



كلية التربية

إدارة: البحوث والنشر العلمي (المجلة العلمية)

=====

**تأثير مستوى طول الاختبار في دقة تقدير معلّمي الصعوبة
والتمييز باستخدام النموذج ثنائي المعلمة في مبحث
الرياضيات للصف الثامن الأساسي**

اعداد

د/ محمد عواد سالم القضاة

وزارة التربية والتعليم الاردنية

masa33198133@gmail.com

«المجلد التاسع والثلاثون- العدد الثاني- فبراير ٢٠٢٣ م»

http://www.aun.edu.eg/faculty_education/arabic

الملخص

هدفت هذه الدراسة لتحديد اثر طول الاختبار على معلمتي التمييز والصعوبة للمفردة باستخدام النموذج الثنائي البارامتر، وتكونت عينة الدراسة من ٩٠٠ طالب، حيث قام الباحث بتطبيق ثلاثة اختبارات رياضيات للصف الثامن الاساسي مكونه من (٦٠، ٤٥، ٣٠) مفردة اختيار من متعدد، مختزله من الاختبار الطويل (٦٠) مفردة، وتوصلت النتائج إلى وجود فروق ذات دلالة إحصائية عند تقدير متوسطات معلمة تمييز المفردة تعزى لاختلاف طول الاختبار، اي ان متوسطات قيم معلمة تمييز المفردة قد تغيرت بتغير طول الاختبار، ولمعرفة الاختلاف بين متوسطات تقدير معلمة تمييز المفردة باختلاف طول الاختبار تبين بعد اجراء المقارنات البعدية ان الفروق لصالح الاختبار الطويل (٦٠ مفردة) والاختبار متوسط الطول (٤٥ مفردة)، وذلك لصالح الاختبار الطويل، و توصلت نتائج الدراسة ايضاً إلى وجود فروق في المتوسطات الحسابية لتقديرات معلمة صعوبة المفردة ذات دلالة إحصائية تعزى لتغير طول الاختبار، اي ان متوسطات تقديرات معلمة صعوبة المفردة قد تغيرت بتغير طول الاختبار. ولمعرفة الاختلاف بين متوسطات تقديرات معلمة صعوبة المفردة باختلاف طول الاختبار تبين بعد اجراء المقارنات البعدية ان الفروق لصالح الاختبار الطويل (٦٠ مفردة) مقابل الاختبار المتوسط (٤٥ مفردة) والقصير (٣٠ مفردة).

الكلمات المفتاحية: صعوبة المفردة، تمييز المفردة، طول الاختبار، النموذج الثنائي البارامتر.

Abstract

This study aimed to determine the effect of the length of the test on the parameters of discrimination and difficulty of the individual using the two-parameter model, and the study sample consisted of 750 students., where the researcher applied three mathematics tests for the ninth grade, consisting of (60, 45, 30) multiple-choice items Shortened from the long test (60) items. The results showed that there were statistically significant differences when estimating the averages of the individual discrimination parameter due to the difference in the length of the test, that is, the average values of the individual discrimination parameter changed with the change in the length of the test, and to know the difference between the averages of the estimation of the individual discrimination parameter with the difference in the length of the test, it was found after making post comparisons that the differences In favor of the long test (60 items) and the medium length test (45 items), And this is in favor of the long test, and the results of the study also found that there are differences in the arithmetic averages of the estimates of the single difficulty parameter with statistical significance due to the change in the length of the test. That is, the averages of the individual difficulty parameter estimates have changed with the change in the length of the test. In order to find out the difference between the averages of the individual difficulty parameter estimations according to the difference in the length of the test, it was found after conducting the dimensional comparisons that the differences are in favor of the long test (60 items) versus the medium test (45 items) and the short test (30 items).

Key words: difficulty of the item, discrimination of the item, the length of the test, the two-parameter model.

مقدمة

كانت الاختبارات وما زالت من أهم الوسائل وأكثرها فاعلية بالنسبة لجميع أطراف العملية التعليمية لما تقدمه من معلومات مفيدة في تحديد ما يمتلكه الأفراد من سمات وقدرات، بما يحقق درجة عالية من الدقة والموضوعية في اختيار الأفراد أو تصنيفهم. وتعد الاختبارات من الوسائل المهمة والضرورية لمعظم عمليات إجراءات التقويم التي تعتمد عليها المؤسسات التعليمية وهي أيضا من الأدوات الرئيسية والفعالة في يد المعلم ذلك أنها تيسر الحصول على معلومات عن التلاميذ سواء عند التخطيط للتعليم الصفّي أو عند تقييم الطرق أو الاستراتيجيات التي يستخدمها في التدريس، وتساعد في معرفة مدى نموهم وتقدمهم أو تأخرهم، ومن ثم رصد مواطن القوة والضعف عند التلاميذ ومراعاة الإجراءات التدريسية أثناء البرنامج التعليمي (أبو ناهية، ١٩٩٤). ورغم تقدم العلوم التربوية والنفسية عامة وعلم القياس خاصة إلا أنه تبيين قصور استخدام الاختبارات في قياس قدرات الطلبة وتحديد مستوياتهم، حيث وجهت لنظرية القياس التقليدية العديد من الانتقادات بسبب اعتمادها على معيار معين في تفسير الدرجات، هذا المعيار هو معيار الجماعة في الاختبارات جماعية المعيار، ومعيار المحك أو مستوى الإتقان في الاختبارات محكية المرجع. ومن أهم هذه الانتقادات عدم تحقيقها لموضوعية القياس المتمثلة في اعتماد القياس على خصائص الاختبار المستخدم (صدق، ثبات، تمييز) أي اختلاف نتيجة القياس باختلاف الاختبار المستخدم وباختلاف مستوى جماعة المختبرين (عينة الاختبار)، (عبد الوهاب، ٢٠٠٧). هذا ما دفع العديد من الباحثين والمهتمين بالقياس إلى البحث عن بديل للنظرية التقليدية، ألا وهو بناء نظرية حديثة تتسم بالمرونة والقدرة على تلافي المشكلات التي تعاني منها أساليب القياس التقليدية وقد انبثقت عن نظرية الاستجابة للمفردة مجموعة من نماذج الاستجابة للمفردة لتتناسب السمات المختلفة المراد قياسها وأنواع وخصائص المفردات المستخدمة في هذا القياس. وهذه الخصائص هي معامل صعوبة المفردة difficulty Item على أنه نسبة الطلاب الذين أجابوا إجابة صحيحة عن المفردة من بين أفراد العينة في ضوء النظرية الكلاسيكية، وفي ضوء نماذج نظرية الاستجابة للمفردة، هي نقطة على متصل القدرة التي تقابل احتمالية الإجابة ٠.٥، أي أنها نقطة على متصل القدرة تمثل إجابة الفرد عن المفردة إجابة صحيحة، وتميز المفردة وهو عبارة عن قدرة المفردة على إظهار الفروق بين مستويات السمة أو القدرة التي يقيسها المقياس، ويعبر عنه بميل Slope المنحني المميز للمفردة الاختبارية (علام، ٢٠٠٥). وأجريت العديد من الدراسات لتحديد العوامل المؤثرة في تقدير معالم المفردات وقدرات الأفراد. حيث قام هيلين وليساك ودراسجو (Hulin, Lisak & Drasgo, 1982) بدراسة أثر حجم العينة وطول الاختبار في دقة تقدير معالم المفردات ومعلم القدرة

للأفراد باستخدام النموذج ثنائي وثلاثي المعلمة. و لتحقيق هدف الدراسة تم استخدام أسلوب المحاكاة وتوليد أربعة مستويات من حجم العينة ٢٠٠، ٥٠٠، ١٠٠٠، ٢٠٠٠ فرد، وثلاثة مستويات لطول الاختبار ١٥، ٣٠، ٦٠ مفردة، أظهرت نتائج الدراسة انه عند استخدام النموذج ثلاثي المعلمة كان معامل الارتباط بين معلمة التمييز الحقيقي و المقدر ٠.٣٦ عندما كان طول الاختبار ١٥ مفردة وحجم عينة ٢٠٠ فرد، في حين كان معامل الارتباط يساوي ٠.٨٦ عندما كان طول الاختبار ٦٠ مفردة وحجم العينة ٢٠٠٠ فرد، كما أشارت النتائج إلى أن معامل الارتباط بين معلمة الصعوبة الحقيقية و المقدرة بلغ ٠.٧٤ عندما كان طول الاختبار ١٥ مفردة وحجم العينة ٢٠٠ فرد، وازداد هذا المعامل ليصبح ٠.٨٦ عند استخدام حجم عينة ٢٠٠٠ فرد و طول اختبار ٦٠ مفردة.

ويعتبر طول الاختبار من اهم العوامل التي تؤثر في ثبات الاختبار طبقاً للنظرية التقليدية، وهناك دراسات لدراسة أثر طول الاختبار على دقة تقديرات معالم المفردات وقدرة الأفراد عند استخدام النماذج ثنائية التدرج وخصوصاً النموذج اللوجستي ثنائي المعالم. حيث أجرى باستري (Bastari, 2000) دراسة هدفت إلى بحث اثر طول الاختبار في مادة اللغة الانجليزية في مراحل تعليمية متعددة وحجم العينة ونمط المفردة ثنائية الدرجة الصواب والخطأ، ومتعددة الدرجات اسئلة الاجابة القصيرة، على تقدير معالم المفردة باستخدام النموذج ثنائي المعلمة، ولدراسة أثر طول الاختبار استخدمت الدراسة اختبارين الاول يتكون من ٣٠ مفردة والثاني يتكون من ٥٠ مفردة من نوع الاختيار من متعدد، واسفرت النتائج أنه كلما زاد طول الاختبار كلما زادت دقة التقدير لمعلمي الصعوبة والتمييز للمفردة.

وهناك دراسات استخدمت أسلوب المحاكاة لتوليد البيانات، لدراسة أثر طول الاختبار على دقة تقديرات معالم المفردات وقدرة الأفراد عند استخدام النماذج ثنائية التدرج. منها دراسة الدرايع (٢٠٠١) التي هدفت إلى التحقق من فاعلية النموذج اللوغاريتمي نموذج راش، في دقة تقدير معلمة الصعوبة للمفردة عند تغير مستوى طول الاختبار، واستخدام عدد مفردات الاختبار ٢٥، ٥٠، ٣٠٠ مفردة، وتم تطبيق بيانات مولدة باستخدام أسلوب المحاكاة، وبينت الدراسة وجود فروق جوهرية في دقة تقدير قدرة الفرد تعزى إلى متغير طول الاختبار. وفيما يتعلق بدقة تقدير معلمة الصعوبة، بينت النتائج بأن هنالك فروقاً جوهرية تعزى لطول الاختبار في دقة تقدير معلمة صعوبة المفردة.

وهدفت دراسة الشمراني (٢٠١٦) إلى فحص أثر حجم العينة وطول الاختبار على تقدير معالم المفردة والخطأ المعياري باستخدام النموذج الثلاثي المعلمة، ولتحقيق هدف الدراسة تم توليد بيانات ثلاث عينات بأحجام مختلفة (٢٠٠، ٦٠٠، ١٠٠٠)، وثلاثة مستويات لطول الاختبار (٣٠، ٤٠، ٦٠) مفردة، باستخدام برنامج Wengen3. وتوصلت نتائج الدراسة إلى وجود فروق ذات دلالة إحصائية عند تقدير متوسطات معلم الصعوبة تعزى لاختلاف حجم العينة، ولا توجد فروق تعزى لطول الاختبار أو للتفاعل بين حجم العينة وطول الاختبار. وكانت النتائج بينت وجود فروق عند تقدير معلم التمييز تعزى إلى اختلاف أحجام العينات، وكانت الفروق لصالح حجم العينة المتوسطة والكبيرة. وعدم وجود فروق في متوسطات معلم التمييز تعزى إلى طول الاختبار ولا يوجد أثر للتفاعل بين طول الاختبار وحجم العينة، وأشارت النتائج إلى أن تقدير معلم التمييز يزداد بزيادة حجم العينة وكذلك بزيادة طول الاختبار. وتوصلت أيضاً إلى وجود فروق عند تقدير معلمة التخمين تعزى إلى حجم العينة، والي طول الاختبار وكانت لصالح حجم العينة الصغيرة، وهذا يعني أن معامل التخمين يرتفع كلما قل حجم العينة، وأشارت النتائج أن معامل التخمين يرتفع كلما قل عدد مفردات الاختبار.

وفي دراسة بني عطا (٢٠١٤) التي هدفت إلى تقصي دقة تقدير النموذج اللوجستي ثلاثي المعلمة لمعالم المفردة وقدرة الأفراد، في ضوء تغير طول الاختبار. ولتحقيق الهدف من الدراسة ولدت بيانات ثنائية الاستجابة بواقع ٥٠ مرة لستة مستويات من طول الاختبار ١٠، ٢٥، ٥٠، ٧٥، ١٠٠، مفردة من خلال استخدام برنامج WINGEN وباستخدام برنامج Bilog- Mg حُلَّت البيانات المولدة. كشفت نتائج الدراسة عن وجود أثر ذي دلالة إحصائية لطول الاختبار في دقة تقديرات معالم المفردات وقدرة الأفراد. وكشفت النتائج أيضاً أن الوسط الحسابي لقيم RMAS لمعالم المفردات وقدرة الأفراد أخذ بالتناقص عندما زاد طول الاختبار على ٥٠ مفردة.

وتعد دقة تقدير معالم المفردات والأفراد من القضايا المهمة في نظرية الاستجابة للمفردة وكذلك لحجم العينة وطول الاختبار أثره في دقة التقدير، وبسبب أن غالبية الدراسات تعتمد غالباً على بيانات غير تجريبية، هذا ومن خلال إطلاع الباحث على الأطر النظرية الخاصة بهذا المجال، تبين عدم وجود محكات دقيقة وأسس علمية واضحة يمكن من خلالها تحديد العدد الأمثل للمفردات التي يمكن من خلالها تقدير القدرة المستهدف قياسها من الاختبار وتقدير معالم المفردة، فإن الباحث يرى إمكانية محاولة لتحديد عدد المفردات الاختبارية اللازمة في الاختبار، وهذا ما حدا بالباحث إلى دراسة واستقصاء هذه الفكرة من خلال دراسة تجريبية.

مشكلة الدراسة:

تعد الاختبارات التحصيلية الوسيلة الأكثر استخداماً لقياس تحصيل الطلبة في جميع مراحل التعليم، وأكثر انماط هذه الاختبارات استخداماً هو اختبار الاختيار من متعدد، ويسعى معد هذه الأنواع من الاختبارات إلى توفير كافة الظروف التي تجعل القياس أكثر دقة وموضوعية، خاصةً العدد الأمثل من المفردات، ويهدف هذا البحث إلى معرفة تأثير طول الاختبار التحصيلي على خصائص مفرداته (معلمي الصعوبة والتمييز) في النموذج الثنائي للمعلمين، حيث تكمن مشكلة الدراسة في الإجابة على الاسئلة الآتية:

١. ما تأثير طول الاختبار على معلم تمييز المفردة في النموذج الثنائي؟

٢. ما تأثير طول الاختبار على معلم صعوبة المفردة في النموذج الثنائي؟

أهداف الدراسة:

تهدف هذه الدراسة إلى:

التعرف على تأثير طول الاختبار على معلم تمييز المفردة في النموذج الثنائي؟

التعرف على تأثير طول الاختبار على معلم صعوبة المفردة في النموذج الثنائي؟

فروض الدراسة

١. لا توجد فروق ذات دلالة إحصائية عند مستوى الدلالة ($\alpha = 0.05$) بين متوسطات معلمة تمييز المفردة في المستويات المختلفة من طول الاختبار باستخدام النموذج الثنائي.

٢. لا توجد فروق ذات دلالة إحصائية عند مستوى الدلالة ($\alpha = 0.05$) بين متوسطات معلمة صعوبة المفردة في المستويات المختلفة من طول الاختبار باستخدام النموذج الثنائي.

أهمية الدراسة

يمكن تحديد أهمية الدراسة في الجوانب التالية

محاولة الإسهام في إلقاء الضوء على مدى أثر طول الاختبار على دقة تقدير معلمي الصعوبة والتمييز، ومحاولة تزويد الباحثين الممارسين المستخدمين لنظرية الاستجابة للمفردة والمهتمين في تطوير الاختبارات باستخدام طول الاختبار المناسب.

المفاهيم الإجرائية للدراسة

طول الاختبار: يعني عدد المفردات التي يتكون منها الاختبار، وفي هذه الدراسة سيتم استخدام ثلاثة اختبارات بأطوال مختلفة.

معلمة الصعوبة: هي نقطة على متصل القدرة التي تقابل احتمالية الإجابة ٠.٥، أي أنها نقطة على متصل القدرة تمثل إجابة الفرد عن المفردة إجابة صحيحة (Umar, 1995).

معلمة التمييز: هي عبارة عن قدرة المفردة على إظهار الفروق بين مستويات السمة أو القدرة التي يقيسها المقياس وهي عبارة عن ميل منحنى خصائص المفردة الذي يحدث عنده تغيير في اتجاه المنحنى (inflexion) الذي يقابل الصعوبة على متصل القدرة (صلاح علام، ٢٠٠٥).

النموذج ثنائي المعلمة: هو أحد نماذج الاستجابة للمفردة، وفيه تختلف المفردات في قيم معالم الصعوبة والتمييز.

مجتمع وعينة الدراسة:

تكون مجتمع الدراسة من جميع طلاب الصف الثامن الأساسي في المدارس الحكومية في مديريات التربية والتعليم (مديرية تربية قسبة الكرك، مديرية تربية لواء المزار الجنوبي، مديرية تربية لواء القصر) في محافظة الكرك، وتكونت عينة الدراسة من ٧٥٠ طالباً من طلاب الصف الثامن الأساسي في تلك المديريات. وبعد كتابة الاختبارات بصورتها النهائية، ٦٠ مفردة، ٤٥ مفردة، ٣٠ مفردة، وقد تم التنسيق مع مدراء المدارس والمعلمين في المدارس التي اختيرت لتطبيق الاختبار، وقد تم توضيح تعليمات الاختبار وآلية التطبيق. طبقت الدراسة في نهاية الفصل الثاني من العام الدراسي ٢٠٢٠-٢٠٢١ حيث اشترك المعلمون بالإشراف على تطبيق الاختبارات، وذلك لضمان الدقة والجدية في الإجابة على الاختبارات. طبق الاختبارات على عينة عشوائية طبقية من طلبة الصف الثامن الأساسي في محافظة الكرك، بعد تقسيم مجتمع الدراسة إلى طبقات حيث تمثل كل مديرية تربية وتعليم إحدى هذه الطبقات، ومن ثم تم اختيار المدارس التي سيتم التطبيق فيها.

أداة الدراسة

من أجل تحقيق هدف الدراسة تم بناء اختبار تحصيلي في الرياضيات مكون من ٦٠ مفردة من نوع الاختيار من متعدد، لكل مفردة أربعة بدائل، وتم اشتقاق اختبار متوسط ٤٥ مفردة من الاختبار الطويل وفق جدول مواصفات، ثم تم اشتقاق اختبار قصير ٣٠ مفردة من الاختبار المتوسط وفق جدول مواصفات.

أ- **تحديد الغرض من الاختبار:** إن الغرض من بناء الاختبار هو تجهيز مفردات قادرة على قياس تحصيل الطلبة في مبحث الرياضيات للصف الثامن الأساسي للفصل الدراسي الثاني ومن أجل تحقيق هدف الدراسة تم بناء اختبار تحصيلي في الرياضيات مكون من ٧٠ مفردة من نوع الاختيار من متعدد، لكل مفردة أربعة بدائل.

ب- **تحليل المحتوى:** تم تحليل محتوى مبحث الرياضيات الفصل الثاني للصف الثامن الأساسي، ووضع الأهداف المعرفية للوحدات الدراسية التي يتكون منها الكتاب المدرسي الذي يمثل منهاج الرياضيات للصف الثامن الأساسي، وجرى تحليل محتوى المنهاج، والذي يتكون من أربع وحدات، تُدرس للطلبة في الفصل الثاني من العام الدراسي ٢٠٢٠/٢٠٢١.

ت- **بناء جدول المواصفات لمحتوى الاختبار:** في ضوء تحليل المنهاج تم بناء جدول مواصفات اعتماداً على المحتوى والأهداف المعرفية، وتصنيف الأهداف المعرفية تبعاً لتصنيف بلوم إلى مستويات التفكير الدنيا (المعرفة، الاستيعاب، التطبيق) ومستويات التفكير العليا (التحليل، التركيب/ الابتكار، والتقييم)، ثم تم تحديد الأوزان النسبية للوحدات الدراسية ومستويات الأهداف التي تضمنها الاختبار من خلال حجم المادة الدراسية، وعدد الأهداف، وإعداد وتطوير الاختبار تم عرض نتائج التحليل على مشرفي الرياضيات للتحقق من مدى ملائمته للمنهاج.

ث- **بناء مفردات الاختبار:** بالاعتماد على تحليل المحتوى، ولائحة المواصفات، وخبرة الباحث العملية في مجال الرياضيات، تم صياغة مفردات الاختبار في صورة الاختيار من متعدد، وقد حرص الباحث على مراعاة شروط صياغة مفردات جيدة ذات بدائل ملائمة، كما راعت تحديد المستوى المعرفي ومدى تمثيل المفردات للأهداف المقابلة لها مع مراعاة الصحة العلمية لهذه المفردات، وقد بلغ عدد مفردات الاختبار ٧٠ مفردة مرتبطة بمحتوى الفصل الدراسي الثاني لمبحث الرياضيات.

ج- **تطبيق الاختبار على عينة استطلاعية:** تم تطبيق الاختبار على عينة مكونة من ١٠٠ طالب ودراسة الخصائص السيكومترية للاختبار (الصدق، الثبات)، وقد تم تصحيح استجابات الطلبة، وتحليل استجابات العينة باستخدام برنامج الحزم الإحصائية SPSS وبرنامج اكسل، لإيجاد قيم معاملات الصعوبة والتمييز.

معامل صعوبة المفردة: تم اختيار المفردات التي تقع معاملات صعوبتها ما بين ٠.٤ و ٠.٨ ويشير معامل الصعوبة إلى نسبة الطلبة الذين أجابوا على المفردة إجابة صحيحة. وتراوح مستويات الصعوبة للمفردات في الاختبار من ٠.١ إلى ٠.٩١ وهذا يشير إلى أن المفردات متوسطة الصعوبة، وبذلك تم حذف المفردات ٢٥، ٣٢، ٣٣، ٤٥، ٦٦، ٧٠.

معامل تمييز المفردة: ويتم حساب معامل التمييز من خلال نسبة الطلبة الذين أجابوا على المفردة إجابة صحيحة وتقع علاماتهم في الثلث الأعلى، ناقص نسبة الأفراد الذين أجابوا على المفردة إجابة صحيحة وتقع علاماتهم في الثلث الأدنى. وتم اختيار المفردات التي يزيد معامل تمييزها عن ٠.٤.

صدق وثبات الاختبار: تم التحقق من صدق الاختبار وثباته من خلال

صدق محتوى الاختبار: تم عرض الاختبار على أربعة معلمين واثنين من المشرفين التربويين لإبداء آرائهم وملاحظاتهم على مدى ملائمة نتائج تحليل المحتوى والاهداف لجداول المواصفات، ومدى ارتباط كل مفردة بالهدف الذي تقيسه، وملائمة اللغة ووضوحها ودقة التعبير والدقة العلمية والفنية في صياغة المفردات، وإضافة أو حذف أو تعديل ما يروونه مناسباً على عناصر أداة الدراسة، وتم تنفيذ جميع ملاحظاتهم.

ثبات الاختبار: تم تحديد قيم الثبات للاختبار من خلال حساب ثبات الاختبار الطويل (٦٠) مفردة، بطريقتي كرونباخ _ الفا، والتجزئة النصفية كمؤشر على ثبات نتائج الاختبار، حيث كانت قيمة معامل كرونباخ _ الفا ٠.٩٢، ومعامل ثبات بيرسون لنصفي الاختبار ٠.٨٨. وبناءً على ذلك يمكن اعتبار أن الاختبارات الثلاثة مناسبة لجمع البيانات اللازمة للدراسة.

بعد تطبيق الاختبار على الطلبة، تم جمع إجابات الطلبة على الاختبار والبالغ عددهم ٧٥٠ طالباً، وتم تصحيح استجابات الطلبة على الاختبار، ومن ثم إدخالها على برنامج Notepad حيث كانت القيمة ١ تعني الإجابة الصحيحة على المفردة، و ٠ تعني الإجابة الخاطئة على المفردة. ليتم استخدامها في برنامج التحليل الإحصائي الخاص بالنظرية الحديثة Bilog-Mg، لتقدير معلمتي الصعوبة والتمييز لمفردات الاختبار للإجابة على فروض الدراسة.

الأساليب الإحصائية المستخدمة:

١- برنامج التحليل الإحصائي SPSS، للقيام بإيجاد قيم معاملات الصعوبة والتمييز للمفردات وإيجاد ثبات الاتساق الداخلي، كرونباخ الفا، ومعامل الثبات باستخدام طريقة التجزئة النصفية لدراسة الخصائص السيكومترية للاختبار (الصدق، الثبات)، وتحليل التباين لمعلمتي الصعوبة والتمييز.

٢- برنامج التحليل الإحصائي Bilog-Mg واستخدام أسلوب الأرجحية القصوى الهامشية لتقدير قيم معلمتي الصعوبة والتمييز.

٣- الاحصائي بونفيروني: في حال وجود فروق بين المتوسطات لمعالم المفردة عند دراسة تأثير حجم العينة، يستخدم لتحديد مصدر تلك الفروق من خلال عمل مقارنات ثنائية.

نتائج الدراسة وتفسيرها:

قام الباحث باستخدام طريقة الارححية العظمى الهامشية Marginal Maximum Likelihood التي قدمها كل من بوك-اتكن (1981 Bock-Atkin)، ثم قام الباحث باستخدام اختبار تحليل التباين أحادي الاتجاه One Way ANOVA لاختبار الفروق بين قيم معالم المفردة (الصعوبة والتمييز) للنموذج الثنائي لكل اختبار من الاختبارات الثلاثة.

لاختبار صحة الفرض الأول الذي نصه كما يلي "توجد فروق ذات دلالة إحصائية عند مستوى الدلالة ($\alpha = 0.05$) بين متوسطات معلمة تمييز المفردة في المستويات المختلفة من طول الاختبار (قصير، متوسط، طويل) للنموذج الثنائي"، ويوضح جدول (١) الإحصاء الوصفي لمعلمة تمييز المفردة لكل طول من أطوال الاختبارات الثلاثة.

الجدول (١) المتوسطات والانحرافات المعيارية والقيم العظمى والصغرى والمدى لقيم معاملات التمييز للأطوال المختلفة للاختبارات

طول الاختبار	المتوسط	الانحراف المعياري	القيمة الصغرى	القيمة العظمى	المدى
قصير	١.٣٣١	٠.٧٧٣	٠.٢٢٩	٣.٧١٣	٣.٤٨٥
متوسط	١.١١٩	٠.٨٤٢	٠.٠٠٧	٤.٩٦٢	٤.٨٩٢
طويل	١.٥٨٨	١.١٢٢	٠.٠٠١	٤.٥٠٨	٤.٥٠٧

ويتبين من جدول (١) ارتفاع قيم معامل تمييز الاختبار الطويل (١.٥٨٨) مقارنة بالاختبار متوسط الطول (١.١١٩) والاختبار القصير (١.٣٣١) كذلك يلاحظ ارتفاع قيم الانحراف المعياري لقيم معامل التمييز للاختبار الطويل (١.١٢٢) مقابل الانحراف المعياري للاختبار متوسط الطول (٠.٨٤٢)، والاختبار قصير الطول (٠.٧٧٣). لذلك استخدم الباحث اختبار دلالة الفروق بين المتوسطات ويبين جدول (٢) نتائج اختبار دلالة الفرق بين المتوسطات.

جدول (٢) تحليل التباين لتأثير طول الاختبار على قيم معلّم معامل التمييز

مصدر التباين	مجموع المربعات	درجة الحرية	متوسط التباين	قيمة ف	مستوى الدلالة	حجم التأثير η^2
بين المجموعات	٥٠٧٣٨	٢	٢٠٨٦٩	٤٠٣٣٨	٠٠١٥	٠٠١٠٢٩
داخل المجموعات	٨٧٠٣٠٤	١٣٢	٠٠٦٦١			

يتبين من جدول (٢) وجود فروق ذات دلالة إحصائية عند مستوى ٠٠٠٥ بين متوسطات قيم معامل تمييز مفردات الاختبارات الثلاثة ترجع إلى تأثير طول الاختبار وهذه النتيجة اتفقت مع نتائج دراسة محمد الشمراني (٢٠١٦)، فقد كانت قيمة ف = ٤٠٣٣٨، كما أظهرت النتائج أن حجم التأثير صغير، حسب تصنيف كوهين لحجوم التأثير ومستوياتها، فإذا كانت قيمة حجم الاثر تقع ما بين ٠٠٠٦ - ٠٠١٤ فإن حجم التأثير يكون متوسط، فقد كانت قيمة مربع إيتا = ٠٠١٠٢٩، وهي تعني أن ١٠.٢٩% من التباين في قيم معلّم التمييز يمكن تفسيرها بواسطة معرفة طول الاختبار، ولتحديد مصدر الفروق بين متوسطات قيم معاملات تمييز المفردات قام الباحث بإجراء مقارنات ثنائية باستخدام طريقة بونفيروني Bonferroni بين أطوال الاختبارات وجاءت النتائج كما يبينها جدول (٣).

جدول (٣) المقارنات الثنائية بين متوسطات معاملات التمييز باستخدام طريقة بونفيروني

اطوال الاختبارات	الفروق بين المتوسطات	الخطأ المعياري	قيمة ت	الدلالة
طويل - المتوسط	٠٠٤٦٩	٠٠١٦	٢٠٩٢٦	٠٠١٢
طويل - قصير	٠٠٢٥٧	٠٠١٨٢	١٠٤١٥	٠٠٤٧٨
متوسط - قصير	٠٠٢١٢	٠٠١٩٢	١٠١٠٦	٠٠٨١٢

يتبين من جدول (٣) وجود فروق ذات دلالة إحصائية عند مستوى ٠٠٠٥ بين متوسط قيم معاملات تمييز الاختبار الطويل (٦٠ مفردة) والاختبار متوسط الطول (٤٥ مفردة)، وذلك لصالح الاختبار الطويل، فقد كانت قيمة الفرق بين متوسطا معامل التمييز ٠٠٤٦٩، وربما يعود السبب في ذلك زيادة عدد مفردات الاختبار الطويل ووجود مفردات متنوعة قادرة على التمييز بين قدرات الطلبة بشكل واضح، أي أن تباين أداء الطلبة على الاختبار الطويل كان أعلى من تباين أداء الطلبة على الاختبارين الآخرين، مما يدل على أن المفردات ذات الصعوبة المرتفعة

لها قدرة على التمييز بين الطلبة وبالتالي فإن تمييز المفردات سيكون مرتفعاً. في حين لا توجد فروق ذات دلالة إحصائية بين متوسط معامل تمييز الاختبار الطويل (٦٠ مفردة) والاختبار القصير (٣٠ مفردة)، كما لا يوجد فروق ذات دلالة إحصائية بين متوسطا معامل تمييز الاختبار المتوسط الطول والاختبار القصير، فقد كانت قيم مستويات الدلالة عند المقارنة بين الاختبار الطويل والقصير (٠.٤٧٨) أكبر من ٠.٠٥، وبالمثل قيمة الدلالة عند المقارنة بين الاختبار المتوسط والاختبار القصير (٠.٨١٢) أكبر من ٠.٠٥ وربما يعود ذلك إلى تقارب عدد المفردات بين الاختبار الطويل والمتوسط وبين المتوسط والقصير. كما أظهرت النتائج أن أقل فرق كان بين الاختبار القصير والاختبار المتوسط (٠.٢١٢)، يليه الفرق بين الاختبار القصير والاختبار الطويل (٠.٢٥٧)، في حين أن أكبر فرق كان بين متوسط معامل التمييز للاختبار المتوسط والاختبار الطويل (٠.٤٦٩).

نتائج اختبار صحة الفرض الثاني، الذي نصه "لا توجد فروق ذات دلالة إحصائية عند مستوى الدلالة ($\alpha = 0.05$) بين متوسطات معلمة صعوبة المفردة في المستويات المختلفة من طول الاختبار باستخدام النموذج الثنائي". ولاختبار دلالة الفروق بين متوسطات معاملات الصعوبة التي ترجع إلى تأثير طول الاختبار قام الباحث باستخدام اختبار تحليل التباين ذو الاتجاه الواحد، وقد جاءت النتائج يبينها الجدولين (٤، ٥).

جدول (٤) المتوسط والانحراف المعياري والقيم العظمى والصغرى والمدى لقيم معاملات الصعوبة للأطوال المختلفة للاختبارات

المدى	القيمة العظمى	القيمة الصغرى	الانحراف المعياري	المتوسط	طول الاختبار
٤.٦٧٤	٣	١.٦٧٤-	٠.٨١١	٠.٦٧٩	قصير
٣.٣٦٣	٣	٠.٣٦٣-	٠.٥٦٩	٠.٥٢٦	متوسط
٩.٣٨٥	٤.٥٨٧	٤.٧٩٨-	٠.٨٩٥	١.٤٦	طويل

يتبين من جدول (٤) ارتفاع قيم متوسط معلّم الصعوبة للاختبار الطويل مقارنة بالاختبارين المتوسط والقصير، فقد كانت قيم متوسط معامل صعوبة (١.٤٦) للاختبار الطويل، في حين أن متوسط معامل الصعوبة (٠.٦٧٩) للاختبار القصير، و٠.٥٢٦ للاختبار متوسط الطول. كما أظهرت النتائج أن الانحراف المعياري لقيم معلّم الصعوبة كان الأكبر في حالة الاختبار الطويل، فقد كانت قيم الانحراف المعياري للاختبار الطويل ٠.٨٩٥، بينما كانت في حالة الاختبار المتوسط ٠.٥٦٩، والاختبار القصير ٠.٨١١.

جدول (٥) تحليل التباين لتأثير طول الاختبار على قيم معلّم صعوبة المفردة

مصدر التباين	مجموع المربعات	درجة الحرية	متوسط التباين	قيمة ف	مستوى الدلالة	حجم التأثير η^2
بين المجموعات	٢٥.٨٣٢	٢	١٢.٩١٦	٥.٣٣	٠.٠١	٠.٢٥٨
داخل المجموعات	٣٢٠.٠٤٢	١٣٢	٠.٦٦١			

يتبين من جدول (٥) وجود فروق ذات دلالة إحصائية عند مستوى ٠.٠٠١، بين متوسطات معاملات صعوبة الاختبارات ترجع إلى تأثير أطوال الاختبار، فقد كانت قيمة $F=٥.٣٣$ وهي دالة عند مستوى ٠.٠٠١، كما أظهرت النتائج أن حجم التأثير أقرب للمتوسط، حسب تصنيف كوهين لحجوم التأثير ومستوياتها، فإذا كانت قيمة حجم الاثر $\eta > 0.142$ فإن حجم التأثير يكون كبير فقد كانت قيمة مربع إيتا ٠.٢٥٨، أي أن طول الاختبار يفسر حوالي ٢٥.٨% من التباين في قيم معاملات الصعوبة، ولتحديد مصدر تلك الفروق قام الباحث بعمل مقارنات ثنائية بين أطوال الاختبار باستخدام طريقة بونفيروني Bonferroni، وقد جاءت النتائج كما يبينها جدول (٦).

جدول (٦) المقارنات الثنائية لمتوسطات معامل الصعوبة باستخدام طريقة بونفيروني

	الفروق بين المتوسطات	الخطأ المعياري	قيمة ت	الدلالة
طويل - المتوسط	٠.٩٣٤	٠.٠٩٧	٩.٦١٥	٠.٠٠١
طويل-قصير	٠.٧٨٢	٠.١١	٧.٠٩٤	٠.٠٠١
متوسط-قصير	٠.١٥٣	٠.١١٦	١.٣١٤	٠.٥٧٣

يتبين من جدول (٦) وجود فروق ذات دلالة إحصائية عند مستوى ٠.٠٠٠١، بين متوسطات معامل صعوبة الاختبار الطويل (٦٠ مفردة) والاختبار المتوسط (٤٥ مفردة) لصالح الاختبار الطويل، فقد كنت قيمة الفرق بين المتوسطين ٠.٩٣٤، أي أن متوسط معامل صعوبة مفردات الاختبار الطويل أكبر من متوسط صعوبة مفردات الاختبار المتوسط.

ويمكن تفسير ذلك إلى ان زيادة عدد مفردات الاختبار بمفردات ذات خصائص سيكومترية فعالة تزيد من مقدار التباين في الدرجات فقد كانت قيم متوسط معامل صعوبة الاختبار الطويل (١.٤٦)، بالمثل توجد فروق ذات دلالة إحصائية بين متوسط قيم معامل صعوبة مفردات الاختبار الطويل ومعامل صعوبة مفردات الاختبار القصير (٣٠ مفردة) لصالح الاختبار الطويل، فقد كان الفرق بين متوسطي معامل صعوبة الاختبارين ٠.٧٨٣، لصالح الاختبار الأطول أي أن الاختبار الأطول أكثر صعوبة. وعلى الجانب الآخر لا توجد فروق ذات دلالة إحصائية بين متوسط صعوبة الاختبار المتوسط الطول والاختبار القصير، فقد كان الفرق بين المتوسطين ٠.١٥٣ وهذا الفرق غير دالة إحصائياً فقد كانت قيمة ت المعدلة ١.٣١٤، وهي غير دالة إحصائياً. الخلاصة أن مصدر الفروق بين الاختبارات ترجع إلى تأثير الاختبار الطويل.

مناقشة النتائج

توصلت نتائج الدراسة إلى وجود فروق ذات دلالة إحصائية عند تقدير معلمة الصعوبة تعزى إلى تغير مستوى طول الاختبار، أي أن طول الاختبار يؤثر في متوسطات معلّم الصعوبة لصالح الاختبار الطويل، أي إن مفردات الاختبار الطويل أكثر صعوبة من باقي الاختبارات، وتتفق هذه النتيجة مع دراسة الدرايع (٢٠٠١)، حيث بينت النتائج بأن هنالك فروقاً جوهرية تعزى لطول الاختبار في دقة تقدير معلّم صعوبة المفردة. وتتوافق هذه النتيجة أيضاً مع نتيجة باستري (Bastari, 2000) حيث توصل إلى ان زيادة طول الاختبار تزيد من دقة تقدير معالم المفردة. وتوصلت النتائج أيضاً إلى وجود فروق ذات دلالة إحصائية عند تقدير متوسطات معلمة تمييز المفردة تعزى لاختلاف طول الاختبار، أي ان متوسطات قيم معلمة تمييز المفردة قد تغيرت بتغير طول الاختبار، وتوصلت أيضاً نتائج الدراسة إلى وجود فروق في المتوسطات الحسابية لتقديرات معلمة صعوبة المفردة ذات دلالة إحصائية تعزى لتغير طول الاختبار، أي ان متوسطات تقديرات معلمة صعوبة المفردة قد تغيرت بتغير طول الاختبار. كما توصلت نتائج الدراسة أيضاً إلى وجود فروق ذات دلالة إحصائية عند تقدير معلمة التمييز تعزى إلى تغير مستوى طول الاختبار، أي أن طول الاختبار يؤثر في متوسطات معلّم التمييز لصالح الاختبار الطويل، أي إن مفردات الاختبار الطويل أكثر تمييز من باقي الاختبارات.

التوصيات

وبالرغم من ان معظم تطبيقات دوال المعلومات اقتصرت على نماذج الاستجابة للمفردة ثنائية الدرجة، الا أن الباحثين بدئوا بدراسة خصائص هذه الدوال في حالة استخدام نماذج الاستجابة متعددة الاقسام، ونماذج الاستجابة متعددة الابعاد لذا يوصي الباحث.

مراعاة مستوى الاختبار عند تقدير معالم المفردات، خلال بناء الاختبارات بالاعتماد على نماذج الاستجابة للمفردة، وذلك بزيادة عدد المفردات قدر المستطاع، للتقليل من أخطاء القياس وبالتالي زيادة الدقة في القياس.

إجراء دراسة تأثير طول الاختبار على تقدير معالم المفردة وفق النموذج الثلاثي المعلمة. وإجراء دراسة تأثير طول الاختبار على تقدير معالم المفردة لبيانات غير ثنائية متعددة الاستجابة وفق النموذج متعدد التدرج. وإجراء دراسة للمقارنة بين نتائج دراسة تجريبية واخرى محاكاة بالاعتماد على النموذج الثلاثي والثنائي المعلمة.

المصادر والمراجع

- ابو ناهية صلاح (١٩٩٤) القياس التربوي، القاهرة: مكتبة الأنجلو المصرية.
- بني عطا، زايد (٢٠١٤). نقصي دقة تقدير النموذج اللوجستي ثلاثي المعلمة لمعالم الفقرة وقدرة الأفراد في ضوء تغير طول الاختبار وحجم العينة. مجلة جامعة الشارقة للعلوم الانسانية والاجتماعية، ١١ (٧)، ١-٣٧.
- الدرايع، ماهر (٢٠٠١). فعالية النموذج اللوغاريتمي ذي المَعْلَم الواحدة (نموذج راش) في دقة تقدير قدرة الفرد ومعامل صعوبة المفردة باختلاف حجم العينة وطول الاختبار. مجلة دراسات العلوم الانسانية، الاردن، المجلد (٢٨)، ص ص ١٩٧-٢٠٨.
- الشمراي، محمد (٢٠١٦) أثر حجم العينة وطول الاختبار والتفاعل بينهما على تقدير معلمة الصعوبة والتمييز والتخمين والخطأ المعياري باستخدام نظرية استجابة الفقرة (٢٠١٦). مجلة كلية التربية، جامعة طنطا. العدد ٤، ص ص ٢٦٥-٣٠٦.
- صلاح الدين محمود علام (٢٠٠٥). نماذج الاستجابة للمفردات الاختبارية احادية البعد ومتعددة الابعاد وتطبيقاتها في القياس النفسي والتربوي، الطبعة الاولى، القاهرة: دار الفكر العربي.
- عبد الوهاب، محمد (٢٠٠٧). استخدام نماذج الاستجابة للمفردة الاختبارية في تدريج بعض الاختبارات المعرفية. أطروحة دكتوراه، كلية التربية، جامعة المنيا.

المراجع الاجنبية

- Bock, R. D. & Aitkin, M. (1981). Marginal Maximum Likelihood Estimation of Item Parameters: Application of an EM Algorithm. *Psychometrika*, 46(4), pp443-459.
- Hulin, C. L. Iissak, R. L and Drasgow, F. (1982). Recovery of two-and three-Parameter Logistic Item Characteristic Measurements, 6, pp, 249-260.
- Umar, J. (1995): Item Banking in Testing and Assessment, In: Husen, T. & Postlethwaite, T.N. (Eds.) *The International Encyclopedia of Education*, Vol. 5, pp. 126 – 148.
- Bastari, B. (2000). Linking multiple-choice and constructed-response items to a common proficiency scale (Unpublished doctoral dissertation). University of Massachusetts Amherst, USA.