المملكة العربية السعودية
وزارة التعليم
جامعة تبوك
كلية التربية والآداب
قسم التربية وعلم النفس

أثر زمن الاختبار على تقدير معالم الفقرة وثبات الاختبار دراسة مقارنة: بين نظرية استجابة الفقرة والنظرية التقليدية

Impact of Test Time on Estimating Aspects of Items and Test Reliability: A Comparative Study : Item Response Theory and Classical Test Theory

بحث مقدم إلى قسم التربية وعلم النفس كمتطلب تكميلي لنيل درجة الماجستير
في التربية تخصص القياس والتقويم

إعداد الباحث

فرحان علي صالح البلوي

(الرقم الجامعي ٣٦١٠٠٠٣٠٥)

إشراف

د. صبري محمد إسماعيل عبد العال

أستاذ مساعد في القياس والتقويم التربوي

كلية التربية والآداب – جامعة تبوك

١٤٣٩هـ /٢٠١٨ م

الملخص

أثر زمن الاختبار على تقدير معالم الفقرات وثبات الاختبار

دراسة مقارنة: نظرية استجابة الفقرة والنظرية التقليدية

فرحان علي صالح البلوي

جامعة تبوك 2018

هدف البحث الكشف عن أثر زمن الاختبار على تقديرات معالم الفقرة وفق نظرية استجابة الفقرة (ثنائي المعلم) والنظرية التقليدية. ولتحقيق هدف البحث، تم بناء اختبار تحصيلي اختيار من متعدد (٢٨ فقرة بشكله النهائي) في مقرر الرياضيات الصف الخامس. وشُكّلت ثلاث صور للاختبار تختلف في زمن الإجابة. الصورة الأولى الزمن الطويل(٤٠د)، الصورة الثانية الزمن المناسب (٣٠د)، الصورة الثالثة الزمن القصير(٢٠د). وطُبقت الصور الثلاث على عينة مكونة من (٤٥١) طالب، ثم حُلّلت البيانات باستخدام برنامج(BILOG-MG3) وبرنامج (ITEMAN).

واظهرت النتائج اتفاق نظرية استجابة الفقرة والنظرية التقليدية على عدم وجود فروق ذات دلالة إحصائية(α=0.05) في متوسطات تقدير معالم الصعوبة تُعزى للزمن. واتفاق بين نظرية استجابة الفقرة النموذج الثنائي مع النظرية التقليدية فيما يخص معلمة التمييز بوجود فروق ذات دلالة إحصائية في المتوسطات الحسابية تُعزى لعامل الزمن، لصالح الوقت المناسب (٣٠ د) وفق نظرية استجابة الفقرة والنظرية التقليدية. ووجود فروق ذات دلالة إحصائية في دقة تقدير معالم الصعوبة ومعالم التمييز لصالح الصورة الثانية (الوقت المناسب٣٠د) فهي الأكثر دقة في تقدير هذه المعالم وفق نظرية استجابة الفقرة.

كما أظهرت نتائج اختبارات الإحصائي M للفروق بين معاملات الثبات للصور الثلاث والمحسوبة بالطرق التقليدية، وجود فروق ذات دلالة إحصائية في معاملات الثبات لصالح الصورة الثانية الوقت المناسب (٣٠ د). وأن نسبة الاتفاق بين نظرية استجابة الفقرة(ثنائي المعلم) والنظرية التقليدية في فرز الفقرات بناء على معالم الصعوبة والتمييز في الصورة الأولى (الوقت الطويل) ٨٩%، والصورة الثانية (الوقت المناسب) ٩٦%، وفي الصورة الثالثة (الوقت القصير) ٦٨%، وقد خرج البحث بمجموعة من التوصيات أهمها: ينبغي تحديد زمن الاختبار المناسب بدقة عالية لما له من تأثير على خصائص الاختبار والفقرات ودقة تقديرها.

الكلمات المفتاحية: اختبار اختيار من متعدد، النموذج اللوجستي ثنائي المعلم، نظرية استجابة الفقرة، النظرية التقليدية، زمن الاختبار، معالم الفقرة، ثبات الاختبار.

ABSTRACT

Impact of Test Time on Estimating Aspects of Items and Test Reliability: A Comparative Study : Item Response Theory and Classical Test Theory

Farhan Ali Al-Balawi
University of Tabuk 2018

The Study aimed to explore the impacts of test time on estimations of items aspects in context with Item Response Theory (pair parameter), and Classical Test Theory. To achieve the objective of the study, a multi-choice achievement test (28) items in final form was constructed in the fifth grade mathematics course. Three test images were formed, varying in response time. The first image (long time 40 m), the second image the appropriate time (30 m), the third picture short time (20 m). The three images were applied to a sample of (451) students. The responses were then analyzed using (BILOG-MG3) program and (ITEMAN) program. The results showed that there is concordance of Item Response Theory and Classical Test Theory on the absence of statistically significant differences ($\alpha = 0.05$) in the average of the estimation of the characteristics of the difficulty is related to time. The results showed that there is concordance of Item Response Theory and Classical Test Theory related to Discrimination Coefficient on the existence of statistically significant differences ($\alpha = 0.05$) in the average of the estimation of the characteristics of the difficulty is related to time in favor for appropriate time (30 m) according to the two theories. The results showed that there were statistically significant differences in the accuracy of the estimation of the difficulty coefficient and discrimination Coefficient for the second image the appropriate time (30 m). As the results of the M-tests showed differences between the reliability coefficients of the three images and calculated by conventional methods, there were statistically significant differences in the reliability coefficients for the second image time (30 m). The results showed that the ratio Item Response Theory (pair parameter), and Classical Test Theory in the classification of items based on the characteristics of difficulty and discrimination in the first image (long time) 89%, the second image (the appropriate time) 96%, the third image (short time (68%).

Key words: Multiple Choice Test, Logistic Three Parameters Model, Item Response Theory, Classical Test Theory, Time Test, Parameters Ietm, Reliability Test

قائمة المحتويات

المحتوى	رقم الصفحة
الإهداء	أ
الشكر والتقدير	ب
الملخص باللغة العربية	ج
الملخص باللغة بالإنجليزية	د
قائمة المحتويات	هـ
قائمة الجداول	و– ز
قائمة الأشكال	ح
قائمة الملاحق	ط
الفصل الأول : الإطار العام للبحث	١–٧
– المقدمة	١
– مشكلة البحث وأسئلته	٣–٥
– أهداف البحث	٥–٦
– أهمية البحث	٦
– مصطلحات البحث	٦–٧
– محددات البحث	٧
الفصل الثاني : أدبيات البحث	٨–٥٦
– الإطار النظري	٨–٥١
– الدراسات السابقة	٥٢–٥٦
الفصل الثالث : منهجية البحث وإجراءاته	٥٧–٦١
– مجتمع البحث وعينة البحث	٥٧
– أداة البحث (بناء الأداة–صدق الأداة–ثبات الأداة)	٥٨– ٦٠
– الأساليب الاحصائية المستخدمة	٦١
الفصل الرابع : عرض نتائج التحليل	٦٢– ٩٣
الفصل الخامس : ملخص النتائج والتوصيات والمقترحات	٩٤–٩٧
– **النتائج**	٩٤–٩٦
– **التوصيات**	٩٧
– **البحوث المقترحة**	٩٧
قائمة المراجع	٩٨–١٠١
الملاحق	١٠٢– ١٣٨

الفصل الأول : الإطار العام للبحث

مقدمة :

يُعدّ المتعلم(الطالب) هو محور العملية التربوية، وبالتالي فإن الحديث عن القياس والتقويم التربوي ينصبّ أساساً وبصورة مباشرة أو غير مباشرة على نواتج التعلم لدى المتعلم، فتنطوي عملية التقويم لأي برنامج تربوي على اتخاذ قرارات تؤثر بصورة مباشرة أو غير مباشرة على المتعلم نفسه. كأن يكون القرار استمرارية برنامج تربوي معين أو إعادة النظر فيه أو استبداله كلياً بناء على مدى فعالية البرنامج في إحداث تقدم عند المتعلم نحو تحقيق أهداف كانت قد تم إعداداها مسبقاً بالاتجاه المرغوب فيه. كما أنه على الأغلب نجد بأن المعلومات اللازمة لعملية التقويم قد جاءت بطريقة أو بأخرى من المتعلم نفسه ومن غرف الصف بشكل خاص، وبما أن دقة القرار تعتمد على دقة المعلومات المتوفرة، فإن الحصول على معلومات صادقة يشكل حجر الزاوية في عملية التقويم، وأن تقويم الطلاب هو حجر الزاوية في البناء المتكامل لعملية التقويم. (عودة، ٢٠١٠:٢٢)

والتحصيل من أهم الجوانب التي ينبغي تقويمها في المتعلم، ومن أهم وسائل وسائل التقويم التي تمد المعلم بمعلومات وافية عن المستوى التحصيلي للمتعلم هي الاختبارات. (الشيخ وآخرون ،٢٠٠٩ : ١٠٥).

فبإستخدام الاختبارات والمقاييس يحصل المعلم على درجات ومعايير، ومتوسطات، وغير ذلك من القيم العددية.فالإنسان الذي هو موضوع القياس التربوي يعدّ أكثر الكائنات تعقيداً وتتداخل سماته وخصائصه، ويعتمد بعضها على البعض الآخر، مما يجعل الفصل بينها من أجل دراستها ليس أمراً يسيراً، فالعلماء لم يتمكنوا من الفصل بدرجة كاملة بين تأثيرات الوراثة والبيئة مثلاً. وكذلك لسنا متأكدين دائماً إذا كنا نقيس قدرة الطالب على التعلم أم رغبته في التعلم، وغير ذلك.(علّام، ٢٠١٥ ب:٢٨).

وللتأكد من تحقيق الأهداف وامتلاك الطلاب للمهارات والمعارف المطلوبة لابد من استخدام أدوات قياس موضوعية، وأدوات القياس يجب أن تُبنى بطريقة علمية وموضوعية وتحلل نتائجها وتعالج بطريقة علمية موضوعية أيضاً. إلا أن تطوير أدوات القياس مازال يُعدّ مشكلة من المشكلات الرئيسة في مجال القياس والتقويم وتؤكد انستازي (ANstasi ,1985) على أهمية الدراسات التي

تقدم معلومات عن خصائص الفقرات التي تم اختيارها وفق النظرية التقليدية مقارنة بخصائص الفقرات التي تم اختيارها باستخدام نظرية استجابة الفقرة باعتبارها النظرية التي تشكّل الإطار الحالي في اختيار الفقرات (في الشمراني، ٢٠١٤).

هناك مدخلين للقياس التربوي والنفسي هما المدخل التقليدي والذي يتمثل في النظرية التقليدية للاختبارات Classical Test Theory (CTT)، والمدخل الحديث والذي يتمثل في نظرية إستجابة الفقرة Item Response Theory (IRT)، ولكل مدخل افتراضاته ومفاهيمه ومبادئه التي يقوم عليها.

وقد وُجد أن معدّي ومطوري الاختبارات يتجاهلون تلك الافتراضات أثناء بنائها مما يجعلهم عرضة للوقوع في أخطاء القياس التي تنعكس سلباً على النتائج ، والتي تُبنى عليها العديد من القرارات، فيغيب عن الأذهان ضرورة التأكد من أبعاد الاختبار والسمات التي يجب أن يكشف عنها قبل بنائها، من هنا كان لابد التأكيد على هذه الافتراضات وأهميتها للتتحق من صدق البناء وموضوعية القياس (Hattie, 1985).

إنّ كلتا النظريتين، النظرية التقليدية في القياس ونظرية إستجابة الفقرة تؤكدان على افتراض التحرر من السرعة. وإذا تم الانتهاك لهذا الافتراض فإن الاختبار يكون سريعاً، وبالتالي فإن المفحوصين لن يتمكّنوا من الإجابة على جميع فقرات الاختبار ضمن الزمن المحدد، فإما أن يتركوا بعض الفقرات بدون إجابة، أو يخمّنوا إجاباتها وخصوصاً تلك الفقرات التي تقع في نهاية الاختبار، وبالتالي فإن سرعة الاختبار ينتج عنها تقديرات غير دقيقة لمعالم الفقرة، وقدرة الأفراد، وإحصائيات الاختبار (Oshima, 1994).

إن انتهاك افتراض التحرر من السرعة يؤدي إلى انتهاك افتراض أحادية البعد، والسبب أن سرعة الاختبار تُضيف بعداً إحصائياً آخر لما يقيسه الاختبار، أي يوجد عامل آخر في الاختبارات السريعة يمكن استخراجه، وبالتالي يصبح الاختبار السريع متعدد الأبعاد. (Mroch, 2006).

غالباً ما ترجع سرعة الاختبار إلى تأثير حدود الزمن على علامات المفحوصين عندما تكون حدود وقت الاختبار مقيدة إلى الحد الذي لا يجد عنده معظم المفحوصين الوقت الكافي للتفكير و الإجابة على كل فقرة في الاختبار، وعندها يُقال بأن الاختبار سريع "Speeded"، وكذلك يسمى الاختبار سريعاً إلى مدى معين، إذا كان المفحوصين يأخذون علامات أقل من العلامات التي يأخذونها إذا

كان لديهم الوقت الكافي لإنهاء الاختبار، أو أُعطوا وقتاً إضافياً غير محدد لإنهائه (Bridgeman, McBride & Monaghan, 2004) (في المنايصة، ٢٠١٣).

كما عرّف ننلي(Nunnally, 1978)الوقت المناسب بأنه المدة التي تلزم 90% من الأشخاص لاكمال الاختبار. ولقد تضاربت الآراء حول طول الفترة المخصصة للاستجابة على الاختبار من حيث أهميتها وتأثيرها على خصائص الاختبار، حيث يرى ننلي Nunnally أن تقصير ثلث الوقت الملائم للاختبار لا يؤثر في خصائص الاختبار السيكومترية. (في العمري، ١٩٨٥) .

ويذكر رينولدس وليفينقستون (Reynolds&Livingston,2012/2013 :311) أن حدود الوقت في الاختبارات تعكس مقاييس صعوبة وتمييز الفقرة بدرجة كبيرة من خلال موقع الفقرة في الاختبار، وليس الصعوبة الفعلية للفقرة أو قدرتها على التمييز. فالفقرات التي تظهر متأخرة في الاختبار يجيب عنها عدد من الأفراد أقل من عدد الأفراد الذين يجيبون عنها إجابة صحيحة إذا ظهرت مبكرة.

كل ما سبق يقودنا إلى تساؤل :ماهي حدود الوقت التي لا تؤثر سلباً على الخصائص السكيومترية للاختبار، والبناء العاملي، وقدرات الأفراد، والذي يناسب الظروف التطبيقية لإجراء الاختبار وكذلك لا تكون إطالة الوقت هدراً له وزيادة في التكاليف ، ويجعل الاختبار أكثر صدقاً وثباتاً وذلك وفق النظرية التقليدية ونظرية إستجابة الفقرة (ثنائي المعلمة) ، ومدى الإتفاق والإختلاف بين هاتين النظريتين.

مشكلة البحث :

مع تزايد الحاجة للاختبارات الموضوعية بأشكالها المتعددة، وفي المراحل التعليمية المختلفة، وخاصة في الاختبارات العامة، وتطبيقها على عدد كبير من الأشخاص، تزايد الاهتمام باجراءات تقديم وتحديد طول الزمن الملائم للاختبار الذي يحقق العدالة وعدم التحيّز ويحقق الأهداف المنشودة ويناسب الظروف العملية لتطبيق الاختبار، بحيث لا تكون إطالة الوقت هدراً له وزيادة في التكاليف، ولا تقصيره إلى الحد الذي يجعل منه اختباراً غير صادق.(العمري، ١٩٨٥)

تبيّن لنا مما سبق ومن خلال المقدمة أهمية الزمن وعلاقته بافتراض التحرر من السرعة المرتبط بافتراض أحادية البعد ، وأن انتهاك افتراض التحرر من السرعة يؤدي لتقديرات غير دقيقة لقدرات المفحوصين ولمعالم الفقرة وإحصائيات الاختبار.

وبما أن افتراض السرعة له علاقة وثيقة بالزمن المُعطى للاختبار، سوف أحاول بهذا البحث الكشف عن أثر الزمن المُعطى للاختبار على تقدير معالم الفقرة وثبات الاختبار وفق نظرية استجابة الفقرة (ثنائي المعلم) والنظرية التقليدية في القياس ومقارنة النتائج بين النموذجين.

أسئلة البحث:

السؤال الأول : هل توجد فروق دالة إحصائيا في المتوسطات الحسابية لمعاملات صعوبة الفقرة وفق النظرية التقليدية والنموذج اللوجستي ثنائي المعلم تُعزى لزمن الاختبار(٤٠ دقيقة ، ٣٠ دقيقة ، ٢٠ دقيقة) ؟

ويتفرع عنه الأسئلة التالية:

١-هل توجد فروق دالة إحصائيا بين متوسطات معامل الصعوبة لفقرات الاختبار وفق النظرية التقليدية تُعزى لزمن الاختبار(٤٠ دقيقة، ٣٠ دقيقة، ٢٠ دقيقة) ؟

٢-هل توجد فروق دالة إحصائيا بين متوسطات قيم معلمة صعوبة فقرات الاختبار وفق النموذج اللوجستي ثنائي المعلم لنظرية استجابة الفقرة تُعزى لزمن الاختبار(٤٠ دقيقة،٣٠ دقيقة،٢٠ دقيقة)؟

٣-هل توجد فروق دالة إحصائيا بين متوسطات دقة تقدير معلمة صعوبة فقرات الاختبار وفق النموذج اللوجستي ثنائي المعلم لنظرية استجابة الفقرة تُعزى لزمن الاختبار(٤٠ دقيقة، ٣٠ دقيقة، ٢٠ دقيقة) ؟

السؤال الثاني : هل توجد فروق دالة إحصائياً في المتوسطات الحسابية لمعاملات تمييز الفقرة وفقا للنظرية التقليدية والنموذج اللوجستي ثنائي المعلم تُعزى لزمن الاختبار(٤٠ دقيقة، ٣٠ دقيقة، ٢٠ دقيقة) ؟

ويتفرع عنه الأسئلة التالية:

١- هل توجد فروق دالة إحصائياً بين متوسطات معامل تمييز فقرات الاختبار وفق النظرية التقليدية تُعزى لزمن الاختبار(٤٠ دقيقة، ٣٠ دقيقة، ٢٠ دقيقة) ؟

٢- هل توجد فروق دالة إحصائيا بين متوسطات قيم معلمة تمييز فقرات الاختبار وفق النموذج اللوجستي ثنائي المعلم لنظرية استجابة الفقرة تُعزى لزمن الاختبار (٤٠ دقيقة، ٣٠ دقيقة، ٢٠ دقيقة)؟

٣- هل توجد فروق دالة إحصائيا بين متوسطات دقة تقدير معلمة تمييز فقرات الاختبار وفق النموذج اللوجستي ثنائي المعلم لنظرية استجابة الفقرة تُعزى لزمن الاختبار (٤٠ دقيقة، ٣٠ دقيقة، ٢٠ دقيقة) ؟

السؤال الثالث: هل توجد فروق دالة إحصائيا بين معاملات ثبات الاختبار وفقا لكل من النظرية التقليدية ونظرية استجابة الفقرة تُعزى لزمن الاختبار (٤٠ دقيقة، ٣٠ دقيقة، ٢٠ دقيقة) ؟ ويتفرع عنه السؤالين التاليين:

١- هل توجد فروق ذات دلالة إحصائية بين متوسطات معاملات ثبات الاختبار وفق النظرية التقليدية تُعزى لزمن الاختبار (٤٠ دقيقة، ٣٠ دقيقة، ٢٠ دقيقة) ؟

٢- هل توجد فروق ذات دلالة إحصائية بين معاملات ثبات الاختبار وفق النموذج اللوجستي ثنائي المعلم لنظرية استجابة الفقرة تُعزى لزمن الاختبار (٤٠ دقيقة، ٣٠ دقيقة، ٢٠ دقيقة) ؟

السؤال الرابع: هل يُفرز النموذج الثنائي المعلم لنظرية استجابة الفقرة فقرات تختلف في محتواها وعددها عن تلك التي تفرزها المؤشرات التقليدية اعتماداً على معاملات الصعوبة والتمييز لصور الاختبار الثلاثة (٤٠ دقيقة، ٣٠ دقيقة، ٢٠ دقيقة) ؟

أهداف البحث :

١- الكشف عن أثر زمن الاختبار على المتوسطات الحسابية لمعاملات صعوبة الفقرات وفق نظرية استجابة الفقرة (النموذج الثنائي) والنظرية التقليدية .

٢- الكشف عن أثر زمن الاختبار على المتوسطات الحسابية لمعاملات تمييز الفقرات وفق نظرية استجابة الفقرة (النموذج الثنائي) والنظرية التقليدية.

٣- تقدير أثر زمن الاختبار على دقة تقدير معالم الفقرة وفق نظرية استجابة الفقرة (النموذج الثنائي).

٤- التعرف على أثر زمن الاختبار على معاملات الثبات وفق النظريتين.

٥- المقارنة بين النظرية التقليدية ونظرية استجابة الفقرة (النموذج الثنائي) من خلال الكشف عن أثر زمن الاختبار ومعرفة مدى التوافق والاختلاف فيما بينهما.

أهمية البحث :

تكمن أهمية البحث في الكشف عن أثر الزمن المعطى للاختبار على معالم الفقرة وفق النظرية الحديثة من خلال النموذج ثنائي المعلمة والنظرية التقليدية في القياس، وعلى دقة التقدير، ومن الممكن أن يقدم مؤشرا لزيادة دقة أدوات القياس (الاختبار)، حيث تعد الاختبارات التحصيلية من أكثر أدوات القياس انتشاراً وشيوعاً في مختلف المؤسسات التعليمية، وبالتالي ستعطي موثوقية وصدق لنتائج الاختبار التي تبنى عليها قرارات، منها تتعلق بانتقال الطلاب من مستوى دراسي إلى آخر، وبمستوى إتقان الطلاب للمهارات والمعارف التي اكتسبوها أثناء الدراسة، وتحقيق للعدالة مما ينعكس إيجاباً على قدرة الطلاب على المنافسة في سوق العمل، كما يمكن للبحث الحالي أن يساعد القائمين على تطوير الاختبارات في المؤسسات التعليمية على الاهتمام ببنية الاختبار وإعداده وإخراجه بما يتوافق مع الزمن المحدد للاختبار.

مصطلحات البحث:

أولاً: معالم الفقرة حسب نظرية استجابة الفقرة

تقدير المعالم الفقرات : هي عملية التعبير الكمي عن المعالم.

دقة التقدير :هو تعبير يشير لجودة التقدير التي يميزها الاحتمالية الكبيرة في أن التقدير قريب من القيمة الحقيقية، وذلك باستخدام معدل مربعات الأخطاء أو الأخطاء المعيارية في التقدير.

صعوبة الفقرة : أحد معالم دالة الإستجابة للفقرة، وهي تعادل قيمة القدرة التي تناظر احتمال ٥٠% للإجابة الصحيحة عندما تكون قيمة التخمين يساوي صفراً.

تمييز الفقرة : ميل منحنى خصائص الفقرة عند نقطة تقاطع العمود المقام على محور الصادات من نقطة منتصف احتمال الحصول على الإجابة الصحيحة.

معلمة القدرة: قياس السمات الكامنة عند الأفراد وتقع على مقياس متصل السمة.

ثانياً: معالم الفقرة حسب النظرية التقليدية في القياس:

صعوبة الفقرة : نسبة الطلاب الذين أجابوا عن الفقرة إجابة صحيحة إلى مجموع الطلاب الذين تقدّموا للاختبار.

تمييز الفقرة: قيمة معامل الارتباط بين درجة الفقرة والدرجة الكلية للاختبار.

زمن الاختبار: للاختبار ثلاث صور يكمن الاختلاف في طول زمن الاختبار، حيث تم تحديد زمن كل اختبار بعد استشارة المحكمين واطّلاعهم على نموذج الاختبار وبعد تطبيق الاختبار على العينة الاستطلاعية.

اختبار اختيار من متعدد: اختبار يضم عدد من الفقرات، كل فقرة لها أربعة بدائل، بديل واحد منها صحيح، ويُعطى الفرد العلامة 1 عند اختيار البديل الصحيح، و صفر عند اختيار سواه.

ثبات الاختبار: إحصائياً هو نسبة التباين الحقيقي إلى التباين الكلي، ومعامل ثبات الاختبار هو المؤشر الإحصائي على دقة القياس، ونعبر عنه لفظياً بمعامل ارتباط الاختبار مع نفسه.

محددات البحث :

- **محددات ديموغرافية:** اقتصر البحث على طلاب الصف الخامس الإبتدائي في مدارس التعليم العام(الحكومية والأهلية) التابعة لإدارة التعليم في مدينة تبوك
- **محددات مكانية:** المدارس الابتدائية (الحكومية والأهلية) التابعة لإدارة التعليم بمنطقة تبوك بمركز إشراف مدينة تبوك.
- **محددات زمانية:** تم تطبيق الدراسة في الفصل الدراسي الأول للعام الدراسي ١٤٣٨هـ-١٤٣٩هـ
- **محددات موضوعية:** اقتصر البحث على النموذج اللوجستي ثنائي المعلم لنظرية استجابة الفقرة ونموذج النظرية التقليدية في القياس، وعلى بناءاختبار تحصيلي في الفصل الأول (القيمة المنزلية) من منهج مادة الرياضيات المقرر تدريسه للصف الخامس الإبتدائي في المملكة العربية السعودية ، وكانت الفقرات من نوع الاختيار من متعدد بأربعة بدائل.

متغيرات البحث:

المتغير المستقل: زمن الاختبار ويتكون من ثلاثة نماذج.

المتغيرات التابعة: معالم الفقرة (الصعوبة والتمييز)، الأخطاء المعيارية لتقديرات معالم الفقرات، الثبات للنماذج الثلاثة وفق النظريتين.

الأساليب الإحصائية

١- التحليل العاملي الاستكشافي: للكشف عن أحادية البعد.

٢- استخراج معاملات الصعوبة والتمييز لنموذج استجابة الفقرة باستخدام برنامج (Bilog-mg3).

٣- استخراج معاملات الصعوبة والتمييز لنموذج النظرية التقليدية، باستخدام برنامج (ITEMAN).

٤- تحليل التباين الأحادي للكشف عن الفروق في دقة تقدير معالم الفقرات.

٥- تحليل التباين الأحادي لمتوسطات تقديرات معالم الفقرات، باستخدام برنامج (SPSS) للكشف عن الفروق في معالم الفقرات(الصعوبة والتمييز) للنموذج الواحد.

٦- حساب الثبات بمعادلة (KR-20) بطريقة الاتساق الداخلي، وبطريقة التجزئة النصفية بمعادلة سبيرمان للنظرية التقليدية ، وحساب معامل الثبات الامبريقي وثبات الأفراد وثبات الفقرات وفق نظرية استجابة الفقرة.

٧- للكشف عن الفروق بين معاملات الثبات عن طريق احصائي (M).

٨- لحساب نسبة الإتفاق بين نظرية استجابة الفقرة (النموذج الثنائي) والنظرية التقليدية عن طريق المعادلة التالية: نسبة الإتفاق = $\frac{\text{عدد الفقرات المتفق على استبعادها+عدد الفقرات المتفق على بقائها}}{\text{العدد الفقرات الكلي}}$