# تأثيرات طول الاختبار وطريقة التصحيح ونموذج تحليل المفردة على تقديرات القدرة ومعايير التحسين للاختبارات ذات المفردات المختلطة

**إعداد/ دكتورة شيرين فاروق محمد طنطاوي**

**مدرس علم النفس الإحصائي بكلية الآداب – جامعة الفيوم**

**المستخلص:**

هدفت الدراسة إلى البحث في تأثيرات طول الاختبار (١٠، ٢٠، ٣٠ مفردة) وطريقة التصحيح (نسبة التصحيح ثنائي الاستجابة ومتعدد الاستجابة) ونموذج تحليل المفردة (باستخدام مجموعتين من النماذج لتصحيح المفردات المختلطة: المجموعة الأولى (نموذج راش ثنائي الاستجابة Raschونموذج التقدير الجزئي متعدد الاستجابات(PCM، والمجموعة الثانية (النموذج ثنائي البارامتر ثنائي الاستجابات 2PLونموذج الاستجابة المتدرجة متعدد الاستجابات (GRM. على تقديرات القدرة ودالة معلومات الاختبار ومعايير التحسين للاختبارات ذات المفردات المختلطة. تكونت عينة الدراسة من ٣٤٠ طالباً وطالبة بالفرقة الثالثة من كلية الآداب جامعة الفيوم الذين يدرسون مقرر الاختبارات والمقاييس النفسية. ثم تم تحليل جميع الاستجابات للمفردات للتحقق من أول افتراضات نظرية الاستجابة للمفردة وهو افتراض أحادية البعد وذلك باستخدام طريقة التحليل العاملي في برنامج SPSS وذلك بطريقة الارجحية القصوى. تم الوصول إلى عامل واحد لكل مجموعة من المفردات. ولتحليل الاستجابات للاختبارات المختلطة تم استخدام مجموعتين من النماذج وهما المجموعة الأولى (نموذج راش ثنائي الاستجابة Rasch ونموذج التقدير الجزئي متعدد الاستجابات(PCM، والمجموعة الثانية (النموذج ثنائي البارامتر ثنائي الاستجابات 2PLونموذج الاستجابة المتدرجة متعدد الاستجابات (GRM. أشارت النتائج إلى أن كلاً من طول الاختبار ونسبة المفردات التي تم تصحيحها بشكل متعدد الاستجابة كان لها تأثير كبير على كمية معلومات الاختبار الناتجة عن الاختبارات ذات المفردات المختلطة. بشكل عام، أدت الاختبارات التي تم إجراؤها على ١٠٠٪ من المفردات التي تم تصحيحها بشكل متعدد الاستجابات إلى الحصول على أعلى معلومات شاملة. يبدو أن هذا ينطبق بشكل خاص على الممتحنين ذوي تقديرات القدرة المنخفضة. تمت المقارنة بين المجموعتين من النماذج من حيث قيم الخطأ

القياسية لتقديرات القدرة والثبات الهامشية ومؤشرات المطابقة (2LL-). وكانت الفروق بينهما واضحة في معدلات الخطأ المعيارية.

**الكلمات المفتاحية**: طول الاختبار، طريقة التصحيح، نموذج تحليل المفردة، تقديرات القدرة، دالة معلومات الاختبار، معايير التحسين، الاختبارات ذات المفردات المختلطة.

**المقدمة**

أخذ الاهتمام بالقياس شكل كبير طوال القرن العشرين وكذلك القرن الحالي، كما أصبح الاعتماد على المعلومات المشتقة من درجات الاختبار أهمية كبيرة في مجتمعنا المعاصر(Erguven,2014) . حيث تعتمد جميع مستويات التعليم، من مرحلة رياض الأطفال إلى مرحلة الدراسات العليا، ومعظم إجراءات الترخيص المهنية والعديد من طرق التوظيف، اعتمادًا كبيرًا على أداء الاختبار لنشر الفرص وتعزيز المعايير المهنية وضمانها(Adegoke,2013 ). فجودة الاختبار لها أولوية كبيرة لمصممي الاختبارات، وكذلك الذين يعتمدون على درجات الاختبار في اتخاذ القرار. الآن أكثر من أي وقت مضى، حيث أن للاختبارات دور مهم في قياس القدرة وأن ما ينتج عنها من نتائج تعتبر مقاييس دقيقة لقدرة الممتحن. تركز المعايير المستخدمة لتحديد جودة الاختبار بشكل عام على مجالات تصميم الاختبار وتقنيات تحليل الاختبار وتفسير درجات الاختبار(Murphy,2021) .

تتأثر جودة تصميم الاختبار بالعديد من العناصر بما في ذلك الشكل والطول وإجراءات التطبيق والبناء والصدق وطريقة وضع الدرجات. يتمثل العنصر الأكثر أهمية في المفردات المختارة لتكوين الاختبار(Oliveri and von Davier,2014 ). أكثر نوعي المفردات شيوعًا هما "الاستجابة المحددةselected response " و "الاستجابة المركبة constructed response". تتطلب مفردات الاستجابة المحددة من الممتحَنين التعرف على إجابة واحدة واختيارها من مجموعة الاستجابات المقدمة (أي مفردات الاختيار من متعدد والمطابقة والصواب والخطأ). تحتوي هذه المفردات بشكل عام على مخطط تقدير الدرجات "ثنائي الاستجابة dichotomous " يسمح بنتيجتين محتملتين لكل استجابة للمفردة، "صحيحة" أو "غير صحيحة". في المقابل، تجمع مفردات الاستجابة المركبة معلومات حول المعرفة غير المكتملة التي قد يمتلكها الممتحن، من خلال مطالبة الممتحن بإنشاء استجابة لمفردة ما (على سبيل المثال، مفردات استجابة مفتوحة، وإجابة قصيرة، ومفردات مقالة). تمنح مفردات

(E-mail of the correspondig: jcps@art.bsu.edu.eg)

الاستجابة المركبة "مصداقية جزئية" على مقياس هرمي للصواب باستخدام مقدر درجات المعروف باسم تقدير الدرجات "متعدد الاستجابات polytomous ". Von) .Davier et al.,2019)

بمجرد تطبيق الاختبار وجمع الاستجابات، توجد العديد من الأساليب التي يمكن استخدامها لتحليل استجابات المفردات. يشار إلى هذه الأساليب عمومًا باسم "النماذج الرياضية mathematical models " وتستخدم لتقدير قدرة الممتحن وخصائص المفردة. يشار إلى هذه العمليات الحسابية باسم "التقديراتestimates " بسبب خطأ القياس المتأصل المرتبط بجميع اختبارات القدرة المعرفية. ويؤدي تقليل هذا الخطأ إلى تحسين دقة القياس للاختبار وكذلك تقديرات خصائص المفردات (Cohen,2018).

توجد نماذج رياضية مختلفة لنظرية الاستجابة للمفردة (IRT) لتحليل استجابة المفردات بهدف تقدير قدرة الفرد وخصائص المفردة ومنها نموذج راش Rasch (Asriadi & Hadi,2021). تم تطبيق هذا النموذج على استجابات مسجلة بشكل ثنائي، ولكن توجد أيضًا إصدارات موسعة منه للتعامل مع البيانات متعددة الاستجابات وهو نموذج التقدير الجزئيPCM. ويطبق نموذج راش تقنية لوجستية، كما يحتفظ نموذج راش بمعلمات مفردة معينة ثابتة (Che Lah, Tasir& .Jumaat,2021)

غالبًا ما تقارن نماذج نظرية الاستجابة للمفردة بشأن قدرتهم على تقدير كفاءات الممتحن ومعلمات المفردة عند تحليل الاستجابة للمفردات الثنائية أو المتعددة (Rahayu et al.,2021). ومع ذلك، لم تتم مقارنتها على نطاق واسع بشأن قدرتها على تحليل الاختبارات ذات المفردات المختلطة. تتوفر نماذج رياضية مختلفة في نظرية الاستجابة للمفردة لتحليل الاستجابة للمفردات التي تمتلك طرق تقدير مختلفة للدرجات، ولكن ليس ضمن الاختبارات ذات المفردات المختلطة.

**مشكلة الدراسة**

غالبًا ما يطبق المعلمون الاختبارات ذات المفردات المختلطة. يستفيد الطلاب عندما تتاح لهم الفرصة للاستجابة لمجموعة متنوعة من أنواع المفردات ومن المعروف أن الاشكال المختلفة للمفردات تختلف في نوع المعلومات التي يتم جمعها. في الآونة الأخيرة، تم اقتراح نظرية مفادها أن الاختبارات ذات المفردات المختلطة توفر

فرصـــة للاســـتفادة من المزايا الكامنة للأشـــكال المختلفة للمفردات في التحليل (Triwik,2021).

يتم تطبيق نماذج نظرية الاســـتجابة للمفردة في مواقف التقييم على نطاق واســـع للمساعدة في تحليل الاختبار وإجراءات تعديل الاختبار(Stewart et al.,2018) . ومع ذلك، يوجد تحد بين الأســـاليب المتاحة لتحليل أنماط الاســـتجابة من الاختبارات ذات المفردات المختلطة. معظم تقنيات القياس النفســـي المتاحة مخصـصـــة للاستخدام مع اســـتجابات الممتحنين للمفردات ذات الدرجات المماثلة. حتى الآن، تم إجراء التحليل النفســـي بشـكل حصـــري تقريبًا على مجموعات من مفردات الاختبار التي كانت من نفس نوع النتيجة (كلها ثنائية أو كلها متعددة). بالإضـافة إلى ذلك، فإن معاملات ثبات وصـدق نظرية القياس الكلاسـيكية (Classical Test Theory ;CTT) تسـتند إلى أن جميع مفردات الاختبار هي المفردات ذات النسـق الواحد، على سـبيل المثال خيارات متعددة ولا تنطبق بســـهولـة على الاختبـارات ذات المفردات المختلطـة ( Tseng& Wang,2021).

كما أن نماذج نظرية الاسـتجابة للمفردة لهما افتراضـــات يجب أن تلبيها البيانات قبل تطبيق النماذج واعتبار النتائج صـحيحة.. يعتبر نموذج راش معلمة التمييز المشـتركة عبر جميع المفردات صـارمًا ويصعب الوفاء به، والذي بدوره قد يؤدي إلى نتائج أقل صـدقا عند تحليل الاختبارات ذات المفردات المختلطة. إن من فوائد اسـتخدام نموذج راش السـماح بمقياس لوجسـتي مشـترك لصـعوبة المفردة وقدرة الشـخص. غالبًا ما تمت مقارنـة نمـاذج نظريـة الاســـتجـابـة للمفردة على تقديراتهم لمعلمـات المفردة ومســـتويـات كفـاءة الممتحن، ولكن ليس على قدرتهمـا على تحليل الاختبـارات ذات المفردات المختلطة(Wang, Drasgow and Liu,2016) . يجب أيضًا أن يعتمد اختيار الطريقة المثلى لتحليل الاســـتجابات للاختبارات ذات المفردات المختلطة على تطبيق النموذج المناسب لنوع البيانات أو "مطابقة" النموذج وافتراضات النموذج للبيانات.

**تفترض الدراســـة الحالية** أن نســـبة المفردات ثنائية الاســـتجابة ومتعددة الاســـتجابة وطول الاختبار وطريقة تقدير الدرجات وطريقة التحليل تؤثر على معايير التحســـين في الاختبارات ذات المفردات المختلطة. معايير التحسـين، بدورها، لها تأثير ملموس عبر شكل المفردات المختلطة على ما يلي: ثبات تقديرات القدرة، ومعلومات المفردة، ومعلومات الاختبار، وتقديرات الكفاءة، ومعلمات المفردة.

**تسعى الدراسة الحالية الى الإجابة على الأسئلة البحثية الآتية:**

(E-mail of the correspondig: jcps@art.bsu.edu.eg)

١. كيف تؤثر نسبة المفردات ثنائية الاستجابة ومتعددة الاستجابة عبر طول الاختبار ونماذج التحليل على تقديرات القدرة؟

٢. كيف تؤثر نسبة المفردات ثنائية الاستجابة ومتعددة الاستجابة عبر طول الاختبار ونماذج التحليل على دالة معلومات الاختبار (TIF)؟

٣. كيف تؤثر نسبة المفردات ثنائية الاستجابة ومتعددة الاستجابة عبر طول الاختبار ونماذج التحليل على التحسن الإجمالي للاختبار؟

٤. هل توجد فروق بين مجموعتي التصـــحيح: المجموعة الأولى (نموذج راش ثنائي الاســـتجابة Raschونموذج التقدير الجزئي متعدد الاســـتجابات(PCM، والمجموعة الثانية (النموذج ثنائي البارامتر ثنائي الاستجابات 2PLونموذج الاستجابة المتدرجة متعدد الاستجابات (GRM في قدرتها على تحليل الاختبارات ذات المفردات المختلطة في قدرتها على تحليل الاختبارات ذات المفردات المختلطة؟

**أهداف الدراسة**

١. الكشف عن تأثير نسبة المفردات ثنائية الاستجابة ومتعددة الاستجابة عبر طول الاختبار ونماذج التحليل على تقديرات القدرة.

٢. معرفة تأثير نســـبة المفردات ثنائية الاســـتجابة ومتعددة الاســـتجابة عبر طول الاختبار ونماذج التحليل على دالة معلومات الاختبار (TIF).

٣. معرفة تأثير نســـبة المفردات ثنائية الاســـتجابة ومتعددة الاســـتجابة عبر طول الاختبار ونماذج التحليل على التحسن الإجمالي للاختبار.

٤. المقارنة بين مجموعتين باختلاف طريقة التصـــحيح: المجموعة الأولى (نموذج راش ثنائي الاســـتجابة Raschونموذج التقدير الجزئي متعدد الاســـتجابات(PCM، والمجموعة الثانية (النموذج ثنائي البارامتر ثنائي الاستجابات 2PLونموذج الاستجابة المتدرجة متعدد الاستجابات (GRM في قدرتها على تحليل الاختبارات ذات المفردات المختلطة.

**أهمية الدراسة:**

تصميم الاختبار، وأساليب التحليل، وقضايا التحسين كلها عوامل حاسمة تؤثر على تقديرات القدرة، ومعلومات الاختبار، وفي النهاية ثبات تقديرات القدرة التي ينتجها أي اختبار. من المهم استكشاف ومقارنة:

(أ) أشكال الاختبار للاختبارات ذات المفردات المختلطة.

(ب) النماذج الرياضية المستخدمة لتحليل الاختبارات ذات المفردات المختلطة.

(ج) أساليب التحسين لتصميم وتحليل الاختبارات ذات المفردات المختلطة. لصالح أولئك الذين يخضعون للاختبارات، والذين يعدّون الاختبارات، وأولئك الذين يستخدمون درجات الاختبار لاتخاذ القرار، تحتاج هذه المجالات إلى مزيد من الاهتمام في مجال أساليب القياس النفسي.

**أشكال الاختبار:**

لقد أتاح التقييم ثنائي الاستجابة فرصة لتطبيق اختبارات أطول من أجل زيادة ثبات الدرجة مع الاستفادة من كفاءة خيارات وضع الدرجات بشكل آلي. في الآونة الأخيرة، تم التعرف على وضع الدرجات متعددة الاستجابة على أنها تقدم شكلاً أكثر دقة لقياس مستوى المهارة الذي يمتلكه الأفراد. مع فئات الاستجابة المتعددة وعتبات الفئات، يسمح وضع الدرجات متعددة الاستجابة بتقدير عدد أكبر من المعلمات. عادةً ما يحتوي الاختبار ذات المفردات المختلطة على نسبة كبيرة من المفردات ذات الدرجات ثنائية الاستجابة ونسبة صغيرة من المفردات ذات الدرجات متعددة الاستجابة. من المهم إجراء مزيد من البحوث حول مشكلات أشكال الاختبار التي تقدر قدرة الممتحن على أفضل وجه (Berger, 1998) مع تزايد شعبية الاختبارات ذات المفردات المختلطة، من المهم إجراء بحوث في مجال كيفية وضع الدرجات المجّمعة.

**النماذج الرياضية:** إن التحليل الإحصائي للاختبارات ذات المفردات المختلطة معقد. تم إجراء التحليل من خلال استخدام العديد من الإجراءات الحالية، ومع ذلك، فإن معظمها ينطوي على عملية فصل مجموعات أنواع المفردات ذات الدرجات المتشابهة، وتحليلها ثم إنشاء طريقة لدمج النتائج. تعتمد الاستنتاجات المتعلقة بالنماذج بشكل كبير على الأساليب المطبقة في التحليل والطريقة الذاتية في ترجيح النتائج والجمع بينها. لم يتم إجراء مقارنة واضحة بين هذه الأساليب في الاختبارات ذات المفردات المختلطة. لم يتم تأكيد طريقة لتحليل مجموعة كاملة من أنواع المفردات

(E-mail of the correspondig: jcps@art.bsu.edu.eg)

المتنوعة على أنها متفوقة إلى الأن، وتعتبر معلمات المفردة من القضايا الحاسمة التي يجب تقييمها(Rezapour et al.,2021) . من المهم ملاحظة أن كمية المعلومات المتاحة هي الأكبر بالنسبة للمفردات التي يتم وضع درجات لها على مقياس مستمر، على الرغم من أن البحث المتعلق بالعدد الأمثل لفئات الاستجابة للمفردات متعددة الاستجابة لم يكن حاسماً. تعد مساهمات معلومات الاختبار الشاملة للمفردات متعددة الاستجابة وثنائية الاستجابة، بالنسبة لبعضها البعض، قضية يجب البحث فيها ودراستها (Carlson, 1996).

**أساليب التحسين:**

قد يؤدي الجمع بين نوعين من المفردات في اختبار واحد إلى الاستفادة من المزايا التي يقدمها كل منهما، طالما أن أساليب التحليل المطبقة تستوعب الجمع بينهما. ركزت معظم الدراسات حول تحسين تصميم الاختبار وإجراءات التحليل التي أجريت في الماضي على الاختبارات ذات المفردات المماثلة. تعتبر افتراضات النموذج الإحصائي ومعايير التحسين لهذا النموذج هما المطلبان الرئيسيان لتصميم الاختبار الأمثل (Berger ,1998.) إذا تم تطبيق طرق التحسين نحو وضع إرشادات صحيحة لاختيار المفردات، فيمكن إجراء اختبارات تقلل التباين في كل من معلمات المفردة وتقديرات القدرة (Berger ,1998). كما تعد نظريات التحسين وأشكال المفردات المختلطة كلاهما جديدًا إلى حد ما وبالتالي لم يتم الجمع بينهما على نطاق واسع.

**تعريف المصطلحات:**

**١. ثنائي الاستجابة Dichotomous:**

مفردات الاختبار التي يتم تصحيحها باستخدام طريقة التصحيح الثنائية: إما صحيح (١) أو غير صحيح (٠).

**٢.متعددة الاستجابة polychotomous:**

مفردات الاختبار التي يتم تصحيحها على مقياس هرمي للصواب، مثل مع معيار تقييم الدرجات، حيث يمكن الحكم على كل إجابة على أنها أكثر صحة نسبيًا أو أقل صحة على النحو الذي تحدده مجموعة من المعايير والمحتوى المحدد. يتم بعد ذلك تخصيص درجة مقياس، على سبيل المثال **"٣" على مقياس يتراوح من ١ إلى ٤. (Hambleton, Swaminathan and Rogers, 1991).**

**٣. الاختبارات ذات المفردات المختلطة Mixed-item format test**:

أداة أو اختبار للتقييم يحتوي على مزيج من نوعين من المفردات التي تتطلب طرق تصحيح مختلفة، على سبيل المثال، مفردات الاختيار من متعدد (ثنائي الاستجابة) والمفردات التي تتطلب طريقة تصحيح جزئية (متعددة الاستجابة) Swaminathan، (Rogers & Hambleton ,1991)

**٤. التصميم الأمثل للاختبار Optimal Test Design** :

توفر كل من افتراضات النموذج الإحصائي ومعايير التحسين لهذا النموذج هما المتطلبان الرئيسيان للتصميم الأمثل للاختبار (Berger ,1998)

**٥. معايير التحسين للاختبارات ذات المفردات المختلطة Optimality Criteria for the Mixed-Item Format Test:**

في الاختبار المركب، يتم تحقيق ذلك عندما:

(أ) يتوفر ثبات عالي للدرجات التي ينتجها كل اختبار فرعي (تتعلق عادةً بطول الاختبار).

(ب) الثبات الناتج عن الدرجات من الاختبار المركب يتجاوز الثبات الناتج عن أي من الاختبارين الفرعيين وحده.

(ج) إمكانية تطبيق الاختبار المركب (Downing,2002).

**الإطار النظري:**

**أشكال الاختبار وإجراءات التصميم:**

القدرات المعرفية، التي يشار إليها عادة باسم القدرات "الكامنة" أو "السمات"، لا يمكن قياسها مباشرة، لذلك، تم تصميم الاختبارات وطرق التحليل للمساعدة في تقدير القدرات التي يمتلكها الممتحنون. تعتمد دقة هذه التقديرات على تصميم وبناء الاختبار والتحليل والتفسير. تتضمن ممارسات الاختبار الشائعة إدارة الاختبارات الكتابية المكونة من أنواع مختلفة من العناصر بما في ذلك الاختيار من متعدد والمقال والمطابقة والاستجابة المفتوحة والإجابة القصيرة. كما تزداد شعبية تقييمات الأداء التي تتطلب عادةً من الأفراد تنفيذ مهام محددة. بدأ البعض في النظر في هذه التقييمات

(E-mail of the correspondig: jcps@art.bsu.edu.eg)

لتوفير مقاييس "أصلية" للقدرة نظرًا لشكلها. يعتبر هذا النهج في التقييم جديدًا إلى حد ما ويحتاج إلى عمالة كثيفة إلى حد ما لإدارته، وبالتالي فهو ليس شائعًا في الممارسة (Kline,2005).

يمكن تحديد جودة أداة التقييم بعدة طرق. يعد البناء والتصميم الدقيق لأداة الاختبار نفسها أمرًا بالغ الأهمية لتحديد درجة الثقة في النتيجة وصلاحيتها. تصميم أداة التقييم هو إجراء متعدد الخطوات يتكون من أربع خطوات رئيسية كما حددها رايت، ستون (Wright and Stone,1979). وهي كالتالي: تحديد المتغير المراد قياسه بالاختبار بوضوح؛ بناء المفردات التي تقيس هذا المتغير؛ إثبات أن المفردات هي مقياس دقيق للمتغير؛ وتقييم أنماط الاستجابة للاتساق المتوقع. هذه العملية عبارة عن عملية دائرية، فبمجرد تقييم أنماط الاستجابة، يمكن استخدام النتائج وذلك لمراجعة وتحسين المفردات وتصميم الاختبار للإدارات المستقبلية.

تتمثل الخطوة الأولى في تصميم أداة التقييم في التحديد الواضح للمتغير المراد قياسه بواسطة الاختبار. يعد تحديد المتغير ضروريًا لبناء مفردات ذات جودة تعكس السمة المرغوبة عند الإجابة بشكل صحيح. هذا يساعد على القضاء على الغموض في تصميم الاختبار. يعتمد ثبات وصدق الدرجات الناتجة على التحديد الواضح للمتغير المراد قياسه. من المهم أن يتم تصميم المفردات لقياس السمة المستهدفة بدقة. وهذا ما يسمى "أحادية البعد"، وهو مهم بشكل خاص عندما يتم دمج أشكال المفردات المختلفة في تقييم واحد. يجب معالجة فروق الأبعاد بين المفردات متعددة الاستجابة وثنائية الاستجابة قبل إجراء أي تحليلات (Carlson, 1996).

الخطوة الثانية، بناء المفردة، بهدف كتابة المفردة التي تقيس المتغير المحدد. عندما تكون مفردة الاختبار مبنية جيدًا ووجود مقاييس صادقة للسمة المستهدفة، فإن درجات الاختبار وأنماط استجابة الممتحن توفر دليلًا واضحًا فيما يتعلق بتقديرات مستوى السمات. تؤثر العديد من العوامل على جودة الاختبار ويحتاج الأشخاص الذين يصممون الاختبارات أو يتخذون قرارات تستند إلى درجات الاختبار إلى فهم واضح لهذه العوامل. بحيث تنتج الاختبارات درجات تعتبر تقديرات لسمات صادقة وأن يتم تفسير هذه الدرجات بعناية (Linn, 1990).

الخطوة الثالثة في تصميم الاختبار: إثبات أن المفردات التي تم إنشاؤها حديثًا هي مقياس دقيق للمتغير المستهدف. إذا تم تحديد المتغير المستهدف جيدًا وتم كتابة المفردات بعناية، فيجب إجراء اختبار ميداني مناسب للمفردات بحيث يمكن تحليل

الاستجابات. بمجرد الموافقة على مجموعة من المفردات، يتم دمجها بعد ذلك لإجراء اختبار. من أجل تأكيد صدق التقييم، يتم إجراء التحقيقات في المحتوى وبناء الصدق. تساعد التحليلات الارتباطية مع الدرجات من التقييمات الأخرى التي تقيس نفس المتغير المستهدف في دعم الصدق التقاربى لدرجات الاختبار (Marnat & Wright,2016).

تتضمن الخطوة الرابعة في تصميم الاختبار إجراء تقييم سليم لأنماط الاستجابة لتحديد أي تناقضات وحلها ودعم ثبات درجات الاختبار، على سبيل المثال، يجب تصميم المفردات في الاختبار بحيث تكون أحادية البعد (أي أنها تقيس بناء واحد أو سمة واحدة). التحليل العاملي مفيد في تقييم مجموعة من مفردات الاختبار للأبعاد الأحادية. بالإضافة إلى ذلك، يمكن تقييم المفردات الفردية لمقدار المعلومات التي تجمعها. معلومات المفردات قابلة للقياس ويمكن تقييم كل مفردة لمساهمتها الفردية في إجمالي معلومات الاختبار. يعتبر الاعداد الجيد المفردة ومعلومات الاختبار أمرًا مهمًا لإنشاء اختبار مصمم على النحو الأمثل(Kline, 2005) ).

**نماذج التصحيح ثنائي الاستجابة:**

يعد تطوير مفردات الاختيار من متعدد المخصصة للتقييم ثنائي الاستجابة مدخلا شائعاً في القياس. التصحيح ثنائي الاستجابة له مزايا وعيوب. وتتمثل إحدى المزايا في أن وقت الاستجابة المطلوب لتطبيق الاختبار يكون ضئيلًا نسبيًا لكل مفردة بحيث يكون المطورون قادرين على التخطيط لتطبيق اختبارات أطول. يزيد طول الاختبار من ثبات ودقة تقديرات القدرة عن طريق تقليل مقدار الخطأ في القياس. في الثلاثينيات من القرن الماضي، أصبح من الممكن إكمال التصحيح الموضوعي لاستجابات الاختيار من متعدد في جزء بسيط من الوقت والتكلفة اللازمتين للتصحيح اليدوي. زاد هذا التقدم التكنولوجي من جاذبية طريقة الاختبار والتصحيح. انخفضت التكاليف المرتبطة بالحصول على هذه التكنولوجيا باستمرار وأصبح هذا النهج متاحًا الآن لمعظم معلمي الفصول الدراسية (صلاح الدين محمود علام، ٢٠١٨).

توجد مساوئ لهذا النهج التجريبي على الرغم من شعبيته المتزايدة. لا يعتبر استخدام مفردات الاختيار من متعدد طريقة تقييم "أصيلة" من قبل البعض لأنه يؤكد فقط قدرة الممتحن على استدعاء المعلومات أو التعرف عليها بدلاً من إظهار المعرفة والقدرة. عيب آخر هو أن التخمين من قبل الممتحنين أمر لا مفر منه ويمكن أن يضيف متغير إزعاج كبير للتحليل وتقديرات القدرة (صلاح الدين محمود علام ،٢٠٠٥).

(E-mail of the correspondig: jcps@art.bsu.edu.eg)

يعتمد تصحيح المفردات ثنائية الاستجابة على افتراض أن الشخص إما يعرف كل شيء أو لا يعرف شيئًا عن محتوى كل مفردة. لذلك، عندما يتم دمج المفردات المكتوبة بدرجات متفاوتة من الصعوبة لبناء اختبار، يتم استخدام عدد المفردات التي تمت الإجابة عليها بشكل صحيح لتحديد درجة الشخص. يكمن الخطر المرتبط بكل الدرجات ثنائية الاستجابة في أن الاستجابة الصحيحة لجميع المفردات في اختبار سهل للغاية يمكن أن تعني شيئًا مختلفًا تمامًا عن الاستجابة بشكل صحيح لـ ٥٠٪ من المفردات في اختبار صعب للغاية. بشكل عام، تتجاوز نظرية الاستجابة للمفردة النسبة المئوية البسيطة للحساب الصحيح وتحلل الاستجابات والصعوبة النسبية لكل عنصر. ثم يتم استخدام هذه المعلومات لتحديد القدرة التقريبية لكل ممتحن (Kaplan,2012).

**نماذج التصحيح متعدد الاستجابة:**

أوضحت تحليلات الاستجابات غير الصحيحة أن تحليل الاستجابات غير الصحيحة المختلفة قد أدى إلى معلومات مفيدة والتصحيح ثنائي الاستجابة يجعل هذه المعلومات غير قابلة للوصول من خلال تجميع جميع الاستجابات غير الصحيحة في فئة واحدة. يمكن تصحيح المفردات ذات فئات الاستجابة متعددة الاستجابة بشكل متعدد. يوفر تصحيح المفردات متعددة الاستجابة مقياسًا لقياس المعرفة الجزئية لكل مفردة. يقترح وانج ولي (Wang and Lee ,1998) أن أفضل قياس لاكتساب المعرفة غير المكتملة هو بنية هرمية مفترضة توفرها استجابات للمفردات متعددة الاستجابة.

كانت البيانات المنشورة غير متسقة وغير حاسمة فيما يتعلق بتفوق أساليب نظرية الاستجابة للمفردة متعددة الاستجابة على أساليب نظرية الاستجابة للمفردة ثنائية الاستجابة. توجد علاقة مباشرة بين وعي الباحثين بعملية اكتساب المعرفة والحاجة إلى التقييم لتعكس القدرة بشكل أكثر دقة. للحصول على معلومات على مستوى المفردة حول قدرة الشخص على عكس النطاق المستمر لاكتساب المعرفة، يجب على كل مفردة قياس القدرة على مقياس مستمر. وتتراوح هذه المقاييس عمومًا بين القيم المطلقة للخطأ والصحيح من خلال استخدام "معيار التقييم" على سبيل المثال، وتتراوح في النقاط الممنوحة من "١" إلى "٤". تحدد نماذج وضع الدرجات مستويات الكفاءة أو القدرة بناءً على المعايير التي تم الوفاء بها على مستويات مختلفة من المعرفة أو المهارة المثبتة. في هذه الحالة، قد تمثل الدرجة "١" دليلًا محدودًا على القدرة على الكتابة بينما تمثل الدرجة "٤" قدرة كبيرة في الكتابة. يؤكد توثيق

المعرفة الجزئية أن المقاييس الدقيقة للمعرفة أو القدرة لم يعد من الممكن حصرها في استخدام المفردات ثنائية الاستجابة التقليدية(von Davier et al.,2019) .

وصف كارلسون (Carlson,1996) المفردات التي يتم تصحيحها بشكل متعدد الاستجابة بأنها فئات مرتبة متعددة تفترض وجود بنية هرمية في فئة الصعوبات. قد يُظهر الممتحن معرفة جزئية بكل مفردة يتم تطبيقها، مما يسمح بجمع المزيد من المعلومات أكثر من المفردات ثنائية الاستجابة. يمكن تسليم مقياس القدرة، الذي يتطلب أربع مفردات ثنائية الاستجابة، من خلال مفردة واحدة متعددة الاستجابة بأربع فئات هرمية. تسمح هذه العملية بجمع كمية متساوية من البيانات من خلال استخدام عدد أقل من المفردات. لقد تبين أن الاختبارات المكونة بالكامل من المفردات التي يتم تصحيحها بشكل متعدد الاستجابة تجمع ما بين ضعفين وثلاثة أضعاف كمية المعلومات مثل الاختبارات التي تحتوي على مفردات ثنائية الاستجابة ( Donoghue, 1994).

**الجمع بين أنواع المفردات:**

تم ذكر الاختبار ذات المفردات المختلطة كثيراً في الأبحاث الحديثة، وتحديداً أن المفردات ثنائية الاستجابة ومتعددة الاستجابة تكمل بعضها البعض عند دمجها في تقييم واحد مما يؤدي إلى تحسين درجة الثبات والصدق وتقليل التكلفة ( Ericikan, Schwarz, Julian, Burket, Weber and Link, 1998) تكمن الميزة الرئيسية لاستخدام نهج الاختبار ذات المفردات المختلطة لاختبار التصميم في فرصة جمع معلومات أكثر اتساقًا على جميع مستويات القدرة مع تحسين طول الاختبار. لقد وثقت العديد من الدراسات مزايا وعيوب كلا النوعين من المفردات، وقد يؤدي دمجها في اختبار واحد إلى التغلب على بعض العيوب الموثقة. درس وينر وتيسين Wainer (and Thissen,1993) مساهمة المعلومات النسبية وطول الاختبار ووقت التطبيق للمفردات متعددة الاستجابة وثنائية الاستجابة، حيث اعتبرا المفردات متعددة الاستجابة على أنها "غير فعالة" استنادًا إلى الوقت المستثمر في بناء المفردة، وتصحيح المفردات، والتطبيق والتكلفة مع وجود معلومات أقل في الدقيقة للتطبيق من المفردات ثنائية الاستجابة. ربما ساهمت نقطتان في بحثهم في هذا الاستنتاج: لقد قاما بتحليل بيانات استجابة الاختبار الميداني وانتهكوا افتراض الاستقلال المحلي مما سمح بمشاكل متعددة الأبعاد. بشكل عام، تعد مساهمات معلومات الاختبار الخاصة بالمفردات متعددة الاستجابة وثنائية الاستجابة بالنسبة لبعضها البعض مشكلة يجب

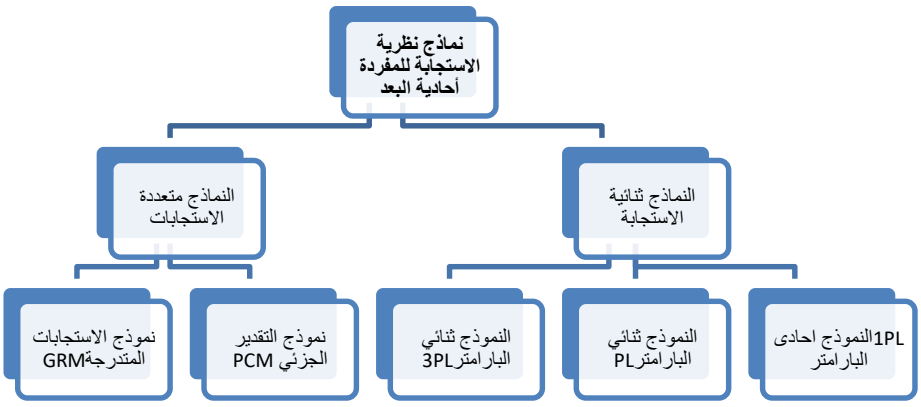
(E-mail of the correspondig: jcps@art.bsu.edu.eg)

البحث فيها بشكل أكبر (Carlson, 1996). جمع البيانات وتصحيحها بشكل ثنائي الاستجابة هما طريقتان تقللان من كمية المعلومات المتاحة لتقدير قدرة الممتحن. توفر مفردات الاختبار التي تجمع المزيد من البيانات حول المعرفة والقدرة، والتي يتم تصحيحها على مقياس متعدد الاستجابة، معلومات أكثر دقة حول قدرة الممتحنين.

قد يؤدي تصحيح "الكل أو لا شيء" لمفردات الاختيار من متعدد إلى تحريف المعرفة الحقيقية أو قدرة الممتحنين. هذه الاستجابات التي تم تصحيحها في شكل ثنائي الاستجابة على أنها صحيحة بنسبة ١٠٠٪ أو ١٠٠٪ غير صحيحة تتجاهل نظرية اكتساب المعرفة الجزئية (Wang & Lee, 1998). تحد بيانات مفردات الاختبار التي لا يمكن تصحيحها وتحليلها إلا على المستوى ثنائي الاستجابة من المعلومات التي يمكن الحصول عليها من خلال التحليل. يتم اكتساب المعرفة أو القدرة الجزئية على طريق المعرفة الكاملة. لذلك، يجب أن يتم اكتساب المعرفة على نطاق مستمر وتراكمي كما تم قياسه بواسطة نماذج التصحيح متعددة الاستجابة. تمتلك المفردات متعددة الاستجابة فئات متعددة مرتبة تفترض بنية هرمية في فئة الصعوبات (Carlson, 1996).

**نماذج نظرية الاستجابة للمفردة أحادية البعد:**

يوضح الشكل التالي نماذج نظرية الاستجابة للمفردة أحادية البعد كما يلي:




**نماذج نظرية الاستجابة للمفردة أحادية البعد لصلاح علام (٢٠٠٥: ٦٨)**

**نماذج نظرية الاستجابة للمفردة ثنائية الاستجابة لصيغة التصحيح:**

هناك العديد من نماذج التحليل المختلفة المصنفة في نطاق نظرية الاستجابة للمفردة. تتضمن كل هذه النماذج مقياسًا لقدرة الشخص (ثيتا) والاهتمام بمعلمات المفردات الأخرى المتعلقة بالقياس. تُستخدم النماذج اللوجستية ذات المعلمات الواحدة والثنائية والثلاثية (1PL, 2PL, 3PL) بشكل شائع لتحليل استجابات المفردات ثنائية الاستجابة في إطار نظرية الاستجابة للمفردة. تمت تسمية هذه النماذج على اسم عدد المعلمات التي لها حرية التغيير في التحليل. تتضمن هذه المعلمات مؤشر التمييز (أ) ومعامل صعوبة المفردة (ب) ومعامل التخمين (ج) (محمد طالب دبوس، ٢٠١٦).

يعد نموذج راش Rasch من ضمن نماذج نظرية الاستجابة للمفردة، تم تصميمه بواسطة جورج راش (Georg Rasch,1960) لتحليل الاستجابة للمفردات ثنائية الاستجابة عن طريق تقدير قدرة الشخص وصعوبة المفردة بشكل منفصل. اشتق جورج راش هذا النهج لتحليل المفردات لغرض نمذجة سلوك الاختبار على وجه التحديد على مستوى المفردة ولتحليل الاستجابات المسجلة ثنائية الاستجابة. إن استخدام "الإحصائيات الكافية" في حساب تقديرات المفردة والأفراد يلغي الترابط بينهما. كما يطبق نموذج راش أسلوب لوجستي لتقدير معلمات المفردة وقدرات الشخص في قياسات لوغاريتمية نسبية. يتيح ذلك مقارنة قدرة الشخص وصعوبة المفردة على مقياس مشترك. كما يتضمن نموذج راش Rasch مقاييس لقدرة الشخص وصعوبة المفردة، مع الاحتفاظ بمعلمات المفردة الأخرى التمييز (١) على طول الاختبار والتخمين قيمة ثابتة وهي (٠) عبر جميع المفردات وهذا ما يجعله مختلف عن النموذج احادي البارامترPL-1 حيث يعطى الأخير قيمة تقريبية لمعلم التمييز يضعها البرنامج المستخدم في التحليل لجميع المفردات وكذلك يثبت معلم التخمين لجميع المفردات. تعتبر الافتراضات التي يتطلبها راش Rasch صارمة ويصعب الوفاء بها، لذا يُعتقد أن مجموعات البيانات القليلة تلبي بشكل كاف افتراضات التحليل(Kaplan,2012) . يحتوي نموذج Rasch على ثلاث صفات تجعله أكثر جاذبية في الاستخدام من بعض النماذج الأخرى: سهولة الاستخدام بسبب عدد أقل من المعلمات، ومشكلات تقدير أقل بسبب عدد المعلمات الأقل، والموضوعية المحددة فيما يتعلق بتقدير المفردة ومعلمات القدرة، والذي كان سبب إنشائه (van der Linden,2016).

هناك العديد من الافتراضات المتعلقة بنماذج نظرية الاستجابة للمفردة التي يجب التحقق منها في البيانات قبل تطبيق الإجراءات. يمكن أن يساعد المستوى الذي يتم تلبية هذه الافتراضات من خلال البيانات في تحديد النموذج الأنسب وإلى أي درجة

(E-mail of the correspondig: jcps@art.bsu.edu.eg)

من الدقة يمكن تفسير النتائج. وهذه الافتراضات هي (أحادية البعد والاستقلال الموضعي والسرعة ومنحنى خصائص المفردات (محمد محمد فتح الله ،٢٠١١).

**الافتراض الأول أحادية البعد :Unidimentionality**

يقصد بأحادية البعد أن تكون جميع المفردات الاختبارية متجانسة أي تقيس خاصية واحدة فقط وأن هناك قدرة كامنة وحيدة هي التي تفسر أداء الفرد في الاختبار , ويعد افتراض أحادية البعد أقوى افتراضات النظرية , فمن الصعب وجود قدرة واحدة تقيسها جميع مفردات اختبار ما بسبب وجود عوامل كثيرة مؤثرة في أداء الفرد على الاختبار كالتخمين الإجابة الصحيحة وقلق الاختبار والقدرة على العمل بسرعة والدافع ، حيث أن هذه العوامل وغيرها تؤثر في استجابة الفرد على الاختبار بالإضافة إلى القدرة الأساسية التي اعد الاختبار لقياسها ولكن بسبب توفر القدرة المسيطرة لدى جميع الأفراد والتي يعزى إليها الأداء بشكل عام وقلة تأثير العوامل الأخرى وعدم ثبات تأثيرها فإن القياس يرد إلى هذه القدرة الأساسية بينما تعد العوامل الأخرى أخطاء في تحليل نماذج الاستجابة للمفردة (Sinharay&Holland ,2006,429 ) **الافتراض الثاني الافتراض الثاني الاستقلال الموضوعي Local independence:**

يعد افتراض الاستقلال الموضعي من الافتراضات الأساسية في كلا من النظرية الكلاسيكية للاختبارات CTTونظرية الاستجابة للمفردة IRT حيث تفترض النظرية الكلاسيكية أن أخطاء القياس مستقلة إحصائيا بين الفقرات المختلفة وبذلك تكون مستقلة في الاختبار ككل وكذلك في نظرية الاستجابة للمفردة يجب أن تكون الاستجابات لأي فقرتين مختلفتين مستقلين إحصائيا عن بعضهما للمفحوصين عند مستوى القدرة، أي أن استجابة المفحوص لمفردة ما لا تتأثر إيجابا أو سلبا باستجاباته على مفردة أخرى ((Lee ,2012

فيعبر الاستقلال الموضعي عن احتمال الإجابة الصحيحة للفرد على مفردة الاختبار يكون مستقلا عن أجابته على أي مفردة أخرى في الاختبار، حيث أن نماذج منحنى خصائص المفردة Item characteristic curveاللوغاريتمية والمنحنى التراكمي الطبيعي يؤكد أن احتمال النجاح على المفردة يعتمد على بارامترات المفردة وقدرة المفحوص ولا شيء أخر (Mandal, 2002 , 36).

كما يقصد بهذا الافتراض أن استجابات الفرد للبنود في الاختبار مستقلة أي لا تؤثر استجابة الفرد لأحدى المفردات على استجاباته للمفردات الأخرى ويتضح هذا في:

تحرر القياس من توزيع العينة المستخدمة sample free وهذا يعنى ثبات تقدير كل من قدرة الفرد وصعوبة المفردة واستقرارهما بالرغم من اختلاف عينة الأفراد المستخدم في الاستجابة طالما أنها عينة ملائمة (Anastsi & warena , 2015,223).

**الافتراض الثالث منحنى خاصية المفردة: Item Characteristic Curve (ICC):**

يعرف منحنى خاصية المفردة بأنه دالة رياضية تربط بين احتمال إجابة المفحوص على الفقرة إجابة صحيحة وبين السمة أو القدرة التي تقيسها المفردة أو يقيسها الاختبار الذي يحتوي على هذه الفقرة وتسمى هذه الدالة بمنحنى خاصية المفردة والتي تمثل بانحدار غير خطى لاحتمال الإجابة الصحيحة على المفردة بدلالة القدرة التي يتم قياسها بالاختبار ويعبر عنها بالصيغة التالية:

$$Pi\ (\theta) = F\ (\theta - bi)$$

(Θ) هي قدرة الفرد, ( b) معلم صعوبة الفقرة (i ). (أمينة كاظم ,١٩٨٨, ٤٤٠).

وتفترض نماذج الاستجابة للمفردة أحادية البعد وجود متصل Continuum للسمة المراد قياسها، وأنه يمكن تقدير احتمال اجابة فرد إجابة صحيحة على مفردة اختبار إذا علمنا موقعة على هذا المتصل، ويعبر عن ذلك رياضيا أن الاختبار يعد بمثابة راسم فوقي Mapping من عينة الأفراد المختبرين إلى متصل السمة الكامنة واحتمال الإجابة الصحيحة يكون دالة تزايدية لمواقع الأفراد على متصل السمة وبذلك يزداد احتمال توصل الفرد إلى الإجابة الصحيحة على مفردة الاختبار بازدياد مقدار السمة لديه.

ويتضح مفهوم المنحنى المميز للمفردة إذا قارنا بين النظريتين التقليدية CTT ونظرية الاستجابة للمفردة IRT فكلاهما يعتمد على فرضية مؤداها أن موقع الفرد على متصل سمة كامنة معينة يحكم احتمال أجابته إجابة صحيحة عن مفردة اختبار يقيس هذه السمة ولكنهما يختلفان في كيفية تحديد هذا الموقع. وارتباطه باحتمال الإجابة الصحيحة على المفردة أي أن الفرق بينهما يتعلق بشكل وخصائص الدالة التي تحدد هذا القيد فالنظرية الكلاسيكية تفترض أن منحنى الدالة المميزة للمفردة لا يختلف باختلاف مفردات الاختبار وتمثل هذه الدالة بالمنحنى الاعتدالي التراكمي

(E-mail of the correspondig: jcps@art.bsu.edu.eg)

لمتغير عشـوائي ويفترض في هذا المنحنى أن احتمال الإجابة الصـحيحة عن مفردة اختبارية دالة تزايدية مطردة لموقع الفرد على متصل السمة الكامنة (صلاح علام , ٢٠١١ , ٦٩٧).

**الافتراض الرابع: التحرر من عامل السرعة في الإجابة Speedness:**

أن نماذج الاختبارات التي تكون ملائمة لنماذج الاستجابة للمفردة لا يتم إعدادها تحت شرط السرعة أي أن إخفاق بعض الأفراد في الإجابة على بعض مفردات الاختبار يرجع إلى تجانس قدراتهم وليس إلى عامل السرعة في إجاباتهم .وهذا الافتراض لا يعلن عنه كأحد افتراضات نماذج الاستجابة للمفردة في معظم الأحوال , حيث أنه متضمن في الافتراض الخاص بأحادية البعد وذلك لأنه لو اعتبرت السرعة عاملا مؤثرا في أداء الفرد على الاختبار ففي هذه الحالة يكون هناك سمتين تؤثران في الأداء وهما سرعة الأداء والسمة المقاسة من خلال محتوى الاختبار.

ويتم التحقق من هذا الافتراض عن طريق مقارنة تباين الفقرات المحذوفة مع تباين الفقرات التي تمت الإجابة عنها خطأ، أو من خلال مقارنة علامات الاختبار عند تحديد الوقت مع علامات نفس الاختبار عند ترك الوقت دون تحديد، أو من خلال التحقق النسبة المئوية للمفحوصين الذين أجابوا عن جميع الفقرات والتحقق من النسبة المئوية للمفحوصين الذين أكملوا ٧٥ % من الاختبار، أو عدد الفقرات التي أجاب عنها ٨٠% من المفحوصين (أمينة كاظم وآخرون,١٩٩٦ ; صلاح علام , ٢٠٠٥ ; مايا بركات ,٢٠١٠; ياسين الشواورة , ٢٠١٣ , آمنه إبراهيم أبو حجر, ٢٠١١).

**نماذج نظرية الاستجابة للمفردة متعددة الاستجابة لصيغة التصحيح:**

**(١) نموذج التقدير الجزئي Partial Credit Model (PCM):**

طور ماسترز (Masters,1982) ويعد هذا النموذج تعميم لنموذج راش للاستجابات الثنائية ليتناسب مع المفردات التي تتطلب عدد من الخطوات للاستجابة عليها، حيث يستخدم في تحليل مفردات الاختبارات، التي تشمل المفردات الاختبارية التي تتطلب مجموعة من الخطوات مثل المسائل الحسابية أو أسئلة المقال، وفيها لا تكون الاستجابة صحيحة أو خاطئة، إنما تقدر درجات للاستجابة الجزئية الصحيحة والذي يعبر عنه بالعلاقة الرياضية الاتية:

$$P_{ix}(\theta) = \frac{e^{\sum_{j=0}^{x}(\theta-\delta_{ij})}}{\sum_{i=0}^{m}(e^{\sum_{j=0}^{x}(\theta-\delta_{ij})})}$$

**٢) نموذج الاستجابة المتدرجة Graded Response Model (GRM):**

يعتبر هذا النموذج امتداد للنموذج ثنائي المعلم والذي طوره ساميجاما (Samejima,1969). ويعد هذا النموذج ملائما لاستجابات المفردات مرتبة الفئات، ويستخدم في التنبؤ باستجابة الفرد المحتملة لفئة معينة والتي تتسق مع مستواه النمائي، حيث يهدف الى تحديد موضع هذه الفئة على متصل السمة الكامنة، ويتم تصنيف استجابات الافراد لكل مفردة وبشكل متدرج، ولحساب مقدار احتمال الاستجابة لكل فئة يتطلب خطوتين

منحنيات عددها يساوي عدد العتبات الفارقة باستخدام النموذج الثنائي كالتالي:

$$P^{*}_{ix}(\theta) = \frac{e^{a_i(\theta-\beta_{ij})}}{1 + e^{a_i(\theta-\beta_{ij})}}$$

حيث j = 1 ,2,.....m

P*ix(θ) تسمى منحنيات مميزة حدية ,boundary response curves ويتطلب هذا النموذج تقدير منحنى مميز حدى لكل عتبة فارقة بين فئات الاستجابات, ويمكن تفسير قيم بارامترات ($\beta ij$) على أنها تمثل مستوى السمة المطلوبة لكى تتخطى الاستجابة العتبة الفارقة jباحتمال قدره ٥٠ . ) , فالمفردة تعالج على أنها سلسلة من الفئات الثنائية ويستخدم النموذج ثنائي البارامتر في تقدير كل هذه الفئات الثنائية , تحت شرط أن ميل المنحنيات الحدية متساوي داخل كل مفردة , وبعد الحصول على هذه التقديرات يحسب الاحتمال الفعلي لفئات الاستجابات وفقا للمعادلة التالية :

$$Pix(\theta) = P^{*}ix - P^{*}i(x+1) \qquad x = 1,2,3,4$$

وتسمى بالمنحنيات الاستجابة للفئات Category response curveوهي تعبر عن احتمال استجابة الفرد في فئة معينة مشروطا على مستوى السمة الخاصة به (صلاح علام , ٢٠٠٥ , ٧٥).

تختلف النماذج متعددة الاستجابة عن النماذج ثنائية الاستجابة في أيهما أكثر ملائمة بناءً على استجابات الافراد. ويمكن تصحيح فئات الاستجابة متعددة الاستجابة بشكل

(E-mail of the correspondig: jcps@art.bsu.edu.eg)

هرمي وينتج عنها معلمات يشـار إليها باسـم "عتبات الفئة". لكل مفردة فئة m + 1، هناك معلمات حد m تفصـل بين الفئات *a(k)*. كما هو الحال مع النماذج ثنائية الاستجابة، يمكن تحليل الاستجابة للمفردات لإنتاج تقديرات لقدرة الشخص ومعلمات المفردة. "مع الاهتمام المتزايد بـ ‘‘القياس النفسي "، يجب الاهتمام بنماذج نظرية الاسـتجابة للمفردة التي يمكنها التعامل مع التصـحيح متعدد الاسـتجابة، نظرًا لأن القياس الحقيقي مرتبط باختبار الأداء وتصـحيح أو وضـع درجات لأداء الممتحن" (Hambleton, Swaminathan, & Rogers, 1991, p.153.)

**تحسين الاختبارات ذات المفردات المختلطة:**

**تصميم الاختبار:**

بالنظر إلى التحول النموذجي في التحليل من تقدير مسـتويات السـمات اسـتنادًا إلى مجموع درجات الاختبار إلى الاستجابة للمفردات، فقد ذكر أنه يجب التخلي عن فكرة درجة الاختبار "الثبات" لصالح القياسات الأكثر ملاءمة لمعلومات المفردة والاختبار (Thissen, 1991.) تم وصف معلومات المفردة والاختبار باستخدام معادلات فيشر Fisher’s formulas بواسـطة بيرنبوم Birnbaum الذي بدأ العمل في مجال تحليل المفردات في الخمسـينيات من القرن الماضـي. يعتمد تطبيق وظائف المعلومات في مراحل إنشـاء الاختبار واختيار المفردة على مسـاهمة المفردة في معلومات الاختبار الإجمالية. عند فحص الكميات النسـبية من المعلومات التي تم جمعها بواسـطة أنواع المفردات المختلفة، اقترح بيرنبوم ترجيح مسـاهمات المفردات في مجموع درجات الاختبار بناءً على مسـاهماتها في إجمالي معلومات الاختبار. يمكن تنفيذ نهج الترجيح هذا من خلال اسـتخدام مقياس درجات. طبق بيرنبوم أيضًـا منهجه في الترجيح على النظريات المثلى لتسجيل درجات الاختبار(Nunnally& Bernstein,1994) .

تتضـمن جهود تصـميم الاختبار معالجة العوامل التالية: شـكل أو صـيغة المفردة واختيار المفردة وطول الاختبار وطرق تصـحيح الدرجات وإجراءات تحسـين الاختبار والجداول الزمنية والمخاوف المتعلقة بالميزانية. تشـمل الجهود المبذولة لتحسين تصميم الاختبار كتابة المفردات التي تقيس بوضوح القدرة المستهدفة والتي تجمع الحد الأقصى من المعلومات بكفاءة عبر جميع مستويات القدرة. قد يكون شكل المفردات المختلطة هو التصميم الأمثل للاختبار لأنه قد يستفيد من مزايا وعيوب كلا النوعين من المفردات (Carlson, 1996). يعد تحقيق التوافق الجيد للنموذج المحدد

مع البيانات التي تم جمعها أمرًا بالغ الأهمية أيضًـا إذا كانت درجات نمط الاسـتجابة ستحقق تقديرات القدرة ثيتا المثلى (Thissen, Nelson and Swygert, 2001).

**معلومات المفردة والاختبار:**

وجد كارلسـون (Carlson ,1996) أن المفردات التي يتم تصـحيحها بشـكل متعدد الاسـتجابة تميل إلى توفير قدر أكبر نسـبيًا من المعلومات مقارنة بالمفردات ثنائية الاسـتجابة وقدمت معلومات عبر نطاق أوسـع من تقديرات الكفاءة مقارنة بالمفردات ثنائية الاسـتجابة. ووجد أيضًـا أن المفردات ثنائية الاسـتجابة تبدو وكأنها تنتج منحنى معلومـات يتطـابق بشـكل كبير مع منحنى إتقـان الممتحن أكثر من منحنى المفردات متعددة الاستجابة. بالنظر إلى الوقت المستغرق في كتابة المفردات متعددة الاستجابة والاسـتجابـة لها وتصـحيحها، يبدو أن معلوماتهم المتزايدة تكون مكلفة، لذلك عمم كارلسـون النتائج التي توصـل إليها أن الاختبارات التي تجمع بين أنواع متعددة من المفردات ستستفيد من مزايا كل منها مع تقليل العيوب.

وجد دون وجو (Donoghue,1994) أن المفردات متعددة الاسـتجابة قد تجمع ما بين ٢,١ إلى ٣,١ مرة من المعلومـات مثل المفردات ثنـائيـة الاسـتجـابـة. هذا هو الدعم للتوصـية بترجيح المفردات متعددة الاسـتجابة الأثقل في التقييم. عندما تم تصـحيح المفردات متعددة الاسـتجابة بشـكل ثنائي، انخفض مقدار المعلومات التي تم جمعها بشـكل كبير ولكنها كانت لا تزال أكبر من تلك التي تم إنتاجها باسـتخدام المفردات ثنائية الاسـتجابة بدقة. تسـتغرق المفردات متعددة الاسـتجابة مزيدًا من الوقت والمال لتطبيقها وتصحيحها.

**معايير التقييم للتحسين:**

كنتيجـة للفوائد الموثقة المنسـوبـة إلى كل من المفردات ثنـائية الاسـتجابة ومتعددة الاستجابة، من المهم إعطاء اعتبارات لتصميم الاختبارات التي تحتوي على مجموعة من هذه الأنواع من التصـحيح أو وضـع الدرجـات. يقترب الاختبـار ذو المفردات المختلطة من التحسـين عند اختيار تصـميم الاختبار وطرق التحليل في وقت واحد. يؤثر تصـميم الاختبار وطرق التحليل أيضًـا على الكمية المتاحة من معلومات المفردة والاختبار (Ercikan et al., 1998; Si, 2002). أشـارت دراسـة الأمثلية أو التحسـين Optimality التي أجراهـا بليـك (Plake ,1993) إلى أن أفضـل النتـائج التي يتم

(E-mail of the correspondig: jcps@art.bsu.edu.eg)

الحصـول عليها بواسـطة مفردات الاختبار هي تلك "الأفضـل" التي تتطابق مع قدرة الممتحن. من المتوقع أن تقدر المفردات التي تم تصـــحيحها بشـــكل متعدد القدرات بأعلى مستوى من الدقة.

لاحظ بيرجرBerger (١٩٩٨) أن بحث تصـميم الاختبار الأمثل يعتمد على النموذج الإحصـــائي المفترض المطبق والمعيار الأمثل لهذا النموذج، مثل تلبية الافتراضـــات والبيانات المناسـبة للنموذج. سـتكون النتيجة المرجوة أكثر كفاءة في تقدير السـمات والمعلمـات (أي تبـاين أقل في التقديرات). هنـاك نوعـان من التحـديـات التي لاحظهـا بيرجر في تحقيق التفاؤل. يتعلق التحدي الأول بتصميم الاختبار الأمثل واعتماده على وظيفة معلومات فيشـر. تعتمد هذه الوظيفة على قيم ثيتا، والتي تكون غير معروفة في البداية في أي عملية تصـميم اختبار. يمكن حل هذا الاعتماد باستخدام إجراء متسلسل محوســب ومعقد لتقدير المفردات التي يتم فيها تحديث تقديرات معلمة المفردة في كل خطوة. التحدي الثـاني لتحقيق الأمثل هو تـأثير عدد الفئـات وعدد معلمـات الفئـة على إجراء تصـــميم الاختبار. لم يتم تحديد العدد الأمثل للفئات في الأدبيات. لاحظ بيرجر أيضًـــا أن هنـاك حـاجـة إلى مزيد من البحث فيما يتعلق باختيار المفردة الأمثل لتقدير القدرة الفعال للأفراد.

**اجراءات الدراسة:**

**عينة الدراسة:**

تكونـت عينـة الـدراســـة من ٣٤٠ طـالبا وطالبة بالفرقة الثالثة من كلية الآداب جامعة الفيوم.

**أداة الدراسة:**

من أجل تحقيق هدف الـدراســـة، تم بنـاء اختبـار، مكون من ٣٠ مفردة، يطلق عليـه الاختبـار ذي المفردات المختلطـة، حيث يتكون من ٧٠٪ ثنـائي الاســـتجابـة و٣٠٪ متعدد الاستجابة.

**تحليل الاشكال المختلطة للمفردات:**

تم تحليل جميع الاستجابات للمفردات من أجل أحادية البعد في برنامج SPSS باستخدام إجراء التحليل العاملي بطريقة الارجحية القصوى. تم الوصول الى بناء عامل واحد لكل مجموعة من المفردات. لتحليل الاستجابات للمفردات ثنائية الاستجابة، تم استخدام نموذج راش Rasch من خلال تطبيق النموذج اللوجستي ذو المعلمة الواحدة. ولتحليل الاستجابات للمفردات متعددة الاستجابة في المدخل القائم على نموذج راش Rasch، تم تطبيق نموذج التقدير الجزئي.

**الثبات:**

تم حساب الثبات باستخدام معامل ألفا كرونباخ للاتساق الداخلي Cronbach's α، وجاءت قيمة معامل الثبات($\alpha = 0.82$) .

**الصدق:**

تم حساب صدق المحتوى للمقياس من قبل مجموعة من ١٠ خبراء، قيّموا مدى ملاءمة كل مفردة باستخدام مقياس Likert المكون من أربع نقاط (حيث يمثل ١ "غير ذي صلة" و٤ يمثل "ذو صلة كبيرة")، وقدموا اقتراحاتهم وتعليقاتهم. تم الحكم على البنود الـ ٣٠ بأنها وثيقة الصلة نوعا ما أو وثيقة الصلة بشكل كبير. تم حساب مؤشر صدق المحتوى على مستوى المفردة (I-CVI = 0.90).

**التصميم:**

تم استخدام تصميم تحليل التباين الأحادي ANOVA ذي التأثير الثابت ١١ × ٣ × ٢ مع أحد عشر شكلا لتصحيح أو وضع درجات الاختبار (النسب المنظمة للمفردات ثنائية الاستجابة ومتعددة الاستجابة)، وثلاثة أطوال للاختبار (قصير = ١٠ مفردة، متوسط = ٢٠ مفردة وطويلًا = ٣٠ مفردة) ومجموعتين من النماذج الرياضية لنظرية الاستجابة للمفردة. يتكون أول أشكال التصحيح المنظم للاختبار الأحد عشر من ١٠٠٪ تصحيح متعدد الاستجابة للمفردة مع تصحيح ٠٪ ثنائي الاستجابة، متبوعًا بنسبة ٩٠٪ تصحيح متعدد الاستجابة مع ١٠٪ ثنائي الاستجابة، وهكذا حتى الوصول إلى التصحيح الحادي عشر الذي يتكون من ٠٪ تصحيح متعددة الاستجابة و١٠٠٪ ثنائية الاستجابة. تم استخدام مجموعتين من النماذج لتصحيح المفردات المختلطة: المجموعة الأولى (نموذج راش تنائي الاستجابة Raschونموذج التقدير

(E-mail of the correspondig: jcps@art.bsu.edu.eg)

الجزئي متعدد الاستجابات(PCM، والمجموعة الثانية (النموذج ثنائي البارامتر ثنائي الاستجابات 2PLونموذج الاستجابة المتدرجة متعدد الاستجابات (GRM نتج عن التصميم ثلاثي العوامل ٦٦ مجموعة بيانات كما هو موضح في الجدول ١.

**الجدول (١) مجموعات العوامل لتصميم الدراسة**

| | | النسبة المئوية للمفردات ثنائية الاستجابة / متعددة الاستجابة | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| طول الاختبار | نماذج التحليل | 100/0 | 90/10 | 80/20 | 70/30 | 60/40 | 50/50 | 40/60 | 30/70 | 20/80 | 10/90 | 0/100 |
| 10 | المجموعة الاولى | 1001 | 901 | 801 | 701 | 601 | 501 | 401 | 301 | 201 | 101 | 1 |
| | المجموعة الثانية | 1001 | 901 | 801 | 701 | 601 | 501 | 401 | 301 | 201 | 101 | 1 |
| 20 | المجموعة الاولى | 1002 | 902 | 802 | 702 | 602 | 502 | 402 | 302 | 202 | 102 | 2 |
| | المجموعة الثانية | 1002 | 902 | 802 | 702 | 602 | 502 | 402 | 302 | 202 | 102 | 2 |
| 30 | المجموعة الاولى | 1003 | 903 | 803 | 703 | 603 | 503 | 403 | 303 | 203 | 103 | 3 |
| | المجموعة الثانية | 1003 | 903 | 803 | 703 | 603 | 503 | 403 | 303 | 203 | 103 | 3 |

**بناء مجموعات البيانات:**

تم اختيار مجموعات الاستجابة المكونة من عشرة مفردات وعشرين مفردة بشكل عشوائي لبناء اختبارات متوسطة وقصيرة في الطول. استنادًا إلى أطوال الاختبار التي تمت ملاحظتها خلال مراجعة الأدبيات والنظر في طول فترة تطبيق المفردات النموذجية التي تم تصحيحها باستخدام نموذج تقييم متعدد الاستجابة، تم اعتبار الاختبار المكون من ١٠ مفردات قصيرًا، وكان الاختبار المكون من ٢٠ مفردة متوسط الطول وتم اعتبار الاختبار المكون من ٣٠ مفردة طويلاً. تم بعد ذلك تكرار مجموعات الاستجابة ١٠ و٢٠ و٣٠ مفردة ١١ مرة وتم اختيار نسبة محددة مسبقًا من المفردات بشكل عشوائي من كل اختبار وإعادة تصحيحها او وضع درجات لها بشكل ثنائي الاستجابة. تم بعد ذلك تحديد أحد عشر شكلا متناسبًا عبر ثلاثة أطوال للاختبار من خلال النسبة المئوية للمفردات في الاختبار الذي تم تصحيحه بشكل متعدد الاستجابة وثنائي الاستجابة، على التوالي (١٠٠٪ / ٠٪، ٩٠٪ / ١٠٪، ٨٠٪ / ٢٠٪، ٧٠٪ / ٣٠٪، ٦٠ ٪ / ٤٠٪، ٥٠٪، ٥٠٪، ٤٠٪ / ٦٠٪، ٣٠٪ / ٧٠٪، ٢٠٪ / ٨٠٪، ١٠٪ / ٩٠٪ و٠٪ / ١٠٠٪.). لإعادة تصحيح المفردات متعددة الاستجابة التي تراوحت من ١ إلى ٤، استند التقسيم ثنائي الاستجابة إلى معايير أن

الاستجابات في أعلى فئة استجابة متعددة الاستجابة (٤) فقط تم ترميزها بشكل ثنائي على أنها "صحيحة" بينما الاستجابات في جميع الفئات الدنيا (١،٢ و٣) تم ترميزها على أنها "غير صحيحة". لذلك، تم الحصول على كل درجة ثنائية الاستجابة عن طريق إعادة ترميز ١ و٢ و٣ كـ "٠" و٤ كـ "١" (٠، ١). نتج عن هذه المعالجات في البيانات أحد عشر اختبارًا ذي مفردات مختلطة، يتكون كل منها من نسبة محددة مسبقًا من البيانات ثنائية الاستجابة ومتعددة الاستجابة. أسفرت هذه المجموعات من استجابات ذان الاشكال المختلطة عبر ثلاثة أطوال اختبار عن ٣٣ مجموعة بيانات من أشكال الاستجابة التي تم تحليلها باستخدام مجموعتين من النماذج لتصحيح المفردات المختلطة: المجموعة الأولى (نموذج راش للاستجابات الثنائية ونموذج التقدير الجزئي للاستجابات المتعددة)، والمجموعة الثانية (النموذج ثنائي البارامتر للاستجابات الثنائية ونموذج الاستجابة المتدرجة للاستجابات المتعددة).

**اجراءات الدراسة:**

قبل تطبيق الاختبار، تم إخطار المشاركين عن الهدف من الدراسة، وذكروا طوعية انهم يرغبون ويوافقون على الاشتراك في الدراسة. لضمان استجابة المشاركين للمفردات بأمانة، تم إخبارهم بعدم كتابة اسماءهم في ورقة الاختبار. كما تم إخبارهم أنه لا ينبغي أن يهتموا بأي شيء يتعلق بمشاركتهم في الدراسة وأن إجاباتهم تكون لأغراض بحثية فقط وستبقى سرية. تم إدخال جميع البيانات في ملف SPSS.

**نتائج الدراسة:**

افترضت الدراسة الحالية أن طريقة التصحيح (نسبة المفردات ثنائية الاستجابة ومتعددة الاستجابة) وطول الاختبار وطريقة التحليل تؤثر على معايير التحسين في الاختبارات ذات المفردات المختلطة. معايير التحسين، بدورها، لها تأثير ملموس عبر الاختبارات ذات المفردات المختلطة على ما يلي: الثبات، ومعلومات المفردة، ومعلومات الاختبار، وتقديرات القدرة، ومعلمات المفردة. وفيما يلي عرض للأسئلة البحثية على النحو الاتي:

**السؤال الأول: كيف تؤثر نسبة المفردات ثنائية الاستجابة ومتعددة الاستجابة عبر طول الاختبار ونماذج التحليل على تقديرات الكفاءة؟**

تم إجراء أحد عشر انوفا ANOVA ثلاثة X اثنين (واحد لكل مستوى نسبة) على ٦٦ مجموعة من تقديرات ثيتاTheta مع طول الاختبار والنموذج كمتغيرات مستقلة. تم

(E-mail of the correspondig: jcps@art.bsu.edu.eg)

تقييم التأثيرات الرئيسية والتفاعلية للأهمية من خلال مقارنة قيم F المرصودة بقيمة F الحرجة. يتم تحديد قيمة F الحرجة من خلال درجات الحرية لكل تأثير ومستوى ألفا الذي تم تعيينه عند 05. = α. تمت مقارنة كل F تمت المرصودة بـ F الحرجة (٠,٠٥، ١، ∞) = ٣,٨٤ لاختبار أهمية نموذج وF حرجة (٠,٠٥، ٢، ∞) = ٣,٠٠ لاختبار دلالة طول الاختبار والتفاعل بين النموذج وطول الاختبار. تم تقييم حجم التأثير أيضا.

حجم التأثير هو نسبة مجموع المربعات بين المجموعات والمجموع الإجمالي للمربعات، وفي هذه التحليلات تمثل نسبة التباين في تقديرات ثيتا التي يفسرها الفرق بين المجموعات. تم قياس تقديرات حجم التأثير باستخدام قيم التربيعية الجزئية وتم تقييمها بناءً على ما ذكره كوهين Cohen (1988) كمرجع عام لحجم التأثير لاختبارات F في أنوفا ANOVA (صغير = ٠,١، متوسط = ٠,٢٥ وكبير = ٠,٤). تم حساب حسابات القدرة المرصودة أيضًا باستخدام SPSS وتم تضمينها في نتائج انوفا ANOVA المجدولة.

كانت قيم F أنوفا المحسوبة لمستويات النسبة من ٠٪ إلى ٥٠٪ للدرجات متعددة الاستجابة غير دالة لجميع التأثيرات الرئيسية والتفاعلية. تراوحت قيم F هذه من ٠,٠٠٣ إلى ٢,٠٥٤ لجميع التأثيرات. كما كانت قيمة F للأنوفا ANOVA لمستوى نسبة واحدة (٩٠٪ متعدد الاستجابة) دالة للتأثير الرئيسي للنموذج المرصود = ٣,٩٧٧ بحجم تأثير ٠,٠٠١ وقوة تساوي ٠,٥٢٤. أدت ANOVAs التي أجريت على مستويات نسبة ٦٠٪ إلى ١٠٠٪ وضع درجات متعددة الاستجابة جميعها إلى قيم F دالة للتأثير الرئيسي لطول الاختبار. تراوحت قيم F هذه من Fالمرصودة = ٦,٨١٣ إلى F المرصودة = ١٨,٤٥٥ مع بقاء حجم التأثير مساويًا أو أقل ٠,٠٠٦ والقوة المتبقية تساوي أو تزيد عن ٠,٩٢١. لم يؤدى أي من أنوفاANOVAs الاحد عشر الى قيمة F دالة لتأثير تفاعل طول النموذج حسب الاختبار. النتائج معروضة في الجداول 2و3 و٣ و٤ و٥.

**الجدول (2) تحليل التباين ثنائي الاتجاه ثابت التأثيرات على تقديرات ثيتا: النسبة ٦٠٪ (النموذج، طول الاختبار)**

| المصدر | م.ع. | د.ح. | م. ع. | ف | قيمةP | مربع ايتا الجزئى | المعلمة غير الرئيسة | القوة المرصودة (a) |
|---|---|---|---|---|---|---|---|---|
| النموذج المصصح | (b)33.662 | 5 | 6.732 | 7.869 | 0. 001 | 0. 007 | 39.347 | 1 |
| التقاطع | 43.157 | 1 | 43.157 | 50.446 | 0. 001 | 0. 008 | 50.446 | 1 |
| النموذج | 0.503 | 1 | 0.503 | 0.588 | 0.443 | 0. 000 | 0.588 | 0. 120 |
| الطول | 31.577 | 2 | 15.789 | 18.455 | 0. 001 | 0. 006 | 36.91 | 1 |
| النموذجX الطول | 1.582 | 2 | 0.791 | 0.925 | 0.397 | 0. 000 | 1.849 | 0. 211 |
| الخطأ | 5127.947 | 5994 | 0.856 | | | | | |
| الكلى | 5204.766 | 6000 | | | | | | |
| الكلى المصحح | 5161.609 | 5999 | | | | | | |

**a** محســوبة باســتخدام الفا =٠,٠٥ ، b ر التربيعية =٠,٠٠٧، ر التربيعية المعدلة =٠,٠٠٦

**الجدول (3) تحليل التباين ثنائي الاتجاه ثابت التأثيرات على تقديرات ثيتا: النســبة ٧٠٪ (النموذج، طول الاختبار)**

| المصدر | م.ع. | د.ح. | م. ع. | ف | قيمةP | مربع ايتا الجزئى | المعلمة غير الرئيسة | القوة المرصودة (a) |
|---|---|---|---|---|---|---|---|---|
| النموذج المصصح | (b) 14.064 | 5 | 2.813 | 2.832 | 0. 015 | 0. 002 | 14.159 | 0. 843 |
| التقاطع | 1.936 | 1 | 1.936 | 1.95 | 0. 163 | 0. 000 | 1.95 | 0. 287 |
| النموذج | 0.018 | 1 | 0.018 | 0.018 | 0.893 | 0. 000 | 0.018 | 0. 052 |
| الطول | 13.535 | 2 | 6.767 | 6.813 | 0. 001 | 0. 002 | 13.626 | 0. 921 |
| النموذجX الطول | 0. 512 | 2 | 0.256 | 0.258 | 0.773 | 0. 000 | 0.515 | 0. 091 |
| الخطأ | 5953.929 | 5994 | 0.993 | | | | | |
| الكلى | 5969.93 | 6000 | | | | | | |
| الكلى المصحح | 5967.993 | 5999 | | | | | | |

**a** محســوبة باســتخدام الفا =٠,٠٥ ، b ر التربيعية =٠,٠٠٢، ر التربيعية المعدلة =٠,٠٠٢

**الجدول (4) تحليل التباين ثنائي الاتجاه ثابت التأثيرات على تقديرات ثيتا: النســبة ٨٠٪ (النموذج، طول الاختبار)**

| المصدر | م.ع. | د.ح. | م. ع. | ف | قيمةP | مربع ايتا الجزئى | المعلمة غير الرئيسة | القوة المرصودة (a) |
|---|---|---|---|---|---|---|---|---|
| النموذج المصحح | (b) 17.553 | 5 | 3.511 | 3.553 | 0. 003 | 0. 003 | 17.763 | 0. 923 |
| التقاطع | 2.6 | 1 | 2.6 | 2.631 | 0. 105 | 0. 000 | 2.631 | 0. 368 |
| النموذج | 0.026 | 1 | 0.026 | 0.026 | 0.872 | 0. 000 | 0.026 | 0. 053 |
| الطول | 17.465 | 2 | 8.733 | 8.837 | 0. 000 | 0. 003 | 17.675 | 0. 972 |
| النموذجX الطول | 0. 062 | 2 | 0.031 | 0.031 | 0.969 | 0. 000 | 0.063 | 0. 055 |
| الخطأ | 5922.938 | 5994 | 0.988 | | | | | |
| الكلى | 5943.091 | 6000 | | | | | | |
| الكلى المصحح | 5940.491 | 5999 | | | | | | |

**a** محســوبة باســتخدام الفا =٠,٠٥ ، b ر التربيعية =٠,٠٠٣، ر التربيعية المعدلة =٠,٠٠٢

**الجدول (5) تحليل التباين ثنائي الاتجاه ثابت التأثيرات على تقديرات ثيتا: النســبة ٩٠٪ (النموذج، طول الاختبار)**

| المصدر | م.ع | د.ح | م.ع | ف | P قيمة | مربع ايتا الجزئى | المعلمة غير الرئيسة | القوة المرصودة (a) |
|---|---|---|---|---|---|---|---|---|
| النموذج المصحح | (b)26.230 | 5 | 5.246 | 5.877 | 0 | 0.005 | 29.386 | 0.995 |
| التقاطع | 4.66 | 1 | 4.66 | 5.22 | 0.022 | 0.001 | 5.22 | 0.627 |
| النموذج | 3.55 | 1 | 3.55 | 3.977 | 0.046 | 0.001 | 3.977 | 0.514 |
| الطول | 21.055 | 2 | 10.527 | 11.794 | 0 | 0.004 | 23.588 | 0.995 |
| النموذج X * الطول | 1.625 | 2 | 0.813 | 0.91 | 0.402 | 0 | 1.821 | 0.208 |
| الخطأ | 5350.28 | 5994 | 0.893 | | | | | |
| الكلى | 5381.169 | 6000 | | | | | | |
| الكلى المصحح | 5376.51 | 5999 | | | | | | |

**a** محســوبة باســتخدام الفا =٠,٠٥ ، b ر التربيعية =٠,٠٠٥، ر التربيعية المعدلة =٠,٠٠٤

**الجدول (6) تحليل التباين ثنائي الاتجاه ثابت التأثيرات على تقديرات ثيتا: النسبة ١٠٠٪ (النموذج، طول الاختبار)**

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| القوة المرصودة(a) | المعلمة غير الرئيسة | مربع إيتا الجزئى | P | | م.ع | | م.ع | |
| | | | قيمة | ف | | د.ح | | المصدر |
| 0.979 | 23.707 | 0.004 | 0 | 4.741 | 4.8 | 5 | (b)24.000 | |
| | | | | | | | | النموذج المصحح |
| 0.338 | 2.377 | 0 | 0.123 | 2.377 | 2.407 | 1 | 2.407 | التقاطع |
| 0.114 | 0.545 | 0 | 0.46 | 0.545 | 0.552 | 1 | 0.552 | النموذج |
| 0.993 | 22.491 | 0.004 | 0 | 11.245 | 11.385 | 2 | 22.769 | الطول |
| 0.104 | 0.671 | 0 | 0.715 | 0.335 | 0.34 | 2 | 0.679 | النموذج X * الطول |
| | | | | | | | | |
| | | | | | 1.012 | 5994 | 6068.224 | الخطأ |
| | | | | | | 6000 | 6094.631 | الكلى |
| | | | | | | 5999 | 6092.225 | الكلى المصحح |

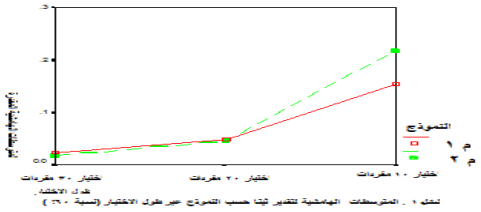
**a** محسوبة باستخدام الفا =٠,٠٥ ، b ر التربيعية =٠,٠٠٤، ر التربيعية المعدلة =٠,٠٠٣

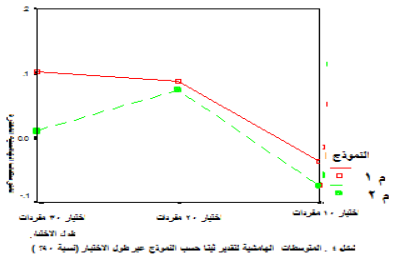
أشارت هذه التحليلات إلى أنه مع زيادة نسبة المفردات متعددة الاستجابة إلى ما بعد ٥٠٪ من الاختبار، أثر طول الاختبار على تقديرات القدرة على الرغم من أن حجم التأثير لكل مقارنة كان ضئيلًا. تم إجراء مزيد من التحليلات للكشف عن تأثير المجموعات ضمن أطوال الاختبار ونماذج التحليل. الأشكال من ١-٦ تعرض بيانياً التأثيرات التي تركتها مجموعات العاملين two factors على تقديرات ثيتا theta estimates.

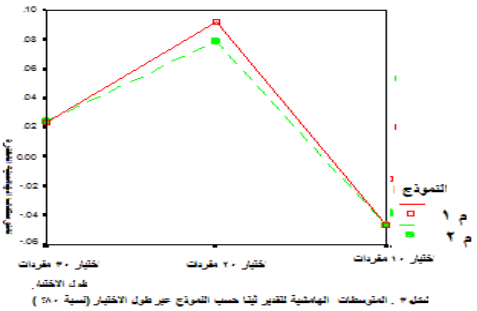
أجريت الاختبارات البعدية بعدما تبين دلالة"ف" F لتحديد اتجاه الفروق باستخدام اختبار شافيه Scheffe المتجانس لمجموعات فرعية عند مستوى ألفا يساوي ٠,٠٥. في نسبة ٦٠ ٪ من المفردات متعددة الاستجابة، كانت اختبارات الطول المكونة من ١٠ و٢٠ مفردة متشابهة في المتوسطات الهامشية. ومع ذلك، كان أداء طول الاختبار المكون من ٣٠ مفردة مختلفًا بشكل دال عن الاختبارات الأقصر (الشكل ١). عند

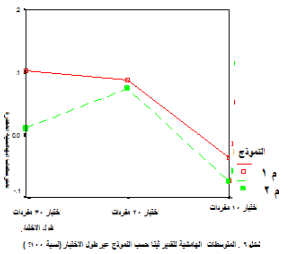
(E-mail of the correspondig: jcps@art.bsu.edu.eg)

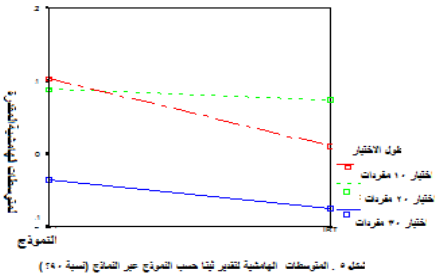
مستوى نسبة ٧٠٪ و٨٠٪، كانت الاختبارات المكونة من ٢٠ مفردة و٣٠ مفردة مختلفة بشكل دال (الشكل ٢ والشكل ٣). عند مستوى نسبة ٩٠٪، اختلف الاختبار المكون من ٣٠ مفردة اختلافًا دالا في المتوسط الهامشي عن الاختبارين الأقصر (الشكل ٤) واختلف نموذجي المجموعتين (المجموعة الأولى والمجموعة الثانية) اختلافًا دالا عبر أطوال الاختبار (الشكل ٥). عند مستوى النسبة ١٠٠٪، اختلفت النسب الهامشية لاختبار ٢٠ مفردة اختلافًا دالا عن الاختبارين الآخرين (الشكل ٦).











**السؤال الثاني: كيف تؤثر نسبة المفردات ثنائية الاستجابة ومتعددة الاستجابة عبر طول الاختبار ونماذج التحليل على وظيفة معلومات الاختبار (TIF)؟**

تعرض الجداول ٦ و٧ و٨ معلومات الاختبار عبر مستويات ثيتا، ومجموع المعلومات وتحسين النسبة (بين نماذج التحليل) لكل مجموعة تصميم. توضح هذه الجداول أنه بالنسبة لجميع أطوال الاختبار، بشكل عام، تكون المعلومات أعلى في منتصف توزيع القدرة ثيتا والمعلومات عبر ثيتا وبشكل عام لها علاقة مباشرة بنسبة المفردات متعددة الاستجابة، ويحسن تحليل المجموعة الثانية (النموذج ثنائي البارامتر للاستجابات الثنائية ونموذج الاستجابة المتدرجة للاستجابات المتعددة). وكان إجمالي المعلومات أكثر للمجموعة الأولى (نموذج راش للاستجابات الثنائية ونموذج التقدير الجزئي للاستجابات المتعددة).

من المعلومات الواردة في الجداول 7 و8 و9، من الواضح أن الاختبارات التي تحتوي على نسبة أكبر من المفردات المصححة بشكل متعدد الاستجابة تجمع معلومات أكثر من تلك التي تم تصحيحها بشكل ثنائي الاستجابة. يُلاحظ هذا بشكل خاص في مستويات القدرة المنخفضة حيث كانت أعلى المعلومات دائمًا تقريبًا عند مستوى متعدد الاستجابة بنسبة ١٠٠٪. في مستويات القدرة الأعلى، تميل أعلى المعلومات إلى أن تكون أقل من ١٠٠٪. تعرض الأشكال ٧ و٨ و٩ مستويات المعلومات عبر توزيع ثيتا لاختبارات بطول ١٠ و٢٠ و٣٠ مفردة.

(E-mail of the correspondig: jcps@art.bsu.edu.eg)

## جدول (7) توزيع معلومات الاختبار وجمعها وتحسينها (اختبارات من ١٠ مفردات)

| النموذج | التعداد% | معلومات على مدى مستويات ثيتا | | | | | | | | | مجموع المعلومات | تحسين الكمية |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 2.0- | 1.5- | 1.0- | 0.5- | 0.0 | 0.5 | 1.0 | 1.5 | 2.0 | | |
| المجموعة الأولى | 0 | 1.8 | 3.5 | 3.5 | 4.5 | 5.6 | 6.7 | 6.5 | 6.5 | 5.6 | 44.2 | %11.5 |
| المجموعة الأولى | 10 | 2.2 | 3.2 | 4.4 | 5.3 | 6.1 | 6.9 | 6.7 | 6.7 | 5.7 | 47.2 | %11.0 |
| المجموعة الأولى | 20 | 2.3 | 3.7 | 5.7 | 6.9 | 7.3 | 7.4 | 6.9 | 6.8 | 5.8 | 52.8 | %11.9 |
| المجموعة الأولى | 30 | 2.4 | 4.2 | 6.9 | 8.4 | 8.4 | 7.9 | 7.1 | **6.9** | **5.9** | 58.1 | %11.2 |
| المجموعة الأولى | 40 | 2.5 | 4.3 | 7.1 | 8.9 | 9.3 | 9.1 | 7.7 | 6 | 5 | 59.9 | %4.0 |
| المجموعة الأولى | 50 | 2.8 | 4.6 | 7.3 | 9.1 | 9.5 | 9.2 | 7.8 | 6 | 5 | 61.3 | %4.9 |
| المجموعة الأولى | 60 | 3.5 | 5.2 | 7.7 | 9.2 | 9.5 | 9.2 | 7.8 | 6 | 5 | 63.1 | %3.2 |
| المجموعة الأولى | 70 | 4.1 | 5.6 | 8 | 9.4 | 9.6 | 9.3 | 7.8 | 6 | 5 | 64.8 | %0.5 |
| المجموعة الأولى | 80 | 4.1 | 5.7 | 8.2 | 9.8 | 10.1 | 9.8 | 8.2 | 6.1 | 4.7 | 66.7 | |
| المجموعة الأولى | 90 | 4.3 | 6 | 8.5 | 10.1 | **10.4** | **9.9** | 8.1 | 5.8 | 4.4 | 67.5 | %1.8 |
| المجموعة الأولى | 100 | **5.4** | **6.8** | **9** | **10.3** | **10.4** | 9.8 | 8 | 3.8 | 4.4 | **67.9** | |
| المجموعة الأولى | 0 | 1.8 | 2.7 | 4.1 | 5.3 | 5.7 | 5.6 | 5.2 | 4.6 | 4.1 | 39.1 | |
| المجموعة الأولى | 10 | 2.1 | 3.2 | 4.8 | 6.2 | 6.5 | 5.9 | 5 | 4.4 | 3.9 | 42 | |
| المجموعة الأولى | 20 | 2.3 | 3.5 | 5.5 | 7.6 | 8.1 | 6.6 | 5 | 4.2 | 3.7 | 46.5 | |
| المجموعة الأولى | 30 | 2.5 | 3.9 | 6.2 | 9 | 9.9 | 7.6 | 5.2 | 3.9 | 3.4 | 51.6 | |
| المجموعة الأولى | 40 | 2.5 | 4 | 6.3 | 9.2 | **10.4** | **8.9** | 7 | 5.3 | 3.9 | 57.5 | |
| المجموعة الأولى | 50 | 3.5 | 5.1 | 7.1 | 9.2 | 9.7 | 8.1 | 6.5 | 5.1 | 4 | 58.3 | |
| المجموعة الأولى | 60 | 4.5 | 6.2 | 7.8 | 9.4 | 9.6 | 8 | 6.4 | 5.2 | 4 | 61.1 | |
| المجموعة الأولى | 70 | 5.4 | 7.2 | 8.7 | 9.6 | 9.3 | 7.8 | 6.7 | 5.4 | 4.4 | 64.5 | |
| المجموعة الأولى | 80 | 5.4 | 7.2 | 8.6 | **9.5** | **9.5** | 8.3 | 7.2 | **6.3** | **4.9** | 66.9 | %0.3 |
| المجموعة الأولى | 90 | 5.3 | 7 | 8.4 | **9.5** | 9.6 | 8.6 | 7.4 | 6 | 4.5 | 66.3 | |
| المجموعة الأولى | 100 | **6.3** | **8.1** | 9 | **9.5** | 9.5 | 8.6 | **7.5** | 6.1 | 4.5 | **69.1** | %1.7 |

ملحوظة: القيم المكتوبة بخط عريض هي أعلى قيمة معلومات ضمن النموذج والنسبة والقدرة.

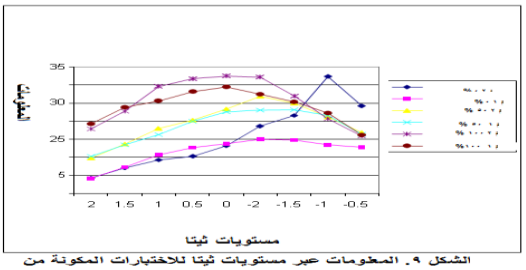
## جدول (8) توزيع معلومات الاختبار وجمعها وتحسينها (اختبارات من ٢٠ مفردات)

| التوزيع | الحصاد % | عدد مستويات ثيتا | | | | | | | | | مجموع المعلومات | تحسين الكمية |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 2.0- | 1.5- | 1.0- | 0.5- | 0.0 | 0.5 | 1.0 | 1.5 | 2.0 | | |
| المجموعة الثانية | 0 | 3.4 | 5.3 | 7 | 8 | 9.4 | 11.5 | 11.6 | 11.1 | 9.7 | 77.0 | %14.2 |
| المجموعة الثانية | 10 | 4.3 | 5.7 | 7.2 | 8.2 | 9.7 | 11.9 | 12.1 | 11.1 | 9.4 | 79.6 | %6.7 |
| المجموعة الثانية | 20 | 4.9 | 6.7 | 8.4 | 9.4 | 10.6 | 12.2 | 11.9 | 10.8 | 9.8 | 84.1 | %8.5 |
| المجموعة الثانية | 30 | 5.9 | 7.3 | 8.8 | 9.6 | 10.9 | 12.7 | 12.8 | **11.7** | 9.7 | 89.4 | %3.2 |
| المجموعة الثانية | 40 | 6.6 | 8.1 | 9.9 | 10.7 | 11.5 | 12.8 | 12.7 | **11.7** | 9.8 | 93.8 | %5.7 |
| المجموعة الثانية | 50 | 7.3 | 9 | 10.7 | 11.3 | 11.8 | 12.6 | 12.4 | 11.5 | 9.8 | 96.4 | %6.1 |
| المجموعة الثانية | 60 | 7.7 | 9.9 | 12.2 | 12.8 | 12.6 | 12.6 | 12.3 | 11.4 | **10.2** | 101.7 | %4.6 |
| المجموعة الثانية | 70 | 7.7 | 10.1 | 12.8 | 13.9 | 13.8 | 13.7 | 13 | 11.1 | 9.1 | 105.2 | %0.7 |
| المجموعة الثانية | 80 | 8.3 | 10.5 | 13.1 | 14 | 13.9 | 13.7 | **13** | 11.1 | 9.1 | 106.7 | %0.6 |
| المجموعة الثانية | 90 | 8.8 | 11 | 13.5 | 14.4 | 14.3 | **14.1** | **13.3** | 11.1 | 8.8 | 109.3 | %0.6 |
| المجموعة الثانية | 100 | **9.6** | **11.7** | **13.9** | **14.7** | **14.5** | **14.1** | 13.2 | 11.1 | 8.7 | **111.5** | %2.7 |
| المجموعة الأولى | 0 | 3.3 | 5.3 | 7.4 | 8.9 | 9.4 | 9.6 | 9.1 | 7.8 | 6.6 | 67.4 | |
| المجموعة الأولى | 10 | 4.9 | 7 | 8 | 8.7 | 9.2 | 9.9 | 10.1 | 9.3 | 7.5 | 74.6 | |
| المجموعة الأولى | 20 | 5 | 7.1 | 9 | 10.4 | 10.9 | 10.5 | 9.6 | 8.2 | 6.6 | 77.5 | |
| المجموعة الأولى | 30 | 7.3 | 9.1 | 9.4 | 9.8 | 10.2 | 10.5 | 10.7 | **10.5** | **9.1** | 86.6 | |
| المجموعة الأولى | 40 | 7.8 | 9.2 | 9.9 | 11 | 11.7 | 11.3 | 10.4 | 9.4 | 8 | 88.7 | |
| المجموعة الأولى | 50 | 8.3 | 9.9 | 10.9 | 11.9 | 12.2 | 11.2 | 9.9 | 8.9 | 7.7 | 90.9 | |
| المجموعة الأولى | 60 | 8.6 | 10.5 | 12 | 13.6 | 13.9 | 12.3 | 10.1 | 8.7 | 7.5 | 97.2 | |
| المجموعة الأولى | 70 | 8.9 | 10.9 | 12.4 | **14.4** | **15.2** | **13.6** | 11.4 | 9.7 | 8 | 104.5 | |
| المجموعة الأولى | 80 | 10.1 | 12.1 | 13.3 | **14.4** | 14.7 | 13.1 | 11 | 9.4 | **8** | 106.1 | |
| المجموعة الأولى | 90 | 10.7 | 12.6 | **13.5** | 14.3 | 14.4 | 13 | 11.3 | 10.1 | 8.7 | **108.6** | |
| المجموعة الأولى | 100 | **11.1** | **12.8** | 13.4 | 13.9 | 14.1 | 13.1 | **11.5** | 10.1 | 8.6 | **108.6** | |

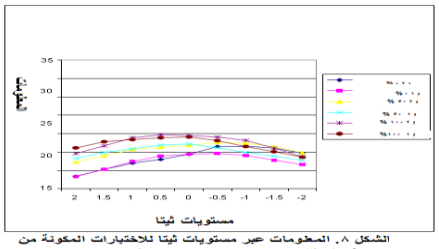
(E-mail of the correspondig: jcps@art.bsu.edu.eg)

## جدول (9) توزيع معلومات الاختبار وجمعها وتحسينها (اختبارات من ٣٠ مفردات)

| النموذج | التصاعد% | مدى مستويات ثيتا | | | | | | | | | | مجموع المعلومات | تحسين النسبة |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 2.0- | 1.5- | 1.0- | 0.5- | 0.0 | 0.5 | 1.0 | 1.5 | 2.0 | | | |
| المجموعة الثانية | 0 | 4.3 | 7.1 | 9.3 | 10.3 | 13.2 | 18.7 | 21.6 | 32.4 | 24.3 | | 141.2 | %34.7 |
| المجموعة الثانية | 10 | 5 | 8.6 | 11.8 | 12.9 | 15 | 18.8 | 20.3 | 27.7 | 22.7 | | 142.8 | %25.0 |
| المجموعة الثانية | 20 | 6.5 | 10 | 14.1 | 15.7 | 17.6 | 20.8 | 20.8 | 21.2 | 20.6 | | 147.3 | %13.0 |
| المجموعة الثانية | 30 | 7.5 | 10.7 | 14.7 | 16.6 | 18.9 | 22.1 | 21.6 | **20.4** | 18.9 | | 151.4 | %7.1 |
| المجموعة الثانية | 40 | 9.1 | 12.2 | 16.3 | 18.6 | 21.8 | 25.4 | 23.6 | **20.2** | 16.7 | | 163.9 | |
| المجموعة الثانية | 50 | 9.8 | 13.7 | 18 | 20.3 | 23.4 | 26.9 | 25 | **21.2** | 16.9 | | 175.2 | %4.9 |
| المجموعة الثانية | 60 | 11.7 | 14.9 | 19.3 | 21.4 | 29.7 | 28.2 | **35.1** | 23.1 | 16.5 | | 199.9 | %11.6 |
| المجموعة الثانية | 70 | 13.5 | 17.6 | 22.7 | 24.5 | 26.2 | 27.7 | 24.9 | 20.9 | 16.7 | | 194.7 | %4.1 |
| المجموعة الثانية | 80 | 16.3 | 21.3 | 28.4 | 30.8 | 31.7 | 31.6 | **26.7** | 20.9 | 16.4 | | 224.1 | %6.8 |
| المجموعة الثانية | 90 | 16.1 | 20.2 | 23.2 | 28.4 | 30.7 | 28 | 25.5 | 22.1 | 15.6 | | 209.8 | |
| المجموعة الثانية | 100 | 17.9 | **22.8** | **29.6** | **31.8** | **32.5** | 32.2 | 27 | 20.7 | **15.9** | | **230.4** | %5.7 |
| المجموعة الأولى | 0 | 4.2 | 7.3 | 10.7 | 12.7 | 13.8 | 15 | 14.8 | 13.5 | 12.8 | | 104.8 | |
| المجموعة الأولى | 10 | 4.7 | 8.3 | 12.2 | 14.9 | 16 | 16.2 | 15.4 | 13.8 | 12.7 | | 114.2 | |
| المجموعة الأولى | 20 | 6.9 | 9.8 | 13.4 | 17.1 | 18.9 | 18.9 | 17.3 | 15.2 | 12.8 | | 130.3 | |
| المجموعة الأولى | 30 | 8.6 | 10.9 | 13.6 | 17.1 | 19.4 | 20 | 19.9 | 17.8 | 14.1 | | 141.4 | |
| المجموعة الأولى | 40 | 16.3 | 13.2 | 15 | 17.8 | 20.5 | 22 | 23.1 | 21.2 | 15.6 | | 164.7 | %0.5 |
| المجموعة الأولى | 50 | 10.3 | 13.5 | 16.3 | 20 | 22.6 | 23.1 | 23.2 | 21.6 | 16.4 | | 167 | |
| المجموعة الأولى | 60 | 13.8 | 16.6 | 17.9 | 21.2 | 23.9 | 24 | 23.6 | 21.8 | 16.4 | | 179.2 | |
| المجموعة الأولى | 70 | 14.9 | 18.3 | 20.5 | 24.1 | 25.9 | 24.3 | 22.7 | 20.6 | 15.8 | | 187.1 | |
| المجموعة الأولى | 80 | 16.1 | 20.2 | 23.2 | 28.4 | 30.7 | 28 | 25.5 | 22.1 | 15.6 | | 209.8 | |
| المجموعة الأولى | 90 | 18.9 | 23.6 | 25.7 | 28.7 | 29.5 | 26.5 | 24.3 | 21.6 | 16 | | 214.8 | %2.4 |
| المجموعة الأولى | 100 | 19.3 | 23.9 | 25.6 | 28.2 | 29.5 | 27.5 | 25.4 | 22.3 | 16.2 | | 217.9 | |



الشكل ٩. المعلومات عبر مستويات ثيتا للاختبارات المكونة من ٣٠ مفردة (النسب ٢٠ و ٥٠ و ١٠٠ ٪)



الشكل ٨. المعلومات عبر مستويات ثيتا للاختبارات المكونة من ٢٠ مفردة (النسب ٢٠ و ٥٠ و ١٠٠ ٪)

يعرض الجدول( ٩) النسبة المئوية للزيادة في مستويات المعلومات عند زيادة طول الاختبار من ١٠ إلى ٢٠ مفردة لكلا النموذجين بثلاثة نسب كحد أقصى (٠٪، ٥٠٪، ١٠٠٪ تصحيح متعدد الاستجابة.) يمكن ملاحظة أن زيادة طول الاختبار من ١٠ إلى ٢٠ مفردة، أو مضاعفة الطول، يزيد متوسط المعلومات عبر النماذج ونسبة المفردات متعددة الاستجابة بنسبة ٧١٪. يبدو أن هذا المعدل مستقر إلى حد ما عبر كلا النموذجين ونسب تصحيح مرتفعة. يعرض الجدول ١٠ النسب المئوية المحسوبة عند زيادة طول الاختبار من ٢٠ إلى ٣٠ مفردة. النسبة المئوية للزيادة في المتوسط هي ٨٢٪ مع أدنى قيمة لراش ٠٪ متعدد الاستجابة عند ٥٤٪ ونظرية الاستجابة للمفردة١٠٠% متعدد الاستجابة الأعلى عند ١٠٤٪. يعرض الجدول ١١ النسب المئوية المحسوبة عند زيادة طول الاختبار من ١٠ إلى ٣٠ مفردة (ثلاثة أضعاف). تزيد هذه الزيادة في طول الاختبار من المعلومات الممكنة المتاحة بنسبة ٢٠٧٪ في المتوسط عبر جميع النماذج والنسب، وتتراوح من ١٦٧٪ لراش ٠% متعدد الاستجابة الى٢٥٤% لنظرية الاستجابة للمفردة ١٠٠% متعدد الاستجابة.

**جدول (١٠). زيادة النسبة المئوية في المعلومات من اختبار ١٠ إلى ٢٠ مفردة**

| | 2- | 1.5- | 1- | 0.5- | 0 | 0.5 | 1 | 1.5 | 2 | متوسط الزيادة |
|---|---|---|---|---|---|---|---|---|---|---|
| IRT-0% | %89 | %51 | %100 | %78 | %68 | %72 | %78 | %71 | %73 | %76 |
| Rasch-0% | %83 | %96 | %80 | %68 | %65 | %71 | %75 | %70 | %61 | %74 |
| IRT-50% | %161 | %96 | %47 | %24 | %24 | %37 | %59 | %92 | %96 | %71 |
| Rasch-50% | %137 | %94 | %54 | %29 | %26 | %38 | %52 | %75 | %93 | %66 |
| IRT-100% | %78 | %72 | %54 | %43 | %39 | %44 | %65 | %192 | %98 | %76 |
| Rasch-100% | %76 | %58 | %49 | %46 | %48 | %52 | %53 | %66 | %91 | %60 |
| الاجمالى | | | | | | | | | | %71 |

(E-mail of the correspondig: jcps@art.bsu.edu.eg)

**جدول ( ١١). زيادة النسبة المئوية في المعلومات من اختبار ٢٠ إلى ٣٠ مفردة**

| | 2- | 1.5- | 1- | 0.5- | 0 | 0.5 | 1 | 1.5 | 2 | متوسط الزيادة |
|---|---|---|---|---|---|---|---|---|---|---|
| IRT-0% | %26 | %34 | %33 | %29 | %40 | %63 | %86 | %192 | %151 | %73 |
| Rasch-0% | %27 | %38 | %45 | %43 | %47 | %56 | %63 | %73 | %94 | %54 |
| IRT-50% | %34 | %52 | %68 | %80 | %98 | %113 | %102 | %84 | %72 | %78 |
| Rasch-50% | %24 | %36 | %50 | %68 | %85 | %106 | %134 | %143 | %113 | %84 |
| IRT-100% | %86 | %95 | %113 | %116 | %124 | %128 | %105 | %86 | %83 | %104 |
| Rasch-100% | %74 | %87 | %91 | %103 | %109 | %110 | %121 | %121 | %88 | %100 |
| الاجمالى | | | | | | | | | | %82 |

**جدول(١٢). زيادة النسبة المئوية في المعلومات من اختبار ١٠ إلى ٣٠ مفردة**

| | 2- | 1.5- | 1- | 0.5- | 0 | 0.5 | 1 | 1.5 | 2 | متوسط الزيادة |
|---|---|---|---|---|---|---|---|---|---|---|
| IRT-0% | %139 | %103 | %166 | %129 | %136 | %179 | %232 | %398 | %334 | %202 |
| Rasch-0% | %133 | %170 | %161 | %140 | %142 | %168 | %185 | %193 | %212 | %167 |
| IRT-50% | %250 | %198 | %147 | %123 | %146 | %192 | %221 | %253 | %238 | %196 |
| Rasch-50% | %194 | %165 | %130 | %117 | %133 | %185 | %257 | %324 | %310 | %202 |
| IRT-100% | %231 | %235 | %229 | %209 | %213 | %229 | %238 | %445 | %261 | %254 |
| Rasch-100% | %206 | %195 | %184 | %197 | %211 | %220 | %239 | %266 | %260 | %220 |
| الاجمالى | | | | | | | | | | %207 |

للإجابة على السؤال البحثي الثاني، أشارت الحسابات إلى وجود علاقة مباشرة بين النسب الأعلى للمفردات متعددة الاستجابة ومقدار المعلومات المتاحة التي ينتجها الاختبار. أشارت جميع التحليلات إلى أن طول الاختبار يلعب دورًا رئيسيًا في تحديد مقدار المعلومات التي يتم جمعها من المفردات. يبدو هذا واضحًا لأن معلومات الاختبار الإجمالية هي مجموع كل قيم معلومات المفردة. مع وجود المزيد من المفردات، يتوفر المزيد من المعلومات.

**السؤال الثالث: كيف تؤثر نسبة المفردات ثنائية الاستجابة ومتعددة الاستجابة عبر طول الاختبار ونماذج التحليل على التحسن الإجمالي للاختبار؟**

تم استخدام ثلاثة معايير لتقييم أمثلية الاختبار عبر مجموعات التصميم في هذه الدراسة: الثبات الهامشي، تحليلات الأنوفا ANOVA لمعدلات الخطأ المعياري لكل مجموعة تصميم، ومستويات المعلومات. يعرض الجدول ١٣ قيم الثبات الهامشي عبر مجموعات التصميم. داخل كل اختبار طول يزداد الثبات عبر نسب المفردات متعددة الاستجابة. وتجدر الإشارة أيضًا إلى أنه عبر جميع أطوال الاختبار الثلاثة، زاد الثبات مع إطالة الاختبار. بالنظر إلى معاملات الثبات الهامشية عبر النماذج، فإن الاختلافات بين مجموعتي النماذج لا تكاد تذكر.

تتألف المعايير الثانية التي تم تحليلها لتحديد أمثلية الاختبار من ١١ انوفا 2 × 3 (واحد لكل مستوى نسبة) أجريت على ٦٦ مجموعة من معدلات الخطأ المعيارية بطول الاختبار والنموذج كمتغيرات مستقلة. تم تقييم التأثيرات الرئيسية والتفاعلية للأهمية من خلال مقارنة قيم F المرصودة بقيمة F الحرجة. يتم تحديد قيمة F الحرجة من خلال درجات الحرية لكل تأثير ومستوى ألفا الذي تم تعيينه عند 05. = α. تمت مقارنة كل F تمت مرصودة مع F حرجة ٠,٠٥, ٢, ∞) = ٣,٨٤ لاختبار أهمية النموذج وF حرجة (٠,٠٥، ٢، ∞) = ٣,٠٠ لاختبار أهمية طول الاختبار والتفاعل بين النموذج وطول الاختبار. تم تقييم حجم التأثير أيضا. حجم التأثير هو نسبة مجموع المربعات بين المجموعات والمجموع الإجمالي للمربعات، وفي هذه التحليلات تمثل نسبة التباين في تقديرات الخطأ المعياري الموضح بالفرق بين المجموعات. وكما ورد سابقًا، تم قياس تقديرات حجم التأثير بقيم مربع إيتا الجزئية وتم تقييمها وفقًا لمعايير كوهين (صغير = ٠,١، متوسط = ٠,٢٥ وكبير = ٠,٤). تم حساب حسابات القدرة المرصودة أيضًا باستخدام SPSS وتم تضمينها في نتائج الانوفا ANOVA المجدولة

(E-mail of the correspondig: jcps@art.bsu.edu.eg)

## جدول (١٣). الثبات الهامشي عبر طول الاختبار ونسبة التصحيح والنموذج

| المفردة | %البول | م١ | م٢ | الفرق |
|---|---|---|---|---|
| | %0 | 0.78 | 0.78 | 0 |
| | %10 | 0.8 | 0.81 | 0.01- |
| | %20 | 0.82 | 0.84 | 0.02- |
| | %30 | 0.84 | 0.85 | 0.01- |
| | %40 | 0.86 | 0.86 | 0 |
| 10 | %50 | 0.86 | 0.87 | 0.01- |
| | %60 | 0.87 | 0.87 | 0 |
| | %70 | 0.87 | 0.87 | 0 |
| | %80 | 0.88 | 0.88 | 0 |
| | %90 | 0.88 | 0.88 | 0 |
| | % 100 | 0.88 | 0.88 | 0 |
| | %0 | 0.88 | 0.88 | 0 |
| | %10 | 0.89 | 0.89 | 0 |
| | %20 | 0.89 | 0.9 | 0.01- |
| | %30 | 0.9 | 0.9 | 0 |
| | %40 | 0.9 | 0.91 | 0.01- |
| 20 | %50 | 0.91 | 0.91 | 0 |
| | %60 | 0.91 | 0.92 | 0.01- |
| | %70 | 0.92 | 0.92 | 0 |
| | %80 | 0.92 | 0.92 | 0 |
| | %90 | 0.92 | 0.92 | 0 |
| | % 100 | 0.92 | 0.92 | 0 |
| | %80 | 0.92 | 0.92 | 0 |
| | %90 | 0.92 | 0.92 | 0 |
| | % 100 | 0.92 | 0.92 | 0 |
| | %0 | 0.92 | 0.92 | 0 |
| | %10 | 0.93 | 0.93 | 0 |
| | %20 | 0.94 | 0.94 | 0 |
| | %30 | 0.94 | 0.94 | 0 |
| | %40 | 0.95 | 0.95 | 0 |
| 30 | %50 | 0.95 | 0.95 | 0 |
| | %60 | 0.95 | 0.96 | 0.01- |
| | %70 | 0.96 | 0.96 | 0 |
| | %80 | 0.96 | 0.96 | 0 |
| | %90 | 0.96 | 0.96 | 0 |
| | % 100 | 0.96 | 0.96 | 0 |

أدت الانوفا ANOVA المحسوبة لجميع مستويات النسبة الأحد عشر لثلاثة تأثيرات إلى ٣٠ قيمة F دالة للنموذج (١١)، وطول الاختبار (١١) وتفاعل طول النموذج على حدة (٨) مع الاستثناءات الثلاثة فقط: تأثيرات التفاعل لـ ٠ لم تكن نماذج النسبة ٠٪ و٥٠٪ و١٠٠٪ دالة. كانت نسب ٠٪ و١٠٠٪ هي الاختبارات الوحيدة التي تم تحليلها والتي لم تكن اختبارات ذات مفردات مختلطة حيث تم تسجيل مستوى النسبة ٠٪ بشكل ثنائي الاستجابة بنسبة ١٠٠٪ وتم تصحيح نموذج النسبة ١٠٠٪ بشكل متعدد الاستجابة. تراوحت قيم F الدالة من تحليلات الأنوفا ANOVAs الأخرى من ١١,٠١٩ إلى ٩٥٠,٣١٩ للتأثير الرئيسي للنموذج، من ٢٢٦٥,٢١٣ إلى ٣٢٢١٦,٦٩٨ للتأثير الرئيسي لطول الاختبار، ومن ٤,١٣٩ إلى ٧٣٤٨,١١ لتأثير التفاعل بين النموذج وطول الاختبار. يتم عرض النتائج في الجداول من ١٤ إلى ٢٤.

تراوح حجم التأثير عبر جميع النسب للتأثير الرئيسي للنموذج من ٠,٠٠٢ إلى ٠,٠١٦ مع استثناءين: نسبة ٦٠٪ لها حجم تأثير ٠,١٣٧ ونسبة ٩٠٪ لها حجم تأثير ٠,٥٩١. تراوح حجم التأثير عبر جميع النسب للتأثير الرئيسي لطول الاختبار من ٠,٦٠٨ إلى ٠,٧٨١ مع نفس الاستثناءين: كان لنسبة ٦٠٪ حجم تأثير ٠,٤٣٠ ونسبة ٩٠٪ كان حجم تأثيرها ٠,٩١٥. تراوح حجم التأثير عبر جميع النسب لتأثير تفاعل النموذج حسب الطول من ٠,٠٠٠ إلى ٠,٠١٢ مع نفس الاستثناءين: نسبة ٦٠٪ كان لها حجم تأثير ٠,١٦٠ ونسبة ٩٠٪ لها حجم تأثير ٠,٧١٠.

ظلت القدرة عبر جميع النسب ثابتة بين ٠,٩١٣ و١ للتأثير الرئيسي للنموذج وتساوي ١ لجميع التأثيرات الرئيسية بطول الاختبار. بالنسبة لتأثير التفاعل، تراوحت قوة تأثير التفاعل من ٠,٧٣٣ إلى ١ لثمانية من مستويات النسبة مع انخفاض مستويات النسبة ٠٪ و٥٠٪ و١٠٠٪ أقل بكثير من الباقي عند ٠,٣٧٥ و٠,٣٤٣ و٠,١٤٩ على التوالي.

(E-mail of the correspondig: jcps@art.bsu.edu.eg)

**جدول (١٤). تحليل التباين الثنائي للتأثيرات الثابتة ثنائية الاستجابة على الأخطاء المعيارية: النسبة ٠٪ (النموذج، وطول الاختبار)**

| المصدر | مجموع المربعات | درجات الحرية | متوسط المربعات | ف | P قيمة | مربع ايتا الجزئى | المعلمة غير الرئيسة | القوة المرصودة (a) |
|---|---|---|---|---|---|---|---|---|
| النموذج المصحح | 28.664(b) | 5 | 5.733 | 1990.358 | 0 | 0.624 | 9951.788 | 1 |
| التقاطع | 771.97 | 1 | 771.97 | 268017.996 | 0 | 0.978 | 268017.996 | 1 |
| النموذج | 0.059 | 1 | 0.059 | 20.515 | 0 | 0.003 | 20.515 | 0.995 |
| الطول | 28.595 | 2 | 14.297 | 4963.852 | 0 | 0.624 | 9927.703 | 1 |
| النموذج * الطول | 0.01 | 2 | 0.005 | 1.785 | 0.168 | 0.001 | 3.57 | 0.375 |
| الخطأ | 17.264 | 5994 | 0.003 | | | | | |
| الكلى | 817.899 | 6000 | | | | | | |
| الكلى المصحح | 45.929 | 5999 | | | | | | |

(a) محسوبة باستخدام ألفا 0.05 ؛ b ر التربيعية = ٠,٦٢٤ (ر التربيعية المعدلة = ٠,٦٢٤)

**جدول (١٥). تحليل التباين الثنائي للتأثيرات الثابتة ثنائية الاستجابة على الأخطاء المعيارية: النسبة ١٠٪ (النموذج، وطول الاختبار)**

| المصدر | مجموع المربعات | درجات الحرية | متوسط المربعات | ف | P قيمة | مربع ايتا الجزئية | المعلمة غير الرئيسة | القوة المرصودة (a) |
|---|---|---|---|---|---|---|---|---|
| النموذج المصحح | 27.574(b) | 5 | 5.515 | 2426.215 | 0 | 0.669 | 12131.074 | 1 |
| التقاطع | 703.077 | 1 | 703.077 | 309315.367 | 0 | 0.981 | 309315.367 | 1 |
| النموذج | 0.031 | 1 | 0.031 | 13.484 | 0 | 0.002 | 13.484 | 0.957 |
| الطول | 27.525 | 2 | 13.762 | 6054.656 | 0 | 0.669 | 12109.312 | 1 |
| النموذج * الطول | 0.019 | 2 | 0.009 | 4.139 | 0.016 | 0.001 | 8.278 | 0.733 |
| الخطا | 13.624 | 5994 | 0.002 | | | | | |
| الكلى | 744.275 | 6000 | | | | | | |
| الكلى المصحح | 41.198 | 5999 | | | | | | |

(a) محسوبة باستخدام ألفا ٠,٠٥؛ b ر التربيعية = ٠,٦٦٩ (ر التربيعية المعدلة = ٠,٦٦٩)

**جدول (١٦). تحليل التباين الثنائي للتأثيرات الثابتة ثنائية الاستجابة على الأخطاء المعيارية: النسبة ٢٠٪ (النموذج، وطول الاختبار)**

| المصدر | مجموع المربعات | درجات الحرية | متسوط المربعات | ف | P | مربع ايتا الجزئية | المعلمة غير الرئيسة | القوة المرصودة (a) |
|---|---|---|---|---|---|---|---|---|
| | | | | | قيمة | | | |
| النموذج المصحح | 23.515(b) | 5 | 4.703 | 2296.209 | 0 | 0.657 | 11481.046 | 1 |
| التقاطع | 629.415 | 1 | 629.415 | 307302.91 | 0 | 0.981 | 307302.91 | 1 |
| النموذج المصحح | 0.086 | 1 | 0.086 | 41.979 | 0 | 0.007 | 41.979 | 1 |
| الطول | 23.365 | 2 | 11.683 | 5703.85 | 0 | 0.656 | 11407.7 | 1 |
| النموذج * الطول | 0.064 | 2 | 0.032 | 15.684 | 0 | 0.005 | 31.368 | 0.999 |
| الخطأ | 12.277 | 5994 | 0.002 | | | | | |
| الكلى | 665.207 | 6000 | | | | | | |
| الكلى المصحح | 35.792 | 5999 | | | | | | |

(a) محسوبة باستخدام ألفا ٠,٠٥؛ b ر التربيعية = ٠,٦٥٧ (ر التربيعية المعدلة = ٠,٦٥٧)

**جدول (١٧). تحليل التباين الثنائي للتأثيرات الثابتة ثنائية الاستجابة على الأخطاء المعيارية: النسبة ٣٠٪ (النموذج، وطول الاختبار)**

| المصدر | مجموع المربعات | ف | متوسط المربعات | ف | P | مربع ايتا الجزئية | المعلمة غير الرئيسة | القوة المرصودة (a) |
|---|---|---|---|---|---|---|---|---|
| | | | | | قيمة | | | |
| النموذج المصحح | 19.445(b) | 5 | 3.889 | 1876.586 | 0 | 0.61 | 9382.93 | 1 |
| التقاطع | 578.143 | 1 | 578.143 | 278976.18 | 0 | 0.979 | 278976.18 | 1 |
| النموذج | 0.023 | 1 | 0.023 | 11.019 | 0.001 | 0.002 | 11.019 | 0.913 |
| الطول | 19.272 | 2 | 9.636 | 4649.712 | 0 | 0.608 | 9299.424 | 1 |
| النموذج * الطول | 0.15 | 2 | 0.075 | 36.244 | 0 | 0.012 | 72.487 | 1 |
| الخطأ | 12.422 | 5994 | 0.002 | | | | | |
| الكلى | 610.01 | 6000 | | | | | | |
| الكلى المصحح | 31.867 | 5999 | | | | | | |

(E-mail of the correspondig: jcps@art.bsu.edu.eg)

(a) محسـوبة باسـتخدام ألفا ٠,٠٥؛ b ر التربيعية = ٠,٦١٠ (ر التربيعية المعدلة = ٠,٦١٠)

**جدول (١٨). تحليل التباين الثنائي للتأثيرات الثابتة ثنائية الاسـتجابة على الأخطاء المعيارية: النسبة ٤٠٪ (النموذج، وطول الاختبار)**

| المصدر | مجموع المربعات | درجات الحرية | متوسط المربعات | ف | P قيمة | مربع ايتا الجزئية | المعلمة غير الرئيسة | القوة المرصودة (a) |
|---|---|---|---|---|---|---|---|---|
| النموذج المصحح | 18.955(b) | 5 | 3.791 | 2000.403 | 0 | 0.625 | 10002.015 | 1 |
| التقاطع | 543.057 | 1 | 543.057 | 286552.14 | 0 | 0.98 | 286552.14 | 1 |
| النموذج | 0.042 | 1 | 0.042 | 22.094 | 0 | 0.004 | 22.094 | 0.997 |
| الطول | 18.892 | 2 | 9.446 | 4984.412 | 0 | 0.625 | 9968.824 | 1 |
| الطول * النموذج | 0.021 | 2 | 0.011 | 5.549 | 0.004 | 0.002 | 11.097 | 0.856 |
| الخطأ | 11.359 | 5994 | 0.002 | | | | | |
| الكلى | 573.372 | 6000 | | | | | | |
| الكلى المصحح | 30.315 | 5999 | | | | | | |

(a) محسـوبة باسـتخدام ألفا ٠,٠٥؛ b ر التربيعية = ٠,٦٢٥ (ر التربيعية المعدلة = ٠,٦٢٥)

**جدول (١٩). تحليل التباين الثنائي للتأثيرات الثابتة ثنائية الاسـتجابة على الأخطاء المعيارية: النسبة ٥٠٪ (النموذج، وطول الاختبار)**

| المصدرة | مجموع المربعات | درجات الحرية | متوسط المربعات | F | P قيمة | مربع ايتا الجزئية | المعلمة غير الرئيسة | القوة المرصودة (a) |
|---|---|---|---|---|---|---|---|---|
| النموذج المصحح | 19.671(b) | 5 | 3.934 | 2481.864 | 0 | 0.674 | 12409.319 | 1 |
| التقاطع | 529.345 | 1 | 529.345 | 333932.24 | 0 | 0.982 | 333932.24 | 1 |
| النموذج | 0.101 | 1 | 0.101 | 63.74 | 0 | 0.011 | 63.74 | 1 |
| الطول | 19.565 | 2 | 9.782 | 6171.178 | 0 | 0.673 | 12342.356 | 1 |
| النموذج * الطول | 0.005 | 2 | 0.003 | 1.611 | 0.2 | 0.001 | 3.223 | 0.343 |
| الخطأ | 9.502 | 5994 | 0.002 | | | | | |
| الكلى | 558.518 | 6000 | | | | | | |
| الكلى المصحح | 29.173 | 5999 | | | | | | |

(a) محسـوبة باسـتخدام ألفا ٠,٠٥؛ b ر التربيعية = ٠,٦٧٤ (ر التربيعية المعدلة = ٠,٦٧٤)

**جدول (٢٠). تحليل التباين الثنائي للتأثيرات الثابتة ثنائية الاسـتجابة على الأخطاء المعيارية: النسبة ٦٠٪ (النموذج، وطول الاختبار)**

| المصدر | مجموع المربعات | درجات الحرية | متوسط المربعات | ف | P قيمة | مربع ايتا الجزئية | المعلمة غير الرئيسة | القوة المرصودة (a) |
|---|---|---|---|---|---|---|---|---|
| النموذج المصحح | 14.117(b) | 5 | 2.823 | 1325.302 | 0 | 0.525 | 6626.512 | 1 |
| التقاطع | 561.275 | 1 | 561.275 | 263455.45 | 0 | 0.978 | 263455.45 | 1 |
| النموذج | 2.025 | 1 | 2.025 | 950.319 | 0 | 0.137 | 950.319 | 1 |
| الطول | 9.652 | 2 | 4.826 | 2265.213 | 0 | 0.43 | 4530.427 | 1 |
| النموذج * الطول | 2.441 | 2 | 1.22 | 572.883 | 0 | 0.16 | 1145.766 | 1 |
| الخطأ | 12.77 | 5994 | 0.002 | | | | | |
| الكلى | 588.162 | 6000 | | | | | | |
| الكلى المصحح | 26.887 | 5999 | | | | | | |

(a) محسـوبة باسـتخدام ألفا ٠,٠٥؛ b ر التربيعية = ٠,٥٢٥ (ر التربيعية المعدلة = ٠,٥٢٥)

**جدول (٢١). تحليل التباين الثنائي للتأثيرات الثابتة ثنائية الاسـتجابة على الأخطاء المعيارية: النسبة ٧٠٪ (النموذج، وطول الاختبار)**

| المصدر | مجموع المربعات | درجات الحرية | متوسط المربعات | ف | P قيمة | مربع ايتا الجزئية | المعلمة غير الرئيسة | القوة المرصودة (a) |
|---|---|---|---|---|---|---|---|---|
| النموذج المصحح | 21.211(b) | 5 | 4.242 | 3402.595 | 0 | 0.739 | 17012.977 | 1 |
| التقاطع | 484.489 | 1 | 484.489 | 388599.48 | 0 | 0.985 | 388599.48 | 1 |
| النموذج | 0.058 | 1 | 0.058 | 46.298 | 0 | 0.008 | 46.298 | 1 |
| الطول | 21.138 | 2 | 10.569 | 8477.007 | 0 | 0.739 | 16954.014 | 1 |
| النموذج * الطول | 0.016 | 2 | 0.008 | 6.333 | 0.002 | 0.002 | 12.666 | 0.9 |
| الخطأ | 7.473 | 5994 | 0.001 | | | | | |
| الكلى | 513.174 | 6000 | | | | | | |
| الكلى المصحح | 28.684 | 5999 | | | | | | |

(E-mail of the correspondig: jcps@art.bsu.edu.eg)

(a) محسـوبة باسـتخدام ألفا ٠,٠٥؛ b ر التربيعية = ٠,٧٣٩ (ر التربيعية المعدلة = ٠,٧٣٩)

**جدول (٢٢). تحليل التباين الثنائي للتأثيرات الثابتة ثنائية الاسـتجابة على الأخطاء المعيارية: النسبة ٨٠٪ (النموذج، وطول الاختبار)**

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | مجموع المربعات | | متوسط المربعات | | P | مربع ايتا الجزئية | المعلمة غير الرئيسة | القوة المرصودة (a) |
| المصدر | | درجات الحرية | | ف | قيمة | | | |
| النموذج المصحح | 22.193(b) | 5 | 4.439 | 3779.672 | 0 | 0.759 | 18898.362 | 1 |
| التقاطع | 462.381 | 1 | 462.381 | 393732.7 | 0 | 0.985 | 393732.7 | 1 |
| النموذج | 0.019 | 1 | 0.019 | 16.27 | 0 | 0.003 | 16.27 | 0.981 |
| الطول | 22.163 | 2 | 11.082 | 9436.311 | 0 | 0.759 | 18872.622 | 1 |
| النموذج * الطول | 0.011 | 2 | 0.006 | 4.735 | 0.009 | 0.002 | 9.471 | 0.793 |
| الخطأ | 7.039 | 5994 | 0.001 | | | | | |
| الكلى | 491.614 | 6000 | | | | | | |
| الكلى المصحح | 29.232 | 5999 | | | | | | |

(a) محسـوبة باسـتخدام ألفا ٠,٠٥؛ b ر التربيعية = ٠,٧٥٩ (ر التربيعية المعدلة = ٠,٧٥٩)

**جدول (٢٣). تحليل التباين الثنائي للتأثيرات الثابتة ثنائية الاسـتجابة على الأخطاء المعيارية: النسبة ٩٠٪ (النموذج، وطول الاختبار)**

| المصدر | مجموع المربعات | درجات الحرية | متوسط المربعات | ف | P قيمة | مربع ايتا الجزئية | المعلمة غير الرئيسة | القوة المرصودة (a) |
|---|---|---|---|---|---|---|---|---|
| النموذج المصحح | (b)96.361 | 5 | 19.272 | 17555.889 | 0 | 0.936 | 87779.444 | 1 |
| التقاطع | 581.684 | 1 | 581.684 | 529881.81 | 0 | 0.989 | 529881.81 | 1 |
| النموذج | 9.495 | 1 | 9.495 | 8649.715 | 0 | 0.591 | 8649.715 | 1 |
| الطول | 70.733 | 2 | 35.366 | 32216.698 | 0 | 0.915 | 64433.396 | 1 |
| النموذج * الطول | 16.133 | 2 | 8.067 | 7348.166 | 0 | 0.71 | 14696.332 | 1 |
| الخطأ | 6.58 | 5994 | 0.001 | | | | | |
| الكلى | 684.625 | 6000 | | | | | | |
| الكلى المصحح | 102.941 | 5999 | | | | | | |

(a) محسـوبة باسـتخدام ألفا ٠,٠٥؛ b ر التربيعية = ٠,٩٣٦ (ر التربيعية المعدلة = ٠,٩٣٦)

**جدول (٢٤). تحليل التباين الثنائي للتأثيرات الثابتة ثنائية الاسـتجابة على الأخطاء المعيارية: النسبة ١٠٠٪ (النموذج، وطول الاختبار)**

| المصدر | مجموع المربعات | درجات الحرية | متوسط المربعات | ف | P قيمة | مربع ايتا الجزئية | المعلمة غير الرئيسة | القوة المرصودة (a) |
|---|---|---|---|---|---|---|---|---|
| النموذج المصحح | (b)23.003 | 5 | 4.601 | 4287.854 | 0 | 0.782 | 21439.27 | 1 |
| التقاطع | 446.861 | 1 | 446.861 | 416485.31 | 0 | 0.986 | 416485.31 | 1 |
| النموذج | 0.103 | 1 | 0.103 | 96.287 | 0 | 0.016 | 96.287 | 1 |
| الطول | 22.898 | 2 | 11.449 | 10670.903 | 0 | 0.781 | 21341.806 | 1 |
| النموذج * الطول | 0.001 | 2 | 0.001 | 0.588 | 0.555 | 0 | 1.177 | 0.149 |
| الخطأ | 6.431 | 5994 | 0.001 | | | | | |
| الكلى | 476.295 | 6000 | | | | | | |
| الكلى المصحح ا | 29.434 | 5999 | | | | | | |

(a) محسـوبة باسـتخدام ألفا ٠,٠٥؛ b ر التربيعية = ٠,٧٨٢ (ر التربيعية المعدلة = ٠,٧٨١)

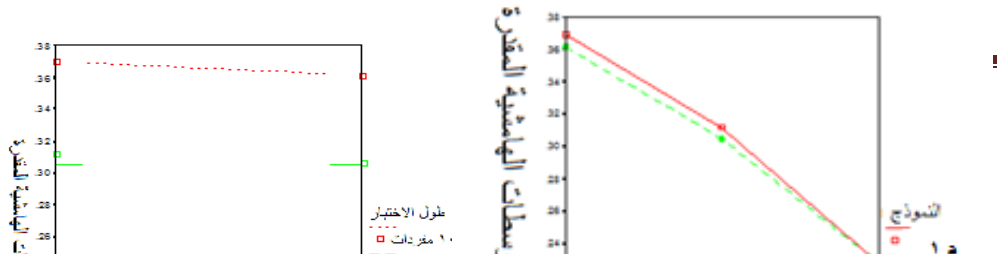
(E-mail of the correspondig: jcps@art.bsu.edu.eg)

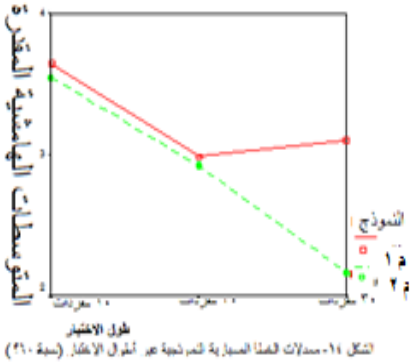
للإجابة على السؤال البحثي الثالث، أشار تحليل المعايير الأولى إلى أن الاختبارات المكونة من ٣٠ مفردة ذات النسب الأعلى من المفردات متعددة الاستجابة بغض النظر عن طول الاختبار أدت إلى معدلات ثبات هامشي أعلى. لا توجد فروق ذات دلالة إحصائية في حسابات الثبات المجموعتين من النماذج.

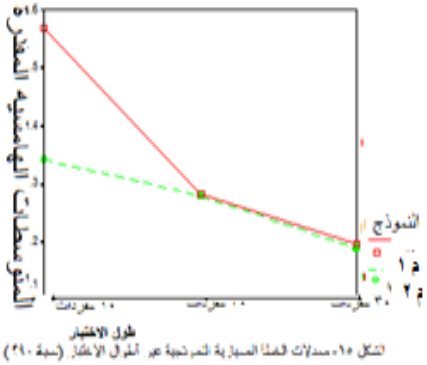
أشارت المعايير الثانية، تحليلات الأنوفا ANOVA لمعدلات الخطأ المعيارية، إلى أن التأثيرات الرئيسية لطول ونموذج الاختبار كانت دالة لجميع مجموعات التصميم و٨ من أصل ١١ اختبارًا للتفاعل. يوضح الشكلان ١٠ و١١ العلاقة النموذجية بين العوامل لمعظم النماذج. أشارت تحليلات أنوفا ANOVAs إلى أن الاختبارات ذات المفردات المختلطة قد حصلت على معدلات خطأ معيارية لثيتا والتي انخفضت مع زيادة طول الاختبار من ١٠ إلى ٢٠ ثم إلى ٣٠ مفردة. تظهر النتائج أيضًا أن معدلات الخطأ تميل إلى أن تكون مختلفة بشكل كبير بالنسبة للمجموعتين من نماذج التحليل في جميع أطوال الاختبار. ومع ذلك، يجب النظر إلى هذه النتيجة في ضوء حقيقة أن حجم التأثير كان ضئيلًا.

ومع ذلك، يجب النظر إلى هذه النتيجة في ضوء حقيقة أن حجم التأثير كان ضئيلًا. يعرض الشكلان ١٢ و١٣ بيانياً العلاقة بين ٦٠٪ و٩٠٪ من المستويات النسبية التي اختلفت في الدلالة عن المستويات الأخرى عبر النماذج. يشير كلا الرقمين إلى أن معدل الخطأ لكل من المجموعتين من نماذج التحليل يختلف بشكل دال عن طول اختبار واحد. في كلتا الحالتين، كان أداء المجموعة الثانية أفضل من المجموعة الأولى لحزمة راش Rasch في تقليل معدلات الخطأ المعيارية لتقديرات ثيتا. يوضح الجدول ٢٥ تقديرات حجم التأثير لجميع مستويات النسب. يُظهر النمط المعروض عبر تقديرات حجم التأثير زيادة في الأخطاء المعيارية. هذه الزيادة ناتجة عن طول الاختبار وزيادة في نسبة التصحيح متعدد الاستجابة.

كانت المعايير الثالثة التي تم تحليلها لتقييم الأمثلية هي مستوى المعلومات عبر أطوال الاختبار والنماذج ومستويات النسبة. بالرجوع إلى الجداول ٦ و٧ و٨، لوحظ أنه بشكل عام، أدت نظرية الاستجابة للمفردة IRT الى معلومات إجمالية أعلى عبر جميع مستويات النسب وحسّنت كمية المعلومات التي تم جمعها مع زيادة طول الاختبار. أيضًا، تم إنتاج كميات أكبر من المعلومات مع زيادة نسبة المفردات متعددة الاستجابة.






**جدول (٢٥). مقارنة حجم التأثير عبر نتائج الخطأ القياسية**

| توليفة التصميم | النموذج | الطول | تفاعل النموذج فى الطول |
|---|---|---|---|
| %0 | 0.003 | 0.624 | 0.001 |
| %10 | 0.002 | 0.669 | 0.001 |
| %20 | 0.007 | 0.656 | 0.005 |
| %30 | 0.002 | 0.608 | 0.012 |
| %40 | 0.004 | 0.625 | 0.002 |
| %50 | 0.011 | 0.673 | 0.001 |
| %60 | *0.137 | *0.430 | *0.160 |
| %70 | 0.008 | 0.739 | 0.002 |
| %80 | 0.003 | 0.759 | 0.002 |
| %90 | *0.591 | *0.915 | *0.710 |
| %100 | 0.016 | 0.781 | 0 |

* تختلف أحجام التأثير للنموذج ٦٠٪ و٩٠٪ عن نمط المخرجات الأخرى لأنها تعكس توزيعات بيانات الاختبار غير العادية

**السؤال الرابع: هل توجد فروق بين مجموعتي التصحيح: أساليب المجموعة الأولى في التحليل ( RASCHنموذج راش ونموذج التقدير الجزئيPCM )والمجموعة**

(E-mail of the correspondig: jcps@art.bsu.edu.eg)

**الثانية ( PL ٢ النموذج ثنائي البارامتر ونموذج الاستجابة المتدرجة(GRM في قدرتها على تحليل الاختبارات ذات المفردات المختلطة؟**

تم تقييم خمسة معايير لمقارنة أساليب المجموعة الأولى (نموذج راش للاستجابات الثنائية ونموذج التقدير الجزئي للاستجابات المتعددة)، والمجموعة الثانية (النموذج ثنائي البارامتر للاستجابات الثنائية ونموذج الاستجابة المتدرجة للاستجابات المتعددة) وهي افتراضات لكل نموذج (إحصائيات المطابقة)، وجودة تقدير الدرجة الحقيقية (تقديرات ثيتا وحجم الأخطاء المعيارية)، والثبات والمعلومات.

البيانات أكثر ملاءمة للمجموعة الثانية (النموذج ثنائي البارامتر للاستجابات الثنائية ونموذج الاستجابة المتدرجة للاستجابات المتعددة). الذي سمح للتمييز بالتنوع. تم تقييم الإحصاء 2LL–، وهو تقدير لإحصاء مربع كاي، للنموذج الملائم للبيانات من خلال مقارنة الفرق $\chi 2$ المرصودة مع $\chi 2$ الحرجة. وكما هو واضح من الجدول ٢٦، توجد فروق دالة بين مطابقة البيانات لنماذج مجموعتين التحليل.

## جدول (٢٦). سلبي ضعف نسبة احتمالية التسجيل (-LL٢)

| المقردات | % البولى | 1م LL2- | L2 LL2- | χ2 الفرق | د.ح | p | مطابقه افضل |
|---|---|---|---|---|---|---|---|
| 10 | %0 | 4943.5- | 5017.5- | 74 | 9 | 0.001 | 1م |
| 10 | %10 | 3573.1- | 3650- | 76.9 | 9 | 0.001 | 1م |
| 10 | %20 | 2355.4- | 2423.6- | 68.2 | 9 | 0.001 | 1م |
| 10 | %30 | 1203.4- | 1247.5- | 44.1 | 9 | 0.001 | 1م |
| 10 | %40 | 11.3 | 6.5- | 17.8 | 9 | 0.05 | 2م |
| 10 | %50 | 714.5 | 677.2 | 37.3 | 9 | 0.001 | 2م |
| 10 | %60 | 1213.3 | 1194.7 | 18.6 | 9 | 0.05 | 2م |
| 10 | %70 | 1843.5 | 1848.1 | 4.6 | 9 | 0.9 | مشابه |
| 10 | %80 | 3221.1 | 3230.9 | 9.8 | 9 | 0.5 | مشابه |
| 10 | %90 | 4891.3 | 4875.6 | 15.7 | 9 | 0.1 | مشابه |
| 10 | %100 | 5438.6 | 5419 | 19.6 | 9 | 0.05 | 2م |
| 20 | %0 | 2323.8 | 2156.1 | 167.7 | 19 | 0.001 | 1م |
| 20 | %10 | 3853 | 3713 | 140 | 19 | 0.001 | 2م |

(E-mail of the correspondig: jcps@art.bsu.edu.eg)

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 2a | 0.001 | 19 | 179.7 | 6559.3 | 6739 | %20 | 20 |
| 1a | 0.001 | 19 | 81.1 | 7738 | 7819.1 | %30 | 20 |
| 2a | 0.001 | 19 | 163.3 | 9339.7 | 9503 | %40 | 20 |
| 2a | 0.001 | 19 | 143.9 | 11496.9 | 11640.8 | %50 | 20 |
| 1a | 0.001 | 19 | 112.8 | 13949.6 | 14062.4 | %60 | 20 |
| 2a | 0.001 | 19 | 44.9 | 16300 | 16344.9 | %70 | 20 |
| 2a | 0.001 | 19 | 50.5 | 17482.7 | 17533.2 | %80 | 20 |
| 1a | 0.02 | 19 | 34.7 | 19525.4 | 19560.1 | %90 | 20 |
| 2a | 0.001 | 19 | 89.5 | 21676.3 | 21765.8 | %100 | 20 |
| 1a | 0.001 | 29 | 318.5 | 6819.5 | 7138 | %0 | 30 |
| 2a | 0.001 | 29 | 297.5 | 10891.7 | 11189.2 | %10 | 30 |
| 2a | 0.001 | 29 | 254.3 | 13371 | 13625.3 | %20 | 30 |
| 1a | 0.001 | 29 | 216.3 | 16174.5 | 16390.8 | %30 | 30 |
| 2a | 0.001 | 29 | 138.1 | 18811.4 | 18949.5 | %40 | 30 |
| 2a | 0.001 | 29 | 197.6 | 22459.7 | 22657.3 | %50 | 30 |
| 1a | 0.001 | 29 | 156.4 | 24347.8 | 24504.2 | %60 | 30 |
| 2a | 0.001 | 29 | 116.6 | 27760.9 | 27877.5 | %70 | 30 |
| 1a | 0.001 | 29 | 1787.8 | 32970.7 | 31182.9 | %80 | 30 |
| 2a | 0.001 | 29 | 1803.3 | 31182.9 | 32986.2 | %90 | 30 |
| 2a | 0.001 | 29 | 96.5 | 36465 | 36561.5 | %100 | 30 |

## مناقشة النتائج:

### تقديرات القدرة:

في هذه الدراســـة، تمت مقارنـة تقديرات ثيتا عبر نماذج تحليل نظرية الاســـتجابة للمفردة IRT عبر ٣٣ شـكلا مختلفًا من أشـكال الاختبار (ثلاثة أطوال اختبار في أحد عشـر نسـباً ثابتة من المفردات متعددة الاسـتجابة وثنائية الاسـتجابة). تم تطبيق نماذج نظرية الاســـتجابة للمفردة IRT على كل مجموعة من مجموعات بيانات الاســـتجابة البالغ عددها ٣٣ لتوليد تقديرات القدرة. ثم تمت مقارنة تقديرات القدرة هذه باسـتخدام أسـلوب تحليل التباين ANOVA العالمية ذات التأثيرات الثابتة. تمت مقارنة التأثيرات الرئيســـية والتفاعلية لنموذج التحليل وطول الاختبار. لوحظت نتائج دالة للاختبارات

ذات النسب العالية من المفردات متعددة الاستجابة. تأثيرات النموذج والتفاعل، بشكل عام، لم تكن دالة.

توجد فروق في تقديرات القدرة عبر أطوال الاختبار عندما تم تصحيح ٦٠ % إلى ١٠٠ % من المفردات بشكل متعدد الاستجابة. يجب النظر إلى هذه الفروق قي ضوء حقيقة أن حجم التأثير على تقديرات القدرة كان ضئيلاً، وعلى مدى المستويات الخمسة من للنسبة، لم ينتج عن طول معين بشكل ثابت تقديرات مختلفة للقدرة. لذلك

بالنسبة للاختبارات ذات المفردات المختلطة، لا يمكن التوصل إلى استنتاجات قوية من هذه التحليلات بأن تقديرات القدرة تختلف بشكل دال بغض النظر عن طول الاختبار أو نموذج التحليل أو نسبة المفردات في الاختبار الذي تم تصحيحه بشكل متعدد الاستجابة. هذه النتيجة هامة عند أخذها في الاعتبار مع النتائج التي توصل اليها ساقي (Si,2002) بأن نماذج الاستجابة للمفردة تختلف بشكل دال في قدرتها على استعادة تقديرات القدرة الأصلية. إلا أن تحليلات ساقي Si لم تكن في الاختبارات ذات المفردات المختلطة ولكن على مجموعات استجابة ثنائية الاستجابة أو متعددة الاستجابة. كما أن نماذج نظرية الاستجابة للمفردة IRT التي تمت مقارنتها هنا أدت إلى تقديرات قدرة مماثلة لأن أشكال وضع الدرجات للاختبار كانت مختلطة.

**معلومات الاختبار:**

أظهرت هذه الدراسة أنه في الاختبارات ذات المفردات المختلطة، فإن طول الاختبار ونسبة المفردات التي تم تصحيحها بشكل متعدد الاستجابة لهما علاقات مباشرة مع كمية معلومات الاختبار المتاحة. بشكل عام، تتيح الاختبارات الأطول التي تحتوي على نسب أكبر من المفردات متعددة الاستجابة مزيدًا من المعلومات عبر توزيع مستويات القدرة. هذا ملحوظ بشكل خاص في مستويات القدرة المنخفضة حيث كانت أعلى المعلومات دائمًا تقريبًا عند مستوى نسبة ١٠٠٪ للمفردات متعددة الاستجابة. بالنسبة للطلاب ذوي المستويات الأعلى من القدرة، تميل المعلومات الأعلى إلى أن تكون على مستويات متناسبة مع درجات متعددة الاستجابة أقل من ١٠٠٪. نتجت أعلى معلومات اختبار إجمالية عن تصحيح متعدد الاستجابة بنسبة ١٠٠٪ عبر جميع أطوال الاختبار وعبر كلا نموذجي التحليل المطبقين في هذه الدراسة.

تشير الأدبيات (Ayanwale & Adeleke,2020) الى أن المفردات متعددة الاستجابة تسمح للممتحنين بالحصول على تقدير جزئي لإثبات المعرفة الجزئية. تميل المفردات

(E-mail of the correspondig: jcps@art.bsu.edu.eg)

ثنائية الاستجابة إلى تجاهل وجود معرفة جزئية بدلاً من تحديد جميع الاستجابات في سياق "الكل أو لا شيء"(Ayanwale,2021 ). تسمح الاستجابات متعددة الاستجابة باستمرارية القياس لمستويات القدرة، وبالتالي توفر فرصة لجمع المزيد من المعلومات من النطاق المتوسط للقدرة على كل مفردة. من الواضح أن هذا المدخل، بدوره، سيسمح بجمع المزيد من المعلومات. يجب أن يخدم الوعي بهذه النتيجة مصممي الاختبارات الذين يسعون جاهدين لقياس القدرة الحقيقية للممتحنين بشكل أكثر دقة من خلال تصميم مفردات اختبار أكثر فعالية.

بالنسبة للمفردات متعددة الاستجابة بشكل خاص، فإن التكاليف التي يتكبدها وقت تطبيق الاختبار، وكتابة المفردات، والاختبار الميداني للمفردات وتصحيح المفردات، جنبًا إلى جنب مع التطبيق العملي لأطوال الاختبار، تعتبر مهمة في موازنة الرغبة في الحصول على مستويات معلومات عالية. ومع ذلك، فقد تم البحث في التطورات التكنولوجية الحديثة، مثل تصحيح المقالات الإلكترونية ( Shermis & Berstein, 2003; Shermis, Koch, Page, Keith & Harrington, 2002) وتبين أنها فعالة بنفس القدر إن لم تكن أكثر فعالية من التصحيح اليدوي.

**التحسين:**

يعتمد تحسين الاختبار على تعظيم كل من افتراضات النموذج الإحصائي ومعايير التحسين لهذا النموذج، والتي تتضمن التوافق بين النموذج والبيانات. حدد الإحصاء الملائم الذي تم تحليله (LL٢-) أن البيانات تتلاءم مع النماذج المختارة من نظرية الاستجابة للمفردة IRT بشكل أفضل في اختبار ١٠ مفردات مع تصحيح متعدد الاستجابة بنسبة ٤٠٪. أظهرت نماذج المجموعة الثانية (النموذج ثنائي البارامتر ونموذج الاستجابة المتدرجة) مطابقة أفضل لمعظم مجموعات البيانات عند مقارنتها بنماذج المجموعة الأولى.

تم تحليل معدلات الخطأ المعيارية لتقديرات القدرة وفقًا لمعايير التحسين وتم الحصول على نتائج دالة عبر طول الاختبار بأحجام تأثير كبيرة. كما لوحظت التأثيرات على الأخطاء المعيارية من تأثيرات النموذج وتفاعل طول النموذج تلو الآخر لكن أحجام التأثير لكليهما كانت صغيرة. مع استثناءات قليلة، اختلفت معدلات الخطأ القياسية في جميع أطوال الاختبار وكلا النموذجين. انخفضت الأخطاء المعيارية مع زيادة طول الاختبار مما أظهر علاقة غير مباشرة أو عكسية. بالنسبة لهذه المقارنات، زاد حجم تأثير طول الاختبار على الخطأ المعياري مع زيادة نسبة

المفردات متعددة الاستجابة. كان هذا صحيحًا لجميع أطوال الاختبار الثلاثة وكانت قيم القدرة العالية موجودة أيضًا.

تبين زيادة نتائج الثبات عبر أطوال الاختبار ومستويات النسبة. حقق الاختبار المكون من ٣٠ مفردة عند التصحيح المتعدد الاستجابة بنسبة ١٠٠٪ أعلى معدلات الثبات عبر كلا النموذجين. تتوافق هذه النتيجة مع دراسة لو ووينج ( Lau and Wang,1998) اللذين وجدا أن الجمع بين نوعين من المفردات في تقييم واحد يعزز ثبات التقييم. وتجدر الإشارة إلى أن مستويات المعلومات قدمت مقاييس أكثر دالة لتحسين الاختبار ذي المفردات المختلطة.

**المقارنة بين المجموعتين من نماذج التحليل لنظرية الاستجابة للمفردة IRT:**

بشكل عام، لم يتم ملاحظة فروق ذات دلالة إحصائية بين أداء المجموعتين على التحليلات في هذه الدراسة. بشكل عام عبر تحليلات المعلومات في هذه الدراسة، أدى استخدام تحليل المجموعة الثانية من النماذج في الاختبارات ذات المفردات المختلطة إلى تحسين طفيف في معلومات الاختبار المتاحة مقارنة بالمجموعة الأولى من النماذج عبر نطاق القدرات لمعظم أطوال الاختبار ومستويات النسبة للتصحيح متعدد الاستجابة. يبدو أيضًا أن المجموعة الثانية استفادت أكثر من الزيادة في طول الاختبار فيما يتعلق بالمعلومات التي تم جمعها. الاختلافات في الثبات وإحصائيات الملاءمة بين المجموعتين لا تذكر. وتتفق هذه النتيجة مع ما توصل اليها بيرجر Berger (1998). تميل معدلات الخطأ المعيارية إلى أن تكون أعلى بالنسبة للمجموعة الاولى في جميع أطوال الاختبار التي تم تحليلها في هذه الدراسة.

**بحوث مقترحة**

لا توجد طريقة تحليل فائقة معروفة لتحليل الاختبار ذي المفردات المختلطة. بحثت هذه الدراسة في نماذج نظرية الاستجابة للمفردة IRT المستخدمة في الدراسة وكيفية مقارنتها عبر ثلاثة أطوال اختبار ومجموعة من أشكال المفردات المختلطة. كانت تحليلات هذه الدراسة كافية في إنتاج النتائج التي تناولت أسئلة البحث المطروحة هنا، ومع ذلك، هناك حاجة إلى مزيد من التحليل في المجالات التالية:

- النظر في صعوبة الاختبار عبر مجموعة متنوعة من خيارات التصحيح المختلطة،

(E-mail of the correspondig: jcps@art.bsu.edu.eg)

• النظر في درجات اختبار الممتحن عبر مجموعة متنوعة من خيارات التصحيح المختلطة،

• تطوير طرق أكثر ملاءمة لتحليل ثبات ومطابقة النماذج الإحصائية للاختبارات ذات المفردات المختلطة،

• اختبار قوة النماذج عند انتهاك الافتراضات (السرعة والتخمين) في الاختبار ذات المفردات المختلطة،

• متابعة استخدام التصحيح الآلي للمفردات متعددة الاستجابة.

• البحث في المزيج المثالي بين المفردات ثنائية الاستجابة ومتعددة الاستجابة التي تعطي درجات اختبار تتسم بدرجة عالية من الثبات والصدق.

• بوجه عام، يجب إجراء مزيد من البحث في مجالات صعوبة الاختبار، ودرجات اختبار الممتحنين، والتصحيح الآلي للتقدير الجزئي جنبًا إلى جنب مع مقارنة مع المقاييس النفسية التقليدية الأخرى وكيفية معالجتها للتحديات المتعلقة بالاختبارات ذات المفردات المختلطة.

**المراجع:**

صلاح الدين محمود علام (٢٠٠٥). *نماذج الاستجابة للمفردات الاختبارية احادية البعد ومتعددة الابعاد وتطبيقاتها في القياس النفسي والتربوي*، الطبعة الاولى، القاهرة :دار الفكر العربي.

صلاح الدين محمود علام (٢٠١٨). *الاختبارات والمقاييس التربوية والنفسية*. الأردن، دار الفكر.

محمد طالب دبوس (٢٠١٦). استخدام نظرية الاستجابة للفقرة في بناء فقرات اختبار محكي المرجع في الرياضيات بفقرات ثنائية الاستجابة ومتعددة الاستجابة وفق النموذج اللوجستي ثنائي المعلم. **مجلة جامعة النجاح للأبحاث - العلوم الإنسانية**، مج٣٠, ع٧، ص ص ١٤٥٣ – ١٤٨٠

محمد محمد فتح الله (٢٠١١). تكامل مدخلي القياس محكي المرجع **CRM** ونظرية الاستجابة للمفردة **IRT** في تقييم فاعلية برنامج تدريبي لتنمية كفايات بناء

الاختبارات التحصـــيلية لدى معلمي التعليم قبل الجامعي. **مجلة التربية**، جامعة الأزهر - كلية التربية، ع١٤٦, ج٦، ٤٥٩ - ٥٢٥

Adegoke, B. A. (2013). Comparison of Item Statistics of Physics Achievement Test using Classical Test and Item Response Theory Frameworks. ***Journal of Education and Practice***, *4*(22), 87-96.

Asriadi, M., & Hadi, S. (2021). Implementation of item response theory at final exam test in physics learning: Rasch model study. ***Proceedings of the 6th International Seminar on Science Education*** *(ISSE 2020)*, *541*(Isse 2020), 336–342. https://doi.org/10.2991/assehr.k.210326.048

Ayanwale, M.A. & Adeleke, J.O. (2020). Efficacy of Item Response Theory in the Validation and Score Ranking of Dichotomous Response Mathematics Achievement Test. Bulg. J. Sci. Educ.Policy, 14(2):260–285. Accessed: May 31, 2021. Available: ***https://www.academia.edu/45182779/***

Ayanwale, M. A. (2021). Calibration of Polytomous Response Mathematics Achievement Test Using Generalized Partial Credit Model of Item Response Theory. *EDUCATUM* ***Journal of Science, Mathematics and Technology***, *8*(1), 57-69. https://doi.org/10.37134/ejsmt.vol8.1.7.2021

Berger, M. P. (1998). Optimal design of tests with dichotomous and polytomous items. ***Applied Psychological Measurement***, 22(3) 248-258.

Bond, T. G., & Fox, C. M. (2015). *Applying the Rasch model:* ***Fundamental measurement in the human sciences*** (3rd). Routledge.

Bond, T.G., Yan, Z., & Heene, M. (2020). *Applying the Rasch Model****: Fundamental Measurement in the Human Sciences*** *(4th Ed.)*. New York: Routledge.

Bozdağ, H. C., & Türkoğuz, S. (2021). A rasch model analysis of primary school students'conceptual understanding levels of the concept of light.

(E-mail of the correspondig: jcps@art.bsu.edu.eg)

*IOJPE: **International Online Journal of Primary Education***, *10*(1), 160–179.

Calderón, C.; Beyle, C.; Véliz-García, O.; Bekios-Calfa, J. (2021). Psychometric properties of Addenbrooke's Cognitive Examination III (ACE-III): An item response theory approach. *PLoS ONE*, 16, e0251137.

Carlson, J. E. (1996, April). *Information provided by polytomous and dichotomous items on certain NAEP instruments.* Paper presented at the annual meeting of the American Educational Research Association, New York, NY.

Che Lah, N. H., Tasir, Z., & Jumaat, N. F. (2021). Applying alternative method to evaluate online problem-solving skill inventory (OPSI) using Rasch model analysis. ***Educational Studies***, *00*(00), 1–23. https://doi.org/10.1080/03055698.2021.1874310

Cohen, R. (2018) Psychological Testing and Assessment*: An **Introduction to Tests and Measurement.*** 9th edition, McGraw-Hill Education, 2 Penn Plaza, New York, NY 10121

Donoghue, J. R. (1994). An empirical examination of the IRT information of polytomously scored reading items under the generalized partial credit model. ***Journal of Educational Measurement***, 31(4) 295-311.

Downing, S. M. (2002). Construct-irrelevant variance and flawed test questions: Do multiple-choice item-writing principles make any difference? ***Academic Medicine,*** *77*(10), s103-104.

Ercikan, K., Schwarz, R. D., Julian, M. W., Burket, G. R., Weber, M. M. & Link, V. (1998). Calibration and scoring of tests with multiple-choice and constructed-response item types. ***Journal of Educational Measurement***, 35(2), 137-154.

Erguven, M. (2014). Two Approaches to Psychometric Process: Classical Test Theory and Item Response Theory. ***Journal of Education***, *2*(2), 23-30.

Hambleton, R. K. (1989). Principles and selected applications of item response theory. R.L. Linn (Ed.), ***Educational Measurement:*** Third Edition, (pp. 147-200). American Council on Education and Macmillan Publishing Company.

Hambleton, R.K.; Swaminathan, H.; Rogers, H.J. (1991). **Fundamentals of Item Response Theory**; Sage: Thousand Oaks, CA, USA.

Kaplan, R. (2012). **Psychological Testing: Principles**, **Applications,** and Issues 8th Edition. Wadsworth, Cengage Learning.

Kline, T. J. B. (2005). *Psychological testing: A practical approach to design and evaluation.* Sage Publications, Inc.

Linn, R.L. (1990). Has item response theory increased the validity of achievement test scores? ***Applied Measurement in Education***, 3(2), 115-141.

Marnat, G and Wright, M (2016). **Handbook of psychological assessment** Sixth Edition. John Wiley & Sons, Inc.

Masers, G.N. (1982). A rasch model for partial credit scoring. ***Psychometrika*** 47(2), 149-174.

Murphy, P. (2021). **Score Interpretation Guide for Educators.** State of New Jersey, Department of Education

Nunnally, J., & Bernstein, I.H. (1994**). Psychometric Theory** *(3rd Ed.)*. New York: McGraw-Hill.

Oliveri, M. E., and M. von Davier (2014). Toward increasing fairness in score scale calibrations employed in international large-scale assessments, ***International Journal of Testing,*** 14(1): 1-21.

Plake, B. (1993). Applications of educational measurement: Is optimum optimal? ***Educational Measurement****: Issues and Practices*, 5-10.

Rahayu, W., Putra, M. D. K., Rahmawati, Y., Hayat, B., & Koul, R. B. (2021). Validating an indonesian version of the what is happening in this class? (WIHIC) questionnaire using a multidimensional rasch

(E-mail of the correspondig: jcps@art.bsu.edu.eg)

model. *International Journal of Instruction*, *14*(2), 919–934. **https://doi.org/10.29333/iji.2021.14252a**

Rezapour, M.; Cuccolo, K.; Veenstra, C.; Ferraro, F.R. (2021). An Item Response Theory to Analyze the Psychological Impacts of Rail-Transport Delay. *Sustainability*, 13, 6935. https://doi.org/10.3390/ su13126935

Si, C. B. (2002). **Ability estimation under different item parameterization and scoring models**. Unpublished doctoral dissertation, University of North Texas.

Stemler, Steven E. and Naples, Adam (2021). Rasch Measurement v. Item Response Theory: Knowing When to Cross the Line," Practical Assessment, Research, and Evaluation: Vol. 26, Article 11. Available at: **https://scholarworks.umass.edu/pare/vol26/iss1/11**

Stewart, J., Zabriskie, C., Devore, S., & Stewart, G. (2018). Multidimensional item response theory and the force concept inventory**.** ***Physical Review Physics Education Research***, *14*(1), 10137. https://doi.org/10.1103/PhysRevPhysEducRes.14.010137

Thissen, D. (1991). **MULTILOG User's Guide: Multiple, Categorical item Analysis and Test Scoring using item Response Theory***; version 6.0.* Scientific Software International Chicago: Illinois.

Thissen, D., Nelson, L. and Swygert, K. (2001). Item response theory applied to combinations of multiple-choice and constructed-response items- Approximation method for scaled scores. In Thissen, D. & Wainer, H. (Eds**.),** ***Test Scoring***, (pp. 293-341). Mahwau, New Jersey: Lawrence Erlbaum Associates, Inc.

Triwik, J. (2021) Rasch analysis on item response theory: Review of model suitability. AIP Conference Proceedings **2326**, 020017 (2021); https://doi.org/10.1063/5.0040305 Published Online: 08 February 2021

Tseng, M. C., & Wang, W. C. (2021). The q-matrix anchored mixture rasch model. ***Frontiers in Psychology,*** *12*(March), 1–9. https://doi.org/10.3389/fpsyg.2021.564976

van der Linden, W.(2016). **Handbook of Item Response Theory** Volume 2: Statistical Tools. Chapman and Hall/CRC.

Van der Linden, W. J. (Ed.). (2018). **Handbook of Item Response Theory**, *Three Volume Set*. CRC Press.

von Davier, M., Yamamoto, K., Shin, H. J., Chen, H., Khorramdel, L., Weeks, J., Davis, S., Kong, N., & Kandathil, M., (2019). *Evaluating item response theory linking and model fit for data from PISA 2000–2012*. ***Assessment in Education: Principles, Policy, and Practice***, *26*(4), 466-488. doi:10.1080/0969594X.2019.1586642

Wainer, H. & Thissen, D. (1993). Combining multiple-choice and constructed- response test scores: toward a marxist theory of test construction. ***Applied Measurement in Education***, 6(2), 103-118.

Wang W, Drasgow F and Liu L (2016) Classification Accuracy of Mixed Format Tests: A Bi-Factor Item Response Theory Approach. ***Front. Psychol***. 7:270. doi: 10.3389/fpsyg.2016.00270

Wright, B. D., & Stone, M. H. (1979). **Best test design: Rasch measurement**. Chicago, IL: Mesa Press.

(E-mail of the correspondig: jcps@art.bsu.edu.eg)

## Abstract

This study investigated the effects of test length, scoring schema and item analysis model on the ability estimates, test information levels and optimization criteria of mixed item format tests. SPSS software across ability estimates and standard errors of ability estimates using a 3 x 11 x 2 fixed factorial ANOVA was used to make the comparison between two sets of models (first group consists of Rasch model and Partial Credit Model, second group consists of Two Parameter Logistic Model and Graded Response Model of item analysis procedures. Effect sizes and power were reported for each procedure. Scheffe post hoc procedures were conducted on significant factors. Test information was analyzed and compared across the range of ability levels for all 66-design combinations. The results indicated that both test length and the proportion of items scored polytomously had a significant impact on the amount of test information produced by mixed item format tests. Generally, tests with 100% of the items scored polytomously produced the highest overall information. This seemed to be especially true for examinees with lower ability estimates. Optimality comparisons were made between IRT models procedures based on standard error rates for the ability estimates, marginal reliabilities and fit indices (-2LL). The only significant differences reported involved the standard error rates for both two sets procedures. This result must be viewed in light of the fact that the effect size reported was negligible.

**Keywords:** test length, scoring schema, item analysis model, the ability estimates, test information function, optimization criteria, mixed item format tests.

(E-mail of the correspondig: jcps@art.bsu.edu.eg)