

Psychometric Characteristics of the Central Mathematics Test for the First Intermediate Grade in Al-Madinah Almunawarah's schools

Ibrahim Bin Abdul-Rahman Al-Harbi

College of Education || Taibah University || KSA

Abstract: The present study aimed at identifying the psychometric characteristics of the central mathematics test For the first grade, which was applied during the second semester of 1438/1439 AH, The study covered objective questions of (40) multiple choice clauses and (10) true or false questions, in order to achieve the objectives of the study, (372) students were selected among (12484) students, After being introduced and analyzed through the (SPSS) program, Version 25, the study found a number of results as the stability coefficient of the Alpha Cronbach (0.89) which means that it is very high, There was found that the validity of the test was very high and that was supported by strong indicators. In view of the characteristics of the questions, the difficulty coefficient of the statements of the questions ranged between (0.30- 0.82), Their ability to discrimination factor is very acceptable, with most of them ranging from (0.20 – 0.58), The distractors were very effective in attracting student attention, which emphasizes the quality of the psychometric properties of the statements of the questions. Based on the results of this study, it is shown that the quality of the psychometric characteristics of this test is very high. The researcher recommended expanding the application of the test optimization project (central tests) to include all the subjects in the intermediate stage, taking into consideration the availability of indicators of the validity of the content, And training those in charge of this project and teachers in general on the skill of formulating highly effective vocabulary in multiple choice questions, and the researcher recommends that teachers use the materials more to build a question bank, which can be used later in preparing pictures equivalent to the tests, Which was not addressed in this study, and training those in charge of this project and teachers in general on the skill of formulating highly effective vocabulary in multiple choice questions, and the researcher recommends that teachers use the materials more to build a question bank, which can be used later in preparing pictures equivalent to the tests.

Keywords: psychometric characteristics, test.

الخصائص السيكومترية لاختبار الرياضيات المركزي للصف الأول المتوسط بمنطقة المدينة المنورة

إبراهيم بن عبد الرحمن بن نجيم الحربي

كلية التربية || جامعة طيبة || المملكة العربية السعودية

المخلص: هدفت الدراسة الحالية إلى التعرف على الخصائص السيكومترية لاختبار الرياضيات المركزي للصف الأول متوسط، والذي تم تطبيقه خلال الفصل الدراسي الثاني للعام 1439/1438 هـ، وشملت الدراسة على الأسئلة الموضوعية (40) فقرة خيار من متعدد و(10) فقرات أسئلة (الصحيح والخطأ)، ولتحقيق أهداف الدراسة تم جمع نماذج إجابات (372) طالباً، تم اختيارهم بطريقة عشوائية من إجمالي (12484) طالباً. وبعد إدخالها وتحليلها عن طريق برنامج (SPSS) الإصدار (25)، توصلت الدراسة إلى مجموعة من النتائج أبرزها ما يلي: ارتفاع معامل الثبات المقدر بمعادلة ألفا كرونباخ حيث بلغ (0.89)، توافر مؤشرات كبيرة تدل على صدق البناء للاختبار،

وبالنظر إلى خصائص الفقرات تراوحت صعوبة الفقرات بين (0.30- 0.82)، وكانت قدرتها على التمييز مقبولة حيث تراوحت معظمها بين (0.20- 0.58)، وأما المشتتات فقد كانت ذو فعالية في جذب انتباه الطلاب مما يؤكد على جودة الخصائص السيكومترية للفقرات. وبناء على ما أظهرت هذه الدراسة من نتائج تبرز جودة الخصائص السيكومترية لهذا الاختبار، فقد أوصى الباحث بالتوسع في تطبيق مشروع تجويد الاختبارات (الاختبارات المركزية) ليشمل جميع المواد الدراسية في المرحلة المتوسطة مع الأخذ بالاعتبار التأكد من توفر مؤشرات صدق المحتوى الذي لم تتطرق له هذه الدراسة، وتدريب القائمين على هذا المشروع والمعلمين بشكل عام على مهارة صياغة مفردات ذات فعالية عالية في أسئلة الاختبار من متعدد، كما يوصي الباحث بالاستعانة بمعلمين المواد بشكل أكبر لبناء بنك أسئلة، يتم الاستفادة منه لاحقاً في أعداد صور مكافئة للاختبارات.

الكلمات المفتاحية: الخصائص السيكومترية، الاختبار، القياس.

مقدمة:

تعتمد عمليتي القياس والتقويم التربوي على أدوات ومقاييس تقدم بيانات كمية تتيح للتربويين فهم الظواهر التربوية بشكل دقيق ومحدد، وتعتبر الاختبارات والمقاييس التربوية جزءاً هاماً من أجزاء العملية التربوية، فمن خلالها يستطيع المعلم الحكم على مدى تحقق أهداف البرامج التربوية التعليمية التي يقوم بتدريسها، كما أنها تفيد المتعلم في التعرف على مستوى تحصيله الدراسي من خلال أدائه في الاختبار.

كما تشكل الاختبارات التربوية أدوات قياس أساسية، تستخدم لتقويم الطلبة للحصول على معلومات يمكن الاعتماد عليها في التوصل إلى أحكام وقرارات متعلقة بتقويم تحصيل الطلبة واستعداداتهم وتشخيص مواطن الضعف والقوة عندهم أو انتقاء الطلبة وتصنيفهم وإرشادهم إلى أنواع التعليم المختلفة ووضعهم في المسار التربوي المناسب وفي التوظيف والتنبيؤ وتحديد مستويات الطلبة (الجلالي، 2011).

وتلعب الاختبارات بأنواعها المختلفة دوراً مهماً في تقدم حضارات الكثير من الأمم، فكثير من القرارات المهمة التي لها تأثير كبير ومباشر في حياة أفراد المجتمع العلمية والعملية، يتم اتخاذها بناء على نتائج اختبارات معينة (دودين، 2006).

ولعل أول تجربة تاريخية للاختبارات هي التجربة الصينية التي قام بها الصينيون منذ (2200) ما قبل الميلاد، حيث كانوا أول من ابتكر الاختبارات وطبقوها عملياً في اختيار موظفي الدولة. (Gregory، 2004).

وفي وقتنا الحاضر يزداد استخدام الاختبارات في أمريكا في اتخاذ قرارات عالية المخاطر على مستوى كل ولاية، حيث تحرص كل ولاية على تطوير وتحسين اختباراتهم بشكل مستمر، حتى تكون نتائجها أكثر مصداقية. (Livingston، Reynolds، 2012).

ونظراً لأهمية القرارات المترتبة على نتائج الاختبارات، فإنه لا بُد من بناء اختبارات تتصف بالدقة والموضوعية في القياس، بمعنى توفر عناصر الفاعلية للاختبارات في إعطاء بيانات يمكن الوثوق بها في التوصل إلى أحكام واستنتاجات وقرارات صحيحة لها ما يبررها، وعناصر فعالية الاختبار تكون في جانبين مترابطين، يتصل الأول بخصائص أو دلالات تصف الاختبار بمجمله، وتعبر عن درجة الثقة في الدرجات أو البيانات المتحققة من تطبيقه، أما الثاني فيتصل بدلالات الفعالية عن مكونات الاختبار أي فقراته ومدى إسهامها في فعالية الاختبار بمجمله، وعناصر فعالية الاختبار تتمثل في الصدق والثبات، وعناصر فعالية الفقرات تتمثل في معاملات الصعوبة والتمييز والتخمين (الكيلاني وآخرون، 2009).

ومازال النوع الشائع استخدامه في أغلب أنظمة التعليم يكاد ينحصر في الاختبارات المدرسية، فإنه من المهم التركيز على الرفع من جودة الاختبارات المستخدمة والتي لا يتمتع معظمها بقدر مناسب من السمات الفنية الضرورية كالصدق والثبات (الدوسري، 2000).

ولهذا تسعى وزارة التعليم إلى تطوير أنظمة الاختبار لتناسب الوظيفة التقويمية التي وضعت من أجلها، ولتحقيق ذلك تتبن العديد من الخطط التي تحاول التغلب على المشكلات المتعلقة بالاختبارات، ومن أبرز أولويات تلك الخطط: تطوير البنية الأساسية للأجهزة والإدارات المختصة بالتقويم ودعمها بالكوادر المؤهلة، وتطوير النظم والإجراءات الفنية والعملية المتبعة في إعداد وبناء الاختبارات وتصحيحها والاستفادة من نتائجها. ولسد الاحتياج من الاختبارات القادرة على تحقيق الهدف التي وضعت من أجله تنفذ وزارة التعليم العديد من المبادرات في ذلك ولعل من أهم المشروعات هو تطبيق ما يسمى بـ(الاختبارات المركزية) التي تهدف إلى تجويد مخرجات التحصيل الدراسي في التعليم العام، وهي اختبارات تحصيلية تكون بديلة للاختبارات التي يعدها المعلمون لطلابهم في نهاية الفصل الدراسي، وذلك لرفع مستوى تحصيل الطلاب، من خلال دور المعلم في شمول كامل وحدات المقرر الدراسي ودعم تكامل الدور التعليمي والإشرافي في تحسين عمليات التدريس والتقويم للمقررات الدراسية (الإدارة العامة للاختبارات والقبول، 1439).

ومن باب تكامل الأدوار وإسهاماً في تحسين العملية التربوية، جاءت هذه الدراسة لتقدير جودة الخصائص السيكومترية لاختبار الرياضيات المركزي للصف الأول متوسط، الذي تم تطبيقه في إدارة تعليم المدينة المنورة للعام 1439/1438 هـ.

مشكلة الدراسة وتساؤلاتها:

تلعب الاختبارات المدرسية دوراً أساسياً في حياة الطالب العلمية في جميع مراحل التعليم، وبناءً على نتائجها يتحدد مستقبل الطالب وتتخذ بحقه الكثير من القرارات، ومازالت الاختبارات الوسيلة الأساسية في تقويم العملية التربوية، للتعرف على ما حققه الطالب من أهداف تعليمية، والاختبارات تُعد كأداة من أدوات التقويم التربوي الأقدم والأكثر شيوعاً وقبولاً عند أولياء الأمور (أبو فودة ويونس، 2012).

وقد أجرى النفيعي (2016) دراسة هدفت إلى التعرف على مدى إتقان أعضاء مشروع الاختبارات التحصيلية لمهارات بناء الاختبارات، وذلك على مستوى أعضاء المشروع ككل في جميع المناطق التعليمية والبالغ عددهم (146) عضواً، وتم عمل اختبار تشخيصي محكي المرجع، يقيس مهارات بناء الاختبارات التحصيلية، وهو عبارة عن (6) أبعاد رئيسية تقيسها (60) مفردة، وكانت النتائج تُبين تدني مستوى أعضاء المشروع لمهارات بناء الاختبارات التحصيلية وذلك على مستوى المهارات ككل وعلى مستوى كل بُعد من الأبعاد، حيث كان عدد أعضاء المشروع الذين بلغوا مستوى الإتقان في مهارات بناء الاختبارات التحصيلية (11) عضواً، وبنسبة (7.53%) من المجموع الكلي للأعضاء، وكان ترتيب منطقة المدينة المنورة التعليمية السابع من بين المناطق العشرة.

كما أجرى ستملر (stemler، 2007) دراسة في الولايات المتحدة الأمريكية هدفت إلى الكشف عن المهارات التي يمتلكها مشرفو المرحلة الأساسية في بناء الاختبارات التحصيلية، وقدرتهم على نقل المهارة إلى المعلمين، تكونت عينة الدراسة من (77) مشرفاً ومشرفة ويعملون في المدارس الحكومية العامة الأساسية، وتم اختيارهم بناء على حضورهم لدورات تدريبه في إعداد الاختبارات وتدريب المعلمين، وأظهرت نتائج الدراسة أن معرفة المشرفين بمبادئ بناء الاختبارات كانت متوسطة وخصوصاً في مجالي إعداد جدول المواصفات وانتقاء المادة التعليمية.

وبالنظر إلى الاختبارات المركزية التي يطبقها مشروع تجويد الاختبارات في منطقة المدينة المنورة فهي مسؤولة عن مصير ما يزيد عن 12 ألف طالب من طلاب المرحلة المتوسطة وهم يمثلون جميع طلاب الصف الأول متوسط في منطقة المدينة المنورة (قسم الاختبارات والقبول بإدارة تعليم المدينة)، وكون هذا المشروع حديث (تجويد الاختبارات).

ومن خلال خبرة الباحث وعمله كمعلم في وزارة التعليم فقد قام بعمل مقابلات مع المسؤولين في أقسام الاختبارات والقبول في إدارات التعليم، فتبين عدم وجود دراسات حديثة تنطبق لهذه الاختبارات فكان لابد من وجود دراسات تقييمية وتحليلية لهذه الاختبارات للتأكد من خصائصها السيكومترية، حتى يتم في المستقبل تلافي جوانب القصور " إن وجدت" ومن ثم تعميم هذه الدراسة كنموذج عملي لبقية الاختبارات المركزية المطبقة في جميع إدارات التعليم بالمملكة العربية السعودية.

أسئلة الدراسة:

وعلى ضوء ما سبق، فإنه يمكن تحديد مشكلة الدراسة الحالية في الأسئلة الآتية:

- 1- ما قيمة معامل الثبات للاختبار المركزي لمقرر الرياضيات في المرحلة المتوسطة؟
- 2- هل الاختبار المركزي للرياضيات يُميز بين متوسط درجات الفئة العليا والفئة الدنيا من الطلاب؟
- 3- هل توجد علاقة ارتباطية ذات دلالة إحصائية بين درجات كل سؤال ودرجة الاختبار ككل كمؤشر على صدق الاختبار؟
- 4- ما جودة معاملات الصعوبة لكل فقرة من فقرات الاختبار المركزي؟
- 5- ما جودة معاملات التمييز لكل فقرة من فقرات الاختبار المركزي؟
- 6- ما مدى فاعلية المشتتات (البدايل الخاطئة) لكل فقرة من فقرات الاختبار من متعدد في الاختبار المركزي؟

أهداف الدراسة:

1. تقدير الخصائص السيكومترية للاختبارات المركزية التي تُعدها الإدارة العامة للتعليم بمنطقة المدينة من خلال حساب كلاً من الصدق والثبات لها.
2. تحليل بنود فقرات الاختبارات المركزية من حيث معاملات التمييز والصعوبة لكل فقرة من فقرات الاختبار.
3. تقدير فعالية المشتتات (البدايل الخاطئة) المستخدمة في كل فقرة من فقرات (الاختبار من متعدد) في الاختبارات المركزية.

أهمية الدراسة:

- 1- تقدم للمسؤولين في وزارة التعليم بشكل عام، وفي منطقة المدينة المنورة بشكل خاص معلومات تفصيلية عن جودة بناء الاختبارات المركزية، بحيث يمكن اتخاذ قرارات تصحيح أو تعديل للألية التي يتم من خلال إعداد وتطبيق هذا الاختبار.
- 2- تستخدم أساليب إحصائية لتحليل بنود الاختبارات لا تستخدم كثيراً في الدراسات الإحصائية المماثلة وخاصةً فيما يتعلق بالأساليب المستخدمة للتعرف على مؤشرات توافر الصدق التجريبي " الإحصائي " Empirical Validity " (مجيد، 2014).
- 3- تحسين وتطوير العملية التعليمية من خلال المساهمة في تحسين جودة الاختبارات المركزية في المرحلة المتوسطة.

مصطلحات الدراسة:

الخصائص السيكومترية للاختبار:

هي تلك الخصائص الضرورية والمتعلقة بالصدق والثبات والتي يتم حسابها بعد تجريب الاختبار على عينة ممثلة للمجتمع الذي طبق عليه الاختبار، وتعتمد جودة الاختبار على مدى توافر بيانات مناسبة لهذه الخصائص (مراد وسليمان، 2012).

المشروع العام لتجويد الاختبارات (الاختبارات المركزية):

هو مشروع وزارى تم تنفيذه في اختبارات نهاية الفصل الدراسي الثاني في المرحلة المتوسطة للعام 1438/1439 هـ، في كل إدارات التعليم بالمملكة ماعدا إدارات الحد الجنوبي، حيث تقوم لجنة برئاسة مدير التعليم بالمنطقة ومكونة من مجموعة من المعلمين المتميزين وعدد من المشرفين التربويين المتخصصين، بإعداد أسئلة مادة تحريرية واحدة في كل صف من صفوف المرحلة المتوسطة، وموحدة للبنين والبنات، وتكون بديلة عن أسئلة المعلمين التي يضعونها في نهاية الفصل الدراسي، سعياً من الوزارة ممثلة في الادارة العامة للاختبارات والقبول في تجويد الاختبارات ورفع مستوى تحصيل الطلاب من خلال دور المعلم في شمول كامل وحدات المقرر الدراسي، وكانت مادة الرياضيات في الصف الأول متوسط هي المادة التي تم اختيارها لذلك العام، ولهذا قام الباحث بدراستها.

حدود الدراسة:

- الحدود الموضوعية: حُددت هذه الدراسة بالموضوع التي تناولته وهو دراسة الخصائص السيكومترية لاختبار الرياضيات المركزي للصف الأول متوسط (الفصل الدراسي الثاني) للعام 1438/1439 هـ من خلال اجابات الطلاب على الأسئلة الموضوعية التي شملت الاختيار من متعدد (40 فقرة) وأسئلة الصح والخطأ (10 فقرات).
- الحدود المكانية: جميع المدارس المتوسطة للبنين في إدارة التعليم بمنطقة المدينة المنورة.

2. الإطار النظري والدراسات السابقة

أولاً- الإطار النظري:

تعريف الاختبار التحصيلي:

عرفه (Lehman, Mehrens, 1975) بأنه اختبار يقيس مدى تحصيل الفرد لشيء من الأشياء أو اكتسابه لمعرفة محددة أو مهارات معينة وفي الغالب تحدث هذه الأشياء نتيجة تدريب وتعليم مخطط ومنظم. وقد أشار علام (2012) أن مفهوم التحصيل يستخدم للإشارة إلى درجة أو مستوى النجاح الذي يحققه الطالب في أي مجال دراسي. و بالنظر إلى (Sax, 1980) فقد عرف الاختبار التحصيلي بأنه اختبار صمم ووضع خصيصاً لقياس التعلم المدرسي أو التعلم النظامي. ومن التعريفات الشهيرة في الاختبار التحصيلي تعريف (Ebel, 1979) حيث قال أن الاختبار التحصيلي هو اختبار صمم بقصد قياس إلمام الطالب في معرفة محددة أو براعته في مهارات معينة.

الخصائص السيكومترية للاختبارات:

إن الاختبار أو المقياس ينبغي أن تتوفر فيه بعض الخصائص السيكومترية الأساسية ومن أهمها ثبات درجاته وصدقه (مجيد، 2014)

الثبات Reliability

نشأ مفهوم الثبات نتيجة لظهور أعداد كبيرة من الأدوات والاختبارات التي تقيس مختلف نواحي السلوك الإنساني، وذلك بعد الحرب العالمية الثانية، حيث أصبح لكل مفهوم أو مصطلح عدد كبير من الاختبارات التي تقيسه، والسؤال الذي طرح نفسه هو أي من هذه الاختبارات أجود وأفضل من حيث دقته في قياس الظاهرة التي يراد قياسها، وهذا ما جعل شركات القياس تتجه إلى إيجاد المعايير الأساسية التي يحتكم إليها لتحديد جودة الاختبار، وكان من أهم المعايير والتقنيات مفهوم الثبات، حيث بدأ علماء القياس في حسابه بطرائق عديدة تقوم الفكرة الأساسية لمعظمها على معامل الارتباط (مجيد، 2014: 121).

ويشير مفهوم الثبات بالمعنى العام للدلالة على مدى اعتماد الفروق في درجات الاختبار على الفروق الحقيقية في السمة أو الخاصية المقاسة من جهة وأخطاء القياس العشوائية أو الناتجة عن الصدفة من جهة أخرى. فالثبات يعني درجة بقاء نتائج الطلبة مستقرة، أو ثابتة، عبر عدة مرات من التطبيق، بمعنى حصول الطالب على الدرجة نفسها إذا أدى الاختبار في وقتين مختلفتين، أو صححت من قبل عدة مصححين، أو عندما تكون هناك صور متكافئة Nitko، Brookhart، (2018).

ويمكن تعريف الثبات بأنه الجزء من تباين درجة الاختبار، الذي يُعزى إلى الفروق الفردية في الدرجة الحقيقية (علام، 2013 ص 168).

الصدق The Validity:

يُعد صدق الاختبار الخاصية الأكثر أهمية بين خصائص الاختبار الجيد، فالثبات شرط ضروري للاختبار، ولكنه ليس مؤشراً كافياً عن صدق الاختبار، ويكون الاختبار صادقاً إذا كان يقيس ما وضع لقياسه، أي إذا حقق الغرض الذي صمم من أجله (مراد، سليمان، 2012).

إن المفاهيم المجردة مثل الذكاء، القلق، يمكن أن تقاس من خلال مجموعة من المؤشرات تعبر عن كل منها، ويكون الاختبار صادقاً كلما كانت مؤشرات (مفرداته) تعبر عنه، وتعكس بدقة المفاهيم التي وضع الاختبار من أجل قياسها (عمر وآخرون، 2010: 189).

العلاقة بين الثبات والصدق

قبل أن نبحث في العلاقة بين الصدق والثبات يجب أن ندرك أن هناك علاقة أكيدة بين صدق الاختبار وثباته، وخاصة أن كلا المفهومين يبحثان في مدى كفاءة الاختبار ومناسبة للمسلمات الرئيسية لنظرية القياس. فمفهوم الثبات يتعلق بمدى استقرار درجات الاختبار، بينما الصدق يتعلق بالهدف والغرض الذي بني من أجله الاختبار، فالثبات يتأثر بالأخطاء العشوائية وأما الصدق يتأثر بجميع الأخطاء، ولا بد أن نلاحظ أن المقياس الصادق لا بد وبالضرورة أن يكون ثابتاً، بينما قد نجد مقياس ثابتاً، ولكنه غير صادق، فمثلاً اختبار ذكاء للأطفال بالرغم من أنه يتمتع بدرجة عالية من الثبات إلا أنه في حال استخدامه لقياس ذكاء الكبار، فلن يكون صادقاً في النتائج المستخرجة، ولا يستطيع أن يعطي دلالات وتفسيرات صحيحة عن ذكاء الكبار.

لماذا نحلل فقرات الاختبار:

يقصد بتحليل فقرات الاختبار استخراج معاميل الصعوبة، ومعاميل التمييز، وتحديد فاعلية " المموهات " في فقرات الاختبار من متعدد.

ويعتمد ثبات وصدق الاختبار على جودة المفردات في الاختبار، فإذا استطعت تحسين جودة كل مفردة، فسوف تُحسن جودة الاختبار ككل (Livingston، Reynolds، 2012).

وقد ذكر كلاً من (أبو فودة ويونس، 2012) و(الخياط، 2010)، مجموعة من الفوائد لهذه الاجراءات، ومنها:

1- من خلال تحليل فقرات الاختبار يستطيع المعلم أن يحدد مواطن القوة والضعف عند طلابه، وتشخيص الأسباب سواء كانت فنية في الاختبار أو تربوية في طريقة التدريس، فمثلاً إذا وجد المعلم أن أكثر الأخطاء في الفقرات التي تقيس التطبيق، فربما كان ذلك عائداً أن المعلم لم يكن يعطيهم الوقت الكافي للتدريب وحل التمارين.

2- من خلال تحليل الفقرات، يستطيع المعلم أن يحصل على فقرات جيدة من حيث الصعوبة والتمييز، وبالتالي الاحتفاظ بها، والاستفادة منها في الاختبارات التحصيلية القادمة.

3- العثور على الفقرات الضعيفة، فنعمل على حذفها أو تحسينها.

4- إن التعرف على درجات الصعوبة يحدد مواقع الفقرات في الاختبار، فإذا كانت الفقرة درجة صعوبتها عالية (أي الفقرة سهلة)، تأتي هذه الفقرة في أول الاختبار، والعكس إذا كانت درجة صعوبتها منخفضة (أي الفقرة صعبة).

5- أن تحليل فقرات الاختبار يكشف أيضاً عن فعالية البدائل (المموهات)، فالبديل الذي لم يختاره أحد يعاد النظر فيه، وكذلك البديل الذي يتم اختياره من مرتفعي التحصيل أكثر من منخفضي التحصيل.

ثانياً- الدراسات السابقة:

تعددت الدراسات العلمية التي تتعلق بالاختبارات التحصيلية بشكل مباشر أو غير مباشر، حيث تناولت هذه الدراسات بحث المشكلات المتعلقة بالاختبارات من خلال زوايا متعددة، كالأنظمة وإدارة الاختبار، والارتباط بالأهداف التعليمية، والقدرة على قياس المستويات التعليمية المتعددة لبلوم، وفي هذه الدراسة سيتم عرض بعض الدراسات المتعلقة بجودة الاختبارات بشكل عام أو الدراسات التي تطرقت لبعض المعايير الشائعة عند دراسة الخصائص السيكومترية للاختبار. وكما نعلم أن إدارة الاختبار هي جزء مهم ومؤثر في الخصائص السيكومترية للاختبار وخاصة ما يتعلق بثبات نتائج الاختبار.

- فقد هدفت دراسة أحمد (2017)، إلى تقويم أداء مديري المدارس الثانوية في إدارة الامتحانات من وجهة نظر المدرسين، واستخدمت هذه الدراسة المنهج الوصفي التحليلي. تم تحديد مجتمع البحث الذي تكون من (3981) مدرساً ومدرسة، واختيرت عينة البحث بالطريقة الطبقيّة العشوائية، إذ بلغ عدد أفراد العينة (274) من مدرسي الثانوية. وتم بناء استبانة تضمنت (60) فقرة موزعة بين مجالات الدراسة. وتأكّدت الباحثة من صدقها وثباتها. حُللت بيانات البحث باستعمال برنامج (SPSS)، وقامت الباحثة بتطبيق الاستبانة في الفصل الأول من العام الدراسي 2015-2016، وتوصلت إلى نتائج أهمها: أن مستوى أداء مديري المدارس في إدارة الامتحانات كان مرتفعاً بنحوٍ عام ويُعبر عن مهارة ومقدرة.

- وأجرى Atalmis (2014) دراسة هدفت إلى فحص حالتين لفقرات الاختبار من متعدد تمثلت الحالة الأولى في دراسة أثر وجود البديل (لا شيء مما ذكر) بين بدائل الفقرات، في حين بحثت الحالة الثانية دراسة العدد

الأنسب لبدائل فقرات الاختيار من متعدد. تم اختبار هاتين الحالتين تجريبياً باستخدام فقرات تحتاج إلى خطوة واحدة للحل وفقرات أخرى تحتاج لأكثر من خطوة واحدة للحل وبيان أثرها على الخصائص السيكمومترية لفقرات الاختبار (الصعوبة والتمييز) وأثرها على ثبات الاختبار. تكونت عينة الدراسة من (1500) طالب في الولايات المتحدة وتركيا، لطلبة الصفين السابع والثامن، وتمثلت خطوات الدراسة بثلاث نماذج وهي نماذج الاختيار من متعدد المكون من أربعة بدائل، والنموذج الثاني اختيار من متعدد من ثلاثة بدائل، والنموذج الثالث اختيار من متعدد من ثلاثة بدائل قسمت فقرات الاختبار إلى قسمين: قسم يحل بخطوة واحدة، والقسم الثاني يحتاج إلى أكثر من خطوة للحل. حللت النتائج باستخدام نظرية استجابة الفقرة، وبينت النتائج: أن نوع النموذج المستخدم وعدد الخطوات لا تؤثر في معامل الصعوبة ومعامل التمييز والثبات للاختبار، ورجح الباحث هذه النتيجة إلى صغر حجم عينة الفقرات، مع وجود فروق غير دالة في معامل الصعوبة لنموذج (لشيء مما ذكر) حيث كان أكبر من معامل الصعوبة لباقي النماذج في الفقرات التي تتطلب خطوة واحدة للحل لكن بفرق غير دال إحصائياً.

- وهدفت دراسة الشريف (2013) إلى تقصي الخصائص السيكمومترية لاختبار تحصيلي في الرياضيات بصورتيه (الورقية، المحوسبة)، لطلبة الصفوف الأولى والثاني والثالث من المرحلة المتوسطة في المملكة العربية السعودية، ولتحقيق هدف الدراسة، تم تطبيق ثلاثة اختبارات تحصيلية بصورتها (الورقية، المحوسبة) كل اختبار مؤلف من (30) فقرة من نوع الاختبار من متعدد من أربعة بدائل، لقياس مهارات حل المسألة الرياضية، والمفاهيم العلمية المتضمنة في وحدات الإحصاء والاحتمالات، وذلك في العام الدراسي (2011/ 2012)، وتكونت عينة الدراسة من (958) طالباً من طلبة صفوف الأولى والثاني والثالث من المرحلة المتوسطة، تم اختيارهم بالطريقة العشوائية التطبيقية العنقودية من مدارس مركز الاشراف التربوي بالروابي التابعة لإدارة التربية والتعليم بالرياض في المملكة العربية السعودية، وأشارت نتائج الدراسة إلى وجود فرق ذو دلالة إحصائية عند مستوى الدالة (0.05) بين متوسطي معاملي صعوبة فقرات الاختبار التحصيلي بصورتيه الورقية والمحوسبة يعزى لصورة الاختبار المحوسب وذلك في كافة الصفوف.

- وفي دراسة أبكر (2012) هدفت هذه الدراسة للتعرف على مدى مطابقة الاختبار الموحد لمادة الرياضيات في الحلقة الثانية من التعليم الأساسي بولاية الخرطوم للمواصفات العلمية، وذلك بتحليل وتقويم أسئلة اختبار العام الدراسي (2011/2012) استخدم الباحث المنهج الوصفي، وأخضع الاختبار الموحد لمادة الرياضيات لمصفوفة العالم ديفز وعزز النتائج بأراء عينة معلمي مادة الرياضيات يبلغ عددهم (60) معلماً بالاستبانة وبعد أن عولجت البيانات، كانت أهم النتائج أن نسب توزيع الأسئلة لم توافق مطابقته للمواصفات العلمية.

- وأجرى يعقوب وأبو فودة (2012) دراسة هدفت إلى معرفة أثر مخالفة قواعد صياغة فقرات الاختيار من متعدد في الخصائص السيكمومترية للفقرات (الصعوبة، والتمييز)، والخصائص السيكمومترية للاختبار (الثبات، والصدق). ولتحقيق ذلك تم بناء اختبار من نوع الاختيار من متعدد في مبحث الرياضيات، تكون من (40) فقرة ذات أربعة بدائل، ثم جرى صياغة فقرات الاختبار في نموذج آخر مخالف لقواعد الصياغة. جرى تطبيق نموذجي الاختبار على عينة مكونة من (300) طالب، بواقع (150) طالباً لكل نموذج، وفي ظروف صفيحة عادية. وبعد إجراءات التطبيق حسبت قيم معاملات صعوبة الفقرات وتمييزها في نموذجي الاختبار، وكذلك معاملات صدقها وثباتها. فأظهرت النتائج أن الفرق كان دالاً إحصائياً بين صعوبة وتمييز الفقرات التي لم تحتوي المخالفات في قواعد الصياغة، أو التي احتوتها، وبينت النتائج أن الفرق بين قيم معاملي ثبات الاتساق الداخلي في نموذجي الاختبار كان دالاً إحصائياً، وكان الفرق بين قيم معاملي صدقها دالاً أيضاً.

- ومن الدراسات ذات الصلة المباشرة بموضوع الدراسة، دراسة الزواهره (2010) التي هدفت إلى استخراج الخصائص السيكومترية لاختبار الرياضيات المستوى الثالث في بطارية اختبارات القدرات الواسعة (WRA) للصفوف الخامس والسادس والسابع في محافظة عجلون. وتكونت عينة الدراسة من (900) طالباً وطالبة من طلبة الصفوف الخامس والسادس والسابع في المدارس الحكومية في محافظة عجلون، ولتحقيق غرض الدراسة طبق المقياس على العينة الاستطلاعية لمعرفة وضوح التعليمات والفقرات وإيجاد معاملات الصدق المرتبط بالتحك مع علاماتهم في مادة الرياضيات في نهاية الفصل الدراسي الأول (2009/2010). وبلغت قيمة للمقياس ككل (0.415) وبلغت (0.25، 0.28، 0.57) بالترتيب للصفوف (الخامس والسادس والسابع)، كما أظهرت نتائج الدراسة أن هناك تدرج في معاملات الصعوبة للفقرات عبر الصفوف، حيث بلغت بين (0.53 - 0.67) للصف الخامس، و(0.55 - 0.84) للصف السادس، و(0.67 - 0.88) للصف السابع. كما أظهرت النتائج أن معاملات التمييز كانت تتراوح بين (0.31 - 0.57) للصف الخامس، و(0.26 - 0.50) للصف السادس، و(0.18 - 0.46) للصف السابع، وبذلك كانت جميعها موجبة وذات قدرة تمييز مقبولة، كم تم إيجاد معاملات الثبات من خلال الاتساق الداخلي باستخدام معادلة كرونباخ ألفا وكانت القيم للاختبارات بالترتيب (0.82 - 0.87 - 0.87) للصفوف الثلاث، وبذلك فالمقياس يتمتع بدرجة عالية من الثبات.
- وأجرت القضاة (2010) دراسة هدفت إلى بناء اختبار تحصيلي في مبحث الرياضيات للصف التاسع الاسامي (الفصل الدراسي الأول) وتقدير خصائصه السيكومترية وفق النموذج ثنائي المعلمة للنظرية الحديثة والنظرية الكلاسيكية، وتكونت مجموعة أفراد الدراسة من (1000) طالبة وطالب من طلبة الصف التاسع موزعين على مديرتين من المديرية التابعة لوزارة التربية والتعليم الاردنية، ثم كتابة (120) فقرة اختبارية من نوع الاختبار من متعدد، موزعة على ثلاث نماذج بواقع (40) فقرة لكل نموذج، وتم ربط النماذج بـ (10) فقرات رابطة، وبعد تحليل الاجابات وفق كل نظرية، تراوح متوسط معامل صعوبة الفقرات وفق النظرية الكلاسيكية ما بين (0.53 - 0.65) ومعامل التمييز (0.51 - 0.64) وتراوحت متوسط معلمة الصعوبة للفقرات وفق النظرية الحديثة ما بين (0.59 - 0.72) ومتوسط معلمة التمييز (1.10 - 1.16). وبين نتائج الدراسة أن عدد الفقرات التي تم اختيارها وفق النظرية الكلاسيكية كانت (112) فقرة، مقابل (116) فقرة انسجمت وفق نموذج النظرية الحديثة، كما دلت الدراسة على وجود درجة عالية من التوافق بين نموذج المعلمتين والنظرية الكلاسيكية في الحكم على الفقرات من حيث الابقاء عليها او استبعادها بناء على مؤشرات الصعوبة والتمييز.

التعليق على الدراسات السابقة:

- تختلف الدراسات السابقة عن بعضها البعض رغم تناولها جميعاً للاختبارات، فنجد أن دراسة أحمد (2017) ركزت على جانب الإدارة المدرسية لهذه الاختبارات، وكون الاختبار طبق داخل المدارس فلا بد الأخذ بالحسبان أن إدارة الاختبار هي جزء لا يتجزأ من جودة الاختبار، فهي مؤثر قوي على ثبات نتائج الاختبار، وقد أسفرت الدراسة إلى أداء مرتفع بنحو عام في إدارة المدرسة للاختبارات.
- ونجد أن بعض الدراسات تطرقت لجودة الخصائص السيكومترية للاختبار وهو ما يرتبط مباشرة بالدراسة كما في دراسة الزواهري (2010) عندما تطرق للخصائص السيكومترية لاختبار الرياضيات المستوى الثالث في بطاريات القدرات الواسعة (WRA) وتوصلت الدراسة إلى نتائج جيدة بخصوص معاملي الصعوبة والتمييز والثبات ككل للاختبار.

- وأما دراسة القضاة (2010) جاءت لتحسم الجدل المنتشر بخصوص جودة تقديرات النظرية التقليدية للاختبار مقارنة بالنظرية الحديثة للاختبار، فقلت توصلت أنه هناك توافق بدرجة عالية بين تقديرات نموذج ثنائي المعلمة في النظرية الحديثة وتقديرات النظرية التقليدية في الحكم على الفقرات بناء على مؤشرات الصعوبة والتمييز.
- وجاءت دراسة ابكر (2012) تؤكد على عدم مطابقة أسئلة الاختبار الموحدة في مادة الرياضيات في ولاية الخرطوم من وجهة نظر المعلمين لمواصفات الاختبار الجيد.
- وأما دراسة يعقوب وأبو جودة (2012) فقد تطرقت لأثر مخالفة فقرات الاختبار من متعدد على الخصائص السيكومترية للاختبار ولل فقرات، وقد أظهرت أن مخالفة الشروط الواجب توافرها في الأسئلة الاختبار من متعدد تؤثر على صدق وثبات الاختبار وكذلك على مؤشرات الصعوبة والتمييز للفقرات.
- وقد هدف دراسة Atalmis (2014) إلى معرفة تأثير نوعية (البدائل الخاطئة) وعدد البدائل الخاطئة على جودة الخصائص السيكومترية للاختبار، وأظهرت أن هناك تأثير ولكن لم يكن ذو دلالة إحصائية ويرجع السبب كما ذكر الباحث إلى صغر حجم العينة.

من خلال استعراض ما سبق يتبين لدى الباحث أن الدراسة الحالية تتميز بما يلي:

- الدراسة الأولى التي تطرقت للخصائص السيكومترية للاختبارات المركزية التي تم تعميمها على جميع إدارات التعليم بالمملكة العربية السعودية من خلال البحث في مصادر البيانات والتحدث مع رؤساء أقسام القبول والاختبارات، كون الاختبارات المركزية حديثة التطبيق.
- جميع الدراسات السابقة تطرقت إلى جانب من جوانب الخصائص السيكومترية للفقرات، فبعضها تطرق لمعالم الصعوبة والتمييز، والآخر إلى فعالية البدائل، والدراسة الحالية تجمع بين كل هذه الخصائص وهي الصعوبة والتمييز وفعالية البدائل.
- الدراسة الحالية تتطرق إلى ما يسمى بالصدق التمييز للاختبار ككل بمعنى مقدرة الاختبار على التمييز بين مرتفعي التحصيل ومنخفضي التحصيل للاختبار ككل وليس للفقرات فقط كما هو موجود بالدراسات السابقة.
- الدراسة الحالية تُعنى بدراسة واقع معمول به، وتناقشه من عدة زوايا إحصائياً باستخدام النظرية الكلاسيكية شائعة الاستخدام في الواقع.

3. منهجية الدراسة وإجراءاتها

منهجية الدراسة

استخدم الباحث المنهج المسحي (Survey Research) وهو أحد أشهر أنواع البحوث والدراسات الوصفية الذي يستخدم في المجالات التربوية، وقد عرفه إبراهيم وأبو زيد (2012، ص208) بأنه " محاولة بحثية منظمة لتقرير وتحليل ووصف الوضع الراهن لموضوع أو ظاهرة أو نظام أو جماعة، بهدف الوصول إلى معلومات وافية ودقيقة.

مجتمع الدراسة:

يقصد بالمجتمع (Population) المجموعة الكلية (Universal Set) من العناصر التي يسعى الباحث إلى أن يعمم عليها النتائج ذات العلاقة بالمشكلة المدروسة (ملحم، 2010)، وبالنسبة للدراسة الحالية فإن مجتمع الدراسة يتكون من الأسئلة المركزية للرياضيات للصف الأول متوسط بإدارة التعليم بمنطقة المدينة المنورة.

عينة الدراسة:

أسئلة الرياضيات المركزية التابعة لإدارة التعليم بالمدينة المنورة للعام 1438/1439 هـ الفصل الدراسي الثاني. وقام الباحث بجمع عينة من اجابات الطلاب البالغ عددهم (12484) طالباً، وكانت العينة عبارة عن (372) طالباً تم أخذها بالطريقة العنقودية العشوائية من خلال اختيار مدارس عشوائياً وأخذ جميع النماذج الموجودة لديهم، وقد استخدم الباحث في تحديد عدد النماذج معادلة Morgan، (1970).

$$N = \text{حجم المجتمع} = D = \text{نسبة الخطأ}$$

$$z = \text{الدرجة المعيارية المقابلة لمستوى الدلالة 0.95 وتساوي 1.96}$$

$$n = \frac{\left(\frac{z}{d}\right)^2 \times (0.50)^2}{1 + \frac{1}{N} \left[\left(\frac{z}{d}\right)^2 \times (0.50)^2 - 1\right]}$$

أداة الدراسة:

استخدم الباحث كراسة إجابة الطلاب على هذه الأسئلة من خلال المدارس.

4. عرض النتائج ومناقشتها:

هدفت الدراسة الحالية إلى التعرف على الخصائص السيكومترية لاختبار الرياضيات المركزي للصف الأول متوسط (الفصل الدراسي الثاني)، للعام 1438/1439 هـ، في منطقة المدينة المنورة. وفيما يلي عرض للنتائج ومناقشتها حسب تساؤلات الدراسة:

1- ما قيمة معامل الثبات للاختبار المركزي لمقرر الرياضيات في الصف الأول متوسط؟

يوضح الجدول التالي: معامل الثبات لاختبار الرياضيات المركزي للصف الأول متوسط (الفصل الدراسي الثاني) للعام 1438/1439 هـ، والمقدر بطريقة ألفا كرونباخ

جدول (1): معامل الثبات للاختبار المركزي المقدر بطريقة ألفا كرونباخ

ألفا كرونباخ	معامل الثبات
0.89	الاختبار المركزي

ويتضح من خلال الجدول أعلاه ارتفاع معامل الثبات للاختبار المركزي للرياضيات للصف الأول متوسط (الفصل الدراسي الثاني) للعام 1438/1439 هـ، ورغم عدم وجود قيمة ثابتة متفق عليها للتعبير عن ارتفاع أو انخفاض قيمة معامل الثبات، إلا أن قيمة معامل الثبات التي تبلغ أكثر من (0.80)، تُعد مقبولة في كثير من الاختبارات التحصيلية التي تطبق جماعياً أو فردياً، فمثلاً اختبار كاليفورنيا التحصيلي، وهو مجموعة من الاختبارات التحصيلية التي تُطبق جماعياً، وتستخدم عادة في المدارس العامة، وقيمة معاملات ثبات معظم درجاتها تزيد عن (0.80) (Livingston, Reynolds, 2012)، وكذلك اختبار متروبوليتان التحصيلية Metropolitan Achievement Test، فهي تستخدم هذه أيضاً في مدارس التعليم العام، وقيمة معاملات ثبات درجاتها أكثر من (0.80) (الجلالي، 2011).

ويمكن تفسير ارتفاع معامل الثبات في الاختبار المركزي للرياضيات عبر عدد من العوامل لعل أبرزها:

- مناسبة عدد فقرات الاختبار للزمن المخصص للإجابة.
- عدم تجانس مجموعة المفحوصين، حيث يزداد معامل الثبات كلما ازدادت تباين الأفراد (القرني وآخرون، 2012).
- موضوعية التصحيح، حيث تزداد معاملات الثبات في الأسئلة الموضوعية.

- مستوى الصعوبة المتوسط لمعظم الفقرات الاختبارية، حيث كلما زادت صعوبة الاختبار، كلما زاد التخمين العشوائي في الاجابات، مما يزيد الأخطاء العشوائية وهذا بدوره يقلل من ثبات درجات الاختبار (علام، 2012). ولأن معامل الثبات المُقدر للاختبار ذو قيمة مرتفعة، فإن ذلك يُمكن من الانتقال بثقة كبيرة إلى النتائج المتعلقة بالتساؤل الثاني للدراسة والمتعلق بأحد مؤشرات الصدق للاختبار وهو ما يسمى (الصدق التمييز)، حيث تمت صياغة التساؤل الثاني كما يلي:

2- هل الاختبار المركزي للرياضيات قادراً على التمييز بين متوسط درجات الفئة العليا والفئة الدنيا من الطلاب؟ بما أن مفهوم الصدق يرتبط بإيراد أكبر عدد من الشواهد التي تدل على قدرة الاختبار على تحقيق ما وضع لقياسه، ولأن الدراسة تهتم ببعض أنواع الصدق التي يمكن تقديرها بطرائق كمية، فإن الصدق التمييزي من الأنواع الشهيرة المستخدمة إحصائياً في تقدير مؤشر الصدق من خلال دراسة الفروق بين المجموعات، وللتعرف على دلالة الفروق الإحصائية بين متوسطي درجتي التحصيل في الاختبار لدى الفئة العليا والفئة الدنيا من الطلاب والإجابة على تساؤل الدراسة تم استخدام اختبار (ت) لعينتين مستقلتين (Independent-samples t- test)، بعد التأكد من افتراضاتها.

جدول (2) نتيجة تطبيق الاختبار الإحصائي

المجموعات	العدد	المتوسط	الانحراف المعياري	قيمة ت	مستوى الدلالة
مرتفعي التحصيل	100	42.03	3.778	44.492	0.000
منخفضي التحصيل	100	18.43	3.723		

بما أن القيمة المحسوبة ت تساوي (44.492)، بمستوى دلالة (0.000) وبمقارنتها مع مستوى الدلالة الإحصائية $\alpha = (0.05)$ ، نستنتج وجود فروقات جوهرية بين متوسطي درجات الفئة العليا والفئة الدنيا، وبالتالي فإن الاختبار له القدرة على التمييز بين مرتفعي التحصيل ومنخفضي التحصيل، وهو أحد مؤشرات صدق الاختبارات، وبما أن هناك دلائل ومؤشرات كثيرة تستخدم للإشارة على صدق الاختبارات، ولا يكتفي بدليل واحد، فإن من المنطقي التأكد من أدلة أخرى وهو ما يقودنا للانتقال للإجابة على التساؤل الثالث في الدراسة، وهو يعتبر أحد مؤشرات الصدق التكويني، ويسمى بصدق الاتساق الداخلي Internal Consistency وتمت صياغة التساؤل كما يلي:

3- هل توجد علاقة ارتباطية ذات دلالة إحصائية بين درجات كل سؤال ودرجة الاختبار ككل كمؤشر على صدق الاختبار؟

ذكر فرج (2007) بأنه يسمى صدق البناء الداخلي، وهي علاقة درجات بنود الاختبار بالدرجة الكلية، وتدل معاملات الارتباط هذه، أن هذه البنود تقيس شيئاً مشتركاً. ومن خلال الجدول التالي يتضح لنا معامل الارتباط بين كل مفردة من مفردات الاختبار المركزي والدرجة الكلية للاختبار.

جدول (3) معامل الارتباط بين كل فقرة والدرجة الكلية

رقم الفقرة	معامل الارتباط								
1	.610**	11	.491**	21	.543**	31	.216**	41	.262**
2	.458**	12	.253**	22	.378**	32	.445**	42	.331**
3	.542**	13	.457**	23	.393**	33	.399**	43	.154**
4	.383**	14	.321**	24	.532**	34	.431**	44	.037
5	.468**	15	.398**	25	.536**	35	.392**	45	.415**

رقم الفقرة	معامل الارتباط								
6	.550**	16	.429**	26	.412**	36	.516**	46	.333**
7	.533**	17	.340**	27	.414**	37	.396**	47	.409**
8	.435**	18	.414**	28	.282**	38	.537**	48	.296**
9	.531**	19	.387**	29	.392**	39	.336**	49	.381**
10	.387**	20	.605**	30	.514**	40	.305**	50	.144**

** (Correlation is significant at the 0.01 level (2-tailed))

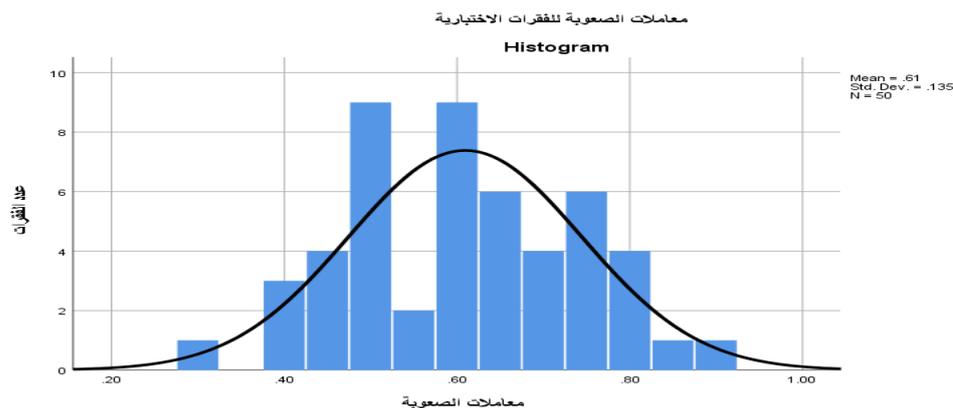
يتضح من الجدول أعلاه أن جميع معاملات الارتباط بين الفقرة الاختبارية والدرجة الكلية ذات دلالة إحصائية عند مستوى $\alpha = 0.01$ ، ماعدا الفقرة رقم (44)، فلم تكن ذات دلالة إحصائية، وهذا مؤشر قوي على صدق البناء التكويني للاختبار.

وبعد الإجابة على تساؤلات الدراسة المتعلقة بالصدق والثبات، لأبد من التعرض للخصائص المتعلقة بتحليل وجودة المفردات، وتمت صياغة عدد من الأسئلة المتعلقة بذلك، ولعل أول هذه الأسئلة ما يتعلق بصعوبة الفقرة الاختبارية، فتمت صياغته كما يلي:

4- ما جودة معاملات الصعوبة لكل فقرة من فقرات الاختبار المركزي في مقرر الرياضيات للصف الأول متوسط؟ وللإجابة على هذا السؤال تم حساب معاملات الصعوبة للفقرات الاختبارية كما هو موضح في الجدول التالي:

جدول (4) معاملات الصعوبة لكل فقرة من فقرات الاختبار

رقم الفقرة	معامل الصعوبة								
1	0.70	11	0.79	21	0.60	31	0.48	41	0.77
2	0.75	12	0.52	22	0.50	32	0.46	42	0.76
3	0.73	13	0.49	23	0.68	33	0.56	43	0.61
4	0.67	14	0.91	24	0.74	34	0.30	44	0.40
5	0.47	15	0.67	25	0.65	35	0.51	45	0.68
6	0.47	16	0.40	26	0.48	36	0.51	46	0.75
7	0.52	17	0.60	27	0.63	37	0.45	47	0.66
8	0.68	18	0.83	28	0.65	38	0.53	48	0.60
9	0.52	19	0.38	29	0.81	39	0.60	49	0.82
10	0.82	20	0.59	30	0.58	40	0.59	50	0.60



نلاحظ من خلال الجدول السابق والرسم البياني ما يلي:

1- عدد الفقرات التي معامل صعوبتها أقل أو يساوي (0.40)، أي بمعنى صعوبتها عالية، (4) فقرات، بنسبة (8%) من إجمالي فقرات الاختبار.

2- عدد الفقرات التي معامل صعوبتها يتراوح بين (0.41-0.70)، أي بمعنى صعوبتها متوسطة، (34) فقرة، بنسبة (68%) من عدد فقرات الاختبار.

3- عدد الفقرات التي معامل صعوبتها أكثر من (0.70)، أي بمعنى صعوبتها منخفضة (فقرات سهلة)، (12) فقرة، بنسبة (24%) من إجمالي فقرات الاختبار.

وقد اختلف متخصصي القياس في تحديد مدى قبول الفقرة الاختبارية (صلاحيتها أو رفضها) نسبة إلى معامل صعوبتها، منهم من جعل مستوى القبول محصوراً بين (0.20-0.80)، ومنهم من حددها بين (0.30-0.70) (الجلالي، 2011).

وبالنظر إلى نتائج هذه الدراسة فقد كان معامل الصعوبة لجميع الفقرات محصوراً بين (0.30-0.82)، وهو مستوى متوسط، وقد اتفقت هذه النتيجة مع دراسة الزواهري (2010) في الحد الأعلى لمعامل الصعوبة. ويمكن تفسير عدم سهولة الفقرات الاختبارية من خلال صعوبة صياغة فقرات الاختبار من متعدد وفقرات (الصحيح والخطأ) وخاصة في المواد العلمية كالرياضيات.

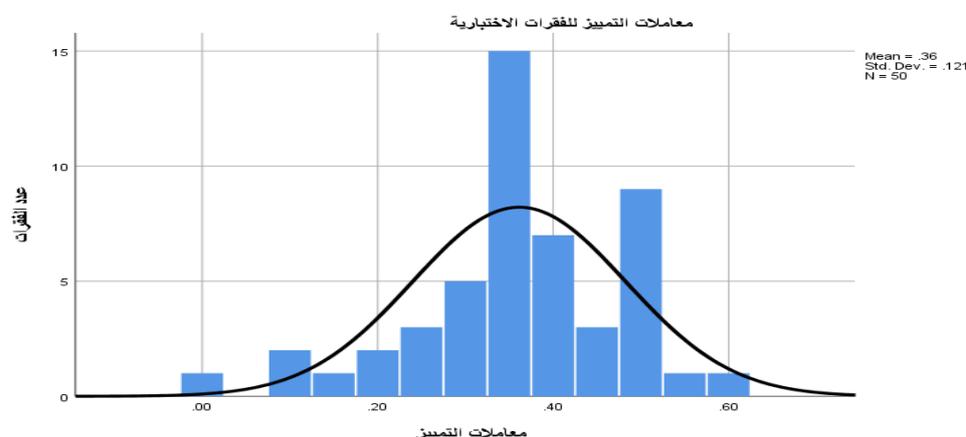
وبعد عرض نتائج معاملات الصعوبة، لا يُد من النظر في الخاصية الأخرى من الخصائص السيكومترية للفقرات وهي ما تسمى بتمييز الفقرات الاختبارية، ولتوضيح قدرة الفقرات الاختبارية على التمييز في الاختبار المركزي، تمت صياغة التساؤل الخامس من أسئلة الدراسة وهو كالتالي:

5- ما جودة معاملات التمييز لكل فقرة من فقرات الاختبار المركزي في مقرر الرياضيات للصف الأول متوسط؟
تم حساب معاملات التمييز عن طريق الارتباط الثنائي المتسلسل Point biserial correlation، كما هو موضح بالجدول التالي:

جدول (5) معاملات التمييز لكل فقرة من فقرات الاختبار

رقم الفقرة	معامل التمييز								
1	0.58	11	0.46	21	0.51	31	0.17	41	0.22
2	0.42	12	0.20	22	0.33	32	0.40	42	0.29
3	0.51	13	0.41	23	0.35	33	0.35	43	0.10
4	0.34	14	0.29	24	0.50	34	0.39	44	0

رقم الفقرة	معامل التمييز								
5	0.43	15	0.36	25	0.50	35	0.35	45	0.37
6	0.51	16	0.39	26	0.37	36	0.48	46	0.29
7	0.49	17	0.29	27	0.37	37	0.35	47	0.37
8	0.39	18	0.38	28	0.23	38	0.50	48	0.25
9	0.49	19	0.34	29	0.36	39	0.29	49	0.35
10	0.35	20	0.57	30	0.47	40	0.26	50	0.09



من خلال الجدول السابق والرسم البياني يمكن ملاحظة ما يلي:

- 1- الفقرات التي معامل تمييزها من (0 - 0.19) عبارة عن (4) فقرات، بنسبة (8%) من فقرات الاختبار، وتعتبر فقرات ضعيفة التمييز وينصح بحذفها.
 - 2- الفقرات التي معامل تمييزها من (0.20 - 0.39) عبارة عن (29) فقرة، بنسبة (58%) من فقرات الاختبار، وتعتبر فقرات مقبولة التمييز وينصح بتحسينها.
 - 3- الفقرات التي معامل تمييزها أعلى من (0.39) عبارة عن (17) فقرة، بنسبة (34%) من فقرات الاختبار، وتعتبر ذات تمييز جيد، ويمكن الاحتفاظ بها.
- نلاحظ من خلال نتيجة هذه الدراسة أن (92%) من الفقرات الاختبارية تراوح معامل تمييزها (0.20 - 0.58)، وهو تمييز مقبول. وعند النظر إلى الفقرة الاختبارية التي حظيت بأقل معامل تمييز (0) والتي يجب حذفها هي ذات الفقرة التي لم تكن ذو دلالة إحصائية في مؤشر الصدق (فقرة رقم 44)، بل أيضاً كانت من الفقرات الصعبة. وهذا ينبه أن الفقرة ربما بها مشكلة ما، أدت إلى تساوي الإجابة عليها من الفئتين (منخفضي التحصيل مرتفعي التحصيل). ولا يمكن أن تكتمل دراسة الخصائص السيكومترية المتعلقة بجودة الفقرة الاختبارية دون التطرق إلى دراسة مموهات أو (مشتتات) كل فقرة اختبارية، وهو بدوره يؤدي إلى الانتقال إلى السؤال السادس والأخير من أسئلة الدراسة وهو كالتالي:
- 6- ما مدى فاعلية المشتتات لكل فقرة من فقرات الاختبار من متعدد في الاختبار المركزي في مقرر الرياضيات للصف الأول متوسط؟

يمكن الإجابة على هذا السؤال عن طريقتين هما:

أولاً: حساب القدرة التمييزية للخيارات المدرجة مع كل فقرة اختبارية:
 للتعرف على معاملات التمييز للخيارات الصحيحة ل فقرات الاختيار من متعدد (40) فقرة (التي يجب أن تكون موجبة التمييز)، وكذلك للمشتتات (التي يجب أن تكون سالبة التمييز)، تم حساب مؤشر التمييز، ويتضح ذلك من خلال الجدول التالي.

جدول (6) معاملات التمييز للخيارات المدرجة مع كل فقرة اختيارية

رقم الفقرة	البدائل				رقم الفقرة	البدائل			
	أ	ب	ج	د		أ	ب	ج	د
1	-0.25	-0.23	-0.24	0.72	21	-0.11	-0.29	-0.32	0.72
2	-0.11	0.5	-0.23	-0.16	22	0.41	-0.23	-0.1	-0.08
3	0.6	-0.39	-0.19	-0.02	23	-0.27	0.46	-0.11	-0.08
4	-0.08	-0.19	0.43	-0.16	24	-0.12	-0.15	0.58	-0.31
5	-0.1	0.59	-0.27	-0.22	25	-0.33	-0.15	0.67	-0.19
6	-0.26	-0.2	0.69	-0.23	26	-0.18	0.55	-0.17	-0.2
7	-0.2	-0.19	0.69	-0.3	27	0.54	-0.25	-0.21	-0.08
8	-0.15	-0.28	-0.08	0.51	28	0.36	-0.13	-0.14	-0.09
9	0.68	-0.46	-0.07	-0.15	29	0.38	-0.04	-0.2	-0.14
10	-0.09	-0.16	0.35	-0.1	30	-0.15	-0.23	-0.24	0.62
11	0.53	-0.17	-0.15	-0.21	31	0.04	-0.22	0.28	-0.1
12	-0.1	0.29	-0.14	-0.05	32	-0.17	-0.13	-0.21	0.51
13	-0.27	0.56	-0.11	-0.18	33	-0.02	0.49	-0.22	-0.25
14	0.22	-0.07	-0.03	-0.12	34	-0.12	-0.22	-0.13	0.47
15	-0.2	-0.21	0.46	-0.05	35	-0.12	-0.15	0.49	-0.22
16	-0.27	0.57	-0.19	-0.11	36	-0.37	-0.2	0.65	-0.08
17	0.37	-0.17	-0.12	-0.08	37	-0.07	0.5	-0.29	-0.14
18	-0.1	-0.17	-0.13	0.4	38	-0.14	-0.38	0.68	-0.16
19	-0.21	-0.22	0.47	-0.04	39	-0.32	-0.07	-0.04	0.43
20	0.78	-0.19	-0.13	-0.46	40	0.41	-0.2	-0.17	-0.04

*الخلية المضللة هي الإجابة الصحيحة.

يوضح الجدول السابق مجموعة من النقاط لعل أهمها أن:

- جميع قيم معاملات التمييز للبدائل الصحيحة موجبة التمييز. وهذا يؤكد أن عدد الذين أجابوا عليها من مرتفعي التحصيل أعلى من منخفضي التحصيل.
- معظم قيم معاملات التمييز للمشتتات سالبة التمييز (عدد الذين أجابوا عليها من الفئة الدنيا أكبر من الفئة العليا من التحصيل)، ماعدا مشتت واحد في الفقرة رقم (31)، الخيار (أ) كان تمييزه موجب، حيث تم اختياره من عدد أكبر من الطلاب أصحاب التحصيل الأعلى.

ثانياً- توضيح نسب الطلاب الذين اختاروا المشتت:

بحساب نسب الطلاب الذين اختاروا كل بديل من البدائل المقترنة مع كل فقرة اختبارية، يمكن ملاحظة قدرة جميع البدائل (الصحيحة والخاطئة) على جذب أكثر من (5%) من عدد الطلاب على الأقل.

جدول (7) النسب المئوية للطلاب الذين اختاروا كل بديل

رقم الفقرة	البدائل				رقم الفقرة	البدائل			
	أ	ب	ج	د		أ	ب	ج	د
1	9.4	10.2	9.9	70.4	21	8.9	12.1	18.5	60.5
2	5.1	75.3	11.8	7.8	22	50.3	26.1	15.1	8.6
3	72.6	17.2	8.3	1.9	23	18.3	67.7	9.4	4.6
4	8.6	14.2	67.5	9.7	24	6.2	7	73.7	13.2
5	16.4	46.8	18.8	18	25	18.3	10.2	64.8	6.7
6	20.2	18.5	46.8	14.5	26	8.1	48.4	12.6	30.9
7	12.1	15.9	52.2	19.9	27	63.2	20.2	11.8	4.8
8	6.2	14.2	11.8	67.7	28	64.5	20.2	11.3	4
9	51.9	34.7	8.3	5.1	29	80.6	1.3	9.4	8.6
10	4	9.9	81.7	4.3	30	7	15.6	19.9	57.5
11	78.8	6.5	6.2	8.6	31	20.4	20.7	47.6	11.3
12	25	51.9	17.7	5.4	32	18.8	15.6	19.6	46
13	27.2	48.9	16.4	7.5	33	13.7	56.2	13.4	16.7
14	91.1	3.5	8	4.6	34	26.3	30.4	13.4	29.8
15	8.3	15.9	66.9	8.9	35	12.6	9.1	50.8	27.4
16	26.6	39.8	18	15.6	36	22.8	17.7	51.3	8.1
17	60.2	11.8	9.1	18.8	37	15.9	45.4	24.7	14
18	3	6.2	8.1	82.8	38	10.5	25.5	53.5	10.5

البدايل				رقم الفقرة	البدايل				
د	ج	ب	أ		د	ج	ب	أ	
59.9	7.3	8.9	23.9	39	39.2	37.6	8.6	14.5	19
9.7	19.9	11.3	59.1	40	24.5	7.5	9.4	58.6	20

*الخلية المظللة هي الإجابة الصحيحة.

النسبة المئوية الاقل من (5%).

نلاحظ من خلال الجدول السابق للتعرف على المشتتات غير الجاذبة أن:

- يوجد (10) بدائل فقط من مجموع البدائل الخاطئة (120) بديل، غير قادرة على جذب انتباه نسبة كافية من الطلاب، حيث بلغت نسبة اختيارها أقل من (5%)، وهي كما يتضح من خلال الجدول التالي:

جدول (8): نسبة البدائل الأقل من 5%

رقم الفقرة	البديل	النسبة %	رقم الفقرة	البديل	النسبة
3	د	1.9	18	أ	3
10	أ	4	23	د	4.6
10	د	4.3	27	د	4.8
14	د	4.6	28	د	4
14	ب	3.6	29	ب	1.3

بالنظر إلى نتائج تحليل فعالية المشتتات نجد أن هذه النتائج تدل على كفاءة المشتتات الموضوعية في جذب انتباه الطلاب، حيث جذبت معظم البدائل نسبة من الطلاب تفوق 5%. وتشير نتائج الدراسة إلى أن أغلب معاملات التمييز المقدرة للمشتتات هي معاملات سالبة مما يؤكد على أن نسبة منخفضي التحصيل الذين اختاروا البديل الخاطئ أكبر من نسبة مرتفعي التحصيل الذين اختاروا هذا البديل، وهو ما يدل على مؤشر كفاءة هذه البدائل. وأخيراً دلت نتائج دراسة فعالية المشتتات على أن المشتت (د) في الفقرات الاختبارية هو المشتت الأقل قدرة على جذب النسبة الكافية من الطلاب. ويمكن تفسير ذلك من خلال صعوبة صياغة بدائل ذات كفاءة عالية.

خلاصة الدراسة:

هدفت الدراسة الحالية إلى التعرف على الخصائص السيكومترية لاختبار الرياضيات المركزي للصف الأول متوسط، والتي تم تطبيقه خلال الفصل الدراسي الثاني للعام 1439/1438 هـ، وشملت الدراسة على الأسئلة الموضوعية حيث كانت عبارة عن (40) فقرة خيار من متعدد، و(10) فقرات أسئلة (الصح والخطأ)، وللإجابة على تساؤلات الدراسة المتعلقة بالخصائص السيكومترية، تم تحليل نماذج إجابات (372) طالباً تم اختيارهم بطريقة عنقودية عشوائية من إجمالي (12484) طالباً، وهم جميع طلاب الصف الأول متوسط بمنطقة المدينة المنورة التعليمية، وبعد جمع البيانات وتحليلها عن طريق برنامج (SPSS) الإصدار (25). توصلت الدراسة إلى مجموعة من النتائج والتي يمكن تلخيصها في النقاط التالية:

- 1- ارتفاع معامل ثبات الاختبار المركزي للرياضيات المقدر بمعادلة ألفا كرونباخ للاتساق الداخلي، وهو الأمر الذي يمكن تفسيره بعدد من العوامل منها جودة الخصائص السيكمومترية المتعلقة بالفقرة، ومناسبة زمن الاختبار لعدد الفقرات المدرجة فيه، وموضوعية التصحيح.
- 2- أشارت الاختبارات التي استخدمت للتعرف على دلالات الصدق البنائي (اختبارات للتعرف على الفروق في المجموعات بين مرتفعي التحصيل ومنخفضي التحصيل، ومعاملات بيرسون بين درجات الفقرة والدرجة الكلية)، إلى وجود مؤشرات كبيرة تدل على توافر الصدق البنائي للاختبار المركزي للرياضيات.
- 3- مناسبة صعوبة الفقرات الاختبارية حيث كانت معظم معاملات الصعوبة للفقرات الاختبارية متوسطة الصعوبة.
- 4- نتيجة مناسبة مستويات الصعوبة هو الأمر الذي أدى إلى زيادة قدرتها على التمييز بين مجموعتي الطلاب مرتفعي التحصيل، والطلاب منخفضي التحصيل، كما هو واضح من خلال معاملات التمييز لمعظم الفقرات كانت مرتفعة.
- 5- أكدت نتائج تحليل فعالية البدائل المصاحبة لكل فقرة اختيارية قدرة معظم البدائل والمشتتات على جذب انتباه الطلاب، كما أثبتت تحليل معاملات التمييز للمشتتات أن هذه المعاملات سالبة التمييز.

توصيات الدراسة:

- 1- التوسع في تطبيق هذا المشروع (تجويد الاختبارات) على جميع المواد الدراسية في المرحلة المتوسطة.
- 2- مناقشه جدوى استخدام مفردات الاختيار من متعدد، ومواصفاتها، وأنواعها، مع إيجابياتها وسلبياتها، ومتى لا ينبغي استخدامها.
- 3- تدريب القائمين على هذا المشروع والمعلمين بشكل عام على مهارة صياغة مفردات ذات فعالية عالية في أسئلة الاختيار من متعدد.
- 4- الاستعانة بمعلمين المواد بشكل أكبر لبناء بنك أسئلة، يتم الاستفادة منه لاحقاً في أعداد صور مكافئة للاختبارات.
- 5- الاهتمام بالدورات التدريبية التي تعمل على تطوير الاختبارات، وصناعة الاختبار الجيد، وخاصة أولئك الذين يتعاملون مع مشروع (تجويد الاختبارات).
- 6- تقديم ورش عمل تتعلق بالتحليلات الإحصائية لنتائج الاختبارات وتفسيرها.

قائمة المراجع

أولاً- المراجع بالعربية:

- إبراهيم، محمد عبدالرازق وأبو زيد، عبد الباقي عبدالمنعم (2012). مهارات البحث التربوي. (ط3)، عمان: دار الفكر.
- ابكر، ثريا محمد (2012). تقويم أسئلة الاختبار الموحد لمادة الرياضيات بحلقة الثانية من التعليم الاساسي بولاية الخرطوم، رسالة ماجستير غير منشورة، جامعة النيلين، السودان.
- أبو فودة، باسل خميس ويونس، نجاتي أحمد (2012). الاختبارات التحصيلية. (ط1)، عمان: دار المسيرة للنشر والتوزيع.

- أحمد، مي فيصل، (2017). تقويم اداء مديري المدارس الثانوية في ادارة الامتحانات من وجهة نظر المدرسين، جامعة بغداد، كلية التربية للعلوم الصرفة - ابن الهيثم، مجلة الاستاذ، العدد (221).
- الجلاي، لمعان مصطفى (2011)، التحصيل الدراسي. (ط1)، عمان: دار المسيرة.
- الخياط، ماجد محمد (2010)، أساسيات القياس والتقويم في التربية. (ط1)، عمان: دار الراجحة.
- دودين، حمزة (2006)، مشكلات الطلاب في الاختبارات وطرق علاجها، الامارات العربية المتحدة: مكتبة الفلاح للنشر والتوزيع.
- الدوسري، إبراهيم مبارك (2000). الإطار المرجعي للتقويم التربوي. (ط2)، الرياض: مكتب التربية العربي لدول الخليج.
- زواهره، ريم (2010). الخصائص السيكومترية لاختبار الرياضيات (المستوى الثالث) في بطارية اختبارات القدرات الواسعة (WRAT) على عينة من طلبة المدارس الحكومية في محافظة عجلون، رسالة ماجستير غير منشورة، الجامعة الاردنية، عمان، الاردن.
- الشريف، نايف شجاع هزاع (2013). الخصائص السيكومترية لصورتي اختبار في الرياضيات للمرحلة المتوسطة: دراسة مقارنة. رسالة ماجستير غير منشورة، كلية التربية، جامعة اليرموك.
- علام، صلاح الدين محمود (2012)، الاختبارات والمقاييس التربوية والنفسية. (ط2)، عمان: دار الفكر.
- علام، صلاح الدين محمود (2013)، اتقان القياس النفسي الحديث. (ط1)، عمان: دار الفكر.
- عمر، محمود وفخرو، حصه والسبيعي، تركي وتركي، آمنه (2010)، القياس النفسي والتربوي. (ط1)، عمان: دار المسيرة.
- القرني، علي عبدالخالق، الدوسري، إبراهيم مبارك، المحرزي، راشد والخروصي، حسين علي (2012). التقييم التربوي للطلبة. الرياض: مكتبة التربية العربي لدول الخليج.
- القضاة، بسمه عواد (2010). بناء اختبار تحصيلي في الرياضيات للصف التاسع الاساسي وتقدير خصائصه السيكومترية وفق النظرية الكلاسيكية ونموذج المعلمتين للنظرية الحديثة، رسالة ماجستير غير منشورة، جامعة مؤتة، الاردن.
- الكيلاني، عبدالله زيد وعدس، عبدالرحمن والتقي، أحمد (2009)، القياس والتعليم في التعلم والتعليم. (ط3)، جامعة القدس المفتوحة.
- مجيد، سوسن شاكر (2014). أسس بناء الاختبارات والمقاييس النفسية والتربوية. (ط3)، عمان: مركز دبيونو لتعليم التفكير.
- مراد، صلاح أحمد وسليمان، أمين علي (2012). الاختبارات والمقاييس في العلوم النفسية والتربوية. (ط2)، القاهرة: دار الكتاب الحديث.
- ملحم، سامي محمد (2010)، مناهج البحث في التربية وعلم النفس.. (ط6)، عمان: دار المسيرة.
- النفيعي، عبدالرحمن عبدالله (2016)، دراسة تقويمية لتعرف مدى إتقان أعضاء مشروع الاختبارات التحصيلية لمهارات بناء الاختبارات التحصيلية، جامعة القاهرة، كلية الدراسات العليا، مجلة العلوم التربوية، العدد 24 (1).
- يعقوب، أبو فودة (2012)، أثر مخالفة قواعد صياغة فقرات الاختيار من متعدد على الخصائص السيكومترية للاختبار وفقراته، مجلة جامعة دمشق للعلوم التربوية والنفسية، العدد 28. سوريا.

ثانياً- المراجع الأجنبية:

- Atalmis, E. H. (2014). The Impact of the Item Types and Number of Solution Steps of Multiple-Choice Items on Item Difficulty and Discrimination and Test Reliability (Doctoral dissertation, University of Kansas).
- Ebel, R. L. (1979). Essentials of educational measurement. New Jersey, Prentice.
- Gregory, R. J. (2004). Psychological testing: History, principles, and applications. Allyn & Bacon.
- Krejcie, R.V., & Morgan, D. W. (1970). Determining sample size for research activities. Educational and psychological measurement, 30(3), 607-610.
- MEHRENS, W. A. L., & IRVIN, J. (1975). MEASUREMENT AND EVALUATION IN EDUCATION AND PSYCHOLOGY/WILLIAM A. MEHRENS AND IRVIN J. LEHMANN (No. LB1051. M43 1991.).
- Nitko, A. J., & Brookhart, S. M. (2018). Educational assessment of students. Englewood Cliffs, NJ: Merrill.
- Reynolds, C. R., & Livingston, R. B. (2012). Mastering modern psychological testing: Theory and methods. Pearson Education.
- Sax, G., & Newton, J. W. (1980). Principles of educational and psychological measurement and evaluation.
- Stemler, S.(2007). Educational supervisors competences in constructing achievement test sin ohio, Ohio Journal of education, 1(1): 30-45.