رفاد للدراسات والأبحاث
www.refaad.com

المجلة الدولية للدراسات التربوية والنفسية
المجلد٤- العدد٢-٢٠١٨، ص ٢٥٢- ص ٢٧٦
e-ISSN 2520-4149 , p-ISSN 2520-4130

أثرزمن الاختبارعلى تقدير معالم الفقرات وثبات الاختبار دراسة مقارنة: نظرية استجابة الفقرة والنظرية التقليدية

فرحان علي صالح البلوي

ماجستير في القياس والتقويم التربوي- وزارة التعليم
Farhan36qiyas@hotmail.com

صبري محمد إسماعيل عبدالعال

أستاذ مساعد القياس النفسي- جامعة تبوك وجامعة المنوفية
sabdelaal@ut.edu.sa

الملخص:

هدف البحث الكشف عن أثر زمن الاختبار على تقديرات معالم الفقرة وفق نظرية استجابة الفقرة والنظرية التقليدية. حيث تم بناء اختبار تحصيلي اختيار من متعدد (٢٨) فقرة في مقرر الرياضيات. وشُكّلت ثلاث صور للاختبار تختلف في زمن الإجابة. وطُبقت على عينة مكونة من (٤٥١) طالباً، ثم حُلّلت البيانات باستخدام برنامجي(BILOG-MG3) و(ITEMAN).

وأظهرت النتائج اتفاق نظرية استجابة الفقرة والنظرية التقليدية على وجود فروق ذات دلالة إحصائية(α=0.05) في متوسطات تقدير معالم التمييز تُعزى للزمن لصالح الوقت المناسب (٣٠ د). 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 المحسوبة بالطرق التقليدية، وجود فروق ذات دلالة إحصائية في معاملات الثبات لصالح الصورة الثانية (٣٠ د). وأن نسبة الاتفاق بين النظريتين في فرز الفقرات بناء على معالم الصعوبة والتمييز في الصورة الأولى ٨٩%، والصورة الثانية ٩٦%، وفي الصورة الثالثة 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، نظرية استجابة الفقرة، النظرية التقليدية، زمن الاختبار، معالم الفقرة، ثبات الاختبار.

المقدمة:

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

فباستخدام الاختبارات والمقاييس يحصل المعلم على درجات ومعايير، ومتوسطات، وغير ذلك من القيم العددية. فالإنسان الذي هو موضوع القياس التربوي يعدّ أكثر الكائنات تعقيدًا، وتتداخل سماته وخصائصه، ويعتمد بعضها على البعض الآخر، مما يجعل الفصل بينها من أجل دراستها أمرًا معقدًا، فالعلماء لم يتمكنوا من الفصل بدرجة كاملة بين تأثيرات الوراثة والبيئة مثلًا. وكذلك لسنا متأكدين دائمًا 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 أيضًا. إلا إن 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 مدخلان 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 سلبًا 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 للتحقق من صدق البناء وموضوعية القياس (Hattie, 1985).

ومن افتراضات النظرية التقليدية في القياس ونظرية إستجابة الفقرة، افتراض التحرر من السرعة. فإذا تم انتهاكه فإن الاختبار يكون سريعاً، وبالتالي فإن المفحوصين لن يتمكّنوا من الإجابة عن جميع فقرات الاختبار ضمن الزمن المحدد، فإما أن يخمّنوا إجاباتها وخصوصًا الفقرات التي تقع في نهاية الاختبار، أو يتركوا بعض الفقرات دون إجابة، وبالتالي يترتب على سرعة الاختبار تقديرات غير دقيقة لمعالم الفقرة، وقدرة الأفراد، ومعالم الاختبار) Oshima, 1994).

إن افتراض التحرر من السرعة متضمن في افتراض أحادية البعد، فسرعة الاختبار تُضيف بعدًا إحصائيًا آخر لما يقيسه الاختبار، وبالتالي فإن 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 عن كل فقرة في الاختبار، وعندها يُقال بأن الاختبار سريع "Speeded"، وكذلك يسمى الاختبار سريعًا إلى مدى معين، إذا كان المفحوصون 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 نللي(Nunnally, 1978)الوقت المناسب بأنه المدة التي تحتاجها 90% من الأشخاص لإكمال الاختبار. ولقد تضاربت الآراء حول طول الفترة المخصصة للاستجابة في الاختبار من حيث أهميتها وتأثيرها على خصائص الاختبار، حيث يرى نللي Nunnally أن تقصير ثلث الوقت الملائم للاختبار لا يؤثر في خصائص الاختبار السيكومترية. (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.

وتذكر كل من أنستازي ويورينا (Anastasi &Urina,1997/2015:134) مجرد استخدام حدود للزمن ليس مؤشرًا لاختبار سرعة. فإذا انتهى جميع المفحوصين في حدود الزمن المعين، فإن سرعة الإجابة لا تلعب أي دور في تحديد الدرجات. وربما تُتّخذ النسبة المئوية من الأشخاص الذين لم يتمكنوا من الانتهاء من الاختبار كمؤشر أوّلي للسرعة مقابل القوة. غير أنه حتى إذا لم ينته أي مفحوص من الاختبار، فإن دور السرعة رُبّما يكون ضئيلاً. فمثلًا، إذا أكمل مفحوص ٤٠ فقرة من بين ٥٠ فقرة يشتمل عليها الاختبار، فإن الفروق الفردية في السرعة تكون غائبة تمامًا، على الرغم من أنه لم يكن لدى أي مفحوص الوقت لمحاولة الإجابة عن جميع الفقرات.

كل ما سبق يقودنا إلى تساؤل :ماهي حدود زمن الاختبار التي لا تؤثر سلبًا 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 صدقًا وثباتًا وذلك وفق النظرية التقليدية ونظرية إستجابة الفقرة (ثنائي المعلمة) ، ومدى الاتفاق والاختلاف بين هاتين النظريتين.

فقد أجرت بورتوليز و كراوس وبونير (Portolese, Krause, & Bonner, 2016) دراسة بعنوان: هل يتحسن أداء الطلاب بزيادة وقت الاختبار؟ وهدفت الدراسة للكشف عن تأثير زيادة وقت الاختبار على الأداء في الاختبارات. وقد أُجريت الدراسة على (٢٣٤) من طلبة الجامعات بأحد المقررات وتكون الاختبار من (٥٧٦) فقرة ، ثم أُجري اختبار آخر لنفس العينة بفترة أطول للإجابة عن الاختبار. وأوضحت نتائج هذه الدراسة أن زيادة وقت الاختبار لم تحسن من أداء الطلاب.

وأجرى تالينتو- ميلر و جو و هان (Talento-Miler,gou& Han,2013) دراسة هدفت لمعرفة تأثير زمن الاختبار على أداء المفحوصين في اختبار اللغة الإنجليزية، وقد أُجريت الدراسة على أكثر من ١٥ مجموعة مختلفة الجنس والعرق. تكون الاختبار من ٣ أقسام من الاختبارات المكيفة، لكل قسم تم تحديد ٧٥ دقيقة كحد أقصى للإجابة. ويوفر البرنامج زمن الإستجابة لكل متدرب. وأبرز نتائج الدراسة أن نسبة المفحوصين الكوريين الذين أتمّوا الاختبار قبل نفاذ الوقت ٨٧% كأقل مجموعة من المجموعات ١٥ ، إلّا إنه لا توجد فروق ذات دلالة إحصائية.

وقد أجرى كل من جينسين و بيري وكومر (Jensen, Perry & Kummer, 2013) دراسة بعنوان تأثير طول الاختبار على التحصيل والإرهاق الذهني، تكونت العينة من الطلاب الجامعيين في أحد الجامعات الخاصة الكبرى، حيث طُبّق على العينة اختبار في مقرر الأحياء، واستخدم في هذه الدراسة تصميم شبه تجريبي غير متكافئ من أجل جمع البيانات. وأوضحت النتائج أن أداء الطلاب في الاختبارات الأطول يكون أفضل، وأن طول الاختبار لم يؤد إلى انخفاض الأداء بسبب ظروف الإرهاق.

أجرى منايصة (٢٠١٣) دراسة هدفت إلى استقصاء أثر زمن الاختبار على دقة تقديرات معالم الفقرات والأفراد وفق نماذج نظرية استجابة الفقرة، حيث تم بناء اختبار تحصيلي تكوّن من ٣٠ فقرة بشكله النهائي، وشكّلت ثلاث صور للاختبار تختلف فقط في زمن الاختبار فأعطي الطلبة ٤٥ دقيقة للإجابة عن فقرات الاختبار بالصورة الأولى، و ٦٠ دقيقة للإجابة عن فقرات الاختبار بالصورة الثانية، و ٧٥ دقيقة للصورة الثالثة، وطبّقت صور الاختبار على عينة مكونة من (١٥٠٠) طالب؛ بواقع ٥٠٠ طالب وطالبة لكل صورة من صور الاختبار، وأظهرت النتائج وجود فروق ذات دلالة إحصائية بين متوسطات الأخطاء المعيارية في تقدير معالم الصعوبة؛ حيث كانت في الصورة الأولى أكثر دقة ثم الصورة الثانية وأخيراً الصورة الثالثة وفق النموذج ثنائي المعالم. كما أظهرت النتائج وجود فروق ذات دلالة إحصائية بين متوسطات الأخطاء المعيارية في تقدير معالم التمييز للفقرات؛ حيث كانت التقديرات لمعالم تمييز فقرات الصورة الثانية أكثر دقة ثم الصورة الثالثة وأخيراً الصورة الأولى، وكما أظهرت النتائج وجود فروق ذات دلالة إحصائية بين متوسطات الأخطاء المعيارية في تقدير معالم القدرة للأفراد وفق النموذج الثنائي المعالم أكثر دقة لقدرات أفراد الصورة الأولى للاختبار منها لتقديرات أفراد الصورة الثانية والثالثة له؛ مع عدم وجود فروق ذات دلالة إحصائية بين الصورة الثانية والصورة الثالثة.

وأجرى ماندينجان وبريجمان وكاهلان ليتيوسيس وترابانى (Mandinach ,bridgeman,Cahalan-Laitusis&Trapani,2005) دراسة كان هدفها بيان أثر تمديد زمن الاختبار في أداء المفحوصين في اختبار (SAT) وكان عدد المهتمين بالمشاركة من ٢٥٠٠ إلى ٣٢٠٠ طالب من ٤٧ مدرسة عامة و ١٣مدرسة مستقلة و ١٢ مدرسة دينية من ١٩ ولاية وذلك بإعطاء المفحوصين وقتا معياريا، وإعطائهم زمنا ونصف الزمن المعياري مع استراحات أو دونها، أو ضعف الوقت المعياري دون جزء مخصص للاستراحة بين جزئي الاختبار. ويمكن تلخيص نتائج هذه الدراسة كالآتي: استفاد المفحوصون ذوو القدرة المتدنية قليلا من الوقت الإضافي ، ومن المعروف أن أي زمن يمكن وضعه لا يحسن من أداء المفحوصين إذا كانوا لايملكون المعرفة أو المهارة. وساعدت الاستراحات في اختبار المفحوصين بكافة مستوياتهم المختلفة من القدرة . واستفاد المفحوصون ذوو القدرة المتوسطة والعالية من تمديد الوقت.

وفي دراسة لبريجمان وكلاين وهيسنجر (Bridgeman , Cline & Hessinger,2004) التي هدفت إلى التحقق من أن سرعة الاستجابة تلعب دورا مهما في اختبارات(GRE) الاختبار العام (General Test) ، وذلك بإضافة جزء من الاختبار في قسم المصطلحات أو في القسم الكمي من اختبار(GRE) ، إما بنفس الوقت المعياري المحدد، أو بمرة ونصف المرة من الوقت المعياري المحدد. دلت النتائج على العلامات التي تم الحصول عليها من (١٥٩٤٨) مفحوصا أشارت إلى أن الوقت الإضافي رفع حوالي (٧) علامات في مقياس العلامات(٨٠٠-٢٠٠) لعلامات المفحوصين في جزء المصطلحات والجزء الكمي ، وكان المفحوصون ذوو القدرة المتدنية الأكثر تأثرا بتمديد زمن الاختبار ، حيث كانت علامتهم أعلى قليلا عند تمديد زمن الاختبار.

وهدفت دراسة واين (Wayne,2003) إلى مراجعة عدة دراسات سابقة تناولت تأثير اختلاف زمن اختبار المواضيع الكتابية في اختبار (SAT II) على درجات الطلبة بمختلف مستويات قدراتهم. وأهم النتائج التي توصل لها بأن الدراسات التي تم مراجعتها كان محورها ثبات وصدق وصعوبة اختبارات المواضيع الكتابية المعطاة ضمن ظروف زمنية مختلفة، حيث أشارت إلى أن إعطاء المفحوصين مزيدًا من الوقت من الممكن أن يرفع من درجاتهم بالاختبار، ولكن لا يؤثر في ترتيب الطلبة.

وفي دراسة لباورز وفاولز(powers&Fowles,1996) هدفت إلى فحص أثر اختلاف الزمن المحدد في الجزء الكتابي من اختبار(GRE) على أداء المفحوصين . وقد تكونت عينة الدراسة من (٣٠٠) مفحوص من طلبة الدراسات العليا شاركوا في كتابة موضوعين مختلفين ضمن زمنين مختلفين الأول : الوقت المحدد المعياري(٤٠) دقيقة والثاني : (٦٠) دقيقة ؛ أي تمديد الوقت المعياري. وجد الباحثان بأنه يمكن تحسين أداء المفحوصين على الاختبار عندما يعطوا (٦٠) دقيقة بدلا من (٤٠) دقيقة لإكمال الاختبار.

مشكلة البحث :

تذكر انستازي ويورينا (Anastasi &Urina,1997/2015:134) أن مجرد استخدام حدود للزمن ليس مؤشراً لاختبار سرعة، فإذا انتهى جميع المفحوصين في حدود الزمن المعين، فإن سرعة الإجابة لا تلعب أي دور في تحديد الدرجات، وربما تُتّخذ النسبة المئوية من الأشخاص الذين لم يتمكنوا من الانتهاء من الاختبار كمؤشر أوّلي للسرعة مقابل القوة.

تبيّن لنا مما سبق ومن خلال المقدمة أهمية الزمن وعلاقته بافتراض التحرر من السرعة المرتبط بافتراض أحادية البعد. وكما يذكر منايصة (٢٠١٣) أن غالبًا ماهنالك فرصة لانتهاك افتراض السرعة وأن انتهاكه 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 من خلال التحكم بزمن الاختبار.

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 انتشارًا وشيوعًا في مختلف المؤسسات التعليمية، وبالتالي ستعطي موثوقية وصدقا 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 معالم الفقرات : هي عملية التعبير عن المعالم كمياً.

دقة التقدير : هو تعبير عن جودة التقدير التي يميزها الاحتمالية الكبيرة في أن التقدير قريب من القيمة الحقيقية، باستخدام الأخطاء المعيارية في التقدير.

صعوبة الفقرة : هي نقطة على مقياس القدرة، عندما يكون احتمال الاستجابة الصحيحة يساوي ٠,٥٠ في غياب التخمين.

تمييز الفقرة : هو المعلم الذي يحدد ميل أو شدة انحدار منحنى خصائص الفقرة.

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 عددا من الفقرات، كل فقرة لها أربعة بدائل، بديل واحد منها صحيح، ويُعطى الفرد الدرجة واحد 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 بناء اختبار تحصيلي في الفصل الأول (القيمة المنزلية) من منهج مادة الرياضيات المقرر تدريسه للصف الخامس الابتدائي في المملكة العربية السعودية، وكانت الفقرات من نوع الاختيار من متعدد بأربعة بدائل.

مجتمع البحث:

تكون مجتمع البحث من طلاب الصف الخامس الابتدائي بمنطقة تبوك في المدارس الحكومية والأهلية خلال الفصل الدراسي الأول لعام ١٤٣٨-١٤٣٩ هـ. والبالغ عددهم (٥١٨٩) طالب بعد استبعاد مدارس: (قسم تحفيظ القرآن الكريم، والتعليم الأجنبي، والتربية الفكرية).

عينة البحث:

تكونت عينة البحث من (٤٥١) طالباً، تم اختيار العينة بالطريقة العشوائية العنقودية من طلاب الصف الخامس الابتدائي، المنتظمين في مدارس التعليم العام، التابعة للإدارة العامة للتعليم بمنطقة تبوك للعام الدراسي١٤٣٨/١٤٣٩هـ. حيث تم اختيار(١٠) مدارس، كما هو مبيّن بالجدول رقم (٣)

الجدول (١): توزيع أفراد عينة البحث من حيث المدارس وعدد الطلاب

العدد	اسم المدرسة	عدد الطلاب	النسبة المئوية من العينة الكلية
١	جابر بن حيان	٣٠	% ٦,٧
٢	الخالدية	٣٦	% ٨
٣	عبد الله بن رواحه	٣٦	% ٨
٤	عثمان بن عفان	٧٧	% ١٧,١
٥	عقبة بن نافع	٣٧	% ٨,٢
٦	محمد بن القاسم	٥٢	% ١١,٥
٧	معن بن زائدة	٢٨	% ٦,٢
٨	الملك عبد العزيز	٧٣	% ١٦,٢
٩	منارات تبوك	٤٠	% ٨,٨
١٠	مؤتة	٤٢	%٩,٣
	المجموع	٤٥١	% ١٠٠

أداة البحث:

تم بناء اختبار تحصيلي في مادة الرياضيات للوحدة الأولى (القيمة المنزلية) للصف الخامس الابتدائي، حيث تمت صياغة (٣٥) فقرة من نوع الاختيار من متعدد بأربعة بدائل لكل فقرة ، ولكل فقرة إجابة واحدة صحيحة ، ولبناء هذا الاختبار استرشدت بالأسس العامة المتبعة في بناء الاختبارات التحصيلية. وفيما يلي الخطوات والإجراءات التنفيذية لأداة الدراسة :

أ- بناء الأداة

١. تحديد الغرض من الاختبار؛ وهو قياس تحصيل طلاب الصف الخامس في مادة الرياضيات الفصل الأول (القيمة المنزلية).
٢. تحليل المحتوى وكتابة الأهداف السلوكية وعمل جدول مواصفات بحيث تم فيه ربط مستويات الأهداف بمحتوى المادة الدراسية موضوع الاختبار.
٣. كتابة (٣٥) فقرة من نوع اختيار من متعدد أربعة بدائل، لكل فقرة بديل صحيح واحد فقط ، مع الأخذ بالاعتبار الأسس والمعايير الفنية في صياغة هذه الفقرات، حيث روعيت مطابقة كل فقرة للهدف الذي يقيسه من حيث المحتوى ومستوى الهدف، وألا تعتمد إجابة إحدى الفقرات على الإجابة على فقرات أخرى (افتراض الاستقلال الموضعي).
٤. عرض تحليل المحتوى والأهداف وأسئلة الاختبار على (١٨) مُحَكّما من أهل الخبرة والاختصاص للحكم من خلال الاستبيان المعد للتحكيم، وتم الأخذ بآراء المحكمين، حيث قدموا مجموعة من الملاحظات والاقتراحات التي بناء عليها جرى تعديل بعض الفقرات. وفي النهاية أجمع المحكمون على أن الفقرات تؤدي الغرض من الاختبار، وتمثل المجال السلوكي الذي يتم قياسه.
٥. إعداد ورقة تعليمات للاختبار و ورقة نموذج الإجابة.

ب- التجريب الأولي للاختبار:

تم تطبيق الأداة بصورتها الأولية والتي تكوّنت من (٣٥) فقرة على عينة استطلاعية مكونة من (٧٦) طالباً خارج عينة البحث، وذلك لتحديد الزمن المناسب للطلاب للإجابة عن فقرات الاختبار، والتأكد من وضوح الصياغة اللغوية والكشف عن الفقرات الغامضة. وتم ادخال البيانات على برنامج (ITEMAN) الخاص بالنظرية التقليدية لحساب معامل الصعوبة لكل فقرة، من خلال نسبة الطلاب الذين أجابوا عن الفقرة إجابة صحيحة إلى الطلاب الذين حاولوا الإجابة على الفقرة، واستخراج القدرة التمييزية لكل فقرة من خلال معامل ارتباط بوينت بايسيريال Biserial Correlation بين درجة الفقرة والدرجة الكلية على الاختبار، كما في الجدول رقم (٢)

الجدول (٢): معاملات الصعوبة والتمييز لفقرات الاختبار للعينة الاستطلاعية وفق النظرية التقليدية

رقم الفقرة	معامل الصعوبة	معامل التمييز	رقم الفقرة	معامل الصعوبة	معامل التمييز	رقم الفقرة	معامل الصعوبة	معامل التمييز
١	٠,٨٠	٠,٦٤	١٣	٠,٥٧	٠,١٧	٢٥	٠,٦١	٠,٦٠
٢	٠,٦٧	٠,٤٢	١٤	٠,٦١	٠,١٥	٢٦	٠,٥٣	٠,٥٩
٣	٠,٥١	٠,٨٠	١٥	٠,٥٧	٠,٣٨	٢٧	٠,٣٢	٠,٥٢
٤	٠,٢٨	٠,٢٠-	١٦	٠,٥٩	٠,٤٩	٢٨	٠,٤٣	٠,٤١
٥	٠,٦٤	٠,٧٣	١٧	٠,٤٩	٠,٤٣	٢٩	٠,٣٣	٠,٤٤
٦	٠,٥٥	٠,١٦	١٨	٠,٢٥	٠,٤١	٣٠	٠,٥٠	٠,٤٨
٧	٠,٧٥	٠,٥١	١٩	٠,٢٩	٠,٤٨	٣١	٠,٥٧	٠,٥٣
٨	٠,٥٩	٠,٣٦	٢٠	٠,٤٢	٠,٥٤	٣٢	٠,٥٣	٠,٤٠
٩	٠,٥٣	٠,٣٢	٢١	٠,٥٠	٠,٤٧	٣٣	٠,٣٧	٠,٤٧
١٠	٠,٥٧	٠,٤٤	٢٢	٠,٦٧	٠,٥١	٣٤	٠,٢٢	٠,٠٥
١١	٠,٧٠	٠,٤٨	٢٣	٠,٦٦	٠,٣٨	٣٥	٠,٥٥	٠,٤٧
١٢	٠,٣٦	٠,٠٥	٢٤	٠,١٧	٠,٠٩-			

أعلى قيمة لمعامل الصعوبة	٠,٨٠	أعلى قيمة لمعامل التمييز	٠,٨٠
أقل قيمة لمعامل الصعوبة	٠,١٧	أقل قيمة لمعامل التمييز	٠,٢٠-
المتوسط الحسابي لمعاملات الصعوبة	٠,٥١	المتوسط الحسابي لمعاملات التمييز	٠,٤٠

يتضح من الجدول رقم (٢) أن معاملات الصعوبة للفقرات في الاختبار التحصيلي بعد تطبيقه على العينة الاستطلاعية تراوحت قيمتها بين (٠,٨٠ و ٠,١٧) بمتوسط حسابي مقداره (٠,٥١). بينما تراوحت قيم معاملات التمييز بين (٠,٨٠ و - ٠,٢٠) بمتوسط حسابي مقداره (٠,١٧). وتم حذف ٧ فقرات لتدنّي معاملات تمييزهما، والفقرات المحذوفة هي: (٤، ٦، ١٢، ١٣، ١٤، ٢٤، ٣٤). وتم الاحتفاظ ببقية الفقرات البالغ عددها ٢٨ فقرة.

كما تم حساب معامل الثبات بمعادلة كيودر-ريتشاردسون ٢٠ (KR-20)، وبطريقة التجزئة النصفية من خلال معادلة سبيرمان براون بثلاثة طرق(التجزئة العشوائية، النصف الأول والنصف الثاني، الفردي والزوجي) كما في جدول رقم (٣):

الجدول (٣): معاملات ثبات الاختبار (العينة الاستطلاعية)

الطريقة	KR-20	طريقة التجزئة النصفية		
		S-B Random	S-B First-Last	S-B Odd-Even
قيمة معامل الثبات	٠,٨٨	٠,٨٨	٠,٨٥	٠,٨٩

يتضح من الجدول رقم (٣) أن قيم معاملات الثبات لأداة البحث مرتفعة، وهذا يدل على أن أداة البحث تتمتع بدرجات عالية من الثبات والاتساق الداخلي.

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 الاتفاق =

$$\frac{\text{عدد الفقرات المتفق على استبعادها+عدد الفقرات المتفق على بقائها}}{\text{العدد الفقرات الكلي}}$$

النتائج وتفسيرها:

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

ويتفرع عن هذا السؤال 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) ؟

للإجابة عن هذا السؤال تم حساب معاملات الصعوبة لكل فقرة لصور الاختبار الثلاث وفق نموذج النظرية التقليدية كما يوضحها الجدول رقم (٤)

جدول (٤): معاملات الصعوبة لكل فقرة لصور الاختبار الثلاث وفق النظرية التقليدية

رقم الفقرة	الصورة الأولى ٤٠ د	الصورة الثانية ٣٠ د	الصورة الثالثة ٢٠ د
١	٠,٩٦	٠,٨٣	٠,٩٥
٢	٠,٦٨	٠,٧١	٠,٦٩
٣	٠,٥٩	٠,٥٥	٠,٤٩
٤	٠,٤٨	٠,٦١	٠,٥١
٥	٠,٧٩	٠,٧٣	٠,٧٥
٦	٠,٥٧	٠,٥٥	٠,٤٨
٧	٠,٤٨	,٥٧	٠,٤٨
٨	٠,٥١	٠,٥٠	٠,٤١
٩	٠,٧٠	٠,٦٨	٠,٦٨
١٠	٠,٥٨	٠,٥٧	٠,٦٨
١١	٠,٤٤	٠,٥٧	٠,٦١
١٢	٠,٣٨	٠,٤٤	٠,٤٧
١٣	٠,٢٧	٠,٢٥	٠,٣١
١٤	٠,٢٣	٠,٢٤	٠,١٩
١٥	٠,٦٤	٠,٤٩	٠,٥٣
١٦	٠,٥٢	٠,٥٠	٠,٥٦
١٧	٠,٥٠	٠,٥٩	٠,٥٥
١٨	,٦٩	٠,٦١	٠,٧٥
١٩	٠,٥٠	٠,٤٩	٠,٤٧
٢٠	٠,٥٠	٠,٥٥	٠,٤٩
٢١	٠,٢٦	٠,٢٤	٠,١٦
٢٢	٠,٢٩	٠,٣٧	٠,٢٧
٢٣	٠,٢٧	٠,٢٨	٠,٣٧
٢٤	٠,٥٢	٠,٤٥	٠,٥١
٢٥	٠,٥١	٠,٥٦	٠,٤٠
٢٦	٠,٤٤	٠,٤٥	٠,٣٤
٢٧	٠,٣٠	٠,٣٣	٠,٢٧
٢٨	٠,٥٤	٠,٥٥	٠,٥١

تم حساب المتوسطات الحسابية والانحرافات المعيارية لمعاملات الصعوبة وفق النظرية التقليدية في كل صورة من صور الاختبار، كما في الجدول رقم (٥).

الجدول (٥): المتوسطات الحسابية والانحرافات المعيارية لمعاملات الصعوبة وفق النظرية التقليدية

الصورة	العدد	المتوسط الحسابي	الانحراف المعياري
الصورة الأولى ٤٠ د	٢٨	٠,٥٠٧	٠,١٧١
الصورة الثانية ٣٠ د	٢٨	٠,٥٠٨	٠,١٤٨
الصورة الثالثة ٢٠ د	٢٨	٠,٤٩٣	٠,١٧٧

يتبيّن من الجدول رقم (٥) أن متوسطات معاملات الصعوبة لصور الاختبار الثلاث متقاربة جداً (حوالي ٠,٥٠) ويشير كل من (عودة،٢٠١٠: ٢٨٧) (علام،٢٠١٥: ٢٥٣)(النبهان،٢٠٠٤: ١٠٣) أن أفضل معامل صعوبة لاختبار الاختيار من متعدد الذي يشمل أربعة بدائل هو(٠,٦٣) ، من ذلك نستدل على أن الاختبارات بالصور الثلاث صعوبتها فوق المتوسط بناء على المعايير التقليدية. ويتضح أن اختبار الصورة الثالثة (٢٠ د) هو الأصعب بفارق ضئيل ثم الصورة الأولى (٤٠د) ثم الصورة الثانية (٣٠ د).

وللكشف عن الدلالة الإحصائية للفروق بين المتوسطات الحسابية لمعاملات الصعوبة وفق النظرية التقليدية لصور الاختبار باختلاف الزمن، تم استخدام تحليل التباين الأحادي والنتائج كما في الجدول رقم (٦).

جدول (٦): نتائج تحليل التباين الأحادي للكشف عن الفروق بين المتوسطات الحسابية لمعاملات الصعوبة وفق النظرية التقليدية

مصدر التباين	مجموع المربعات	درجات الحرية	متوسط المربعات	F	الدلالة الإحصائية
زمن الاختبار	٠,٠٠٤	٢	٠,٠٠٢	٠,٠٦٩	٠,٩٣٣
الخطأ	٢,٢٣١	٨١	٠,٠٢٨		
الكلي	٢,٢٣٥	٨٣			

يتبيّن من الجدول رقم (٦) عدم وجود فروق ذات دلالة إحصائية عند مستوى (٠,٠٥) بين المتوسطات الحسابية لمعاملات الصعوبة وفق نموذج النظرية التقليدية في صور الاختبار الثلاث تُعزى للزمن.

وتتفق هذه النتائج مع ماذكره ننلي (Nunnally,1978) أن تقصير زمن الاختبار المناسب بمقدار الثُلث لا يؤثر في خصائص الاختبار السيكومترية.

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

للإجابة عن هذا السؤال تم حساب معاملات الصعوبة لكل فقرة لصور الاختبار الثلاث وفق نموذج النموذج اللوجيستي ثنائي المعلم لنظرية استجابة الفقرة كما يوضحها الجدول رقم (٧)

الجدول (٧): معاملات الصعوبة لكل فقرة لصور الاختبار الثلاث وفق نظرية استجابة الفقرة

رقم الفقرة	الصورة الأولى ٤٠ د	الصورة الثانية ٣٠ د	الصورة الثالثة ٢٠ د
١	-٢,٩٠	-١,٣٤	-٣,٢٦
٢	-٠,٧٦	-٠,٨٨	-١,١٩
٣	-٠,٥٦	-٠,٢٦	٠,٠٣
٤	,٠٩٦	-٠,٤٤	-٠,١٠
٥	-٢,١٢	-١,٢٤	-١,٥٦
٦	-٠,٣٦	-٠,٢٧	٠,٠٩٤
٧	-٠,٠١	-٠,١٣	٠,٠٣٧
٨	-٠,١١	٠,٠١	٠,٦٩

٩	-١,٠٥	-٠,٩١	-٠,٩٦
١٠	-٠,٤٣	-٠,٣٤	-٠,٥٦
١١	٠,٢٢	-٠,٣٠	-٠,٥٢
١٢	٠,٥٨	٠,٣٣	٠,٣٩
١٣	١,٢٩	١,٣٠	١,١٩
١٤	١,٧٢	١,٢٥	١,٦٦
١٥	-٠,٧٠	٠,٠٣	-٠,٢٥
١٦	-٠,١٥	٠,٠٠	-٠,٣٠
١٧	-٠,٠٨	-٠,٣٦	-٠,٢٨
١٨	-١,٢٤	-٠,٥٢	-١,٤٥
١٩	-٠,٠٨	٠,٠٥	٠,١١
٢٠	-٠,٠٣	-٠,٢٤	٠,٠٠
٢١	١,٤١	١,٠٩	٢,٠٤
٢٢	١,١٧	٠,٨٢	١,٧٧
٢٣	١,٣٦	١,٢٠	٠,١٤
٢٤	-٠,٢٠	٠,٢١	- ٠,٠٨
٢٥	-٠,٠٧	-٠,٢٦	٠,٤٢
٢٦	٠,٣٥	٠,٢٥	٠,٨٥
٢٧	١,١٢	٠,٧٩	١,٣٢
٢٨	-٠,٣٧	-٠,٢٩	-٠,٠٩

تم حساب المتوسطات الحسابية والانحرافات المعيارية لمعاملات الصعوبة وفق نظرية استجابة الفقرة في كل صورة من صور الاختبار، كما في الجدول رقم (٨)

جدول (٨): المتوسطات الحسابية والانحرافات المعيارية لمعاملات الصعوبة وفق نظرية استجابة الفقرة

الصورة	العدد	المتوسط الحسابي	الانحراف المعياري
الصورة الأولى ٤٠ د	٢٨	- ٠,٠٦٩	١,٠٣٤
الصورة الثانية ٣٠ د	٢٨	- ٠,٠١٧	٠,٧٠٦
الصورة الثالثة ٢٠ د	٢٨	- ٠,٠٠٥	١,١٠٦

يتبيّن من الجدول رقم (٨) أن متوسطات معاملات الصعوبة لصور الاختبار الثلاث متقاربة جداً وتقترب من الصفر، ويدل ذلك على أنها متوسطة الصعوبة كما يذكر أيالا (Ayala,2009/2017,p17) أن الفقرات التي قيمتها حوالي الصفر تعتبر متوسطة الصعوبة. ويتضح أن اختبار الصورة الثالثة (٢٠ د) هو الأصعب ثم الصورة الثانية (٣٠ د) ثم الصورة الأولى (٤٠ د).

وللكشف عن الدلالة الإحصائية للفروق بين المتوسطات الحسابية لمعاملات الصعوبة وفق نظرية استجابة الفقرة لصور الاختبار باختلاف زمن الاختبار، تم استخدام تحليل التباين الأحادي والنتائج كما في الجدول رقم (٩)

الجدول (٩): تحليل التباين الأحادي للكشف عن الفروق بين المتوسطات الحسابية لمعاملات الصعوبة وفق نظرية استجابة الفقرة

مصدر التباين	مجموع المربعات	درجات الحرية	متوسط المربعات	F	الدلالة الإحصائية
زمن الاختبار	٠,٠٦٥	٢	٠,٠٣٢	٠,٠٣٥	٠,٩٦٦
الخطأ	٧٥,٣٥٦	٨١	٠,٩٣		
الكلي	٧٥,٤٢	٨٣			

يتضح من الجدول رقم (٩) عدم وجود فروق ذات دلالة إحصائية عند مستوى (٠,٠٥) بين المتوسطات الحسابية لمعاملات الصعوبة وفق نظرية استجابة الفقرة في صور الاختبار الثلاث تُعزى للزمن.

نجد أن هناك اتفاق بين نظرية استجابة الفقرة النموذج الثنائي مع النظرية التقليدية فيما يخص معلمة الصعوبة بعدم وجود فروق ذات دلالة إحصائية تُعزى لعامل الزمن.

١- **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) ؟**

من خلال برنامج (Bilog-Mg3) تم حساب الخطأ المعياري في تقدير معلم صعوبة كل فقرة من الفقرات لصور الاختبار الثلاث كما في الجدول رقم (١٠)

الجدول (١٠): الأخطاء المعيارية في تقدير معلمة الصعوبة

رقم الفقرة	الخطأ المعياري في تقدير معلمة الصعوبة			رقم الفقرة	الخطأ المعياري في تقدير معلمة الصعوبة		
	٤٠ د	٣٠ د	٢٠ د		٤٠ د	٣٠ د	٢٠ د
١	٠,٠٨٨	٠,١٩٤	١,٣٣٩	١٥	٠,١٩٣	٠,١٦٥	٠,٢٥٣
٢	٠,١٦٧	٠,١٨٤	٠,٣٥٤	١٦	٠,١٣١	٠,١٧٥	٠,١٢٠
٣	٠,٢٣٢	٠,١٨٩	٠,٢٢٣	١٧	٠,١٦٦	٠,١٥١	٠,١٧٥
٤	٠,٢٤١	٠,١٣٢	٠,١٨٣	١٨	٠,٣٧٥	٠,٢٠١	٠,٤٠٥
٥	٠,٦٦٨	٠,٢٨٦	٠,٤٤٨	١٩	٠,٢٢٤	٠,١٤٦	٠,٢٠٦
٦	٠,١٥٨	٠,٢٠٠	٠,٢٥٧	٢٠	٠,٢٠٦	٠,١٧٠	٠,٢٤٧
٧	٠,١٢٦	٠,١٨٧	٠,١٧٧	٢١	٠,٣٥٢	٠,١٩١	٠,٤٤٤
٨	٠,٣٥١	٠,١٢١	٠,١٢٨	٢٢	٠,٣٢٩	٠,٢٧٩	٠,٥٣٧
٩	٠,٢٦٧	٠,٢٢٧	٠,٢٥٦	٢٣	٠,٣٩٤	٠,٢٧٧	١,١١٩
١٠	٠,١٧٢	٠,١٩٨	٠,١٥٦	٢٤	٠,٣٠٣	٠,١٦٥	٠,٢٧٤
١١	٠,١٨١	٠,١٦٨	٠,١٤٠	٢٥	٠,١٨٤	٠,١٥٤	٠,٢٢١
١٢	٠,٢٢٩	٠,٢٠٧	٠,١٦٨	٢٦	٠,٢٨٠	٠,٢٠٠	٠,٣٢٥
١٣	٠,٣٤٢	٠,٢٧٢	٠,٣٧٧	٢٧	٠,٣٣٠	٠,٢٠٢	٠,٣٨٤
١٤	٠,٤٣٢	٠,٢٢٥	٠,٣٤٤	٢٨	٠,١٣٥	٠,٢١٤	٠,١٩٤

تم حساب المتوسطات الحسابية والانحرافات المعيارية للأخطاء المعيارية لتقديرات معلم الصعوبة لفقرات كل صورة من صور الاختبار، كما هو موضح في جدول رقم (١١)

الجدول (١١): المتوسطات الحسابية والانحرافات المعيارية للأخطاء المعيارية لتقديرات معلم الصعوبة

الصورة	العدد	المتوسط الحسابي	الانحراف المعياري
الصورة الأولى ٤٠ د	٢٨	٠,٢٩٥	٠,١٩٥
الصورة الثانية ٣٠ د	٢٨	٠,١٩٩	٠,٠٤٠
الصورة الثالثة ٢٠ د	٢٨	٠,٣٣٧	٠,٢٧٦

يتبين من الجدول (١١) أن المتوسطات الحسابية للأخطاء المعيارية لتقديرات معالم الصعوبة لفقرات كل صورة من صور الاختبار أن الصورة الثانية (٣٠ د) هي الأقل أخطاء وبفارق كبير عن بقية الصور الأخرى، ثم الصورة الأولى(٤٠ د)، والأكثر أخطاء في تقدير معالم صعوبة الفقرات الصورة الثالثة (٢٠ د).

ولفحص الفروق في دقة تقديرات معالم الصعوبة، تم استخدام تحليل التباين الأحادي للأخطاء المعيارية لهذه التقديرات لكل صورة من صورة الاختبار، والجدول (١٢) يوضح ذلك.

الجدول (١٢): تحليل التباين لفحص الفروق في دقة تقديرات معالم الصعوبة

مصدر التباين	مجموع المربعات	درجات الحرية	متوسط المربعات	F	الدلالة الإحصائية
زمن الاختبار	٠,٢٨٢	٢	٠,١٤١	٣,٦٦٢	٠,٠٣٠
الخطأ	٣,١١٧	٨١	٠,٠٣٨		
الكلي	٣,٣٩٨	٨٣			

يتضح من الجدول رقم(١٢) وجود فروق ذات دلالة إحصائية في دقة تقدير صعوبة الفقرات في صور الاختبار الثلاث تُعزى للزمن، ولذلك تم استخدام طريقة المقارنات البعدية بين متوسطات الأخطاء المعيارية لتقدير معالم الصعوبة للصور الثلاث للاختبار لمعرفة أي الفروق دالة إحصائياً ويتبين ذلك من الجدول رقم (١٣)

الجدول (١٣): المقارنات البعدية بين متوسطات الأخطاء المعيارية لتقدير معالم الصعوبة

المقارنات	الفرق بين المتوسطات	الخطأ	الدلالة الإحصائية
٣٠*٤٠	٠,٩٥٩	٠,٠٥٢	٠,١٩٤
٢٠*٤٠	٠,٠٤٣-	٠,٠٥٢	٠,٧١٩
٢٠*٣٠	٠,١٣٩- *	٠,٠٥٢	٠,٠٣٥

يتضح من الجدول رقم (١٣) أن الوقت المناسب (الصورة الثانية ٣٠ د) هي الأكثر دقة في تقدير معالم صعوبة الفقرات من بقية الصور، وتوجد فروق ذات دلالة إحصائية في دقة تقدير معالم الصعوبة بين الصورة الثانية (٣٠ د) والصورة الثالثة (٢٠د) لصالح الصورة الثانية (٣٠ د). وكذلك توجد فروق بين الصورة الثانية (٣٠ د) والصورة الأولى (٤٠ د) لصالح الصورة الثانية ولكن غير دالة إحصائياً. وتوجد فروق في دقة التقدير بين الصورة الأولى(٤٠ د) والصورة الثالثة (٢٠ د) لصالح الصورة الأولى ولكن الفرق غير دال إحصائياً.

وتتفق هذه النتيجة مع دراسة أوشيما(Oshima,1994) ودراسة منايصة(٢٠١٣) في أن زمن الاختبار له تأثير على دقة تقدير معالم صعوبة الفقرات، وتختلف مع دراسة منايصة في الصورة الأكثر دقة، حيث تؤكد نتائج هذه الدراسة أن الوقت الطبيعي (المناسب) هو الأكثر دقة في تقدير معالم الصعوبة، بينما في دراسة منايصة الوقت القصير هو الأكثر دقة في تقدير معالم صعوبة الفقرات.

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) ؟**

للإجابة عن هذا السؤال تم حساب معاملات التمييز لكل فقرة لصور الاختبار الثلاث وفق نموذج النظرية التقليدية كما يوضحها الجدول رقم (١٤)

الجدول (١٤): معاملات التمييزلكل فقرة لصور الاختبار الثلاث وفق النظرية التقليدية

رقم الفقرة	الصورة الأولى (٤٠ د)	الصورة الثانية (٣٠ د)	الصورة الثالثة (٢٠ د)
١	٠,٢١	٠,٤٦	٠,١٥٨
٢	٠,٤٤	٠,٤٦	٠,٢٤٥
٣	٠,٣٢	٠,٤٠	٠,٣١١
٤	٠,٣٠	٠,٥٥	٠,٣٣٢
٥	٠,١٦	٠,٣٤	٠,٢٤١
٦	٠,٤٤	٠,٣٦	٠,٢٥٠
٧	٠,٥٤	٠,٤٠	٠,٣٩
٨	٠,١٢	٠,٣٥	٠,١٧
٩	٠,٣٣	٠,٣٦	٠,٢٨
١٠	٠,٣٧	٠,٣٨	٠,٤٠
١١	٠,٤٠	٠,٤٦	٠,٤١
١٢	٠,٣٩	٠,٣٦	٠,٤١
١٣	٠,٣٥	٠,٣٨	٠,٢٦
١٤	٠,٣٣	٠,٤٤	٠,٤٢
١٥	٠,٣٨	٠,٤٦	٠,٢٣
١٦	٠,٥١	٠,٤٣	٠,٤٩
١٧	٠,٤٠	,٥٠	٠,٣٥
١٨	٠,٢١	٠,٣٩	٠,٢٥
١٩	٠,٣٢	٠,٥٣	٠,٣٣
٢٠	٠,٣٤	٠,٤٤	٠,٢٣
٢١	٠,٣٥	٠,٥٣	٠,٣٧
٢٢	٠,٣٣	٠,٣١	٠,٢٠
٢٣	٠,٣٢	٠,٣٧	٠,١٤
٢٤	٠,١٣	٠,٤٧	٠,١٩
٢٥	٠,٣٦	٠,٤٨	٠,٣٥
٢٦	٠,٢٥	٠,٣٨	٠,٣٣
٢٧	٠,٣٥	٠,٣٥	٠,٣٢
٢٨	٠,٤٧	٠,٣٤	٠,٣٤

تم حساب المتوسطات الحسابية والانحرافات المعيارية لمعاملات التمييز وفق النظرية التقليدية في كل صورة من صور الاختبار، كما في الجدول رقم (١٥).

الجدول (١٥): المتوسطات الحسابية والانحرافات المعيارية لمعاملات التمييز وفق النظرية التقليدية

الصورة	العدد	المتوسط الحسابي	الانحراف المعياري
الصورة الأولى ٤٠ د	٢٨	٠,٣٣٥	٠,١٠٣
الصورة الثانية ٣٠ د	٢٨	٠,٤١٩	٠,٠٦٣
الصورة الثالثة ٢٠ د	٢٨	٠,٢٩٩	٠,٠٨٩

يتضح من الجدول رقم (١٥) أن متوسطات معاملات التمييز وفق النظرية التقليدية تتفاوت حسب كل صورة، فنجد أن متوسط معامل التمييز بالصورة الثانية (٣٠ د) هو الأعلى ثم الصورة الأولى (٤٠ د) وأخيراً الصورة الثالثة (٢٠ د). وللكشف عن الدلالة الإحصائية للفروق بين المتوسطات الحسابية لمعاملات التمييز وفق النظرية التقليدية لصور الاختبار باختلاف الزمن، تم استخدام تحليل التباين الأحادي والنتائج كما في الجدول رقم (١٦)

الجدول (١٦): تحليل التباين الأحادي للكشف عن الفروق بين متوسطات معاملات التمييزوفق النظرية التقليدية

مصدر التباين	مجموع المربعات	درجات الحرية	متوسط المربعات	F	الدلالة الإحصائية
زمن الاختبار	٠,٢١٤	٢	٠,١٠٧	١٤,٣٢١	٠,٠٠٠
الخطأ	٠,٦٠٦	٨١	٠,٠٠٧		
الكلي	٠,٨٢	٨٣			

يتضح من الجدول رقم (١٦) وجود فروق ذات دلالة إحصائية بين متوسطات معاملات التمييز تُعزى للزمن، ولذلك تم استخدام طريقة المقارنات البعدية بين متوسطات معاملات التمييز للصور الثلاث للاختبار لمعرفة أي الفروق دالة إحصائياً ويتبين ذلك من الجدول رقم (١٧)

الجدول (١٧): المقارنات البعدية بين متوسطات معاملات التمييزوفق النظرية التقليدية

المقارنات	الفرق بين المتوسطات	الخطأ	الدلالة الإحصائية
٣٠*٤٠	٠,٠٨٤-*	٠,٠٢٣	٠,٠٠٢
٢٠*٤٠	٠,٠٣٧	٠,٠٢٣	٠,٢٩١
٢٠*٣٠	٠,١٢١ *	٠,٠٢٣	٠,٠٠٠

يلاحظ من الجدول رقم (١٧) أن الفرق بين متوسطي معاملات التمييز كان دال إحصائياً بين الصورتين الأولى(٤٠د) والثانية (٣٠ د) ولصالح الصورة الثانية، أي أن فقرات الصورة الثانية أكثر قدرة على التمييز بين الطلاب من فقرات الصورة الأولى. كما أشارت النتائج إلى وجود دلالة إحصائية للفرق بين متوسطات معاملات التمييز وفق النظرية التقليدية للصورتين الثانية (٣٠ د) والثالثة (٢٠ د) لصالح الصورة الثانية. في حين أن الفرق بين متوسطات معاملات التمييز في الصورة الأولى (٤٠ د) والصورة الثالثة (٢٠ د) غير دالة إحصائياً.

تختلف النتائج المتعلقة بمعلمة التمييز مع ما ذكره نلي(Nnnnally,1978) حيث إن عامل الزمن أدى لوجود فروق ذات دلالة إحصائية بين متوسطات معاملات التمييز في كلا النموذجين.

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

للإجابة عن هذا السؤال تم حساب معاملات التمييز لكل فقرة لصور الاختبار الثلاث وفق النموذج اللوجستي ثنائي المعلم لنظرية استجابة الفقرة كما يوضحها الجدول رقم (١٨)

الجدول (١٨): معاملات التمييز لكل فقرة لصور الاختبار الثلاث وفق نظرية استجابة الفقرة

رقم الفقرة	الصورة الأولى ٤٠ د	الصورة الثانية ٣٠ د	الصورة الثالثة ٢٠ د
١	١,١٧	١,٨٩٨	٠,٩٨٩
٢	١,٤٤	١,٣٧٢	٠,٧٧٤
٣	٠,٨٣	٠,٩٨٥	٠,٧٩٠
٤	٠,٧٢	١,٦٥٠	٠,٩٩٩
٥	٠,٦٥	٠,٩٦٥	٠,٧٧١
٦	١,٢٢	٠,٩٠٤	٠,٦٦٨
٧	١,٦٩	٠,٩٧	١,١٤
٨	٠,٤٧	٠,٨٣	٠,٥١
٩	١,٠٢	١,٠١	٠,٩٥
١٠	١,١١	٠,٩٤	١,٣١
١١	١,١٦	١,١٨	١,٤٧
١٢	٠,٩٦	٠,٨٩	١,١٣
١٣	٠,٨٧	١,٠٣	٠,٧٣
١٤	٠,٨١	١,٢٣	١,٠٥
١٥	١,١٢	١,١٥	٠,٦٨
١٦	١,٦٠	١,٠٧	١,٧٣
١٧	١,١٣	١,٣٧	١,٠٥
١٨	٠,٦٩	١,٠١	٠,٨٥
١٩	٠,٧٨	١,٤٦	٠,٩٤
٢٠	٠,٨٦	١,١٢	,٧٠٢
٢١	٠,٨٥	١,٦١	٠,٩٣
٢٢	٠,٨٦	٠,٧٧	٠,٥٩
٢٣	٠,٨٥	٠,٩٦	٠,٥٢
٢٤	٠,٥٦	١,١٦	٠,٦١
٢٥	٠,٩٩	١,٢٦	٠,٨٧
٢٦	٠,٦٦	٠,٩٥	٠,٨٤
٢٧	٠,٨٦	١,١٢	٠,٨٢
٢٨	١,٥٢	٠,٨٥	٠,٩٣

تم حساب المتوسطات الحسابية والانحرافات المعيارية لمعاملات التمييز وفق نظرية استجابة الفقرة في كل صورة من صور الاختبار، كما في الجدول (١٩)

الجدول (١٩): المتوسطات الحسابية والانحرافات المعيارية لمعاملات التمييزوفق نظرية استجابة الفقرة

الصورة	العدد	المتوسط الحسابي	الانحراف المعياري
الصورة الأولى ٤٠ د	٢٨	٠,٩٧٨	٠,٣٠٨
الصورة الثانية ٣٠ د	٢٨	١,١٣٣	٠,٢٦٩
الصورة الثالثة ٢٠ د	٢٨	٠,٩٠٦	٠,٢٧٦

يتضح من الجدول رقم (١٩) أن متوسطات معاملات التمييز وفق نظرية استجابة الفقرة النموذج ثنائي المعلم تتفاوت حسب كل صورة، فنجد أن متوسط معامل التمييز بالصورة الثانية (٣٠ د) هو الأعلى ثم الصورة الأولى (٤٠ د) وأخيراً الصورة الثالثة (٢٠ د).

وللكشف عن الدلالة الإحصائية للفروق بين المتوسطات الحسابية لمعاملات التمييز وفق نظرية استجابة الفقرة لصور الاختبار باختلاف الزمن، تم استخدام تحليل التباين الأحادي والنتائج كما في الجدول رقم(٢٠).

الجدول (٢٠): تحليل التباين الأحادي للكشف عن الفروق بين المتوسطات الحسابية لمعاملات التمييزوفق نظرية استجابة الفقرة

مصدر التباين	مجموع المربعات	درجات الحرية	متوسط المربعات	F	الدلالة الإحصائية
زمن الاختبار	٠,٧٥٤	٢	٠,٣٧٧	٤,٦٤٣	٠,٠١٢
الخطأ	٦,٥٧٥	٨١	٠,٠٨١		
الكلي	٧,٣٢٩	٨٣			

يتضح من الجدول رقم (٢٠) وجود فروق ذات دلالة إحصائية بين متوسطات معاملات التمييز تُعزى للزمن، ولذلك تم استخدام طريقة المقارنات البعدية بين متوسطات معاملات التمييز للصور الثلاث للاختبار لمعرفة أي الفروق دالة إحصائياً ويتبين ذلك من الجدول (٢١).

الجدول (٢١): نتائج المقارنات البعدية لمتوسطات معاملات التمييزوفق نظرية استجابة الفقرة النموذج الثنائي

المقارنات	الفرق بين المتوسطات	الخطأ	الدلالة الإحصائية
٣٠*٤٠	٠,١٥٥-	٠,٠٧٦	٠,١٣٤
٢٠*٤٠	٠,٠٧٣	٠,٠٧٦	٠,٦٣٧
٢٠*٣٠	٠,٢٢٧ *	٠,٠٧٦	٠,٠١٥

يلاحظ من الجدول رقم (٢١) أن الفرق بين متوسطي معاملات التمييز كان دال إحصائياً بين الصورتين الثانية(٣٠ د) والثالثة (٢٠ د) ولصالح الصورة الثانية، أي أن فقرات الصورة الثانية أكثر قدرة على التمييز بين الطلاب من فقرات الصورة الثالثة. في حين أن بقية الفروق بين متوسطات معاملات التمييز غير دالة إحصائياً.

نجد أن هناك اتفاق بين نظرية استجابة الفقرة النموذج الثنائي مع النظرية التقليدية فيما يخص معلمة التمييز بوجود فروق ذات دلالة إحصائية تُعزى لعامل الزمن، وكان هنالك اتفاق بالنتائج أيضاً، حيث إن الفرق لصالح الوقت المناسب (٣٠ د) وفق النموذجين .

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) ؟

للإجابة عن هذا السؤال تم حساب الخطأ المعياري في تقدير معلم التمييز لكل فقرة من الفقرات لصور الاختبار الثلاث من خلال برنامج (Bilog-Mg3) كما في الجدول رقم (٢٢)

الجدول (٢٢): الخطأ المعياري في تقدير معلم التمييز لكل فقرة من الفقرات لصور الاختبار الثلاث

رقم الفقرة	الخطأ المعياري في تقدير معلمة التمييز			رقم الفقرة	الخطأ المعياري في تقدير معلمة التمييز		
	٤٠ د	٣٠ د	٢٠ د		٤٠ د	٣٠ د	٢٠ د
١	٠,٥٨١	٠,٤٦٠	٠,٤٩١	١٥	٠,٢٩١	٠,٢٢٥	٠,١٧٦
٢	٠,٣٢٣	٠,٢٦٨	٠,٢٢٤	١٦	٠,٣٤٢	٠,٢١٨	٠,٤٤٦
٣	٠,٢٠٦	٠,٢٠٥	٠,٢٠٤	١٧	٠,٢٥٨	٠,٢٦٦	٠,٢٥٧
٤	٠,١٦٥	٠,٣٥٢	٠,٢٣٩	١٨	٠,١٧٩	٠,٢٠٢	٠,٢٦٣
٥	٠,٢١٢	٠,٢١٤	٠,٢٢٦	١٩	٠,١٩٢	٠,٢٧٧	٠,٢٥٦
٦	٠,٢٧٧	٠,١٨٥	٠,١٦٤	٢٠	٠,٢٠١	٠,٢٢١	٠,١٩٩
٧	٠,٣٧٧	٠,١٩٩	٠,٢٩٦	٢١	٠,١٩٣	٠,٣٠٤	٠,٢١٥
٨	٠,١٢٧	٠,١٧٣	٠,١٤٤	٢٢	٠,١٩٧	٠,١٦٩	٠,١٥٥
٩	٠,٢٩٤	٠,٢٢١	٠,٢٤٧	٢٣	٠,١٩٨	٠,٢٠٨	٠,٥٢٢
١٠	٠,٢٧١	٠,١٩١	٠,٢٧٠	٢٤	٠,١٤٦	٠,٢٢٦	٠,١٦٦
١١	٠,٢٥٦	٠,٢٢٠	٠,٣٩٢	٢٥	٠,٢٣٩	٠,٢٦١	٠,١٧٦
١٢	٠,٢٣٠	٠,١٨٤	٠,٢٥٤	٢٦	٠,١٧٠	٠,١٩٣	٠,٢١٩
١٣	٠,١٩٧	٠,٢١١	٠,١٨٢	٢٧	٠,٢١٠	٠,٢٢٣	٠,٢٠٤
١٤	٠,٢٠٠	٠,٢٦١	٠,٢٣٢	٢٨	٠,٣٥٧	٠,١٨٥	٠,٢٥١

كما تم حساب المتوسطات الحسابية والانحرافات المعيارية للأخطاء المعيارية لتقديرات معلم التمييز لفقرات كل صورة من صور الاختبار، كما هو موضح في جدول رقم (٢٣)

الجدول(٢٣): المتوسطات الحسابية والانحرافات المعيارية للأخطاء المعيارية لتقديرات معالم التمييز

الصورة	العدد	المتوسط الحسابي	الانحراف المعياري
الصورة الأولى ٤٠ د	٢٨	٠,٢٤٦	٠,١٩٥
الصورة الثانية ٣٠ د	٢٨	٠,٢٣٣	٠,٠٤٠
الصورة الثالثة ٢٠ د	٢٨	٠,٢٥٣	٠,٢٧٦

يتبين من الجدول رقم (٢٣) أن المتوسطات الحسابية للأخطاء المعيارية لتقديرات معالم التمييز لفقرات كل صورة من صور الاختبار أن الصورة الثانية (٣٠ د) هي الأقل أخطاء بين بقية الصور الأخرى أي أنها الأكثر دقة في تقدير معلم التمييز، ثم الصورة الأولى(٤٠ د)، والأكثر أخطاء في تقدير معالم صعوبة التمييز الصورة الثالثة (٢٠د).

ولفحص الفروق في دقة تقديرات معالم التميز، تم استخدام تحليل التباين الأحادي للأخطاء المعيارية لهذه التقديرات لكل صورة من صورة الاختبار، والجدول رقم (٢٤) يوضح ذلك.

الجدول (٢٤): تحليل التباين لفحص الفروق في دقة تقديرات معالم الصعوبة

مصدر التباين	مجموع المربعات	درجات الحرية	متوسط المربعات	F	الدلالة الإحصائية
زمن الاختبار	٠,٠٠٦	٢	٠,٠٠٣	٠,٣٨٩	٠,٦٧٩
الخطأ	٠,٥٧٩	٨١	٠,٠٧		
الكلي	٠,٥٨٥	٨٣			

يشير الجدول رقم (٢٤) إلى عدم وجود فروق ذات دلالة إحصائية في دقة تقدير معالم التمييز بين صور الاختبار تُعزى لزمن الاختبار.

وتختلف هذه النتيجة مع دراسة منايصة (٢٠١٣) التي بيّنت وجود فروق ذات دلالة إحصائية في دقة تقدير معلم التمييز تُعزى لزمن الاختبار. وتتفق معها بالصورة الأكثر دقة في تقدير معلم التمييز(الوقت المناسب).

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) ؟

للإجابة عن هذا السؤال يلزمنا حساب معاملات الثبات لكل صورة من صور الاختبار وفق الطرق التقليدية.

تم حساب معاملات الثبات لصور الاختبار وفق النظرية التقليدية بأربع طُرق كما يوضحها الجدول رقم (٢٥)، وهذه الطرق هي:

١. معادلة كيودر-ريتشاردسون ٢٠ (KR-20)
٢. التجزئة النصفية (العشوائية) باستخدام معادلة سبيرمان- براون (S-B Random).
٣. التجزئة النصفية (النصف الأول- الثاني) بمعادلة سبيرمان- براون (S-B First-Last)
٤. التجزئة النصفية (فردي- زوجي) بمعادلة سبيرمان- براون (S-B Odd-Even)

الجدول (٢٥): معاملات الثبات وفق النظرية التقليدية لصور الاختبار

الطريقة	KR-20	التجزئة العشوائية	التجزئة (الأول-الثاني)	التجزئة (فردي-زوجي)
الصورة الأولى ٤٠ د	٠,٨٢	٠,٨١	٠,٨١	٠,٨٤
الصورة الثانية ٣٠ د	٠,٨٨	٠,٨٨	٠,٨٤	٠,٩١
الصورة الثالثة ٢٠ د	٠,٧٩	٠,٨١	٠,٦٨	٠,٧٩

يتضح من الجدول رقم (٢٥) اتفاق الطرق المستخدمة لحساب معامل الثبات وفق النظرية التقليدية أن الصورة الثانية (الوقت المناسب ٣٠ د) هي الأعلى ثباتاً، ثم الصورة الأولى (الوقت الطويل ٤٠ د)، وأخيراً الصورة الثالثة (الوقت القصير ٢٠ د).

ولفحص دلالة الفروق بين معاملات الثبات لصور الاختبار، تم عمل المقارنات الثنائية عن طريق الإحصائي M ، من خلال المعادلة التالية :

$$M = \frac{J-1}{1}\{\sum_{k=1}^{k} B_k - \frac{\{\sum_{k=1}^{k} B_k \left(1-r_k\right)^{\frac{-1}{3}}\}}{\{\sum_{k=1}^{k} B_k (1-r_k)^{\frac{-2}{3}}\}}\}$$

حيث تُحسب B_k من المعادلة التالية: $B_{k=\frac{(9n_k-11)^2}{n_k}}$

حيث ، J: عدد فقرات الاختبار، r_k: معامل الثبات، n_k : عدد أفراد العينة

حيث تم استخراج قيمة M لكل مقارنة بين صور الاختبار لجميع الطرق المستخدمة لحساب الثبات كما في الجدول رقم (٢٦)

الجدول(٢٦): قيم الإحصائي M للمقارنات الثنائية بين معاملات الثبات لصور الاختبار الثلاث

	40*30	40*20	30*20
KR-20	*٥,١١	١,٠٦	*١٠,٧
التجزئة العشوائية	*٧,١١	٠,٠٣	*٦,٢٤
التجزئة(الأول-الثاني)	١,٠٦	*٩,٦٤	*١٦,٨٣
التجزئة (فردي-زوجي)	*٨,٧٤	٣,٥٩	*٢٢,٩٢

• ذات دلالة إحصائية عند ٠,٠٥

من قيم M المحسوبة في الجدول رقم (٢٦) ومقارنتها بالقيمة الحرجة لتوزيع كاي تربيع (X^2) عند درجة حرية واحدة = ٣,٨٤ ، نجد أن بعض قيم M تجاوزت القيمة الحرجة، وبالتالي يدل ذلك على وجود فروق ذات دلالة إحصائية عند (٠,٠٥)، وذلك يدل على أن معاملات الثبات وفق الطرق التقليدية تتأثر بعامل الزمن، والفروق كالتالي:

١. معامل الثبات KR-20:

وجود فروق ذات دلالة إحصائية بين الصورة الأولى والصورة الثانية، لصالح الصورة الثانية. وجود فروق ذات دلالة إحصائية بين الصورة الثانية والصورة الثالثة، لصالح الصورة الثانية.

٢. معامل ثبات التجزئة النصفية (العشوائي) (S-B Random):

وجود فروق ذات دلالة إحصائية بين الصورة الأولى والصورة الثانية، لصالح الصورة الثانية. وجود فروق ذات دلالة إحصائية بين الصورة الثانية والصورة الثالثة، لصالح الصورة الثانية.

٣. معامل ثبات التجزئة النصفية (النصف الأول- الثاني) (S-B First- Last):

وجود فروق ذات دلالة إحصائية بين الصورة الأولى والصورة الثالثة، لصالح الصورة الأولى. وجود فروق ذات دلالة إحصائية بين الصورة الثانية والصورة الثالثة، لصالح الصورة الثانية.

٤. معامل ثبات التجزئة النصفية (فردي- زوجي) (S-B Odd- Even):

وجود فروق ذات دلالة إحصائية بين الصورة الأولى والصورة الثانية، لصالح الصورة الثانية. وجود فروق ذات دلالة إحصائية بين الصورة الثانية والصورة الثالثة، لصالح الصورة الثانية.

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

للإجابة عن هذا السؤال يلزمنا حساب معاملات الثبات لكل صورة من صور الاختبار وفق نظرية استجابة الفقرة

١. حساب معاملات الثبات للصورة الأولى (٤٠ د):

أ- معامل ثبات الأفراد (عينة الدراسة):

تراوحت قيم معلم الصعوبة للفقرات في الصورة الأولى بين (-٢,٩٠) و (١,٧٢) بوسط حسابي (-٠,٠٦٩)، وانحراف معياري (١,٠٣٤)، وقد تراوحت قيم الخطأ المعياري في تقدير معلم الصعوبة للفقرات بين (٠,١٣) و (١,٠٩) بوسط حسابي (٠,٢٩٥) وانحراف معياري (٠,١٩٥).

معامل الفصل للأفراد (G) = $\frac{\text{الانحراف المعياري لقيم معلم الصعوبة للفقرات}}{\text{الوسط الحسابي للخطأ المعياري في تقدير معلم الصعوبة}}$ = ٣,٥١

معامل ثبات الأفراد للصورة الأولى = $\frac{G^2}{1+G^2}$ = ٠,٩٢

ب- معامل ثبات الفقرات:

بلغت قيمة الوسط الحسابي لقدرات المفحوصين (-٠,٠٤٣) بانحراف معياري (١,٠٧)، كما بلغت قيمة الوسط الحسابي للخطأ المعياري في تقدير القدرة (٠,٤٤) بانحراف معياري (٠,١٩).

معامل الفصل للفقرات (G) = $\frac{\text{الانحراف المعياري لقيم تقدير القدرة}}{\text{الوسط الحسابي للخطأ المعياري في تقدير القدرة}}$ = ٢,٤٣

معامل ثبات الفقرات للصورة الأولى = $\frac{G^2}{1+G^2}$ = ٠,٨٦

ج- معامل الثبات التجريبي:

بلغت قيمة معامل الثبات التجريبي للصورة الأولى (٠,٨٣) وفق برنامج (Bilog-Mg3)

١- حساب معاملات الثبات للصورة الثانية (٣٠ د):

أ- معامل ثبات الأفراد (عينة الدراسة):

تراوحت قيم معلم الصعوبة للفقرات في الصورة الأولى بين (-١,٣٤) و (١,٣٠) بوسط حسابي (-٠,٠١٧)، وانحراف معياري (٠,٧٠٦)، وقد تراوحت قيم الخطأ المعياري في تقدير معلم الصعوبة للفقرات بين (٠,١٣) و (٠,٢٩) بوسط حسابي (٠,١٩٩) وانحراف معياري (٠,٠٤).

معامل الفصل للأفراد (G) = الانحراف المعياري لقيم معلم الصعوبة للفقرات / الوسط الحسابي للخطأ المعياري في تقدير معلم الصعوبة = ٣,٥٥

معامل ثبات الأفراد للصورة الأولى = $\frac{G^2}{1+G^2}$ = ٠,٩٣

ب- معامل ثبات الفقرات:

بلغت قيمة الوسط الحسابي لقدرات المفحوصين (٠,٠٠٦) بانحراف معياري (١,١٢)، كما بلغت قيمة الوسط الحسابي للخطأ المعياري في تقدير القدرة (٠,٣٩٩) بانحراف معياري (٠,١٦).

معامل الفصل للفقرات (G) = الانحراف المعياري لقيم تقدير القدرة / الوسط الحسابي للخطأ المعياري في تقدير القدرة = ٢,٨١

معامل ثبات الفقرات للصورة الأولى = $\frac{G^2}{1+G^2}$ = ٠,٨٩

ج- معامل الثبات التجريبي:

بلغت قيمة معامل الثبات التجريبي للصورة الثانية (٠,٨٧) وفق برنامج (Bilog-Mg3)

٢- حساب معاملات الثبات للصورة الثالثة (٢٠ د):

أ- معامل ثبات الأفراد (عينة الدراسة):

تراوحت قيم معلم الصعوبة للفقرات في الصورة الثالثة بين (-٣,٢٦) و (٢,٠٤) بوسط حسابي (-٠,٠٠٥)، وانحراف معياري (١,١٠٦)، وقد تراوحت قيم الخطأ المعياري في تقدير معلم الصعوبة للفقرات بين (٠,١٢) و (١,٣٤) بوسط حسابي (٠,٣٣٨) وانحراف معياري (٠,٢٧٦).

معامل الفصل للأفراد (G) = الانحراف المعياري لقيم معلم الصعوبة للفقرات / الوسط الحسابي للخطأ المعياري في تقدير معلم الصعوبة = ٣,٢٧

معامل ثبات الأفراد للصورة الأولى = $\frac{G^2}{1+G^2}$ = ٠,٩١

ب- معامل ثبات الفقرات:

بلغت قيمة الوسط الحسابي لقدرات المفحوصين (- ٠,٠١٦) بانحراف معياري (١,١٢)، كما بلغت قيمة الوسط الحسابي للخطأ المعياري في تقدير القدرة (٠,٤٧٣) بانحراف معياري (٠,٢٢).

معامل الفصل للفقرات (G) = الانحراف المعياري لقيم تقدير القدرة / الوسط الحسابي للخطأ المعياري في تقدير القدرة = ٢,٣٧

معامل ثبات الفقرات للصورة الأولى = $\frac{G^2}{1+G^2}$ = ٠,٨٥

ج- معامل الثبات التجريبي:

بلغت قيمة معامل الثبات التجريبي للصورة الثانية (٠,٨٢) وفق برنامج (Bilog-Mg3)

الجدول (٢٧): يوضح معاملات الثبات وفق نظرية استجابة الفقرة لصور الاختبار الثلاث

	الثبات التجريبي	ثبات الأفراد	ثبات الفقرات
الصورة الأولى ٤٠ د	٠,٨٣	٠,٩٢	٠,٨٦
الصورة الثانية ٣٠ د	٠,٨٧	٠,٩٣	٠,٨٩
الصورة الثالثة ٢٠ د	٠,٨٢	٠,٩١	٠,٨٥

يتضح من الجدول رقم (٢٧) اتفاق الطرق المستخدمة لحساب معامل الثبات وفق نظرية استجابة الفقرة أن الصورة الثانية (الوقت المناسب ٣٠ د) هي الأعلى ثباتاً، ثم الصورة الأولى (الوقت الطويل ٤٠ د)، وأخيراً الصورة الثالثة (الوقت القصير ٢٠ د).

ولفحص فروق معاملات الثبات بين صور الاختبار، تم عمل المقارنات الثنائية عن طريق الإحصائي M من خلال المعادلة التالية:

$$M = \frac{J-1}{1}\{\sum_{k=1}^{k} B_k - \frac{\{\sum_{k=1}^{k} B_k \left(1-r_k\right)^{\frac{-1}{3}}\}}{\{\sum_{k=1}^{k} B_k (1-r_k)^{\frac{-2}{3}}\}}$$

حيث تُحسب B_k من المعادلة التالية:

$$B_{k=\frac{(9n_k-11)^2}{n_k}}$$

حيث ، J: عدد فقرات الاختبار، r_k: معامل الثبات، n_k : عدد أفراد العينة

وتم استخراج قيمة M لكل مقارنة بين صور الاختبار لجميع الطرق المستخدمة لحساب الثبات كما في الجدول (٢٨)

الجدول(٢٨): قيم الإحصائي M للمقارنات الثنائية بين معاملات الثبات لصور الاختبار الثلاث

	40*30	40*20	30*20
الثبات التجريبي	٢,٧٦	٠,٣٣	٣,٧٦
ثبات الأفراد	٠,٦٣	٠,٥٠	٢,٢٥
ثبات الفقرات	٢,٠٦	٠,١٦	٣,٤٢

من الجدول رقم (٢٨) يتضح أن جميع قيم M المحسوبة من خلال المعادلة لم تتجاوز القيمة الحرجة لتوزيع كاي تربيع (X^2) عند درجة حرية واحدة = ٣,٨٤ ، ومن ذلك نجد أن الفروق بين معاملات الثبات بين صور الاختبار غير دالة إحصائياً، أي أن عامل الزمن ليس له تأثير ذو دلالة إحصائية على معاملات الثبات المحسوبة وفق نظرية استجابة الفقرة.

تتفق معاملات الثبات المحسوبة بالطرق التقليدية مع معاملات الثبات المحسوبة وفق نظرية استجابة الفقرة على أن معاملات ثبات الصورة الثانية (الوقت المتوسط ٣٠ د) هي الأعلى ثباتًا، ثم معاملات ثبات الصورة الأولى (الوقت الطويل ٤٠ د)، وأخيرًا معاملات ثبات الصورة الثالثة(الوقت القصير ٢٠ د). ويكمن الاختلاف بينهما في دلالة الفروق، حيث يتضح لنا أن الفروق بين معاملات الثبات لصور الاختبار بالنظرية التقليدية ذات دلالة إحصائية، بينما الفروق في معاملات الثبات لصور الاختبار بنظرية استجابة الفقرة غير دالة إحصائياً.

السؤال الرابع: هل يُفرز النموذج اللوجستي ثنائي 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) ؟

للإجابة عن هذا السؤال تم حساب معاملات الصعوبة والتمييز لكل فقرة من فقرات صور الاختبار وفق نظرية استجابة الفقرة (ثنائي المعلم) والنظرية التقليدية. حيث تم اعتماد المعيار التالي لحذف الفقرات أو بقائها:

١- نظرية استجابة الفقرة: أي فقرة معامل صعوبتها خارج المدى (-١,٥٠ ، ١,٥٠) تُحذف. أي فقرة معامل تمييزها أقل من (٠,٤٠) تُحذف.

٢- النظرية التقليدية: أي فقرة معامل صعوبتها خارج المدى (٠,٢٠ ، ٠,٨٠) تُحذف. أي فقرة معامل تمييزها أقل من (٠,٢٥) تُحذف.

أولاً: الصورة الأولى(٤٠ د):

تم حساب معاملات الصعوبة والتمييز لكل فقرة من فقرات الصورة الأولى للاختبار وفق النموذج اللوجستي ثنائي المعلم والنظرية التقليدية كما في الجداول رقم (١٠) و(١٣) و(٢٠) و(٢٤)

وبناء على المعيار المحدد فإن الفقرات المستبعدة وفق النموذج اللوجستي ثنائي المعلم وعددها ٤ هي: (١،٥، ١٤، ٢٤) أي ما نسبته ١٤%. وأن الفقرات المستبعدة وفق النظرية التقليدية وعددها ٥ وهي: (١ ، ٥، ٨، ١٨، ٢٤) أي ما نسبته ١٨%. واتفق النموذج الثنائي مع المؤشرات التقليدية على حذف ٣ فقرات هي:(١، ٥، ٢٤)، وعلى مطابقة ٢٢ فقرة وهي (٢، ٣، ٤، ٦، ٧، ٩،١٠، ١١، ١٢،١٣ ، ١٥،١٦ ،١٧، ١٩، ٢٠، ٢١، ٢٢، ٢٣، ٢٥، ٢٦، ٢٧، ٢٨).

ويمكن حساب نسبة الاتفاق بين النموذجين بالصورة الأولى (٤٠ د) من خلال المعادلة التالية:

$$\text{نسبة الاتفاق} = \frac{\text{عدد الفقرات المتفق على استبعادها} + \text{عدد الفقرات المتفق على بقائها}}{\text{العدد الفقرات الكلي}} = \frac{٣+٢٢}{٢٨} = ٨٩\%$$

أي أن ٨٩% من الفقرات (٢٥ فقرة) اتفق النموذجان في الحكم عليها، حيث اتفقا على استبعاد ٣ فقرات وبقاء ٢٢ فقرة من فقرات اختبار الصورة الأولى (٤٠ د)، أما النسبة المتبقية وهي ١١% من فقرات الاختبار (٣ فقرات) فقد اختلفا بالحكم عليها، فقرتين استبعدتها المؤشرات التقليدية ولم يستبعدها النموذج الثنائي، وفقرة واحدة استبعدها النموذج الثنائي ولم تستبعدها المؤشرات التقليدية.

ثانياً: الصورة الثانية (٣٠ د):

تم حساب معاملات الصعوبة والتمييز لكل فقرة من فقرات الصورة الأولى للاختبار وفق النموذج اللوجستي ثنائي المعلم والنظرية التقليدية كما في الجداول رقم (١٠) و(١٣) و(٢٠) و(٢٤)

وبناء على المعيار المحدد فإن النموذج للوجستي ثنائي المعلم في الصورة الثانية (٣٠د) لم يستبعد أي فقرة. وأن الفقرات المستبعدة وفق النظرية التقليدية فقرة واحدة (رقم ١) أي ما نسبته ٤%. واتفق النموذج الثنائي مع المؤشرات التقليدية على مطابقة ٢٧ فقرة وهي(٢، ٣، ٤، ٥، ٦، ٧،٨، ٩،١٠، ١١، ١٢،١٣، ١٤،١٥،١٦، ١٧، ١٨، ١٩، ٢٠، ٢١، ٢٢، ٢٣، ٢٤، ٢٥، ٢٦، ٢٧، ٢٨).

ويمكن حساب نسبة الاتفاق بين النموذجين بالصورة الثانية (٣٠د) من خلال المعادلة التالية:

$$\text{نسبة الاتفاق} = \frac{\text{عدد الفقرات المتفق على استبعادها} + \text{عدد الفقرات المتفق على بقائها}}{\text{العدد الفقرات الكلي}} = \frac{٢٧+٠}{٢٨} = ٩٦\%$$

أي أن ٩٦% من الفقرات (٢٧ فقرة) اتفق النموذجان في الحكم عليها، حيث اتفقا على بقاء ٢٧ فقرة من فقرات اختبار الصورة الثانية (٣٠ د)، أما النسبة المتبقية وهي ٤% من فقرات الاختبار (فقرة واحدة) فقد اختلفا بالحكم عليها، استبعدتها المؤشرات التقليدية ولم يستبعدها النموذج الثنائي الذي لم يستبعد أي فقرة من فقرات الصورة الثانية (٣٠ د) للاختبار.

ثالثاً: الصورة الثالثة (٢٠د):

تم حساب معاملات الصعوبة والتمييز لكل فقرة من فقرات الصورة الأولى للاختبار وفق النموذج اللوجستي ثنائي المعلم والنظرية التقليدية كما في الجداول رقم (١٠) و(١٣) و(٢٠) و(٢٤)

وبناء على المعيار المحدد فإن الفقرات المستبعدة وفق النموذج اللوجستي ثنائي المعلم وعددها ٤ هي: (١،٥، ١٤، ٢١) أي ما نسبته ١٤%. وأن الفقرات المستبعدة وفق النظرية التقليدية وعددها ١١ وهي: (١ ،٢ ،٥، ٨ ،١٤، ١٥، ١٨،٢٠، ٢٢، ٢٣، ٢٤) أي ما نسبته ٣٩%. واتفق النموذج الثنائي مع المؤشرات التقليدية على حذف ٣ فقرات هي:(١، ٥، ١٤)، وعلى مطابقة ١٦ فقرة وهي (٣، ٤، ٦، ٧، ٩ ،١٠، ١١، ١٢، ١٣ ، ١٦ ،١٧، ١٩، ٢٥، ٢٦، ٢٧، ٢٨).

ويمكن حساب نسبة الاتفاق بين النموذجين بالصورة الثانية (٢٠ د) من خلال المعادلة التالية:

$$\text{نسبة الاتفاق} = \frac{\text{عدد الفقرات المتفق على استبعادها} + \text{عدد الفقرات المتفق على بقائها}}{\text{العدد الفقرات الكلي}} = \frac{١٦+٣}{٢٨} = ٦٨\%$$

أي أن ٦٨% من الفقرات (١٩ فقرة) اتفق النموذجان في الحكم عليها، حيث اتفقا على استبعاد ٣ فقرات وبقاء ١٦ فقرة من فقرات اختبار الصورة الثالثة (٢٠ د)، أما النسبة المتبقية وهي ٣٢% من فقرات الاختبار (٩ فقرات) فقد اختلفا بالحكم عليها، ٨ فقرات (٢، ٨، ١٥، ١٨، ٢٠، ٢٢، ٢٣، ٢٤) استبعدتها المؤشرات التقليدية ولم يستبعدها النموذج الثنائي، وفقرة واحدة (٢١) استبعدها النموذج الثنائي ولم تستبعدها المؤشرات التقليدية.

يتبيّن مما سبق أن نسبة الاتفاق بين النظريتين في فرز الفقرات بناء على معاملات الصعوبة والتمييز تتأثر بزمن الاختبار، حيث تكون نسبة الاتفاق بينهما عالية جداً في الوقت المناسب، وتنخفض عند زيادة وقت الاختبار، أما عند تقليص وقت الاختبار بمقدار الثُلث تكون نسبة الاتفاق منخفضة جدًا، وذلك بسبب دقة تقدير هذه المعالم من خلال الأخطاء المعيارية فكما ظهر بالنتائج أن الصورة الثانية (الزمن المناسب) هي الأكثر دقة، ثم الصورة الأولى (الزمن الطويل) ثم الصورة الثالثة (الزمن القصير).

ومن ذلك يتضح لنا أهمية زمن الاختبار وتأثيره على معالم الفقرات ودقة تقديرها، فالزمن المناسب للاختبار الذي يتناسب مع عدد الفقرات وخصائص المفحوصين تكون فيه دقة التقدير لمعلمة الصعوبة ومعلمة التمييز أعلى من دقة تقديرها في الزمن القصير والزمن الطويل بسبب قلة الأخطاء المعيارية في تقدير هذه المعالم في الزمن المناسب.

أما بالنسبة لتفسير أن ثبات الاختبار في الزمن المناسب حسب النظرية التقليدية هو الأعلى ثباتًا، ووجود فروق ذات دلالة إحصائية لصالح الزمن المناسب، إنه تم حساب الثبات بطرق الاتساق الداخلي وكما يذكر يذكر(الشيخ وآخرون ،٢٠٠٩: ٢٣٦) تعتمد هذه الطريقة في حساب الثبات على الاتساق في أداء الطلاب على الاختبار من فقرة لأخرى، فكلما زاد الاتساق بين هذه الفقرات زاد ثبات الاختبار ككل. وبالتالي فإن زمن الاختبار له تأثير على اتساق هذه الفقرات من خلال أداء الطلاب، فالزمن القصير يؤثر على أداء الطالب وخصوصًا في الفقرات الأخيرة مما يقلل من اتساق هذه الفقرات، والزمن الطويل كذلك قد يكون له تأثير على أداء الطالب من خلال تغيير بعض الإجابات الصحيحة بإجابات خاطئة بسبب طول فترة الاختبار مما يقلل من اتساق فقرات الاختبار وبالتالي تأثيرها على الثبات.

التوصيات:

في ضوء النتائج التي توصلت إليها الدراسة الحالية، يمكن الخروج بالتوصيات التالية:

١. ينبغي تحديد زمن الاختبار المناسب بدقة عالية وفق الأسس والمعايير والالتزام به، لما له من تأثير على خصائص الاختبار والفقرات ودقة تقديرها.
٢. إطالة زمن الاختبار عن الوقت المناسب له تأثير سلبي على خصائص الاختبار والفقرات ودقة تقديرها، بالإضافة إلى أن فيه هدرًا لوقت وجهد الطالب والمراقب، وزيادة حالات الغش. كما لوحظ من خلال تصحيح الصورة الأولى (الوقت الطويل) كثير من الحالات التي يتم فيها تغيير اختيار الإجابة الصحيحة بإجابة خاطئة وتأثيرها على درجة الطالب، ويؤكد ذلك المتوسط الحسابي للدرجات الكلية بالصورة الثانية (الوقت المناسب) هو الأعلى، ثم الصورة الأولى (الوقت الطويل)، وأخيراً الصورة الثالثة (الوقت القصير).
٣. تقصير زمن الاختبار عن الوقت المناسب له تأثير سلبي على خصائص الاختبار والفقرات ودقة تقديرها، وقد يؤدي لارتفاع القلق عند الطلاب وبالتالي يؤثر على أدائهم.
٤. تنويع أساليب التقويم في المرحلة الابتدائية واستخدام الاختبارات الموضوعية وخاصة (الاختيار من متعدد) في الصفوف العليا.
٥. توعية الطلاب بأهمية الاختبارات، وتدريبهم على إدارة وقت الاختبار.

البحوث المقترحة:

١. الكشف عن أثر زمن الاختبار على تحصيل طلاب المرحلة الجامعية.
٢. الكشف عن تأثير زمن الاختبار على أداء الطلاب في الاختبارات المقالية أو الاختبارات الشفوية.

المراجع:

أولاً: المراجع العربية:

١. أيالا، دي (ترجمة الكيلاني، عبدالله ؛ البرصان، إسماعيل) (٢٠١٧). النظرية والتطبيق في نظرية استجابة الفقرة. الرياض: جامعة الملك سعود.
٢. الشمراني، محمد (٢٠١٤). استخدام نظرية استجابة الفقرة والنظرية التقليدية في تقدير الخصائص الإحصائية السيكومترية لاختبار مهارات التفكير والتعلم لطلبة السنة التحضيرية الجامعية .التربية (جامعة الأزهر) – مصر. م ٢. ع ١٥٨ . ص ص ٧١٧- ٨٠٢ .
٣. الشيخ ، تاج السر؛ أخرس، نائل، وعبدالمجيد، بثينة (٢٠٠٩) . القياس والتقويم التربوي .ط ٥ . الرياض : مكتبة الرشد.
٤. علام ، صلاح الدين (٢٠١٥ أ)، القياس والتقويم التربوي والنفسي . ط ٦ . القاهرة : دار الفكر العربي.
٥. علام، صلاح الدين (٢٠١٥ ب) ، القياس والتقويم في العملية التدريسية. ط ٥ . عمّان : دار المسيرة.
٦. العمري، حمزة (١٩٨٥) . أثر طول وقت الاختبار التحصيلي على أداء المفحوصين عليه و على خصائصه السيكومترية عند مفحوصين مختلفين في درجات مخاطرتهم و مستوى قلقهم العام ، رسالة ماجستير غير منشورة. جامعة اليرموك، إربد .
٧. عودة ، أحمد سليمان (٢٠١٠) ، القياس والتقويم في العملية التدريسية. ط٤. إربد : دار الأمل.
٨. كروكر، ليندا و ألجينا، جيمس (ترجمة الحموري، هند و دعنا، زينات) (٢٠١٧). مدخل إلى نظرية القياس التقليدية والمعاصرة. عمّان: دار الفكر.
٩. منايصة، هيثم (٢٠١٣) .أثر زمن الاختبار على دقة تقديرات معالم الفقرات والأفراد وفق نماذج نظرية استجابة الفقرة ، رسالة دكتوراه غير منشورة. جامعة اليرموك، إربد .
١٠. النهان ، موسى (٢٠٠٤) ، أساسيات القياس في العلوم السلوكية. ط ١. عمّان : دار الشروق.

ثانياً: المراجع الاجنبية:

[1] B. Bridgeman, F. Cline & J. Hessinger, (2004). Effect of Extra Time on Verbal and Quantitative GRE Scores. Applied Measurement In Education. Vol.17.No. 1,pp. 25-37.
[2] J. Hattie, (1985). Methodology review: Assessing unidimensionality of tests and items. Psychological Measurement.Vol. 9.No. 2 ,pp.139-164.

[3] J. Jensen, D. Berry & T. Kummer, (2013). Investigating the Effects of Exam Length on Performance and Cognitive Fatiague. PLOS ONE.Vol.8.No.8,pp.1-9.

[4] E.B. Mandinach, Bridgeman, B., Cahalan-Laitusis, C., & Trapani, C. (2005). The impact of extended time on SAT test performance. research report no. 2005-8. ETS RR-05-20 College Board. 250 Vesey Street, New York, NY 10281. Retrieved from: https://search.proquest.com/docview/1773216240?accountid=142908

[5] A.A. Mroch, (2006). A randomized response deletion approach for reducing the effects of test speededness on item response theory item parameter estimates. Doctoral dissertation.University of Wisconsion- Madison.

[6] T.C. Oshima, (1994). the effect of speededness on parameter estimation in item response theory. Journal of Educational measurement.Vol. 31.No.1,pp. 200-219.

[7] L. Portolese, J. Krause & J. Bonner, (2016). Timed Online Tests: Do Students Perform Better With More Time?. American Journal of Distance Education, 30(4), 264-271.Powers, D.E., & Fowles, M.E. (1996).Effect of applying different time limits to a proposed GRE writing test. Journal of Educational Measurement.Vol. 33.No.4,pp.433-452.

[8] E. Talento-Miller, F. Guo & K.T.Han, (2013). Examining Test Speededness by Native Language. International Journal Of Testing.Vol. 13.No.2,pp. 89-104.

رفاد للدراسات والأبحاث
www.refaad.com

المجلة الدولية للدراسات التربوية والنفسية
المجلد٤- العدد٢- ٢٠١٨، ص ٢٥٢- ص ٢٧٦
e-ISSN 2520-4149 , p-ISSN 2520-4130

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

Farhan Ali Saleh Al - Balawi

Master of Educational Measurement and Evaluation - Ministry of Education
Farhan36qiyas@hotmail.com

Sabri Mohamed Ismail Abdel Aal

Assistant Professor of Psychometry - Tabuk University and Menoufia University
sabdelaal@ut.edu.sa

Abstract:

The research aimed to explore the impacts of test time on estimations of items aspects in context with Item Response Theory, and Classical Theory. Where a multi choice test of (28) items was constructed. Three test images were formed, varying in response time. Were applied sample of students. Then analyzed .The results showed that there is concordance of Theories on the existence statistically significant differences in the average of the estimation of discrimination Coefficient and the accuracy of the difficulty coefficient. As the results of the M-tests showed differences between the reliability coefficients of the three images. Also, the ratio between two theories in the classification of items based on the characteristics of difficulty and discrimination in the images respectively 89%, 96% and 68%. As research results, we recommend test time to be carefully selected as its impaction on test characteristics, items and estimation accuracy .

Keywords: Classical Test Theory, Item Response Theory, Multiple Choice Test, Parameters Item, Time Test, Reliability Test