IQ Tests And Scores Do Not Measure An Interval On A Continuum Of Intelligence

Stephanie E. King

Abstract

Psychometrics has alluded to implications of and represented IQ scores under a *normal distribution*, i.e. Gaussian or bell curve, without proof there is any quantitative data or *sufficient statistics* that are *absolutely continuous* with regard to a measure-theoretic continuum of presupposed latent g. We provide with statistical research elucidating the *continuous formula* for σ (*standard deviation*) has been used since at least the 1940's in lieu of the discrete version which is the appropriate one for the qualitative observations in IQ testing. Provided is a proof to show that the currently assumed continuity fails the Radon-Nikodym theorem which categorically renders the Gaussian for IQ invalid. A better interpretation of the Rasch model applicability is given with regard to its logit and probabilistic dichotomy with a classical example. Following is another proof supplying the argument that the test format fails the Kolmogorov axioms and definition of a σ-field. IQ is robustly shown to not be a measure-theoretic *measure* for the multiple item response test format (multiple choice) based on the current assumptions. This paper highlights historiographic connectionism errors between leaps about biometrics and twin studies with the misinterpretation of mathematical *continuity* as well as the absurdity of making analogies to Euclidian dimensional measurements for IQ. Its similarities to cult psychology are elucidated with passages and quotes.

Introduction

Social scientists have no taxonomy for the ontological measurement of intelligence (Cattell, 1971; Jensen, 1967a) and no constrained definition of it despite the efforts of psychometrists for over a century. Does a Gaussian of discrete IQ scores prove a *continuous function* of a latent trait (Hambleton & Cook, 1977) or a continuum of intelligence? Statistical rules and definitions, calculus, and measure theory are used to prove and systematically destroy the pseudoscience surrounding the format of a multiple choice answer IQ, so-called 'cognitive ability' test, e.g. Raven's Progressive Matrices (RPM, any type) (Raven, 1938; Raven, 2008), and its so-called measure of g (Cattell, 1943; Jensen, 1967b; Jensen, 1972; Jensen, 1997a; Jensen, 1997b; Jensen, 1998; Raven, 2008; Spearman, 1927; Wechsler, 1944; Wechsler, 1958). The instrument consists of a set of questions (items) that cannot be determined to soundly and theoretically (quantitatively) measure the interval level of singular latent g for general intelligence proposed by Charles Spearman or mental age related "intelligence quotient" (IQ), invented by Alfred Binet, Théodore Simon, and William Stern (Binet, 1898; Binet &

Simon, 1916; Cattell, 1971; Kovacs & Pléh, 2023). "Psychometrics without it loses its basic prop." (Wechsler, 1944)

The "general ability" latent trait g by Spearman is thought to be comprised of two abilities; "meaning making" (*eductive*) and reproductive (*recall*) ability (Spearman, 1927; Raven, 2008), or fluid and crystallized intelligence (Cattell, 1943; Cattell, 1971; Horn & Cattell, 1966; Kline, 1991). These abilities (whichever pair) are actually confounded for intelligence and should be considered polytomous for IQ which is oft overlooked and evaded, but it should have been evident to psychometrists by the outcome of the most gifted child prodigies not getting the highest possible scores on their IQ tests that there was no true measurement. These individuals typically have shown early childhood (before age 5) onset of scholastic talent, autodidactic reading or other complex abilities, domain specific abilities, i.e. music and computer programming, working memory, and attention to detail, but do not usually get the highest "measure" of g with an IQ test (Feldman, 1993; Ruthsatz & Urbach, 2012; Ruthsatz, Ruthsatz-Stephens, & Ruthsatz, 2014).

A *statistic* $T(X)$ is sufficient if no other information can be calculated to provide further what we know about a parameter θ, i.e. $\theta = (\mu, \sigma^2)$ of the sample (Fisher, 1922). By the Radon-Nikodym theorem (Halmos & Savage, 1949; Nikodym, 1930; Radon, 1913) a sufficient statistic exists for a *conditional probability* only when the *probability measures* involved are *absolutely continuous* and when technically *dominated* (Halmos & Savage, 1949, Tao, 2011). For a valid Gaussian model this condition is met for *continuous random variables* measured but psychometrics has never proven this prerequisite is established for IQ. Latent g is paradoxically assumed analogous to Euclidian unidimensional length which is a continuous random variable on a Borel σ-field (Borel, 1909) with an absolutely continuous measure (Kolmogorov, 1956). No sufficient statistics for latent g exist for a measure appropriate for a Gaussian since discrete methods have largely been ignored for the calculations of the *normal standard variable* and σ and the assumptive continuous versions have supplanted them which allows the conclusion that the Gaussian model is an insufficient fabricated inference (artifact) for IQ and has no appropriate implications for g in/of any sample.

Nomenclature

For the corrective purposes of this paper some definitions which have been widely abused (Mishra et al., 2018) in the field of statistical research overlapping the fields of biomedical and psychometrics are defined. We have a more foundational and mathematically logical basis for our usages which should be considered highly generalizable.

Data is information contained in facts gathered or observed as well as its limitations or what is dichotomously not contained in the data about a population of people, objects, attributes, qualities, values, or events. It can be quantitative, or qualitative, or both depending on how it is used in a conclusion. We denote the differences between describing the data itself and how its inference is about the subject matter is valid. If we have quantitative data, i.e. height measurements of people, this would be considered qualitative if we were proposing ideas based on this data that are indirect to the quantitative data gathered. Assuming taller men are more successful in marriage and testing this hypothesis would require a qualitative experiment or study. The results would yield a quantitative premise, a qualitative hypothesis, and a mixed data type conclusion.

A *variable* is an abstraction of some particularly defined data which can be calculated, counted, constrained, denoted, or observed by the quantitative or qualitative value, attribute, or by some measurement of it. For our purposes a measure and a measurement are not synonymous. In statistics random variables are either continuous or discrete.

Probability has two types; classical and frequentist. The classical probability is a set of known possibilities (outcomes) for an event which are assumed perfectly random, fair, and unbiased. An example is a fair coin flip which has two possible outcomes with equal probability. The frequentist probability is the count of occurrences of some particular outcome in a set of items, people, events, etc. which can come from a random sample of a population of those aforementioned, or can be randomly generated, or can be contained in a known population within certain parameters. A Bernoulli trial is often used to count the existence of certain outcomes or attributes existing in populations, outcomes of data, random events, etc. Probabilities are countable finite results that if exist for theoretically infinite events or outcomes with known classical probabilities can become more robust mathematically. There are severe limitations on what they can provide based on the variable type they are applied to and accordingly constraints on the measure theoretic measures.

Randomness is a process of unordered unbiased outcomes or selection given a set of parameters. This classically pertains to probability. Statistics of events, outcomes, populations, etc. which are not probabilities *per se* can also be created by utilizing random processes to generate examples data which information may be extracted from, taken, or observed.

Continuous random variables are considered to be on a continuum which is formally on the real number line for some function or even the line itself. Data from a continuous variable may be discretized or taken on arbitrary units of information about the quantity contained in a real measurement. There are no actual points and the continuum is considered more of a single entity. So, when a value at a point is described we are denoting an abstraction of data for a quantitative variable at that value. Continuous random variables are always quantitative, but they are not all the quantitative variables.

Discrete random variables are finite but can be represented as fractions and can be rational numbers that take on infinite decimal values on a continuum. This however, does not change the discrete categorization of them, i.e. classical probabilities. Mishra et al. mistakenly asserted that discrete "values cannot be expressed or presented in the form of a decimal" (Mishra et al., 2018) which is categorically false for classical probabilities which are discrete, e.g. any 1/3 probability. These events are not discretized, but may be conceptualized as on a continuum of the discrete events which allows a limit with respect to the probabilities to be utilized. Any nominally counted quantity, i.e. number of births given by an individual, is discrete even if it may increase and any proportions with respect to sufficient statistics (counted values or occurrences) as nominal amount per each member or event is also discrete even if in the form of a continuous fraction.

Nominal values are numerical and can be for discrete or continuous random variables, values, or quantities, i.e. for continuous it could be the number of total individuals over 6 ft. tall in a sample or how many individuals who are between 5 and 6 ft tall are over a certain threshold in that range.

Ordinal values are the nominal placement in a hierarchical ranking based on quantities or values which can be based on continuous or discrete data, or another abstractly formulated value. These can be strictly quantitative or a combination of quantitative and qualitative. An ordinal ranking that may improve, i.e. a chess ranking, is also discrete. Finite discrete random variables are not to be confused with the concept of determinism, but they are related. Discrete random variables do not easily transmute into continuous random variables, and for this paper we assert that they may not decide the existence of a proposed continuum even if one may or does exist without the proper prerequisites for a continuous random variable already proven for that assumption.

A *measure* is a set theoretic and probabilistic constrained quantity that meets the requirements of the formal definition of a measure. They are either atomic (discrete) or non-atomic (continuous). They are for classical and frequentist probabilities and/or physical measurements whether Euclidian or Cartesian.

Measurement is a quantitative system of arbitrary value on a continuum usually for a physical phenomenon of n dimensions, but is taken or observed for a continuous variable whether considered random or not. We do not considered unconstrained units valid for taking measurements.

Overview

This article confronts the massive mathematical mistake made in Psychology that has been allowed to persist the assumption of a normal Gaussian for IQ presumed to be the deviated IQ (Wechsler, 1944) that is mathematically invalid as a representation model for actual IQ scores. David Wechsler "standardized" IQ with an invariant scale (Jensen, 1967a) using Hull's method for a scoring mechanism but was not particularly convinced by IQ tests as notions of measurement, or their suitability for a Gaussian, and was well aware of the lack of units (Beaujean et al., 2018; Wechsler, 1944; Wechsler, 1952). Wechsler himself somewhat understood the limitations with regard to his treatment of the scores, and he never supported the oversimplification based on assumptions toward quantitative measurement. He did however, use the wrong formula assumptions in the scoring of IQ tests in multiple ways. Standard deviations are not a unit and displacement of a discrete formula with a continuous one is addressed as artifact.

Wechsler noted the *method of inverse frequencies* (Wechsler, 1952) which is the assumption of inferred difficulty for a task based on how many fail to perform it (the more difficult, the fewer successes) (Fisher, 1922) which he regarded as generally an invalid substitution for constrained real measurement. This was the correct observation but persists in item response theory (IRT). He called the argument *petitio principii* when one arrives at having to assume the arbitrary unit equivalence one has attempted to establish. It is also known as begging the question or restating the premise as the argument in lieu of proof. IRT is addressed here showing the item difficulty and person ability used in the Rasch model (RM) (Douglas & Wright, 1986; Rasch, n.d.-a; Rasch, n.d.-b) assumption is a false dichotomy based on invalid substitution. Neglect of theoretical statistics (Fisher, 1922) does not resolve the issue of discrete data being presumed continuous for confounded latent variables recorded with qualitative testing. A new continuous interpretation of the RM for a human guessing binary (Bernoulli) coin flips is constructed with the conditional probability being set by using the *strong law of large numbers*.

The IQ test itself is not a σ-field for a measure space of a continuum. Insofar as measures for the test score, the conditional probabilities on the power set of answers makes the binary answer counts of the multiple choice tests purely atomic (discrete) measures, but nothing more concerning the implications of any test itself having a profound eductive measure of hypothetical meaning-making for some absent absolutely continuous X_g which has never been proven to exist on Borel set intervals or a σ-field and is known to be merely assumed, i.e. in IRT. The conflation of measure with an inherent technical lack of it at interval level for psychometrics has been aptly addressed elsewhere (Beaujean et al., 2018; Michell, 1997; Wechsler, 1952).

Furthermore, the ethical and legal implications of continuing IQ testing under the current assumptions of measure or lack thereof is touched upon, and the future of psychometrics is predicted as untenable without mathematically valid notions which constrain measurement values and units owing a brief legal recount *ipso facto* that IQ testing is a violation of our rights as employees since ”intelligence” purported measured is in fact not measured by them, and will not be involved in the majority of job roles if the questions themselves are not directly related. Psychologists are beseeched to stop using the informal term ’measure’ for the assumption of a continuum without absolute continuity with a measure on an infinite Borel σ-field. Math is not something we can have people taking for granted or worse ignoring outright which is what the evidence shows is the current state of ”quantitative” psychology, a misnomer.

The Gaussian In Psychometrics

The titan of hierarchical hereditary notions lacking the requisite underpinnings, Arthur Jensen states, ”Without assuming approximate normality of the population distribution of intelligence, IQ scores cannot be treated as other than an ordinal, or rank-order, scale.” (Jensen, 2001). Just like any other kind of competitive performance ranks, i.e. chess ranking or bridge (card game) ranking, it takes habitually self systemized, environmental facilitation, motivated study, and/or deliberate practice time to achieve higher possible marks or improve a competitive status despite underlying talent by either winning games or demonstrating strong performance in competition (Glickman & Jones, 1999) for a ranked task. It has been shown independently that repeated use of practice tests invested as preparation results experimentally in higher scores even on IQ tests (Bors & Vigneau, 2003), e.g. RPM, and that training of working memory can help improve transferable fluid intelligence, i.e. the ability to solve new problems despite lack of prior knowledge (Jaeggi et al., 2008). It should be noted that the ranking devised by tests is only a competition for each specific test. The separate tests do not function as a complete system despite the notion that certain people score high on all of them.

> An I.Q. merely tells you how much better or worse, or how much above or below the average any individual is, when compared with persons of his own age. What that average represents we really do not know.

(Wechsler, 1944)

This author managed to find a book called *Modern Psychometrics, The Science of Psychological Assessment* (Rust & Golombok, 2014), but it offers no advancement of psychometric math. However, it does deeply discount the ND as appropriate for IQ with correct intuition revealing the appropriate suspicion that the ND was presumed correct and not ever proven for IQ test scores. It is referred to as a 'cult'.

> IQ-style scores are best avoided by psychometrists today. They have become something of a cult, and their extrapolations bear very little relation to normal scientific processes. ...
>
> All of these standardization techniques (the z score, the T score, the sten, the stanine and the 'IQ' format) make the assumption that scores for the general population already have a ND.

Robert Lifton empirically describes the 'totalist' milieu of cults as making the 'sacred science' of *ideology*; "that is, any set of emotionally-charged convictions about man and his relationship to the natural or supernatural world". Doctrine is imposed over the person subjugated (Lifton, 1989).

> The totalist milieu maintains an aura of sacredness around its basic doctrine or ideology, holding it as an ultimate moral vision for the ordering of human existence ...
>
> ... it greatly simplifies the world and answers a contemporary need to combine a sacred set of dogmatic principles with a claim to a science embodying the truth about human behavior and human psychology. ...
>
> The underlying assumption is that doctrine/ideology is ultimately more valid, true and real than any aspect of actual human character or human experience and one must subject one's experience to that "truth".

E.L. Thorndike et al. give us *Form A* for the 'approximate curve' and diminish its use as 'orthodox doctrine' which was a critique, not an endorsement. Throughout the text they respect discrete distributions of IQ test scores. They give the formula for it and relate the fashion of the Gaussian for intellect in use to quantitative assumptions and leaps originating with Sir Francis Galton (Thorndike et al., 1927).

$$y = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-x^2}{2\sigma^2}}$$

> The orthodox doctrine is that the form of distribution of intellect in human beings of the same sex and age ... exercises about the same amount of influence on intellect as any other, and being a surface enclosed by a curve approximating the normal probability curve ...
>
> This doctrine was urged by Francis Galton, on the basis partly of analogy with the facts in the case of certain bodily dimensions, and partly of his own shrewd observations of human abilities. Since his day it has gained very wide acceptance. ...

> It is partly because some assumption had to be made in one investigation after another for purposes of quantitative treatment, and this assumption was about as safe as any other one assumption, and much easier to operate with. Hence we gradually slid into the habit of using the doctrine. This fashion became so strong that in recent years psychologists have assumed symmetry, even though the units taken at their face value produced a markedly skewed distribution. ...
>
> We should then be very skeptical of a priori assumptions of Form A as the form of distribution of intellect in human beings ...

Variables were owed continuity with respect for the ND by Galton, the father of eugenics (Galton, 1877), but then he made a fanatical leap about the continuous measurement of intelligence which was adopted by Jensen who admits to be in it for 'negative eugenics' toward 'mental retardation' (Jensen, 1967a). What appears to have happened across conceptual drift in the field for a century and a half is that mathematical ingenuity was lost on people who feigned benevolence in its stead and committed themselves to semantic overlaps rather than expand their mathematical competence. The assumptions are syllogistic assertions and competitive pseudoscience is popularized. The notion of mathematical measurement is not constrained and utilized by psychometrics at present despite claims to the contrary. On the one hand, Galton would be so proud, and on the other, possibly ashamed that the mathematical liberties taken point back to the inferiority of the unprovable psychometrics.

Paul Kline supplants the notion of a continuous X_g required for the ND with 'polygenetic' handwaving instead of acknowledging he has no idea how to arrive at non-confounded notions of g or what to do mathematically without a continuous X_g being provable. He pretends psychometrics is a system of trait measurement when what it adheres to is correlations of items with probabilities, some misapplied log odds of how often they are answered correctly which cannot measure "objectivity" (Douglas & Wright, 1986; Rasch, n.d.-b), feigned continuity (Wechsler, 1944), and lack of proof *per se* that latent trait(s) involved aren't confounded (Spearman, 1927).

> It is assumed that intelligence has a normal distribution as is the case with most polygenetic variables.

(Kline, 1991)

The Handbook of Psychological Assessment states that IQ scores are not innate, yet gives Wechsler results a comparison to other tests visualizing a ND (Groth-Marnat, 2009) which is a mental playground paradoxically conjured by the field despite David Wechsler who was unwilling to adhere to the extrapolated theory indicated by the Gaussian for IQ (Wechsler, 1944).

> Some authors also believe that the resulting frequency curve ought to be Gaussian or as nearly Gaussian as possible. This requirement seems to be a result of the widespread belief that mental measures usually distribute themselves according to the normal curve of error. ...
>
> The distribution of our measures, however, is not Gaussian. ...
>
> A curve fitted to this data would be in the form of Pearson's Type IV.

Pearson's Type IV is non empirical and so is the discrete Gaussian construct for IQ. The calculations of standard deviation and *mean* can be done on discrete random variables, but are altered slightly in the correct formulations. The normal standard variable which can transmute onto several different distributions is

$$\mathrm{Z} = \frac{X - \mu}{\sigma}$$

Using Hull's method (Wechsler, 1944) an inflection point is attained by multiplying the standard variable by a "second standard" and adding a "second mean" to it in order to "normalize" on an absolute μ, which is not how a real mean is calculated or how a ND is formulated (Jensen, 1967a). In the 3rd ed. of Wechsler's *Measurement of Adult Intelligence* there is nothing but backpedaling in regard to use of the ND, and he gives no visual diagram of a ND in the book. There's not even a visually approximated (pseudo) ND. Here is Hull's method,

$$X_2 = M_2 + \frac{\sigma_2}{\sigma_1}(X_1 - M_1)$$

The correct standardized random variable format for discrete events or binomial distributions is below, but Wechsler didn't use it. This author believes his mistake is a primary source of much of the present day confusion.

$$\mathrm{Z} = \frac{X - np}{\sqrt{npq}}$$

That is $\mu = np$ and n is the number of events or trials. In the classical sense p is the probability which defines one possible outcome of the event and q is $1-p$. For an IQ test item the number of possible outcomes would be the sum of the answer choices $P(x_a) = \frac{1}{\sum_{n=1}^{N} X_n}$ for each discrete multiple choice question X and one outcome would be the probability of a single outcome $P(x_i) = \frac{1}{X}$. We have to combine the frequentist interpretation for scoring the test as a set of events in order to obtain this probability. This would be done with the binomial distribution formula.

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

In this formula p is our classical probability for each item and n is the number of items. This is new to psychometrics; using a measure for "measure". We have shown that sufficient statistics presumed do not exist for the continuous variable formulas combined with Hull's method on test scores and the Gaussian construct assumed for over a century. The discrete methods yield very different results for parameters without the need for the backwards construction of a continuum which cannot be proven with the test score.

The test probability for the successful items of an individual's outcome is now $P(X = k)$ because we have taken the probability of the total correct answers. If we want to understand this in terms of item failure as probabilities we can see which items were incorrect with respect to all test takers by plugging in different counts to the binomial again. The p is still the same and the n is now how many tests were given. The k is the number of times a particular item was answered correctly. This has to be done for each item across all tests taken which would yield the $P(X = k)$ for each item. From there we would see the probability of each item being answered correctly for the test sample space. We would not be measuring item difficulty, it would be the probability of each item having been answered correctly is theoretically smaller as the items become more difficult, but the classical probability remains the same for each one on a multiple choice test. The accurate statistical model of the test is by and large avoided across the field as more or less an insufficient fantasy about what can be inferred ensues and it is built upon a pillar of eugenics.

The frequentist approach can be used on the *ad hoc* answered set of questions which are determined correct or incorrect and counted. This is where a probability paradox known as the paradox of inverse probability (Fisher, 1922) occurs for a multiple choice test in the theory; its "measurement" is *post hoc*, i.e. after an imperfect observation, not precisely during the events, i.e. *a posteriori*. The paradox result is that the probabilities don't make the continuum of the imagined measure of a latent trait exist on the set of questions. It is quite the opposite. The probabilities of continuous random variables would make the continuum between independent events exist for the measures and the test would serve as the necessary function of measurement.

The use of the ND burgeoned across the field without any regard to the lack of explanations for absence of a valid continuous variable. The formula for the standard deviation is given below assuming the variable is a continuous random variable, but unlike IQ it will really have to be one or the math isn't valid, and that's saying that IQ math is not congruent with the reality of the limitations in the assumptions and psychometrists are lying about what it means as science generally speaking. The standard deviation can be calculated.

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N}(x_i - \mu)^2}{N}}$$

The discrete standard deviation first requires multiplying the value of the x by its probability, $P(x_i) = \frac{1}{\sum_{n=1}^{N} X_n}$, where X_n is the number of answer choices for question n and N is the number of test items. Doing this for multiple choice tests confounds the probability, however. Each item has a probability of $P(i) = \frac{1}{n}$, where n is the number of answer choices. Thus, the proper way to formulate the real standard deviation by discrete items with multiple probabilities would be to take each correctly answered item multiplied by $P(x_i) = \frac{1}{Nn_i}$ into the formula for σ. Let $y_i = P(x_i) \cdot x_i = \frac{x_i}{Nn_i}$ denote the probability-weighted score for item i. The mean of these weighted scores is

$$\mu = \frac{1}{N}\sum_{i=1}^{N} y_i = \frac{1}{N}\sum_{i=1}^{N} \frac{x_i}{Nn_i}$$

Substituting y_i for x_i in the standard sample variance formula $\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2$ and assuming binary values of 0 and 1 for incorrect and correct x_i we obtain

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N}\left(\frac{x_i}{Nn_i} - \mu\right)^2}{N}}$$

The 'orthodox doctrine' is assumed and instituted across the field by a misinterpretation. The fantasy of constraining intelligence in the field doesn't stop there. What is mathematically meant by an approximation of a measurement of a continuous variable, i.e. height, is that the units are arbitrary and lie on an interval for a probability of X on the closed interval $[a,\ b]$ if discretized owing that the probability of the discretized values on a continuum is 0.

Even if a continuous random variable, X, is somehow by some quantitative (Michell, 1997; Michell, 2001; Michell, 2005) notion entailed in IQ testing the scoring is nondeterministic and in essence, a variable non-deterministic outcome requiring a σ-field beforehand to be a probability measure. This means IQ scores are a specialized ordinal competitive ranking that can change with practice just like a chess ranking, SAT, or LSAT, and the score outcome doesn't define or determine anything as a continuous probability measure due to the lack of constrained measurable sets, on *measurable space* $(\Omega, \mathcal{F})$, that satisfy the Kolmogorov axioms (Kolmogorov, 1956).

It must first be proven g is a continuous measure theoretic measure, i.e. a *Lebesgue measure* or equivalent (Tao, 2011), or it doesn't go into a PDF or under the ND. Binomial distributions of size may also be symmetric or closely resemble the normal curve without assuming discrete implies continuous; the classic psychological blunder. These can be visually normally approximated by a pseudo-function which is called *correction for continuity*, but it does not contain the continuous X or transform the discrete variable into it, and it is well established that if you have binary data, i.e. test answers that are correct or incorrect, that the binomial is the right representation and can be used on the individual test collected or on the entire set of test answers or randomized scores collected which are counted and not presumed to be on a continuum without a proof of absolute continuity or a continuous function across their set. The binomial histogram is approximated with a superimposed normal curve by using an interval of $[\mu - 3\sigma, \mu + 3\sigma]$ and the Central Limit Theorem (CLT) which doesn't magically change the properties of the underlying probabilities despite the foolish optimism of certain coalitions of psychologists.

The conditional probability used to "estimate" the allegedly continuous variable variance of IQ ability already broke this rule two-fold. Making the "measure" only defined by dependent item "difficulty", i.e. based on how many people get it right independently of each other, and discrete answers of binary values, i.e. right or wrong, cannot substitute for measurement of units. The probability of the test score or an individual item landing on the presupposed intelligence point on a person's alleged continuum of g for their intellectual profile is 0. Psychologists count scores, compare them for answers, pretend they measure intelligence with conditional probabilities, and then plot the wrong distribution and use artificial metrics. No finite sample has a frequency curve, they have histograms (Fisher, 1922). "There is no such absolute measure of likelihood." (Fisher, 1922)

Even mathematics journals fall for the normalized psychometric assumptions without any proof (Grabinski & Klinkova, 2020). Grabinski & Klinkova would have been right if they had a proven variable type and measure to have a Gaussian and might be the coveted Fields medal recipients. Where is their proof of the Gaussian for IQ? In order to evaluate a Gaussian distribution of intelligence on an interval without the language of calculus you first need a continuous random variable, X_g, and a measure for it which means the calculus cannot truly be bypassed. "It is a result of a random mix of the genes of the two parents. Due to the central limit theorem such randomness leads to a Gaussian distribution. It is very hard to imagine that any other mechanism creates a Gaussian distribution." (Grabinski & Klinkova, 2020). The Radon-Nikodym theorem creates it with a measure of a continuous variable X and calculus which is absent in psychometrics.

The Rasch Model Error

A *conditional joint density function* such as required for Bayes' theorem (Koch, 2006) is also required for the RM to be mathematically sound, but we show precisely how this requisite is not met. Assuming continuous random variables for probabilities has been feigned as ”objectivity” in ’scientific measurement’ (Douglas & Wright, 1986; Rasch, n.d.-a; Rasch, n.d.-b) for the RM. Georg Rasch applied an integral to a nominal and discrete *cumulative distribution function* (CDF) for mean test scores knowing it was paradoxical. According to Rasch the discrete CDF step function can be substituted into an integral as if it's the continuous case CDF without first proving the continuity of the random variables. The logic for this mathematical atrocity is that the mean value of test scores is not to be understood as a mean of evenly distributed weights for each test result, i. Presto chango, you have scores without any measurement of difficulty magically assumed to be continuous measures. Merely taking a mean of discrete scores does not establish a continuous random variable which is necessary for a PDF and ND or a conditional joint density function.

$$M\{a_{ji}\} = \frac{\theta_j \varepsilon_i}{1 + \theta_j \varepsilon_i}$$

The mean value score for the *jth* person on the *ith* test is considered equivalent to the probability of answering θ ”ability” with ε_i ”easiness” of the test. It is stated by Rasch that the $H(\theta)$ is the CDF across persons tested. The mean raw scores are technically appropriate for *discontinuous frequency distributions* (Fisher, 1922) and assumed to be continuous below.

$$\frac{1}{n} \sum_{j=1}^{n} M\{a_{ji}\} = \frac{1}{n} \sum_{j=1}^{n} \frac{\theta_j \varepsilon_i}{1 + \theta_j \varepsilon_i}$$

The substitution is invalid.

$$\int_0^\infty \frac{\theta\varepsilon}{1 + \theta\varepsilon} \, dH(\theta) = \frac{\alpha e^\beta}{1 + \alpha e^\beta}$$

He misused the *mean value theorem for integrals* by not first establishing a continuous function on $\mathbb{R}$ in his memo (Rasch, n.d.-a).

Theorem (Mean Value Theorem for Integrals)**.** *If f is continuous on the closed interval $[a, b]$ and H is a distribution function on $[a, b]$, then there exists $c \in [a, b]$ such that*

$$\int_a^b f(\psi) \, dH(\psi) = f(c)\big(H(b) - H(a)\big)$$

From there he continued with a Laplacian proof based on nothing but unfounded assumptions that render the entire model inapplicable, and he knew it. He contrived linearity of the log odds without regard to first proving the limit of ability is established for general difficulty in the test without regard to other examples of ability that are equally as difficult or complex that the test taker is capable of performing outside the test. It established a false dichotomy in a probability without a real proof of any continuum being constrained by the test items. The fraudulent proof based on assuming discrete and continuous CDFs were equivalent was embellished further and became a pillar of psychometrics.

> We base our proof on the principle of objectivity. These proofs lead to the same conclusion: the model which connects observations and measures must be one in which the log-odds of an indicative response is governed entirely by a linear function of its parameters, as in
>
> $$log\left[\frac{p}{1-p}\right] = \beta - \delta$$
>
> The derivation of objective measurement from the observation of indicative events shows that reliance on the observation of qualities does not make social science measurement inferior to physical science measurement. ...
>
> ... the probability models developed by Rasch are the only psychometric models which produce the measurement objectivity necessary for scientific comparisons. ...
>
> ... The idea that scientific observations begin as quantities is an illusion produced by familiarity with the measurement models on which the success of physical science is based.

(Douglas & Wright, 1986)

If the complexity of a question increases in tandem with the difficulty, then the linear relationship of item difficulty is not on R and would only exist on the derivative for the unknown continuous function of ability which is non linear. It's precisely this ability that is not being measured by the test. A continuum of this ability cannot be atomic points that are probabilities of probabilities merely assumed linear. Complexity classes in computation (Knuth, 1976) already hold that various algorithms are non linear as difficulty (hardness) increases. An item is compatible with an algorithm in abstraction for the steps required to solve it. Algorithms in the same class have time and space complexity accounted for regardless of their relative size. Tasks performed by them become measurable with regard to this size. A test item cannot prove that an individual is incapable of performing tasks of the same item complexity class elsewhere. These tasks could be bigger or smaller, some may be quite particular and involve different abilities in the brain or even be highly specialized.

A better interpretation of the RM probability is guessing heads or tails for a sequence of fair coin flipping. The notion of measuring luck is perplexing and challenging. We know that guessing the outcome of a flip can be decidedly strategic, dictated, pseudorandom, or merely decided, but we can see this as a function in time or a moment in time. Each discrete Bernoulli trial is paired with its informational moment for time and guess, and each "continuous" guess for its moment in time and event in time despite the events being independent. This cannot exceed the probability of the event outcome guessed even if the coin flips have runs of either possible outcome. We know this by the Weak Law of Large Numbers for Bernoulli given below. The epsilon is not the same variable as in Rasch's formulations, just the same Greek letter.

$$\lim_{n\to\infty} P\left(\left|\frac{X}{n}\right| \geq \varepsilon\right) = 0$$

With time as a consideration that independent discrete random events with known probability occur on, i.e. fair coin flips, we can make attempts to match the outcomes by timing information and deciding the result of pseudorandom events, i.e. guessing. This produces the log odds of guessing the binary probability outcome. The revised model doesn't need to know your "ability" to give the right guess. It accounts for a guess X in time of fair coin flip outcome δ_X because the determined guess information is required before the event information becomes known, how many successes, i.e. "abilities", over the time the event sequence occurs, and the total failures, i.e. "difficulties", which will increase as the probability of guessing more correctly in time decreases.

$$P\Big(\lim_{n\to\infty} \mu \mid \bar{X}_n\Big) \geq \frac{e^{X-\delta_X}}{1+e^{X-\delta_X}}$$

The probability can only be 1/2 or less that you will guess correctly, but the limit on the running total of the set of correct guesses over time is what the model is parameterizing. If the sequence of these trials is given to a guesser in a time interval then X becomes any guess made which we assume you cannot perfectly predict the outcome with and the y can act like a conditional probability of 1 given an guess over a pseudorandom set of guesses. The classical probability of guessing correctly is technically 1/4 for our coin flip because we have two events however, one doesn't really occur without the other. The two events with probability of 1/2 are combined to produce an expected outcome. The probability of merely guessing without the second event is 0.

Effectively guessing only one of the possible outcomes each time for each trial flip will then increase your probability two-fold because the probability of the event outcome is still 1/2 whereas you multiply your own probability by that when you fail to race to the bottom of it. You will approach the limit in our formula which is going to be the same of the random event being guessed. This is possible because the guessing is not purely random. It is controlled and "unfair". Having the ability to create the probability makes a continuum which is linear to the random trial event. If we apply this idea to a multiple choice test, the linearity only exists if the test taker guesses randomly which we can infer isn't exactly how most test takers take the test. The opposite assumption is assumed true in IRT due to false interpretations of calculus and probability with respect to linearity. The neglect of information theory is also a very real issue.

A real interval for a conditional probability with continuous random variables is given below. If item difficulty is with respect to an ability of a continuous variable, they must show this in their proofs without the continuity being merely assumed by the shape of the 'S' curve for the test scores when the paradoxical probabilities are exponentiated.

$$P(a < Y < b \mid x < X < x+dx) = \int_a^b f(y \mid x < X < x+dx)\, dy$$

Our coin flipping requires a limit on the Bernoulli trial which then creates the theoretical interval on R to infinity. This is valid whereas finite tests of discrete items are not.

The paradoxical 'binary continuity' of answers in the RM afforded to IQ tests (Raven, 2008) is a bad recursion, an infinite loop based on a false linearization of a trait they never proved was a continuous X or linear to difficulty, and thus their model is the ultimate subjectivity machine. They simulate an 'interval' of calculus and a 'measure' of probability without that trait variable having absolute continuity (Kolmogorov, 1956) *per se* in the person, in the test items, or in general, and without proof the person doesn't have the ability to answer certain questions by figuring them out even if they are more difficult than some they got wrong. They also lack proof whereas other tasks the person may be capable of are just as complex or even more so than the test. Again, they have no proven continuum of gradual "intelligence" or of a limit imposed on the ability to answer correctly based on ability. Luck is involved as we have shown. There's an incorrect discount of it assumed by the RM. The paradox Rasch introduced has been resolved.

Raven would argue that the item difficulty alone is sufficient to draw the lines at a person's cognitive abilities, without having an actual measure of difficulty, information, or intelligence (Raven, 2008). The probability of guessing an answer correctly is $1/n$, for n number of possible answers. It's not $1/2$ for knowing how to do it if guessed correctly. It's $1/n$ for a correct guess with or without knowing how to get it right. The PMF below is for a discrete conditional probability that can be used on correct test answers given the number of questions.

$$P(X = x_k \mid B) = \frac{P\{(X = x_k) \cap B\}}{P(B)}, \quad P(B) > 0$$

With the RM, they sidestep the lack of a linear relationship and again, assume they have an absolutely continuous set of test items to show a perfect linear relationship which of course, they can't. There's a key mathematical difference between linearizing and a perfect linear relationship. You shouldn't be linearizing an artificial signal without a prior proven linear relationship. People can get harder questions correct and less hard questions incorrect and be perfectly capable of having figured those out or even the most difficult problem out, but not some others *per se*, and that should hold true for actions and abilities untested.

Psychometric Calculus

Consider Lebesgue measures $\lambda(\cdot)$ (Halmos & Savage, 1949; Kolmogorov, 1956, Tao, 2011), which consist of the necessary prerequisites for standard units of measurement across dimensions, i.e. length, volume, etc. If the Ω of the measurable space is intelligence but the IQ test is various sets of questions and answers that are not on a σ-field, then you are not consistent with measure theory by "standardizing" a bell curve or approximating a ND and estimating the test data no matter how many scores you obtain or what their shape resembles prior to the approximate normalization. Essentially, an integral for nothing but correct test answer conditional probability per se is pseudo formalized and then informally pretended to constrain "intelligence" as a "measure" without any proof of the necessary σ-field. The Radon-Nikodym theorem also deals with this integral to show you do not get a continuous X_g for free from a conditional probability integral of a discrete variable. Whereas, it has been feigned outright, and the following is proven fraudulent as we here make meaning of it and its non-measure.

> Stated in the simplest possible terms, the objective was to create a set of items (RPM) whose level of difficulty would increase in such a way that everyone would get all the items up to the most difficult they could solve right and fail to solve all the more difficult items. This would be the exact equivalent of a meter stick

> or tape measure where everyone passes every centimetre mark up to that which indicates their height and then fails to reach all the subsequent marks.

(Raven, 2008)

Psychometrics needs the Lebesgue measure (Tao, 2011) as a unit in order to calculate the bell curve argument or else nothing but a discrete conditional probability is being calculated and then feigned as having the robust axioms with semantics required otherwise. As to why the function and proof of continuous measure will probably not arise there is one most obvious reason. The σ-fields for the IQ test questions don't exist and cannot exist on Ω' of theoretical intelligence for an individual's intellectual profile, i.e. general competencies. Referring to the probabilities of discrete items and of answering combinations and permutations of them correctly and concluding the assumption it means X_g for free is subterfuge.

An appropriate analogy for what IQ tests are purported to do is throwing darts with respect to g and not taking physical measurement, i.e. distance to the dart board, speed, velocity of the dart, or size of the board. However, throwing a dart is acceptable on a σ-field because the dartboard is a continuum and you can assume the darts land on it as events, but there is no such proven continuum of g for test items landing on it. You precisely do not get to make this leap with a mere conditional probability of correct answers despite the semantic "variance" of ability to get the easier versus hard questions correct (Raven, 2008). Throwing darts is more measurable than an IQ test, and the test is like throwing darts but not knowing where the dart board is located or pretending that you have a board to throw at in the first place. The test might as well be parallel to it and have no chance of hitting it directly. Psychologists aim for a presuppositional continuum of something and don't really understand how a measure of it works with IQ tests, but then pretend to measure how and where the darts land with respect to the bullseye which was known to be assumed. The scoring of g with IQ always involves multiple dimensions that are feigned to be irreducible. Big O manages this complexity for specific tasks, but it won't do it for confounded g. It will require reducibly complex computational analysis to describe g.

The calculus should apply to a single score as well as all scores in a given test administered to any number of people; a measure is a measure on an individual or not the whole group of test takers. Wechsler provided a trap door to practitioners with his own safer assumption that IQ tests wouldn't really work on everyone, and wrote that the psychologists should assume they don't work on individuals they don't appear to work on, but this notion has been nearly buried and largely ignored. If they are not absolutely continuous, they are not fit for a ND.

The $f(x)$ is the PDF over the $d\lambda$, Lebesgue measure on the real number line, $\mathbb{R}$ (Tao, 2011). If we refrain from pretending statistics are calculus we may proceed with our proof by the Fundamental Theorem of Calculus for Lebesgue and the Radon-Nikodym theorem (Halmos, & Savage, 1949; Nikodym, 1930; Radon, 1913) without assuming we have an actual lambda for IQ, however necessary to perform this derivation. They make the following formula necessary in their arguments but never use it and have no $d\lambda$ for it. *A necessary and sufficient condition that ($P \ll \lambda$) is that there exists a non negative function f on X*

$$P(A) = \int_A f(x)\, d\lambda$$

for every A in S (Halmos & Savage, 1949). If P is absolutely continuous with respect to λ ($P \ll \lambda$), then the derivative of this measure is the PDF $f(x)$. By the Radon-Nikodym derivative,

$$f(x) = \frac{dP}{d\lambda}(x)$$

If X is on a normally distributed density function $f(x)$ we can set our Lebesgue derivative equal to our derivative for our formula for the normalized $f(x)$.

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

$$\int_{-\infty}^{\infty} f(x)\, dx = 1$$

The derivative is undefined if there is no lambda implying there is no defined ND without a lambda. You cannot divide by zero to obtain a valid $f(x)$. By substitution for the Lebesgue derivative we get the assumption that we are still working with a continuous variable to which we can apply the Euler derivative in order to obtain $X \sim \mathcal{N}(\mu, \sigma^2)$.

$$\frac{dP}{d\lambda}(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

$$dP = \phi(x)\, d\lambda$$

We obtain the equivalent of the Lebesgue calculus derivative for the ND which isn't legal if the Lebesgue measure for continuous X doesn't exist, it would be undefined because there's no division by 0. Maybe psychologists think this is the formula for g that will eventually be proven for IQ, but since they didn't do the work, this article can let them in on the secret; we just showed exactly what you need and exactly what you can't have with IQ test data. Without first proving the $d\lambda$ or equivalent, you have no sufficient statistic for the Gaussian of allegedly continuous IQ as a measure of g. The *law of weak convergence* is another particular way this idea that the model doesn't meet provable assumptions is approached.

Theorem (Kolmogorov, 1956, Ch. III, §5)**.** *If the sequence* $x_1, x_2, \ldots, x_n, \ldots$ *converges in probability to* x*, the corresponding sequence of distribution functions* $F_n(a)$ *converges at each point of continuity of* $F(a)$ *to the distribution function* $F(a)$ *of* x*.*

This means even if the test items are probabilities and a valid continuous distribution is the desired goal that the items must be continuous with each other which you cannot merely reverse engineer with discrete items.

Corollary 1. *If the continuum of ability is a real function on* $\mathbb{R}$ *as measured on a Borel* σ*-field of intervals then it has absolute continuity with the test item difficulty if the test is itself a dominated measure. This cannot automatically be inferred with discrete probabilities despite variance and conditional item difficulty. The general ability "measure" is not proven as tested and assumed by IQ, and there is no scientific consensus about how to take a measurement of it with appropriate instruments or constraints.*

The $P(A)$ for an IQ test of discrete random answers does not substitute for the $P(A)$ in our Lebesgue calculus, which we assumed it had to at the beginning of this disproof in order to demonstrate the current state-of-the-art. This is the fundamental flaw in relevant psychometrics; the failure to constrain an absolutely continuous variable or function as a probability measure. Jensen described the positive correlation of the mental ability test questions as ”an inexorable fact of nature” (Jensen, 1997a) and had no good nature or measure theoretic probability to support him.

Theorem. *If f is differentiable at a, then f is continuous at a.*

We have shown no such proven continuous function exists for test items and there is none proven for a trait g so, we have the following argument. The derivative of taking a measure on a non negative integral to obtain a human's continuous capacity for intelligence is another continuous function or variable that represents the rate of a latent trait at the established point on the continuum for the individual's measurement. If the calculus of measurement were utilized on a latent trait there would be the most area under the curve over the interval of highest measure. In the event of the derivative being equal to 1 there is the event of the latent trait occurrence.

The PDF is a curve that is smooth for the set containing the counted instances of the measured continuous variable, not for the underlying random events which is why the probability is 0 at every point, i.e. there are no actual points on a continuum and yet they are infinite and infinitesimal. This is the beauty of calculus. There's a probability of 0 for any value of the function $f(x)$ being the value of X, the variable. By putting IQ on a ND, psychologists have applied the notion that they are not assuming the score is valid at any point without even realizing that's what it meant to have it on an ND. They were much better off with Poisson and binomial, but they wanted the illusion of a biometric.

There seems to be specialized knowledge or specific skill involved in IQ testing of g. Who is to say that it isn't too narrow to be construed as general intelligence? It would adhere to the above corollary if measurable. This is to say that the current variable is indeed pseudo random yet merely a trial since there's no way to make a test taker perform exactly the same way twice even if they want to, they may choose not to do so by dint of their own free will or even make a fearful anxious mistake and answer wrongly after getting the item previously correct. Thus, the test is not a direct observation of constrained measurable capacity. It allows for the set of answers for each question in the set of questions to contain false answers. However, the correct answers may be guessed at random and a score is then determined based on not having actually solved certain problems.

A continuous measurement could effectively eliminate this problem of a ”fuzzy interval” assumption by having difficulty increase by a lambda and not allowing for any pseudorandom answering. A question and answer format for this is not only daunting, it is severely improbable because you effectively have to construct the test based on information that is a constant defined in a real equation and have a model of the test and of the purported trait as measures with proof, and the multiple choice test outcomes are not that for a continuum of their own volition. Even the most sophisticated neuroscience cannot do this as of yet.

If we were to do the equivalent of the IQ paradigm method to height, we would take a random sample or a collection of the population possibly controlled for by age. Then we would ”measure” the first amount of length of their height without a unit but start at a random place for each person without any real system of measurement and pretend to have one. The "item variance" of our "measure" would be whether if someone had the item we constructed in their height, then how often those people with at least that have all the other possible outcomes. We wouldn't even be sure our lines were headed in the right direction, they may as well be perpendicular to the desired measurement and possibly intersect each other not adding up to the real length of the person.

Measurement means the continuum exists and there are no gaps or "guesstimates" for part of the result. An IQ test being compared to a height measurement is effectively saying you can measure someone's height by leaving random lengths out of it, not know where the 0 begins the distance or length, and then maybe just compare your guess for the total with other people who were falsely measured this way to see how that data compares. The irony is of course that the more powerful analogy would have been Cartesian and not Euclidian due to the more generalizable assumptions for brain activity, so that is to say they absolutely feigned the wrong analogous continuum of false measure.

Top-Down Probability

The requirement of absolute continuity by Radon-Nikodym across the set of IQ test items gives us for free that a continuous function of coherent non-confounded latent "intelligence" for the test measure to represent is not possible for the test results to signify if questions are potentially left unanswered or items are selected in the multiple choice format. A probability measure of the score itself doesn't validate the currently unsubstantiated and disprovable semantics of the allegedly underlying continuum. The implications are severe. There's no continuum of unidimensional ability being modeled by the test. It has merely been assumed all along.

The discrete variable requirement of the deterministic correct answer from a set of potential choices on IQ tests does distinguish the variables on the sets of possible answers and the sets of correct answers from each other with the correct answers being contained as a subset of all possible answers or contained by the set of all answers. This notion can also be applied to the sets of all correct answers of all different IQ tests and sets of all possible answers with subsets on a sample space. What this means is that you can't mathematically separate the tests as distinct from each other if they measure on an interval the same continuous random variable. The concept aids disproving that IQ tests are measures on measurable spaces and thus, g, is not a measure constrained by IQ tests on a real interval, that ordinal level "rankings" (Spearman, 1927) are non measurable, i.e. not applicable mathematical measures to the individual test takers, and that fluid intelligence as well as general intelligence complexity for tasks are not defined by IQ or g. Big O can give us accurate abstract measures that surpass any ability as to "measured" IQ test scores.

The probability space would be $(\Omega, \mathcal{F}, P)$ in measure theory, but IQ is derived from the psychometric flaw shown previously by presuming a measure space $(\Omega, \mathcal{F}, \mu)$ by falsely inferring measure theory and induction for an automatically continuous X_g for free which is fraudulent. We can deduce that Borel sets, $\mathcal{F} : \mathcal{B}([0,1])$ on the interval $[0,1]$ for X_g are undefined since we only have a step function for the conditional probability of different discrete scores on the IQ test. We must resort to a probability measure if there is any, and as we shall see, there is no adherence to the Kolmogorov axioms or the properties of an alleged σ-field of an IQ test. Taking the power set for the test of finite answers can be an atomic $\mathcal{F}$ of a score, but the test itself is really nothing beyond that score. It's just an algebra for the set of answers for performance on the test and doesn't itself constrain the supposed continuums of latent trait(s) involved.

By the Kolmogorov Axioms let $(\Omega, \mathcal{F}, P)$ define a probability space. A function $P : \mathcal{F} \to [0, 1]$ is a probability measure if and only if:

1. **Non-negativity.** For every event $A \in \mathcal{F}$,

$$P(A) \geq 0$$

2. **Normalization.** The probability of the entire sample space is unity,

$$P(\Omega) = 1$$

3. **Countable additivity (σ-additivity).** For any countable sequence of mutually disjoint events $\{A_i\}_{i=1}^{\infty}$ where $A_i \cap A_j = \varnothing$ for $i \neq j$,

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

Now we prove that the alleged probability measure of an IQ test cannot be established for intelligence as an underlying psychometric measure with the $P(\Omega)$ by the definition of the probability measure for a continuous random variable X. If we have a probability measure, then for any event $A \in \mathcal{F}$

$$1 = P(\Omega) = P(A \cup A^c) = P(A) + P(A^c)$$

$$\Rightarrow P(A^c) = 1 - P(A)$$

The first part of this proof establishes that A is an event and also a set on Ω which cannot have a larger subset by measure than A, and for any set on Ω, the probability cannot be greater than Ω. Let's presume an IQ test can exist on Ω as a discrete case probability measure, but we are about to show why it cannot in fact do that. Now, g as an abstract could be presumed to exist here as with a conditional probability and not be necessarily in the space for the set union to be a valid measure, but in short, the IQ event set will not be dominated and the test format is not measure theoretic if even a single question may be replaced with a different one or randomly guessed correctly.

You must have a set of all answers which is greater than the number of questions for a multiple choice test with several answers to choose from meaning the correct answers are a subset of the set of all answers. Where are the questions? You cannot then have a set of answers contained by the set of the questions. In the following diagram there are B total answers and then A correct answers and thus fewer correct answers than the total alleged measure of B. Why can't the questions be in the set of answers? The set of questions has to be disjoint from the set of answers unless all the correct answers are given on the test. If we assume the set of questions is C not shown on the following diagram,

$$1 - P(C^c) = P(C) \Rightarrow P(A) + P(B) = P(C^c)$$

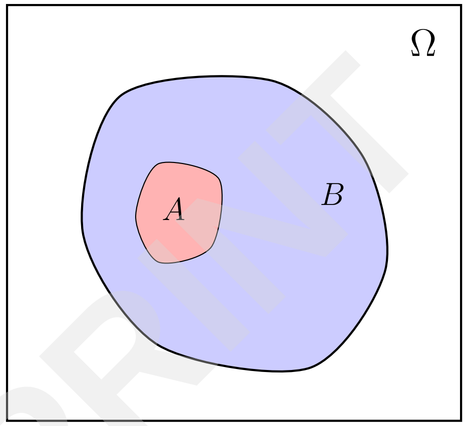

Figure 1: Nested sets on Ω: correct answers A as subset of all answers B.

$$A \subseteq B \Rightarrow \mu(A) \leq \mu(B) \Rightarrow A \leq \Omega$$

However, we already know there are more wrong choices than right choices. Therefore,

$$A = C < \Omega$$

and what you need for a measure is

$$A \cup C = \Omega$$

for some event $P(A) \Rightarrow P(C)$ which you cannot construct with a written or digital multiple choice test. The IQ test can have skipped answers or events that are not completed which means you have no σ-field as a probability measure by way of the countable additivity and normalization axioms; your variety of different questions can go randomly unanswered and therefore your alleged continuous "interval level" approximate measure of some latent trait variable is not valid for the test format.

We may have the disjoint sets of correct answers and then questions, but recall that the set of all answers cannot contain the questions. If we have the set of wrong answers and it were disjoint from the set of questions, the conditional probability would break and we would have a set of answers for all questions instead of subsets for each one. If we had the sets of questions and correct answers on Ω, we would have the following diagram where A and B are equivalent hypothetical probability measures for different discrete variables (events). They can remain disjoint in this interpretation because no one has to even so much as read the questions in order to give a randomly guessed answer.

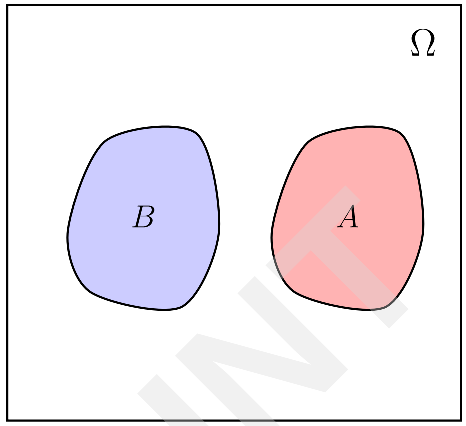

Figure 2: Disjoint sets on Ω: questions A and correct answers B as equivalent hypothetical measures.

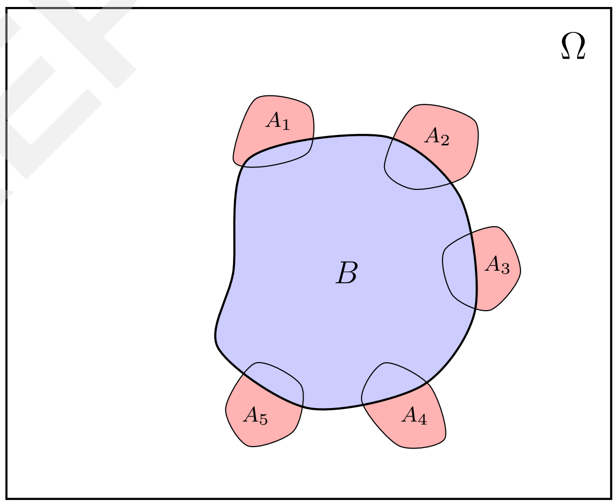

Figure 3: Question-answer intersections on Ω: singleton sets $\{A_1 \dots A_5\}$ intersecting the set of all possible answers $\{B\}$.

If we have disjoint question and answer sets that intersect with the set of all possible answers and arbitrarily they would be disjoint since they are replaceable items and there's a pairing of the event of each question with the correct outcome, we have the above singleton sets $A_1 ... A_5$ intersecting the set of all possible answers B which contains the correct answers as intersections. Making a joined set of all questions and correct answers that intersect B won't help the argument that there's absolute continuity or countable additivity of the test model, and neither would creating the subsets of wrong answer choices in B containing the correct answer for each question. Remembering that we cannot have the Ω be less than a measurable set we know that the Ω for an IQ test cannot be the sets of only questions and correct answers.

It means the set doesn't adhere to a σ-field if first, any question is potentially left unanswered or can possibly not be answered, i.e. the event doesn't occur, and second, that only if the test score is perfect is g even possible to measure with absolute continuity otherwise you have uncertainty and no normalization of the probability which means with respect to the ND; it must satisfy the Kolmogorov axioms. We don't have a σ-field for a measure of g unless g is theoretically proven and modeled on $\mathbb{R}$, intervals with Borel sets, and our IQ test probability measure is absolutely continuous with it and dominated by it.

Ethical And Legal Grounds

Scant experts have had the correct intuition that what psychometrics purports to "measure" for g are non-measurable attributes (Michell, 1997).

> Summing up the work done on mental measurements, we must conclude that for most abilities we have as yet no adequate system of mensuration, and though the results of mental measurements may frequently be expressed by numbers, the quantities so expressed cannot be compared one with another in the same sense as physical quantities may be compared.

(Wechsler, 1952)

Pretending the psychological verbiage of misrepresentations of data ought allow for theoretical applications of psychometrics has to stop. It appears that g has been considered overstated for test answer conditional probabilities since at least 1916.

> If a hierarchy can be formed the existence of a General Factor is said to be proved. Such a statement is however incorrect or at least misleading and it is to this point that my paper is devoted. I propose to show that an excellent hierarchy can be made with Specific and Group Factors only, without a General Factor. ...
>
> It has here been shown that a certain set of correlation coefficients, which we know to contain no General Factor, would be claimed by Professor Spearman as giving further support to the existence of such a factor.

(Thomson, 1916)

Psychologists are using false proofs of assumptions and analogies based on making their own semantic devices into mathematical artifacts without the appropriate level of mathematical training generally speaking, and in some cases have shown correct advanced proofs with the incorrect leaps attached to them (Douglas & Wright, 1986). Jordan Peterson got away with claiming false ideological rhetoric and *ad hominem* accusations which appear to this author to be a scapegoating tactic in multiple publicly available lectures to his students including,

> That's a great estimator of IQ; working memory. Neuropsychologists like to think that working memory is something in and of itself. That's because they don't know a damn thing about psychometrics generally speaking. There's almost no difference between fluid intelligence and working memory, and the reason for that is all measures of intellectual function collapse into g.

(Peterson, 2017, 1:12:28). According to Robert Lifton's *Eight Criteria For Thought Reform* which examines cult dynamics,

> A reverence is demanded for the ideology/doctrine, the originators of the ideology/ doctrine, the present bearers of the ideology/doctrine ...
>
> Repetitiously centered on all-encompassing jargon ...
>
> Since the group has an absolute or totalist vision of truth, those who are not in the group are bound up in evil, are not enlightened, are not saved, and do not have the right to exist.

(Lifton, 1989).

Binet (Binet & Simon, 1916) gives us the original IQ formula.

$$\text{IQ} = \frac{\text{mental age}}{\text{chronological age}}$$

Peterson, a firm believer in Binet having "measured" IQ, has executed a political strategy allegedly owing him absolute free speech on campus. Michell has a quite wonderful and well researched alternative to delusions about Binet's merit which elucidates his work as never having measured intelligence (Michell, 2012). In 2017 Peterson claimed in his course lecture that Binet did measure it despite the test in question being considered for an ordinal scale (Terman, 1949), and after having done no real work to arrive at his self assured conclusive drivel (Peterson, 2017, 5:57) he stated,

> Intelligence is actually quite measurable. We've been able to measure it since roughly the beginning of the twentieth century. Binet originated the first ... reliable and valid IQ tests... There isn't anything that social scientists have been able to measure more successfully than IQ.

There should be policy implications as the demand for psychometric literacy increases due to mounting evidence that IQ is not understood across Psychology which wielded eminence and ordinal status as the equivalence of intellectual success and modeled nothing useful or reproducible with regard to 'intelligence'.

Jensen admits the heritability studies of IQ only have variance (Jensen, 1972) to work with for assumptions of measurement without distinguishing between discrete and continuous cases which are very different. He purports that IQ test scores are a window via statistical variance to the genetic processes that can be evidenced by certain output, e.g. milk production in cows, height, but the lie about IQ is currently caught in the trap of measure theory. Volume and length are valid measurements by Lebesgue measures which transduce into continuous X variables with appropriate calculus (Tao, 2011). Jensen also describes heritability (Galton, 1870) as a technical term which describes the measurement of genetic variation of a trait or phenotype by the variance in a population. Of course, his precious variance on a ND was never valid and there was never a direct observation of genetic human intellectual ability measurable with IQ tests.

The psychometric architects feigned continuous X, sufficient statistics, statistical measurements, and a unidimensional model of general ability, g. Jensen himself purports that IQ can only be on an interval scale if it measures intelligence under a ND, and he never proved it measured intelligence, but defended it as heritable trait data to be used in population descriptions obviously betting that he would one day be proven right. That is not how real science is applied. He claimed psychologists have no use for an absolute scale yet defended IQ as having the same scale properties of height, volume, and weight which have an "absolute scale" with no apparent comprehension of the underlying math of units of measurement. He appeals to the authority of geneticists who have written textbooks and found no fault with IQ testing without citing any or naming anyone, and he then goes on to defend the government spending billions of dollars on studies of achievement differences; "where the action is" (Jensen, 1972).

Jensen purports that IQ is so impressive for genetical analysis because it behaves like "measures of continuous physical traits" (Jensen, 1972). If you have any valid Lebesgue measure, you can take for granted that all values are contained in a continuum on $\mathbb{R}$ for Borel interval(s) of the measurement which may become a minimal sufficient statistic used under a ND (Halmos & Savage, 1949), and this is how they behave for measurable physical traits, and precisely not what can be done with IQ testing.

Twin studies (Jensen, 1967b; Jensen, 1997b) have been used as substitute "evidence" in lieu of the appropriate proof of real measure and have aided the concept of false heritability of IQ test results. The assumption of continuous X_g in the field of psychology has stagnated society's competence about intelligence for over a century. The United States Armed Forces (Jensen, 1967a) was Jensen's target of eminence and has fallen for his fantasy that he was the expert on the ultimate domain of expertise. Extant psychometric instruments have failed to provide mathematically sound measurement results even for the U.S. military. The phenomenon is not well understood, but this article demonstrates the mistakes previously made with a few solutions and ideas going forward. This author assumes the U.S. Armed Forces and intelligence community prefers to understand how to accurately conceive of a measure intelligence.

Eminence has been considered some form of evidence of genius prior to IQ tests. We all know of noteworthy individuals who have made their mark on the world, but cannot predict what leads to it (Terman, 1926). Perhaps pretend playing and manufacturing ego can induce illusory valence which fools people. IQ tests also fool people.

> One cannot emphasize too often the importance of the statistical concept of intelligence for the science of mental classification. It was first introduced by Galton when he defined genius. A genius, according to Galton, was a man "who [because of his eminent work] achieved the position of one in each million."

(Wechsler, 1944)

Quite possibly just feeling "right" emotionally regulates folks across the field, but there's essentially no such thing as emotional correctness to suffice for a sound probabilistic or statistical model that is maybe not so well understood by people who prefer their own comfort level and self assurance aided by their gullibility. Psychologists who can't even understand their own behavior for lack of scientific merit (Galton, 1891) are an incompetent bunch and might have other unethical tendencies as well. They will maybe be shocked by this notion and their licensure ought to be jeopardized (American Psychological Association, 2017). It is in fact quite shocking what psychologists have gotten away with for so long under the guise of Science. Jordan Peterson claimed in his 2017 lectures,

> One of the things I have to tell you about IQ research is that if you don't buy IQ research, you might as well throw away the rest of Psychology, and the reason

> for that is first of all, the psychologists who developed intelligence testing were among the early psychologists who instantiated the statistical techniques all psychologists use to verify and test their hypotheses. So, you end up throwing the baby out with the bath water.

(Peterson, 2017, 49:00). Peterson appeals to authority without an explanation for the mathematical leap(s) in psychometrics, feigns generalized mathematical competence, and makes excuses for his orthodox indoctrination of his ideology. He is seemingly compelled by eminence and what is 'proper' behavior and decorum in his general stance as a public intellectual. It is not real science that motivates such individuals. It is obviously eminence and *per se* status, legacy, and proximity to the originators.

Maybe there exists an ongoing raging politically fueled debate among the intellectual elite as to the rank correlation coefficient or a linear regression that will suffice for an actual measure of a continuous X_g. I'd love to see this magical psychometric proof that transforms a discrete variable into a continuous one while skipping the measure theory altogether. That would be truly a mathematical breakthrough potentially owing someone a Fields Medal. Consider Peterson who might want to be considered a proponent of this approach (Peterson, 2017, 1:22:10),

> If you talk to psychologists they'll wave their hands and they'll say, ”Well, I don't really believe in IQ.” It's like well, ok, you know we're not talking about ... paranormal phenomena. They're not things you get to believe in or not. There are rules for defining what constitutes something that's real from a scientific perspective and a psychometric perspective. So, you don't get to say, ”Well, I'll apply one definition of reality because I don't like IQ, and then I'll use the same definition to justify my own constructs.” ... If you're a postmodernist you can do that. If you're a person who actually thinks scientifically you don't get to play that game.

Peterson suggested in 2017 that John Carroll should be required reading for all psychologists. Carroll appears to be on a bandwagon with the notion that claims of IQ being pseudoscience (Carroll, 1997) are fueled by the idea that these people assume there are no differences in individual abilities. Of course, this is a political straw man due to resentment at the failure of the education system among other poor reasonings. What this author sees is a failure of academia at the highest echelons to provide grounded science to the population it uses as test subjects. Carroll sided with Richard Herrnstein's and Charles Murray's *The Bell Curve* (Herrnstein & Murray, 1994) which is solely written upon the assumption and of the implications based on IQ being appropriate for a ND. Herrnstein admits in his *Crimson* interview he sided with Jensen and his twin studies analyses (Harvard Crimson, 1971, September 22).

There is a map of cognitive distortion now unfolded. These men possess many ways of not being very mathematically rigorous. Being counterintuitive stands out so much that maybe it should be considered a latent personality trait that hinders their intelligence but never was a detriment to their perceived eminence or academic success which is so unfortunate for the rest of us. In a lecture from 2025 for Peterson Academy, an unaccredited platform, John Vervaeke claimed (Vervaeke, 2025, 1:04:55),

> There's overwhelming (scientific) consensus... that g is a thing. It's not a statistical artifact. So, it's pointing at something. So, I'm not going to go into that debate because I think that debate has largely been resolved by the scientific community.

Vervaeke is feigning a scientific consensus in a for-profit lecture owned by Jordan Peterson and his daughter. Lifton would track this under Mystical Manipulation (planned spontaneity) (Lifton, 1989).

> Legitimizes the deception used to recruit new members and/or raise funds, and the deception used on the "outside world"

The comparison to a cult is by no means exhaustive in the present article. There are many more similarities, overlaps, and exact matches within the empirical definition of a cult to be studied for the proponents of the Gaussian of IQ especially for Peterson, Jensen, Herrnstein, and Murray.

This article is a *radical* (Herrnstein's definition) (Harvard Crimson, 1971, September 22) attempt to get at the roots of what is fundamentally wrong with IQ testing, and has successfully done the work Herrnstein and Jensen refused to do while being provoked by leftists to come up with an explanation for genetic IQ claims or cease and desist. The academic believers in IQ were actually social extremists who played the victim seamlessly when intuitive people revolted at the leap toward pseudoscience being applied to them at the expense of the taxpayer and at federally funded universities. I can only imagine Herrnstein's dismay if Harvard or *The Atlantic* had realized the induction he was provoking was reprehensible rather than mathematical by not rewarding him for his status and giving him deference for free. Paranoia ensued.

To this day, *The Atlantic* does not reward Science over style, and neither does The Berggruen Institute with its annual essay contest. In his article *IQ* for *The Atlantic (Monthly)*, Herrnstein referred to IQ as a yardstick, "The measurement of intelligence is one of the yardsticks by which we may assess the growing meritocracy," (Herrnstein, 1971). Clearly he was making a mockery of himself in the grand scheme of mathematics and will never be a wiser legacy. Anil Seth's winning Berggruen essay for 2025 penned that humans are not computers (Seth, 2025) when in fact computational complexity (Knuth, 1976) is potentially the correct abstract mechanism of intelligence that offers humans the potential to understand performance of any task while preserving the beauty of diversity across individuals' abilities considered necessary for a complex system of possible behaviors within many intelligences and their combinatorics. Unlike Spearman's g we do not neglect revelation toward the inward nature of Time and Space (Spearman, 1927). In the new chapter of psychometrics, which this author has started, there is still much to be said and observed.

Governing the perception of intelligence with IQ tests is totalitarianism and pretending otherwise is also totalitarianism. No one owes these people the proof they didn't produce or an apology for not falling for it politically. Hannah Arendt unwittingly described this very dynamic with,

> The ideal subject of totalitarian rule is not the convinced Nazi or the convinced communist, but people for whom the distinction between fact and fiction (i.e.,

> the reality of experience) and the distinction between true and false (i.e., the standards of thought) no longer exist.

(Arendt, 1973, p. 474).

It would be "on the nose" to claim she meant IQ tests were valid due to binary answer scoring and the detractors were totalitarian, but this author assumes that is the interpretation by many due to their narrow minded conformance and inabilities or possibly whatever the traits are that grease the wheels of mass atrocities. She didn't mean Nazis and Communists never ignored facts, by the way, it was digested as more generalized. Making everyone in society functionally have to adhere to IQ tests while it is pretended that it is mathematically resolved or scientifically proven via pseudoscience acting as propaganda is the definition of totalitarian here as it parallels the notion of mass control. Having us fall for false psychometrics as human subjects is mass control. Society requires the field adhere to better abstraction and ethical boundaries be reinforced since its members are considered free game for the sample space.

The Crimson unmasks where a popular accusation toward IQ detractors came from with a paranoid quote from passive aggressive Herrnstein (Harvard Crimson, 1971, September 22),

> "America should not be handling important matters of social policy on the basis of imaginary facts," according to Herrnstein. Asked to what he referred specifically, Herrnstein mentioned a recent Supreme Court decision which barred an employer from giving intelligence tests to janitors on the ground that such a job did not involve the use of intelligence and such a test was used discriminatorily against blacks.
>
> "I doubt that such a thing is true. It's been my experience that every job requires a certain amount of intelligence. What I object to is not Congress's right to legislate equality of opportunity but equally of outcome," he added.

The case was of course *Griggs v. Duke Power Co.* circa 1971 which is summarized.

> Even if there is no discriminatory intent, an employer may not use a job requirement that functionally excludes members of a certain race if it has no relation to measuring performance of job duties. Testing or measuring procedures cannot be determinative in employment decisions unless they have some connection to the job.

Since then, IQ tests have remained intact as artifact driven evidence and used as deciders by employers with the argument that they measure intelligence on an "interval" level and certain roles require or desire certain interval placement. This appears to be a potential case of psychological placebo effect.

Jensen, of course, fails to come up with the notions of fear of participating, intimidation by competing, low morale, psychological effects of racism, or even intuition toward better mathematical representation to bolster any argument against IQ being synonymous with intelligence. If test takers think they will be punished collectively or that there is a conspiracy against them, or if they don't believe the test really measures anything, we can't really account for how that will turn out as a set of scores, but the outlook should not be good and we should expect reluctance and even deliberate whether intended or not lack of effort in actions people are not entirely capable of controlling in their cognition. What if a 'dominant' racial group has the ability to essentially subsume the IQ test results of the minority due to using psychology on them and no one can explain how that behaves *en masse* in the minds and decisions of the population? We may in fact be living in The Twilight Zone.

> Before we go overboard in deploring the fact that our disadvantaged minority groups fail to clear many of the hurdles we set up for many jobs, including service in the Armed Forces, we should determine whether the educational and mental test hurdles that stand at the entrance of many of these jobs are actually relevant. Wittingly or unwittingly, they may be only instruments of social or racial discrimination. ...
>
> If we measure ability by IQ tests or by school achievement in academic subjects, I think the picture is a gloomy one for society as a whole and especially for some of our minorities, especially our largest minority, the Negroes. ...
>
> A few years ago I would have said that the IQ tests are not very good, that they are so culturally biased in favor of the white middle-class population ...
>
> ... no amount of equality of opportunity or improvement of educational facilities will result in equality of achievement or in any improvement of the chances for the Negro population to compete on equal terms.

(Jensen, 1967a)

The Supreme Court will probably side with this article as evidence that there are indeed violations of their case law due to what it reveals and because IQ testing interpretations have been proven to be violations of the American Psychological Association code of ethics (American Psychological Association, 2017) and those maintained across the lifespan of the ruling mentioned here are prolific. The APA has extensive ethics codes that require members to not only halt their own circus but correct and retract their publications if necessary. Psychologists governed by it are not allowed to operate beyond the boundaries of their competence and they must seek training in order to bring them up to speed. They also cannot keep lying via analogies. This means they have to learn appropriate probability, calculus, and measure theory to form new dimensions for qualitative evaluation, use varying distributions with appropriate formulas, and establish provable psychometric equations that fit the claims they wish to make.

IQ is not synonymous with anything biological, but discovering what is involved and genetically inherited by monozygotic twins is a nice problem even without assuming continuous latent variables and a ND for general intelligence is valid for the genome (Jensen, 1997b). Gaussian IQ adherence was referred to as a 'cult' by Rust & Golombok for a reason and briefly examined here with Lifton's criteria in this section. It's not ethical to feign proof of intelligence measures, apply them to real people in society or on the earth wherever these individuals exist, and then attempt to influence broader systems with assumptions in the progeny of all encompassing pseudo mathematical eugenics thought to be Darwinian. Does it really matter how ordinal you feel your rank should be as a practitioner if you can't prove your own concept that governs your ranking? The concern that it fosters the idea of supremacy is ethically valid.

Conclusion

This paper has demonstrated that IQ test scores are discrete random variables for which no Radon-Nikodym derivative with respect to Lebesgue measure exists, and therefore no valid PDF can be constructed from them. The Gaussian applied to IQ scores by psychometrists is not a density over a proven continuous variable X_g and has never been shown to satisfy the prerequisite of absolute continuity required by the Radon-Nikodym theorem. The neglected notion of the discrete standard deviation and binomial distribution are the correct models for test score handling. Neither yields a measure of g for free, and without first establishing a sufficient statistic for a continuous g on a Borel σ-field, the Gaussian model for IQ is formally indefensible.

An appropriate application based on the RM assumptions was constructed for guessing classically random binary events. The test format itself was shown to violate the Kolmogorov axioms: unanswered or randomly guessed items preclude the normalization and countable additivity required for a valid probability measure on a σ-field.

The ethical and legal consequences of IQ testing under its current assumptions have been shown to be untenable under *Griggs v. Duke Power Co.* and inconsistent with the APA code of ethics. The misdirection in media and popular science has been addressed with historiographic account which reveals a pattern of begging the question and prematurely optimistic conclusions rather than progress in Science. A comprehensive data-driven computationally complex meta-analysis of the psychometric IQ literature, deep statistical research into Jordan Peterson's academic claims as an educator and public intellectual, cognitive complexity theory that un-confounds g, and further analysis of the racism perpetuated by the ND assumptions are reserved for subsequent work.

References

1. American Psychological Association. (2017). *Ethical principles of psychologists and code of conduct* (2017 ed.). https://www.apa.org/ethics/code/ethics-code-2017.pdf

2. Arendt, H. (1973). *The origins of totalitarianism* (New ed.). Harcourt Brace Jovanovich.

3. Beaujean, A. A., Benson, N. F., McGill, R. J., & Dombrowski, S. C. (2018). A misuse of IQ scores: Using the dual discrepancy/consistency model for identifying specific learning disabilities. *Journal of Intelligence*, *6*(3), 36. https://doi.org/10.3390/jintelligence6030036

4. Binet, A. (1898). La mesure en psychologie individuelle. *Revue Philosophique de la France et de l'Étranger*, *46*, 113–123.

5. Binet, A., & Simon, T. (1916). *The development of intelligence in children* (E. S. Kite, Trans.). Williams & Wilkins.

6. Borel, É. (1909). Les probabilités dénombrables et leurs applications arithmétiques. *Rendiconti del Circolo Matematico di Palermo*, *27*(1), 247–271.

7. Bors, D. A., & Vigneau, F. (2003). The effect of practice on Raven's Advanced Progressive Matrices. *Learning and Individual Differences*, *13*(4), 291–312.

8. Carroll, J. B. (1997). Psychometrics, intelligence, and public perception. *Intelligence*, *24*(1), 25–52. https://doi.org/10.1016/S0160-2896(97)90012-X

9. Cattell, R. B. (1943). The measurement of adult intelligence. *Psychological Bulletin*, *40*(3), 153–193. https://doi.org/10.1037/h0059973

10. Cattell, R. B. (1971). *Abilities: Their structure, growth, and action*. Houghton Mifflin.

11. Douglas, G. A., & Wright, B. D. (1986). *The two category model for objective measurement* (Memorandum No. 34). MESA Psychometric Laboratory, Department of Education, University of Chicago.

12. Feldman, D. H. (1993). Child prodigies: A distinctive form of giftedness. *Gifted Child Quarterly*, *37*(4), 188–193.

13. Fisher, R. A. (1922). On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society A*, *222*(594–604), 309–368. https://doi.org/10.1098/rsta.1922.0009

14. Galton, F. (1870). *Hereditary genius: An inquiry into its laws and consequences*. D. Appleton & Company.

15. Galton, F. (1877). *Typical laws of heredity*. William Clowes and Sons.

16. Galton, F. (1891). *Hereditary genius*. D. Appleton.

17. Glickman, M. E., & Jones, A. C. (1999). Rating the chess rating system. *Chance*, *12*(2), 21–28.

18. Grabinski, M., & Klinkova, G. (2020). Scrutinizing distributions proves that IQ is inherited and explains the fat tail. *Applied Mathematics*, *11*, 605–618.

19. *Griggs v. Duke Power Co.*, 401 U.S. 424 (1971). https://supreme.justia.com/cases/federal/us/401/424/

20. Groth-Marnat, G. (2009). *Handbook of psychological assessment*. John Wiley & Sons.

21. Halmos, P. R., & Savage, L. J. (1949). Application of the Radon-Nikodym theorem to the theory of sufficient statistics. *The Annals of Mathematical Statistics*, *20*(2), 225–241.

22. Hambleton, R. K., & Cook, L. L. (1977). Latent Trait Models and Their Use in the Analysis of Educational Test Data. *Journal of Educational Measurement*, *14*(2), 75–96. http://www.jstor.org/stable/1434009

23. Harvard Crimson. (1971, September 22). Herrnstein in the Atlantic predicts American "meritocracy". *The Harvard Crimson*. https://www.thecrimson.com/article/1971/9/22/herrnstein-in-the-atlantic-predicts-american/

24. Herrnstein, R. J. (1971, September). I.Q. *The Atlantic Monthly*, *228*(3), 43–64.

25. Herrnstein, R. J., & Murray, C. (1994). *The bell curve: Intelligence and class structure in American life*. Free Press.

26. Horn, J. L., & Cattell, R. B. (1966). Refinement and test of the theory of fluid and crystallized general intelligences. *Journal of Educational Psychology*, *57*, 253–270.

27. Jaeggi, S. M., Buschkuehl, M., Jonides, J., & Perrig, W. J. (2008). Improving fluid intelligence with training on working memory. *Proceedings of the National Academy of Sciences*, *105*(19), 6829–6833.

28. Jensen, A. R. (1967a, October). *How much can we boost IQ and scholastic achievement?* [Speech, California Advisory Council of Educational Research annual meeting, San Diego, CA]. ERIC. https://files.eric.ed.gov/fulltext/ED023722.pdf

29. Jensen, A. R. (1967b). Estimation of the limits of heritability of traits by comparison of monozygotic and dizygotic twins. *Proceedings of the National Academy of Sciences*, *58*(1), 149–156.

30. Jensen, A. R. (1972). The IQ controversy: A reply to Layzer. *Cognition*, *1*(4), 427–452.

31. Jensen, A. R. (1997a). The psychometrics of intelligence. In H. Nyborg (Ed.), *The scientific study of human nature* (pp. 221–239). Pergamon.

32. Jensen, A. R. (1997b). The puzzle of nongenetic variance. In R. J. Sternberg & E. L. Grigorenko (Eds.), *Intelligence, heredity, and environment* (pp. 42–88). Cambridge University Press.

33. Jensen, A. R. (1998). *The g factor: The science of mental ability*. Praeger.

34. Jensen, A. R. (2001). Misleading caricatures of Jensen's statistics. *Chance*, *14*(4), 22–24.

35. Kline, P. (1991). *Intelligence: The psychometric view.* Routledge.

36. Knuth, D. E. (1976). Big omicron and big omega and big theta. *ACM SIGACT News*, *8*(2), 18–24.

37. Koch, K. R. (2006). Bayes' theorem. *Bayesian Inference with Geodetic Applications* (pp. 4-8). Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/BFb0048702

38. Kolmogorov, A. N. (1956). *Foundations of the theory of probability.* Chelsea Publishing Co.

39. Kovacs, K., & Pléh, C. (2023). William Stern: The relevance of his program of 'differential psychology' for contemporary intelligence measurement and research. *Journal of Intelligence*, *11*(3), 41. https://doi.org/10.3390/jintelligence11030041

40. Lifton, R. J. (1989). *Thought reform and the psychology of totalism: A study of "brainwashing" in China* (2nd ed., Chap. 22). University of North Carolina Press. https://cultrecover.com/sites/default/files/pdfs/lifton8criteria.pdf

41. Lord, F. (1952). A theory of test scores. *Psychometric Monographs*, *7*, x, 84.

42. Michell, J. (1997). Quantitative science and the definition of measurement in psychology. *British Journal of Psychology*, *88*(3), 355–383. https://doi.org/10.1111/j.2044-8295.1997.tb02641.x

43. Michell, J. (2001). Teaching and misteaching measurement in psychology. *Australian Psychologist*, *36*, 211–218. https://doi.org/10.1080/00050060108259657

44. Michell, J. (2005). The logic of measurement: A realist overview. *Measurement*, *38*(4), 285–294.

45. Michell, J. (2012). Alfred Binet and the concept of heterogeneous orders. *Frontiers in Psychology*, *3*, Article 261. https://doi.org/10.3389/fpsyg.2012.00261

46. Mishra, P., Pandey, C. M., Singh, U., & Gupta, A. (2018). Scales of measurement and presentation of statistical data. *Annals of Cardiac Anaesthesia*, *21*(4), 419–422. https://doi.org/10.4103/aca.ACA_131_18

47. Nikodym, O. (1930). Sur une généralisation des intégrales de M. J. Radon. *Fundamenta Mathematicae*, *15*, 131–179.

48. Peterson, J. B. (2017, April 18). *2017 Personality 18: Biology & traits: Openness/intelligence/creativity* [Video]. YouTube. https://www.youtube.com/watch?v=D7Kn5p7TP_Y

49. Radon, J. (1913). Theorie und Anwendungen der absolut additiven Mengenfunktionen. *Sitzungsberichte der Kaiserlich-Königlichen Akademie der Wissenschaften zu Wien, Mathematisch-Naturwissenschaftliche Klasse*, *122*(IIa), 1295–1438.

50. Rasch, G. (n.d.-a). *Estimation of parameters and control of the model for two response categories* (Memo No. 196y) [Courtesy of L. Wølk Olsen & S. Kreiner]. https://www.rasch.org/memo196y.pdf

51. Rasch, G. (n.d.-b). *On objectivity and models for measuring* (Memo No. 196z) [Lecture notes edited by J. Stene; courtesy of L. Wølk Olsen & S. Kreiner]. https://www.rasch.org/memo196z.pdf

52. Raven, J. (2008). The Raven progressive matrices tests: Their theoretical basis and measurement model. In J. Raven & J. Raven (Eds.), *Uses and abuses of intelligence: Studies advancing Spearman and Raven's quest for non-arbitrary metrics* (pp. 17–68). Royal Fireworks Press.

53. Raven, J. C. (1938). *Progressive matrices: A perceptual test of intelligence*. H. K. Lewis.

54. Rust, J., & Golombok, S. (2014). *Modern psychometrics: The science of psychological assessment*. Routledge.

55. Ruthsatz, J., & Urbach, J. B. (2012). Child prodigy: A novel cognitive profile places elevated general intelligence, exceptional working memory and attention to detail at the root of prodigiousness. *Intelligence*, *40*(5), 419–426.

56. Ruthsatz, J., Ruthsatz-Stephens, K., & Ruthsatz, K. (2014). The cognitive bases of exceptional abilities in child prodigies by domain: Similarities and differences. *Intelligence*, *44*, 11–14.

57. Seth, A. (2025). The mythology of conscious AI. *Noema Magazine*. https://www.noemamag.com/the-mythology-of-conscious-ai/

58. Spearman, C. (1927). *The abilities of man: Their nature and measurement*. Macmillan.

59. Tao, T. (2011). *An introduction to measure theory*. American Mathematical Society. https://www.stat.rice.edu/~dobelman/courses/texts/qualify/Measure.Theory.Tao.pdf

60. Terman, L. M. (Ed.). (1926). *Genetic studies of genius* (Vol. 2). Stanford University Press.

61. Terman, L. M. (1949). The Binet-Simon Intelligence Scale. In W. Dennis (Ed.), *Readings in general psychology* (pp. 331–337). Prentice-Hall. https://doi.org/10.1037/11352-047

62. Thomson, G. H. (1916). A hierarchy without a general factor. *British Journal of Psychology*, *8*(3), 271.

63. Thorndike, E. L., Bregman, E. O., Cobb, M. V., Woodyard, E., & the Staff of the Division of Psychology of the Institute of Educational Research of Teachers College, Columbia University. (1927). *The measurement of intelligence*. Teachers College, Columbia University.

64. Vervaeke, J. (2025, October 3). *Introduction to intelligence: Lecture 1 (Official)* [Video]. Peterson Academy. YouTube. https://www.youtube.com/watch?v=alzy7HqFI_w&t=3s

65. Wechsler, D. (1944). *The measurement of adult intelligence* (3rd ed.). The Williams & Wilkins Company.

66. Wechsler, D. (1952). *The range of human capacities*. The Williams & Wilkins Company.

67. Wechsler, D. (1958). *The measurement and appraisal of adult intelligence*. Williams & Wilkins Co. https://doi.org/10.1037/11167-000