مجلة العلوم التربوية، المجلد 32، العدد (3)، ص ص 455 - 483، الرياض (2020م/ 1441هـ)

استخدام مؤشر متوسط معلومات المفردة الاختبارية
في الحصول على صورة مختصرة من اختبار المصفوفات المتتابعة (العادي)

محمد منصور محمد الشافعي[1]، **ووائل محمد أحمد مبارك**[2]

جامعة نايف العربية للعلوم الأمنية

(قدم للنشر في 16/ 08/ 1440هـ؛ وقبل للنشر في 17/ 03/ 1441هـ)

المستخلص: استهدفت الدراسة الحصول على صورة مختصرة من اختبار المصفوفات المتتابعة العادي باستخدام مؤشر متوسط معلومات بند الاختبار الناتج عن تحليلات النموذج اللوغاريتمي الثلاثي المعلم، وقد سعت الدراسة إلى تحقيق ذلك الهدف اعتمادًا على عينة مكونة من (662) طالبًا من طلاب التعليم الابتدائي والمتوسط والأول الثانوي، وقد استخدمت الدراسة في المعالجات الإحصائية الخاصة بها الفروق بين تقديرات القدرة المشتقة من الاختبار الكلي (المرجعي) المكون من (60) مفردة والاختبارات الأخرى الفرعية مختلفة الأطوال التي انتقيت بنودها في ضوء مؤشر متوسط معلومات البند، والأخطاء المعيارية المشتقة من تحليل بنود الاختبار الكلي والاختبار الفرعي المختلف من حيث الطول، وقد انتهت نتائج الدراسة إلى اختزال الصورة الأصلية للاختبار إلى (45) مفردة تعطي نفس تقديرات القدرة التي حُصل عليها من الاختبار الأصلي، وأوصت باستخدامها مستقبلا توفيرًا في الجهد والوقت.

الكلمات المفتاحية: دالة معلومات المفردة الاختبارية – النموذج اللوغاريتمي ثلاثي المعلم – اختبار المصفوفات المتتابعة – صورة مختصرة.

Use the test item information function to obtain a brief form Of the regular progressive matrices test

Mohamed M. alshafei[1], and Wael M. mobark[2]

Naif Arab University for Security Sciences

(Received 21/04/2019; accepted 14/11/2019)

Abstract: The study aimed to obtain a brief form of the normal matrices test by using the average index of the test item information resulting from the analyzes of the three-parameter logarithmic model. The study sought to achieve this objective based on a sample of 662 students in primary, intermediate and secondary education, Differences between power estimates derived from the total test (reference) of 60 items and the other sub-tests of different lengths that were selected in the light of the average index of the item information and the standard errors derived N analysis of the total test items and sub-test is different in terms of length, and the results of the study concluded that the reduction of the original image of the test to (45) item gives the same ability estimates obtained from the original test recommended using a shortcut from the test image in the future savings in time and effort.

Keywords: Information function Test item - Logarithmic model Tri - parameter - Sequential matrices test - Short picture.

---

(1) أستاذ الإحصاء والقياس بجامعة نايف العربية للعلوم الأمنية.

البريد الإلكتروني: e-mail: Alshafey610@yahoo.com

(2) أستاذ القياس والتقويم المساعد بجامعة الملك سعود.

البريد الإلكتروني: e-mail: wmobark@ksu.edu.sa

(1) Professor of Statistics and Measurement at Naif Arab University for Security Sciences.

(2) Assistant Professor of Measurement and Evaluation at King Saud University.

مقدمة:

تعد اختبارات الورقة والقلم وسيلة من الوسائل المهمة التي يعول عليها في قياس وتقويم قدرات الأفراد في شتى مناحي ومجالات القياس، فهي ركيزة من ركائز العمليات التقويمية التي تستهدف معرفة المستوى التحصيلي للطلاب؛ إذ نقف من خلالها على مدى تحقيق الأهداف السلوكية أو النواتج التعليمية، كما أنها تستخدم كأدوات قياس لمختلف الجوانب العقلية المعرفية والنفسية والمهارية، فمن خلالها يمكن تحديد ما يتمتع به الفرد من إمكانات لتحقيق العديد من الأغراض في مختلف المجالات والميادين (Bond & Fox, 2010).

ولقد شهد ميدان القياس النفسي والسلوكي بعض التطورات والاتجاهات الحديثة والبارزة تستهدف بناء الاختبارات والمقاييس في العديد من المجالات النفسية والتربوية والتحصيلية (Aimee, 2008)، ومن بين هذه الاتجاهات والتطورات الحديثة استخدام النظرية الحديثة للقياس والمتمثلة في نظرية الاستجابة للمفردة (IRT)، (Bond & Fox, 2010) حيث ارتبطت هذه النظرية بالعديد من النماذج الرياضية اللوغاريتمية التي تختلف الافتراضات الخاصة بها باختلاف البيانات الاختبارية والتي قد تتأثر بالمتغيرات المختلفة مثل الصعوبة، التمييز، التخمين (Wells, Subcoviak & Serlin, 2010).

كما اقترن بظهور هذه الاتجاهات الحديثة للقياس تطوير العديد من أدوات القياس والاختبارات (Crall, 2006: 566) منها الاختبارات التقليدية الخطية، ومنها أيضًا اختبارات المواءمة التكيفية المحوسبة (CAT) (computer adaptive testing) (Kingsbury & Zara, 2002)، كما ظهرت الاختبارات التفريدية (adaptive tailored tests) (individualized tests والاختبارات المفصلة (Aimee, 2008), والمبرمجة (programmed) (سلامة، 2006) وذلك نتيجة لاستخدام النظرية الحديثة للقياس وتطبيقاتها المختلفة التي امتدت لتشمل العديد من ميادين القياس (Wang, 2011).

ومن بين الطرق المستخدمة في اختيار المفردة في هذه النوعية من الاختبارات الموائمة طرق مؤشرات دوال معلومات المفردات الاختبارية (Information function). كما توفر هذه النوعية من الاختبارات تقديرًا لقدرة الأفراد بأقل عدد من المفردات (Schmitt, 2010) ونظرًا لأن بناء هذه النوعية من الاختبارات يتطلب العديد من الإجراءات والعمليات الإحصائية المعقدة والتي تستهدف إنشاء بنكٍ للأسئلة في الخطوات الأولى التمهيدية، وكذلك توافر بعض البرمجيات الخاصة بسحب الأسئلة من البنك وأيضًا برامج أخرى لتقديم تلك الاختبارات على شاشة الحاسوب (Georgiadou & Anastasios, 2006)، فإن الدراسة الحالية ترى أن

استخدام مؤشر متوسط المعلومات للمفردات الاختبارية في انتقاء مفردات اختبار المصفوفات المتتابعة العادي وتقليص عدد هذه المفردات إلى الحد المناسب قد يكون له نفس إمكانية الاختبار الموائم (التكيفي) في تقدير ذكاء المفحوصين بأقل عدد من المفردات الاختبارية، كما أنه يوفر الجهد والوقت- لكل من المفحوص والقائم بتقدير الدرجات - التي يمكن أن يُستنفد مع عدد كبير من المفردات التي قد يحتوى عليها الاختبار بصورته الأولية التي تحتوي عددًا كبيرًا من المفردات وقد وقع اختيار الباحثين لهذا الاختبار نظرًا لأنه أحد اختبارات العالمية المشهورة والتي تستخدم استخدامًا واسعًا الانتشار في قياس مستوى ذكاء الأفراد لمختلف الأغراض التعليمية والتربوية.

طول الاختبار وعلاقته بدقة تقديراته للقدرة:

يرتبط طول الاختبار بدقة تقديرات القدرة حيث أشارت دراسة (Torre, 2008) إلى تأثر دقة تقديرات القدرة بفعل طول الاختبار، وقد توصلت الدراسة إلى أن دقة تقديرات القدرة ترتبط إيجابيًّا بطول الاختبار، هذا ومن خلال خبرات الباحثين الحالية في مجال القياس وإعداد الاختبارات، واطلاعهما 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، وبصفة خاصة الاختبار التقليدي الخطي (الورقة والقلم)، والتي يمكن أن تقدر غالبًا من خلال واضع الاختبار بطريقة ذاتية، ونظرًا لأن الاتجاهات الحديثة للقياس وما آلت إليه من اختبارات حاسوبية موائمة تعتمد على بعض المؤشرات الإحصائية في انتقاء المفردات الخاصة بها، ومن بينها مؤشر متوسط معلومات المفردة، الذي يمكن الحصول عليه من تحليلات النماذج اللوغاريتمية المرتبطة بنظرية الاستجابة للمفردة (IRT)، وحيث إن هذه النوعية من الاختبارات تستخدم في تقدير قدرات الأفراد بأقل عدد من المفردات، فإن الباحثين الحاليين يريان إمكانية استخدام مؤشر متوسط معلومات المفردة الاختبارية الذي يعطي مؤشرًا للمعلومات على متصل المتغير المستهدف من القياس بخلاف مؤشر أقصى معلومات المفردة الذي يعطي معلومات عند مستوى واحد محدد فقط على متصل المتغير موضوع القياس، وذلك في تقليص عدد المفردات الاختبارية لأي اختبار تقليدي خطي، بما لا يؤثر في تقديرات القدرة التي حصل عليها من الاختبار الأصلي (الكلي)، وبما لا يؤثر على الخصائص السيكومترية والإحصائية للاختبار وهذا ما حدا بالباحثين إلى دراسة واستقصاء هذه الفكرة البحثية المطروحة على بساط هذا البحث.

وقد كانت هناك مؤشرات مشجعة بالفعل من

خلال دراسة استطلاعية أجراها الباحثان على عينة مكونة من (30) طالبًا من طلاب التعليم المتوسط استهدفت تحديد إمكانية استخدام متغير متوسط معلومات المفردة، في خفض العدد الكلي لمفردات اختبار في مقرر الجبر يتكون من (36) مفردة إلى عدد أقل من المفردات، أي اشتقاق صورة قصيرة من نفس الاختبار يمكن من خلالها الحصول على نفس تقديرات القدرة التي يمكن الحصول عليها من الاختبار الكلي الأصلي، واتضح من خلال النتائج الاستطلاعية أن عددًا قدره (20) مفردة من الاختبار الكلي أدت إلى الحصول على نفس التقديرات التي حصل عليها من الاختبار الكلي، بما يوحي إمكانية استخدام قيم متغير متوسط معلومات المفردة في تقليص واختزال العدد الكلي لمفردات أي أداة اختبارية، وقد دعمت هذه النتائج توجه الباحثين لإجراء دراستها الأساسية بهدف تنفيذ المزيد من الإجراءات للحصول على مزيد من الأدلة التي تدعم فكرتها البحثية.

نماذج الاستجابة للمفردة: Item response models

هناك عدد من النماذج في نظرية الاستجابة للمفردة IRT models، غير أن هناك ثلاثة نماذج شائعة الاستخدام في البيانات المصححة ثنائيًا وهي النماذج الأحادية والثنائية والثلاثية المعلمة، وحيث إن الدراسة الحالية سوف تعتمد في إجراءاتها المختلفة على تحليلات النموذج الثلاثي المعلمة فسوف يُتناول هذا النموذج بشيء من الإيضاح كما يلي (Hambelton & Swaminathan, 1987: 46):

النموذج ثلاثي المعلمة: Three-Parameter Logistic Model

يفترض النموذج أن بيانات الاختبار تتفاوت في صعوبتها وقوتها التمييزية وكذلك التخمين (De Ayala, 2009: 123-130)، ويمكن التعبير عن النموذج ثلاثي المعلمة بالمعادلة التالية (Edward, 2010: 470):

$$P(\theta) = c + \frac{1-c}{1+e^{-1.7a(\theta-b)}}$$

حيث ترمز (a)، (b)، (c) للمعالم المميزة للمفردة حيث يمثل المعلمة (a) القوة التمييزية للمفردة أما (b) فهو يمثل صعوبة المفردة والمعلمة (c) يمثل الخط التقاربي الأسفل lower asymptote للمنحنى المميز للمفردة وهو بارامتر التخمين وهذا المعلمة مؤشر لاحتمالية الفرد ذي القدرة المنخفضة أن يجيب عن المفردة بصورة صحيحة، عن طريق التخمين العشوائي ويمثل بالمعلمة.

دالة المعلومات للنموذج الثلاثي المعلمة

وتقدر دالة المعلومات لبنود الاختبار في هذا النموذج بالصيغة التالية (Edward, 2010: 470):

$$I_i(\theta) = a^2\left[\frac{Q_i(\theta)}{P_i(\theta)}\right]\left[\frac{P_i(\theta)-c^2}{(1-c^2)}\right]$$

حيث $Pi\ (\theta)$ هي احتمال حدوث الاستجابة الصواب للفرد الذي قدرته (θ) على المفردة (i)، C تقدير متغير التخمين على المفردة، a تقديرًا لمتغير تمييز المفردة، Qi هي احتمال حدوث الاستجابة غير الصحيحة حيث:

$$Qi = 1.0 - Pi\ (\theta)$$

ويمكن إيجاد دالة المعلومات الكلية للاختبار باستخدام العلاقة التالية (De Ayala, 2009: 144):

$$I(\theta) = \frac{1}{\sigma^2_{\hat{}}(\theta)} = \sum_{j=1}^{l} I_j(\theta)$$

حيث $I\ (\theta)$ هي دالة معلومات الاختبار.

$\sigma^2_e(\theta)$ هي مربع الخطأ المعياري لتوزيع تقديرات القدرة (θ). أو تباين خطأ تقدير القدرة.

$Ij(\theta)$ هي دالة معلومات المفردة الاختبارية، L عدد مفردات الاختبار.

ويتضح من العلاقة السابقة بين تقدير دالة المعلومات والخطأ المعياري لتقدير القدرة أنها علاقة عكسية، ويمكن الحصول على ثلاثة معالم لدوال معلومات مفردات الاختبار باستخدام النموذج الثلاثي المعلمة (أقصى معلومات للمفردة الاختبارية، مستوى القدرة المناظر لأقصى معلومات للمفردة، متوسط معلومات المفردة الاختبارية عند المستويات المختلفة من القدرة) وسوف تقتصر الدراسة في إجراءاتها على معلمة متوسط معلومات المفردة الاختبارية على متصل القدرة من خلال تحليلات النموذج الثلاثي المعلمة، وسوف يُتناول هذا المعلمة في الفقرة التالية بشيء من الإيجاز (Zimowski, Mislevy, 2007):

معلم متوسط المعلومات Average Information

وهو يشير إلى متوسط المعلومات للمفردة الاختبارية على متصل القدرة المقاسة ومن ثم تكون هذه المفردة أكثر دقة للقياس عند هذا المستوى (Rose & Friesc, 2008)، هذا وقد تميز هذا النموذج الثلاثي المعلمة بالدقة في مجال تقدير الخصائص الإحصائية للمفردات الاختبارية وفي مجال تقديرات القدرة للأفراد، فقد أشارت العديد من الدراسات مثل دراسة (Aimee, 2008) إلى دقة استخدام النموذج اللوغاريتمي ثلاثي المعلمة إذا ما قورن بكل من النموذجين الأحادي والثنائي المعلمة في انتقاء مفردات اختبار القبول لكليات الطب بالجامعات الأمريكية، كما أشارت دراسة العكايلة (2007) إلى أن متوسط تقديرات الخطأ المعياري للقياس المقدر باستخدام النموذج الثلاثي المعلمة أدنى من متوسط الخطأ المعياري للقياس المقدر بالنظرية التقليدية، مما يشير إلى الدقة العالية لتقدير النموذج الثلاثي المعلمة، كما أثبتت دراسة (DeMars, 2001) أن هناك أفضلية لاستخدام النموذج الثلاثي المعلمة لمعرفة الفروقات الجماعية المبنية على نظرية الاستجابة للمفردة إذا ما قورن بكل من النموذجين (الأحادي والثنائي).

إجراءات انتقاء المفردة:

إن إجراء انتقاء المفردة للاختبار هو عملية تخضع

لضوابط وشروط تتحدد في ضوء الأهداف والمستويات المعرفية المستهدفة من الأداة الاختبارية (Penfield, 2008)، وبصفة عامة هناك أسلوبان رئيسان لانتقاء المفردات (Azevedo, Bolfarine & Andrade, 2011):

الأسلوب الأول هو المعلومات القصوى: يقوم هذا الإجراء حسبما يذكر (Lord, 1980) بانتقاء المفردة التي تعطي أقصى معلومات سيكومترية ممكنة عندما يجيب المفحوص على السؤال، وبعد الإجابة عنها تُنتقى المفردة التالية ذات المعلومات الأقصى أيضًا وهكذا.

الأسلوب الثاني هو الأسلوب البييزي: ويعتمد على تقليص من التباين التالي المتوقع من تقديرات القدرة إلى أدنى حد ممكن. وقد بُني الأسلوب البييزي على التباين البعدي المتوقع بافتراض توزيع قبلي في كل مرحلة من القياس والمدخل البييزي يستخدم التباين البعدي لتلك القدرة كمحك لانتقاء المفردة؛ حيث تنتقى المفردة ذات التباين البعدي المتوقع والأقل (Azevedo, 2011).

مشكلة الدراسة:

لقد تبين للباحثين في عملهما المهني في مجال القياس، ومن خلال مسحهما للأدبيات ذات الصلة بهذا المجال عدم عثورهما على أطر تنظيرية تفيد بوجوب احتواء الاختبار احتواء عامًا – وبصفة خاصة اختبار الورقة والقلم – على حد أدنى من المفردات يمكن أن يعطي نفس تقديرات القدرة المقاسة والتي يمكن الحصول عليها باستخدام عدد أكبر من المفردات، ونظرًا لظهور اتجاهات حديثة للقياس تتمثل في مؤشر دالة معلومات المفردة، وهو المؤشر الذي يعطي مؤشرات للمعلومات عن قدرة المستجيب للاختبار على متصل القدرة المستهدف قياسها (Siang & Fritz 2006)، ونظرًا لأن نفس المؤشر يعد أحد الطرق التي يُعتمد عليها في اختيار وانتقاء مفردات بعض الاختبارات التواؤمية المحوسبة (Altaf, 2011) والتي تتسم باحتوائها على عدد أقل من المفردات إذا ما قورنت بالاختبارات الخطية التقليدية وهي الاختبارات الأكثر شيوعًا وانتشارًا في مجال القياس والتقويم، جاء اهتمام الباحثين الحاليين بالقضية البحثية المطروحة بالدراسة الحالية والتي تتمثل في استقصاء إمكانية استخدام هذا المؤشر في اختزال العدد الكلي لمفردات اختبار رافن للمصفوفات المتدرجة (العادي) إلى عدد أقل يمكن أن يعطي نفس تقديرات القدرة التي يمكن الحصول عليها من الاختبار الأصلي الكلي الذي يحتوي عدد أكبر من المفردات... ومن ثم فإن مشكلة الدراسة يمكن أن تتلخص في الإجابة عن السؤال العام الآتي:

هل تختلف جوهريًا تقديرات القدرة المشتقة من الاختبار الكلي للدراسة وتقديرات القدرة المشتقة من الاختبارات الفرعية التي تحتوي أعدادًا أقل من المفردات (المسحوبة من نفس الاختبار)، والتي تمتلك قيمًا أعلى

لمتوسط المعلومات عبر متصل القدرة المقاسة؟ ويتفرع من هذا السؤال الأسئلة الفرعية الآتية:

1 – هل تختلف جوهريًا تقديرات القدرة المتناظرة للعينة الكلية المشتقة من كل من الاختبار الكلي وتقديرات القدرة المشتقة من الاختبار الفرعي باختلاف مستوى متوسط المعلومات تمتلكها المفردات المحتواة بالاختبار الفرعي (مرتفعة/ منخفضة)؟

2 – هل تختلف جوهريًا تقديرات القدرة المتناظرة للعينة الكلية المشتقة من كل من الاختبار الكلي والاختبارات الفرعية المشتقة من نفس الاختبار وتحتوي أعدادًا متفاوتة من حيث عدد المفردات التي تتصف بامتلاكها لأعلى متوسط لقيم المعلومات الموزعة على الأبعاد الخمسة للاختبار؟

3 – ما أقل عدد من المفردات الاختبارية التي تمتلك مستوى أعلى من متوسط المعلومات التي يمكن أن تعطي نفس تقديرات القدرة التي يمكن الحصول عليها من الاختبار الكلي؟

4 – هل تختلف الخصائص السيكومترية والإحصائية للاختبارات الفرعية مختلفة الطول، التي تتضمن مفردات تمتلك أعلى متوسط للمعلومات على الأبعاد الخمسة للاختبار عن الخصائص السيكومترية والإحصائية للاختبار الأصلي المرجعي؟

الأهمية النظرية للدراسة:

الدراسة الحالية قد تسفر عن نتائج يمكن أن تثري الإطار النظري في مجال القياس والاختبارات، وبصفة خاصة ما يتعلق بإمكانية خفض عدد المفردات الاختبارية، كما تهتم أيضًا بتحليل اختبار المصفوفات المتتابعة العادي لجون رافن باستخدام أحد النماذج الرياضية اللوغارتمية المرتبطة بنظرية الاستجابة للمفردة (النموذج الثلاثي المعلمة)، وهذا له أهميته النظرية والعلمية في مجال القياس العقلي بوجه عام وبصفة خاصة في مجال اختزال مفردات الاختبارات المشهورة عالميًا إلى أقل عدد من المفردات يمكن أن تعطي نفس نتائج القياس وبنفس الدقة.

الأهمية التطبيقية للدراسة:

تحاول الدراسة الكشف عن الحجم المناسب لعينة فرعية من المفردات – في ضوء متغير متوسط معلومات للمفردة الاختبارية، والذي يمكن أن يعطي نفس التقديرات للقدرة العقلية العامة للأفراد الذي يمكن الحصول عليها من العينة الكلية لمفردات الاختبار وفقًا لنموذج التحليل المستخدم، وهذا له أهميته التطبيقية في مجال الاستخدامات واسعة الانتشار لهذا الاختبار في قياس القدرة العقلية العامة، وأيضًا له أهميته التطبيقية في بناء وإعداد صور مختصرة من مختلف أنواع الاختبارات وأدوات القياس الأخرى التي يمكن أن توفر للباحثين والمبحوثين مزيدًا من الجهد والوقت.

أهداف الدراسة:

تهدف الدراسة إلى الكشف عن مدى تأثر تقديرات القدرة باختلاف مستوى متوسط معلومات (مرتفع – منخفض) تمتلكها المفردات المحتواه بالاختبار المستخدم. وكذلك الكشف عن مدى تأثر تقديرات القدرة للأفراد باستخدام اختبارات مختلفة الأطوال تحتوي مفردات تمتلك أعلى متوسط للمعلومات على متصل القدرة المقاسة. كما تستهدف الكشف عن أقل عدد من المفردات الاختبارية التي يمكن اشتقاقها من الاختبار المستخدم بالدراسة، والتي يمكن أن تعطي نفس تقديرات القدرة في ضوء متغير متوسط معلومات المفردة الاختبارية.

حدود الدراسة:

1- تقتصر الدراسة الحالية على نوعية من الاختبارات وهي الاختبارات الخطية التقليدية (اختبار الورقة والقلم)، وهو اختبار المصفوفات المتتابعة العادي لجون رافن.

2- تقتصر الدراسة الحالية على استخدام النموذج الثلاثي المعلمة في تحليل البيانات الخاصة بها، ودالة المعلومات المرتبطة به.

إجراءات الدراسة:

تضمنت الإجراءات الآتية:

عينة الدراسة:

لقد وقع اختيار الباحثين على عينة الدراسة المتاحة من طلاب التعليم العام الابتدائي والمتوسط والأول الثانوي بالمدارس التابعة لمنطقة وسط الرياض التعليمية كما هو موضح بالجدول الآتي:

جدول (1): توزيع عينة الدراسة الأساسية على المراحل التعليمية والعمرية.

مسلسل	المرحلة التعليمية	الصف الدراسي	عدد أفراد العينة
1	ابتدائي	السادس	80
2	متوسط	الأول	170
3		الثاني	198
4		الثالث	106
5	ثانوي	الأول	108
المجموع		العينة الكلية	662

أدوات الدراسة:

استخدمت الدراسة اختبار المصفوفات المتتابعة (العادي) «لجون رافن»، وهو أحد الاختبارات العالمية المشهورة في قياس القدرة العقلية العامة والمتحررة من أثر الثقافة، ويتكون الاختبار في صورته العادية من خمس مجموعات من الأسئلة (أ، ب، ج، د، هـ) حيث تحتوي كل مجموعة على اثنتي عشرة مشكلة (مفردة) وبذلك يتكون الاختبار في جملته من (60) مفردة مرتبة في سياق متدرج من حيث الصعوبة. والبيانات التي يمكن الحصول عليها من تطبيق الاختبار هي بيانات ثنائية التقسيم حيث تقدر الاستجابة الصحيحة بالدرجة (1) والخاطئة بالدرجة (0)، لكل مفردة من مفردات الاختبار (صفوت فرج، 1980 ;Raven, 1989 ; فؤاد

أبو حطب، 1996).

وقد وقع اختيار الباحثين على هذا الاختبار نظرًا لإمكانية استخدامه مع مدى واسع من الفئات العمرية المختلفة من (6 سنوات إلى 65 سنة) (معوض، 2007) وقد تحقق من الشروط والخصائص السيكومترية للاختبار من خلال العديد من الدراسات والأبحاث العربية والأجنبية التي استخدمت هذا الاختبار، كما تحققت الدراسة أيضًا من ثبات الاختبار باستخدام معادلة كيودار- ريتشاردسون (صيغة 21) وبلغت (0.98) وهي الصيغة التي تناسب المفردات ثنائية الإجابة، كما استخدم الباحثان المجموع الكلي لدرجات الطلاب في اختبارات مقررات السنة الدراسية السابقة كمحك خارجي لصدق الاختبار على عينة استطلاعية قوامها (157) من طلاب الصفوف الثلاثة من المرحلة الثانوية بواقع عدد [(42) من الصف الثالث المتوسط + (61) من الصف الثاني، (54) طالب من الصف الأول]، وقد بلغ قيمة معامل ارتباط بيرسون بين الاختبار المحك (0.88) ومن ثم كانت هناك ثقة في دقة البيانات التي حصل عليها باستخدام هذا الاختبار.

الأساليب الإحصائية المستخدمة:

- التحليل العاملي للتحقق من افتراض أحادية البعد، تحليل تباين القياس المتكرر للإجابة عن السؤال الأول.
- الخطأ المعياري لقياس القدرة للإجابة عن السؤال الأول والثاني.
- معادلة النسبة الحرجة لدلالة الفرق بين نسبتين المعدلة باختبار بنفروني للمقارنات المتعددة) للتغلب على تضخم الخطأ من النمط الأول (ألفا) نتيجة المقارنات المتعددة للنسب المئوية وهي مدرجة بخيارات:
- اختبار مربع كاي للإجابة عن الأسئلة الأول والثاني - معامل ارتباط بيرسون - معامل ألفا كرونباخ - مقاييس النزعة المركزية - مقاييس التشتت إحصاءة الالتواء- إحصاءة التفرطح- اختبار «كلمجروف- سميرنوف» للإجابة عن السؤال الرابع.

إجراءات الإجابة عن تساؤلات الدراسة:

- تحليل مفردات اختبار الدراسة (الاختبار الكلي) باستخدام النموذج اللوغاريتمي ثلاثي المعلمة لتحديد الأفراد وكذلك المفردات غير الملائمين لاستكمال إجراءات التحليل، وفقًا لمؤشر اختبار حسن المطابقة (اختبار مربع كاي) وذلك باستخدام البرنامج الإحصائي BILOG –MG3. (Zimowski, et.al., 2007) .
- حذف واستبعاد المفردات والأفراد غير الملائمين للتحليلات وفقًا للمؤشر المذكور سابقًا.
- إعادة التحليل مرة أخرى على الأفراد وكذلك المفردات الملائمة بهدف الحصول على تقديرات لكل من صعوبات ومتوسط المعلومات لمفردات الاختبار،

وكذلك الحصول على التدريج المرجعي الخاص بتقديرات القدرة لكل فرد من أفراد عينة الدراسة الملائمين للتحليلات.

– التحقق من توافر افتراضات نظرية الاستجابة للمفردة: أحادية البعد، استقلالية القياس، السرعة.

– تقسيم الاختبار الكلي – باستخدام مقياس الوسيط – إلى اختبارين فرعيين يحتوي أحدهما على المفردات التي تمتلك أعلى قيم لمؤشر متوسط المعلومات، والآخر يحتوي على المفردات التي تمتلك أدنى قيم لمؤشر متوسط المعلومات.

– ترتيب مفردات الاختبار – 60 مفردة تبعًا لأبعاده الخمسة وفقًا لمتغير أعلى متوسط للمعلومات، وذلك بهدف اشتقاق اختبارات فرعية مختلفة الأطوال تتضمن مفردات تمتلك أعلى متوسط لقيم المعلومات تمثل أبعاد الاختبار تبعًا للخطوات الآتية:

– اختيار عدد (4) مفردات تمتلك أعلى متوسط للمعلومات من كل بعد من أبعاد الاختبار الخمسة، ومن ثم يكون عدد مفردات الاختبار الفرعي الناتج (20) مفردة. ثم تكرار نفس الخطوة السابقة ولكن بانتقاء مفردات مختلفة العدد [(5)، (6)، (7)، (8)، (9)، (10)، (11)] من كل بعد من أبعاد الاختبار الخمسة، ومن ثم تتكون الاختبارات الفرعية الناتجة عن هذه الخطوات وتتضمن مفردات أعدادها (25، 30، 35، 40، 45، 50، 55) على الترتيب.

– تحليل كل اختبار فرعي، لإيجاد التدريج الجديد لتقديرات القدرة لكل فرد من أفراد عينة الدراسة والمشتق من التحليلات الناتجة عن استخدام كل اختبار فرعي على حدة.

إجراءات المعالجة الإحصائية للإجابة عن السؤال الأول:

– إيجاد مطلق الفروق في تقديرات القدرة المتناظرة لكل فرد من أفراد العينة المشتقة من تحليل الاختبار الكلي والأخرى الناشئة من الاختبارين الفرعيين (مرتفع / منخفض) المعلومات.

– إيجاد مجموع الخطأين المعياريين المشتقين من تحليل الاختبار الكلي وكل اختبار فرعي على حدة (مرتفع / منخفض) المعلومات لكل فرد من أفراد العينة.

– تحديد دلالة الفرق بين تقديرات القدرة المشتقة من الاختبار الكلي والاختبار الفرعيين (مرتفع/ منخفض) المعلومات باستخدام القاعدة الآتية: إذا كان الفرق بين تقديري القدرة المشتقين من الاختبار الكلي والآخر الفرعي (مرتفع / منخفض) المعلومات أكبر من أو يساوي مجموع الخطأين المعيارين لهما اعتبر ذلك الفرق فرقًا جوهريًا. (Kingsbury & Zara, 2002)

– إيجاد النسبة المئوية للفروق الجوهرية (الدالة) في تقديرات القدرة المشتقة من الاختبار الكلي وكل اختبار

فرعي (مرتفع / منخفض) المعلومات.

– إيجاد دلالة الفرق بين كل نسبتين مئويتين متناظرتين باستخدام اختبار (z) لدلالة الفرق بين نسبتين.

– إيجاد النسبة المئوية للفروق الجوهرية (الدالة) في تقديرات القدرة المشتقة من الاختبار الكلي وكل الاختبارين الفرعيين، فإذا تجاوزت هذه النسبة القيمة (0.05) دل ذلك على وجود فروق جوهرية (Kingsbury & Zara, 2002).

– إيجاد دلالة الفروق بين نسبة تجاوز تقديرات القدرة بين الاختبار الكلي والاختبار الفرعي المحدد (مرتفع) ونسبة تجاوز تقديرات القدرة بين الاختبار الكلي والاختبار الفرعي آخر (المنخفض من حيث المعلومات).

– استخدام تحليل تباين القياس المتكرر لتقديرات القدرة المشتقة من الاختبار الكلي المرجعي والاختبارين الفرعين (المرتفع/ المنخفض) المعلومات. وذلك كأسلوب إحصائي – إجراء آخر– للتحقق من صحة الناتج التي تم التوصل إليها من الإجراء السابق.

إجراءات الإجابة عن السؤال الثاني:

– اتباع نفس الخطوات المذكورة سابقًا ولكن باستبدال الاختبارين الفرعيين (مرتفعي ومنخفضي) المعلومات بالاختبارات الفرعية التي تحتوي على أعداد مختلفة من المفردات التي تمتلك أعلى متوسط للمعلومات تتراوح أعدادها من (20 إلى 55 مفردة).

– التحقق من مدى اختلاف أو تكافؤ الخصائص السيكومترية (الثبات، الصدق) والخصائص الإحصائية للاختبارات الفرعية مع الخصائص السيكومترية والإحصائية للاختبار الكلي المرجعي.

– إيجاد معامل ثبات الاتساق الداخلي باستخدام معامل ألفا كرونباخ لكل من الاختبار الكلي المرجعي والاختبارات الفرعية الأخرى المشتقة من نفس الاختبار والمختلفة الأطوال.

– إيجاد الانحراف المعياري وأيضًا متوسط الأخطاء المعيارية لكل من تقديرات القدرة المشتقة من الاختبار الكلي المرجعي والاختبارات الفرعية المشتقة منه والمختلفة من حيث الطول كإجراء للتحقق من تكافؤ الخصائص الإحصائية للاختبار الكلي والاختبارات الفرعية.

– التحقق من تكافؤ توزيعات القدرة لكل من الاختبار الكلي وكذلك الاختبارات الفرعية اختبار كلمجروف – سميرنوف.

– وسوف يتحقق الباحثان من تكافؤ الإحصاءات المتناظرة المستخرجة من الاختبار الكلي والأخرى الفرعية حسب القاعدة: إذا كان الفروق بين الإحصاءات المتناظرة أقل من أو تساوي النسبة (0.05)

اعتبرت إحصاءات متكافئة أما إذا ازدادت الفروق عن تلك النسبة اعتبرت إحصاءات غير متكافئة (Kingsbury & Zara, 2002).

نتائج الدراسة:

1- نتائج تحليل مفردات اختبار الدراسة باستخدام النموذج الثلاثي المعلمة: أسفرت نتائج تحليل البيانات الخاصة بمفردات اختبار الدراسة باستخدام نموذج التحليل الثلاثي المعلمة من خلال نتائج تحليل اختبار الدراسة على عينة الدراسة الأساسية كما هو مبين بنفس الجدول أن متوسط معلومات مفردات الاختبار تراوحت بين (0.005، 0.635).

2- نتائج التحقق من افتراضات نظرية الاستجابة للمفردة: أسفرت نتائج التحليلات الخاصة بتوافر افتراضات نظرية الاستجابة للمفردة بالاختبار المستخدم عما يلي:

- **افتراض أحادية البعد:** لقد أسفرت التحليلات الخاصة بأسلوب التحليل العاملي الاستكشافي لمفردات الاختبار المستخدم بالدراسة عن النتائج الموضحة بالجدول التالي:

جدول (2): الجذور الكامنة للعوامل التي استخرجت من التحليل العاملي الاستكشافي للاختبار المستخدم بالدراسة.

العوامل	الجذور الكامنة		
	الكلي	النسبة المئوية للتباين المفسر	التباين المفسر التراكمي
1	10.288	21.146	21.146
2	3.177	6.962	28.108
3	2.631	6.052	34.16
4	1.475	4.125	38.285
5	1.094	3.677	41.962

ويتضح من الجدول السابق ووفقا لمحك كايزر أن الاختبار يحتوي على خمسة عوامل يزيد الجذر الكامن لكل منها عن القيمة (1)، وقد تتلخص هذه العوامل في القدرة العقلية العامة التي تتمثل في العامل الأول وأربعة عوامل أخرى قد تمثل الأبعاد الأربع المحتواة بالاختبار، وحيث إن الظاهرة الإنسانية ظاهرة معقدة لا تتسم بالبساطة الشديدة التي لا تجعلها تختزل في عامل واحد، فقد تبين للباحث من خلال الاطلاع على الأطر النظرية الخاصة بالتحقق من أحادية البعد للاختبار أن وجود عامل واحد مهيمن يكفي لإثبات أحادية البعد للاختبار حتى مع وجود عوامل أخرى، حيث ذكر خبراء القياس ومنهم ريكاس (Recase, cited in Hattie, 1985) أنه إذا فسر العامل الأول أكبر تباين، فإن ذلك يعد مؤشرًا على أحادية البعد، بشرط أن تباينه يكون أكبر من أو

تساوي (20٪) كما أن مؤشر لورد (Lord, 1980) يشير إلى تحقق خاصية أحادية البعد إذا كانت نسبة الجذر الكامن للعامل الأول إلى نسبة الجذر الكامن للعامل الثاني أكبر من (2) مما يعني أن الاختبار متشبعا بعامل واحد مهيمن، كما أن تمثيل Scree plot لأحادية البعد يبدو من خلال وجود انحدار شديد بين قيم الجذر الكامن بين العامل الأول والثاني في حين يقل هذا الانحدار انحدارًا واضحًا بين الجذور الكامنة للعوامل الأخرى يشير أيضًا إلى مؤشر آخر لتمتع الاختبار بأحادية البعد (Gessaroli & De Champlain, 2005).

- **افتراض الاستقلالية الموضعية**: ويعني هذا الافتراض عدم اعتماد المفردات بعضها على البعض الآخر وقد أسفرت نتائج التحليلات الخاصة باختبار هذا الافتراض عن عدم تجاوز أي مفردة من مفردات الاختبار لمؤشر الملاءمة الإحصائية (مربع كاي) فقد جاء جميع مستويات الدلالة الخاصة بهذا المؤشر أكبر من 0.05 وهو مؤشر يشير إلى حسن مطابقة التوزيع الملاحظ لبيانات المفردة للتوزيع المتوقع، وكذلك مؤشر البواقي Residual حيث لم تتجاوز أي مفردة حدود الملاءمة الإحصائية وهي (±2) (Zimowski, 2007) مما يفيد بتوافر افتراض الاستقلالية لمفردات الاختبار.

- **افتراض السرعة**: ويعني هذا الافتراض أن الأسئلة المتروكة في نهاية الاختبار ينبغي أن تكون تركت بسبب عدم قدرة المفحوص في الإجابة عنها وليس بسبب أنه لم يكن هناك وقت كاف لمحاولة المفحوص أن يجيب عنها (كاظم، 1988: 92)، وقد راعى الباحثان في تطبيق اختبار الدراسة إعطاء وقتا مفتوحا للإجابة عن الأسئلة لحين انتهاء كل المفحوصين من محاولة الإجابة عن جميع فقرات الاختبار، ومن ثم فلم يكن لعامل السرعة أي تأثير في تحديد نوع الاستجابة على فقرات الاختبار مما يعني تحقق افتراض السرعة.

3- نتائج التحليلات الخاصة باشتقاق الاختبارين الفرعيين (المرتفع – المنخفض) لقيم متوسط المعلومات:

أسفرت نتائج الإجراء الخاص بهذا الإجراء الحصول على مفردات مرتفعة وأخرى منخفضة لاختبار الدراسة من حيث قيم متوسط المعلومات، مفردات الاختبارين: الاختبار المرتفع المتضمن مفردات يزيد متوسط معلوماتها عن قيمة الوسيط (0.32)، والاختبار المنخفض والمتضمن مفردات يساوي أو يقل متوسط معلوماتها عن (0.32)، وكلاهما يحتويان على مفردات تنتمي إلى الأبعاد الخمسة للاختبار.

النتائج الخاصة بإجراءات الإجابة عن السؤال الأول من أسئلة الدراسة:

وقد استخدمت الدراسة إجراءين مختلفين للإجابة عن هذا السؤال الأول من أسئلة الدراسة والذي ينص على: هل تختلف جوهريًا تقديرات القدرة

المتناظرة للعينة الكلية المشتقة من كل من الاختبار الكلي وتقديرات القدرة المشتقة من الاختبار الفرعي باختلاف مستوى متوسط المعلومات تمتلكها المفردات المحتواة بالاختبار الفرعي (مرتفعة / منخفضة)؟ الإجراء الأول يتضمن استخدام اختبار النسبة الحرجة لدلالة الفرق بين نسبتين مئويتين والإجراء الثاني تضمن استخدام اختبار دلالة تباين القياس المتكرر.

أولًا: النتائج الخاصة بالإجراء الأول: لقد أسفرت التحليلات الخاصة بهذا الإجراء عن البيانات الموضحة بالجدول الآتي:

جدول (3): تكرارات الفروق الدالة بين تقديرات القدرة المشتقة من الاختبار الكلي وكل من الاختبارين الفرعين (منخفض ومرتفع المعلومات) والنسب المئوية لها ودلالة هذه الفروق باستخدام اختبار (z).

الفروق في تقديرات القدرة المتناظرة	العدد الكلي للفروق	تكرارات الفروق الجوهرية بين تقديرات القدرة المتناظرة	نسبة الفروق الجوهرية إلى العدد الكلي للفروق	النسبة المئوية	قيمة اختبار (z)	مستوى الدلالة
بين الاختبار الكلي والاختبار المرتفع	662	15	0.022	2.2	24.74	0.00
بين الاختبار الكلي والاختبار المنخفض	662	55	0.083	8.3		

يتضح من الجدول (3) أن عدد تكرارات الفروق الدالة في تقديرات القدرة المشتقة من الاختبار الكلي والاختبار الفرعي المتضمن للمفردات التي تمتلك أعلى مستويات قيم متوسط معلومات جاءت لعدد (15) فردًا من جملة (662) فرد بنسبة مئوية قدرها (2.2٪) أما الفروق الدالة في تقديرات القدرة المشتقة من الاختبار الكلي والاختبار الفرعي المتضمن للمفردات التي تمتلك أدنى مستويات قيم متوسط المعلومات كانت لعدد (55) فرد بنسبة مئوية قدرها (8.3٪) وقد بلغت قيمة اختبار (z) لفرق هاتين النسبتين القيمة (24.74) وهي قيمة دالة عند مستوى دلالة قدره (0.01)، وذلك في صالح النسبة المئوية الأكبر، وهي (8.3٪)، وقد اتضح من الجدول السابق أيضًا وجود فروق جوهرية عند مستوى دلالة (0.00) بين النسبة المئوية لتكرارات الفروق الدالة بين تقديرات القدرة المتناظرة المشتقة من الاختبار الكلي والأخرى المشتقة من الاختبار مرتفع المعلومات والنسبة المئوية لتكرارات الفروق الدالة بين تقديرات القدرة المتناظرة المشتقة من الاختبار الكلي والأخرى المشتقة من الاختبار المنخفض المعلومات، وتشير هذه النتيجة تقديرات القدرة المشتقة من الاختبار المرتفع المعلومات تتكافأ مع تقديرات القدرة المشتقة من الاختبار الكلي المرجعي بشكل يفوق تكافؤ تقديرات القدرة يفوق

تكافؤ تقديرات القدرة المشتقة من الاختبار المنخفض المعلومات مع تقديرات القدرة المشتقة من الاختبار الكلي المرجعي.

ثانيًا: النتائج الخاصة بالإجراء الثاني: لقد أسفرت التحليلات الخاصة بهذا الإجراء والذي اعتمد على استخدام تحليل تباين القياس المتكرر عن البيانات الموضحة بالجدول الآتي:

جدول (4): الإحصاءات الوصفية لتقديرات القدرة المشتقة من الاختبار الكلي المرجعي والاختباران الفرعيان (المرتفع – المنخفض) في قيم متوسط المعلومات.

الاختبار المستخدم في تقديرات القدرة	متوسط تقديرات القدرة	الانحراف المعياري	عدد أفراد العينة
الاختبار الكلي المرجعي	63.34	5.77	662
الاختبار المرتفع	63.12	5.9	
الاختبار المنخفض	56.28	7. 73	

يتضح من الجدول (4) أن هناك فروقًا ظاهرية بين متوسطات قيم تقديرات القدرة (مقدرة بوحدة الوات) لعينة الدراسة والمشتقة من الاختبارات الثلاثة وهي: الاختبار الكلي والاختباران الفرعيان (المرتفع والمنخفض في قيم متوسط المعلومات) المشتقان من نفس الاختبار الكلي، حيث بلغت أكبر قيمة لمتوسط تقديرات القدرة القيمة (63.12) تقريبًا، وهي مشتقة من الاختبار الفرعي الأول (الذي يحتوي مفرداته التي تمتلك أعلى قيم لمتوسط المعلومات) في حين بلغت أقل قيمة (56.28) تقريبا واشتقت من الاختبار الفرعي الثاني (الذي يحتوي مفرداته تمتلك أدنى قيم لمتوسط المعلومات) بينما بلغت تقديرات القدرة المشتقة من الاختبار الكلي المرجعي (63.34) تقريبًا.

جدول (5): دلالة اختبار ماكولي للتحقق من شرط الدورية لقياسات تقديرات القدرة الثلاثة المشتقة من الاختبار الكلي المرجعي والاختباران الفرعيان (المرتفع والمنخفض) من حيث قيم متوسط المعلومات.

قيمة اختبار ماكولي	مربع كاي التقريبية	درجات الحرية	مستوى الدلالة
0.595	178.011	2	0.00

يتضح من الجدول (5) أن قيمة اختبار ماكولي للدورية بلغت (0.595) تقريبًا وهي قيمة دالة حيث بلغت قيمة الدلالة المقترنة بها القيمة (0.00) ما يشير إلى عدم تجانس القياسات الثلاثة الناتجة عن التحليل باستخدام الاختبار الكلي والاختبارين الفرعيين ومن ثم عدم تحقق شرط الدورية مما يستوجب استخدام اختبار «جرين هاوس – جيسر» Greenhouse-Geisser عند اختبار دلالة التباين بين القياسات الثلاث وهو موضح بالجدول الآتي:

جدول (6): دلالة تباين القياسات الثلاث المشتقة من الاختبار الكلي والاختباران الفرعيان (من حيث قيم متوسط المعلومات.

مصدر التباين	الاختبار	مجموع المربعات	درجات الحرية	متوسط المربعات	قيمة (ف)	مستوى الدلالة
بين القياسات	Sphericity Assumed	616.526	2	308.263	321.509	0.00
	Greenhouse-Geisser	616.526	1.424	433.071	321.509	
داخل القياسات	Sphericity Assumed	659.655	688	0.959		0.00
(الخطأ)	Greenhouse-Geisser	659.655	489.724	1.347		

يتضح من الجدول السابق أن قيمة اختبار التباين باستخدام طريقة «جرين هاوس – جيسر» بلغت (322) تقريبا وهي قيمة دالة حيث بلغت قيمة الدلالة الخاصة بها (0.00) مما يشير إلى وجود فروق جوهرية بين متوسطات القياسات الخاصة بتقديرات القدرة المشتقة من الاختبارات الثلاثة المستخدمة بالدراسة وللتعرف على اتجاه الفروق استخدم جدول المقارنات الثنائية Pairwise Comparisons بطريقة LSD، اكتفى الباحثان بالتعرف على دلالة الفرق بين تقديري القدرة المشتق من الاختبار الكلي من جانب وتقديرات القدرة المشتقة من الاختبارين الفرعيين (المرتفع – المنخض) من حيث قيم متوسط المعلومات من جانب آخر، وهي موضحة بالجدول الآتي:

جدول (7): المقارنات الثنائية بين متوسطات تقديرات القدرة المشتقة الاختبار الكلي المرجعي والاختباران الفرعيان (المرتفع والمنخفض) من حيث قيم متوسط للمعلومات باستخدام اختبار (LSD).

طرف المقارنة الأول	طرف المقارنة الثاني	الفرق بين متوسطي طرفي المقارنة (الأول – الثاني)
الاختبار الكلي (المرجعي)	الاختبار المرتفع	0.0342
	الاختبار المنخفض	16.619870*

*** دال عند مستوى (0.01)**

يتضح من الجدول رقم (7) أن هناك فروقًا جوهرية بين متوسطي قيم تقديرات القدرة المشتقة من الاختبار الكلي المرجعي وتقديرات القدرة المشتقة من الاختبار المنخفض من حيث أعلى متوسط معلومات في صالح الاختبار الكلي المرجعي، في حين لا توجد فروق جوهرية بين متوسطات تقديرات القدرة الاختبار الكلي المرجعي والاختبار المرتفع من حيث أعلى متوسط معلومات، مما يشير إلى أن الاختبار المرتفع الذي تمتلك

مفرداته أعلى قيم لمتوسط المعلومات يعطي نفس التقديرات التي يمكن الحصول عليها من الاختبار الكلي المرجعي، وتختلف جوهريًا تقديرات القدرة المشتقة من الاختبار الفرعي الثاني الذي تمتلك مفرداته أقل قيم لمتوسط المعلومات مع تقديرات القدرة التي يمكن الحصول عليها من الاختبار الكلي المرجعي، وتعد هذه النتيجة مقبولة من الناحية العلمية والمنطقية نظرًا لأن الاختبار الكلي يحتوي في مجمل مفرداته على مفردات تمتلك أعلى متوسط معلومات تمثل الأبعاد الخمس للاختبار، ومن ثم فإن التقديرات الخاصة بقدرات الأفراد المتلقين لهذا الاختبار يمكن أن تتشابه أو تتقارب مع نفس التقديرات التي يمكن الحصول عليها مع الاختبار الفرعي المشتق من الاختبار الكلي (المرجعي) والذي يحتوي فقط مفردات من الاختبار المرجعي تمتلك أعلى متوسط للمعلومات، وهذا ما أشار إليه (Kingsbury & Zara, 2002) من أن استخدام أسلوب متوسط المعلومات في انتقاء المفردات الاختبارية يوفر أعلى كفاءة للاختبار مما يؤدي إلى تقديرات للقدرة أكثر دقة. كما أن هذه النتيجة تتسق مع ما أشار إليه (Wang, & Vispoel, 1998) بأن اختيار المفردة وفقًا لمتغير متوسط للمعلومات يعطي أقل خطأ قياس عند مستوى القدرة المستهدفة، وهذا مما يمكن تفسيره بدقة تقديرات القدرة، كما تتسق هذه النتيجة مع نتائج دراسة (Schmitt, 2010)، التي توصلت إلى أن نتائج تقديرات القدرة التي يُحصَل عليها باستخدام اختبارات المواءمة الحاسوبية تعطي تقديرات أفضل للقدرة مع أقل عدد من المفردات يجرى انتقاؤه في ضوء متغير أقصى معلومات، وهذا ما أشارت إليه بعض الدراسات الأخرى مثل دراسة (Siang & Fritz, 2006) التي توصلت إلى إمكانية تقدير قدرة المفحوص باستخدام عدد قليل من المفردات الاختبارية التي تمتلك أقصى معلومات وتتسق مع قدرة المفحوص،كما اتسقت النتائج التي توصلت إليها دراسة . (Aimee, 2008) مع تلك النتيجة حيث أشارت نتائج هذه الدراسة إلى أن النقص في عدد مفردات الاختبار الموائم المحوسب والتي تتمتع بامتلاكها لأقصى معلومات لا يؤثر سلبًا في دقة تقديراته للقدرة.

النتائج الخاصة بالإجابة على السؤال الثاني من أسئلة الدراسة: لقد استخدمت الدراسة في الإجابة عن السؤال الثاني من أسئلة الدراسة والذي ينص على: هل تختلف جوهريًا تقديرات القدرة المتناظرة للعينة الكلية المشتقة من كل من الاختبار الكلي والاختبارات الفرعية التي تحتوي على أعداد متفاوتة من المفردات التي تتصف بامتلاكها لأعلى قيم لمتوسط دوال المعلومات عند مستويات مختلفة من متغير متوسط المعلومات؟ الإجراءات الآتية:

- إجراء اشتقاق مفردات الاختبارات الفرعية مختلفة الأطوال والتي تمتلك مفرداتها أعلى قيم لمتغير متوسط معلومات على ثماني فترات مختلفة على متصل نفس المتغير وفقًا لعدد المفردات، وقد أسفر استخدام هذا الإجراء عن النتائج الموضحة بالجدول رقم (8).

جدول (8): توزيع مفردات الاختبار الكلي على الاختبارات الفرعية الثمانية المختلفة من حيث عدد المفردات.

الاختبار الفرعي	عدد المفردات التي تمتلك أعلى متوسط معلومات	إجمالي عدد مفردات الاختبار
الأول	11	55
الثاني	10	50
الثالث	9	45
الرابع	8	40
الخامس	7	35
السادس	6	30
السابع	5	25
الثامن	4	20

ويتبين من الجدول أن المفردات المحتواة بجميع الاختبارات الفرعية الثمانية التي لها أعلى قيم لمتغير متوسط المعلومات موزعة على الأبعاد الخمسة للاختبار التي تتدرج من حيث مستوى صعوبتها، وذلك مما يزيد الثقة في استخدام الاختبارات الفرعية كاختبارات بديله تغطي الاختبار الأصلي بأبعاده الخمسة.

جدول (9): التكرارات والنسب المئوية للفروق الجوهرية بين تقديرات القدرة المشتقة من الاختبار الكلي (المرجعي) والأخرى المشتقة من الاختبارات الفرعية المحددة بالدراسة.

الفروق في تقديرات القدرة المشتقة من اختباري المقارنة	تكرارات الفروق الجوهرية	نسبة الفروق الجوهرية إلى جملة الفروق (662 فرقا)	النسبة المئوية
الاختبار الكلي والاختبار الفرعي الأول (55) مفردة	صفر	صفر	صفر
الاختبار الكلي والاختبار الفرعي الثاني (50) مفردة	2	0.0043	0.4
الاختبار الكلي والاختبار الفرعي الثالث (45) مفردة	18	0.037	4
الاختبار الكلي والاختبار الفرعي الرابع (40) مفردة	45	0.067	6.7
الاختبار الكلي والاختبار الفرعي الخامس (35) مفردة	48	0.073	7.3

تابع/ جدول (9).

الفروق في تقديرات القدرة المشتقة من اختباري المقارنة	تكرارات الفروق الجوهرية	نسبة الفروق الجوهرية إلى جملة الفروق (**662** فرقا)	النسبة المئوية
الاختبار الكلي والاختبار الفرعي السادس (30) مفردة	69	0.104	10.4
الاختبار الكلي والاختبار الفرعي السابع (25) مفردة	58	0.087	8.7
الاختبار الكلي والاختبار الفرعي الثامن (20)	55	0.083	8.3

يتضح من الجدول (9) أن النسب المئوية لتكرارات الفروق الجوهرية بين تقديرات القدرة المشتقة من الاختبار الكلي (المرجعي) والتقديرات المشتقة من كل اختبار فرعي على حدة لم تتجاوز النسبة المئوية (٪ 0.05) للاختبارات التي تحتوي على مفردات يتراوح بين (45) إلى (55) مفردة مما يدل على تكافؤ التقديرات المشتقة من هذه الاختبارات مع نظائرها المشتقة من الاختبار الكلي المرجعي في حين تجاوزت الفروق الجوهرية بين التقدير المشتق من الاختبار الكلي (المرجعي) والتقديرات المشتقة من الاختبارات التي تتراوح مفرداته بين (20) إلى (40) مفردة، وهو ما يعني عدم تكافؤ تقديرات القدرة المتناظرة المشتقة من الاختبار المربع وهذه الاختبارات، وعلى الرغم من عدم توقع الباحثين لهذه النتائج الخاصة بإجراءات الإجابة عن السؤال الثاني من أسئلة الدراسة، فإنها تعد نتائج مقبولة؛ إذ إن الاختبار الكلي يحتوي من بين مفرداته على مفردات تمتلك أعلى متوسط للمعلومات، ومن ثم فإن التقديرات الخاصة بقدرات الأفراد المستجيبين لهذا الاختبار يمكن أن تتقارب مع نفس التقديرات التي يمكن الحصول عليها مع الاختبار الفرعي المشتق من الاختبار الكلي (المرجعي) والذي يحتوي فقط من بين مفردات الاختبار المرجعي تلك التي تمتلك أعلى متوسط للمعلومات وهذا ما أشار إليه (Kingsbury & Zara, 2002) من أن استخدام أسلوب أقصى معلومات يوفر أعلى كفاءة للاختبار والذي يؤدي إلى تقديرات للقدرة أكثر دقة، كما أن هذه النتائج تتسق مع ما أشار إليه (Wang, & Vispoel, 1998) بأن اختيار المفردة وفقًا لمتغير أقصى معلومات يعطي أقل خطأ قياس عند مستوى القدرة المستهدفة، وهذا مما يمكن تفسيره بدقة تقديرات القدرة، كما تتسق هذه النتائج مع نتائج تقديرات القدرة التي يُحصَل عليها باستخدام اختبارات المواءمة الحاسوبية التي تعطي تقديرات أفضل للقدرة مع أقل عدد من المفردات وهذا ما أشار إليه (Schmitt, 2010). كما دعمت النتائج التي توصلت إليها دراسة (Aimee, 2008) تلك النتيجة حيث أشارت نتائج هذه الدراسة إلى أن النقص في عدد مفردات الاختبار الموائم المحوسب والتي تتمتع بامتلاكها لأقصى معلومات لا

يؤثر سلبًا في صدق الاختبار أو دقة تقديراته للقدرة، في حين جاءت هذه النتائج على عكس ما أشارت إليه دراسات (Goodman, 2011) والتي أظهرت نتائجها أنه كلما قل طول الاختبار فإن ذلك يمكن أن يؤثر سلبيًا على تقدير القدرة تأثيرًا عامًا وكذلك دراسة (de la Torre, 2008) التي توصلت إلى أن دقة تقديرات القدرة ترتبط إيجابيًا (طرديًا) بطول الاختبار، وأيضًا دراسة (Wavg, 2011) التي أشارت إلى أنه عند استخدام مفردات ثنائية التقسيم يجب زيادة طول الاختبار، وقد يعزو الباحثان سبب اختلاف هذه النتائج عن نتائج تلك الدراسات، إلى أن تلك الدراسات لم تعتمد في إجراءاتها على أسلوب اختيار المفردات فالاختبارات مختلفة الأطوال في ضوء متغير متوسط المعلومات المستخدم بالدراسة الحالية.

ثالثًا: نتائج الإجراءات الخاصة بالإجابة عن السؤال الثالث من أسئلة الدراسة: أسفرت نتائج الدراسة الخاصة بالإجابة عن السؤال الثالث من أسئلة الدراسة والذي ينص على: ما أقل عدد من المفردات الاختبارية التي تمتلك أعلى متوسط للمعلومات التي يمكن أن تعطي نفس تقديرات القدرة التي يمكن الحصول عليها من الاختبار الكلي؟

تعتمد الإجابة الخاصة بهذا السؤال على النتائج التي تُوصّل إليها من خلال الإجراءات التي استخدمت للإجابة عن السؤال الثاني من أسئلة الدراسة أن أقل عدد من المفردات التي أدت إلى نفس تقديرات القدرة التي اشتقت من الاختبار الكلي المرجعي هي المفردات المتضمنة بالاختبار الفرعي الخامس والتي بلغ عددها (45) مفردة وذلك عند مستوى دلالة قدره (0.05) وهي تقريبًا حوالي 0.75 من عدد مفردات الاختبار الكلي المرجعي والبالغ عددها (60) مفردة، وعلى الرغم من غرابة هذه النتيجة فإنها جاءت ممنهجو وإجراءاتها علمية اعتمدت عليها الدراسة الحالية في الإجابة عن هذا السؤال، كما جاءت في طيات مناقشة النتائج التي تعلقت بالإجابة عن السؤال الثاني، وتعد هذه النتيجة مقبولة نظرًا لأن مفردات الاختبار الذي يحتوي (45) مفردة يتضمن المفردات التي تمتلك أعلى متوسط للمعلومات من بين مفردات الاختبار المرجعي، ومن ثم تقاربت التقديرات الخاصة بقدرات الأفراد المستجيبين لهذا الاختبار مع نفس التقديرات التي يمكن الحصول عليها من الاختبار الكلي وهذا ما أشار إليه (Kingsbury & Zara, 2002) من أن استخدام أسلوب أقصى معلومات يوفر أعلى كفاءة للاختبار والذي يؤدي إلى تقديرات للقدرة أكثر دقة. كما أشار (Wang & Vispoel, 1998) إلى أن اختيار المفردة وفقًا لمتغير متوسط المعلومات يعطي أقل خطأ قياس عند مستوى القدرة ومن ثم دقة تقديرات القدرة، وبناء على ذلك جاءت تقديرات القدرة المشتقة من عدد (45) مفردة اختبارية متسقة مع نفس

التقديرات التي اشتقت من عدد (60) مفردة يتكون منها الاختبار الكلي (المرجعي) وهذا ما تأكد من قبل بعض الدراسات التي اهتمت بدقة تقديرات القدرة باستخدام اختبارات موائمة حاسوبية تتضمن عددًا أقل من المفردات فقد أشارت نتائج هذه الدراسات إلى أن تقديرات القدرة التي يُحصَل عليها باستخدام اختبارات المواءمة الحاسوبية تعطي تقديرات أفضل للقدرة مع أقل عدد من المفردات وهذا ما تحقق مع دراسة (Schmitt, 2010). في حين اختلفت هذه النتائج مع ما أشارت إليه دراسات: (de la Torre, 2008; Goodman, 2011; Wells, et. al., 2011) التي أشارت إلى التأثير السلبي للاختبارات القصيرة في دقة تقديرات القدرة للأفراد، كما اختلفت كذلك مع نتائج دراسة (Wang, 2011) التي أشارت إلى أنه عند استخدام مفردات ثنائية التقسيم ينبغي زيادة عدد مفردات الاختبار، وقد يرجع سبب اختلاف هذه النتائج التي حُصِل عليها من خلال الدراسة الحالية عن نتائج تلك الدراسات المشار إليها، إلى أن تلك الدراسات لم تعتمد في إجراءاتها على انتقاء المفردات فالاختبارات مختلفة الأطوال في ضوء متغير أعلى متوسط للمعلومات.

رابعاً: نتائج الإجراءات الخاصة بالإجابة عن السؤال الرابع من أسئلة الدراسة.

أسفرت نتائج الدراسة الخاصة بالإجابة عن السؤال الرابع من أسئلة الدراسة والذي ينص على: **هل تختلف الخصائص السيكومترية والإحصائية للاختبارات الفرعية مختلفة الطول والتي تتضمن مفردات تمتلك أعلى متوسط للمعلومات على فترات مختلفة من القدرة المقاسة وفقا للمقياس التساعي عن الخصائص السيكومترية والإحصائية للاختبار الأصلي المرجعي؟**

أولاً: النتائج التي تتعلق بالشق الأول من السؤال والذي يتعلق بالخصائص السيكومترية.

لقد اعتمد في التحقق من الخصائص السيكومترية (الثبات والصدق) للاختبار الكلي والاختبارات الفرعية على معامل ألفا كرونباخ للتحقق من ثبات الاتساق الداخلي، وكذلك معامل ارتباط بيرسون بين تقديرات القدرة المشتقة من الاختبار الكلي (المرجعي) والاختبارات الفرعية، وقد أسفرت نتائج التحليل عما يلي:

(أ) نتائج التحليلات الخاصة بالثبات:

لقد أسفرت التحليلات الخاصة باستخدام معامل ألفا كرونباخ لحساب ثبات الاتساق الداخلي لمفردات الاختبار الكلي والاختبارات الفرعية الثمانية عن البيانات الموضحة بالجدول الآتي:

جدول (10): معامل ألفا لتقدير ثبات الاتساق الداخلي.

الاختبار	عدد المفردات	معامل ألفا		الاختبار	عدد المفردات	معامل ألفا
الاختبار الكلي	60	0.90				
الاختبار الفرعي الأول	55	0.90		الاختبار الفرعي الخامس	35	0.82
الاختبار الفرعي الثاني	50	0.88		الاختبار الفرعي السادس	30	0.78
الاختبار الفرعي الثالث	45	0.87		الاختبار الفرعي السابع	25	0.72
الاختبار الفرعي الرابع	40	0.84		الاختبار الفرعي الثامن	20	0.50

يتضح من النتائج المبينة بالجدول السابق أن معامل ثبات الاتساق الداخلي للاختبار الكلي (المرجعي) والثلاث اختبارات الفرعية الأولى تراوحت بين (0.90 إلى 0.87) حيث تفيد هذه القيم بتمتع الاختبار الكلي والثلاثة اللفرعية الأولى بثبات مرتفع، وهي قيم لم تتجاوز الفروق بينها وبين قيمة ألفا المستخرجة من الاختبار الكلي الحد (0.03)، مما يعني عدم وجود فروق جوهرية بين تقديرات معاملات الثبات لهذه الاختبارات الفرعية وتقدير معامل الثبات المستخرج من الاختبار الكلي المرجعي، في حين يتضح من الجدول أن قيمة معامل ألفا قد انخفض بشدة للاختبار الفرعي الرابع حتى الثامن،حيث تراوحت قيم معامل ألفا (0.84, إلى 0.50) على الترتيب، وقد تجاوز الفرق بين قيمتي معامل ثبات ألفا للاختبار الكلي ومعامل الثبات كل اختبار فرعي من الاختبارات الخمس الأخيرة السابقة قيمة (0.05) وهو المستوى التقريبي المقبول رياضيًا والذي يفيد بعدم فروق جوهرية بين النتائج الرقمية (Lord, 1980) مما يعني وجود فروق جوهرية بين تقديرات معاملات الثبات لهذه الاختبارات الفرعية الخمس الأخيرة وتقدير معامل الثبات المستخرج من الاختبار الكلي المرجعي، ومن ثم فإن الاختبارات الفرعية التي تضمنت المفردات التي يتراوح حجمها بين (45 إلى 55) مفردة تتمتع بنفس مستوى ثبات الاتساق الداخلي الذي يتمتع به الاختبار الكلي المرجعي، وذلك على عكس الحقيقة القائلة بأن ثبات الاختبار يتأثر بطول الاختبار.

(ب) نتائج التحليلات الخاصة بالصدق:

لقد أسفرت التحليلات الخاصة باستخدام معامل ارتباط بيرسون لتقديرات القدرة المتناظرة المشتقة من تحليلات الاختبار الكلي المرجعي (كاختبار محكي)، وتحليلات الاختبارات الفرعية الثمانية عن النتائج الموضحة بالجدول الآتي:

جدول (11): معاملات الارتباط بين تقديرات القدرة المقدرة من كل اختبار فرعي وتقديرات القدرة المقدرة من الاختبار الكلي.

الاختبارات الفرعية	عدد المفردات	معامل ارتباط القدرة المقدرة من الاختبار الفرعي مع تقدير القدرة المقدر من الاختبار الكلي
الفرعي الأول	55	.99**
الفرعي الثاني	50	.98**
الفرعي الثالث	45	.98**
الفرعي الرابع	40	.94**
الفرعي الخامس	35	.90**
الفرعي السادس	30	.89**
الفرعي السابع	25	.81**
الفرعي الثامن	20	.69**

**** دال عند مستوى (0.01)**

يتضح من النتائج المبينة بالجدول (11) أن معامل الارتباط بين تقديرات القدرة المشتقة من تحليلات مفردات الاختبار الكلي (المرجعي) والأخرى المشتقة من مفردات الاختبارات الفرعية الثلاثة الأولى مختلفة الأطوال تجاوزت القيمة (0.95) وتراوحت قيمها بين القيمتين (0.98 إلى 0.99) تقريبًا وهي دالة عند مستوى دلالة (0.01) مما يشير إلى وجود علاقة طردية قوية بين التقديرات المشتقة من كل من الاختبار الكلي (المرجعي) والذي يحتوي على (60) مفردة، والاختبارات الفرعية (الأول والثاني والثالث) والتي تحتوي على مفردات حجمها (55، 50، 45) على التوالي، في حين قل معامل الارتباط عن القيمة (0.95) بين التقديرات المشتقة من كل من الاختبار الكلي (المرجعي) والذي يحتوي على (60) مفردة، والاختبارات الفرعية (الرابع حتى الثامن)، حيث تراوحت القيم الخاصة بمعاملات الارتباط لها بين القيمتين (0.94 إلى 0.69)، مما يعني أن الاختبارات الفرعية الثلاثة الأولى تقيس ما يقيسه الاختبار الكلي المرجعي كاختبار محكي إذا ماقورنت بالاختبارات الفرعية (الرابع حتى الثامن)، وهذه النتائج تفيد بأن نتائج التحليلات التي أسفرت عنها الاختبارات الفرعية الثلاثة والتي تحتوي على مفردات حجمها (55، 50، 45) يمكن قبولها بدرجة عالية من الثقة إذا ما قورنت بالاختبارات الفرعية الأخرى والتي تحتوي على أعداد أقل من المفردات.

ثانيًا: النتائج التي تتعلق بالشق الثاني من السؤال الرابع والذي يتعلق بالخصائص الإحصائية.

لقد استخدم الباحثان في التحقق من الخصائص الإحصائية للاختبار الكلي والاختبارات الفرعية

(الثمانية) مختلفة الأطوال وتحتوي على مفردات لها أعلى متوسطات للمعلومات على متصل القدرة المقاسة – مقاييس النزعة المركزية، ومقاييس التشتت وكذلك المؤشرات الخاصة بتوزيع تقديرات القدرة (الالتواء) وقد أسفرت نتائج التحليل عما يلي:

جدول (12): الإحصاءات الوصفية لتقديرات القدرة بوحدة الوات لعينة الدراسة الناتجة من تحليل الاختبار الكلي المرجعي والاختبارات الفرعية الثمانية مختلفة الطوال.

الإحصاءات الوصفية	الاختبار الكلي المرجعي	الاختبارات الفرعية							
		الأول	الثاني	الثالث	الرابع	الخامس	السادس	السابع	الثامن
عدد المفردات	60	55	50	45	40	35	30	25	20
المتوسط	63.34	63.33	63.37	63.39	58.36	56.67	54.08	46.79	42.56
الخطأ المعياري للمتوسط	.518	.517	.524	.535	.542	.575	.632	.574	1.22
الوسيط	65.12	65.10	65.09	65.08	60.88	60.03	55.56	50.06	45.2
المنوال	65	65	65	65	63	60	62	50	41
الانحراف المعياري	11.13	11.10	11.27	11.45	11.65	12.36	13.59	12.34	15.23
الالتواء	1.03-	1.06-	1.02-	1.04-	1.10-	2.31-	2.05-	3.054	3.45
الخطأ المعياري للالتواء	.016	.015	.018	.019	.214	.304	.302	1.15	1.45

يتضح من خلال البيانات الموضحة بالجدول (12)، وجود تشابه وتقارب شديد جدًا بين جميع قيم الإحصاءات الوصفية لكل من بيانات تقديرات القدرة لعينة الدراسة المشتقة من الاختبار الكلي (المرجعي) والأخرى المشتقة من الاختبارات الفرعية من الأول وحتى الثالث وهي الاختبارات التي تتضمن المفردات (45 إلى 55) مفردة حيث لم يتسع الفارق بين متوسط توزيع تقديرات القدرة المشتقة من الاختبار الكلي وأي قيمة لمتوسط توزيعات تقديرات القدرة المشتقة من تلك الاختبارات الفرعية عن القيمة (0.05) بوحدة الوات، كما لم يتسع الفارق بين القيمة المنوالية لتوزيع الاختبار الكلي وأي قيمة منوالية لتوزيع تقديرات القدرة المشتقة من الاختبارات الفرعية المشار إليها عن القيمة فجميعها تساوي (65) ولم يتسع الفارق بين قيمة وسيط توزيع الاختبار الكلي قيمة وأي قيمة لوسيط توزيعات الاختبارات الفرعية المشار إليها عن القيمة (0.05) تقريبًا بوحدة الوات، في حين اتسع الفارق بين تلك المؤشرات المتناظرة بين (المتوسط والوسيط والمنوال)

ومثيلاتها المشتقة من تحليلات الاختبار الكلي المرجعي والأخرى المشتقة من الاختبارات الفرعية الرابع- الثامن) مما يتجاوز (0.05)، مما يعني أن الاختبارات التي تحتوي على عدد من المفردات تمتلك أعلى متوسط معلومات يتراوح بين (45 إلى 55) مفردة تتكافأ من حيث الخصائص الإحصائية مع الاختبار الكلي المرجعي، أما الاختبارات التي تقل عدد مفرداتها عن (45) مفردة فلا تتكافأ خصائصها الإحصائية مع خصائص الاختبار الكلي المرجعي، أما فيما يتعلق بالتحقق من اعتدالية توزيعات تقديرات القدرة المشتقة من الاختبار الكلي وجميع الاختبارات الفرعية، فقد أسفر استخدام التحليلات الخاصة باختبار **كلمجروف – سميرنوف** لاختبار اعتدالية تلك التوزيعات عن النتائج الموضحة بالجدول الآتي:

جدول (13): قيم اختبار كلمجروف – سميرنوف لاعتدالية توزيع تقديرات القدرة المشتقة من الاختبار الكلي وجميع الاختبارات الفرعية مختلفة الأطوال.

الاختبار الإحصائي ومستوى الدلالة	الاختبار الكلي (المرجعي) (60) مفردة	الاختبار الفرعي							
		الأول (55) مفردة	الثاني (50) مفردة	الثالث (45) مفردة	الرابع (40) مفردة	الخامس (35) مفردة	السادس (30) مفردة	السابع (25) مفردة	الثامن (20) مفردة
Kolmogorov-Smirnov Z	0.077	0.070	0.081	0.095	1.885	2.51	3.09	3.54	3.98
Asymp. Sig. (2-tailed)	.0.620	0.101	0.117	0.089	0.044	0.037	0.022	0.005	0.001

يتضح من النتائج المبينة بالجدول (13) أن جميع قيم اختبار كلمجروف- سميرنوف والخاصة بتوزيع تقديرات القدرة لعينة الدراسة التي اشتقت من الاختبار الكلي والذي يحتوي (60) والاختبارات الفرعية التي تتضمن مفردات تتراوح بين (60 إلى 45) مفردة ليست دالة إحصائيا، في حين جاءت قيم نفس الاختبار دالة مع الاختبارات الفرعية التي تتضمن أعداد أقل تتراوح بين (40 إلى 20) مفردة، مما يشير إلى أن أقل طول للاختبارات الفرعية هو (45) مفردة يمكن أن يمتلك توزيع القدرة الخاصة به نفس التوزيع الاعتدالي الذي يتمتع به توزيع القدرة للاختبار الكلي، كما يتضح وجود تشابه وتقارب بين معاملات الثبات والصدق والخصائص الإحصائية الخاصة بالاختبار الكلي المرجعي والاختبارات الفرعية (60 إلى 45) مفردة، وتعد هذه النتيجة أيضًا نتيجة غير متوقعة نظرًا للتفاوت الشديد في عدد المفردات المتضمنة بتلك الاختبارات الفرعية والاختبار الكلي، حيث لم تتسق هذه النتيجة مع بعض الدراسات ومنها دراسة (Kahraman, N., Thompson, T., 2011) والتي أظهرت أن ثبات الاختبارات الفرعية الأطول كان مرتفعًا مقارنة بثبات الاختبارات الفرعية

الأقل طولًا، كما اتفقت مع دراسة (Jon A., Geoffrey F., & Aurelio J., 2016)، وكذلك دراسة (Carretta, Thomas R. 2010) التي أشارت إلى أن قيمة معامل ثبات الاختبار تتاثر سلبًا بقصر طول الاختبار، وأيضًا دراسة (Uttl, Bob, 2009) التي توصلت إلى أن ازدياد طول الاختبار يؤدي إلى زيادة ثبات وصدق الاختبار. ودراسة (Wells, Craig S. et. al., 2010) التي أفادت بالتأثر السلبي لمعاملات الثبات والصدق للاختبار بالاختبارات التي تتضمن أعدادًا أقل من المفردات.

وتتسق هذه النتيجة مع بعض الدراسات الأخرى مثل دراسة (Wim V., Carolijn O., Peter v., Nikki L., Martin V., & Jelle J., 2013). التي أشارت إلى تكافؤ الخصائص الإحصائية لصورة مختصرة من اختبار المصفوفات المتدرجة المعياري مع الصورة الأصلية من نفس الاختبار) وكذلك دراسة (Klaas, S., 2009) التي أشارت إلى اختلاف الخصائص السيكومترية والإحصائية للاختبارات بصورة واضحة باختلاف حجم عينة المفردات المتضمنة بها التي تقل عن العدد الكلي لمفردات الاختبار الأصلي بنسبة تزيد عن (70٪)، وهذا ما يتفق إلى حد ما مع نتائج الدراسة الحالية والتي أثبتت عدم تأثر الخصائص الإحصائية للاختبارات التي تجاوزت النسبة الخاصة بعدد مفرداتها بنسبة (75٪) من العدد الكلي لمفردات الاختبار الأصلي. في حين اتسقت هذه النتيجة مع ما أشارت إليه بعض الأدبيات التي أشارت إلى تمتع الاختبارات الموائمة المحوسبة بالعديد من المزايا من أهمها أن الاختبار الموائم يكون عادة أقصر من الاختبار التقليدي الخطي بنسبة قد تصل إلى (50٪) ويعطي قياسًا أكثر دقة، كما أن ثبات الاختبار الموائم المحوسب الذي يحتوي على عدد (10) مفردات يزيد على ثبات الاختبار التقليدي الذي يقيس نفس المتغير ويشتمل على (25) مفردة (Siang, C. & Fritz D., 2006, 241).

كما أن النقص في عدد مفردات الاختبار الموائم المحوسب لا يؤثر سلبًا في صدق الاختبار (ay, 1997, vispol, 1993)، ويمكن تفسير هذه النتيجة بأن الدراسة الحالية اعتمدت في اختيارها لمفردات الاختبارات الفرعية مختلفة الأطوال على متغير متوسط المعلومات، وهو أحد الأساليب التي استخدمت بالدراسات التي اهتمت بانتقاء المفردات الاختبارية المحوسبة والتي حققت نتائج مشابهة لهذه النتيجة، في حين لم تعتمد الدراسات السابقة التي أشارت إلى التأثر السلبي لمعاملات الثبات والصدق بقصر الاختبار في إجراءاتها على متغير متوسط معلومات المفردة الاختبارية في اختيار المفردات الخاصة بها، ومن ثم جاءت النتائج الخاصة بها على العكس من النتائج التي توصل إليها من خلال هذه الدراسة.

وقد ترجع تلك النتائج التي توصلت إليها

الدراسة الحالية نتيجة استخدام دالة المعلومات للمفردة الاختبارية إلى الأساس الرياضي لهذه الدالة الذي على العلاقة العكسية مع الخطأ المعياري لتقدير القدرة، وهذا ما يبدو واضحًا في العلاقة الرياضية التي تجمع بين قيمة دالة المعلومات والخطأ المعياري للتقدير.

ومن ثم فإن مؤشر متوسط معلومات المفردة الاختبارية يعطي أدق تقدير لقدرة الفرد، وحيث أن جميع المفردات الاختبارية المتضمنة بالاختبارات مختلفة الأطوال تمتلك قيم لمتوسط المعلومات تمتد على متصل القدرة المقاسة، فإن ذلك من شأنه أن يحقق أعلى مستويات دقة قياس لقدرات الأفراد وهذا ما يفسر ثبات تقديرات القدرة لعينة الدراسة على اختبارات تتراوح أعداد الاختبارات المتضمنة بها بين (45 إلى 60) مفردة، أما عدم تحقق نفس ثبات تقديرات القدرة المشتقة من اختبارات تتضمن أعدادًا من المفردات تقل عن (45) مفردة فقد يفسر بعدم التمثيل الجيد لهذه الأعداد القليلة من المفردات للأبعاد الخمسة الخاصة بالاختبار، والتي مثلت من خلال مفردات الاختبارات مختلفة الأطوال والتي تتضمن المفردات المتضمنة بها (45) مفردة فأكثر.

توصيات الدراسة:

- ضرورة استخدام مؤشر متغير متوسط المعلومات على متصل القدرة المقيسة التي يمكن الحصول عليه من تحليلات النموذج اللوغاريتمي ثلاثي المعلمة كإجراء أساسي في إجراءات بناء الاختبارات المختلفة، وبصورة خاصة قبل إعداد الصورة النهائية للاختبار مباشرة.

- استخدام مؤشر دوال معلومات المفردات في اختزال بعض أنواع الاختبارات الأخرى والتي تحتوي نوعًا مختلفًا من المفردات مثل اختبارات المقال أو المفردات ذات مستويات الاستجابة المتدرجة.

- استخدام مؤشر دوال معلومات المفردات في اختزال مفردات الاختبارات المشهورة عالميًا مثل اختبارات الذكاء والقدرات العقلية واختبارات الشخصية.... والحصول منها على صور أخرى مختصرة.

- عقد مقارنة بين أقل عدد من المفردات الاختبارية - من نوع ثنائي التقسيم والتي يمكن أن تنتج من استخدام مؤشرات دوال المعلومات المستخرجة من تحليلات نماذج اللوغاريتمية المختلفة (الأحادي والثنائي والثلاثي المعلمة) والتي يمكن أن تعطي نفس تقديرات القدرة.

***

قائمة المصادر والمراجع

أولاً: المراجع العربية:

البرصان، إسماعيل (2006). *أثر عدد مراحل القياس التكيفي وعدد أسئلة كل مرحلة في تقدير القدرة والخطأ المعياري في التقدير باستخدام أسئلة ثنائية ومتعددة التدريج*. رسالة دكتوراه غير منشورة. جامعة عمان العربية للدراسات العليا.

العكايلة، عبد الناصر (2007). دراسة مقارنة بين النظرية الكلاسيكية والنظرية الحديثة في القياس في كشف وتقدير الخطأ المعياري في اختبارات القدرات المعرفية لطلبة المرحلة الأساسية في الأردن. رسالة دكتوراه غير منشورة، جامعة القاهرة، مصر.

كاظم، أمينة (1988). دراسة نظرية نقدية حول التفسير الموضوعي للسلوك (نموذج راش)، ط1، الكويت، مؤسسة الكويت للتقويم العلمي.

معوض، خليل (2007). القدرات العقلية، مصر، القاهرة، دار المعارف.

ثانياً: المراجع الأجنبية:

Aimee, M. (2008). A comparison of (CAT) scores based on based on different Measurement Models for test lets, *Educational and psychological measurement, 31*(3), 371-398.

Altaf, S. (2011*).* Bayesian Analysis Of The Davidson Model For Paired Comparison With Order Effect Using Noninformative Priors*, Pakistan Journal of Statistics*, 27(2), 171-185.

Azevedo, C., Bolfarine, H., & Andrade, D. (2011).Bayesian inference for a skew-normal IRT model under the centred parameterization. *Computational Statistics & Data Analysis*, 55(1), 353-365.

Bond, T., & Fox, C. (2010). *Applying the Rasch model: Fundamental measurement in the human sciences.* Mahwah, NJ: Lawrence Erlbaum Associates.

Carretta, T. (2010). Short-Term Retest Reliability of an Experimental U.S. Education Collage Candidate Selection Test Battery. *International Journal of Aviation Psychology*, 2(3), 161.

Crall, J. (2006). Classical test theory and item response theory analyses of multi-item scales assessing parents' perceptions of their children's dental care. *Medical Care*, 44(2), 560–568.

De Ayala, R. (2009).*The Theory and Practice of Item Response Theory*. NY. the Guilford press.

De la Torre, J., & Hao Song, Y. (2011). A Comparison of Four Methods of IRT Subscoring. *Applied Psychological Measurement*, 35(4), 296-316.

De la Torre, J.(2008). Multidimensional Scoring of Abilities: The Ordered Polytomous Response Case*, Applied Psychological Measurement,* 32 (5), 355-370.

DeMars, C.(2001**).** Group differences based on IRT scores: Does the model mater*? Educational and psychological measurement,* 61(1): 60-70.

Edward, H. (2010). Interpretation of the Three-Parameter Testlet Response Model and Information Function. *Applied Psychological Measurement*, 34 (7): 467-482.

Georgiadou, E., Triantafillou, E., & Anastasios, A. (2006). Evaluation parameters for computer-adaptive testing. *British Journal of Educational Technology*, 37(2), 261-278.

Gessaroli ME & De Champlain A. (2005). Test dimensionality In: Everitt BS, Howell DC, *Eds. Encyclopedia of Statistics in Behavioral Science*. Chi Chester: John Wiley & Sons

Goodman, J. (2011). Identification of Differential Item Functioning in Assessment Booklet Designs With Structurally Missing Data**.** *Educational & Psychological Measurement***,** 71(1), 80-94.

Hambleton, R. (1987).*Item Response Theory Principles And Applications*: N.Y., Kluwer. Nijh off Publishing.

Hattie, J., & Methodology, R. (1985). Assessing Unidimensionality of Tests and Items, *Applied Psychological Measurement,* 9(2), 139-164.

Jason, M., Edward, M.,& Walter, W. (2009). Item Position and Item Difficulty Change In An IRT-Based Common Item Equating Design , *Applied Measurement In Education*, 22(3), 38–60,

Kingsbury, G., & Zara, A. (2002). Item Selection in Computerized Adaptive Testing: Should More Discriminating Items be Used First? *Journal of Educational Measurement*, 33(2), 221-255.

Liden, W., & Pushley, P.(2003). *item selction and ability estimation in adaptive testing. Computerized adaptive testing: theory and practice*, kluwer academic publishers.

Lidwine, B. (2006). Improving the Quality and Applicability of the Dutch Scales of the Communication Profile for the Hearing Impaired Using Item Response Theory. *Applied Psychological Measurement*, 3(2), 98-122.

Lord, F. (1980). *Applications of item response theory to practical testing problems.* Hillsdale, NJ: Erlbaum.

Magis, D**. (**2011). A Test-Length Correction to the Estimation of Extreme Proficiency Levels*. Applied Psychological Measurement*, 35(2), 91-109.

Marinagi, C., & Kaburlasos, G. (2010). Bayesian Decision Theory for Multi-Category Adaptive Testing, *Educational & Psychological Measurement*, 20(1), 37-59.

Penfield, D. (2008). Applying Bayesian Item Selection Approaches to Adaptive Tests Using Polytomous Items. *Applied Measurement in Education*, 17(1), 1-20.

Rose, B., Bjorner, J., Becker, F., & Friesc. E. (2008). Evaluation Of A Preliminary Physical Function Item Bank Supported The Expected Advantages Of The Patient-Reported Outcomes Measurement Information System (Promis). *Journal of Clinical Epidemiology*, 61(4), 17- 33.

Schmitt, T. A. (2010). A Monte Carlo Simulation Investigating the Validity and Reliability of Ability Estimation in Item Response Theory with Speeded Computer Adaptive Tests. *International Journal of Testing*, 10(3), 230-261.

Siang, C., & Fritz D. (2006). How Big Is Big Enough? Sample Size Requirements for CAST Item Parameter Estimation, *Applied Psychological Measurement*, 19(3), 241–255.

Uttl, B. (2009). Measurement of Individual Differences. *Psychological Science (Wiley-Blackwell),* 16(6), 460-467.

Wang, T., & Vispoel, W. (1998). Properties of ability estimation methods in computerized adaptive testing. *Journal of Educational Measurement*, 35(2), 109-135.

Wang, W., & Liu, C. (2011). Computerized Classification Testing Under the Generalized Graded Unfolding Model**,** *Educational & Psychological Measurement*, 71(1), 114-128**.**

Wells, C., Subcoviak, M., & Serlin, R. (2010). The Effect of Item Parameter on Examinee Ability Estimates. *Educational & Psychological Measurement*, 64(1), 33-52.

Wim V. , Carolijn O., Peter v., Nikki L., Martin V., & Jelle J.(2013). The Shortened Raven Standard Progressive Matrices: Item Response Theory– Based Psychometric Analyses and Normative Data. *Assessment*, 20(1), 48–59.

Zimowski, M., Muraki, E., Mislevy, R., & Bock, R. (2007). *BILOG -MG 3 for Windows*, Chicago: Scientific Software Inc.

❊ ❊ ❊