

رؤية جديدة في إعداد الاختبارات

د. محمد الشافعي
أستاذ القياس والإحصاء

أهداف حلقة النقاش

- الإلمام بمشكلات الاختبارات الخطية التقليدية.
- الإلمام بالأسس العلمية والاتجاهات المختلفة في إعداد الاختبارات.
- الإلمام بأنواع الاختبارات التي أجري إعدادها في ضوء الاتجاهات الحديثة لإعداد الاختبارات.
- ترويد المشاركون بالمعارف العلمية المتطورة التي تستخدم في بناء الاختبارات الخطية.
- التعرف على دالة معلومات المفردة الاختبارية التي يعتمد عليها كإجراء جديد في اتقاء المفردة الاختبارية.
- الإلمام بالإجراءات التي تمكن واضع الاختبار من تحديد العدد المناسب من بنود الاختبار.
- الإلمام بمفهوم كفاءة الاختبار.
- عرض تجربة في إعداد اختبار موضوعي وفقاً لرؤية جديدة.

جمهور المستفيدون من حلقة النقاش

- أعضاء وعضوات الهيئة التدريسية بجامعة الملك سعود .
- طلاب وطالبات الدراسات العليا بجامعة الملك سعود .

مقدمة

■ الاختبارات الخطية التقليدية

■ هي اختبارات الورقة والقلم، وهي إحدى أهم الوسائل التي يعتمد عليها في قياس وتقويم قدرات الأفراد .

■ فمن خلالها يمكن تحديد ما يتمتع به الفرد من إمكانيات وقدرات ولذا تستخدم بشكل واسع الانتشار لتحقيق العديد من الأهداف في مختلف المجالات والميادين، سواء كانت هذه الميادين دراسية تعليمية أو مهنية أو عسكرية أو إكلينيكية

■ وعلى الرغم من استخدام هذه الاختبارات لفترات طويلة من الزمن إلا أنها ما زالت تعاني من المشكلات التي أفقدتها موضوعيتها لذا وجهت لها العديد من الانتقادات من أهمها:

الانتقادات التي وجهت للاختبارات الخطية التقليدية

- عدم وجود خطية للقياس .
- تعلق إحصاءات المفردات بعينة المتحنيين .
- تغير معنى ومضمون مفردات الاختبار بمضي الزمن .
- عدم توافر أحادية للقياس .
- تقيد الدرجة الكلية للاختبار بمفردات هذا الاختبار .

ونظرا لوجود هذه المشكلات فقد بذلت الكثير من الجهود من قبل المهتمين بمجال القياس السلوكي في احتوائها وقد أسفرت هذه الجهود في الآونة الأخيرة عن بعض الاتجاهات الحديثة في القياس، وبالفعل فقد شهد ميدان القياس النفسي والسلوكي بعض التطورات البارزة التي استخدمت في بناء الاختبارات والمقاييس التي تتصف بالدقة والموضوعية بصورة تشبه إلى حد بعيد الموضوعية التي تتمتع بها أدوات القياس المستخدمة في مجال القياس الفيزيقي .

الاتجاهات الحديثة للقياس وبناء الاختبارات

- ومن بين هذه الاتجاهات المتطورة التي أدت إلى تحقيق الموضوعية المنشودة في مجال القياس (النظرية الحديثة للقياس) والمتمثلة في نظرية الاستجابة للمفردة (**IRT**) حيث ارتبطت هذه النظرية بالعديد من النماذج الرياضية اللوغاريتمية التي تختلف الافتراضات الخاصة بها باختلاف البيانات الاختبارية .
- كما اقترن بظهور هذه الاتجاهات الحديثة للقياس تطوير العديد من أدوات القياس والاختبارات منها الاختبارات الخطية التقليدية، ومنها أيضا الاختبارات الموائمة الكمبيوترية (**CAT**)، كما ظهرت الاختبارات التفريدية والاختبارات المفصلة أو المبرجة، وذلك نتيجة لاستخدام النظرية الحديثة للقياس وتطبيقاتها المختلفة التي امتدت لتشمل العديد من ميادين القياس .

الاختبار الموائم

- يختلف هذا النوع عن الاختبار التقليدي الخطي وهو الذي يتم من خلاله تلقي كل الأفراد لجميع المفردات الاختبارية وبنفس الترتيب الخاص بها بغض النظر عن مستوى قدرة الفرد متلقي الاختبار .
- وقد عرف الاختبار الموائم بأنه الاختبار الذي يفصل لكل فرد وفقاً لمستوى قدرته، بحيث يبدو وكأنه اختبار لكل فرد على حدة، مما يؤدي في النهاية الحصول على أكبر قدر من المعلومات عن قدرة ذلك الفرد .
- الاختبار الموائم هو الاختبار الذي يوائم بين صعوبة المفردات وقدرات الأفراد، بحيث تكون صعوبة المفردات مناسبة لمستوى قدرات الأفراد .

بعض الاعتبارات والحقائق الخاصة بالاختبار الموائم

- لقد أظهرت العديد من الدراسات أن الاختبار الموائم عادة يكون أقصر من الاختبار التقليدي الخطي بنسبة (50%) ويعطي قياساً أكثر دقة .
- كما أن ثبات الاختبار الموائم الذي يحتوي على عدد (15) مفردة يزيد على ثبات الاختبار التقليدي الذي يقيس نفس المتغير ويشتمل على (35) مفردة .
- كما أن هذا النقص في عدد مفردات الاختبار الموائم والتي تتمتع بامتلاكها لأقصى معلومات لا يؤثر سلباً في صدق الاختبار أو دقة تقديراته للقدرة .
- أشارت العديد من الأدبيات أن الاختبارات الموائية تُنشط المفحوص وتُسثير دافعيته للأداء دون إحساسه بالتعب والملل الذي ينتج أحياناً في مواقف تطبيق الاختبارات التقليدية الخطية .

تابع بعض الاعتبارات والحقائق الخاصة بالاختبار الموائم

- ومن بين أشهر الأساليب المستخدمة في اختيار المفردة في هذه النوعية من الاختبارات الموائمة هي أسلوب أقصى معلومات (Maximum Information). كما توفر هذه النوعية من الاختبارات تقديراً لقدرة الأفراد بأقل عدد من المفردات الاختبارية.
- وقد أشارت الأدبيات إلى تفوق أسلوب أقصى معلومات على الأساليب الأخرى في انتقاء مفردات الاختبار الموائمة لأنه يعطي أقل خطأ قياس عند مستوى القدرة المستهدفة.

رؤية جديدة للاختبار الخطي

- ونظراً لأن بناء هذه النوعية من الاختبارات تتطلب العديد من الإجراءات والعمليات الإحصائية المعقدة والتي تستهدف إنشاء بنكا للأسئلة في الخطوات الأولى التمهيدية، وكذلك توافر بعض البرمجيات الخاصة بسحب الأسئلة من البنك وبرامجيات أخرى لتقديم تلك الاختبارات على شاشة الحاسوب ، فإنه بالإمكان:
- استخدام متغير أقصى معلومات للمفردات الاختبارية في انتقاء مفردات الاختبار التقليدي الخطي - والذي يقدم عادة في صورة ورقة و قلم - وتقليص عدد هذه المفردات إلى الحد المناسب قد يكون له نفس إمكانية الاختبار الموائم (التكفي) في تقدير قدرات المفحوصين بأقل عدد من المفردات الاختبارية .
- كما أنه يوفر الجهد والوقت - لكل من المفحوص والقائم بتقدير الدرجات - الذي يمكن أن يستفد مع عدد كبير من المفردات التي قد يحتوي عليها الاختبار التقليدي بصورته الأولية التي قد تحتوي عدد كبير من المفردات .

دالة المعلومات INFORMATION FUNCTION

- تعتبر دالة المعلومات لب نظرية الاستجابة للمفردة، ويقصد بها مقدار ما تعطيه المفردة من معلومات عند مستويات القدرة المختلفة، والمفردة التي يكون بها بارامتر (a) عالياً، وبارامتر (c) منخفضاً توفر معلومات أكثر من (a) المنخفضة و (c) المرتفعة.
- ويعطي نموذج الاستجابة المفردة الثلاثي البارمتر افضل وأدق قيم لدالة معلومات للمفردة الاختبارية إذا ما قورن بالنموذج الثنائي والأحادي البارمتر ، حيث أن يعتمد في تقديره لمعلم دوال معلومات المفردة على متغيرات الصعوبة والتمييز والتخمين ، وعدم الاقتصار على التمييز أو الصعوبة فقط كما يعتمد النموذج الثنائي البارمتر والصعوبة فقط كما يعتمد الأحادي البارمتر.

دالة المعلومات للنموذج الثلاثي البارامتر

وتقدم دالة المعلومات لبنود الاختبار في هذا النموذج بالصيغة التالية:

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

حيث $P_i(\theta)$ هي احتمال حدوث الاستجابة الصواب للفرد الذي قدرته (θ) على المفردة (i) ،

c تقديرا متغير التخمين على المفردة ، a تقديرا المتغير تمييز المفردة

، Q_i هي احتمال حدوث الاستجابة غير الصحيحة حيث : $Q_i = 1 - P_i(\theta)$

اشتقاق بارمترات دالة معلومات المفردة الاختبارية

- تستخدم تحليلات بعض البرامج الإحصائية في استخراج قيم دالة معلومات المفردة، ومن أشهر البرامج هذه البرامج برنامج BILOG-MG حيث يسفر استخدام هذا البرنامج عن تقدير ثلاث بارمترات لدالة المعلومات هي:
 - أقصى معلومات للمفردة الاختبارية **Maximum Information**.
 - مستوى القدرة المناظر لأقصى معلومات للمفردة. **Point of Maximum Information**.
 - متوسط معلومات المفردة الاختبارية عند المستويات المختلفة من القدرة. **Average Information**.
- ويعتمد على بارمتر أقصى معلومات يمكن أن تقدمه المفردة عند المستويات المختلفة من القدرة المقاسة بالاختبار، وذلك للحصول على أفضل المفردات التي يمكن أن تعطي أكبر قدر من المعلومات عن قدرة الفرد عند كل مستوى من المستويات المختلفة من القدرة المقاسة. وذلك في اختبارات المواثمة.

المصطلحات

■ بارمتر أقصى معلومات Maximum Information

وهو يشير إلى أقصى ارتفاع لمنحنى دالة المعلومات للمفردة الاختبارية عند مستوى معين من السمة أو القدرة المقاسة، وبالتالي تكون هذه المفردة أكثر دقة للقياس عند هذا المستوى.

■ كفاءة الاختبار

تعني كفاءة الاختبار **efficiency** أن يعطي أفضل معلومات ممكنة عن الممتحن بأقل عدد ممكن من المفردات .

عصف ذهني

- هل هناك طول مناسب للاختبار؟
- هل هناك علاقة بين طول الاختبار ودقة تقدير القدرة المقاسة؟

طول الاختبار ودقة تقديرات القدرة المقاسة

- **علاقة طول الاختبار بدقة تقديرات القدرة**
- الاختبار الأكثر طولاً هو الذي يحتوي على عدد أكبر من المفردات .
- اهتمت بعض الدراسات بالكشف عن العلاقة بين متغير طول الاختبار ودقة تقديراته للقدرة المقاسة وبعض خصائصه السيكومترية والإحصائية .
- فقد تبانت العديد من الدراسات التي تناولت الكشف عن تلك العلاقة ، فقد أظهر البعض منها أن **تقدير القدرة يتأثر إيجاباً بطول الاختبار وأظهر البعض الآخر تأثيراً سلبياً** ومن بين هذه الدراسات:
- دراسة (de la Torre, 2008) إلى تأثير دقة تقديرات القدرة بفعل عوامل مثل عدد من فئات الاستجابة وطول الاختبار ، وذلك باستخدام بيانات حقيقية تم الحصول عليها من عينة من طلاب الجامعة بلغ حجمها (920) ، وقد توصلت الدراسة إلى أن دقة تقديرات القدرة **ترتبط إيجابياً بطول الاختبار وكلما إنزادت فئات الاستجابة .**

تابع طول الاختبار ودقة تقدير القدرة المقاسة

■ كما توصلت دراسة (Bob, 2009) التي اجريت على عينة جامعية حجمها (208) طالب بالمرحلة الثانوية إلى أن الاختبارات التي تعتمد على التذكر الأكثر طولاً، يكون لها تأثيرات سلبية على دقة قياس القدرة على العكس من الاختبارات الأخرى التي تقيس قدرات عقلية مختلفة وتتمتع بصفة الطول، كما توصلت النتائج إلى أن مؤشرات الثبات والصدق أظهرت تقديرات أدق كلما إنرداد طول الاختبار إلى الطول المناسب.

■ دراسة (Kahraman, Thompson, 2012) والتي استهدفت مقارنة تقديرات قدرات عينة الدراسة من خلال استخدام اختبارات مختلفة الأطوال تحتوي مفردات الاختبار أحادية ومتعددة البعد وذلك باستخدام النموذج الثلاثي البارمتر كنموذج نظرية الاستجابة للمفردة لتقدير القدرة، وقد أظهرت النتائج أن تقديرات القدرة الأكثر دقة ارتبطت بالاختبارات الأكثر طولاً.

واقع تحديد طول الاختبار

- يبدو من خلال تفقد الأدبيات السابقة أن القضية البحثية التي استهدفت دراسة العلاقة بين طول الاختبار وتقديرات القدرة انطلقت من عدم وجود إجراء علمي دقيق وواضح يمكن من خلاله تحديد العدد المناسب لبنود الاختبار والذي يمكن من خلاله تقدير القدرة بنفس الدقة التي يمكن الحصول عليها من عدد أكبر من المفردات .
- كما أن الأطر النظرية الخاصة ببناء الاختبار المتاحة حالياً لم تشير إلى الكيفية أو المنهجية التي يمكن من خلالها تحديد العدد المناسب من بنود الاختبار ، وبالتالي في حال تواجد هذا الإجراء وتلك الأطر النظرية ما كانت هناك حاجة وداعي لبحث هذه النوعية من القضايا .

الأسس العلمية لتحديد العدد المناسب لفقرات الاختبار

- هذا ومن خلال خبرات الباحث المتحدث في مجال القياس وإعداد الاختبارات ، وإطلاعة 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) ، والتي يمكن أن تقدر غالبا من خلال واضع الاختبار بطريقة ذاتية .

الرؤية الجديدة

■ ونظراً لأن الاتجاهات الحديثة للقياس وما آلت إليه من اختبارات حاسوبية موائمة تعتمد على بعض المؤشرات الإحصائية في اتقاء المفردات الخاصة بها، ومن بينها مؤشر أقصى معلومات، الذي يمكن الحصول عليه من تحليلات النماذج اللوغاريتمية المرتبطة بنظرية الاستجابة للمفردة (IRT).

■ وحيث أن هذه النوعية من الاختبارات تستخدم في تقدير قدرات الأفراد بأقل عدد من المفردات ، فإن المتحدث يرى إمكانية استخدام نفس المؤشر في تقليص عدد المفردات الاختبارية لأي اختبار تقليدي خطي، بما لا يؤثر على تقديرات القدرة التي تم الحصول عليها من الاختبار الأصلي (الكلبي)، وبما لا يؤثر على الخصائص السيكومترية والإحصائية للاختبار وهذا ما حدا بالباحث إلى دراسة واستقصاء هذه الفكرة من خلال تجربة ميدانية.

تساؤلات

- 1 - هل تختلف جوهريا تقديرات القدرة المتناظرة للعينة الكلية المشتقة من كل من الاختبار الكلي والاختبار الفرعي باختلاف مستوى أقصى قيم للمعلومات تمتلكها المفردات المحتواه بالاختبار الفرعي (مرتفعة -منخفضة) ؟
- 2 - هل تختلف جوهريا تقديرات القدرة المتناظرة للعينة الكلية المشتقة من كل من الاختبار الكلي والاختبارات الفرعية المشتقة من نفس الاختبار وتحتوي على أعداد متفاوتة من حيث عدد المفردات التي تمتلك أقصى قيم للمعلومات عند مستويات مختلفة من القدرة (وفقا لمقياس التساعيات) ؟
- 3 - ما أقل عدد من المفردات الاختبارية التي تمتلك أقصى معلومات التي يمكن أن تعطي نفس تقديرات القدرة التي يمكن الحصول عليها من الاختبار الكلي ؟
- 4 - هل تختلف الخصائص السيكومترية والإحصائية للاختبارات الفرعية مختلفة الأطوال والتي تتضمن مفردات تمتلك أقصى معلومات على فترات مختلفة من القدرة المقاسة وفقا للمقياس التساعي عن الخصائص السيكومترية والإحصائية للاختبار الأصلي المرجعي ؟

الإجراءات المتبعة في الإجابة عن التساؤلات

- ١- تحليل مفردات الاختبار الكلي (أداة التجربة) المتضمن لجميع المفردات بلا استثناء، باستخدام النموذج اللوغاريتمي ثلاثي البار متر على عينة مكونة من (١٥٢٠) فرد لتحديد الأفراد وكذلك المفردات غير الملائمين لاستكمال إجراءات التحليل، وفقاً لمؤشر اختبار حسن المطابقة (اختبار مربع كاي) والذي يشير إلى دلالة مطابقة التوزيع المتوقع لقدرات الأفراد مع التوزيع الملاحظ لها، وكذلك توزيع صعوبات البنود المتوقعة مع توزيع الصعوبة الملاحظ
- تعتبر قيمة مربع كاي الدالة مؤشراً لعدم ملائمة الفرد للتحليلات ونفس الشيء من مع كل مفردة من مفردات الاختبار.

تابع الإجراءات

- 2 - حذف واستبعاد المفردات والأفراد غير الملائمين للتحليلات وفقاً للمؤشر المذكور في الخطوة السابقة.
- 3 - إعادة التحليل مرة أخرى على الأفراد وكذلك المفردات الملائمة بهدف الحصول على تقديرات لكل من صعوبات وأقصى معلومات مفردات الاختبار، وكذلك الحصول على التدرج المرجعي الخاص بتقديرات القدرة لكل فرد من أفراد العينة الملائمين للتحليلات.

تابع الإجراءات

- 4 -التحقق من توافر افتراضات نظرية الاستجابة للمفردة في بيانات الاختبار الكلي وهي افتراضات أحادية البعد ، استقلالية القياس ، السرعة، واستبعاد التحقق من افتراض توافري المنحنيات المميزة للبند الذي يعني امتلاك مفردات الاختبار لقوة تمييز متساوية ، حيث ان هذا الافتراض يجرى التحقق منه فقط عند استخدام النموذج اللوغاريتمي أحادي المعلم (نموذج مراهي) (Chon, K.; Lee, W.; Dunbar, S.,2010:326) والتجربة الحالية تعتمد على نموذج اخر لايقوم على هذا الافتراض وهو النموذج اللوغاريتمي ثلاثي البار متر.
- 5 -تقسيم الاختبار الكلي - باستخدام مقياس الوسيط - إلى اختبارين فرعيين يحتوي أحدهما على المفردات التي تمتلك أعلى قيم لمؤشر أقصى معلومات (فيكون الاختبار مرتفع المعلومات) ، والاخر يحتوي على المفردات التي تمتلك أدنى قيم لمؤشر أقصى معلومات (فيكون الاختبار منخفض المعلومات) . مع مراعاة أن تمثل الأسئلة في كل منهما المجالات الأربع (الأبعاد الأربعة للاختبار).

تابع الإجراءات

5- تقسيم المدى الكلي لصعوبة مفردات الاختبار الكلي إلى تسع مستويات مختلفة من حيث الصعوبة (والتي تقابل نفس مستويات القدرة المقاسة بالاختبار ولهذا السبب يقاس مستوى الصعوبة للمفردة ومستوى القدرة للمستجيب عنها بنفس الوحدة (Hambelton & Swaminathan , 1985:65) اعتمادا على المقياس التساعي ، وقد اختير هذا المقياس نظرا لإعتماده من قبل العديد من الكليات الجامعية في تقدير المستوى التحصيلي للطلاب في مختلف المقررات الدراسية حيث يجرى تقدير مستوى الطالب التحصيلي وفقا لتسعة مستويات متباينة وهي (أ+، أ، ب+، ب، ج+، ج، د+، د، هـ) ، وتتميز هذه المستويات بحساسيتها في إبراز الفروق الفردية للطلاب على متصل أي قدرة بشكل واضح دقيق .

6- ترتيب مفردات الاختبار - 80 مفردة - التي تمثل كل فترة من الفترات التسع المختلفة ، وذلك وفقا لمتغير أقصى معلومات ، وذلك بهدف اشتقاق اختبارات فرعية مختلفة الأطوال تتضمن مفردات تمتلك أقصى قيم للمعلومات تمثل الفترات التسعة المختلفة للقدرة المقاسة من الاختبار ، تباعا للخطوات التالية:

تابع الإجراءات

- اختيار (8) مفردات تمتلك أقصى معلومات تمثل المجالات الأربعة المتضمنة بالاختبار من بين المفردات المناظرة لكل فترة من فترات القدرة التسع، فيكون عدد مفردات الاختبار الفرعي الناتج (72) مفردة.
- تكرر نفس الخطوة السابقة ولكن بانتقاء مفردات مختلفة العدد [(1) ، (2) ، (3) ، (4) ، (5) ، (6) ، (7)] من كل فترة من فترات القدرة التسع، وبالتالي تكون الاختبارات الفرعية الناتجة عن هذه الخطوات تتضمن مفردات أعدادها (63 ، 54 ، 45 ، 36 ، 27 ، 18 ، 9) على الترتيب.
- ٧ - تحليل كل اختبار فرعي ناتج عن كل خطوة سابقة، لإيجاد التدرج الجديد لتقديرات القدرة لكل فرد من أفراد عينة التجربة والمشتق من التحليلات الناتجة عن استخدام كل اختبار فرعي على حدة.
- ٨ - إيجاد الفروق في تقديرات القدرة المناظرة لكل فرد من أفراد العينة المشتقة من تحليل الاختبار الكلي، والأخرى الناشئة من كل اختبار فرعي على حدة.

تابع الإجراءات

- ٩- إيجاد مجموع الخطأين المعيارين المشتقين من تحليل الاختبار الكلي وكل اختبار فرعي على حدة لكل فرد من أفراد العينة البحثية
- ١٠- تحديد دلالة الفرق بين تقديرات القدرة المشتقة من الاختبار الكلي والاختبار الفرعي باستخدام القاعدة التالية:
 - إذا كان الفرق بين تقديري القدرة المشتقين من الاختبار الكلي والآخر الفرعي أكبر من أو يساوي مجموع الخطأين المعيارين لهما اعتبر ذلك فرقاً جوهرياً ، أما إذا قل اعتبر ذلك فرقاً غير جوهرياً .
- ١١- إيجاد النسبة المئوية للفروق الجوهرية (الدالة) في تقديرات القدرة المشتقة من الاختبار الكلي وكل من الاختبارين الفرعيين (مرتفع ومنخفض) المعلومات .
- ١٢- إيجاد دلالة الفرق بين كل نسبتين مؤيتين متناظرتين باستخدام اختبار (z) لدلالة الفرق بين نسبتين .
- ١٣- إيجاد النسبة المئوية للفروق الجوهرية (الدالة) في تقديرات القدرة المشتقة من الاختبار الكلي وكل من اختبار فرعي على حدة الناتج من الخطوة رقم (5) فإذا تجاوزت هذه النسبة القيمة (0.05) دل ذلك على وجود فروق جوهرية بين تقديرات القدرة المشتقة من الاختبار الكلي والآخر الفرعي .

تابع الإجراءات

- ١٤- إيجاد دلالة الفروق بين نسبة تجاوز تقديرات القدرة بين الاختبار الكلي والاختبار فرعي محدد ونسبة تجاوز تقديرات القدرة بين الاختبار الكلي والاختبار فرعي آخر.
- ١٥ - استخدام تحليل تباين القياس المتكرر لتقديرات القدرة المشتقة من الاختبار الكلي المرجعي والاختبارات الفرعية الثمانية التي تمتلك مفرداتها أقصى معلومات على مستويات القدرة التسعة والمختلفة من حيث الطول. وذلك كأسلوب إحصائي - إجراء آخر - للتحقق من صحة الناتج التي توصل إليها من الإجراء السابق.
- ١٦ - التحقق من مدى اختلاف أو تكافؤ الخصائص السيكومترية (الثبات، الصدق) والخصائص الإحصائية للاختبارات الفرعية مع الخصائص السيكومترية والإحصائية للاختبار الكلي المرجعي وذلك من خلال الإجراءات التالية والتي تعتمد على التحليلات الكيفية لبعض المؤشرات الإحصائية:
- إيجاد معامل ارتباط بيرسون لتقديرات القدرة المشتقة من الاختبار الكلي بكل من تقديرات القدرة المشتقة من كل اختبار فرعي مستخدم بالتجربة (الثماني اختبارات فرعية). باعتبار أن تقديرات القدرة المشتقة من الاختبار المرجعي محكا خارجيا لتقديرات القدرة المشتقة من الاختبارات الفرعية مختلفة الأطوال كمؤشر على صدق المحك الخارجي.

النتائج

■ النتائج الخاصة بإجراءات الإجابة عن السؤال الأول :

■ وقد استخدم إجراءين مختلفين للإجابة عن هذا السؤال الأول والذي ينص على " هل تختلف جوهريا تقديرات القدرة المتناظرة للعينة الكلية المشتقة من كل من الاختبار الكلي وتقديرات القدرة المشتقة من الاختبار الفرعي باختلاف مستوى أقصى قيم للمعلومات تمتلكها المفردات المحتواه بالاختبار الفرعي (مرتفعة - منخفضة) ؟

■ **الإجراء الأول :** يتضمن استخدام اختبار النسبة المحرجة لدلالة الفرق بين نسبتين مؤيتين .

■ **الإجراء الثاني :** يتضمن استخدام اختبار دلالة تباين القياس المتكرر .

نتائج الإجراء الأول : وهو استخدام اختبار النسبة الحرجة.

تكرارات الفروق الدالة بين تقديرات القدرة المشتقة من الاختبار الكلي وكل من الاختبارين الفرعيين (منخفض ومرتفع المعلومات) والنسب المئوية لها ودلالة هذه الفروق باستخدام اختبار (Z)

| مستوى الدلالة | قيمة اختبار (z) | النسبة المئوية | نسبة عدد نسبة الفروق الجوهرية إلى العدد الكلي للفروق | تكرارات الفروق الجوهرية بين تقديرات القدرة المتناظرة | العدد الكلي للفروق | الفروق في تقديرات القدرة المتناظرة |
|---------------|-------------------|----------------|--|--|--------------------|--------------------------------------|
| 0.000 | 16.5 | .8 | 0.008 | 12 | 1520 | بين الاختبار الكلي والاختبار المرتفع |
| | | 3.2 | 0.032 | 49 | 1520 | بين الاختبار الكلي والاختبار المنخفض |

الإجراء الثاني : استخدام اختبار دلالة تباين القياس المتكرر.

■ وقد أسفر هذا الإجراء عن وجود دلالة بين تباين القياسات الثلاث:

دلالة تباين القياسات الثلاث المشتقة من الاختبار الكلي والاختباران الفرعيان (المرتفع والمنخفض) من حيث قيم أقصى معلومات

| مصدر التباين | الاختبار | مجموع المربعات | درجات الحرية | متوسط المربعات | قيمة (ف) | مستوى الدلالة |
|-------------------------|--------------------|----------------|--------------|----------------|----------|---------------|
| بين القياسات | Sphericity Assumed | 4555.257 | 2 | 2277.628 | 426.199 | .000 |
| | Greenhouse-Geisser | 4555.257 | 1.140 | 3995.164 | 426.199 | .000 |
| | Lower-bound | 4555.257 | 1.000 | 4555.257 | 426.199 | .000 |
| داخل القياسات (الخطأ) | Sphericity Assumed | 16181.763 | 3028 | 5.344 | | |
| | Greenhouse-Geisser | 16181.763 | 1726.252 | 9.374 | | |
| | Lower-bound | 16181.763 | 1514.000 | 10.688 | | |

المقارنات الثنائية بين متوسطات تقديرات القدرة المشتقة الاختبار الكلي المرجعي والاختبارات الفرعية باستخدام اختبار (LSD)

| الفرق بين متوسطى طرفى المقارنة (الأول - الثانى) | طرف المقارنة الثانى | طرف المقارنة الأول |
|---|------------------------|--------------------|
| .095 | الاختبار الفرعى الأول | الاختبار الكلي |
| .190- | الاختبار الفرعى الثانى | |
| .083- | الاختبار الفرعى الثالث | |
| .139- | الاختبار الفرعى الرابع | |
| .605- | الاختبار الفرعى الخامس | |
| .689-* | الاختبار الفرعى السادس | |
| 2.857-* | الاختبار الفرعى السابع | |
| 2.933-* | الاختبار الفرعى الثامن | |

نتائج إجراءات الإجابة عن السؤال الثاني

- لقد استخدمت التجربة في الإجابة عن السؤال الثاني والذي ينص على " هل تختلف جوهريا تقديرات القدمرة المتناظرة للعينة الكلية المشتقة من كل من الاختبار الكلي والاختبارات الفرعية المشتقة من نفس الاختبار وتحتوي على أعداد متفاوتة من حيث عدد المفردات التي تتصف بامتلاكها لأقصى قيم لدوال المعلومات عند مستويات مختلفة من القدمرة (وفقا لمقياس التساعيات) ؟" عدد من الإجراءات وهي على التوالي:
- أ- إجراء اشتقاق مفردات الاختبارات الفرعية مختلفة الأطوال والتي تمتلك مفرداتها أعلى قيم لمتغير أقصى معلومات على تسع فترات مختلفة على متصل القدمرة المقاسة وفقا للمقياس التساعي .
- وقد أسفر استخدام هذا الإجراء عن النتائج الموضحة بالجدول التالي: والذي يوضح نتائج التحليل الخاصة بنقاط مقياس التساعيات لتحديد تسعة مستويات مختلفة على متصل القدمرة المقاسة بالاختبار الكلي (المرجعي) .

قيم التساعيات

قيم نقاط مقياس التساعيات للقدرة المقاسة بالاختبار الكلي (المرجعي)

| التساعي | قيمة التساعي |
|---------|--------------|
| الأول | 0.603 |
| الثاني | 1.088 |
| الثالث | 1.211 |
| الرابع | 1.416 |
| الخامس | 1.564 |
| السادس | 1.959 |
| السابع | 2.092 |
| الثامن | 2.829 |

توزيع مفردات الاختبارات الكلية على الاختبارات الفرعية وفقا لمتغير أقصى معلومات

المفردات المحتواة بالاختبارات الفرعية التي تمتلك أقصى معلومات والموزعة على ابعاد الاختبار الكلي

| عدد المفردات | | | | الاختبار الفرعي | |
|----------------|--------|--------|-------|-----------------|--------|
| أبعاد الاختبار | | | الكلي | | |
| الرابع | الثالث | الثاني | | الأول | |
| 11 | 12 | 29 | 20 | 72 | الأول |
| 10 | 11 | 26 | 16 | 63 | الثاني |
| 9 | 9 | 22 | 14 | 54 | الثالث |
| 7 | 5 | 19 | 14 | 45 | الرابع |
| 3 | 4 | 18 | 11 | 36 | الخامس |
| 3 | 4 | 13 | 7 | 27 | السادس |
| 2 | 3 | 9 | 4 | 18 | السابع |
| 1 | 1 | 4 | 3 | 9 | الثامن |

نتائج إجراء استخدام اختبار (Z) لنسب الفروق الجوهرية بين تقديرات القدرة المشتقة من الاختبار الكلي والاختبار الفرعي المستخدم

■ انظر جدول النتائج (بصيغة Execl)

نتائج الإجراء الثاني وهو استخدام تحليل القياس المتكرر في الكشف عن دلالة الفروق بين تقديرات القدرة المشتقة من الاختبار الكلي والخرى المشتقة من الاختبارات الفرعية

■ وقد أسفر استخدام هذا الأسلوب عن النتائج الموضحة بالمجداول التالية:

نتائج تحليل تباين القياس المتكرر لتقديرات القدرة المشتقة من كل من الاختبار الكلي والاختبارات الفرعية مختلفة الأطوال

| مصدر التباين | الاختبار | مجموع المربعات | درجات الحرية | متوسط المربعات | قيمة (ف) | مستوى الدلالة |
|-------------------------|--------------------|----------------|--------------|----------------|----------|---------------|
| بين القياسات | Sphericity Assumed | 276.831 | 8 | 34.604 | 3.444 | 0.001 |
| | Greenhouse-Geisser | 276.831 | 1.992 | 138.948 | 3.444 | 0.036 |
| | Lower-bound | 276.831 | 1 | 276.831 | 3.444 | 0.069 |
| داخل القياسات (الخطأ) | Sphericity Assumed | 4500.977 | 448 | 10.047 | | |
| | Greenhouse-Geisser | 4500.977 | 111.571 | 40.342 | | |
| | Lower-bound | 4500.977 | 56 | 80.375 | | |

المقارنات الثنائية بين متوسطات تقديرات القدرة المشتقة الاختبار الكلي المرجعي والاختبارات الفرعية باستخدام اختبار (LSD)

| الفرق بين متوسطى طرفى المقارنة (الأول - الثانى) | طرف المقارنة الثانى | طرف المقارنة الأول |
|---|------------------------|--------------------|
| .095 | الاختبار الفرعى الأول | الاختبار الكلي |
| .190- | الاختبار الفرعى الثانى | |
| .083- | الاختبار الفرعى الثالث | |
| .139- | الاختبار الفرعى الرابع | |
| .605- | الاختبار الفرعى الخامس | |
| .689-* | الاختبار الفرعى السادس | |
| 2.857-* | الاختبار الفرعى السابع | |
| 2.933-* | الاختبار الفرعى الثامن | |

* دال عند مستوى (٠,٠٥)

■ ثالثاً : نتائج الإجراءات الخاصة بالإجابة عن السؤال الثالث:

■ أسفرت نتائج التجربة الخاصة بالإجابة عن السؤال الثالث والذي ينص على " ما أقل عدد من المفردات الاختبارية التي تمتلك أقصى معلومات التي يمكن أن تعطي نفس تقديرات القدمرة التي يمكن الحصول عليها من الاختبار الكلي؟ "

■ يتبين من النتائج السابقة أن أقل عدد من المفردات (٣٦) مفردة تمتلك أقصى معلومات عند المستويات المختلفة للقدمرة المقاسة

نتائج الإجراءات الخاصة بالإجابة عن السؤال الرابع

- أسفرت نتائج التجربة الخاصة بالإجابة عن السؤال الرابع والذي ينص على: هل تختلف الخصائص السيكومترية والإحصائية للاختبارات الفرعية مختلفة الطول والتي تتضمن مفردات تمتلك أقصى معلومات على فترات مختلفة من القدمة المقاسة وفقاً للمقياس التساعي عن الخصائص السيكومترية والإحصائية للاختبار الأصلي المرجعي؟

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

لقد اعتمدت الإجراءات في التحقق من الخصائص السيكومترية (الثبات والصدق) الاختبار الكلي والاختبارات الفرعية (الثمانية) مختلفة الأطوال والمشتقة من نفس الاختبار - تحتوي على مفردات لها أقصى قيم للمعلومات على متصل القدمرة المقاسة - على:

■ **معامل ألفا كرونباخ:** للتحقق من ثبات الاتساق الداخلي .

■ **وكذلك معامل ارتباط بيرسون:** بين تقديرات القدمرة المشتقة من الاختبار الكلي (المرجعي) والاختبارات الفرعية مختلفة الأطوال المشتقة من نفس الاختبار باعتبار أن الاختبار الكلي هو اختباراً محكياً يمكن الاعتماد عليه لتقدير صدق هذه الاختبارات الفرعية .

■ **ويبدو ذلك أمراً منطقياً ، نظراً لأن التجربة الحالية تسعى للكشف عن مدى اتساق الخصائص السيكومترية للاختبار الكلي مع الاختبارات الفرعية المشتقة من نفس الاختبار الكلي والتي تمتلك مفرداتها أقصى قيم للمعلومات وتختلف من حيث أطوالها، وذلك للإجابة عن الشق الأول من هذا السؤال وقد أسفرت نتائج التحليل عما يلي:**

نتائج التحليلات الخاصة بالثبات:

- لقد اسفرت التحليلات الخاصة باستخدام معامل ألفا كرونباخ لحساب ثبات الاتساق الداخلي لمفردات الاختبار الكلي والاختبارات الفرعية الثمانية المشتقة من هذا الاختبار عن البيانات الموضحة بالمجدول التالي:

معامل ألفا لتقدير ثبات الاتساق الداخلي

| الاختبار | عدد المفردات | معامل ألفا لثبات للاتساق الداخلي |
|------------------------|--------------|----------------------------------|
| الاختبار الكلي | 80 | 0.84 |
| الاختبار الفرعي الأول | 72 | 0.84 |
| الاختبار الفرعي الثاني | 63 | 0.84 |
| الاختبار الفرعي الثالث | 54 | 0.83 |
| الاختبار الفرعي الرابع | 45 | 0.82 |
| الاختبار الفرعي الخامس | 36 | 0.79 |
| الاختبار الفرعي السادس | 27 | 0.68 |
| الاختبار الفرعي السابع | 18 | 0.57 |
| الاختبار الفرعي الثامن | 9 | 0.41 |

نتائج التحليلات الخاصة بالصدق:

لقد اسفرت التحليلات الخاصة باستخدام معامل ارتباط بيرسون لتقديرات القدرة المتناظرة المشتقة من تحليلات الاختبار الكلي المرجعي، وتحليلات الاختبارات الفرعية الثمانية عن النتائج الموضحة بالجدول التالي:

معاملات الارتباط بين تقديرات القدرة المقدره من كل اختبار فرعي وتقديرات القدرة المقدره من الاختبار الكلي

| معامل ارتباط القدرة المقدره من الاختبار الفرعي مع تقدير القدرة المقدر من الاختبار الكلي | عدد المفردات | الاختبارات الفرعية |
|--|--------------|--------------------|
| .991** | 72 | الفرعي الأول |
| .981** | 63 | الفرعي الثاني |
| .964** | 54 | الفرعي الثالث |
| .939** | 45 | الفرعي الرابع |
| .917** | 36 | الفرعي الخامس |
| .858** | 27 | الفرعي السادس |
| .797** | 18 | الفرعي السابع |
| .658** | 9 | الفرعي الثامن |

** دال عند مستوى (0.01)

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

■ لقد استخدمت التجربة في التحقق من الخصائص الإحصائية للاختبار الكلي والاختبارات الفرعية (الثمانية) مختلفة الأطوال والمشتقة من نفس الاختبار -تحتوي على مفردات لها أقصى قيم للمعلومات على متصل القدمرة المقاسة- مقاييس النزعة المركزية (المتوسط ، الوسيط ، المنوال ، أقل تقدير للقدمرة ، أعلى تقدير للقدمرة) ، ومقاييس والتشتت (المدى ، الإنحراف المعياري ، التباين) وكذلك المؤشرات الخاصة بتوزيع تقديرات القدمرة (الإلتواء ، التفرطح) وقد أسفرت نتائج التحليل عما يلي :

نتائج الإحصاءات الوصفية لتقديرات القدرة المشتقة من تحليلات الاختبارات الفرعية على عينة الدراسة

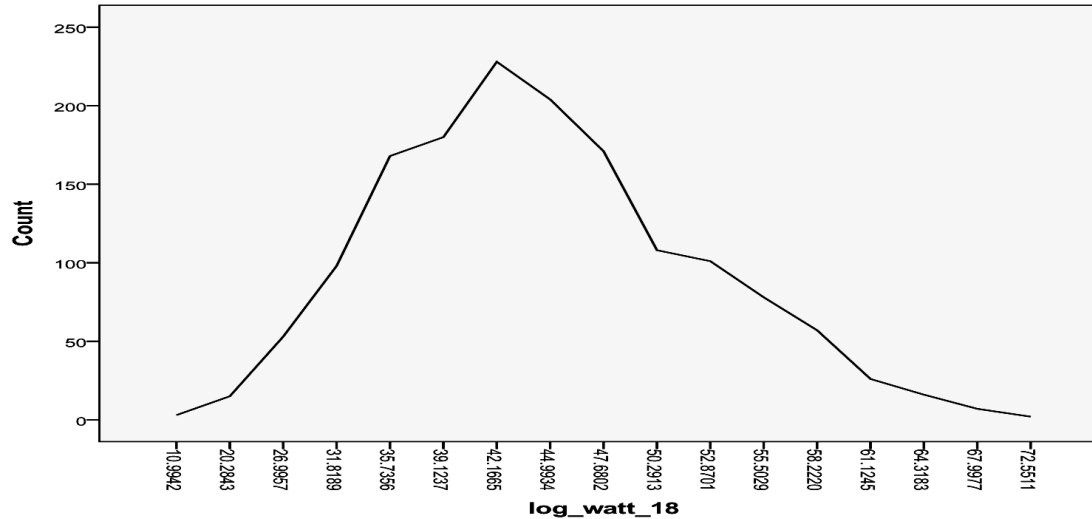
| الاختبارات الفرعية | | | | | | | | الاختبار الكلبي المرجعي | الإحصاءات الوصفية |
|--------------------|--------|--------|--------|--------|--------|--------|-------|-------------------------------|-------------------------|
| الثامن | السابع | السادس | الخامس | الرابع | الثالث | الثاني | الأول | | |
| 9 | 18 | 27 | 36 | 45 | 54 | 63 | 72 | 80 | عدد المفردات |
| 46.5 | 45.9 | 43.5 | 42.8 | 42.7 | 42.6 | 42.5 | 42.5 | 42.4 | المتوسط |
| 0.27 | 0.22 | 0.2 | 0.19 | 0.19 | 0.19 | 0.19 | 0.18 | 0.17 | الخطأ المعياري للمتوسط |
| 39.3 | 40 | 42.8 | 42.3 | 42 | 42 | 42 | 42 | 41.87 | الوسيط |
| 40.9 | 41.2 | 38.1 | 41.6 | 41.5 | 42.4 | 42.3 | 41.9 | 41.89 | المنوال |
| 10.5 | 8.7 | 7.7 | 7.5 | 7.5 | 7.4 | 7.3 | 7 | 6.98 | الإحتراف المعياري |
| 109.6 | 75.7 | 58.6 | 56.9 | 56.7 | 54.9 | 53.3 | 49 | 44.63 | التباين |
| - | -2 | 1.21 | 0.069 | 0.09 | 0.05 | 0.021 | 0.077 | 0.015 | الإلتواء |
| 0.091 | 0.09 | 0.09 | 0.063 | 0.06 | 0.06 | 0.063 | 0.063 | 0.06 | الخطأ المعياري للإلتواء |
| 0.216 | 0.34 | 0.36 | - | - | -0.1 | - | -0.23 | - | التفرطح |
| 0.126 | 0.13 | 0.13 | 0.126 | 0.13 | 0.13 | 0.126 | 0.126 | 0.13 | الخطأ المعياري للتفرطح |

النتائج الخاصة بخصائص منحى تقديرات القدرة المشتقة من جميع الاختبارات مختلفة الأطوال (من حيث الاعتدالية)

قيم ودلالة اختبار كلمجروف - سميرنوف لاعتدالية توزيع تقديرات القدرة المشتقة من الاختبار الكلي وجميع الاختبارات الفرعية مختلفة الأطوال

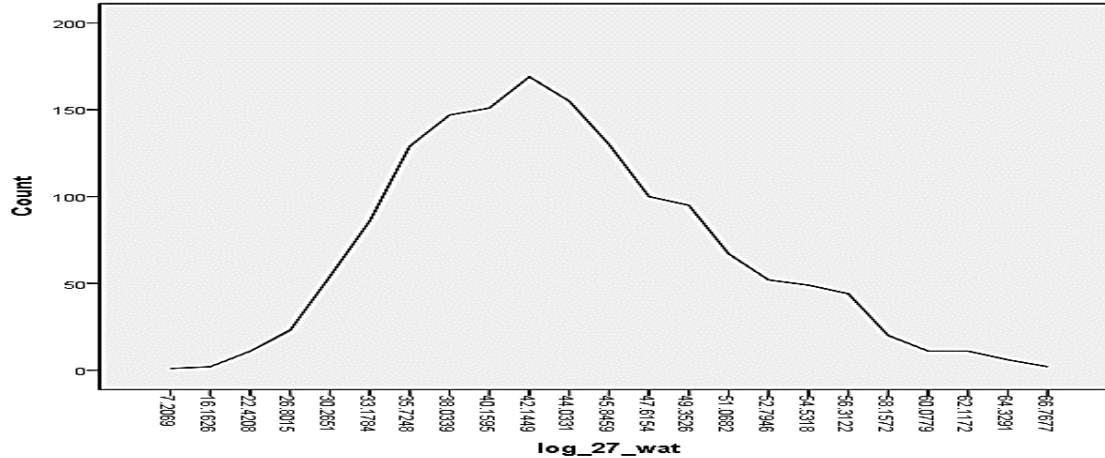
| الاختبار الفرعي الثامن (9) مفردة | الاختبار الفرعي السابع (18) مفردة | الاختبار الفرعي السادس (27) مفردة | الاختبار الفرعي الخامس (36) مفردة | الاختبار الفرعي الرابع (45) مفردة | الاختبار الفرعي الثالث (54) مفردة | الاختبار الفرعي الثاني (63) مفردة | الاختبار الفرعي الأول (72) مفردة | الاختبار الكلي (المرجعي) (80) مفردة | الاختبار الإحصائي ومستوى الدلالة |
|----------------------------------|-----------------------------------|-----------------------------------|-----------------------------------|-----------------------------------|-----------------------------------|-----------------------------------|----------------------------------|-------------------------------------|----------------------------------|
| 0.196 | 0.172 | 0.17 | 0.144 | 0.098 | 0.087 | 0.079 | 0.073 | 0.024 | Kolmogorov-Smirnov Z |
| 0.058 | 0.081 | 0.097 | 0.188 | 0.321 | 0.511 | 0.675 | 0.782 | .0.980 | Asymp. Sig. (2-tailed) |

التمثيلات البيانية لمنحنى تقديرات القدرة المشتقة من الاختبار الكلي والاختبار المكون من (٣٦) مفردة



التمثيل البياني لتوزيع تقديرات قدرات عينة التجربة من
التحليل بالاختبار الفرعي المكون من (80) مفردة

تابع: التمثيلات البيانية لمنحنى تقديرات القدرة المشتقة من الاختبار الكلي والاختبار المكون من (٣٦) مفردة



التمثيل البياني لتوزيع تقديرات قدرات عينة التجربة من التحليل
بالاختبار الفرعي المكون من (36) مفردة

توصيات

- من خلال النتائج التي تم التوصل إليها يمكن وضع التوصيات التالية:
- ضرورة استخدام مؤشر دوال المعلومات للمفردات الاختبارية (متغير أقصى معلومات على متصل القدرة المقاسة) التي يمكن الحصول عليه من تحليلات النموذج اللوغاريتمي ثلاثي البارمتر **Three Parameter Logistic Model (3-PL)** كإجراء أساسي في إجراءات بناء الاختبارات المختلفة، وبصورة خاصة قبل إعداد الصورة النهائية للاختبار مباشرة، مما يؤدي إلى الحصول على صورة نهائية مختصرة للاختبار توفر الكثير من الوقت والجهد لكل من مستخدم الاختبار وكذلك الجمهور المستهدف من الاختبار، كما أنها تيسر سهولة التعامل مع عدد أقل من البيانات الذي ييسر بدوره في عمليات التصحيح وتقدير الدرجات.

النهائية