أثر حجم العينة على خصائص الاختبار التحصيلي ومعادلة درجاته

إعداد

د. حسان غازي العمري
جامعة جدارا

د. احمد يوسف قواسمة
جامعة اليرموك

مجلة بحوث التربية النوعية – جامعة المنصورة
عدد (٢٦) – يوليو ٢٠١٢

أثر حجم العينة على خصائص الاختبار التحصيلي ومعادلة درجاته

إعداد

د . حسان غازي العمري * د . احمد يوسف قواسمة **

ملخص البحث:

هدفت هذه الدراسة الى معرفة اثر حجم العينة على الخصائص السيكومترية للاختبار وخصائص فقراته ومعادلة درجاته، حيث تم بناء اختبار في القدرة الرياضية للصف الثامن من (٨٠) فقرة تحققت فيه مظاهر الصدق الظاهري وصدق البناء ، وبثبات قدره (٠,٩١) تم تطبيقه على عينة من ٣٢٠٠ مفحوصا من الجنسين ، تم ادخال استجاباتهم باستخدام البرنامج الاحصائي spss ثم تم سحب (٦) عينات جزئية عشوائيا بحجوم (١٦٠٠ ، ٨٠٠، ٤٠٠، ٢٠٠، ١٠٠، ٥٠) على التوالي .اشارت نتائج الدراسة الى ان الخصائص المتمثلة بالثبات والتمييز وصعوبة الفقرات لا تختلف باختلاف حجم العينة المسحوبة عشوائيا اذا زاد حجمها عن ١٠٠ ، كما اشارت نتائج الدراسة الى ان تطبيق الاختبار على عينات اقل من ١٠٠ يهدد صدق محتوى الاختبار وذلك لارتفاع نسبة الفقرات المحذوفة وفق مؤشري الصعوبة والتمييز لتصل الى حوالي ٣٩٪ من فقرات الاختبار الاصل.

الكلمات المفتاحية: العينة العشوائية ، الاختبار، حجم العينة ، الثبات ، الصعوبة والتمييز .

* جامعة جدارا
** جامعة اليرموك

The Effect of the Sample Size on the Test and it's Items characteristics and equating scores

Dr.Hassan Alomari \ Jadara Uiniversity
Dr.Ahmad Qawasmeh \ Yarmouk University

Abstract

This study aimed to determine the impact of sample size on the test properties and the characteristics of its items and equating scores. So a test in MATH ability of the eighth grade of 80 items were constructed with good face, construct validity, and (0.91) reliability. This test has been applied to a sample of 3200 examinees, and their responses entered using the statistical program spss.

Then 6 samples of sizes (1600, 800, 400, 200, 100, 50) was randomly withdrawn. findings indicated that the test reliability and item's discrimination and the difficulty do not vary according to the size of the sample , especially if the sample size more than 100 examinees , also the results of the study indicated that the application of the test on samples less than 100 examinees threatening the content validity of the test because of the high proportion of deleted items according to difficulty and discrimination indices which reach 40% of the original test items.

Keywords:

Random sample, test, sample size, reliability, difficulty and discrimination.

أثر حجم العينة على خصائص الاختبار التحصيلي ومعادلة درجاته

إعداد

د . حسان غازي العمري * د . احمد يوسف قواسمة **

المقدمة:

تعتبر الاختبارات التحصيليه واحدة من وسائل التقويم المتنوعة، ومن اهم الأدوات التي يستند إليها في اتخاذ القرارات ورسم السياسات والتخطيط في العملية التربوية. وعليه فإن جودة تلك الاختبارات وتمتعها بخصائص سيكومترية جيدة ركيزة أساسية تصب في تحسين تلك السياسات ورفع مستوى دقة القرارت التربوية.

إن الخصائص السيكومترية للاختبار وخصائص فقراته تتأثر بعوامل عدة أهمها حجم عينة الدراسة، فالمتمعن في عملية تقدير مؤشرات الاختبار من صدق وثبات، وتقدير لمؤشرات فقراته من صعوبة وتمييز وغيرها، يجد أن جميعها تعتمد على عينة المفحوصين الذين يجرّب أو يطبق عليهم الاختبار، فجميع هذه المؤشرات يتم تقديرها مباشرة من خلال استجابات العينة على فقرات الاختبار، ولذلك تعتبر هذه المؤشرات والخصائص غير مستقلة عن عينة التجريب أو التطبيق .

وعليه فإن العديد من الباحثين في مجال تطوير الاختبارات والمقاييس يطرحون تساؤلا مستمرا حول الحجم المناسب للعينة (ملحم، ٢٠٠٢)، وهو تساؤل ليس من السهل الإجابة عنه، فقد أشار فيرجسون(Ferguson, 1989) في هذا الصدد إلى عدم إمكانية الحصول على إجابة لمثل هذا التساؤل، بل وأن معرفة الحجم المناسب أمر لا يخلو من التعقيد ويعتمد على العديد من العوامل مثل: الكلفة المادية، وتوفر أفراد العينة، والجهد والوقت اللازمين لعملية جمع البيانات وغيرها. فيما أشار رسل (Russel,2001) إلى أن هناك العديد من الأسئلة التي يجب إثارتها قبل السؤال عن حجم العينة مثل: ما هو هدف الدراسة؟ وكيف تخطط لعملية القياس؟ وهل تتوفر أدوات اخرى؟ وكم الوقت المتوفر؟ وما هو متغير الاستجابة؟ وما هي القيود العملية في الدراسة؟ وغيرها من التساؤلات الأخرى.

خلفية الدراسة:

تختلف وجهات النظر حول الحجم المناسب للعينة من اجل الحصول على خصائص سيكومترية جيدة للاختبار والفقرات بشكل عام، فقد أشار البعض أن الحجم المناسب لتلك العينة يجب أن يتراوح بين نسبتي(٥٪- ١٠٪) من حجم المجتمع أو بحجم لا يقل عن ٤٠٠ مفحوص، على أن

* جامعة جدارا
** جامعة اليرموك

يتم اختيار الافراد لتلك العينة وفق أسلوب يحد من خطأ الصدفة الذي ينعكس على درجة تمثيل تلك العينة للمجتمع (Henrysoon,1971).

وفي الأدب المتعلق بخاصية ثبات الاختبار تحديدا، يشير كلاين (Kline,1986) إلى ان أقل حجم للعينة يجب أن لا يقل عن ٣٠٠ مفحوصا وهو الحجم الذي اطلق عليه كل من نانللي وبريستين(Nunnally&Bernstien,1994) وسيجال(Segall,1994) بالعينة الصغيرة، في حين حدد كارتر(Charter, 2003) العينة الصغيرة بـ ٤٠٠ مفحوص من أجل الحصول على تقدير غير متحيز لثبات المجتمع من ثبات العينة، أما دراسة يوردكول (Yurducul, ٢٠٠٨) فقد أشارت إلى أن الحد الأدنى اللازم لحجم العينة لتقدير معامل الثبات مرتبط بالقيمة الأولية المحسوبة للجذر الكامن (Eigen Value) والناتجة من التحليل العاملي للمكونات الأساسية. فإذا كانت قيمة الجذر الكامن الأول اكبر من (٦)، فان معامل كرونباخ الفا (الثبات) في هذه الحالة يكون مقدرا جيدا (robust Estimator) للقيمة المقابلة له في مجتمع هذه المعاملات حتى ولو كان حجم العينه (٣٠) مفحوصا، وأما اذا كانت قيمة الجذر الكامن ما بين (٦- ٣) فان حجم العينة المطلوب يجب أن لا يقل عن (١٠٠) مفحوص كي يتم الحصول على تقدير غير متحيز لقيمة الفا المقدرة لثبات المجتمع.

أما على مستوى الفقرات وعند القيام بتقدير مؤشرات خصائص الفقرة وخصوصا مؤشر تمييزها اعتمادا على مجموعة الأداء المرتفع بنسبة ٢٧٪ ومجموعة الأداء المنخفض بنسبة ٢٧٪ من المفحوصين، فإنه يوصى بأن لايقل حجم تلك النسب عن ٥٥ مفحوصا، مما يعني أن حجم العينة كاملا سيبلغ ٢٠٠ مفحوصاً وذلك كي يتم الحصول على أفضل تباين بين المجموعتين (Anastasi,1988). أما كروكر والجينا (Krocker&ALgina,1986) فقد أشارا إلى أن الحجم المناسب لعينة المفحوصين لغايات حساب مؤشرات خصائص الاختبار وفقراته ينبغي أن لا يقل عن (٢٠٠) مفحوصا لأن ذلك يقلل من حجم أخطاء القياس إلى الحد الأدنى، فيما أشار نانالي (Nunnally,67) إلى أن حجم العينة متعلق بعدد فقرات المقياس، اذ يجب ان يتراوح الحجم من (٥- ١٠) أضعاف عدد الفقرات، أما في النظرية الحديثة للقياس فأوضح أن حجم العينة اللازم يختلف باختلاف النموذج المستخدم (احادي المعلم ، ثنائي المعلم ، ثلاثي المعلم) ويتراوح من ٢٠٠- ١٠٠٠ مفحوص.

في التحليل العاملي والذي يستخدم بشكل كبير في الكشف عن خصائص الفقرات وارتباطاتها، وتحديد عدد عوامل الاختبار أو المقياس، فالمتوفر في الادب النظري من دراسات يشير إلى أن هناك اتجاهان في تحديد الحد الأدنى لحجم العينة المطلوب، فالاتجاه الأول يشير إلى أن حجم العينة يحدد بشكل مستقل عن اي متغيرات، فيما يربط الاتجاه الثاني حجم العينة بعدد الفقرات التي يتكون منها الاختبار أو المقياس. ففي الاتجاه الأول أشار كل من جورسوش (Gorsuch, 1983) وكلاين (Kline ,1979) إلى أن الحجم اللازم للعينة يجب أن لا يقل عن ١٠٠، فيما أكد هيوشيسون وسوفرونيو (Hutcheson & Sofroniou , 1999) إلى أن الحجم يجب أن لا يقل عن ١٥٠ خصوصا عندما تكون المتغيرات مرتبطة مع بعضها البعض بشكل غير مرتفع، أما جيلفورد

(Guilford ,1954) وماكولوم وزملاؤه (MacCallum, Widaman, Zhang & Hong, 1999) فقد اقترحوا بأن لا ينقص حجم العينة عن ٢٠٠ مفحوصا، فيما اقترح كاتل Cattell (1978,) أن يكون الحد الادنى ٢٥٠ مفحوصا، مع احتمالية أن يصل هذا الحد إلى ٥٠٠ مفحوصا وهو ما أشار له (Comreyg & Lee ,1992).

أما في الاتجاه الثاني والذي يربط حجم العينة بعدد الفقرات فقد أشار هير وزملاؤه Hair, (Anderson, Tatham & Black ,1995) إلى أن النسبة بين عدد الفقرات وعدد المفحوصين يجب أن تكون ١:٢٠ فيما بين نانالي (Nunnally, 1978) إلى أن النسبة يجب أن تكون ١٠ مفحوصين لكل فقرة، وهو ما أكده بريانت ويارنولد (Bryant and Yarnold, 1995) فيما أشار جروش (Gorsuch, 1974) إلى أن النسبة يجب أن لا تقل عن ١:٥ بين الفقرات والمفحوصين.

الدراسات السابقة:

ان المتصفح للدراسات المنشورة في هذا المجال يجد شحا واضحا في الدراسات التي تتعلق تحديدا باثر حجم العينة على الخصائص السيكومترية للاختبار وخصائص فقراته تحديدا سواء العربية منها أو الاجنبية، اذ أن معظمها تلك الدراسات قد تناول أثر حجم العينة على المقاييس النفسية أو التربوية وليس الاختبارات التحصيلية بالذات، اذ لم يحصل الباحثان على اية دراسة عربية أو اجنبية تتعلق بأثر حجم العينة على الاختبار التحصيلي ولذلك فسيتم عرض الدراسات المتعلقة بأثر حجم العينة على خصائص المقاييس بشكل عام.

لقد قام الميامي(٢٠٠٥) بدراسة هدفت إلى التعرف على أثر حجم العينة وطول المقياس في خصائصة السيكومترية، والمتمثلة بالثبات والصدق والتوزيع الطبيعي وحساسية المقياس، استخدم في دراسته مقياسا لسمة الصبر تم تطبيقه على عينات بحجوم مختلفة(٦٠٠، ٥٠٠، ٤٠٠، ٣٠٠، ٢٠٠، ١٠٠) فردا، توصل فيها إلى أن زيادة حجم العينة يؤدي إلى تحسن خصائص المقياس، وفي مختلف الأنماط للمقياس. إضافة إلى ذلك، فقد بينت النتائج إلى أنه كلما زاد طول المقياس كلما قلت الحاجة إلى اعتماد حجم كبير للعينة.

وفي دراسة الكبيسي (١٩٩٥) لأثر حجم العينة وحجم المجتمع الإحصائي في مؤشر معامل التمييز لفقرات المقاييس النفسية، تم تطبيق مقياس الثقه بالنفس على عينات بأحجام مختلفة (٤٠٠، ٣٠٠، ٢٠٠، ١٠٠) تم اختيارها من مجتمع من(٢٠٠٠) طالب وطالبة. ومن ثم اختيار عينات أخرى بنفس الحجوم من مجتمع اكبر مؤلف من (٥٠٠٠) طالبا وطالبة. أشارت نتائج الدراسة إلى أن أفضل حجم عينه فيما يتعلق بمؤشر تمييز الفقرة، كان الحجم(٣٠٠) في المجتمع الأكبر والمؤلف من(٥٠٠٠) طالبا وطالبة ، والحجم(١٠٠) في المجتمع المؤلف من (٢٠٠٠) طالبا وطالبة.

كما وأجرى جافالي (Javali,2011) دراسة حول أثر حجم العينة عل تقديرات الثبات بإسلوب الاتساق الداخلي لكل من معامل الثبات الفا، وجاما ، واوميجا، توصل فيها إلى أن أفضل تقدير لكل منها تم ملاحظته في العينات التي لا تقل عن ٥٠ ولذلك وضحت الدراسة أنه لتقدير الثبات لمقياس من ٥ فقرات فاكثر فانه يلزمنا عينة من ٥٠ مستجيبا على الاقل.

كما وأجرى دالتون وكوزير (Dalton & Cosier,1988) دراسة لمقارنة الخصائص السيكومترية لمقياس المواطنة وخصائص فقراته باستخدام عينات مختلفة الحجوم، والعمر بالنسبة للمستجيبين ولدى ثقافات مختلفة حيث تم إعداد مقياس لأغراض جمع البيانات اللازمة، وقد أظهرت النتائج غياب الفروق الدالة احصائياً والمتعلقة بمؤشرات صدق الفقرة والصدق العاملي للمقياس الذي تم استخدامه باختلاف حجم العينة.

كذلك اجرى باريت وكلاين (Barrett and Kline,1981) دراسة استخدمت فيها عينتين تجريبيتين كبيرتين، ثم تم سحب عينات جزئية من احجام مختلفة من العينتين الكبيرتين واجراء تحليل عاملي للعينات الجزئية لمقارنة النتائج الخاصة بالعينات الجزئية مع العينات الكبيرة، أشارت نتائج دراستهما إلى أن الوضع يكون جيدا عندما تكون العينة الجزئية مكونة من ٤٨ فردا في المقاييس التي تتكون من ١٦ فقرة وكذلك في العينة التي تتكون من ١١٢ فردا في المقاييس المكونة من ٩٠ فقرة .

كذلك قام هنسون وروبرت (Henson & Roberts 2006) بدراسة تحليلية ل ٦٠ بحثا تم نشرها حول التحليل العاملي في ٤ مجلات هي: *Educational and Psychological Measurement, Journal of Educational Psychology, Personality and Individual Differences*, and *Psychological Assessment*. توصلا فيها إلى أن الحد الأدنى للعينة في تلك الابحاث كان ٤٢ فردا، فيما أشارت دراسة فابرجار وزملاؤه Fabrigar, Wegener, MacCallum, and Strahan ,1999)) والتي قدم فيها تحليلا للبحوث المنشورة في مجلتي *Journal of Personality and Social Psychology (JPSP)* and *Journal of Applied Psychology (JAP)*. إلى أن ما نسبته ١٩٪ من البحوث في مجلة JPSP وما نسبته ١٤٪ من البحوث في مجلة JAP لم تقل العينات فيها عن ١٠٠ مفحوصا.

مشكلة الدراسة وأهميتها :

لقد بينت خلفية الدراسة إلى انه لا توجد إجابة واضحة حول حجم العينة اللازم لتطوير الاختبار، وأنها قضية ما زالت خلافية وتتعدد فيها الاراء، الأمر الذي يجعلها قضية مفتوحة للمزيد من البحث، فالأبحاث التي كتبت في هذا الصدد شحيحة وقد أكد ماكالوم وزملاؤه (MacCallum, Widaman, Zhang & Hong, 1999) بأن الأبحاث التي قدمت توصيات حول حجم العينة في مجال العلوم السلوكية قليلة جدا، بالإضافة لذلك فإن تعدد الاراء قد جعل تخفيض حجم العينة امرا مستباحا ومقبولا من قبل الباحثين بذرائع كثيرة منها ضيق الوقت، والجهد اللازم، والتكلفة، وجمع البيانات، بل إن هناك ميلاً ذاتيا واضحا عند الباحثين تجاه تخفيض حجم العينة عن الحدود الدنيا الموصى بها وهو ما أكده كروكر والجينا (Crocker&Algina,1986) حيث أشارا إلى أن معظم الباحثين وطلبة الدراسات العليا بشكل خاص والذين يطورون أدوات واختبارات في أطروحاتهم يميلون إلى الاعتماد على عينات صغيرة مع أن الحد الأدنى المرغوب به للعينة ٢٠٠ مفحوصا.

ولذا يبرز تساؤل مهم هنا يمثل مشكلة الدراسة الرئيسة وهو: ما الذي سيختلف في خصائص الاختبار وفقراته لو كان هناك التزاما من الباحثين بالحدود الدنيا (أو اكثر) لحجم العينة ؟. فالواقع يشير الى أن الكثير من الباحثين ليس لديهم القناعة الكافية بضرورة زيادة حجم العينة لتوقعهم بأن الأمر ليس له تأثير، وعليه فيمكن ان تصاغ اسئلة الدراسة كما يلي:

١. ماثر حجم العينة في استقرار مؤشر صعوبة الفقرة للاختبار؟
٢. ما أثر حجم العينة في استقرار مؤشر تمييز الفقرة للاختبار؟
٣. ما اثر حجم العينة على عدد الفقرت المحذوفة من الاختبار وفق حدود مؤشري الصعوبة والتمييز؟
٤. ما أثر حجم العينة في تقدير معامل الثبات للاختبار؟
٥. ما اثر حجم العينة في معادلة درجات الاختبار؟

أما أهمية هذه الدراسة فتتمثل في أنها تتعلق بأثر حجم العينة على الاختبار التحصيلي بالذات والذي ندرت الدراسات حول اثر حجم عينة المفحوصين على خصائصه وخصائص فقراته، بالإضافة إلى أن الاختبارات التحصيلية مرتبطة في عملية تطويرها بجداول مواصفات تمثل صدق محتوى الاختبار، ولذا فان التغيير الناتج عن حجم عينة المفحوصين قد يؤدي إلى تغير في خصائص الفقرات وبالتالي تغييرا في احتمالات قبول الفقرة أو حذفها من الاختبار الأمر الذي قد يهدد صدق محتوى الاختبار التحصيلي والذي يمثل أحد أهم خصائصه .

الطريقة والإجراءات:

مجتمع الدراسة وعينتها:

تكون مجتمع الدراسة من ٣٢٠٠ طالبا وطالبة من طلبة الصف الثامن في المدارس الحكومية في مدينة اربد مناصفة بين مدارس الذكور والإناث، وقد اعتبر هذا المجتمع بمثابة العينة الكبيرة التي تم سحب عينات فرعية منها. فيما تكونت عينة الدراسة من ٦ عينات فرعية وفق الحجوم (١٦٠٠،٨٠٠،٤٠٠،٢٠٠،١٠٠،٥٠) تم سحبها بطريقة عشوائية بتقسيم العينة إلى النصف في كل مرة باستخدام برنامج الحزمة الإحصائية للعلوم الإنسانية SPSS.

أداة الدراسة (الاختبار):

تم تطوير أداة هذه الدراسة والمتمثلة باختبار تحصيلي في مادة الرياضيات لقياس القدرة الرياضية لدى طلبة الصف الثامن الأساسي، حيث تألف الاختبار بصورته الأولية من (١٠٠) فقرة ، تم إعدادها من خلال الاسترشاد بخطوات بناء الاختبار التحصيلي إلي أشار لها ملمان وجري (Millman & Gree,1989) وذلك بهدف قياس درجة امتلاك طلبة الصف الثامن للمهارات الأساسية في مجالات (الأعداد، الكسور، الهندسة، الإحصاء).

ولتحقيق ذلك فقد تم الإطلاع على نتاجات التعلم من الصف الأول الأساسي إلى الصف الثامن الأساسي من قبل الباحثين والمتوفرة في دليل المعلم، وتم بناء جدول المواصفات من بعدين، البعد

الأول ويمثل الموضوعات التي يتضمنها المحتوى، والبعد الثاني ويمثل مستويات النتاجات لتلك لموضوعات وهي ثلاثة مستويات (المعرفه، الفهم والتطبيق، القدرات العقلية العليا). بعد ذلك تم تحديد الأوزان لكل موضوع من بين الموضوعات وفق عدد النتاجات في كل منها، ووزن كل مستوى من بين المستويات حيث تم تحديد (٢٠٪ للمعرفة و٥٠ ٪ للفهم والتطبيق و٢٠٪ للقدرات العقلية العليا)، ومن ثم توزيع فقرات الاختبار على الخلايا الناتجة من التقاطع فيما بينهما، بحيث تكوّن الاختبار بصورته الأولية من (١٠٠) فقرة من نوع الاختيار من متعدد، لكل فقرة ٥ بدائل، واحد منها فقط يمثل الإجابة الصحيحة لتلك الفقرة.

صدق الأداة(الاختبار) وثباتها:

للتحقق من دلالات صدق محتوى الاختبار، ثم عرضه على مجموعة المختصين في مجال تدريس الرياضيات من معلمين ومشرفين في تدريس الرياضيات وقد بلغ عددهم (٩) محكمين ويحملون الدرجة الجامعية الأولى أو الثانية (البكالوريوس أو ماجستير) وبخبرات تعليمية لا تقل عن عشرة سنوات بالاضافة إلى ثلاثة من المختصين في القياس والتقويم.

وقد طلب منهم تقديم ملاحظاتهم حول فقرات الاختبار من حيث وضوحها، وخلوها من الأخطاء وبيان درجة تمثيل الفقرة لما تقيسه وتغطيه من المجال التعليمي المراد قياسه ، كما وتم الطلب منهم ابداء الرأي فيما يتعلق بالجوانب الفنية في إخراج الاختبار، وتقديم ما يرونه مناسبا من ملاحظات أخرى تتعلق بالتعديل أو الإضافة أو الحذف لأي فقرة.

وبناء على نتائج التحكيم، فقد تم حذف (٩) فقرات وتعديل فقرات أخرى ليصبح الاختبار من ٩١ فقرة ، كما وتم تقدير الزمن اللازم للاختبار من قبل المحكمين بـ (٩٠ دقيقة). بعد ذلك تم تطبيقه على عينة مؤلفة من (٣٠٠) طالبا وطالبة تم اختيارهم عشوائيا من مجتمع الدراسة وذلك لغايات تحري ثبات الاختبار، والكشف عن خصائص فقراته، والتعرف إلى الملاحظات التي قد تظهر أثناء عملية التطبيق والتحقق من خلوه من الاخطاء الطباعية واللغوية وما يتعلق بتعليمات الاختبار وإخراجه والتعرف إلى الوقت اللازم فعليا للاستجابة على فقراته. ثم حُلّلت استجابات أفراد هذه العينة لغايات حساب مؤشر الصعوبة ومؤشر التمييز للفقرات ومعامل الثبات للاختبار، حيث تم شطب "١١" فقرة لعدم تحقيق شرط مدى مؤشر الصعوبة(٠,٧- ٠,٣) ومؤشر التمييز(٠,٣- ١) . وبعد شطب هذ العدد من الفقرات بلغ معامل ثبات الاختبار ٠,٩١ باستخدام معادلة كودرشارتسون KR20 ، وبذلك أصبح الاختبار بصورته النهائية من (٨٠) فقره.

تطبيق الصورة النهائية للاختبار:

تم تطبيق الاختبار بصورته النهائية والمؤلفة من (٨٠) فقرة على عينة الدراسة الكلية والبالغة (٣٢٠٠) طالبا وطالبة، وتم الاستعانة بمشرفي مادة الرياضيات في الفصل الدراسي الأول لعام ٢٠١٠/ ٢٠١١ للمساهمة في عملية التطبيق.

متغيرات الدراسة:

تضمنت الدراسة المتغيرات التالية:

المتغير المستقل:

حجم العينة: وهو عبارة عن عدد المفحوصين الذين يخضعون لتطبيق الاختبار المعد كأداة في هذه الدراسة، ويقع حجم العينة في (٦) قيم هي (١٦٠٠،٨٠٠،٤٠٠،٢٠٠،١٠٠،٥٠).

المتغيرات التابعة: هناك ٥ متغيرات تابعة هي:

- مؤشر معامل الصعوبة للفقرة.
- مؤشر معامل التمييز للفقرة.
- عدد الفقرات المحذوفة وفق مؤشري الصعوبة والتمييز للفقرة .
- معامل الثبات للاختبار .
- العلامات المئينية المقابلة للعلامات الخام في كل اختبار.

تعريف المصطلحات:

مؤشر صعوبة الفقرة : نسبة الاستجابات الصحيحة للفقرة من قبل مجموعة المفحوصين.

مؤشر تمييز الفقرة : معامل الارتباط المصحح بين اداء مجموعة المفحوصين على الفقرة وعلى الاختبار بعد حذف تلك الفقرة منه.

ثبات الاختبار : قيمة الثبات التي تم الحصول عليها باستخدام معادلة كودر رتشاردسون KR20

المعالجة الإحصائية:

تم سحب ٦ عينات عشوائية من عينة الدراسة الكلية والمؤلفه من (٣٢٠٠) طالبا وطالبة وبحجوم هي (١٦٠٠، ٨٠٠،٤٠٠،٢٠٠، ١٠٠، ٥٠) تم سحبها باستخدام البرنامج الإحصائي SPSS (ولذلك تعتبر هذه الدراسة من الدراسات الاسترجاعية حيث تم أخذ العينات الصغيرة بعد أن أنجزت الكبيرة)، بعد ذلك تم تحليل الاستجابات على فقرات الاختبار على مستوى كل عينه، حيث تم تقدير قيم مؤشري الصعوبة والتمييز لكل فقرة، ومعاملات الثبات، والعلامات المئينية المقابلة للعلامات الخام، كما واستخدمت الاحصائيات الوصفية وتحليل التباين لمعرفة الفروق في الخصائص للفقرات والاختبار في ضوء أحجام العينات الفرعية.

محددات الدراسة: تتحدد نتائج هذه الدراسة بأن العينات الفرعية المسحوبة هي عينات عشوائية

ولذلك فان نتائجها تتحدد في هذا الإطار، على عكس ما يميل له معظم الباحثين في استخدام العينات المتيسرة تقليلا للوقت والجهد والتكلفة. كما وتتحدد نتائج هذه الدراسة بخصائص الفقرات والاختبار والمقدَرة وفق النظرية الكلاسيكية في القياس.

نتائج الدراسة:

يتناول هذا الجزء عرضا لنتائج الدراسة التي تم التوصل إليها وسوف يتم عرضها تبعا لتسلسل أسئلتها:

للإجابة عن السؤال الأول والمتعلق باثر حجم العينة على مؤشر صعوبة الفقرات في الاختبار، يظهر جدول (١) قيم الوسط الحسابي والانحراف المعياري والمدى لقيم معاملات الصعوبة لفقرات الاختبار تبعاً لحجم العينة، حيث تشير القراءة الأولية للنتائج بأن هناك تقاربا في متوسطات معاملات الصعوبة المقدرة في تلك العينات، فقد بلغ المتوسط الحسابي لصعوبة

جدول (١): الإحصاءات الوصفية لقيم معاملات الصعوبة باختلاف حجم العينة

الإحصاء/ حجم العينة	٥٠	١٠٠	٢٠٠	٤٠٠	٨٠٠	١٦٠٠	٣٢٠٠
المتوسط الحسابي	.39	.39	.38	.39	.37	.38	.38
المدى	.80	.69	.75	.78	.74	.75	.75
الانحراف المعياري	.17	.16	.15	.15	.15	.15	.15

الفقرات ما بين القيمتين (٠,٣٩ _ ٠,٣٧). وبانحراف معياري تراوح بين (٠,١٧ _ ٠,١٥) . وبالنظر إلى متوسطات معاملات الصعوبة في الجدول وتوزيع قيمها حسب حجم العينات نجد أنه لا يوجد اي ترتيب نمطي أو تسلسلي للمتوسطات تبعا لحجم العينة الا أنه يمكن القول أنها متقاربة جدا بحيث أن الفرق بين أعلى متوسط وأدنى متوسط هو فقط(٠,٠٢)، وللكشف فيما إذا كانت هناك فروق بين هذه المتوسطات لمعاملات الصعوبة في الاختبارات السبعة، تم إجراء تحليل التباين الأحادي لقيم معاملات الصعوبة لفقرات الاختبار تبعاً لحجم العينة. والجدول (٢) يظهر نتائج هذا التحليل.

جدول (٢): نتائج تحليل التباين الاحادي لمتوسطات معاملات الصعوبة وفق حجم العينة

مصدر التباين	مجموع المربعات	درجات الحرية	متوسط المربعات	ف	الدلالة الاحصائية
بين العينات	.03	6	.005	.243	.94
داخل العينات	11.95	484	.025		
الكلي	11.98	490			

حيث تبين نتائج تحليل التباين الاحادي في هذا الجدول أن قيمة الدلالة الإحصائية هي ٠,٩٤ اي أنه لا توجد دلالة إحصائية على مستوى (α= ٠,٠٥) للفروق بين متوسطات صعوبة فقرات الاختبارات المسحوبة تبعا لحجم العينة، مما يعني عدم وجود أثر لحجم العينة في معاملات صعوبة الفقرات.

وللإجابة عن السؤال الثاني حول أثر حجم العينة على مؤشر التمييز للفقرة، تم إيجاد مؤشرات التمييز لفقرات الاختبار في كل عينة من عينات الدراسة وحساب المتوسط لهذه المؤشرات بالإضافة للمدى والانحراف المعياري لها والتي يظهرها جدول (٣).

جدول (٣): الإحصاءات الوصفية لمؤشرات التمييز للفقرات وفقا لحجم العينة

احصاء الفقرة / حجم العينة	٥٠	١٠٠	٢٠٠	٤٠٠	٨٠٠	١٦٠٠	٣٢٠٠
الوسط الحسابي لمؤشر تمييز الفقرات	.25	.28	.26	.31	.27	.29	.28
المدى	.75	.58	.50	.56	.48	.46	.48
الانحراف المعياري	.16	.13	.15	.12	.11	.11	.11

حيث يظهر الجدول أن قيم الوسط الحسابي لمؤشرات تمييز فقرات الاختبار قد تراوحت ما بين القيمتين(0.25_0.31). وبإنحراف معياري تراوح بين (0.11_0.16). وبالنظر إلى المتوسطات نجد أنه لا تظهر اي علاقة أو أي نمط بين حجم العينة ومتوسط تمييزها حيث نجد مثلا أن متوسط مؤشرات تمييز الفقرات في العينة ١٦٠٠ قد بلغ ٠,٢٩ وهو اعلى منه في العينة الكلية ٣٦٠٠ والذي بلغ فقط ٠,٢٨، وأن اعلى متوسط لمؤشر التمييز قد كان في العينة ٤٠٠ اذ بلغ ٠,٣١.

وللمقارنة بين متوسطات قيم هذه المؤشرات فقد تم تحويل قيمها إلى ما يقابلها من قيم (Zp) الفشرية، ومن ثم أجريت عملية تحليل التباين الاحادي لهذه القيم، والجدول رقم (٤) يظهر نتائج هذا التحليل.

جدول (٤):نتائج تحليل التباين لمتوسطات مؤشرات تمييز الفقرات تبعاً لحجم العينة

مصدر التباين	مجموع المربعات	درجات الحرية	متوسط المربعات	ف	الدلالة الاحصائية
بين المجموعات	٠,١٤	6	.024	١,٤٤.	.197
داخل المجموعات	٧,٩٧	٤٨٣	.017		
الكلي	٨,١٠	٤٨٩			

حيث يشير الجدول إلى أن قيمة الدلالة الإحصائية للفروق بين متوسطات مؤشرات تمييز الفقرات قد بلغت ٠,١٩٧ مما يعني عدم وجود فروق ذات دلالة إحصائية بينها على مستوى (α= ٠,٠٥) الأمر الذي يعني عدم تأثير حجم العينة على مؤشر تمييز الفقرة.

وفيما يتعلق بالسؤال الثالث والمتعلق بعدد الفقرات المحذوفة من الاختبار لدى كل من عينات الدراسة، فقد تم حساب مؤشري الصعوبة والتمييز لكل فقرة ومن ثم محاكمة هذه القيم للمؤشرات اعتمادا على القيم المقبولة لمعاملات الصعوبة ومعاملات التمييز لتلك الفقرات، حيث تم حذف الفقرة التي يخرج مؤشر صعوبتها عن المدى ٠,٣- ٠,٧ أو يخرج مؤشر تمييزها عن المدى (٠,٣- ١)، ويبين الجدول رقم (٥) عدد الفقرات المحذوفه لدى كل عينة من عينات الدراسة.

جدول (٥):عدد الفقرات المحذوفة من الاختبار وفقا لمؤشر الصعوبة ومؤشر التمييز

مؤشر الفقرة/ حجم العينة	٥٠	١٠٠	٢٠٠	٤٠٠	٨٠٠	١٦٠٠	٣٢٠٠
عدد الفقرات المحذوفة حسب مؤشر الصعوبه	٢٣	١٩	١٨	١٨	١٨	١٨	١٩
عدد الفقرات المحذوفة حسب مؤشر التمييز	٣٢	١٦	٢١	١٢	١٣	١٣	١٥

ويظهر من الجدول أن عدد الفقرات المحذوفه حسب مؤشر الصعوبة قد بلغ ٢٣ فقرة في العينة التي حجمها ٥٠ مما يعني حذف ما نسبته ٢٩٪ من عدد فقرات الاختبار الأصلي، فيما تقاربت عدد الفقرات المحذوفة في باقي العينات فكان حوالي ١٩ فقرة، كما ويشر الجدول كذلك إلى أن عدد الفقرات المحذوفة وفق مؤشر التمييز قد كان الأكبر في العينة التي حجمها ٥٠ أيضا اذ بلغ عدد الفقرات المحذوفة ٣٢ فقرة اي ما نسبته ٤٠٪ من عدد فقرات الاختبار الأصلي، وهو تقريبا ضعفي عدد الفقرات المحذوفة في العينة التي حجمها ١٠٠، مما يعني أن تطبيق الاختبار على عينة من ٥٠ هو تهديد مباشر للصدق الداخلي للاختبار والمتمثل بصدق محتوى الاختبار نتيجة لحذف هذه النسبة العالية من الفقرات والتي وصلت الى ٤٠٪.

أما فيما يتعلق بإجابة السؤال الرابع حول اختلاف قيم معاملات الثبات للاختبارات وفق حجم العينة، فقد تم تقدير معامل الثبات باستخدام معامل الثبات كوردر ريتشاردشون KR20 كما هو مبين في الجدول رقم (٦) أدناه.

جدول (٦): قيم معاملات الثبات KR20 للاختبارات وفقا لحجم العينة

معامل الثبات/ حجم العينة	٥٠	١٠٠	٢٠٠	٤٠٠	٨٠٠	١٦٠٠	٣٢٠٠
KR20	.85	.87	.86	.89	.87	.88	.87

ويتضح من الجدول أن قيم الثبات المقدرة متقاربة جدا بفارق بسيط بلغ ٠,٠٤ وقد سجلت العينة ٥٠ الثبات الأدنى من بينها حيث وصلت قيمة الثبات فيها إلى ٠,٨٥، كما ويتضح من الجدول أنه من الصعب تحديد اتجاه في اختلاف تلك القيم أو إظهار علاقة نمطية مع حجم العينة. ولكن يبين هذا الجدول أن أقل قيمة لثبات الاختبار كانت لدى العينة الأصغر حجماً (٥٠) مفحوصا ، وأعلاها لدى العينة ذات الحجم (٤٠٠) وهي ليست العينة الأكبر .

وفيما يتعلق بالإجابة عن السؤال الخامس حول معادلة درجات الطلبة على الاختبارات، يشير الجدول (٧) أدناه إلى الإحصاءات الوصفية لمستويات أداء الطلبة على الاختبارات حيث يتضح من الجدول أن أعلى متوسط أداء كان للطلبة ضمن العينة الصغيرة ٤٠٠ حيث بلغ المتوسط ٢٧,٩، كما وكان أدنى متوسط للعينة ٨٠٠ حيث بلغ ٢٦,٣ ، كما ويشير الجدول إلى انخفاض الانحراف المعياري في العينة ٥٠ نسبة إلى باقي العينات حيث بلغ ٥٠ في حين كان في باقي العينات ٦٠ فاكثر.

جدول ٧ : الاحصائيات الوصفية للعلامة الكلية للطلبة على الاختبار تبعاً لحجم العينة.

الاحصائيات الوصفية/حجم العينة	٥٠	١٠٠	٢٠٠	٤٠٠	٨٠٠	١٦٠٠	٣٢٠٠
المتوسط الحسابي	٢٧,٤	27.25	26.9	27.9	26.3	26.9	26.8
المدى	٩,٧	10.5	10.1	11.4	10.2	10.7	10.5
الانحراف المعياري	٥٠	٦٠	61	65	65	65	65

كما ويشير الجدول ٨ إلى معادلة الدرجات في الاختبارات الستة وفقا للعلامة المئينية المقابلة للعلامة الخام في الاختبار حيث يمكن من الجدول إيجاد العلامة المقابلة بين أي زوجين من الاختبارات ويظهر من الجدول أن الاختلافات في الرتب المئينية برزت في العينة ٥٠ بشكل ملفت للنظر.

جدول٨ :معادلة الدرجات لدى عينات الدراسة إعتماداً على القيم المئينية.

المئين/ حجم العينة	٥٠	100	200	400	800	1600	3200
٥	١٩	14	14	14	14	14	14
١٠	١٧	17	16	16	١٦	16	16
٢٥	19.7	20	20	19	19	19	19
٥٠	25	25	25	25	24	24	24
٧٥	33	34	33	35	32	33	33
٩٠	43	43	41	43	41	43	42
٩٥	47.4	47	47	50	48	48	48

مناقشة النتائج:

لقد أشارت النتائج في إطارها العام إلى عدم وجود فروق جوهرية في خصائص الفقرات وخصائص الاختبار بشكل عام تبعا لعينة المفحوصين التي يتم اختيارها بطريقة عشوائية عن خصائص المجتمع الذي سحبت منه، فلم تختلف خصائص الفقرات في العينات عنها في المجتمع المحسوبة منه، كما لم تختلف خصائص الاختبار في العينات عنها في المجتمع المسحوبة منه مما يعني أن العينات العشوائية قد تتمتع بدرجة من الدقة تصل في تمثيلها لمجتمع الدراسة إلى ١٠٠٪ وهذا يتفق مع ما أشار له بنتي(Pentti,2007) حيث أوضح أن العينة العشوائية تخلو من التحيز المنتظم وهي ممثلة للمجتمع، وهذا يعني بأن النتائج التي نحصل عليها من العينة العشوائية مطابقة للمجتمع تماما، كما وأشار إلى أن العينة صغيرة الحجم والمختارة بطريقة عشوائية يمكن أن تكون أكثر تمثيلا من عينة كبيرة وهو ما اتفق مع نتائج هذه الدراسة، حيث أشارت النتائج هنا الى أن بعض العينات الصغيرة قد قدمت مؤشرات اقرب إلى المجتمع من العينات الكبيرة في بعض المؤشرات مثل التمييز والثبات حيث طابقت العينة ١٠٠ مؤشراتها لمؤشرات المجتمع أكثر من العينة ١٦٠٠ مثلا. كما وتتفق النتائج مع دراسة ميشل (Micheal, 2010) والتي توكد أن حجم العينة قد لايكفل دقة تمثيلها للمجتمع الا اذا كانت العينة عشوائية، اذ أنها العينة العشوائية ممثلة للمجتمع بدرجة عالية باستمرار سواء كانت كبيرة أو صغيرة، ومع ذلك فانه يجب أن نشير هنا إلى أن العينات الكبيرة كانت الأكثر محافظة على الفقرات من عملية الحذف نسبة إلى العينات الصغيرة التي ارتفع فيها اعداد الفقرات المحذوفة، فقد بلغ عدد الفقرات المحذوفة وفق مؤشر التمييز مثلا ١٣ فقرة (١٦٪) في العينة ١٦٠٠ بينما وصل في العينة ٥٠ إلى ٣٢ فقرة (٤٠٪) وهو أمر ملفت للنظر مما يشير بشكل واضح إلى أن تطبيق أوتجريب الفقرات على عينات اقل من ١٠٠ من الممكن أن يشكل خطرا على

صدق محتوى الاختبار نتيجة للفقرات التي سيتم حذفها وهذا يتفق مع ما يشير له كروكر والجينا (Krocker&ALgina,1986) ويكسدوكول (Yuxducol ،٢٠٠٨) وانستازي (Anastasi, 1988) وجروش (Gorsuch 1983) وكلاين (Kline ,1979)، كما ويتفق مع دراسة فابرجار وزملاؤه (Fabrigar, Wegener, MacCallum, and Strahan ,1999) بضرورة تكبير عينة التجريب وأن لا يقل حجم عينة التقنين عن ١٠٠ بعكس البعض الآخر الذي يوصي بإمكانية التجريب على عينات صغيرة تصل إلى ٥٠ أو اقل . وبهذا فان توجه بعض الباحثين إلى تخفيض حجم العينة إلى أقل من الحدود الموصى بها قد يؤدي إلى نتائج غير دقيقة في تحري الخصائص السيكومترية سواء على مستوى الفقرات أو على مستوى الاختبار ككل، كما تجدر الأشارة هنا إلى أنه عندما اتضح في هذه الدراسة بأن الحجم الأقل الذي يمكن التعامل معه هو ١٠٠ مفحوص وليس ٥٠ كما أوصت به بعض الدراسات الأخرى، فإننا نقصد بذلك تحديدا أن هذه العينة يجب أن تكون عينة عشوائية تامة وليس عينة متيسرة أو مقصودة أو غيرها من العينات غير الاحتمالية الاخرى كما يفعل معظم الباحثين .وعليه فانه يمكن القول أنه إذا كان الحد الأدنى المطلوب لتحري خصائص الاختبار وخصائص فقراته في عينات عشوائية تامة لا يقل عن ١٠٠ فإنه من الضروري جدا أن يتضاعف هذا الرقم في حالة العينات غير الاحتمالية.

وعليه توصي الدراسة بـ:

١. استخدام برنامج ال SPSS في سحب العينات العشوائية لأنه يوفر عينات مطابقة تماما في خصائصها لمجتمع الدراسة.

٢. ضرورة أن تزيد عينة التجريب في الاختبارات التحصيلية عن ١٠٠ مفحوص لضمان عدم حذف عدد كبير من الفقرات وبالتالي المحافظة على صدق محتوى الاختبار.

٣. إجراء دراسات مماثلة باختيار عينات غير احتمالية من حجوم مختلفة للكشف عن أثرها في خصائص الاختبار وخصائص فقراته.

المصادر والمراجع:

- الكبيسي، كامل (١٩٩٥) ، أثر اختلاف حجم العينة والمجتمع الاحصائي في القدرة التمييزية لفقرات المقاييس النفسية : دراسة تجريبية . جامعة بغداد ، كلية التربية.
- الياحي، امل (٢٠٠٥) أثر حجم العينة وطول الاختبار في الخصائص السيكومترية للمقاييس النفسية . اطروحة دكتوراه غير منشورة . جامعة بغداد.

المراجع الأجنبية:

- *Anastasi,A.(1986).Psycological testing (3rd ED) .New York .Macmillan.*
- *Barrett, P. T., & Kline. P. (1981). The observation to variable ratio in factor analysis. Personality Study in Group Behavior, 1, 23-33.*
- *Bryant, F. B., & Yarnold, P. R. (1995). Principal components analysis and exploratory and confirmatory factor analysis. In L. G. Grimm & R R. Yarnold (Eds.), Reading and understanding multivariale statistics (pp. 99-136). Washington, DC: American Psychological Association.*
- *Cattell, R. B. (1978). The Scientific Use of Factor Analysis. New York: Plenum.*
- *Charter,R.A.(1999). Study Samples Are Too Small to Produce Sufficiently Precise Reliability Coefficient .The Journal of General Psychology, 130,117-129*
- *Comrey, A. L., & Lee, H. B. (1992). A first Course in Factor Analysis. Hillsdale, NJ: Erlbaum.*
- *Crocker & Algina , (1986).Introduction to Classical and Modern Test Theory. New York .holt Rinehart and Winston.*
- *Datlon , D & Cosier ,R. (1988). Psychometric Properties of the Organization Citizenship Behavior Scale. Educational and Psychological Measurement*
- *Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. Psychological Methods, 4, 272-299.*
- *Ferguson , A (1989). Statistical Analysis in Psychology and Education, N .Y , Me Craw – Hill.*
- *Guilford, J. P. (1954). Psychometric methods (2nd ed.). New York: McGraw-Hill.*
- *Hair, J. F. J., Anderson, R. E., Tatham, R. L., & Black,W. C. (1995). Multivariate data analysis (4th ed.). Saddle River, NJ: Prentice Hall.*
- *Heniysoon, S .(1971) . Correction of Item – Total Correlation In Item Analysis , Psychometric . V28 ,N03 .*

- *Henson, R. K., & Roberts, J. K. (2006). Use of exploratory factor analysis in published research: Common errors and some comment on improved practice. Educational and Psychological Measurement, 66, 393-416.*
- *Hutcheson, G., & Sofroniou, N. (1999). The multivariate social scientist: Introductory statistics using generalized linear models. Thousand Oaks, CA: Sage Publications.*
- *Javali S B, Gudaganavar N .(2011). Effect Of Varying Sample Size In Estimation Of Reliability Coefficients Of Internal Consistency. Article URL: http://www.webmedcentral.com/article_view/1572*
- *Kline, P. (1979). Psychometrics and psychology. London: Acaderric Press.*
- *Kline,p .(1989).A handbook of test construction. Introduction to psychometric design .New York. Methune & company .*
- *MacCallum, R. C., Widaman, K. F., Zhang, S., & Hong S. (1999). Sample size in factor analysis. Psychological Methods, 4, 84-99.*
- *Maloney , P&Word , P. (1980).Psychological Assessment A conceptual Approach , N.Y , Oxford University Press.*
- *Millman.J, Green.J.(1989).The Specification and Development of Test Achievement and Ability .Edited by: Robert. L.Linn.*
- *Noruis, M. J. (2005). SPSS 13.0 Statistical Procedures Companion. Chicago: SPSS, Inc.*
- *Nunnally, J. C. (1978). Psychometric theory (2nd Ed.). New York: McGraw-Hill.*
- *Nunnally.J.(I967).Psychometric theory .New York: MicGraw –Hill.*
- *Segall.D.O.(1994). The Reliability of linearly equated tests. Psychometrika ,59,361- 375.*
- *Yurducal , H.(2008).Minimum Sample Size for Cronbach 's Cofficient Alpha .Journal of Education.V35:397-405.*