أثر اختلاف طريقة انتقاء المفردات على دقة تقدير القدرة والكفاءة النسبية للاختبار التحصيلي

Impact of Difference in Item Selection on the Accuracy of Ability Estimation and the Relative Efficiency of Achievement Test

الباحثة

شرف بنت حامد الأحمدي

أستاذ القياس والتقويم المساعد

كلية التربية- جامعة طيبة

ملخص البحث:

هدفت الدراسة إلى مقارنة أثر اختلاف طريقة انتقاء المفردات على دقة تقدير القدرة والكفاءة النسبية للاختبار التحصيلي، وتم إعداد اختبار تحصيلي في مقرر التقويم التربوي مكون من (٦٠) مفردة من نوع الاختيار من متعدد، وتم تطبيقه على (٤٠٩) طالبة من كلية التربية، وتم تدريج المفردات باستخدام النموذج اللوغارتمي ثلاثي المعلم، وتم اشتقاق ثلاثة اختبارات من الاختبار الكلي بانتقاء مفرداتها وفقاً لطريقة:(أقصى قدرة، وأقصى معلومات، والتمييز التقليدي)، ويتكون كل منها من (٢٠) مفردة، وتم تدريج كل منها بالنموذج اللوغارتمي الثلاثي المعلم. وكان من أهم النتائج، ما يلي:

أن الاختبارات الثلاثة تتكافأ في خصائصها السيكومترية ومؤشراتها الاحصائية مع الاختبار الكلي.

لا توجد فروق دالة بين تقديرات القدرة المقدرة بالاختبار الكلي وتقديرات القدرة المقدرة بالاختبارات الثلاثة.

توجد فروق دالة في دقة تقديرات القدرة لصالح الاختبار الثاني (أقصى معلومات)، والثالث (التمييز التقليدي) بالنسبة للاختبار الأول (أقصى قدرة).

لا توجد فروق دالة في دقة تقديرات القدرة بين الاختبار الثاني (أقصى معلومات)، والثالث (التمييز التقليدي).

تم ترتيب الاختبارات بحسب الكفاءة النسبية من الأعلى دقة على النحو التالي: الاختبار الثاني (أقصى معلومات)، ثم الاختبار الثالث (التمييز التقليدي)، وأخيراً الاختبار الأول (أقصى قدرة)، مع ملاحظة وجود تقارب كبير بين مؤشرات الاختبار الثاني (أقصى معلومات)، والاختبار الثالث (التمييز التقليدي).

الكلمات المفتاحية: طرق انتقاء المفردات، دالة المعلومات، مؤشر التمييز التقليدي، دقة تقدير القدرة، الكفاءة النسبية.

Abstract:

The study aimed at comparing impact of difference in item selection procedures on ability estimation precision and Relative Efficiency of achievement test. An achievement test composed of (60) multiple choices items set within Educational Evaluation course and applied to (409) student at faculty of Education. Item calibration was done using Three-Parameters Logistic Model. Three tests were derived from the overall test upon theta at maximum information (θmax), maximum information (I(θmax) and classical discrimination (Dclss). Each one composed of (20) items that were calibrated by (3PLM) model.

Results:

The three tests are equivalent in psychometric properties, statistical indicators and (θ estimates) with the overall test.

No differences in ability estimation between the overall test and three tests.

There are differences in the accuracy of ability estimation for the second test (I(θmax) and third (Dclss) comparing first (θmax).

No differences in accuracy of ability estimation between the second test (I(θmax) and third (Dclss).

Tests were arranged upon values of (RE) as of the highest accuracy as follows: the second (I(θmax), third (Dclss) and lastly the first (θmax) with wide convergence between the indicators of the second (I(θmax) and third (Dclss).

Key Words: Item Selection Procedures, Information Function, Classical Discrimination Index, Accuracy of Ability Estimation, Relative Efficiency.

المقدمة والخلفية النظرية:

لقد تطورت مفاهيم القياس النفسي تطوراً كبيراً وقد ساعد ظهور نظريات، مثل: نظرية القياس التقليدية ClassicalTest Theory ونظرية الاستجابة للمفردة Item Response Theory في ذلك التطوير من خلال استحداث أساليب متنوعة وملائمة لمعظم الأغراض التي يتضمنها موقف القياس بدءاً بمرحلة بناء الاختبار مروراً بتحليل المفردات وصولا للتفسيرات واتخاذ القرار حول الموضوع المستهدف بالقياس، وكان من بين الأساليب المتنوعة التي قدمتها النظريتان في مجال تجويد الاختبارات التربوية أساليب وإجراءات انتقاء المفردات _في الاختبارات الخطية الورقية أو المحوسبة أو الاختبارات التكيفية المحوسبة _ من بين تجمع للمفردات التي تم إعدادها لقياس سمة نفسية أو عقلية أو مهارية(البرصان، ٢٠١٨؛ Yao، ٢٠١٩)، حيث يركز القائمون ببناء الاختبارات اهتمامهم المبدئي على صياغة مفردات تغطي مجال السمة، ومن ثم التحقق من تغطيتها للنطاق المستهدف بالأساليب التحكيمية الكيفية والكمية لضمان صدق محتوى كل مفردة وصولا لصدق الاختبار ككل، وبعد تجريب الاختبار أو تطبيقه لأول مرة يتم تحليل مفرداته بالأساليب الإحصائية، بهدف الكشف عن مؤشرات المفردات للاحتفاظ بالجيد منها وحذف المفردات الغامضة أو المربكة أو بالغة الصعوبة أو السهولة والتي لا تسهم بأي تباين في درجات الاختبار ولا تميز بين مستويات القدرة المستهدفة بالتصنيف(Price, 2016)، ومن ثم يأتي دور طرق انتقاء المفردات في انتقاء مفردات معينة من تجمع المفردات الكلي الذي تم إعدادها، وهذا الانتقاء يأتي بهدف اشتقاق اختبارات لأغراض معينة، مثلاً: للتشخيص، أو للتسكين، بحيث تكون تلك الاختبارات المشتقة تقدم تقدير دقيق للقدرة في المدى المستهدف من القدرة (Murphy et al., 2010 ; Yao, 2019). أي أن طرق انتقاء المفردات تعمل على استبعاد بعض المفردات من بين نطاق المفردات التي تم تجريبها وإيداعها في تجمع المفردات، وذلك بحسب الهدف من انتقاء المفردات(الربايعة، ٢٠١٩)، أيضاً تُفيد طرق انتقاء المفردات في التعرف على الأسباب التي تجعل المفردة غير صالحة لهدف ما، واستبقاء المفردات التي تفي بخصائص تحقق الغرض من استخدام الاختبار في مجالات التقويم المختلفة، كما تساهم في خفض عدد مفردات الاختبار إلى عدد أقل من المفردات تقدم تقديرات مماثلة للقدرة المقاسة بالاختبار الكامل(Lunz, & Bergstrom, 1994؛ Halkitis, 1998؛ Kalender, 2012; Eggen & Straetmans, 2000

الشافعي، ٢٠١٢)، وقد تعددت طرق انتقاء المفردات في إطار النظرية التقليدية ومن أشهرها طريقة معامل التمييز، ومعامل الصعوبة (الصبحي، ٢٠٠١؛ الدلالعة، ٢٠٠٤)، أما بالنسبة لنظرية الاستجابة للمفردة فهي متنوعة ولكن من أشهرها طرق انتقاء المفردات المتعلقة بدالة معلومات المفردة؛ ممثلة بأقصى قيمة لمعلومات المفردة (هياجنة،٢٠١١؛ الشافعي، ٢٠١٢؛ Yao, 2019)، وبمستوى القدرة المقابل لأقصى معلومات (هياجنة، ٢٠١١)، مع الأخذ في الاعتبار أن قدرة الفرد من العوامل المهمة عند انتقاء المفردات، ويتم ذلك بانتقاء المفردة ذات الصعوبة المقابلة لمستوى القدرة، وعند ما تكون هناك أكثر من مفردة عند مستوى قدرة معين يتم تحديد معيار آخر للإنتقاء ما بين تلك المفردات، مثل: دالة المعلومات، أو قيمة معلم التمييز (البرصان، ٢٠١٨).

وقد تم تطبيق طرق انتقاء المفردات في العديد من الدراسات سواء عند إعداد الاختبارات التكيفية المحوسبة أو الاختبارات الخطية المحوسبة أو التقليدية، منها على سبيل المثال: دراسة (Eggen & Straetmans, 2000؛ والصبحي، ٢٠٠١؛ Gierl et al., 2001 ؛ والدلالعة، ٢٠٠٥؛ والعمري، ٢٠٠٥؛2010 Murphy et al., ؛ وهياجنة، ٢٠١١؛ والشافعي، ٢٠١٢؛ Kalender, 2012؛ ومرشود، ٢٠١٤؛ والموسوي،٢٠١٦؛ ومحمد، ٢٠١٦؛ والجمل وآخرون، ٢٠١٧؛ والبرصان، ٢٠١٨؛ وسخيم، ٢٠١٨؛ والربايعة، ٢٠١٩؛ Yao, 2019).

وقد اهتمت بعض تلك الدراسات بمقارنة طرق انتقاء المفردات من حيث فعاليتها في إعطاء تقديرات أدق لمعالم المفردات والأفراد، ومنها على سبيل المثال: دراسة تشن وأنكينمان Chen & Ankenman (2004) التي قارنت بين ثلاث طرق من طرق أقصى معلومات (فيشر، فيشر بتوزيع بعدي، دالة كولباك بتوزيع بعدي)، والطريقة العشوائية، وأظهرت أن طرق أقصى معلومات على اختلافها قدمت اختبارات جيدة حتى مع الاختبارات الأقل من (١٠) مفردات.

ودراسة الدلالعة (٢٠٠٥) التي قارن فيها بين ثلاث طرق لانتقاء مفردات اختبار في تأثيرها على قيمة دالة المعلومات، وهي القدرة المناظرة للقيمة القصوى للمعلومات، والطريقة العشوائية، والطريقة الكلاسيكية (أعلى تمييز). وأظهرت النتائج ان الاختبارات التي تم انتقاءها وفق طريقة القيمة القصوى للمعلومات تعطي أكبر قيمة لدوال معلومات الاختبارات تليها الطريقة الكلاسيكية (أعلى تمييز) ثم الطريقة العشوائية. في حين كانت أكبر قيمة للخطأ المعياري في تقدير القدرة هي للاختبار الذي اعتمد على الطريقة العشوائية وأقلها للاختبار

الذي اعتمد على القدرة المناظرة للقيمة القصوى للمعلومات. كما أظهرت النتائج أن الكفاءة النسبية للاختبار الذي اعتمد على القيمة القصوى للمعلومات كانت أكبر ما يكون عند مستوى القدرة القصوى مقارنة مع الطريقة العشوائية.

وأيضاً قارن هياجنة (٢٠١١) بين أثر ثلاث طرق لانتقاء المفردات على دالة معلومات الاختبار، وهي طريقة: قيمة أقصى قدرة، وقيمة أقصى معلومات عند مستوى قدرة محدد، والطريقة العشوائية. وأظهرت النتائج ان الاختبارات التي تم انتقاءها وفق طريقة أقصى معلومات والطريقة العشوائية تزيد من قيمة دوال معلومات الاختبار أكثر من طريقة أقصى قدرة. كما أوضحت النتائج أن أخطاء تقدير معالم مفردات الاختبار التي تم انتقاءها وفق طريقة أقصى معلومات والطريقة العشوائية أقل من أخطاء تقدير معالم مفردات الاختبار وفق طريقة أقصى قدرة.

وطبق الشافعي (٢٠١٢) طريقة أقصى معلومات لانتقاء تسعة اختبارات بأطوال مختلفة تراوحت ما بين (٩) إلى (٧٢) مفردة من اختبار أوتيس لينون للذكاء المستوى المتوسط، المكون من (٨٠) مفردة واستطاع أن يختصر أطوال الاختبار إلى عدد أقل من المفردات تصل إلى (٣٦) وتقدم نفس تقديرات القدرة التي تم الحصول عليها من الاختبار الأصلي. في هذا الصدد أيضاً قام الجراح (٢٠١٥) باستقصاء أثر حجم العينة وصعوبة المفردة وتمييزها على دالة المعلومات والخطأ المعياري حسب النموذج الثنائي المعلم في نظرية الاستجابة للمفردة. وأظهرت النتائج أنه لا توجد فروق دالة احصائياً بين متوسطات قيم دالة المعلومات تعزى لمتغير حجم العينة، في حين أظهرت النتائج فروق دالة احصائياً بين متوسطات قيم دالة المعلومات تعزى لمتغيري معامل صعوبة وتمييز المفردة. كما أظهرت النتائج وجود علاقة عكسية بين نسبة الخطأ المعياري وكلٍ من حجم العينة ومعامل التمييز، وعلاقة طردية مع معامل الصعوبة عند ما يتم تقدير القدرة (θ) باستخدام الاختبار.

كما أجرى البرصان (٢٠١٨) دراسة هدفت إلى تقصي فاعلية خمس استراتيجيات ثانوية لاختيار مفردات الاختبارات التكيفية متعددة المراحل إضافة للاستراتيجية الرئيسة القائمة على صعوبة المفردة وتواؤمها مع قدرة المفحوص، هذه الاستراتيجيات الخمس تعتمد على: دالة المعلومات، معلم التمييز، أقل معلم تخمين، إحصائي المطابقة، والاختيار العشوائي. وقد أظهرت النتائج أن الاستراتيجية المعتمدة على إحصائي المطابقة هي الأفضل في كل المراحل

الثلاث تليها طريقة الاختيار العشوائي ثم معلم التمييز. ولم تأخذ الاستراتيجيات القائمة على دالة المعلومات ومعلم التخمين أفضلية في كافة المراحل.

مما سبق؛ ونظرا لما توفره نظرية الاستجابة للمفردة ونماذجها المختلفة في تدريج الاختبارات من عدالة وموضوعية ودقة القياس، ولما لطرق انتقاء المفردات سواء القائمة على افتراضات النظرية التقليدية أو نظرية الاستجابة للمفردة من أهمية في الكشف عن المفردات الأكثر دقة في قياس السمة، ولكون الباحثة لم تجد أي دراسة قارنت بين طرق: (أقصى قدرة، وأقصى معلومات، والتمييز التقليدي)، كما لم تجد أي دراسة سابقة استخدمت برنامج Xcalibre عند تدريج تجمع المفردات أو عند إيجاد المؤشرات التي على أساسها تتم إجراءات الانتقاء، حيث أن برامج التحليل الاحصائي تختلف في دقة تقدير معالم المفردة والفرد (Wang, 2018) وقد أكد كل من وايس ومينين (٢٠١٢) Weiss & Minden إلى أن قيم معالم المفردات والأفراد المقدرة ببرنامج Xcalibre أقرب إلى القيم الحقيقية مقارنة ببرنامج Bilog mg، عليه فإن هذه الدراسة ستحاول استكمال جهود الدراسات السابقة من خلال إجراء مقارنة بين ثلاث طرق لانتقاء المفردات، طريقتان منها تعتمد على دالة معلومات المفردة وفقاً لنظرية الاستجابة للمفردة وهما: متغير مستوى القدرة المقابل لأقصى معلومات للمفردة، ومتغير أقصى معلومات للمفردة، وطريقة تعتمد على مؤشرات المفردة وفقاً للنظرية التقليدية وهي التمييز التقليدي لاستقصاء أثر تلك الطرق على دقة تقدير القدرة ورفع الكفاءة النسبية للاختبار.

طرق انتقاء المفردات:

تهدف طرق انتقاء المفردات وإجراءاتها إلى رفع مستوى دقة الاختبار في قياس القدرة المستهدفة بالقياس، وتلك الطرق تتدرج إلى قسمين بحسب النظرية التي تتبنى مفاهيمها وتوظف إجراءاتها في عملية الانتقاء، وهما النظرية التقليدية في القياس ونظرية الاستجابة للمفردة، وسيتم شرح الطرق في إطار كل من النظريتين بشي من الايجاز، فيما يلي:

أولاً: طرق انتقاء المفردات وفقاً للنظرية التقليدية في القياس:

النظرية التقليدية في القياس تقوم على أساس نظري قوي يشمل مفاهيم نظرية وبُنى جبرية دعمت مفاهيم القياس وبنيت عليها جميع التطبيقات العملية للنظرية، مثل تقدير السمات والقدرات، والخصائص السيكومترية لأدوات القياس، ويشار إليها أيضا بنظرية "الدرجة الحقيقية"، والنظرية تقوم على افتراض جوهري يتمثل في أن الدرجة الملاحظة للفرد على

اختبار ما يمكن اعتبارها تتألف من جزأين، هما: الدرجة الحقيقية التي تعبر عن مستوى الفرد، بالإضافة إلى أخطاء القياس التي تتعلق بطبيعة موقف القياس ومكوناته، علماً بأن خطأ القياس يمكن أن يكون بالزيادة أو النقصان (كروكر وألجينا، ٢٠١٧) ولها عدة افتراضات، وهي أن: الدرجة الحقيقية للفرد لابد أن تكون قيمة ثابتة لأنها تمثل القدرة الحقيقية للفرد، وخطأ القياس المتوقع للفرد عند ما يطبق عليه اختبار ما عدد لانهائي من المرات يساوي صفر. ونسبة الارتباط بين الدرجة الحقيقية للفرد على اختبار ما مع خطأ القياس يساوي صفر، والدرجة الحقيقية لا يمكن معرفتها وقياسها إنما يتم الاستدلال عليها أو تقديرها من خلال الدرجة الكلية للفرد على اختبار ما، ويمكن تقدير الدرجة الحقيقية للفرد من خلال الوسط الحسابي للدرجات الكلية للفرد على اختبار ما عند تطبيقه عدداً لانهائي من المرات(Price, 2016)، وعندما يتم تطبيق اختبارين متوازيين لنفس السمة فإن الارتباط بين خطأ القياس بين الاختبارين يساوي صفر (كروكر وألجينا، ٢٠١٧)، ولا يوجد ارتباط بين الدرجات الحقيقية وأخطاء القياس لعدد من الأفراد على نفس الاختبار، أي أنه ليس من الضروري أن تكون أخطاء القياس للأفراد من ذوي الدرجات الحقيقية العالية أقل منها عند ذوي الدرجات الحقيقية المتدنية (النعيمي، ٢٠١٥)، وأن خطأ القياس هو خطأ عشوائي، وإن الخطأ المنتظم متعلق بصدق الاختبار، إلا أن الخطأ العشوائي هو الذي يحدد دقة المقياس أو ما يسمى بثبات درجات الاختبار (النعيمي، ٢٠١٥ ؛ مرشود، ٢٠١٤)، ويتساوى تباين خطأ القياس لجميع الأفراد الذين يطبق عليهم الاختبار(كروكر وألجينا، ٢٠١٧).

وتعتمد معظم طرق انتقاء المفردات التي تتبع النظرية التقليدية على نتائج التحليل الاحصائي للاستجابة الملاحظة على المفردة، للتعرف على الخصائص السيكومترية لكل مفردة، مثل معامل: (الصعوبة، والتمييز، وفاعلية المشتتات، والتخمين)، لكشف المفردات غير الجيدة التي لا تحقق هدف الاختبار مثلا؛ من حيث درجة الصعوبة أو الغموض، وبناء على المؤشرات التي يتم جمعها حول كل مفردة يتم تحسين جودة الاختبار من خلال عملية الانتقاء حيث يتم اتخاذ قرار بالاحتفاظ بالمفردة أو تحسينها أو حذفها وصولا لاختبار تتسم درجاته بالصدق والثبات في قياس السمة أو القدرة المستهدفة (محمد، ١٩٩٩)، وبما أنه في الدراسة الحالية سيتم تطبيق اختبار تحصيلي بمفردات ثنائية التقدير(٠، ١) من نوع الاختيار من متعدد، لذا سيتم التطرق لطرق انتقاء المفردات وفقاً لمؤشرات مفردة الاختيار من متعدد، والتي من أهمها:(الصعوبة، والتمييز) وسيتم شرحها بإيجاز، وهي كما يلي:

صعوبة المفردة: تُعبر عن نسبة الذين أجابوا إجابة صحيحة، من المجموع الكلي وهي نسبة تتراوح قيمتها من (صفر) إلى(١)، وكلما زادت قيمة النسبة المئوية كانت المفردة أكثر سهولة (نيتكو وبروخارت،٢٠١٢)، وتختلف النسبة المناسبة باختلاف طبيعة الاختبار هل هو من نوع الإجابات المفتوحة؟ أم من نوع الاختيار من متعدد؟ وكم عدد البدائل في كل مفردة اختيار من متعدد؟، أيضاً تتوقف على الغرض من الاختبار هل صمم لأغراض الفحص أو للمفاضلة (الترشيح) لمنح دراسية؟، أو لتقييم التمكن؟، وإذا تم افتراض أن حساب مؤشر الصعوبة يهدف للوصول إلى تقييم دقيق حول مدى انجاز الفرد عند مستوى قدرة معينة، فإن المفردات التي تقترب قيمة معامل صعوبتها من (١) أو(٠) لا تقدم أي معلومات حول قدرة الفرد لأنها لا تسهم بأي معلومات حول الفروق الفردية، بينما أنه كلما اقتربت قيمة معامل الصعوبة من القيمة (٠.٥٠) فإنها ستسهم بأقصى معلومات حول الفروق بين الافراد لأنها ستساهم في رفع قيمة معامل التمييز بينهم عند جميع مستويات القدرة، وستساهم بفعالية في تباين درجات الاختبار مما سينعكس على صدق وثبات الدرجات، أي أنه من الأفضل على العموم انتقاء المفردات التي تنتشر قيم مؤشر الصعوبة لها حول المتوسط (انستازي ويوريان، ٢٠١٥).

أما مؤشر تمييز المفردة: فهو يمثل الفرق بين نسبة الذين أجابوا عن السؤال إجابة صحيحة في المجموعة العليا، ونسبة الذين أجابوا عن السؤال إجابة صحيحة في المجموعة الدنيا (نيتكو وبروخارت،٢٠١٢)؛ وقد تم تطوير ما يزيد عن خمسين مؤشر مختلف لتمييز المفردات وتختلف في كون بعضها يستخدم في الاختبارات الثنائية الاستجابة وبعضها مع المتصلة، كما أن تلك الطرق تختلف في الإجراءات والافتراضات إلا أن معظم مؤشرات تمييز المفردات تقدم نتائج متماثلة تقريباً، ومن أشهر الطرق في حالة الاختبارات ثنائية الاستجابة؛ كما ذكر سابقاً يتم عن طريق مؤشر التمييز Index of Discrimination بحساب الفرق بين نسبة الذين أجابوا إجابة صحيحة في المجموعتين العليا والدنيا، وأيضاً عن طريق حساب الارتباط بمعامل Phi Coefficient، وهاتان الطريقتان تقدم نتائج متحيز للمفردات متوسطة الصعوبة وتعطي معها أكبر قيمة للارتباط (انستازي ويوريان، ٢٠١٥)، ويتم حسابه أيضاً عن طريق الارتباط الثنائي المتسلسل Biserial Correlation؛ الذي يعبر عن معامل الارتباط بين الدرجة على المفردة والدرجة الكلية على الاختبار باعتبار الدرجة الكلية كتقدير للقدرة، ويبين كيفية إسهام كل مفردة في الاختبار ككل (مهرنز ولهمن، ٢٠٠٣؛ محمد، ١٩٩٩)، خاصة إذا تم التحقق من صدق التكوين الفرضي للاختبار وأنه يقيس سمة واحدة عندها يتم

اعتبار الدرجة الكلية محك مناسب لإنتقاء المفردات وهذه الطريقة تقدم نتائج تمييز مستقلة عن صعوبة المفردة (أنستازي ويوريان، ٢٠١٥). ويفيد مؤشر التمييز في التعرف على مدى قدرة المفردة في التفريق بين الأفراد الذين حصلوا على درجات عالية (المجموعة العليا)، ووالافراد الذين حصلوا على درجات منخفضة (المجموعة الدنيا)، وتتراوح قيمة هذا المؤشر من (+١) إلى (−١) (نيتكو وبروخارت، ٢٠١٢)، ويجب أن تكون قيمة التمييز مرتفعة وإيجابية ولا تقل عن القيمة (٠.٢٠+)، وقيمة التمييز السالبة قد تشير إلى أخطاء في المفردة، أو إلى تدريس دون المستوى أو إلى المصادفة، كما أن هناك توصية في حالة كون محتوى الاختبار يتكون من عدة موضوعات فيجب انتقاء المفردات التي تميز ضمن كل موضوع وليس على مستوى الاختبار ككل، بمعنى أن يتم الاختيار مع أخذ متغير تصنيف الموضوعات إلى جانب التمييز، لضمان توازن المحتوى للاختبار (مهرنز ولهمن، ٢٠٠٣).

مما سبق؛ وعلى الرغم من سهولة حساب مؤشرات المفردة التقليدية خاصة عند تطوير الاختبارات في المجال التربوي والاكاديمي إلا أنه لابد من ملاحظة أن مؤشرات المفردة في النظرية التقليدية تتأثر بخصائص عينة الأفراد التي يطبق عليها الاختبار، مما يَحُد من إمكانية تعميم النتائج على عينات أخرى، كما أن المقارنة بين الأفراد في مستوى القدرة تقتصر فقط على تطبيق نفس المفردات أو مفردات مكافئة، أي أنه لا يمكن المقارنة بين مستويات القدرة إذا أجاب الأفراد على مفردات مختلفة(كروكر وألجينا، ٢٠١٧). بالإضافة إلى ما أشار إليه مرشود (٢٠١٤) من أن طرق انتقاء المفردات في ضوء النظرية التقليدية تفترض تساوي أخطاء القياس لجميع الأفراد وهذا غير منطقي حيث أنه يحدث فروق بين الافراد في اتساق الإجابة قد يرجع لمستوى القدرة التي يقيسها الاختبار أو لعوامل أخرى، وهذا يبرر استخدام طرق بديلة أو مكملة لطرق النظرية التقليدية كطرق نظرية الاستجابة للمفردة التي قدمت إجراءات تساهم في تلافي نواحي القصور التي تضمنتها إجراءات النظرية التقليدية.

ثانياً: طرق انتقاء المفردات وفقاً لنظرية الاستجابة للمفردة:

تعتبر نظرية الاستجابة للمفردة من الاتجاهات الحديثة في مجال القياس النفسي لمعالجة المفردات وتصميم الاختبارات، وهي لا تعتمد عند تقدير درجات الأفراد على عدد الإجابات الصحيحة، انما تعتمد على نمط إجابة الفرد بناء على دالة رياضية مستندة على نظرية الاحتمالات تسمى دالة الاستجابة للمفردة (Item Response Function) (محمد، ٢٠١٦)، وتتميز النظرية ببعض المزايا الرئيسة من أهمها خاصية اللاتغاير (Invariance) وتعني بقاء خاصية معينة ثابتة بتغير المعطيات التي تكونها (Price, 2016). فهي تفترض أن تقديرات قدرة الأفراد متحررة من خصائص المفردات التي تم استخدامها لتقدير القدرة، كما أن تقديرات معالم المفردات متحررة من خصائص الأفراد الذين تم استخدام استجاباتهم في تقدير المعالم، كما أنها تُمكن من الحصول على معامل احصائي يقيس مدى دقة تقدير قدرة كل فرد من خلال مفردات الاختبار (Hambleton & Swaminathan,1985)، كما أنه لابد من التحقق من افتراضات نظرية الاستجابة للمفردة كي يمكن الإفادة من تطبيقاتها المختلفة حيث ترتكز على مجموعة من الافتراضات ترتبط بالبيانات المتعلقة بالاختبارات، ويتم بناءً عليها تحديد النموذج الملائم للتحليل وهي، افتراض: (١) أحادية البعد: ويقصد به وجود قدرة واحدة تفسر أداء الفرد على الاختبار، (٢) الاستقلال الموضعي: ويقصد به أن تكون استجابات الفرد على مفردات مختلفة في الاختبار مستقلة استقلالاً احصائياً بحيث لا تؤثر استجابة الفرد لإحدى المفردات على استجابته لمفردات أخرى، (٣) منحنى خصائص المفردة: وهو دالة رياضية متزايدة تربط بين احتمال نجاح الفرد في إجابته عن مفردة ما وبين القدرة التي تقيسها مجموعة المفردات التي يشتمل عليها الاختبار، (٤) التحرر من السرعة: ويقصد به أن سبب فشل الفرد في إجابته عن مفردة ما ناتج عن عدم قدرته عن الإجابة على هذه المفردة وليس بسبب عدم وصوله لهذه المفردة نتيجة لقصر زمن الاختبار (علام، ٢٠٠٥).

ونظرية الاستجابة للمفردة تملك عدة نماذج تختلف بحسب طبيعة بيانات تقدير الدرجات، وتنقسم النماذج لقسمين، هي: نماذج ثنائية التدريج (Dichotomous)، أو نماذج متعددة التدريج (Polytomous)، ومن أشهر نماذج النظرية ثنائية التدريج، النماذج اللوغارتمية المعلمية (الأحادي، والثنائي، والثلاثي)، وفي الدراسة الحالية لكون طبيعة البيانات ثنائية والمفردة من نوع اختيار من متعدد سيتم تطبيق النموذج اللوغارتمي ثلاثي المعلم

Three-Parameters Logistic Model (3PLM)، ويمكن التعبير رياضيا عنه بالمعادلة رقم (١)

:(Hambleton & Swaminathan,1985)

$$P_i(\theta) = c_i + (1 - c_i)\frac{e^{Da_i(\theta - b_i)}}{1 + e^{Da_i(\theta - b_i)}} \ldots (i = 1,2,3,\ldots,n) \ldots (1)$$

حيث إن:

$P_i(\theta)$: هي احتمال أن يجيب أي فرد يتم اختياره عشوائياً ذو قدرة (θ) إجابة صحيحة على المفردة (i).

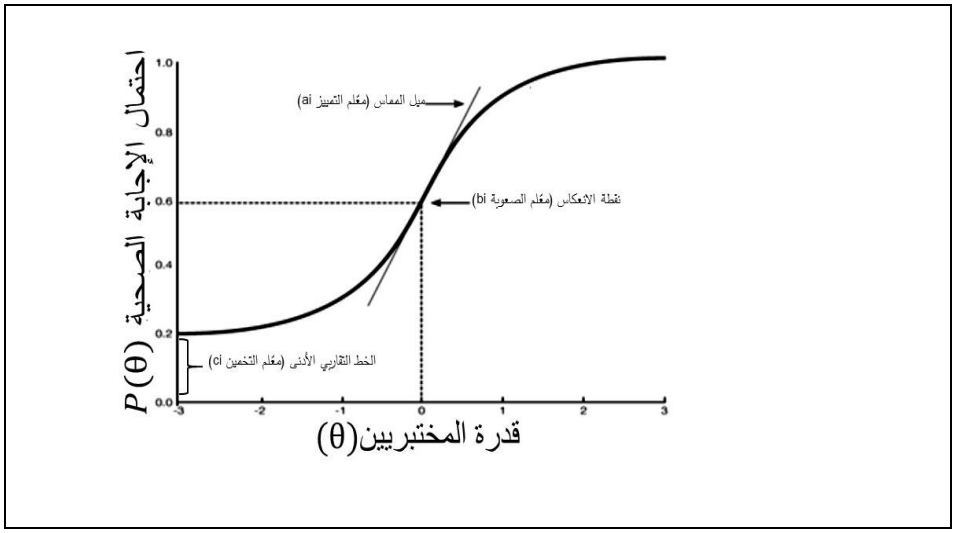
D: عامل التدريج يساوي 1.7.

b_i: معلم صعوبة المفردة i.

a_i: معلم تمييز المفردة i.

c i: معلم التخمين للمفردة i.

وبيانياً بالشكل رقم (١) الذي يوضح منحنى خصائص المفردة (ICC) Item Characteristic Curve للنموذج اللوغارتمي الثلاثي المعلم، والذي يعبر عن احتمالات الإجابة الصحيحة عن المفردة عند مستويات القدرة المختلفة، حيث يتضح من الشكل أن النموذج يفترض أن المفردة يمكن تقديرها من خلال ثلاثة معالم، وهي: معلم الصعوبة (bi) وهو يعبر عن نقطة انعكاس على متصل القدرة (θ) وتتراوح قيمته العملية من (+٣) إلى (−٣)؛ ومعلم التمييز (ai) هو ميل المماس لمنحنى خصائص المفردة عند نقطة الانعكاس (bi) وتتراوح قيمتها العملية من (صفر) إلى (+٢)؛ ومعلم التخمين (ci) وهو الخط التقاربي لمنحنى خصائص المفردة، وهو يعبر عن قدرة المختبر للوصول للإجابة الصحيحة عن طريق التخمين العشوائي (بيكر، ٢٠١٠؛ الشافعي، ٢٠١٢) وتتراوح قيمته العملية ما بين (صفر) إلى (٠.٣٠)، أما قدرة الفرد (θ) فهي تعبر عن اللوغارتم الطبيعي المرجح وتتراوح قيمتها العملية من (+٣) إلى (−٣)، مع ملاحظة أن القيم النظرية لكل من معلم القدرة والصعوبة والتمييز تتراوح ما بين (+∞) إلى (−∞)، بينما معامل التخمين تتراوح قيمته النظرية ما بين (صفر) إلى(١) (Hambleton, Swaminathan,1985).



شكل رقم (١): منحنى خصائص المفردة للنموذج اللوغارتمي الثلاثي المعلم (Fotaris&Mastoras,2014)

وتعتمد طرق انتقاء المفردات القائمة على نظرية الاستجابة للمفردة على خصائص المنحنى المميز للمفردة من خلال إمكانية تقدير احتمالات الإجابة الصحيحة على المفردة عند جميع مستويات القدرة (θ) بدلالة معالم المفردة (ai, bi, ci) ودالة معلومات المفردة Item Information Function (IIF) وهي تعبر عن دالة رياضية تصف العلاقة بين مستوى القدرة ومقدار المعلومات التي تقدمها المفردة (حرز الله، ٢٠٠٤)، وذلك اذا تمت مطابقة البيانات مع النموذج المستخدم وتحققت خاصية اللاتغاير، إلا أنه في حالة تطبيق الاختبارات تكون تلك المعالم مجهولة ويكون المعلوم فقط الاجابة التي قدمها كل مختبر عن كل مفردة من تلك المفردات، ولكي يتم تقدير القدرة يتم تطبيق مجموعة من الطرق الإحصائية، مثل تقدير: الارجحية القصوي (Maximum Likelihood)، والارجحية القصوى المشتركة (Maximum Likelihood Joint)، والارجحية القصوى المشروطة (Conditional Maximum Likelihood)، والارجحية القصوى الهامشية (Marginal Maximum Likelihood) (هياجنة، ٢٠١١)، والأساليب البيزية (Bayesian techniques) وتشمل، الأساليب البيزية التالية: أوين (Owen)، والبعدي المتوقع (expected a posteriori)، والبعدي الأقصى (maximum a posteriori) (نور الدين، ٢٠٠٢؛ علام، ٢٠٠٥) فتنتج تقديرات احتمالية للقدرة وهنا يصبح من الممكن تحديد كمية المعلومات (أقصى ارتفاع لمنحنى

دالـة معلومـات المفـردة) عنـد أي مسـتوى للقـدرة (θ). ونتيجـة لـذلك يمكـن حسـاب مقـدار المعلومات بناء على مفردة واحدة في أي مستوى للقدرة ويرمز لها بالرمز $\mathbf{I_i(\theta)}$ حيث تشير i إلـى المفـردة، وبمـا أن الاختبـار يتكـون مـن عـدد مـن المفـردات فـإن دالـة المعلومـات للاختبار(Test Information Function) تمثل حاصل جمع المعلومات التي تقدمها تلك المفردات عند كل مستوى من مستويات القدرة، وسيكون المستوى العام لدالـة معلومـات الاختبـار أعلـى مـن دالـة معلومـات المفـردة الواحـدة (بيكـر، ٢٠١٠) والمعادلـة رقـم (٢) تمثـل **معادلة دالة معلومات الاختبار:**

$$\mathbf{I(\theta) = \sum_{i=1}^{n} I_i(\theta) \dots\dots (2)}$$

حيث أن:

$\mathbf{I(\theta)}$ مقدار معلومات الاختبار عند مستوى القدرة($\mathbf{\theta}$)

$\mathbf{I_i(\theta)}$ مقدار معلومات المفردة i. عند مستوى القدرة($\mathbf{\theta}$)

$\mathbf{n}$ عدد مفردات الاختبار

أي أنـه يمكـن التوصـل لدالـة المعلومـات للاختبـار ككـل مـن خـلال قيـاس فعاليـة الاختبـار فـي التمييـز بـين مسـتويات القـدرة لـدى الأفـراد. وتنـتج مـن مجمـوع دوال المعلومـات للمفردات التي يتكون منها الاختبار، ويمكن حسابها من خلال التعويض في المعادلة رقم(٢) بالصيغة الرياضية في المعادلة رقم(٣)(Hambleton & Swaminathan,1985):

$$\mathbf{I(\theta) = \sum_{i=1}^{n} \frac{[p_i'(\theta)]}{[P_i(\theta)Q_i(\theta)]} \dots\dots (3)}$$

حيث أن:

$\mathbf{I(\theta)}$: هي دالة معلومات الاختبار.

(θ): معلم القدرة للمفحوصين.

$\mathbf{p_i(\theta)}$: دالة استجابة المفردة i.، وتعبر عن احتمال الإجابة الصحيحة

$\mathbf{p_i'(\theta)}$: المشتقة الأولى لدالة استجابة المفردة i..

$\mathbf{Q_i(\theta) = 1 - pi\ (\theta)}$ وتعبر عن احتمال الإجابة الخاطئة

ويتم تعريف مقدار المعلومات للمفردة بحسب النموذج المستخدم، حيث يأخذ منحنى دالة المعلومات للمفردة الشكل الجرسي، أي أن دالة معلومات المفردة، تُمثل منحنى يبين كمية المعلومات التي تملكها المفردة عند جميع مستويات القدرة (محمد، ١٩٩٩؛ هياجنة، ٢٠١١)، وتقوم المفردة بقياس القدرة بدقة أعظم في مستوى القدرة (θ) المقابل لصعوبة المفردات (bi) (بيكر، ٢٠١٠). والأمر مختلف بالنسبة للنموذج اللوغارتمي الثلاثي المعلم حيث أن أقصى قيمة لدالة المعلومات في ضوءه لا يمكن إيجادها مباشرة كما في النموذج الأحادي والثنائي ولكنها تعبر عن الدالة المقابلة لأقصى قدرة (θmax) من خلال المعادلة رقم(٤) :(Hambleton & Swaminathan,1985)

$$\theta_{max} = b_i + \frac{1}{D_{a_i}} \ln\left[\frac{1}{2} + 1/2\sqrt{1+8c_i}\right] \dots\dots\dots (4)$$

حيث أن:

(θ_{max}) قيمة أقصى معلومات القدرة (θ) والذي تقدم المفردة أقصى معلومات لها عند هذه النقطة من متصل القدرة

b_i: معلم الصعوبة للمفردة i.

a_i: معلم تمييز المفردة i.

c i: معلم التخمين للمفردة i.

ومن المعادلة السابقة رقم (٤) يتضح أن أقصى قيمة لدالة المعلومات ثابتة للنموذج الأحادي، أما النموذج الثنائي فإنها تتناسب طردياً مع مربع قيمة معلم التمييز (ai)، أما للنموذج اللوغارتمي الثلاثي المعلم، فيمكن التعبير عنها رياضياً بالمعادلة رقم(٥) :(Hambleton & Swaminathan,1985)

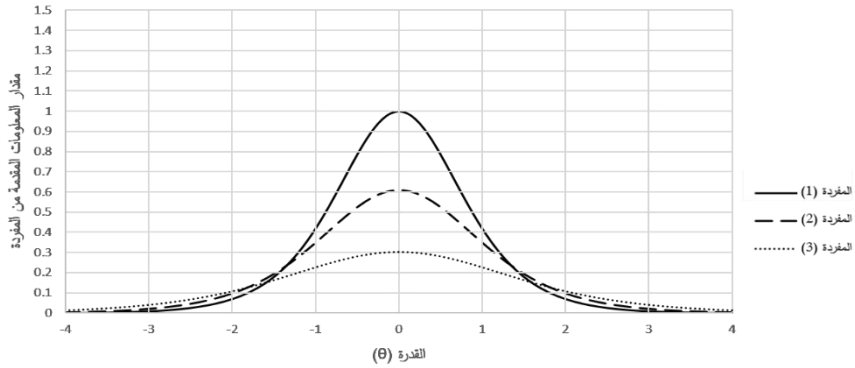
$$I(\theta, u_i)_{max} = \frac{D^2 a_i^2}{8(1-c_i^2)}\left[1 - 20c_i - 20c_i^2 + (1+8c_i)^{\frac{3}{2}}\right] \dots (5)$$

يتضح من المعادلة رقم (٥) علاقة دالة معلومات الاختبار بمعالم كل مفردة من المفردات التي يتكون منها الاختبار. ويمكن تفسير هذه العلاقة نظرياً، كما يلي:

علاقة دالة المعلومات للمفردة بمعلمة الصعوبة (bi) : تقوم المفردة بقياس القدرة أو السمة الكامنة بأعلى دقة في مستوى القدرة (θ) المناظر لمعلم صعوبة المفردة(bi)، وينخفض مقدار المعلومات التي تقدمها المفردة كلما ابتعد مستوى القدرة عن صعوبة المفردة

واقترب من الصفر في طرفي متصل القدرة أو السمة (بيكر، ٢٠١٠)، فعلى سبيل المثال؛ الشكل رقم (٢) يمثل منحنى دالة المعلومات لثلاث مفردات مختلفة ولكن متساوية في قيمة معامل الصعوبة (b1=b2=b3=0)، ويمكن ملاحظة أن تلك المفردات تعطي أعلى كمية معلومات لذوي القدرة المتوسطة ($\theta = 0$)، حيث أن مقدار المعلومات يرتفع كلما اقتربت قيمة صعوبة المفردة من مقدار قدرة الفرد، وينخفض ببعد قيمتي صعوبة المفردة ومقدار القدرة عن بعضهما (محمد، ١٩٩٩؛ Hambleton, et al. ,1991)

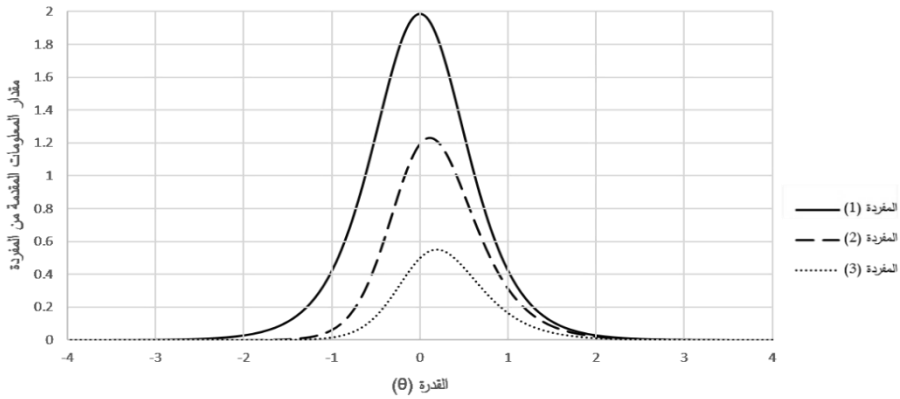
أما بالنسبة لعلاقة دالة المعلومات للمفردة بمعلمة التمييز (ai) فكلما زادت القدرة التمييزية للمفردة زاد مقدار المعلومات التي تقدمها المفردة على متصل السمة، ويقل مقدار المعلومات المقدمة من المفردة عندما تنخفض قدرتها التمييزية بين فئات القدرة (المرتفع، والمتوسط، والمنخفض)، ويمكن توضيح ذلك على سبيل المثال؛ من خلال الشكل رقم (٢) أيضاً الذي يمثل منحنى دالة المعلومات لثلاث مفردات مختلفة في قيمة معامل التمييز بحيث(a1= 2, a2= 1.5, a3=1)، ولكنها متساوية في قيمة معامل الصعوبة (b1=b2=b3=0)، ومن الشكل نجد أن تلك المفردات تعطي أعلى كمية معلومات لذوي القدرة المتوسطة ($\theta = 0$)، ويمكن أيضاً ملاحظة أن كمية المعلومات تزيد بزيادة قيمة معامل التمييز حيث بلغت قيمتها (I(θ)= 1, 0.6, 0.3) على الترتيب، أي أن قيمة دالة المعلومات ترتفع بصورة عامة كلما ارتفعت قيمة معامل التمييز (محمد، ١٩٩٩)، وهي تتناسب طردياً مع مربع قيمة معلم التمييز للمفردة (Hambleton, et al. ,1991)، أما عن علاقة دالة المعلومات للمفردة بمعلمة التخمين (ci) فمن خلال التخمين يحصل الأفراد ذوي القدرات المنخفضة على درجات أعلى من مستوى قدرتهم الحقيقي، وبذلك يؤثر التخمين على دالة المعلومات بأنه يقلل من كمية المعلومات التي تقدمها المفردة عن القدرة الحقيقية للفرد (بيكر، ٢٠١٠، ١٠٦).



شكل رقم (٢): منحنى دالة المعلومات لثلاث مفردات مختلفة في معلم التمييز (a1= 2, a2= 1.5, a3=1)، ومتساوية في معلم الصعوبة (b1=b2=b3=0) (إعداد الباحثة)

فيمكن توضيح ذلك على سبيل المثال من خلال الشكل رقم (٣) والذي يمثل منحنى دالة المعلومات لثلاث مفردات مختلفة في قيمة معامل التخمين بحيث أن (c1= 0, c2= 0.25, c3=0.60)، ولكنها متساوية في قيمة معامل الصعوبة والتمييز (a1,2,3=b1,2,3)، فيمكن ملاحظة أن تلك المفردات تعطي أعلى كمية معلومات للمفردة رقم (١) لأنها منعدمة التخمين (c1= 0)، ويمكن أيضاً ملاحظة أن كمية المعلومات تزيد بانخفاض قيمة معامل التخمين حيث بلغت قيمتها (0.5), (1.2), (2)= $I(\theta)$ على الترتيب، أي أن قيمة دالة المعلومات ترتفع بصورة عامة كلما اقتربت معلمة التخمين من الصفر، أي أن العلاقة بينهما عكسية(محمد، ١٩٩٩؛ Hambleton, et al.,1991).

يتضح مما سبق؛ أن تقدير مستويات القدرة المختلفة يتأثر بدالة المعلومات للمفردة أو للاختبار ككل؛ ويوضح الشكل السابق رقم(٣) ثلاث دوال معلومات لمستوى القدرة (θ) لثلاث مفردات منفصلة حيث تكون المفردة رقم(١) أفضل مفردة لأنها تمكنت من كمية معلومات أكبر عن القدرة أو السمة المتوفرة لدى المفحوصين تساوي(2.0) ثم المفردة رقم (٢) ورقم(٣) قدمت كمية معلومات تساوي (٠.٥، ١.٢) على الترتيب.



شكل رقم (٣): منحنى دالة المعلومات لثلاث مفردات مختلفة في قيمة معلم التخمين (c1= 0, c2= 0.25, c3=0.60) ومتساوية في معلم الصعوبة والتمييز (a1,2,3=b1,2,3) (إعداد الباحثة)

ومن الأهمية بمكان، عند تفسير دالة المعلومات للاختبار أن يتم ذلك مع الاخذ في الاعتبار العلاقة بين مقدار المعلومات وتغير تقدير القدرة والتي يمكن ترجمتها إلى مقدار الخطأ المعياري لتقدير القدرة $SE(\theta)$ ، وهو مرتبط عكسياً مع الدقة لأي نقطة على متصل القدرة (θ) (محمد، ١٩٩٩)، وهذا الخطأ يتغير بتغير مستوى القدرة(θ) وليس متساو لجميع أفراد العينة كما هو الحال في النظرية التقليدية للقياس، ويمكن التعبير عنه بالمعادلة رقم (٦) (بيكر، ٢٠١٠):

$$SE(\theta) = \frac{1}{\sqrt{I(\theta)}} \dots\dots\dots (6)$$

حيث أن:

$SE(\theta)$: مقدار الخطأ المعياري لتقدير القدرة عند مستوى القدرة (θ)

$I(\theta)$: دالة معلومات الاختبار

وتتم مقارنة دوال المعلومات للاختبارات من خلال إيجاد الكفاءة النسبية (Relative Efficiency)، من خلال المعادلة رقم (٧) (بيكر، ٢٠١٠):

$$RE(\theta) = \frac{IA(\theta)}{IB(\theta)} \dots\dots\dots (7)$$

حيث أن:

$RE(\theta)$: الكفاءة النسبية للاختبار (A) بالنسبة للاختبار (B).

$IA(\theta)$: دالة معلومات الاختبار (A)

$IB(\theta)$: دالة معلومات الاختبار (B)

وتستخدم قيم الكفاءة النسبية للاختبار للمقارنة بين دوال المعلومات لاختبارين، وذلك من أجل تقييم كفاءة اختبارين في قياسهم لنفس السمة. فبزيادة مقدار المعلومات تزيد إمكانية تقدير القدرة بدقة أعلى أي ستقترب كل التقديرات من القيم الصحيحة بطريقة مقبولة (بيكر، ٢٠١٠)، مما سبق تتضح أهمية انتقاء المفردات عند ما تُطبق طرق دقيقة تعمل على زيادة كمية المعلومات مما يسهم في رفع كفاءة الاختبار وينعكس إيجاباً على دقة تقدير القدرة.

مشكلة الدراسة:

إن الاختبارات التحصيلية الخطية؛ هي الوسيلة الأكثر استخداماً عند تحديد وتشخيص مستوى قدرة الطلاب، وإذا تم التسليم بأن الغرض الأساسي من بناء الاختبار هو تقدير قدرة الأفراد بطريقة تسهل تصنيفهم وبالتالي تقويمهم على أساسها، فإن ذلك لا يتم إلا إذا تميز الاختبار بالدقة من خلال انتقاء المفردات الجيدة الملائمة للهدف من الاختبار (مهرنز ولهمن، ٢٠٠٣).

ونتائج تحليل مفردات الاختبار التحصيلي بإيجاد مؤشراتها باستخدام نظريتي القياس أو أحداهما يساعد في الحكم على دقة الاختبار الذي تم تطبيقه على الطلاب كما يساهم في إعداد تجمع المفردات item pool للانتقاء منه لتجويد الاختبار إذا طبق في المستقبل، وقد أضاف توفر برامج الحاسوب إمكانات للحصول على مؤشرات متنوعة لانتقاء المفردات بدقة وسرعة فائقة(Wang, 2018)، وبقي المجال مفتوحاً لتضافر الجهود البحثية لمواصلة جمع الأدلة حول جودة تلك المؤشرات عند انتقاء المفردات المناسبة لتقدير القدرة محل القياس بغرض الوصول إلى الدقة المرجوة في القياس النفسي والتربوي. وعليه، ستتصدى الدراسة الحالية إلى الإجابة عن السؤال الرئيس التالي:"ما أثر اختلاف طريقة انتقاء المفردات على دقة تقدير القدرة والكفاءة النسبية للاختبار التحصيلي؟".

من خلال الإجابة عن الأسئلة الفرعية التالية:

ما مدى اختلاف الخصائص السيكومترية والإحصائية للاختبار الكلي والاختبارات الثلاثة المشتقة منه بانتقاء مفرداتها وفقاً لطريقة:(أقصى قدرة، وأقصى معلومات، والتمييز التقليدي)؟

هل توجد فروق جوهرية بين قيم تقديرات قدرة أفراد العينة التي تم تقديرها من الاختبار الكلي ومن الاختبارات الثلاثة التي تم انتقاء مفرداتها وفقاً لطريقة:(أقصى قدرة، وأقصى معلومات، والتمييز التقليدي)؟

هل تختلف دقة تقدير قدرة الأفراد تبعاً لاختلاف طريقة انتقاء مفردات الاختبار التحصيلي: (أقصى قدرة، وأقصى معلومات، والتمييز التقليدي)؟

ما مدى اختلاف الكفاءة النسبية للاختبار ككل في تقدير قدرة الأفراد لكل اختبار من الاختبارات الثلاثة التي تم انتقاء مفرداتها وفقاً لطريقة:(أقصى قدرة، وأقصى معلومات، والتمييز التقليدي)؟

أهداف الدراسة:

تهدف الدراسة إلى ما يلي:

تحديد مدى اختلاف الخصائص السيكومترية والإحصائية للاختبار الكلي والاختبارات الثلاثة المشتقة منه بانتقاء مفرداتها وفقاً لطريقة:(أقصى قدرة، وأقصى معلومات، والتمييز التقليدي).

الكشف عن الفروق بين قيم تقديرات قدرة أفراد العينة التي تم تقديرها من الاختبار الكلي ومن الاختبارات الثلاثة التي تم انتقاء مفرداتها وفقاً لطريقة:(أقصى قدرة، وأقصى معلومات، والتمييز التقليدي).

تحديد مدى إختلاف دقة تقدير قدرة الأفراد تبعاً لاختلاف طريقة انتقاء مفردات الاختبار التحصيلي: (أقصى قدرة، وأقصى معلومات، والتمييز التقليدي).

تحديد مدى اختلاف الكفاءة النسبية للاختبار ككل في تقدير قدرة الأفراد لكل اختبار من الاختبارات الثلاثة التي تم انتقاء مفرداتها وفقاً لطريقة: (أقصى قدرة، وأقصى معلومات، والتمييز التقليدي).

أهمية الدراسة:

الأهميـة النظريـة؛ تتحدد في كون الدراسة توضح إجراءات انتقاء مفردات الاختبار التحصيلي بثلاث طرق، وهي:(أقصى قدرة، وأقصى معلومات، والتمييز التقليدي)، وتقدم معلومات حول مدى فعاليتها في دقة تقدير قدرة الأفراد والكفاءة النسبية للاختبار وقد تسفر الدراسة عن نتائج حول جودة هذه مؤشرات مما قد يُعد إضافة في مجال طرق انتقاء المفردات.

الأهميـة التطبيقيـة؛ تسعى الدراسة إلى تحديد الأهمية النسبية لثلاثة طرق لانتقاء المفردات من الاختبار الكلي(تجمع المفردات)، وهي طريقة:(أقصى قدرة، وأقصى معلومات، والتمييز التقليدي) للكشف عن أدق هذه الطرق في تقدير قدرات الأفراد بنفس دقة الاختبار الكلي، ليتم تطبيقه في اشتقاق صور مختلفة(نماذج) من الاختبار الكلي، بعدد أقل من المفردات؛ بحيث أن كل صورة تملك مؤشرات مشابهة للاختبار الكلي مما يوفر الوقت والجهد مع ضمان دقة القياس، وذلك قد يعود بالنفع خاصةً في المجال التربوي حيث يكون الأستاذ في الغالب مضطر لصياغة العديد من الصور الاختبارية لعدة شُعب Classes في فترات مختلفة أثناء تدريس المقرر الواحد، وكذلك في المواقف البحثية التي تتطلب توفر صور اختبارية مختلفة.

مصطلحات الدراسة:

طرق انتقاء المفردات:Item Selection Procedures وهي الإجراءات التي سيتم تطبيقها في الدراسة لإنتقاء المفردات من الاختبار الكلي، بثلاثة طرق، وتوجد نتائج مؤشر كل طريقة ضمن نتائج تحليل البيانات وفقاً لنظرية الاستجابة للمفردة باستخدام النموذج اللوغارتمي ثلاثي المعلم من خلال مخرجات برنامج (Xcalibre 4.2)، وهي:

أقصى قدرة (θmax) Theata at Maximum Information: وهو أحد مؤشرات دوال معلومات مفردات الاختبار؛ ويُعبر عن مستوى القدرة المقابل لأقصى معلومات تقدمها المفردة، ويتم الانتقاء إجرائياً باختيار (٢٠) مفردة من بين مفردات الاختبار التحصيلي المؤلف من (٦٠) مفردة اعتماداً على أعلى قيمة لمتغير أقصى قدرة مقابلة لأقصى معلومات بحيث تغطي مدى القدرة المقاسة، وستتم الإشارة للاختبار الذي تم انتقاء مفرداته بهذه الطريقة باختبار (أقصى قدرة).

أقصى قيمة معلومات (I(θmax)) Maximum Information: وهو أحد مؤشرات دوال معلومات مفردات الاختبار؛ ويُعبر عن أقصى معلومات(أقصى ارتفاع لمنحنى دالة المعلومات) يمكن أن تقدمها المفردة عند مستوى معين من القدرة، ويتم الانتقاء إجرائياً باختيار (٢٠) مفردة من بين مفردات الاختبار التحصيلي المؤلف من (٦٠) مفردة اعتماداً على أعلى قيمة لمتغير أقصى معلومات بحيث تغطي مدى القدرة المقاسة، وستتم الإشارة للاختبار الذي تم انتقاء مفرداته بهذه الطريقة باختبار(أقصى معلومات).

أعلى قيمة لمؤشر التمييز التقليديClassical Discrimination: وهي إحدى مؤشرات تحليل المفردة في النظرية التقليدية في القياس، وتم الإشارة إليه بالتمييز التقليدي لتمييزه عن بارامتر التمييز في نظرية الاستجابة للمفردة(Guyer & Thompson, 2014)، وقيمته تُعبر عن قيمة معامل الاتباط الثنائي المتسلسل بين درجة المفردة الثنائية(٠، ١) والدرجة الكلية للاختبار، باعتبار أن الدرجة الكلية تستخدم لتقدير القدرة(مهرنز ولهمن، ٢٠٠٣)، ويتم الانتقاء إجرائياً باختيار (٢٠) مفردة من بين مفردات الاختبار التحصيلي المؤلف من (٦٠) مفردة اعتماداً أعلى قيمة لمؤشر التمييز التقليدي في المستوى المحدد على متصل القدرة المقاسة، وستتم الإشارة للاختبار الذي تم انتقاء مفرداته بهذه الطريقة باختبار(التمييز التقليدي).

دقة تقدير القدرة ($\hat{\theta}$) Accuracy of Ability Estimation: تقاس دقة تقدير القدرة ($\hat{\theta}$) بعدة مؤشرات إحصائية وسيتم قياسها إجرائياً في الدراسة من خلال حساب الخطأ المعياري لتقدير القدرةStanderd Error of Estimation ويعبر عن الانحراف المعياري لتقديرات القدرة حول خط الانحدار عند مستوى ثابت من مستوى القدرة الحقيقي، ويحسب الخطأ المعياري بإيجاد قيمة مقلوب الجذر التربيعي لقيمة دالة المعلومات(بيكر، ٢٠١٠).

دالة معلومات المفردة Item Information Function: يقصد بها مقدار طول فترة الثقة في تقدير القدرة للفرد، حيث أنها تحدد كمية المعلومات التي تزود بها المفردة عند تقدير قدرات المفحوصين (Hambleton & Swaminathan, 1985)، وعرفها نور الدين(٢٠٠٢) بأنها مقدار ما تعطيه المفردة من معلومات عند المستويات المختلفة على متصل القدرة، ودالة المعلومات مشروطة على القدرة ويتم حسابها عن طريق مقلوب تباين الخطأ المعياري لتقدير القدرة.

دالة معلومات الاختبار :Test Information Function تُعبر هذه الدالة بعلاقة منحنية بين متغيرين أحدهما قدرة الأفراد على المحور الأفقي والآخر مقدار المعلومات التي يقدمها الاختبار على المحور الرأسي، أي أنها تمثل كمية المعلومات التي يقدمها الاختبار عند جميع مستويات القدرة (علام، ٢٠٠٥)، وتعرف إجرائيا بأنها حاصل جمع دوال معلومات جميع المفردات التي يتكون منها الاختبار.

الكفاءة النسبية للاختبار :Relative Efficiency هي مؤشر يستخدم لمقارنة دوال المعلومات للاختبارات وذلك عند ما يتم تقدير القدرة (θ) باستخدام اختبارات مختلفة، وتتم المقارنة من خلال إيجاد الكفاءة النسبية لاختبار كمقدر للقدرة عند مستوي معين بالنسبة لاختبار الآخر، من خلال قسمة قيمة دالة المعلومات للاختبار الأول على قيمة دالة المعلومات للاختبار الثاني(بيكر، ٢٠١٠).

محددات الدراسة: تتحدد الدراسة بنوع الاختبار التحصيلي المكون من أسئلة موضوعية ثنائية التقدير(١،٠) بغض النظر عن محتوى الاختبار في أي مقرر، وبالنموذج المستخدم في تدريج تجمع المفردات وهو النموذج اللوغارتمي ثلاثي البارامتر، وبالطرق التي تم انتقاء المفردات بها، وهي:(أقصى قدرة، وأقصى معلومات، والتمييز التقليدي)، وبالبرامج والأساليب الإحصائية المستخدمة.

إجراءات الدراسة:

منهج الدراسة:

تقارن الدراسة بين ثلاث طرق في انتقاء المفردات لتحديد فعاليتها في دقة تقدير قدرة الأفراد ورفع كفاءة الاختبار التحصيلي، وبالتالي فالمنهج المناسب للإجابة عن أسئلة الدراسة هو المنهج الوصفي التحليلي المقارن.

عينة الدراسة:

تكونت عينة الدراسة من(٤٠٩) طالبة من طالبات كلية التربية بجامعة طيبة الدارسات لمقرر التقويم التربوي (EDPS 314) بمرحلتي البكالوريوس والدبلوم التربوي في عامي (٢٠١٧-٢٠١٨)، منهن (١٠٦) طالبة بمرحلة البكالوريوس بنسبة (٢٦%)، و(٣٠٣) طالبة بالدبلوم التربوي بنسبة (٧٤%).

البرامج والأساليب الإحصائية المستخدمة في تحليل البيانات:

برنامج Xcalibre 4.2 (Guyer & Thompson, 2014)، تحليل استجابات عينة الدراسة على مفردات الاختبار التحصيلي وفقاً لإجراءات النظرية التقليدية في القياس مثل: حساب معامل الصعوبة ومعامل التمييزو معامل ثبات ألفا لكرنباخ، ووفقاً لنظرية الاستجابة للمفردة باستخدام النموذج اللوغارتمي الثلاثي المعلم لتقدير معلم(الصعوبة، والتمييز، والخمين) وتقدير قدرة الأفراد والخطأ المعياري للتقدير، وقيمة اختبار Z للبواقي لملائمة بيانات المفردات للنموذج وقيمة دالة المعلومات للمفردة ودالة المعلومات للاختبار وإيجاد متغير مستوى القدرة المقابلة لأقصى معلومات، ومتغير أقصى معلومات للمفرد.

برنامج الحزم الإحصائية SPSS 25.0 (IBM SPSS 25.0, Armonk, NY, USA): لحساب الإحصاءات الوصفية، والتحليل العاملي الاستكشافي، ومعامل ارتباط بيرسون، ومعامل التجزئة النصفية لجتمان، تحليل التباين للقياسات المتعددة، والرسوم البيانية.

أداة الدراسة:

عند إعداد أداة الدراسة كان الهدف أن تمثل الأداة أي اختبار تحصيلي من نوع اختيار من متعدد بأربع بدائل استجابة، فتم اختيار اختبار تحصيلي من مقرر التقويم التربوي (314 EDPS)، وذلك لكون المقرر أحد المقررات التي تُدرس كمتطلب كلية بمرحلة البكالوريوس والدبلوم التربوي مما يتيح فرصة التطبيق على عينة مناسبة من حيث العدد، ولخبرة الباحثة في تدريس المقرر مع إمكانية الاستعانة ببعض الأساتذة ممن لديهم خبرة في تدريس المقرر. وقد قامت الباحثة بإعداد مفردات اختبارية (تجمع مفردات) مكون من (٦٠) مفردة من نوع الاختيار من متعدد بأربع بدائل، ويتم تقدير درجات الاختبار بطريقة ثنائية (٠، ١)، وقد روعي عند إعداد الاختبار معايير الاختبارات التحصيلية، وخاصة معايير صياغة مفردات الاختيار من متعدد.

أدلة صدق التكوين الفرضي للاختبار:

تم التحقق من أدلة صدق التكوين الفرضي من خلال بناء المفردات وفقاً لجدول المواصفات في محتوى الوحدة الدراسية المستهدفة بمحتوى الاختبار، وهي: وحدة (إعداد الاختبار التحصيلي الجيد والتحقق من خصائصه السيكومترية وتحليل المفردات وإخراج الاختبار) بحيث يغطي جدول المواصفات الأهداف السلوكية حسب تصنيف بلوم، كالتالي:(التذكر(٦) مفردات، الفهم(١١) مفردة، التطبيق(٢٠) مفردة، التحليل(٧) مفردات،

التركيب(٨) مفردات، التقييم(٨) مفردات). وبعد إعداد الصيغة النهائية للمفردات تم عرض الاختبار على (٣) أساتذة لديهم خبرة في تدريس المقرر وأشاروا بمناسبة المفردات لمحتوى الوحدة وطبيعة الأهداف المرجو تحقيقها.

نتائج الدراسة:

التحقق من افتراضات نظرية الاستجابة للمفردة لبيانات الاختبار الـ (٦٠) مفردة

تم تطبيق الاختبار المكون من (٦٠) مفردة على عينة الدراسة المكونة من (٤٠٩) طالبة، ومن ثم تم تحليل البيانات للتحقق من افتراضات نظرية الاستجابة للمفردة تمهيداً لتدريج الاختبار باستخدام النموذج اللوغارتمي الثلاثي المعلم، والنتائج كما يلي:

أحادية البعد: تم التحقق من افتراض أحادية البعد لمفردات الاختبار باستخدام التحليل العاملي، والجدول رقم(١) يوضح نتائج اختبار كفاية العينة واختبار بارتلت.

جدول رقم (١): اختبار كفاية العينة واختبار بارتلت (ن=٤٠٩)

مقياس كايزر–مير– أولكن لكفاية العينة		**0.957**
اختبار بارتلت	**قيمة مربع كاي التقريبية**	**8013.242**
	درجات الحرية	**1653**
	قيمة مستوى الدلالة	**0.00***

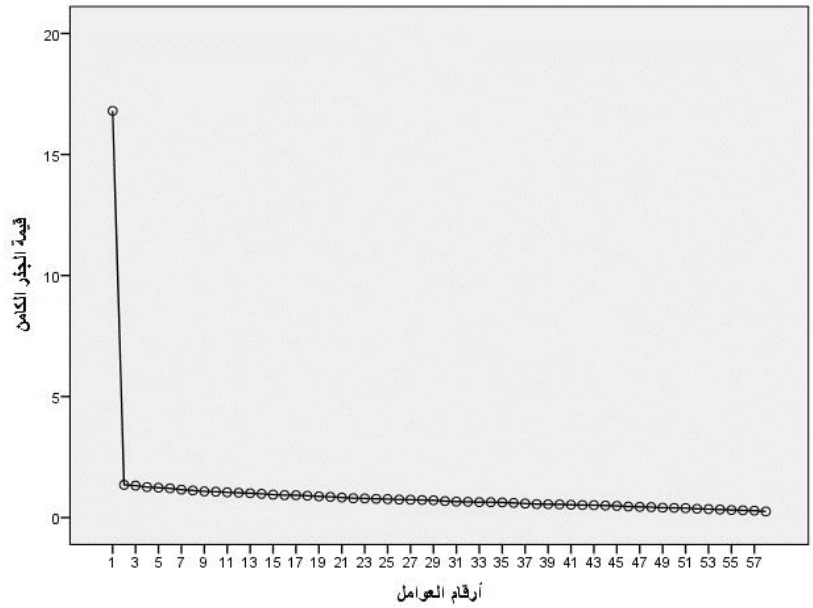
* دالة عند مستوى الدلالة ($\alpha \leq 0.05$)

وتم التحقق من كفاية حجم العينة من خلال محك كايزر حيث كانت قيمته تساوي (0.957) وهي قيمة أكبر من المحك الذي اعتمده كايزر لكفاية حجم العينة، وهو(0.50) مما يشير إلى أن أنماط الارتباطات بين المفردات منتظمة (Feldet, 2017). كما أن قيمة اختبار بارتلت لفحص مصفوفة الوحدة تساوي(8013.242) وهي دالة عند مستوى الدلالة ($\alpha \leq 0.05$) مما يدل على أن مصفوفة معاملات الارتباط تختلف عن مصفوفة الوحدة (Feldet, 2017) وبالتالي يمكن الوثوق بنتائج التحليل العاملي. والجدول رقم (٢) كذلك الشكل رقم (٤) يوضحان هذه النتيجة.

جدول رقم (٢) قيمة الجذر الكامن ونسبة التباين المفسر للعاملين الأول والثاني الناتجة من التحليل العاملي للاختبار المكون من (٦٠) مفردة

العوامل	الجذور الكامنة*		
	الكلي	نسبة التباين	النسبة التراكمية
1	16.801	28.967	28.967
2	1.355	2.336	31.303

***تم الاكتفاء برصد نتائج العوامل الأول والثاني فقط لتوضيح النسبة بينهما**



شكل رقم(٤): التمثيل البياني لقيم الجذر الكامن للعوامل الناتجة من رسم منحنى أقصى انحدار

يمكن ملاحظة أنه من خلال نتائج الجدول رقم(٢)، والتمثيل البياني لقيم الجذر الكامن برسم منحنى أقصى انحدار حسب شكل رقم(٤) تبين وجود عامل واحد مهيمن لكامل الاختبار بقيمة جذر كامن تساوي (16.801) ويفسر (28.967 %) من التباين الكلي، وقيمة جذر كامن للعامل الثاني تساوي (1.355) ويفسر (2.336%) من التباين الكلي؛ وبقية

الجذور الكامنة المتتالية متقاربة(Hambleton, et al., 1991, 56)، كما أن حاصل قسمة قيمة الجذر الكامن للعامل الأول على قيمة الجذر الكامن للعامل الثاني تساوي (12.40) وهي قيمة أكبر من (2)(Hambleton & Swaminathan, 1985; Lord, 1980)، مما يشير إلى تحقق أحادية البعد للاختبار التحصيلي.

الاستقلال الموضعي: يتحقق الاستقلال الموضعي بتحقق افتراض أحادية البُعد ضمنياً للاختبار المكون من (٦٠) مفردة، وذلك وفقاً لما أشار إليه كل من (Lord, 1080؛ Hambleton & Swaminathan, 1985؛ Hambleton, et., 1991)

المنحنيات المميزة للمفردة: تم التحقق من هذا الافتراض من خلال مخرجات برنامج (Xcalibre 4.2) حيث أنها أوضحت وجود رسم للمنحنيات المميزة لكل مفردة من مفردات الاختبار المكون من (٦٠) مفردة.

التحرر من السرعة: يتم التعرف على التحرر من عامل السرعة من خلال التعرف على عدد الطلاب الذين لم ينتهوا من الاجابة عن مفردات الاختبار (Hambleton & Swaminathan, 1985)، واختبار مقرر التقويم التربوي حُدد له زمن للإجابة بمقدار ساعتين وهو وقت كافٍ. وعند مراجعة الاستجابات وجدت (٣) بيانات مفقودة في المفردات ذات الرقم (١٩)، و(٣٢)، و(٥٠) وهي تقع في منتصف الاختبار وليس آخره؛ وقد يعود سبب عدم الإجابة إلى انخفاض في مستوى القدرة وليس لعامل الوقت؛ وهذا قد يؤيد تحرر الاختبار من عامل السرعة.

نتائج تحليل مفردات الاختبار الـ (٦٠) مفردة باستخدام النموذج اللوغارتمي ثلاثي المعلم

تم التحقق من مطابقة البيانات للنموذج المُستخدم عن طريق إحصاء الملائمة بقيمة بواقي z (zResidual Statistic Fit)، عند مستوى دلالة ($\alpha \leq 0.05$) لجميع المفردات، وقد وُجد أن: جميع المفردات كانت مطابِقة للنموذج اللوغارتمي الثلاثي المعلم ما عدا المفردتان رقم (٢) ورقم (٥٧)، وبالتالي سيتم حذفها، أما من ناحية مُطابقة الأفراد للنموذج فإن برنامج يتوفر له خاصية حذف الأفراد غير المطابقين، وقد كان جميع الافراد مطابقين للنموذج؛ لذلك لم يتم حذف استجابات أي فرد. ومن ثم تمت إعادة تحليل بيانات (٥٨) مفردة بعد حذف المفردتين غير المطابقة، وذلك بهدف تقدير معالم (التمييز، والصعوبة، والتخمين) للنموذج اللوغارتمي الثلاثي المعلم، ونتائج التحليل للاختبار المكون من (٥٨) موضحة في الجدول رقم (٣):

جدول رقم (٣): قيمة البواقي Z (z Resid) كإحصاء لملائمة المفردات للنموذج اللوغارتمي الثلاثي المعلم وتقدير معالم(التمييز، والصعوبة، والتخمين) لمفردات الاختبار (i*=58) وعدد الأفراد(n**=409)

رقم	A	B	C	أقصى قدرة	أقصى معلومات	بواقي Z***
1	1.63	0.36	0.12	0.44	1.50	1.20
2*	–	–	–	–	–	–
3	1.32	0.23	0.13	0.3	0.97	0.50
4	1.42	0.94	0.22	1.06	0.94	0.72
5	1.21	1.22	0.30	1.4	0.59	1.45
6	1.27	-0.10	0.24	0.04	0.73	0.66
7	1.71	0.41	0.19	0.5	1.47	1.48
8	1.15	0.64	0.22	0.8	0.62	1.43
9	1.37	0.39	0.22	0.52	0.88	.581
10	1.06	0.44	0.24	0.62	0.51	1.13
11	1.32	0.69	0.21	0.82	0.83	0.71
12	1.67	0.73	0.20	0.84	1.36	0.87
13	1.72	0.72	0.19	0.82	1.48	1.36
14	0.90	0.60	0.23	0.8	0.38	1.05
15	1.25	0.05	0.23	0.2	0.72	0.60
16	1.18	0.10	0.23	0.24	0.65	0.60

جدول رقم (٣): قيمة البواقي Z (z Resid) كإحصاء لملائمة المفردات للنموذج اللوغارتمي الثلاثي المعلم وتقدير معالم(التمييز، والصعوبة، والتخمين) لمفردات الاختبار (i*=58) وعدد الأفراد(n**=409)

رقم	A	B	C	أقصى قدرة	أقصى معلومات	بواقي Z***
17	1.11	0.70	0.22	0.86	0.58	0.70
18	1.43	0.38	0.21	0.5	0.98	0.85
19	1.01	0.03	0.24	0.22	0.46	0.51
20	1.11	0.89	0.22	1.04	0.58	0.75
21	1.05	0.45	0.24	0.62	0.50	0.49
22	1.14	0.50	0.24	0.66	0.59	1.35
23	1.54	0.47	0.21	0.58	1.14	0.93
24	1.61	1.32	0.18	1.42	1.32	1.37
25	1.62	1.00	0.18	1.1	1.32	1.40
26	1.04	0.63	0.22	0.8	0.50	1.07
27	1.02	1.28	0.21	1.44	0.49	1.08
28	1.31	1.43	0.18	1.54	0.87	1.46
29	1.16	0.29	0.23	0.44	0.62	1.79
30	1.42	0.10	0.22	0.22	0.95	0.74
31	1.05	0.47	0.23	0.64	0.51	0.74
32	1.10	1.08	0.21	1.24	0.58	0.97
33	0.92	0.26	0.25	0.4	0.37	1.03

جدول رقم (٣): قيمة البواقي Z (z Resid) كإحصاء لملائمة المفردات للنموذج اللوغارتمي الثلاثي المعلم وتقدير معالم(التمييز، والصعوبة، والتخمين) لمفردات الاختبار (i*=58) وعدد الأفراد(n**=409)

رقم	A	B	C	أقصى قدرة	أقصى معلومات	بواقي Z***
34	1.30	0.96	0.20	1.08	0.82	1.71
35	1.10	0.75	0.24	0.92	0.55	0.32
36	1.22	-0.46	0.24	-0.32	0.68	0.27
37	1.20	1.01	0.21	1.16	0.70	0.94
38	1.28	0.61	0.23	0.74	0.75	0.36
39	1.32	0.81	0.22	0.94	0.82	0.51
40	1.24	0.94	0.19	1.08	0.76	1.29
41	1.21	0.90	0.22	1.04	0.68	1.21
42	1.33	0.99	0.22	1.12	0.83	0.60
43	1.48	0.82	0.17	0.92	1.13	1.85
44	1.37	0.79	0.22	0.92	0.87	0.53
45	1.53	0.46	0.20	0.58	1.13	1.05
46	1.14	0.49	0.23	0.64	0.60	0.61
47	1.37	1.78	0.21	1.9	0.91	1.18
48	1.54	0.37	0.20	0.48	1.16	1.02
49	1.15	0.19	0.23	0.34	0.61	0.57
50	1.08	-0.07	0.24	0.1	0.53	0.56

جدول رقم (٣): قيمة البواقي Z (z Resid) كإحصاء لملائمة المفردات للنموذج اللوغارتمي الثلاثي المعلم وتقدير معالم(التمييز، والصعوبة، والتخمين) لمفردات الاختبار (i*=58) وعدد الأفراد(n**=409)

رقم	A	B	C	أقصى قدرة	أقصى معلومات	بواقي Z***
51	1.06	0.38	0.23	0.54	0.52	0.46
52	1.25	1.36	0.19	1.48	0.78	1.17
53	1.31	1.34	0.20	1.46	0.83	0.94
54	1.51	0.31	0.22	0.42	1.08	0.70
55	1.62	1.05	0.17	1.14	1.36	1.82
56	1.25	-0.20	0.24	-0.06	0.71	0.63
57*	-	-	-	-	-	–
58	1.50	0.60	0.20	0.72	1.10	1.02
59	1.24	1.52	0.21	1.66	0.73	1.02
60	1.22	1.49	0.19	1.62	0.74	1.26

* مفردات غير مطابقة: **لا يوجد أفراد غير مطابقين *** جميع

من الجدول رقم (٣) يمكن ملاحظة أن قيمة بواقي z (z Residual Statistic Fit)، غير دالة عند مستوى دلالة ($\alpha \leq 0.05$) لجميع المفردات، أي أن جميع المفردات كانت مطابِقة للنموذج اللوغارتمي الثلاثي المعلم أي أنه مناسب لتحليل البيانات، فيما كانت القيمة القصوى لمعلم الصعوبة (1.7886)، والقيمة الدنيا له (-0.4684)، أما معلم التمييز فقد كانت القيمة القصوى له (1.726)، وبقيمة دنيا تساوي (0.9091). وأخيراً معلم التخمين الذي وُجد أن القيمة القصوى له (0.3054)، وبقيمة دنيا تساوي (0.1278) وبالتالي نستنتج أن جميع قيم معالم النموذج جاءت ضمن الحدود المقبولة(نورالدين، ٢٠٠٢). في حين كانت قيمة أقصى كمية معلومات مقدمة من المفردات

تقدّر بالقيمة (1.5077) وتقابل القدرة (0.44)، وأقل قيمة لكمية المعلومات التي تقدمها المفردات تساوي(0.3767) عند القدرة المقدرة بالقيمة(0.46).

نتائج تحليل مفردات الاختبار الـ (٥٨) مفردة باستخدام النظرية التقليدية

تم تحليل استجابات (٤٠٩) من الطالبات على الاختبار المكون من (٥٨) مفردة، حسب النظرية التقليدية، لتقدير معاملي الصعوبة والتمييز، والنتائج موضحة في الجدول رقم (٤).

جدول رقم(٤): معاملي (الصعوبة والتمييز) حسب النظرية التقليدية للقياس لمفردات الاختبار(i=58) وعدد الأفراد(n=409)

رقم المفردة	معامل الصعوبة	معامل التمييز	رقم المفردة	معامل الصعوبة	معامل التمييز	رقم المفردة	معامل الصعوبة	معامل التمييز
١	0.4817	0.63	٢١	0.5306	0.463	٤١	0.4205	0.438
٢	–	-	٢٢	0.5232	0.4765	٤٢	0.3961	0.4402
٣	0.5232	0.5779	٢٣	0.4768	0.634	٤٣	0.3472	0.6446
٤	0.3936	0.4871	٢٤	0.2543	0.4407	٤٤	0.4352	0.4985
٥	0.3888	0.4168	٢٥	0.3252	0.5707	٤٥	0.4768	0.635
٦	0.6504	0.5268	٢٦	0.4768	0.4731	٤٦	0.5159	0.4841
٧	0.4645	0.6954	٢٧	0.3447	0.3739	٤٧	0.2421	0.2039
٩	0.5159	0.5678	٢٩	0.5526	0.5257	٤٩	0.5697	0.5391
١٠	0.533	0.4708	٣٠	0.5844	0.6085	٥٠	0.6357	0.5048
١١	0.4425	0.5348	٣١	0.5086	0.4949	٥١	0.5403	0.4858

رقم المفردة	معامل الصعوبة	معامل التمييز	رقم المفردة	معامل الصعوبة	معلم التمييز	رقم المفردة	معامل الصعوبة	معامل التمييز
١٢	0.4108	0.5881	٣٢	0.3716	0.4326	٥٢	0.2836	0.4105
١٣	0.3912	0.6376	٣٣	0.5917	0.4094	٥٣	0.3007	0.3827
١٤	0.4963	0.4275	٣٤	0.3692	0.5043	٥٤	0.5257	0.6157
١٥	0.6015	0.5615	٣٥	0.4719	0.4311	٥٥	0.2885	0.6086
١٦	0.5892	0.5503	٣٦	0.7188	0.5206	٥٦	0.6626	0.5474
١٧	0.4572	0.4859	٣٧	0.3716	0.4699	٥٧	-	-
١٨	0.4988	0.622	٣٨	0.489	0.4908	٥٨	0.4377	0.6197
١٩	0.6235	0.4607	٣٩	0.4303	0.4927	٥٩	0.2861	0.32
٢٠	0.4132	0.4572	٤٠	0.3594	0.5336	٦٠	0.2567	0.4075

ويتبين من الجدول رقم (٤) أن قيم معامل الصعوبة لجميع المفردات تراوحت ما بين (٠.٢٤٢) و(٠.٧١٨). فيما كانت قيم معامل التمييز ما بين (٠.٢٠٣) و(٠.٦٩٥)، بمتوسط صعوبة للاختبار ككل يساوي (٠.٤٥٦) ومتوسط تمييز يساوي (٠.٥٠٧). وهي قيم مقبولة للصعوبة والتمييز، وبالنظر لقيمة متوسط الصعوبة يعتبر الاختبار متوسط الصعوبة تقريباً (نيتكو وبروخارت، ٢٠١٢؛ انستازي ويوريان، ٢٠١٥).

نتائج انتقاء مفردات الاختبارات الثلاثة المكونة من (٢٠) وفقاً لطريقة أقصى قدرة، وطريقة أقصى معلومات، وطريقة انتقاء المفردات حسب التمييز التقليدي

تم تقسيم مدى صعوبة مفردات الاختبار الكلي (التي تقابل نفس مدى القدرة المقاسة بالاختبار وتقاس بنفس وحدة القياس اللوجت) إلى أربعة فترات (إرباعيات) متساوية(هياجنة، ٢٠١١) بناء على نتائج تحليل الاختبار الكلي، ونتائج هذا الاجراء موضحة في الجدول رقم (٥):

جدول رقم (٥): قيم نقاط مقياس الارباعيات للقدرة المقاسة بالاختبار الكلي (58 مفردة)

الارباعي	الأول	الثاني	الثالث
قيمة الارباعي	0.095	0.657	1.220

تم انتقاء المفردات لكل من الاختبارات الثلاثة، بناء على الفترات التي تم تحديدها في الجدول رقم(٥) بحيث يكون المجموع الكلي لمفردات كل اختبار يساوي(20) مفردة وذلك لأن هذه النسبة تمثل(١:٣) وهي ضمن المدى المناسب المستخدم في بنوك الأسئلة(هياجنة، ٢٠١١)، بحيث تغطي هذه المفردات مدى القدرة المقاسة، وفي كل اختبار؛ تم اختيار (6) مفردات للفترات الداخلية، و(٤) مفردات للفترات الطرفية. والجدول رقم(٦) يوضح أرقام مفردات الاختبارات الثلاثة التي سيتضمنها كل اختبار فرعي مكون من (20) مفردة، موزعة على مستويات القدرة الأربع، مع ملاحظة أن المفردات التي تم انتقاءها في كل اختبار تُعد ممثلة للسمة المقاسة بإعتبار أن الاختبار الكلي التي تم اشتقاقها منه يقيس من الاساس سمة واحدة، وقد اتضح ذلك من نتائج التحليل العاملي لمفردات الاختبار الكلي التي تم عرضها عند التحقق من أحادية البعد، كما سيتم التحقق من أحادية البعد لكل اختبارمن الاختبارات الثلاثة.

الجدول رقم (٦): يوضح أرقام مفردات الاختبارات الثلاثة المكونة من (20) مفردة عند المستويات الأربعة للقدرة (الارباعيات)

مستويات القدرة	مدى القدرة	طريقة الانتقاء لمفردات الاختبار	عـــــدد المفردات	أرقام المفردات
الأول	$-0,468 \leq \emptyset \leq 0,095$	أقصى قدرة	4	15- 16- 19- 50
		أقصى معلومات	4	6- 15- 36- 56
		التمييز التقليدي	4	6- 15- 16- 56
الثاني	$0,095 < \emptyset \leq 0,657$	أقصى قدرة	6	14- 22- 26- 38-46-58
		أقصى معلومات	6	1- 7- 23- 45- 48- 58
		التمييز التقليدي	6	1- 18- 23- 45-48- 54
الثالث	$0,657 < \emptyset \leq 1,220$	أقصى قدرة	6	5- 25- 32- 37- 42- 55
		أقصى معلومات	6	4- 12-13-25-43-55
		التمييز التقليدي	6	12- 13- 25- 40- 43-55
الرابع	$1,220 < \emptyset \leq 1,788$	أقصى قدرة	4	28- 47- 59- 60
		أقصى معلومات	4	24- 28- 47- 53
		التمييز التقليدي	4	27-52- 53- 59

التحقق من افتراضات نظرية الاستجابة للمفردة للاختبارات الثلاثة التي تم انتقاء مفرداتها من الاختبار الكلي

أحادية البعد: تم التحقق من افتراض أحادية البعد لمفردات الاختبارات الثلاثة؛ باستخدام التحليل العاملي، والجدول رقم (٧) يوضح نتائج اختبار كفاية العينة وبارتلت

جدول رقم (٧): اختبار كفاية العينة واختبار بارتلت لنتائج التحليل العاملي للاختبارات الثلاثة

الاحصاءة		الاختبار		
		أقصى قدرة	وأقصى معلومات	قيمة التمييز
مقياس كايزر-مير-أولكن لكفاية العينة		0.920	0.957	0.957
اختبار بارتلت	قيمة مربع كاي التقريبية	1589.423	2497.449	2475.179
	درجات الحرية	190	190	190
	قيمة مستوى الدلالة	0.00*	0.00*	0.00*

* دالة عند مستوى الدلالة ($\alpha \leq 0.05$)

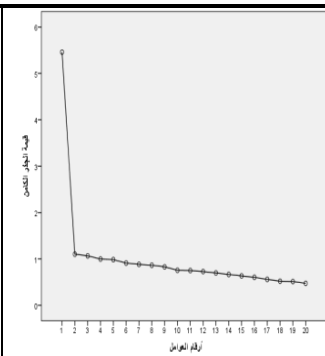
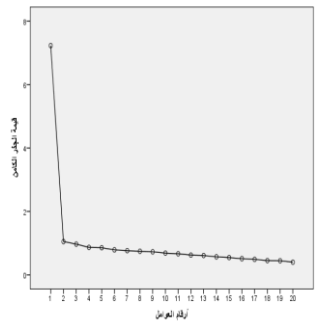
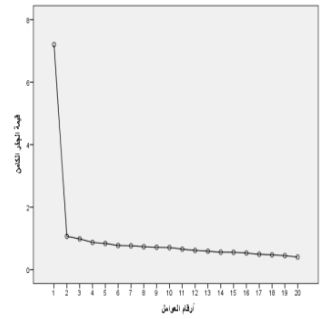
وتم التحقق من كفاية حجم العينة من خلال محك كايزر حيث كانت قيمته للاختبارات الثلاثة لتي تم انتقاء مفرداتها وفقاً لطريقة:(أقصى قدرة، وأقصى معلومات، والتمييز التقليدي) تساوي (0.920، 0.957، 0.957) على الترتيب، وهي قيمة أكبر من المحك الذي اعتمده كايزر لكفاية حجم العينة، وهو (0.50) مما يشير إلى أن أنماط الارتباطات منتظمة بالنسبة للاختبارات الثلاثة (Feldet, 2017). كما أن قيمة اختبار بارتلت لفحص مصفوفة الوحدة تساوي (1589.423، 2497.449، 2475.179) على الترتيب، وهي دالة عند مستوى الدلالة ($\alpha \leq 0.05$) مما يدل على أن مصفوفة معاملات الارتباط تختلف عن مصفوفة الوحدة (Feldet, 2017) وبالتالي يمكن الوثوق بنتائج التحليل العاملي. والجدول رقم (٨) يوضح قيمة الجذر الكامن ونسبة التباين المفسر للعاملين الأول والثاني من العوامل الناتجة من التحليل العاملي للاختبارات الثلاثة.

جدول رقم(٨) قيمة الجذر الكامن ونسبة التباين المفسر للعاملين الأول والثاني من العوامل الناتجة من التحليل العاملي للاختبارات الثلاثة

الاختبارات	العوامل	الجذور الكامنة*		
		الكلي	نسبة التباين	النسبة التراكمية
الاختبار الأول (أقصى قدرة)	1	5.458	27.291	27.291
	2	1.103	5.514	32.806
الاختبار الثاني (أقصى معلومات)	1	7.229	36.145	36.145
	2	1.050	5.252	41.398
الاختبار الثالث (التمييز التقليدي)	1	7.201	36.004	36.004
	2	1.068	5.341	41.344

*تم الاكتفاء برصد نتائج العامل الأول والثاني فقط لتوضيح النسبة بينهما

ومن خلال نتائج الجدول رقم(٨)، والتمثيل البياني لقيم الجذر الكامن برسم منحنى أقصى انحدار حسب شكل رقم(٥) تبين وجود عامل واحد مهيمن لكل اختبار من الاختبارات الثلاثة لتي تم انتقاء مفرداتها وفقاً لطريقة:(أقصى قدرة، وأقصى معلومات، والتمييز التقليدي) بقيمة جذر كامن تساوي (5.458، 7.229، 7.201) ويفسر (27.291%، 36.145%، 36.004%) من التباين الكلي لكل اختبار على الترتيب، وقيمة جذر كامن للعامل الثاني تساوي (1.103، 1.050، 1.068) ويفسر (5.514%، 5.252%، 5.341%) من التباين الكلي لكل اختبار على الترتيب ؛ وبقية الجذور الكامنة المتتالية متقاربة القيمة(Hambleton, et al., 1991, 56)، كما أن حاصل قسمة قيمة الجذر الكامن للعامل الأول على قيمة الجذر الكامن للعامل الثاني تساوي (4.95، 6.88، 6.74) لكل اختبار على الترتيب، وهي قيمة أكبر من (2)(Hambleton & Swaminathan, 1985; Lord, 1980)، مما يشير إلى تحقق أحادية البعد للاختبارات الثلاثة.

		
الاختبار الأول (أقصى قدرة)	**الاختبار الثاني (أقصى معلومات)**	**الاختبار الثالث (التمييز التقليدي)**

شكل رقم (٥): التمثيل البياني لقيم الجذر الكامن للعوامل الناتجة من رسم منحنى أقصى انحدار للاختبارات الثلاثة

الاستقلال الموضعي: يتحقق الاستقلال الموضعي بتحقق افتراض أحادية البُعد ضمنياً للاختبارات الثلاثة، وذلك وفقاً لما أشار إليه كل من (Lord, 1080؛ Hambleton & Swaminathan, 1985؛ Hambleton, et., 1991)

المنحنيات المميزة للمفردة: تم التحقق من هذا الافتراض من خلال مخرجات برنامج (Xcalibre 4.2) حيث أنها أوضحت وجود رسم للمنحنيات المميزة لكل مفردة من مفردات الاختبار.

التحرر من السرعة: يتم التعرف على التحرر من عامل السرعة من خلال التعرف على عدد الطلاب الذين لم ينتهوا من الاجابة عن مفردات الاختبار (Hambleton & Swaminathan, 1985)، واختبار مقرر التقويم التربوي حُدد له زمن للإجابة بمقدار ساعتين وهو وقت كافٍ. وعند مراجعة الاستجابات وجدت (٣) بيانات مفقودة في المفردات ذات الرقم (١٩)، و(٣٢)، و(٥٠) مع العلم بأنها تقع ضمن مفردات الاختبار التي تم انتقاء مفرداته وفقاً لأقصى قدرة وهي تقع في منتصف الاختبار الكلي وليس آخره؛ وقد يعود سبب عدم الإجابة إلى انخفاض في مستوى القدرة وليس لعامل الوقت؛ وهذا قد يؤكد تحرر الاختبارات الثلاثة من عامل السرعة.

مطابقة البيانات للنموذج المُستخدم: تم التحقق من المطابقة بين البيانات والنموذج اللوغارتمي الثلاثي المعلم عن طريق قيمة بواقي z (z Residual Statistic Fit)، عند مستوى دلالة ($\alpha \leq 0.05$) للاختبارات الثلاثة، وقد وُجد أن: جميع المفردات كانت مطابِقة للنموذج اللوغارتمي الثلاثي المعلم.

الإجابة عن أسئلة الدراسة

ما مدى اختلاف الخصائص السيكومترية والإحصائية للاختبار الكلي والاختبارات الثلاثة المشتقة منه بانتقاء مفرداتها وفقاً لطريقة:(أقصى قدرة، وأقصى معلومات، والتمييز التقليدي)؟ وللاجابة عن السؤال الأول؛ تم عرض النتائج المتعلقة بالخصائص السيكومترية، ومن ثم عرض النتائج المتعلقة بالخصائص الإحصائية، كما يلي:

أ: عرض النتائج المتعلقة بالخصائص السيكومترية: تم تحليل الاختبار الكلي والاختبارات الثلاثة المشتقة منه التي تم انتقاء مفرداتها وفقاً لطريقة:(أقصى قدرة، وأقصى معلومات، والتمييز التقليدي)؛ من خلال حساب المؤشرات السيكومترية وفقاً للنظرية التقليدية، وبتحليل جميع الاختبارات باستخدام النموذج الثلاثي المعلم وفقاً لنظرية الاستجابة للمفردة ، حيث تم حساب قيم ثبات معامل ألفا لكرنباخ لجمع أدلة تتعلق بالاتساق الداخلي، وكذلك تم حساب معاملات الارتباط لبيرسون بين تقديرات قدرة الافراد من الاختبار الكلي وتقديرات القدرة من الاختبارات الثلاثة؛ حيث تم اعتبار الاختبار الكلي اختباراً محكياً لتقدير صدق درجات

الاختبارات حيث أن الدراسة تسعى للتحقق من مدى اتساق الخصائص السيكومترية للاختبار الكلي مع الخصائص التي تمتلكها الاختبارات الثلاثة المشتقة منه، والجدول رقم (٩) يوضح تلك النتائج:

جدول رقم(٩): قيم معاملات الثبات للاختبارات وقيم معامل ارتباط بيرسون بين القدرة المقدرة من الاختبار الكلي والقدرة المقدرة من الاختبارات الثلاثة ن=409.

الاختبار	عدد المفردات	معامل ألفالكرنباخ	معامل جتمان التجزئة النصفية	معامل الارتباط
الكلي	**58**	**0.956**	**0.953**	–
الأول (أقصى قدرة)	**20**	**0.856**	**0.830**	**0.944***
الثاني (أقصى معلومات)	**20**	**0.905**	**0.906**	**0.965***
الثالث (التمييز التقليدي)	**20**	**0.904**	**0.906**	**0.957***

* دالة عند مستوى الدلالة ($\alpha \leq 0.05$)

يتضح من الجدول رقم(٩) أن قيمة معامل ألفا لكرنباخ لدرجات الاختبار الكلي تساوي(0.956) مما يدل على أن درجات الاختبار تملك ثبات مرتفع، كما أن قيم ثبات درجات الاختبارات الثلاثة التي تم انتقاء مفرداتها وفقاً لطريقة:(أقصى قدرة، وأقصى معلومات، والتمييز التقليدي) تساوي(0.856، 0.905، 0.904) على الترتيب، كما أن معامل الثبات بطريقة التجزئة النصفية لجتمان لإختلاف قيم التباين بين نصفي الاختبار يساوي(0.953، 0.830، 0.906، 0.906) للاختبار الكلي والاختبارات الثلاثة على الترتيب، وهي على العموم معاملات ثبات جيدة، وعند مقارنتها بقيمة الثبات للاختبار الكلي يمكن ملاحظة أن قيم معامل الثبات للاختبارين اللذين تم انتقاء مفرداتهما وفقاً لأقصى معلومات، ووفقاً لقيمة معامل التمييز التقليدي متقاربة مع قيمة معامل الثبات للاختبار الكلي ولم تتجاوز الفروق بين القيم القيمة (0.05) (الشافعي، ٢٠١٢)، مما قد يعد مؤشراً على أنها تملك نفس درجة الاتساق الداخلي التي يمتلكها الاختبار الكلي. بينما كان مقدار الفرق بين قيمتي معامل ألفا للاختبار الكلي والاختبار الذي تم انتقاء مفرداته وفقاً لأقصى قدرة أعلى من القيمة(0.05) (الشافعي، ٢٠١٢).

وبالعودة للجدول رقم(٩) يمكن ملاحظة أن قيم معامل ارتباط بيرسون بين قيم تقدير قدرات الأفراد بتطبيق الاختبار الكلي وقيم تقدير قدرات الأفراد بتطبيق الاختبارات الثلاثة المشتقة منه بانتقاء مفرداتها وفقاً لطريقة:(أقصى قدرة، وأقصى معلومات، والتمييز التقليدي) تساوي(0.944، 0.965، 0.957) على الترتيب، وهي علاقة موجبة وقوية بين تقديرات القدرة بالاختبار الكلي المكون من (58) مفردة وبالاختبارات التي يتكون كل منها من (20) مفردة، مما يدل على أن الاختبارات تقيس نفس السمة التي يقيسها الاختبار الكلي، وذلك يؤيد أن درجات الاختبارات الثلاثة تملك أدلة على صدق التكوين الفرضي.

أيضا بالعودة إلى الجدول رقم(٨) والجدول رقم(٢) السابقين نجد أن نتائج التحليل العاملي لدرجات الاختبار الكلي والاختبارات الثلاثة كشفت عن وجود عامل واحد مهيمن على باقي العوامل وقد تشبعت عليه جميع المفردات، وكانت التشبعات دالة وفقا لمحك كايزر بأن تكون القيمة (≤ 0.3) في كل من الاختبار الكلي والاختبارات الثلاثة المشتقة منه بانتقاء مفرداتها وفقاً لطريقة:(أقصى قدرة، وأقصى معلومات، والتمييز التقليدي)على الترتيب، مما يشير أيضاً إلى صدق التكوين الفرضي للاختبار الكلي والاختبارات الثلاثة المشتقة منه.

ب: عرض النتائج المتعلقة بالخصائص الإحصائية: تم التحقق من الخصائص الاحصائية للاختبار الكلي والاختبارات الثلاثة التي تم انتقاء مفرداتها وفقاً لطريقة:(أقصى قدرة، وأقصى معلومات، والتمييز التقليدي) بحساب الاحصاءات الوصفية لتقديرات القدرة، وهي: مقاييس النزعة المركزية، ومقاييس التشتت، والنتائج موضحة في الجدول رقم(١٠)

جدول رقم(١٠): الاحصاءات الوصفية لتقديرات قدرة الأفراد من الاختبار الكلي ومن الاختبارات الثلاثة المشتقة منه (ن=409).

الاحصاءات الوصفية	الاختبارات			
	الكلي	الأول (أقصى	الثاني(أقصـــــــى	الثالث (التمييز
عدد المفردات	58	20	20	20
المتوسط	-0.026	−0.015	-0.001	0.001
الخطـــــــأ	0.051	0.047	0.049	0.048
الوسيط	0.044	−0.118	-0.092	-0.119
الانحـــــراف	1.032	0.961	0.995	0.990
التباين	1.065	0.924	0.981	0.981
المدى	4.08	3.91	3.66	3.62

يتضح من الجدول رقم(١٠) أن قيم الاحصاءات الوصفية للاختبار الكلي والاختبارات الثلاثة المشتقة منه بانتقاء مفرداتها وفقاً لطريقة:(أقصى قدرة، وأقصى معلومات، وبحسب قيمة مؤشر التمييز التقليدي) جاءت متقاربة حيث أن الفروق بين قيم مؤشرات الاحصاء الوصفي(المتوسط، والوسيط، والانحراف المعياري، والتباين) للاختبارات الثلاثة والاختبار الكلي لم تتجاوز (0.01، 0.07، 0.07، 0.14) على الترتيب، مما يدل على أن الاختبارات الثلاثة متكافئة تقريباً من حيث المؤشرات الاحصائية مع الاختبار الكلي الذي اشتقت منه، كما أنه بمقارنة قيمتي كل من المتوسط والوسيط في كل اختبار من الاختبارات الثلاثة، يتضح أن توزيع تقديرات القدرة قريب من الاعتدالية، وذلك يؤيد أن الاختبارات تملك توزيع قدرة قريب من توزيع القدرة الخاص بالاختبار الكلي الذي اشتقت منه. مما سبق؛ يمكن استنتاج أن الاختبارات الثلاثة التي تم اشتقاقها من الاختبار الكلي عن طريق انتقاء مفرداتها وفقاً لطريقة:(أقصى قدرة، وأقصى معلومات، والتمييز التقليدي) تتكافأ في خصائصها السيكومترية (الصدق والثبات) ومؤشراتها الاحصائية مع الاختبار الكلي.

هل توجد فروق جوهرية بين قيم تقديرات قدرة أفراد العينة التي تم تقديرها من الاختبار الكلي ومن الاختبارات الثلاثة التي تم انتقاء مفرداتها وفقاً لطريقة:(أقصى قدرة، وأقصى معلومات، والتمييز التقليدي)؟، وللاجابة عن السؤال الثاني، تم إجراء تحليل التباين للقياسات

المتعددة لتقدير القدرة من الاختبار الكلي والتقديرات المقدرة من الاختبارات الثلاثة، والنتائج موضحة في الجدول رقم(١١):

جدول رقم (١١): دلالة اختبار ماكولي لقيم تقدير القدرة بالاختبار الكلي وبالاختبارات الثلاثة.

قيمة اختبار	مربع كاي	درجات الحرية	مستوى الدلالة*
0.637	183.643	5	* 0.00

* مستوى الدلالة ($\alpha \leq 0.05$)

يتضح من الجدول رقم (١١) أن قيمة اختبار مربع كاي التقريبية تساوي (183.643) وهي قيمة دالة إحصائياً عند مستوى الدلالة ($\alpha \leq 0.05$)، مما يشير إلى وجود فروق دالة بين تجانس القياسات الأربعة الناتجة عن استخدام الاختبار الكلي والاختبارات الثلاثة مما يدل على عدم تحقق شرط الدورية وبالتالي سيتم تطبيق الاختبار الاحصائي "جرين هاوس–جيسر" كتصحيح لقيمة "ف" باستخدام معامل "إبسلون"، والجدول رقم (١٢) يوضح ذلك.

جدول رقم (١٢): تحليل تباين القياسات المتعددة لقيم تقديرات القدرة من الاختبار الكلي والمقدرة من الاختبارات الثلاثة (ن=409)

مصدر التباين	الاختبار الاحصائي	مجموع المربعات	درجات الحرية	متوسط المربعات	قيمة (f)	مستوى الدلالة *
بين القياسات	Greenhouse-Geisser	0.387	2.270	0.171	2.325	0.091
داخل القياسات(الخطأ)	Greenhouse-Geisser	67.9 10	926.149	0.073		

* مستوى الدلالة ($\alpha \leq 0.05$)

يتضح من الجدول رقم (١٢) أن قيمة اختبار قرين هاوس–جيسر تساوي (2.325) وهي قيمة غير دالة إحصائياً عند مستوى الدلالة الاحصائية ($\alpha \leq 0.05$) مما يدل على عدم وجود فروق دالة بين متوسطات تقديرات القدرة المقدرة من الاختبار الكلي والتقديرات المقدرة من

الاختبارات الثلاثة التي تم انتقاء مفرداتها وفقاً لطريقة:(أقصى قدرة، وأقصى معلومات، والتمييز التقليدي).

ومما سبق؛ يمكن استنتاج أن تقديرات القدرة بالاختبارات الثلاثة التي تم انتقاء مفرداتها وفقاً لطريقة:(أقصى قدرة، وأقصى معلومات، والتمييز التقليدي) مماثلة للتقديرات التي قدمها الاختبار الكلي.

هل تختلف دقة تقدير قدرة الأفراد تبعاً لاختلاف طريقة انتقاء مفردات الاختبار التحصيلي: (أقصى قدرة، وأقصى معلومات، والتمييز التقليدي)؟ وللإجابة عن السؤال الثالث؛ تم تحديد دقة تقدير قدرة الأفراد من خلال مؤشر الخطأ المعياري لقياس القدرة (نور الدين، ٢٠٠٢) والجدول رقم (١٣) يوضح المتوسط والانحراف المعياري لقيم الأخطاء المعيارية لتقدير القدرة المقدرة باستخدام النموذج اللوغارتمي الثلاثي المعلم في تحليل مفردات الاختبارات الثلاثة التي تم انتقاءها وفقاً لطريقة:(أقصى قدرة، وأقصى معلومات، والتمييز التقليدي)

جدول رقم (١٣): المتوسط والانحراف المعياري لقيم الأخطاء المعيارية لتقدير القدرة باستخدام الاختبارات الثلاثة(ن=409)

الاختبار	المتوسط	الانحراف المعياري
الأول (أقصى قدرة)	**0.494**	**0.115**
الثاني (أقصى معلومات)	**0.443**	**0.131**
الثالث (التمييز التقليدي)	**0.440**	**0.128**

يتضح من الجدول رقم (١٣) أن هناك تقارب في قيم متوسطات الاخطاء المعيارية لتقديرات القدرة الناتجة عن تحليل الاختبار الثاني(أقصى معلومات) والاختبار الثالث(التمييز التقليدي) حيث بلغت قيمتها (0.443، 0.440) بانحراف معياري قدرة (0.131، 0.128) على الترتيب، بينما كانت قيمة المتوسط الناتجة عن تحليل الاختبار الأول(أقصى قدرة) تمثل أعلى قيمة حيث بلغت (0.494) بانحراف معياري(0.115). ولتحديد دلالة الفروق في قيم الاخطاء المعيارية لتقديرات القدرة الناتجة عن استخدام الاختبارات الثلاثة تم التحقق من دلالة تحليل التباين للقياسات المتعددة، والنتائج موضحة في الجدول رقم (١٤):

الجدول رقم (١٤): يوضح دلالة اختبار ماكولي لقيم أخطاء تقدير القدرة بالاختبارات الثلاثة.

قيمة اختبار ماكولي	مربع كاي التقريبي	درجات الحرية	مستوى الدلالة*
0.891	47.036	2	*0.00

* مستوى الدلالة ($\alpha \leq 0.05$)

يتضح من الجدول رقم (١٤) أن قيمة اختبار مربع كاي تساوي (47.036) وهي قيمة دالة إحصائياً عند مستوى الدلالة ($\alpha \leq 0.05$)، مما يشير إلى وجود فروق دالة بين تجانس القياسات الثلاثة الناتجة عن استخدام الاختبارات الثلاثة مما يدل على عدم تحقق شرط الدورية وبالتالي سيتم تطبيق الاختبار الاحصائي "جرين هاوس–جيسر" كتصحيح لقيمة "ف" باستخدام معامل "إبسلون"، والجدول رقم (١٥) يوضح ذلك.

جدول رقم (١٥): تحليل تباين القياسات المتعددة لقيم الاخطاء المعيارية لقيم تقديرات القدرة بالاختبارات الثلاثة ن=409

مصدر التباين	الاختبار الاحصائي	مجموع المربعات	درجات الحرية	متوسط المربعات	قيمة (F)	مستوى الدلالة *
بين القياسات	Greenhouse-Geisser	.750	1.8 03	.416	158.813	.000
داخل القياسات (الخطأ)	Greenhouse-Geisser	1.926	735.706	.003		

* مستوى الدلالة ($\alpha \leq 0.05$)

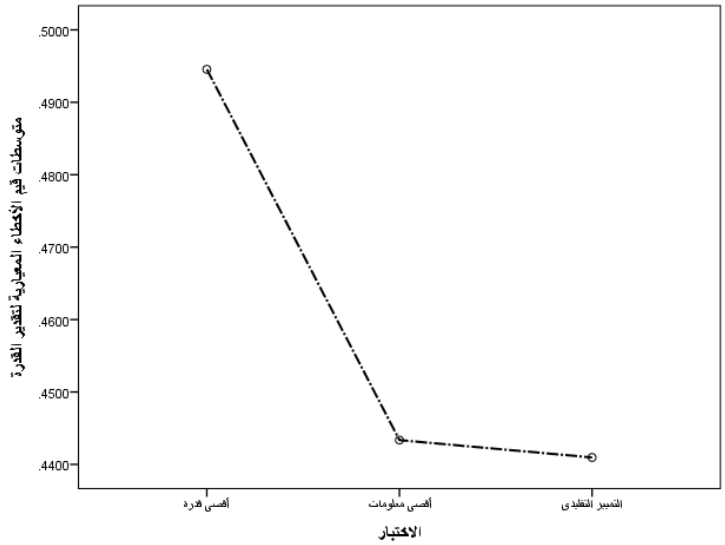
يتضح من الجدول رقم (١٥) أن قيمة اختبار قرينهاوس–جيسر تساوي (158.813) وهي قيمة دالة إحصائياً عند مستوى الدلالة الاحصائية ($\alpha \leq 0.05$) مما يدل على وجود فروق جوهرية بين متوسطات الاخطاء المعيارية لتقديرات القدرة الناتجة عن استخدام الاختبارات الثلاثة التي تم انتقاء مفرداتها وفقاً لطريقة:(أقصى قدرة، وأقصى معلومات، والتمييز التقليدي)، ولمعرفة وجهة الفروق سيتم تطبيق طريقة (LSD) لإجراء المقارنات الثنائية والجدول رقم (١٦) يوضح النتائج.

جدول رقم (١٦) اختبار (LSD) للمقارنات الثنائية لدلالة الفروق بين متوسطات قيم الاخطاء المعيارية لتقديرات القدرة الناتجة من الاختبارات الثلاثة

عنصر المقارنة الأول	عنصر المقارنة الثاني	الفرق بين متوسطي عناصر المقارنة	مستوى الدلالة *
الاختبار الأول (أقصى قدرة)	الاختبار الثاني (أقصى معلومات)	0.051	0.00 *
	الاختبار الثالث (التمييز التقليدي)	0.054	0.00 *
الاختبار الثاني (أقصى معلومات)	الاختبار الثالث (التمييز التقليدي)	0.002	1.00

* مستوى الدلالة ($\alpha \leq 0.05$)

يتضح من الجدول رقم(١٦) وجود فروق دالة إحصائياً عند مستوى الدلالة ($\alpha \leq 0.05$) بين متوسطي قيم الاخطاء المعيارية لتقديرات القدرة الناتجة عن استخدام الاختبار الأول(أقصى قدرة) وكل من الاختبار الثاني(أقصى معلومات)، والاختبار الثالث(التمييز التقليدي)، حيث بلغت قيمة الفروق(0.051، 0.054) على الترتيب، وجاءت لصالح متوسط قيم الاخطاء المعيارية لتقديرات القدرة الناتجة عن استخدام الاختبار الأول(أقصى قدرة)، أيضاً تشير النتائج إلى عدم وجود فروق دالة إحصائياً عند مستوى الدلالة ($\alpha \leq 0.05$) بين متوسطي قيم الاخطاء المعيارية لتقديرات القدرة الناتجة عن استخدام االاختبار الثاني(أقصى معلومات) والاختبار الثالث(التمييز التقليدي)، حيث بلغت قيمة الفروق(0.002). والشكل رقم(٦) يوضح هذه النتيجة.



الشكل رقم (٦): متوسطات قيم الاخطاء المعيارية لتقدير القدرة الناتجة عن الاختبارات الثلاثة التي تم انتقاء مفرداتها وفقاً لطريقة:(أقصى قدرة، وأقصى معلومات، والتمييز التقليدي)

من الشكل رقم (٦) يمكن ملاحظة أن قيمة متوسط الاخطأ المعيارية لتقديرات القدرة التي تعود للاختبار الأول (أقصى قدرة) كانت أعلى بشكل واضح من متوسطي قيم الاخطاء المعيارية لتقديرات القدرة الناتجة عن استخدام الاختبار الثاني(أقصى معلومات) والاختبار الثالث (التمييز التقليدي).

مما سبق يمكن استنتاج أن أعلى قيمة دالة لمتوسط الاخطأ المعيارية لتقديرات القدرة تعود للاختبار الأول (أقصى قدرة)، مما يدل على أن تقديرات القدرة باستخدام الاختبار الأول (أقصى قدرة) كانت أقل دقة من التقديرات التي قدمها الاختباران: الثاني (أقصى معلومات) والثالث (التمييز التقليدي)

السؤال الرابع: ما مدى اختلاف الكفاءة النسبية للاختبار في تقدير قدرة الأفراد لكل اختبار من الاختبارات الثلاثة التي تم انتقاء مفرداتها وفقاً لطريقة:(أقصى قدرة، وأقصى معلومات، والتمييز التقليدي)؟

للاجابة عن السؤال الرابع، تم عرض نتائج أهم الخصائص السيكومترية والإحصاءات الوصفية لمعلمات المفردة في كل اختبار من الاختبارات الثلاثة، وهي موضحة في الجدول رقم (١٧)

جـدول رقـم (١٧): أهـم الخصـائص السـيكومترية والإحصـاءات الوصـفية لمعلمـات الاختبارات الثلاثة بعد تحليلها بالنموذج اللوغارتمي الثلاثي المعلم (20) مفردة و(409) فرد

الخصائص	الإحصائي	الاختبار		
		الأول (أقصى قدرة)	الثاني (أقصى	الثالث (التمييز
دوال المعلومات للمفردات	المتوسط	0.581	0.993	0.932
	الانحراف	0.209	0.272	0.298
	(أقل – أعلى)	(1.022)-(0.217)	(1.501)-(0.603)	(1.5131)-(0.427)
القدرة	المتوسط	0.016	0.001-	0.001
	الانحراف	0.962	0.995	0.990
	(أقل – أعلى) قيمة	(2.484) – (1.426)	(2.278)- (1.383-)	(2.2461)-1.374-)
التمييز	المتوسط	1.085	1.417	1.365
	الانحراف	0.206	0.179	0.213
	(أقل – أعلى)	(1.412)-(0.605)	(1.712)-(1.158)	(1.716)-(0.948)
الصعوبة	المتوسط	0.840	0.685	0.680
	الانحراف	0.600	0.614	0.570
	(أقل – أعلى)	(2.006)-(0.068-)	(1.691)-(0.480)	(1.730)-(0.290)
التخمين	المتوسط	0.211	0.208	0.205
	الانحـــــراف	0.049	0.041	0.020
	(أقل – أعلى) قيمة	(0.325)- (0.097)	(0.150)-(0.346)	(0.239)-(0.167)

من الجدول رقم (١٧) يمكن ملاحظة أن أعلى قيمة لمتوسط دوال معلومات الاختبار كانت للاختبار الاختبار الثاني (أقصى معلومات) ثم الاختبار الثالث (التمييز التقليدي)، وأخيراً الاختبار الأول (أقصى قدرة)، وأعلى قيمة لمتوسط معامل صعوبة الاختبار كانت للاختبار الأول (أقصى قدرة) وبقيم متساوية تقريباً لكل من الاختبار الثاني (أقصى معلومات)

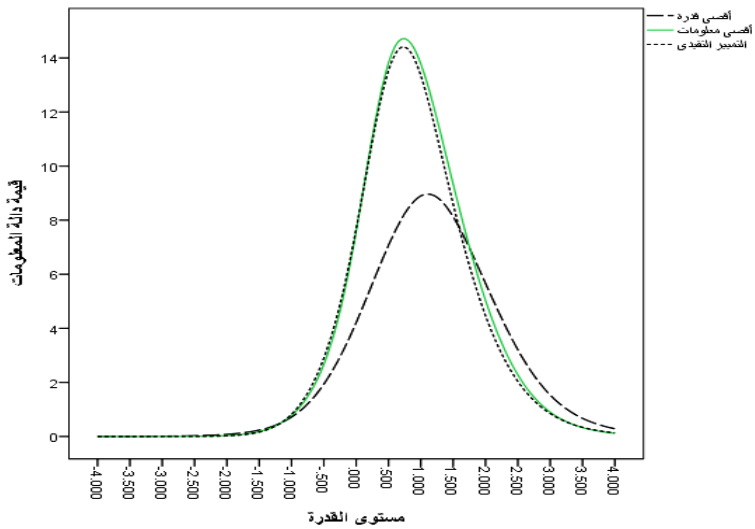
والاختبار الثالث (التمييز التقليدي)، وأعلى معلم تمييز كان للاختبار الثاني (أقصى معلومات). أما قيم متوسط تقدير القدرة للمفردات وكذلك معلم التخمين فقد جاءت متقاربة للاختبارات الثلاثة.

وتم عرض نتائج قيم دالة المعلومات للاختبار ومقياس الخطأ المعياري المشروط لتقدير القدرة ومقدار الكفاءة النسبية لكل اختبار من الاختبارات الثلاثة التي تم انتقاء مفرداتها وفقاً لطريقة:(أقصى قدرة، وأقصى معلومات، والتمييز التقليدي)، والنتائج موضحة في الجدول رقم(١٨).

جدول رقم (١٨): قيم دالة المعلومات للاختبار ككُل ومقياس الخطأ المعياري المشروط لتقدير القدرة ومقدار الكفاءة النسبية لكل اختبار من الاختبارات الثلاثة

طريقة انتقاء المفردات	قيمة دالة المعلومات	مقياس الخطأ المعياري	مستوى القدرة المقابلة	الكفاءة النسبية $\frac{\theta \max}{I(\theta \max)}$	الكفاءة النسبية $\frac{\theta \max}{Dmax}$	الكفاءة النسبية $\frac{I(\theta \max)}{Dmax}$
أقصى قدرة	8.967	0.334	1.100	0.7658	0.7860	1.0273
أقصى معلومات	11.719	0.261	0.750			
التمييز التقليدي	11.407	0.263	0.750			

من الجدول رقم (١٨) يمكن ملاحظة أن أعلى كمية معلومات عن الأفراد كانت تساوي (11.719) للاختبار الثاني (أقصى معلومات)، يليه الاختبار الثالث (التمييز التقليدي) بقيمة (11.407)، وأخيراً كانت أقل كمية معلومات مقدمة من الاختبار الأول (أقصى قدرة) حيث بلغت كمية المعلومات المقدمة منه (8.967)، ويمكن توضيح هذه النتيجة من الشكل رقم (٧).



شكل رقم (٧): دالة المعلومات للاختبارات الثلاثة

يتضح من الشكل رقم (٧) أن الفرق في كمية المعلومات التي يقدمها الاختبار جاءت لصالح الاختبار الثاني (أقصى معلومات) والثالث(التمييز التقليدي) بمقدار متقارب جدا مقارنة بالاختبار الأول عند جميع مستويات القدرة فيما عدا عند مستوى القدرة العليا من (2+) إلى(4+) فقد أظهر الاختبار الأول(أقصى قدرة) كفاءة أعلى. كما أن الاختبارين الثاني والثالث قد قدما أعلى قيمة لدالة المعلومات عند نقطة مقابلة للقيمة (٠.٧٥٠) على متصل القدرة، بينما قدم الاختبار الأول(أقصى قدرة)أعلى قيمة لدالة المعلومات عند نقطة مقابلة للقيمة(١.١٠٠).

وبالعودة للجدول رقم (١٨) نجد أن أقل قيمة لمقياس الخطأ المعياري المشروط لتقدير القدرة، تساوي (0.261، 0.263، 1.100) للاختبار الثاني (أقصى معلومات)، وللاختبار الثالث (التمييز التقليدي) وللاختبار الأول (أقصى قدرة) على الترتيب. وبمقارنة قيم الكفاءة النسبية بين كل اختبارين في قياسهما للسمة نجد أن: قيمة الكفاءة النسبية بين الاختبار الأول(أقصى قدرة) والاختبار الثاني (أقصى معلومات) تساوي(0.7658)، وقيمتها بين الاختبار الأول(أقصى قدرة) والاختبار الثالث(التمييز التقليدي) تساوي(0.7860)، وبين الاختبار الثاني (أقصى معلومات) والاختبار الثالث(التمييز التقليدي) تساوي(1.0273). وبالتالي يمكن ترتيب الاختبارات المكونة من (20) مفردة، بحسب قيم كل من دالة معلومات الاختبار والكفاءة النسبية من الأعلى دقة على النحو التالي: الاختبار الثاني (أقصى

معلومات)، والاختبار الثالث(التمييز التقليدي)، الاختبار الأول(أقصى قدرة)، مع ملاحظة وجود تقارب كبير بين مؤشرات الاختبار الثاني (أقصى معلومات)، والاختبار الثالث(التمييز التقليدي).

مناقشة النتائج:

في هذه الدراسة تم إعداد اختبار تحصيلي مكون من(٦٠) مفردة، من نوع الاختيار من متعدد بأربع بدائل في مقرر التقويم التربوي وتم التحقق من خصائصه السيكومترية وتم تدريجه وفقاً للنموذج اللوغارتمي الثلاثي المعلم، ومن ثم تم القيام بتجريب استخدام ثلاث طرق لانتقاء (٢٠) مفردة من تلك المفردات، وهي: طريقتي (أقصى قدرة) و(أقصى معلومات) من طرق انتقاء المفردات في نظرية الاستجابة للمفردة، وطريقة(التمييز التقليدي) هذه من طرق النظرية التقليدية في القياس، بهدف التعرف على أثرها في دقة تقدير القدرة والكفاءة النسبية للاختبار التحصيلي. وكان من

أهم النتائج، ما يلي:

أولاً: نتائج مقارنة خصائص الاختبار الكلي مع الاختبارات الثلاثة التي تم اشتقاقها من الاختبار الكلي عن طريق انتقاء مفرداتها وفقاً لطريقة:(أقصى قدرة، وأقصى معلومات، والتمييز التقليدي) وجد أن الاختبارات الثلاثة المكونة من (٢٠) مفردة تتكافأ في خصائصها السيكومترية (الصدق والثبات) ومؤشراتها الاحصائية مع الاختبار الكلي. كما أن النتائج أشارت إلى أنه لا توجد فروق دالة بين تقدير القدرة التي قدمها الاختبار الكلي وتقدير القدرة بالاختبارات الثلاثة. أي أن خصائص الاختبارات وتقديرات القدرة لم تتأثر بالفروق في عدد المفردات بين الاختبارات الثلاثة والاختبار الكلي والتي كانت بنسبة(١ :٣)، كما لم تتأثر جوهرياً باختلاف طرق انتقاء المفردات. كما أنها أوضحت بأن الاختبارات الثلاثة تملك توزيع قدرة قريب من الاعتدالية ومشابهه تقريباً لتوزيع القدرة الذي يملكة الاختبار الكلي وأن الاختبارات الثلاثة غطت مدى للقدرة قريب من المدى الذي تمت تغطيته بالاختبار الكلي حيث كانت قيم مدى القدرة تساوي(4.08، 3.91، 3.66، 3.62) للاختبار الكلي والاختبارات الثلاثة :(أقصى قدرة، وأقصى معلومات، والتمييز التقليدي) على الترتيب، وتجدر الإشارة هنا للتقارب الملحوظ بين مدى القدرة للاختبار الكلي والاختبارالثاني(أقصى معلومات).

وهذه النتيجة تختلف إلى حدٍ ما مع الأدب السيكومتري الذي أشار إلى أن هناك علاقة طردية بين طول الاختبار وجودة خصائصه السيكومترية(انستازي ويوريان، ٢٠١٥)، إلا أنها

تتسق من ناحية أخرى مع الهدف الرئيس الذي من أجله تم تطوير طرق وإجراءات دقيقة لإنتقاء المفردات للتوصل إلى أعلى قيمة ممكنة لدالة معلومات الاختبار؛ أي زيادتها إلى الحد الأعلى الذي يسمح بتقدير قدرة المختبرين بطريقة دقيقة على تناسب توزيع قدرتهم على متصل السمة. ولعل هذا ما يبرر الاختلاف في النتجة الحالية حيث أن الاختبارات الثلاثة لا يمكن النظر إليها بأنها اختبارات أقصر وحسب، بل لا بد من الأخذ في الحسبان أنه قد تم انتقاء مفرداتها وفقاً لطرق أثبتت فاعليتها في رفع جودة الاختبار واختزال مفرداته. وهذه النتيجة تتفق مع نتائج دراسة (Chen & Ankenman, 2004) التي قارنت بين ثلاث طرق من طرق أقصى معلومات(فيشر، فيشر بتوزيع بعدي، دالة كولباك بتوزيع بعدي)، والطريقة العشوائية، وأظهرت أن طرق أقصى معلومات على اختلافها قدمت اختبارات جيدة حتى مع الاختبارات الأقل من (١٠) مفردات.

وهذه النتيجة تتفق أيضاً مع ما توصلت إليه دراسة الشافعي(٢٠١٢) التي طبقت طريقة أقصى معلومات لانتقاء تسعة اختبارات بأطوال مختلفة تراوحت ما بين (٩) إلى(٧٢) مفردة من اختبار أوتيس لينون للذكاء المكون من (٨٠) مفردة واستطاع اختصار أطوال الاختبار إلى عدد أقل من المفردات تصل إلى (٣٦)مفردة، وتقدم نفس تقديرات القدرة التي تم الحصول عليها من الاختبار الأصلي

ثانياً: نتائج مقارنة الاختبارات الثلاثة من حيث دقة تقدير القدرة والكفاءة النسبية؛ أشارت النتائج إلى وجود فروق في دقة تقديرات القدرة بالاختبارات الثلاثة التي تم انتقاء مفرداتها وفقاً لطريقة:(أقصى قدرة، وأقصى معلومات، وأعلى قيمة تمييز). وجاءت الفروق لصالح **الاختباران:** الثاني(أقصى معلومات) والثالث(التمييز التقليدي)، حيث أن تقديرات القدرة باستخدام الاختبار الأول(أقصى قدرة) كانت أقل دقة من التقديرات التي قدمها الاختبار الثاني(أقصى معلومات)، وكذلك كانت أقل دقة من التقديرات التي قدمها الاختبار الثالث(التمييز التقليدي)، حيث أن أعلى قيمة لمتوسط الخطأ المعيارية لتقدير القدرة تعود للاختبار الأول(أقصى قدرة). بينما لا توجد فروق بين دقة تقديرات القدرة التي قدمها **الاختباران:** الثاني(أقصى معلومات) والثالث(التمييز التقليدي)، وبالنسبة لتقدير الكفاءة النسبية للاختبارات الثلاثة، أسفرت النتائج عن إمكانية ترتيب الاختبارات الثلاثة المكونة من (٢٠) مفردة، بحسب الكفاءة النسبية من الأعلى دقة على النحو التالي: الاختبار الثاني(أقصى معلومات)، والاختبار الثالث(التمييز التقليدي)، الاختبار الأول(أقصى قدرة)، مع ملاحظة وجود تقارب كبير بين مؤشرات الاختبار الثاني(أقصى معلومات)، والاختبار الثالث(التمييز

التقليدي). ويمكن تفسير نتائج المقارنة هذه من كون مقدار الخطأ المعياري في تقدير القدرة ($SE(\theta)$) يرتبط بعلاقة عكسية مع كمية معلومات الاختبار(Moghadamzadeh et al., 2011). وهذا بدوره يؤثر أيضاً على ثبات الاختبار من العلاقات التالية(بيكر، ٢٠١٠؛ انستازي ويوريان، ٢٠١٥):

$$SE(\theta) = \frac{1}{\sqrt{I(\theta)}} \dots (8)$$

$$r_{xx} = 1 - SE(\theta)^2 = 1 - \left(\frac{1}{I(\theta)}\right) \dots\dots (9)$$

حيث أن: r_{xx} معامل ثبات الاختبار و $I(\theta)$ دالة معلومات الاختبار و $SE(\theta)$ القدرة لتقدير المعياري الخطأ

وقد جاءت قيم الخطأ المعياري لتقدير القدرة كالتالي: (0.494، 0.443، 0.440) وكمية المعلومات: (8.967، 11.719، 11.407)، وقيم معامل ثبات ألفا لكرنباخ: (0.856، 0.905، 0.904) لكل من الاختبارات الثلاثة: (أقصى قدرة، وأقصى معلومات، والتمييز التقليدي) على الترتيب، مما يوضح كون الفروق في الدقة جاءت لصالح الاختباران: الثاني(أقصى معلومات) والثالث(التمييز التقليدي). أيضاً يوضح كون الكفاءة النسبية جاءات لصالح الاختبار الثاني(أقصى معلومات)، ثم الاختبار الثالث(التمييز التقليدي)، وأخيراً الاختبار الأول(أقصى قدرة). أيضا يمكن ملاحظة التقارب الواضح بين مؤشرات الاختبار الثاني(أقصى معلومات)، والاختبار الثالث(التمييز التقليدي).

وقد جاءت طريقة انتقاء المفردات حسب قيمة أقصى معلومات في المرتبة الأولى، عند مقارنة خصائصه بخصائص الاختبار الكلي حيث يملك مدى للقدرة قريب من مدى القدرة للاختبار الكلي، ومن جانب آخر تفوق الاختبار الثاني(أقصى معلومات) على الاختبار الأول(أقصى قدرة)، والاختبار الثالث(التمييز التقليدي)، من حيث كمية دالة المعلومات للاختبار؛ ودقة تقدير القدرة والكفاءة النسبية للاختبار، وهذا قد يعود إلى طريقة انتقاء المفردات نفسها، حيث تم انتقاء المفردات التي تحمل أعلى قيمة لدالة المعلومات في كل مستوى من مستويات القدرة (الأرباعيات)؛ والذي بدوره يؤدي إلى تناقص قيمة الخطأ المعياري لتقدير القدرة(دي إيالا، ٢٠١٧)، وبالتالي سيحقق الاختبار أعلى دقة لتقدير قدرة الافراد.

وهذه النتيجة تتسق مع ما أكده هاملتون وآخرون (١٩٩١) Hambleton, et al., من أن الطريقة القوية والفعالة لانتقاء مفردات الاختبار ومقارنة الاختبارات في إطار نظرية الاستجابة للمفردة تقدم من خلال معلومات المفردة؛ حيث أنه في هذه الطريقة يتم انتقاء المفردة على أساس معلومات الاختبار والتي تعد المؤشر الأساسي لدقة القياس في نظرية الاستجابة للمفردة والتي هي عكس الخطأ المعياري لتقدير القدرة(بيكر، ٢٠١٠). وكما عبر عنها هاميلتون وكوك (Hambelton & Cook, 1977) فإنه يمكن من خلال هذه الطريقة تحديد كمية المعلومات لكل مفردة عند مستوى قدرة محدد ومن ثم الوصول إلى الشكل المستهدف لكمية المعلومات للاختبار ككل. أي الوصول إلى أكبر كمية معلومات مقدمة من الاختبار عند جميع مستويات القدرة للمختبرين.

من جانب آخر؛ اتضح من نتائج الدراسة الحالية مناسبة طريقة أقصى معلومات جميع قدرات المختبرين باختلاف العينة المراد تطبيق الاختبار عليها، حيث أظهرت دقة عالية مماثلة للاختبار الكلي(يملكان نفس قيمة الثبات) وغطت نفس مدى القدرة تقريباً. ومما يؤيد هذه النتيجة أن طريقة أقصى معلومات هي الطريقة المتبعة في انتقاء المفردات في الاختبارات الحاسوبية التكيفية التي أثبتت كفاءتها في الدراسات السابقة وهي في الأساس صُممت للتحسين من كفاءة الاختبارات التحصيلية(Weiss & Kingsbury, 1984; Lunz, & Bergstrom, 1994). أيضاً أثبتت هذه الاختبارات التكيفية والمعتمدة على طريقة أقصى معلومات في انتقاء مفرداتها جدارتها في عملية القياس؛ لحساسيتها الجيدة لاستجابات المختبرين مختلفي القدرة (Lunz & Bergstrom, 1994؛ Halkitis, 1998).

وهي بالتالي طريقة جيدة لعمل الاختبارات الخطية الحاسوبية أو التقليدية ولكن بنفس الأسلوب المتبع من خلال الاختبارات الحاسوبية المتكيفة؛ لمحاولة الوصول إلى نفس دقتها في تقدير قدرات المختبرين، لأن طريقة المعلومات القصوى تعطي أقل خطأ ممكن في تقدير القدرة عند مستوى القدرة المستهدف للمفردة؛ وذلك لكون الخطأ في تقدير القدرة يتناسب عكسياً مع كمية المعلومات المقدمة، مع إمكانية توفير درجة ثبات عالية بأقل عدد ممكن من المفردات(Riley et al., 2007 ؛Lunz, & Bergstrom, 1994; Halkitis, 1998).

وبالتالي يمكن بتطبيق طريقة أقصى معلومات الوصول إلى مستوى دقة الاختبارات التكيفية التي تتميز بكونها تقيس بدقة بحسب قدرة المختبرين(Murphy et al., 2010 ; Yao, 2019)؛ أي أن كل مفردة تم اختيارها تكون مُصممة أو تم بنائها في الأساس لتقدير قدرة محددة، وبعد الانتهاء من عمليات التقدير يكون تجمع المفردات مُحمل بالمفردات

المتنوعة للقدرات المختلفة، وعادةً تتطلب هذه الطريقة أسئلة أقل للحصول على نفس الكمية من المعلومات (Riley et al., 2007) وهذا يحقق الكفاءة في الاختبار المعطى، وهذا بالضبط ما وجده كل من(Eggen & Straetmans, 2000; Kalender, 2012) حيث توصلوا إلى أن اختيار المفردة الأكثر دقة في تقدير السمة للأفراد قد وفر درجة أعلى من الدقة في التقدير بالرغم من استخدام عدد مفردات أقل، وقد أوصوا باستخدام طريقة المعلومات القصوى لانتقاء المفردات. وكذلك في دراسة ياو (2019) Yao الذي استخدم طريقة المعلومات القصوى في انتقاء مفردات اختبار متعدد الأبعاد ووجد أنها تعطي أقل قيمة لخطأ التقدير. وبذلك تكون هذه الطريقة مشابهة لطريقة انتقاء المفردات في الاختبارات التكيفية.

وقد وُجد أيضاً من خلال نتائج هذه الدراسة؛ أن طريقة انتقاء المفردات المعتمدة على مؤشر التمييز التقليدي جاءت في المرتبة الثانية وكانت متقاربة جدا مع طريقة أقصى معلومات، وقد يعود ذلك لأنه تم انتقاء جميع المفردات عالية التمييز بعد ترتيبها تنازلياً في كل فترة من مستويات القدرة (الارباعيات) بغض النظر عن مستوى القدرة المقابل لهذه المفردات والتي يقابلها مؤشر الصعوبة أيضاً، كما أن مؤشر التمييز هنا قد تم حسابه عن طريق معامل الارتباط الثنائي المتسلسل والذي يقدم نتائج مستقلة عن الصعوبة(انستازي ويوريان، ٢٠١٥)، وبذلك جاء الاختبار ذو ال(٢٠) مفردة بأعلى مستوى تمييز؛ مما أدى إلى أن جميع مفردات الاختبار اشتركت مع بعضها عن طريق مؤشر التمييز للتوصل إلى درجة كلية مُميزة لجميع فئات المختبرين (Kelley & Linacre, 2002). وبالتالي أصبح الاختبار مميزاً بين الفئات المختلفة من ذوي القدرة المرتفعة أو المتوسطة أو المنخفضة وتصنيفهم تبعاً لذلك، وهذا بالضبط ما أشار إليه كلارك وواطسون (Clark & Watson, 1995) في عملية انتقاء المفردات للاختبار من أجل تحقيق الصدق حيث أجمعوا أن أحد أهم الأسس في ذلك هو قوة المفردات على التمييز بشكل مناسب بين المجموعات. وبذلك تكون المفردات التي تعطي أقصى كمية معلومات عن الفرق في مستوى التحصيل، هي ذاتها التي تُساهم بارتفاع ثبات الاختبار (Ebel & Frisbie, 1986)، وبما أن هدفنا هو زيادة ثبات(دقة) الاختبار لأقصى درجة مُمكنة لذلك يجب اختيار أكثر المفردات تَمييزاً فيه.

وأما عند النظر للنتيجة التي أشارت إلى وجود تقارب كبير بين مؤشرات الاختبار الثاني(أقصى معلومات)، والاختبار الثالث(التمييز التقليدي). فقد يعود إلى أن طريقة أقصى معلومات ترتبط بعلاقة طردية مع مؤشر التمييز؛ لأنه يتم الاعتماد هنا على كمية دالة المعلومات للمفردة فإذا أعطت المفردات تمييزاً عالياً عند مستوى قدرة معين؛ فإن دالة

المعلومات للاختبار ككل ستكون عالية والعكس صحيح(نور الدين، ٢٠٠٢) أيضاً لأنه_في هذه الدراسة_ عند الانتقاء تم تقسيم مستوى القدرة إلى إرباعيات. وبالتالي زادت إمكانية جعل جميع المفردات تَعمل بطريقة جيدة في مستويات القدرة المختلفة والتي من الممكن لمُعد الاختبار أن يحددها إما إرباعيات أو تساعيات أو على حسب هدفه من القياس(هياجنة، ٢٠١١؛ الشافعي، ٢٠١٢)، وتتفق هذه النتيجة مع ما أشار إليه لونغ وآخرون Leung et al. (2002) من أن استخدام المفردات الأكثر تمييزاً عند انتقاء المفردات يقلل أخطاء التقدير.

أما طريقة أقصى قدرة فقد أظهرت دقة أقل في تقدير قدرات المختبرين وجاءت في المرتبة الأخيرة، وقد يعود ذلك إلى طريقة الانتقاء ذاتها حيث أنها تعتمد على انتقاء المفردات وفق معيار أعلى قيمة لتدرج القدرة المقابلة لموقع قمة دالة المعلومات في مستوى القدرة المحدد بغض النظر عن مدى ارتفاع تلك القمة، ومن ناحية أخرى لكون هذه الطريقة تتأثر بشكل أساسي بمعلم صعوبة المفردة باعتباره انعكاس للقدرة. وبالنظر إلى قيم صعوبة مفردات الاختبار الأول(أقصى قدرة) نجد أنها تتراوح ما بين (-٠.٠٠٦) إلى(٢.٠٠٦) بمتوسط (0.840) وانحراف معياري(0.600) أي أن الاختبار يميل إلى الصعوبة لأن قيم معلم الصعوبة غطت مدى القدرة الأعلى من المتوسط فقط.

أيضاً بالعودة إلى شكل(٧) تتضح هذه النتيجة حيث تبين أنه عند رسم دوال المعلومات للاختبارات الثلاثة جاء الفرق في كمية المعلومات التي يقدمها الاختبار لصالح الاختبار الثاني(أقصى معلومات) والثالث(التمييز التقليدي) بمقدار متقارب جدا مقارنة بالاختبار الأول(أقصى قدرة) عند جميع مستويات القدرة فيما عدا عند مستوى القدرة العليا من (2+) إلى(4+) فقد أظهر الاختبار الأول(أقصى قدرة) كفاءة أعلى، وذلك قد يعود لسببين أولاً؛ لأن الاختبار الأول (أقصى قدرة) يميل إلى الصعوبة أكثر من الاختبارين الاخرين، ثانياً؛ قد يعود لطريقة انتقاء المفردات التي تعتمد على اختيار أعلى نقاط على المتصل في كل مستوى من مستويات القدرة(إرباعي).

ومن الممكن زيادة فعالية هذه الطريقة إذا كان تجمع المفردات يحتوي على مفردات متجانسة في خصائصها السيكومترية وكمية المعلومات في كل مستوى من مستويات القدرة، لكن إذا تنوعت مؤشرات المفردات التي تغطي مدى معين من القدرة، فإن هذه الطريقة قد تعطي دقة أقل، نظراً لتأثر عملية القياس بمصادر الخطأ(Moghadamzadeha et al., 2011, 1367).

وختاماً؛ يمكن استنتاج أن الاختبارات الثلاثة استطاعت تقدير القدرة بنفس جودة الاختبار الكلي، ولكن بالتقصي الاحصائي لمؤشرات الدقة (الخطأ المعياري لتقدير القدرة،

وقيمة دالة المعلومات) أمكن المفاضلة بينها لصالح طريقة أقصى معلومات وطريقة التمييز التقليدي؛ وعلى العموم؛ تتفق نتائج الدراسة الحالية مع نتائج دراسة (Gierl et al. 2001)، ودراسة هياجنة(٢٠١١) التي بحثت في أثر طريقة انتقاء المفردات (أقصى قدرة، وأقصى معلومات، العشوائية) على كفاءة الاختبارات، وأظهرت النتائج أن طريقة انتقاء المفردات بأقصى معلومات وبالطريقة العشوائية أعطت كمية من المعلومات ودقة في تقدير القدرة أعلى من طريقة أقصى قدرة.

كما تتسق مع نتائج دراسة الدلالعة(٢٠٠٤) التي وجدت أن الاختبار الذي تم انتقاء مفرداته وفقاً لأقصى معلومات قدم أعلى كمية معلومات وأقل خطأ في تقدير القدرة وأعلى كفاء نسبية يليه الاختبار الذي تم انتقاء مفرداته وفقاً لطريقة التمييز التقليدية وأخير الاختبار الذي تم انتقاء مفرداته وفقاً للطريقة العشوائية.

وأيضاً تتفق مع ما أشار إليه الشافعي (٢٠١٢) من أن انتقاء المفردات وفقاً لأقصى معلومات لا ختبارات مختلفة الأطوال يؤدي إلى رفع الكفاءة للإختبار مما يؤدي إلى تقديرات أكثر دقة للقدرة. كما تتفق مع نتائج دراسة الجراح (٢٠١٥) التي أظهرت نتائجها بأنه توجد فروق دالة احصائياً بين متوسطات قيم دالة المعلومات تعزى لمعامل تمييز المفردات. كما أظهرت النتائج وجود علاقة عكسية بين نسبة الخطأ المعياري وكلٍ من حجم العينة ومعامل التمييز، وعلاقة طردية مع معامل الصعوبة عند ما يتم تقدير القدرة (θ) باستخدام الاختبار.

وتختلف النتائج مع ما توصلت إليه دراسة البرصان(٢٠١٨) التي هدفت إلى تقصي فاعلية خمس استراتيجيات ثانوية لاختيار مفردات الاختبارات التكيفية متعددة المراحل إضافة للاستراتيجية الرئيسة القائمة على صعوبة المفردة وتواؤمها مع قدرة المفحوص، هذه الاستراتيجيات الخمس تعتمد على: دالة المعلومات، معلم التمييز، أقل معلم تخمين، إحصائي المطابقة، والاختيار العشوائي. وقد أظهرت النتائج أن الاستراتيجية المعتمدة على إحصائي المطابقة هي الأفضل في كل المراحل الثلاث تليها طريقة الاختيار العشوائي ثم معلم التمييز. ولم تأخذ الاستراتيجيات القائمة على دالة المعلومات ومعلم التخمين أفضلية في كافة المراحل دراسة. وقد يعود هذا الاختلاف لكون طريقة الانتقاء بأقصى معلومات كانت طريقة ثانوية بينما كانت الطريقة الرئيسية على صعوبة المفردة.

التوصيات والدراسات المستقبلية المقترحة:

بناء على أهم النتائج التي توصلت إليها الدراسة، يمكن الخروج بالتوصيات والمقترحات التالية:

- عند انتقاء مفردات الاختبار التحصيلي وفقاً لطرق نظرية الاستجابة للمفردة يوصى باستخدام طريقة أقصى معلومات على متصل القدرة؛ حيث أنه ثبتت فعاليتها وكانت أدق من طريقة أقصى قدرة.
- انتقاء المفردات باستخدام طريقة التمييز التقليدي_ حيث أنها حققت دقة مقاربة لطريقة أقصى معلومات_ خاصة عند إعداد نماذج اختبارية في حالة اختبارات المعلم لسهولة حسابهاعن طريق البرامج الشائعة مثل برنامج Microsoft Excel وغيره من البرامج المتاحة.
- يوصى باستخدام طريقة أقصى معلومات في انتقاء المفردات عند ما تتجانس قدرات المختبرين، حيث أنها تقدم دقة جيدة لتقدير القدرة بصرف النظر عن درجة تباين مدى قدرة المختبرين.
- ضرورة التحقق من تجانس الخصائص السيكومترية وكمية المعلومات قبل انتقاء المفردات بطريقة أقصى قدرة حيث أن دقتها قد تقل بوجود تباين في خصائص المفردات المكونة لتجمع المفردات.
- إجراء دراسة لأثر اختلاف حجم العينة في دقة تقدير معالم المفردة والكفاءة النسبية للاختبار التحصيلي الذي تم انتقاء مفرداته وفقاً لطريقتي التمييز التقليدي وأقصى معلومات.
- إجراء دراسة لأثر اختلاف النموذج المستخدم لتدريج تجمع المفردات (الأحادي، الثنائي، الثلاثي) في دقة تقدير القدرة للاختبار التحصيلي الذي تم انتقاء مفرداته وفقاً لطريقتي التمييز التقليدي وأقصى معلومات.

قائمة المراجع

أولاً :المراجع العربية

الأسمر، مصطفى محمد عبدالله (٢٠٠٧). الخصائص السيكومترية لصورتي اختبار في الرياضيات فقرات إحداهما ثنائية التدريج والأخرى متعددة التدريج وفق نظريتي القياس التقليدية والحديثة (رسالة دكتوراه غير منشورة). جامعة عمان العربية. الأردن.

أنستازي، أ؛ يوريان، س. (٢٠١٥). القياس النفسي (ترجمة، صلاح الدين محمود علام). دار الفكر. (٢٠١٥/١٩٩٧)

البرصان، إسماعيل سلامة (٢٠١٨). تقصي فاعلية خمس استراتيجيات ثانوية في اختبار فقرات القياس التكيفي متعدد المراحل. المجلة التربوية الدولية المتخصصة، ٧(١٠)، ٤٠-٥٣

بيكر؛ فرانك ب. (٢٠١٠). أسس نظرية الاستجابة للمفردة (ترجمة، عبد الرحمن الطريري، والسيد أبو هاشم، وسوسن شلبي) جامعة الملك سعود للنشر. (٢٠١٠/٢٠٠١).

الجراح، خلدون محمد فهد (٢٠١٥). أثر حجم العينة وصعوبة الفقرة وتمييزها على دالة المعلومات والخطأ المعياري حسب نظرية الاستجابة للفقرة: النموذج ثنائي المعلمة (رسالة ماجستير غير منشورة). جامعة مؤتة. الأردن.

الجمل، هبة محمد إبراهيم؛ طاحون، حسين حسن حسين؛ دسوقي، شيرين محمد أحمد؛ والناغي، هبة إبراهيم محمد علي (٢٠١٧). تدريج بنك أسئلة مادة الرياضيات لتلاميذ الصف الثالث الإبتدائي باستخدام النموذج اللوغاريتمي الثلاثي المعلم. مجلة كلية التربية، (٢١)، ٨٣٧ -٨٦٧

حرز الله، علية محمد (٢٠٠٤). بناء بنك أسئلة في الرياضيات والتحقق من فاعليته في انتقاء فقرات اختبار محكي المرجع في مستوى امتحان شهادة الدراسة الثانوية العامة في الأردن (رسالة دكتوراه غير منشورة). جامعة عمان العربية. الأردن

الخطيب، عبد الله يونس فؤاد (٢٠١٣). مدى التوافق بين النظرية التقليدية في القياس ونظرية الاستجابة للفقرة في انتقاء فقرات اختبار احكام التلاوة والتجويد (رسالة ماجستير غير منشورة). جامعة اليرموك. الأردن

الدلالعة، محمد محمود سليمان (٢٠٠٥). مقارنة فاعلية ثلاث طرق في انتقاء فقرات اختبار رياضيات محكي المرجع في المرحلة الأساسية (رسالة دكتوراه غير منشورة). جامعة عمان العربية. الأردن

دي إيالا، ر. ج. (٢٠١٧). النظرية والتطبيق في نظرية الاستجابة للمفرد (ترجمة، عبد الله زيد الكيلاني، واسماعيل سلامة البرصان). دار جامعة الملك سعود للنشر. (٢٠١٧/٢٠٠٩)

الربابعة، صفاء محمد أحمد (٢٠١٩). بناء بنك أسئلة في أحكام التلاوة والتجويد باستخدام النموذج أحادي المعلمة (رسالة ماجستير غير منشورة). جامعة اليرموك. الأردن.

سخيم، خالد أحمد حسن (٢٠١٨). تصور مقترح لإعداد بنك أسئلة في ضوء جدول المواصفات وخصائص الاختبار الجيد: تطبيقا على مقرر العلوم الصف السادس بمرحلة التعليم الأساسي باليمن (رسالة ماجستير غير منشورة). جامعة القرآن الكريم والعلوم الإسلامية. اليمن.

الشافعي، محمد منصور محمد (٢٠١٢) اختزال مفردات الاختبارات التقليدية الخطية في ضوء متغيري أقصى قيمة للمعلومات وعدد المفردات. مجلة كلية التربية،٢٣(٩٠)، ١٠٢ – ١٧٢

الشريف، خالد حسن بكر (٢٠١٨). فاعلية نموذج راش في تدريج مفردات اختبار مقرر مهارات التعلم والتفكير على عينة من طلاب جامعة الملك فيصل: دراسة مقارنة مع النظرية الكلاسيكية. مجلة العلوم الإنسانية والاجتماعية، (٥٠)، ٨٩-١٧٤

الصبحي، محمد علي حميد (٢٠٠١). بناء اختبار محكي المرجع لقياس الكفايات الرياضية في المفاهيم الهندسية للمرحلة الابتدائية في مدارس مكة المكرمة(رسالة ماجستير غير منشورة). جامعة أم القرى. السعودية.

علام، صلاح الدين محمود (٢٠٠٥). نماذج الاستجابة للمفردة الاختبارية أحادية البعد ومتعددة الأبعاد وتطبيقاتها في القياس النفسي والتربوي. دار الفكر العربي.

العمري، حسان غازي (٢٠٠٥). مقارنة ثلاث طرق لانتقاء المفردات في نظرية الاستجابة للمفردة (دراسة في تدريج الأفراد) (رسالة دكتوراة غير منشورة). جامعة اليرموك. الأردن.

كروكر، ليندا؛ الجينا، جيمس (2017). مدخل إلى نظرية القياس التقليدية والمعاصرة (ترجمة، هند الحموري، وزينات دعنا). دار الفكر. (٢٠١٧/2008)

ليفنجستون، رونالد ب.؛ رينولدز، سيسيل ر. (٢٠١٣). إتقان القياس النفسي الحديث (ترجمة، صلاح الدين محمود علام). دار الفكر. (٢٠١٣/٢٠١٢)

محمد، شحتة عبد المولى عبد الحافظ (١٩٩٩). تقويم بناء الاختبارات المرجعية للمحك / المعيار في ضوء نظرية الاستجابة للمفردة (رسالة دكتوراه غير منشورة). جامعة عين شمس. مصر.

محمد، محمد جاد محمد (٢٠١٦). دراسة سيكومترية في دقة بناء بنك أسئلة في مادة الرياضيات باستخدام بعض نماذج نظرية الاستجابة للمفردة. عالم التربية، ١٧(٥٤)، ٢٦٣ -٢٦٩

مرشود، محمد فايق سالم (٢٠١٤). بناء بنك أسئلة محوسب في الإحصاء لطلبة الكليات الإنسانية في جامعة النجاح الوطنية – فلسطين باستخدام نموذج موكن اللابارامتري (رسالة دكتوراه غير منشورة). جامعة اليرموك، الأردن.

مهرنز، وليم؛ ولهمن، ايفن (٢٠٠٤). القياس والتقويم في التربية وعلم النفس (ترجمة، هيثم الزبيدي). دار الكتاب الجامعي. (٢٠٠٣/٢٠٠٤)

الموسوي، نعمان محمد صالح (٢٠١٦). تطوير بنك أسئلة في مقرر التقويم التربوي للمرحلة الجامعية باستخدام النظرية الكلاسيكية للقياس ونظرية الاستجابة للمفردة. مجلة كلية التربية، ٦٤(٤)،٦٣٥-٦٧٦

نيتكو، انثوني؛ بروخارت، سوزان (٢٠١٢). التقييم التربوي للطلبة (ترجمة، علي القرني، إبراهيم الدوسري، راشد المحرزي، وحسين الخروصي) مكتب التربية العربي لدول الخليج. (٢٠٠٧/٢٠١٢)

النجار، نبيل جمعه صالح (٢٠٠٨). بناء بنك أسئلة في الثقافة الحاسوبية للمرحلة الثانوية في الأردن باستخدام نظرية استجابة الفقرة نموذج راش والنظرية الكلاسيكية. مجلة كلية التربية، ١(٣٢)، ٤٩٣-٥٢٦

النعيمي، عز الدين عبدالله (٢٠١٥). معالم الفقرات والأفراد وخاصية اللاتغير فى الإختبارات الوطنية لضبط جودة التعليم فى الأردن: مقارنة بين النظرية الكلاسيكية والنظرية الحديثة فى القياس. مجلة اتحاد الجامعات العربية للتربية وعلم النفس،١٣ (١)، ١٣٦ – ١٥٥

نور الدين، محمد أمين صبري (2002). فاعلية استخدام الاختبار الموائم باستخدام الحاسب في تقدير قدرة الأفراد وتحديد الخصائص السيكومترية للمقياس (رسالة دكتوراه غير منشورة). جامعة عين شمس. مصر.

هياجنة، علي ذيب حمد (٢٠١١). أثر طريقة انتقاء الفقرات في دالة معلومات الاختبار (رسالة دكتوراه غير منشورة). جامعة اليرموك. الأردن

ثانياً: المراجع الانجليزية

Allam, Salah al-Din Mahmoud (2005). One-Dimensional and Multi-Dimensional Item Response Models for the Test and Their Applications in Psychological and Educational Measurement. Dar Alfikr Alearabi.

Al-Amri, Hassan Ghazi (2005). A Comparison of Three Item Selection Methods in the Item Response Theory (Study in Personnel Calibration) (Ph. D, Dissertation). Yarmouk University. Jordan.

Anastasia, a; Yurian, S. (201٥). Psychometrics (Salah El-Din Mahmoud Allam, Trans.). Dar Al-Fikr. (١٩٩٧/201٥)

Al –Asmar, Mustafa Moh'd (2007). The Psychometric Characteristics of Two Forms of Mathematics Test, with Dichotomous Items in One Form and Polytomous Items in the Other, According to the Classical and Modern Test Theory. (Ph. D, Dissertation). Amman Arab University. Jordan.

AL-Bursan, Ismael Salameh (2018). The Investigation of the Effectiveness of Five Secondary Strategies in the Item Selection of Multi–Stage Adaptive Testing. The International Interdisciplinary Journal of Education, 7 (10), 40-53

AL-Dalala, Muhammad Mahmoud Suleiman (2005). Comparing the Effectiveness of Three Item Selection Methods for A Criterion-Referenced Math Test for the Fundamental Stage (Ph. D, Dissertation). Amman Arab University. Jordan.

AL-Shareef, Khalid Hassan Bakr (2018). The Effectiveness of Using Rasch Model in Calibration of Test Descriptors for the Learning and Thinking Skills Course Applied to a Sample of Undergraduate Students at King Faisal University Comparative Study with Classical Test Theory. Journal of Humanities and Social Sciences, (50), 89-174

Al-Jamal, Heba Muhammad Ibrahim; Tahoun, Hussein Hassan Hussein; Desouki, Sherine Muhammad Ahmed; & Al-Naghi, Heba Ibrahim Muhammad Ali. (2017). Calibration the Mathematics Question Bank for Third Grade Students Using Three-Parameters Logistic Model. Journal of the College of Education, (21), 837-867.

AL –Jarah, Khaldun Mohammad (2015) The Effect of Sample Size and Item Difficulty and Discrimination on Information Function and Standard Error by Using Item Response Theory (Tow - Parameter Logistic Model) (Master Dissertation). Mu'tah University. Jordan

AL-khateeb, Abdallah Yoines Fo'Ud (2013). The Extent of Compatibility between the Classical Theory of Measurement and Item Response Theory in Item Selection of Ttest in Provisions of Recitation and Tajweed (Master Dissertation). Yarmouk University. Jordan.

AL-Musawi, Noman Muhammad Saleh (2016). Developing A Question Bank in the Educational Evaluation Course for the Undergraduate Level

Using the Classic Theory of Measurement and the Item Response Theory. Journal of the College of Education, 64 (4), 635-676

AL-Najjar, Nabil Juma Saleh (2008). Developing A Question Bank in the Computer Culture of High School in Jordan, Using Item Response Theory, Rush Model and Classical Theory. Journal of the College of Education, 1 (32), 493-526.

Al-Nuaimi, Ezz Al-Din Abdullah (2015). Parameters of Items, Individuals and the Invariance Characteristic of the National Tests for Controlling the Quality of Education in Jordan: A Comparison between the Classical Theory and the Modern Theory of Measurement. Journal of the Union of Arab Universities for Education and Psychology, 13 (1), 136 - 155.

Al-Shafei, Muhammad Mansour Muhammad (2012). Reducing the Item of the Linear Traditional Tests in Light of the Variables of Maximum Information and Number of Items. Journal of the College of Education, 23(90): 102 - 172

Al-Subhi, Muhammad Ali Hamid (2001). Development A Criterion Reference Test to Measure Mathematical Competencies in Engineering Concepts for the Elementary Stage in Makkah Al-Mukarramah Schools (Master Dissertation). Umm Al Qura University. Saudi.

Baker, F. (20١٠). The Basics of Item Response Theory, (Abdul Rahman Al-Tariri; Al-Sayyid Abu Hashem & Sawsan Shalabi, Trans.), King Saud University for Publishing. Saudi. (2001/2010)

Chen, S. Y., & Ankenman, R. D. (2004). Effects of Practical Constraints on Item Selection Rules at the Early Stages of Computerized Adaptive Testing. Journal of Educational Measurement, 41(2), 149-174

Clark, L. A. & Watson, D. (1995). Constructing Validity: Basic Issues in Objective Scale Development. Psychological Assessment, 7, 309-319

Crocker, L.& ALgina, J. (2017). Introduction to Classical and Modern Test Theory (Hind Al-Hammouri & Zeenat Dana, Trans.). Dar Al-Fikr. (2008/2017)

De Ayala, R. J. (2017). The Theory and Practice of Item Response Theory (Abdullah Al-Kilani & Ismail Al-Barasan, Trans.). King Saud University Dar for Publishing.

Ebel, R.L. & Frisbie, D.A. (1991). Essentials of Educational Measurement (5th ed.). Prentice-Hall, Englewood Cliffs.

Eggen, T. & Straetmans, G. (2000). Computerized Adaptive Testing for Classifying Examinees into Three Categories. Educational and Psychological Measurement, 60(5),713–734

Feldet, A. (2017). Discovering Statistics Using IBM SPSS Statistics (5th ed.). SAGE Publications Ltd.

Fotaris, P., & Mastoras, T. (2014). LMS Assessment: Using IRT Analysis to Detect Defective Multiple-Choice Test Items. International Journal of Technology Enhanced Learning, 6(4), 281-296.

Gierl, M. J.; Henderson, D.; Jodoin, M. & Klinger, D. (2001). Minimizing the Influence of Item Parameter Estimation Errors in Test Development: A Comparison of Three Selection Procedures. The Journal of experimental education, 69(3), 261-279

Guyer, R., & Thompson, N.A. (2014). User's Manual for Xcalibre Item Response Theory Calibration Software, version 4.2.2 and later. Assessment Systems Corporation.

Halkitis, P. N. (1998). The Effect of Item Pool Restriction on The Precision of Ability Measurement for A Rasch-Based CAT: Comparisons to Traditional Fixed Length Examinations. Journal of outcome measurement, 2(2), 97-122.

Hambleton, R.K., & Cook, L.L. (1977). Latent Trait Models and Their Use in the Analysis of Educational Test Data. Journal of Educational Measurement, 14(2), 75-94

Hambleton, R. K., & Van der Linden, W. J. (1982). Advances in Item Response Theory and Applications: An Introduction. Applied Psychological Measurement, 6, 373-378

Hambleton, R.K., & Swaminathan, H. (1985). Item Response Theory Principle and Application. Kluwer Nijhoff Publishing.

Hambleton, R. K., Swaminathan, H., & Rogers H. J. (1991). Fundamentals of Item Response Theory. Sage Publications, Inc.

Hayajneh, Ali Theep (2011). The Effect of Item Selection Procedures on Test Information Function (ph. D, Dissertation). Yarmouk University. Jordan.

Herzalla, Alia Mohammed (2004). Perparing an Item Bank Mathmatics and Ascertaining its Efficecy in Item Selection of a Criterion - Referenced Test at the General Secondary Examination Level in Jordan (Ph. D, Dissertation). Amman Arab University. Jordan.

Kalender, I. (2012). Computerized Adaptive Testing Administration for Entrance to Higher Education. The Journal of Higher Education, 2, 13-19

Kelley, T.; Ebel, R. & Linacre, J. (2002). Item Discrimination Indices. Rasch Measurement Transactions, 16, 883-884

Leung, C. K.; Chang, H. H., & Hau, K. T. (2002). Item Selection in Computerized Adaptive Testing: Improving the A-Stratified Design with the Sympson-Hetter Algorithm. Applied Psychological Measurement, 26(4), 376-392.

Livingston, Ronald B. & Reynolds, Cecil R. (2013). Mastering Modern Psychological Testing (Theory &Methods). (Salahuddin Allam, Trans.). Dar Al Fikr. (2012/2013)

Lord, F. M. (1981). Applications of Item Response Theory to Practical Testing Problems. Routledge.

Lunz, M. E. & Bergstrom, B. A. (1994). An Empirical Study of Computerized Adaptive Test Administration Conditions. Journal of Educational Measurement, 31(3), 251-263

Marshoud, Muhammad Faik Salem (2014). Development of A Computerized Question Bank in Statistics for Students of Humanitarian Colleges at An-Najah National University - Palestine, Using the Non-Parametric Locus Model (Ph. D, Dissertation). Yarmouk University. Jordan

Mehrenz, W. & Lehmann, I. (2004). Measurement and Evaluation in Education and Psychology (Haitham Al-Zubaidi, Trans.). Dar Alkitab Aljamieuy. (2003/2004)

Moghadamzadeha, Ali; Salehi, Keyvan & Khodaie, Ebrahim (2011). A Comparison the Information Functions of the Item and Test in One, Two and Three Parametric Model of the Item Response Theory (IRT). Procedia - Social and Behavioral Sciences, 29, 1359–1367

Muhammad, Shehta Abd al-Mawla Abd al-Hafiz (1999). Evaluating the Construction of Reference Tests / Criterion Considering the Item Response Theory (Ph. D, Dissertation). Ain-Shams University. Egypt.

Muhammad, Muhammad Gad Muhammad (2016). A Psychometric Study in the Accuracy of Developing A Question Bank in Mathematics, Using Some Models of the Item Response Theory. World of Education, 17(54), 263-269.

Murphy, D. L.; Dodd, B. G., & Vaughn, B. K. (2010). A Comparison of Item Selection Techniques for Testlets. Applied Psychological Measurement, 34(6), 424-437.

Nitco, A. &Bruchhart, S. (2013). Educational Evaluation for Students (Ali Al-Qarni, Ibrahim Al-Dossary, Rashid Al-Mahrazi & Hussein Al-Kharousi, Trans.) Arab Bureau of Education for the Gulf States (2012/2013)

Nour al-Din, Muhammad Amin Sabry (2002). The Effectiveness of Computerized Adaptive Testing in Ability Estimation and Scales Psychometric Properties Determination (Ph. D, Dissertation). Ain-Shams University. Egypt.

Price, L. R. (2016). Psychometric methods: Theory into Practice. The Guilford Press.

Rababaa, Safaa Muhammad Ahmad (2019). Developing A Question Bank in the Provisions of Reading and Tajweed Using the One-Parameter Model (Master Dissertation). Yarmouk University. Jordan.

Riley, B. B.; Conrad, K. J.; Bezruczko, N. & Dennis, M. L. (2007). Relative Precision, Efficiency and Construct Validity of Different Starting and Stopping Rules for A Computerized Adaptive Test: The GAIN Substance Problem Scale. Journal of Applied Measurement, 8(1), 48-64.

Sakhim, Khaled Ahmed Hassan (2018). A Proposed Scenario for Preparing a Question Bank in Light of the Table of Specifications and

Characteristics of a Good Test: An Application to the Sixth Grade Science Course in the Basic Education Stage In Yemen (Master Dissertation). University of the Holy Quran and Islamic Sciences. Yemen.

Spss, I. (2017). IBM SPSS Statistics for Windows, version 25. Armonk, NY: IBM SPSS Corp.

Van der Linden, W. J., & Hambleton, R. K. (2013). Handbook of modern item response theory. Springer Science & Business Media

Wang, J. (2018). Technical Report: Does It Matter Which IRT Software You Use?.AssessmentSystemsCorporation,https://assess.com/docs/Which_IRT_Software.pdf

Weiss, D. J., & Kingsbury, G. G. (1984). Application of Computerized Adaptive Testing to Educational Problems. Journal of Educational Measurement, 21(4), 361-375.

Weiss, D. J., & Von Minden, S. (2012). A Comparison of Item Parameter Estimates from Xcalibre 4.1 and Bilog-MG. Assessment Systems Corporation, http://www.assess.com/docs/Xcalibre_Bilog_Comparison_Report.pdf

Yao, L. (2019). Item Selection Methods for Computer Adaptive Testing with Passages. Frontiers in psychology, 10, 240.