جامعة المنصورة
كلية التربية

اثر حجم العينة في دقة تقدير قدرة الفرد ودالة معلومات الاختبار

إعداد
الباحث/ محمد عواد سالم القضاة

إشراف
أ.د ممدوح عبد المنعم الكناني
أ.د محمد عبد السميع رزق

مجلة كلية التربية – جامعة المنصورة
العدد ١١٠ – إبريل ٢٠٢٠

اثر حجم العينة في دقة تقدير قدرة الفرد ودالة معلومات الاختبار

محمد عواد سالم القضاة

الملخص

هدفت هذه الدراسة لتحديد اثر حجم العينة على تقدير دالة معلومات الاختبار وقدرة الفرد باستخدام النموذج الثلاثي، وتكونت عينة الدراسة من ١٢٣٠طالب، حيث قام الباحث بتطبيق اختبار رياضيات للصف التاسع الاساسي للفصل الدراسي الثاني٢٠١٨-٢٠١٩ مكون من ٦٠ مفردة اختيار من متعدد، على ثلاثة مستويات للعينة(٣٠٠،٤٠٠،٥٣٠) طالب، واستخدم الباحث برنامجي التحليل الاحصائي SPSS , Bilog-mg لتدريج مفردات الاختبارات، توصلت النتائج الى وجود فروق ذات دلالة إحصائية عند تقدير طات دالة معلومات الاختبار تعزى لاختلاف حجم العينة، وتبين بعد اجراء المقارنات البعدية ان الفروق لصالح العينة(٥٣٠) مقابل العينتين (٣٠٠)، (٤٠٠)، أي ان دقة تقدير دالة معلومات الاختبار تزداد بزيادة حجم العينة. وتوصلت نتائج الدراسة ايضاً الى وجود فروق في المتوسطات الحسابية للخطأ المعياري لتقديرات قدرة الافراد ذات دلالة إحصائية تعزى لتغير حجم العينة، وتبين بعد اجراء المقارنات البعدية ان الفروق لصالح العينة(٥٣٠) مقابل العينات الاخرى، وهي ما تؤكدة نتيجة الفروق بين المتوسطات للخطأ المعياري لتقدير قدرة الافراد حيث كان هناك فروق ذات دلالة احصائية بين متوسطات الخطأ المعياري لتقديرات قدرة الافراد لحجم العينة، ولصالح العينة (٥٣٠) الكبيرة، اي ان قيمة الخطأ المعياري لتقديرات قدرة الافراد تتناسب عكسياً مع حجم العينة، وهذا يعود ربما الى ارتباط حجم العينة المعياري، حيث ان زيادة حجم العينة يقلل من الخطأ المعياري للتقدير، وبالتالي تزيد الدقة في تقدير القدرة.

الكلمات المفتاحية: *دالة معلومات الاختبار، قدرة الفرد، دقة التقدير*

Abstract

This study aimed to determine the effect of the sample size on estimating the function of test information and the ability of the individual using the triple model. The study sample consisted of 1230 students, where the researcher applied a mathematics test for the ninth grade for the second semester 2018-2019 consisting of 60 multiple choice items, on three levels of the sample (300,400,530) students، The researcher used the two statistical analysis programs, SPSS, Bilog-mg, to scale the vocabulary of the tests. The results concluded that there were statistically significant differences when estimating the averages of the test information function due to the difference in the sample size. , (400), that is, the accuracy of the

estimation of the test information function increases with the increase in the sample size. The results of the study also found that there are differences in the arithmetic averages for the standard error for the estimates of the individuals 'ability to be statistically significant due to the change in the sample size. Individuals where there were statistically significant differences between the standard error averages for the estimations of individuals ability for the sample size, and for the benefit of the large sample (530), That is, the value of the standard error of the individual's ability estimates is inversely proportional to the sample size, and this is probably due to the correlation of the sample size with the standard error, as increasing the sample size reduces the standard error of the estimate, and therefore increases the accuracy in the power estimate.
Keywords: test information function, individual ability, accuracy of assessment.

مقدمة

يعتبر الهدف الرئيسي لعملية القياس النفسي والتربوي هو اعطاء تقديرات او درجات للأفراد على السمات النفسية والتربوية المختلفة، ومن اجل انجاح التطبيقات العملية للاختبارات والمقاييس النفسية في قياس تلك السمات، يجب ان تعكس التقديرات او الدرجات الناتجة من استخدامها دقة عالية في وضع الافراد على متصل السمة المستهدفة من عملية القياس. ولأن العلامة الناتجة من تطبيق الاختبار تعد الركيزه الاساسية التي يُعتمد عليها في اتخاذ القرارات المتعلقة بالافراد في المواقف العملية، فقد اهتم البحث السيكومتري بقضية صدق علامة الاختبار، من خلال النظر الى صدق علامة الفرد وليس الاقتصار على الصدق من خلال المجموعة (Wright & Stone, 1999,21)، وذلك من اجل الوصول الى أدق التفسيرات للأداء الحقيقي للمفحوص على الاختبار دون أخطاء تقود الى قرارات مضللة.

ونظرا لشيوع الاختبارات من نوع الاختيار من متعدد خصوصا اختبارات الذكاء والتحصيل، فإن استجابات المفحوصين على مثل هذا النوع من الاختبارات قد لا تعكس قدرة المفحوص الحقيقية بسبب عوامل كثيرة مثل الغش، والتخمين، والتباطؤ (Belov & Armstrong, 2010, 36).

ويميز القياس النفسي والتربوي بين مدخلين رئيسين في تصميم وبناء والمقاييس وتحليل البيانات المستمدة منها، هما المدخل التقليدي المتمثل في النظرية الكلاسيكية للاختبارات والمدخل المعاصر المتمثل في نظرية الاستجابة للمفردة. وظلت النظرية الكلاسيكية للاختبار تشكل الاساس النظري والعملي في العلوم السلوكية لسنوات عديدة، وقد استندت هذه النظرية على نموذج بسيط للقياس الذي ينص على ان درجة الفرد الملاحظة في الاختبار تساوي مجموع الدرجة الحقيقية للفرد في السمة المقاسة ودرجة الخطأ في الاختبار. وعلية نجد ان القياس في النظرية الكلاسيكية

يتأثر بخصائص عينة المفحوصين وبصعوبة المفردات التي يشتمل عليها الاختبار (Stone&Yumoto,2004,41).

وفي المقابل تعد نظرية الاستجابة للمفردة طريقه لنمذجة البيانات من خلال نمذجة العلاقة بين المتغير غير الملاحظ وهو قدرة الفرد، واحتمال استجابة الفرد على مفردة الاختبار اجابة صحيحة وهو المتغير الملاحظ. وتتميز هذه النظرية بإمكانية التنبؤ باداء الفرد وتقدير قدرته الكامنة وراء هذا الاداء بغض النظر عن الاداة المستخدمة، كما تقدم مؤشرات إحصائية ثابتة للاختبار ومفرداته لا تختلف باختلاف عينة الافراد، وكما توفر هذه النظرية قياس اكثر ثباتا مع أخطاء اقل للقياس(Nitko, 2001, 409).

وفي ضوء هذه النظريه يمكن تفسير درجة الفرد على الاختبار في ضوء محكات محددة مسبقاً وليس بإرجاعها الى متوسط العينة التي ينتمي اليها الفرد، وبذلك فان نظرية الاستجابة للمفردة تساعد في التغلب على العديد من المشكلات الناتجة عن استخدام النظرية الكلاسيكية في القياس والتقويم التربوي، وتقدم ادوات للقياس تقترب من الادوات المستخدمة في القياس الفيزيائي. (Embretson & Reise, 2000, 85)

مشكلة الدراسة

نظراً لأعتماد نماذج النظرية الحديثة للقياس في التقدير الدقيق الموضوعي للمتغيرات على مطابقة بيانات الاختبار للنموذج المستخدم، وفي محاولة للتعرف على أثر حجم العينة بوصفه عامل مؤثرة على المطابقة، ومن ثم على دقة تقدير قدرة الافراد ودالة معلومات الاختبار، عمدت هذه الدراسة الى التعرف على اثر أختلاف حجم العينة على دقة تقدير كل من قدرة الفرد ودالة معلومات الاختبار بالاعتماد على بيانات ميدانية بإستخدام النموذج ثلاثي المعلمة وبذلك تتحدد مشكلة الدراسة بالتساؤلات الاتية.

أسئلة الدراسة

١- ما تأثير حجم العينة على دقة تقدير قدرة الفرد بإستخدام النموذج الثلاثي.

٢- ما تأثير حجم العينة على دقة تقدير دالة معلومات الاختبار بإستخدام النموذج الثلاثي.

هداف االدراسة

تهدف هذه الدراسة إلى استقصاء أثر حجم العينة على دقة تقدير كل من قدرة الفرد ودالة معلومات الاختبار بإستخدام النموذج ثلاثي المعلمة.

أهمية الدراسة

يمكن تحديد اهمية الدراسة في الجوانب التالية

١- محاولة الاسهام في القاء الضوء على مدى أثر حجم العينة على تقدير معلم قدرة الفرد وتقدير دالة معلومات الاختبار.

٢- توفير معلومات وارشادات تساعد المهتمين في تطوير الاختبارات بإستخدام حجم العينة المناسب.

٣- محاولة تزويد الباحثين الممارسين المستخدمين لنظرية الاستجابة للمفردة بمستويات لحجم العينة المناسب عند تقدير قدرة الفرد ودالة معلومات الاختبار.

المفاهيم الإجرائية للدراسة

حجم العينة: عدد المفحوصين الذين سيطبق عليهم الاختبار، وفي هذه الدراسة سيتم استخدام ثلاث عينات بأحجام مختلفة.

دالة معلومات الاختبار(test information function) اقتران رياضي يعبر عن مجموع دوال المعلومات لمفردات الاختبار(حابس الزبون، ٢٠١٣، ١٣٣٥).

قدرة الفردAbility: اللوغاريتم الطبيعي لمرجح نجاح الفرد على المفردات التي تعبر نقطة صفر التدريج عن صعوبتها وتقدر باللوجيت(أمينة كاظم، ١٩٨٨، ٥٦)، ومن الناحية الإجرائية هو مستوى القدرة لدى الفرد الذي يجيب عن مفردات الاختبار.

دقة التقدير: مصطلح يشير الى جودة التقدير التي يميزها الاحتمالية الكبيرة في ان التقدير قريب من المعالم الحقيقية، وذلك بالاعتماد على الجذر التربيعي لمتوسطات مربعات الانحرافات للفروق بين المعالم المقدرة والمعالم الحقيقية (زايد بني عطا، ٢٠١٤، ٨).

الإطار النظري

تعد نظرية الاستجابة للمفردة الاختبارية IRT أتجاه معاصر في القياس النفسي يسعى إلى تحقيق الموضوعيه وتوفير خطية واستقلالية القياس، وتعتبر من النظريات السيكومترية المعاصرة التي اتضحت فائدتها في التغلب على كثير من مشكلات القياس التي عجزت النظرية الكلاسيكية عن مواجهتها، فهذه النظرية تحاول نمذجة العلاقة القائمة بين مستوى سمة معينة لدى الفرد Trait Level التي يقيسها اختبار معين، واستجابة لمفردة من مفردات الاختبار. ونظراً لأن مستويات السمة غير منظوره بطبيعتها وتُغير في المتغيرات المنظورة، فإنه يتم تقديرها من هذه المتغيرات، أي من الاستجابات الملاحظة. وعادة يُفترض ان السمة التي تنطوي عليها الاستجابات بمثابة قدرةAbility من نوع معين. وعلى العكس من نظرية القياس الكلاسيكية، فإن

نظرية الاستجابة للمفردة تستند إلى افتراضات قوية فيما يتعلق بما يحدث عند استجابة الفرد لمفردات اختباريه، (صلاح الدين علام، ٢٠٠٥، ٥٣).

افتراضات نماذج الاستجابة للمفردة الاختبارية أحادية البعد

تتميز نظرية الاستجابة للمفردة بعدة افتراضات ينبغي ان تتوفر في البيانات المستمدة من حيث يتم اختيار النموذج الملائم للبيانات وفقاً لمدى تحقق هذه الافتراضات في البيانات، حيث ان ذلك يؤدي إلى دقة وإحكام القياس، وقد ثبت انه لا يمكن الاستفادة من مزايا نظرية الاستجابة للمفردة إذا لم يتم التحقق من افتراضاتها في البيانات المراد تحليلها، وكذلك التأكد من اختيار النموذج المناسب لتحليل البيانات، وفيما يلي توضيح لهذه الافتراضات

١-احادية البعدUni-dimensionality

يعرف مفهوم احادية البعد على انه وجود عامل واحد يكمن وراء الأداء في المقياس، وهذا العامل هو القدرة او السمة المقيسة، وأحادية البعد تتحقق في مقياس ما إذا كان التباين بين المفحوصين في الأداء على مفردات المقياس يعود إلى عامل واحد فقط هو القدرة او السمة المقيسة. وقد أشار السيد محمد ابو هاشم (٢٠٠٦، ١٣)، وماجد الخياط (٢٠١٢، ٩٣) إلى أن احادية البعد تعني تجانس مفردات المقياس فيما بينها وقياسها نفس السمة، وان اية مفردة للاستجابة لها نفس الإجراءات والعمليات السلوكية، ولكنها تختلف فيما بينها من حيث تدرج صعوبتها فقط. ولكن العديد من الاختبارات عادةً ما تحتوي على عدة مجالات، وهنالك عاملاً مشتركاً أساسياً بين هذه المجالات وهو قياسها للسمة المراد قياسها. فمن الناحية النظرية فإن مسائل الحساب تتكون من أربع مكونات او أبعاد هي الجمع والطرح والضرب والقسمة. ولكن من الناحية العملية نفترض ان هذه الأبعاد الأربعة تقيس فعلياً موضوعاً واحداً وهو القدرة الحسابية، ويصعب إيجاد اختبار حساب لا يجمع بينهم (راشد الدوسري، ٢٠٠٤، ٤٥).

٢-الاستقلال الموضعي Local Independence

ان افتراض الاستقلال الموضعي يعد افتراضاً رئيساً وحجر الزاوية في كل نماذج نظرية الاستجابة للمفردة، ويؤدي انتهاكه إلى أخطاء سيكومترية في تقدير قدرات الأفراد وبارامترات المفردات (الصعوبة والتمييز والتخمين)، قد يتبعها عدة قرارات خاطئة (Reeve, 2004, 12). ويقصد بالاستقلال الموضعي أن احتمال الإجابة الصحيحة للفرد على مفردة اختباريه يكون مستقلاً عن ناتج إجابته عن أي مفردة في الاختبار، أي عدم تأثر الإجابة عن أي مفردة من مفردات الاختبار بالإجابة عن أي مفردة أخرى تأثراً سالباً أو موجباً.

٣- منحنى خصائص المفردة Item Characteristic Curve

وهو عباره عن علاقة تربط بين احتمال النجاح في الاجابة على المفردة والقدرة التي يقيسها الاختبار الذي يحتوي تلك المفردة، وان هذه العلاقة تتخذ شكل منحنى الترجيح اللوغاريتمي الاحتمالي Logistic Curve ويسمى منحنى خصائص المفردة، حيث يُمثل قدرة المفحوص على المحور السيني واحتمال حصول المفحوص على الإجابة الصحيحة للمفردة يُمثل على المحور الصادي. او هو معادلة انحدار غير خطية لاحتمال اجابة المفردة إجابة صحيحة على القدرة أو السمة التي يقيسها الاختبار. وتعتمد احتمالية الاجابة الصحيحة لمفحوص معين على مفردة ما على شكل المنحني، أي تعتمد على الصعوبة فقط، وهذا يدلل على استقلاليتها عن توزيع القدرة لمجتمع المفحوصين المطبق عليهم الاختبار، أي أن احتمالية الإجابة الصحيحة لا تعتمد على عدد المفحوصين ذوي مستوى القدرة نفسها. ولا على الافراد عند مستويات القدرة الأخرى(Lord, 1980,35) .

٤-افتراض عامل السرعة في الاجابة assumption of speedness

نظراً لكون نظرية الاستجابة للمفردة تركز في دقة الإجابة وتهمل عامل السرعة فان معظم نماذج النظرية تفترض أن عامل السرعة لا يلعب دوراً في الإجابة على المفردة لذلك فعندما لا تتم الإجابة عن المفردة بسبب عامل السرعة وضيق وقت الاختبار فان الأداء على الاختبار لا يكون دالة للقدرة، لأنه عندما تؤثر السرعة في الأداء تكون هناك قدرتان تؤثران في الأداء على الاختبار. وينص هذا الافتراض على ان إجابة المفحوص على المفردة سواء كانت صحيحة ام خاطئة يجب ان تعتمد على مقدار القدرة التي يقيسها الاختبار، وليس على سرعته في الإجابة(Hambelton & Swamina, 1985, 30).

نماذج الاستجابة للمفردة

١- النماذج احادية البعد:Unidimensional Models

ويشير مصطلح احادية البعد الى ان هنالك سمة واحدة متصلة تكمن وراء استجابات الافراد على مفردات المقياس، وان الفروق بين الافراد في هذه السمة هي التي يرجع لها كل التباين بين استجابات الافراد.

٢- النماذج متعددة الابعاد: Multidimensional Models

تفترض هذه النماذج ان هنالك أكثر من بعد يكمن وراء استجابات الافراد على مفردات المقياس، وتعد هذه النماذج امتداد الى لنماذج الاستجابة للمفردة احادية البعد، لتتناسب مع المقاييس التي تتضمن مفردات تقيس تركيبة من السمات او عدد من الابعاد.

أنواع نماذج الاستجابة للمفردة احادية البعد:Unidimensional IRT Models

تنقسم النماذج احادية البعد في ضوء عدد استجابات الافراد على مفردات المقياس الى نماذج ثنائية واخرى متعددة، فاذا كانت الاستجابات ثنائية الدرجة بمعنى ان اجابة الفرد على المفردة قد تكون صحيحة ١ وقد تكون خطأ٠

نماذج الاستجابة للمفردة احادية البعد ثنائية الدرجة وهي:

١- نموذج راش:

النموذج اللوغاريتمي احادي المعلم One-Parameter Logistic Model. وهو من اهم نماذج النظرية الحديثة الذي يتميز بسهولة مقارنته بالنماذج الاخرى من حيث عدد الفروض اللازم توافرها من البيانات لكي يعطي النموذج تقديرات دقيقة، وبسهولة تقدير معلم الصعوبة والقدرة احصائياً، و يشير فيشر Fisher الى ان الدرجة الخام تعبر عن المعلومات الوحيدة المطلوبة في البيانات لتقدير القدرة، أي ان كل الاشخاص الذين يشتركون في درجة ما ستقدر لهم القدرة نفسها (Fisher , 1973 , 100) ويتميز هذا النموذج بأن معلم القدرة يمكن ان يقدر من غير تحيز لأي مجموعة من المفردات المختارة من النطاق الشامل للمفردات المطابقة للنموذج ، وتعتمد دقة القياس على الاختبار الملائم لصعوبة المفردة، وكذلك إمكانية فحص صدق النموذج بشكل مستقل عن قدرة الأفراد وصعوبة المفردات (Bock & Wood, 1971, 203).

ويفترض النموذج تساوي معاملات التمييزai، وانعدام التخمين ci، ويقوم بتقدير صعوبة المفردات bi فقط وتتمثل الصيغة الرياضية لهذا النموذج فيما يلي:

$$p\ i\ (\theta\ ) = \frac{e^{\ D\ (\theta\ -\ b\ i\ )}}{1\ +\ e^{\ D\ (\theta\ -\ b\ i\ )}} \quad \ldots\ldots\ldots\ldots\ldots\ldots (١)$$

bi: معامل الصعوبة للمفردة i.

e: ألأساس اللوغاريتمي الطبيعي = (٢,٧١٨) تقريبا.

Pi (θ): دالة استجابة المفردة و تمثل احتمال إجابة المفحوص ذي القدرة θ على المفردة i اجابة صحيحة.

D: معامل القياس او التدرجscaling factor وهو قيمة ثابته = ١,٧وهذا الثابت يحول المنحنى اللوغاريتمي الى تطابق تقريبي مع المنحنى الطبيعي.

٢-**النموذج اللوغاريتمي ثنائي المعلم** Two-Parameter Logistic Model

يعد نموذج راش حاله خاصة من النموذج اللوغريتمي الثنائي البـــاراميتر والـــذي يـــسمى بنموذج بيرنبوم Birnbaum Model حيث يفترض نموذج راش ان جميع مفردات المقياس لها نفس القدرة التمييزية بين الافراد ذوي المستويات المختلفة، ولكن النمـــوذج الثنـــائي البـــاراميتر يستخدم في حالة لم تكن مفردات الاختبار لها نفس معاملات التمييز، وفيه يتم اضافة معلم جديـــد الى نموذج راش وهو معلم التمييز a وهو عبارة عن ميل منحنى خاصية المفـــردة عنـــد نقطـــة انقلاب المنحنى، وهي النقطة التي يكون فيها احتمال اجابة الفرد عن المفـــردة اجابـــة صـــحيحة يساوي ٠,٥. ويؤكد (Schrodt, 2007, 5) انه كلما زادت قيمة a زاد ميل المنحنى عند نقطـــة الانقلاب وبالتالي زاد تمييز المفردة. وتتمثل الصيغة الرياضية لهذا النموذج فيما يلي:

$$pi\ (\theta) = \frac{e^{Dai\ (\theta - bi)}}{1 + e^{Dai\ (\theta - bi)}} \quad \text{.....(٢)}$$

حيث :ai معلم تمييز المفردة i.

ويعتبر النموذج الثنائي هو الأقرب الى الواقعية والمنطقية من النموذج الأحادي، اذ انه من الصعب إيجاد مفردات في مقياس لسمة ما لها نفس القوة التمييزية بين المستويات المختلفـــة مـــن القدرة (علام، ٢٠٠٥، ٧٢).

٣- **النموذج اللوجستي ثلاثي المعلم** Three Parameter Logistic Model

يعد هذا النموذج هو الحالة العامة للنموذجين الاحادي والثنـــائي البـــاراميتر، اذ ان هـــذا النموذج يصلح لتدريج المفردات ذات معاملات الصعوبة والتمييز المختلفة، والتي يكون احتمـــال اجابة الافراد ذوي المستويات المنخفضة جدا عنها عن طريق التخمين لا يساوي صفر، وبـــذلك يمكن التغلب على مشكلة التخمين في الاختبارات الموضوعية كالصواب والخطأ والاختيار مـــن متعدد. ويسمى بنموذج لورد Lord Model حيث تم إضافة معلماً ثالثاً أطلق علية مَعلـــم الخـــط التقاربي الأدنى او مَعلم التخمين. ويفترض هذا النموذج أن المفردات تتباين من حيـــث درجـــة الصعوبة والتمييز بين المفحوصين، كما أن مستويات التخمين لدى المفحوصين تختلف من مفردة لأخرى، وتأخذ معادلة هذا النموذج الصيغة الرياضية الموضحة فيمـــا يلـــي (Magno, 2009, 11).

$$pi(\theta) = Ci + (1 - Ci)\frac{e^{Dai(\theta - bi)}}{1 + e^{Dai(\theta - bi)}} \quad \text{..... (٣)}$$

حيث تشير Ci الى مَعلم التخمين او ما يطلق علية مستوى شبة الصدفة او الخط التقـاربي الأدنى للمنحنى Lower Asymptote (Yen, 1992, 66). وبالتالي فان احتمال اجابة الأفـراد من ذوي المستويات المنخفضة جدا ليس صفر كما في النموذجين، وإنما هذا الاحتمال فـي هـذه الحالة يساويCi، وبالتالي فان صعوبة المفردة في هذه الحالة هي نقطة على متـصل الـسمة او القدرة يكون عندها احتمال الاجابة عـن هـذه المفـردة يـساوي [(1+ C) /2] (Cuervo & Andrade, 2004, 29). بمعنى انه إذا كان احتمال إجابة الأفراد من ذوي المستويات المنخفضة من القدرة عن مفردة ما يساوي 0.2 فان هذا يعني ان صعوبة المفردة هي نقطة علـى متـصل القدرة يكون احتمال اجابة الفرد عنها مساويا [(1+0.2/2) =0.6].

دالة معلومات المفردة: Item Information Function

لكل مفردة من مفردات الاختبار هناك ما يسمى دالة معلومات المفردةIIF، وهو عبارة عن دالة بين مدى مساهمة المفردة في تحديد القدرة. وبشكل عام، فان المفردات ذات التمييز العـالي تسهم بقوة أكبر في تأكيد دقة القياس من تلك ذات التمييز المتدني، حيث تعطي المفـردة أفـضل مساهمة لها في دقة القياس حول قيمة صعوبتهاb على متصل القدرة، فاذا ما كان المنحنى مزاحا باتجاه اليمين، فهذا يعني ان المفردة صعبة، وإذا كان ارتفاعه كبير فهذا يعني ان تمييز المفـردة عالٍ (حسين عبد النبي، ٢٠١٣، ٢٧).

وبعد اختيار نموذج الاستجابة للمفردة المناسب وتقدير بارامتراته، فانه ينبغي التحقق مـن مدى دقة تقدير قدرة الأفراد، فعندما يتم تقدير الأرجحية القصوى للقدرة، فان دقة تقـدير القـدرة يمكـن التعبيـر عنهـا بدالـة المعلومـات.information function او منحنـى المعلومـات information Curve الذي يمثلها بيانياً. ودالة المعلومات من المفاهيم الاساسـية فـي نظريـة الاستجابة للمفردة. وبالرغم من ان معظم تطبيقات دوال المعلومـات اقتـصرت علـى نمـاذج الاستجابة للمفردة ثنائية الدرجة، الا ان الباحثين بدئوا بدراسة خصائص هذه الـدوال فـي حالـة استخدام نماذج الاستجابة متعددة الاقسام، ونماذج الاستجابة متعددة الابعاد (صلاح علام، ٢٠٠٢، ١١٢).

تعد دالة المعلوماتية المفـردة item information function(IIF) مـن الإحـصاءات المهمة في النظرية الحديثة في القياس، حيث يمكن من خلالها تحديد الخطأ المعياري في التقـدير،

حيث إنه يمكِن من إيجَادها تحديد الخطأ المعياري في التقدير، وذلك بالاعتماد علـــى الأرجحيـــة العظمى لمَعلم القدرة، حيث إن تباين الخطأ في تقدير القدرة يساوي معكوس دالـــة المعلومـــات، وكما أن مصفوفة التباين المشترك covariance matrix variance للتقديرات تساوي معكـــوس مـــصفوفة المعلومـــات information matrix لتقـــديرات معـــالم المفـــردة (Hambleton & Swaminathan, 1991, 56).

ومن الجدير بالذكر أن المنحنى المميز لمفردة اختبارية معينـــة (ICC) لأي مـــن نمـــاذج الاستجابة للمفردة ثنائية الدرجة، أو متعددة الأقسام يمكن إجراء تحويل عليـــه ليـــصبح منحنـــى معلومات المفردة Information Curve (IIC) Item. وهذا المنحنى يبين كميـــة المعلومـــات السيكومترية التي تشتمل عليها مفردة اختبارية عند جميع النقط على متصل السمة الكامنة.

والصيغة الرياضية العامة لمنحنى معلومات المفردة (ICC) كالتالي:

$$I(\theta) = \frac{\left[P'i(\theta)\right]^2}{\left[\left[Pi(\theta)\right]\right]\left[1 - \left[Pi(\theta)\right]\right]} \quad \text{.....(٤)}$$

حيث يشير الرمز $P'i(\theta)$ إلى المشتقة الأولى للدالة المميزة للمفردة عند مستوى قدرة أو سمة معين (θ)، وقيمة هذه المشتقة عند مستوى معين من مستويات القدرة (θ) يمثلها ميل المماس للمنحنى المميز للمفردة عند النقطة (θ)P.

والرمز $Pi(\theta)$ إلى الاحتمال المشروط لإجابة المفردة (i) إجابة صحيحة المستمدة من المنحنى المميز للمفردة (IIC). ولتيسير حساب مقدار المعلومات $I(\theta)$ عند مستوى قدرة معين لأي من النماذج ثنائية الدرجة يمكن إعادة كتابة الصيغة في المعادلة (٤) كالتالي:

$$I(\theta) = \left[ai\ \frac{1 - Pi(\theta)}{Pi(\theta)}\right]\left[\frac{Pi(\theta) - ci^2}{(1 - ci)^2}\right] \quad \text{.....(٥)}$$

حيث يشير الرمز(ai) إلى بارامتر تمييز المفردة (i)

والرمز(ci) إلى بارامتر التخمين (بارامتر شبه الصدفة) للمفردة (i) ففي حالـــة النمـــوذج ثلاثي البارامتر يمكن حساب مقدار المعلومات المشروطة للمفردة. وفي حالـــة النمـــوذج ثنـــائي البارامتر نضع (ci) تساوي صفر في الصيغة في المعادلة (٥) وبذلك تصبح كالتالي:

$$I(\theta) = \mathrm{a}i^2\, Pi(\theta)\left[1 - Pi(\theta)\right] \quad \text{.....(٦)}$$

وفي حالة النموذج أحادي البارامتر (نموذج راش) نضع (ai) تساوي الواحد الصحيح، وبذلك تصبح الصيغة كالتالي:

$$I(\theta)=Pi(\theta)\left[1-Pi(\theta)\right]=Pi(\theta)\cdot Qi(\theta) \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (٧)$$

دالة معلومات الاختبار Test Information Function

تلعب دالة معلومات الاختبار دوراً مهماً في النظرية الحديثة في القياس، إذ يمكن من خلالها تحديد الخطأ المعياري في التقدير، وتتمتع دالة معلومات الاختبار، التي تمثل مجموع دوال معلومات الاختبار مستقلة عن عينة المفحوصين، وبذلك تقدم النظرية الحديثة في القياس مميزات إضافية، فيما يتعلق بزيادة القدرة على تقدير أخطاء القياس. كما أن الخطأ المعياري في التقدير يرتبط عكسياً مع الجذر التربيعي لدالة معلومات الاختبار وفق العلاقة الآتية:

$$S E \;=\; \frac{1}{\sqrt{I\ (\theta\ )}} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (٨)$$

وهذا يعني أن الخطأ المعياري في التقدير، يكون أقل ما يمكن عند مستويات القدرة التي تناظر أقصى معلوماتية. أي ان دقة القياس ترتبط بحجم المعلومات التي تعطيها المفردة، التي تسهم في الاختبار او المقياس ككل. وتتغير دالة معلومات الاختبار بتغير القدرة، وهذا يؤدي الى التغير في الخطأ المعياري في التقدير SEE، والقيمة العظمى لدالة معلومات الاختبار، سوف تقابل اقل قيمة للخطأ المعياري في التقدير (Swaminathan & Hambleton, 1985, 142).

ونظراً لان دالة المعلومات تتباين بتباين مستوى القدرة، فان خبراء القياس اقترحو ان منحنيات معلومات الاختبارات يمكن ان تكون بديلاً لتقديرات الثبات المستخدمة في نماذج القياس الكلاسيكية، وللأخطاء المعيارية للقياس المستخدمة في تفسير درجات الاختبارات. ومن الاستخدامات المفيدة لدالة معلومات المفردة، إمكانية معرفة مدى مساهمة كل مفردة في دالة معلومات الاختبار بشكل مستقل عن المفردات الأخرى للاختبار، فإذا كان لدينا فكرة جيدة عن قدرات مجموعة المفحوصين، فيمكن انتقاء مفردات الاختبار التي تعظم المعلومات التي يقدمها الاختبار في المدى الذي تتوزع فيه قدرات المفحوصين في الاختبار (راجي الصرايرة، وحابس الزبون.٢٠١٧، ١٩٣).

وتستخدم دالة معلومات المفردة ايضاً لاختيار المفردات عند بناء الاختبارات، بالاعتماد على النظرية الحديثة في القياس، على افتراض ان معلومات المفردة تتغير عبر مستويات السمة

المختلفة، وعلية فإنه من الممكن اختيار مفردات تقدم دقة قياس مرتفعة عند نقطة معينة على متصل السمة، كما أن المفردات التي تتضمن معالم تمييز كبيرة، تقدم معلومات أكبر عن قدرة المفحوصين، وبالتالي الحصول على دقة أكبر، وتعطى دالة المعلومات للاختبار من خلال احدى المعادلات الآتية:

$$I(\theta) = \sum_{i=1}^{n} I_i(\theta) \quad \text{.......... (٩)}$$

أسلوب الأرجحية القصوى الهامشية: Marginal Maximum Likelihood في تقدير المعالم.

تعد هذه الطريقة من أكثر الطرق انتشاراً في تقدير معالم المفردة، إذا يتيح بنائها الرياضي تطبيق جميع أنواع نماذج استجابة المفردة، وأيضاً تعد من الطرق الفعالة في حالة الاختبارات الطويلة والقصيرة، بشرط تحقق الافتراضات الأساسية الأربعة التي تقوم عليها نظرية استجابة المفردة. ويتعامل هذا الأسلوب مع مستويات القدرة غير المعلومة، وذلك بالتعبير عن احتمالات أنماط الإجابات بتوقعات Expectations من توزيع مجتمع معين. فالبيانات الاختبارية ينظر إليها على أنها عينة عشوائية مستمدة من مجتمع، غير انه لم تكن هناك إجراءات عملية لتقدير معالم المفردات والأفراد إلى ان تم التوصل الى برنامج حاسوب يجري هذا التقدير في مرحلتين، الأولى تسمى مرحلة التوقع Expectation والثانية تسمى مرحلة التعظيم Maximaization (صلاح علام، ٢٠٠٥: ٩٧).

وفي مرحلة التوقع يتم فيها حساب العدد المتوقع من الأفراد في كل مستوى من مستويات القدرة، وكذلك العدد المتوقع من الأفراد الذين يجيبون إجابة صحيحة على كل مفردة، واستخدام هذه التوقعات في معادلات التقدير Estimation Equations وقد وجد أن تقديرات معالم المفردات يجعل الارجحية أكبر ما يمكن في مرحلة التعظيم Maximaization. وفي مرحلة التوقع الثانية تستخدم معالم المفردات في حساب الارجحية المعدلة للبيانات استناداً إلى التوقعات المعدلة. وعندما يحدث تقارب Convergance لهذه القيم، يتم إيجاد القيم التقديرية النهائية للمعالم وأخطائها المعيارية. ويتميز هذا الأسلوب بميزات متعددة، لعل أهمها أنه يمكن استخدامه في تقدير معالم جميع النماذج أحادية البعد، وكذلك النماذج متعددة الأبعاد. كما أن هذا الأسلوب يتميز بالفعالية Efficiency، سواء كان الاختبار عدد مفرداته قليلاً أو كثيراً. والقيم التقديرية للأخطاء المعيارية الناتجة تتميز بالدقة، ويمكن الحصول على قيم تقديرية لمعالم الأفراد الذين أجابو إجابة صحيحة أو خطأ على جميع المفردات (راجي الصرايرة، ٢٠١٦: ٦ – ٨).

دراسات سابقة

من الدراسات التي تناولت التحقق من دقة التقدير بإستخدام النمـــوذج الثلاثـــي البـــاراميتر دراسة كلير (Keller, 2002) التي هدفت الى تقويم الدقة الاحصائية في تقدير معـــالم المفـــردة والقدرة بإستخدام اسلوب المحاكاة، وقد تم توليد البيانات بإستخدام النموذج الثلاثي عـــن طريـــق ثلاث اختبارات افتراضية تتكون من ١٥، ٢٥، ٤٠ مفردة وعينات مختلفة من حجم العينة ١٠٠، ٢٠٠، ٥٠٠، ١٠٠٠، وقد استخدمت برمجية BILOG-MG لتقدير قدرات المفحوصين ومعـــالم مفردات الاختبار ودالة معلومات الاختبار في التقدير، وأشارت النتائج الى تأثر دقة تقـــدير معلـــم الصعوبة والتمييز باختلاف حجم العينة حيث زادت دقة التقدير بزيادة حجم العينة، في حـــين لـــم تتاثر دقة تقدير التخمين بحجم العينة، كما زادت دقة تقدير دالة معلومات الاختبار بزيـــادة حجـــم العينة.

كما أجرى ايضاُ دي لا توري ويوان (De La Torre & Yuan, 2010) دراسة هـــدفت الى التعرف على اثر حجم العينة على دقة تقدير معلمة القدرة ومعالم المفـــردة فـــي اختبـــارات مطورة حسب نماذج نظرية الاستجابة للمفردة (IRT). واستخدمت الدراسة أسلوب المحاكاة فـــي توليد مجموعة من البيانات ضمن ظروف اختبار مختلفة وتم استخدام طريقة مونتي كارلو، مـــن اجل التعرف الى العلاقات بين معلمة القدرة ومعالم المفردة في الاختبار، وأثر حجم العينة في دقة تقدير معلمة القدرة ومعالم المفردة، وأشارت النتائج الى ان حجم العينة يؤثر على قدرة الاختبـــار في تقدير معلمة القدرة ومعالم المفردة، والى وجود علاقة ارتباطية بين تقدير معلمة القدرة ومعالم المفردة.

وفي دراسة اجراها زايد بني عطا (٢٠١٤) هدفت الدراسة إلى تقصي دقة تقدير النمـــوذج اللوجستي ثلاثي المعلمة لمعالم المفردة وقدرة الافراد، في ضوء تغير طول الاختبار وحجم العينة. ولتحقيق الهدف من الدراسة وُلِّدت بيانات ثنائية الاستجابة بواقع ٥٠ مرة لستة مستويات من طـــول الاختبار١٠، ٢٥، ٥٠، ٧٥، ١٠٠، مفردة وست مـــستويات لحجـــم العينـــة ١٠٠، ٢٥٠، ٥٠٠، ١٠٠٠، من خلال استخدام برنامج WINGEN وبإستخدام برنامج Bilog- Mg حُلِّلَت البيانـــات المولدة لكل خلية من تقاطع مستوى طول الاختبار ومستوى حجم العينة. كشفت نتائج الدراسة عن وجود أثر ذي دلالة إحصائية لكل من طول الاختبار وحجم العينة والتفاعل بينهما في دقة تقديرات معالم المفردات وقدرة الأفراد. وكشفت النتائج أيضا أن الوسط الحـــسابي لقـــيم RMAS لمَعْـــالَم المفردات وقدرة الأفراد أخذ بالتناقص عندما زاد طول الاختبار على ٥٠ مفردة وزاد حجم العينـــة على ٢٠٠٠ وقد أيدت قيم معاملات الارتباط بين المعَالِم المقدرة والحقيقة هذه النتيجة، حيث وصلت إلى الارتباط شبه التام، بالإضافة إلى ذلك كانت قيم التحيز في التقديرات قريبة من الصفر.

وفي دراسة حابس الزبون (٢٠١٣) الهادفة لمعرفة أثر حجم العينة على تقدير دالة المعلومات والخطأ المعياري في تقديرها. ولأغراض الدراسة تم استخدام استجابات ٧٥٠٠ طالب وطالبة من طلبة الصف الثامن الأساسي موزعين عشوائي على خمس مجموعات ٥٠٠،١٠٠٠،١٥٠٠،٢٠٠٠،٢٥٠٠ وأجابوا على الاختبار الوطني لضبط نوعية التعلم لمادة الرياضيات المكون من ٤٠ مفردة من نوع الاختيار من متعدد، وتم استخدام برنامج Bilog-Mg وبرنامج SPSS لايجاد معالم المفردات حسب النموذج الثلاثي، وتقدير دالة المعلومات والخطأ المعياري في تقدير دالة المعلومات، وأظهرت النتائج ان مقدار تقديرات دالة المعلومات يتغير بتغير حجم العينة، إذ يزداد بزيادة حجم العينة، كما أظهرت النتائج ان الخطأ المعياري في تقدير دالة المعلومات يتغير بتغير حجم العينة، اذ يتناقص بزيادة حجم العينة.

وفي دراسة يمان شما (٢٠١٣) الهدفة إلى التعرف على أثر حجم العينة على دقة تقدير صعوبة المفردات وقدرة الافراد بإستخدام نموذج راش ولتحقيق هذه الاهداف تم تطبيق اختبار أوتس- لينون للقدرة العقلية على ثلاث عينات مكونة من (٢٠٠، ٨٠٠، ١٥٠٠) طالب وطالبة من المرحلة الاساسية واستخدام برنامج SPSS وبرنامج Bilog-mg3 لتقدير صعوبة مفردات اداة الدراسة وقدرة أفراد عينة الدراسة وفقاً لكل حجم من حجوم العينة وتوصلت نتائج الدراسة إلى أن دقة تقدير صعوبة المفردات يزداد بزيادة حجم العينة وان اختلاف حجم العينة لا يؤثر على دقة تقدير قدرة الافراد. حيث ان قيم متوسطات الخطأ المعياري لتقديرات قدرة الافراد في كل حجم من حجوم العينة المستخدمة متساوية .

كما أجرى نضال الشريفين (٢٠١٢) دراسة كان الهدف منها كشف أثر طريقة تقدير معالم المفردات وقدرت الافراد على قيم معالم المفردة والخصائص السيكومترية للاختبار، في ضوء تغير حجم العينة. ولتحقيق الهدف من الدراسة تم بناء اختبار تحصيلي في الفيزياء مكون من ٣٣ مفردة من نوع الاختيار من متعدد، وطبق الاختبار على عينة مكونة من ١٠٠٠ طالب وطالبة من طلبة الصف الثاني ثانوي علمي، وحللت النتائج وفق النموذج الثلاثي المعلمة بإستخدام البرمجية Bilog mg3. وبينت النتائج وجود فروق ذات دلالة إحصائية بين متوسطات الأخطاء المعيارية لتقديرات معالم المفردات تعزى للتفاعل بين طريقة التقدير وحجم العينة، ولم تظهر فروق ذات دلالة إحصائية تعزى لمتغير حجم العينة وطريقة التقدير. كما أشارت النتائج إلى وجود فروق ذات دلالة إحصائية بين متوسطات الأخطاء المعيارية لتقديرات القدرة للأفراد تعزى لحجم العينة والتفاعل بين طريقة التقدير وحجم العينة، ولم تظهر النتائج فروق ذات دلالة احصائية تعزى لطريقة التقدير، كما بينت النتائج عدم وجود فروق دالة بين معاملات الثبات

المقدرة وفق نظرية الاستجابة للمفردة عند أحجام العينة المختلفة(١٠٠، ٥٠٠، ١٠٠٠) فرد، وأشارت النتائج الى ان دقة تقديرات معلم القدرة تزداد قي حالة عينة الافراد ذوي القدرة المتدنية عند استخدام طريقة بيز، في حين تزداد الدقة عند مستويات الافراد ذوي القدرة المتوسطة بإستخدام طريقة الارجحية العظمى بغض النظر عن حجم العينة.

فروض الدراسة

١- توجد فروق ذات دلالة إحصائية عند مستوى الدلالة (α= ٠,٠٥) بين متوسطات قيم دالة معلومات المفردة في المستويات المختلفة من حجم العينة بإستخدام النموذج الثلاثي.

٢- توجد فروق ذات دلالة إحصائية عند مستوى الدلالة (α= ٠,٠٥) بين متوسطات الخطأ المعياري لتقدير القدرة في المستويات المختلفة من حجم العينة بإستخدام النموذج الثلاثي

الإجراءات الميدانية

١- مجتمع وعينة الدراسة:

تكون مجتمع الدراسة من جميع طلبة الصف التاسع الأساسي في المدارس الحكومية في مديريات التربية والتعليم (مديرية تربية قصبة الكرك، مديرية تربية لواء المزار الجنوبي، مديرية تربية لواء القصر) في محافظة الكرك للعام الدراسي ٢٠١٨/ ٢٠١٩م. وتكونت عينة الدراسة من ١٢٣٠ طالبا من طلبة الصف التاسع الأساسي في تلك المديريات.

٢- **أداة الدراسة** (اختبار الرياضيات للصف التاسع الأساسي للفصل الدراسي الثاني)

أ- **تحديد الغرض من الاختبار:** إن الغرض من بناء الاختبار هو تجهيز مفردات قادرة على قياس تحصيل الطلبة في مبحث الرياضيات للصف التاسع الأساسي للفصل الدراسي الثاني ٢٠١٨/ ٢٠١٩ م، لتحديد اثر حجم العينة على تقدير قدرة الفرد ودالة المعلومات للاختبار، وتتمتع هذه المفردات بخصائص سيكومترية جيدة وتكون مطابقة لافتراضات نظرية استجابة المفردة (النموذج الثلاثي). ومن أجل تحقيق هدف الدراسة تم بناء اختبار تحصيلي في الرياضيات مكون من ٧٠ مفردة من نوع الاختيار من متعدد، لكل مفردة أربع بدائل، وبعد تطبيق الاختبار على العينة الاستطلاعية (عينة الصدق والثبات) البالغ عددها ١٠٠طالب، وإيجاد الخصائص السيكو مترية للاختبار، من خلال استخدام برنامج الحزم الإحصائية، SPSS وبرنامج Excel وحذف المفردات ذات قيم معاملات صعوبة وتمييز غير مناسبين، كانت الصورة النهائية للاختبار مكونة من ٦٠ مفردة.

ب- **تحليل المحتوى:** تم تحليل محتوى مقرر الرياضيات الفصل الثاني للصف التاسع الأساسي، ووضع الأهداف المعرفية للوحدات الدراسية التي يتكون منها الكتاب المدرسي

الذي يمثل منهاج الرياضيات للصف التاسع الأساسي، وجرى تحليل محتوى المنهاج، والذي يتكون من أربع وحدات، تُدَرس للطلبة في الفصل الثاني من العام الدراسي.

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

د- **بناء مفردات الاختبار:** بالاعتماد على تحليل المحتوى، ولائحة المواصفات، وخبرة الباحث العملية في مجال الرياضيات، تم صياغة مفردات الاختبار في صورة الاختيار من متعدد، وقد حرص الباحث على مراعاة شروط صياغة مفردات جيدة ذات بدائل ملائمة، كما راعيت تحديد المستوى المعرفي ومدى تمثيل المفردات للأهداف المقابلة لها مع مراعاة الصحة العلمية لهذه المفردات، وقد بلغ عدد مفردات الاختبار ٧٠ مفردة مرتبطة بمحتوى الفصل الدراسي الثاني لمبحث مقرر الرياضيات، ثم عرض الاختبار على اثنين من المحكمين أعضاء هيئة تدريس في قسم المناهج في جامعة المنصورة، واثنين من التربويين الأردنيين ممن يحملون درجة الدكتوراة والماجستير في الأساليب وطرق التدريس، لإبداء آرائهم وملاحظاتهم وإضافة او حذف او تعديل ما يرونه مناسباً على عناصر أداة الدراسة ، وتم الاخذ بملاحظاتهم بعين الاهتمام.

ه- **تطبيق الاختبار على عينة استطلاعية:** تم تطبيق الاختبار على عينة مكونة من ١٠٠ طالب ودراسة الخصائص السيكومترية للاختبار (الصدق، الثبات)، وقد تم تصحيح استجابات الطلبة، و تحليل استجابات العينة بإستخدام برنامج الحزم الاحصائية SPSS وبرنامج اكسل، لايجاد قيم معاملات الصعوبة والتمييز. وحذفت ١٠مفردات من الاختبار ذات معاملات الصعوبة والتمييز غير المناسبين، وقد روعي ان لا يكون أفراد العينة الاستطلاعية من ضمن أفراد الدراسة الرئيسة، وقد تم اعتماد المعايير الاتية في انتقاء المفردات.

معامل صعوبة المفردة: تم اختيار المفردات التي تقع معاملات صعوبتها ما بين ٠,٤ و ٠,٨ ويشير معامل الصعوبة الى نسبة الطلبة الذين أجابوا على المفردة إجابة صحيحة. وتراوحت مستويات الصعوبة للمفردات في الاختبار من ٠,١ الى ٠,٩١ وهذا يشير الى ان المفردات متوسطة الصعوبة، وبذلك تم حذف المفردات ٢٥، ٣٢، ٣٣، ٤٥، ٦٦، ٧٠،

معامل تمييز المفردة: تم اختيار المفردات التي يزيد معامل تمييزها عن ٠,٤، وهذا ما ايده صلاح علام (٢٠٠٠: ٢٨٩) فيقول إذا كانت قيمة معامل تمييز المفردة ٠,٤٠ او أكثر، فإن هذا يكون دليلاً على ان المفردة تميز بدرجة جيدة بين المجموعتين الطرفيتين، وإذا تراوحت بين ٠,٢٠- ٠,٤٠ فان تمييز المفردة بين المجموعتين يكون لا باس به، وإذا قلت هذه القيمة عن ٠,٢٠ فان تمييزها يكون ضعيفاً، ام إذا كان تمييزها صفر او سالب، فإنه ينبغي حذف المفردة من الاختبار او مراجعتها مراجعة دقيقة. وتراوحت قيم معاملات التمييز للمفردات في الاختبار ما بين – ٠,٣ الى ٠,٣٣، لذلك حذفت المفردات ٢، ٣، ١٤، ٢٥، ٣٢، ٣٣، ٤٥، ٦٦، ٦٩،٧٠.

صدق وثبات الاختبار: تم التحقق من صدق الاختبار من خلال نوعين من الصدق:

صدق محتوى الاختبار: عرض الاختبار على المحكمين، لابداء ملاحظاتهم على مدى ملائمة نتائج تحليل المحتوى والاهداف لجدول المواصفات، ومدى ارتباط كل مفردة بالهدف الذي تقيسه، وملائمة اللغة ووضوحها ودقة التعبير والدقة العلمية والفنية في صياغة المفردات، وقام المحكمون بإبداء ملاحظاتهم من خلال الاستبانة التي وزعت عليهم، وتم الاخذ بملاحظاتهم بعين الاعتبار.

الصدق التلازمي للاختبار: بعد تطبيق الاختبار على عينة استطلاعية مكونة من (١٠٠) طالب، تم إيجاد معامل الارتباط بين الدرجة الكلية للطالب في الاختبار ودرجته في الرياضيات في الصف الثامن الأساسي. وكانت قيمة معامل الصدق التلازمي ٠,٨٣٣.

ثبات الاختبار: تم تحديد قيم الثبات للاختبار من خلال حساب ثبات الاختبار بطريقتي كرونباخ _ الفا، والتجزئة النصفية كمؤشر على ثبات نتائج الاختبار، حيث كانت قيمة معامل كرونباخ _ الفا **٠,٩٢**، ومعامل ثبات بيرسون لنصفي الاختبار ٠,٨٨. وبناءً على ذلك يمكن اعتبار ان الاختبار اداة مناسبة لجمع البيانات الازمة للدراسة.

و- مرحلة تطبيق الاختبار النهائية: بعد كتابة الاختبار بصورتة النهائية، ٦٠ مفردة، تمت كتابة تعليمات الإجابة على الاختبار وذلك بوضع دائرة حول رمز الإجابة الصحيحة،

والزمن الازم للإجابة، وقد تم التنسيق مع مدراء المدارس والمعلمين في المدارس التي اختيرت لتطبيق الاختبار، وقد تم توضيح تعليمات الاختبار وآلية التطبيق. وطبق الاختبار في نهاية الفصل الثاني من العام الدراسي ٢٠١٨/٢٠١٩ بعد تطبيق الاختبار على الطلبة، تم جمع إجابات الطلبة على الاختبار والبالغ عددهم ١٢٣٠طالباً، وتم تصحيح استجابات الطلبة على الاختبار، ومن ثم إدخالها على برنامج Notepad حيث كانت القيمة ١ تعني الإجابة الصحيحة على المفردة، و٠ تعني الإجابة الخاطئة على المفردة. ليتم استخدامها في برنامج التحليل الاحصائي الخاص بالنظرية الحديثة Bilog، لتقدير قدرة الافراد ودالة المعلومات للإجابة على فروض الدراسة.

الأساليب الإحصائية المستخدمة:

١- برنامج التحليل الاحصائي SPSS، للقيام بايجاد قيم معاملات الصعوبة والتمييز للمفردات وإيجاد ثبات الاتساق الداخلي، كرونباخ الفا، ومعامل الثبات بإستخدام طريقة التجزئة النصفية لدراسة الخصائص السيكومترية للاختبار (الصدق، الثبات)، وتحليل التباين للقدرة ودالة معلومات الاختبار.

٢- برنامج التحليل الاحصائي Bilog-Mg واستخدام اسلوب الأرجحية القصوى الهامشية لتقدير قيم مَعلَم القدرة ودالة معلومات الاختبار.

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

نتائج الدراسة وتفسيرها: يستعرض الباحث في هذا الجزء نتائج الدراسة وتفسيرها من خلال عرض نتائج كل فرض وتفسيره في ضوء الاطلاع على نتائج الدراسات السابقة.

١- وصف نتائج الفرض الاول الذي ينص على انه: " توجد فروق ذات دلالة احصائية عند مستوى الدلالة (α= ٠,٠٥) بين متوسطات معاملات دالة معلومات الاختبار في مستويات حجم العينة(كبير، متوسط، صغير)".وللإجابة على هذا الفرض يجب فحص أثر حجم العينة وفقاً للنموذج الثلاثي في متوسط قيم دالة معلومات الاختبار وذلك عن طريق حساب المتوسطات الحسابية والانحرافات المعيارية لقيم دالة المعلومات لكل حجم عينة ويمثل الجدول (١) الاحصاء الوصفي الوسط الحسابي والانحراف المعياري لقيم دالة المعلومات، في اختبار مقرر الرياضيات للصف التاسع الاساسي.

جدول(١)المتوسطات الحسابية والانحرافات المعيارية لتقديرات قيم دالة معلومات الاختبار تبعاً لمتغير حجم العينة(صغيره، متوسطه، كبيره)

حجم العينة	المتوسط الحسابي	الانحراف المعياري
٣٠٠فرد	٠,٤٢	٠,٢١٧
٤٠٠فرد	٠,٦٥	٠,٣٣٤
٥٣٠فرد	٠,٨٤	٠,٢١٣

يتضح من جدول(١) ان هناك فروقاً في المتوسطات الحسابية لقيم دالة معلومات الاختبار وفقاً لحجوم العينة الثلاثة، حيث تغير متوسط دالة المعلومات من ٠,٤٢ عندما كان حجم العينة٣٠٠فرد الى ٠,٨٤ عندما كان حجم العينة ٥٣٠ فرد، وربما تكون الزيادة في دالة المعلومات بسبب زيادة حجم العينة، ناتج من ان متوسط معلمة التخمين يقترب من الصفر كلما زاد حجم العينة، حيث ان كمية المعلومات تزداد كلما قلت قيمة التخمين، لذلك عند الاعتماد على النموذج الثلاثي يتم اخذ معلمة التخمين بعين الاعتبار مما يقلل من أثر التخمين وبالتالي زيادة دالة معلومات الاختبار. ولتحديد فيما اذا كانت هذه الاختلافات ذات دلالة إحصائية عند مستوى الدلالة(α= ٠,٠٥) تم استخدام تحليل التباين الاحادي للفروق بين المجموعات، ويبين الجدول (٢) نتائج تحليل التباين الاحادي للفروق بين المتوسطات الحسابية لقيم دالة معلومات الاختبار تبعاً لمتغير حجم العينة.

جدول(٢) تحليل التباين الاحادي للفروق بين المتوسطات الحسابية لقيم دالة معلومات الاختبار تبعاً لمتغير حجم العينة(٣٠٠، ٤٠٠، ٥٣٠).

مصدر التباين	مجموع المربعات	درجات الحرية	متوسط المربعات	قيمة F	الدلالة الاحصائية
بين المجموعات	٠,٠٦٦	٢	٠,٠٣٣	٤,٤٥	٠,٠٣٧
داخل المجموعات	١,٣١٤	١٧٧	٠,٠٠٧٤٢		
المجموع	١,٣٥٨	١٧٩			

يتبين من جدول(٢) وجود فروق ذات دلالة احصائية عند مستوى الدلالة(α= ٠,٠٥) بين المتوسطات الحسابية لتقديرات دالة معلومات الاختبار أثر التغير في حجم العينة، حيث كانت قيمة ف (٤,٤٥) والدلالة الاحصائية كانت ٠,٠٣٧ وهي دالة احصائياً، وبناء على ذلك يمكن القول بأن دالة معلومات الاختبار تتأثر بتغير مستوى حجم العينة. ولتحديد مصدر تلك الفروق قام الباحث بإجراء مقارنات ثنائية بين الأحجام المختلفة للعينات، ويبين جدول (٣) نتائج المقارنات الثنائية بإستخدام طريقة شافية.

جدول(٣) نتائج المقارنات البعدية بطريقة شيفيه(Scheffe) للفروق في قيم دالة المعلومات وفقاً لحجم العينة

المقارنات الثنائية	الفرق بين المتوسطين	الدلالة الاحصائية
(٣٠٠-٥٣٠)	٠,٤٢	٠,٠٢٢
(٤٠٠-٥٣٠)	٠,١٩	٠,١١٥
(٣٠٠-٤٠٠)	٠,٢٣	٠,٢٣٠

يتبين من جدول (٣) وجود فروق ذات دلالة إحصائية بين متوسطات دالة المعلومات في العينة ٥٣٠، و العينة ٣٠٠ وقد كانت الفروق لصالح العينة ٥٣٠، في حين لا توجد فروق بين العينة الكبير والمتوسطة ، والعينة المتوسطة والصغيرة، اي ان تقدير دالة المعلومات كان الافضل عند حجم العينة ٥٣٠ فرد، وعدم وجود فروق في تقدير دالة المعلومات عندي حجم العينة الصغيرة والمتوسطة. مما يؤكد على أن مصدر التباين بين العينات هو العينة الكبيرة، التي كان الفرق بينها وبين العينة الصغيره ٠,٤٢ وبينها وبين العينة المتوسطة ٠,١٩. وهذا يؤكد على أهمية حجم العينة في تقدير دالة المعلومات وأننا نحتاج في حالة النموذج الثلاثي عينات كبيرة لا تقل عن ٥٣٠ فرد.

٢- وصف نتائج الفرض الثاني الذي ينص على انه: " توجد فروق ذات دلالة احصائية عند مستوى الدلالة (α= ٠,٠٥) بين متوسطات الخطأ المعياري لتقدير القدرة في حجم العينة(كبير، متوسط، صغير)". وللإجابة على هذا الفرض يجب فحص أثر حجم العينة وفقاً للنموذج الثلاثي في متوسط قيم الخطأ المعياري لتقدير القدرة، وذلك عن طريق حساب المتوسطات الحسابية والانحرافات المعيارية لقيم الخطأ المعياري لتقدير القدرة لكل حجم عينة ويمثل الجدول (٤) الاحصاء الوصفي الوسط الحسابي والانحراف المعياري لقيم تقدير القدرة.

جدول(٤) الاحصاءات الوصفية للخطأ المعياري لتقدير قدرة الفرد تبعاً لمتغير حجم العينة(صغيره، متوسطه، كبيره)

حجم العينة	الوسط الحسابي	الانحراف المعياري
٣٠٠فرد	٠,٣٢٤	٠,١٢٣
٤٠٠فرد	٠,٣٣٧	٠,١١٤
٥٣٠فرد	٠,٣١٢	٠,٠١٧

يتبين من الجدول(٤) وجود فروق ظاهرية بين قيم المتوسطات الحسابية للخطأ المعيـــاري لتقدير قدرة الافراد ناتجة عن اختلاف مستويات حجم العينة، حيث يلاحظ ان اكبر قيمة لمتوســـط الخطأ المعياري لتقدير قدرة الافراد كان (٠,٣٣٧) وانحراف معياري(٠,١٢٣) عند حجم العينـــة (٣٠٠)، واقل قيمه (٠,٣١٢) وانحراف معياري(٠,٠١٧) عند حجم العينـــة(٥٣٠)، ويلاحـــظ ان قيمة الخطأ المعياري تقل كلما زاد حجم العينة. اي ان حجم العينة يتناسب عكسياً مع قيمة الخطأ المعياري، والخطأ المعياري يتناسب عكسياً مع دقة تقدير قدرة الفرد.

جدول(٥) تحليل التباين الاحادي للمتوسطات الحسابية للخطأ المعياري لتقدير قدرة الافراد(θ) تبعاً لمتغير حجم العينة

مصدر البتباين	مجموع المربعات	درجات الحرية	متوسط المربعات	قيمة F	الدلالة الاحصائية
بين المجموعات	١,٢١١	٢	٠,٦٠٥٥	٣,٦٥	٠,٠٤٣
داخل المجموعات	٢٠٣,٦	١٢٢٧	٠,١٦٦		
المجموع	٣٠٤,٨١١	١٢٢٩			

يتبين من الجدول(٥) وجود فروق في المتوسطات الحسابية للخطأ المعياري لتقدير قـــدرة الافراد، حيث كانت قيمة ف= ٣,٦٥ وكانت قيمة الدلالة الاحصائية ٠,٠٤٣ وهي اقل من ٠,٠٥ اي انة توجد فروق في المتوسطات الحسابية للخطأ المعياري لتقدير قدرة الافراد تعـــزى لتغيـــر حجم العينة. ولتحديد مصدر تلك الفروق قام الباحث بإجراء مقارنات ثنائية بين الأحجام المختلفـــة للعينات، ويبين جدول (٦) نتائج المقارنات الثنائية بإستخدام طريقة شافية.

جدول(٦) نتائج المقارنات البعدية بطريقة شيفيه(Scheffe) للفروق في قيم الخطأ المعياري لتقدير القدرة وفقاً لحجم العينة

المقارنات الثنائية	الفرق بين المتوسطين	الدلالة الاحصائية
(٣٠٠ - ٥٣٠)	٠,٠١٢-	٠,٠١٢
(٤٠٠ - ٥٣٠)	٠,٠٢٥-	٠,٠٨٥
(٣٠٠ - ٤٠٠)	٠,٠١٣-	٠,٠٦٠

يتبين من جدول (٦) وجود فروق ذات دلالة إحصائية بين متوسطات الخطـــأ المعيـــاري لتقدير القدرة في العينة ٥٣٠، و العينة ٣٠٠ وقد كانت الفروق لصالح العينة ٥٣٠، في حـــين لا توجد فروق بين العينة الكبير والمتوسطة ، والعينة المتوسطة والـــصغيرة، اي ان قـــيم الخطـــأ

المعياري لتقدير قدرة الافراد كانت الافضل واقل ما يمكن عند حجم العينة ٥٣٠ فرد، وعدم وجود فروق في قيم الخطأ المعياري لتقديرات قدرة الافراد عندي حجم العينة الصغيرة والمتوسطة، مما يؤكد على أن مصدر التباين بين العينات هو العينة الكبيرة، التي كان الفرق في الخطأ المعياري بينها وبين العينة الصغيره -٠,٠١٢ وبينها وبين العينة المتوسطة -٠,٠٢٥. وتُظهر النتائج الى اختلاف دقة تقدير قدرات الافراد باختلاف حجم العينة، حيث يقل الخطأ المعياري كلما زاد حجم العينة، اي تزداد دقة تقدير قدرة الفرد بزيادة حجم العينة.

مناقشة النتائج

مناقشة نتائج الفرض الاول: توصلت النتائج الى وجود فروق ذات دلالة إحصائية عند تقدير متوسطات دالة معلومات الاختبار تعزى لاختلاف حجم العينة، اي ان متوسطات قيم دالة معلومات الاختبار قد تغيرت بتغير حجم العينة، ولمعرفة الاختلاف بين متوسطات تقدير دالة معلومات الاختبار باختلاف حجم العينة تبين بعد اجراء المقارنات البعدية ان الفروق لصالح العينة(٥٣٠) مقابل العينتين (٣٠٠)، (٤٠٠)، وهي ما تؤكدة نتيجة الفروق بين المتوسطات لتقدير دالة المعلومات، حيث كان هناك فروق ذات دلالة احصائية بين متوسطات دالة المعلومات تعزى لحجم العينة لصالح العينة (٥٣٠) الكبيرة، اي ان دقة تقدير دالة معلومات الاختبار تزداد بزيادة حجم العينة. وتتفق هذه النتيجة مع ما توصلت اليه دراسة كلير (Keller, 2002) ودراسة حابس الزبون (٢٠١٣) حيث أظهرت نتائجهما ان مقدار تقديرات دالة المعلومات يتغير بتغير حجم العينة، إذ يزداد بزيادة حجم العينة.

مناقشة نتائج الفرض الثاني حيث توصلت نتائج الدراسة ايضاً الى وجود فروق في المتوسطات الحسابية للخطأ المعياري لتقديرات قدرة الافراد ذات دلالة إحصائية تعزى لتغير حجم العينة، اي ان متوسطات الخطأ المعياري لتقديرات قدرة الافراد قد تغيرت بتغير حجم العينة. ولمعرفة الاختلاف بين متوسطات الخطأ المعياري لتقديرات قدرة الافراد باختلاف حجم العينة تبين بعد اجراء المقارنات البعدية ان الفروق لصالح العينة(٥٣٠) مقابل العينات الاخرى، وهي ما تؤكدة نتيجة الفروق بين المتوسطات للخطأ المعياري لتقدير قدرة الافراد حيث كان هناك فروق ذات دلالة احصائية بين متوسطات الخطأ المعياري لتقديرات قدرة الافراد لحجم العينة، ولصالح العينة (٥٣٠) الكبيرة، اي ان قيمة الخطأ المعياري لتقديرات قدرة الافراد تتناسب عكسياً مع حجم العينة، وهذا يعود ربما الى ارتباط حجم العينة بالخطأ المعياري، حيث ان زيادة حجم العينة يقلل من الخطأ المعياري للتقدير، وبالتالي تزيد الدقة في تقدير القدرة. وهذه النتيجة تتفق

مع ما توصلت الية دراسة زايد بني عطا (٢٠١٤) و دراسة دي لا توري ويوان (De La Torre & Yuan, 2010) التي اشارتا الى زيادة دقة تقدير القدرة بزيادة حجم العينة. الا انها تتعارض مع ما توصلت اليه دراسة يمان شما (٢٠١٣) من ان اختلاف حجم العينة لا يؤثر على دقة تقدير قدرة الافراد، حيث ان قيم متوسطات الخطأ المعياري لتقديرات قدرة الافراد في كل حجم من حجوم العينة المستخدمة متساوية. كما بينت المقارنات البعدية ان حجم عينة ٥٣٠ فرد يمكن ان يعطي تقديرات دقيقة لقدرة الافراد بإستخدام النموذج الثلاثي.

التوصيات

بالرغم من ان تطبيقات دوال المعلومات اقتصرت على نماذج الاستجابة للمفردة ثنائية الدرجة، الا ان الباحثين بدئوا بدراسة خصائص هذه الدوال في حالة استخدام نماذج الاستجابة متعددة الاقسام، ونماذج الاستجابة متعددة الابعاد (صلاح علام، ٢٠٠٢، ١١٢). لذا يوصي الباحث.

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

٢- إجراء دراسة تأثير تغير حجم العينة على قدرة الفرد ودالة المعلومات لبيانات غير ثنائية متعددة الاستجابة وفق النموذج متعدد التدريج.

٣- إجراء دراسة للمقارنة بين دالة المعلومات المحسوبة بالاعتماد على النموذج الثلاثي والثنائي المعلمة.

المراجع

أمينة كاظم(١٩٨٨). دراسة نظرية نقدية حول القياس الموضوعي للسلوك(نموذج راش). سلسلة الكتب المتخصصة، الطبعة الاولى، الكويت.

حابس الزبون (٢٠١٣). أثر حجم العينة على تقدير دالة المعلومات للاختبار والخطأ المعياري في تقديرها بإستخدام النظرية الحديثة في القياس. مجلة جامعة النجاح للأبحاث (العلوم الانسانية) المجلد٢٧(٦)٢٠١٣.

حسين عبد النبي القيسي (٢٠١٣): دقة تقدير معالم الفقرة والقدرة بإستخدام نماذج نظرية استجابة الفقرة البارامترية واللابارامترية باختلاف حجم العينة وطول الاختبار: دراسة مقارنة. رسالة دكتوراه غير منشورة، كلية التربية – جامعة اليرموك.

راجي عوض الصرايرة، وحابس سعد الزبون (٢٠١٧). أثر ترتيب فقرات اختبار الاختيار من متعدد وفقاً لمعالم الفقرة على قدرة المفحوصي ودالة المعلومات للاختبار وفقاً للنموذج ثلاثي المعلمة. مجلة جامعة الحسين بن طلال للبحوث، مجلد ٣ العدد١.

راشد الدوسري (٢٠٠٤). القياس والتقويم التربوي الحديث، عمان، دار الفكر.

زايد بني عطا (٢٠١٤). تقصي دقة تقدير النموذج اللوجستي ثلاثي المعلمة لمعالم الفقرة وقدرة الأفراد في ضوء تغير طول الاختبار وحجم العينة. مجلة جامعة الشارقة للعلوم الانسانية والاجتماعية، ١١(٧)، ١-٣٧.

السيد محمد ابو هاشم (٢٠٠٦).دراسة مقارنة بين النظرية التقليدية ونموذج راش في اختيار فقرات مقياس مداخل الدراسة لدى طلاب الجامعة،مجلة كلية التربية، جامعة الزقازيق، العدد٥٢ ،١-٥٢.

صلاح احمد مراد، وأمين علي سليمان (٢٠٠٥). والمقاييس في العلوم التربوية والنفسية خطوات إعدادها وخصائصها، الكويت، دار الكتاب الحديث.

صلاح الدين محمود علام (١٩٨٦). تطورات معاصرة في القياس النفسي والتربوي. مطابع القبس التجارية: جامعة الكويت.

صلاح الدين محمود علام (٢٠٠٢). القياس والتقويم التربوي والنفسي أساسياته وتطبيقاته وتوجهاته المعاصرة. القاهرة: دار الفكر العربي.

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

ماجد محمد الخياط (٢٠١٢). درجة مطابقة اختبار تحصيلي وفق نموذج راش احادي المعلمة في الكشف عن مستوى المعرفة العلمية في المهارات الرياضية لدى طلبة الصف الثامن الاساسي. مجلة جامعة الاقصى (سلسلة العلوم الانسانية)١٦(١)، ٨٧-٢١٤.

نضال كمال الشريفين (٢٠١٢). أثر طريقة تقدير معالم الفقرة وقدرات الافراد على قيم معالم الفقرة، والخصائص السيكومترية للاختبار، في ضوء تغير حجم العينة. المجلة التربوية، الاردن، العدد٢٦،ص ص١٧٧-٢٣٨.

يمان نزار شما(٢٠١٣). اثر حجم العينة على دقة تقدير صعوبة المفردات وقدرة الافراد بإستخدام نموذج راش. مجلة الاداب، جامعة بغداد، العدد، ١٠٥، ٦٧٣-٦٩٨.

- Belov, D. I. & Armstrong, R. D. (2010). **Automatic detection of answer copying via kullback-leibler divergence and K- index**. Applied psychological Measuement, 34(6), 379-392.
- Bock, R.D. & Wood, R. (1971): Test theory Annual Review of Psychology, 22.
- Crocker, L. & Algina, J. (1986). **Introduction to classical and Modern Test Theory**. New York, CBS College Publishing.
- Cuervo, E.&Andrade, J. (2004)**: Modeling Abilities in 3 –IRT Models**. Revista Colombiana de Estadĭstica, 27(1).27-41.
- De la Torre, Jimmy; Yuan Hong (2010). **Parameter Estimation with Small Sample Size A Higher-Order IRT Model Approach**. Applied Psychological Measurement. 34 (4), p267-285.
- Emberson, S. & Reise, S (2000). **Item Response Theory for Psychologists New Jersey**: Lawrence Earlbaum Associates, Publishers Mahwa.
- Fischer, G. H. (1973)**: The Linear Logistics model as an instrument in educational Research and psychological**, voI .37
- Hambleton, R. K. &Swaminathan, H (1985). **Item Response Theory, Principles and Applications**. Nijhoff Publishing amembers of the Kluwer Academic Publishers Group, Bston. Kluwer.
- Hambleton, R., Swaminathan, H. &Rogers, H. (1991). **Fundamentals of item response theory**, International Educational and Professinonal. Publisher Newbury Park.
- Keller, L., (2002). **Small –sample item parameter estimation in the three parameter logistic model**: Using collateral information. Doctoral dissertations Available from proquest. AA13068572.
- Lord, F. (1980). **Applications of Item Response Theory to Practical testing problems**. New Jersey, Lawrence Erlbaum Associates,
- Magno, C. (2009). **Demonstrating the Difference between. Classical-Test Theory and item Response Theory Using Derived Test Data**. The International Journal of Educational and Psychological Assessment1 (1)1*11.
- Nitko, A. (2001). **Educational Assessment of Students**. Amazon, UAS: Merrill.
- Reeve, B. (2004). **An introduction to modern measurement theory**. Division of fencer control and population science, National Cancer Institute.

- Schrodt, p. (2007**): Inductive Event Data Scaling Using Item Response Theory. Paper prepared for delivery at the Summer Meeting of the Society for political Methodology**, Pennsylvania State University
- Stone, M. &Yumoto, F. (2004). **The Effect of Sample Size ForEstimating Rasch/ IRT Parameters with Dichotomous Items**. Journal of Applied Measurement. Vol. 5, No. 1, pp.48 - 61.
- Wright, B .D. & Stone, M .H. (1999). **Measurement essentials.** Wilmington, Delaware.
- Yen, W. (1992). **Item Response Theory (6th Ed)**. Encyclopedia of Edu ional Reserch. NY: Macmillam