Journal for Re Attach Therapy and Developmental Diversities
eISSN: 2589-7799
2023 August; 6 (9s2): 1414-1428

Clinical Significance in Psychological and Educational Research: Does it Matter?

Yahya Nassar[1]
[1] Cognitive Sciences department, United Arab Emirates University
mohamedrasik007@gmail.com

Received: 19- June -2023
Revised: 02- July -2023
Accepted: 10-August-2023

Abstract

The main aim of this study was to confirm the importance of the concept of clinical significance. Also, this study aimed at discriminating among the concepts statistical, practical and clinical significance. In addition, this study attempted to present the most known methods to estimate the clinical significance of psychological and educational studies. These methods were Jacobson-Truax method (JT), GulliksenLord - Novick method (GLN), Edwards-Nunnally method (EN), Hageman-Arrindell method (HA), and Hierarchical Linear Method (HLM). The current study constraints on the Jacobson-Truax method (JT) and via using hypothetical data, psychological and educational examples were presented to explain how to implement JT method. Moreover, via applying JT method this study attempted to introduce a suggested model to investigate the clinical significance of the treatment on the differences among the groups of the experimental designs. Finally, the results of this study confirmed the value of investigating the clinical significance of the psychological and educational studies because the results could be statistically and practically significant although could not indicate to clinical significance either on the individual or groups level.

Keywords: Clinical significance, statistical significance, practical significance, Jacobson-Truax method.

1. Introduction

Researchers in the field of psychological and educational studies are concerned mainly with extracting the statistical significance of the results of their research, as there is a prevalent idea among researchers in the field of psychological and educational studies that good results are the results that achieve statistical significance and on the contrary, the results are not of value if they are not significant Statistically. The concept of statistical significance indicates that the role of the chance factor in the differences between the averages of the experimental and control groups is less than the allowable limit by the researcher (Tuckman,2005). Studies indicate that the concept or strategy of statistical significance may be misleading in many cases, as the results can be statistically significant as a result of enlarging or amplifying the sample size, but it is also possible to convert non-statistically significant data into a statistical function by increasing the sample size (Nassar,2006). Also, the results of some studies may be statistically significant but not practical in terms of practice (Al-Sayyad, 1989, Nassar,2019) herein began to the talk about another concept in psychological and educational research, which is the practical importance of the results of the so-called effect size (Audi and Khalili, 1988). The concept of effect size essentially indicates that the results of statistical significance are not necessarily of practical significance, as the practical significance of the results of psychological and educational research is examined through certain statistical methods such as "Cohen's D" in the case of dealing with two independent samples or ETA square test if they include The experimental study has two or more means, that is, there are one or more statistical methods that can be used to study the practical significance of the results in psychological and educational research so that they are complementary to those results obtained from statistical tests that study the statistical importance of data in psychological and educational studies (Nassar,2006).

The main topic of the current study is the concept of clinical significance that concept or technique which has been neglected by most of the researchers across the world, especially those who use experimental designs. The main idea of clinical significance is that: the researchers should be aware of the extent to which the used programs or treatments in their research are effective in terms of changing the reality of the case of group members or experimental groups from case to case from the current situation to a better condition either at the individual or group level (Campbell, 2005). For example, conducting an experimental study that aims primarily to reduce the

level of test anxiety for twelfth-grade students in Al Ain through a psychological group counseling program that is applied to members of the experimental group while the control group is not exposed to that program, is an example of studies used for experimental designs in The field of psychological and educational counseling, In such studies, the researcher might examine the statistical significance of the differences between the experimental and control groups at best, he/she may examine the practical significance of the results. However, the question posed here is whether the counseling program was effective, so that the level of test anxiety decreased for the students who were exposed to the program and that they switched from students who suffer from test anxiety to normal students who face achievement tests with confidence and without any unusual symptoms of anxiety. In my opinion, this is the primary goal of studies based on extension or training programs. The clinical significance of the results according to the aforementioned concept is important for understanding and interpreting the results of psychological and educational research. Rather, it may be considered a criterion for Assessing the progress, improvement, or regression that occurred to the individuals of the study sample after applying the treatment (Lambert et al., 2001).

The importance of the concept of clinical significance emerges through its interest in the change that occurred at the individual level thus determining the extent to which each of these individuals benefited from the treatment used in the study. Hence, the clinical significance is concerned with exploring or testing the benefit that can be achieved through treatments used in experimental studies in more useful ways than traditional methods that rely on examining the statistical and practical significance of the research results (Hansen, Lambert, & Forman, 2002).

2. Research Problem

The primary goal of the current study is to present the concept of the clinical significance of the results in psychological and educational studies used for experimental designs and to distinguish it from concepts of statistical significance and practical significance, then review the most important methods used in estimating or calculating the clinical significance of the results of experimental research.

3. Research Questions

The current study tried to answer the following questions:

1- What is the concept of the clinical significance of experimental psychological and educational research results?

2- What is the difference between statistical significance and clinical significance?

3- Is the clinical significance the same as the practical significance of the results, or is there a difference between them?

4- What are the methods or procedures by which to verify the clinical significance of experimental psychological and educational research results?

5- Can the statistical significance be used to examine the clinical significance of the differences between experimental and control groups?

Significance of the study

Considering the primary goal of any statistical analysis is to discover whether there is the statistical significance of the results is a problem in itself. Because, despite the importance of obtaining statistically significant results, that may not be sufficient (Thompson, 2006). As some researchers have pointed out, the concepts of "statistical significance" and "practical significance" have overlapped and were often used incorrectly (Hubbar and Ryan, 2000; Huberty, 2002). In this context, Al-Tubaity's study (2008) indicates that most experimental studies focus on the level of significance to make a decision about accepting or rejecting the null hypothesis without paying attention to using the practical significance of the results. A third method has recently emerged in human studies to interpret research findings, relying mainly on so-called clinical significance (Campbell, 2005). This concept was used to study whether the results obtained in empirical studies are really important and valuable in relation to the topic or phenomenon.

Essentially, the importance of the current study is concentrated as it attempts to distinguish between three types of indices that may be used in psychological and educational studies, especially in the context of experimental designs. That is: statistical significance, practical significance, and clinical significance. The current study also seeks to provide specific statistical methods to examine the clinical significance at the individual level which is the basic difference that distinguishing that method from the statistical and practical significance. as the current study attempted to present or suggest a model or method through which the statistical significance is used to examine the clinical significance between experimental and control groups. probably the primary importance of the current study lies in providing researchers, specifically in the fields of psychology and education, with a new approach (I.e. clinical significance) to interpret the results of their experimental research. It is worth noting in this context that this concept appeared in 1984 and has become one of the basic elements that some specialized psychological and educational journals expect to be considered in interpreting experimental research results whose primary goal is treatment or improvement. (Kendall, 1999; Kendall, Marrs-Garcia, Nath, & Sheldrik, 1999).

Definitions of the concepts:

Clinical importance: The amount of change in the behavior of an individual or group resulting from the effect of treatment, where the results are clinically important when the individual or group shifts from one case to another and is estimated either at the individual or group level through statistical methods, some of which will be presented in the current study.

Statistical Significance: A statistical method used to examine the role or the probability of the chance factor in the relationship between the variables or in the differences between the means, where the results are statistically significant if the actual probability of the chance factor is less than or equal to the highest permissible limit of the chance factor in the relationship. It is determined by the researcher and included in the null and alternative hypotheses, Called the expected level of significance.

Expected Practical Significance or Expected effect Size: The strength of the expected relationship between study variables, which can be estimated through a review of the previous literature.

Actual Practical significance or actual effect size: The amount of the strength of the actual relationship between study variables, which can be estimated through the process of analyzing study data and by using specific statistical methods such as Cohen "D" and Eta square and others (Nassar, 2006).

An analytical study: It is a study based on the analysis of a psychological, social, or natural phenomenon, or a scientific concept, in terms of its definition and clarification of its relationship to the concepts associated with it in different ways, including providing scientific examples in which virtual data is used and analyzed. This method was used in the current study.

4. Methodology

To achieve the objectives of the current study and to answer its research questions, an analytical descriptive approach was used where the concept of clinical significance was reviewed or clarified by distinguishing between this concept on the one hand and the concepts of statistical and practical significance on the other hand. Also, in the current study, an attempt was made to present some of the most used statistical methods in psychological and educational literature to explore the clinical significance of the experimental studies. In this study, hypothetical psychological and educational research examples have been presented in order to clarify the concept of clinical significance and explain how it is estimated, calculated and interpreted, especially for researchers interested in using non-traditional methods to interpret the results of their experimental research. Also, a proposed model was presented that aims to examine the clinical significance of experimental treatments at the group level, that is, to determine the clinical significance of these treatments on experimental and control groups. This model is based on the use of statistical significance, but after classifying the members of the experimental and control groups according to the criteria used in one of the most famous methods used in estimating the clinical significance, which is the Jacobson & Truax method (1991). Then, Chi-square for crosstabulation was used to check whether there is a significant relationship between the variables group membership (Experimental - control) and the classification of the subjects of the sample according to the aforementioned method.

Statistical significance

The concept of statistical significance appeared about three hundred years ago (Thomopson, 2002). However, it was used more intensively in the early twentieth century, especially when using statistical tests such as Chi_ Square, T-test, and ANOVA test. The use of the method of testing the statistical significance of null hypotheses has increased in recent times, and the aim was to make a desion whether the results are statistically significant or not. This method is based on formulating a null hypothesis about the relationship between specific variables within a specific research population. That hypothesis is tested through sample data that is supposed to have similar characteristics with the population from which it was drawn. The concept of the actual level of significance is used to test whether the relationship between the target variables or the difference between the means of the experimental and control groups is statistically significant or not. (Thomopson,1998). The problem with using the statistical significance is that it does not indicate the actual significance of the results. Hence, researchers cannot determine the degree of the practical significance of the results, as it only tells whether the results obtained are due to the chance factor or not. In light of this, the researcher can adopt a position toward the alternative hypothesis that represents his point of view. It is statistically known that this type of significance is directly affected by the size of the sample, meaning that results that are not statistically significant can become statistically significant by increasing the sample size without any change in the characteristics of the statistical distribution of the original data. Consequently, it can be said that the statistical significance is the function of the sample size (Nassar, 2006).

The statistical significance of the research results also depends on the expected level of significance, which is determined by the researcher. Unfortunately, to get a significant result, it's likely that some researchers may change that value based on the actual probability of chance in their results (Sig or P-value). Thus, it can be said that reliance on statistical significance alone in the process of interpreting research results is not sufficient - or even that it may sometimes be misleading.

Practical significance

This concept indicates to explore the strength of the relationship between the variables by using statistical methods less affected by the sample size than those used to examine the null hypotheses. Basically, it is possible to obtain statistically significant results even with no significant differences between the statistical means due to the use of relatively large samples. The statistical explanation for such a case is that as the value of the Standard Error decreases when the sample size increases, the statistical power of the used statistical procedure to examine the null hypothesis increases as well. In other words, a difference of 5, 10, 20, or 30 between the statistical means of the experimental and control groups may be sufficient to obtain statistically significant results even though the practical significance of these differences is not necessarily at the same level (Nassar, 2006). Therefore, there are many educational and psychological journals that do not accept the results of studies that are devoid of reference to practical significance.

The effect size is one of the most famous methods that are used to examine practical significance. As the results might be statistically significant, but not practically significant (Thompson 2006). There are several methods for extracting the value of the effect size. Some of which depend on examining the differences between the averages (such as Cohen's “D”). Other methods use the ratio of the explained variance in the analysis of variance (like eta squared). The primary goal of all methods is to answer the following question: Did the treatment make a practical difference between experimental and control groups, and how much difference did they make? The method of effect size is useful because it is concerned with determining the importance of results at the practical level, and thus exceeds the concept of statistical significance, which is concerned only with whether the results are due to chance or not. As examining the effect size of the statistical results is an essential process in psychological and educational studies the American Psychological Association was considered that procedure as one of the basic elements that must be considered by psychological journals to accept or to reject the publication of these studies (Wilkinson & APA Task Force on statistical Inference, 1999).

Clinical significance

Most of the methods used to estimate the magnitude of the effect size in psychological and educational studies depend on the difference between the experimental and control groups without being concerned with that change

that occurs at the individual level. There is a recent trend in psychological and educational research that indicates the necessity of measuring and observing the level of the effect of experimental treatment on each member of the experimental group instead of dealing with the arithmetic averages that deal with them as a group. The aim of experimental studies in psychological and educational research is to examine whether treatment helps people to change for the better, whether it is in the field of mental disorders, learning difficulties, or even in the field of diagnosis and treatment. The examination of the effect of treatment or its effectiveness is not considered important unless it has brought about the expected and required change. For example, the aim of using counseling or treatment programs to deal with depression disorder in a group of individuals is to change their case from the state of depression into a state of no depression. Likewise, it can be said that the effectiveness of a program aiming at improving children's reading skills is measured by its ability to transfer these children from the weak reading level to a level that is compatible with their peers of the same level, hence they can be reintegrated and returned to regular programs.

The methods used in examining the clinical significance of the results of experimental studies attempt to examine the degree of effectiveness of the treatment used to effect the required change for the target group of individuals (Campbell, 2005). "Jacobson, Follette, and Revenstorf (1984)" were the first to use a specific method to examine the clinical significance of the results. The reason that these researchers have been interested in the concept of clinical significance is due to their feeling that there is some deficiency in the field of studies that are interested in the field of psychotherapy since the field of interest here is the individual and not the group as is the case in statistical and practical significance.

These researchers (Jacobson, Follette, and Revenstorf, 1984) have indicated that the difference between the averages does not give real information about the number of clients who have moved from the level of disorder to the normal level. The method that has been used by those researchers has become essential in the area of the subject of clinical significance for psychological and educational studies. Since that time, many other methods have emerged, which were proposed by many researchers to examine the clinical significance. So that the concept of clinical significance has become one of the main indicators in determining the importance of results not only at the level of groups but also at the level of individuals. In addition, this concept has made a qualitative shift in the field of psychological and educational research that exceeded the statistical and practical significance (effect size), which are no longer sufficient to indicate the importance of the results in experimental studies These researchers (Jacobson, Follette, and Revenstorf, 1984) have indicated that the difference between the averages does not give real information about the number of clients who have moved from the level of disorder to the normal level. The method that has been used by those researchers has become essential in the area of the subject of clinical significance for psychological and educational studies. Since that time, many other methods have emerged, which were proposed by many researchers to examine the clinical significance. So that the concept of clinical significance has become one of the main indicators in determining the importance of results not only at the level of groups but also at the level of individuals. In addition, this concept has made a qualitative shift in the field of psychological and educational research that exceeded the statistical and practical significance (effect size), which are no longer sufficient to indicate the importance of the results in experimental studies. Nevertheless, the concept of clinical significance is often overlapped with the concept of practical significance of results, where both types are often confused (Peterson, 2008)., in fact, the clinical significance is completely different from the practical significance. The following two examples illustrate the difference between these two types of significance in the context of psychological and educational studies.

Psychological example

Suppose you are a school psychologist and asked to deal with a child who suffers from depression symptoms. So, Via searching the literature you found two studies each of which suggested a specific treatment method to cope with the problem of depression in children. Both studies tested the following Null hypothesis: There are no statistically significant differences at 0.05 level of significance between the average degree of depression on the post-test scores of the experimental and control groups. In addition, suppose that the results of each of the studies indicated the possibility of rejecting that null hypothesis, where the value of the actual level of significance was less or equal (0.05). In this case, it can be concluded that both methods used to treat the depression problem in

children were effective. Let us also assume that the researchers in the two studies were also interested in examining the practical significance of the results. To achieve this purpose, the effect size was calculated using Cohen'S d equation. Suppose that the magnitude of the effect size in both studies was equal to (0.9). Referring to the criteria referred to by Cohen, this value can be described as very large, indicating a practical significance of the results in both studies. Although most psychological and educational studies refer at best to the statistical and practical implications of the results. Unfortunately, psychological and educational studies that indicated clinical significance in their results are very rare. Referring to the previous example, it is noted that the results were statistically and practically significant, so, does this mean that the researcher can use either of the two methods as they achieve the same results? Before answering this question, let us assume that the researchers in those studies not only pointed to the statistical significance and practical significance, but they also pointed to the clinical significance in their results, in this case, we can find differences between the results of both studies. Despite the statistical and practical significant differences between the means of the experimental and control groups on the post-test in the first study, they still suffer from the depression problem according to their scores on Beck's depression inventory (the well- known depression scale). Therefore, they still need treatment in order to reach a state of no depression. On the other hand, if the clinical significance results in the second study indicated that 80% of the subjects' post-test scores on the depression scale have been sufficiently decreased so that they moved from depressed case to normal case. Consequently, they no longer need any treatment. Based on the clinical significance results indicated in both studies, the psychologist may choose the second method because it is more clinically effective than the method used in the first study.

Educational Example

Suppose a principal of a primary school has noticed that some fourth-grade students suffer from poor reading skills. A test for all fourth-grade students to measure their reading skills was applied. As the degree of cutting was determined by 50 words per minute as a criterion for determining whether the child is fluent or not fluent in reading. As a result, twenty children were identified with a reading score of less than 50, and need to exposed to a special training program to improve their reading skills. The school's director instructed the English language teacher to choose ten children from a group of children with poor reading so that they are distributed into two groups: an experimental group and a control group. So that the Experimental group is subject to an intensive training program for a period of four weeks aimed at improving their reading skills. While the control group was not exposed to that program. a valid and reliable reading scale was implemented before and after exposing the experimental group to the training program (pre and post-test). The number of words that children of both groups can read before and after the implementation of the program were measured. Table 1 indicates the number of words that members of both groups were able to read correctly in a minute.

Table 1: The number of words that children were able to read correctly during a minute before and after the program

The name	The group	Before the program	After the program
Mohammed	Experimental	40	60
Omar	Experimental	25	47
Hamza	Experimental	45	65
Sawsan	Experimental	28	46
Safaa	Experimental	27	32
Ali	Control	20	24
Qusay	Control	40	46
Zain	Control	18	23

The name	The group	Before the program	After the program
Peaceful	Control	24	20
Fatema	Control	30	30

Using the "t" test for independent samples to examine the significance of the difference between the average scores of the two groups on the post-test scale at 0.05 as a level of significance. The results of this analysis indicated that there are statistically significant differences between the average number of words between the two groups. The descriptive statistics results revealed that the average of number of words the experimental group of children can read correctly is higher than the average of the number of correct words in the control group. Table 2 shows the mean and standard deviations for the number of words that the experimental and control groups were able to read within a minute after the program. Also, this table includes the results of the "t" test to examine the significance of the differences between the two means.

Table 2: A number of words that members of the experimental and control groups were able to read within a minute after the program, and the results of the "t" test to examine the significance of the differences between the two means.

the group	the number	The statistical mean	standard deviation	"t" value	Degrees of freedom	Level of significance
Experimental	5	50	12.98	2.88	8	0.021
Control	5	28.6	10.38			

The results of "t" test indicate that there are statistically significant differences at 0.05 level of significance between the experimental and control groups. However, these results are not sufficient to determine to what extent the subjects in the experimental group got benefits at the individual level, not at the group level, from the program. Therefore, the inferential statistical results may not be enough to make the necessary decisions for improving the reading skill of these individuals. By using practical significance testing methods such as Cohen's "D" test, we obtain an effect size of 1.64 which indicates, according to Stevens (1996), a large effect size. The aforementioned value of the effect size indicates that the training program has not made a statistically significant difference between the experimental and control groups only, yet it was practically significant as well. The importance of the effect size index lies in evaluating the amount of the difference, as it is, in this example, a large and clear difference.

In order to get better understanding level of the effectiveness of the program, especially in terms of its impact on each individual of the experimental group, it is necessary to estimate the clinical significance. Taking into account that the cut-off point that distinguishes between children who are able to read and who cannot read is 50 words per minute, and by referring to the data provided in Table 1, it is clear that two children namely Muhammad and Hamza form the experimental group did not only improve reading but rather they moved from the category of children who are not able to read to the category of children who are reading. While we find that the rest of the children of the experimental group, namely Omar, Sawsan, and Safa have improved their level of reading, but they remained below the indicated cut point, in other words, they did not move to the category of students who are master reading.

It is also noticed through the data in Table 1 that the members of the control group have also achieved an improvement in the reading level, so how can this be explained in the absence of their exposure to the training program? The reason for the increase in the average scores of the individuals of the control group may be due to the regression factor towards the mean. As it is statistically known when dealing with extreme samples on an attribute, it might be observed that the average of their post-test scores regresses toward the mean of the entire sample, this is known statistically as the regression towards the mean.

Methods for estimating clinical significance in psychological and educational studies:

There are several methods for estimating clinical significance. In the current study, the five most important methods used in psychological and educational studies were reviewed to estimate this significance. These methods are the Jacobson-Truax (JT) method, the Gulliksen-Lord-Novick (GLN) method, the Edwards-Nunnally method (EN), and the Hageman-method. Arrindell (HA), and finally the Hierarchical Linear Method (HLM) (Bauer, Lambert, & Nielsen, 2004).

In the current study, the focus was only on the Jacobson-Truax (JT) method for two reasons. First, because this method is the most used in experimental psychological and educational studies, namely those studies whose primary purpose is to bring about change at the individual level as a result of exposure to a specific treatment, and secondly: because this method - despite its simplicity - its indicate an average estimate of the effect of the treatment in the sense that it does not underestimates or overestimates at the same time the effect of that treatment (Bauer et al., 2004). On the other hand, the current study seeks to provide a brief presentation of the other methods mentioned previously to estimate the clinical significance of the experimental results. The main aim is to provide a number of methods that can be used by researchers, especially as this concept, although it has appeared since in 1984, however, it was not mentioned adequately in the psychological and educational studies that were carried out since that date.

Jacobson-Truax (JT) method: This method which had been developed by Jacbson, Follett, & Revenstrof in 1984 is the primary method for estimating clinical significance in psychological and educational studies. This method was revised by Jacobson and Truax in 1991(Jacobson and Truax,1991) and was known as the JT Method. This method is based on two basic steps, where a degree called the cutoff is initially determined, which is a degree in the scale that is supposed to be able to distinguish between individuals who need treatment (Dysfunctional group or clinical group) and ordinary individuals who do not need the same treatment (functional group or nonclinical group). The second step, the so-called Reliability Chang Index is estimated and it is denoted by the symbol (RCI). In light of these two criteria, the individuals of the study sample are classified into four categories: the category of the treated individuals (Recovered group), The category of individuals who have improved (Improved group), the category of individuals who have not changed (Unchanged group), and the category of individuals who have experienced a kind of deterioration or retreat on the measured feature (Deteriorated group).

The techniques of estimating the cutoff point in this method, there are three ways for achieving this according to what Peterson (2008) indicates. These techniques are referred to in the literature by what is called method A, method B and method C. According to method A, the cutoff score can be determined by extracting that degree which deviates by two standard degrees from the mean of the pre-test scores of the population of individuals who still need treatment (Bauer, Lambert, & Nielsen, 2004).

A = The mean of pretest scores (for the population of individuals who need treatment) + 2 × the standard deviation of their pretest scores

In other words, the cutoff point A can be extracted by calculating the arithmetic mean and the standard deviation of the pre-test scores for the population of individuals who need treatment estimated through the sample. Then the mean is summed to the doubled value of the standard deviation of the same population.

Peterson (2008) indicates that the cutoff score B is a score in post-test scores that must fall within the range of the population represented by the functional group (which does not need treatment). The cutoff point B can be calculated from the following formula:

B = Average scores for an ordinary population (no treatment needed) + 2 × standard deviation for same population

or Cutoff B = $M_{nonclinical} + 2\ SD_{nonclinical}$

It is clear that method B assumes the availability of information about the distribution of scores of ordinary individuals on the attribute under study. In the absence of such information, this method cannot be used to determine the cutoff point. However, it is relatively easy for most clients to override this point due to the fact that in many cases there is an overlap between the distribution of scores of the untreated (ordinary) population and the distribution of scores for the population in need of treatment (Bauer, Lambert, & Nielsen, 2004)).

The third cutoff score C is the post-test score that must be closer to the average scores for the ordinary individual's population than the average scores for the individuals who need treatment (Peterson, 2008), or it is the degree that is expected to fall between the average of the scores of the distribution of ordinary individuals' population(nonclinical) and the population of Individuals in need of treatment (clinical). This point can be calculated using the following equation (Bauer, Lambert, & Nielsen, 2004):

$$\text{Cutoff C} = [(SD_{clinical} \times M_{nonclinical} + (SD_{nonclinical} \times M_{clinical})] / (SD_{clinical} + SD$$

The third method (method C) of estimating the cutoff point considered as the best method if the information is available on the distribution of the scores of the nonclinical population (who do not need treatment) and the distribution of scores for the individuals who need to treat, especially if there is an overlap between the two distributions (Bauer, Lambert, & Nielsen, 2004)). Also, this method of estimating the cutoff point is more accurate than the previous two methods, especially as it depends on the relative probability of a degree that is obtained through a specific population versus another population. In other words, this cutoff score, which is obtained through the statistical characteristics of the distribution of the scores of ordinary individuals' population (nonclinical or no need for treatment) and the distribution for the clinical population (in need of treatment), which makes this method relatively more accurate compared to the two methods mentioned earlier (Jacobson, Roberts, Berns, & McGlinchey, 1999). On the other hand, the use of methods A and B, is more logical in the absence of sufficient data to calculate the cutoff point according to method C, and it is worth noting that method A is the best in the absence of sufficient data on the nature of the distribution of scores of the ordinary individuals' population (nonclinical) and the distribution of individuals who need treatment(clinical).

Afterward determining the cutoff point according to any of the previous methods, the next step is to determine the amount of change that took place through the use of the so-called Reliability Change Index - (RCI). The value of this indicator is calculated for each individual and by calculating the difference between his/her scores on the pre and post measurements and then dividing this difference by the standard error of the difference scores (calculated by entire of the study). According to this method, the value of the constant change index may be positive (if the score on the post-measurement is greater than the degree on Pre-measurement)., (or zero) if the degree on the pre-measurement is equal to the degree on the post-measurement) or negative (if the degree on the post-measurement is smaller than the degree on the pre-measurement).

Through the use of the cutoff score and the reliability change index (RCI), the individuals who participated in the sample of the study can be classified into four categories: Recovered individuals: those individuals who have reached the state of recovery and no longer need any treatment, and they are known according to the Jacobson-Troax method as those individuals who crossed the cutoff point and obtained a positive-value of reliability change index., and the group of individuals who improved (Improved): they are those individuals whose post-measurement scores have improved compared to their pre-measurement scores but still need to be treated, and defined according to the Jacobson-Troax method (JT) as those individuals who did not go beyond the specified cutoff point but got a positive-value reliability change index, and the category of individuals who did not change (Unchanged): they are those individuals who have not changed and therefore their degrees still need to be addressed, and they are defined according to the method The Jacobson-Troix method (JT) as those individuals who have a zero reliability change index as a result of their pre-measurement scores being equal to their post-measurement scores, and the group of individuals who degraded or regressed Individuals whose their post-measurement scores were lower than their pre-measurement scores and still need to be treated, are defined according to the Jacobson-Trox method (JT) as those individuals who did not cross the specified cut-off point and obtained a negative reliability change index.

To clarify how the clinical significance is estimated according to the Jackson-Trox method, the data referred to in the educational example (Table 1) related to examining the effectiveness of a training program in improving reading skills in fourth-grade children will be calculated, where the reliability change index(RCI) has been calculated For each of the members of the experimental and control groups and using that indicator and the cutoff point 50 words per minute, each member of the previous hypothetical study was classified into one of the four categories mentioned previously and Table 3 indicates these results.

Table 3: Classification of respondents in the hypothetical study" improving reading skills of fourth-grade children" according to the Jackson-Trox method

The name	the group	The number of words in the telemetry	Reliability change index (RCI)	Classification or description of change
Mohammed	Experimental	60	6.89	Treating
Omar	Experimental	47	7.58	Improvement
Hamza	Experimental	65	6.89	Treating
Sawsan	Experimental	46	6.20	Has improved
Safaa	Experimental	32	1.72	Has improved
Ali	Control	24	1.38	Improvement
Qusay	Control	46	2.07	Improvement
Zain	Control	23	1.72	Improvement
Salma	Control	20	-1.38	Retreat
Fatema	Control	30	.00	Fixed or unchanged

Table 3 reveals that two children from the experimental group have moved to the level of recovery or recovery stage, where it is noticed that the two children Muhammad and Hamza have exceeded the set cutoff point which is 50 words per minute, and the reliability change index was positive. As for the children Omar, Sawsan and Safa - who are the rest of the experimental group - they were classified into the category “improved”. As the reliability change index for each one of them was positive, However, neither one of them has exceeded the set cutoff point which is 50 words per minute on the post-test. Regarding the classification of the individuals of the control group, as three of its members, Ali, Qusai, and Zain were also classified into the "improved" category. Once again, this improvement can be justified by a well-known statistical principle, which is the regression of values towards the mean. To clarify this principle, let us assume that 100 students have applied for a test consisting of 100 items of multiple-choice in the Statistical course, where their average score in this test was 70. Suppose that the range of scores on that test ranged from 30 to 90 and that five students, or 5% of the students, had very low scores 30-40, and in return, the same proportion of students (5 students) got high marks 80-90. Statistically, it is expected in this case that if an equivalent test to the first test is applied to the same individuals after a period of time, the scores of students of the first group will be increase and the scores of the second group will be decreased. So that the scores of students of these two groups approximate the average of the scores (70). This phenomenon is known statistically by "regression toward the mean."

Referring to the results of Table 3, it is noted that the child Fatima - who is a member of the control group - was classified into “the fixed” category (they did not have a change in the measured attribute) and the basis of this classification was that the score of this student on the pre-test was equal to her score on the post-test (see Table 1data) and the value of the reliability change index for her performance was zero. As for the child Salma - who is also from the control group - was classified into the category of "deteriorated". As her score on the post-test was lower than her score on the pre-test (see Table 1 data), and the reliability change index for this student was negative.

Gulliksen-Lord-Novick (GLN) method: This method has been suggested to estimate the clinical significance of the results as an attempt to override or correct the error in the Jacobson-Truax (JT) method. From the viewpoint of some statisticians Like Hsu (1999), who noted that using this method to distinguish between pre-test and post-test scores in determining the value of the reliability change index (RCI) did not consider the probability of the regression of the values towards the mean. To solve this problem Hasso devised a new method for estimating Clinical significance relied mainly on some statistical methods presented by the scientist Gulliksen in 1950 and

Lord and Novic in 1968. Hence the abbreviation GLN is used when referring to this method in estimating the clinical significance. According to the (GLN) method, arithmetic mean for a hypothetical population should first be used, whereby the mean of the pre and post scores for each member of the experimental group is subtracted from that mean as a method to solve the problem of the regression of the values towards the mean, and also according to this method (GLN method), the difference between the pre and post scores should be divided by the standard deviation of the scores of the hypothetical population instead of dividing the difference by the standard error as in the case of the Jackson-Troax method (Peterson, 2008).

Edwards-Nunnally (EN) method: The scientist criticized Speer (1992) the Jackson-Trox method (JT) for the same reason that Hsu (1999) cited that this method may be affected by the problem of the regression of the values towards the mean. Accordingly, Speer (1992) presented another method for assessing clinical significance based on the ideas presented by Edwards (1978) and Nunnally (1965). The idea of this method is mainly based on adjusting the pre-test scores so they approach the average of the pre-test scores. In other words, the pre-test scores of the individual after the amendment approaches more than the average pre-test scores and thus the pre-test scores become more homogeneous. Therefore, the effect of the problem of the regression of the values towards the mean decreases, especially in the case of extreme values. The next step, according to the Edwards-Nanoli method, is to estimate the change in individual performance using the Confidence Interval method based on the modified pre-test scores rather than the observed scores. Regarding this point, Peterson (2008) indicates that the use of the confidence interval method requires a greater difference between the pre and post scores at the individual level comparing with the difference obtained by the Jacobson-Trox method in order to be considered clinically significant.

The Hageman-Arrindell (HA) method: Hageman and Arrindell (1999) suggested that two fundamental adjustments should be made to the method presented by Jacobson and Truax- (JT), 1991). The first amendment is the necessity of using different statistical methods to distinguish more clearly between change at the individual level (CS_{indiv}) and change at the group level (CS_{group}). The second amendment includes agreement with Hsu (1999) and Speer (1992) regarding the necessity of adjusting an equation (JT) to consider the problem of the regression of the values to the mean. To solve these problems, Hageman and Arrindell (1999) proposed the use of two new statistical indicators developed by Cronbach and Gleser (1959) in 1959. The two proposed indicators are the Stabilization of Change Index of the Individual (RC_{indiv}), in which it is necessary to classify the individual with an accuracy of no less than 95%. In the same context, both Hejmann and Arendel (1999) indicated that the clinical significance of the change at the individual-level (CS_{indiv}) can be explained by modifying the method of calculating the cutoff point used in the Jacobson and Truax- (JT) method. That modification relies on the true score and the reliability change index for each individual. So that the clinical significance of the individual is used to classify him/her into one of the following groups: 1- Deteriorated, 2- a group of individuals not reliably changed, 3- a group of individuals who improved but not recovered, 4- a group of beneficiaries treated or who have reached the stage of treatment (Recovered). The indicators of change at the group level (group) and the clinical significance of the group ((group) are among the most important concepts presented by Hageman and Arrindell (1999) on the topic of clinical significance and they proposed statistical methods to calculate each indicator.

Hierarchical Linear Method (HLM) This method was introduced to estimate clinical significance by Speer and Greenbaum in 1995 (Bauer, Lambert, & Nielsen, 2004). This method is mainly based on Growth curve models not on the difference between pre and post scores, as is the case in the methods mentioned previously. It is necessary for this method to obtain at least three measurements of the individual at different stages, after which special equations are used to determine the degree of change at the individual level. In fact, the calculations used in this method are relatively difficult, so there are special statistical programs that are necessary to use to estimate the clinical significance according to this method. Peterson (2008) considers this method to be one of the most flexible and useful methods for estimating clinical significance compared to other traditional methods. It should be noted that this method requires in-depth research that may be outside the framework of the current research objectives.

A suggested model for examining the clinical significance of experimental treatments at the group level:

To achieve the last objective of the current study, the following suggested model aiming at determining the clinical significance of treatments on experimental and control groups. This model depends on the use of statistical significance but after classifying the members of the experimental and control groups based on the criteria presented by the Jacobson-Troax method (Jacobson and Truax- (JT), 1991) to estimate the clinical significance. Then use the Chi-Square for crosstabulation to check whether there is a statistically significant relationship between the group variables (experimental - control) and classify the sample members according to the mentioned method. The hypothetical data contained in the educational example that presented earlier in the current study was used for clarifying the suggested model. Referring to the results presented in Table 3, it is noted that the distribution of the numbers of the members of the experimental and control groups according to their classification based on the Jacobson-Trox method (JT) was as shown in Table 4.

Table 4 :The number of respondents in each category in the hypothetical study "improving the reading skills of children fourth grade" based on the Jacobson-Troax (JT) method

Category / the group	Recovered	Improved	unchanged	Deteriorated
Experimental	2	3	0	0
Control	0	3	1	1

It is clear from Table 4 that two individuals of the experimental group have reached the recovery stage whereas the remaining members of the group (three individuals) have improved. As for the control group, no individual has reached the recovery stage that was expected as a result of the absence of treatment. The rest of the control group members were distributed on the other categories, as follows: three individuals within the category of improvement and one individual within the category of unchanged and one individual within the category of deteriorated.

To examine the clinical significance of treatment at the group level, a Chi-Square for the cross-tabulation test was used to check whether there was a statistically significant relationship between group and classification variables. The results of this analysis indicate that the value of the Chi-square is 4 which is not statistically significant ($\alpha > 0.05$). That means that there is no significant relationship between the group and classification variables. This indicates that the treatment used in the hypothetical study was not sufficiently effective to get clinical significance at the group level. However, the same data was sufficient to obtain statistical and practical significance for the differences between the experimental and control groups. Nonetheless, what if it is assumed that the distribution of the individuals of the two groups was different? so that, the effect of the treatment was more noticeable according to the numbers mentioned in Table 5, then do the results of the chi-square will differ in this case?

Table 5. Hypothetical distribution for the number of individuals of the experimental and a control groups into the different categories based on Jacobson-Troax (JT) method

Category / the group	Therapists	Enthusiasts	Stability (They did not change)	Deteriorating or declining
Experimental	4	1	0	0
Control	0	1	3	1

To examine the relationship between the group and classification variables according to the distribution of the individuals of the experimental and control groups into the four groups according to the data that appear in Table 5 based on the Jacobson-Troax (JT) method, chi-square for the cross-tabulation procedure was used. The results of this analysis indicated that the value of the Chi-square is 8, which is statistically significant ($\alpha \leq 0.05$), which indicates that there is a clinical indication of the effect of treatment at the group level. In other words, the results

differed in terms of clinical significance at the group level compared to the results based on the hypothetical data in Table 4 is because the effect of the treatment was more evident in the second hypothetical example.

It appears through the previous results that when using the method or model proposed through the current study to examine the clinical significance of the results at the group level, it can be concluded that when the results are clinically significant at the group level it will necessarily be statistically and practically significant, but the reverse is not true.

5. Conclusions and Recommendations

The primary goal of the current study is to present the concept of clinical significance to researchers in the fields of psychology, educational sciences, and all fields of human sciences. The study also aimed to highlight the differences among the concepts: statistical significance, practical significance, and clinical significance. In ddition, attempted to provide specific statistical methods to examine the clinical significance of the results of psychological and educational research, which used the experimental approach to achieve its goals or to answer its research questions. The current study also focused on the fact that the method of clinical significance is aiming to monitor whether the change that occurs to the individual after the experiment is sufficient to claim that he/she has moved from a state of illness or suffering from a behavioral, psychological or educational problem to healing or improvement. In many cases, this change might be not sufficient to make these allegations. Rather, this change may be unstable, and it is called by some methods that were reviewed in the context of the current study, "No reliable change" or even in the worst cases, the individual's condition may deteriorate or decline. After treatment than it was before treatment. The current study reviewed some statistical methods to examine the clinical significance, even though the researcher decided to focus on the Jacobson and Truax- (JT) method (Jacobson and Truax 1991). As this method is the most used in the experimental psychological and educational studies, in addition, this method - despite its simplicity - it is according to literature - offers an average estimate of the effect of treatment (i.e., it does not reduce nor amplify the effect of this treatment) (Bauer, Lambert, & Nielsen, 2004). Further, the simplicity of the statistical methods used in this method motivates the researchers to implement that method regardless of their statistical background or skills. As it is possible to easily employ some well-known statistical programs such as the Statistical Package for the Social Sciences (SPSS) to extract the statistical values required by this method to determine the clinical significance of the results.

Worthily speaking that researchers interested in using other methods mentioned or even not mentioned in the context of the current study can refer to some references or studies - some of which were mentioned in the context of the current study - that enable them to achieve the goals of their research. In the same context, it can be noticed that some important statistical concepts that may not be used by the psychological and educational researchers due to the high level of difficulty that embedded within those methods. Hence it is better - from the researcher's point of view - to provide simple statistical methods that are necessary to be appropriate and scientifically acceptable to accomplish the intended experimental psychological and educational research objectives. Through what has been mentioned, the focus of the current study was on the Jacobson-Troax (JT) method without other methods, with reference to the importance of other methods in particular, it aims to avoid some problems or defects in the Jacobson-Troax (JT) method , especially the problem the regression of the values toward the Mean.

Also, one of the most important objectives of the current study was to present a model proposed by the researcher to examine the clinical significance at the group level and not only at the individual level as is the case in all the methods that were reviewed in the context of the current study. To achieve this goal, the Jacobson-Troax (JT) method was used to identify or classify the study sample individuals in the experimental and control groups into one of the four groups, : recovered, improved, unchanged, and deteriorated. In a later step, it was suggested to use Chi-Square for cross-tabulation to examine whether there is a statistically significant relationship between the variables: group and classification. The results obtained through the analysis of hypothetical data indicated the ability of this proposed model to examine the clinical significance - from the researcher's point of view - at the group level in an effective, scientific, and practically acceptable way.

The problem of statistical methods that are widely used in psychological and educational studies is that they deal with averages, standard deviations and other statistical values that are extracted through groups and here they may - and unintentionally - neglect that change that is supposed to happen at the individual level and not at the group

level. So that the change that happened to the individual after being subjected to treatment is monitored as he/she is supposed to move from a disease or suffering from a behavioral or emotional problem to a state of recovery or at least improvement. Unfortunately, a quick review of psychological and educational research that used experimental designs enables to be noted that most of these studies were concerned with the statistical significance and few of them concerned the practical significance of the results and the least indicated this very important concept which is the clinical significance.

Finally, it is important to recommend the necessity of examining the clinical significance of experimental psychological and educational research, whether at the individual or the group level. Consequently, it is not acceptable for psychological and educational researchers to be satisfied with examining the statistical significance or even the practical significance of their research results. As the current study showed that the results of experimental research might be statistically and practically significant, but not clinically significant, especially at the individual level.

References

[1] Al-Sayyad, Abdel-Aty, (1988). The practical Significance and the Sample Size Associated with the Statistical Significance of t-test in Psychological and Educational Research (Evaluation study). In Proceedings of the Educational Research Conference" Between Reality and the Future". Cairo University, Vol 2, (pp.197-233).

[2] Al-Thabiti, Ali bin Hamed. (2008). Scientific research designs and their role in validating the results of educational studies. The Journal of the Arab Gulf Letter, No. 108, Year 29, 1-113.

[3] Bauer, S., Lambert, M., & Nielsen, S. (2004). Clinical Significance Methods: A Comparison of Statistical Techniques. Journal of Personality Assessment, 82(1), 60–70.

[4] Campbell, T. C. (2005). An introduction to clinical significance: An alternative index of intervention effect for group experimental design. Journal of Early Intervention, 27, 210-227.

[5] Cronbach, L., & Gleser, G. (1959). Interpretation of reliability and validity coefficients: remarks on a paper by Lord. Journal of Educational Psychology, 50, 230–237.

[6] Gulliksen, H. (1950). Theory of mental tests. New York, NY: Wiley.

[7] Hageman, W. J., & Arrindell, W. A. (1999). Establishing clinically significant change: increment of precision and the distinction between individual and group level analysis. Behavior Research and Therapy, 37, 1169-1193.

[8] Hansen, N., Lambert, M. J., & Forman, E. M. (2002). Comparisons of clinically significant change in clinical trials and naturalistic practice settings: The dose-effect relationship and its implication for practice. Clinical Psychology: Science and Practice, 9, 329–343.

[9] Hsu, L. M. (1999). A comparison of three methods of identifying reliable and clinically significant client changes: commentary on Hageman and Arrindell. Behaviour Research and Therapy, 37, 1195-1202.

[10] Hubbard, R., & Ryan, P. A. (2000). The historical growth of statistical significance testing in psychology--and its future prospects. Educational and Psychological Measurement, 60, 661-681.

[11] Huberty, C. J. (1999). On some history regarding statistical testing. In B. Thompson (Ed.), Advances in social science methodology (Vol. 5, pp. 1-23). Stamford, CT: JAI Press. Huberty, C. J. (2002). A history of effect size indices. Educational and Psychological Measurement, 62, 227-240.

[12] Jacobson, N. S., Follette,W. C., & Revenstorf, D. (1984). Toward a standard definition of clinically significant change. Behavior Therapy, 17, 308–311.

[13] Jacobson, N. S., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. Journal of Consulting and Clinical Psychology, 59, 12–19.

[14] Kendall, P. C. (1999). Clinical significance. Journal of Consulting and Clinical Psychology, 67, 283-284.

[15] Kendall, P. C., Marrs-Garcia, A., Nath, S. R., & Sheldrick, R. C. (1999).

[16] Normative comparisons for the evaluation of clinical significance. Journal of Consulting and Clinical Psychology, 67, 285–299.

[17] Lambert, M. J., Whipple, J. L., Smart, D.W., Vermeersch, D. A., Nielsen, S.L., & Hawkins, E. J. (2001). The effects of providing therapists with feedback on patient progress during psychotherapy: Are outcomes enhanced? Psychotherapy Research, 11, 49–68.

[18] Lord, F., & Novick, M. (1968). Statistical theories of mental scores. Reading, MA: Addison-Wesley.
[19] McGlinchey, J. B., & Jacobson, N. S. (1999). Clinically significant but impractical? A response to Hageman and Arrindell. Behavior Research and Therapy, 37, 1211-1217.
[20] Odah, Ahmad., and Al-Khalili, Khalil (1988). Statistics for Researcher in Education and Humanities, Amman: Dar Al-Fikr for publication and distribution.
[21] Nassar, Yahya. (2006). Using the effect size to examine the scientific significance of the results in the educational and psychological studies used for the quantitative approach, Journal of Educational and Psychological Sciences - Bahrain, Volume 7 - Second Issue.
[22] Nassar, Yahya. (2019). Using SPSS as Instruction Assisted Program in improving Postgraduate Students' Comprehension to Statistical Concepts, Jordan Journal of Educational Sciences, Vol. 15, No. 2, pp 251-257
[23] Peterson, L. (2008). "Clinical" Significance: "Clinical" Significance and "Practical" Significance are NOT the Same Things. Paper presented at the annual meeting of the Southwest Educational Research Association, New Orleans, February 7.
[24] Speer, D. C. (1992). Clinically significant change: Jacobson and Truax (1991) revisited. Journal of Consulting and Clinical Psychology, 60, 402-408.
[25] Stevens, J.P. (1996). Applied multivariate statistics for the social sciences (3rd ed.). NJ: Lawrence Erlabaum.
[26] Thompson, B. (1998). Encouraging effect size reporting is not working: The etiology of research resistance to changing practices. Paper presented at the annual meeting of the Southwest Educational Research Association (Houston , Tx , January 1998). (Eric Document Reproduction Service No. ED 416214).
[27] Thompson, B. (2006). Foundations of behavioral statistics: An insight-based approach. New York: The Guilford Press.
[28] Thompson, B. (2002). "Statistical", "practical," and "clinical": How many kinds of significance do counselors need to consider? Journal of Counseling and Develoment, 80, 64-80.
[29] Wilkinson, L., & APA Task Force on Statistical Inference. (1999). Statistical methods in psychology journals: Guidelines and explanations. American Psychologist, 54, 594-604.