكافأة والعقاب) : في هذه الطريقة يتم الجمع بين أسلوبي العقاب والمكافأة حيث يعاقب الطالب على الفقرات التي يتوقع انه خمن فيها، ويكافأ على الفقرات التي تركها دون تخمين (الفقرات المحذوفة)، ويتم ذلك وفقا للمعادلة التالية:

$$\text{درجة الطالب المعدلة} = \text{س} + \frac{\text{م}}{\text{ل}} - \frac{\text{خ}}{\text{ل}-١}$$

حيث : س : عدد الفقرات التي أجاب عليها الطالب إجابة صحيحة، خ : عدد الفقرات التي أجاب عليها الطالب إجابة خاطئة، م : عدد الفقرات التي تركها الطالب دون إجابة (المحذوفة)، ل : عدد البدائل (Bush,2001).

الدراســـــات والبحوث المرتبطة:

المحور الأول : الدراسات التي تناولت طرق تقدير الدرجات في اختبارات الاختيار من متعدد:

انطلاقاً من إيمان العديد من الباحثين بوجود بعض أوجه القصور في الطريقة التقليدية المستخدمة في تقدير درجات فقرات اختبارات الاختيار من متعدد، والمتمثل في أنها تشجع الطلبة على التخمين العشوائي، كما أنها لا تسمح بالتمييز بين المستويات المختلفة لدى الطلبة فيما يخص المعرفة التي تقيسها الفقرة الاختبارية (Kurz,1999)، قدمت العديد من الدراسات التي تهدف لاستقصاء عدد من الطرق البديلة في تقدير درجات فقرات اختبارات الاختيار من متعدد، وذلك للكشف عن مدى فاعليتها في تحسين القياس، وزيادة دقته، والرفع من قيم خصائصه السيكومترية، و ومن أوائل هذه الدراسات كانت الدراسة التي أجراها كل من سابرز ووايت Sabers &White (1969) والتي هدفت لبحث أثر طريقة تقدير الدرجات الوزنية لديفز على كلا من الصدق التنبؤي للاختبار وثباته، وذلك بالمقارنة مع طريقة تقدير الدرجات التقليدية، وقد أظهرت النتائج ارتفاع قيمة معامل الثبات بعد استخدام الطريقة الوزنية، إلا أن هذه الزيادات لم تكن ذات دلالة إحصائية ، وكذلك الأمر فيما يتعلق بالصدق التنبؤي فلم يكن هناك أي تأثير ذو دلالة إحصائية يعود لاختلاف طريقة تقدير درجات الاختبار على صدقة التنبؤي. كما أجرى ريبي (1970) Rippeyدراسة كانت بهدف المقارنة بين ثبات درجات خمسة طرق لتقدير الدرجات وهي : طريقة الاحتمال المقترح للإجابة الصحيحة, والطريقة اللوغاريتمية، والطريقة الكروية، والطريقة الإقليدية، وطريقة الاختيار الاستنتاجي، وقد قام الباحث بحساب ثبات درجات كل طريقة وذلك بعد

أن طبق المستوى الاول من اختبار التقدم التعليمي المتسلسل للكتابة وفق كل طريقة من الطرق السابقة وقد أظهرت نتائج الدراسة أن طريقة الاحتمال المقترح للإجابة الصحيحة والتي تعد الأكثر بساطة هي الأعلى في درجة الثبات بمعامل ثبات بلغ (٠,٦٩) ثم الطريقة الإقليدية بمعامل ثبات (٠,٥٨) ثم الطريقة اللوغاريتمة بمعامل ثبات (٠,٥٠) ثم الطريقة الكروية بمعامل ثبات (٠,٤٩) وأخيرا طريقة الاختيار الاستنتاجي بمعامل ثبات(٠,٤٧). كذلك قدم ثومسون Thompson (1976) دراسة جاءت بهدف المقارنة بين الطريقة التقليدية في تقدير درجات فقرات الاختيار من متعدد وبين الطريقة الذاتية والتي يقوم فيها الطالب بتجريب البدائل حتى الوصول للبديل الصحيح Answer-until-correct وذلك من حيث أثر كلا منهما على تحصيل الطلاب، وقلق الطالب، وثبات درجات الاختبار،وقد صمم الباحث أوراق الاختبار بطريقة كيمائية بحيث وضع غطاء كربوني بجوار كل بديل ووضع تحته أما العلامة (0) وذلك عندما يكون البديل خاطئ، أو العلامة (+) وذلك عندما يكون البديل صحيح . وبعد تحليل البيانات أظهرت النتائج أن الطلاب الذين اختبروا بالطريقة الذاتية كانوا أقل قلقا وأكثر تحصيلا من الطلبة الذين اختبروا بالطريقة التقليدية، ألا أن الفروق لم تكن كبيرة، أما فيما يخص ثبات درجات الاختبار فلم تظهر النتائج وجود أية فروق ذات دلالة إحصائية بين المجموعتين في درجة الثبات. كما أجرى (سواقد) ١٩٨٣ دراسة بهدف المقارنة بين أثر ثلاث طرق لتصحيح اختبارات الاختيار من متعدد على الخصائص السيكومترية للاختبار عند مفحوصين من مستويات تحصيل ودرجات مخاطرة مختلفة، وقد استخدمت الدراسة كل من (الطريقة التقليدية، طريقة استخدام معادلة التصحيح لأثر التخمين، طريقة استخدام الاختيار الجزئي)، وذلك بعد أن قام الباحث بإعداد ثلاثة نماذج متكافئة من اختبار يقيس تحصيل الطلاب في مادة العلوم للصف الثالث متوسط حيث تم تطبيق وتصحيح كل نموذج من هذه النماذج وفق تعليمات كل طريقة من الطرق الثلاث، وقد أظهرت النتائج عدم وجود فروق ذات دلالة إحصائية بين درجات المفحوصين ذوي المخاطرة العالية، وذوي المخاطرة المنخفضة عندما تم تصحيح الاختبار بكل من الطريقة التقليدية، طريقة الاختيار الجزئي، في حين كانت هناك فروق ذات دلالة إحصائية بين درجات المفحوصين ذوي المخاطرة العالية، وذوي المخاطرة المنخفضة عندما تم تصحيح الاختبار وفق معادلة التصحيح لأثر التخمين لصالح المفحوصين ذوي المخاطرة العالية، كما أظهرت النتائج أن هناك زيادة ذات دلالة إحصائية في قيمة معامل الثبات عند استخدم طريقتي (الاختيار الجزئي، ومعادلة التصحيح من أثر التخمين) بالمقارنة مع قيمة معامل الثبات عند استخدام الطريقة التقليدية، وفي المجمل أشارت نتائج الدراسة إلى ان طريقة الاختيار الجزئي أفضل من كل من الطريقة التقليدية، وطريقة استخدام معادلة التصحيح من أثر التخمين، وقد فسر الباحث ذلك بتميز هذه الطريقة بأنها تسمح للمفحوص باستغلال معرفته الجزئية، وبقدرتها على ضبط أثر التخمين . كما قدم العدل (١٩٨٦) دراسة بهدف الكشف عن أثر بعض طرق تقدير الدرجات للفقرات على ثبات الاختبارات ذات الاختيار من متعدد، وذلك من خلال الإجابة على التساؤلات التالية: هل تؤثر الطرق المختلفة لتقدير الدرجات على ثبات الاختبار؟، هل يتوقف تأثير الثبات بالطرق المختلفة لتقدير الدرجات على طريقة حساب الثبات؟ هل يتوقف تأثير الثبات بالطرق المختلفة لتقدير الدرجات على نوع الاختبار؟ وقد أظهرت نتائج الدراسة المتعلقة بالاختبار التحصيلي عدم وجود فروق ذات إحصائية بين طريقتي حساب الثبات (معادلة سبيرمان ـ بروان،معادلة ألفا كرونباخ) وكذلك بين طريقتي حساب الثبات (معادلة ألفا كرونباخ وإعادة الاختبار) وذلك بالنسبة لجميع طرق تقدير الدرجات المستخدمة في الدراسة، في حين أظهرت النتائج وجود فروق ذات دلالة إحصائية بين طريقتي حساب الثبات (معادلة سبيرمان ـ بروان، إعادة الاختبار) في طريقتين من طرق تقدير الدرجات هما (طريقة الاحتمال المقترح للإجابة الصحيحة والطريقة اللوغاريتمية)، أما النتائج المتعلقة باختبار التفكير الاستنباطي فأظهرت

أنه لا توجد فروق ذات دلالة إحصائية بين طريقتي حساب الثبات (معادلة سبيرمان- براون، ومعادلة كرونباخ) ووجود فروق ذات دلالة إحصائية بين طريقتي حساب الثبات (معادلة سبيرمان – بروان، وإعادة الاختبار) وكذلك بين طريقتي (معادلة كرونباخ و إعادة الاختبار) وذلك لجميع طرق تقدير الدرجات المستخدمة في الدراسة، وعند المقارنة بين نوعي الاختبار أظهرت النتائج أن تأثر ثبات الاختبار بالطرق المختلفة لتقدير الدرجات في الاختبار التحصيلي يختلف عنه في اختبار التفكير الاستنباطي. كذلك أجرى محمود (١٩٨٦) دراسة بهدف تقصى أثر بعض طرق تقدير الدرجات على ثبات الاختبارات مرجعية المحك ذات الاختيار من متعدد، حيث قام الباحث بإعداد صورتين متكافئتين من اختبار تحصيلي مرجعي المحك في مادة الجبر للصف التاسع وقام بتطبيق كل صورة من صور الاختبار على عينة مكونة ٥٨٠ طالب، وبعد جمع البيانات وتحليلها أظهرت الدراسة أن الثبات المحسوب يختلف باختلاف الطريقة المستخدمة لحساب الثبات، كما أظهرت وجود تأثير مشترك بين طرق تقدير الدرجات وطرق حساب ثبات الاختبارات مرجعية المحك، فقد كشفت النتائج أنه عند استخدام طريقة تقدير الدرجات التقليدية كانت طريقة لفنجستون لحساب الثبات كانت الأفضل فيما لم يكن هناك أية فروق بين طريقة هارس وطريقة هاميلتون والجينا لتقدير معامل ثبات الاختبارات مرجعية المحك، أما عند استخدام طريقة الدرجات الذاتية، فظهر أنه لا توجد فروق بين طريقة لفنجستون وطريقة هارس في تقدير معامل الثبات، فيما كانت طريقة هارس أفضل من طريقة هاميلتون والجينا، وعند استخدام طرق وزن الثقة لتقدير الدرجات فقد أظهرت النتائج أن عند استخدام طريقتي الاحتمال المقترح، والطريقة اللوغاريتمية، فلم يكن هناك أية فروق بين طريقة لفنجستون، وطريقة هارس وطريقة هاميلتون والجينا، في تقدير معامل الثبات، أما عند استخدام طريقة توزيع الثقة فقد أظهرت النتائج أنه لا توجد فروق بين طريقة لفنجستون، وطريقة هارس في حساب الثبات، فيما كانت طريقة لفنجستون أفضل من طريقة هاميلتون والجينا، وطريقة هارس أفضل من طريقة هاميلتون والجينا.كما قام سواقد (١٩٩٢) بدراسة هدفت لاختبار صحة الافتراضات النظرية التي تقوم عليها بعض الطرق المقترحة لتصحيح اختبارات الاختيار من متعدد، والمقارنة بين أثر استخدام هذه الطرق على الخصائص السيكومترية للفقرات والمتمثلة بكل من : معامل صعوبة الفقرة، معامل تمييزها، مؤشر ثبات الفقرة، ومؤشر صدقها، وقد تناولت الدراسة أربعة طرق لتقدير درجات الاختيار من متعدد هي (الطريقة التقليدية، معادلة التصحيح لأثر التخمين، طريق الاختيار الجزئي، وطريقة استبعاد البدائل غير الصحيحة)، ولتحقيق أهداف الدراسة قام الباحث باستخدام اختبارين متكافئين في مادة أساليب القياس أحدهما مقالي والآخر من نوع الاختيار من متعدد وقدم الاختبارين وفق نموذج تجريبي على عينة من طلبة كلية تأهيل المعلمين العالية في أربد وذلك بعد أن قام بتقسيم الطلبة إلى أربع مجموعات متكافئة ثم قدم الاختبار المقالي لجميع طلبة هذه المجموعات الأربع بطريقة متشابهة أما اختبار الاختيار من متعدد فقد قدمه لكل مجموعة وفق تعليمات تختلف عن المجموعات الأخرى وفقا لاختلاف طريقة تقدير الدرجات (التصحيح) الخاصة بكل مجموعة، وقد أظهرت نتائج الدراسة المتعلقة بصحة الافتراضات النظرية، عدم صحة الافتراض النظري الذي تقوم عليه معادلة التصحيح من أثر التخمين، وصحة الافتراض النظري الذي تقوم عليه كل من طريقة : الاختيار الجزئي، واستبعاد البدائل غير الصحيحة، أما النتائج المتعلقة بأثر استخدام الطرق الأربع على الخصائص السيكومترية للفقرات فأظهرت أن أعلى قيمة لمعامل صعوبة الفقرات كانت عند استخدام طريقة تقدير الدرجات التقليدية، فيما جاءت أقل قيمة لمعامل الصعوبة عند استخدام معادلة التصحيح لأثر التخمين، وجاءت أعلى قيم للإحصائيات (معامل التمييز، ومؤشر الثبات، ومؤشر صدق الفقرات) عند استخدام طريقتي الاختيار الجزئي، واستبعاد البدائل فيما كانت أقل قيم لهذه الإحصائيات عند استخدام معادلة التصحيح لأثر التخمين، وقد فسر الباحث هذه النتائج بأنه عند

استخدام كل من الطريقتين (الاختيار الجزئي، واستبعاد البدائل) فإن تعليمات الاختبار وسلم الدرجات الذي يشكل متصل يمتد من (-٣ الى +٣)، يسمح للمفحوص باستخدام معرفته الجزئية، ويمنحه درجة تتناسب مع ما لديه من هذه المعرفة، وفي الوقت نفسه يحد من التخمين العشوائي وهو الأمر الذي لا يتوفر سواء عند استخدام الطريقة التقليدية أو عند استخدام معادلة التصحيح. كذلك أجرى عبد الوهاب (٢٠٠١) دراسة هدفت للمقارنة بين عدة طرق لتقدير الدرجات على صدق اختبار مرجعي المحك من نوع الاختيار من متعدد، وهذه الطرق هي : الطريقة التقليدية، وطريقة الاحتمال المقترح للإجابة الصحيحة، و طريقة الاختيار الاستنتاجي، والطريقة التربيعية، والطريقة التجريبية، وقد أظهر نتائج الدراسة وجود فروق ذات دلالة إحصائية بين قيم معاملات صدق الاختبار تعود لاختلاف طريقة تقدير الدرجات المتبعة، حيث كان الاختبار المصحح وفق الطريقة التجريبية والبالغة (٠,٩١٧) أعلى في قيمة معامل الصدق من كلا من الطريقة التقليدية والبالغة (٠,٨٣٨)، وكذلك الطريقة التربيعية (٠,٣٣١)، كذلك كانت قيمة معامل صدق الاختبار عند تصحيحه وفق طريقة الاحتمال المقترح (٠,٨٩٥) أعلى منها عند تصحيحه وفق الطريقة التقليدية (٠,٨٣٨)،كذلك كانت هناك فروق بين طريقتي الاختيار الاستنتاجي (٠,٩٦٩) والطريقة التقليدية لصالح طريقة الاختيار الاستنتاجي، والتي كانت كذلك أفضل في قيم معامل الصدق من كلا من الطريقة التجريبية، وطريقة الاحتمال المقترح، والطريقة التربيعية (٠,٣٣١). وفي ضوء نتائج الدراسة أوصى الباحث باستخدام بعض طرق تقدير الدرجات الوزنية التي أسهمت في زيادة صدق الاختبار مقارنة بالطريقة التقليدية. كما قدم دودين (٢٠٠٣) دراسة قام فيها باقتراح معادلة جديدة يمكن استخدامها للتصحيح درجات الاختبار من أثر التخمين وذلك في اختبارات الاختيار من متعدد، وتم فيها مراعاة مستوى جاذبية البديل عند حساب درجة السؤال النهائية، كما قام الباحث بالمقارنة بين هذه المعادلة والمعادلة التقليدية المستخدمة لتصحيح أثر التخمين، من حيث تأثير كلا منهما على تقدير مستويات الطلاب، وعلى ثبات درجات الاختبار، وقد أظهرت النتائج أن المعادلة المقترحة كانت لها أفضلية في مكافئة الطلبة على حسن اختيارهم للبديل الأقرب للصواب والأكثر صحة كما أن الدرجات كانت أعلى في قيمة الثبات عند استخدام هذه المعادلة مقارنة بالمعادلة التقليدية . كذلك قامت اللحياني (٢٠٠٩) بدراسة جاءت بهدف الكشف عن أثر بعض طرق تقدير الدرجات (التقليدية، التجريبية، الاحتمالية، المكافأة)على ثبات وصدق درجات اختبار تحصيلي في الرياضات ذي الاختيار من متعدد لدى طالبات الصف الأول الثانوي بمكة المكرمة،، وقد أظهرت نتائج الدراسة أن كلا من الطريقة التقليدية والاحتمالية كانتا هما الأعلى في قيم ثبات درجات الاختبار، فيما كانت طريقة المكافأة هي الأقل في قيمة الثبات، كما أظهرت النتائج أن الصدق التلازمي كان هو الأعلى قياسا بالصدق التنبؤي وذلك لجميع طرق تقدير الدرجات المستخدمة في الدراسة. كذلك أجرى لاو وآخرون (Lau et al.(2011 دراسة على المجتمع الماليزي بهدف الكشف عن فاعلية تطبيق برنامج حاسوبي لتقدير درجات فقرات اختبارات الاختيار من متعدد يستخدم طريقة لتقدير الدرجات تجمع بين أسلوب استبعاد البدائل الخطأ والأسلوب التقليدي لتقدير الدرجات (عدد الإجابات الصحيحة) Number Right Elimination Testing (NRET) وذلك من خلال الإجابة على الأسئلة التالية : ما درجة استخدام الطلبة للتخمين عند تطبيق طريقة تقدير الدرجات المتبعة بالدراسة (NRET) ؟، ما مدى كفاءة طريقة تقدير الدرجات (NRET) في الكشف عن المعرفة الجزئية ؟ ما مدى كفاءة طريقة تقدير الدرجات (NRET) في الكشف عن سوء الفهم في المعرفة لدى الطلبة؟، هل هناك فرق في قيمة ثبات الاختبار عند استخدام الطريقة التقليدية لتقدير الدرجات وطريقة (NRET) المستخدمة في الدراسة؟، وقد أظهرت نتائج الدراسة أنه عند استخدام طريقة تقدير الدرجات (NRET) كانت درجات التخمين أقل ما يمكن، كما أمكن من خلال استخدام هذه الطريقة الكشف عن معرفة الطلبة الجزئية، وفهمهم الخاطئ وذلك لكل فقرة

من فقرات الاختبار، كذلك أظهرت النتائج أن ثبات درجات الاختبار كان أعلى عند استخدام طريقة تقدير الدرجات (NRET) مقارنة بثبات درجاته عند استخدام الطريقة التقليدية. كما أجرى لسيج و آخرون (2013) .Lesage et al دراسة جاءت بهدف تقديم عرض شامل لأهم طرق تقدير درجات الاختيار من متعدد المستخدمة في التعليم العالي وذلك وصولا إلى تحديد نقاط ضعف وقوة كل منها، ولحصر هذه الطرق قام الباحثون بمسح الدراسات تناولت طرق تقدير الدرجات منذ عام ١٩٧٩ لحين إجراء البحث، وقد تم حصر أهم الطرق التي يشيع تناولها في هذه الدراسات والمقارنة بينها من خلال ستة معايير هي : (القاعدة التي يتم في ضوئها تقدير الدرجات، طريقة الإجابة، الثبات والصدق، عامل التخمين، مراعاة المعرفة الجزئية، القضايا التعليمية ذات علاقة)، وقد خرجت الدراسة بعدد من النتائج كان من أهمها ما يلي : بالنسبة للطريقة التقليدية في تقدير الدرجات فقد أشارت الدراسات إلى تدنى ثبات درجاتها بسبب تدخل عامل التخمين، كما يعاب عليها عدم مراعاتها لمعرفة الطالب الجزئية، وأنها تشجع الطلاب على التخمين وتكافئهم عليه، أما طريقة التصحيح من أثر التخمين عن طريق العقاب فقد اشارت الدراسات إلى وجود زيادة طفيفة في معاملي ثبات وصدق الاختبارات التي تقوم عليها إلا أن نتائج الدراسات لم تكن متفقه حول هذه الزيادة، ويؤخذ على هذه الطريقة عدم مراعاتها لمعرفة الطالب الجزئية، وعدم تقبل الطلاب لها، كما أنها لم تمنع الطلاب من التخمين بل أدخلت عاملاً أخراً يمكنه أن يؤثر على تباين الدرجات وهو سلوك التخمين لدى الطالب، أما طريقة التصحيح من أثر الخمين عن طريق المكافأة فقد جاءت نتائج الدراسات متعارضه حول خصائصها السيكومترية (الثبات والصدق)، ويؤخذ عليها أنها لا تراعي المعرفة الجزئية للطالب، وكذلك تأثر تباين الدرجات فيها بسلوك التخمين لدى الطلاب، إلا أن هذه الطريقة كانت أكثر مناسبة للطلاب من طريقة التصحيح لأثر التخمين عن طريق العقاب، أما طرق تقدير الدرجات التي تراعي الاتقان الجزئي (طريقة الحذف، الطريقة التجريبية أو الاختيار الجزئي، طريقة مستويات الثقة) فقد جاءت نتائج الدراسات غير متسقة حول خصائصها السيكومترية إلا أن هذه الطرق تتميز عن سابقاتها بمراعاتها لمعرفة الطالب الجزئية، فيما تشترك مع طرق التصحيح لأثر التخمين في إدخال عامل جديد يمكن أن يؤثر على تباين الدرجات وهو سلوك التخمين لدى الطلاب، كما يؤخذ عليها أن تطبيق اختباراتها وتصحيحها أكثر تعقيدا من الطريقة التقليدية.

المحور الثاني : الدراسات التي قارنت بين الاختبارات التي تستند إلى النظرية الكلاسيكية والاختبارات التي تستند إلى نظرية الاستجابة للمفردة،

أجريت العديد من الدراسات التي حاولت المقارنة بين المدخلين النظريين ومن أهم هذه الدراسات دراسة فان (1998) Fan والتي أظهرت أنه على الرغم من الاختلافات النظرية الواضحة بين النظرية الكلاسيكية، ونظرية استجابة الفقرة إلا أن الفروق العملية بينهما لا تزال بحاجة إلى مزيد من الدراسة، حيث بحثت هذه الدراسة المقارنة التجريبية بين النظريتين في خصائص كل من فقرات الاختبار، وقدرات الطلاب، وفي ثبات إحصائيات الفقرة عبر عينات المشاركين المختلفة، وقد أظهرت نتائج الدراسة أن إحصائيات الأفراد (مستويات قدرات الطلاب) المستمدة من النظرية الكلاسيكية متماثلة بدرجة كبيرة مع إحصائيات الأفراد المستمدة من نماذج نظرية الاستجابة للمفردة الثلاثة، كذلك جاءت مؤشرات صعوبة الفقرة المستمدة من النظريتين متماثلة بدرجة كبيرة. كما قام ستيج (1999) Stage بدراسة هدفت للمقارنة بين معاملات صعوبة وتمييز الفقرات المستمدة من النظرية الكلاسيكية وتلك المستمدة من نظرية الاستجابة للمفردة وذلك عن تطبيق اختبار قبلي، وعند تطبيق اختبار بعدي، وكذلك مقارنة مدى استقرار هذه المؤشرات الإحصائية في الاختبار القبلي، والبعدي وذلك في كل من الإطارين النظريين، وقد أظهرت نتائج الدراسة وجود توافق بين النتائج المستمدة من الإطارين النظريين بدرجة جيدة جدا، الا ان التوافق في تقدير مؤشرات

الصعوبة كان أكبر منه عند تقدير مؤشرات التميز. كذلك أجرى كانتريل Cantrell (1997) دراسة أستخدم فيها تحليل الانحدار لإيجاد العلاقة بين تدريج قدرات الأفراد وفق نموذج راش الأحادي وبين تدريج قدرات الأفراد وفق النظرية الكلاسيكية، ولرسم مخطط الانتشار وخط الانحدار لتوزيع قدرات الأفراد وفق المدخلين النظريين، وقد جاءت النتائج متحدية التوجه النظري القائل بتفوق نموذج راش الأحادي على المدخل الكلاسيكي بسبب تحرر تقدير القدرات وفق هذا النموذج من خصائص كل من الأفراد، والفقرات، حيث أظهرت النتائج إلى أن معامل الارتباط بين تدريج القدرات وفق نموذج راش ودرجات الافراد وفق النظرية الكلاسيكية كان مرتفعاً جداً حيث بلغت قيمته (٠,٩٨٥) وهذا يشير إلى أن تقدير القدرة وفق نموذج راش لم يكن بالفعل متحرراً من خصائص الأفراد والفقرات، أو أن تدريج القدرات وفق المدخل الكلاسيكي له نفس القدر من التحرر. كما أجرى كل من نداليتشاكو وروجرز Ndalichako&Rogers (1997) دراسة بهدف المقارنة بين ثلاث نظريات لتقدير قدرات الطلاب هي النظرية الكلاسيكية، ونظرية الاستجابة للمفردة بنماذجها الأحادي والثنائي والثلاثي، ونظرية الدرجة المحددة، وقد تمت المقارنة من خلال دراسة: (أ) ـ ما إذا كانت النماذج الخمسة (النماذج اللوغاريتمية الثلاثة، والنموذج الكلاسيكي، ونموذج نظرية الدرجة المحددة) ترتب قدرات الطلاب بطريقة متشابهة، (ب) ـ مدى التقارب أو بين الدرجات المقدرة بكل نموذج من النماذج الخمسة مع الدرجات المقدرة من بقية النماذج، وقد أظهرت النتائج وجود تشابه كبير بين النماذج الخمسة في ترتيب الطلاب وفق قدراتهم، إلا أن أكثر النماذج اتفاقا مع نموذج النظرية الكلاسيكية هو النموذج اللوغاريتمي الثلاثيPL3، فيما كان النموذج الأقل اتفاقا مع نموذج النظرية الكلاسيكية هو نموذج نظرية الدرجة المحددة كما أظهرت النتائج أن النموذجين الثلاثي والثنائيPL2& PL3 كانا أكثر النماذج الخمسة اتفاقا في تقدير الدرجات. كذلك قامت جمحاوي (٢٠٠٠) بدراسة هدفت للمقارنة بين النظرية التقليدية ونظرية الاستجابة للمفردة في أربعة متغيرات هي (صعوبة الفقرات، تمييز الفقرات، تدريج القدرات، معامل الثبات)، وأظهرت نتائج الدراسة وجود اتفاق عالي بين النظريتين في تقدير صعوبة الفقرات، في حين كان الاتفاق بين المدخلين متدنيا عند تقدير القوة التمييزية للفقرات وخاصة عند المقارنة بالنموذج اللوغاريتمي الثلاثي والذي بلغت درجة ارتباط مؤشرات التمييز المستمدة منه بنظرياتها المستمدة وفق بالنظرية التقليدية (٠,٢٦) والتي تدل على ارتباط بدرجة ضعيفة، أما فيما يخص توزيع قدرات المفحوصين فقد أظهرت النتائج توافقاً كبيراً بين النظريتين. كما جاءت دراسة ستيج Stage (2003) متسقة إلى حد ما مع نتائج دراسة فان Fan السابقة، حين بحثت ما إذا كانت نماذج نظرية الاستجابة للمفردة قابلة للتطبيق في اختبار الاستعداد المدرسي السويدي المستخدم لقبول الطلبة لمواصلة التعليم العالي في السويد، وقد أظهرت نتائج الدراسة أن مناسبة بيانات الاختبار للنموذج البارامتري كانت محل شك، خاصة عند تطبيق نموذج النظرية الحديثة على الاختبار ككل، كما أظهرت أن تطبيق نماذج النظرية الحديثة لم يقدم أية إضافة لجودة الاختبار. كذلك أجرى وايبيرج Wiberg (2004) دراسة جاءت بهدف معرفة أي من نماذج نظرية الاستجابة للمفردة اللوجستية (الأحادي PL1، أو الثنائي PL2، أو الثلاثي PL3) هو الأكثر ملائمة لاستخدامه في الاختبار النظري لرخص القيادة في السويد، وقد خلصت النتائج إلى أن النموذج اللوجستي الثلاثي المعلم PL3 كان هو الأفضل والأكثر كفاءة لاستخدامه في الاختبار النظري للقدرة على القيادة، وفيما يخص المقارنة بين المؤشرات التي تم الحصول عليها من تحليل الاختبار وفق النظرية الكلاسيكية والحديثة تم التوصل إلى أنه كلا منهما أعطت معلومات قيمة يجب الاستفادة منها وتضمنيها عند تحليل الاختبار النظري لرخص القيادة في السويد. أما في الدراسة التي قدمها كل من اينول ونور Ainol&Noor (2006) فقد كانت نتائجها مؤكدةً تفوق نماذج نظرية الاستجابة للمفردة لاستخدامها، حيث أظهرت النتائج أنه لا يمكن الاعتماد فقط على

المدخل الكلاسيكي في تحليل فقرات الاختبار لأنه لا يمدنا بمعلومات عن الطلاب ذوي المستويات المختلفة في القدرة وهو الأمر الذي أمكن توفره بعد تحليل الفقرات وفق نظرية استجابة الفقرة. كما قدمت الحكماني (٢٠٠٧) دراسة بهدف المقارنة بين النظرية الكلاسيكية للاختبار والنموذج اللوغاريتمي ثنائي البارامتر من حيث تقدير مستويات قدرة الطلاب وفقا لأدائهم في اختبار المصفوفات العادية لرافن، وكذلك من حيث مدى استقرار المؤشرات الإحصائية للفقرات الاختبارية والمتمثلة في صعوبة الفقرات وتمييزها وذلك باختلاف الصفوف الدراسية، وأظهرت نتائج الدراسة بأن هناك تشابهاً كبيراً في ترتيب درجات الطلاب المقدرة وفقا للمدخلين الكلاسيكي والنموذج ثنائي البارامتر، إلا أن قيم درجات الطلاب وفق النموذج الكلاسيكي تختلف بشكل كبير عن القيم المقدرة وفق النموذج الثنائي، كما أظهرت النتائج أن المؤشرات الإحصائية للفقرات الاختبارية المستمدة من النموذج الثنائي كانت أكثر استقرارا من تلك المقدرة باستخدام النظرية الكلاسيكية. وقام عباينة (٢٠٠٨) بدراسة هدفت إلى استقصاء اللاتغير في تقدير إحصائيات الفقرة باستخدام كل من نظرية الاستجابة للمفردة والنظرية الكلاسيكية في القياس وقد أظهرت نتائج الدراسة أن افتراض اللاتغير في تقدير معلمتي الصعوبة والتمييز قد تحقق عند استخدام عينات ذات قدرات متقاربة وذلك عند استخدام النموذج اللوغاريمي اللوجستي الثلاثي المعلمة، في حين لم يتحقق عند استخدام نظرية القياس الكلاسيكية، أما افتراض اللاتغير لتقديرات معلمة التخمين فلم يتحقق. كما أجرى كل من بروجر وسوين Progar & Soean (2008) دراسة جاءت بهدف المقارنة بين النظرية الكلاسيكية والنماذج اللوغاريتمية الأحادي والثنائي والثلاثي وذلك من حيث مدى استقرار إحصائيات الأفراد عبر مجموعات مختلفة من الأفراد، وكذلك مدى استقرار إحصائيات الفقرات عبر مجموعات مختلفة من الأفراد، وقد أظهرت نتائج الدراسة إلى وجود تشابه كبير في درجة اختلاف بارامترات الفقرات عبر مجموعات مختلفة من الأفراد وذلك عند استخدام النظريتين الكلاسيكية أو نظرية الاستجابة للمفردة، كذلك هناك تشابه كبير في درجة اختلاف تقديرات الأفراد عبر مجموعات مختلفة من الفقرات في ضوء كل من النظريتين، إلا أن النتائج أظهرت وجود اختلافات كبيرة في قيم إحصائيات الأفراد والفقرات في ضوء النظريتين. كذلك قام أددوين Adedoyin (2010) بدراسة تجريبية بهدف الكشف عن مدى ثبات تقديرات معالم الأفراد (قدرات الأفراد) استنادا إلى كل من المدخل الكلاسيكي في القياس ومدخل نظرية استجابة الفقرة، وقد أظهرت نتائج الدراسة أن هناك فروقا ذات دلالة احصائية في ثبات تقدير معالم الأفراد لصالح النموذج الثنائي المعلم، كما أظهرت النتائج تدني ثبات تقدير معالم الأفراد عند تقديرها وفق المدخل الكلاسيكي في القياس، وفي ضوء هذه النتائج اوصت الدراسة باستخدام نظرية الاستجابة للمفردة في بناء وتطوير الاختبارات. كما أجرى عبد الوهاب (٢٠١٣) دراسة بهدف المقارنة بين النظرية الكلاسيكية والنموذج الثلاثي البارامتر في معالجة أثر التخمين عند تقدير إحصائيات الفقرات (بارامتري الصعوبة والتمييز) وإحصائيات الأفراد المتمثلة في قدرات عينة من طلاب الفرقة الرابعة والدبلوم العام بكلية التربية بالمنيا على اختبار تحصيلي من نوع الاختيار من متعدد في مادة القياس النفسي، وذلك من خلال معرفة الفروق بين هذه الإحصائيات في ضوء النظرية الكلاسيكية في القياس قبل وبعد تصحيح أثر التخمين، وفي ضوء النموذج الثلاثي البارامتر، وأوضحت النتائج عدم وجود اختلاف في تقدير احصائيات الأفراد والفقرات في ضوء النظرية الكلاسيكية قبل وبعد تصحيح أثر التخمين، ولكن وجد اختلاف واضح في هذه الإحصائيات في ضوء النموذج الثلاثي البارامتر قبل وبعد تصحيح أثر التخمين، مما يشير إلى اختلاف كل منهما في معالجة أثر التخمين عند تقدير الإحصائيات. كذلك قدم أيهان Ayhan (2015) دراسة اهتمت بالمقارنة بين المدخلين الكلاسيكي ونظرية استجابة الفقرة وذلك من خلال بحث مدى إمكانية استخدام الاختبار المحوسب التكيفي Computerized Adaptive Testing (CAT) كبديل لاختبار الورقة والقلم الذي يقوم

على المدخل الكلاسيكي والمستخدم لقبول الطلاب في التعليم العالي في تركيا. وأظهرت نتائج الدراسة وجود ارتباط دال وقوي بين درجات الطلاب في الاختبار المحوسب التكيفي، واختبار الورقة والقلم، وكذلك كان توزيع درجات القدرة متشابه لكل من الاختبارين، إلا أنه في المقابل كان هناك انخفاض كبير في عدد الأسئلة المستخدمة في الاختبار المحوسب مقارنة باختبار الورقة والقلم، حيث أكدت الدراسة أنه يمكن الحصول على نفس درجة ثبات الاختبار في القياس الكلاسيكي لكن بعدد أقل من الأسئلة عند استخدام الاختبار المحوسب التكيفي CAT، وأوصت الدراسة باستخدام الاختبار المحوسب بديلا لاختبار الورقة القلم لاختيار طلبة التعليم العالي في تركيا.

ويتبين من قراءة الدراسات السابقة التي قارنت بين طرق تقدير الدرجات الكلاسيكية أنها تباينت في معايير المقارنة بين هذه الطرق فمثلا استخدمت دراسات (Rippey،1970)، (Thompson,1976)، (واللحياني،٢٠٠٩) معيار معامل الثبات، أما دراسة (عبدالوهاب،٢٠٠١) فاستخدمت معيار معامل الصدق، ودراسة(سواقد،١٩٨٣) استخدمت معاملي الثبات والصدق، ودراسة (سواقد،١٩٩٢) استخدمت عدة معايير منها معامل التمييز، مؤشر الثبات، مؤشر الصدق، إلا أن جميع هذه الدراسات كانت تقوم بالمقارنات حسب المدخل الكلاسيكي في القياس ولم تتطرق أي منها لمعيار درجة الارتباط بأي من نماذج نظرية الاستجابة للمفردة. كما أن هذه الدراسات تباينت نتائجها في الكشف عن أفضل هذه الطرق فمثلا أشارت دراسة (Rippey ,1970) إلى أفضلية طريقة الاحتمال المقترح للإجابة الصحيحة وإلى تدني أفضلية الاختيار الاستنتاجي، في المقابل نجد أن دراسة (اللحياني، ٢٠٠٩) توصي باستخدام طريقة الاختيار الاستنتاجي بالإضافة للطريقة التقليدية كأفضل الطرق التي تناولتها الدراسة، كذلك خرجت دراستي (سواقد،١٩٨٣) و(سواقد، ١٩٩٢) بالتوصية باستخدام الطريقة التجريبية (الاختيار الجزئي) مقارنة بالطريقة التقليدية وهو عكس ما تم التوصل إليه في دراسة (اللحياني، ٢٠٠٩) والتي اشارت إلى أفضلية الطريقة التقليدية على الطريقة التجريبية وذلك وفق نفس المعيار (ثبات الاختبار)، كما تباينت نتائج دراسي (سواقد،١٩٨٣) و(عبدالوهاب،٢٠٠١) وفي أفضلية الطريقتين التقليدية والتجريبية وفق معيار صدق الاختبار، وهذا التباين في النتائج يؤكد على الحاجة إلى إجراء المزيد من الدراسات للكشف عن أفضيلة أي من هذه الطرق.

وبالرجوع إلى الدراسات التي قارنت بين النظرية الكلاسيكية ونظرية الاستجابة للمفردة، نجد أنها كذلك أظهرت تباينا في نتائجها ففي حين أشارت نتائج دراسات (Hambleton& Jones 1993)، و(عباينة، ٢٠٠٨) و(Adedoyin,2010) إلى تحرر إحصائيات الفقرة والأفراد من خصائص العينة والفقرات عند استخدام نظرية الاستجابة للمفردة نجد في المقابل أن هناك دراسات شككت في هذه النتيجة كدراسة (Cantrell, 1997)والتي أشارت إلى أن المدخلين لهما نفس القدر من التحرر، ودراسة (fan,1998) والتي أشارت للتشابه الكبير بين المدخلين في ثبات الإحصائيات. وكذلك نجد أن عدد من الدراسات أوصت باستخدام مدخل نظرية الاستجابة للمفردة في تطوير الاختبارات وتقدير قدرات الطلاب كدراسة (Ainol&Noor,2006)، ودراسة (Adedoyin,2010)، ودراسة (Ayhan,2015)، ويقابل ذلك هناك عدد من الدراسات التي شككت في جدوى ذلك كدراسة (Ndalichako& Rogers,1997)،ودراسة (fan,1998)، ودراسة (Stage,1999)، وهذا التباين كذلك يدعو لإجراء المزيد من الدراسات العملية المقارنة بين هذين المدخلين. كما يلاحظ أن الدراسات التي قارنت بين النظرية الكلاسيكية ونظرية الاستجابة للمفردة وفق متغير معامل الارتباط بين إحصائيات الأفراد أو الفقرات كدراسات (Cantrell,1997)، و(Stage,1999)، و(جمحاوي، ٢٠٠٠) اقتصرت على النموذج التقليدي في تقدير الدرجات باستثناء دراسة (Ndalichako&Rogers,1997) والتي استخدمت نموذج الدرجة المحددة Finite State Score بالإضافة إلى النموذج التقليدي في تقدير الدرجات.

وفي ضوء ما سبق يلاحظ أن الدراسات السابقة التي قارنت بين طرق تقدير الدرجات الكلاسيكية تباينت في نتائجها وفي طرق تقدير الدرجات التي تناولتها، وفي المعايير التي استخدمتها للحكم بأفضلية أي من هذه الطرق، إلا أنها لم تتطرق لمعيار الارتباط بنماذج نظرية الاستجابة للمفردة، وكذلك نجد أن الدراسات التي قارنت بين المدخلين النظريين (الكلاسيكي، والاستجابة للفقرة) تباينت في نتائجها وفي المتغيرات التي درستها إلا ان هذه المقارنات كانت تتم باستخدام الطريقة التقليدية لتقدير الدرجات باستثناء دراسة (عبد الوهاب، ٢٠١٣) والتي اضافة طريقة التصحيح من أثر التخمين، ودراسة (Ndalichako&Rogers,1997) والتي أضافت نموذج الدرجة المحددة، إلا أن أي منهما لم تتطرق لطرق تقدير الدرجات التي تناولتها الدراسة الحالية. لهذا جاءت الدراسة الحالية لإكمال البناء المعرفي وذلك بإضافة متغير لم تتطرق له هذه الدراسات وهو درجة ارتباط إحصائيات الأفراد والفقرات المستمدة من ثلاثة طرق من طرق تقدير الدرجات بنظائرها المستمدة من النموذج اللوغاريتمي الثلاثي المعلم، وفي نفس الوقت تعد الدراسة الحالية امتداد للدراسات التي قارنت بين المدخلين النظريين (الكلاسيكي والاستجابة للفقرة) وذلك من خلال الكشف عن درجة ارتباط تقديرات قدرات الأفراد وإحصائيات الفقرات المستمدة من كل من هذين المدخلين، وبإضافة نموذجين لم تتطرق لهما الدراسات السابقة وهما (الطريقة التجريبية، وطريقة الاحتمال المقترح للإجابة الصحيحة) .

منهج وإجراءات الدراسة:

أولاً: منهج الدراسة: يعرف عبيدات (٢٠٠٣، ٢٤٧) المنهج الوصفي بأنه ذلك المنهج العلمي الذي يعتمد على دراسة الواقع أو الظاهرة كما توجد في الواقع ويهتم بوصفها وصفا دقيقا ويعبر عنها تعبير كيفياً أو تعبيراً كمياً، ويضيف العساف (٢٠١٢، ١٧٧) أن المنهج الوصفي هو أطار عام تقع تحته كل البحوث التي تصف الظاهرة فقط أو التي توضح العلاقة بين المتغيرات أو تهدف لاكتشاف الأسباب الكامنة وراء سلوك معين من معطيات سابقة،وحيث أن الدراسة الحالة تهدف لدراسة الظاهرة كما هي في الواقع والتعبير عنها كميا وكيفيا، وذلك من خلال حساب معاملات الارتباطات بين عدد من المؤشرات الإحصائيات (إحصائيات الفقرة، والأفراد) المستمدة من طرق تقدير الدرجات الكلاسيكية(الطريقة التقليدية، والتجريبية، والاحتمال المقترح للإجابة الصحيحة) وبين المؤشرات الإحصائية المناظرة لها والتي تم حسابها باستخدام النموذج اللوغاريتمي الثلاثي المعلم (معالم الفقرة والأفراد)، ومحاولة تفسير قيم هذه المعاملات في ضوء الإطار النظري والدراسات السابقة لذا فإن منهج الدراسة الذي تم استخدامه هو المنهج الوصفي .

ثانياً: مجتمع الدراسة: تكون مجتمع الدراسة من جميع طلبة الصف الثالث متوسط (بنين) في مدراس مدينة عرر المتوسطة الحكومية والبالغ عددها ٢٨ مدرسة ضمن إطار إدارة تعليم منطقة الحدود الشمالية بالمملكة العربية والسعودية، في العام الدراسي ١٤٣٦/ ١٤٣٧هـ الفصل الدراسي الثاني.

ثالثاً: عينة الدراسة: تم اختيار عينة الدراسة بطريقة العينة العشوائية العنقودية وبما يحقق متطلبات تحليل البيانات في الدراسة الحالية، حيث تم اختيار (١٥) مدرسة عشوائياً من مجموع مدارس مدينة عرعر البالغ عددها (٢٨) مدرسة، وقد بلغ مجموع الطلبة الذين قدمت لهم أدوات الدراسة في هذه المدراس (١٠٦٥) طالب، وبعد استبعاد الأوراق غير المكتملة في أي من نماذج الاختبار أو تلك غير القابلة للتحليل أصبح مجموع عينة الدراسة (٩٠٣) طالباً، حيث تم استبعاد (١٦٢) طالب، وجدول (٣) يوضح أعداد الطلبة الذين تم ادخال درجاتهم ضمن تحليل البيانات وذلك في كل مدرسة من مدراس عينة الدراسة (العينة في صورتها النهائية) :

جدول (٣)

توزيع عينة الدراسة في صورتها النهائية

م	المدرسة	العدد	م	المدرسة	العدد
١	غرناطة	١٠٣	٩	الأمير عبدالله بن مساعد	٥٦
٢	مصعب بن عمير	٦٢	١٠	صلاح الدين الأيوبي	٣٦
٣	المقداد	٥٣	١١	عثمان بن عفان	٣٨
٤	الأمير سلطان بن عبد العزيز	٧٥	١٢	نهاوند	٦٦
٥	الفاروق	٨٥	١٣	تحفيظ القرآن	٢٥
٦	سعيد بن المسيب	٩٢	١٤	الإمام أبو حنيفة	٩٥
٧	النعمان بن مقرن	٤١	١٥	سلمان الفارسي	٥٦
٨	الإمام البخاري	٤٧	المجموع = ٩٠٣ طالب		

رابعاً: أداة الدراسة: انطلاقاً من أهداف الدراسة وسعيا للإجابة على أسئلتها قام الباحث ببناء اختبار من نوع الاختيار من متعدد ذو الأربع بدائل وذلك في الوحدتين الدراستين (المعادلات الخطية، والمتباينات الخطية) من مادة الرياضيات للصف الثالث متوسط - الفصل الدراسي الأول للعام الدراسي ١٤٣٦، ١٤٣٧هـ، وبعد الرجوع لأدبيات بناء الاختبارات التحصيلية قام الباحث باتباع الخطوات التالية :

١ - التخطيط للاختبار : تضمن التخطيط للاختبار القيام بالإجراءات التالية :

أ- تحديد الغرض من الاختبار: يذكر علام (٢٠٠١، ٣٣٤) أن التخطيط الجيد للاختبار يتطلب تحديد الغرض من الاختبار وفيمَ تستخدم نتائجه، وينبغي أن يكون الغرض من الاختبار واضحا لكي يرشد جميع القرارات التالية (علام،٢٠٠٧، ٥٤).

وقد تم تحديد الغرض من الاختبار في ضوء متطلبات الدراسة الحالية وهو قياس تحصيل الطلبة في وحدتي المعادلات الخطية والمتباينات الخطية وتقدير مدى معرفتهم في هاتين الوحدتين وذلك للإجابة على أسئلة الدراسة.

ب- تحديد النطاق السلوكي (المحتوى الدراسي) وتحليله إلى الموضوعات والعناصر الرئيسية: تبدو أهمية هذه الخطوة في كونها الأساس الذي تبنى عليه الفقرات والمجال الذي تشتق منه (النبهان،٢٠٠٤)، وقد قام الباحث بتحديد النطاق السلوكي المستهدف بمحتوى الوحدتين الدراسيتين (المعادلات الخطية والمتباينات الخطية) ضمن منهج مادة الرياضيات للصف الثالث متوسط الفصل الدراسي الأول للعام الدراسي ١٤٣٦/ ١٤٣٧هـ، بعد ذلك قام الباحث بتحليل محتوى الوحدتين اللاتي تم اختيارهما وفقا لما تضمنه كل من كتاب الطالب ودليل المعلم وذلك للكشف عن المحتوى المعرفي المتعلق بكل درس من الدروس.

ج- صياغة الأهداف المعرفية السلوكية: بعد الرجوع إلى أدبيات وشروط صياغة الأهداف السلوكية وتصنيفها، وبالاستعانة بما ورد في دليل المعلم من أهداف تدريسية قام الباحث بصياغة عدد من الأهداف السلوكية المتعلقة بكل وحدة دراسية وذلك في ضوء المحتوى المعرفي لكل موضوع من الموضوعات، وقد بلغ عدد الأهداف السلوكية المتعلقة بالوحدة الأولى (المعادلات الخطية) ١٨ هدفا سلوكا، والمتعلقة بالوحدة الثانية (المتباينات الخطية) ١٩ هدفا سلوكيا، وبعد مراجعتها تم عرضها على عدد من المحكمين المختصين في المناهج وطرق التدريس، والقياس والتقويم، ومشرفي ومدرسي مادة الرياضيات (ملحق رقم ٤، وملحق رقم ٥)، وذلك للتأكد من ملائمتها للمحتوى المعرفي المستهدف، وسلامة صياغتها اللغوية، وعدم تكرراها، وفي ضوء تحليل أراء المحكمين تم حذف بعض الأهداف وإعادة صياغة البعض الآخر لتصبح قائمة الأهداف في صورتها النهائية مكونة (١٧) هدفا سلوكيا في وحدة المعادلات الخطية، و(١٥) هدف سلوكيا في وحدة المتباينات الخطية تم استخدامها لبناء فقرات الاختبار .

د- إعداد الجدول التخطيطي لمحتوى الاختبار (جدول المواصفات) Table Of Specification : يذكر النبهان (٢٠٠٤، ٨٠) أن هناك نماذج متنوعة لتكوين جدول مواصفات اختبارات التحصيل، إلا أن هذه النماذج بوجه عام تتكون من بعدين أساسين يمثل البعد الأول الموضوعات التي يغطيها الاختبار، ويمثل البعد الثاني العمليات الذهنية (الأهداف السلوكية) المرتبطة بالموضوعات، ويتكون من تقاطع الأعمدة والصفوف خلايا يمكن استخدامها لتقدير الأوزان الخاصة بكل موضوع والعمليات الذهنية المرتبطة به (الدوسري،٢٠٠٠، ١٨٠)، وقد قام الباحث ببناء جدول المواصفات في ضوء الوزن النسبي للزمن المستغرق في تدريس كل موضوع، وكذلك في ضوء الوزن النسبي للأهداف في كل مستوى ولكل موضوع، حيث تم تحديد عدد الأسئلة الخاصة بكل موضوع من الموضوعات وفق المعادلة التالية: عدد أسئلة الموضوع = الوزن النسبي للموضوع (الوزن النسبي لعدد حصصه) × العدد الكلي لأسئلة الاختبار، وكذلك تم تحديد عدد الأسئلة الخاصة بكل خلية من خلايا الجدول وفق المعادلة التالية: عدد أسئلة الخلية = الوزن النسبي للموضوع × الوزن النسبي للأهداف × العدد الكلي لأسئلة الاختبار مع مراعاة جبر الكسور وطبيعة كل هدف من الأهداف (ملحق رقم ٦).

ه- تحديد شكل الأسئلة التي سيتضمنها الاختبار : انطلاقاً من غرض الدراسة وأهدافها فقد تم تحديد صيغة الاختيار من متعدد لكتابة جميع فقرات الاختبار، وذلك بعد الرجوع للعوامل التي يجب مراعاتها عند اختيار صيغة الفقرة الاختبارية كالغرض من الاختبار، والوقت المخصص للإجابة، وأعداد الطلبة الخاضعين للاختبار، ومهارة المعلم في صياغة كل نوع، والتسهيلات المادية المتوفرة (مهرنز ولمن، ٢٠٠٣، ٩٤-٩٥).

٢ ـ كتابة فقرات الاختبار وتقييمها : بعد الرجوع إلى قواعد وتعليمات كتابة الفقرات الاختبارية بوجه عام وفقرات الاختيار من متعدد بوجه خاص قام الباحث بدراسة كل هدف من الأهداف السلوكية على حده ومحاولة صياغة أفضل الفقرات التي تقيس ذلك الهدف وذلك في ضوء جدول مواصفات الاختبار، بعد ذلك قام الباحث بإرفاق كل هدف مع الأسئلة التي تقيسه وعرضها على مجموعة المحكمين للتأكد من مدى قياس السؤال للهدف السلوكي المرتبط به، ومدى سلامة

ووضوح صياغة السؤال وإضافة أي مقترحات أو ملاحظات حول أي سؤال، (ملحق رقم٧)، وبعد تحليل أراء المحكمين تم تعديل وإعادة صياغة عدد من الأسئلة لتصبح الأسئلة في شكلها النهائي مكونة من (٥٢) سؤالا وهي التي تم اعتمادها في هذه الدراسة.

٣ ـ الإخراج الأولي للاختبار : بعد الانتهاء من تحليل أراء المحكمين وتعديل فقرات الاختبار في ضوئها تمت كتابة وإخراج النسخة المبدئية من الاختبار والتي تكونت من ثلاث نماذج تختلف فقط في تعليمات الإجابة عن كل نموذج وذلك بحسب طريقة تقدير الدرجات المستخدمة (الطريقة التقليدية، الطريقة التجريبية، طريقة الاحتمال المقترح للإجابة الصحيحة) .

٤ ـ تجريب الاختبار: قام الباحث بتجريب الاختبار على عينة استطلاعية مشابه لخصائص عينة الدراسة مكونة من(٩٥) طالبا الصف الثالث متوسط في مدرستي الإمام مسلم، والقيروان من خارج عينة الدراسة، وذلك بعد أن قام الباحث بشكل مباشر وبدعم من مديري المدرستين بتدريبهم على كيفية الإجابة على كل نموذج من نماذج الاختبار، وقد كان الغرض من تجريب الاختبار ما يلي:

- التحقق من وضوح التعليمات وذلك عند تطبيق الاختبار وفق الطرق الثلاثة (الطريقة التقليدية، الطريقة التجريبية، طريقة الاحتمال المقترح للإجابة الصحيحة)
- التحقق من وضوح الصياغة اللغوية للعينة المستهدفة.
- معرفة الزمن الذي يستغرقه الاختبار

٥ ـ الإخراج النهائي للاختبار : بعد الانتهاء من إجراءات تجريب الاختبار تم التوصل إلى الصيغة النهائية لأسئلة الاختبار (ملحق رقم ٩)، كما تم إخراج ثلاث نماذج لتعليمات الإجابة على النحو التالي:

أ- نموذج الطريقة التقليدية (والذي سيستخدم للنموذج اللوغاريتمي الثلاثي المعلم)

ب- نموذج الطريقة التجريبية.

ج- نموذج طريقة الاحتمال المقترح للإجابة الصحيحة. (ملحق ١٠)

٦ـ إجراءات جمع البيانات (تطبيق الاختبار وتصحيحه): قام الباحث بالإجراءات الآتية لجمع البيانات اللازمة لتحقيق أهداف الدراسة:

١- تحديد مجمع الدراسة،ثم اختيار العينة من مدارس الصف الثالث متوسط وذلك وفق الإجراءات المشار إليها في عينة الدراسة.

٢- إعداد وطباعة أداة الدراسة مع أوراق الإجابات المخصصة لكل نموذج وذلك بحسب أعداد الطلبة في كل مدرسة من المدارس الواردة في عينة الدراسة .

٣- الحصول على الموافقات الخطية الرسمية الموجهة من كلية التربية لمدير تعليم الحدود الشمالية ومنه إلى مديري المدارس المعنية إيذاناً بالتطبيق وتسهيل المهمة. (ملحق رقم ٢، وملحق رقم ٣)

٤- التنسيق مع قسم المناهج طرق التدريس بجامعة الحدود الشمالية وبإشراف مباشر من اثنين من الأساتذة في القسم لتكوين فريق من المساعدين من طلبة قسم المناهج في التدريب الميداني .

٥- زيارة المدارس التي شملتها عينة الدراسة والتحدث مع مديري المدارس حول إجراءات تطبيق الاختبار واختيار الحصص المناسبة للتدريب على التطبيق.

٦- تدريب الطلبة في المدراس الخاضعة للدراسة على كيفية الإجابة وفق تعليمات كل نموذج من نماذج الاختبار، وعدم البدء في التطبيق الفعلي إلا بعد التأكد من المام جميع الطلبة بتعليمات الاجابة.

٧- التطبيق الفعلي للاختبار وتم فيه :

أ- الالتقاء بالطلبة في عينة الدراسة وترتيب مقاعدهم بحيث تبقى مسافة مناسبة بين المفحوص والآخر بما يسمح بالتجوال وتوضيح التعليمات، وإخبارهم بأن عليهم الإجابة عن الاختبار بكل دقة وفقًا للتعليمات، وأن نتائج الاختبار لن تستخدم إلا لأغراض البحث العلمي.

ب- توزيع كراسات الاختبار (أوراق الأسئلة والإجابة) على الطلبة بمعاونة فريق المساعدين، ومن ثم توجيه كلمة للطلبة حول الغرض من الاختبار وحثهم على الإجابة بجديّة وفق التعليمات الخاصة بكل نموذج من نماذج الاختبار.

٨- جمع البيانات من مختلف المدراس وترتيبها واستبعاد البيانات غير القابلة للتصحيح أو غير المكتملة وحصر البقية والتي بلغت (٩٠٣) طالبا.

٩- استخدام الورق المثقب لتصحيح الاختبار ووضع علامة لكل فقرة من فقرات الاختبار بحسب تعليمات تصحيح كل نموذج.

٧ ـ الخصائص السيكومترية للاختبار وإحصائيات فقراته:

أولا: صدق الاختبار: التحقق من صدق الاختبار هو تقييم شامل لمدى دعم الأدلة التجريبية والنظرية والمنطقية لكفاية وملائمة التفسيرات أو الإجراءات التي تبنى في ضوء درجات الاختبار (Messick,1995,741)، وحيث أن أدلة صدق المحتوى هي الأدلة المناسبة للتحقق من صدق الاختبارات التحصيلية (رينالد وليفينجستون،٢٠١٣، ٢٢٦).

لذا قام الباحث بالتحقق من صدق المحتوى للاختبار وذلك من خلال:

١- المخطط التفصيلي للاختبار (جدول المواصفات) وفيه تم توزيع الأسئلة على الموضوعات الدراسية والأهداف المعرفية وذلك بحسب الأوزان النسبية لكل موضوع

٢- تقييم المختصين لعناصر الاختبار ومدى قياسها لمكونات المحتوى والأهداف السلوكية (صدق المحكمين)، بالإضافة إلى تحكيم الأهداف ومدى مناسبتها للمحتوى الدراسي، حيث قام الباحث بتحكيم الاختبار من حيث :

أ- تحكيم الأهداف المعرفية السلوكية من حيث مدى ملائمتها للمحتوى المعرفي المستهدف، وسلامة صياغتها اللغوية والفنية، وعدم تكرارها.

ب- تحكيم الفقرات الاختبارية (الأسئلة) من حيث مدى قياس السؤال للهدف السلوكي المرتبط به، ومدى سلامة ووضوح صياغة السؤال.

وقد شملت لجنة التحكيم على عدد من الأساتذة المختصين في القياس والتقويم، والمناهج وطرق التدريس، بالإضافة إلى عدد من معلمي ومشرفي مادة الرياضات في المرحلة المتوسطة.

ثانياً: ثبات درجات الاختبار: يقصد بثبات درجات الاختبار مدى خلوها من الأخطاء غير المنتظمة التي تشوب عملية القياس، وتهتم طرق تقدير الثبات بتقدير مدى ما يعكسه الاختبار ودرجاته من تباين حقيقي (علام، ٢٠٠٢)، وللكشف عن قيمة ثبات درجات الاختبار وفق طرق تقدير الدرجات

التي تناولتها الدراسة تم استخدام طريقة الاتساق الداخلي وذلك بتطبيق معامل ألفا (α) كرونباخ لصيغ الاختبار الثلاث وقد جاءت النتائج على النحو التالي :

جدول (٤)

قيم معاملات ثبات ألفا كرونباخ لدرجات الاختبار في نماذجه الثلاثة

طريقة تقدير الدرجات	التقليدية	التجريبية	الاحتمال المقترح للإجابة الصحيحة
قيمة معامل ألفا كرونباخ (α)	٠,٨٨	٠,٩٠	٠,٨٩

يلاحظ اتصاف الاختبار بدرجة ثبات مرتفعة وذلك في نماذجه الثلاث مما يدل على قلة أخطاء القياس العشوائية في درجات الاختبار، وهذا يزيد من الثقة في النتائج التي تستمد منه.

ثالثاً : معاملات الصعوبة والتمييز : ١- صعوبة الفقرات Items Difficulty: تعبر صعوبة الفقرة عن نسبة الطلاب الذين أجابوا بشكل صحيح عن الفقرة إلى العدد الكلي للطلاب (الذين اجابوا بالفعل عن تلك الفقرة) (Demars ,2010,4)، وقد تم استخراج قيم معاملات الصعوبة وفق النماذج الثلاث للاختبار، ولتسهيل قراءة وتفسير بيانات الجداول السابقة فقد تم تلخيصها في الجدول التالي :

جدول (٥)

تلخيص بيانات معاملات صعوبة الاختبار وفق طرق تقدير الدرجات الثلاث

طريقة تقدير الدرجات	ملخص معاملات الصعوبة				
	عدد الفقرات	أعلى قيمة	أدنى قيمة	المتوسط	الانحراف المعياري
التقليدية	٥٢	٠,٧٢	٠,٢٦	٠,٤٤	٠,١٠
التجريبية	٥٢	٠,٥٨	٠,٢٧	٠,٤٢	٠,٠٧
الاحتمال المقترح للإجابة الصحيحة	٥٢	٠,٥٧	٠,٢٤	٠,٤١	٠,٠٨

وبقراءة بيانات الجدول السابق، وبالرجوع إلى البيانات الكاملة لمعاملات الصعوبة،يلاحظ أن قيم معاملات الصعوبة للاختبار في نماذجه الثلاث كانت مناسبة، حيث تراوحت بين (٠,٧٢)، و (٠,٢٤) ومن ثم فجميعها تقع ضمن المدى الذي يعد مقبولا لقيم معاملات صعوبة الاختبار والذي يمتد من (٠,٢٠ إلى ٠,٨٠) (Jha,2014,305)، ويتأكد ذلك من خلال قيم متوسطات معاملات الصعوبة والتي بلغت (٠,٤٤) للطريقة التقليدية، و(٠,٤٢) للطريقة التجريبية، و(٠,٤١) لطريقة الاحتمال المقترح للإجابة الصحيحة وهذه القيم قريبة من القيمة المثالية للصعوبة (٠,٥٠) والتي تجعل تباين الفقرة يصل إلى أقصى ما يمكن (النبهان، ٢٠٠٤ ، ١٩٢)، كما يلاحظ أنها كانت أقل من المتوسط مما يدل على الصعوبة النسبية للاختبار.

٢- تمييز الفقرات Items Discrimination: تدل قيمة معامل تمييز الفقرة العالية على أن هذه الفقرة تميز بدرجة عالية بين المختبرين ذوي المستويات المختلفة، والغرض من استخدام هذا المعامل هو الكشف عن مدى مقدرة الفقرة على التمييز بين الطلبة المتمكنون مما يقيسه الاختبار

وبين أولئك غير المتمكنين (Demars,2010)،وقد قام الباحث بحساب قيم معاملات تمييز الفقرات باستخدام الطريقة الارتباطية والتي يتم فيها إيجاد معامل الارتباط بين الدرجة على الفقرة من جهة والدرجة الكلية من جهة أخرى (النبهان، ٢٠٠٤، ٢٠١)ولتسهيل قراءة وتفسير بيانات الجداول السابقة فقد تم تلخصيها في بعد إضافة المتوسط الحسابي والانحراف المعياري لمعاملات التمييز لكل طريقة:

جدول (٦)

تلخيص بيانات معاملات تمييز الاختبار وفق طرق تقدير الدرجات الثلاث

طريقة تقدير الدرجات	ملخص معاملات التمييز				
	عدد	أعلى قيمة	أدنى قيمة	المتوسط	الانحراف
التقليدية	٥٢	٠,٥١	٠,٢١	٠,٣٩	٠,٠٧
التجريبية	٥٢	٠,٦٢	٠,٣٣	٠,٤٨	٠,٠٧
الاحتمال المقترح للإجابة الصحيحة	٥٢	٠,٥٧	٠,٣١	٠,٤٦	٠,٠٧

وبقراءة بيانات معاملات يلاحظ عدم وجود أي معامل تمييز ذو قيمة سالبة، بالإضافة إلى أن جميع القيم للنماذج الثلاث كانت أعلى من (٠,١٩) وهو الحد الأدنى الذي اقترحه أيبل Ebel للإبقاء على الفقرة (الدوسري،٢٠٠٠ ، ٢٢٠)، كذلك نجد أن متوسط قيم معاملات التمييز للطريقتين التجريبية والاحتمال المقترح للإجابة الصحيحة أعلى من القيمة (٠,٤)، مما يدل على فقرات جيدة جدا، فيما كان متوسط معاملات التمييز للطريقة التقليدية مساويا للقيمة (٠,٣٩) والتي تدل على فقرات جيدة وفقا لنفس المعيار.

خامساً: المعالجة الإحصائية للبيانات: تم استخدام البرامج والأساليب الإحصائية التالية لمعالجة نتائج الدراسة :

١- برنامج الحزمة الإحصائية للعلوم الاجتماعية SPSS(22) وذلك لتقدير معاملات ثبات الفا كرونباخ وحساب معاملات ارتباط بيرسون، ومعاملات الصعوبة والتمييز لفقرات الاختبار وفق المدخل الكلاسيكي في القياس، بالإضافة إلى عدد من الإحصائيات الوصفية الأخرى.

٢- برنامج بيلوج – أم جي (3) Bilog-MG لتحليل البيانات وفق نظرية استجابة الفقرة (IRT) وقد استخدمه الباحث لتقدير معلمات الفقرة (الصعوبة – التمييز)، ومعلمة الأفراد (القدرة)، وذلك باستخدام أسلوب الأرجحية العظمى الهامشية MML في النموذج اللوغاريتمي الثلاثي المعلم.

نتائج الدراسة وتفسيرها:

أولاً : التحقق من افتراضات النموذج اللوغاريتمي الثلاثي: حيث أن إجابة أسئلة الدراسة الحالية تتطلب استخدام النموذج اللوغاريتمي الثلاثي المعلم كان لابد قبل البدء في معالجة البيانات إحصائيا من فحص الافتراضات الأساسية لهذا النموذج (أحادية البعد، والاستقلال الموضعي، والتحرر من السرعة) والتي يجب توافرها في البيانات، وللتحقق من افتراض أحادية البعد فقد

أشار هاتي (1985) Hattie إلى عدد من المؤشرات التي تستخدم للدلالة على تحقق هذا الافتراض والتي من أهمها ما يلي :

١- معامل كرونباخ (α) :تم تقدير معامل الثبات باستخدام طريقة ألفا كرونباخ على عينة الدراسة المكونة من ٩٠٣ طالبا، وبلغت قيمة هذا (٠,٨٨) والتي تعد مؤشرا على تحقق أحادية البعد في البيانات.

٢- معاملات الارتباط بين الاستجابة على الفقرة، والاستجابة على الاختبار الكلي: تم حساب معاملات الارتباط بين الاستجابة على الفقرة، والاستجابة على الاختبار الكلي،،حيث تراوحت قيم معاملات الارتباط بين (٠,٢١) و (٠,٥١) بوسط حسابي بلغ (٠,٣٩)، وكانت جميعها ذات دلالة إحصائية عند مستوى الدلالة، (α=٠,٠١) وهذا يدل على أن الفقرات تتشارك في قياس بعد واحد تعبر عنه العلامة الكلية، وكذلك يمكن اعتبار هذه النتيجة مؤشراً على أحادية البعد للمقياس، حيث أشار هاتي (1985) Hattie أنه إذا كان ارتباط معظم الفقرات بالدرجة الكلية للمقياس أعلى من (٠,٢) فإن ذلك مؤشرا على تحقق أحادية البعد، والجدول التالي يعرض قيم معاملات الارتباط بين فقرات المقياس والدرجة الكلية

جدول (٧)

معاملات الارتباط بين فقرات المقياس والدرجة الكلية

رقم الفقرة	معامل الارتباط مع الدرجة الكلية	رقم الفقرة	معامل الارتباط مع الدرجة الكلية	رقم الفقرة	معامل الارتباط مع الدرجة الكلية
1	0.412	19	0.341	37	0.399
2	0.371	20	0.439	38	0.347
3	0.433	21	0.458	39	0.299
4	0.23	22	0.379	40	0.21
5	0.477	23	0.456	41	0.439
6	0.512	24	0.432	42	0.226
7	0.402	25	0.371	43	0.451
8	0.468	26	0.325	44	0.314
9	0.479	27	0.377	45	0.36
10	0.476	28	0.446	46	0.42
11	0.422	29	0.461	47	0.441
12	0.399	30	0.363	48	0.293
13	0.362	31	0.371	49	0.342
14	0.334	32	0.406	50	0.397
15	0.41	33	0.324	51	0.359
16	0.376	34	0.366	52	0.364
17	0.413	35	0.375	المتوسط	0.39
18	0.407	36	0.342		

٣- التحليل العاملي: أ - تم التحقق من افتراضات إ جراء التحليل العاملي على النحو التالي :

- استخرجت القيمة المطلقة لمحدد مصفوفة الارتباط وقد بلغت (٠,٠٠٣١)، وهي أكبر من المحك الذي يساوي (٠,٠٠٠١) وهذا يشير إلى عدم وجود مشاكل في مستوى الارتباط بين المتغيرات (شراز، ٢٠٠٩، ١٥٠).

- تم التحقق من عدم تماثل مصفوفة الارتباط الأصلية من خلال دلالة قيمة كاي تربيع (x^2) لاختبار بارتليت حيث بلغت القيمة (٥٣٤٤,٦١)، وهي دالة إحصائيا عند مسوى دلالة (٠,٠٠٠) .
- تم التحقق من كفاية حجم العينة لإجراء التحليل العاملي عن طريق اختبار كايزر – ماير – أولكن (K.M.O)، وقد بلغت قيمة الاختبار (٠,٩٠٥) وهي أكبر من الحد الأدنى اللازم لإجراء التحليل العاملي الذي يساوي (٠,٥)، وحيث أنها تعدت (٠,٩) فهي تعد قيمة مثالية (شراز، ٢٠٠٩، ١٥١) .

ب- تم إجراء التحليل العاملي لبيانات الاختبار المتعلقة باستجابات عينة الدراسة البالغة (٩٠٣) طالباً وذلك باستخدام طريقة تحليل المكونات الرئيسة (Principle Components Analysis, PCA)، والتدوير تبعاً لمحاور متعامدة (Varimax)، وقد أسفر التحليل العاملي عن (١٤) عاملاً كانت جذورها الكامنة أكبر من (١)، فسر العامل الأول منها (١٥,٣٦%) من التباين، وفسرت بقية العوامل (٤٧,٤٤%) من التباين الكلي، ويبين الجدول التالي قيم الجذور الكامنة (Eigenvalues)، ونسبة التباين المفسر لكل عامل، وكذلك نسبة التباين المفسر التراكمية.

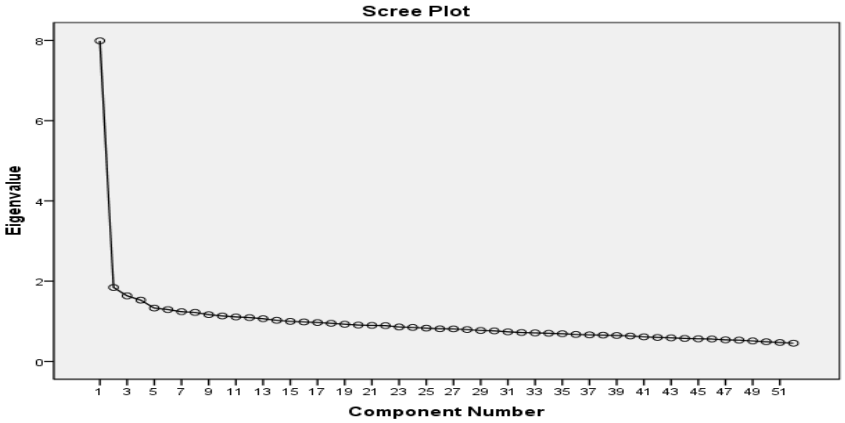
جدول (٨)

قيم الجذور الكامنة ونسبة التباين المفسر

رقم	الجذر	نسبة التباين	نسبة التباين المفسر
1	7.99	15.36	15.36
2	1.84	3.55	18.91
3	1.63	3.14	22.05
4	1.53	2.94	24.99
5	1.33	2.56	27.55
6	1.29	2.48	30.03
7	1.24	2.38	32.42
8	1.22	2.35	34.77
9	1.17	2.24	37.01
10	1.13	2.18	39.19
11	1.11	2.13	41.32
12	1.09	2.10	43.42
13	1.07	2.05	45.47
14	1.02	1.97	47.44

ويظهر من الجدول السابق أن قيمة الجذر الكامن للعامل الأول بلغت ٧,٩٩ ويفسر ما نسبته ١٥,٣٦% من التباين الكلي، وهي قيمة مرتفعة إذا ما قورنت مع قيم الجذور الكامنة لبقية العوامل، أما قيمة الجذر الكامن للعامل الثاني فقد بلغت ١,٨٤ ويفسر ما نسبته ٣,٥٥% فقط من التباين الكلي، أي أن قيمة الجذر الكامن للعامل الأول كانت أكبر من ضعفي قيمة الجذر الكامن العامل الثاني (حيث بلغت النسبة بينهما ٤,٣ أضعاف)، كما يلاحظ بأن نسبة التباين المفسر لكل من العوامل المتبقية جاءت متقاربة، بمعنى أنه يوجد شبه استقرار في نسب التباين المفسر لجميع العوامل باستثناء العامل الأول، وهذا مؤشر على تحقق افتراض أحادية البعد للمقياس أي أن المقياس يقيس سمة واحدة (Hambleton et al,1991; Onder,2007)

٤ ـ يتعزز افتراض أحادية البعد للمقياس المعدّ بالدراسة الحالية من خلال تمثيل الجذور الكامنة بيانياً باستخدام ما يعرف باختبار فحص العوامل (Scree plot) الذي يظهر في الشكل التالي:



بملاحظة التمثيل البياني لقيم الجذور الكامنة أعلاه، يظهر أن الجذر الكامن للعامل الأول مهمين بشكل واضح على الجذور الكامنة لبقية العوامل، حيث يظهر أن هناك تحوّلاً في ميل المنحنى عند العامل الثاني ويبقى الميل متقارباً لبقية العوامل مما يرجح وجود عامل سائد يمكن أن يستدل منه على أحادية البعد لبيانات المقياس .

ونظرا لأن افتراض الاستقلال الموضعي Local Independence يتحقق في حالة تحقق افتراض أحادية البعد (Hambleton& Swaminathan, 1985,11) لذا فإن ما سبق من إجراءات كافي لتحقق افتراض الاستقلال الموضعي. وفيما يخص افتراض التحرر من السرعة Speediness فقد راعى الباحث أثناء تطبيق الاختبار أن يعطي الطلاب الوقت الكافي للإجابة، بحيث أن أكثر من ٩٥ ٪ من الطلاب أنهوا الاختبار قبل الوقت المحدد مما يعني أن عامل السرعة لم يلعب دوراً في تحديد قدرة الطلاب.

ثانياً : عرض نتائج الدراسة وتفسيرها :

السؤال الأول : ما درجة ارتباط تقديرات قدرات الطلاب عند استخدام النموذج اللوغاريتمي الثلاثي المعلم بتقديرات درجاتهم عند استخدام كل من طريقة من طرق تقدير الدرجات الكلاسيكية (طريقة الاحتمال المقترح للإجابة الصحيحة، والطريقة التجريبية)؟

للإجابة عن هذا السؤال تم حساب معامل ارتباط بيرسون بين قدرات الطلبة مقدرة باستخدام النموذج اللوغاريتمي الثلاثي المعلم وتقديرات درجاتهم عند استخدام كل من (الطريقة التجريبية، وطريقة الاحتمال المقترح للإجابة الصحيحة)، وذلك كما في الجدول التالي:

جدول (٩) قيم معاملات ارتباط بيرسون بين تقديرات قدرات الطلاب وفق الطرق الكلاسيكية والنموذج اللوغاريتمي الثلاثي

	التجر ي	الاحتمال المقترح للإجابة الصحيحة	النموذج اللوغاريتمي
التجريبية	١	٠,٨٥٩ **	٠,٧٦ **
الاحتمال المقترح للإجابة الصحيحة		١	٠,٧٤ **
النموذج اللوغاريتمي			١

يظهر من الجدول السابق وجود علاقة طردية قوية بين قدرات الطلاب المقدرة بالنموذج اللوغاريتمي ثلاثي المعلم ودرجاتهم عند تقديرها باستخدام أي من الطرقتين الكلاسيكيتين، إلا أن الطريقة الأكثر ارتباطا بالنموذج اللوغاريتمي في تقدير الدرجات كانت الطريقة التجريبية بمعامل ارتباط بلغت قيمته (٠,٧٦) والتي تشير إلى وجود علاقة ارتباطية قوية ودالة إحصائيا عند مستوى الدلالة (٠,٠١). في حين بلغت قيمة معامل ارتباط الدرجات المقدرة وفق طريقة الاحتمال المقترح للإجابة الصحيحة مع القدرات المقدرة وفق النموذج اللوغاريتمي (٠,٧٤) وهي تشير كذلك إلى وجود ارتباط طردي دال وقوي عند مستوى الدلالة (٠,٠١).

وبإعادة قراءة النتائج السابقة نجد أنها تتفق مع عدد من الدراسات التي سعت للكشف عن طبيعة العلاقة بين تدريج درجات نماذج النظرية الكلاسيكية، ونماذج نظرية الاستجابة للمفردة كدراسات (حجماوي،٢٠٠٠؛الحكماني، ٢٠٠٧؛ Ndalichako&Rogers,1997; fan,1998 ;Cantrell,1997)، فجميع هذه الدراسات اتفقت على وجود ارتباط عالي بين تدريج الدرجات وفق النظرية الكلاسيكية وتدريج القدرات وفق نماذج نظرية الاستجابة للمفردة، إلا ان نتائج الدراسة الحالية أكدت على وجود هذا الارتباط المرتفع بين تقدير الدرجات وفق المدخلين النظريين وذلك باستخدام نموذجين كلاسيكيين لتقدير الدرجات لم تتطرق لهما تلك الدراسات وهما (الطريقة التجريبية، والاحتمال المقترح للإجابة الصحيحة).

السؤال الثاني : ما درجة الاختلاف بين قدرات الطلاب المقدرة وفق النموذج اللوغاريتمي الثلاثي المعلم وقدراتهم المقدرة وفق كل من (طريقة الاحتمال المقترح للإجابة الصحيحة، والطريقة التجريبية)؟

تعد قيمة الارتباط العالية بين نتائج اختبارين مؤشرا على أن هذين الاختبارين يرتبان الأفراد - وفق قدراتهم - بشكل متشابه، إلا أن هذا لا يعني بالضرورة أن هذين الاختبارين يعطيان تقييمات متقاربة لقدرات الأفراد، لهذا تم بحث مؤشرا آخراً للمقارنة بين طرق تقدير الدرجات الداخلة في هذه الدراسة من خلال الكشف عن مدى تقارب أو تباعد هذه الطرق في تقدير قدرات الأفراد، ولإجراء هذه المقارنة تم اتباع الآتي :

١- تحويل الدرجات المستمدة من طرق تقدير الدرجات الثلاثة إلى وحدة قياس موحدة، وهي الدرجات التائية (ت)، وذلك لكي نتمكن من المقارنة الكمية بين درجات الطالب على كل طريقة، وكذلك لتسهيل قراءة وتفسير الدرجات (حيث أن متوسطها الحسابي =٥٠، وانحرافها المعياري= ١٠)، وقد تم تحويل الدرجات من خلال التعويض بالمعادلة التالية: الدرجة التائية (ت) = [ز × ١٠] + ٥٠.

حيث: (ز) هي الدرجة المعيارية وتساوي [الدرجة الخام – المتوسط الحسابي] ÷ الانحراف المعياري . (الشافعي، ٢٠١٤، ٩٣-٩٥).

٢- حساب متوسط الفروق المطلقة بين الدرجات التائية (ت) لكل طريقتين من طرق تقدير الدرجات، وذلك وفق المعادلة التالية:

$$MAD_{xy} = \frac{\sum_{j=1}^{N} |T_{xj} - T_{yj}|}{N}$$ (Ndalichako & Rogers,1997).

حيث:

MAD_{xy} : متوسط الفروق المطلقة بين الدرجات التائية لطريقة تقدير الدرجات x، وطريقة y

T_{xj}: درجة الطالب (j) المستمدة من الطريقة x

T_{yj} : درجة الطالب (j)المستمدة من طريقة تقدير y .

N: عدد الطلاب .

وقد جاءت النتائج كما في الجدولين التاليين :

جدول (١٠)

عدد من الإحصائيات الوصفية للدرجات التائية المستمدة من طرق تقدير الدرجات الكلاسيكية والنموذج اللوغاريتمي

م	طريقة تقدير الدرجات	عدد الفقرات	أقل درجة	أعلى درجة	المدى	الوسط الحسابي	الانحراف المعياري
١	التجريبية	٥٢	٣٤٥٦٧	٨١،١	٤٧،١	٥٠	١٠
٢	الاحتمال المقترح للإجابة الصحيحة	٥٢	٣٣،٥	٨٠،٧	٤٧،٢	٥٠	١٠
٣	النموذج اللوغاريتمي الثلاثي	٥٢	٢٢،٤	٧٥،٧	٥٣،٣	٥٠	١٠

جدول (١١)

متوسط الفروق المطلقة بين درجات كل طريقتين من طرق تقدير الدرجات الداخلة في الدراسة.

مجال المقارنة		عدد الطلاب	متوسط الفروق المطلقة	الانحراف المعياري
النموذج اللوغاريتمي	التجريبية	٩٠٣	٥،٧	٣،٩
	الاحتمال المقترح للإجابة الصحيحة	٩٠٣	٥,٩	٤,٢
التجريبية	الاحتمال المقترح للإجابة الصحيحة	٩٠٣	٣,٩	٣,٩

بقراءة بيانات الجدول السابق يظهر أن طريقة تقدير الدرجات الأكثر قرباً من القدرات المحسوبة وفق النموذج اللوغاريتمي الثلاثي هي الطريقة التجريبية حيث بلغ متوسط الفروق المطلقة بين الطريقتين (٥،٧) درجة، في حين بلغ متوسط الفروق المطلقة بين القدرات المحسوبة وفق النموذج اللوغاريتمي الثلاثي وطريقة الاحتمال المقترح للإجابة الصحيحة (٥,٩) درجة تائية.

وتتسق هذه النتيجة مع ما تم التوصل إليه عند إجابة السؤال الأول من كون الطريقة التجريبية هي الأعلى ارتباطا بالنموذج اللوغاريتمي الثلاثي في تقدير قدرات الطلاب مقارنة بطريقة الاحتمال المقترح للإجابة الصحيحة، كما يلاحظ على هذه النتيجة والتي بلغت فيها الفروق المطلقة بين النموذجين الكلاسيكيين والنموذج اللوغاريتمي (٥،٧)، (٥,٩) درجة، أنها اكبر بشكل واضح مما أظهرته دراسة ندالياتشاكو وروجرز (Ndalichako & Rogers,1997)

والتي أظهرت أن الفروق المطلقة بين الدرجات المقدرة وفق النماذج الكلاسيكية والنماذج اللوغاريتمية أقل من (٢,١٥) درجة، ويعزو الباحث ذلك إلى اختلاف الطرق الكلاسيكية التي استخدمتها الدراسة الحالية مع الطرق الكلاسيكية التي استخدمتها دراسة نداليتشاكو وروجرز.

السؤال الثالث : ما درجات ارتباط قيم معاملات صعوبة الفقرات عند استخدام كل من طريقة الاحتمال المقترح للإجابة الصحيحة، والطريقة التجريبية مع قيم معاملات صعوبة الفقرات عند استخدام النموذج اللوغاريتمي الثلاثي المعلم؟

للإجابة عن هذا السؤال تم حساب معاملات صعوبة الفقرات وفق طرق تقدير الدرجات الكلاسيكية وكذلك قيم حساب بارامتر صعوبة الفقرات وفق النموذج اللوغاريتمي الثلاثي المعلم، والجدول التالي يلخص هذه الإحصائيات، حيث تم تحديد أقل وأعلى قيمة لمعاملات الصعوبة والمتوسطات الحسابية والانحرافات المعيارية وفقا لكل طريقة من طرق التقدير :

جدول (١٢)

ملخص لقيم صعوبة الفقرات المقدرة وفق طرق تقدير الدرجات الكلاسيكية والنموذج اللوغاريتمي الثلاثي

طريقة تقدير الدرجات	عدد الفقرات	أقل قيمة لمعامل الصعوبة		أعلى قيمة لمعامل الصعوبة		المدى بوحدات (Z)	المتوسط	الانحراف المعياري
		بوحدات الصعوبة	بدرجات (Z)	بوحدات الصعوبة	بدرجات (Z)			
التجريبية	٥٢	٠،٢٧	٢,٠٤-	٠,٥٨	٢,١٩	٤،٢	٠,٤٢	٠،٠٧٤
الاحتمال المقترح للإجابة الصحيحة	٥٢	٠,٢٤	٢,١٩ -	٠,٥٧	٢,١٢	٤،٣	٠,٤١	٠,٠٧٨
النموذج اللوغاريتمي الثلاثي	٥٢	٠,٤٣ -	٢,٩٦ -	٢,١٦	٢,٣١	٥،٣	١,٠٣	٠,٤٩

بعد ذلك تم حساب معاملات ارتباط بيرسون بين قيم صعوبة الفقرات المقدرة وفق النموذج اللوغاريتمي الثلاثي وقيم صعوبة الفقرات المقدرة وفق كل طريقة من طرق تقدير الدرجات الكلاسيكية كما في جدول التالي :

جدول (١٣)

معاملات ارتباط بيرسون بين قيم صعوبة الفقرات المقدرة وفق الطرق الكلاسيكية والنموذج اللوغاريتمي الثلاثي

	الطريقة التجريبية	**طريقة الاحتمال المقترح للإجابة الصحيحة**	**النموذج اللوغاريتمي الثلاثي**
الطريقة التجريبية	١	٠,٩٨٣ **	٠,٨٣٣ - **
طريقة الاحتمال المقترح للإجابة الصحيحة		١	٠،٨١٤ - **
النموذج اللوغاريتمي الثلاثي			١

تظهر بيانات جدول السابق وجود علاقة ارتباطية ذات دلالة إحصائية عند مستوى الدلالة ٠,٠١ وبدرجة ارتباط قوية بين قيم معاملات الصعوبة المقدرة وفق النموذج اللوغاريتمي الثلاثي ووفق طريقتي تقدير الدرجات الكلاسيكية، ويلاحظ أن وجود الإشارة السالبة لا يعني العلاقة العكسية بل يعود إلى المنطق المعكوس لقيم معاملات الصعوبة وفق طرق القياس الكلاسيكية، كما تظهر البيانات أن معاملات الصعوبة الكلاسيكية المستمدة من الطريقة التجريبية كانت هي الأكثر ارتباطا ببارامتر صعوبة النموذج اللوغاريتمي الثلاثي وذلك بمعامل ارتباط بلغت قيمته (- ٠,٨٣٣) والتي تشير إلى درجة ارتباط قوية، وذلك بالمقارنة مع طريقة الاحتمال المقترح للإجابة الصحيحة والتي بلغت قيمة معامل ارتباطها بالنموذج اللوغاريتمي الثلاثي (- ٠،٨١٤)

وبوجه عام تتفق هذه النتائج مع الدراسات التي أشارت إلى وجود ارتباط قوي بين تدريج صعوبة الفقرات المقدرة وفق نموذج النظرية الكلاسيكية ونماذج نظرية استجابة الفقرة كدراسة (جمحاوي،٢٠٠٠)، ودراسة (fan.1998)، ودراسة (Stage.1999)، إلا ان الدراسة الحالية أكدت قوة هذا الارتباط بين المدخلين النظريين باستخدام نموذجي لتقدير الدرجات لم تتطرق لهما الدراسات السابقة (طريقة التجريبية، والاحتمال المقترح للإجابة الصحيحة)

السؤال الرابع : ما درجات ارتباط قيم معاملات تمييز الفقرات عند استخدام كل من طريقة الاحتمال المقترح للإجابة الصحيحة والطريقة التجريبية مع قيم معاملات تمييز الفقرات عند استخدام النموذج اللوغاريتمي الثلاثي المعلم؟

للإجابة عن هذا السؤال تم حساب معاملات تمييز الفقرات وفق طرق تقدير الدرجات الكلاسيكية، بالإضافة إلى بارامتر تمييز الفقرات في النموذج اللوغاريتمي الثلاثي المعلم وجدول التالي يخلص هذه الإحصائيات، حيث تم تحديد أقل وأعلى قيمة لمعاملات التمييز والمتوسطات الحسابية والانحرافات المعيارية وفقا لكل طريقة من طرق تقدير الدرجات:

جدول (١٤)

ملخص لقيم تمييز الفقرات المقدرة وفق طرق تقدير الدرجات الكلاسيكية والنموذج اللوغاريتمي الثلاثي

طريقة تقدير الدرجات	عدد الفقرات	أقل قيمة لمعامل التمييز		أعلى قيمة لمعامل التمييز		المدى بوحدات (Z)	المتوسط	الانحراف المعياري
		بوحدات التمييز	بدرجات (Z)	بوحدات التمييز	بدرجات (Z)			
التقليدية	٥٢	٠,١٩٤	٢,٩ ٢-	٠,٥١٢	١,٩٢	٤,٨	٠,٣٩	٠,٠٦٦
التجريبية	٥٢	٠,٣٣٣	٢,٠٥ -	٠،٦٢	٢،٠٤	٤,٠٩	٠,٤٨	٠,٠٧٠
الاحتمال المقترح للإجابة الصحيحة	٥٢	٠,٣١١	٢,٢٣ -	٠,٥٧٣	١،٦٩	٣,٩١	٠,٤٦	٠,٠٦٧
النموذج اللوغاريتمي الثلاثي	٥٢	٠،٥١	٢,٠٧-	٢,٦٥	٣,٠٤	٥,١١	١,٣٧	٠,٤١٨

بعد ذلك تم حساب معامل ارتباط بيرسون بين قيم تمييز الفقرات المقدرة وفق النموذج اللوغاريتمي الثلاثي وقيم تمييز الفقرات المقدرة وفق كل طريقة من طرق تقدير الدرجات الكلاسيكية، وقد جاءت النتائج كما في جدول التالي:

جدول (١٥)

معاملات ارتباط بيرسون بين قيم تمييز الفقرات المقدرة وفق الطرق الكلاسيكية والنموذج اللوغاريتمي الثلاثي

	الطريقة التجريبية	طريقة الاحتمال المقترح للإجابة الصحيحة	النموذج اللوغاريتمي الثلاثي
الطريقة التجريبية	١	**٠,٩١٣	٠,١١٦
طريقة الاحتمال المقترح للإجابة الصحيحة		١	٠,٢٦٨
النموذج اللوغاريتمي الثلاثي			١

تظهر بيانات الجدول السابق أن قوة الارتباط بين الطرق الكلاسيكية والنموذج اللوغاريتمي الثلاثي في تقدير تمييز الفقرات كانت ضعيفة، وذلك بالنسبة لأي من طريقتي تقدير الدرجات، إلا أن طريقة الاحتمال المقترح للإجابة الصحيحة كانت هي الأعلى ارتباطا بالنموذج اللوغاريتمي بمعامل ارتباط بلغت قيمته (٠,٢٦٨) وذلك بالمقارنة بالطريقة التجريبية والتي جاءت بمعامل ارتباط (٠,١١٦)، ويلاحظ أن أي من ارتباط هاتين الطريقتين لم يكن ذو دلالة إحصائية عند مستوى الدلالة ٠,٠٥

وتتفق هذه النتيجة مع ما توصلت إليه دراسة (جمحاوي، ٢٠٠٠) من ضعف الارتباط بين النظرية الكلاسيكية والنموذج اللوغاريتمي الثلاثي في تقدير تمييز الفقرات، وكذلك مع ما توصلت إليه دراسة ستيج (Stage, 1999) من أن الارتباط بين النظرية الكلاسيكية ونظرية الاستجابة للمفردة في تقدير تمييز الفقرات أقل منه في تقدير صعوبة الفقرات، إلا أنها تختلف معها في قوة الارتباط بين تمييز الفقرات وفق النموذجين، ففي حين أظهرت نتائج دراسة (Stage, 1999) أن قوة الارتباط بينهما (متوسطة) نجد أن الدراسة الحالية أظهرت أن قوة الارتباط (ضعيفة)، ويعزو الباحث هذا الاختلاف إلى اختلاف النماذج المستخدمة في دراسة دراسة ستيج Stage والتي طبقت على ا لطريقة التقليدية في تقدير درجات الاختيار من متعدد، في حين استخدمت الدراسة الحالية طريقتي الاحتمال المقترح للإجابة الصحيحة والطريقة التجريبية .

المراجع

أبو جراد، حمدي يونس (٢٠٠٨). استخدام نموذج (راش) في تطوير اختبار كاتل الثالث للذكاء الصورة (أ). *مجلة الجامعة الإسلامية – شئون البحث العلمي والدراسات العليا بالجامعة الإسلامية- غزة- فلسطين*، ١٦ (٢)، ٥٥٥-٥٨٣.

اناستازي، أي، و أورينا، اس (٢٠١٥). *القياس النفسي*.(صلاح الدين علام، مترجم). عمان : دار الفكر .

جمحاوي، إيناس (٢٠٠٠) . *مقارنة خصائص الفقرات وفق النظرية التقليدية ونظرية استجابة الفقرة في مقياس للقدرة الرياضية*.(رسالة ماجستير غير منشورة). كلية التربية، جامعة اليرموك.

الحكماني، رحاب بنت سعيد (٢٠٠٧). مقارنة بين النظرية الكلاسيكية للاختبار ونظرية الاستجابة للمفردة في تقدير قدرات الأفراد ومدى استقرار مؤشرات المفردات الاختبارية. *المجلة التربوية- الكويت*، ٢٣ (٨٩)، ٢٥٣- ٢٥٩.

الدوسري، إبراهيم مبارك (٢٠٠٠). *الإطار المرجعي للتقويم التربوي*، مكتب التربية العربي لدول الخليج : الرياض

دودين، حمزة محمد (٢٠٠٣). طرح معادلة لتصحيح أثر التخمين في اختبارات أسئلة الاختيار من متعدد. *كلية العلوم التربوية، جامعة الإمارات العربية المتحدة، العين*.

رينالد، سي، و ليفينجستون. ار .(٢٠١٣). *إتقان القياس النفسي الحديث - النظريات والطرق*.(صلاح الدين علام، مترجم). دار الفكر: عمان.

زكري، علي محمد (٢٠١٠). *الخصائص السيكومترية لاختبار (أوتيس - لينون) للقدرة العقلية مقدرة وفق القياس الكلاسيكي ونموذج راش لدى طلبة المرحلة المتوسطة بمحافظة صبيا التعليمية*.(رسالة دكتوراه غير منشورة)، كلية التربية، جامعة أم القرى.

سواقد، ساري سليم سلامة (١٩٨٣). *المقارنة بين أثر ثلاث طرق لتصحيح اختبارات الاختيار من متعدد على الخصائص السيكومترية للاختبار وعلى أداء المفحوصين من مستويات تحصيل ودرجات مخاطرة مختلفة على الاختبار*. (رسالة ماجستير غير منشورة)،كلية التربية، جامعة اليرموك.

سواقد، ساري سليم سلامة (١٩٩٢). *اختبار صحة الافتراضات النظرية لطرق التصحيح من أثر التخمين، ومقارنة أثر استخدام هذه الطرق على الخصائص السيكومترية للفقرة*. (رسالة دكتوراه غير منشورة)،كلية الدراسات العليا، الجامعة الأردنية.

الشافعي، محمد منصور محمد. (٢٠١٤) *الإحصاء في البحوث العلمية والإنسانية – أسس وتطبيقات باستخدام برنامج SPSS* . الرياض: مكتبة الرشد.

شراز، محمد صالح (٢٠٠٩). *التحليل الإحصائي للبيانات باستخدام برنامج المجموعة الإحصائية للعلوم الاجتماعية SPSS*. مكة المكرمة : جامعة أم القرى.

الشوربجي، أبو المجد و حسن، عزت .(٢٠١٢) *القياس والإحصاء التربوي والنفسي*. الرياض: مكتبة الرشد.

عبد الرحمن، سعد (١٩٨٣).*القياس النفسي* . الكويت : مكتبة الفلاح.

عباينة، عماد غصاب (٢٠٠٨). استقصاء اللاتغير في تقدير إحصائيات الفقرة باستخدام نظرية الاستجابة للفقرة والنظرية الكلاسيكية في القياس. *مجلة العلوم الانسانية*، ٣٩ ، ١ – ٢٦ .

عبدالوهاب، محمد محمود (٢٠١٠). *استخدام نماذج المفردة الاختبارية في تدريج بعض الاختبارات المعرفية*. (رسالة دكتوراه غير منشورة)، كلية التربية،جامعة المنيا.

عبد الوهاب، محمد محمود (٢٠١٣). المقارنة بين النظرية الكلاسيكية و النموذج ثلاثي البارامتر في معالجة اثر التخمين عند تقدير إحصائيات الأفراد و المفردات في اختبار اختيار من متعدد، *المجلة التربوية ـ الكويت*، ٢٨ (١٠٩)، ١٦٩-٢١١

عبد الوهاب، صلاح شريف (٢٠٠١). أثر بعض الطرق الوزنية لتقدير الدرجات على صدق الاختبارات مرجعية المحك ذات الاختيار من متعدد. *مجلة كلية التربية ببنها*، ١٢ (٤٩)، ٢٠٢-٢٥٥.

عبيدات، ذوقان (٢٠٠٣). *البحث العلمي - مفهومة أدواته أساليبه*. الكويت: إشراقات للنشر والتوزيع.

العساف، صالح أحمد (٢٠١٢). *المدخل إلى البحث في العلوم السلوكية*.ط٢. الرياض: دار الزهراء .

العدل، عادل محمد محمود (١٩٨٦). *أثر بعض طرق تقدير الدرجات للمفردات على ثبات الاختبارات ذات الاختيار من متعدد*.(رسالة ماجستير غير منشورة).كلية التربية، جامعة الزقازيق.

علام، صلاح الدين محمود(١٩٨٦). *تطورات معاصرة في القياس التربوية والنفسي*. جامعة الكويت : مطابع القبس التجارية.

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

علام، صلاح الدين محمود (٢٠٠٢). *القياس والتقويم التربوي والنفسي أساسية وتطبيقاته وتوجهاته المعاصرة*. القاهرة :دار الفكر العربي.

علام، صلاح الدين محمود (٢٠٠٧).*القياس والتقويم التربوي في العملية التدريسية*، عمان :دارس المسيرة.

عبد الحافظ، شحته عبد المولى (١٩٩٩). *تقويم بناء الاختبارات المرجعة إلى المحك المعيار فى ضوء نظرية الاستجابة للمفردة والنظرية التقليدية*.(رسالة دكتوراه غير منشورة). كلية التربية، جامعة عين شمس.

عودة، أحمد سليمان (٢٠٠٢). *القياس والتقويم في العملية التدريسية*. أربد: دار الأمل.

غنيم، أحمد الرفاعي (١٩٨٦). بعض طرق تقدير الدرجات للمفردات ذات الاختيار من متعدد، *مجلة كلية التربية بالمنصورة، مصر*،٧ (٥)، ٧٣-١٠٧.

كاظم، أمينة محمد (١٩٨٨). *دراسة نظرية نقدية حول القياس الموضوعي للسلوك (نموذج راش)*. سلسلة الكتب المتخصصة .الكويت: مؤسسة الكويت للتقدم العلمي.

اللحياني، عفاف راضي (٢٠٠٩). *أثر بعض طرق تقدير الدرجات للمفردات على ثبات وصدق درجات اختبار تحصيلي في الرياضات ذي الاختيار من متعدد لدى طالبات الصف الأول الثانوي بمكة المكرمة*. (رسالة ماجستير غير منشورة).كلية التربية، جامعة أم القرى.

محمود، سليمان محمد سليمان (١٩٨٦). *أثر بعض طرق تقدير الدرجات على ثبات الاختبارات المرجعية ذات الاختيار من متعدد*. (رسالة ماجستير غير منشورة). كلية التربية، جامعة الزقازيق.

ملحم، سامي محمد محمود (٢٠٠٢). *القياس والتقويم في التربية وعلم النفس*.ط ٢. عمان: دار المسيرة للنشر والتوزيع والطباعة.

مهرنز، وليم و لمن، ايرفن (٢٠٠٣). *القياس والتقويم في التربية وعلم النفس* (هيثم كامل الزبيدي وماهر أبو هلالة، مترجمان).العين :دار الكتاب الجامعي.

النبهان، موسى (٢٠٠٤). *أساسيات القياس في العلوم السلوكية*. عمان : دار الشروق للنشر والتوزيع.

النفيعي، عبد الرحمن عبد الله (٢٠١٢). الخصائص السيكومترية لاختبار المصفوفات المتتابعة المتقدم في ضوء نظرية الاستجابة للمفردة الاخبارية، *مجلة التربية* .*جامعة الأزهر*،١٤٧ (٢)، ١٧٥-٢١٤.

الهويدي، زيد (٢٠٠٤). *أساسيات القياس والتقويم التربوي* . العين : دار الكتاب الجامعي للنشر والتوزيع .

Adedoyin, O. (2010). Investigating the invariance of person parameter estimates based on classical test and item response theories. *International Journal of Science, 2* (2), 107-113.

Ainol, M.& Noor,L.(2006): Classical and rasch analyses of dichotomously scored reading comprehension test items. *Malaysian Journal of ELT Research,2*,1-20.

Andra.C. & Magnano.G. (2011). Multiple- choice math test :should we worry about guessing?. *Quaderni di Ricerca in Didattica (Mathematics),21*,235-243.

Ayhan, S.(2015) *Comparability of scores from CAT and paperand pencil implementations of student selection examination to higher education*,(Unpublished master dissertation).Bilkent University,Ankara.

Baradaran. A,(2009). The impact of correction for guessing formula on mc and yes/no vocabulary tests' scores, *The Journal of Applied Linguistics*, 2,(9), 80-98.

Ben-Simon,A,Budescu, D. V., & Nevo, B. (1997). A comparative study of measures of partial knowledge in multiple-choice questions. *Applied Psychological Measurement, 21*, 65-88.

Bradbard, D. A., & Green, S. B. (1986). Use of the coombs elimination procedure in classroom tests. *Journal of Experimental Education, 54*, 68–72.

Bush,m. (2001). A muliple choice test that rewards partial knowledge. *Journal of Further and Higher Education,25* (2), 157-163.

Cantrell, C.E. (1997*).*Item response theory: understanding the one, *Paper presented at the annual meeting of the southwest educational research association, Austin, TX*. (ERIC document reproduction service. No. ED 415 281)

Chang, S.H., Lin, P.C., & Lin, Z. C. (2007). Measures of partial knowledge and unexpected responses in multiple-choice tests. *Educational Technology & Society, 10* (4), 95-109.

De Finetti,B.(1965) Methods for discriminating levels of partial knowledge concerning a test item**,** *British Journal of Mathematical and Statistical Psychology,18*, 87-128.

Demars,C (2010).*Item response theory: understanding statistics measurement*. New York: Oxford University Press.

Fan, X. (1998). Item response theory and classical test theory: an empirical Comparison of their item/person statistics. *Educational and Psychological Measurement, 58*(3**),** 357-382.

Farland, M. Z., Barlow, P. B., Levi Lancaster, T. & Franks, A. S. (2015). Comparison of Answer-Until-Correct and Full-Credit Assessments in a Team-based Learning Course. *Am J Pharm Educ, 79*(2), 21.

Georgiev,N.(2008). Item analysis of C, D and E series from raven′s standard progressive matrices with item response theory two parameter logistic modle. *Europe's Journal of Psychology, 4* (3).

Hattie,J.(1985). Methodology review: assessing unidimensionality of tests and items, *Applied Psychological Measurement, 9* (2), 139- 164.

Hambleton, R. K. & Jonse, R. W (1993).Comparison of classical test theory and item response theory and their applications to test development. *Educational Measurement, 12*(4),38-47**.**

Hambleton, R. K., & Swaminathan,H.(1985). *Item response theory*: Principles and applications, New York: Springer Science + Business Media

Hambelton,R; Swaminathan,S &Rogers,H (1991**).** *Fundamentals of item response theory*, SAGE publications, Newbury Park: The international professional publishers.

James. D (1975) A preliminary study of the reliability and validity of a scoring procedure based upon confidence and partial information, *Journal of Educational Measurement,12* (2),129-133.

Jha,A.S (2014). *Social research methods*. New Delhi. McGraw-Hill Education.

Kansup, W& Hakstian, A. R(1975). A Comparison of Several Methods of Assessing Partial Knowledge in Multiple Choice Tests: I. Scoring Procedures. *journal of educational measurement,12* (4),219-230.

Kim.K.(2015*). Model comparisons among test response theories (TRT) on a reading comprehension test*, (Unpublished doctoral dissertation), Middle Tennessee State University.

Kurz, t. B. (1999). A review of scoring algorithms for multiple -choice tests. *paper presented at annual meeting of the Southwest Educational Research Association, january 21–23, San Antonio, tx.(ed428076) available at:* http://eric.ed.gov/?id=ED428076

Lau.P.N.K., Lau,S.H., Hong,K.S., & Usop, H. (2011). Guessing, partial knowledge, and misconceptions in multiple-choice tests. *Educational Technology& Society,14* (4), 99–110.

Lesage. E, Valcke. M, & Sabbe(2013). Scoring methods for multiple choice assessment in higher education – is it still a matter of number right scoring or negative marking?, *Studies in Educational Evaluation, 39*(3) 188 – 193**.**

Lisa G.O, Carolyn J.R, & Julie A.G,(2002). Applying systems design and item response theory to the problem of measuring information literacy skills. *Chicago, Statement of Ownership, Management, and Circulation.*

Messick, S(1995). Validity of psychological assessment validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, *50*,741-749.

Ndalichako, J.& Rogers, W. T. (1997). Comparison of finite score theory, classical test theory and item response theory in scoring multiple-choice items**,** *Educational and Psychological Measurement,* 57(4),580-589.

Onder,I.(2007). An investigation of goodness of model data fit model Veri Uyumunun Araştirilmasi. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi* (32), 210-220.

Progar, S & Socan, G. (2008). An empirical comparison of item response theory and classical test theory**.** *Psiholoska obzorja / Horizons of Psychology***,** *17*(3), 5-24.

Rand.R.(1981) Solving measurement problems with an answer-until-correct scoring procedure, *Applied Psychological Measurement* ,5(3) **,**399-414.

Rand, R.(1982) Some new results on an answer-until-correct scoring procedure, *Journal of Educational Measuremen,19*(1),67-74.

Rippey, R .M.(1970). A comparison of five different scoring function for confidence tests**,** *Journal of Educational Measurement,7*(3),165-170.

Roger,A.(1971). A comparison of the validities of conventional choice testing and various confidence marking procedures, *Journal of Educational Measurement, 8*(4),297-303.

Sabers,D &White,G.(1969). The effect of differential weighting of individual item responses on the predictive validity and reliability of an aptitude test**.** *Journal of Educational Measurement,6*(2),93-96.

Stage, C.(1999). *A comparison between item analysis based on item response theory and classical test theory*. a study of the SweSAT Sub-test READ. *Educational Measurement 33*,Umea University, Department of Educational Measurement.

Stage,C.(2003). Classical test theory or item response theory: the swedish experience. *Paper written for, and published in Spanish by Centro de Estudios Públicos, Santiago, Chile. Available at* www.cepchile.cl.

Stanley,S.(1975). Correlates of unwarranted confidence in responses to objective test items, *Journal of Educational Measurement,8*(1),15-20.

Thompson. L, (1976): A *study of the effect of an answer-until-correct multiple-choice procedure on mathematics achievement.*(unpublished doctoral dissertation), Oregon State University, Oregon.

Wiberg,M(2004). Classical test theory vs item response theory: an evaluation of the theory test in the swedish driving-license test. *University UMEA. EM No 50*.