

قياس أثر استعمال ثلاثة طرائق مختلفة لتمييز فقرات اختبار مرجعي المحك لمادة القياس والتقويم في معامل ثباته

أ.م.د. عبود جواد راضي

جامعة ولسط - كلية التربية

ABSTRACT

Measuring the effect of using three different Methods to distinguish the items of Referential test for the measure material and Evaluation in its constant coefficient.

Assist.Prof. Dr. Abood Jawad Radhi :University of Wassit/ College of Education /Psychological and Educational Science.

The goal of Recent Research to determine the effect of using three different methods to distinguish the test's items of Referential test of the measure material and evaluation in its Reliability Coefficient. The sample was practiced on a random sample consisted of (٨٠) students (female, male) about (٣٢ male student) and (٤٨ female student) of the fourth class student/Psychological and Educational department/college of Education/University of Wassit. The researcher had prepared a referential test for the measure and Evaluation materials for fourth class students/Psychological and Educational Department according to three stages which is :Analyzing stage , experiment and Construction .It is consisted of ١٤٤ an objective item of multi-choices, and in front of each item for alternates for answer which are(A,B,C,D),and the Psychometric features were resulted for it such validity, reliability and has used the equation (Kruskal - Wailis) to find out the differences in the Reliability coefficients, and analyze the results of the research that referred to that the values of Reliability coefficient for the three branch tests

that includes in this test ,and it doesn't refer to a great differences in the Reliability Coefficient according to the different methods to distinguish its items which is the Brennan (B)index Approach and Coefficient phi and agreement statistic.

This study has resulted that the value of Liability coefficient of referential test to the material of measure and Evaluation never affected by using three different methods to distinguish between its items.

الفصل الأول

التعريف بالبحث

مشكلة البحث :

أحس الباحث بمشكلة البحث الحالي من خلال خبراته المتراكمة في مجال التدريس الجامعي لأكثر من (٢٠) سنة إذ أسندت إليه مهام تدريس مادة علم النفس التربوي ومادة القياس والتقويم التي تتضمن فصولاً كاملة وموضوعات مقررة عن خصائص التقويم الجيد ومنها موضوع التمييز والثبات ومن خلال اطلاعه على كثير من مصادر علم النفس التربوي والقياس والتقويم ذات العلاقة باختصاصه وأيضاً من خلال اطلاعه على نتائج بعض الدراسات السابقة كدراسة الزهراني، ٢٠١٠ ودراسة عبدالله ٢٠١١ ودراسة القاطعي ٢٠١٢. وكذلك من خلال محاوراته ولقاءاته مع بعض المختصين في مجال علم النفس التربوي والقياس والتقويم بخصوص موضوع طرائق تمييز فقرات الاختبارات مرجعية المحك ومعامل الثبات ومن خلال مشاركته في كثير من لجان المناقشات العلمية لرسائل الماجستير وأطاريح الدكتوراه في اختصاص علم النفس التربوي والقياس والتقويم أتضح أن هناك عدة طرائق مختلفة في تمييز فقرات الاختبارات مرجعية المحك منها طريقة معامل التمييز لمجموعة الدارسين (المتقنين) وغير الدارسين (غير المتقنين) . وطريقة معامل فاي (ϕ) وطريقة معامل التمييز القبلي - البعدي ومعامل استجابة الفقرة. (item Berk , ٢٠١٠:١١١) Response) إذ يرى (براون Brown ٢٠٠١) إن هذه الطرائق تتباين بتباين المنطلقات النظرية التي تنبثق منها ، فبعض هذه الطرائق يعتمد على النظريات التقليدية في تمييز الفقرات إحصائياً وبعضها الآخر يعتمد على النظريات المعاصرة في القياس (Brown. ٢٠٠١; ١٤) فالطرائق التقليدية المتبعة في

تمييز الفقرات قد تكون غير ملائمة للاختبارات مرجعية المحك (Harris. ٢٠١٣: ١٠٥) ويرى (أوجر ٢٠٠٢ Ojer) أن الطرائق المعاصرة المتبعة في تمييز الفقرات إحصائياً قد تعطي نتائج طيبة ومثمرة وصادقة وموضوعية بالنسبة للخصائص الإحصائية لكل فقرة من فقرات الاختبار مرجعي المحك. (Ojer ٢٠٠٢: ٣٥) في حين يرى (بوفام ٢٠٠٥ Popham) أن هناك تباين واضح في أهداف وإجراءات الطرائق التقليدية في تمييز الفقرات إحصائياً ومنطلقاتها النظرية (Popham ٢٠٠٥: ٤٤) ويقترح (سيلفا Silva ٢٠٠٥) ضرورة تباع أساليب وطرائق جديدة ومعاصرة لتمييز فقرات الاختبارات مرجعية المحك بما ينسجم مع مبادئ النظريات المعاصرة في القياس والتقويم (٦٠-٥٩: Silva. ٢٠٠٥). وتشير أدبيات القياس والتقويم إلى أن هناك نوعين من الاختبارات مصنفة وفق معيار الاختبار هما الاختبارات مرجعية المعيار (Referenced tests) والاختبارات مرجعية المحك

(Criterion - Referenced tests) (النعيمي، ٢٠١٤: ٢٥٨) وعلى الرغم من الأهمية الكبيرة والفوائد العديدة للاختبارات مرجعية المعيار إلا أنها تعرضت لكثير من الانتقادات بخصوص أهدافها وإجراءاتها النظرية والتطبيقية (الطريري ٢٠٠٧: ٩١) ويشير بيرك Berk (٢٠١٠) إلى أن هناك كثيراً من الأسباب التي جعلت الاختبارات مرجعية المعيار غير ملائمة وغير مفيدة في مجال القياس والتقويم التربوي (Berk, ٢٠١٠: ٩٩) في حين أشارت نتائج دراسة (عبد الله، ٢٠١١) إلى أن الاختبارات مرجعية المحك هي اتجاه معاصر وذات فائدة كبيرة وأهمية واضحة وأنها أكثر ملاءمة وانسجاماً لمتطلبات القياس والتقويم التربوي المعاصرة من حيث أهدافها وإجراءاتها النظرية والتطبيقية (عبدالله، ٢٠١١: ٤٩) ويرى الباحث أن كثرة الانتقادات الموجهة إلى الاختبارات مرجعية المعيار قد أدت إلى ظهور اتجاهات معاصرة في مجال القياس والتقويم التربوي وهي الاختبارات مرجعية المحك . وبناء على ما تقدم يرى الباحث أن هذا التباين في طرائق تمييز الفقرات قد يؤدي إلى حدوث تباين في معامل الثبات نتيجة استعمال طرائق متباينة في تمييز فقرات الاختبار مرجعي المحك ، لذا ارتأى الباحث الخوض في غمار هذا الموضوع المهم الذي تتمثل مشكلة البحث فيه بالإجابة عن التساؤل الآتي :- ما أثر استعمال ثلاثة طرائق مختلفة لتمييز فقرات اختبار مرجعي المحك لمادة القياس والتقويم في معامل ثباته؟

أهمية البحث :-

إن عملية تحليل الفقرات هو أسلوب علمي منظم وهادف يصمم للحصول على بيانات ومعلومات واضحة ومحددة ودقيقة وموضوعية تتعلق بكل فقرة من فقرات الاختبار المراد إعداده أو بناءه أو تصميمه . وهذه البيانات والمعلومات يمكن الاستفادة منها في تحديد الفقرات الغامضة أو المربكة أو غير الفاعلة من أجل مراجعتها أو حذفها أو استبعادها أو تحسينها وإعادة صياغتها من جديد , وأنتقاء أفضل الفقرات المتوافرة لتضمينها في الصيغة النهائية للاختبار . وتعد عملية تحليل الفقرات أمراً ضرورياً لتحسين الاختبارات وبخاصة الاختبارات التحصيلية سواء المقننة أو الصفية التي يعدها المدرس لطلبته

(علام , ٢٠٠٦: ١١٢) وتساعد عملية تحليل الفقرات أيضاً في تنمية مهارات القائمين بأعداد وتصميم الاختبارات والمقاييس نتيجة الفحص الدقيق والمستنير للفقرات والتعرف على جوانب الضعف أو الخلل التي ربما تجعل بعض هذه الفقرات غير صالحة واستبقاء الفقرات الصالحة أو الجيدة التي تفي بخصائص تحقق الغرض من استعمال الاختبار أو المقياس في مجالات التقويم المختلفة وإعادة استعمالها إذا تطلب الأمر ذلك (علام , ٢٠١١ : ٢٦٧) . وتوضح لنا عملية تحليل الفقرات كيفية تطوير فقرات الاختبار من خلال إجراء مصمم الاختبار للمراجعة الفنية للفقرات وتحسينها بحيث تسهم كل فقرة أسهاماً إيجابياً فيما يقيسه الاختبار وبما يجعلها صادقة في قياس ما يرمي إليه الاختبار (الأنعمي , ٢٠١٤ : ٢٨٦) وأيضاً تتيح هذه العملية الفرص الكثيرة أمام معد أو مصمم الاختبار لتقويم جهوده وقدراته على أنتقاء أو اختيار أفضل الفقرات التي تؤدي الغرض الذي وضع من أجله ذلك الاختبار والعمل على تطوير هذه الجهود والقدرات لديه . (عبد المجيد وساجدة , ٢٠١٣ : ١٣٨) وأن الغرض الأساس من تحليل فقرات أي اختبار هو التحقق من بعض الخصائص التي يتكون منها ذلك الاختبار ومن أبرز هذه الخصائص التي يجب توافرها في الفقرات التي يتكون منها ذلك الاختبار هي القوة التمييزية بين المستويات العليا والدنيا في الظاهرة أو القدرة أو السمة أو الخاصية التي يقيسها الاختبار .(عبد المجيد وساجدة , ٢٠١٣ : ١٤٣) وأن القوة التمييزية أو معامل التمييز تعطينا مؤشرات عن مستوى الثقة والاطمئنان بقدرة فقرات الاختبار عن الكشف الدقيق عن الفروق الفردية بين الأفراد المفحوصين في الخاصية أو السمة أو الظاهرة التي يهدف الاختبار لقياسها .(Gronlund, ٢٠١٢: ٢٣٣) . وتفيدنا أيضاً هذه العملية في اتخاذ القرار النهائي بخصوص وضعية فقرات الاختبار من حيث بقائها أو حذفها واستبعادها أو تعديل صياغتها وتحسينها فاستناداً إلى معيار القوة التمييزية يتخذ مصمم الاختبار الموقف الخاص أو القرار النهائي لكل فقرة من فقراته , فالفقرة المقبولة هي تلك التي

تتراوح قوتها التمييزية أكبر من (٠.١٩) بحسب معيار أيبيل (Ebel) الذي يعد أفضل معيار لتحديد القوة التمييزية للفقرات. (النعيمة، ٢٠١٤: ٢٨٦) إذ يعد معامل تمييز الفقرة أو قدرته على التمييز دليلاً على صدقه وخاصة إذا كان الأمر ينطوي على مقارنة طرفي القدرة التي تقيسها الفقرة (عبد الرحمن، ١٩٩٨: ٢١) وان حساب قوة تمييز كل فقرة (power of Discrimination) سيهيئ الطريقة الواضحة والدقيقة والموضوعية للحصول على اختبار صادق في حالة ارتفاع معاملات تمييز تلك الفقرات (٢١: ٢٠٠٥، Silva). ويمكن الاستفادة من نتائج عملية تحليل الفقرات إحصائياً في التحقق والتأكد من معامل الصعوبة والسهولة ومعامل التمييز ومدى فاعلية بدائل الإجابات في فقرات الاختيار من متعدد (عبد المجيد وساجدة ٢٠١٣: ١٤٣). ويتطلب بناء الاختبار التحصيلي الموضوعي الذي يتضمن فقرات الاختيار من متعدد إجراء تحليل الفقرات لمعرفة مدى صعوبة أو سهولة كل فقرة ومدى فاعليتها وقدرتها على التمييز في الفروق الفردية للصفة أو الخاصية المراد قياسها (الامام وآخرون، ٢٠٠٥: ١٠٦) وتعد الاختبارات مرجعية المحك من الاتجاهات المعاصرة في القياس التربوي إذ تفسر درجة المتعلم في ضوء مدى تحقيق أهداف محددة مسبقاً (علي، ٢٠٠١: ٥١-٥٢) وتسعى هذه الاختبارات إلى تحديد مستوى الطالب (المتعلم) بالنسبة إلى محك (مستوى) معين ومحدد مسبقاً دون الرجوع إلى أداء فرد آخر أو أداء الأفراد الآخرين (العزاوي، ٢٠٠٨: ٨٦) ولذا فأنها لا تهتم بالمركز النسبي للفرد بين زملائه في الصف الواحد وإنما تهتم بمقارنته درجته في الاختبار بمستوى أداء (محك) يدل على المستوى المطلوب والمقبول لهذا الأداء فإذا كانت درجة الفرد (المتعلم) تفوق أو تعادل مستوى الأداء (المحك) المحدد فإن الفرد المتعلم يكون قد حقق المستوى المطلوب والمرغوب فيه إما إذا كانت درجة أقل من المستوى المحك المحدد مسبقاً فإن أداءه يعد ضعيفاً وغير مقبول. (علام، ٢٠٠٠، ٢٦٢). ويمكن استعمال الاختبارات مرجعية المحك في وصف أداء الفرد (المتعلم) وتصنيف مستوى إتقانه أو تمكنه لمجموعة من الكفايات الأدائية الخاصة بأداء المهارات أو الفعاليات (الدوسري، ٢٠١٢: ١٠٨) وتساعد نتائج الاختبارات مرجعية المحك على تشخيص حالات التفوق الدراسي والتخلف الدراسي وصعوبات التعلم من قبل الهيئات التعليمية والتدريسية لأن الأداء يقارن بالمستوى المطلوب والمقبول (المحك) وليس بمتوسط الجماعة أو الأقران (النعيمة، ٢٠١٤: ٢٦٠) وهذا المستوى (المحك) يسهم بتطوير الاختبارات الصفية التي يقوم المدرس أو المعلم بأعدادها والعمل على زيادة وتطوير مستوى الكفايات الخاصة بأداء المهارات أو القدرات أو الاستعدادات تبعاً لدرجة التطور الحاصل في مختلف أصناف العلم والمعرفة (القاطعي، ٢٠١٢: ٢١). وتعد مادة القياس والتقويم من المواد الدراسية

الأساسية والمهمة ضمن المناهج الدراسية في كليات التربية بجميع أقسامها الدراسية لكونها تزود الطلبة بمعلومات وبيانات كافية وواقية عن المبادئ الأولية والأساسية في الاختبارات التحصيلية بأنواعها وتمدهم بمعلومات كثيرة عن خصائص الاختبار الجيد والتي من أبرزها هي الموضوعية (objectivity) والملاءمة (Relevance) والشمولية (Compehcnsion) والصدق (Validity) والسهولة (Praeticability) والوضوح (Clearence) والثبات (Reliability) (النعيمي، ٢٠١٤: ٢١٢) .

وللثبات أهمية خاصة في اختيار واستعمال اختبار ما لغرض معين إذ يشير إلى مستوى الثقة الذي يمكننا ان نضعه في نتائج اختبار اتنا فالغاية الرئيسة لاختبار الأفراد (المفحوصين) هو الوصول إلى أحكام أو قرارات تتعلق بهم ونطمئن إليها . لذا يجب أن تقوم هذه الأحكام والقرارات بدورها على أدوات ووسائل دقيقة وموضوعية في قياسها وموثوق فيها فإذا كانت أداة قياسنا غير ثابتة فأنا نشك في الأحكام المعتمدة عليها (عمر وآخرون، ٢٠١٠: ٢١٥) . وصمم معامل الثبات

(Reliability Coefficient) ليقدّم تقديراً كمياً (رقمياً) لاتساق القياس ودقته لذلك فمعامل الثبات قد يؤدي غرضين الأول أنه يعبر عن دقة الاختبار نفسه كأداة للقياس والثاني أنه يقدم تقديراً للاتساق (Consistency) في أداء المفحوص على الاختبار (عمر وآخرون، ٢٠١٠: ٢١٥) فالاختبار الذي يتميز بالاتساق يعد شرطاً ضرورياً للاختبار الجيد (Nitko، ١٩٨٣: ٨٦) ومن الحقائق الإحصائية والمنطقية في القياس والتقويم هي " أن الاختبار لا يمكن إن يكون صادقاً إلا إذا كان ثابتاً فمعرفة معامل الثبات لاختبار معين في موقف معين يسمح لنا بمعرفة الحدود التي لا يمكن للصدق أن يتخطاها في هذا الموقف وأيضاً أن معرفة معامل الثبات تسمح لنا بمعرفة الخطأ المحيط بالاختبار أي درجة الدقة والثقة التي يمكن في ضوءها أن تفسر الدرجة التي حصل عليها الفرد المفحوص " (Brad Field ، ٢٠٠٨: ١١١) وان ثبات درجات الاختبارات يمكن أن يؤثر في نوعية القرارات المتعلقة بالأفراد أو الجماعات (علام، ٢٠٠٦: ٨٩).

لذا تتجلى أهمية هذا البحث في جانبيه النظري والتطبيقي في الآتي :-
١- الأهمية النظرية :-

١- يمثل هذا البحث إضافة نوعية جديدة للدراسات العراقية والعربية في مجال القياس والتقويم وإثراء وأغناء الجانب النظري في هذا المجال نظراً لقلّة الدراسات التي بحثت في هذا الموضوع المهم على مستوى الدراسات العربية وبالأخص الدراسات العراقية .

- ٢- يكتسب هذا البحث أهميته من أهمية دراسة طرائق تمييز الفقرات التي تزودنا بمعلومات متكاملة ومناسبة وواضحة ودقيقة وموضوعية عن كل فقرة من فقرات الاختبار المراد تصميمه أو أعداده بهدف الإبقاء على هذه الفقرات أو حذفها أو تعديل مضمونها أو تحسينها بما ينسجم مع محتوى أو مضمون ذلك الاختبار .
- ٣- تنبثق أهمية هذا البحث من أهمية الاختبارات مرجعية المحك التي تقدم لنا معلومات متكاملة ومناسبة تتعلق بمستوى كفاية وتمكن الفرد في مجال معين وأيضاً تفيدنا في تقويم الأفراد والبرامج التربوية والتعليمية المختلفة .
- ٤- يكتسب هذا البحث أهميته من أهمية مادة القياس والتقويم التي تعد مادة دراسية أو مقرر دراسي مهم يدرس في جميع الأقسام الدراسية في كليات التربية ومنها قسم العلوم التربوية والنفسية.
- ٥- تنبثق أهمية هذا البحث من أهمية التعرف على معامل ثبات الاختبار إذ أن أي أداة قياس ينبغي أن تتوافر فيها بعض الخصائص الفنية والسايكومترية الأساسية لعل من أهمها الثبات والذي يعد من المفاهيم الأساسية في القياس التربوي

ب- الأهمية التطبيقية :-

- ١- يسهم هذا البحث في تعريف طلبة الدراسات العليا في اختصاص علم النفس التربوي والقياس والتقويم الذين يرومون أعداد أو بناء أو تصميم اختبارات مرجعية المحك بالطرائق الفاعلة في تمييز فقرات هذه الاختبارات والتي قد تؤدي إلى زيادة أو ارتفاع معامل الثبات فيها .
- ٢- يسهم هذا البحث في تبصير مصممي الاختبارات مرجعية المحك بأبرز مزايا وعيوب هذه الاختبارات كي يتمكنوا من مراعاتها ووضعها في الحساب عند بناء الاختبارات مرجعية المحك.

هدف البحث :- يرمي البحث الحالي إلى التعرف على أثر استعمال ثلاثة طرائق مختلفة لتمييز فقرات اختبار مرجعي المحك لمادة القياس والتقويم في معامل ثباته .

حدود البحث :- يقتصر إجراء البحث الحالي على ما يأتي :

- ١- طلبة الصف الرابع من كلا الجنسين (ذكور , إناث) في / قسم العلوم التربوية والنفسية بجامعة واسط المستمرين بالدراسة الصباحية للعام الدراسي ٢٠١٣/٢٠١٤ .

٢- مادة القياس والتقويم للصف الرابع في قسم العلوم التربوية والنفسية / كلية التربية / جامعة واسط

٣- ثلاثة طرائق مختلفة لتمييز فقرات اختبار مرجعي المحك لمادة القياس والتقويم وهي :

أ. طريقة هارس وسبكوفياك لمعامل التوافق المرجعي Harris and

Subkovaik Method for Agreement Statistic Index .

ب. طريقة برينان لمعامل التمييز The Brennan Method for

Discrimination Index

ج. طريقة فاي لمعامل التمييز The Phi method for Discrimination

Index

تحديد للصلحات :-

أولاً :- القياس : Measurement

عرفه كل من : * براد فيلد (Bradfield) ٢٠٠٨ بأنه ((عملية تحديد النواحي الكمية المرتبطة بأبعاد السمة أو الظاهرة المراد قياسها ليتسنى وصفها بدقة)) (Bradfield. ٢٠٠٨:٩٨) . * مادوس (Madaus) ٢٠١٠ : بأنه ((عملية تحديد الجوانب الكمية التي تصف سمة أو خاصية أو ظاهرة معينة طبقاً لقواعد محددة تحديداً جيداً)) (Madaus. ٢٠١٠:٤٨)

التعريف النظري :- يتبنى الباحث تعريف (براد فيلد Bradfield) ٢٠٠٨ لكونه ينسجم مع أهداف البحث الحالي وإجراءاته .

ثانياً :- التمييز Discrimination عرفه كل من : * الأمام وآخرون ، ٢٠٠٥ ، بأنه ((قدرة الفقرة على تحديد مستوى الفروق الفردية بين الأفراد المفحوصين الذين يملكون الصفة أو يعرفون الإجابة الصحيحة وبين الذين لا يملكون الصفة المقاسة أو لا يعرفون الإجابة الصحيحة لكل فقرة من فقرات الاختبار)) (الامام وآخرون ، ٢٠٠٥:١١٤) . * النعيمي (٢٠١٤) بأنه ((مدى قدرة الفقرة عن إمكانية قياس الفروق الفردية بين الأفراد المتقنين (الممتازين) في الخاصية أو السمة أو الظاهرة التي يقيسها الاختبار وبين الأفراد غير المتقنين (الضعاف) في تلك الخاصية أو السمة السمة أو الصفة بواسطة فقرات الاختبارات مرجعية المحك)) . (النعيمي،

٢٠١٤ : ٢٨٣). * فورآن (Foraan) ٢٠١٣ : بأنها ((العملية التي فيها تحديد قدره فقرات الاختبار عن الكشف الدقيق لمستوى الفروق الفردية بين المفحوصين في الخاصة أي السمة أو الظاهرة المراد قياسها من خلال إجراء فحص أو اختبار استجابات المفحوصين عن كل فقرة من فقرات الاختبار لكي يستطيع الباحث انتقاء الفقرات المناسبة التي تميز بين المفحوصين المتميزين (المتكئين من الإجابة الصحيحة) والضعفاء (غير المتكئين من الإجابة الصحيحة) في تلك السمة أو الخاصية في الاختبار بصيغته النهائية وحذف الفقرات غير المميزة أو تعديلها)) (Foraan. ٢٠١٣: ١٨٩- ١٩٠)

* علام (٢٠١١) : بأنه ((مدى إمكانية قياس الفروق الفردية بواسطة فقرات الاختبارات مرجعية المحك (علام, ٢٠١١ : ٢٧٧) .

التعريف النظري :- يتبنى الباحث تعريف فورآن (Foraan) ٢٠١٣ لكونه ينسجم مع أهداف البحث الحالي وإجراءاته .

التعريف الإجرائي :- هو مدى قدرة فقرات اختبار مرجعي المحك لمادة القياس والتقويم عن الكشف لمستوى الفروق الفردية وقياسها بين طلبة الصف الرابع المتقنين وغير المتقنين من خلال إجراء فحص لأستجاباتهم عن كل فقرة من فقرات الاختبار باستعمال الطرائق الثلاثة الآتية :

١. طريقة هارس وسبكوفايك للتوافق المرجعي : Harris and

Subkovaik Method for Agreement Statistic

وهي عبارة عن معرفة احتمالية التوافق (الاتساق أو الانسجام) بين مستوى الإجابات عن الفقرة ومستوى التمكن (الإتقان) (متقنين وغير متقنين) وتعتمد هذه الطريقة على بيان عدد المفحوصين المتقنين (المتكئين) الذين أجابوا عن الفقرة إجابة صحيحة وعدد المفحوصين غير المتقنين (الضعفاء أو غير المتكئين) الذين أجابوا عن الفقرة إجابة خاطئة .

ب. طريقة برينان لمعامل التمييز The Brennan Method for

Discrimination Index

وهي عبارة عن الفرق في صعوبة الفقرة بين مجموعة المتقنين (المتميزين) أو المتكئين الذين أجابوا عن الفقرة إجابة صحيحة ومجموعة غير المتقنين (الضعفاء أو غير المتكئين) الذين أجابوا عن الفقرة أجا به صحيحة .

ج. طريقة فاي لمعامل التمييز The Phi method for Discrimination

Index

وهي عبارة عن توافق (أنساق) الفقرة والاختبار في تصنيف المفحوصين إلى مجموعتين هما مجموعة الإتقان أو المتفنين أو المتمكنين ومجموعة غير المتفنين أو غير المتمكنين (الضعفاء) وتعتمد هذه الطريقة على عدد الإجابات الصحيحة والخاطئة للمتفنين وعدد الإجابات الصحيحة والخاطئة لغير المتفنين

ثالثاً : الاختبار مرجعي المحك Criterion – Referenced test عرفه كل من :-

* بوفام (Popham) ٢٠١٠ بأنه ((ذلك الاختبار الذي يستعمل لتحديد وتفسير مستوى أداء الفرد وموقعه بالنسبة لمجال سلوكي (Behavioral Domain) معين ومحدد مسبقاً بدقة وموضوعية ويتضمن هذا المجال السلوكي مجموعة من المهارات والمعارف التي يؤديها المفحوص في الموقف الاختباري (Popham. ٢٠١٠:٤) . * كريكوري ٢٠١٢ Gregory بأنه :- ((ذلك الاختبار الذي يسعى إلى تحديد وضع أداء الفرد في الاختبار (مجال سلوكي معين) وينسب ذلك الأداء إلى مجموعة محددة ومعينة من المحكات السلوكية كالكفايات الأدائية للاختبار نفسه ويمثل المحك هنا مستويات الأداء المقبولة واللازمة للقيام بمهمة معينة)) . (Gregory. ٢٠١٢:٢١) .

التعريف النظري :- يتبنى الباحث تعريف كريكوري (Gregory ٢٠١٢) لكونه ينسجم مع أهداف البحث الحالي وإجراءاته .

التعريف الإجرائي :- هو اختبار يقيس مجموعة من المهارات (الكفايات) التي تتضمنها مادة القياس والتقويم ومنها المهارات النظرية (الوصفية والمنطقية) والعملية أو التطبيقية (الأدائية) .

رابعاً : مادة القياس والتقويم :- عرفه الباحث إجرائياً بأنها مقرر دراسي محدد من قبل اللجنة القطاعية لكليات التربية ولأقسام العلوم التربوية والنفسية ويتكون هذا المقرر من إحدى عشر فصلاً ويسعى إلى تعريف طلبة الصف الرابع في قسم العلوم التربوية والنفسية بأبرز أسس القياس والتقويم وتزويدهم بمهارات بناء وأعداد بعض أدوات ووسائل القياس والتقويم وكيفية استعمالها في الحياة العملية المدرسية

خامساً: معامل الثبات :- Reliability Coefficient عرفه كل من : * سبكوفايك (subkoviak) (٢٠٠٢) " بأنه تقدير كمي لاتساق الاختبار ودقته كأداة للقياس في أداء الفرد المفحوص عن ذلك الاختبار والمتعلق باتساق تصنيف الأفراد (المفحوصين) وفقاً لدرجة تمكنهم من نطاق سلوكي معين تقيسه فقرات الاختبار " (Subkoviak. ٢٠٠٢:١) * علام ٢٠١١. بأنه ((تقدير كمي لاتساق الاختيار ودقته لاتخاذ قرارات بشأن تصنيف الأفراد المفحوصين إلى متقنين وغير متقنين اعتماداً على درجاتهم في الاختبار في ضوء درجة فاصلة)) (علام, ٢٠١١ : ٢٧٠) .
التعريف النظري : يتبنى الباحث تعريف (علام), ٢٠١١ لكونه ينسجم مع أهداف البحث الحالي وإجراءاته .
التعريف الإجرائي :- هي الدرجة التي يتوصل إليها الباحث من خلال استعمال معامل الاتساق (الاتفاق أو التوافق) لسبكوفايك (Pc) Subkoviak Coefficient of Agreement لحساب قيمة معامل ثبات الاختبار مرجعي المحك بعد إجابة أفراد عينة البحث عن فقرات هذا الاختبار .
سادساً : الدرجة الفاصلة : Class Interval أو درجة القطع Cut-off Score عرفها كل من :-

*فوران ٢٠١٢ Foraan بأنها الدرجة التي يصنف الأفراد (المفحوصين) في ضوءها إلى ناجحين (متقنين) أو راسبين (غير متقنين) في اختبار ما كشرط للإتقان (٤٩ : ٢٠١٢ . Foraan) : * شانون Shannon ٢٠١٢ بأنها ادني مستوى للأداء المطلوب والمقبول والمحدد مسبقاً كشرط للإتقان أو الإتقان (Shannon. ٢٠١٢ : ١١٢)

التعريف النظري : يتبنى الباحث تعريف فوران Foraan ٢٠١٢ لكونه ينسجم مع أهداف البحث الحالي وإجراءاته
التعريف الإجرائي : بأنها الدرجة التي في ضوءها يمكن تصنيف الأفراد (المفحوصين) إلى متقنين (متمكنين) وغير متقنين (غير متمكنين) في الاختبار المستعمل في هذا البحث وفقاً لطريقة أنجوف (Angoff) في تحديد هذه الدرجة .

الفصل الثاني

خلفية نظرية ودراسات سابقة

أولاً : خلفية نظرية :- نبذة تاريخية عن نشوء الاختبارات مرجعية المحك :

ساهمت مقالة جليزر (Glaser) ١٩٦٣ حول الاختبارات مرجعية المحك في ظهور فلسفة بناء الاختبارات التحصيلية إذ حفزت المشتغلين بالقياس والتقويم إلى الاهتمام بتطوير أساليب التقويم باستعمال الاختبارات مرجعية المحك (Gronlund, ٢٠١٢: ٥).

ويمكن أرجاع مصطلح (الاختبارات مرجعية المحك) إلى العالم جليزر وتعد البداية الحقيقية لهذه الاختبارات في بداية السبعينيات في القرن العشرين على يد كل من جيمس بوفام (James Popham) وهوسك (HOSIK) اللذين حددا استراتيجيات ومضامين ومبادئ القياس مرجعي المحك , الأمر الذي أدى إلى زيادة اهتمام علماء القياس بهذا النوع من الاختبارات (علام : ٢٠٠٦ : ٥٣) وترتب على ذلك ظهور حركة بحثية نشطة وفاعلة وكبيرة منذ ذلك الوقت حتى الوقت الحاضر إذ اهتمت بدراسة الجوانب النظرية -السيكو مترية الخاصة بهذا المنحى الجديد (إبراهيم , ٢٠١٠ : ٩٩).

استعمالات الاختبارات مرجعية المحك :

- ١- تقويم أداء الفرد (المفحوص) كل حسب قدراته واستعداداته .
- ٢- قياس إتقان (تمكن) المهارات أو الكفايات الأساسية لمجال معين سواء كان مجالاً تعليمياً أو مهنيّاً .
- ٣- تحديد مستوى إتقان الأفراد (المفحوصين) للمتطلبات المحددة مسبقاً للبدء في درس جديد أو مادة جديدة .
- ٤- الحكم على سلوك الفرد (المفحوص) بالنسبة إلى قدراته واستعداداته وإمكانياته الذاتية ، لا بالنسبة لقدرات وإمكانيات الآخرين وهذا يمثل قيمة تربوية كبيرة لتعديل وتغيير في بعض البرامج التربوية والتعليمية (علام ، ٢٠١١ : ٣٢٩)
- ٥- تستعمل الاختبارات مرجعية المحك في التعليم المبرمج واختبارات الكفايات و الإِتقان (التمكن) .
- ٦- معرفة وتحديد المستويات المختلفة لأداء الأفراد (المفحوصين) مما يمكن الهيئات التدريسية من تشخيص حالات التفوق الدراسي وحالات التأخر الدراسي وتشخيص الصعوبات والمشاكل التي قد يواجهها بعض المتعلمين .

- ٧- تنظيم المتعلمين في مجموعات (Groups) من أجل التدريس على أساس ما يعرفونه وما لا يعرفونه .
- ٨- التعرف على المعلومات والمهارات والكفايات الأساسية التي ينبغي أن يعرفها الطالب في مادة دراسية معينة أو مجال معين بصرف النظر عما يعرفه أقرانه أو أفراد مجموعة معيارية .
- ٩- تطوير الاختبارات الصفية التي يعدها المدرس لطلبته لتزيد من فاعليتها في تقويم كفاياتهم ومكتسباتهم (عمر واخرون ، ٢٠١٠ : ١٠٢) .
- ١٠- أن الاختبارات مرجعية المحك تمثل رسائل جديدة تمكن المدرسين من الحصول على معلومات أكثر تفصيلاً وتحديداً عن مستوى أداء المتعلم كفرد متميز بحيث يمكن استعمالها في اتخاذ القرارات التربوية المتعلقة بالمتعلمين وكذلك في تقويم البرامج التربوية والتعليمية (الطريزي ، ٢٠٠٧ : ٣٨) .
خصائص الاختبارات مرجعية المحك :
- ١- أن الاختبارات مرجعية المحك تعمل على تحديد مستوى أداء الفرد (المفحوص) في ضوء مستويات معينة من الإتقان (التمكن) محددة مسبقاً
- ٢- أن الاختبارات مرجعية المحك تركز على مجموعة محددة ومعينة من سلوكيات المتعلم مما يجعل المدرس قادراً على أن يحدد ما إذا كان طلبته أتقنوا (تمكنوا) مهارة أو كفاية ما أم لم يتقنوها (لم يتمكنوا منها) .
- ٣- أن فقرات الاختبارات مرجعية المحك تكون ذات صعوبة تتناسب مع المهام التعليمية المطلوبة والمحددة مسبقاً .
- ٤- أن درجات الاختبارات مرجعية المحك تكون على شكل نسب مئوية من الإجابة الصحيحة أو على شكل مؤشر يدل مدى التمكن أو الإتقان من التعلم أو من عدمه .
- ٥- ترتبط الاختبارات مرجعية المحك بالتعليم الفردي أو التدريس على شكل مجموعات صغيرة (Small groups) .
- ٦- تستعمل الاختبارات مرجعية المحك في الاختبارات التحصيلية فقط وليس في الاختبارات النفسية .
- ٧- أن مدى التباين (الفروق الفردية أو الاختلافات) يصبح أقل ما يمكن أن يكون عليه عندما يتقن (يتمكن) جميع الطلبة ، العناصر الأساسية للموقف التعليمي .
- ٨- يتم تفسير درجة الفرد (المفحوص) بالرجوع إلى الدرجة الفاصلة أو درجة القطع وليس بالرجوع إلى درجات الزملاء أو الأقران في الصف ذاته لان أداء كل

فرد (مفحوص) يتم مقارنته بمستوى أداء محدد ومعين مسبقاً وليس له علاقة بأداء زملائه في الصف الدراسي الذي ينتمي إليه (عبدالله , ٢٠١١ : ٩٩)
طرائق تمييز فقرات الاختبارات مرجعية المحك :-

١- طريقة معامل التمييز القبلي – البعدي Preinstruction Postinstruction

measurements Method

وتسمى هذه الطريقة أحيانا بمقياس حساسية الفقرات أو المفردات لعملية التدريس أو التعليم واستحدثت من قبل العالمين كوكس (Cox) وفرجاس (Varges) في عام (١٩٦٣) وأصبحت من الطرائق الرئيسة في تمييز فقرات الاختبارات مرجعية المحك (Berk, ٢٠١٠: ٥٣) ويعتمد هذا المعامل على تطبيق الاختبار مرتين على مجموعة واحدة من الطلبة (المفحوصين) أحدهما قبل التعليم أو التدريس (قبلي) والأخرى عقب الانتهاء منه (بعدي) وتكوين مصفوفة تدون في خلاياها الدرجة التي حصل عليها كل فرد (مفحوص) في كل فقرة من فقرات الاختبار وتعطى الدرجة (١) إذا كانت الإجابة عن الفقرة أجابه صحيحة والدرجة (صفر) إذا كانت الإجابة عن تلك الفقرة أجابه خاطئة أو متروكة ويفضل إعطاء وقت كاف للأفراد المفحوصين للإجابة عن جميع الفقرات (علام , ١٩٩٥ : ١٥٨) ويحسب هذا المعامل بطرح نسبة الأفراد المفحوصين الذين أجابوا عن الفقرة إجابة صحيحة في الاختبار أو التطبيق القبلي من نسبة الأفراد المفحوصين الذين أجابوا عن تلك الفقرة أجابه صحيحة في الاختبار أو التطبيق البعدي (عمر واخرون ، ٢٠١٠ : ٧٠) . ومن النقد الموجه إلى هذه الطريقة هو صعوبة تطبيق الاختبار البعدي (الإعادة) ولكن يمكن التغلب على مشكلة الإعادة (التطبيق البعدي) عن طريق بناء أو أعداد مجموعتين متكافئتين من الفقرات أو أطالة المدة الزمنية الفاصلة بحدود مناسبة وبشرط أن لا يسمح لعوامل النضج بالتأثير على الدرجات (الطريبي ، ٢٠٠٧ : ١٥٩) .

٢- طريقة معامل التمييز لمجموعة الدارسين (المتلقين للتعليم) وغير الدارسين (غير المتلقين للتعليم) .

The Unin Structed – Instructed Groups Approach وهي من الطرائق المستعملة في تمييز فقرات الاختبارات مرجعية المحك وتعتمد على اختيار مجموعتين مختلفتين من الأفراد في وقت واحد إذ يمكن اختيار المجموعة الأولى من الأفراد (المفحوصين) الذين تلقوا تعليماً فاعلاً في أحد الصفوف الدراسية وإنهم حققوا أهداف الوحدة التعليمية التي يقسيها الاختبار أما المجموعة الثانية فيمكن اختيارها من بين الطلبة (المفحوصين) الذين لم يتلقوا تعليماً في هذه الوحدة التعليمية

ويطبق عليهم الاختبار في الوقت نفسه . ويجب أن تكون هاتان المجموعتان (الدارسين وغير الدارسين) متشابحتين قدر الإمكان في مستوى القدرة وفي جميع الخصائص الأخرى ذات العلاقة بموضوع وطبيعة الاختبار ويتم حساب معامل التمييز لكل فقرة بطرح نسبة الأفراد المفحوصين الذين أجابوا إجابة صحيحة في مجموعة غير الدارسين من نسبة الأفراد المفحوصين الذين أجابوا إجابة صحيحة في مجموعة الدارسين (١٠٢ : ٢٠٠٥ . popham). ومن أبرز الانتقادات التي وجهت إلى هذه الطريقة هو صعوبة تعريف المحك الذي جرى في ضوئه اختيار كل من المجموعتين (الدارسين وغير الدارسين) تعريفاً إجرائياً وكذلك مشكلة التأكد والتحقق من عملية تكافؤ وتشابه هاتين المجموعتين في الخصائص ذات العلاقة بموضوع وطبيعة الاختبار (الزهراني : ٢٠١٠ : ٨٩) .

٣- طريقة معامل التوافق المرجعي Agreement Statistic Method

تعد هذه الطريقة من أحدث طرائق حساب معامل التمييز الخاصة بتحليل فقرات الاختبارات مرجعية المحك ويهدف معامل التوافق المرجعي إلى معرفة احتمالية التوافق (الاتساق أو الانسجام) بين نواتج فقرة معينة ونواتج الاختبار ككل وهو مشابه في قاعدته النظرية لطريقة معامل استجابة الفقرة (Latent Trail Statistic) . وتعتمد هذه الطريقة على تطبيق الاختبار مرة واحدة (دفعه واحدة) على مجموعة واحدة من الأفراد المفحوصين ومن ثم يتم تصنيف أفراد هذه المجموعة إلى صنفين أو مجموعتين هما مجموعة الأفراد المتقنين ومجموعة الأفراد غير المتقنين بناءً على مدى تحقيقهم للمستوى (المحك) المطلوب للإتقان (التمكن) . ويحسب معامل التوافق المرجعي وفق معادلة هارس (Harris) وسبكوفياك (Subkoviak) والتي تنص على ما يأتي :-

$$\frac{أ + د}{ن}$$

معامل التوافق المرجعي = $\frac{أ + د}{ن}$ إذ أن (أ) يشير إلى عدد الأفراد (المفحوصين) المتقنين الذين أجابوا عن الفقرة إجابة صحيحة وأن (د) يشير إلى عدد الأفراد (المفحوصين) غير المتقنين الذين أجابوا عن الفقرة إجابة خاطئة وأن الرمز (ن) يشير إلى العدد الكلي للأفراد (المفحوصين) في كلا المجموعتين ويمكن عد الفقرة جيدة وفق معامل التوافق المرجعي إذا كان الفرق بين الحد الأدنى لمعامل التوافق المرجعي ومعامل التوافق المرجعي أكبر من أو يساوي (٠.٠٥) (Subkoviak , ٢٠٠٢ : ١٧)

٤- طريقة معامل فاي Phi Method تهدف هذه الطريقة إلى تحديد درجة التوافق في التصنيف بين الفقرة والاختبار للأفراد المفحوصين وتعتمد هذه الطريقة على تطبيق الاختبار مرة واحدة على مجموعة واحدة من الأفراد (المفحوصين) ويتم اختيار وتحديد درجة فاصلة (درجة القطع) تمثل مستوى الإتقان وتتحدد فاعلية الفقرة بقدرتها على التمييز بين المفحوصين عند درجة فاصلة ومحددة على الدرجة الكلية للاختبار (الأحمدي ، ٢٠١٢ : ٨٥) .

ويتم إيجاد معامل فاي عن طريق الجدول الثنائي (٢×٢) إذ يبين هذا الجدول عدد الإجابات الصحيحة والخاطئة للمتقنين وعدد الإجابات الصحيحة والخاطئة لغير المتقنين (الشربيني ، ٢٠١٠ : ٣٢) وتعد الفقرة جيدة وفق معامل فاي إذا كانت قيمته أكبر من أو تساوي (٠.٣٠)

(Harris, ٢٠٠٩:٤١) .

٥- طريقة معامل التمييز / برينان أو معامل التمييز (ب) The Brnnan Index Method

أو The (B) Index Approach

وهي من الطرائق الحديثة المستعملة في تمييز فقرات الاختبارات مرجعية المحك إذ يتم حساب معامل التمييز فيها من خلال إيجاد الفرق في تمييز الفقرة بين مجموعة المتقنين (المجموعة العليا) ومجموعة غير المتقنين (المجموعة الدنيا)

وفق معادلة برينان Brennan التي تساوي

$$\frac{أ}{ب} = \frac{ن١}{ن٢}$$

إذ أن الرمز (أ) يشير إلى عدد المتقنين الذين أجابوا عن الفقرة إجابة صحيحة وأن الرمز (ب) يشير إلى عدد غير المتقنين الذين أجابوا عن الفقرة إجابة صحيحة وأن الرمز (ن١) يشير إلى عدد المتقنين وان الرمز (ن٢) يشير إلى عدد غير المتقنين وهنا لا بد من الإشارة إلى أن (ن١) يمكن أن تكون مساوية أو مختلفة عن (ن٢) وتعد هذه الفقرة جيدة وفق معامل التمييز ب (برينان) إذا كانت قيمته أكبر من أو تساوي

(٠.٢٠) (Lin, ٢٠٠٨:٣٤-٣٥) .

٦- طريقة معامل الارتباط الثنائي التسلسل الحقيقي – الأصلي)

بوينت بايسيريال) . Point Biserial Coefficient Correlation

يشير معامل الارتباط المستخرج بهذه الطريقة إلى القوة أو القدرة التمييزية للفقرة من خلال حساب العلاقة الارتباطية بين درجة كل فقرة والدرجة الكلية للاختبار إذ يفترض أن الذين يجيبون عن الفقرة أجابه صحيحة هم من فئة أو مجموعة الأفراد المتقنين بينما الذين يجيبون عن الفقرة أجابه خاطئة هم من فئة أو مجموعة الأفراد غير المتقنين (Browen and Hudson, ٢٠١١: ٣٩) ويهتم هذا النوع من معاملات الارتباط في دراسة العلاقة بين متغيرين أحدهما يقع في المستوى الفاصل أو النسبي والآخر يقع في المستوى الاسمي مثل متغير الجنس أو النوع أي أن يكون متغير ثنائي بشكل طبيعي (حقيقي) غير مصطنع مثل متغير الجنس (ذكور , إناث) أو طبيعية الإجابة (صح , خطأ) مثل العلاقة بين متغير الجنس ومتغير الذكاء أو الطول أو الوزن أو التحصيل ويحسب معامل الارتباط هذا من خلال المعادلة الآتية

$$\frac{\bar{x}_1 - \bar{x}_2}{sd} \sqrt{pq}$$

(النعيمي , ٢٠١٤ : ١٥١ – ١٥٢) .

٧- طريقة معامل استجابة الفقرة . Laten Trail Statistic Method

يعتمد معامل استجابة الفقرة على المبدأ النظري الذي ينص على ((أن الفرد المفحوص الذي يملك قدرة عالية لديه احتمالية أعلى للإجابة عن الفقرة أجابه صحيحة)) وطور العالم لندن (Vander Linden) انحدار هذا المعامل ليستعمل مع مفهوم درجة القطع (الدرجة الفاصلة) إذ أستبدل مفهوم الدرجة الفاصلة (درجة القطع) بما يقابلها من مستوى القدرة (القوة) (ق) وأن مفهوم التمييز في نظرية استجابة الفقرة يعبر عنه بمدى انحدار الفقرة وأن الفقرة تعد جيدة حينما يكون أقصى انحدارها مقابلاً للقدرة الفاصلة (القاطعي , ٢٠١٤ : ٥٣) .

ثبات الاختبارات مرجعية المحك :-

إن الاختبارات مرجعية المحك تستعمل عادة لأغراض تصنيف الطلبة إلى (متقنين وغير متقنين) وبيان درجة تمكنهم من المجال أو النطاق السلوكي المراد قياسه والمحدد مسبقاً لذا فإن ثبات هذا النوع من الاختبارات تطلق عليه تسمية

ثبات التصنيف لاختبارات الإتقان أو ثبات تقدير درجات الأفراد في نطاق سلوكي معين (Ojer, ٢٠١٠: ٨٢).

وتوصل العلماء في هذا المجال إلى عدد من الطرائق لحساب ثبات الاختبارات مرجعية المحك إذ يمكن تصنيفها إلى مجموعتين هما:
المجموعة الأولى : وهي تمثل ثبات تقديرات الأفراد المفحوصين في نطاق سلوكي معين ومن أبرز هذه الطرائق هي :

١- معامل لفنجستون (Living Ston) : أن هذه الطريقة تعتمد على أسس ومفاهيم النظرية التقليدية (الكلاسيكية) وترتكز على معرفة انحراف درجة الفرد عن متوسط العينة إلا أنها تستعمل الدرجة الفاصلة (درجة القطع) بدلاً من المتوسط الحسابي للعينة فبدلاً من إيجاد انحراف درجات الأفراد المفحوصين عن متوسط (معدل) العينة تحسب انحرافات درجات الأفراد المفحوصين عن الدرجة الفاصلة المحددة مسبقاً . (علام , ٢٠٠٥ : ٢٦٤) .

٢- معامل برينان (Brennan) أو معامل الاعتمادية : إن هذه الطريقة تعتمد على مفاهيم ومبادئ نظرية إمكانية التعميم لكرونباخ (Cronbach) إذ أوضح كرونباخ أن هناك نوعان من تباين الخطأ أحدهما يتعلق بتفسير البيانات المستمدة من الاختبارات مرجعية المحك وأن معامل التعميم يقصد به هنا هو النسبة بين تباين الدرجة الشاملة (الكلية) والتباين المتوقع الدرجة الملاحظة وأطلق عليه برينان تسمية معامل الاعتمادية (علام , ١٩٩٥ : ٢٦٦-٢٦٧) .

المجموعة الثانية : وهي تمثل ثبات تصنيف الأفراد المفحوصين إلى مجموعات بحسب درجة إتقانهم (تمكنهم) من نطاق سلوكي معين (التصنيف للإتقان) أن هذه المجموعة تهتم بحساب أتساق قرارات التصنيف إلى متقنين وغير متقنين استناداً إلى درجة فاصلة كما تهتم بتقدير أخطاء التصنيف الموجبة والسالبة ، إذ أن الخطأ الموجب يحدث إذا كان المستوى الحقيقي للفرد يقل عن الدرجة الفاصلة في حيث يكون أداءه على الاختبار مساوٍ للدرجة الفاصلة أو يزيد عنها وبذلك يصنف على أنه متقن أما الخطأ السالب فيحدث إذا كان المستوى الحقيقي للفرد يساوي أو يزيد عن الدرجة الفاصلة إلا أنه يحصل في الاختبار على درجة أقل من الدرجة الفاصلة وبذلك يصنف على أنه غير متقن ويجب أن تكون درجة التصنيف غير مشوبة بأي من نوعي الخطأ (الموجب والسالب) (علام , ١٩٩٥ : ٢٧٠) . وتصنف المعاملات التي تضمها هذه المجموعة إلى قسمين هما :

١- معاملات ثبات التصنيف للإتقان من تطبيقين : إذ يتطلب تطبيق الاختبار نفسه مرتين أو تطبيق صيغتين متوازيتين للاختبار نفسه ومن هذه المعادلات،

معامل اتفاق (أتساق) هامبلتون (Hambleton) ومعامل نوفيك (

Novick) ومعامل كارفر (Karvir) ومعامل كبا (Kappa)

ب-معامل ثبات التصنيف للإتقان (التمكن) من تطبيق واحد (مرة واحدة)
ويكفي لحسابه توافر بيانات من تطبيق واحد للاختبار مثل معامل اتفاق

سبكوفياك (Subkoviak)

وسيتم استعراض معامل الاتفاق لسبكوفياك (Subkoviak) بشكل مفصل
لأن الباحث أعتمد هذا المعامل في البحث الحالي ، لاستخراج معامل ثبات
الاختبار مرجعي المحك لمادة القياس والتقويم وكالاتي :

معامل الاتفاق لسبكوفياك Subkoviak Coefficient of Agreement

تعتمد هذه الطريقة على مدى أتساق (انسجام) تصنيف الأفراد المفحوصين
إلى مجموعتين أحدهما مجموعة متقنة والمجموعة الأخرى غير متقنة
باستعمال نتائج تطبيق الاختبار مرة واحدة فقط . وتهدف هذه الطريقة إلى
تقدير (حساب) مقدار احتمال تصنيف كل فرد تصنيفاً صحيحاً لـ (ف)
وجمع هذه المقادير الاحتمالية وإيجاد متوسطها ، وبذلك نحصل على معامل
اتفاق خاص بالمجموعة التي يطبق عليها الاختبار ويرمز له بالرمز (ل)
ويتم إيجاد معامل الاتفاق (سبكوفياك) وفق المعادلة الآتية :

$$ل س = م ج - ل س (ف)$$

إذ يرمز (س) إلى الدرجة الفاصلة في الاختبار وأن (ن) ترمز إلى العدد الكلي
لفقرات الاختبار وأن [ل س (ف)] ترمز إلى معامل الاتفاق الخاص بتصنيف بكل
فرد مفحوص (وهو مقدار احتمال التصنيف المتسق لهذا الفرد المفحوص حين
يطبق عليه اختبارين متوازيين) .
ويمكن كتابة هذه المعادلة بصيغة أخرى وكالاتي :

$$ل س (ف) = [ل (س ف \leq س)] + [١ - ل (س ف \leq س)]$$

وإذا كان عدد فقرات الاختبار قليلاً فأن القيمة التقديرية لدرجة الفرد في المجال أو
النطاق السلوكي باستعمال المقدار (ل ف) لا تكون دقيقة ولذلك يفضل في هذه
الحالة استعمال أسلوب الانحدار إذ أنه يعطي قيمة تقديرية أكثر دقة ويمكن الحصول
على هذه القيمة باستعمال صيغة المعادلة الآتية :

$$L = r \left(\frac{S_f}{n} \right) + (r-1) \left(\frac{S_s}{n} \right)$$

إذ أن الرمز (r) يشير إلى معامل ثبات الاختبار باستعمال طريقة كيودر - ريتشاردسون ٢١ (K.R. ٢١) وأن الرمز (S) يشير إلى متوسط نسبة عدد الإجابات الصحيحة للأفراد المفحوصين الذين طبق عليهم الاختبار . (Subkoviak , ٢٠٠٢ : ٢١- ٢٨) .

ثانياً : دراسات سابقة :-

- ١- دراسة الزهراني ٢٠١٠ الموسوعة (مقارنة طرق تقدير الثبات في القياس محكي المرجع) .
هدفت هذه الدراسة إلى إجراء مقارنة بين ثمان طرائق لتقدير الثبات في القياس محكي المرجع وهي معامل أتفاق (هاينا) ومعامل أتفاق (سبكوفياك) ومعامل أتفاق (كبا) (لهاينا) ومعامل أتفاق (كبا) (لسبكوفياك) ومعامل بيتا (B) ومعامل كيودر ريتشاردسون ٢٠ (K.R. ٢٠) ومعامل كيودر ريتشاردسون ٢١ (K.R. ٢١) ومعامل (لفنجستون) وطبقت هذه الدراسة على عينة عشوائية مكونة من (٦٥٩) تلميذاً وتلميذة في الصف السادس الابتدائي في محافظة جدة (في المملكة العربية السعودية) وأستعمل في هذه الدراسة اختبار محكي المرجع لقياس الكفايات الرياضية لمادة الرياضيات للصف السادس الابتدائي مكون من (٦٠) فقرة موضوعية وتوصلت هذه الدراسة إلى عدة نتائج منها : أ- أن أفضل قيمة للدرجة الفاصلة والتي قد تسهم في تقليل تأثير عوامل الصدفة وأخطائها في قرارات التصنيف تقع في منتصف توزيع الدرجة الكلية . ب- أن ليس هناك اختلاف أو فروق عملية بين الطرائق الثمانية التي استعملت في هذه الدراسة في معامل الثبات من حيث اتجاهات التغيير وشدته في قيم معاملات الثبات في ظل تغير قيم التباين وعدد أفراد العينة وعدد الفقرات والدرجة الفاصلة (الزهراني ، ٢٠١٠ : ٣٧) .
- ٢- دراسة عبدالله ٢٠١١ الموسومة (مقارنة ثلاث طرائق لحساب اختبار مرجعي المحك لمادة الجبر للصف الأول المتوسط) .
هدفت هذه الدراسة إلى مقارنة ثلاث طرائق لحساب ثبات اختبار مرجعي المحك لمادة الجبر للصف الأول المتوسط وهذه الطرائق هي لفنجستون وهاينا

وسبكوفياك واستعملت أطوال مختلفة من الاختبار وعينات مختلفة ودرجات فاصلة مختلفة وتكون هذا الاختبار من (١٩٠) فقرة وطبق على عينة عشوائية مكونة من (٨١٨) طالباً وطالبة ومن أبرز نتائج هذه الدراسة هي أنه يمكن عد الدرجتين الفاصلتين (٧٠% و٨٠%) درجتين جيدتين ومقبولتين للتصنيف من أجل تحديد مستوى التمكن (الإتقان) إذ أن الاختبارات ذات الأطوال المختلفة وذات الأحجام المختلفة من العينات تعطي قيم معاملات ثبات مرتفعة عند هذه الدرجات الفاصلة كما أنها تدل على تحقيق نسبة كبيرة من الأهداف المرغوب قياسها وأن طريقة (هاينا) في حساب معاملات الثبات تعطي نتائج دقيقة وغير متحيزة أو أقل تميزاً من طريقة (لفنجستون سبكوفياك) كما أن طريقة (لفنجستون) أظهرت دلالات كثيرة لفاعليتها في معامل الثبات وفي مواقع متعددة ولأحجام مختلفة ولأطوال مختلفة للاختبارات ولكنها تتأثر بالدرجات الفاصلة القريبة من المتوسط وعادة ما تكون من (٤٠%) إلى (٦٠%) أما عندما يؤخذ حجم العينة بنظر الاعتبار فيمكن استعمال معامل سبكوفياك إذ تكون العينة كبيرة وطول الاختبار يتراوح بين (٣٢) فقرة و(٩٦) فقرة واستعمال معامل لفنجستون حينما يكون حجم العينة صغيراً (وتتراوح ما بين (٣٠-٤٠) فرداً لطول الاختبار ذاته أما حينما تكون العينة متوسطة الحجم ويتراوح عدد أفرادها ما بين (٥٠-٦٠) فرداً فيمكن استعمال معامل (هاينا) لطول الاختبار ذاته. (عبدالله ، ٢٠١١ : ١٩).

٣-دراسة شانون Shannon ٢٠١٢ الموسومة ((العلاقة الارتباطية بين طريقة استجابة الفقرة وأربع طرائق لتمييز فقرات اختبار محكي المرجع)) هدفت هذه الدراسة إلى معرفة العلاقة الارتباطية بين طريقة استجابة الفقرة وأربع طرائق تمييز الفقرات وهي معامل التمييز (ب) (برينان) ، ومعامل فاي ، فاي العظمى ، ومعامل التوافق المرجعي وطبقت على عينة مكونة من (٢٢٢) طالباً وطالبة وأستعمل اختبار محكي المرجع لمادة علم النفس التربوي مكوناً من (٨٠) فقرة موضوعية متعددة الخيارات وكان من أبرز نتائجها أن هناك ارتباطاً عالياً وكبيراً بين معامل فاي واستجابة الفقرة وكان متوسط القيم هو (٠.٩٦) إذ تم حساب الارتباط بين استجابة الفقرة وكل من الطرائق الأربعة عند درجات فاصلة مختلفة وهي (٧٠% ، ٧٥% ، ٨٠%) ووجد أن أعلى قيمة كانت بين استجابة الفقرة وبين معامل فاي ثم معامل (ب) ومعامل التوافق ومعامل فاي / فاي العظمى إذ بلغت قيمة متوسط معاملات الارتباط بين استجابة الفقرة ومعامل فاي تساوي (٠.٩٦) وبين استجابة الفقرة ومعامل التوافق المرجعي تساوي (٠.٧٨) وبين استجابة الفقرة ومعامل (ب،برينان) يساوي (٠.٨٦) وبين استجابة الفقرة ومعامل فاي(٠.٥٢) وبين استجابة الفقرة وفاي العظمى تساوي (٠.٥٣). (Shannon ، ٢٠١٢ : ٢٠).

- موازنة الدراسات السابقة مع الدراسة الحالية :

تباينت الدراسات السابقة فيما بينها من حيث مشكلة البحث وأهدافه وحدوده ومجتمع البحث وعينته والوسائل الإحصائية المستعملة في تحليل البيانات وعرض النتائج ومناقشتها إذ تراوح أكبر حجم للعينة في دراسة (عبدالله ، ٢٠١١) إذ كان (٨١٨) فرداً وأقل حجم للعينة في دراسة (شانون ٢٠١٢) إذ كان (٢٢٢) فرداً أما حجم عينة البحث الحالي فكانت (٨٠) طالباً وطالبة من الصف الرابع في قسم العلوم التربوية والنفسية / كلية التربية / جامعة واسط وتباينت هذه الدراسات في أدواتها إذ استعملت دراسة (الزهراني ٢٠١٠) اختبار محكي المرجع لمادة الرياضيات للصف السادس الابتدائي مكوناً من (٦٠) فقرة موضوعية ذات الاختيار من متعدد واستعملت دراسة (عبدالله ٢٠١١) اختباراً محكي المرجع لمادة الجبر للصف الأول المتوسط مكوناً من (١٩٠) فقرة موضوعية ذات الاختيار من متعدد في حين استعملت دراسة (شانون ٢٠١٢) اختباراً محكي المرجع لمادة علم النفس التربوي مكوناً من (٨٠) فقرة موضوعية ذات الاختيار من متعدد أما البحث الحالي فأستعمل اختبار محكي المرجع لمادة القياس والتقويم للصف الرابع في قسم العلوم التربوية والنفسية / كلية التربية / جامعة واسط والمكون من (١٤٤) فقرة موضوعية ذات الاختيار من متعدد إذ وضعت أمام كل فقرة أربعة بدلائل للإجابة .

-جوانب الاستفادة من الدراسات السابقة : تتمثل جوانب الاستفادة من الدراسات السابقة في الآتي :

- ١- تحديد مشكلة البحث الحالي وتحديد مجتمع البحث وعينته وأدواته
- ٢- مقارنة نتائج الدراسات السابقة بنتائج البحث الحالي .

الفصل الثالث

منهج البحث وإجراءاته

منهج البحث : أستعمل الباحث المنهج الوصفي المقارن في بحثه الحالي لأنه يسعى إلى قياس أثر استعمال ثلاثة طرائق مختلفة لتمييز فقرات اختبار مرجعي المحك لمادة القياس والتقويم في معامل ثباته وأستعمل منهج الدراسات المقارنة لأنه ركز على ثلاثة طرائق مختلفة لتمييز فقراته ووصف أثرها في معامل ثباته إذ إن هذا المنهج يقوم على وصف العلاقات والمؤثرات التي توجد بين الظواهر وتفسيرها وتحليلها (٤٣ : ١٩٨٣، Nitko) .

مجتمع البحث : تكون مجتمع البحث من (١٢٠) طالباً وطالبة بواقع (٤٨) طالباً و(٧٢) طالبة في الصف الرابع / قسم العلوم التربوية والنفسية / كلية التربية / الدراسة الصباحية ، وتتراوح أعمارهم بين (٢٢-٢٤) سنة وينحدرون من بيئات ثقافية واقتصادية واجتماعية متقاربة إلى حد كبير أو تشير نتائج دراسة (الوائلي ، ٢٠١٢) إلى شيوع صفة التقارب الاجتماعي والثقافي والاقتصادي بين أغلب أبناء أفضية ونواحي ومدن محافظة واسط (الوائلي ، ٢٠١٢ : ٧) والجدول (١) يوضح ذلك .

جدول (١)

يوضح عدد طلبة الصف الرابع في قسم العلوم التربوية والنفسية في كلية التربية /

جامعة واسط بحسب متغير الجنس (ذكور ، إناث)

النسبة المئوية	المجموع الكلي	عدد الطالبات / الإناث	عدد الطلاب / الذكور	الجنس الصف
٤٠%	٤٨	/	٤٨	الرابع في قسم العلوم التربوية والنفسية
٦٠%	٧٢	٧٢	/	
١٠٠%	١٢٠	٧٢	٤٨	المجموع الكلي

عينة البحث : اختيرت عينة البحث البالغ عددها (٨٠) طالباً وطالبة بواقع (٣٢) طالباً و(٤٨) طالبة بصورة عشوائية من مجتمع البحث البالغ عدده (١٢٠) طالباً وطالبة وبذلك بلغت نسبة عينة البحث إلى مجتمع البحث (٦٦.٦٦%) وتتراوح أعمارهم بين (٢٢-٢٤) سنة وينحدرون من بيئات ثقافية واجتماعية واقتصادية متقاربة إلى حد كبير والجدول (٢) يوضح ذلك .

جدول (٢)

يوضح عدد أفراد عينة البحث موزعة بحسب متغير الجنس

(ذكور ، إناث)

النسبة المئوية	المجموع الكلي	عدد الطالبات / الإناث	عدد الطلاب / الذكور	الجنس الصف
٤٠%	٣٢	/	٣٢	الصف الرابع في
٦٠%	٤٨	٤٨	/	قسم العلوم التربوية والنفسية
١٠٠%	٨٠	٤٨	٣٢	المجموع الكلي

أداة البحث : استعمل الباحث اختبار مرجعي المحك لمادة القياس والتقييم للصف الرابع / قسم العلوم التربوية والنفسية والذي قام بأعداده الباحث الدكتور عبود جواد في عام ٢٠١٢/٢٠١٣ وفق ثلاث مراحل هي :
أولاً: مرحلة التحليل : وتتكون من الخطوات الآتية :

- ١- تحديد المحتوى (المضمون) : ينقسم محتوى مادة القياس والتقييم إلى قسمين أساسيين هما القسم (الجزء) النظري والقسم (الجزء) العملي التطبيقي .
- ٢- تحديد المهارات أو الكفايات الأساسية : صنفت هذه المهارات (الكفايات) إلى قسمين هي مهارات أو كفايات نظرية ومهارات (كفايات) عملية - تطبيقية .

٣- صياغة الأهداف : تم صياغة (٢٤) هدفاً رئيساً بواقع (١٢) هدفاً للمهارات النظرية و(١٢) هدفاً للمهارات العملية-التطبيقية لكونها متساوية في الأهمية النسبية التي حددها الخبراء الذين استعان بهم الباحث .

٤- صياغة الأهداف السلوكية : تم صياغة (١٤٤) هدفاً سلوكياً موزعة على (٢٤) هدفاً رئيساً بواقع (٧٢) هدفاً سلوكياً للمهارات النظرية بنوعها الوصفية بواقع (٣٦) فقرة والمنطقية بواقع (٣٦) فقرة و(٧٢) هدفاً سلوكياً للمهارات العملية التطبيقية وتم عرضها على (١٦) خبيراً وأعتمد الباحث نسبة موافقة (١٢) خبيراً فأكثر معياراً لصلاحيته هذه الأهداف السلوكية وهم الذين يشكلون نسبة (٧٥%) من مجموع الخبراء لأن الفرق بين هذا العدد من الموافقين وبين عدد غير الموافقين وهم (٤) خبراء بدلالة إحصائية عند مستوى (٠.٠٥) وباستعمال اختبار مربع كاي ، إذ تكون قيمة مربع كاي الجدولية (النظرية) التي تساوي (٣.٨٤١) بدرجة حرية (١)

وفي ضوء ملاحظاتهم تم تعديل صياغة بعض الأهداف السلوكية ولم يتم حذف أي هدف سلوكي ما يدل على أنها كانت دقيقة وواضحة وسهلة التطبيق وشاملة بجميع مفردات هذه المادة من وجهة نظر هؤلاء الخبراء .

ثانياً : مرحلة البناء : أعتد الباحث طريقة بوفام (Popham) في بناء الاختبار وفق الخطوات الآتية :

١- أعداد مواصفات الأسئلة : قام الباحث بتصميم يتكون من مواصفات كل هدف من الأهداف الرئيسية البالغ عددها (٢٤) هدفاً رئيساً ومستويات الجانب أو الهدف المعرفي وفق تصنيف بلوم الذي يضم ست مستويات في هذا الجانب وهي (المعرفة ، الفهم ، التطبيق ، التحليل ، التركيب ، التقويم) ومواصفات السؤال المكونة من (المثير والاستجابة) (الفقرة وبدائل الإجابة عنها) ومثال إجرائي (عملي) وبذلك حسب الوزن النسبي للأهداف الرئيسية والفرعية في المحتوى .

٢- صياغة فقرات الاختبار : تم صياغة (١٤٤) فقرة موضوعية من اختيار من متعدد ووضع أمام كل فقرة أربعة بدائل للإجابة هي (أ،ب،ج،د) لقياس (٢٤) هدفاً رئيساً والذي تتمثل في (١٤٤) هدفاً سلوكياً وبواقع (٧٢) فقرة لقياس المهارات النظرية (الوصفية والمنطقية) بواقع (٣٦) فقرة لكل منها و(٧٢) فقرة لقياس المهارات العملية (التطبيقية) وتعطى درجة واحدة للإجابة الصحيحة ودرجة (صفر) للإجابة الخاطئة وبحسب رأي الخبراء إذ يرون أن الأهمية النسبية (الوزن النسبي) للمهارات النظرية والعملية تكون متساوية في هذه الأهمية لذا كان عدد الفقرات المخصصة لهاتين المهارتين (النظرية والتطبيقية) كان متساوياً.

٣- كتابة تعليمات الاختبار :- بعد أن تمت صياغة فقرات الاختبار قام الباحث بوضع تعليمات الاختبار وهي على نوعين :

أ- التعليمات الخاصة بشروط استعمال الاختبار وإجراءات تطبيقه وآلية تصحيحه والظروف الملائمة للاستعمال الأفضل والأمثل له.

٤- وأيضاً قام الباحث بالتحقق من صدق المحتوى عن طريق معرفة مدى ارتباط الفقرة بالهدف الذي يقابلها إذ تم أعداد أستبانة خاصة لهذا الغرض عرضت على (١٦) خبيراً الذين أكدوا على جودة وصحة وملائمة جميع فقرات هذا الاختبار للمحتوى (المضمون) المراد قياسه في البحث الحالي . كما تم حساب معامل الاتفاق لفقرات الاختبار فبلغ (٧٥%) وهذا يدل على أن جميع فقرات الاختبار كانت صادقة إلى حد كبير في تمثيلها للنطاق أو المجال السلوكي المراد قياسه في البحث الحالي .

٥- ثبات الاختيار : تم حساب ثبات الاختبار بصيغته الأولية بمعادلة كيوذرر يتشاردسون (٢٠) إذ طبق الاختبار على عينة عشوائية مكونة من (٦٠) طالباً وطالبة بواقع (٢٤) طالباً و(٣٦) طالبة من طلبة الصف الرابع في قسم

العلوم التربوية والنفسية / كلية التربية إذ بلغت قيمة معامل ثبات الاختبار ككل تساوي (٠.٨٢١) في حين بلغت قيمة معامل ثبات الاختبار الفرعي الأول (المهارات النظرية – الوصفية) تساوي (٠.٨٢٤) بينما بلغت قيمة معامل الثبات الاختبار الفرعي الثاني (المهارات النظرية المنطقية) تساوي (٠.٨٢١) أما قيمة معامل ثبات الاختبار الفرعي الثالث (المهارات العملية – التطبيقية) تساوي (٠.٨٣٠).

٦. تحديد الدرجة الفاصلة : أعتمد الباحث طريقة أنجوف (Angoff) في تحديد الدرجة الفاصلة لكل اختبار فرعي وكانت على النحو الآتي (٧٥%) للمهارات النظرية الوصفية التي يقيسها الاختبار الفرعي الأول و(٧٥%) للمهارات المنطقية والتي يقيسها (الاختبار الفرعي الثاني) و(٧٥%) للمهارات العملية – التطبيقية التي يقيسها (الاختبار الفرعي الثالث) وتهدف هذه الطريقة الى التمييز بين الأفراد المتقنين وغير المتقنين نظراً لكفاءتها وسهولة فهمها ويسر تطبيقها (علام ، ١٩٩٥ ، ١١٨) وحددت هذه الدرجات الفاصلة بالتعاون مع (١٦) خبيراً في مادة القياس والتقويم من ذوي الخبرة الاختصاص في هذه المادة الدراسية . إذ قام الباحث بنفسه بتوزيع الاستمارات عليهم إذ تضمنت هذه الاستمارة جميع الفقرات وطريقة الإجابة عنها وكيفية تعبئتها وملئها ومثالاً توضيحياً عن ذلك الأمر . وبعد انتهاء الخبراء من الإجابة عن فقرات هذه الاستمارة تم جمعها وتفرغ بياناتها إذ أستخرج متوسط النسب لكل خبير ومن ثم جمعت المتوسطات التي بلغت (١٢٠٠) وقسمت على عدد الخبراء المشاركين في تعبئة هذه الأسئلة والبالغ عددهم (١٦) خبيراً ($1200 \div 16 = 75\%$) وهذا يعني أن الدرجة الفاصلة هي (٧٥%) من فقرات الاختبار ككل قد أعتمدها الباحث في هذه الدراسة وبما أن الاختبار مكون من (١٤٤) فقرة وأن الدرجة الفاصلة هي (٧٥%) فان الفرد المفحوص الذي يحصل على درجة (١٠٨) فما فوق يعد فرداً متقناً لهذه المادة الدراسية وهذه القيمة (الدرجة الفاصلة) ناتجة من ضرب عدد فقرات الاختبار (X) ضرب قيمة الدرجة الفاصلة وهي هنا تساوي (٠.٧٥) أي (٧٥%) . والفرد المفحوص الذي يحصل على أقل من (١٠٨) يعد فرداً غير متقن لهذه المادة الدراسية والجدول (٣) يوضح ذلك .

جدول (٣)

يوضح درجة الإتقان في ضوء الدرجة الفاصلة للاختبار

ت	أسم الاختيار	محتوى الاختيار	عدد الفقرات	قيمة الدرجة الفاصلة	درجة الإتقان في ضوء الدرجة الفاصلة
١	الفرعي الأول	المهارات النظرية الوصفية	٣٦	%٧٥	٢٧
٢	الفرعي الثاني	المهارات النظرية المنطقية	٣٦	%٧٥	٢٧
٣	الفرعي الثالث	المهارات العملية- التطبيقية	٧٢	%٧٥	٥٤
فقرات الاختيار ككل			١٤٤	%٧٥	١٠٨

ثالثاً: مرحلة التجريب :

١- التطبيق على عينة استطلاعية : قام الباحث بتطبيق الاختبار على عينة عشوائية مكونة من (٢٦) طالباً وطالبة بواقع (٨) طلاب (ذكور) و(١٨) طالبة (أنثى) لغرض التأكد من مدى وضوح صياغة الفقرات وحساب الزمن المناسب لتطبيق الاختبار بصيغته النهائية إذا ظهرت النتائج أن جميع الفقرات كانت دقيقة ومفهومة وواضحة الصياغة وان متوسط الزمن المناسب للتطبيق هو (٦٠) دقيقة إذ سبق وأن حصل الباحث على موافقة رئاسة قسم العلوم التربوية والنفسية على منحه أكثر من ساعة لأجراء بحثه على هذه العينة وللتحقق من جودة الفقرات تم استعمال ثلاثة طرائق مختلفة لإيجاد معامل التمييز وهي معامل بيرينان (ب) ومعامل فاي ومعامل التوافق المرجعي كما تمت دراسة فعالية المشتتات (فاعلية البدائل الخاطئة) لفقرات الاختبار الموضوعي ذات الأربعة بدائل للإجابة ونتج عن ذلك تعديل بعضها واستبدال بعض المشتتات غير الجيدة .

وسبق للباحث أن أستخرج صدق المحتوى إذ بلغت قيمة معامل الاتفاق (%٧٥) كما تم حساب ثبات الاختبار ككل بطريقة سبكوفياك وكان معامل

الثبات يساوي (٠.٨٢٩٤) كما تم حساب الدرجة الفاصلة وهي (٧٥%) أي (٠.٧٥) ومن أبرز إجراءات التطبيق النهائي هي ما يأتي :

١- قيام الباحث بنفسه بتطبيق الاختبار على عينة البحث التي تم جمعها في قاعة دراسية واحدة بعد التأكد من أن جميع أفراد عينة البحث قد أخذوا أو درسوا جميع مفردات هذه المادة وتم التحقق من هذه المسألة من خلال الاستفسار من الطلبة والتدريسيين بهذا الخصوص .

٢- تم توزيع كراسة الاختبار وتعليمات الإجابة عنها على عينة البحث وتم توضيح كافة الأمور بخصوص الفقرات والإجابة عنها .

٣- حرص الباحث على تهيئة الأجواء المناسبة للطلبة قبل البدء للإجابة وبذلك أصبح الاختبار جاهزاً بصيغته النهائية إذ تكون من ثلاثة اختبارات فرعية تضم (١٤٤) فقرة موضوعية ووضع أمام كل فقرة أربعة بدائل للإجابة (اختبار الاختيار من متعدد) لقياس نوعين من المهارات (الكفايات) هي المهارات النظرية الوصفية والمنطقية والمهارات العملية – التطبيقية لمادة القياس والتقويم للصف الرابع في قسم العلوم التربوية والنفسية كلية التربية وتعطى درجة واحدة للإجابة الصحيحة ودرجة (صفر) للإجابة الخاطئة إذ طبق الباحث هذا الاختبار بصيغته النهائية لمعرفة أثر استعمال ثلاثة طرائق مختلفة لتمييز فقرات هذا الاختبار على معامل ثباته .

ب- التطبيق على العينة الأساسية (التطبيق النهائي) . قام الباحث بتطبيق الاختبار على عينة البحث المكونة من (٨٠) طالباً وطالبة بواقع (٣٢) طالباً و(٤٨) طالبة من طلبة الصف الرابع في قسم العلوم التربوية والنفسية لغرض تحقيق هدف البحث الحالي والتحقق من صحة فرضيته .الوسائل الإحصائية : أستعمل الباحث الحقيبة الإحصائية للعلوم الاجتماعية (Spss) في تحليل بيانات ونتائج البحث الحالي إحصائياً ومن أبرز المعادلات المستعملة في هذا البحث ما يأتي :

١- معادلة كروسكال – واليز Kruskal – Wallis للكشف عن الفروق في معاملات الثبات.

٢- معامل سبكوفياك (Pc) Subkoviak Coefficient of Agreement

لحساب قوة معامل ثبات الاختبار مرجعي المحك لمادة القياس والتقويم
معامل التمييز - برينان (ب) The Brennan (B) Index
Approach لتمييز الفقرات .

٣ - معامل فاي Coefficient Phi لتمييز الفقرات .

٤ - معامل التوافق المرجعي Agreement Statistic لتمييز الفقرات .

الفصل الرابع

عرض النتائج ومناقشتها

لغرض تحقيق هدف البحث لحالي التي يص على ما يأتي ((التعرف على أثر استعمال
ثلاثة طرئق مختلفة لتمييز فقرات اختبار مرجعي المحك لمادة القياس والتقويم في
معامل ثباته))

قام البحث باستعمال معادلة سبكوفياك (Pc) Subkoviak Coefficient of
Agreement

لحساب ثبات الاختبار مرجعي المحك لمادة القياس والتقويم (أداة البحث) بعد تطبيق
البحث هذا الاختبار على عينة البحث لمعرفة أثر استعمال ثلاثة طرئق مختلفة لتمييز
فقراته في معامل ثباته والجدول (٥) يوضح ذلك .

جدول (٥)

يوضح نتائج معادلة سبكوفياك لحساب معاملات ثبات الاختبار مرجعي المحك
(أداة البحث) .

الاختبار ككل	معامل التوافق المرجعي A.S	معامل فاي Phi	معامل برينان Br	نوع معامل التمييز أسم الاختبار
٠.٨٢٨٦	٠.٨٣٢	٠.٨٢٩	٠.٨٢٥	الاختبار الفرعي الأول / المهارات الوصفية
٠.٨٢٩	٠.٨٣٤	٠.٨٢٤	٠.٨٢٩	الاختبار الفرعي الثاني / المهارات المنطقية
٠.٨٣٠٦	٠.٨٣٣	٠.٨٢٥	٠.٨٣٤	الاختبار الفرعي /المهارات العملية
٠.٨٢٩٤	٠.٨٣٣٠	٠.٨٢٦٠	٠.٨٢٩٣	الاختبار ككل

وباستعمال معادلة كروسكال- واليز التي تعتمد على عدد القيم ومتوسط الرتب ودرجة الحرية وقيمة كاي تربيع (ك^٢) للكشف عن دلالة الفروق بين طرائق التمييز الثلاثة التي أتبعها الباحث في هذه الدراسة ومن خلال ملاحظة الجدول (٦)

جدول (٦)

نتائج استعمال معادلة كروسكال - واليز في الكشف عن الفروق بين قيم معاملات الثبات لاختبار مرجعي المحك لمادة القياس والتقويم تبعاً لاختلاف طرائق تمييز فقراته .

ت	طريقة التمييز	(عدد القيم الممثلة) (عدد الاختبارات الفرعية الثلاثة والاختبار ككل)	متوسط الرتب	درجة الحرية	قيمة مربع كاي		مستوى الدلالة (٠.٠٥)
					المحسوبة	الجدولية	
١	معامل بيرينان	٤	١٣,٥٠٠	٢	٠,٩٥٤	٥,٩٩	غير دالة لا توجد فروق
٢	معامل فاي	٤	١٣,٥٠٠				
٣	معامل التوافق المرجعي	٤	٢٧,٠٠				

يتضح أن قيمة كاي المحسوبة تساوي (٠.٩٥٤) وهي أقل من قيمة مربع كاي الجدولية (النظرية) التي تساوي (٥.٩٩) وبدرجة حرية (٢) وبمستوى دلالة إحصائية (٠.٠٥) وبهذا تقبل الفرضية الصفرية التي تنص على ما يأتي ((لا توجد فروق دالة إحصائية في قيم معامل ثبات اختبار مرجعي المحك لمادة القياس والتقويم نتيجة لاستعمال ثلاثة طرائق مختلفة لتمييز فقراته)) وترفض الفرضية البديلة التي تنص على ما يأتي ((توجد فروق دالة إحصائية في قيم معامل ثبات اختبار مرجعي المحك لمادة القياس والتقويم نتيجة لاستعمال ثلاثة طرائق مختلفة لتمييز فقراته)) . ويتضح من ملاحظة هذين الجدولين إن قيم معاملات ثبات هذه الاختبارات الفرعية الثلاثة التي تضمنها الاختبار مرجعي المحك لمادة القياس والتقويم (أداة البحث) لا تشير إلى وجود فروق كبيرة في قيم معامل ثبات هذا الاختبار تبعاً لاختلاف طرائق تمييز فقراته ويلاحظ أن هذه النتيجة تنحصر في الطرائق المستعملة في البحث الحالي وهي معامل تمييز بيرينان (B) ومعامل فاي (Phi) ومعامل التوافق المرجعي (A.S) ويرى الباحث أن هذه النتيجة طبيعية ومنطقية ومقبولة إذ تشير نتائج دراسة (رويد)

(Roid) أن هناك اتفاق ملحوظ بين هذه الطرائق الثلاثة بصفة خاصة والطرائق الأخرى بصفة عامة في اختيار الفقرات وبالتالي ينتج عنه عدم اختلاف في قيم معامل ثبات الاختبار الناتج من استعمال كل طريقة في تمييز الفقرات (١٥٢-٢٠١٢: Roid). أن هذه النتيجة تتفق مع نتائج دراسة كل من الزهراني ، (٢٠١٠) وتختلف مع دراسة عبدالله (٢٠١١) ودراسة شانون (٢٠١٥) Shannon).

الاستنتاجات : في ضوء نتائج البحث الحالي توصل الباحث إلى الاستنتاج الآتي:
أن قيمة معامل ثبات اختبار مرجعي المحك لمادة القياس والتقويم لا تتأثر باستعمال ثلاثة طرائق مختلفة لتمييز فقراته .

التوصيات :

- ١- يوصي الباحث بضرورة استعمال هذه الطرائق في تمييز الاختبارات مرجعية المحك في مواد دراسية أخرى عندما تتوفر المعرفة الكافية لدى الباحثين باستعمال هذه الطرائق التي قد تعطي نتائج متشابهة ومقاربة إلى حد كبير في معامل ثبات الاختبارات مرجعية المحك .
- ٢- ضرورة توعية العاملين في ميدان التربية والتعليم بأهمية الاختبارات مرجعية المحك وأهمية تمييز فقراتها .

المقترحات :

استكمالاً لنتائج البحث الحالي وتطويراً له يقترح الباحث إجراء الدراسات اللاحقة الآتية :

- ١- إجراء دراسة معرفة أثر اختلاف طرائق تمييز فقرات الاختبار مرجعي المحك لمواد دراسية متنوعة أخرى على معامل ثباته وأنتقاء فقراته مع الأخذ بنظر الاعتبار تأثير الدرجة الفاصلة إضافة إلى المتغيرات الأخرى كحجم العينة وطول الاختبار .
- ٢- إجراء دراسة لمقارنة طرائق تمييز الفقرات المشتقة من نظرية استجابة الفقرة بالطرائق الأخرى لتمييز فقرات الاختبارات مرجعية المحك في أنتقاء الفقرات ومعامل الثبات .
- ٣- إجراء دراسة مقارنة الطرائق الثلاثة للتمييز (برينان وفاي والتوافق المرجعي) بالطرائق التقليدية – الاعتيادية كسبيرمان براون و بيرسون والاختبار

التائي (T. test) وبقية معاملات الارتباط للتمييز ومقارنتها بالطرائق المشتقة من نظرية استجابة الفقرة في أنتقاء الفقرات ومعامل الثبات.

المصادر

أ- العربية

- ١- إبراهيم ، محمود سالم (٢٠١٠) ، دراسة سيكومترية مقارنة لطرق حساب معامل ثبات الاختبارات مرجعية المحك ، دار الفكر العربي ، القاهرة ، مصر .
- ٢- الأمام ، مصطفى محمود وآخرون (٢٠٠٥) التقويم والقياس ، دار الحكمة للطباعة والنشر ، بغداد ، العراق .
- ٣- الدوسري ، إبراهيم مبارك (٢٠١٠) أطار مرجعي في التقويم التربوي بدول مجلس التعاون الخليجي ، مكتب التربية العربي لدول الخليج العربي .
- ٤- الزهراني ، صالح عبد العزيز (٢٠١٠) مقارنة طرق تقدير الثبات في القياس محكي المرجع ، رسالة ماجستير غير منشورة ، كلية التربية ، جامعة أم القرى ، مكة المكرمة ، المملكة العربية السعودية.

- ٥- الطرييري ، عبد الرحمن (٢٠٠٧) القياس النفسي والتربوي ، نظرياته ، أسسه تطبيقاته ، مكتبة الرشد ، الرياض ، المملكة العربية السعودية .
- ٦- عبدالله خالد بن إبراهيم (٢٠١١) مقارنة ثلاث طرائق لحساب اختيار مرجعي المحك لمادة الجبر للصف الأول المتوسط ، الرياض ، السعودية .
- ٧- عبد المجيد ، نبيل عبد الغفور وساجدة جبار لفته (٢٠١٣) القياس والتقويم ، ط١ دار الكتب والوثائق ، بغداد ، العراق .
- ٨- عبد الرحمن ، سعد (١٩٩٨) القياس النفسي ، النظرية والتطبيق ، ط٣ دار الفكر العربي ، القاهرة ، مصر .
- ٩- العزاوي ، رحيم يونس كرو (٢٠٠٨) القياس والتقويم في العملية التدريسية ، ط١ ، دار دجلة للطباعة والنشر ، الأردن .
- ١٠- علام ، صلاح الدين محمود (١٩٩٥) الاختبارات التشخيصية مرجعية ا لمحك في المجالات التربوية والنفسية والتدريسية ، ط١ دار الفكر العربي ، القاهرة ، مصر .
- ١١- التربوي والنفسي أساسياته المعاصرة ، دار الفكر العربي ، القاهرة ، مصر . (٢٠٠٠) القياس والتقويم وتطبيقاته وتوجهاته
- ١٢- الحددين في تقدير درجة القطع للاختبار محكي المرجع (دراسة أحصائية) المجلة العربية للعلوم الانسانية ، جامعة الكويت ، دولة الكويت . (٢٠٠٥) أستخدام النموذج ذي
- ١٣- التربوية والنفسية ، ط١ ، دار الفكر العربي ، القاهرة ، مصر . (٢٠٠٦) الاختبارات والمقاييس
- ١٤- التربوي والنفسي ، ط٥ ، دار الفكر العربي ، القاهرة ، مصر . (٢٠١١) القياس والتقويم أساسياته وتطبيقاته وتوجهاته المعاصرة ،
- ١٥- علي ، صلاح شريف (٢٠٠١) ، تصميم وتقييم فاعلية برنامج تعليمي لكفايات بناء الاختبارات التحصيلية مرجعية المحك لدى معلمي العلوم لمرحلة التعليم الأساسي ، رسالة دكتوراه غير منشورة ، جامعة الأزهر ، كلية التربية ، جمهورية مصر العربية .
- ١٦- عمر ، محمود أحمد واخرون (٢٠١٠) ، القياس النفسي والتربوي ، ط١ ، دار المسيرة للنشر والتوزيع والطباعة ، الأردن .

١٧- الفاطمي ، عبدالله بن علي (١٩٩٣) دراسة مقارنة لبعض طرق تحليل بنود الاختبارات مرجعية المحك وفعاليتها في الاختبار ، دراسات تربوية ، المجلد الثامن ، الجزء (٥٠).

١٨- النعيمي ، مهند محمد عبد الستار (٢٠١٤) ، القياس النفسي في التربية وعلم النفس ، دار الكتب والوثائق ، بغداد ، العراق .

١٩- الوائلي ، حسين علوان (٢٠١٢) المستوى الاجتماعي والثقافي والاقتصادي في محافظتي واسط وميسان ، دراسة مقارنة ، جامعة البصرة .
ب- الأجنبية

٢٠- Berk .R.A (٢٠١٠) . Item Analysis in R.A. Criterion Referenced Measurement

The State of the Art Gohns Hopkins University Press .

٢١- BradField. (٢٠٠٨) Educational Testing .N.Y. U.S.A .

٢٢- Brown. J.D (٢٠٠١) Criterion .Referenced Testing in Two Academic Reading Courses Cambrig Uni Press .

٢٣- Foraan . (٢٠١٢) The Guide To Taking Test .N.J. Prentice. Hall .

٢٤- Gregory .F (٢٠١٢) Classroom Testing .N.Y.Johnwiley.

٢٥- Gronlaud.N.C (٢٠١٢) .Measurement and Evaluation in Teaching .N.Y. Macmillan Publishing CO.ink .

٢٦- Harris .D.J. (٢٠٠٩) .Item Analysis Ashort .Cut Statistic For Mastery Test .

University of Wisconsin – Madison .

٢٧- ----- . (٢٠١٢) . Item Selection For Mastery Tests . Acomparision of Three Procedures . University of Wisconsin Madison .

٢٨- Lin. Hui – Fen .(٢٠٠٨) Acomparision of Three Item Selection Method in Criterion Referenced Tests . Unversity of North Texas . U.S.A .

٢٩- Madaus .G (٢٠١٠) .Educational Evaluation . Kluwer Academic Publishers .U.S.A .

- ٣٠- Nitko .A (١٩٨٣) Educational Tests and Measurement . Anintroduction .N.Y. Harcourt Bracejovanvick . inc
- ٣١- Ojer. L . (٢٠٠٢) Acomparison of Select Indicas That Estimate Reliability of Criterion Referenced Test . University of Kansas .U.S.A .
- ٣٢- Popham . W.J (٢٠٠٥) Educational Evaluation Englwood.CliffsN.J. Prentice .Hall .U.S.A .
- ٣٣- ----- (٢٠١٠) . Measurement and Evaluation Holt Rinehart and Winston .U.S.A .
- ٣٤- Roid . G (٢٠١٢) Acomparison of two Item Selection Procedures for Building criterion – Refferenced Tests . Eric no .ED. U.S.A .
- ٣٥- Shannon .G.A. (٢٠١٢) An Application of item Response Theory in The Comparison of Four Conventional item . Discrimination Indices for Criterion – Referenced Tests. Jem. Vol, ٢٤ .No: ٤ .
- ٣٦- Silva .S.J (٢٠٠٥) Acomparison of Traditional Approaches and item Response Approaches to The Problem of Item Selection For Criterion Referenced Measurement . Ericno . ED. U.S.A .
- ٣٧- Subkoviak .M. (٢٠٠٢) . Acomparison of Reliability Estimates From Single and Double Administration of Criterion Referenced Test . N.J. Prentice . Hall .U.S.A .

