COMMON LANGUAGE EFFECT SIZE
– A valuable step towards a more comprehensible presentation of statistical information?

Johan Lindh
Handledare: Anders Sjöberg

Undersökning och rapportskrivning – 22,5 hp, HT2018

STOCKHOLMS UNIVERSITET
PSYKOLOGISKA INSTITUTIONE
COMMON LANGUAGE EFFECT SIZE
– A valuable step towards a more comprehensible presentation of statistical information?

Johan Lindh

To help address the knowledge gap between science and practice this study explores the possible positive benefits of using a more pedagogical effect size estimate when presenting statistical relationships. Traditional presentation has shown limitations with major downsides being that scientific findings are misinterpreted or misunderstood even by professionals. This study explores the possible effects of the non-traditional effect size estimate Common Language Effect Size (CLES) on different training outcomes for HR professionals. This study also explores the possible effect of cognitive system preference on training outcomes. Results show no overall effect of CLES on either training outcomes or cognitive system. A positive effect of CLES on training outcome is found at the subfactor level showing a significant effect. The results can be interpreted that non-traditional effect size estimates have a limited effect on training outcomes. This small but valuable piece to bridge the gap of knowledge is discussed.

This study aims to contribute to the ongoing mission to bridge the gap of knowledge between science and practice within the field of industrial and organizational psychology (Aguinis, et al., 2017; Briner, & Rousseau, 2011; Cascio, 1995; Levy, 2017; Toaddy, 2015). Keeping this in mind, the core of scientific findings is often the evidence or indication of one or several relationships between different variables. In turn, these relationships are presented in a numeric format that can prove to be hard to understand and interpret (Cortina & Landis, 2009). This study asks the question; “Can the presentation of results be made more understandable with a small easy intervention?”. The aim being to contribute to the overall goal of sharing knowledge by further exploring the positive effects (Brooks, Dalal, & Nolan, 2014) of the non-traditional effect size estimate Common Language Effect Size (CLES) created by McGraw & Wong (1992). More in detail, this study examines if correlations presented using the common Pearson correlation coefficient ($r$) supplemented with CLES in a training course for psychological assessment tools will have a positive effect on the training outcomes for a group of HR professionals.

Effect size indicators

An effect size indicator is a descriptive statistic that is meant to convey a substantive interpretation of significance (Cohen, 1988). The estimate provides information regarding the magnitude of relationships and/or effects (Ellis, 2010). When reporting empirical research findings, it is considered good practice to report effect sizes (APA, 2010) in fields such as medical research and social sciences where they are often key to the interpretation of the results. Effect sizes are thus reported alongside inferential statistics as a standard part of the results (Fritz, Morris & Richler, 2012). However, there are many different ways of presenting an effect size, and several different kinds to choose from, the most commonly used being the Pearson correlation coefficient ($r$). $r$ can in short be described as a numeric estimate of a linear correlation between two different variables. The value of $r$ can vary from -1 to +1 where -1 represents a perfect negative correlation, 0 being no correlation and +1 a perfect positive correlation (Ellis, 2010). $r$ has historically been criticized for both being hard to understand without having any prior statistical knowledge as well as underestimating the true value of
effects (Cortina & Landis, 2009). The main point being that ‘small’ or ‘weak’ correlations (typically a value of $r < .2$) are often overlooked but can still provide a great value depending on the nature and context of the subject. If this problem is linked to the presentation of $r$, and more specifically how $r$ is generally interpreted, it becomes an obstacle in the process of reaching out with scientific information. It is of course a real challenge to make scientific information understandable and available to the masses or general public, in most cases though, the information only needs to reach a certain industry or a more limited number of recipients. Regardless, the core problem still remains the same if the shared information is easy to misinterpret and/or requires advanced prior knowledge, the information risk missing the target audience or recipient. Therefore, when considering the amount of research that might be beneficial for a wider audience, any intervention able to contribute towards the goal of making information available and understandable should be considered.

The sharing and availability of knowledge is key to enable both the scientific community and the general public to progress and advance. This means that a gap in knowledge between science and the public should be in both parties best interest to examine further and ultimately address to continue to develop and progress. A recent meta-analytic study by Bosco, Aguinis, Singh, Field and Pierce (2015) show that the general benchmarks established by Cohen (1988) and commonly used when interpreting $r$ (as a small, medium or large effect) do not correctly describe the findings in applied psychology. This is an example of when results presented only using $r$ seem to make knowledge is harder to understand as well as easier to underestimate and/or misinterpret. Again this poses a risk that valuable knowledge stays within the scientific community and the public or interested recipients miss out on important results. The study by Bosco et al. (2015) also indicate that experienced professionals are having difficulties interpreting $r$. This is particularly noteworthy since the presentation of the results is at the core of any scientific findings and one would be inclined to think that professionals would have no trouble at all with such a common estimate such as $r$. Considering this, an overall need to make scientific findings more accessible and understandable seem to exist, both within the community itself as well as towards other intended recipients.

The knowledge gap centered around $r$ is by no means unique for the field of applied psychology, but rather a problem that researchers face in numerous fields that base findings on similar statistical calculations. Misinterpretation of scientific results is, of course, not by any means limited to being unable to understand statistical information. It is a complex area in its own right where the presentation of results constitutes a small part of all the possible affecting factors. Within applied psychology there have been cases where the collected evidence has been strong as well as presented in a clear and understandable way, and yet the community showed reluctance towards the new findings. Meehl (1954) is a good example of this as he was one of the first within the field of psychology to criticize the clinical interpretation and prediction process used in psychiatry for its inability to incorporate statistical findings. A critique that is still valid today when applied to the area of industrial and organizational psychology. The critique has been directed at many different areas within psychology in history, although today the field of industrial and organizational psychology stand out for showing a striking reluctance for change. Addressing this problem has been the focus, mainly or partial, of several researchers (Grove & Meehl, 1996; Kleimunzt, 1990; Kuncel, 2008; Schmidt, 2006; Vrieze & Grove, 2009) throughout the years. Researchers (Highhouse, 2008; Kuncel & Rigdon, 2012) also expressed discontent towards the inability of the scientific community to communicate research findings in a comprehensible way, one of the major obstacles being translating science into practice while keeping it understandable. Highhouse (2008) argues that one of the greatest contributions of applied psychology lies within the area of selection decision aids, and that these are underutilized and undervalued by recruitment, selection and HR professionals. According to Highhouse (2008) this is partly due to the inability of scientists to communicate the
advantages of such tools in a clear and comprehensible way. Other and more critical factors include overvaluing experience and the belief that one can achieve a high, near perfect level of prediction through an intuitive assessment. A study by Priem and Rosenstein (2000) suggests that managers simply may not be aware of research findings as well as do not feel the need to further increase their knowledge regarding industrial and organizational psychology. This might seem surprising as the consequences of not being well informed and/or using psychological assessment tools (such as general mental ability and personality tests) incorrectly can be severe. Schmidt and Hunter (2000) present several examples of where failure to use intelligence tests for selecting employees led to performance declines and injuries among many other negative outcomes. One can conclude that there is indeed a big difference between the current science and how the field applies industrial and organizational psychology in everyday work life. To bridge this gap several issues and obstacles must be dealt with along the way.

To present scientific findings in an understandable way is indeed a complex problem to solve in itself that includes overcoming many obstacles. One might be tempted to try and solve them all with a single intervention but considering all the different possible scenarios and factors that affect learning and knowledge sharing, this seems implausible. A different approach is to make several smaller interventions that contribute to the overall goal. One such small intervention that might prove valuable to help bridge the knowledge gap is to complement $r$ with additional effect size measurements to help the recipient both interpret as well as better understand the result. The central idea behind this being that an individual who understands the information is more likely to learn, remember and apply it by themselves at a later stage. If the field of industrial and organizational psychology is going to change then the majority of its practitioners have to apply current science to a much greater degree than today. One way to make scientific findings more attractive is to center around the practical implications, this makes it more understandable and perceived as more relevant for the recipient, who in turn has an easier time contextualizing the new knowledge.

Focusing on the practical implications of research findings has been key in the development of additional effect size measurements since the development of $r$. The main reason behind this being the concept that a more recognizable format focusing on the practical importance of a result will be easier to interpret and understand. The Publication Manual of the American Psychological Association (APA, 2010) include the presentation of effect size measurements paired with confidence intervals (if possible) since the release of its revised 5th edition in 2001. The different available options of effect size estimates mean researchers must carefully consider which to include depending on the design and method of the study (Fritz, Morris & Richler, 2012). Thus, it has become more important for researchers to choose effect size estimates that complement the research and the study in a good way.

Common Language Effect Size indicator

In response to the critique of the traditional use of $r$ the Common Language Effect Size indicator was developed by McGraw & Wong (1992) and Dunlap (1994) contributed by extending the CLES to include bivariate normal correlations ($CLES_r$). The idea being that it would be easier for a recipient (especially nonscientists) to understand an effect size when presented as a probability, a concept more familiar to the general public including decision-makers and potential clients and/or customers. The CLES estimate would thus help bridge the gap of knowledge between the scientific community and the public as the recipient would not require prior advanced statistical knowledge to interpret the effect or result. CLES has shown to be a more understandable format than the traditional Pearson correlation coefficient when used to communicate the effectiveness of a specific intervention or training program (Brooks, Dalal, & Nolan, 2014). Worth noting though, is that few studies regarding the effects of CLES have been
conducted at this point in time. A recent Swedish study by Rapp (2018) showed promising results regarding presenting $r$ paired with CLES estimates. This spawned the idea to contribute and further investigate the results and potential of CLES in this study.

In contrast to $r$, CLES is presented as a probability which in some cases result in the need to transform continuous data to group data. For instance, working with selection this might result in splitting a variable such as job performance into two groups, above average and below average performance. The CLES probability is expressed as a probability percentage that a score (chosen randomly) from one distribution will be greater than one sampled from another distribution (McGraw & Wong, 1992). Whereas, $CL_{R}$ estimates the probability when randomly sampling two scores, that the one with a higher score on the first variable will also have a higher score on the second variable (Dunlap, 1994). As an example, a correlation of $r=.30$ would translate into roughly a 60% probability expressed as a CLES estimate. The basic idea in this study is to complement $r$ with CLES to provide a better basis for understanding and ultimately learning. By pairing the two estimates they might contribute to one another with each of their unique attributes.

The idea is that CLES will act as a complement to $r$ and help the recipient to better understand the presented statistical information. This results in the formulation of the hypothesis below:

**Hypothesis 1:** A statistical effect communicated in terms of CLES paired with $r$ in a training setting will have a positive effect on the measured training outcomes compared to when only presenting $r$.

Cognitive systems

CLES might prove to provide an important piece of the puzzle to make statistical information more understandable and relatable. However, as mentioned before CLES alone does not fully address the challenge of making advanced statistical concepts easy to understand. Reyna & Brainerd (2008) found that many adults have difficulties with decision-making when faced with ratio concepts, like risk, probability, fractions and proportions. Considering this, A CLES estimate would still be regarded as hard to understand and/or comprehend but it would still be considered easier than $r$. In addition, Gigerenzer (1994) found that probabilities presented in a numerical format (e.g., .01) where perceived as harder to understand than when presented in a frequency format (e.g., 1 in 100). If applied to $r$ and CLES it would imply that there is a difference in level of difficulty between the two estimates based on the format of presentation. Both Tversky and Kahneman (1974) and Gigerenzer (1994) also point out that when making decisions we tend to misinterpret ratio concepts and information due to a variety of irrational and biased factors affecting the judgement process. We are inclined to believe that our underlying cognitive processes regarding decision-making and judgement are both highly refined and reliable when in fact it seems to be quite the opposite. Humans struggle with rational decision-making and for us to succeed, it seemingly requires both cognitive effort and conscious action to do so (Kahneman, 2011). It would seem then that not only do people in general struggle with a numeric format, but also find it difficult to be able to fully use the available information regarding ratio concepts in a decision-making process. Considering this, it may provide a small piece of relevant information to consider when investigating the potential effects of CLES.

A possible partial explanation might be found in the dual-process theory of our different cognitive systems, described by Stanovich and West (2000) as System 1 and System 2. The theory characterizes System 1 as an automatic process that demands minimal cognitive effort and is described as being intuitive, heuristic and personal. System 2 however, demands
cognitive effort and is defined as a controlled process that is analytical, depersonalized and uses rational strategy. These two systems activate in different contexts and when facing problems and decisions, they also have different pros and cons (Kahneman, 2003). According to Kahneman (2003) both systems (System 1 often labeled as Intuition and System 2 as Analysis) cannot be active at the same time, resulting in one system being dominant in a specific situation. Changing or activating the other system would require a conscious action to do so, and with System 1 being automatic by nature, decision-making done by System 1 is hard to evaluate in the moment it occurs by the individual. The pattern and frequency of use of the different systems differ between individuals, but System 1 seems to constitute for the majority of cognitive processing (Kahneman, 2003). There have been discussions regarding whether the two cognitive systems are two opposite poles of a single dimension, or if they are two different uncorrelated systems (Hodgkinson, Sadler-Smith, Sinclair, & Ashkanasy, 2009). In a recent meta-analysis (Wang, Highhouse, Lake, Petersen, and Rada, 2017) it was concluded that Intuition and Analysis are uncorrelated. Thus, this supports the theory that the two cognitive systems work independently of each other and are not simultaneously active. This would imply that each individual automatically manages cognitive resources and energy by switching between the cognitive systems when deemed necessary depending on contextual factors.

Considering the dual-processing theory, having information regarding a preference for Intuition or Analysis might prove valuable when presenting advanced statistical information. Due to the complex nature of statistics, one would be inclined to think that an active System 2 (Analysis) would be preferable when attempting to acquire new advanced knowledge. Especially if one is to evaluate the possible value of a probability, for instance, a CLES estimate. As mentioned before, to determine the ‘true’ value based on a probability estimate can be a hard task in itself. In the study by Brooks, Dalal and Nolan (2014) the different tasks presented to the respondent are centered around evaluating effectiveness when it is presented in different ways. This current study does not involve any element of evaluation of value or any similar task that needs to be completed by the participants. Unfortunately, the individual’s active usage of the cognitive systems cannot be controlled in this particular quasi experimental setting. It is however, possible to gather information regarding whichever system is more dominant or preferred by an individual. To be able to measure cognitive system preference paired with further examination into the possible effects of CLES, it was concluded that this study needs two different instruments to do so. One to evaluate the training intervention and one to assess the individual’s cognitive system.

**Cognitive systems effects of training**

Training programs and courses for organizations and employees has shown to have several benefits such as keeping up with modern technology, organizational performance and innovation (Aguinis & Kraiger, 2009). Researchers (Giargreco, Carugati & Sebastiano, 2010; Grossman & Salas, 2011) found that professional training is critical to organizational success. Considering these effects, it is easy to see the appeal for companies to invest in training programs and courses for its employees and why it is so common (Aguinis & Kraiger, 2009; Grossman & Salas, 2011). Trainings within the field of industrial and organizational psychology as well as HR are no exception, with an abundance of different training programs and courses available. The particular training within the scope of this study contains, as mentioned before, advanced statistical information that the participants might be unfamiliar with, both in content and format. Not being able to understand or prior knowledge being required might become obstacles that ultimately have a negative impact on the training outcome. One would then be inclined as a trainer to achieve an optimal setting to enable the positive effects of training, especially when its contents are at advanced level or above. Considering both the advanced content and the possible effect of the training, as well as the two cognitive systems, it would seem reasonable to believe that the cognitive system Analysis could
relate to a positive training effect. This due to the analytic and rational components of Analysis and the training that focus on acquisition of new knowledge and the application of it respectively. The reasoning behind this possible relationship results in the formation of the second hypothesis below:

**Hypothesis 2:** The cognitive system Analysis will have a positive effect on the measured training outcomes.

**Method**

**Participants**
For this study it was deemed relevant to approach HR professionals to participate since they represent a key group when it comes to putting theory into practice. HR professionals also represents the first line recipient of scientific findings within the area of industrial and organizational psychology. To be able to target this audience a private company was approached, the company provide HR and recruitment trainings and are based in Stockholm, Sweden. The company agreed to cooperate and help with the collection of data for this study. The company is well-known in the field of organizational and industrial psychology in Sweden and carry out several different training programs and courses. The particular training course chosen for intervention in this study is an entry-level training course in selection and the use of recruitment tools, such as personality and general mental ability tests. Some of the participants may have prior statistical knowledge but the training course itself assumes no prior knowledge on the subject.

A total of 96 HR professionals participated in the study, one participant was excluded due to incomplete data. The majority were women (n=68) and one participant reported ‘other’ regarding gender. All (n=96) were actively working with recruitment and selection at the time of participation in the study. Reported work experience in the current role was distributed followingly; Less than 1 year, 30,2%; 1-2 years, 24%; 3-4 years, 26%; and 4 or more years, 19,8%. Highest level of completed education was distributed across three levels; Middle/junior high or high school, 5,2%; Less than 3 years of post-secondary education (college, university), 64,6%; and lastly 3 or more years of post-secondary education (college, university), 30,2%. All participants signed up for a training course centered around psychological assessments and recruitment as a part of further training in their profession as HR specialists. Groups of participants were randomly assigned to one of the two groups CLES-group (experiment group, n=45) or r-group (control group, n=51).

**Design and procedure**
This study used a between-groups design where the two groups of individuals complete a training course in selection and recruitment tools that contains statistical information. The course was provided by a private company based in Stockholm, Sweden. The trainings were conducted during both the spring and fall of 2018. Groups of participants have signed up for the training to further enhance their knowledge as HR professionals. The content of the two-day training course is centered around a 155-page PowerPoint presentation identical for the two groups apart from one slide where Schmidt and Hunter’s (1998) research findings are presented in a table. The table contains both the type of test or intervention as well as the r value showing its correlation with prediction of job performance. In the CLES-group the table containing r estimates also show corresponding CLES estimates calculated using Dunlap’s (1999) method (i.e. GMA; r=.51; CLES 67%). The trainer also explains in short how to read the CLES estimate, the same being done for r in both groups (since no prior statistical knowledge is assumed). The participants both listen to the presentation and get hands-on training regarding
the different assessment tools during the two-day period. Participants who signed up for training were randomly assigned to one of the two different conditions. The Questionnaire for Professional Training Evaluation (Q4TE, supplemented with additional cognitive system items) pen-and-paper survey is presented by the instructor upon completion of the training. Two different instructors performed the training on the different occasions and they both used the same identical presentation.

Ethics and privacy
The quasi experimental design of this study involves an intervention that the participants are unaware of when signing up for the training course. To withhold information towards the participants can be considered unethical, and considerations have been made with this in mind when designing this study. There are two main cases to consider; (1) the participants are unaware of the study and the experiment when signing up for the training course; and (2) the participants are unaware of the intervention during the training course. In both cases it was deemed acceptable to withhold information due to the intervention being minor, non-intrusive as well as participation in the study being voluntary and the individual can easily opt out. Transparency, voluntariness, anonymity and the participants privacy are upheld by both the trainer and the survey itself stating that participation is optional, no personally identifiable information (either by itself or in combination) is gathered, and lastly that the reason for collecting data is a part of a master thesis in Psychology at Stockholm University. Participants were also informed that any questions regarding the study can be directed to the trainer or the author of this study.

Measures
The Questionnaire for Professional Training Evaluation (Q4TE) created by Grohmann & Kauffeld (2013) paired with ten additional items (measuring cognitive system preference) has been chosen to cater this study’s needs. The ten additional items originate from the General Decision-Making Style Inventory, GDMS by Scott and Bruce (1995) and have been adapted to Swedish along with the Q4TE. The Swedish Q4TE inventory adaptation done by Rapp (2018) serves as the tool for evaluating training outcomes and is applicable across different contents and organizations (Grohmann & Kauffeld, 2013). The Q4TE is based on Kirkpatrick’s four level framework (Kirkpatrick & Kirkpatrick, 2006) that is widely used for summative evaluation (Blau, Gibson, Bentley & Chapman, 2012). The four levels of the framework consist of: (1) Reaction, the participants’ emotional reaction to the training; (2) Learning, gaining procedural, methodological and expert knowledge along with an attitude change through training; (3) Behavior, the application of the contents of the training; and (4) Results, the procedural, methodological and expert knowledge along with an attitude change through training. The Q4TE four-factor structure (Reaction, Learning, Behavior and Results) was deemed to be helpful when analyzing the results of the CLES intervention as well as any potential relationships regarding cognitive style. The Q4TE also provides a possibility for repeated measures to evaluate the effects of a training intervention over time and if it led to change in behaviors (Grohmann & Kauffeld, 2013). The Swedish adaptation of the Q4TE inventory by Rapp (2018) (Reaction, α=.845; Learning, α=.591; Behavior, α=.672; Results, α=.651) was used to evaluate the training course paired with the 10-item questionnaire based on GDMS by Scott and Bruce (1995) also adapted to Swedish. All items in the survey, 21 in total, follow a standard 5-point Likert scale item design (1=Strongly disagree; 5=Strongly agree). The items of the Q4TE are all related to the training and a sample item (belonging to Learning in this case) would be “After the training, I know substantially”. Whereas the cognitive system items are framed in a recruitment context, for instance; an item belonging to Intuition, “I believe that it is important to trust your gut feeling when hiring new people”; and one belonging to Analysis, “When assessing an individual’s suitability for work it is important to use methods that have been proven scientifically.”
Data analysis
A multivariate analysis of covariance (MANCOVA) was conducted with group as an independent variable, cognitive systems as a covariate, and reaction, learning, behavior and results as dependent variables. MANCOVA extends the capabilities of analysis of variance (ANOVA) by assessing multiple dependent variables simultaneously and the multivariate analysis could also combine both categorical and continuous independent variables. Statistics to analyze the different relationships between the two groups (CLES and \( r \)), cognitive system (Intuition and Analysis) and the four factors of the Q4TE. Cohen’s \( d \) (Cohen, 1988) was also calculated as a supplement to the results of the MANCOVA.

Results
The quasi experiment of this study was designed to further explore the possible positive effects of supplementing \( r \) with CLES, as well as a possible relationship between a training outcome and cognitive system among HR professionals. The 96 participants were distributed between the two groups \( r \) (n=51) and CLES (n=45). The descriptive statistics for the different groups (\( r \) and CLES), Q4TE factors, Intuition and Analysis supplemented with Cohen’s \( d \) are presented in Table 1. The mean values for the CLES group showed slightly increased values for all Q4TE factors compared to the \( r \)-group, the increased mean values of the CLES group range from .1 to .25.

The MANCOVA analysis conducted was aimed at answering both of the study’s hypotheses. The first hypothesis (H1: A statistical effect communicated in terms of CLES paired with \( r \) in a training setting will have a positive effect on the training compared to when only presenting \( r \)) yielded non-significant results (F(4, 91)=1.221, \( p=.307 \), Wilks’ \( \Lambda =.949 \), \( \eta^2=.051 \)) for the overall model. In order to estimate the effects of the individual factors at the univariate level further analysis was conducted that showed one significant effect (F(1, 94)=4.676, \( p=.033 \), \( \eta^2=.047 \)) between group and the factor Results (related to the organizational impact of the training). The results of the MANCOVA for all four training factors and group (\( r \) and CLES) are presented in Table 2. The low level of explained variance indicates that the CLES intervention in the experiment group had a very small, if any, effect. These results indicate that the first hypothesis of this study could not be supported fully.

To answer to the second hypothesis (H2: The cognitive system Analysis will have a positive training effect) the MANCOVA analysis conducted aimed to explore the relationships between Intuition, Analysis and a possible training effect. The analysis showed non-significant results for the overall model regarding both Intuition (F(5, 90)=2.296, \( p=.052 \), \( \eta^2=.113 \)) and Analysis (F(5, 90)=2.117, \( p=.071 \), \( \eta^2=.105 \)). At the univariate level one significant positive effect was shown between Intuition and the factor Learning (F(1, 94)=5.128, \( p=.026 \), \( \eta^2=.054 \)). The complete results of the MANCOVA analysis regarding cognitive system and the training factors are presented in Table 3. The results suggest that the individual’s cognitive system does not have an overall effect regarding the training. These results do not show support for the second hypothesis of this study.
Table 1: Descriptive statistics and Cohen’s d for the different factors of the Q4TE, Intuition and Analysis (r, n=51; CLES, n=45; Intuition, N=96; Analysis, N=96).

<table>
<thead>
<tr>
<th>Group</th>
<th>Mean</th>
<th>SD</th>
<th>Cohen’s d</th>
<th>Lower bound</th>
<th>Upper bound</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Cohens’s d</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>95% Confidence Interval</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reaction</td>
<td>4.2647</td>
<td>0.49095</td>
<td>0.205</td>
<td>-0.197</td>
<td>0.607</td>
</tr>
<tr>
<td>CLES</td>
<td>4.3667</td>
<td>0.50452</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Learning</td>
<td>4.2157</td>
<td>0.82450</td>
<td>0.152</td>
<td>-0.249</td>
<td>0.554</td>
</tr>
<tr>
<td>CLES</td>
<td>4.3444</td>
<td>0.86171</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Behavior</td>
<td>4.2451</td>
<td>0.47300</td>
<td>0.251</td>
<td>-0.151</td>
<td>0.653</td>
</tr>
<tr>
<td>CLES</td>
<td>4.3778</td>
<td>0.58539</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Results</td>
<td>3.8824</td>
<td>0.56104</td>
<td>0.442</td>
<td>0.036</td>
<td>0.848</td>
</tr>
<tr>
<td>CLES</td>
<td>4.1333</td>
<td>0.57472</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intuition</td>
<td>1.4042</td>
<td>0.52071</td>
<td>5.994</td>
<td>5.331</td>
<td>6.657</td>
</tr>
<tr>
<td>Analysis</td>
<td>4.4828</td>
<td>0.50642</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2: Multivariate analysis of covariance between the two experiment groups, r and CLES, and the Q4TE factors (r, n=51 and CLES, n=45).

<table>
<thead>
<tr>
<th></th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
<th>η²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reaction</td>
<td>0.249</td>
<td>1</td>
<td>0.249</td>
<td>1.005</td>
<td>0.319</td>
<td>0.011</td>
</tr>
<tr>
<td>Learning</td>
<td>0.396</td>
<td>1</td>
<td>0.396</td>
<td>0.556</td>
<td>0.458</td>
<td>0.006</td>
</tr>
<tr>
<td>Behavior</td>
<td>0.421</td>
<td>1</td>
<td>0.421</td>
<td>1.506</td>
<td>0.223</td>
<td>0.016</td>
</tr>
<tr>
<td>Results</td>
<td>1.506</td>
<td>1</td>
<td>1.506</td>
<td>4.676</td>
<td>0.033</td>
<td>0.047</td>
</tr>
</tbody>
</table>
Table 3: Multivariate analysis of covariance between the cognitive systems Intuition and Analysis and the Q4TE factors (N=96).

<table>
<thead>
<tr>
<th>Cognitive System</th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
<th>$\eta^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall model</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intuition</td>
<td>2.914</td>
<td>5</td>
<td>0.583</td>
<td>2.296</td>
<td>0.052</td>
<td>0.113</td>
</tr>
<tr>
<td>Analysis</td>
<td>2.564</td>
<td>5</td>
<td>0.513</td>
<td>2.117</td>
<td>0.071</td>
<td>0.105</td>
</tr>
<tr>
<td>Reaction</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intuition</td>
<td>0.421</td>
<td>1</td>
<td>0.421</td>
<td>1.659</td>
<td>0.201</td>
<td>0.018</td>
</tr>
<tr>
<td>Analysis</td>
<td>0.198</td>
<td>1</td>
<td>0.198</td>
<td>0.819</td>
<td>0.368</td>
<td>0.009</td>
</tr>
<tr>
<td>Learning</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intuition</td>
<td>1.302</td>
<td>1</td>
<td>1.302</td>
<td>5.128</td>
<td>0.026</td>
<td>0.054</td>
</tr>
<tr>
<td>Analysis</td>
<td>0.469</td>
<td>1</td>
<td>0.469</td>
<td>1.937</td>
<td>0.167</td>
<td>0.021</td>
</tr>
<tr>
<td>Behavior</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intuition</td>
<td>0.363</td>
<td>1</td>
<td>0.363</td>
<td>1.429</td>
<td>0.235</td>
<td>0.016</td>
</tr>
<tr>
<td>Analysis</td>
<td>0.065</td>
<td>1</td>
<td>0.065</td>
<td>0.269</td>
<td>0.605</td>
<td>0.003</td>
</tr>
<tr>
<td>Results</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intuition</td>
<td>0.573</td>
<td>1</td>
<td>0.573</td>
<td>2.258</td>
<td>0.136</td>
<td>0.024</td>
</tr>
<tr>
<td>Analysis</td>
<td>0.227</td>
<td>1</td>
<td>0.227</td>
<td>0.936</td>
<td>0.336</td>
<td>0.010</td>
</tr>
</tbody>
</table>

Discussion

The main purpose of this study was to further investigate the possible benefits of CLES as a more understandable and pedagogical format for presentation of statistical findings. The secondary purpose was to explore if the cognitive systems Analysis might have an effect regarding the different training settings ($r$ only or $r$ paired with CLES) containing statistical information. Due to the gap in knowledge between science and practice within the field of industrial and organizational psychology (Grove & Meehl, 1996; Highhouse, 2008; Kleinmuntz, 1990; Kuncel, 2008; Kuncel & Rigdon, 2012; Schmidt, 2006; Vrieze & Grove, 2009), active HR professionals were chosen as the target sample for this study. A private company based in Stockholm, Sweden providing training in recruitment and selection cooperated regarding the data collection. The groups of participants were assigned randomly to either the standard training or one where the presentation of $r$ estimates was supplemented with CLES estimates. Upon completion of the training the participants were asked to fill in the Q4TE (Grohmann & Kauffeld, 2013), a training evaluation coupled with an adapted GDMS (Scott and Bruce, 1995) questionnaire measuring cognitive systems (Intuition and Analysis). The data was analyzed using a multivariate analysis of covariance and showed no significant results for any of the overall models and thus, no support for any of the study’s two hypotheses; (1) A statistical effect communicated in terms of CLES paired with $r$ in a training setting will have a positive effect on the training compared to when only presenting $r$; (2) The cognitive system Analysis will have a positive effect on the Q4TE factors Learning and Behavior. However, two significant effects on training outcome was found at the univariate level, the first being related to the Q4TE factor Results and the different groups using $r$ and CLES, and the second being between the Q4TE factor Learning and the cognitive system Intuition. The results of this study show no support for the overall models, looking closer at the results however, it might still be premature to dismiss possible effects and benefits of the CLES estimate.

The results regarding the cognitive systems Intuition and Analysis in relation to training outcomes are more ambiguous in the sense that the results do not reflect the theoretical assumption. It is possible that the results are a consequence of the Swedish adaptation. Another possible explanation might be that any underlying effect connected to the training only appears
over time when it has been fully integrated by the participant. In this study the participants filled in the Q4TE questionnaire upon completion of the training and no follow-up evaluation has been conducted. Regardless, any further investigation into this possible relationship will most surely benefit from a robust and high-quality instrument when assessing the cognitive systems.

Focusing on the first hypothesis and considering the results of CLES not showing any overall effect in this study, as opposed to the clear positive effect shown by Brooks, Dalal & Nolan (2014), one of the major differences is the design of the studies. The Brooks, Dalal & Nolan (2014) study focused heavily on the presentation and evaluation of different non-traditional effect size estimates in a very hands-on context for the participants. This is, as mentioned before, quite different from the design of this study that used a very minimal intervention and also a different sample (HR professionals vs undergraduate students). Choosing active HR professionals as a group to participate in the study was deemed highly relevant considering this study’s overall aim to help bridge the knowledge gap within industrial and organizational psychology. Being a sample of convenience, it has both pros and cons, but considering this it is important to conduct research in the field as well as in a more controlled traditional experimental setting. It is also likely that research conducted using a more representative sample in the sense of active work professionals, can be perceived as more relevant when presented to the non-scientific community. This aligns with the overall aim to bridge the gap between science and practice by combating the critique of scientific experiments and settings not accurately representing the “real world”.

In the absence of a large CLES related effect, one can also note that it did not have a negative impact. This might sound obvious, and the fact that it does is a very good argument to why it should be mandatory to present appropriate effect size estimates as a key part of any findings. It is a small piece of information that we still do not know how much of an impact it has but we know it does not have a negative effect when it comes to result presentation. This is of course not limited only to CLES but rather something that could be applied to how all non-traditional effect size estimates are treated. A small piece that provides value and requires no extra effort to include in the presentation, that is how the effect size estimate should be considered and used. A new set of recommendations and standards could be very helpful as a starting point. The scientific community (and the industrial and organizational psychology field especially) need to do everything possible to reach out to the “real world” of practice. As mentioned before, many small interventions might prove to provide great value when combined to create a more pedagogical format, something science should actively encourage and strive towards. The collective pursuit of knowledge and the work done by scientists should be promoted and shared with those who benefit from it. This is, perhaps, a utopian idea but it can surely be helpful nonetheless to provide direction for the continuous development. Fact is that the current situation in the world of HR practice is dire in many cases, and beliefs falsified by science a long time ago still thrive. A study conducted by (Rynes, Colbert, & Brown, 2002) tell a grim story of common beliefs regarding effective practices, one of them being using graphology as an instrument for selection with the belief that it is a reliable and valid method. In the end it is the job applicants and candidates that suffer, as well as the companies themselves for failing to apply knowledge that in some cases, have been proved and well documented for several decades. As one of the foundations of science is to spread knowledge as well as a mission that should be actively pursued by the scientific community. In the field of industrial and organizational psychology, sadly, it is needed more than ever.

This study aimed to help bridge the knowledge gap between science and practice, albeit with a very small contribution. There are indeed indications that non-traditional effect size estimates have merit and provide something unique that is needed as we move forward. Not only to make scientific findings more understandable for recipients outside of the scientific community but
within it as well. The findings of Bosco et al. (2015) remind us that we have work to do regarding the evaluation and interpretation of both traditional methods as well as new. The continuous rapid technical development within the scientific field might provide new ways to present data. But until then the non-traditional effect size estimates can provide part of that function. There is no doubt that bridging the knowledge gap requires several interventions and efforts. Researchers Rynes, Bartunek, and Daft (2001) suggest (among several things) that the scientific community should approach practice in a very real way by engaging in meetings, summits and cross-field forums in a more active way. This might provide a very valuable channel of communication between the fields that indeed seems to be needed. At the moment the field of practice seem to be in a state of stagnation, stuck in relying on old methods. The key missing component from that perspective being change. In turn an important part of change is acquiring new knowledge and applying it i.e., learning. A certain degree of openness and curiosity also help of course and perhaps practice is slowly going to get there if they discover the benefits. Pfeffer and Sutton (2000) provide insights on the subject of the knowledge gap in terms of knowing but failing with the application of that knowledge. Guidance and concrete application are a conceivable mission that the scientific community can help with going forward. One such thing is providing literature that is easier to understand and more easily available. Cooper and Locke’s (2000) handbook for HR practitioners is such an example that could be yet another important piece to bridge the ever-present gap of knowledge.

**Limitations**

In this study two main limitations have been discovered during the work process. The first is that there is a degree of uncertainty connected to the Swedish adaptations of the measuring instruments (both Q4TE and GDMS), especially the Q4TE show low internal consistency, and secondly, the size of the intervention in this quasi experiment design being quite small (1 out of 155 slides in the training contain CLES estimates). It is possible that these factors have a negative effect on the positive effects of CLES showed by Brooks, Dalal & Nolan (2014). Another possibility is of course, simply that the effect of CLES (all or most of it) is lost when it is not the subject of the main focus, but rather a small part of a larger context. One way to answer this question could, for example, have been to ask the participants in the experiment group how they perceived the CLES estimate after the completed training.

**Suggestions for further research**

Although no support was found in this study that CLES is perceived as more understandable format for presenting a correlation, non-traditional effect size estimates definitely have merits that should be the subject of further research and development. There is the potential of helping bridge the knowledge gap by turning something abstract and advanced into something that is perceived as tangible and valuable. Since there are several different non-traditional effect size estimates further research into which (if any) show a greater effect or merit than others. The area of sharing and making knowledge available is key for the future of the field of industrial and organizational psychology and every little bit that helps is worth looking into further.

**Conclusions**

In summary, with the overall aim to lessen the knowledge gap present between science and practice in the field of industrial and organizational psychology, this study set out to investigate the potential positive effects of CLES in a training setting for HR professionals. A contribution focused on investigating if the non-traditional effect size estimate might provide a valuable piece towards making scientific findings more understandable. Although this study found no support for the overall model, the most plausible explanation for this is presumably the design and especially the minor intervention. For the effect to reach the level where it can be measured reliably, it definitely needs more exposure. Future research is recommended to focus on the intervention and to give it a more central role in the design. For example, one could give HR
professionals tasks to evaluate different statements similar to those of the Brooks, Dalal & Nolan (2014) study centered around perceived value, understandability and effectiveness. Another could be to explore in-depth how understandable a recipient perceives the different available non-traditional effect size estimates. Effect sizes as a subject show promise and merit going forward, and this study aimed to contribute to the research into the effects of the non-traditional effect size estimate CLES in a field study. Further research, both lab and field studies, are needed to draw any final conclusions regarding the positive effects of non-traditional effect size estimates. Further research is deemed valuable for both science and practice.

References


Gigerenzer, G. (1994). Why the distinction between single-event probabilities and frequencies is important for psychology (and vice versa). In G. Wright & P. Ayton (Eds.), *Subjective probability* (pp. 129–161). New York, NY: Wiley.


