Five Minute Science Lesson: Measurement Theory and Lie Detection

Article · January 2016

CITATIONS
0

READS
4

1 author:

Raymond I Nelson

48 PUBLICATIONS **68** CITATIONS

SEE PROFILE

Five Minute Science Lesson: Measurement Theory and Lie Detection

Raymond Nelson[1]

1. Raymond Nelson is a psychotherapist and polygraph examiner who has conducted several thousand polygraph examinations. He has expertise in working with perpetrators and victims of sexual crimes and other abuse and violence. Mr. Nelson has expertise in statistics and data analysis and is one of the developers of the OSS-3 scoring algorithm and the Empirical Scoring System. He is a researcher for Lafayette Instrument Company (LIC), a developer and manufacturer of polygraph and life-science technologies, and is also a past-President of the American Polygraph Association (APA), currently serving as an elected Director. Mr. Nelson teaches and lectures frequently throughout the United States and internationally, and has published numerous studies and papers on all aspects of the polygraph testing, including the psychological and physiological basis, test data analysis, faking/countermeasures, interviewing and question formation and test target selection. Mr. Nelson has been involved in policy development at the local, state, national and international levels in both polygraph and psychology, and has testified as an expert witness in court cases in municipal, district, appellate, superior and supreme courts. Mr. Nelson is also the academic director of the International Polygraph Training Center (IPTC). There are no proprietary or commercial interests and no conflicts of interest associated with the content of this publication. The views and opinions expressed in this publication are those of the author and not necessarily those of the APA, LIC or IPTC. Mr. Nelson can be reached at raymond.nelson@gmail.com.

It is sometimes said that it is not possible to actually measure a lie. Simplistic and concrete thinkers, and those opposed to the polygraph test, are content to end the discussion at this point and offer the impulsive and erroneous conclusion that scientific tests for lie detection and credibility assessment are not possible. This conclusion is erroneous, a *non-sequitur*, because many areas of science involve the quantification of phenomena for which direct physical measurement is not possible. The theory of the polygraph test, and lie detection and credibility assessment in general, in fact does not involve the measurement of deception or truth-telling. Nor does it involve the measurement, or recording, of fear or any other specific emotion. This publication attempts to introduce and orient the reader to measurement theory and its application to the problem of the polygraph and scientific lie detection or credibility assessment testing.

The analytic theory of the polygraph is that greater changes in physiological activity are loaded at different types of test stimuli as a function of deception or truth-telling in response to the relevant target stimuli (Nelson, 2015a, 2016; Senter, Weatherman, Krapohl & Horvath, 2010). In the absence of an analytic theory or hypothesis of polygraph testing, polygraph theories have previously been expressed in terms intended to describe the psychological process or mechanism responsible for reactions to polygraph test stimuli. Although much has been learned about the recordable physiology associated with deception and polygraph testing, less work has been done to investigate psychological hypotheses about deception. In general, the psychological basis of the polygraph is presently assumed to involve a combination of emotional, cognitive and behaviorally conditioned factors (Handler, Shaw & Gougler, 2010; Handler, Deitchman, Kuczek, Hoffman, & Nelson, 2013; Kahn, Nelson & Handler, 2009). The analytic theory of polygraph testing implies that there are physiological changes associated with deception and truth-telling, and that these changes can be recorded, analyzed, and quantified through the comparison responses to different types of test stimuli. Comparison and quantification are objectives central to measurement theory. Application of measurement theory to the polygraph test will require at least a basic understanding of measurement theory.

Types of measurement

Stevens (1946) attempted to provide a framework for understanding types of measurement. At that time, part of the intent was to clarify the selection of statistical and analytic methods associated with different types of measurement data. It was evident almost immediately that the selection of statistical was a more complex endeavor than could be characterized by the reduction of the array of data types and scientific questions to a small set of categories.

Nominal scales are without any rank order meaning (e.g., *cat, mouse, dog, ostrich, zombie, robot*). Mathematical transformation of nominal items is not possible. *Ordinal* measurements have rank order meaning but have imprecise meaning about the distance between items (e.g., knowing the first, second and third place winners of an ostrich race does not provide information about the difference in race times). Some mathematical transformations are possible with ordinal measurements, with the requirement that they preserve the ordinal information and meaning. *Interval* scale measurement have both rank order meaning and provide meaningful information about the difference between items. However, the zero point of an ordinal scale is arbitrary and therefor meaningless. A classical teaching example for the arbitrariness of an interval-scale zero point is a temperature scale for which we have both the Fahrenheit and Celsius scales with different arbitrary zero points, and no expectation that zero means that there is no temperature or no heat to be measured. *Ratio* measurements include combination of rank order meaning and interval distance meaning along with the notion of a non-arbitrary zero point. In ratio scales measurements zero means none (e.g., no difference). Later, Stevens (1951) offered a set of prescriptions and proscriptions as to the type of statistics that are appropriate for each type of data.

The most common form of criticism of Stevens have focused on the fact that it is unnecessarily restrictive (Velleman & Wilkinson, 1993), resulting in the overuse of non-parametric methods that are known to be less efficient than parametric methods (Baker, Hardyck, & Petrinovich, 1966; Borgatta & Bohrnstedt, 1980), and that the type of analysis should be determined by the research question to be asked (Guttman, 1977; Lord, 1953; Tukey, 1961). Luce (1997) asserted directly that measurement theorists today do not accept Stevens' overly broad definition of measurement. Nevertheless, Stevens's work provides a useful introduction to the conceptual language and problems of measurement theory.

Measurement theory

Measurement theory is an area of science concerned with the investigation of measurability and what makes measurement possible. Helmholtz (1887) began the tradition of scientific and philosophical inquiry into measurement theory by asking the question *"why can numbers be assigned to things"*, along with other questions such as *"what can be understood from those numbers"*? According to Campbell (1920/1957), *measurement is the process of using numbers to represent qualities*. In general, the properties of measurable phenomena must in some ways resemble the properties of numbers. Later work by Suppes (1951) on the differences between measurable and un-measurable phenomena and began to formalize the tradition of measurement theory by clarifying our understanding of the requirements for measurement and gave rise to a modern representational theory of measurement (Diez, 1997; Suppes, 2002; Suppes & Zinnes, 1963; Suppes, Krantz, Luce, & Tversky, 1989; Niederee, 1992). Stated simply, the representational theory of measurement involves the assignment of numbers to physical phenomena such that empirical or observable relationships are preserved.

The existence of order (rank order) relationships between measurable objects is central to the requirements for the measurability of any phenomena. We must be able to quantify one instance of the phenomena as have greater magnitude than another. Another central requirement of measurable phenomena is that there must be a way of combining measurable objects in a way that is analogous to mathematical addition. This is, the addition of measurable phenomena must have a sensible physical interpretation. These are among the main differences between measurable and un-measurable phenomena.

For example: measurements can be applied to physical phenomena such as a person's height, weight, and blood pressure. This is possible because these things involve physical phenomena: the linear or unitized distance from head to toe, the gravitational force on a person's physical mass, and the unitized pressure required to overcome and occlude arterial pressure relative to a reference point such as average atmospheric pressure at sea level (i.e., 29.92inHg or 760mmHg). These phenomena can be combined in ways that are in some way analogous to numerical addition. That is, there is some coherent physical interpretation to additive combinations of different instances of these physical phenomena. Time limited events can also be measured. For example: if a person jumps into the air two times and if we mark the physical height of each jump and

then combine the two distances, then this is also analogous numerical addition. However, attempts to record physiological changes to polygraph stimuli does not necessarily conform to these requirements for rank order relationships and additivity. The details of how recorded polygraph data can result in the quantification of deception and truth-telling are addressed in the remainder of this publication.

Firstly, it has long been established that responses to polygraph stimuli cannot be taken or interpreted directly as a measurement of deception. Nor can responses to polygraph stimuli be interpreted as a recording or measurement of fear or any other specific emotion. Responses to polygraph stimuli are a form of proxy or substitute data for which there is a relationship or correlation with deception and truth-telling. The reactions and recorded data themselves are neither deception nor truth-telling per se. Secondly, although it may be possible to interpret rank-order the relationships between test stimuli according to the magnitude of response, polygraph recording instrumentation today has not been designed to provide data that satisfy the additivity requirement for measurement data. In other words, attempts to make any sensible additive combination of the actual response data within each of the respiration, cardio, electrodermal and vasomotor sensors is neither intended or established. Instead, polygraph data must be transformed to a more abstracted form before it can be further analyzed and interpreted as to their meaning. Polygraph scoring and analysis algorithms, whether manual or automated, are intended to accomplish and facilitate such transformation, analysis and interpretation.[1]

Fundamental and derived measurements

Some measurements can be referred to as *fundamental* and require no previously measured phenomena to achieve their determination. The main requirement for a fundamental measurement is that there are some physical phenomena for which there is

1. A major difference between manual an automated polygraph analysis algorithms is that manual scoring protocols were developed during a time when field practitioners did not have access to and were unfamiliar with use of powerful microcomputers. Manual scoring algorithms therefore rely on mathematical transformations that are, of necessity, very simple, if not somewhat blunt. Earlier versions of manual scoring protocols did not make use of normative reference distributions, statistical corrections or confidence intervals. Another major difference is that manual scoring protocols accomplish feature extraction tasks – the extraction of signal information from other recorded information and noise – using subjective visual methods. Automated analysis algorithm will make use of more advanced statistical methods, and will rely on objective and automated feature extraction methods that are less vulnerable to subjective interference.

some quantity that can be understood as either more or less (e.g. *is it heavy*) as opposed to phenomena that are better understood as all-or-nothing (e.g., *is it an ostrich*). If we have two ostriches, it makes some sense to ask a question such as which ostrich is heavier because there is meaningful intuition around the idea that some ostriches are heavier. But it does not make sense to ask the question which is more an ostrich, because there is no meaningful intuition that can be gained from its answer. Being an ostrich is a property, not a quantity. The weight of an ostrich is also a property, and this illustrates that some properties can also be quantities. The physical phenomena of weight or heaviness can be quantified to achieve greater precision than simply saying *very* heavy or *very very* heavy when attempting to compare the weight of two ostriches. Without the use of numerical quantities, two different observers might reach two different conclusions about which ostrich is heavier no matter how we attempt to use our descriptive adjectives. Different observers are more likely to reach similar conclusions when using measurements vs. the alternative of not using measurements. The use of measurements permits us to think about, understand, describe and plan the world around us with greater precision, which is to say greater *reproducibility*.

When a measurement is not intended or not expected to be a precise or exact quantity it is sometimes referred to as an *estimate*. Probabilities, because they are not expected to be exact, are estimates. Although some may use or express the notion of probabilities subjectively, reproducibility of computational probability estimates is an important difference between the scientific and unscientific use of the concept of probability.

Some measurements can be thought of as *derived*, because these are achieved not through the direct quantification of a physical phenomenon, but through the comparison of an unquantified physical phenomenon with another known physical phenomenon. In principle, we can measure an *unknown* distance if we have some other distances and angles that are already *known*. For example, if we place a set of satellites in orbit around the earth we can calculate and know the locations of those satellites relative to a set of objects for which the locations are known on the earth. Then, if we have some means of receiving information from the satellites with known locations, we can use the information from the satellites to calculate and measure our own location if our location is unknown. This would be like older practices in which if we can calculate the location of objects in the solar system according to a system of counting or quantifying the number of days since a previously observed event, then we can use the location of

the object in the solar system. And the location of objects in the solar system could be used, along with a defined system of scientific and mathematical rules, to measure or quantify our current location on the earth. Another example of a derived measurement is the measurement of blood pressure, for which we use our knowledge about atmospheric pressure to quantify our assessments of cardio pressures during the systolic and diastolic phases of the cardiac cycle.

Scientific testing as a form of (probabilistic) measurement

As it happens, many interesting and important phenomena cannot be either observed directly or are not subject to physical measurement. This is sometimes because the phenomenon of interest is amorphous (without physical substance), and sometimes because the information does not conform to the order and additivity requirements of measurement. If we want to improve the precision of our assessment and decisions for these phenomena we will need to rely not on measurements but on scientific tests that quantify a phenomenon of interest using statistics and probability theory. Nelson (2015b) provided a description of how a polygraph test, and tests in general, can be thought of as a single subject science experiment. Scientific tests can also be thought of as a form of probabilistic measurement, in which statistical and probability theories are used to quantify a phenomenon that is not amenable to actual measurement.

An example of scientific testing as a form of probabilistic measurement is the testing measurement of amorphous and un-measurable psychological phenomena such as *personality* and *intellectual functioning*, during which an observed quantity of data from an individual is compared mathematically to a known quantity in the form of normative reference distribution, or probability reference model, that characterizes our knowledge of what we expect to observe. Reference models can be calculated empirically, through statistical sampling methods, and can also take the form of theoretical reference distributions that characterize our working theories about how the universe, or some small part, works by relying only on facts and information that are subject to mathematical and logical proof.

In the case of the polygraph test – for which the basic analytic theory holds that greater changes in physiological activity will be loaded for different types of test stimuli as a function of deception and truth-telling in response to the relevant stimuli – it is not the comparison of relevant and other test questions that forms the basis of our conclusions. Instead, it is the comparison of differences in reactions to relevant and other test questions to a reference distribution that anchors our knowledge about the expected

differences in responses to relevant and other questions among deceptive or truthful persons. Ideally, other questions would have the potential to evoke cognitive and emotional activity of similar quality, though perhaps different in magnitude, then the relevant target stimuli. However, it is not necessary that other questions have similar ecological value compared to the relevant stimuli to be a useful and effective basis for statistical comparison. An example of this can be seen in the use of directed-lie-comparison (DLC) questions, for which Blalock, Nelson, Handler & Shaw (2011) provided a summary of the research on their effectiveness (and for which the name DLC should not be taken to imply that response to these questions are actual lies).

Scientific tests as a form of prediction

If we want to quantify or improve the accuracy or precision associated with our assessments and conclusions about future events that have not yet occurred – assuming we want to quantify our conclusions now without waiting for the event to occur – then we are once again attempting to quantify a phenomenon that is not amenable to direct observation or measurement. For this we need a *test*, with which we can make probabilistic conclusions about the future outcome. Tests used in this way can be thought of as a form of scientific *prediction*. It is not a form of magic or divination. It is a form of probabilistic modeling.

An example of the quantification of a future event is the measurement or quantification of risk level for some hazardous event – for which it is implicit that the future event has not yet occurred and therefore cannot be physically quantified or observed. Yet another example, involving the prediction of a future event, will be the quantification of an outcome for an election that has not yet occurred. Both examples – risk outcomes and election outcomes – can involve a future event for which the associated value is binary (e.g., an event has or has not occurred, or an election has been won or not won). At any single point in time, the event has either occurred or has not occurred. We might, at times, want to simply wait to observe the result to achieve a deterministic conclusion. Deterministic observation of an outcome would, of course, obviate any need for testing and quantification.

A notable difference between the prediction of risk events and scheduled outcomes is that election outcomes can be expected to occur at a scheduled point in time, at which time it is possible to observe the result. After the scheduled event the outcome is a matter of fact, not probability. Prior to the scheduled event, the outcome can be thought of as a probability, such that there are some factors that are associated with the different possible outcomes. A goal of scientific predic-

tion involves the identification these associated factors so that they can be characterized as random variables and used to develop a predictive test or model.

Probabilities associated with the outcomes of scheduled events that have not yet occurred can be thought of as the proportion of outcomes that would occur a certain way, given the random variables that influence the outcome, if it were possible to observe the event over numerous repetitions. Effectiveness or precision of a test as a predictive model will depend on our ability to correctly understand the random variables related to the possible outcomes. Ultimately, the outcome will be a certainty, and not a probability. Prior to the outcome occurrence, it remains a probability or prediction. When prediction errors occur, their causes can be due either to random variation, or to misunderstanding and mischaracterizing the random variables related to the possible outcomes.

Some types of outcomes are expected to occur at an unknown time, or they may not occur at all for very long periods of time. We can think of these outcomes as probabilities. For example: *what is the probability that a known criminal offender will re-offend*, or *what is the probability of an earthquake in Mexico City*, or *what is the probability of a flood*? These events can also be regarded as certainties after they have occurred, and are also subject to some relationship with related factors that are associated with their occurrence. As with other prediction models, identification and characterization of the associated factors is an important objective in the development of risk assessment or risk prediction models. Probabilities associated with risk prediction outcomes can be thought of in terms of frequencies, such that high probability events occur with greater frequency, while low probability events occur with lower frequency.

Nearly everything – including events for which our intuition tells us the likelihood is very low – can thought of as a probability. This can, at times, be taken to absurdity. For example: *what is the probability of a zombie horde attack*, or *what is the probability of a robot apocalypse*? For these extreme examples our intuition tells us the probability is either absolute zero or essentially zero, but we can still engage some imagination as to the factors that could become associated with their occurrence. If we expand the period under consideration, then the probabilities associated with rare events can become conceivably greater. For example: *what is the probability that an ostrich will fall from the sky*? If we expand our dimensions for time and location to the notions of ever and anywhere, we can intuitively understand some non-zero probability associated with an ostrich falling from the sky, along with the kinds of factors that might

be associated with its possible occurrence (e.g., emergency ostrich airlift from a flooded ostrich farm).

Quantification of future events such as hazards or election outcomes requires that we treat the future outcome in the same manner as any other amorphous phenomena that we may wish to quantify. We treat the future outcome as a probability. Quantification of an outcome is useful only when it is a future outcome – an outcome that has not yet occurred. If information exists, and is available for observation or measurement, then the outcome is not amorphous but is a physical phenomenon. Direct observation or measurement of a future outcome will require that we wait until the future point in time. Until then, if we want to try to predict a future outcome that has not yet occurred we will need to rely on probabilities to describe the amorphous future event. Similarly, observation or measurement of a past event will require that some physical phenomena from the event are available for observation or measurement. If we wish to quantify a past event for which no physical phenomena are available, then we will once again need to rely on probability theory to quantifying the amorphous phenomena.

A famous quotation of unknown Danish authorship during the years 1937-1938 states, [in English] *"It is difficult to make predictions, especially about the future."* This simple and humorous quotation reminds us that predictions of all kinds are inherently imperfect, including predications based on scientific test data. Probabilistic conclusions are inherently imperfect. Indeed, they are not expected to be perfect. Probabilistic conclusions are expected only to quantify the margin of uncertainty associated with a conclusion. Statistical predication is an inherently probabilistic and statistical endeavor for which any conclusion is both probably correct and probably incorrect. Conclusions about deception or truth-telling, despite the desire for certainty and infallibility, will be inherently probabilistic and inherently imperfect.

Conclusion: scientific polygraph tests as a form of statistical classification

Polygraph test results can be thought of a form of prediction that some other evidence exists and can be identified as a basis of evidence to confirm or refute a test result. A simpler and more general way to think about these tests will be as a form of statistical *classification*. Like other scientific tests, statistical tests intended for classification are not expected to be perfect, infallible or deterministic. Neither are statistical classifications expected to provide the same level of precision as an actual measurement of a physical phenomenon. Like other probabilistic endeavors, scientific

tests intended for classification are expected only to quantify the margin of uncertainty or level of confidence that can be attributed to a conclusion. Most importantly, the method for statistical quantification should be accountable and the results should be reproducible by others. The ultimate measure of effectiveness of a statistical test is not in the achievement of perfection or infallibility, but in the observation of correct and incorrect real-world classifications that conform to our calculated probability estimates.

If the basic analytic theory of the polygraph test is incorrect – if no physiological changes are correlated with differences between deception and truth-telling – if all physiological activity in mere random chaos with regard to deception and truth-telling, then humans have virtually no chance of ever known if they are being lied with any precision greater than random chance. The only way to protect oneself from deception will be to remain cynical and suspicious of all, while trusting no-one. Although perhaps tempting, this will be unrealistic and unsustainable over time. On the other hand, if it is correct that some changes in physiological activity are associated with deception and truth-telling at rates significantly greater than chance, then it is only a matter of time before technologists, engineers, mathematicians, statisticians and data analysts devise some means to increase the availability of useful signal information amid the chaotic noise of other physiological activities and exploit those signals with some new form of scientific credibility assessment or lie detection test.

If the polygraph test is ultimately an interrogation and not a scientific test, then measurement theory is of no concern and no consequence to the polygraph profession. But in this case, people will begin to turn to other scientific methodologies when they desire a scientific test for credibility assessment, and the polygraph test may eventually be replaced. On the other hand, if the polygraph test is a scientific test, then it will serve the interests of all for polygraph professionals to become familiar with the basics of measurement theory and the discussion of scientific polygraph test results, including categorical conclusions about deception and truth-telling and conclusions about countermeasures, using the conceptual language of measurement and probability theories. Polygraph conclusions are not physical measurements; they are probability estimates.

In the absence of probabilistic thinking applied to the polygraph test, there will be an impulse for some to engage in naïve and unrealistic expectations for deterministic perfection. There will also be a desire or impulse for some to feign infallibility, due to superior professional wizardry or skill, and this

can for a time appear to be an effective marketing strategy. But feigned infallibility will lead to confusion and frustration when it is inevitably observed that testing errors can, and do, occur. A temporary corrective solution to this frustration will be to find fault with the professional, not the test – thereby restoring the false assumption of infallibility, so long as we avoid those less competent wizards less competent experts. Although gratifying for a time, this type of approach is unscientific, and will be unsustainable in the context real-world experience and scientific evidence.

Polygraph test result should be understood and described like other scientific test results, using the conceptual language of statistical probabilities. Expression of purportedly scientific conclusions, including conclusions about deception and truth-telling and conclusions about the use of countermeasure, without the use of probability metrics will invite accusation that polygraph is mere subjective pseudoscience cloaked in overconfidence. A scientific approach to polygraph testing will recognize that the task of any test is to quantify a phenomenon probabilistically when direct observation or physical measurement are not possible, and to recognize and make accountable use of the potential for testing error when deciding what value to place upon and how to use or rely upon the test result.

Like other scientific tests, polygraph tests are intended to make probabilistic classifications of deception and truth-telling in the absence of an ability to directly observe or physically measure the issue of concern. If physical phenomena were available for observation or measurement, then a scientific test would not be needed. Because deception and truth-telling are amorphous constructs, scientific lie detection and credibility assessment are, ultimately, epistemological concerns that are sometimes the subject of complex and important philosophical questions such as: *what does it mean to say that something is true*, and *what kind of things can be said to be true?* Although deeply interesting, these must be the subject of another publication.

References

Baker, B.O., Hardyck, C.D., & Petrinovich, L.F. (1966). Weak measurements vs. strong statistics: An empirical critique of S.S. Stevens' proscriptions on statistics. Educational and Psychological Measurement, 26, 291-309.

Blalock, B., Nelson, R., Handler, M. & Shaw, P. (2011). A position paper on the use of directed lie comparison questions in diagnostic and screening polygraphs. Police Polygraph

Digest, 2011, 2-5.

Borgatta, E.F., & Bohrnstedt, G.W. (1980). Level of measurement - once over again. Sociological Methods and Research, 9, 147-160.

Campbell, N. (1920/1957). Foundations of science: the philosophy of theory and experiment. New York: Dover.

Diez, J.A. (1997). A hundred years of numbers. An historical introduction to measurement theory 1887–1990. Studies in History and Philosophy of Science Part A 28 (2), 237-265.

Guttman, L. (1977). What is not what in statistics. The Statistician, 26, 81-107.

Handler, M., Shaw, P. & Gougler, M. (2010). Some thoughts about feelings: a study of the role of emotion and cognition in polygraph testing. Polygraph, 39(3), 139-154.

Handler, M., Deitchman, G., Kuczek, T., Hoffman, S., & Nelson, R. (2013). Bridging emotion and cognition: a role for the prefrontal cortex in polygraph testing. Polygraph 42(1), 1-17.

Helmholtz, H.V. (1887). Zahlen und messen, erkenntnistheoretish betrachtet, in H.v. Helmholtz, Schriften zur Erkenntnistheorie, 70-108. Engl. Trans., Numbering and measuring from an epistemological viewpoint", in H.V. Helmholtz, (1977). Epistemological Writings, 72-114, Dordrecht: Reidel.

Kahn, J., Nelson, R. & Handler, M. (2009). An exploration of emotion and cognition during polygraph testing. Polygraph, 38, 184-197.

Lord, F. M. (1953). On the statistical treatment of football numbers. American Psychologist, 8(12), 750- 751.

Luce, R. D. (1997). Quantification and symmetry: commentary on Michell 'Quantitative science and the definition of measurement in psychology'. British Journal of Psychology, 88, 395–398.

Nelson, R. (2015a). Scientific basis for polygraph testing. Polygraph, 44(1), 28-61.

Nelson, R. (2015b). Polygraph as a single-subject scientific experiment. APA Magazine, 48(5) 101-106.

Nelson, R. (2016). Scientific (analytic) theory of polygraph testing. APA Magazine, 49(5), 69-82.

Niederee, R. (1992). What do numbers

measure? A new approach to fundamental measurement. Mathematical Social Sciences 24. 237-276.

Senter, S., Weatherman, D., Krapohl, D., &Horvath, F. (2010). Psychological set or differential salience: a proposal for reconciling theory and terminology in polygraph testing. Polygraph 39(2), 109-117.

Stevens, S. S. (1946). On the theory of scales of measurement. Science, 103(2684). 677–680.

Stevens, S.S. (1951). Mathematics, measurement, and psychophysics. In S.S. Stevens (Ed.), Handbook of experimental psychology. New York: Wiley.

Suppes, P. (1951). A set of independent axioms for extensive quantities. Portugaliae Mathematica 10, 163-172.

Patrick Suppes (2002). Representational measurement theory. In J. Wixted & H. Pashler (eds.), Stevens' Handbook of Experimental Psychology. Wiley.

Suppes, P. and J. Zinnes (1963). Basic measurement theory. in Handbook of Mathematical Psychology vol. 1 (ed. D. Luce, R.R. Bush and E. Galanter). 3-76.

Suppes, P., Krantz, D., Luce, D. and A. Tversky (1989). Foundations of Measurement vol. 2. New York: Academic Press.

Tukey, J.W. (1961). Data analysis and behavioral science or learning to bear the quantitative man's burden by shunning badmandments, in The Collected Works of John W. Tukey, vol. III (1986), Lyle V. Jones (ed), Belmont, CA, Wadsworth, Inc. 391–484.

Velleman, P. F., & Wilkinson, L. (1993). Nominal, ordinal, interval, and ratio typologies are misleading. The American Statistician (1993), 47(1), 65-72.