

Generalization of Results in Psychological and Educational Research: From Quality Conditionality to Statistical Ritualism

Chaib Mahfoud ^{1*}, Mansouri Yahia ²

¹University of Algiers 2 – Abou El Kacem Saâdallah – (Algeria), mahfoud.chaib@univ-alger2.dz

² University of biskra – Mohamed Khider – (Algeria), yahia.mansouri@univ-biskra.dz

Received: 28 / 09 / 2025

Accepted: 22 / 11 / 2025

Published: 30 / 11 / 2025

Abstract:

This article examines the critical shift in the process of generalizing results in psychological and educational research from being a hallmark of research quality to becoming a ritualistic practice based solely on statistical significance. Such reliance is often pursued without meeting the methodological requirements that should precede statistical generalization and without giving due attention to practical significance (effect size and correlation magnitude). The paper further identifies key gaps frequently overlooked by researchers, illustrating with examples the limitations of statistical significance in capturing the realistic and practical meaning of research findings. In addition, it analyzes the risks posed by certain statistical inferences that may lead to decisive outcomes in research contexts, particularly when extreme scores influence psychological and educational interventions

Keywords: Quality; Statistical Significance.

*Corresponding author

1. Research problem

Generalization is one of the foundational concepts in scientific research, as it plays a central role in building theories and ensuring their applicability across diverse contexts. It represents a systematic process through which findings derived from experiments and studies conducted on relatively small samples are transformed into broader conclusions that can be extended to larger populations or different settings. This process enables researchers to better understand human behavior, mechanisms of learning, and other psychological and educational phenomena moving from narrow contexts to more inclusive ones.

The generalizability of results in psychological and educational research is thus regarded as a key dimension of scientific validity. As Hammersley (1999) emphasized, scientific theories lose much of their significance if they cannot be transferred beyond the specific situations in which they were initially developed. Similarly, Howitt and Cramer (2016) argue that psychology is inherently oriented toward generalization, since psychological findings are presumed to apply across individuals regardless of their societies or environments. In this sense, psychological principles are treated as universal laws of human behavior.

Linguistically, *generalization* is defined as issuing a universal judgment derived from specific cases (Contemporary Arabic Language Dictionary, 2008; cited in Ayesha, 2016). Methodologically, Mayring (2007) views it as the move from the particular to the general, or from specific observations to broader laws and formulations—a process that every psychological and educational theory seeks in order to secure external validity and withstand falsification.

In practice, statistical generalization in psychological and educational sciences typically occurs through the rejection of the null hypothesis, thereby supporting the alternative hypothesis and confirming the presence of a relationship or difference between variables. This allows the researcher to generalize the findings statistically to the research population, essentially completing the process of scientific induction. However, the act of induction extends far beyond formal science. From a cognitive psychology perspective, Kahneman and Tversky's work shows that humans naturally rely on heuristics—mental shortcuts based on generalization—to save cognitive effort. While efficient, these heuristics frequently lead to errors in judgment. Everyday reasoning provides clear examples: assuming that all children are mischievous, that mathematics is always difficult, or that all politicians are corrupt. Such sweeping

generalizations are often grounded in a few anecdotal cases and can easily result in biased or misleading conclusions.

This illustrates the danger of hasty generalizations: limited examples cannot provide a reliable foundation for sound theory. Populist generalizations, in particular, often rest on unstable grounds, relying more on estimation than measurement. By contrast, generalization rooted in rigorous scientific and methodological foundations remains indispensable for safeguarding the credibility of knowledge. On such a basis, prediction—considered a higher-order scientific goal than description—can be achieved. Furthermore, research should not be confined to generalization as an end in itself. It should progress toward informed decision-making that is effective, efficient, and minimizes the risks of error, while also providing sound explanations for relationships between variables—whether descriptive, predictive, or causal. In this way, scientific knowledge advances cumulatively and coherently.

It is worth noting, however, that some psychologists still advocate the study of single cases or relatively small samples to explore human behavior in depth, as commonly practiced in qualitative research (Howitt & Cramer, 2016). Such approaches contribute to quality in psychological research but are often accompanied by caution regarding the generalization of findings, even under strict methodological conditions.

Ultimately, the quality of any scientific study is compromised when it fails to meet the conditions of generalization. This process cannot be reduced to statistical significance alone but must rest on a broader set of methodological and statistical requirements. Unfortunately, many researchers continue to rely excessively on significance testing as the sole benchmark for generalization, neglecting the contexts and conditions that give findings their true meaning. This paper, therefore, seeks to highlight this problematic shift: the transformation of generalization from a quality criterion into a ritualistic practice overly dependent on statistical significance.

2. Key Concepts

2.1. Scientific Research

Babbie (2020), in *The Practice of Social Research*, defines scientific research as “a systematic intellectual process of collecting and analyzing data in a methodical way to reach scientific conclusions, focusing on accuracy, replication, and objectivity to ensure the reliability of results.” In contrast, Van Dalen

describes it as “the careful and critical attempt to find solutions to the problems that trouble and perplex human beings”.

For Popper (1959), scientific research is conceived as a continuous process of testing theories and seeking to falsify them rather than confirm them (Zeitouni, 2020).

From these definitions, it becomes evident that scientific research presupposes the existence of a specific problem that motivates the researcher to investigate it in an organized and systematic manner. Its quality and effectiveness can be assessed through the rigorous application of the scientific method, aiming to explain the problem and ultimately arrive at facts and solutions.

2.2. Generalization

The *Oxford Dictionary* (1989) defines **generalization** as the process of forming general concepts derived from specific cases, the capacity of such concepts to be generalized, or a scientific fact that can be generalized.

In terminological usage, **generalization** refers to the process by which conclusions drawn from the observation of a limited number of cases are extended to encompass all other unobserved cases that fall within the same category (Al-Annsi, 2024).

2.3. Quality

Quality is defined as “a set of conditions and specifications that must be fulfilled in the educational process in order to meet the needs of its beneficiaries and to produce outcomes characterized by competence in addressing the requirements of society.”

2.3.1. Criteria of Research Quality

High-quality scientific research is distinguished as a systematic process that seeks truth in order to provide appropriate solutions to scientific, social, or applied problems. It is also a logical process in which the researcher advances step by step toward solving a problem in a coherent and consistent manner.

Moreover, it is an empirical process reflected in its observations, procedures, implementation, and the application of results. Research is reliable, repeatable, and expected to yield similar outcomes. In this sense, it constitutes a process aimed at updating, refining, or expanding human knowledge.

It is further characterized by objectivity, seriousness, and caution, requiring the researcher to possess significant expertise in planning, conducting, and evaluating the study, while striving to minimize subjectivity as much as possible (Abdullah & Laaradj, 2023).

Table (1)

Research Quality Criteria according to Gurol Yokus & Hakan Akdag (2002)

Quality Condition	Elements
Originality	Research title; Research problem; Results and analysis
Scientific Rigor	Appropriate methodology; Adherence to ethics; Avoidance of bias
Comprehensiveness	Use of recent studies; Consideration of supporting and opposing views; Use of primary studies
Presentation	In-text citations and references; Reporting statistical values; Quality of language
Contribution	Filling a gap in the literature; Offering practical solutions; Scientific impact
Relevance	Suitability of research population; Appropriateness of data collection tools; Relevance of statistical analysis

Source: (Abdullah & Laaradj, 2023 ; Gurol & Hakan, 2002, pp. 516-527).

2.4. Statistical Significance

Cohen (1994) defines statistical significance as *“the probability that the effect observed in the studied sample is real and not due to chance”*. According to Wasserstein and Lazar (2016), in their statement for the American Statistical Association (ASA), statistical significance refers to the likelihood that a result is due to random variation.

In summary, statistical significance can be understood as a mathematical indicator used to determine whether research results are genuine or merely the product of chance, and thus whether they can be generalized to the statistical population. Conventionally, if the p -value is less than 0.05, the null hypothesis is rejected, indicating that the result is statistically significant (Neyman & Pearson, 1933). However, this threshold does not necessarily imply that the results are practically important or strong; it only signals the presence of a statistically testable relationship.

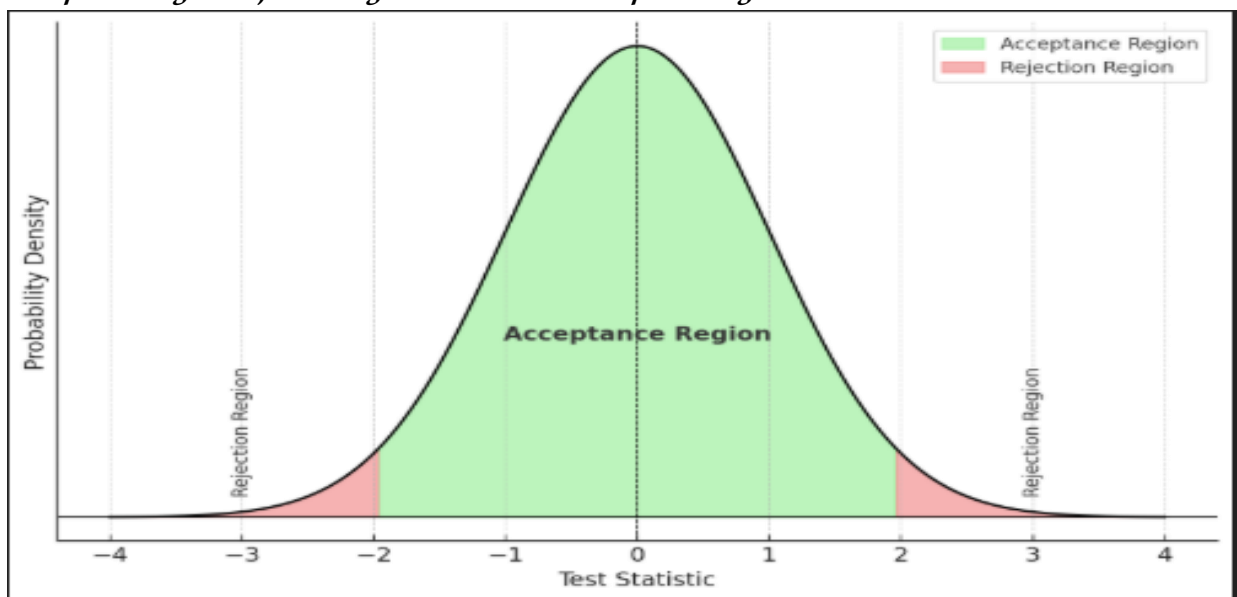
It should also be emphasized that statistical power analysis is a complementary procedure to significance testing, as it estimates the probability of committing a Type II error (accepting the null hypothesis when it is actually false). Thompson defined power as *“the probability that a test will correctly reject a false null hypothesis.”* Yet, as Quetelet and Shaver observed, this approach is rarely applied by researchers, despite the fact that it incorporates essential considerations such as effect size and sample size. Thompson further argued that when statistical power is high, even non-significant results can still contribute meaningfully to the accumulation of knowledge. Similarly, Lay maintained that any researcher who accepts the null hypothesis without examining the power of the test assumes responsibility for the risk of committing a Type II error.

Effect size also referred to as practical significances another critical factor directly influencing statistical power. It represents the numerical difference between the true population parameters and those specified under the null hypothesis. As a key element of the effect size index, it is closely tied to statistical power: the greater the effect size, the higher the probability of detecting a true effect.

In this sense, effect size is the other side of statistical significance. While significance indicates whether a difference exists, effect size reflects the magnitude and practical usefulness of that difference. Therefore, statistical significance alone is insufficient for sound scientific decision-making; researchers must also calculate practical significance by estimating the strength of the relationship between dependent and independent variables.

Figure (1)

A representing the rejection region of H^0 and the acceptance region of H^0



The diagram of acceptance and rejection regions for H_0 illustrates that statistical significance appears at the extreme values (tails) of the distribution, whether in a one-tailed or two-tailed test. If the calculated value falls within the acceptance region, the null hypothesis is retained, indicating no relationship between the variables. This demonstrates that statistical significance is essentially linked to rarity and extremity, since significant values occur in the distribution's tails, far from the mean of the sample distribution, a concept captured by the standard error.

3. Manifestations of the Ritualistic Use of Statistical Significance

3.1. The Manifestation of Non-Random Sampling (Pseudo-Randomness)

The best strategy for achieving statistical generalization is to rely on random sampling (Polit et al., 2010). Random sampling ensures that each member of the statistical population has an equal and independent chance of being included in the study. If the sample is not selected randomly, a fundamental principle of inferential statistics is violated, and any generalizations made to the population will lack validity (Abu Allam, 2006). Statistically, random sampling guarantees the independent distribution of errors under treatment conditions, allowing us to estimate the margin of chance error during sampling. By contrast, non-random sampling does not allow the researcher to determine the margin of error in the study.

In practice, however, researchers often fail to apply true random sampling, whether intentionally or unintentionally, while still claiming to fulfill the condition of randomness and ensure a fair distribution of sampling errors. Yet, upon closer inspection of their methods, flaws often emerge that make generalization to the population impossible. It is important to remember that the population is the true target of research, since samples are not studied for their own sake but rather as a means to investigate the characteristics of the broader population. As Abu Allam (2006) emphasized in an article cited by Kroum and Kroum (2022): "If research findings applied only to the group under investigation, we could never benefit from previous studies in any field. Imagine how slow the advancement of knowledge would be if that were true".

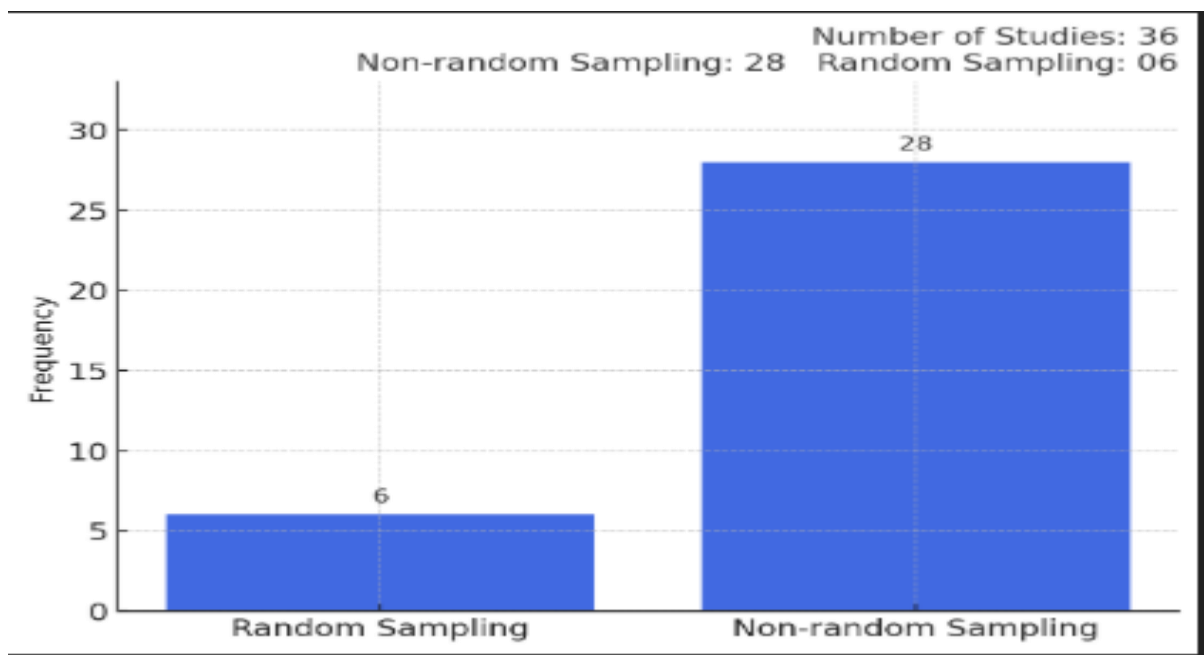
Still, Abu Allam's acknowledgment of the importance of generalization remains incomplete unless embodied in strict methodological techniques. This serves as a reminder for all researchers, including graduate students writing theses, that they are not immune to this trap. For example, in a study conducted by Ayesha (2016) on the problems of generalization in psychological and educational research, covering 95

master's theses and doctoral dissertations, the findings revealed that 28 of them applied random sampling incorrectly. She concluded:

"When reviewing the criteria followed in sampling methods, we find that so-called 'random' sampling did not actually follow scientific procedures. In the remaining theses, what was labeled as random sampling was, in reality, accidental sampling". This can be illustrated in the following figure:

Figure (2)

Bar chart showing the frequency of correct vs. incorrect random sampling



From the chart, we observe that out of 36 studies claiming to use random sampling, only 6 were methodologically valid. Ayesh carried out deeper analysis by examining the conditions that these theses failed to meet. She identified major shortcomings, such as the failure to retrieve all distributed questionnaires, which undermines randomness and thus generalization. This connects to what she described as a "*methodological necessity*"—namely, that all selected participants must actually participate in the study without attrition. She argued:

"Although it is reasonable to expect that not all distributed questionnaires will be returned, the researcher should anticipate this issue in advance. For example, if a researcher wants a final sample of 200 participants assumed to represent the population, it is preferable to distribute to a larger number so that excluding invalid or unreturned questionnaires does not jeopardize randomness".

With respect to the requirement of clearly defining the research population, the study further revealed that 13 theses failed to specify the population from which the sample was drawn, neglecting essential characteristics such as age, gender, or social status. Since defining the population is the main step preceding random sampling, Ayesh attributed this failure to the researchers' weak methodological background. From a psychometric perspective, one could argue:

"The more homogeneous the population, the greater the study's capacity for generalization, provided that the sample is truly representative of that population".

Shaver (1992), cited by Bahash (2023), reinforced this point: "The necessity of randomness becomes even clearer when we ask: To what extent does the sample statistic represent the population parameter as defined in the null hypothesis? Without randomness, we cannot truthfully answer this question using standard significance tests".

As Ayesh (2016) further emphasized: "Sampling must begin from the characteristics of the population to which we intend to generalize. Yet many researchers ignore this and jump directly to choosing a sample by convenience, and only later attempt to generalize their findings".

This is precisely where the ritualistic practice emerges: researchers treat statistical significance as the ultimate criterion for generalization, overlooking the fact that true statistical generalization requires deriving results from a randomly chosen, well-defined sample that adequately represents the population (Mills et al., 2010).

3.2. The Manifestation of Bias Toward Statistical Significance

Carver (1987) defined statistical significance as indicating rarity or scarcity, meaning that statistical significance essentially represents "*the rare*" (cited in Bahash, 2023).

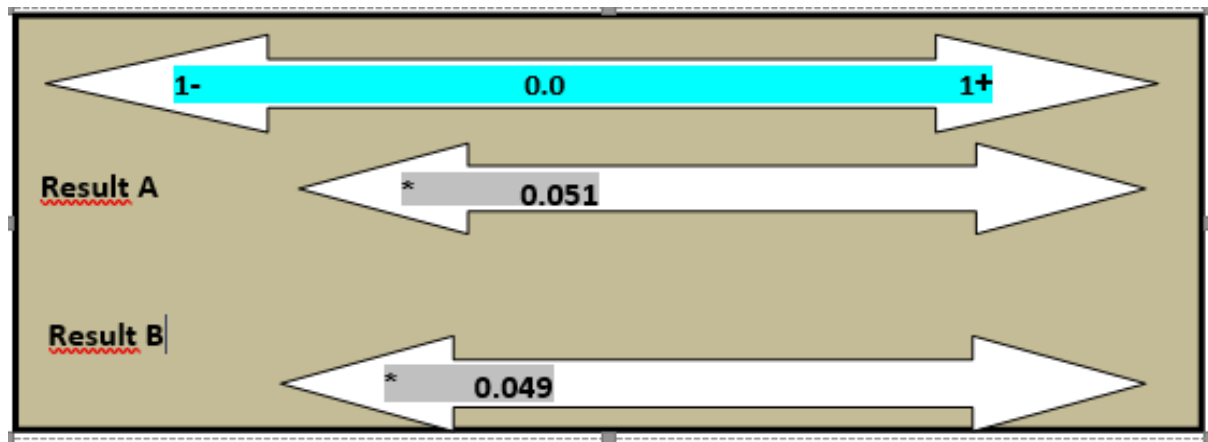
One notable form of bias is the sacralization of the critical value (p-value). Hopkins (1997) questioned the origin of the conventional threshold, asking:

"What is so special about 0.05? In truth, nothing. Someone once decided this value was reasonable, and now we are stuck with it. It has become sacred as the cutoff point for determining significance, giving researchers license to publish if $p < 0.05$."

Example of bias toward statistical significance (Hopkins, 1997)

Figure (3)

Diagram comparing two study outcomes: one with $\alpha = 0.051$ and another with $\alpha = 0.049$.



From the figure, we observe two different results: one with $\alpha = 0.051$ and the other with $\alpha = 0.049$. The difference between them is minuscule (0.002), yet it raises a critical question: Which result is publishable? According to McLain (1997), cited in Bahash (2023):

“The 0.05 threshold has become an arbitrary gatekeeper of practical knowledge. Results that fail to meet the cutoff are deemed unimportant. While this gives the illusion of scientific objectivity, it is actually meaningless because the cost of this objectivity is meaning itself”.

Bias also manifests in ignoring effect size and correlation magnitude—the core of practical significance. Effect size is a central factor influencing statistical power, as it represents the actual numerical difference between true population parameters and those specified under the null hypothesis. In this sense, effect size is directly tied to statistical power: the larger the effect size, the stronger the test.

Effect size complements statistical significance by adding practical and applied meaning. It reflects the extent to which research findings can be applied in real-life contexts. Thus, statistical significance alone is insufficient for sound decision-making; effect size must also be considered.

Another form of bias arises when studies yield non-significant results, often leaving researchers perplexed. For example, in a doctoral dissertation titled *Exam Anxiety and Its Relation to Academic Achievement Among High School Students*, results from a sample of 400 students revealed no significant relationship. The researcher admitted being shocked, since all previous studies had reported significance. However, closer inspection revealed critical contextual factors: the study was conducted in a support center providing intensive tutoring and psychological assistance, and it took place well before final

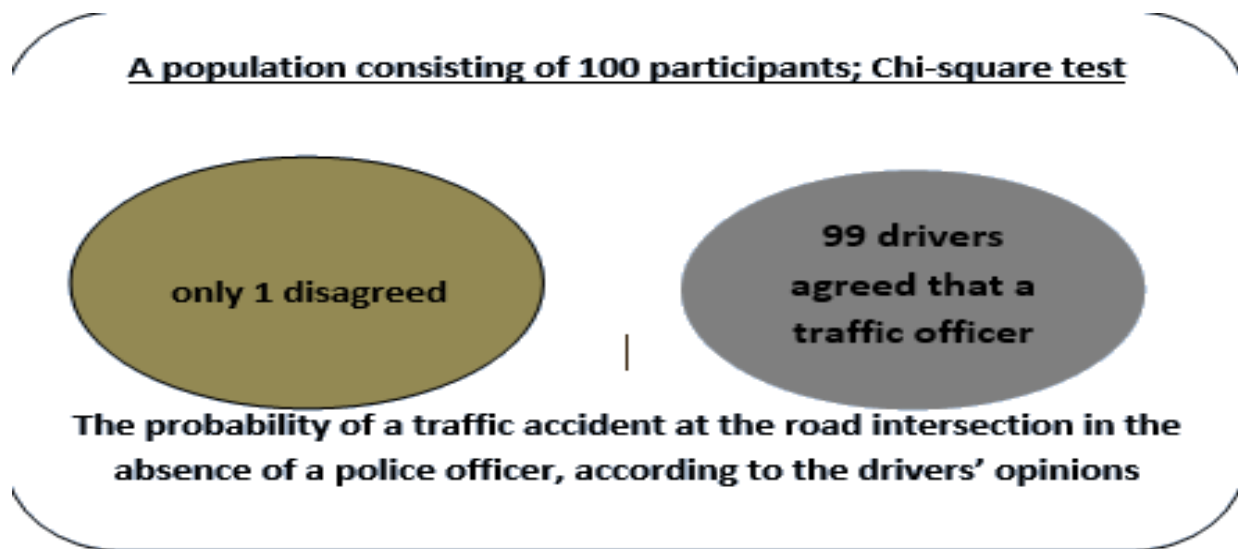
exams—when exam anxiety had not yet peaked. Under these conditions, non-significant results were not surprising. Yet, rather than interpreting them meaningfully, the researcher succumbed to the ritual of expecting significance.

This illustrates how researcher subjectivity becomes entangled with statistical rituals, regardless of whether significance is present or absent.

Another example of the dangers of relying solely on statistical significance:

Figure (4)

Diagram: Chi-square test results for the opinions of 100 drivers about the presence of a traffic officer at an intersection.



In this hypothetical study, 99 drivers agreed that a traffic officer was unnecessary, while only 1 disagreed, citing the risk of reckless driving. Statistically, the result would be highly significant ($p \leq 0.01$), seemingly allowing for generalization. Yet, if such a result were taken as justification for eliminating traffic officers, the consequences could be catastrophic.

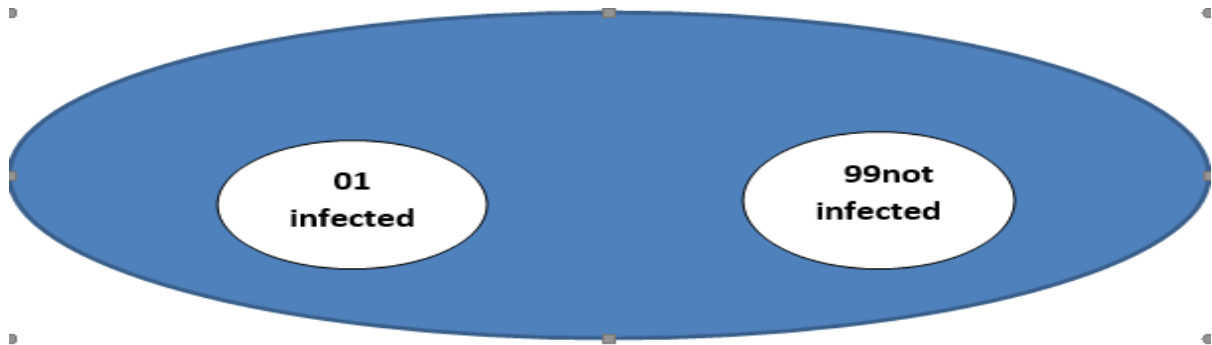
The critical questions should instead be:

- What is the probability of an accident without the officer present?
- How does this probability change when the officer is present?

This example underscores the danger of ritualistic dependence on quantitative data alone, while disregarding qualitative considerations. The essential issue is not merely whether significance exists, but what meaning and implications lie behind it.

Figure (5)

Represents the results of COVID-19 infection tests on a sample of 100 individuals using the Chi-square test (infected / not infected).



In this other hypothetical study, we see the danger of relying on interpretations of statistical significance in phenomena that include rare observations, as such cases make the p-value highly misleading and may lead to misguided decisions. For instance, imagine that 99 out of 100 individuals tested were not infected with COVID-19, while only one individual was diagnosed with the disease. Any researcher relying solely on inferential statistics would reject the null hypothesis (H_0), which denies differences in infection rates among individuals. The results would indicate significance at a level below $\alpha = 0.01\%$, despite the fact that there was practically no spread of infection.

This makes the compass of significance highly deceptive, leading us to affirm that “statistical significance hears but does not see.” This is a warning all researchers must heed. In other words, rare or infrequent observations can alter the course of events in many phenomena. If we accept that one infected individual could transmit the virus to all others within days, then the outcomes of statistical significance alone must be reconsidered.

This point was emphasized by Babbitt (2000), as cited in Al-Dhahir et al. (2010), in their article *The Reality of Statistical and Practical Significance and the Power of Statistical Tests Used in Master's Theses on Talent and Creativity*. They noted that perspectives on statistical significance have evolved: it now provides little useful information for real research questions, and the binary outcome of significance (accept/reject) no longer satisfies many scientists. The null hypothesis is increasingly used to assess the magnitude of effects rather than simply their existence (Al-Dhahir et al., 2010).

The process of interpreting whether a relationship exists between two variables is more important than the relationship itself. Interpretation is what gives scientific knowledge its meaning. Regardless of how strong the statistical associations may appear; they are insufficient for decision-making unless examined under questions about the extent to which rare or limited observations can shift results. Ignoring this means chasing meaningless associations confirmed only by statistical significance such as linking intelligence and height in a student sample, or age and academic achievement in youth. Many student theses fall into this trap, producing trivial statistical correlations that ultimately lead to dead ends. If, for example, we assume a correlation between the increase in the number of mosques and the rise in crime rates, or between the number of schools and the increase in bullying, this reinforces the claim that “statistical significance hears but does not see.” The problem deepens when regression analysis—an extension of Pearson’s correlation—is used to predict relationships. This shifts us from mere description to prediction, leading to flawed claims such as “building more mosques predicts higher crime” or “constructing more schools predicts more bullying.”

The reality we seek requires us to question the type of relationship: descriptive or causal. Descriptive relationships are confined to statistical significance, while causal ones demand the conditions of the experimental method. To strengthen the interpretation of descriptive correlations, researchers must seek mediating variables that clarify the true relationship. For instance, it is illogical to accept a direct causal link between the number of mosques and crime without experimental verification. Descriptively, the mediating factor is population density. Once identified, a logical inference emerges: increased density leads to more mosques and to higher crime. The same holds true for schools and bullying, where the mediator is student population size. Thus, the explanation gains coherence.

Such misleading correlations merely reflect mathematical-statistical formulations imposed on psychological and educational sciences. It is unreasonable to accept a positive correlation between variables that inherently hold opposing meanings (e.g., mosque-building and crime, schools and bullying) without accounting for mediating factors. This underscores the need for researchers to choose the most appropriate methods for relevant problems.

The search for relationships and their generalization recalls Karl Popper’s epistemology of scientific knowledge and his confrontation with the problem of induction. Popper rejected causal generalization of phenomena, denying that relationships between variables could be viewed as causal in the strict sense. He

emphasized temporal succession, arguing: “When we look outside ourselves toward external things and reflect on causes, we cannot discover from a single case of power or necessary connection any property that ties an effect to its cause and makes one the inevitable outcome of the other. All we find is that one event follows another in reality: the movement of the first billiard ball is accompanied by the movement of the second. That is all that appears to the senses. The mind perceives no inner impression of causal necessity.” (Zeitouni, 2020).

From the researcher’s perspective, Popper’s idea carries more serious implications than it seems at first glance. Reinterpreted, it suggests that levels of description and explanation constantly overlap: every event can simultaneously be descriptive and explanatory. For instance, if there is a correlation between the expansion of employer-employee relationships and employee negligence, one can describe the result (“expanded relationships were observed”) and explain it (“expanded relationships led to increased negligence”). This raises the question: where does description end and explanation begin?

If negligence results in lower productivity, negligence becomes a cause after initially being a consequence of workplace relationships. If lower productivity leads to bankruptcy, and bankruptcy leads to layoffs, we must then ask: Who pushed the first billiard ball? And what if we discover that the expanded relationships were due to the employer’s highly extroverted personality, which made him lenient with requests? Then the search for a primary cause becomes misplaced—like searching for a needle in a haystack.

Researchers must remain alert to this complexity. They should avoid superficial understandings of social phenomena and not stop at statistical generalizations without probing deeper interconnections among variables. Every research project must aim to approach its subject matter from multiple angles. Returning to the negligence example, we would inevitably find a subset of employees who remain disciplined despite the employer’s leniency. So, what kind of generalization can we claim? And since every generalization informs decisions, what kind of decisions are implied—if, for example, the decision were to dismiss all employees?

3.3. The Appearance of Bias Toward Large Samples

Studies that rely on large sample sizes often yield statistically significant results, since sample size itself becomes one of the main indicators determining whether a correlation is significant or not. In other words, one might argue that *statistical significance means large sample size*. This was emphasized by several researchers in psychological statistics, including Thompson (1989), Wilkerson and Lawson, and

Snyder et al. (1992), who maintained that sample size is the primary factor determining whether results turn out significant. Consequently, the null hypothesis cannot be sustained when large samples are used, as they inevitably produce variability within the sample, and therefore significant differences. Large observations inevitably lead to this outcome, which supports the argument that one can question the meaningfulness of statistically significant results at $p = 0.06$ with a sample of 40, compared to the same study yielding $p = 0.04$ with a sample of 400 participants.

This was a key point raised by Daniel (1998), who demonstrated that minor, non-significant differences between two groups in a small sample become significant in large samples, as the latter provide researchers with a sufficient number of cases (Bahash, 2023). Similarly, Kriss et al., in their book *Research Methods in Clinical and Counseling Psychology*, affirmed that statistical significance by itself does not tell us much. The problem lies in the arbitrariness of the traditional significance threshold of $p < 0.05$. This issue becomes more pronounced with large samples: a result considered statistically significant is one whose probability of occurring by chance is less than 1 in 20 trials. Yet a result at $p = 0.049$ is not substantially stronger than one at $p = 0.051$, although the first is reported as significant while the second is not. This undermines the logical justification for the conventional cutoff of 1 in 20—it might just as well be 1 in 25 or 1 in 18.

A more critical issue is that any effect will become statistically significant if the sample is large enough, since no null hypothesis can ever be completely true. As a result, one may obtain statistically significant findings that are scientifically meaningless. For example, when testing a new treatment for depression, the mean difference between two groups may only be two points on the Beck Depression Inventory. While this may reach statistical significance with a large sample, it would be clinically insignificant if both groups remain severely depressed (Shinar & Boulhbal, 2021).

Moreover, research aimed purely at statistical generalization risks becoming like a horse pulling a cart that is about to lose its balance. Every field has examples of this. Take, for instance, Gardner's theory of multiple intelligences, which restored value to many individuals whom statistical significance had labeled as failures. For example, when comparing the mean scores of a sample of school dropouts with the population mean of their peers, statistically significant differences between the two would still brand them as failures. Yet, a closer look at the skills and trades mastered by today's youth reveals an entirely different reality. A car breakdown, for instance, will not be fixed by statistical significance, but by that very dropout

whom statistics declared a failure. Likewise, an electrical, heating, or plumbing problem will not be solved by the majority, but by the few, contrary to what statistical significance had implied.

4. Conclusion

Statistical significance captures only a fragment of truth in psychological and educational sciences. Human behavior does not always mirror underlying traits, making generalization a complex and uncertain endeavor. Scientific reality resembles a broken mirror: each fragment offers only a partial reflection, and statistical inference is but one of these fragments.

Rarity, though often overlooked, holds profound meaning. Across history, religions, and civilizations, lasting change has typically arisen from the few rather than the many. This reminds us that statistical “rarity” should not be dismissed as trivial but considered a potential driver of meaningful insights.

Today, generalization in psychological and educational research is increasingly problematic, as scattered, intervening, and extraneous variables multiply. What seems generalizable today may be contested tomorrow. Thus, researchers must approach generalization with humility, methodological rigor, and a recognition of the limits of statistical significance in capturing the complex realities of human experience.

Reference

- Abdullah, N., & Laaradj, L. (2023). Criteria for evaluating the quality of scientific research from the perspective of the academic community: An applied study. *Journal of Applied Sciences*. 18(1). 1-22. <https://asjp.cerist.dz/en/article/217404>
- Abu Allam, R. (2006). *Statistical analysis in psychological, educational, and social sciences: Inferential statistics* (Ed.). Dar Al-Nashr Liljamia
- Aiche, S. (2016). The Problem of Generalization in Psychological and Educational Research: An Evaluation Study of University Thesis at the Universite of Oran. *Journal of University of Babylon, Humanities* 26(6). 127-146. <https://www.researchgate.net/publication/325854867>
- Al-Annsi, A. (2024). *Generalization in scientific research*. Retrieved from <https://fr.scribd.com/document/763563230>

- Bahash, A. (2020). *Statistical and practical significance of psychological and educational research hypotheses*. *Al-Jami' Journal of Psychological Studies and Educational Sciences*, 5(1), 371-404. <https://search.emarefa.net/detail/BIM-987048>
- Carver, R. P. (1978). The case against statistical significance testing. *Harvard Educational Review*, 48(3), 378–399. <https://doi.org/10.17763/hear.48.3.t490261645281841>
- Cohen, J. (1994). The earth is round ($p < .05$). *American Psychologist*, 49(12), 997–1003. <https://doi.org/10.1037/0003-066X.49.12.997>
- Hammersley, M. (1999). *What's wrong with ethnography?* (Ed.). Routledge. <https://doi.org/10.4324/9781351038027>
- Hopkins, W. G. (1997). *A new view of statistics*. Retrieved from <http://www.sportsci.org/resource/stat/>
- Howitt, D., & Cramer, D. (2016). *Introduction to statistics in psychology* (6th Ed.). Pearson.
- Kroum, A., & Kroum, K. (2022). The sample and the problem of representation and generalization in social and human research. *Oases Journal for Research and Studies*, 15(1), 916–930.
- Makhlouf, S., & Ghrabi, K. (n.d.). Quality in institutions: A theoretical study. *Haqa'iq Journal of Psychological and Social Studies*, 4(15), 42-53. <https://asjp.cerist.dz/en/article/128015>
- Polit, D. F., & Beck, C. T. (2010). *Essentials of nursing research: Appraising evidence for nursing practice* (7th Ed.). Lippincott Williams & Wilkins.
- Popper, K. (1959). *The logic of scientific discovery* (M. Al-Baghdadi, Trans.). Arab Thought Foundation.
- Shinar, S., & Boulhbal, A. (2021). Statistical significance and statistical test power. *Journal of Islamic Sciences and Civilization*, 6(2).
- Wasserstein, R. L., & Lazar, N. A. (2016). The ASA's statement on p-values: Context, process, and purpose. *The American Statistician*, 70(2), 129–133. <https://doi.org/10.1080/00031305.2016.1154108>
- Zeitouni, C. (2020). Epistemology of the growth of scientific knowledge in the philosophy of Karl Popper. *Al-Ibrahimi Journal of Social and Human Sciences*, 4(1), 9-43. <https://asjp.cerist.dz/en/article/184639>