185
المجلد 10 العدد 2 يونيو 2009

استراتيجية مقترحة لبناء اختبارات تشخيصية محكية المرجع

د. هند عبد المجيد الحموري
قسم علم النفس التربوي
الجامعة الهاشمية-الأردن

د. أحمد إسماعيل الكحلوت
عمادة الدراسات التربوية
الجامعة العربية المفتوحة-الكويت

استراتيجية مقترحة لبناء اختبارات تشخيصية محكية المرجع

د. هند عبد المجيد الحموري
قسم علم النفس التربوي
الجامعة الهاشمية-الأردن

د. أحمد إسماعيل الكحلوت
عمادة الدراسات التربوية
الجامعة العربية المفتوحة-الكويت

الملخص

هدفت هذه الدراسة إلى تقديم استراتيجية لبناء اختبارات تشخيصية محكية المرجع، تستند إلى منحى المواصفات التفصيلية، والشاملة في وصف المجال السلوكي. وتتضمن ثلاث خطوات رئيسة: الأولى، إعداد مواصفات تفصيلية شاملة للاختبار من حيث: تحديد الغرض من الاختبار، ومجال محتواه، والمجموعة التي سيطبق عليها، وفريق العمل؛ وتحليل الموضوع المستهدف تحليلاً تفصيلياً، وتنظيم عناصر المعرفة المرتبطة بالموضوع، والتحقق من صدق التحليل، وإعداد مواصفات تفصيلية للفقرة. والخطوة الثانية، هي بناء الفقرات الاختبارية باستخدام بناء "الفقرة النموذج". وكانت الخطوة الثالثة مراجعة الفقرات الاختبارية، وتصديقها.

الكلمات المفتاحية: اختبار تشخيصي، اختبار محكي المرجع، استراتيجية، مواصفات الفقرة، الفقرة النموذج

* تاريخ تسلم البحث: ٤/١٠/٢٠٠٧م
* تاريخ قبوله للنشر: ١٢/٥/٢٠٠٨م

A Proposed Strategy for Constructing Criterion-Referenced Diagnostic Test

Dr. Hind Alhammouri
Dept. of Educational Psychology
The Hashemite University

Dr. Ahmed Kahlout
Educational Studies Deanship
Open Arab University- Kuwait

Abstract

The aim of this study was to introduce a strategy for constructing criterion-referenced diagnostic tests based on the Comprehensive and Detailed Specifications Approach for describing the behavioral domain. The strategy consisted of three main steps. First, describing the behavioral domain in a detailed and comprehensive manner. This step includes identifying the aim and context the test, in addition to the working group and the targeted population. Also, it includes a thorough analysis the domain to be tested and hierarchic organization of its components, proving the validity of the hierarchy, and preparing detailed item specifications. Second, constructing items using the "model item". Finally, reviewing, revising, and validating the items.

Key words: diagnostic test, criterion-referenced test, strategy, item specification, model item.

188
المجلد 10 العدد 2 يونيو 2009

استراتيجية مقترحة لبناء اختبارات تشخيصية محكية المرجع

د. هند عبد المجيد الحموري
قسم علم النفس التربوي
الجامعة الهاشمية-الأردن

د. أحمد إسماعيل الكحلوت
عمادة الدراسات التربوية
الجامعة العربية المفتوحة-الكويت

مقدمة

يعاني ميدان التربية، في العالم بعامة، من مشكلات حاسمة تتعلق بتعلم الطلبة. ومن الملاحظ أن معظم ما يجري في مدارس العالم اليوم، تحت اسم التقويم التربوي، لا يجدي في التعلم والتعليم. فمعظم الاختبارات التي تطبق في المدارس لا تلبي حاجات المتعلمين، إذ قلما يحصل الطلبة على تغذية راجعة، تكون سريعة وسليمة (الصيداوي، ٢٠٠٤).

والأردن ليس أحسن حالاً من غيره من الدول، إذ يشير الأدب التربوي، الذي تناول واقع أداء الطلبة فيه، إلى أن هناك ثمة ضعفاً واضحاً في أدائهم في الرياضيات. ولم يقتصر هذا الضعف على أصناف المعرفة الرياضية المتقدمة من نظريات ومسائل، وإنما تعدى ذلك إلى ضعف في المهارات الرئيسة، المتعلقة بمفاهيم الأعداد، والكسور، والعمليات الأربع (أبو لبدة، ٢٠٠٣؛ الرواشدة وزملاؤه، ٢٠٠٠؛ عنابي، ١٩٩٩). وقد كشفت الدراسة الدولية الثالثة للرياضيات والعلوم، التي عقدت في العام ٢٠٠٣ عن وجود ضعف واضح في مادة الرياضيات عند الطالب الأردني بخاصة، والعربي بعامة. فقد تبين أن متوسط أداء الطلبة الأردنيين عام ٢٠٠٣ لم يختلف اختلافاً دالاً إحصائياً عما كان عليه عند تطبيق الدراسة ذاتها عام ١٩٩٩، هذا بالإضافة إلى أن أداءهم أقل من متوسط الأداء الدولي، في العامين اللذين أجريت فيهما الدراسة. وبينت الدراسة أيضاً أن نسبة الطلبة الأردنيين، الذين وصلوا إلى مستوى الأداء المتقدم، لم تتجاوز ١٪، في حين كانت نسبة نظرائهم الذين وصلوا إلى مستوى الأداء المنخفض ٦٠٪ (الرفيع وزملاؤه، ٢٠٠٧).

وفي ضوء هذا الواقع التربوي، ومنذ العام ٢٠٠٣، بدأت وزارة التربية والتعليم في الأردن بالتعاون مع البنك الدولي، والصندوق العربي بعملية إصلاح تربوي، تمثلت في تبنيها مشروع تطوير التعليم نحو اقتصاد المعرفة (Education Reform for Knowledge Economy). ومن العناصر الرئيسة في رسالة هذا المشروع "تطوير، وإدارة نظام تربوي، يركز على التميز والإتقان، ويستثمر موارد بشرية تتمتع بقدر عالٍ من إتقان كفايات التعلم الأساسية"؛ بمعنى أن التعلم الإتقاني أحد المخرجات الرئيسة، التي يسعى هذا المشروع لبلوغه (وزارة التربية والتعليم، ٢٠٠٦).

ويعد التعلم الإتقاني تقدماً تربوياً، وإن لم يكن إصلاحاً تربوياً جذرياً كاملاً، فهو يمثل استراتيجية قابلة للنمو، والتطور، والتكامل مع استراتيجيات التعلم الأخرى (الصيداوي، ٢٠٠٤). وعليه، فقد أضحى التقويم التشخيصي عماد التعلم الإتقاني، إذ إنه الدعامة الأساسية، التي لا بد منها للارتقاء بالمتعلمين إلى درجة الإتقان المطلوبة على الأقل.

ويتطلب التقويم التشخيصي جملة من الأساليب، والأدوات التي تتنوع ما بين اللااختبارية والاختبارية. وتشمل الأولى ملاحظة أداء الطالب، والتقارير، والعروض، والمداخلات التي يقدمها، والسجل الصحي، وسجل التوجيه التراكمي، وإجراء دراسة الحالة، ومناقشة الحالات. في حين تشتمل الثانية على الامتحانات والاختبارات موضوعية التصحيح، ومنها الاختبارات التشخيصية والتحليلية. وعلى الرغم من أن الأساليب اللااختبارية توفر قدراً من المعلومات المهمة المتعلقة بأداء الطالب، إلا أنها قد تفتقد الدقة والشمول؛ لأنها تتطلب وقتاً صويلاً في التطبيق، وتحليل النتائج النوعية المتحصلة من تحليل أداء كل طالب؛ ومن ثمَّ، يصعب تطبيقها على عدد كبير من الطلبة. ومن هنا، فقد جاءت الأساليب الاختبارية التي تتيح فرصة لتشخيص مواطن قوة الطلبة وضعفهم، وربما اكتشاف بعض أسبابها في وقت قصير نسبياً، شريطة أن تكون موثوقة، محكمة البناء. كما أنها تمتاز عن غيرها من أساليب التشخيص الأخرى، بأنها مرتبطة بميادين ومهارات معينة، ولذلك، تمهد لعملية اتخاذ التدابير الوقائية، والتصحيحية، أو العلاجية والتنشيطية اللازمة، فهي ألصق بعمل معلم الصف من غيرها من الاختبارات.

تركز الاختبارات التشخيصية على أجزاء معينة من موضوع دراسي معين، لتكشف عن مكامن القوة والضعف في اكتساب المتعلم للقدرات والمهارات التي تمثلها تلك الأجزاء من ذلك الموضوع الدراسي. ويقتضي ذلك تقسيم الموضوع إلى وحدات تعلمية شبه مستقلة، وتحديد العلاقات القائمة بين مختلف الأجزاء التي تؤلف الموضوع المعين، ثم وضع الفقرات الصادقة المناسبة، والكافية بحسب مستويات المعرفة، والسلوك، والأداء المنشود (Hambleton, 1982). فتشخيص أداء الطلبة، وتحديد مواطن ومجالات ضعفهم وقوتها يتطلب بحثاً تفصيلياً عن مصادر الأخطاء، ولا بد من أن يستند إلى أدوات تشخيصية، محكمة البناء، تتطلب مواقف تمكن المستخدم من الوقوف على هذه الأخطاء، ودرجة شيوعها، وأهميتها للتعلمات اللاحقة. وعليه، توفر الاختبارات التشخيصية معلومات أكثر تحديداً عن تعلم الطلبة مما توفره الاختبارات الصفية.

لقد تضمنت عملية الإصلاح التربوي في الأردن إصلاحاً لنظام التقويم. فقد تم إنشاء مديرية للامتحانات، والاختبارات تضم من بين أقسامها قسم "التقويم التشخيصي". ومن الأهداف التي يعمل هذا القسم على بلوغها بناء اختبارات للتشخيص، تساعد على مقارنة أداء الطالب مع أداء طلبة مدرسته ومنطقته، وتعرف درجة امتلاك الطلبة للمهارات الأساسية

(وزارة التربية والتعليم، ٢٠٠٦). بمعنى أنهم يعتمدون التقويم معياري المرجع؛ للتحقق من بلوغ الهدف الأول، والتقويم محكي المرجع للتحقق من بلوغ الهدف الثاني.

ومن المعلوم أن أكثر أنماط التقويم مساهمة في التخطيط وبناء العمليات الخاصة بالتعلم والتعليم الصفي، هو التقويم محكي المرجع؛ لأنه يهتم بقياس، وتشخيص ما حققه المتعلم من معارف ومهارات، وما اكتسبه من اتجاهات، وسلوكات تتعلق بمحتوى دراسي، أو تعليمي معين في ضوء محك أداء مطلق، وبالكشف عن صعوباته التعلمية، ومواطن الضعف لديه (Murphy & Davidshofer, 2001). وتفسر الدرجات على هذا النوع من الاختبارات بناء في الأداء على المهمة المقيسة؛ بمعنى أن تحديد موقع المتعلم بالنسبة لمجال سلوكي محدد تحديداً تاماً يتم بالاعتماد على مجموعة من القواعد الصريحة، التي تسهل عملية تفسير النتائج على الاختبار (Berk, 1982). ومن ثمَّ، تستخدم درجات الاختبار لتصنيف المفحوصين في فئات إتقان متمايزة (Crocker & Algina, 1986)، وذلك من خلال تقسيم مقياس علامة المجال إلى فئات إتقان متمايزة عددها (ك) فئة، تحدد بنقاط قطع عددها (ك−١)؛ والأكثر شيوعاً التقسيم إلى فئتين: متقن وغير متقن؛ الأمر الذي يمكّن من تشخيص مواطن ضعف الفرد، أو قوته في ضوء القرارات المتعلقة بتصنيفه. والاختبارات التي تستخدم لهذا الغرض تسمى اختبارات تشخيصية محكية المرجع.

غير أن المتابع للجهود التي تبذل في مديرية الامتحانات والاختبارات تلك، يلاحظ غياباً واضحاً لأدوات القياس التشخيصية محكية المرجع الدقيقة، التي تساعد على الوقوف على الأسباب الحقيقية للضعف. وفي الحقيقة، فإن استخدام ما يتم بناؤه من اختبارات تشخيصية يستخدم، في أحسن حالاته، لمراقبة تعلم الطلبة بشكل عام؛ إذ لم يتم إجراء تقويم تشخيصي بالمفهوم الوارد في الأدب التربوي الذي تم تبنيه في هذا البحث، حتى الآن.

وعلى بالرغم من أهمية الاختبارات التشخيصية محكية المرجع، فإنه يلاحظ ندرة توافرها في الوطن العربي، إذ يوكل أمر إعدادها، في الغالب، لمدرسين، لا تتوافر لديهم الإمكانيات، والقدرات اللازمة لإعدادها (علام، ٢٠٠١)، أو لطلبة الدراسات العليا كجزء من رسالة الماجستير أو أطروحة الدكتوراه (أبو ريدة، ١٩٩٣؛ أحمد، ١٩٩٥؛ أحمد، ٢٠٠٤؛ الريماوي، ١٩٩٠؛ العابدي، ٢٠٠٣). ومع ذلك، فقد ظهرت محاولات عربية (علام، ٢٠٠١؛ الشريفين، ٢٠٠٦) لبنائها، إلا أن هذه المحاولات ما زالت متواضعة، أو أنها لم تخرج من المنطقة التي أجريت فيها. فغالباً ما تهتم ببناء اختبار تشخيصي لمهمة محددة، وليس بتطوير استراتيجية لبناء هذا النوع من الاختبارات.

وقد تعزى ندرة توافر الاختبارات التشخيصية محكية المرجع إلى ندرة الخبرات العربية في مجال تطوير هذا النوع من الاختبارات، من حيث معرفة: كيفية بنائه، وما الاختبارات، والتحليلات الإحصائية المناسبة لكل خطوة من خطوات بنائه، وكيفية تفسيرها، وكيفية

استخدامه، والقرارات التي يمكن أن تصدر عنه. فعملية بناء الاختبارات التشخيصية محكية المرجع تحتاج تعاون العديد من المؤسسات التربوية من جامعات، ومراكز بحوث، ومراكز القياس والتقويم؛ كما تتطلب تضافر جهود فريق من الاختصاصيين في مختلف العلوم التربوية الذين لديهم مهارات وكفايات عالية، واستراتيجيات محددة واضحة لبنائها؛ وتوافر مراجع وأدلة مناسبة تساعدهم على تصميمها وبنائها (الصيداوي، ٢٠٠٤). وفضلاً عن ذلك، فإنها تتطلب توافر مجال سلوكي واضح يتفق عليه المقيمون (Hambleton & Sireci, 1997; Hambleton & Swaminathan, 1985). هذا، بالإضافة إلى أنّ المؤسسات العالمية المتخصصة في تطوير الاختبارات لا تقدم خبراتها العملية إلا بأسعار مرتفعة. وفضلاً عن ذلك، فهي تتطلب كلفة مادية، وجهداً كبيرين، إذ يتطلب تمثيل المجال السلوكي عدداً كبيراً من الفقرات الاختبارية (Popham, 1982)؛ كما أن تحليلها يحتاج إلى مهارات قلما تتوافر لدى العاملين في بنائها (Murphy & Davidshofer, 2001).

ومن هنا تبدو الحاجة ملحة لطرح استراتيجية محددة لبناء الاختبارات التشخيصية محكية المرجع المحكمة، تصمم في إطار خطة شاملة، ومتكاملة لتقويم أداء الطالب، وذلك ضمن إجراءات الإصلاح المطلوبة بإلحاح في نظام التقويم في الأردن، يمكن أن يتبعها مصممو هذا النوع من الاختبارات؛ كي يتمكن مستخدموها من اتخاذ القرارات الصائبة في ضوء الغايات المقصودة.

هدف الدراسة

في ضوء ما سبق، فقد سعت هذه الدراسة إلى تقديم استراتيجية لبناء اختبارات تشخيصية محكية المرجع، يمكن استخدامها من قبل المراكز المتخصصة التي تتوافر فيها الإمكانات المادية والبشرية اللازمة لبناء هذا النوع من الاختبارات. وعليه، فقد قام الباحثان بتطوير استراتيجية خاصة لبناء اختبارات تشخيصية محكية المرجع، تستند إلى منحى المواصفات التفصيلية والشاملة في وصف المجال السلوكي، تتلخص في اتباع الخطوات الآتية: إعداد مواصفات تفصيلية شاملة للاختبار، وبناء الفقرات الاختيارية، ومراجعة الفقرات الاختيارية وتصديقها.

وفيما يأتي توضيح لكل منها:

أولاً: إعداد مواصفات تفصيلية شاملة للاختبار، وقد تم ذلك وفق الخطوات الآتية:

١) التخطيط للاختبار، الذي تطلب:

أ – تحديد الغرض من الاختبار، وهو تشخيص مواطن ضعف طلبة المرحلة الأساسية وقوتهم، وأسبابها المحتملة.

ب – تحديد مجال المحتوى، وقد اختير موضوع قسمة الكسور، ذلك أن عدة دراسات

أشارت إلى وجود ضعف عند الطلبة فيها (أبو لبدة، ٢٠٠٣؛ الرواشدة وزملاؤه، ٢٠٠٠؛ الرفيع وزملاؤه، ٢٠٠٧؛ الحموري والكحلوت، ١٩٩٩؛ عنابي، ١٩٩٩؛ الكحلوت والحموري، ١٩٩٩).

ج – تحديد المجموعة التي سيطبق عليها الاختبار، وقد اختير الصف الخامس الأساسي ليكون مجتمع الدراسة، لكون منهاجه يغطي موضوع الكسور العادية، والعمليات المتعلقة بها. وبسبب كون بناء الاختبار يتطلب عينات متعددة، كل منها يستخدم لغرض معين، فقد تم تحديد العينات التي ستستخدم في بناء الاختبار وفق ما يأتي:

١. العينة الرئيسة التي سيتم تطبيق الاختبارات عليها. تكونت من ست شعب من الصف الخامس، نصفهم من الذكور، وقد تم اختيار هذه الشعب بالطريقة العشوائية، تضم هذه الشعب ١٧٧ طالباً وطالبة، منهم ٩١ طالباً.

٢. عينة التحقق من صحة الفرضيات المتعلقة بمستويات التحليل في البناء الهرمي (سنأتي على ذكرها لاحقاً)، تكونت من أربع شعب من مجتمع الدراسة، وخارج عينتها الرئيسة، تم اختيارها بالطريقة العشوائية، تضم ١٨٢ طالباً وطالبة.

٣. عينة التحقق من فاعلية البناء الهرمي (سنأتي على ذكرها لاحقاً)، تكونت من شعبتين من مجتمع الدراسة، وخارج عينتها الرئيسة، تم اختيارهما بالطريقة العشوائية، تضمان ٦٥ طالباً وطالبة.

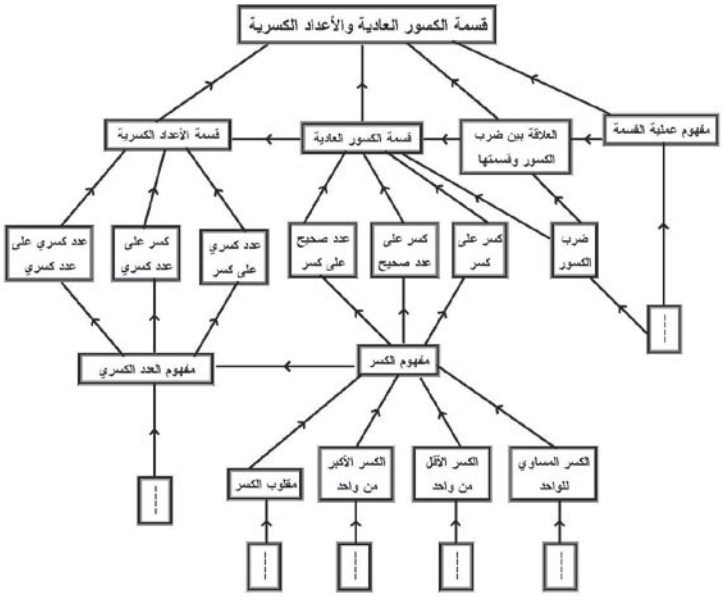
٤. عينة حصر أخطاء الطلبة (سنأتي على ذكرها لاحقاً)، وقد تكونت من:

شعبتين من الصف السادس، وشعبتين من الصف الخامس، تم اختيارهما عشوائياً.

أوراق امتحانات خمسين طالباً من الصف الخامس من الفئات العليا والوسطى والدنيا من أربع شعب من المجتمع، وخارج العينة الرئيسة، نصف هذه الشعب من الإناث. وقد تم تحديد طلبة الفئات الثلاث كما يأتي:

* تم الاختيار، بطريقة عشوائية، لِ (١٥) طالباً من الطلبة الذين متوسط علاماتهم في الرياضيات أكبر من ٩٠٪، على أنهم فئة عليا.
* والذين متوسط علاماتهم أقل من ٤٠٪، على أنهم الفئة الدنيا.
* والذين متوسط علاماتهم أقل من ٧٠٪ وأكبر من ٥٠٪ على أنهم الفئة الوسطى.
* عينة وضوح "الفقرات النموذج" (سنأتي على ذكرها لاحقاً)، تكونت من شعبة من مجتمع الدراسة وخارج عينتها الرئيسة، تم اختيارها بالطريقة العشوائية، تضم ٢٧ طالباً.
* عينة دراسة تشابه استجابات مجموعة الطلبة على كل أزواج "الفقرات النموذج" (سنأتي على ذكرها لاحقاً)، تكونت من شعبتين من مجتمع الدراسة، وخارج عينتها الرئيسة، تم اختيارهما بالطريقة العشوائية، تضمان ٥٠ طالباً وطالبة، منهم ٢٧ من الذكور.
* عينة تحديد زمن الاختبار، وتكونت من شعبتين من المجتمع، وخارج عينة الدراسة، تم

اختيارهما بطريقة عشوائية، تضمان ٤٥ طالباً وطالبة.
د- تحديد فريق العمل، وقد تم اختيار ثلاثة من المختصين في الرياضيات واثنين من المختصين في بناء الاختبارات، وثلاثة من مدرسي الرياضيات الأكفَاء للصف الخامس، واثنين من المشرفين التربويين لمادة الرياضيات.
٢) تحليل موضوع قسمة الكسور وتنظيمه، وقد تضمن الخطوتين الآتيتين:
أ. تحليل موضوع قسمة الكسور تحليلاً تفصيلياً
ب. تنظيم عناصر المعرفة المتعلقة بموضوع قسمة الكسور تنظيماً هرمياً
وفيما يأتي توضيح لهاتين الخطوتين.
١. تحليل موضوع قسمة الكسور تحليلاً تفصيلياً، وتحديد عناصر المعرفة المتعلقة به. إذ قام الباحثان بتحليل موضوع قسمة الكسور العادية الوارد في كتيب "منهاج الرياضيات" للصفوف من الأول إلى السادس، والكتب المدرسية، وأدلة المعلم لتلك الصفوف، تحليلاً متعمقاً تفصيلياً، وشاملاً، يبين جميع عناصر المعرفة الرياضية المتعلقة بهذا الموضوع من مفاهيم، وخوارزميات، وتعميمات، ومسائل.
٢. تنظيم عناصر المعرفة المتعلقة بموضوع قسمة الكسور تنظيماً هرمياً، وقد تم وفق ما يأتي:
- تحديد المتطلبات السابقة من عناصر المعرفة الرياضية (المهارات الفرعية) اللازم توافرها لدى المتعلم كي يحقق مهارة "قسمة الكسور العادية".
- ترتيب المهارات الفرعية ترتيباً منطقياً، بحيث يكون إتقان المتعلم لواحدة، أو مجموعة منها متطلباً أساساً لإتقان مهارات أخرى أعلى مرتبة. ومثال على ذلك، عملية قسمة الكسور العادية (وهي في مستوى عالٍ من الهرم) تتطلب أن يكون الطالب متقناً لما يأتي: قسمة كسر على عدد صحيح، وقسمة عدد صحيح على كسر، وقسمة كسر عادي على كسر عادي، ومعرفة العلاقة بين عمليتي ضرب الكسور وقسمتها، وجميعها في نفس المستوى، ويسبقها في الهرمية المهارات الآتية: مفهوم عملية قسمة الكسور، ومفهوم ضربها، وضرب الكسور العادية؛ وهكذا.
- وضع هذا التحليل على شكل شجرة، توضح البنية الهرمية للموضوع. وبما أن الشكل الناتج مركب، ومعقد جداً فقد تم الاكتفاء بجزء من البنية الهرمية، التي سيتم وفقها توضيح استراتيجية بناء اختبارات تشخيصية محكية المرجع، كما في الشكل رقم (١).



الشكل رقم (١)
البنية الهرمية لموضوع قسمة الكسور

٣) التحقق من صدق التحليل، وصدق البناء الهرمي، وقد تضمن الخطوات الآتية:
أولاً: التحقق من صدق البناء الهرمي لعناصر الموضوع.
ثانياً: التحقق من صحة الفرضيات المتعلقة بمستويات التحليل في البناء الهرمي.
ثالثاً: التحقق من فاعلية البناء الهرمي.
رابعاً: التحليل الدقيق للمهارات الجزئية.
خامساً: تحديد مواصفات الفقرات في المجال السلوكي.
وفيما يأتي توضيح لكل منها:
أولاً: التحقق من صدق البناء الهرمي لعناصر الموضوع، وقد تطلب ذلك:
– عقد عدة لقاءات مع فريق العمل بهدف تحقيق فهم مشترك للمفاهيم، والمصطلحات المتعلقة بعملية الصدق، من مثل: مواصفات الفقرة، ومواصفات المثير والاستجابة، ودرجة ارتباط الفقرة بمواصفاتها والعكس، وصدق المحتوى، وطريقة كتابة الفقرات، ذات الاختيار من عدة خيارات؛ ذلك أن طرق استنباط الأحكام وإجراءات صدق المحتوى عملية مفصلية، وأساسية للاختبارات التشخيصية محكية المرجع (Olagunju, 2004).
– عرض البناء الهرمي المتعلق بعملية قسمة الكسور على فريق العمل، والطلب منهم الحكم عليه من حيث درجة:
• صحة العلاقات التي تربط بين عناصر البناء الهرمي.

· انتمائها إلى الموضوع.
· دقة التسلسل الهرمي للعناصر.

وذلك باستخدام مقياس رباعي يبين الدرجات الآتية من: الانتماء/الصحة/الدقة: (غير منتمية/غير صحيحة/غير دقيقة)، (منتمية/صحيحة/دقيقة إلى درجة ما)، (منتمية/صحيحة/دقيقة بشكل مناسب)، (منتمية جداً/صحيحة جداً/دقيقة جداً) حيث أعطيت القيم ١، ٢، ٣، ٤ على الترتيب.

– تفريغ النتائج لكل من البنود الأربعة السابقة، ومن ثم تحليلها بعد تحويل التقديرات إلى تقديرات ثنائية الدرجة، كما في الجدول الآتي رقم (١).

الجدول رقم (١)

تفريغ نتائج المقياس الرباعي للحكم على صدق البناء الهرمي

		المحكم ١	
		ارتباط ضعيف (التقدير ١، ٢)	ارتباط قوي (التقدير ٣، ٤)
المحكم ٢	ارتباط ضعيف (التقدير ١، ٢)	A	B
	ارتباط قوي (التقدير ٣، ٤)	C	D

– استخدام المعادلة رقم (١) (Gregory, 2007, 123) لإيجاد درجة الارتباط بين تقديرات كل محكِّم من فريق العمل:

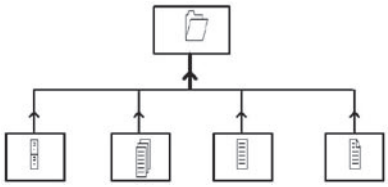
$$\text{درجة الارتباط} = \frac{D}{A+B+C+D} \quad \text{معادلة رقم (١)}$$

حيث تشير (A) إلى عدد العناصر، أو العلاقات بين عناصر المعرفة الرياضية الأعلى والأدنى مرتبة، التي رأى المحكمان الأول والثاني أنها ضعيفة، (B) إلى عدد العناصر، أو العلاقات التي رأى المحكم الأول أنها قوية، في حين رأى المحكم الثاني أنها ضعيفة، (C) عدد العناصر أو العلاقات التي رأى المحكم الأول أنها ضعيفة، في حين رأى المحكم الثاني أنها قوية، (D) عدد العناصر أو العلاقات التي رأى المحكمان الأول والثاني أنها قوية.

– حساب صدق البنية الهرمية من خلال إيجاد معدل جميع الأزواج الممكنة لمعاملات الارتباط لتقديرات المحكمين وفق المعادلة رقم (١)، وعلى كلٍ من البنود الثلاثة السابقة، وقد تراوح معدل تقديراتهم بين ٠,٩٠ – ٠,٩٤

ثانياً: التحقق من صحة الفرضيات المتعلقة بمستويات التحليل في البناء الهرمي:

نظراً لصعوبة التعامل مع البنية الهرمية الكلية، فقد اكتفى الباحثان بذكر الإجراءات المتعلقة بقسمة الكسور العادية، كما وردت في الهرم في المستويين الثاني والثالث من الأعلى، ويبين الشكل (٢) جزءاً من مخطط البنية الهرمية الذي سيجرى عليه توضيح صحة الفرضية.



الشكل رقم (٢)
جزء من مخطط البنية الهرمية الذي سيجرى عليه توضيح صحة الفرضية

والفرضية التي تم اختبارها هي: "كي يحقق الفرد الهدف رقم (١) يجب أن يكون قادراً على تحقيق الأهداف ذوات الأرقام (٢) و (٣) و(٤) و (٥)". فإذا اجتاز الطالب درجة القطع على أي من الاختبارات التي تقيس أيًّا من هذه الأهداف، فإنه يكون قد حقق الهدف الذي يقيسه ذلك الاختبار، ويحصل على الدرجة (١)، وغير ذلك، فإنه يحصل على الدرجة (٠). ويبين الجدول رقم (٢) بعض النواتج الممكنة (الأزواج المرتبة) للدرجات المتعلقة بالأهداف.

الجدول رقم (٢)
بعض النواتج الممكنة للدرجات المتعلقة بالأهداف

مستوى الهدف	البنية الهرمية	تحقق/عدم تحقق الهدف				
الثالث (١)		١	١	٠	٠	٠
الثاني (٢، ٣، ٤، ٥)		١١١١	٠١١١	١١١١	٠١١١	٠٠٠٠
الناتج		(١،١)	(٠،١)	(١،٠)	(٠،٠)	(٠،٠)

وقد تم استخدام هذا الجدول؛ لإيجاد ثلاث نسب مختلفة، تفيد في التحقق من صحة الفرضية السابقة أعلاه، والنسب هي: نسبة الاتساق، ونسبة كفاية الترتيب، ونسبة الاكتمال. ولاختبار الفرضية، اتبعت الإجراءات الآتية:

١– إعداد خمسة اختبارات، يتكون كل منها من عشر فقرات على كل من أصناف المعرفة الواردة في الشبكة (البناء الهرمي)؛ حيث تم إعداد اختبار على كل من: قسمة الكسور العادية (رقم ١)، العلاقة بين ضرب الكسور وقسمتها (رقم ٢)، قسمة كسر على كسر (رقم ٣)، قسمة عدد صحيح على كسر (رقم ٤)، قسمة كسر على عدد صحيح (رقم ٥). وقد اتبعت الإجراءات المنهجية في إعداد الاختبارات من كتابة الأهداف، وبناء جدول مواصفات لكل منها، وبناء فقرات من نمط الاختيار من أربعة خيارات على كل هدف؛ وتحكيم كل من الأهداف، وجدول المواصفات، ودرجة انتماء الفقرات للأهداف؛ ومن ثم تعديل الفقرات في ضوء التغذية الراجعة من المحكمين؛ وإخراج كل من هذه الاختبارات؛ وقد كان الزمن المخصص لكل منها خمس عشرة دقيقة.

٢- اختيار أربع شعب عشوائياً، من مجتمع الدراسة، وخارج عينتها الرئيسة، اثنتان منهما من الإناث، وقد كان عدد الذكور ٩٠ طالباً، وعدد الإناث ٩٤ طالبة.

٣- تطبيق هذه الاختبارات مجتمعة على هذه العينة، حيث طبق الاختباران (١)، (٢) في جلسة واحدة، مدتها ثلاثون دقيقة في اليوم الأول، وبعد يومين طبقت الاختبارات رقم (٣، ٤، ٥) في جلسة أخرى مدتها خمس وأربعون دقيقة. ثم صححت الاختبارات، وذلك بإعطاء الإجابة الصحيحة علامة واحدة، والإجابة الخطأ صفراً.

٤- تحديد نقاط القطع التي تعبر عن تحقيق الهدف الذي يقيسه كل اختبار من الاختبارات الخمسة، وفق ما سيرد في بند (٤) من المراجعة الإمبريقية، وقد كانت كما يأتي: ٧٠٪، ٩٠٪، ٨٠٪، ٨٠٪، ٨٠٪ لكل من الاختبارات ذوات الأرقام (١، ٢، ٣، ٤، ٥) على الترتيب.

٥- تحديد نسبة الطلبة الذين حققوا الأهداف الواردة في الاختبارات الخمس السابقة.

٦- حساب أعداد الطلبة الآتية:

* الذين حققوا الهدف الذي يقيسه اختبار رقم (١) (في المستوى الأعلى)، وحققوا جميع متطلباته السابقة (أهداف الاختبارات رقم٢، ٣، ٤، ٥) (N (١،١)).

* الذين حققوا الهدف الذي يقيسه الاختبار رقم (١)، ولم يحققوا أحد الأهداف، على الأقل، التي تقيسها الاختبارات ذوات الأرقام (٢، ٣، ٤، ٥) (N(١، ٠)).

§ الذين لم يحققوا الهدف الذي يقيسه الاختبار رقم (١)، وحققوا جميع الأهداف التي تقيسها الاختبارات ذوات الأرقام (٢، ٣، ٤، ٥) (N (٠،١)).

* والذين لم يحققوا الهدف الذي يقيسه الاختبار رقم (١)، ولم يحققوا أحد الأهداف، على الأقل، التي تقيسها الاختبارات ذوات الأرقام (٢، ٣، ٤، ٥) (N (٠،٠)).

ويبين الجدول الآتي رقم (٣) توضيحاً لكيفية حساب هذه القيم، حيث يشير العدد (١) إلى أن مجموعة الطلبة حققت الهدف، والعدد (٠) إلى أنها لم تحققه.

الجدول رقم (٣)
مجموعات الطلبة الذين حققوا/لم يحققوا الهدف الرئيس والأهداف الفرعية

رقم مجموعة الطلبة	نوع الهدف/رقم الهدف					الناتج
	الرئيس	الفرعي (المتطلبات السابقة)				
	١	٢	٣	٤	٥	
المجموعة ١	١	١	١	١	١	N(١،١)
المجموعة ٢	١	٠	١	١	٠	N(١،٠)
المجموعة ٣	٠	١	١	١	١	N(٠،١)
المجموعة ٤	٠	١	٠	٠	١	N(٠،٠)

٧- تقدير النسب الآتية: نسبة الاتساق (Consistency ratio)، ونسبة كفاية الترتيب (Adequacy Ratio)، ونسبة الاكتمال (Completeness ratio) (علام، ٢٠٠١، ص٧٢-

٧٤). حيث تختبر نسبة الاتساق ما إذا كان تحقيق الفرد للمهارة الرئيسة (الهدف الرئيس) يتضمن تحقيقه جميع متطلباتها السابقة؛ في حين تختبر نسبة كفاية الترتيب ما إذا كان تحقيق الفرد للمتطلبات السابقة أو بعضها يتضمن تحقيق المهارة الرئيسة؛ وأما نسبة الاكتمال فتحدد قوة اعتماد كل مستوى من مستويات الأهداف على الآخر، وتدل هذه القيمة على نسبة عدد الأفراد الذين حققوا المهارة الرئيسة إلى عدد الذين لم يحققوا أحد متطلباتها السابقة، على الأقل. وتبين المعادلات الآتية ذوات الأرقام (٢)، (٣)، (٤) معادلات هذه النسب.

نسبة الاتساق= $\frac{N(1,1)}{N(1,1)+N(1,0)}$ المعادلة رقم (٢)

نسبة كفاية الترتيب= $\frac{N(1,1)}{N(1,1)+N(0,1)}$ المعادلة رقم (٣)

نسبة الاكتمال= $\frac{N(1,1)}{N(1,1)+N(0,0)}$ المعادلة رقم (٤)

وبتطبيق المعادلات السابقة كانت قيم نسب الاتساق وكفاية الترتيب والاكتمال كما يأتي، وعلى الترتيب: ٠,٩١ ، ٠,٨٧ ، ٠,٨٩؛ ولأن هذه القيم أكبر من ٠,٨٥، فإنها تعد مؤشرات مناسبة لصدق البناء الهرمي، المتعلق بمهارة قسمة الكسور العادية (Walbesser, 1968).

ثالثاً: التحقق من فاعلية البناء الهرمي، وذلك عن طريق تقدير نسبة الطلبة الذين حققوا المهارة الرئيسة ممن تلقوا التعليم لهذه المهارة مسترشدين بهذا البناء الهرمي، وذلك وفق الإجراءات الآتية:

- تم اختيار شعبتين من شعب الصف الخامس من خارج عينة الدراسة، إحداهما من شعب الذكور، والأخرى من شعب الإناث، عشوائياً، تضمان ٦٥ طالباً وطالبة.
- طلب من معلمي الرياضيات لهاتين الشعبتين، أن يدرسا طلبتهما موضوع قسمة الكسور العادية، مسترشدين بالبناء الهرمي الوارد في الشكل رقم (١) أثناء التدريس، وذلك بمراعاة التعلمات السابقة، اللازمة لتحقيق الهدف والواردة في المخطط.
- بعد انتهاء التدريس، طبق الاختبار رقم (١) على طلبة الشعبتين المذكورتين.
- حُسبت نسبة الطلبة الذين حققوا المهارة الرئيسة، فكانت ٨٨ ٪ تقريباً. وتعد هذه القيمة مؤشراً جيداً على فعالية البناء الهرمي (Walbesser, 1968).

رابعاً: التحليل الدقيق للمهارات الجزئية، وقد تضمن تنفيذ الآتي:

أ. التحليل الدقيق لكل هدف رئيس في البنية الهرمية إلى أهدافَ سلوكية أكثر تحديداً.

تم تحديد الأهداف السلوكية الفرعية التي تتناول نواتج تعلمية، يتوقع من الطلبة تحقيقها، لكل هدف رئيس، وهي بمجملها تعطي مؤشراً على تحقيق الهدف العام؛ بحيث تعبر هذه

النواتج عن الأداء تعبيراً دقيقاً. وعليه، تم تحليل المتطلبات السابقة اللازمة لمهارة قسمة الكسور العادية. وبما أن هذه العملية طويلة جداً، وتحتاج إلى وقت وجهد، وبهدف توضيح خطوات الإستراتيجية واختصارها، فقد اكتفى الباحثان بتضمين المهارات الأدنى مرتبة والأكثر تحديداً لمهارة قسمة كسر على عدد صحيح.

ب. التحقق من دقة التحليل

تم التحقق من دقة التحليل بعرضه على فريق العمل، الذين طلب منهم تقدير درجة دقة التحليل على مقياس رباعي على النحو الآتي: دقيق جداً (٤) ، دقيق (٣)، الدقة متوسطة (٢)، الدقة ضعيفة (١)، ومن ثم استخدمت المعادلة رقم (١) وبالكيفية ذاتها، حيث تراوح معدل التقديرات بين ٠,٩١ – ٠,٩٣، وهي قيم جيدة تدل على دقة التحليل.

خامساً: تحديد مواصفات الفقرات في المجال السلوكي

١. تحديد سمات خيارات الفقرات

وقد تم ذلك من خلال حصر أخطاء الطلبة المحتمل وقوعها؛ كي يتم تضمينها في الاختبار، حيث قام الباحثان بما يلي:

– حضور أربع حصص صفية في موضوع قسمة الكسور العادية، اثنتان في الصف الخامس، ومثلهما في الصف السادس؛ بهدف تعرف الأخطاء التي يرتكبها الطلبة أثناء قيامهم بحل مسائل على موضوع القسمة.

– تحليل ٥٠ ورقة من أوراق الامتحانات التي أعدها المعلمون في موضوع قسمة الكسور العادية لخمسين طالباً وطالبة من الصف الخامس، وذلك بهدف تعرف الأخطاء التي يقع فيها الطلبة، وقد شملت عملية التحليل أوراقاً من مستويات الطلبة الثلاثة (عليا ووسطى ودنيا)؛ على النحو الآتي: (١٥) ورقة من الفئة العليا، و(٢٠) ورقة من الفئة المتوسطة، و(١٥) ورقة من الفئة الدنيا. وقد اعتمد تحديد الفئات وفق متوسط علاماتهم في الرياضيات ذكره.

– مقابلة ١٠ من طلبة الصف الخامس من الفئة الوسطى، والفئة الدنيا، لتعرف أسباب تقديم الإجابات الخطأ التي ظهرت في أوراق امتحاناتهم.

وقد توصل الباحثان نتيجة للإجراءات السابقة، إلى تحديد قائمة تضمنت خمسة وعشرين خطأ مختلفاً. وقد تم استخدامها كخيارات للفقرات الاختبارية.

٢. تحديد مواصفات الفقرات في المجال السلوكي

تم إعداد مواصفات لكل فقرة في المجال السلوكي، وقد تضمنت مواصفات الفقرة: الهدف الذي تقيسه الفقرة؛ وعدد الفقرات التي تقيس الهدف؛ بغرض الحصول على تمثيل مناسب للمجال؛ والوصف التفصيلي لسمات المثيرات، وسمات الاستجابة الصحيحة؛ ومواصفات خيارات أو مموهات الفقرة اعتماداً على قائمة الأخطاء التي تم التوصل إليها في الخطوة السابقة (Crocker & Algina, 1986; Popham, 1981). وقد كان تضمين

سمات المثيرات، والاستجابات، والخيارات في المواصفات أمراً حاسماً؛ لأن الفرد الذي يقوم ببناء الاختبار يعتمد عليها اعتماداً أساسيا (الجلبي، ٢٠٠٥). وبعد تحديد مواصفات الفقرات تم مراجعتها، وذلك من خلال:

– عرض المواصفات على فريق العمل بهدف تعرف درجة مناسبة المواصفات للأهداف، ووضوحها، ودقتها، ومناسبة عدد الفقرات المقترحة.

– تعديل المواصفات وفق التغذية الراجعة التي زود المحكمون بها. ويبين الجدول رقم (٤) مواصفات الفقرات التي تقيس الهدف: " يقسم كسراً أقل من واحد صحيح، بسطه ومقامه أقل من ١٠ على عدد صحيح أقل من ١٠ " بعد أن أجمع عليها المحكمون.

الجدول رقم (٤)
مواصفات الفقرات التي تقيس الهدف: "يقسم كسراً أقل من واحد صحيح بسطه ومقامه أقل من ١٠ على عدد صحيح أقل من ١٠" بعد أن أجمع عليها المحكمون

المهارة الفرعية	قسمة كسر على عدد صحيح
الهدف السلوكي لكل هدف رئيس	يقسم كسراً أقل من واحد صحيح بسطه ومقامه أقل من ١٠ على عدد صحيح أقل من ١٠
عدد الفقرات المقترحة	١٠ فقرات من نمط الاستجابة المختارة (اختيار من أربعة خيارات)
سمات المثير	سمات الاستجابة
١. يجب أن تتطلب الفقرة قسمة كسرين عاديين ٢. الكسور في كل فقرة يجب أن تكون مكتوبة بشكل أفقي ٣. بسط المقسوم ومقامه أوليان فيما بينهما وكلاهما أقل من ٩ ٤. العدد الصحيح (المقسوم عليه) يتكون من منزلة واحدة ٥. هناك حاجة لإجراء خطوتين على الأقل قبل التوصل إلى الإجابة ٦. يجب ألا يزيد كل من بسط الكسر الناتج ومقامه على ثلاثة أرقام ٧. يجب ألا يستخدم الكسر أكثر من مرة واحدة	١. الشكل: كل الكسور يجب أن تكون مكتوبة على شكل كسر عادي (بسط ومقام) ٢. الخيارات يجب أن يكون هناك أربعة خيارات لكل فقرة هي: أ) الاستجابة الصحيحة ب) أحد الخيارات يجب أن يعكس خطأ ناتجاً من قسمة كل من بسط الكسر ومقامه على العدد الصحيح ج) أحد الخيارات يجب أن يعكس خطأ ناتجاً من طرح بسط الكسر من العدد الصحيح (بغض النظر عن إشارة العدد) ثم وضع المقام كما هو دون تغيير. د) أحد الخيارات يجب أن يعكس خطأ ناتجاً من قلب المقسوم والمقسوم عليه ثم ضرب النواتج في بعضها هـ) يجب ألا يوضع الخيار الآتي: "لا شيء مما ذكر" ٣. الناتج: ناتج العملية يوضع على شكل كسر

ثانياً: بناء الفقرات الاختبارية

قام الباحثان ببناء فقرات اختبارية من نوع الاختيار من متعدد على كل هدف سلوكي ينبثق عن مهارة "قسمة الكسور العادية" الواردة في شكل (١)، وفق الخطوات الآتية:

١) بناء فقرات بوصفها أمثلة على مواصفات الفقرات على أساس تجريبي؛ للمساعدة في تقصي أية مشكلة متوقعة في المواصفات (Berk, 1982; Crocker & Algina, 1986)، وهذه الفقرات يمكن استخدامها كنماذج يتم محاكاتها، وقد اصطلح على تسميتها في هذا البحث "الفقرة النموذج". وقد قام كلا الباحثين ببناء "فقرة نموذج" على كل من مواصفات الفقرات، وكلِّ هدف من الأهداف المراد قياسها، وذلك بهدف تسهيل مهمة إعداد الفقرات التي ستتضمنها الصورة النهائية للاختبار. وكمثال على "الفقرة النموذج" التي مواصفاتها كما في الجدول رقم (٤) تم بناء الفقرة الآتية:

ضع دائرة حول رمز الإجابة الصحيحة لناتج العملية الآتية:

١) =

أ) ب) ج) د)

٢) عرض هذه "الفقرات النموذج"، ومواصفات الفقرة المناظرة على فريق العمل؛ للحكم على درجة التطابق بين الفقرات ومواصفاتها، والحكم أيضاً على درجة تمثيل الفقرات للمجال.

٣) تقويم وضوح " الفقرات النموذج"، وذلك من خلال تجريبها على طلبة شعبة مكونة من ٢٧ طالباً من خارج عينة الدراسة، وتبين نتيجة التجريب أن صياغة "الفقرات النموذج" واضحة.

٤) بسبب وجود مواصفات تفصيلية لكل الفقرات في المجال السلوكي، فقد كان من الممكن دراسة تشابه استجابات مجموعة الطلبة على كل أزواج "الفقرات النموذج" التي تم بناؤها وفق المواصفات التفصيلية للفقرة. وعليه، فقد تم اختيار شعبتين، عشوائيا، تضمان خمسين طالباً وطالبةً، من خارج عينة الدراسة. ومن ثم تطبيق أزواج "الفقرات النموذج" عليهم. وبعد تصحيح الاستجابات، تم تصنيف الاستجابات في جدول رباعي، يبين العلاقة بين الاستجابات على كل زوجين من "الفقرات النموذج" كما في الجدول رقم (٥) الآتي الذي يبين مثالاً على زوجين، فقط، من "الفقرات النموذج" (الفقرة١، الفقرة ٢).

الجدول رقم (٥)
استجابات الطلبة على زوجين من "الفقرات النموذج"

		الفقرة ١	
		+	−
الفقرة ٢	+	a	b
	−	c	d

ويبين جدول (٦) توزيع تكرارات استجابات الطلبة على هذين الزوجين من "الفقرات النموذج"

الجدول رقم (٦)
توزيع تكرارات استجابات الطلبة على الزوجين السابقين من "الفقرات النموذج"

		الفقرة ١	
		+	−
الفقرة ٢	+	٢٣	٤
	−	٢	٢١

حيث يمثل العدد (٢٣) في الجدول رقم (٦) عدد الطلبة الذين أجابوا إجابة صحيحة على زوجي "الفقرات النموذج"، في حين يمثل العدد (٢) عدد الطلبة الذين أجابوا إجابة صحيحة على الفقرة (١) وإجابة خاطئة على الفقرة (٢)، وهكذا.

وقد تمت دراسة تشابه استجابات مجموعة الطلبة على كل أزواج "الفقرات النموذج" وفق النماذج والأغراض الآتية:

١. بغرض معرفة إذا كانت كل "الفقرات النموذج" تقيس المحتوى نفسه، كما عرف في المواصفات التفصيلية للفقرات، فقد تم استخدام اختبار الاستقلال الإحصائي χ^2، كما في المعادلة رقم (٥) الآتية (Crocker & Algina, 1986, p.332)، لاختبار الفرضية الصفرية الآتية: "استجابات الطلبة على زوجي "الفقرات النموذج" مستقلة عن بعضها".

$$\chi^2 = \frac{n(ad-bc)^2}{(a+b)(c+d)(b+d)(a+c)}$$ معادلة رقم (٥)

وتقارن قيمة χ^2 الناتجة مع قيمة χ^2 الحرجة بدرجة حرية واحدة. وفي المثال أعلاه، كانت قيمة χ^2 = ٤,٣٦، في حين كانت قيمتها الحرجة بدرجة حرية واحدة =٣,٨٤، على مستوى دلالة ٠,٠٥. وعليه، نرفض الفرضية الصفرية، أي أن الاستجابات على "الفقرتين النموذج" معتمدة على بعضها. وهذا يدعم التفسير بأن زوجي "الفقرات النموذج" يقيسان نفس المهارة أو المحتوى.

٢. وبعد أن تبين أن هنالك تشابها في استجابات الطلبة، فقد رغبنا في حساب إحصائي يصف درجة هذا التشابه. وعليه فقد تم حساب نسبة الاتفاق في عدد الطلبة الذين أجابوا إجابة صحيحة أو خاطئة على زوجي "الفقرات النموذج" باستخدام المعادلة رقم (٦) الآتية (Crocker & Algina, 1986,p334):

المعادلة رقم (٦)
$$p = \frac{(a+d)}{n}$$

حيث تشير (P) إلى نسبة اتفاق الطلبة الذين استجابوا بطريقة متسقة على زوجي "الفقرات النموذج". وباستخدام البيانات الواردة في الجدول (٦)، كانت قيمة نسبة الاتفاق= ٠,٨٨، بمعنى أنه يوجد أداء متسق بين ٨٨٪ من الطلبة على زوجي "الفقرات النموذج".

٣. وفضلاً عن ذلك، فقد تم اختبار الفرضية الآتية: "صعوبة زوجي "الفقرات النموذج" في مجتمع المفحوصين متساوية". وبعبارة أخرى، هل الفرق الملاحظ في الصعوبة صغير جداً، بحيث يمكن عزوه إلى أخطاء المعاينة؟ وقد تم استخدام اختبار χ^2 الذي اقترحه هاريس وبيرلمان (Harris & Pearlman, 1977) الوارد في المعادلة رقم (٧):

المعادلة رقم (٧)
$$\chi^2 = \frac{(|b-c|-1)^2}{b+c}$$

وتقارن قيمة χ^2 الناتجة مع قيمة χ^2 الحرجة بدرجة حرية واحدة. وباستخدام البيانات الواردة في الجدول (٦)، كانت قيمة χ^2= ٠,١٧، في حين كانت قيمتها الحرجة بدرجة حرية واحدة=٣,٨٤، على مستوى دلالة ٠,٠٥. وعليه، نقبل الفرضية الصفرية، أي نفشل في إيجاد فرق دال إحصائياً في صعوبة زوجي "الفقرات النموذج"، بمعنى أن صعوبة زوجي "الفقرات النموذج" متساوية، إحصائياً.

وفي ضوء ما سبق، تبين من أدلة اتفاق الفقرات أن هناك درجة من الاتفاق في استجابات الطلبة على أزواج "الفقرات النموذج"، وهذا يعد مؤشراً على صدق بناء الفقرات.

٥) من أجل التأكد من أن الاختبار الذي سيتم إعداده، يمثل مواصفات الفقرات الاختبارية، قام كل باحث منفرداً مسترشداً بالوصف التفصيلي لفقرات الاختبار، ومستفيداً من "الفقرات النموذج" التي أعدت سابقاً، بإعداد اختبار من نوع الاختيار من أربعة خيارات يضم اثنتي عشرة فقرة اختبارية، تقيس الهدف السلوكي السابق نفسه: "يقسم كسراً أقل من واحد صحيح، بسطه ومقامه أقل من ١٠ على عدد صحيح أقل من ١٠".

ثالثاً: مراجعة الفقرات الاختبارية وتصديقها

للتأكد من مدى صلاحية فقرات الاختبارين المعدين للكشف عن مواطن الضعف، ومواطن القوة لدى الطلبة، فقد تمت مراجعتها مراجعة منطقية، معتمدة على تقديرات فريق

العمل، ومراجعة إمبريقية، تستند إلى تحليل نتائج تطبيق الاختبارات، وفيما يأتي تبيان لكيفية مراجعة هذه الفقرات:

١) المراجعة المنطقية للفقرات، وقد تضمنت الخطوتين الآتيتين:

أ- التأكد من مواءمة الفقرات لمعايير بناء الفقرات المتبعة في الأدب التربوي

ب- التحقق من اتساق الفقرات الاختبارية مع مواصفات الفقرة

وفيما يأتي توضيح لكل منهما.

أ) التأكد من مواءمة الفقرات لمعايير بناء الفقرات المتبعة في الأدب التربوي؛ وقد تمت مراجعة الفقرات الاختبارية وفق معايير رويد وهالادينا (Roid & Haladyna , 1982, p203-206) الآتية:

- أن يكون الطول في جذر السؤال وليس في الخيارات
- معظم المعلومات تكون في الجذر
- البعد عن العبارات المنفية
- توافق الخيارات مع الجذر من الناحية اللغوية.
- تساوي طول الإجابة الصحيحة مع طول الخيارات.
- عدم وجود تلميحات تساعد على الإجابة.
- وجود تعليمات واضحة للإجابة
- جودة اللغة والسلامة العلمية.

وقد تحقق الباحثان من توافر هذه المعايير؛ إذ إن جميع الفقرات الاختبارية قد تم بناؤها على شكل عبارات رياضية واضحة ومختصرة، كما تم وضع تعليمات واضحة للإجابة عنها.

ب) التحقق من اتساق الفقرات الاختبارية مع مواصفات الفقرة (الصدق المنطقي)

وقد تم فحص درجة اتساق الفقرات الاختبارية مع مواصفات الفقرة بثلاث طرق:

الطريقة الأولى: للتحقق من مدى اتفاق فريق العمل حول درجة اتساق الفقرات مع مواصفات الفقرة، تم اتباع الإجراءات التي اقترحها أولاجونجو (Olagunju, 2004) الآتية:

- صُممت استمارة، تضم الأهداف التي يقيسها الاختبار، والفقرات الاختبارية التي تقيس كل هدف، ومواصفات كل فقرة من حيث شكلها، والكلمات، والأعداد المستخدمة فيها، والسلوك المستهدف.
- طُلب من كل محكم وضع الرقم (+١) إذا كان متأكداً أن الفقرة الاختبارية تتفق مع أي من هذه المواصفات، والرقم (صفر) إذا كان غير متأكد، والرقم (-١) إذا كان متأكداً أن الفقرة الاختبارية لا تتفق مع أي من المواصفات المتعلقة بها. ويبين الجدول رقم (٧) مخطط أداة تقدير ارتباط الفقرة بالهدف وبمواصفاتها التفصيلية.

الجدول رقم (٧)
مخطط أداة تقدير ارتباط الفقرات بالأهداف وبمواصفاتها التفصيلية

الهدف	الفقرة الاختيارية	معايير تقدير الفقرات											
		الشكل			الكلمات المستخدمة			الأعداد			السلوك		
		١	٠	-١	١	٠	-١	١	٠	-١	١	٠	-١
١	١												
	٢												
	...												
٢	١												
	٢												
...	...												

– حُسب معامل الاتساق((Index of Consistency (IOC) لكل فقرة في كلا الاختبارين اللذين أعدهما الباحثان، من خلال تطبيق المعادلة الآتية رقم (٨) (Roid & Haladyna, 1982, 210).

$$IOC = \frac{(N-1)S_1 - S_2 + S_1}{2(N-1)n}$$

المعادلة رقم (٨)

حيث تمثل:

N : عدد الأهداف التي يقيسها الاختبار ، S_1 : مجموع تقديرات فريق العمل لهدف واحد وفقرة واحدة

S_2 : مجموع تقديرات فريق العمل لكل الأهداف، ولفقرة واحدة ، n : عدد أفراد فريق العمل.

وقد بينت النتائج أن معاملات اتساق الفقرات الاختبارية مع الأهداف التي تنتمي لها قد تراوحت بين ٠,٨٤ و، ٠,٩٥ لكلا الاختبارين، وهي أعلى من معاملات اتساق الفقرة مع أي هدف آخر في الاختبارين، وفضلاً عن ذلك فإن قيم معظم معاملات اتساق الفقرات الاختبارية مع الأهداف الأخرى التي لا تنتمي إليها سالبة، أما القيم الموجبة فجميعها أقل ٠,٢٨.

الطريقة الثانية: تم التحقق من أن جميع الفقرات في كلا الاختبارين، لها نفس الوزن في قياس المجال السلوكي للاختبار، حيث طلب من أعضاء فريق العمل تقدير أهمية كل فقرة في قياس المجال السلوكي للاختبار، وذلك من خلال الإجابة على مقياس خماسي الدرجة (يتراوح ما بين أوافق بشدة (٥) إلى أعارض بشدة (١))، وقد بينت النتائج أن جميع الفقرات الاختبارية، وفي كلا الاختبارين، قد حازت على درجة من الأهمية في وصف المجال السلوكي تراوح بين ٤ و ٥ درجات، مما يشير إلى أهمية كل الفقرات. وتعد هذه النتائج دليلاً آخر على اتساق الفقرات الاختبارية مع مواصفات الفقرة.

الطريقة الثالثة: غالباً ما يتم تقييم صدق المحتوى باستخدام متوسط تقديرات المحكمين لكل فقرة من فقرات الاختبار. ولكون تقديرات المحكمين غير موزعة توزيعاً سوياً، فقد استخدمت فترة ثقة، غير متماثلة لمتوسط مجتمع متغير مقياس التقدير، اقترحها بنفيلد

(Penfield, 2003)، حيث تبين المعادلتان رقم (٩) ورقم (١٠) قيمتي الحدين الأدنى والأعلى للفترة، على الترتيب.

الحد الأدنى لفترة الثقة $M - z\sqrt{\frac{k\pi_l(1-\pi_l)}{n}}$ المعادلة رقم (٩)

الحد الأعلى لفترة الثقة $M - z\sqrt{\frac{k\pi_u(1-\pi_u)}{n}}$ المعادلة رقم (١٠)

حيث:

$$M = \frac{\sum_{i=1}^{n} X_i}{n}$$

$$\pi_l = \frac{2pnk + z^2 - z\sqrt{4nkp(1-p)+z^2}}{2(nk+z^2)}$$

$$p = \frac{\sum_{i=1}^{n} X_i}{nk}$$

$$\pi_u = \frac{2pnk + z^2 + z\sqrt{4nkp(1-p)+z^2}}{2(nk+z^2)}$$

n : عدد المحكمين

k : أعلى قيمة ممكنة لتقدير المحكمين في سلم تقدير الفقرة ضمن المقياس.

قيمة z التي تناظر فترة ثقة ٩٠٪ هي ١,٦٤، والتي تناظر ٩٥٪ فترة ثقة تساوي ١,٩٦

وقد قام الباحثان بفحص الفرضية البحثية الآتية: {μ> 2, α=0.05: H}، حيث μ متوسط تقديرات مجتمع فريق العمل (المحكمين)، أما القيمة (٢) فتمثل متوسط قيم سلم التقدير. ولفحص هذه الفرضية، تم فحص الفرضية الصفرية الآتية: {μ> 2, α=0.05: H}، وذلك من خلال حساب قيمتي الحدين الأدنى والأعلى لفترة ثقة محددة، ولكل فقرة اختبارية باستخدام المعادلتين رقم (٩، ١٠)، فإذا كان الحد الأدنى لفترة الثقة أكبر من القيمة الثابتة (٢) الواردة في الفرضية الصفرية، يتم قبول الفرضية البحثية؛ وإلا فترفض (Penfield & Miller, 2004).

ويبين الجدول الآتي رقم (٨) مثالاً توضيحياً لاستخدام معادلة بنفيلد لأربع فقرات من الاختبارين وفترة ثقة ٩٠٪ لمتوسط المجتمع (μ)، والقرارات المتعلقة برفض الفرضية البحثية أو قبولها.

الجدول رقم (٨)
مثال توضيحي لاستخدام معادلة بنفيلد لأربع فقرات

الاختبار	رقم الفقرة	تكرار التقديرات					متوسط تقديرات المحكمين (M)	٩٠٪ فترة ثقة		القرار
		٠	١	٢	٣	٤		الحد الأدنى	الحد الأعلى	
١	١	١	٠	٤	٣	٢	٢,٥	١,٩٩	٢,٩٥	رفض
	٢	٠	٠	٠	٤	٦	٣,٦	٣,٠٩	٣,٨٦	قبول
٢	١	١	٢	٣	٣	١	٢,١	١,٦١	٢,٥٨	رفض
	٢	٠	٠	٣	٥	٢	٢,٩	٢,٣٨	٣,٣٠	قبول

وفي ضوء ذلك، فقد حذفت الفقرات التي لم تحقق الفرضية المذكورة، وكان عددها أربعاً (فقرتان متناظرتان من كل اختبار)، وبقي عشر فقرات في كل اختبار. ومن ثم وضعت تعليمات الإجابة، ومعلومات تتعلق بالمفحوص، وتم إخراج الاختبار. ولتحديد الزمن اللازم، فقد تم تطبيق أحد الاختبارين على شعبتين، من خارج عينة الدراسة، إحداهما من الذكور، تضمان خمساً وأربعين طالباً وطالبة. وحسب متوسط الزمن الذي استغرقوه في الإجابة فكان اثنتي عشرة دقيقة؛ ومن ثمّ قدر الزمن اللازم لهذا الاختبار بأربع عشرة دقيقة.

٢) المراجعة الإمبريقية

يرى ميسك (Messick, 1993) أن الأدلة التي تزودنا بها المراجعة المنطقية غير كافية للثقة بصدق الفقرات الاختبارية، فقد تزودنا عمليات تحليل درجات الطلبة على الاختبار بنتائج مغايرة. لذلك قام الباحثان بجمع أدلة كمية مرتبطة بتفسير الاستجابة على الفقرة الاختبارية، ومقارنتها بالأدلة المنطقية بهدف التزويد بأدلة عن صدق الاختبار وثباته وصعوبة كل من فقراته، وتمييزها، وفاعلية خيارات الفقرات. وقد تم اتباع الخطوات الآتية:

١) دمج الاختبارين المعدين من الباحثين في اختبار واحد بترتيب عشوائي، وذلك من أجل توفير ظروف متشابهة للتطبيق، والتأكد من أن إجابات الطلبة ليست عشوائية.

٢) تطبيق الاختبار الناتج على عينة الدراسة البالغ عددها ١٧٧ طالباً وطالبة.

٣) التأكد من عدم وجود إجابات عشوائية، حيث درست استجابات الطلبة، ومن ثم استبعدت إجابات ١٢ طالباً وطالبة، بسبب ظهور عدم توافق في إجاباتهم على فقرات الاختبارين الأصليين.

٤) تحديد درجة القطع، وتم ذلك بطريقتين:

· طريقة أنجوف: طُلب من فريق العمل وضع تقديرات عامة لدرجة القطع لكل اختبار، ثم حسب معدل هذه التقديرات (Murphy & Davidshofer, 2001, p105)، وقد بلغت هذه الدرجة ٧٧٪ تقريباً.

· طريقة ندلسكي: طُلب من كل محكم من فريق العمل تقدير عدد الخيارات التي يمكن أن يستبعدها طلبة الفئة غير المتقنة في كل فقرة اختبارية، وحُسب مجموع مقلوب عدد الخيارات المتبقية، ثم حُسب معدل تقديرات المحكمين (Crocker, & Algina, 1986, p 212-213)، فكانت درجة القطع ٧٢٪ تقريباً.

وقد اعتمد الباحثان الدرجة ٧٥٪ لتكون درجة القطع للاختبارات التشخيصية باعتبارها مساوية، تقريباً، لمعدل درجتي القطع المقدرة من الطريقتين.

٥) التوصل إلى مؤشرات الثبات:

تم التوصل إلى مؤشرات ثبات الدرجات بالطريقتين الآتيتين:

أ) ثبات اتساق قرار التصنيف باستخدام معامل كوهين كابا ((Cohen's Kappa (K))،

حيث تم التحقق من ثبات قرارات التصنيف الإتقاني للأفراد (متقنون أو غير متقنين) على الاختبارين اللذين أعدهما الباحثان، وفق درجة القطع التي تم تبنيها في هذا البحث (٧٥٪). ويبين الجدول رقم (٩) الهيكل التنظيمي للطريقة التي تم بها التوصل إلى دلالات الثبات باستخدام إحصائي كوهين كابا (K).

الجدول رقم (٩)
الهيكل التنظيمي لطريقة التوصل إلى دلالات الثبات باستخدام معامل كوهين كابا

الاختبار الثاني	الاختبار الأول		المجموع
	متقن	غير متقن	
متقن	A	G	F
غير متقن	H	B	E
المجموع	C	D	I

تم تقدير معامل كوهين كابا باستخدام المعادلة رقم (١١) (Shultz & Whitney, 2005, 75) الآتية:

معادلة رقم (١١)
$$K = \frac{P_o - P_{ex}}{1 - P_{ex}}$$

حيث تمثل P_o نسبة الاتفاق في التصنيف الملاحظ، وتساوي (A+B).

P_{α} نسبة الاتفاق المتوقع في التصنيف، وتساوي (C× F+D ×E)

وقد بلغت قيمة معامل كوهين كابا ٠,٨٦، عند درجة قطع ٧٥٪، ويستدل من ذلك على وجود درجة جيدة من الاتفاق في تصنيف الأفراد على الاختبارين (Berk, 1982). أي أن الاختبارين يصنفان الأفراد بطريقة متشابهة.

ب) ثبات تقدير درجات الأفراد على المجال السلوكي، الذي يتطلب معرفة انحراف القيمة التقديرية لدرجة الفرد في مجال سلوكي معين عن درجة قطع معينة، وقد استخدم معامل ثبات ليفنجستون Livingston وفق المعادلة رقم (١٢) (Feldt & Bernan 1993, 140) الآتية:

معادلة رقم (١٢)
$$\left[\hat{k}^2 = 1 - \frac{1}{1 - n_i}\left[\frac{\sum \bar{x}_i(1-\bar{x}_i)/n_i - s^2(\bar{x}_p)}{(\bar{x}-\lambda)^2 + s^2(\bar{x}_p)}\right]\right]$$

حيث:

$\hat{k}^2$ تشير إلى معامل ليفنجستون.

$$s^2(\bar{x}_p) = \frac{(\bar{x}_p - \bar{x})^2}{n_p}$$

$\bar{X}_p$ متوسط درجات الفرد (p) على فقرات الاختبار، $\bar{X}$ المتوسط الكلي للاختبار، $\ddot{e}$ درجة القطع، $\bar{X}_i$ متوسط الدرجات على الفقرة رقم (i)

وقد بلغ معامل الثبات للاختبار ٠,٩١، وفي ضوء ما سبق، يمكن القول إن الاختبار يتمتع بدرجة جيدة من الثبات.

٦) التوصل إلى مؤشرات الصدق

بما أن الاختبار التشخيصي محكي المرجع يتضمن مجموعة من الاختبارات التشخيصية الجزئية التي يقيس كل منها مهارة فرعية، أو هدفاً سلوكياً محدداً؛ فإن تفسير نتائج هذه الاختبارات قد يكون مضللاً، إذا لم تكن هذه الاختبارات أحادية البعد (Hattie, 1985)، وقد اقترح نانداكومار (Nandakumar, 1994) استخدام التحليل العاملي لفحص أحادية البعد، وفيما يأتي نتائج هذا التحليل لأحد الاختبارات، الذي أعده أحد الباحثين الموضح في الجدول رقم (١٠):

الجدول رقم (١٠)
التحليل العاملي للاختبار

العامل	الجذر الكامن Eigenvalues	التباين المفسر	التباين المفسر التجميعي	تشبعات الفقرات على العامل الوحيد
الأول	٥,١٢٣	٥١,٢٢٦	٥١,٢٢٦	٠,٤٢٥
الثاني	٠,٩٦٣	٩,٦٣		٠,٣٦١
الثالث		٠,٨١٦		٠,٣٤٦
الرابع		٠,٧٢١		٠,٦١٠
الخامس		٠,٥٧٩		٠,٥٨٨
السادس		٠,٥٧٩		٠,٥٢٤
السابع		٠,٥٤٧		٠,٤٧٦
الثامن		٠,٣٨٩		٠,٥٤٧
التاسع		٠,٣٢٣		٠,٥٩٦
العاشر		٠,٢٩٥		٠,٦٥٠

يتبيّن نتائج التحليل العاملي للاختبار في الجدول رقم (٩) وجود عامل واحد فقط، قيمة جذره الكامن تزيد على واحد، وأن هذا العامل قد فسر نسبة جيدة من التباين (٥١,٢٢٦ ٪)، كما تبين النتائج أن تشبعات الفقرات على هذا العامل تزيد على الحد الأدنى لتشبعات الفقرات على عوامل الاختبار، وهو ٠,٣٠، ويدل ذلك على أن جميع فقرات الاختبار لها ارتباط جيد مع هذا العامل. وتعد هذه النتائج دليلاً على أحادية البعد لهذا الاختبار.

٧) تقدير فاعلية الفقرات

تم استخراج دلالات عن فاعلية الفقرات بإيجاد معاملات صعوبة كل فقرة من الفقرات الاختبارية، وتمييزها، وفاعلية خياراتها. وفيما يأتي عرض لهذه الدلالات.

١- معامل الصعوبة: ويعبر عنه بنسبة الطلبة الذين أجابوا عن الفقرة الاختبارية إجابة صحيحة.

٢- معامل التمييز: ويعبر عنه بنسبة الطلبة المتقنين، الذين أجابوا عن الفقرة إجابة صحيحة، مطروحا منها نسبة الطلبة غير المتقنين، الذين أجابوا عن الفقرة إجابة صحيحة، حيث تم تحديد الفئة المتقنة بالفئة التي تجاوزت درجات أفرادها على الاختبار درجة القطع (٧٥٪) أما الفئة غير المتقنة، فهي الفئة لم تتجاوز درجات أفرادها على الاختبار درجة القطع.

٣- فاعلية الخيارات: وتمثل درجة تمييز الخيار بين فئتين من الطلبة (الفئة المتقنة، والفئة غير المتقنة) حيث تم تقدير فاعلية كل خيار بالطريقة التي حسب بها معامل تمييز الفقرة الاختبارية نفسها.

ويبين الجدول (١١) معاملات الصعوبة ومعاملات التمييز وفاعلية خيارات كل فقرة اختبارية، حيث رتبت الخيارات في الجدول ترتيباً أبجدياً، بعد استبعاد الإجابة الصحيحة.

الجدول رقم (١١)
معاملات الصعوبة، ومعاملات التمييز، وفاعلية خيارات الفقرات الاختبارية

رقم الفقرة	معامل الصعوبة	معامل التمييز	فاعلية الخيارات		
			الأول	الثاني	الثالث
١	٠,٦١	٠,٥٧	-٠,١٦	-٠,٢٠	-٠,٢٠
٢	٠,٦٥	٠,٥٥	-٠,١٨	-٠,١٩	-٠,١٨
٣	٠,٥٥	٠,٤٩	-٠,١٥	-٠,٠٧	-٠,١٥
٤	٠,٥٥	٠,٦٩	-٠,٣٩	-٠,١٠	-٠,٢٠
٥	٠,٦١	٠,٦٧	-٠,٠٦	-٠,٤٥	-٠,١٦
٦	٠,٦٣	٠,٦٠	-٠,٣٩	-٠,١٠	-٠,١١
٧	٠,٦٥	٠,٥٧	-٠,١٢	-٠,٣١	-٠,١٤
٨	٠,٥٧	٠,٦٤	-٠,٠٨	-٠,٥٠	-٠,٠٦
٩	٠,٥٧	٠,٧٤	-٠,٠٦	-٠,٥١	-٠,١٨
١٠	٠,٦١	٠,٦٨	-٠,١٢	-٠,٢٥	-٠,٣١

ويظهر الجدول أن صعوبة الفقرات الاختبارية قد تراوحت بين ٠,٥٥ و ٠,٦٥، ويستدل من ذلك على أن جميع هذه الفقرات متوسطة الصعوبة، إذ إن الصعوبة المثالية لفقرات اختبار من نوع الاختيار من متعدد، عدد خياراته أربعة، هي ٠,٦٢٥، (Hambleton & Sireci, 1997) وتعد هذه الدرجات من الصعوبة ملائمة لاختبار تشخيصي.

كما يظهر الجدول أن جميع معاملات تمييز الفقرات الاختبارية تزيد على (٠,٤٠) ويستدل من ذلك على أن تمييز جميع الفقرات مرتفع (Berk, 1982)، وأنها قادرة على التمييز بين الفئة المتقنة، والفئة غير المتقنة.

وفيما يتعلق بفاعلية الخيارات، فقد بينت النتائج أن جميع معاملات تمييز هذه الخيارات سالبة، ومتفاوتة في قدرتها التمييزية، حيث تراوحت من -٠,٠٦ إلى ٠,٥١، مما يدل على أن نسبةالأفراد الذين اختاروا الخيار الخطأ من المجموعة المتقنة أقل من نسبة نظرائهم من

المجموعة غير المتقنة، التي اختارت الخيار عينه، وأن نسبة الذين اختاروا أياً من الخيارات الخطأ أقل من نسبة الذين اختاروا الإجابة الصحيحة؛ وهذه النتائج تتفق مع الشروط التي يجب توافرها في خيارات الفقرة الاختبارية ،كما وصفها بيرك (Berk, 1982).

وفي الختام، فقد تم تقديم استراتيجية لبناء اختبارات تشخيصية محكية المرجع، تستند إلى منحى المواصفات التفصيلية، والشاملة في وصف المجال السلوكي، وتتضمن ثلاث خطوات رئيسة: الأولى، إعداد مواصفات تفصيلية شاملة للاختبار من حيث: تحديد الغرض من الاختبار، ومجال محتواه، والمجموعة التي سيطبق عليها، وفريق العمل؛ وتحليل الموضوع المستهدف تحليلاً تفصيلياً، وتحديده عناصر المعرفة المرتبطة بالموضوع وتنظيمها، والتحقق من صدق التحليل، وإعداد مواصفات تفصيلة للفقرة. في حين كانت الخطوة الثانية بناء الفقرات الاختبارية، وإعطاء أدلة على صدق بنائها من خلال بناء "الفقرة النموذج". في حينكانت الخطوة الثالثة مراجعة الفقرات الاختبارية وتصديقها، حيث تم القيام بمراجعة منطقية ومراجعة إمبريقية للفقرات، وأعطيت أدلة على ثبات الاختبار، وصدقه، وتقدير فاعلية فقراته. وعليه، يمكن القول إن الاختبار، الذي يتم بناؤه باستخدام هذه الاستراتيجية المقترحة، يمكن أن يكون صالحاً لتشخيص مواطن القوة والضعف في المهارة التي يقيسها. وفي ضوء ذلك، يوصي الباحثان باستخدام هذه الاستراتيجية في مراكز القياس والتقويم في الجامعات، ومؤسسات التعليم المتعددة لبناء اختبارات تشخيصية محكية المرجع في موضوعات، ومواد أخرى؛ للتحقق من فاعلية هذه الاستراتيجية، بالإضافة إلى تطبيق الاختبارات التي يتم بناؤها وفق هذه الاستراتيجية على الفئات المستهدفة، ودراسة قدرتها على التشخيص.

المراجع

أبو ريدة، توفيق (١٩٩٣). **تشخيص مواطن ضعف طلبة الصفوف الابتدائية الأربعة الأولى في المفاهيم والمهارات الأساسية في مادة الرياضيات بالمدارس الأردنية**. رسالة ماجستير غير منشورة، الجامعة الأردنية، عمان، الأردن.

أبو لبدة، خطاب (٢٠٠٣). **الأخطاء الرياضية عند االطلبة الأردنيين في الدراسة الدولية الثالثة للرياضيات والعلوم–إعادة**. عمان: المركز الوطني لتنمية الموارد البشرية.

أحمد، مازن (٢٠٠٤). **علاقة جنس طالب الصف السادس الأساسي باكتساب المفاهيم والمهارات الجبرية والمهارات الحسابية الأساسية في محافظة جنين**. رسالة ماجستير غير منشورة، جامعة النجاح، نابلس.

أحمد، هبة (١٩٩٥). **تشخيص مواطن الضعف والقوة لدى طلبة الصف السادس والسابع والثامن في وحدة المعادلات**. رسالة ماجستير غير منشورة، الجامعة الأردنية، عمان، الأردن.

الجلبي، سوسن (٢٠٠٥). **أساسيات بناء الاختبارات والمقاييس النفسية والتربوية**، (ط١). دمشق: مؤسسة علاء الدين للطباعة.

الحموري هند وأحمد الكحلوت، (١٩٩٩). تطور مفهوم الكسر لدى طلبة الصفوف من الرابع إلى الثامن. **دراسات. العلوم التربوية، ٢٦**(١)، ١٦٥-١٨٦

الرفيع، أحمد؛ سكاف، أنطون؛ أبو لبدة، خطاب؛ الخضري، سليمان؛ ساسي، محمد؛ مطر، محمد (٢٠٠٧). **نتائج الدول العربية المشاركة في الدراسة الدولية لتوجهات مستويات التحصيل في الرياضيات والعلوم "TIMSS 2003"**. عمان: النول الدولية للدعاية والإعلان.

الرواشدة، إبراهيم؛ والثوابية، أحمد؛ والغرايبة، أحمد؛ وإسماعيل، رابح؛ وأبو أصفر، رزق؛ والرقب، سعيد؛ ومهيدات، عبد الحكيم؛ والديك، محمود؛ والعمري، يوسف (٢٠٠٠). **التقرير الإحصائي للاختبار الوطني لضبط نوعية التعليم للمرحلة الأساسية للعام الدراسي ١٩٩٩/٢٠٠٠**. الأردن: وزارة التربية والتعليم.

الريماوي، هالة (١٩٩٠). **تشخيص الأداء الرياضي لدى طلبة الصفوف الإعدادية في اختبار متعدد المستويات**. رسالة ماجستير غير منشورة، الجامعة الأردنية، عمان، الأردن.

الشريفين، نضال (٢٠٠٦). الخصائص السيكومترية لاختبار محكي المرجع في القياس والتقويم التربوي وفق النظرية الحديثة في القياس النفسي والتربوي. **مجلة العلوم التربوية والنفسية، ٧**(٤)، ٧٩-١٠٩.

الصيداوي، أحمد (٢٠٠٤). **التقويم التربوي المستقبلي: من التشخيصي إلى التكويني إلى الأدائي إلى الحقيقي**. بيروت: مكتب اليونسكو الإقليمي للتربية في الدول العربية.

العابدي، شرحبيل (٢٠٠٣). **تشخيص صعوبات التعلم في الهندسة لدى طلبة الصف الثامن الأساسي في حل المسائل الهندسية وطرق علاجها من وجهة نظر مدرسي الرياضيات في تربية عمان الأولى**. رسالة ماجستير غير منشورة، جامعة عمان العربية للدراسات العليا، عمان.

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

عنابي، حنان (١٩٩٩). **مستوى إتقان طلبة الصف الرابع الأساسي في الأردن للكفايات الأساسية في مبحث الرياضيات-تقرير رقم ١**. عمان: المركز الوطني لتنمية الموارد البشرية.

الكحلوت، أحمد وهند الحموري، (١٩٩٩). مدى إتقان طلبة صفوف الرابع إلى السادس في محافظة/العاصمة عمان مفهوم الكسر. **دراسات. العلوم التربوية، ٢٦**(٢)، ٤٥٤-٤٦٩.

وزارة التربية والتعليم (٢٠٠٦). الخطة الاستراتيجية لوزارة التربية والتعليم خلال الفترة ٢٠٠٦/٢٠٠٠. المملكة الأردنية الهاشمية: **منشورات إدارة البحث والتطوير التربوي**.

Berk, R. (1982). **Criterion-referenced measurement: The state of the art.** London: The Johns Hopkins University Press.

Crocker, L. & Algina, J. (1986). **Introduction to classical and modern test theory.** New York: Holt, Rinehart and Winston, Inc.

Feldt L.& Bernan R. (1993). Reliability. In Linn (ed.). **Educational measurement.** American Council on Education, Series on Higher Education. ORYX Press. P140.

Gregory, R. (2007). **Psychological testing, history, principles, and applications,** (5th ed). Pearson Education, Inc. USA.

Hambleton, R & Sireci, S. (1997). Future directions for norm-referenced and criterion-referenced achievement testing, **International Journal of Research, Educational Testing and Assessment: Lessons From the Past Directions for the Future, 27**(5), 379-393

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

Hambleton, R.K. (1982). Advances in criterion-referenced testing technology. C.R. Reynolds & T.B. Gutkin (Eds.), **The handbook of school psychology,** 351-379. NY: Wiley, pp.

Harris, C. & Pearlman, A. (1977). Conventional significance tests and indices of agreement or association. In C. Harris, A. Pearlman, and R. Wilcox (Eds.). **Achievement test items-Methods of study.** Los Angeles: Center for the study of evaluation, University of California.

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

Messick, S. (1993). Validity. In Linn (ed.). **Educational Measurement.** American Council on Education, Series on Higher Education. ORYX Press.

Murphy, K. & Davidshofer, C. (2001). Psychological Testing: Principles and Applications. New Jersey: Prentice Hall.

Nandakumar, R. (1994). Assessing dimensionality of a set of item responses-Comparison of different approaches. **Journal of Educational Measurement. 31,** 17-35.

Olagunju. A. (2004). Automatic generation and content validation of mathematics achievement test items. **National Tech Prep Network Conference,** Minneapolis, 10/15/2004.

Penfield, R. (2003). A score method of constructing asymmetric confidence intervals for the mean of a rating scale item. **Psychological Methods, 8,** 109-163.

Penfield, R., Miller, J. (2004). Improving validation studies using an asymmetric confidence interval for the mean of experts ratings. **Applied Measurement in Education, 17**(4), 359-370.

Popham, W (1982) Domain specification strategies. In Berk (ed). **Criterion-referenced measurement.** The State of the Art, London: The Johns Hopkins University Press.

Popham, W. (1981). **Modern educational measurement.** N.J, Prentice-Hall, Inc., Englewood Cliffs.

Roid, G. & Haladyna, T. (1982). **A technology for test-item writing.** New York: Academic Press.

Shultz, K., & Whitney, D. (2005). **Measurement theory in action: Case studies and exercises.** California: Sage Publications Inc.

Walbesser, H. (1968). **An evaluation model and its application.** Washington, D. C.: American Association for the Advancement of Science.