

استعمال الانموذج ثنائي المعلم في بناء اختبار تعليم التفكير لدى طلبة أقسام العلوم**التربوية والنفسية****زهراء رعد عبد الرسول****أ.د. صفاء طارق حبيب****كلية التربية ابن رشد / جامعة بغداد / قسم العلوم التربوية والنفسية****ملخص البحث:**

هدفت الدراسة إلى استعمال الانموذج ثنائي المعلم في بناء اختبار تشخيصي محكي المرجع في مادة تعليم التفكير لدى طلبة قسم العلوم التربوية والنفسية، ولتحقيق هذا الهدف تم بناء اختبار مكون من (٨٠) فقرة من نوع الاختيار من متعدد بأربعة بدائل، وتكونت عينة الدراسة من (٦٠٤) طالباً وطالبة موزعين على جامعات (بغداد، المستنصرية، العراقية، الكوفة، واسط، كربلاء، القادسية) من طلبة الصف الرابع في كليات التربية، وللدراسة الصباحية، للعام الدراسي (٢٠٢١ / ٢٠٢٢) تم اختيارهم بالطريقة الطبقيّة العشوائية، وحللت الفقرات باستعمال برنامج لغة الأوامر لاستجابة الفقرة (ICL) لتدريج الفقرات على وفق الانموذج ثنائي المعلم، وتحققت الباحثة من افتراض أحادية البعد باستعمال التحليل العاملي بطريقة المكونات الرئيسية لكايزر وحساب العلاقة بين درجة الفقرة والدرجة الكلية للاختبار، كما تم التأكد من فرض تحرر الاختبار من عامل السرعة فضلاً عن فرض تدني عامل التخمين بالتأكد من أداء الافراد الأقل قدرة على الفقرات الصعبة وقد جاءت الافتراضات مطابقة للافتراضات، ولتحديد درجة القطع للاختبار استعملت الباحثة طريقة المجموعات المتضادة، فقد بلغت درجة القطع (٥١) وهي تمثل نسبة (٦٨,٥ ٪)، وتم التحقق من افتراضات هذا الانموذج بدلالة مطابقة جميع فقرات الاختبار للانموذج عدا (٥) فقرات لم تكن مطابقة، واصبح الاختبار بصورته النهائية يتكون من (٧٥) فقرة، وظهرت النتائج ان جميع الفقرات والاختبار الكلي يتمتع بخصائص قياسية جيدة، وتوصلت الباحثة الى مجموعة من الاستنتاجات والتوصيات والمقترحات.

الكلمات المفتاحية:

(الاختبار التشخيصي محكي المرجع، الانموذج ثنائي المعلم، تعليم التفكير)

أولاً : مشكلة البحث:

ترتبط بالجانب المعرفي في التحصيل الدراسي مشاكل وصعوبات عدة والتي تختلف تبعاً لاختلاف فلسفة القياس والتقويم والهدف منه، سواء كانت هذه المشاكل مرتبطة بأدوات التقويم (الاختبارات التحصيلية)، أو مشاكل مرتبطة بالمعايير التي يستند إليها في تفسير درجة الطالب التي يحصل عليها في اختبار ما، أو مشاكل مرتبطة بتقويم الاداء، والقياس هنا يعتبر بمثابة الخطوة الاولى في عملية تقويم التحصيل الدراسي، وللقياس أدواته التي يتم بها تقدير أداء الطالب في صورة كمية (سريوة، ٢٠١٦ : ٣).

ومن بين هذه المشاكل في تفسير الدرجات المستمدة من أدوات القياس في التقويم معياري المرجع والذي يعتمد بشكل أساسي على طبيعة الجماعة المرجعية التي تستخدم في الحصول على المعيار، حيث يجب أن تكون خصائص هذه الجماعة مماثلة قدر الإمكان لخصائص مجموعة المختبرين من حيث المستوى العمري والنوع والصف الدراسي والمنطقة الجغرافية. وذلك نظراً لأنه لا يمكننا الموازنة بين الأفراد على أساس الدرجات الخام التي حصل عليها كل منهم في الاختبار، فإن هذه الدرجات لا تقدم لنا أي معلومات عن مكانة الفرد في السمة أو القدرة التي يقيسها الاختبار. وتصبح الدرجة التي يحصل عليها الفرد بلا معنى ويصعب تفسيرها إذا لم نقم بإرجاعها إلى المعيار (علام، ١٩٨٦ : ١٠٥).

وعلى الرغم من وفرة الدراسات والبحوث التي طبقت النموذج الأحادي المعلم لكن هناك صعوبة في بناء فقرات اختبار تحقق افتراض تساوي التمييز لجميع الفقرات التي يتكون منها الاختبار حتى وإن حاولنا ذلك إلا أنه

ضمن مجالات ضيقة جداً من القدرة، وعلى العكس نلاحظ قلة الدراسات التي طبقت النموذج الثلاثي المعلم لانه يحتاج الى عينات كبيرة جداً بالإضافة الى ان هناك صعوبة في تقدير بارامترات الفقرة وبخاصة بارامتر التخمين ومن هنا استعملت الباحثة النموذج ثنائي المعلم (حبيب والحمامي، ٢٠١٨ : ١٠٣)

ويستخدم التقويم المحكي كأداة اختبار تشخيصي لتقويم البرنامج التعليمي وذلك لزيادة فاعليته لوجود عيوب في التدريس؛ لأنه يشخص مواطن القصور أو الثغرات في أداء الطالب الذي يكون سبباً محتملاً في أي جزء من أجزاء البرنامج التعليمي سواء أكان في المنهج أو في طريقة التدريس أو حتى في الاختبار نفسه أو في التفاعل الصفّي و يُعدّ التوجه إلى استخدام الاختبارات المحكية توجه جديد في عمليتي التعليم والتعلم ترك العديد من الآثار الايجابية بما في ذلك تحديد جوانب القوة والضعف في تعلم الطلاب وكذلك تحديد مدى إتقانهم للمهام المطلوبة منهم ، وتحديد آليات العلاج فيما لو فشلوا بإتقانها ضمن أساليب متنوعة منها ما يتعلق بتعديل أساليب وطرائق التدريس وفي ضوء الاستفادة من إيجابيات الاختبارات المحكية بناءً على التغذية الراجعة المشجعة في استخدام هذه الطريقة في تقويم تعلم الطالب (سرقوبة، ٢٠١٦ : ٢٠).

وبناءً على ما تم ذكره فإن مشكلة البحث الحالي تتحدد في عدم وجود اختبار تشخيصي محكي المرجع في مادة تعليم التفكير لطلبة أقسام العلوم التربوية والنفسية في كليات التربية، مبني على وفق نظرية القياس الحديثة ونماذج القياس المنبثقة عنها (النموذج ثنائي المعلم) على وجه التحديد. لما يتمتع به هذا الاختبار من دقة في قياس الخصائص التربوية والنفسية للأفراد.

ثانياً: أهمية البحث:

يؤكد المهتمون بالتعليم والتربويون على ضرورة الاهتمام بجودة طرق القياس والتقويم التي تساهم في اتخاذ قرارات موضوعية على أسس علمية، وعلى اعتبار ان التقويم هو جزءاً من نسيج النظام التعليمي، والذي من خلاله يتم الوقوف على مدى تحقيق أهداف النظام التعليمي، بما يقدم من تغذية راجعة مستمرة تساعد في تعديل وتطوير النظام وترديد من كفاءته ومن ثم نوعيته وجودة مخرجاته، وقد يرجع ضعف النظام التعليمي (المدخلات _ العمليات _ المخرجات) إلى ضعف في أدوات والوسائل المستخدمة، وبالتالي فإن الاهتمام بالتقويم وأدواته يعتبر مدخلاً جيداً لإصلاح التعليم رفع مستوى جودة مخرجاته، يتم ذلك من خلال تقويم جميع مكونات النظام والعناصر المرتبطة به بما في ذلك عناصر الموقف التعليمي الاساسية والثانوية، وبما ان الطالب هو أحد العناصر الاساسية فيجب أن يشتمل التقويم على جميع جوانب شخصية المتعلم (المعرفية _ العقلية _ الجسمية) (غنيم، 2003 : ٩٦).

وتعتبر الاختبارات التربوية والمقاييس النفسية التربوية جزءاً مهماً من أجزاء العملية التعليمية، حيث يمكن للمدرس من خلالها الحكم على مدى تحقيق أهداف البرامج التعليمية والتدريبية التي يقوم بتدريسها، كما أنها تفيد الطالب أيضاً في التعرف على مستواه في التحصيل الدراسي من خلال أدائه في الاختبار، وتتعدد استخدامات الاختبارات في المؤسسات التربوية فبعضها ما يستخدم للقياس وبعضها ما يستخدم للتقويم وبعضها ما يستخدم للتشخيص (التميمي، ١٩٩٩ : ١٤٧).

وقد جاءت النظرية الحديثة بمنطلقات تتصدى للعديد من السلبيات التي تشوب النظرية الكلاسيكية للقياس (عيد ، ٢٠٠٤ : ٢٣٥)، وأن أهمية النظرية الحديثة في القياس تأتي من تركيزها على المستوى الذي يصل إليه الفرد في ادائه، وتقدير اكتسابه أو تحصيله، وأن استخدام النظرية الحديثة يجعل خصائص الفقرة (صعوبة، تمييز، تخمين) مستقلة عن عينة الافراد المستخدمة لتقدير هذه الخصائص، وكذلك تجعل تقدير قدرات الافراد مستقلة عن فقرات الاختبارات المستخدمة للحصول على هذه التقديرات، وكذلك تحسين دقة وثبات النتائج من خلال تحديد وحذف الافراد والفقرات التي لا تتوافق مع للنموذج المستخدم (Hambleton, 1989, :187)، وهذا هو الاهتمام المتعارف عليه للمقاييس والاختبارات التربوية المعاصرة.

وتبرز أهمية هذه النظرية بتعرف سمات استجابات الافراد ومميزاتهم لفقرات الاختبار، وهذه المميزات او السمات يمكن تسميتها بالسمات الكامنة، وسميت بالكامنة أي المخفية او غير المرئية لأنه لا يمكن ملاحظتها بشكل مباشر وانما نستدل على وجودها من خلال ملاحظة سلوك الفرد عن طريق استجاباته على فقرات الاختبار (علام، ١٩٨٦ : ١٠٢).

واشارت دراسة (دبوس، ٢٠١٦) إلى استخدام نظرية الاستجابة للفقرة في بناء تجمع فقرات اختبار محكي المرجع في الرياضيات وفق النموذج اللوجستي ثنائي المعلم، وقد أظهرت النتائج تحقق افتراضات نظرية الاستجابة

للفقرة. كما وبينت نتائج التحليل فيما يتعلق بمطابقة الفقرات الثنائية التدريج للنموذج الثنائي المعلم أن جميع الفقرات كانت مطابقة للنموذج ثنائي المعلم (دبوس، ٢٠١٦ : ١٤٥٣).

ولأن مادة تعليم التفكير تعتبر من المواد الحديثة التي بدء تدريسها في الجامعات، كما أن أغلب الدراسات التي بُنيت وفقاً لنظرية الاستجابة للفقرة - على حد علم الباحثة- استخدمت الأنموذج أحادي المعلمة (نموذج راش)، ومن هنا فقد ارتأت الباحثة استخدام الأنموذج ثنائي المعلم في بناء اختبار تشخيصي في مادة تعليم التفكير.

ثالثاً: أهداف البحث:

يهدف البحث الحالي إلى بناء اختبار تشخيصي محكي المرجع في مادة تعليم التفكير وفق النموذج ثنائي المعلم لدى طلبة الجامعة.

رابعاً: حدود البحث:

يقتصر البحث الحالي على طلبة الصف الرابع في قسم العلوم التربوية والنفسية في كليات التربية للعلوم الإنسانية في الجامعات العراقية (بغداد، المستنصرية، العراقية، واسط، كربلاء، الكوفة، القادسية) للعام الدراسي (٢٠٢١ - ٢٠٢٢) للدراسة الصباحية.

خامساً: تحديد المصطلحات:

الاختبار التشخيصي محكي المرجع (CRT- Criterion- Referenced Test):

عرفه كل من:

بابام وهيوسك (١٩٦٩، Popham and Husek):

الاختبار الذي لا يعتمد على تفسيره لدرجة الطالب على خصائص مجموعته التي طبق عليها الاختبار ولكنه يعتمد على موازنة أداء الطالب بمحك أو مستويات أداء متوقعة ومحددة مسبقاً، ويتم تحديد هذه المستويات في ضوء الأهداف السلوكية المراد قياسها وتقيس هذه الاختبارات مدى تحقيق التلميذ لهذه الأهداف السلوكية وبذلك يمكن التحقق من كفاءة كل طالب من اكتسابه للمهارات والمعارف الموجودة ومن ثم تشخيص نواحي الضعف والقوة (٣) : (Popham and Husek، ١٩٦٩).

عبد الحافظ (١٩٩٩):

يهتم بالكشف عن مواطن الضعف عند التلميذ ومصادر الخطأ وأسباب وقد تكون جماعية أو فردية، ولكل منهما فائدته وأهميته، فالاختبارات الجماعية توقف المعلم على نسبة شيوع الأخطاء بين التلاميذ، فهي تستخدم على نطاق واسع وتطبق على عدد كبير من الطلاب (عبد الحافظ، ١٩٩٩ : ٢٩٧).

وقد تبنت الباحثة تعريف بابام وهيوسك (١٩٦٩، Popham and Husek)، تعريفاً نظرياً للاختبار التشخيصي محكي المرجع كونه يتناسب مع ما يهدف إليه البحث الحالي.

اما التعريف الاجرائي لمصطلح الاختبار التشخيصي محكي المرجع فهو يشير الى:

عينة من الفقرات من نوع الاختيار من متعدد بواقع أربعة بدائل أحدها صحيح والثلاثة الباقية خاطئة أعدت وفق افتراضات نظرية الاستجابة للفقرة وتحليل نتائج استجابات الطلبة باستخدام الأنموذج ثنائي المعلم بما يحقق دقة وموضوعية القياس وتشخيص درجة اتقان طلاب الصف الرابع في اقسام العلوم التربوية والنفسية من خلال موازنة أدائهم بدرجة القطع المحددة في ضوء الأهداف المرسومة.

الأنموذج ثنائي المعلم:

عرفه كل من:

هامبلتون واخرون (١٩٩١، Hambleton et al):

وهو الأنموذج الذي تُمثلُ الفقرة فيه بواسطة دالة الترجيح اللوغاريتمي، معتمداً على بارامتر جديد وهو بارامتر التمييز بالإضافة لبارامتر الصعوبة المستخدم في النموذج الأحادي (نموذج راش) (Hambleton et al, 1991): 12.

علام (٢٠٠١):

هو أحد نماذج نظرية الاستجابة للفقرة والذي يفترض أن كلا من معاملي الصعوبة والتمييز متغيران، وأن التخمين لجميع الفقرات يساوي صفراً (علام، ٢٠٠١ : ١٢٦). وقد تبنت الباحثة تعريف هامبلتون وآخرون (Hambleton et al, ١٩٩١)، تعريفاً نظرياً للنموذج ثنائي المعلم كونه يتناسب مع ما يهدف إليه البحث.

اما التعريف الاجرائي لمصطلح الأنموذج ثنائي المعلم فهو يشير الى:

جميع التحليلات الإحصائية التي ستجريها الباحثة لفقرات الاختبار التشخيصي محكي المرجع ابتداءً من التحقق من الافتراضات الأساسية لنظرية الاستجابة للفقرة ومطابقة الفقرات بمعلميها (الصعوبة ، والتمييز) للمديات المقبولة، فضلاً عن التحقق من الخصائص القياسية للاختبار.

تعليم التفكير:

المادة التي تدرس في كليات التربية في المرحلة الرابعة من أقسام العلوم التربوية والنفسية في كليات التربية في الجامعات العراقية كافة، وهي من المواد الأساسية التي تزود الطالب بمجموعة متكاملة من المعارف والمهارات عن التفكير ومكوناته ومحدداته، وأنواعه والمهارات المرتبطة به، وبرامج التفكير المختلفة (الهيئة القطاعية للعلوم التربوية، ١٩٩٠ : ٥٩).

الاختبارات محكية المرجع:

إنَّ المتنبع لأدبيات القياس التربوي والنفسية يلاحظ تعدد وتنوع تعريفات الاختبارات مرجعية المحك، وربما يرجع ذلك إلى حداثة مفهوم الاختبارات محكية المرجع (مرجعية المحك) وتعدد أنواع هذه الاختبارات، كما يرجع إلى عدم الاتفاق حول مفهوم المحك الذي تنسب إليه درجات الأفراد في هذه الاختبارات، إذ يرى بعض علماء القياس أن المحك يعني نطاق شامل من المعارف والمهارات المحددة تحديداً جيداً بحيث يمكن موازنة أداء الفرد في الاختبار بهذا النطاق أن نعرف ما يستطيع أن يؤديه هذا الفرد وما لا يستطيع أن يؤديه، أي أن الموازنة هنا تكون في ضوء الفقرات الإختبارية ذاتها، فكل مفردة يشتمل عليها الاختبار تمثل معلومة أو مهارة ما ينبغي أن يكتسبها الفرد المختبر لكي ينجح في أداء عمل معين، فقائد السيارة ينبغي أن يتعرف على دلالة علامات المرور الموضحة على الطريق، وعدم تمكنه من ذلك يعني لا يمتلك المهارة الضرورية للقيادة الآمنة، ولا يستطيع اكتساب ما يترتب على ذلك من مهارات تالية (علام، ١٩٩٥ : ٢٤).

بينما يرى البعض الآخر أن المحك يشير إلى مستوى الأداء أو درجة القطع، فعندما يود المعلم أن يحقق تلاميذه مستوى إتقان معين، فإنه ينبغي أن يجيب كل منهم إجابة صحيحة عن نسبة مئوية من فقرات الاختبار يتم تحديدها مسبقاً، وبعد روبرت جلاسر أول من طور مصطلح الاختبارات محكية المرجع، وقد أضاف الكثير من الباحثين في العقدين الأخيرين السبعينات والثمانينات رؤى جديدة حول المفهوم، فتجمع لدينا عدد كبير من البيانات والدراسات عن الاستخدامات والنواحي الفنية المرتبطة بمفهوم الاختبار القياس محكي المرجع، وأنَّ هذه الدراسات والبيانات المتعددة أدت إلى تطور القياس محكي المرجع ليجمع معنى مزدوجاً فمن ناحية يعني إرجاع أداء الفرد إلى معيار ما وهو مستوى أداء محدد وتفسر درجة الطالب في الاختبار بصورة مطلقة وليست بصورة تقريبية (أبو ناهية، ١٩٩٤ : ٥٢).

خصائص الاختبارات محكية المرجع:

تعتمد الاختبارات محكية المرجع على تقويم أداء المتعلم في ضوء محك معين يأخذ مستوى الطالب بعين الاعتبار وقد شاع استخدام هذه الاختبارات حديثاً في مجال التربية والتعليم.

وتصنف هذه الاختبارات بعدد من الخصائص أهمها ما يأتي:

١. مبنية على عدد من الأهداف السلوكية.
٢. أنها معدة بحيث تكون على درجة عالية من الملائمة لأنَّ محتوياتها ترتبط بأهدافها.

٣. عبارة عن عينات ممثلة للسلوك أو الأداء الحقيقي للأفراد، يمكن تفسير الأداء عليها في ضوء درجات فاصلة محددة مسبقاً (منسي، ٢٠٠١ : ٢٢٤).

استعمالات الاختبارات محكية المرجع:

تستعمل الاختبارات في العادة لعمل قرارات حول قدرات ومعارف الطلبة ومن القرارات التي يتخذها المعلمون أو الإدارة التربوية بناء على نتائج الاختبارات رسوب أو نجاح التلاميذ في مادة صعبة أو إخضاع الطلبة ذوي التحصيل المنخفض لبرنامج علاجي معين أو تصنف التلاميذ إلى تخصصات معينة أو في تغيير إستراتيجيات التدريس التي يتبناها المعلم، وتختلف أغراض ما قام به براون (١٩٩٦) من مزوجة بين غرض الاختبار ونوع الاختبار سواء كان محكي أو معياري، (عبابنة، ٢٠٠٩ : ٢٥). والجدول (١) يبين تلك المزوجة:

جدول (١)

استعمالات الاختبارات محكية المرجع والاختبارات معيارية المرجع

غرض الاختبار	نوع الاختبار محكي او معياري
اختبار قبول في برنامج معين الاستعداد لدخول برنامج تدريبي	معياري المرجع: إذ نريد إيجاد المستوى المناسب للتلميذ في البرنامج من خلال المقارنة مع عدد من الطلبة الآخرين
تعيين	معياري المرجع: إذ نرغب بمقارنة أداء الفرد بأداء كل فرد في المجموعة
تشخيصي	محكي المرجع: إذ نرغب بقياس جوانب محددة في معرفة الطالب تكون مرتبطة بأهداف البرنامج
قياس انجاز الطالب (التحصيل)	محكي المرجع: لأننا نرغب في قياس نقاط محددة في معرفة التلميذ وهذه النقاط مرتبطة بأهداف البرنامج

مجالات استعمال الاختبارات محكية المرجع:-

أ. **تقويم التعلم الإثنائي:** حيث أنه لا يقتصر مفهوم المحك على وصف المجال السلوكي فحسب، بل يتضمن تحديد مستوى الإثنان (الأداء) على شكل تقدير عددي (أن لا تقل نسبة الإجابة الصحيحة في المهمة عن (٨٠ %) حتى يصنف التلميذ أنه متقن.

ب. تستخدم الاختبارات محكية المرجع لاتخاذ قرارات تعليمية مثل:

ماذا يعرف المتعلم؟ أين يمكن أن يكون الطالب في سلسلة التعلم؟ وما هو التعلم الذي يلزم التلميذ؟ وذلك من خلال اتخاذ قرارات قبلية التي يتم إجراؤها في بداية البرنامج التعليمي.

ج. **تستخدم في التقويم التكويني:** أشار بلوم إلى أن نطبق الاختبار محكي المرجع عند نهاية كل وحدة، وتحليل النتائج يظهر لنا جوانب القوة والضعف لدى المتعلمين واقتراح طرق علاجية للتغلب على القصور. ويتم بعدها استخدام صورة ثانية للاختبار محكي المرجع، بعد فترة زمنية كافية.

د. **يستخدم في التقويم القبلي:** حيث نستطيع هنا قياس المهارات الأساسية كونها متطلبات سابقة لتعلم جديد، ولتحديد نوع ومستوى البرنامج المناسب مع قدرات التلميذ، إضافة لتحديد المستوى القاعدي للتدريس أين تبدأ؟ وبعد التدريس أين وصلت؟.

هـ. **في مجالات التشخيص:** حيث يبنى الاختبار بحيث يتناول نماذج من المهمات يكون احتمالية الخطأ فيها كبيراً، إضافة لوجود عينة ممثلة من المهمات الأساسية.

و. **تقويم التحصيل في البرامج المبنية على النتائج:** بعض النتائج تكون بسيطة وأخرى معقدة، والأداء على الاختبار يعبر عن النتائج التي أتقنها الطالب، وهي بالضرورة مجال تحصيلي معين، وتحديد مستوى الأداء المقبول، يرتبط بالقرار المستهدف من تطوير الاختبار، وهل النتائج التعليمية تم تطويرها أم لا ؟ وتعتبر عملية تكوينية تشخيصية في آن واحد هدفها التأكد من تحقيق النتائج المستهدفة في البرنامج التعليمي (عبابنة، ٢٠٠٩ : ٢٧).

أنواع الاختبارات محكية المرجع:

١. **الاختبارات مرجعية الهدف:** هي تلك الاختبارات التي تبنى على مجموعة من الأهداف التعليمية المصاغة صياغة سلوكية، وتكون هناك مزاجية بين فقرات الاختبار وهذه الأهداف، غير أنَّ النطاق السلوكي الذي تمثله هذه الأهداف لا يكون محدداً، وبالتالي فإن الفقرات التي يشتمل عليها الاختبار تكون قليلة العدد نسبياً، لأنها لا تمثل النطاق الشامل للفقرات الممكنة أن تقيس مجموعة الأهداف، وتطبق هذه الاختبارات عادة عند الانتهاء من وحدة تعليمية أو وحدة نسقية معينة، بهدف تصنيف المختبرين في مجموعتين إحداهما حققت الأهداف والأخرى لم تحققها في ضوء نسبة مئوية محددة من الفقرات التي ينبغي الإجابة عنها إجابة صحيحة والتعرف على الأهداف التي لم يستطع كل منهم تحقيقها (علام، ١٩٨٥ : ٢٤).
٢. **الاختبارات مرجعية النطاق:** تبنى هذه الاختبارات بتحديد نطاق سلوكي شامل من المهام أو المهارات أو المتطلبات تحديداً واضحاً دقيقاً، وتنتقي الفقرات التي يشتمل عليها الاختبار من هذا النطاق إنتقاء عشوائياً أو بطرق معاينات عشوائية طبقية، وتستخدم درجات هذه الاختبارات في الحصول على تقديرات إحصائية لاحتمال إجابة فرد أو مجموعة من الأفراد عن فقرات النطاق الشامل الذي تمثله الإجابة الصحيحة على فقرات الاختبار في وقت معين، وهذا يساعد في تعميم نتائج الاختبار على النطاق الشامل (علام، ٢٠٠١ : ٢٥).
٣. **اختبارات التمكن أو الإتقان:** تعتمد هذه الاختبارات على تحديد مدى اكتساب فرد معين للسلوك الذي يهدف البرنامج التعليمي أو التجريبي إلى تنميته لديه، فاختبار التمكن أو الإتقان يساعد في اتخاذ قرارات تتعلق بإتقان الفرد لهدف تعليمي أو مهارة معينة أو نطاق من المهارات المحددة، ولذلك يمكن أن يكون هذا الاختبار مرجعي النطاق السلوكي للأداء الذي يقيسه الاختبار (علام، ٢٠٠١ : ٢٥).

خطوات بناء الاختبارات محكية المرجع:

أولاً: تحديد المحتوى المراد قياسه: يعتمد ذلك على طبيعة وحدود ذلك المحتوى، فإذا كان المحتوى المراد قياسه محدداً (مثل وحدة دراسية معينة، أو موضوع دراسي واضح المعالم) فإنه يمكن الاكتفاء بمعرفة مكونات هذه الوحدة أو هذا المحتوى، أما إذا كان المحتوى المراد دراسته واسعاً وكبيراً فإنه من المستحسن تقسيمه إلى موضوعات فرعية مرتبطة ببعضها البعض بحيث يمكن قياسها كوحدة. ويتطلب ذلك بناء عدة اختبارات كل منها يرتبط بأحد هذه الموضوعات الفرعية. ولا توجد قاعدة ثابتة لتقسيم المحتوى المراد قياسه، ولكن يجب أن يسمح هذا التقسيم ببناء اختبار أو اختبارات تكون فقراتها بمثابة عينة ممثلة تمثيلاً كافياً للموضوعات المختلفة التي يشتمل عليها المحتوى. كما يجب مراعاة طبيعة الطلاب المختبرين، والفترة الزمنية اللازمة لتعليم هذا المحتوى.

ثانياً: تحديد الأهداف العامة التي سيقاسها الاختبار: إن لكل محتوى أهداف مرتبطة به تهدف إلى قياس مدى تحققها لدى الطلاب. ومثال ذلك الهدف العام يطبق الطالب الطريقة الاستنباطية في التفكير عند حل المسائل الرياضية " وهذا الهدف العام يعبر عن نواتج متوقعة من عملية التعليم، ولكنها تحتاج إلى إعادة صياغة حتى تعبر عن أنشطة سلوكية تكون دليلاً على أن الطالب قد طبق الطريقة الاستنباطية وما إلى ذلك.

ثالثاً: تحليل الأهداف العامة إلى أهداف جزئية: ونعني بذلك إعادة صياغة الأهداف العامة عن طريق وصف عينة من الأهداف السلوكية التي يمكن أن تتخذ دليلاً على تحقق كل هدف من الأهداف العامة. وعادة يواجه مصمم الاختبار بعض الصعوبات عند إجراء هذا التحليل. ويعتمد هذا على النطاق السلوكي المراد قياسه (الجلبي، ٢٠٠٥ : ١٧٢).

رابعاً: تحديد مواصفات النطاق السلوكي الذي يقيسه الاختبار: عند الحصول على الأهداف الإجرائية من التحليل السابق قد لا تكون واضحة بدرجة تسمح بتحديد نطاق فقرات الاختبار التي تقيس السلوك المطلوب. وبالطبع يصعب اختيار عينة ممثلة لفقرات الاختبار من النطاق إذا كان هذا النطاق الذي يقيس هدفاً معيناً غير واضح. وقد سبق وأن ذكرنا أن أداء الفرد في عينة من فقرات الاختبار التي تقيس هدفاً معيناً يستخدم في تقدير درجة إتقانه للنطاق الشامل للفقرات التي تقيس هذا الهدف. لذلك فإنه يجب تعريف وتحديد هذا النطاق الشامل للفقرات تحديداً جيداً يمكن سحب عينات ممثلة من هذه الفقرات. فإذا كان النطاق الشامل يصعب تحديده تحديداً كاملاً، فإنه يمكن اعداد مواصفات الاختبار باستخدام طريقة بابام التي تم عرضها في تحديد النطاقات السلوكية. أما إذا كان النطاق الشامل حدوده معرفاً تعريفياً جيداً فإنه يمكن استخدام صيغ الفقرات التي اقترحها هايفلي.

خامساً: تكوين فقرات الاختبار: وتشمل الخطوات الآتية:

١. اختيار أنواع الفقرات المناسبة لقياس الأهداف: بعد تحديد مواصفات النطاق السلوكي يجب اختيار أفضل أنواع الفقرات التي تقيس السلوك الذي تم تحديده في المواصفات الخاصة بكل هدف. فمن المعلوم ان بعض أنواع الفقرات تصلح لقياس أهداف معينة بدرجة أفضل من غيرها. فمثلا تصلح فقرات الاختيار من متعدد، وفقرات المزاوجة، والصواب والخطأ، لقياس التذكر والفهم وأحيانا التطبيق، بينما تصلح فقرات المقال لقياس القدرة على تنظيم المعلومات، والاستنتاج والتفسير، وإعادة صياغة الأفكار.... وهكذا.

٢. تحديد العدد المناسب من الفقرات: في هذه الخطوة نحاول تحديد العدد المناسب من الفقرات التي تقيس الأهداف الممثلة في النطاق السلوكي الذي سبق تحديده (مجيد، ٢٠١٤ : ١٩٠).

٣. كتابة فقرات الاختبار: وهذه الخطوة تحتاج إلى عناية كبيرة، ففقرات الاختبار محكي المرجع تبنى على أساس مواصفات النطاق السلوكي الذي سبق اعداده. لذلك يجب أن تقيس هذا النطاق بدرجة كبيرة من الدقة، وان يكون مستوى صعوبة كل مفردة مناسبة لمستوى صعوبة الهدف الذي تقيسه، ولمستواه المعرفي. وان تكون عينة الفقرات ممثلة للنطاق السلوكي للأهداف. كما يجب ان تراعى الأصول الفنية في كتابة الأنواع المختلفة للفقرات، ويعتمد اختيار شكل الفقرة التي ستستخدم في الاختبار على طبيعة السمة او الهدف التعليمي الذي يهتم الاختبار بقياسه، وهناك نوعان رئيسيان من الفقرات هما: الفقرات الموضوعية والفقرات المقالية ويتضمن كلا النوعين اشكالا مختلفة من الفقرات، ويعود السبب في تسمية الفقرات الموضوعية بهذا الاسم إلى توفر الموضوعية في تصحيح هذا النوع من الفقرات بالاستناد إلى مفتاح تصحيح معياري، وعلى اية حال فإن الموضوعية ليست مطلقة، إذ أنها ترتبط بعملية التصحيح، وعلى ذلك فإن كلا النوعين يتوفر هما قدر معين من عدم الموضوعية، ومن الناحية الفنية فإن موضوعية الفقرة ترتبط بتصحيح الفقرة وليس ببناء تلك الفقرة. وسنفرق بين الأنواع المختلفة للفقرات من خلال الكيفية التي يتم بها الإجابة عن الفقرة، فهل يتم الإجابة عن الفقرة بالاختيار من عدة إجابات أو أن الإجابة تحتاج إلى إنشاء من قبل المفحوص؟ (عبابنة، ٢٠٠٩ : ٦٧).

مميزات وعيوب الاختبارات محكية المرجع:

مميزات الاختبارات محكية المرجع:

١. انها اختبارات اعدت في الاصل لكي تستخدم في اتخاذ قرارات عن مستويات تمكن الطالب بالنسبة لموضوع محدد لذلك نجدها تعرف باسم اختبارات التمكن او الكفاءة.
٢. نظرا لكون الاختبارات مطلقة المرجع تستخدم لتحديد التمكن او الكفاءة بالنسبة لموضوع ما لذا نجدها تتطلب من المفحوص الالتزام الكامل بكل ما تتضمنه من تفاصيل واجراءات وتعليمات.
٣. تتطلب وضع الدرجة الفاصلة لتحديد مستوى التمكن او الكفاءة والدرجة الفاصلة عبارة عن نقطة او رتبة يتم على أساسها قبول ما فوقها ورفض ما دونها من نتائج او أشياء مثل ناجح او راسب وكذلك مقبول او مرفوض.

عيوب الاختبارات محكية المرجع:

١. مشكلة تحديد المعيار.
٢. مشكلة تقدير الصدق والثبات.
٣. تحتاج الى متخصصين مدربين على اعدادها.
٤. مشكلة اختبار عينة سلوكية ممثلة للمهارة المطلوب قياسها (عبابنة، ٢٠٠٩ ، ٣١).

نظرية الاستجابة للفقرة (IRT) Item Response Theory

أدت الجهود التي بذلها علماء القياس النفسي والتربوي الى تطوير نظرية معاصرة في القياس النفسي والتربوي أطلق عليها نظرية الاستجابة للفقرة (IRT) Item Response Theory ويندرج تحت هذه النظرية مجموعة من النماذج تسمح بالقياس الموضوعي، وقد تصدت هذه النظرية للكثير من المشكلات التي تواجه النظرية الكلاسيكية في القياس، ويُفضل الكثير من علماء القياس النفسي أمثال هامبلتون ولورد تسمية هذا الاتجاه في القياس النفسي بنظرية الاستجابة للفقرة (Item Response Theory) لانها تربط بين احتمال الاستجابة الصحيحة للفرد على فقرة اختبارية معينة ، وبين خصائص هذه الفقرة ، ولهذا السبب تُستخدم النماذج الرياضية لتصف هذه العلاقة بين الاداء الملاحظ للأفراد على الاختبار وتلك السمات الكامنة (السامرائي والخفاجي، ٢٠١٢ : ٩٨).

بدأ يعول في السنوات الاخيرة على نظرية الاستجابة للفقرة في تحليل فقرات الاختبار، هذه النظرية تعتمد على اداء الفرد على كل فقرة في الاختبار لإعطاء تقدير إحصائي لقدرة الفرد التي يقيسها الاختبار (McIntir & Miller, 2000 : 216).

ونظرية الاستجابة للفقرة تقوم على اساس بعض المسلمات، وهذه المسلمات هي:

١. يمكن التنبؤ بأداء الفرد في اي اختبار بمجموعة من العوامل ، يطلق عليها السمات او القدرات الكامنة.
٢. يمكن وصف العلاقة بين اداء الفرد على اي فقرة اختبارية ومجموعة السمات او القدرات الكامنة التي يُفترض انها تؤثر في أدائه على هذه الفقرة بدالة طردية ويطلق عليها دالة خصائص الفقرة او منحني خاصية الفقرة (ICC) إذ ان هذه الدالة تحدد الافراد الذين حققوا درجات مرتفعة في السمات التي لها توقعات احتمالية عالية للاجابة الصحيحة للفقرة من المختبرين او المفحوصين الذين حققوا درجات منخفضة على السمات (Hambleton , 1989 : 148) ، إذ ان هذه الدالة تحدد احتمالات الاجابة الصحيحة للمستويات المختلفة للفقرة، فالافراد ذوو القدرة المرتفعة يمتلكون احتمالات للاجابة الصحيحة عن الفقرات اعلى من احتمالات الافراد ذوي القدرة المنخفضة ، اي كلما زادت قدرة الفرد زاد احتمال اجابته الصحيحة على الفقرة ، لذلك تعد هذه النظرية والتي كان يُطلق عليها من قبل نظرية السمات الكامنة (Latent Trait Theory) ، او نظرية المنحني المميز (Item Characteristics Curve) (Fraley , et al , 2000 : 351).

ويتميز استخدام نظرية الاستجابة للفقرة عن النظرية الكلاسيكية للاختبار بميزتين واضحتين هما:

١. يسمح للباحثين والاختصاصيين في مجال القياس التربوي والنفسي بتصنيف الخاضعين للاختبار بشكل اكثر دقة فيما يتعلق بنماذج استجاباتهم وفيما يمتلكونه من قدرات وسمات.
٢. ان استخدام تقديرات نظرية الاستجابة للفقرة يسمح بعمل استقرار لدرجات المهتمين والمستخدمين في المستقبل بينما لا تسمح درجات النظرية الكلاسيكية للاختبار بعمل استقرار للمستخدمين في المستقبل (إسماعيل، ٢٠٠٧ : ١٨).

بعض المفاهيم المتعلقة بنظرية الاستجابة للفقرة:

فيما يلي استعراض لمعنى بعض المفاهيم الاساسية المتعلقة بنظرية الاستجابة للفقرة ونماذجها:

١. صعوبة الفقرة

هي نقطة على متصل السمة الكامنة تمثل احتمالية الوصول الى الاجابة الصحيحة تساوي (٠,٥٠) وهذه النقطة تقابل نقطة انقلاب المنحني، ويرى لورد (Lord , 1980) ان معامل الصعوبة في ظل نظرية الاستجابة للفقرة يشير الى مقدار القدرة اللازمة لكي يصبح احتمال اعطاء الاجابة الصحيحة لفقرة ما (٠,٥٠)، وكلما زادت قيمة معامل صعوبة الفقرة حدثت اراحة للمنحني المميز للفقرة جهة اليمين ، وكلما قلت قيمته ازيح المنحني المميز للفقرة جهة اليسار.

٢. تمييز الفقرة

تُقدر قدرة الفقرة على التمييز بمعامل التمييز الذي يمكن التعبير عنه بالميل النسبي للمنحني المميز للفقرة على محور السمة عند نقطة الانقلاب ، وكلما زادت درجة التواء المنحني المميز زادت قيمة معامل التمييز ، وبذلك يشير تمييز الفقرة الى معدل التغير في احتمال الاستجابة الصحيحة للافراد على الفقرة بالنسبة لمستوى القدرة ، وعلى ذلك فان احسن الفقرات تمييزاً هي تلك الفقرات متوسطة التمييز ، واحسن ميل للمنحني المميز للفقرة عندما تكون زاوية ميله على محور السمة (٤٥°)، عندئذٍ تتأرجح قيمة ميل المنحني حول القيمة المثلى على متصل السمة وهي قيمة واحد (كاظم ، ١٩٨٨ : ٧٣).

٣. تخمين الفقرة

يُقاس تخمين الفقرة باحتمال الاجابة المناسبة (نقطة على المحور الراسي) عند المستويات المنخفضة جدا من القدرة، وهذا البارامتر يستخدم في اسئلة الاختيار من متعدد ، وهو يُمثل حدوث الاستجابة الصحيحة بمحض الصدفة، وقيمة بارامتر التخمين تمثل الجزء الذي يقطعه المنحني المميز للفقرة من المحور الراسي الذي يمثل الاجابة الصحيحة للفقرة ، إذ ان الفرد الذي يقع على مستوى قدرة منخفض من متصل القدرة ، يتوقع ان يخمن الاجابة الصحيحة على فقرة الاختيار من اربعة بدائل باحتمال (٠,٢٥) ، وهذا يعد مكان البدء في حال وجود التخمين ،

وبالتالي يكون الجزء الذي يقطعه المنحنى المميز للفقرة مع المحور الاحتمالي يمثل (٠,٢٥) (Rust & Golmbok, 1999, 57).

افتراضات نظرية الاستجابة للفقرة :

١. افتراض احادية البعد

يُشير مصطلح الأبعاد إلى عدد السمات الكامنة المسؤولة عن أداء الأفراد في الاختبار والسمة مفهوم يستخدم لوصف سلوك الأفراد وهو عبارة عن تجمع من السلوك المتداخل والمتربط بصورة متكاملة ، وهذا يعني ان السمة ليست صفة منفردة ، وإنما تُعدُّ مفهوماً مجرداً غير ملموس، لذا يُعد تحديد وتعريف السمات المراد قياسها من الخطوات الأساسية في القياس السلوكي، وتفترض معظم نماذج نظرية الاستجابة للفقرة ان هناك سمة او قدرة واحدة فقط كافية لتفسير وتوضيح الفروق بين أداءات الأفراد على الاختبار، وتُسمى تلك النماذج بالنماذج أحادية البعد، اما النماذج التي تفترض وجود أكثر من قدرة واحدة تكمن خلف أداء الفرد وتسمى نماذج متعددة الأبعاد (Weiss & Yoes, 1994, 72).

٢. الاستقلال الموضوعي

ويُقصد به ان استجابات الفرد لل فقرات المختلفة مستقلة استقلالاً إحصائياً، وهذا يعني ان أداء الفرد على الفقرة لا يتأثر بأدائه على الفقرات الأخرى في الاختبار، ولا يتأثر أيضاً بأداء الأفراد الآخرين الذين يؤدون هذا الاختبار، اي ان تقدير معالم الفقرة لا يعتمد على تقديرات معالم الفقرات المكونة للاختبار، ولا يعتمد على تقديرات قدرات الأفراد الذين يجيبون عن هذه الفقرات، وكذلك تقدير القدرة (السمة) للفرد لا يعتمد على قدرة الأفراد الآخرين الذين يؤدون الاختبار، ولا على تقديرات معالم الفقرات التي يؤدونها، اي يُتوقع ان نحصل على نفس تقدير القدرة وذلك من اجل اي مجموعة جزئية من الفقرات طالما انها ملائمة وكذلك من اجل اي عينة من المختبرين او المفحوصين، يُتوقع ان تكون تقديرات معالم الفقرة نفسها، وبالتالي فان الاستقلال الموضوعي يتطلب ان يكون احتمال الاجابة الصحيحة او الخاطئة للفرد على فقرة ما مستقلاً عن اي فقرة أخرى عند مستوى مُعين من القدرة، والعوامل المؤثرة في احتمال استجابة الفرد على فقرة معينة هي فقط السمة الكامنة التي تقيسها الفقرة، ومعالم الفقرة التي يُجيب عليها الفرد (Hambleton, 1989 : 136).

٣. افتراض منحنى خاصية الفقرة

وهي من الافتراضات الأساسية التي تقوم عليها نظرية الاستجابة للفقرة التي تتناول سمة احادية البعد، وهو دالة رياضية تربط بين احتمال نجاح الفرد في اجابة الفقرة وبين السمة او القدرة التي تقيسها مجموعة فقرات، او يقيسها اختبار يحوي هذه الفقرة، اي هو دالة غير خطية لانحدار درجة الفقرة على السمة الكامنة، او القدرة التي يقيسها الاختبار والاختلاف الرئيس بين نماذج الاستجابة للفقرة يعتمد على الصيغة الرياضية لمنحنى خاصية الفقرة (Hambleton, 1989, 151).

وهناك اجماع بان منحنى خاصية الفقرة يمكن تحديده من خلال اربعة متغيرات ، وهي قدرة الفرد وثلاثة متغيرات تتعلق بالفقرة والتي تدعى معالم (بارامترات) الفقرة، وهذه المعالم هي معامل الصعوبة ، ومعالم التمييز ، ومعامل التخمين، وصعوبة الفقرة تقابل بنقطة على متصل السمة الكامنة او القدرة يمر فيها منحنى خاصية الفقرة عند القيمة الاحتمالية (٠,٥)، وبالتالي فان صعوبة الفقرة تقابل نقطة على متصل السمة او القدرة يكون فيها احتمال الاجابة الصحيحة يساوي احتمال الاجابة الخاطئة ويساوي (٠,٥) (Aiken, 1998 : 53).

٤. افتراض عامل السرعة Speedness :

تفترض نظرية الاستجابة للفقرة ان عامل السرعة لا يلعب دوراً في الإجابة عن فقرات الاختبار، بمعنى اخفاق المختبرين او المفحوصين في اجابة فقرات الاختبار يرجع الى انخفاض قدراتهم وليس الى تأثير عامل السرعة على اجاباتهم ، ويمكن تقدير ما اذا كان عامل السرعة قد لعب دورا في الاجابة وذلك عن طريق معرفة عدد المختبرين او المفحوصين الذين لم يتمكنوا من الاجابة عن جميع فقرات الاختبار الذي اجري عليهم ، وعندما تكون السرعة احد العوامل المؤثرة في الاداء على الاختبار ، فان هناك قدرتين تؤثران في هذا الاداء وهما القدرة التي يقيسها الاختبار ، والقدرة الخاصة بسرعة الاداء ، وهذا انتهاك لافتراض احادية البعد.

الدراسات السابقة:

جدول (٢) الدراسات السابقة العربية والاجنبية

اولاً: الدراسات العربية		ت
استخدام نظرية الاستجابة للفقرة في بناء فقرات اختبار محكي المرجع في الرياضيات بفقرات ثنائية التدرج ومتعددة التدرج وفق النموذج اللوجستي ثنائي المعلم	العنوان	١
محمد طالب دبوس (٢٠١٥)	اسم الباحث والسنة	
جامعة الاستقلال \ فلسطين	مكان اجراء الدراسة	
استخدام نظرية الاستجابة للفقرة في بناء تجمع فقرات اختبار محكي المرجع وفق النموذج اللوجستي ثنائي المعلم	أهداف البحث	
٥٠٢ طالباً وطالبة،	مجتمع البحث وعينته	
بناء تجمع من فقرات اختبارات محكية المرجع في الرياضيات من نوع الأسئلة الاختيار من متعدد مؤلف من (٥٠) فقرة لكل فقرة (٤) بدائل	أداة البحث	
SPSS، Bilog MG3، MULTILOG، 7.	الوسائل الإحصائية	
تحقق افتراضات نظرية الاستجابة للفقرة وتم التحقق من صدق وثبات الاختبار باستخدام الانموذج الثنائي المعلم	أبرز النتائج	
بناء اختبار محكي المرجع في الرياضيات باستخدام نظرية الاستجابة للمفردة وفق النموذج اللوجستي ثنائي المعلم لطلاب الصف الخامس	العنوان	٢
عبد الله بن محمد السلامي (٢٠١٨)	اسم الباحث والسنة	
جامعة تبوك / المملكة العربية السعودية	مكان اجراء الدراسة	
استخدام نظرية الاستجابة للمفردة في بناء اختبار محكي المرجع في الرياضيات وفق النموذج اللوجستي ثنائي	أهداف البحث	
٤٠٠ طالباً وطالبة، في الصف الخامس الابتدائي التابعة لإدارة تعليم تبوك	مجتمع البحث وعينته	
بناء اختبار محكي المرجع في الرياضيات مكون من ٤٠ فقرة	أداة البحث	
spss - Bilog MG3	الوسائل الإحصائية	
بينت نتائج التحليل فيما يتعلق بمطابقة فقرات الاختبار ثنائية التدرج للنموذج الثنائي المعلم ان الفقرات كانت مطابقة للنموذج الثنائي	أبرز النتائج	
ثانياً: الدراسات الأجنبية		
Education in Estimates Based on Classical Test and Item Response Theories	العنوان	٣
التحقق من الثبات في التقديرات القائمة على الاختبار الكلاسيكي ونظرية الاستجابة للفقرة		
اديدوين Adedoyin (2010)	اسم الباحث والسنة	
نيجيريا	مكان اجراء الدراسة	
التحقق من ثبات تقدير معالم الافراد في النظرية التقليدية ونظرية الاستجابة للفقرة (النموذج ثنائي المعلمة)	أهداف البحث	
تكونت عينة البحث من ٥٠٠٠ طالباً وطالبة	مجتمع البحث وعينته	
الورقة الأولى من اختبار شهادة الثانوية العامة في الرياضيات المكون من ٤٠ فقرة	أداة البحث	
اظهرت النتائج ان هناك فروق ذات دلالة إحصائية في ثبات تقدير معالم الافراد ولصالح النموذج ثنائي المعلمة	أبرز النتائج	

العنوان	Classical test theory versus item response theory: An evaluation of the comparability of item analysis results نظرية الاختبار الكلاسيكية مقابل نظرية استجابة العنصر: تقييم قابلية مقارنة نتائج تحليل العنصر
اسم الباحث والسنة	Onn (٢٠١٣) , اون
مكان اجراء الدراسة	نيجيريا
أهداف البحث	مقارنة النظرية الكلاسيكية ونظرية الاستجابة للفقرة باستخدام النموذج ثنائي المعلم
مجتمع البحث عينته	تكونت عينة البحث من 69 طالباً وطالبة
أداة البحث	اعداد اختبار في مادة الفيزياء مكون من ٥٠ فقرة
أبرز النتائج	اظهرت النتائج ان معامل الثبات للنموذج ثنائي المعلمة اعلى من معامل الثبات في النظرية الكلاسيكية

ويمكن توضيح جوانب الافادة من الدراسات السابقة بتقديم مؤشرات واضحة حول مشكلة البحث كما ساعدت على صياغة الاهداف بدقة، وقد وضحت الدراسات السابقة الطرائق المستعملة في توصيف مجتمع الدراسة واساليب اختيار العينة، كما تبين من خلال عرض الدراسات السابقة الخطوات الواجب اتباعها للتحقق من افتراضات النموذج وحساب الخصائص القياسية والبرامج الاحصائية المستعملة، فضلاً عن مناقشة نتائج البحث الحالي في ضوء الدراسات السابقة.

منهج البحث:

ان المنهج الوصفي المقارن هو المنهج الملائم لموضوع البحث الحالي وذلك لوصف خصائص الاختبار وفق الانموذج ثنائي المعلم على وفق نظرية القياس الحديثة.

مجتمع البحث:

ويتكون مجتمع البحث الحالي من طلبة المرحلة الرابعة في قسم العلوم التربوية والنفسية في كليات التربية لدى جامعات (بغداد ، المستنصرية ، العراقية، القادسية، واسط، كوفة، كربلاء، ميسان، البصرة، ذي قار) للعام الدراسي (٢٠٢٢-٢٠٢١)، وقد بلغ عدد الطلبة الإجمالي (٩١٩) طالب وطالبة.

عينة البحث :

حددت الباحثة العينة بالطريقة الطبقيّة العشوائية من كليات التربية قسم العلوم التربوية والنفسية من كل من جامعة (بغداد – المستنصرية – العراقية – واسط – كربلاء – الكوفة – القادسية) والتي بلغ عددها (٦٠٤) طالباً وطالبة بنسبة (٦٥,٧%) من المجموع الكلي لمجتمع البحث.

إجراءات بناء الاختبار:

١. **تحديد المحتوى الدراسي:** قامت الباحثة بتحديد المحتوى الدراسي لمادة تعليم التفكير من خلال الرجوع إلى ما هو مقرر في هيئة اللجنة القطاعية الخاصة بمقررات مناهج كلية التربية/ جامعة بغداد، وذلك لغرض تحديد الفصول ثم الموضوعات في مادة تعليم التفكير.

٢. **صياغة الأهداف السلوكية:** قامت الباحثة بصياغة الأهداف السلوكية لفصول المادة المحددة والمستويات المعرفية الثلاثة الاولى في تصنيف بلوم (المعرفة، الفهم، التطبيق)، اذ بلغت الأهداف (١١٥) هدفاً سلوكياً موزعة على المستويات الثلاثة وعلى النحو التالي: (٦٥) هدفاً سلوكياً لمستوى المعرفة وبلغ وزنه (٥٦%)، و(٣٠) هدفاً لمستوى الفهم وبلغ وزنه (٢٦%)، و(٢٠) هدفاً لمستوى التطبيق وبلغ وزنه (١٧%).

٣. **اعداد الخارطة الاختبارية:** لبناء اختبار تشخيصي محكي في مادة تعليم التفكير لدى طلبة الجامعة تم إعداد جدول المواصفات ببعدي المحتوى والأهداف السلوكية، اذ تم سحب ٨٠ هدف سلوكي بطريقة عشوائية أي ما يمثل (٧٠%) من مجموع الأهداف السلوكية، وتم توزيع الأهداف السلوكية والبالغة (٨٠) هدف سلوكي الموضوعات لفصول المادة للمحتوى المحدد، وبحسب الأهمية النسبية لكل فصل.

٤. **صياغة الفقرات والبدايل:** ولجأت الباحثة إلى الأسئلة من نوع الاختيار من متعدد ذي البدائل الاربعة في بناء فقرات الاختبار جميعها، وذلك لان هذه الصيغة الأكثر شيوعاً واستخداماً في الاختبارات التشخيصية محكية المرجع التي يتم تحليلها على وفق الانموذج ثنائي المعلم ونظرية الاستجابة للفقرة الاختبارية.

٥. **التحقق من صدق فقرات الاختبار:** اعتمدت الباحثة قيمة مربع كاي (٢٤) محك لقبول الفقرة او رفضها التي توازي نسبة (٨٠ %) من نسبة اتفاق الخبراء، تم إجراء بعض التعديلات على الفقرات الاختبارية والاهداف السلوكية والمستويات المعرفية التي تقيس الفقرات الاختبارية لتكون أكثر ملائمة من الناحية الفنية والعلمية، مع ملاحظة عدم حذف اي فقرة اختبارية من الاختبار

٦. **تجربة وضوح الفقرات والتعليمات:** قامت الباحثة بتطبيق الاختبار بصيغته الأولية على عينة بلغت (٣٠) طالب وطالبة اختيروا بطريقة عشوائية من طلاب الصف الرابع قسم العلوم التربوية والنفسية من جامعة ميسان، بتسجيل وقت انتهاء كل طالب حتى اخر طالب، وكان المتوسط للزمن الذي استغرق للإجابة عن الاختبار بأكمله هو (٣٥) دقيقة

٧. **تصحيح الاختبار:** اعتمدت الباحثة أسلوب التصحيح اليدوي باستعمال مفتاح التصحيح المثقب.

٨. تجربة التحليل الاحصائي:

التحقق من افتراضات النموذج (ثاني المعلم، PLM2) :

إن الافتراضات الرئيسة لنظرية الاستجابة للفقرة الاختبارية وهي : افتراض أحادية البعد ، وكذلك الاستقلال الموضوعي والمطابقة للمنحنى المميز للفقرة ، فمن الضروري جداً قبل استعمال النموذج المعتمد (ثنائي المعلم، PLM2) في التحليل الإحصائي، التحقق من الافتراضات الخاصة بالنموذج وعلى النحو الاتي :

أولاً: التحقق من افتراض أحادية البعد :

إن مصطلح أحادية البعد له معنيان متميزان. أولاً ، يمكن تفسيره بالمعنى النفسي بالإشارة إلى القدرة الأساسية التي تؤثر على الأداء. ثانياً ، يمكن تعريفه على أنه خاصية قياس نفسية تشير إلى بُعد قياس أساسي واحد، (McNamara 1996: 271). ويعني أن مجموعة من الفقرات في الاختبار تقيس قدرة واحدة فقط . وفي الواقع ، فمن الناحية العملية ، فإن مطلب أحادية البعد هو أن يكون عاملاً مهماً، (Hambleton and Cook, 1977: 77) وينظر إلى افتراض أحادية البعد باعتباره يمثل وضعاً مثالياً مماثلاً لافتراض تجانس التباين في تحليل التباين. فأنتهاك بدرجة ما لافتراض أحادية البعد قد يكون يؤدي إلى إشكالية، (Ayala, 2015; 23).

١. **التحليل العاملي :** للتحقق من الخطوة الأولى أخضعت الباحثة فقرات الاختبار التشخيصي المحكي في مادة تعليم التفكير للتحليل العاملي ، وذلك للتثبت من البنية العاملية للاختبار ، وعليه أخضعت الاستمارات التي أجري عليها التحليل الإحصائي للفقرات والبالغة (٦٠٤) اجابة للتحليل العاملي ، وبعد الاعتماد على نتائج التحليل العاملي باستخدام أسلوب تحليل المكونات الأساسية وبعد إجراء التدوير المتعامد بطريقة الفارماكس (أقصى التباين) لكاي، تم الحصول على عامل واحد وجذر كامن أكثر من واحد عدد صحيح بحسب طريقة الحدود الدنيا لكتمان. بلغ الجذر الكامن للاختبار (٢٦,٢٢٨) وأسهم في تفسير (٣٢,٧٨٥) من التباين الكلي، إذ تعد هذه الطريقة أن العامل الدال أو الجذر الكامن الذي يمكن تفسيره يساوي أو يزيد عن الواحد عدد صحيح (عبد الخالق ، ١٩٨٣ : ١١٨) ، وعدت جميع فقرات الاختبار التشخيصي المحكي لمادة تعليم التفكير متشعبة بالعامل العام طبقاً لمعيار جيلفورد ونسبة أكثر من (٠,٣٠) (Guilford, 1998: 500).

٢. **الاتساق الداخلي :** لقد حسبت الباحثة معامل الارتباط بين درجة كل فقرة والدرجة الكلية للاختبار باستعمال معامل ارتباط بونت بايسيرال ، وباستعمال عينة التحليل الإحصائي البالغ عددها (604) طالب وطالبة، وبعد مقارنة قيم معاملات الارتباط المحسوبة للاختبار التشخيصي المحكي لمادة تعليم التفكير بين درجة الفقرة والدرجة الكلية للاختبار ، وتبين إن علاقة جميع الفقرات لكل الاختبارات بالدرجة الكلية دالة إحصائياً. وهي دلالة على إن فقرات الاختبار متسقة فيما بينها لقياس الخاصية ، مما يدل على تحقق الفرض الأول (أحادية البعد) من افتراضات نظرية الاستجابة للفقرة الاختبارية.

ثانياً: التحقق من افتراض الاستقلال المحلي :

وحقق هذا الافتراض من خلال نتائج التحليل العاملي بواسطة ، الجذر الكامن والتباين المفسر وتشبع الفقرات بعامل واحد ، وكذلك من خلال معاملات الارتباط للفقرة بالدرجة الكلية ، وعليه تعد هذه النتائج مؤشرات تحقق افتراض الاستقلال المحلي من خلال التحقق من احادية البعد .

ثالثاً: المنحني المميز للفقرة :

من خلال استعمال برنامج (ICL) المحوسب في تحليل البيانات عينة البحث ، فقد تم الحصول على المنحنيات المميزة لجميع فقرات الاختبار التشخيصي المحكي لمادة تعليم التفكير من خلال مخرجات هذا البرنامج.

رابعاً : التحرر من عامل السرعة :

قد بلغت نسبة الطلبة الذين أكملوا جميع الاختبارات هي (٨٧%)، وان نسبة الفقرات التي تمت الإجابة عليها هي (٩٢%)، وهذا يعني أن الاختبار متحرر من عامل السرعة ، وعليه يمكن القول ان البيانات ملائمة للتحليل وفقاً لنموذج ثنائي المعلم PLM2 .

خامساً : فرض تدني عامل التخمين :

قد أختارت الباحثة مجموعة من الأفراد بلغ عددهم (٦٣) طالباً وطالبة من ذوي الأقل قدرة بالدرجة الكلية ويمثلون نسبة (١٠%) للتأكد من عدم تضمين الانموذج ثنائي المعلم لأثر التخمين ، ومن ثم دراسة أدائهم على الفقرات الأكثر صعوبة وذلك باستعمال الحقيبة الإحصائية (SPSS)، ومقارنة نسب أولئك الأفراد الأقل قدرة الذين أجابوا إجابة صحيحة على تلك الفقرات الصعبة بالقيمة النظرية للتخمين العشوائي، والذي تبلغ قيمتها (0.25) في حالة البدائل الأربعة كما في الاختبار الحالي.

سادساً : التحقق من ملائمة البيانات للأنموذج الثنائي المعلم PLM2 :

ولغرض الكشف عن أجابات عينة البحث على جميع فقرات الاختبار التشخيصي المحكي لمادة تعليم التفكير أي الافراد الذين كانت جميع إجاباتهم اجابات صحيحة لكل الفقرات ، وكذلك الافراد الذين كانت جميع اجاباتهم اجابات خاطئة لكل فقرات الاختبار والتي تسمى بالبيانات الصفرية ، وتسمى البيانات التامة اي الفقرات التي أجاب عنها كل أفراد العينة إجابة صحيحة ، وقد قامت الباحثة بتدقيق البيانات الخاصة بأجابات أفراد العينة على اختبارات فليب كارتر للرشاقة العقلية ، ولم تجد الباحثة مثل هذه الحالات المذكورة اعلاه. وبعد ذلك قامت الباحثة بادخال بيانات افراد عينة التحليل البالغة (٦٠٤) طالباً وطالبة، على الاختبار التشخيصي المحكي لمادة تعليم التفكير والتي يمكن من خلالها الحصول على تقديرات دقيقة لعمليات التقدير لقيمة مربع كاي لأختبار حسن المطابقة، وكذلك التحقق من فائدة الأنموذج على التنبؤ بالدرجات الفعلية للاختبار، فإذا كانت قيمة مربع كاي المحسوبة اكبر من القيمة الجدولية لمربع كاي تكون دالة وتحذف الفقرة ، والجدول (٣) يظهر قيم مربع كاي (٢٤) لفقرات الاختبار للحكم على مدى ملائمتها الأنموذج ثنائي المعلم PLM2.

الجدول (٣)

قيم معلم الصعوبة ومعلم التمييز وقيم مربع كاي لفقرات الاختبار التشخيصي المحكي لمادة تعليم التفكير على مدى ملائمة الأنموذج ثنائي المعلم PLM2

ت الفقرة	معامل الصعوبة	الخطأ المعياري	معامل التمييز	الخطأ المعياري	قيمة مربع كاي	درجة الحرية
١	٠,٦٩٤	٠,١٨٠	0.904	0.158	١٩,٣٣٢	18
٢	٠,٤٢٣-	٠,١٢٨	1.243	0.232	١٣,٤٣٩	18
٣	٠,٩٠٥	٠,١٩٦	1.526	0.278	٧,٤٢٧	18
٤	١,٩٩٢-	٠,١٢٠	1.241	0.206	٢٠,٦٦١	18
٥	١,١١٣	٠,٢١٣	1.416	0.422	١٩,٢٨٩	18
٦	٠,٧٢٧	٠,١٨٢	1.207	0.209	١٨,٣٥١	18
٧	٠,٧٦٠	٠,١٨٥	1.091	0.193	٢٢,٩١٧	18
٨	٠,٦٦٢	٠,١٧٨	1.289	0.220	١١,٨٧٨	18

18	١٧,٢٤٦	0.168	0.963	٠,١٨٥	٠,٧٦٠	٩
18	١٠,٤٧٥	0.160	0.950	٠,١١٥	١,٢٣٩-	١٠
18	٢٢,١٠٤	0.223	1.420	٠,١١٥	١,١٩٩-	١١
18	١٥,٧١٣	0.157	0.892	٠,١٩٠	٠,٨٣١	١٢
18	١٥,٠٩٢	0.176	1.191	٠,١٧٦	٠,٦٣٠	١٣
18	١٦,٣٦٠	0.319	1.451	٠,١٩٦	٠,٩٠٥	١٤
18	٢٢,٥١٢	0.312	1.897	٠,١٩٣	٠,٨٦٧	١٥
18	١٨,١٣٣	0.190	1.256	٠,١٨٠	٠,٦٩٤	١٦
18	٨,٥٤٠	0.169	1.244	٠,١٩٣	٠,٨٦٧	١٧
18	٢٠,٤١٨	0.240	1.515	٠,١١٧	١,٨١٠-	١٨
18	١٥,٩٨٧	0.193	1.400	٠,١٩٣	٠,٨٦٧	١٩
18	١٧,٤١٠	0.298	1.١98	٠,١٢٠	٠,٧٨٨-	٢٠
18	١٢,٩٠٢	0.260	1.٤98	٠,١٦٥	٠,٤٥٦	٢١
18	٢٥,٢٤٨	0.211	1.203	٠,١٢١	٢,٠٦٤-	٢٢
18	١٩,٥٨٧	0.275	1.٣70	٠,١٨٢	٠,٧٢٧	٢٣
18	١٦,٩٥٩	0.366	1.٨01	٠,١٩٣	٠,٨٦٧	٢٤
18	١٠,٧٩٥	0.238	1.٢27	٠,١١٧	١,٧٩٧-	٢٥
18	١٠,٠٧٢	0.158	0.904	٠,١٥٦	٠,٣٠٢	٢٦
18	٢٤,٦٧٤	0.232	1.243	٠,١٧٤	٠,٦٠٠	٢٧
18	١٠,٩١٢	0.278	1.526	٠,١٣٢	٠,٢٧١-	٢٨
18	٢٠,٢٣٤	0.206	1.241	٠,١٨٧	٠,٧٩٥	٢٩
18	٢١,٠٤٩	0.422	1.416	٠,١١٥	١,٢١٣-	٣٠
18	١٥,١١٩	0.209	1.207	٠,١٢٠	٠,٧٥٩-	٣١
18	١٨,٥٠٩	0.193	1.091	٠,١٢٩	٠,٣٩٠-	٣٢
18	١٧,٠٢١	0.220	1.289	٠,١٢٣	٠,٦١٢-	٣٣
18	١١,٧٨٤	0.168	0.963	٠,١٢٩	٠,٣٩٠-	٣٤
18	١٢,٧٦٢	0.160	0.950	٠,١٣٥	٠,١٨١-	٣٥
18	٧,٨٠٣	0.223	1.420	٠,١٢٧	٠,٤٣٩-	٣٦
18	١٩,٠٧٠	0.157	0.892	٠,١٢٤	٠,٥٦٦-	٣٧
18	١٠,٧٣٧	0.176	1.191	٠,١٢٥	٠,٥١٩-	٣٨
18	١٥,٢٧٣	0.319	1.451	٠,١٣٥	٠,١٨١-	٣٩
18	١٨,٩١٥	0.312	1.897	٠,١٣٧	٠,١٢٥-	٤٠
18	٧,٥٢٢	0.190	1.256	٠,١٢٧	٠,٤٣٩-	٤١
18	١٨,٦٤٥	0.169	1.244	٠,١٢٥	٠,٥١٩-	٤٢
18	٧,٦٣٧	0.240	1.515	٠,١٢٦	٠,٥٠٣-	٤٣

18	١٣,٠٢٠	0.193	1.400	٠,١٣٣	٠,٢٥٣-	٤٤
18	٢٠,٨٢٧	0.298	1.998	٠,١٣٧	٠,١٢٥-	٤٥
18	١٤,٦٥٦	0.158	0.904	٠,١٣٤	٠,٢٣٥-	٤٦
18	١٧,٥٤٢	0.232	1.243	٠,١٣٤	٠,٢١٧-	٤٧
18	١٢,٥٩٥	0.278	1.526	٠,١٢٩	٠,٣٩٠-	٤٨
18	١٨,٢٣٧	0.206	1.241	٠,١٣٣	٠,٢٥٣-	٤٩
18	١٤,٧٦٨	0.422	1.416	٠,١٣١	٠,٣٢٣-	٥٠
18	١٤,٨٣٨	0.209	1.207	٠,١٣٥	٠,١٨١-	٥١
18	١٥,٠٧١	0.193	1.091	٠,١٣٨	٠,١٠٦-	٥٢
18	١٣,٥٧٢	0.220	1.289	٠,١٣٥	٠,١٨١-	٥٣
18	١١,٠٢٩	0.168	0.963	٠,١٣٢	٠,٢٧١-	٥٤
18	١٠,٩٢٨	0.160	0.950	٠,١٤٢	٠,٠٠٨-	٥٥
18	١٣,٣٨٩	0.158	0.904	٠,١٣٣	٠,٢٥٣-	٥٦
18	١٤,٩١٧	0.232	1.243	٠,١٢٥	٠,٥٣٥-	٥٧
18	٩,٨٨٠	0.278	1.526	٠,١٣٧	٠,١٤٤-	٥٨
18	٢٢,٤٩٦	0.206	1.241	٠,١٤٠	٠,٠٤٨-	٥٩
18	٢٢,٠٠٨	0.422	1.416	٠,١٣٥	٠,١٨١-	٦٠
18	١٥,٣٠٦	0.209	1.207	٠,١٣٤	٠,٢٣٥-	٦١
18	٦,٣١١	0.193	1.091	٠,١٣٢	٠,٢٧١-	٦٢
18	١٧,٨٤٨	0.220	1.289	٠,١٢٥	٠,٥١٩-	٦٣
18	٢٥,٣٠٧	0.168	0.963	٠,١٢٥	٠,٥٣٥-	٦٤
18	١٠,٩٨٥	0.160	0.950	٠,١٣٢	٠,٢٧١-	٦٥
18	١٦,٨٧٦	0.223	1.420	٠,١٤٢	٠,٠٠٨-	٦٦
18	١٥,٥٦٠	0.157	0.892	٠,١٤٠	٠,٠٤٨-	٦٧
18	٨,٤٢١	0.176	1.191	٠,١٣٢	٠,٢٨٨-	٦٨
18	٢٠,٤٢٥	0.319	1.451	٠,١٢٨	٠,٤٢٣-	٦٩
18	١٤,١٢٩	0.312	1.897	٠,١٢٧	٠,٤٧١-	٧٠
18	٩,٨٤٠	0.190	1.256	٠,١٢٨	٠,٤٢٣-	٧١
18	١٩,٦٤٨	0.169	1.244	٠,١٤٣	٠,٠١٢	٧٢
18	١٣,٣١٨	0.240	1.515	٠,١٢٩	٠,٣٩٠-	٧٣
18	١٠,٥٥٤	0.193	1.400	١,٠٠٢	١,٣٧٤	٧٤
18	١٤,١٨٢	0.298	1.٣98	٠,٢٥٠	١,٤٨٤	٧٥
18	٩,٢٤٠	0.260	1.٥98	٠,١٥٤	٠,٢٥٣	٧٦
18	١٥,١٤٧	0.211	1.203	٠,١٣٣	٠,٢٥٣-	٧٧
18	١١,٤٤٩	0.275	1.٥70	٠,٣٨٣	١,٤٠٧	٧٨

٧٩	٠,٤٣٠	٠,١٦٤	1.٢01	0.366	١٣,٣٣٦	18
٨٠	١,٢٧٠	٠,٣٥٩	1.727	0.238	١٦,٦٩٣	18

يتبين من الجدول (٣) على أن جميع فقرات الاختبار مطابقة للنموذج ثنائي المعلم PLM2 (صعوبة وتمييز) حيث لم تظهر فروق دالة أحصائياً بين توزيع إجابات الأفراد .

سابعاً: استقلالية القياس :

إن استقلالية القياس تدل على قدرة العينة والتي تستجيب للاختبار ، من حيث كون قدرة الفرد لا تستند على قدرة الأفراد الآخرين الذين يجيبون عن الاختبار ، وتقدير صعوبة الفقرة لا تستند على قدرة الأفراد الذين يجيبون عن الاختبار ، ولغرض تحقيق استقلالية القياس قامت الباحثة بحساب تقدير صعوبة الفقرة مستقلاً عن باقي فقرات الاختبار ، وتقدير قدرة الأفراد بشكل مستقل عن باقي فقرات الاختبار التي يجيب عنها أفراد العينة، وتحققت الباحثة من هذين الجانبين وفق الآتي :

استقلال القياس عن قدرة العينة التي تؤدي الاختبار ولغرض تحقيق ذلك :

١. جزئت عينة التحليل الإحصائي للاختبارات وتم تقسيم العينة الكلية والبالغة (٦٠٤) فرداً إلى عينة مرتفعة المستوى (فوق الوسيط) ، وإلى عينة منخفضة المستوى (تحت الوسيط) ، وبالأعتماد على ملف الدرجات .
٢. وتم تحليل النتائج بالأعتماد على إستجابات أفراد كل عينة عن الاختبارات ، لغرض حساب صعوبة الفقرات وأخطائها المعيارية وتقديرات القدرة وأخطائها المعيارية.

ولغرض التحقق من التكافؤ الإحصائي تم إجراء مقارنة لمعلمي (الصعوبة ، القدرة) لاختبار مادة تعليم التفكير المحكي ، وتم اشتقاق النتائج من خلال تحليل العينة الكلية وتحليل العينتين (مرتفعة المستوى ، منخفضة المستوى). وتعد تقديرات الصعوبة متكافئة إحصائياً فلم يتجاوز الفرق بين أي من التقديرين مجموع الخطأ المعياري لكل منهما. وتبين أن جميع فقرات الاختبار كانت مكافئة للتقديرات الإحصائية المناظرة لها، وأيضاً كانت جميع الفروق أقل من مجموع الخطأ المعياري للتقديرين وهذا يعني تكافؤ التقديرات المتناظرة في تحليل العينة الكلية من خلال وصفها (تقديرات مرجعية) عدا الفقرات (٢٢ ، ٤٦ ، ٥٥ ، ٦٧ ، ٧٣) فلم تكن تقديراتها الإحصائية متكافئة ، وتلك المشتقة من أداء العينتين سواء كانت المرتفعة او المنخفضة المستوى، وهذا يدل على (تحرر صعوبة الفقرات من قدرة العينة التي طبقت عليها الاختبارات). وتم استخراج تقديرات القدرة المقابلة لكل درجة كلية محتملة والمشتقة من أداء العينة الكلية والمرتفعة المستوى والمنخفضة المستوى ، وأخطائهما المعيارية، وكما موضح في الجدول.

ثامناً : تغيير صفر التدرج بوحدات اللوجيت :

إن صفر التدرج يتغير بعد حذف الفقرات غير المطابقة يختلف عنه قبل حذفها . أي تحدث إزاحة لهذا الصفر تؤثر في تدرج صعوبة الفقرات وفي تقديرات قدرات المفحوصين، وبالطبع فهذا لا يعني الاختلاف في دلالتها الكمية ، إنما حدوث إزاحة لتدرج الفقرات وتدرج قدرات الأفراد (كاظم ، ١٩٨٨ : ١٠٠-١٠١) يؤثر حذف الفقرات غير المطابقة للنموذج الثنائي في متوسط صعوبة فقرات الاختبار ، ولما كان متوسط صعوبة الفقرات في البرنامج أو أي تحليل للفقرات هو صفر التدرج ، لذلك أعيد تحليل بيانات الاختبار مرة أخرى باستخدام البرنامج المحوسب بعد حذف الفقرات غير المطابقة للنموذج ، وحُدد صفر جديد للتدرج هو متوسط صعوبات الفقرات المتبقية في الاختبار البالغ عددها (٧٥) فقرة .

وارتأت الباحثة الاعتماد على وحدة الواط (WAT)، الذي يمثل التدرج المنوي الذي قدمه ماسترز (Masters, 1984) لكونه أكثر التدرجات ألفة في أغلب مجالات القياس . ويمكن تحويل تقديرات صعوبة الفقرات وقدرات المفحوصين من وحدة اللوجيت إلى وحدة الواط

الجدول (٤)

تقدير صعوبة الفقرات مقدرة بوحدة اللوجيت ووحدة الواط للاختبار بصورته النهائية

ت الفقرة	معلم صعوبة الفقرة		ت الفقرة	معلم صعوبة الفقرة		الخطأ المعياري	
	وحدة اللوجيت	وحدة ألواط		وحدة اللوجيت	وحدة ألواط	وحدة اللوجيت	وحدة ألواط

Wat	Logit	Wat	Logit		Wat	Logit	Wat	Logit	
٢	٠,١٨٤	٥٤	٠,٦٦٤	٣٩	١	٠,١٢٦	٤١	٠,٢٦٤	1
١	٠,١٣٨	٤٩	٠,٠٨٤-	٤٠	١	٠,١٢٨	٤٦	٠,٣٨٣-	2
١	٠,١٢٨	٤٦	٠,٤٠٠-	٤١	٢	٠,١٩٦	٦٠	٠,٩٥٠	3
١	٠,١٢٦	٤٥	٠,٤٨٠-	٤٢	٢	٠,١٤٥	٥٩	٠,٨٧٩	4
١	٠,١٢٦	٤٥	٠,٤٦٤-	٤٣	٢	٠,٢١٣	٦٣	١,١٥٩	5
١	٠,١٣٣	٤٨	٠,٢١٢-	٤٤	٢	٠,١٨٣	٥٨	٠,٧٧٢	6
١	٠,١٣٨	٤٩	٠,٠٨٤-	٤٥	٢	٠,١٨٥	٥٩	٠,٨٠٥	7
١	٠,١٣٤	٤٨	٠,١٩٥-	٤٦	٢	٠,١٩٦	٦٤	١,٢٣٤	8
١	٠,١٣٥	٤٨	٠,١٧٧-	٤٧	٢	٠,١٨٥	٥٩	٠,٨٠٥	9
١	٠,١٢٩	٤٦	٠,٣٥٠-	٤٨	١	٠,١١٥	٣٧	١,٢٠٥-	10
١	٠,١٣٣	٤٨	٠,٢١٢-	٤٩	١	٠,١١٦	٣٧	١,١٦٥-	11
١	٠,١٣١	٤٧	٠,٢٨٢-	٥٠	٢	٠,١٩٠	٥٩	٠,٨٧٦	12
١	٠,١٣٦	٤٨	٠,١٤٠-	٥١	١	٠,١٢٢	٤٤	٠,٢١١	13
١	٠,١٣٩	٤٩	٠,٠٦٥-	٥٢	٢	٠,١٩٦	٦٠	٠,٩٥٠	14
١	٠,١٣٦	٤٨	٠,١٤٠-	٥٣	١	٠,١٣٤	٦٤	١,٢٦٧	15
١	٠,١٤٩	٥٣	٠,٩٤١	٥٤	١	٠,١١٦	٣٧	١,١٧٨-	16
١	٠,١٣٤	٥٦	٠,٧٥٤	٥٥	١	٠,١٢١	٤٢	٠,٧٢٢-	17
١	٠,١٣٣	٤٨	٠,٢١٢-	٥٦	١	٠,١٢٩	٤٦	٠,٣٥٠-	18
١	٠,١٢٥	٣٧	٠,٣٤١	٥٧	١	٠,١٢٤	٤٤	٠,٥٧٣-	19
١	٠,١٣٧	٤٩	٠,١٠٣-	٥٨	١	٠,١٢٩	٤٦	٠,٣٥٠-	20
٢	٠,٢٦٤	٣٤	٠,٢٥٤-	٥٩	١	٠,١٣٦	٤٨	٠,١٤٠-	21
١	٠,١٣٦	٤٨	٠,١٤٠-	٦٠	١	٠,١٢٨	٤٦	٠,٤٠٠-	22
١	٠,١٣٤	٤٨	٠,١٩٥-	٦١	١	٠,١٢٤	٣٢	٠,٢٣١-	23
١	٠,١٣٣	٤٨	٠,٢٣٠-	٦٢	١	٠,١٢٦	٤٥	٠,٤٨٠-	24
٢	٠,١٥٧	٥٥	٠,٦٨٧	٦٣	١	٠,١٤٣	٥٢	١,٠٢٥	25
٢	٠,١٤٥	٥٢	٠,٥٢٦	٦٤	١	٠,١٣٨	٤٩	٠,٠٨٤-	26
١	٠,١٣٣	٤٨	٠,٢٣٠-	٦٥	١	٠,١٢٨	٤٦	٠,٤٠٠-	27
٢	٠,١٤٢	٥٠	٠,٠٣٤	٦٦	١	٠,١٢٦	٤٥	٠,٤٨٠-	28
٢	٠,١٤١	٥٠	٠,٠٠٦-	٦٧	١	٠,١٢٦	٤٥	٠,٤٦٤-	29
١	٠,١٣٢	٤٧	٠,٢٤٨-	٦٨	١	٠,١٣٣	٤٨	٠,٢١٢-	30
١	٠,١٢٨	٤٦	٠,٣٨٣-	٦٩	١	٠,١٣٨	٤٩	٠,٠٨٤-	31
١	٠,١٢٧	٤٥	٠,٤٣٢-	٧٠	١	٠,١٣٤	٤٨	٠,١٩٥-	32
١	٠,١٢٨	٤٦	٠,٣٨٣-	٧١	١	٠,١٣٥	٤٨	٠,١٧٧-	33
٢	٠,١٤٣	٥١	٠,٠٥٤	٧٢	١	٠,١٢٩	٤٦	٠,٣٥٠-	34
١	٠,١٢٩	٤٦	٠,٣٥٠-	٧٣	١	٠,١٣٣	٤٨	٠,٢١٢-	٣٥
٢	٠,١٥٢	٥٩	٠,٩٩٧	٧٤	١	٠,١٣١	٤٧	٠,٢٨٢-	٣٦
٢	٠,٢٥٠	٦٧	١,٥٣٠	٧٥	١	٠,١٣٦	٤٨	٠,١٤٠-	٣٧
					١	٠,١٣٩	٤٩	- ٠,٠٦٥	٣٨

الخصائص القياسية لاختبار التشخيصي المحكي لمادة تعليم التفكير على وفق نظرية الاستجابة للفقرة :
اولاً : صدق الاختبار:

وقد تحققت الباحثة من الصدق على وفق نظرية الاستجابة للفقرة للاختبار التشخيصي المحكي لمادة تعليم التفكير وعلى النحو الآتي :

- **الصدق الوصفي:** وقد تحقق هذا النوع من الصدق من خلال عرض الاختبار التشخيصي المحكي لمادة تعليم التفكير على مجموعة من الخبراء بغية ابداء ارائهم فيها، وقد أعتمدت نسبة (١٠٠%) من آراء الخبراء وأجريت التعديلات التي طلبها الخبراء على فقرات الاختبار.
- **الصدق الوظيفي:** ولغرض تحقيق الصدق الوظيفي استعمل الباحث مؤشرات ملائمة الفقرات للأنموذج المستعمل، حيث تم استعمال احصائيات قيم مربع كاي وهو الاسلوب المستعمل في برنامج المحسوب، وهو البرنامج الذي استعملته الباحثة في تحليل بياناته، ووفقاً لنتائج هذا المؤشرات الاحصائية تم الحكم على مدى ملائمة الفقرات او عدم الملائمة لحذفها او الابقاء عليها.

ثانياً: ثبات الاختبار:

إن أدوات القياس المتضمنة فقرات من نوعية جيدة يمكن أن تكون أكثر ثباتاً هذا ما تؤكد نظرية الاستجابة للفقرة (علام ، ٢٠٠٥ : ٥٦). وتوجد عدة مؤشرات لتحديد ثبات الاختبار تبعا لنظرية الاستجابة للفقرة، وعليه ستعتمد الباحثة على مؤشر نسبة التباين لتقدير ثبات الاختبار:

الجدول (٥)

قيم معامل الثبات لاختبار مادة تعليم التفكير وفقا لمؤشر نسبة التباين

الاختبار	مصدر التباين	الانحراف المعياري للتقدير	تباين تقدير القدرة 2_{TS}	الانحراف المعياري للخطأ	تباين الخطأ للتقدير $2_{E\sigma}$	معامل الثبات (r)
التشخيصي المحكي لمادة تعليم التفكير	بين الأفراد	٠,٩٧٨	٠,٩٤١	٠,٠٣٤	٠,٠٠١	٠,٨٩٨

● **تحديد درجة القطع :**

ولتحديد درجة القطع للاختبار التشخيصي المحكي لمادة تعليم التفكير أستعملت الباحثة طريقة المجموعات المتضادة وبذلك تم الحصول على درجة قطع الاختبار التشخيصي المحكي لمادة تعليم التفكير من تقاطع المنحنيات فقد بلغت درجة القطع (٥١) وهي تمثل نسبة (٦٨,٥%).

الوسائل الإحصائية :

اولاً : استعملت الباحثة الحقيبة الإحصائية للعلوم الاجتماعية SPSS لاستخراج :

١. لاستخراج التحليل العاملي بطريقة المكونات الاساسية، (Principles Component) مع اعادة التحليل بطريقة الفاريماكس (Varimax) للتحقق من احادية البعد .
٢. معامل ارتباط (بوينت بايسيريال) لحساب ارتباط درجة الفقرة بالدرجة الكلية للاختبارات.
٣. معادلة الخطأ المعياري لتقدير الخطأ المعياري للمقياس الذي بني وفقاً لنظرية القياس التقليدية وأنموذج نظرية السمات الكامنة بالاعتماد على معاملي الثبات المحسوب وفقاً لأسلوبين.
٤. اختبار مربع كاي Chi-Square لحسن المطابقة في حساب مطابقة فقرات لانموذج ثنائي المعلم (PLM2) .

٥. الاختبار التائي لعينة واحدة One Sample t.test لمعرفة الفرق بين متوسط درجات أفراد العينة ودرجة القطع الاختبار التشخيصي محكي المرجع لمادة تعليم التفكير لطلبة الصف الرابع.
٨. قانون مربع أيتا لتحديد حجم الأثر الخاص بأتقان طلبة الصف الرابع لمادة تعليم التفكير.
- ثانياً : برنامج لغة الأوامر لأستجابة الفقرة (ICL) ، (Item response command language) لتحليل بيانات البحث وفق نظرية الاستجابة للفقرة للأنموذج ثنائي المعلم (PLM2)

نتائج البحث وتفسيرها ومناقشتها :

- لتحقيق الهدف الأول الذي ينص على (بناء اختبار تشخيصي محكي المرجع لمادة تعليم التفكير لطلبة الصف الرابع لقسم العلوم التربوية والنفسية في كليات التربية في الجامعات العراقية) ولتحقيق الهدف الأول قامت الباحثة بجميع الإجراءات والخطوات العملية اللازمة لبناء الاختبارات وفقاً لنظرية الاستجابة للفقرة، وكما موضح في الفصل الثالث. بدءاً من تحديد المحتوى الدراسي وتحديد الأهداف السلوكية، ومن ثم إجراءات التحليلات المنطقية والإحصائية للتحقق من افتراضات نظرية الاستجابة للفقرة، وتحديد الخصائص القياسية لل فقرات والاختبار الكلي، وقد أعتمدت برنامج لغة الأوامر لأستجابة الفقرة (ICL) ، (Item response command language) لتحليل بيانات البحث وفق نظرية الاستجابة للفقرة للأنموذج ثنائي المعلم (PLM2)، وقد تبين ان جميع الفقرات والاختبار الكلي يتمتع بخصائص قياسية جيدة، وتم الاستبقاء على جميع الفقرات، وبذلك تم تحقيق هذا الهدف.

الاستنتاجات :

١. مناسبة الأنموذج الثنائي المعلم PLM2 في بناء الاختبار التشخيصي المحكي المرجع لمادة تعليم التفكير أداة للبحث الحالي، وذلك من خلال مطابقة بيانات الاختبارات لافتراضات الأنموذج.
- ٢.فاعلية البرنامج الإحصائي برنامج لغة الأوامر لأستجابة الفقرة (ICL) ، (Item response command language) في إجراء التحليلات الإحصائية لبيانات الاختبار، وذلك لتقدير مطابقة هذه البيانات لافتراضات الأنموذج كما ورد في النقطة السابقة، فضلاً عن استقراره في معايرة وتدرج الفقرات وقدرات الأفراد على متصل واحد للسمة الكامنة، وتحديد الخصائص القياسية للفقرات والمقياس الكلي، وذلك من خلال ما تقدم من مخرجات نصية وبيانية لهذه التحليلات.
- ٣.مناسبة الاختبار التشخيصي محكي المرجع لمادة تعليم التفكير لطلبة الصف الرابع في الجامعات العراقية الحكومية وذلك لاتساق قدرات أفراد العينة وموقع الفقرات في تحقيق هدف الاختبار، ومن خلال تقارب نقطتي الأصل لمتصل السمة او قدرة الأفراد، والمعروف كذلك بتوزيع تقديرات قدرة المجتمع (θ). وان اقتراب التوزيعين من التوزيع الاعتدالي يشير الى تمثيل عينة البحث لمجتمع البحث، باستثناء استبعاد بعض الأفراد غير المطابقين.
٤. يستبعد الطلبة ذوي الاستجابات غير مطابقة لآنموذج التحليل، يحقق عدة أغراض منها: تحقيق مطابقة أفضل لبيانات الاختبار مع توقعات الأنموذج، والتوصل الى تقديرات أفضل لمعالم الفقرات والأفراد، وبالتالي تحقق عينة البحث تمثيل أفضل لمجتمع البحث.

التوصيات:

١. استعمال الاختبار التشخيصي محكي المرجع لمادة تعليم التفكير لطلبة الصف الرابع في قسم العلوم التربوية والنفسية كأداة تحصيلهم الدراسي وتشخيص مشاكلهم الدراسية، وعدم الاقتصار على العينة الحالية.
٢. استعمال البرنامج الإحصائي البرنامج الإحصائي برنامج لغة الأوامر لأستجابة الفقرة (ICL) ، (Item response command language) في إجراء التحليلات الإحصائية لبيانات الاختبارات والمقاييس.

المقترحات:

١. إضافة متغير آخر يتعلق بخصائص العينة بدراسات أخرى مثل (خصائص توزيع العينة، او حجم العينة). وباستعمال التصميم القياسات المتكررة المختلط. او أي تصميم عاملي، وبحسب مستويات المتغيرين.

٢. إضافة متغير آخر يتعلق بخصائص الفقرات بدراسات أخرى مثل (ترتيب البدائل، او عدد بدائل الاستجابة، او نوع البدائل) وباستعمال أي تصميم عاملي، وبحسب مستويات المتغيرين.

المصادر:

١. أبو ناهية، صلاح الدين محمد (١٩٩٤). القياس التربوي. مكتب الانجلو المصرية، القاهرة، مصر.
٢. اسماعيل وآخرون، محمد (١٩٩٤). معايير نمو طفل ما قبل المدرسة. الدراسات النفسية، المجلد الثاني، المجلس القومي للطفولة والأمومة، القاهرة، مصر.
٣. التميمي، خالد بن حسن شيبان (١٩٩٩). اثر كل من نوع المحكم وطول الاختبار على تحديد درجة القطع لاختبار محكي المرجع يقيس الكفايات الرياضية في العمليات الحسابية على الاعداد بالصف السادس ابتدائي بمدينة جدة. رسالة ماجستير، جامعة ام القرى، مكة المكرمة، المملكة العربية السعودية.
٤. الجلي، سوسن (٢٠٠٥). اساسيات بناء الاختبارات والمقاييس النفسية والتربوية. علاء الدين للطباعة والنشر، دمشق، سوريا.
٥. دبوس، محمد طالب (٢٠١٦). استخدام نظرية الاستجابة للفقرة في بناء تجمع فقرات اختبار محكي المرجع في الرياضيات وفق النموذج اللوجستي ثنائي المعلم. مجلة جامعة النجاح للعلوم الانسانية، مجلد ٣٠، العدد ٧.
٦. السامرائي والخفاجي، محمد أنور محمود وأحمد محمد شاكر (٢٠١٢). بناء اختبار تحصيلي محكي المرجع في مادة علم النفس الخواص لطلبة أقسام العلوم التربوية والنفسية. العدد (٢٠٣).
٧. سرقوبة، عبد السلام عوض (٢٠١٦). بناء اختبار تشخيصي محكي المرجع في موضوعات من مقرر القياس والتقويم في جامعة عمر المختار. كلية التربية البيضاء، المجلة الليبية العالمية، ليبيا.
٨. السلامي، عبد الله بن محمد (٢٠١٨). بناء اختبار محكي المرجع في الرياضيات باستخدام نظرية الاستجابة للمفردة وفق النموذج اللوجستي ثنائي المعلم لطلاب الصف الخامس. مركز الارشاد النفسي والتربوي، كلية التربية، أسيوط.
٩. عيابة، عماد غضاب (٢٠٠٩). الاختبارات محكية المرجع فلسفتها وأسس تطويرها. دار المسيرة للنشر، عمان، الأردن.
١٠. عبد الحافظ، شحته عبد المولى (١٩٩٩). تقويم بناء الاختبارات المرجعة إلى محك، معيار في نظرية الاستجابة للمفردة والنظرية التقليدية. أطروحة دكتوراه، كلية التربية، جامعة عين شمس، مصر.
١١. عبد الخالق، أحمد محمد (١٩٨٣). الأبعاد الأساسية للشخصية، ط ٦، الإسكندرية، دار المعرفة الجامعية.
١٢. علام، صلاح الدين محمود (١٩٨٥). الاختبارات التشخيصية مرجعية المحك في المجالات التربوية والنفسية والتدريبية. دار الفكر العربي، القاهرة، مصر.
١٣. _____ (١٩٨٦). تطورات معاصرة في القياس النفسي والتربوي. مطابع القبس التجارية، الكويت.
١٤. _____ (٢٠٠١). الاختبارات التشخيصية مرجعية المحك في المجالات التربوية والنفسية والتدريبية. دار الفكر العربي للطبع والنشر، القاهرة، مصر.
١٥. _____ (٢٠٠٥). نماذج الاستجابة للمفردة الاختبارية احادية البعد ومتعددة الابعاد وتطبيقاتها في القياس النفسي والتربوي. القاهرة، دار الفكر العربي.
١٦. عيد، غادة خالد (٢٠٠٤). الدرجة الحقيقية باستخدام نظرية السمات الكامنة والنظرية الكلاسيكية - دراسة سيكومترية. جامعة ام القرى، العدد ٢، المجلد ١٦.
١٧. غنيم، محمد أحمد إبراهيم (٢٠٠٣). الاتجاهات الحديثة في بحوث مشكلات تقويم التحصيل الدراسي. بحث مقدم إلى اللجنة العلمية الدائمة لعلم النفس التربوي والصحة النفسية " مستوى الأساتذة "، جامعة الزقازيق، مصر.
١٨. كاظم، امينة محمد (١٩٨٨). دراسة نظرية نقدية حول القياس النفسي للسلوك وفق أنموذج راش، في أنور الشراوي وآخرون، اتجاهات معاصرة في القياس والتقويم النفسي والتربوي. القاهرة، مكتبة الأنجلو المصرية.

١٩. مجيد، سوسن شاكر (٢٠١٤). **اسس بناء الاختبارات والمقاييس النفسية والتربوية**. الطبعة الثالثة، مركز ديبونو، عمان، الاردن.
٢٠. منسي، محمود عبد الحكيم (٢٠٠١). **التقويم التربوي**. دار المعرفة الجامعية للطبع والنشر والتوزيع.
21. Adedoyin, O. O. (2010). **Investigating the invariance of person parameter estimates based on classical test and item response theories**. International Journal of Educational Sciences, 2(2), 107-113.
22. Aiken, L. R. (1998). **Tests and examinations: Measuring abilities and performance**. Journal of personality and Social psychology.
23. Hambleton & Swaminathan ,H.(.(1989). **Item Response Theory Principles And Applications** .Kluwer-Nijh of publishing , Boston ,U.S.A.
24. _____, Swaminathan, H., & Rogers, H. J. (1991). **Fundamentals of item response theory** (Vol. 2). Sage.
25. McIntire , S . A . & Miller ,L . A . (2000) .**Foundations Of Psychological Testing** . New York ,McGraw-Hill Companies.
26. Onn, D. (2013, May). **Classical test theory versus item response theory: An evaluation of the comparability of item analysis results**. Joint Admissions and Matriculation Board, 1-23.
27. Popham, W. J., & Husek, T. R. (1969). **IMPLICATIONS OF CRITERION-REFERENCED MEASUREMENT** 1, 2. Journal of Educational Measurement, 6(1), 1-9.
28. Rust , J & Golombok ,K ,S (1999) .**Modern Psychometrics The Science Of Psychological Assessment** .British Library Cataloging Inpublication Data.
29. Weiss , D . J . & Yoes , M.E .(1994) . Item Response Theory, IN: H ambleton, R.k.& Zaal , J.N.(Eds): **Advances In Educational and psychological Testing : Theory And Applications** .,Kluwer Academic Publishers, Boston.

Abstract:

The present study aims at using dual criterion model in constructing a diagnostic criterion- Referenced test in the course of teaching thinking for the students at the Department of Educational and psychological Sciences. To achieve this aim, a test of (80) MCQ items with four alternatives has been constructed. The sample of the present study consists of (604) male and female students distributed on the universities of : (Baghdad, Al-Mustansiriya, Al-Iraqiya, Al-Kufa, Wasit, Karbala, Al-Qadisiyah). from fourth-grade students in the colleges of education, and for the morning study, for the academic year (2021/ 2022), The sample was selected by the stratified random method with a proportional distribution. The items have been statistically analyzed by using the (ICL) program in order to grade the items based two-parameter model. The researcher verified the hypothesis of one-dimensionality by using factor analysis by the principal components Kaizer method, and computing the relationship between the item scone and the test total score. In addition, the assumption of freedom from speed factor as well as the decrease of the Guessing factor have been verified through assuring the individuals performance who were of lower ability on the difficult items. As a result, the items have been matched to the assumptions above. To identify the cutting score of the test, the opposite groups method has been used, where the score has been found out to be (51) representing (68.5 %). The assumptions of this model have been verified by the significance of matching all the test items with the model itself except five items which were not matched. The final version of the test consists (75) items. Result of the data analysis show that all the items and the whole lest have good psychometric characteristics. Based on the results, the researcher reached a set of conclusions, recommendations and suggestions.