Cypriot Journal of Educational Sciences

Volume 15, Issue 2, (2020) 246-258

www.cjes.eu

Cohen's criteria for interpreting practical significance indicators: A critical study

Ibrahim Said Humaid AlWahaibi*, (IIUM), College of Education, Department of Educational Psychology, International Islamic University (Malaysia), Muscat, P.O. 524, Oman https://orcid.org/0000-0001-5911-7762

Dawood Abdul Malik Yahya AlHadabi, Department of Education, International Islamic University (Malaysia), (IIUM), , Kuala Lumpur, Malaysia

Hussain Ali Talib AlKharusi, Dean of Admission and Registration, Sultan Qaboos University, Muscat, Oman https://orcid.org/0000-0002-5822-9188

Suggested Citation:

AlWahaibi, I.S.H., AlHadabi, D.A.M.Y. & AlKharusi, H.A.T. (2020). Cohen's criteria for interpreting practical significance indicators: A critical study. *Cypriot Journal of Educational Science. 15*(2), 246–258. https://doi.org/10.18844/cjes.v15i2.4624

Received September 2, 2019; revised February 5, 2020; accepted April 2, 2020.
Selection and peer review under responsibility of Prof. Dr. Huseyin Uzunboylu, Higher Education Planning, Supervision, Accreditation and Coordination Board, Nicosia, Cyprus.

Abstract

The present study aimed at clarifying the various shortcomings of the Cohen's criteria for the interpretation of the values of the practical significance indicators. The hypothetical data were used for two experimental and control groups and calculating the paired-samples *t*-test. To clarify the inadequacy of Cohen's criteria in interpreting practical significance indicators, it was compared with another criterion which is Black's Modified Gain Ratio. Through the compatibility of mathematical equations to calculate the practical significance and the values of the interpretations of the correlation coefficient, the present study suggested that a criterion for the practical significance should be as follows: small when the values of the index (*d*) are less than (0.631), medium when the values are between 0.631 and 1.50 and large when the values are equal to or greater than (1.51). The study showed the justifications that distinguish this criterion from the Cohen criterion.

Keywords: Cohen's, criteria, practical, significance.

---

*ADDRESS FOR CORRESPONDENCE: **Ibrahim Said Humaid AlWahaibi,** International Islamic University (Malaysia), Muscat, P.O. 524, Oman. *E-mail address*: ihsalwahaibi@gmail.com / Tel.: +968 99224380

1. Introduction

Due to the shortcomings of statistical significance tests in results interpretation, many statistical methods were found to support the statistical significance and serve as a complementary aspect. Among them practical significance or pedagogical significance, which had a role in liberating researchers from adhering to dual answer to research questions, which enables them to answer the effect magnitude of independent variables in dependent variables, rather than confining to the existence of differences (or not) when the null hypothesis is true.

1.1. The concept of 'practical significance'

The term 'practical significance' appeared in 1901 by scientist Karl Pearson. However, it was not clear in practice until it was known by Ronald Fisher in 1925, where he suggested that researchers complete the statistical significance test in the analysis of variance with the ratio of correlation or eta, which measures the strong correlation between independent and dependent variables (Kirk, 1996). Statistically significant results do not necessarily have a crucial practical significance (Nassar, 2017). Practical significance seeks to clarify the practical and actual importance of the results regardless of being statistically function or not, by calculating the size difference between the two study groups and measuring effectiveness. The practical significance or effect size is defined as the strong relationship between the independent variable (s) and dependent variable (s), or the size of differences between the levels of the independent variable in the dependent variable (George, Jeffrey & Robert, 2006).

Although practical significance indicators have emerged since the 1940s (Huberty, 2002), their use is still limited. Meta-analysis, as well as computer statistical programs, do not calculate most of these indicators (Coe, 2002).

These indicators are rarely discussed in most statistical books. Therefore, the American Educational Research Association (AERA) in its editions since 1994 recommended that researches and studies should be accompanied by 'practical significance of effect size' and the 'qualitative interpretation of the effect' (AERA, 2006, p. 5). Besides, several foreign scientific journals (23) insisted on the requirement to include practical indications in research studies (Pedersen, 2003, p. 311; Vacha-Haase & Thompson, 2004, p. 473). Unfortunately, the researcher did not reach out to the Arab scientific journals that follow the policy of including indications of practical significance in research.

1.2. Importance of practical significance indicators

There is a need to use practical significance indicators to achieve three main objectives (Coe, 2002; Ellis, 2010; Kirk, 2003; Lakens, 2013; Nandy, 2012; Tomczak & Tomczak, 2014).

1.2.1. Study the practical importance of research and study results, and emphasise test results of the null hypothesis

Since the statistical significance does not indicate the practical significance of the results, the results may be statistically significant, but they are not practical, or the size of their effect is limited. As mentioned earlier, the statistical significance is very much affected by the size of the sample. Aktas and Keskin (2013, pp. 583–584) stated an example to clarify the role of sample size, which reached (1 million) in the presence of statistical significance although the difference between the two groups is almost zero (0.0046). This means that the two groups have identical data. The researcher used data from a study—he analysed statistically in his centre 'Al-Khwarizmi Center for Educational Studies and Consultations'—which consisted of (300) individuals. He replicated data to (7,200) individuals, with a slight modification for some of them. Then, he conducted a test to compare the two proposed groups. The outcomes showed that the test results (2.679) which is statistically significant (0.007) though the difference between the two groups did not reach the correct one (1.00) where it reached (0.08) only.

On the other hand, non-statistical significance results may have practical significance and are characterised by medium or large effect sizes. However, because of the small-sized sample led to non-statistical significance. Nonetheless, Fan (1999) warns against explanation on the large effect size because it could be due to chance.

Al-Qudhah (2016) and Mahmoud (2003) pointed out that Huston (1993) claimed that practical significance measures the extent of the existence of the studied phenomenon in society, through a continuous scale with a numerical value starting with zero which represents the absence of the phenomenon in society, and its increase indicates the degree of its existence.

Therefore, it is necessary to combine the results of statistical significance and practical significance as they are complementary to each other. Consequently, relevant results, statistically and practically significant can be trusted and reliable in decision-making (Sun, Pan & Wang, 2010).

1.2.2. Comparing the results of research studies conducted in the same fields

One of the advantages of practical significance is to use in comparing the results of studies conducted in a particular field (Aron & Aron, 1994) and judge which is better. The best study is where the effect size is higher. The practical significance can be used in meta-analysis by combining estimates of different impact sizes for each study to give the best overall estimate of the magnitude of impact for studies in a given area, and comparing results from different sources correctly done by different authors. The practical significance is critical to research combinations and metadata that incorporate quantitative results from studies for different related phenomena (Tomczak & Tomczak, 2014). The statistical significance tests based on sample size do not allow such comparisons.

Therefore, it is necessary to emphasise the necessity of spreading the knowledge of effect indicators among researchers. Not using any of these indicators combined with statistical significance tests negatively affects the conduct of meta-analysis and accumulation of knowledge over the years (Sun et al., 2010).

1.3. Determine the sample size

When planning a new study, statistical strength analysis based on the practical significance of the previous studies can be used to determine the average sample size required by the study to obtain statistically significant results at a given significance level (Lakens, 2013; Tomczak & Tomczak, 2014).

1.3.1. Factors affecting the indications of practical significance

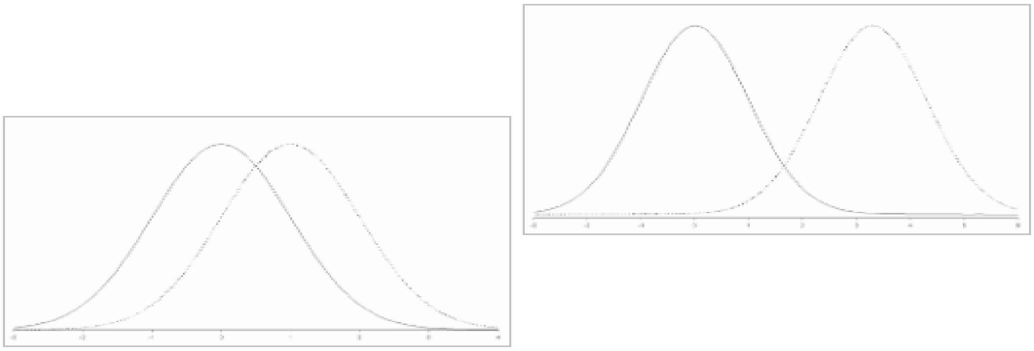
There are many factors that directly affect the indicators of practical significance, including the standard deviation used, ideally, the use of the standard deviation of the control group gives more accurate results, since the members of this group were not affected by the procedures of the experiment, but if the control group is not very large; the standard deviation becomes less accurate. In some studies where there is no control group such as two experiments to see which is better, choosing a standard deviation for one of the two groups reduces the accuracy of the results of the practical significance or the effect size. Therefore, the most appropriate solution is using a pooled estimate 'Pooled' standard deviation, the average standard deviations of the experimental and control groups, as in Eq. (1), (Coe, 2002, p. 10).

$$SD_{pooled} = \sqrt{\frac{(N_E - 1)SD_E^2 + (N_C - 1)SD_C^2}{N_E + N_C - 2}} \quad (1)$$

NE: Number of experimental groups, Nc: Number of control groups

SDE2: standard deviation of the experimental group, SDc2: standard deviation of the control group

Another factor that influences the accuracy of the practical significance indicators is the moderate distribution of the two study groups (experimental and control). If one of the groups (or both) does not have a normal distribution, the effect size or practical significance will not be accurate which leads to false interpretations. Coe (2002, p. 13) stated an example of how dangerous the uneven distribution in finding inaccurate values for effect size indicators, as in Figure 1:



(a) (b)

Figure 1. Normal and abnormal distribution at effect size = 1

From Figure 1, Normal distribution (a) and the abnormal distribution (b), the difference between the experimental and control groups in graph (b) appears large significant. In terms of the amount of overlap, in (Chart b) 97% of the 'experimental' group is above the control group, which is high compared to the true overlap between the two groups at an effect size equal to (1.00), which is only 84%, as in Table 1. This difference represents the risk of using practical significance indicators when the distribution is not normal (Coe, 2002). On the other hand, factors affect the accuracy practical significance indicators.

Due to the fact that the indicators of practical significance are mathematical equations and when used are reached numerical values, like statistical significance tests; and there must be criteria for the interpretation and explanation of these numerical values, in order to reach an interpretation of the results of the study to build appropriate statistical decisions, so several interpretations of the values of indicators have aroused. The most famous was the Cohen standard in 1969, in which the effect size (*d*) was classified as small, medium and large (0.2, 0.5 and 0.8), respectively, in which the values of the other indicators were interpreted as the ETA (η^2). McLean (1995) developed another criterion for interpreting Effect Size (*d*). Considering that the magnitude of the effect is small if the value of (*d*) (less than 0.5), average if equal (0.5–1.0), and large if (greater than 1.0) (McLean, O'Neal & Barnette, 2000). Slavin and Fashola (1998) suggested that the effect size should be equal to or more than 0.25 as evidence for effectiveness. The US Department of Education, represented by the Joint Publishing Review Committee (JDRP), has its standard, where the effectiveness of innovative educational program projects requires that the effect size be at least (1.00). Although multiple Criteria for interpreting the values of practical significance indicators have become common, Cohen's criteria has become most common applicable for most researchers—if not all of them—such as (Al-Barqi, 2012; Al-Darabee, 2003; Al-Haddad, 2006; Al-Jawda, 2004; Al-Maliki, 2018; Al-Qudhah, 2016; Christopher, 2006; Ibrahim, 2000; Lakens, 2013; Mahmoud, 2003; Nandy, 2012; Nassar, 2006; Olejnik & Algina, 2000; Tomczak & Tomczak, 2014).

Barnette and McLean (1999) do a study to determine empirically-based criteria for the interpretation of practical significance values. The study aimed to determine the extent to which effect sizes differ by chance, and how standard effect sizes correspond to Cohen's criteria for interpreting small, medium and large effect sizes. The standard effect is random or uniform across group numbers and sample sizes, and whether standard effect sizes can be predicted using degrees of freedom, the number of groups and sample sizes. Monte Carlo procedures have been used to create uniform effect sizes in case of one-way analysis of variance with a number of groups ranging from (2 to 10) groups with sample sizes from (5 to 100), data have been repeated (5,000) times, and the results found that the use of Cohen's criteria (1988) to judge the practical significance is risky. Equations that can be used to predict standard effect sizes were developed, using the number of groups and sample size. The prediction equations were very accurate, and this study provided a better alternative to assess the empirical standard for practical significance than the use of arbitrary and static criteria that often use practical significance as small, medium, or big. Barnette and McLean (1999) argued that the interpretation of practical significance varies from study to study based on specific variables as described in their equations, which have come to predict the standard practical significance. Although this method is good and has its characteristics, the existence of fixed criteria for the interpretation of practical significance is important in meta-analysis to compare the results of studies conducted in a particular field. Also, the existence of a clear criterion for the interpretation of the importance of practical significance helps non-specialised researchers in statistics in interpreting the results of their studies easily.

McLean et al. (2000) presented a working paper at the annual meeting of the Educational Research Association, entitled 'Are All Effect Sizes Created Equal?' They calculated the effect sizes for the results of a national standardized test that was applied for grades 4, 6 and 8 in 749, 574 and 464 schools, respectively. The sample size reached 120,149 individuals and raw data converted to standard scores and normal curve equivalent. The study found that measurement or data type makes a difference when calculating effect sizes, so we should not have a one-size-fits-all rule for interpreting effect sizes, and we should take the type of results with other factors into account when interpreting effect size.

1.3.2. The problem of the study

Based on the above, the problem of the study is addressed through research the following questions:

Question 1: How accurate is Cohen's criteria for interpreting practical significance indicators?

Question 2: Are there alternative solutions to Cohen's criteria for interpreting practical significance values?

1.3.3. Objectives of the study

The present study aims to:

1) Determine the accuracy of Cohen's criteria to interpret the indications of practical significance.

2) Propose alternative solutions to Cohen's criteria for interpreting practical significance values.

1.3.4. The importance of the study

This study seeks to clarify the shortcomings in Cohen's criteria to explain the values of practical significance, which researchers have been using for a long time. Additionally, this study came in response to calls for studies that claim inaccuracy of Cohen's criteria, and this criterion created statistical problems that caused the amplification of the simple difference between treatments. This study may be useful in finding the most appropriate criterion for the interpretation of the values of practical significance.

1.3.5. Definition of terms

Practical Significance: The practical significance represents the magnitude of the effect. Cohen (1988) defines it as the parameter that estimates the degree of departure from the null hypothesis (Vera Garcia, 2017). Pallant (2010) defines it as a set of mathematical equations used to find the differences between arithmetic averages for levels of the independent variable. Or the amount of total variance caused by the independent variable in the dependent variable.

Cohen's criteria: A criterion founded by the statistical scientist Cohen (1962) to interpret the values of the index of practical significance of statistical tests that look for differences between groups as tests (T). This criterion consists of three levels (small) if the index of the effect size is less than (0.5), medium (0.5–0.8) and large (0.8 and above) (Al-Maliki, 2018). Cohen generalised this criterion to various other indicators to calculate the magnitude of the impact or practical significance of statistical tests, such as the index of the ETA square (μ^2), and the Omega square (w^2), and other indicators.

1.3.6. Methodology and procedures of the study

To answer the study questions and to achieve its objectives, a descriptive and analytical approach was used to review the reasons for the shortcomings of Cohen's criteria in interpreting the indications of practical significance. The hypothetical data were used for two experimental and control groups and calculating the paired-samples *t*-test. To clarify the inadequacy of Cohen's criteria in interpreting practical significance indicators, it was compared with another criterion which is Black's Modified Gain Ratio. Therefore, to obtain an appropriate standard for the interpretation of practical significance values, alternative solutions have been provided for Cohen's criteria by making use of interpretations of the correlation coefficient values and linked to the values of '*d*' corresponding index, which Cohen clarified in his book (Cohen, 1988).

2. Results

2.1. The answer to the question 1

2.1.1. How accurate is the Cohen criteria for interpreting practical significance indicators?

From the theoretical literature, it is clear to the researcher that the classification of Cohen's criteria cannot be blindly adopted in various research fields; and it is merely a heritage adopted by the researchers in the manner in which they began to study effect size for the following reasons, (Ialongo, 2016; McLean et al., 2000; Pedersen, 2003):

- Cohen has based his criterion on the phenomena he observed in his research area of behavioural sciences. Cohen stated an example of the average age of girls in the United States of America (Cohen, 1988, p. 26), and therefore this classification is difficult to apply to other areas. This critique was not recent, but only a while ago, when Coe (2002) stated that (Glass, McGaw & Smith, 1981) were among the critics of this approach and that Cohen himself acknowledges the risk of using these terms (Small, Medium and Large) out of context.
- The criteria of Cohen (1969) and (Slavin & Fashola, 1998) were based on experiences and wisdom only (McLean et al., 2000).
- Cohen also admitted that these values (0.2, 0.5, and 0.8) are related to the specific context and style of a particular research situation (Barnette & McLean, 1999; Coe, 2002; Cohen, 1988). Additionally, Cohen did not talk about the interpretation of effect size if less than (0.2).
- Barnette and McLean (1999) proved with convincing evidence that the uses of Cohen's criteria (1988) to judge practical significance were risky.
- Lakens (2013) quoted from Thompson (2007) as saying that the values upon which Cohen based his interpretations of effect size are arbitrary and should not be interpreted strictly.
- Cohen and other researchers attempted to interpret the different effect magnitude values by assuming that they are similar to the *Z*-score. Although this 'percentile' interpretation seems appropriate, it is costly and improper because it is based on the assumption that the baseline distribution is normal (Coe, 2002; Lalongo, 2016), which researchers cannot control. In the standard normal distribution where about 99.9% of the data fall between 3 and 3 standard degrees, the magnitude of the effect size values have no upper limit. It may reach a value of (10) or more, and therefore the researcher declares to study this similarity more accurately and in-depth and provide examples to verify the validity. Table 1 illustrates the relationship between effect size and the percentage of the percentile, and the correlation coefficient.

Table 1. The relationship between effect size, percentile grade and correlation coefficient.

Effect size	Percentile grade (%)	Non-overlap (%)	Correlation coefficient	Effect size	Percentile grade (%)	Non-overlap (%)	Correlation coefficient
0.0	50.0	0.0	0.000	1.6	94.5	73.1	0.625
0.1	54.0	7.7	0.050	1.8	96.4	77.4	0.669
0.2	57.9	14.7	0.100	2.0	97.7	81.1	0.707
0.3	61.8	21.3	0.148	2.2	98.6	84.3	0.740
0.4	65.5	27.3	0.196	2.4	99.2	87.0	0.768
0.5	69.1	33.0	0.243	2.6	99.5	89.3	0.793
0.6	72.6	38.2	0.287	2.8	99.7	91.2	0.814
0.7	75.8	43.0	0.330	3.0	99.9	92.8	0.832
0.8	78.8	47.4	0.371	3.2	99.9	94.2	0.848
0.9	81.6	51.6	0.410	3.4	>99.95	95.3	0.862
1.0	84.1	55.4	0.447	3.6	>99.95	96.3	0.874
1.2	88.5	62.2	0.514	3.8	>99.95	97.0	0.885
1.4	91.9	68.1	0.573	4.0	>99.95	97.7	0.894

'Analysis of statistical power', Cohen, 1988, p. 22).

When Cohen explanations are studied as shown in Table 1, there is a contradiction between them and the explanations of the correlation coefficient. For example, if the correlation coefficient is small (0.24), the effect size is medium (0.5), and if the correlation coefficient is medium (0.45), the effect size is very large (1.00).

- Deep scientific mobility in recent years, in statistics and on-going discussions in old statistical societies and specialised statistical journals, indicate that the practical significance or magnitude of the case, such as statistical significance or '*p*' values, suffers from statistical problems and its misuse to determine the effectiveness of treatments (Hassan, 2008). The biggest problem is the small values that represent the medium effect size or the large effect size, defined by Cohen '*d*' (0.5, 0.8) respectively (Pogrow, 2019). These values amplify the slight difference between processors.
- Cohen criteria for Interpreting Effect Size Values contradicts the interpretation of Black's Modified Gain Ratio which looks for the effect size of the *t*-test for two correlated samples through Eq. (2) (Al-Obaidi, 2019):

$$\text{E.SB}=\frac{y-x}{z-x}+\frac{y-x}{z} \qquad (2)$$

Where *y*: mean of post-measurement, *x*: mean of pre-measurement, *z*: the total score of the measuring instrument.

The value of the Blacks Gain Index (E.SB) ranges from the value (0–2), and the effect size is considered effective if the gain index is a value of (1.2) (Al-Obaidi, 2019). The researcher calculated hypothetical data for a test of two correlated groups (using two-dimensional measurements and follow-up of an experimental group in a program).

Table 2. Results for paired-samples *t*-test of hypothetical data

Experimental group	*N*	*M*	SD	Df	*r*	*T*	*p*-value
Post- measurement	20	1.50	0.513	19	−0.349	−3.005	0.007
Follow-up measurement		2.75	1.618				

When applying the Eq. (2) of the Black's gain index (E.SB) to the test results (*T*) for two correlated samples in Table 2, we find that the value is (0.61). This result indicates the size of the average effect, but when using Cohen index (*d*) in Eq. (3) (Al-Obaidi, 2019). The value of the index (*d*) is (1.10):

$$d = \frac{Y}{s_d} \times \sqrt{2(1-r)} \quad (3)$$

where: Y: mean of differences between pre and post measurements,

SD: Standard deviation of differences,

r: correlation coefficient between pre and post measurements

However, when calculating the value of Eta squared (η^2) by Eq. (4) (Al-Maliki, 2018; Tomczak & Tomczak, 2014):

$$\eta^2 = \frac{t^2}{(t^2 + df)} \quad (4)$$

where: df: degrees of freedom for the t-test in case of two independent samples t-test (N + 1, N−2), and case of the paired-samples t-test (N-1)

The value of (η^2) is then converted to the Cohen Index (d) using Eq. (5) (Al-Dawy, 2006; Al-Maliki, 2018):

$$d = \frac{2\sqrt{\eta 2}}{\sqrt{1-\eta 2}} \quad (5)$$

Cohen Index (d) was 1.38, so using Cohen's interpretations of effect size values (1.10, 1.38) obtained from Eqs. (4) and (5) leads to a contradiction between the interpretation of the Black's Gain Index obtained from the Eq. (3), where the Cohen explanations refer to a large effect size, since both values −1.10, 1.38—bigger than (0.8), while Black's explanation of effect size (medium) where the value (0.61) is less than (1.2).

Based on the above, the researcher considers reviewing the Cohen criterion for the interpretation of values of practical significance, to avoid the previous shortcomings, and to find balanced values in the interpretation of practical significance.

2.2. Answer to the question2

2.2.1. Are there alternative solutions to the Cohen criterion for interpreting practical significance values?

Cohen, in his criterion for interpreting effect size levels, determined that a small difference between the averages of two study groups by only half a standard deviation is considered by a medium effect size, and a difference of eight-tenths of a standard deviation—(0.8)—he calls a significant effect size, although there is no upper limit. Based on the fact that this criterion is based on unpractical experiences only, and that Cohen is concerned about the generalisation of this criterion. The researcher considers that there is no proportionality between the values of the practical significance indicators and Cohen explanations (Small, Medium and Large).

Therefore, by looking at the theoretical literature and relying on certain axioms in statistics such as the impossibility of exceeding correlation coefficient the correct one (1.00), and conducting some statistical applications; the researcher proposes interpretations of the values of indicators of practical significance differ from the Cohen's interpretations which adopted by researchers in their studies more than half a century ago. The researcher relied on his criterion on the following:

1. The correlation coefficient has a maximum value and the correct one (1.00) cannot be exceeded.
2. Thus, the period in which the correlation coefficient is small, the effect size will be small, and so on.
3. The correlation coefficient is known to be 'small' if it does not exceed (0.3), 'average' if it is (0.5) and 'large' if exceeded (0.7) (Al-Shayeb, 2009; Rumsey, 2011).
4. The following equation used for converting effect size (*d*) to Bi-serial correlation coefficient (Cohen, 1988, p. 23):

$$r = \frac{d}{\sqrt{d^2 + 4}} \quad (6)$$

5. Table 3 illustrates suggested explanations for practical indications.

Table 3. The explanation proposed by the researcher for the effect size

Correlation coefficient	Effect size values '*d*'	Explanation of the practical significance suggested by the researcher
0–0.30	0–0.630	Small
0.31–0.60	0.631–1.50	Medium
0.61–1.00	1.51 and above	Large

Table 3 shows the modified explanations proposed by the researcher for the effect magnitude values, where it is 'small' if the values of '*d*' do not exceed (0.63), 'average' in the period (0.631–1.50) and 'large' if the value exceeds (1.50).

2.2.2. Justification for the proposed interpretation of the practical significance

The researcher relied on several justifications for his proposed explanations for the impact magnitude values, the most important ones are the following:

First: the equation for converting effect size (*d*) to the correlation coefficient of the Be-serial as in the previous Eq. (6).

Second: Cohen (1988), in his book titled 'Statistical Power Analysis' (pp. 77–82), pointed to effect sizes of the correlation coefficient; he explained that the effect size was small when ($0 \leq r \leq 0.29$) and medium when ($0.30 \leq r \leq 0.49$), and large when ($0.50 \leq r \leq 1.00$). This interpretation fits with the explanation proposed by the researcher for the values of the effect size. However, Cohen's interpretation of the effect sizes of the correlation coefficients contradict the values reported by himself in Table 1, where he mentioned the effect of correlation coefficient is median at (0.3) and in Table 1, the effect of the correlation coefficient is moderate—($d = 0.5$)—when the correlation coefficient is equal to (0.243). Cohen discussed his justifications for the values of the correlation coefficient which contradicts his previous explanations of '*d*' values in Table 1. He justified that the correlation coefficient mentioned in Table 1 is the Bi-serial correlation coefficient (r_p) point. The explanations previously given for correlation coefficients are shown in Table 1 is the Bi-serial correlation coefficient of normal distribution. The equation for converting (r_p) to (r_b), as defined by Eq.. (7) (Cohen, 1988, p. 82).

$$r_b = 1.253\, r_p \qquad (7)$$

So that the effect size values (small, medium and large) are equal to the values of the correlation coefficients (small, medium, large), so that the small effect size ($d = 0.2$) is equal to the small correlation coefficient ($r = 0.3$), and so on. However, as it is known that the correlation coefficients of different types do not exceed the correct one (1.00), and Cohen equation to convert (r_p) to (r_b)—Eq. (3)—is incorrect, because when the value of the correlation coefficient (r_p) is equal to (0.814)—which is equal to the value of ($d = 2.8$) as in Table 1—the value of the correlation coefficient (r_b) is equal to (1.02) and this value exceeds the maximum value of the correlation coefficient which is the correct one (1.00).

Third: Assuming that the basic distribution follows the normal distribution, the non-overlap ratios of the explanations assumed by the researcher find that they are balanced at the subdivisions (small, medium and large). When the effect size is small (0.63 or less), the non-overlap ratio is approximately 40% as in Table 1. When the effect size is large (1.51 or more), the non-overlap ratio is (70.7%). Thus the non-overlap ratios (0.00%–100.00%) are divided into semi-equal divisions when taking the explanations suggested by the researcher for the effect size values. Although, Cohen subdivisions [small (0.2), medium (0.5) and large (0.8)] do not correspond to logic in terms of non-overlap ratios between the two total distributions. When the effect size is (0.2), the non-overlap ratio is only 14.7%, the non-overlap ratio is (33%) when the effect size is (0.5). Whereas according to Cohen's criterion for large effect size (0.8) the non-overlap ratio is (47.4%) which did not exceed (50%).

Fourth: according to Table 3, the criterion proposed by the researcher fits and corresponds to Black's interpretation, where Cohen (*d*) values using Eqs. (4) and (5) fall within the range (0.631–1.50), which the researcher interpreted the size of a medium effect, and this explanation is consistent with the interpretation of BlackBerry.

Fifth: the researcher's suggestion converges with the standards of the US Department of Education's JDRP, which requires the effectiveness of innovative educational program projects that must be at least (1) (McLean et al., 2000).

Sixth: The researcher did not take the interpretation proposed by McLean (1995), who divided the size of the effect into (small, medium and large) with values (less than 0.5, 0.5–1.0, greater than 1) respectively. Although, McLean criterion closer to the researcher's explanation, his divisions (small, medium and large) does not fit with the correlation coefficients of mentioned earlier.

Seventh: The deep scientific mobility during recent years in statistics and the continuous discussions in the ancient statistical societies and specialised statistical journals indicate that the practical significance or magnitude of the case, such as statistical significance or *p* values, suffers from statistical problems and its misuses to determine the effectiveness of treatments (Hassan, 2008). The largest of these issues is the small values that represent 'the average effect size' or the 'large effect size', which Cohen identified (0.5 and 0.8), respectively (Pogrow, 2019, p. 225). These values amplify the slight difference between processors. Consequently, the researcher suggested interpretation of effect size values in Table 2 contributes to the treatment of this problem where small and large effect size values are equal (0.631 and 1.51), respectively.

2.2.3. Researcher suggested practical significance levels for other indicators

2.2.3.1. First: Indicators (η^2, ω^2, f)

Cohen (1988) indicated that the value of the index (*d*) can be converted to index (*f*) using the following equation:

$$d = 2f \qquad (8)$$

The conversion of the contrast ratio (η^2) to the effect size (*f*) and vice versa can also be converted by the following equations:

$$f = \sqrt{\frac{\eta^2}{1-\eta^2}} \qquad (9)$$

$$\eta^2 = \frac{f^2}{1+f^2} \qquad (10)$$

Suggested practical significance levels for other indicators:

For the omega square index (ω^2)

Nassar (2006) indicated that there is no criterion for interpreting its value, and the same Eta criterion is often used. (η^2, ω^2, f) are:

Table 4. Practical significance Levels by indicators (η^2, ω^2, f)

Indicator	Practical significance levels		
	Small	Medium	Large
D	$0 \leq d < 0.631$	$0.631 \leq d < 1.51$	$d \geq 1.51$
f	$0 \leq f < 0.316$	$0.316 \leq f < 0.755$	$f \geq 0.755$
η^2	$0 \leq \eta^2 < 0.091$	$0.091 \leq \eta^2 < 0.363$	$\eta^2 \geq 0.363$
ω^2	$0 \leq \omega^2 < 0.091$	$0.091 \leq \omega^2 < 0.363$	$\omega^2 \geq 0.363$

2.2.3.1 Second: indicators (r, R^2, F^2)

The practical significance of the Pearson correlation coefficient is explained by the value of the correlation coefficient (r) itself or by the squared of its value (R^2) (Al-Maliki, 2018). The practical significance of the independent variables in the multiple regression model is obtained by Eq. (11) (Cohen, 1988):

$$F^2 = \frac{R^2}{1 - R^2} \quad (11)$$

where: R^2: is the value of the coefficient of determination

Therefore, the suggested practical significance levels for the indicators (r, R^2, F^2) are:

Table 5. Practical significance Levels by indicators (r, R^2, F^2)

Indicator	Practical significance levels		
	Small	Medium	Large
D	$0 \leq d < 0.631$	$0.631 \leq d < 1.51$	$d \geq 1.51$
R	$0 \leq r < 0.31$	$0.31 \leq r < 0.61$	$r \geq 0.61$
R^2	$0 \leq R^2 < 0.096$	$0.096 \leq R^2 < 0.372$	$R^2 \geq 0.372$
F^2	$0 \leq F^2 < 0.106$	$0.106 \leq F^2 < 0.593$	$F^2 \geq 0.593$

Based on what has been mentioned earlier, the researcher emphasises that the most appropriate fields to use his suggested explanations of effect size in psychological, educational and social studies only because a semi-experimental approach is often used there and the accuracy of the results do not reach (100%). Also, we realise that the practical significance depends on costs and benefits. For example, a small effect size (d = 0.6) does not necessarily mean that the treatment method used is inefficient, but the costs and application efforts must be taken into account. It is preferable to apply medium-effect-sized method rather than big-sized- one. Educational and scientific decision-making process should rely on the results of meta-analyses studies, not on the results of the effect values of individual studies.

For the results of medical studies that apply the experimental method and rely on rigorous experimental designs, it is possible to rely on Cohen's criteria of the effect size.

References

AERA. (2006), *Standards for reporting on empirical social science research in AERA publications*. Retrieved from https://www.aera.net/Portals/38/docs/12ERv35n6_Standard4Report%20.pdf

Aktas, A. & Keskin, B. (2013). *Statistical power analysis* (pp. 578–587). The 7th International Days of Statistics and Economics, Prague, Czech Republic, (September 19-21, 2013). https://msed.vse.cz/files/2013/224-Keskin-Burak-paper.pdf

Al-Barqi, T. H. (2012). *The reality of statistical significance and practical significance of research published in the Journal of Umm Al-Qura University for Educational, Social and Human Sciences* (pp 1425–1430). (Unpublished master's thesis) College of Education, Umm Al-Qura University, Mecca, Saudi Arabia.

Al-Darabee, M. Y. (2003). The reality of statistical and practical connotations and the strength of statistical tests used in the research of Mu'tah Journal for Research and Studies - Humanities and Social Sciences Series. *Mutah Research and Studies - Humanities and Social Sciences, 18*(1).

Al-Haddad, S. A. M. (2006). *The strength of statistical tests and the reality of statistical significance and practical significance in the research of the Journal of the National Authority for Scientific Research in the Great Jamahiriya.* (Unpublished master's thesis) Faculty of Graduate Studies, Omdurman University, Omdurman, Sudan.

Al-Jawda, M. M. S. (2004). *Statistical significance, practical significance and test force of research published in the Yarmouk Research Journal of Humanities and Social Sciences Series (1985–2001).* (Unpublished master's thesis) Faculty of Graduate Studies, Yarmouk University, Irbid, Jordan.

Al-Maliki, F. A. O. (2018). *The reality of statistical significance and practical significance and the strength of statistical testing (applied study on the research of educational journals in Saudi universities)* (Unpublished doctoral dissertations) Umm Al-Qura University, Mecca, Saudi Arabia.

Al-Obaidi, A. A. K. (2019). The magnitude of the impact and its applications in the methodology of educational, psychological and medical research (1st ed.). Amman, Jordan: Wael Publishing & Distribution House.

Al-Qudhah, M. A. S. (2016). *The degree of compatibility between the result of statistical significance and practical significance in the results of statistical analyzes in master theses at the Faculty of Educational Sciences at Mu'tah University*. (Unpublished master's thesis) Mu'tah University, Mutah, Jordan.

Al-Shayeb, A. H. (2009). *Fundamentals of educational research* (1st ed.). Amman, Jordan: Wael Publishing House.

Aron, A. & Aron, E. N. (1994). *Statistics for psychology*. Englewood Cliffs, NJ: Prentice-Hall, 07632.

Barnette, J. J. & McLean, J. E. (1999). *Empirically based criteria for determining meaningful effect size*. Paper presented at the annual meeting of the MidSouth Educational Research Association, Point Clear, AL.

Christopher, A. S. (2006). Practical significance: The use of effect sizes in school counseling. *School Counseling, 9*(5), 403–406. https://doi.org/10.1177/2156759X0500900406

Coe, R. (2002, September 12–14). *It's the effect size, stupid: What effect size is and why it is important*. Paper presented at the Annual Conference of the British Educational Research Association, Exeter, UK: University of Exeter. https://www.leeds.ac.uk/educol/documents/00002182.htm

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Ellis, P. D. (2010). *The essential guide to effect sizes statistical power, meta-analysis, and the interpretation of research results*. New York, NY: Cambridge University Press.

Fan, X. (1999, November 3–6). *Statistical significance and effect size: Two sides of a coin*. Paper presented at the Annual Meeting of the American Evaluation Association, Orlando, FL. Retrieved from https://files.eric.ed.gov/fulltext/ED435702.pdf

George, A. M., Jeffrey, A. G. & Robert, J. H. (2006). *Understanding and evaluating research in applied and clinical settings*. Mahwah, NJ, London, UK: Lawrence Erlbaum Associates, Publishers.

Huberty, C. J. (2002). A history of effect size indices. *Educational and Psychological Measurement, 62*(2), 227–240. doi:10.1177/0013164402062002002

Ialongo, C. (2016). Understanding the effect size and its measures. *Biochemia Medica, 26*(2), 150–163. doi:10.11613/BM.2016.015

Ibrahim, I. R. I. H. (2000). *The reality of statistical significance and practical significance and the strength of statistical tests used in the researches of the Journal of Educational Sciences Studies in the years (1997-1998) at the University of Jordan*. (Unpublished master's thesis) Faculty of Graduate Studies, University of Jordan, Amman, Jordan.

Kirk, R. E. (1996). Practical significance: A concept whose time has come. *Educational and psychological measurement, 56*(5), 746–759. doi:10.1177/0013164496056005002

Kirk, R. E. (2003). The importance of effect magnitude, In S. F. Davis (Ed.), *Handbook of research methods in experimental psychology* (pp. 83–105). Oxford, UK: Blackwell.

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for *t*-tests and ANOVAs. *Frontiers in Psychology, 4*, 863. doi:10.3389/fpsyg.2013.00863

Mahmoud, F. M. M. H. (2003). *The reality of statistical significance, practical significance and test force for the statistical tests used in the master theses issued by the Faculty of Educational Sciences at Mutah University*. (Unpublished master's thesis) Mutah, Jordan: Faculty of Graduate Studies, Mu'tah University.

McLean, J. E. (1995). *Improving education through action research: a guide for administrators and teachers*. Thousand Oaks, CA: Corwin Press, Inc.

McLean, J. E., O'Neal, M. R. & Barnette, J. J. (2000). *Are all effect sizes created equal?* Paper presented at the Annual Meeting of the Mid-South Educational Research Association 28th, Bowling Green, KY: ERIC Document Reproduction Service No. ED 448188. https://files.eric.ed.gov/fulltext/ED448188.pdf

Nandy, K. (2012). Understand and quantifying effect size. Oakland, CA: University of Californian.

Nassar, Y. H. B. (2006). Use effect size to examine the practical significance of results in quantitative studies. *Journal of Educational and Psychological Sciences, 7*(2), 35–59*.* Bahrain. (in Arabic)

Nassar, Y. H. B. (2017). Clinical significance of psychological and educational research used for experimental designs: An analytical study. *Journal of Educational and Psychological Studies, 11*(2). Sultan Qaboos University. College of Education, Sultanate of Oman. (in Arabic). https://www.squ.edu.om/Portals/53/College%20Journal/2017/JEPS%20Vol.%2011%20Issue%202.pdf?ver=2017-06-22-123815-320

Olejnik, S. & Algina, J. (2000). Measures of effect size for comparative studies: Applications, interpretations, and limitations. *Contemporary Educational Psychology, 25*(3), 241–286. doi:10.1006/ceps.2000.1040

Pallant, J. (2010). *SPSS survival manual: A step by step guide to data analysis using SPSS* (4th ed.). Maidenhead, UK: Berkshire; New York, NY: Open University Press, c2005.

Pedersen, S. (2003). Effect sizes and "What If" analyses as supplements to statistical significance tests. *Journal of Early Intervention, 25*(4), 310–319. doi:10.1177/105381510302500406

Pogrow, S. (2019). How effect size (practical significance) misleads clinical practice: The case for switching to practical benefit to assess applied research findings. *The American Statistician, 73*(Sup 1), 223–234. doi:10.1080/00031305.2018.1549101

Rumsey, D. J. (2011). *Statistics for dummies* (2nd ed.). Hoboken, NJ: Wiley Publishing, Inc.

Slavin, R. E. & Fashola, O. S. (1998). *Show me the evidence: Proven and promising programs for America's schools*. Thousand Oaks, CA: Corwin Press.

Sun, S., Pan, W. & Wang, L. L. (2010). A comprehensive review of effect size reporting and interpreting practices in academic. *Journal of Educational Psychology, 102*(4), 989. doi: 10.1037/a0019507

Tomczak, M. & Tomczak, E. (2014). The need to report effect size estimates revisited. An overview of some recommended measures of effect size. *Trends in Sport Sciences, 1*(21), 19–25. http://tss.awf.poznan.pl/files/3_Trends_Vol21_2014__no1_20.pdf

Vacha-Haase, T. & Thompson, B. (2004). How to estimate and interpret various effect sizes. *Journal of Counseling Psychology, 51*(4), 473–481. doi:10.1037/0022-0167.51.4.473

Vera Garcia, F. (2017). Effect-size reporting in Mexican psychology journals: What it says about the quality of research within the field. *Ciencia Ergo Sum, 24*(3), Retrieved from http://www.redalyc.org/articulo.oa?id=10452159006