On the Ontology of the True Score: From Repeated Testing to Conditional Expectation

Bruno D. Zumbo
University of British Columbia, Vancouver, Canada
Università Cattolica del Sacro Cuore, Milan, Italy (affiliate)

Paper prepared for the Invited Symposium titled *From Static Traits To Situational Constructs, Dynamic Systems, And Test Validation: Rethinking Measurement Through Context And Innovation* at the International Congress of Applied Psychology, Florence, Italy. July 24, 2026

Abstract
The repeated testing thought experiment has long served as the conceptual foundation of classical test theory. It defines the true score as the hypothetical mean obtained from an infinite number of administrations of the same test under identical conditions. Although typically presented as an intuitive heuristic, this thought experiment embodies a particular ontology of psychological measurement. It defines the true score as a property of the distribution of hypothetical observations rather than as a quantity derived from an explicit theory of the construct or the measurement process. This paper examines the conceptual commitments embedded in the repeated testing framework and argues that they have shaped psychometric thinking far beyond the definition of reliability. In particular, the framework assumes stable constructs, incorporates systematic, reproducible influences into the true score, and relegates construct theory to considerations external to the formal mathematics of test theory. These implications are contrasted with the operator-theoretic formulation of test theory, in which the true score is defined as a conditional expectation and therefore depends explicitly on an information structure. This reformulation replaces repeated observation with probabilistic projection. It aligns the mathematical foundations of test theory with contemporary explanation-focused theories of construct validity.

Keywords: classical test theory, true score, repeated testing, conditional expectation, construct validity, operator-theoretic test theory

About the author: Bruno D. Zumbo is a Distinguished University Scholar & Professor, and he also holds the Tier 1 Canada Research Chair in Psychometrics and Measurement. Affiliate: Metodi di ricerca e Tecniche di Analisi (MERITEA) Research Group, Università Cattolica del Sacro Cuore, Milan, Italy. Address correspondence to: bruno.zumbo@ubc.ca

Author note: I wish to thank Professor Annamaria Di Fabio (University of Florence), who, in her role as President of the Scientific Committee and Program International Conference of Applied Psychology ICAP 2026, and co-Vice President International Conference of Applied Psychology ICAP 2026, invited me to organize this symposium that aligns with the invitation to deliver a ICAP keynote address.

Introduction

This paper is a contribution to the Invited Symposium titled *From Static Traits to Situational Constructs, Dynamic Systems, and Test Validation: Rethinking Measurement Through Context and Innovation* at the 31st International Congress of Applied Psychology (ICAP) held in Florence, Italy, 21st-25th July, 2026. The purpose of the symposium was, in essence, to challenge the dominant view of static, decontextualized models of psychological measurement and, in response, to promote alternative views, such as an in vivo (Zumbo, 2015, 2017) and process-oriented perspective.

The focus of this paper is psychometric methods and particularly classical test theory (CTT), which, for much of the past 125 years, has been the dominant framework for test construction, psychometric analysis, and the use of technology for test delivery in the health, educational, psychological, and social science research and professional practice.

I contend, however, that what is referred to as CTT is typically either poorly described or not described in reports of psychometric research in the applied measurement and assessment literature and in psychological research more generally. Lord and Novick (1968) introduced their widely accepted and adopted CTT description. We will see later in this note that applications of their view of CTT is driven by a semantic model based on idealized repeated testing that provides the interpretive description of the syntactic (mathematical) model of CTT that is usually described as serving an expository or pedagogical purpose, but otherwise is inconsequential to the development and use of CTT.

This paper argues that the repeated testing framework embodies an implicit ontology of measurement. Its assumptions extend well beyond statistical convenience and shape how psychometricians conceptualize constructs, score variation, and measurement goals. These commitments become particularly evident when contrasted with the operator-theoretic formulation of test theory (Zumbo, 2026a, 2026b), in which the true score is defined as a conditional expectation rather than hypothetical repetition.

Definition of Classical Test Theory

Following Zimmerman and Zumbo (2001), Kroc and Zumbo (2020), and Zumbo and Kroc (2019), classical test theory is not an empirical model but a formal structure defined by

$$X = T + E, \text{ where } T = \mathbb{E}[X \mid \sigma(f)],$$

where X denotes the observed score, T the true score (the conditional expectation of X given the σ-algebra $\sigma(f)$), $E = X - T$ the error score, f the assignment-to-individual function, and $\sigma(f)$ the σ-algebra generated by f. From this definition, two properties that are mathematical consequences (not assumptions) follow immediately:

$$\mathbb{E}(E) = 0, \text{ and } \mathrm{Cov}(T, E) = 0,$$

where $\mathbb{E}(\cdot)$ and $\mathrm{Cov}(\cdot)$ denote the expectation and covariance operators, respectively. As we see, CTT is developed by construction, thus definitional rather than assumptive. What the classical literature sometimes describes as "axioms" (mean-zero error, uncorrelated true and error scores) are, in fact, theorems derived from the formal definition above.

Novick (1966) and Lord and Novick (1968) provide a detailed demonstration of how true and error scores, under very general conditions, may be constructed (defined) so that they satisfy the assumptions of the classical theory. These definitions are purely syntactic, though Lord and

Novick (1968, pp. 28-32) provide parallel semantic interpretations (the "repeated testing with memory erasure" thought experiment) that generalize the classical model described to date in the psychometric literature.

Many presentations of classical test theory have developed the theory by focusing on a measurement error model for a single individual. Lord and Novick (1968, particularly Chapter 2) adopt this strategy by introducing a repeated-testing thought experiment in which one individual is hypothetically tested an infinite number of times under equivalent conditions. The resulting probability distribution over repeated measurements provides the foundation for defining the individual's true score as the expected observed score.

The CTT model proposes:

$$X = T + E,$$

with an additional (strong) assumption that, for every individual (distinct unit in the study population), the individual's true score equals the average of their observed scores upon repeated applications of the measurement (test). More formally, for every individual i in the study population, there is a collection of events denoting all possible applications of the test to that individual, $\sigma(i)$. Then, we require

$$T = E\left(X \mid \sigma(i)\right), \text{ for every } i.$$

The exchangeability captured by the condition

$$T = E\left(X \mid \sigma(i)\right).$$

It is very strong: errors balance on every individual. This condition is the defining feature of classical test theory.

The CTT model proposes:

$$X = T + E.$$

The theory is subsequently extended from the single-individual setting to a population of individuals. This formulation can be enriched and rigorously grounded in the language of Hilbert spaces and operator theory, as elaborated in Zimmerman (1975) and Zimmerman and Zumbo (2001).

Transitioning to the semantic interpretation, it is noteworthy that while Lord and Novick (1968) established the pure mathematical definitions of true and error scores, Anastasi (1983) cautioned against reifying these formal definitions into fixed, unchangeable human traits. She argued that a psychometric score merely represents a behavioral sample at a single point in time. This perspective aligns closely with Zumbo and Kroc's (2019) assertion that measurement is fundamentally a contextual choice rather than an empirical discovery.

Lord and Novick Offer a Semantic Interpretation of CTT

Few ideas have been more influential in psychometric theory than the "*repeated testing with memory erasure*" thought experiment. Introduced as the conceptual foundation of classical test theory (CTT), it provides an intuitive interpretation of the true score as the average score that would be obtained if the same individual were tested an infinite number of times under identical conditions, and their memory erased between test administrations.

Because this description is presented early in virtually every treatment of CTT, it is often regarded as little more than a pedagogical device. Nevertheless, the repeated testing framework is considerably more than an illustration. It determines what the true score is, what error is, what constitutes reliability, and, more subtly, what kinds of explanations are regarded as relevant to psychological measurement.

The Repeated Testing Thought Experiment

The repeated-testing framework asks us to imagine administering the same test to the same individual infinitely many times under identical conditions, with memory erased after each test administration. Let $X_1, X_2, X_3, \ldots$ denote the resulting sequence of observed scores. The person's true score is defined as the limiting average; this is the average

$$T = lim_{(n\to\infty)}(1/n)\Sigma_{(i=1)}^{n}X_i$$

which is interpreted as "the limiting value of the (average) score that a person would obtain as the length of the test increased; that is, as the average score on a test consisting of infinitely many replications (Lord & Novick, 1968, p. 28).

Equivalently, one could describe the $T = \mathbb{E}(X)$, where the expectation operation, denoted $\mathbb{E}(\cdot)$ is taken over the hypothetical distribution generated by repeated administrations. The observed score decomposes into $X = T + E$, where $E = X - T$ represents measurement error.

Conceptually, the framework assumes that the underlying attribute remains perfectly stable while random measurement influences fluctuate across repeated administrations. The mathematics is coherent, but the substantive interpretation is restrictive. In most testing contexts, performance reflects memory, learning, fatigue, practice, attention, interpretation, and context.

Take-home message: The "*repeated testing with memory erasure*" thought experiment is more than an expository or pedagogical device but has **shaped the thinking of several generations of psychologists and psychometricians regarding the basic entities of test theory and particularly our understanding of a true score**. It determines what the true score is, what error is, what constitutes reliability, and, more subtly, what kinds of explanations are regarded as relevant to psychological measurement.

The "repeated testing" thought experiment is one of the conceptual cornerstones of classical test theory (CTT), even though it is almost never realizable in practice. It was developed to provide an intuitive interpretation of the true score and measurement error, and it remains influential despite the fact that modern probability theory has largely replaced it with a conditional expectation formulation.

The Hidden Ontology of Repeated Testing

The repeated testing framework is usually introduced as a definition. In reality, it also specifies what kinds of entities are presumed to exist. The framework assumes that the individual possesses a fixed quantity that persists unchanged across all hypothetical repetitions.

Measurement does not construct or explain this quantity; it merely samples from a distribution centered on it. Consequently, the true score becomes an ontological primitive. It is assumed to exist independently of any substantive explanation of how responses are generated.

Take-home message: The thought experiment, therefore, embeds three major commitments.
First, constructs are treated as stable properties rather than dynamic processes.
Second, measurement is conceived as repeated observation of an already existing quantity.
Third, explanation becomes unnecessary for defining the true score. These assumptions have shaped classical psychometric thinking for nearly a century.

The Construct Disappears from the Mathematics

Perhaps the most striking consequence of repeated testing is that the construct itself does not appear in the formal mathematics. The definition $T = \mathbb{E}(X)$ contains only observed scores and their probability distribution. Neither intelligence, depression, anxiety, quality of life, nor any other psychological construct appears explicitly in the mathematical model.

The construct motivates the development of the test, but once the probability model is established, it disappears from the theory. This disappearance of the construct is a remarkable feature of classical test theory. The mathematics concerns observed scores rather than the mechanisms that generate them.

Error Without Explanation

Within the repeated testing framework, measurement error is defined statistically rather than substantively. Any deviation from the hypothetical long-run average becomes an error. This definition intentionally avoids questions concerning the causes of variation.

Temporary fatigue, guessing, misunderstanding, contextual influences, response styles, linguistic ambiguity, and cognitive complexity all become indistinguishable components of a single residual quantity.

The theory partitions variation. It does not explain variation.

Systematic Influences Become Part of the True Score

The repeated testing framework distinguishes only between stable and random components. Suppose every administration of a questionnaire consistently induces acquiescence bias or is systematically influenced by item wording complexity. Because these influences occur on every hypothetical administration, they contribute to the long-run average itself.

Consequently, reproducible systematic influences become incorporated into the true score. This implication is rarely acknowledged but has important consequences for construct validity. Stability alone does not distinguish construct-relevant variation from construct-irrelevant variation.

Repeated testing averages systematic bias; it does not identify it.

Consequences for Construct Development

These observations reveal why construct theory remains largely external to classical test theory. The repeated testing framework assumes that the construct has already been adequately conceptualized before measurement begins.

Questions concerning

- response processes,
- cognitive mechanisms,
- linguistic interpretation,
- contextual influences,
- construct representation,

lie outside the formal statistical model.

Consequently, construct development becomes an activity separate from test theory itself. The mathematics describes score variation without explaining why that variation exists.

From Descriptive to Explanatory Measurement

Modern validity theory (Zumbo, 2023a, 2023b) has progressively shifted attention away from descriptive regularities toward explanatory models. Construct validity is no longer understood merely as demonstrating statistical consistency but as developing and testing explanations for observed score variation. Reliability coefficients, factor structures, and validity coefficients become evidential rather than foundational (Zumbo, in press). They support explanatory hypotheses rather than replacing them. This transition represents a fundamental change in the scientific aims of psychometrics.

Conditional Expectation as an Alternative Ontology

Operator-theoretic test theory (Zimmerman & Zumbo, 2021; Zumbo, 2026a, 2026b) replaces the repeated testing thought experiment with a definition grounded directly in probability theory.

Definition (True Score). Let $X \in L^2(\Omega, \mathcal{F}, P)$, and let $\mathcal{G} \subseteq \mathcal{F}$ be a sub-σ-algebra.

The true score of X defined with respect to $\mathcal{G}$ is,

$$T := \mathbb{E}(X \mid \mathcal{G}),$$

equivalently,

$$T = P_{\mathcal{G}}X,$$

where $P_{\mathcal{G}}$ denotes the orthogonal projection onto the closed subspace of $\mathcal{G}$ -measurable random variables, that is, it is the orthogonal projection from $L^2(\Omega, \mathcal{F}, P)$ onto $L^2(\Omega, \mathcal{G}, P)$.

Let $\mathcal{G}$ be a sub-σ-algebra representing the information preserved by the measurement model. This probability structure formalizes how individuals are linked to possible observed scores.

Definition (Error Score). The error score is the projection residual $\varepsilon = X - T$, with

$$\mathbb{E}(\varepsilon \mid \mathcal{G}) = 0.$$

The sub-σ-algebra $\mathcal{G}$ encodes the information structure relative to which systematic variation is defined. It generates the true-score subspace. $L^2(\Omega, \mathcal{G}, P)$, and the conditional expectation $P_{\mathcal{G}}X$ **isolates the systematic component** of the observed score. The residual $\varepsilon = X - T$ represents variation not explained by that information structure.

Take-home message: These definitions require no hypothetical sequence of repeated observations. Instead, the true score is the optimal predictor of the observed score given a specified information structure.

This formulation separates the projection operation from the substantive interpretation of $\mathcal{G}$. The sub-σ-algebra is not itself an operator; it is a collection of measurable events that specifies which person-, situation-, ecological-, or contextual-level features count as systematic for the measurement problem at hand (Anastasi, 1983; Steyer, 1988, 1989; Steyer & Schmitt, 1990; Steyer et al., 1992; Steyer et al., 2015; Zumbo, 2007, 2017, 2023a, 2023b; Zumbo et al., 2015). The corresponding operator $P_{\mathcal{G}}$ projects observed scores onto the closed subspace generated by that information structure.

At this point, it bears repeating that while Lord and Novick (1968) established the pure mathematical definitions of true and error scores, Anastasi (1983) cautioned against reifying these mathematical constructions into fixed, unchangeable human traits. She argued that a psychometric score represents a behavioral sample at a single point in time. This view aligns with Zumbo and Kroc's (2019) assertion that measurement is fundamentally a contextual choice rather than an empirical discovery.

Different test-theoretic frameworks define $\mathcal{G}$ differently. Across frameworks, the choice of $\mathcal{G}$ encodes theoretical commitments by distinguishing construct-relevant variation from residual variation. In a narrow latent-variable formulation, $\mathcal{G}$ may represent events generated by the latent attribute of interest, such as ability, proficiency, or trait level. In an ecological formulation, $\mathcal{G}$ may also include person-in-situation features that explain item responding and score interpretation. **Context, therefore, does not merely distort measurement; it helps define the measurement process**.

Thus, Zumbo's operator-theoretic framework treats, for example, true scores, error scores, and reliability as information-relative objects. Conditioning on $\mathcal{G}$ identifies the predictable

component of the measurement process, while the residual captures variation outside that structure. This formulation provides a rigorous basis for variance decomposition and score precision while keeping the intended construct and interpretive context explicit.

Implications for Construct Theory

The change from repeated testing to conditional expectation is more than a mathematical reformulation.

As we see in the Appendix, Lord and Novick's Chapter 2 is best understood as a **constructive derivation with definitional milestones rather than as an axiomatic theory.** The repeated-testing thought experiment is the constructive mechanism that generates the probability model, from which the true score, error score, variance decomposition, and reliability are subsequently defined and derived. The latter operator-theoretic formulation preserves this constructive spirit while replacing the repeated-testing construction with the more general construction of conditional expectation as an orthogonal projection.

It changes the ontology of measurement.

The true score is no longer viewed as the average of an infinite number of hypothetical observations. Instead, it becomes a projection determined by the information that the measurement theory regards as relevant.

Consequently, different substantive theories of the construct may imply different information structures and, therefore, different true-score operators.

Construct theory is no longer external to test theory mathematics. It becomes part of the mathematical definition itself.

The true score is therefore not an intrinsic property of the person independent of theory. Rather, it is defined relative to an explicitly specified probabilistic representation of the measurement process.

Discussion

The repeated-testing thought experiment has played an indispensable role in the development of classical test theory. Nevertheless, it also established an implicit ontology that continues to shape psychometric reasoning.

By defining the true score through hypothetical repeated observation, the framework privileges stability over explanation, descriptive partitioning over causal understanding, and probability distributions over theories of response generation.

The operator-theoretic formulation provides an alternative foundation. By defining the true score as a conditional expectation, it replaces hypothetical repetition with probabilistic projection. It makes explicit that measurement depends upon a theory of the information relevant to the construct.

Seen from this perspective, the transition from repeated testing to conditional expectation parallels a broader transition within psychometrics—from descriptive empiricism toward explanatory empiricism (Zumbo, 2023b). Reliability, validity, and construct theory become components of a unified inferential framework whose objective is not merely to summarize score variation but to explain it.

References

Anastasi, A. (1983). Evolving trait concepts. *American Psychologist, 38(2)*, 175–184. https://doi.org/10.1037/0003-066X.38.2.175

Kroc, E., & Zumbo, B.D. (2020). A transdisciplinary view of measurement error models and the variations of X= T+ E. *Journal of Mathematical Psychology, 98*, 102372. https://doi.org/10.1016/j.jmp.2020.102372

Lord, F.M., & Novick, M.R. (1968). *Statistical theories of mental test scores*. Addison-Wesley.

Novick, M. R. (1966). The axioms and principal results of classical test theory. *Journal of Mathematical Psychology, 3(1)*, 1–18. https://doi.org/10.1016/0022-2496(66)90002-2

Steyer, R. (1988). Conditional expectations: An introduction to the concept and its applications in empirical sciences. *Methodika, 2*, 53-78.

Steyer, R. (1989). Models of classical psychometric test theory as stochastic measurement models: representation, uniqueness, meaningfulness, identifiability, and testability. *Methodika, 3*, 25-60.

Steyer, R., & Schmitt, M. (1990). Latent state-trait models in attitude research. *Quality & Quantity, 24*, 427–445. https://doi.org/10.1007/BF00152014

Steyer, R., Ferring, D., & Schmitt, M.J. (1992). States and traits in psychological assessment. *European Journal of Psychological Assessment, 8(2)*, 79–98.

Steyer, R., Mayer, A., Geiser, C., & Cole, D. A. (2015). A theory of states and traits—Revised. *Annual Review of Clinical Psychology, 11*, 71–98. https://doi.org/10.1146/annurev-clinpsy-032813-153719

Zimmerman, D.W. (1975). Probability spaces, Hilbert spaces, and the axioms of test theory. *Psychometrika, 40(3)*, 395-412. https://doi.org/10.1007/BF02291765

Zimmerman, D. W., & Zumbo, B. D. (2001). The geometry of probability, statistics, and test theory. *International Journal of Testing, 1(3 4)*, 283 303. https://doi.org/10.1080/15305058.2001.9669476

Zumbo, B. D. (2007). Validity: Foundational Issues and Statistical Methodology. In C.R. Rao & S. Sinharay (Eds.), *Handbook of statistics* (Vol. 26, pp. 45–79). Elsevier.

Zumbo, B. D. (2015, November*). Consequences, side effects and the ecology of testing: Keys to considering assessment "in vivo"* [Plenary address]. Annual Meeting of the Association for Educational Assessment – Europe (AEAEurope), Glasgow, Scotland. URL: https://youtu.be/0L6Lr2BzuSQ

Zumbo, B. D. (2017). Trending away from routine procedures, toward an ecologically informed in vivo view of validation practices. *Measurement: Interdisciplinary Research and Perspectives, 15(3-4)*, 137–139. https://doi.org/10.1080/15366367.2017.1404367

Zumbo, B.D. (2023a). *Validity theories, frameworks and practices in using tests and measures: an over-the-shoulder look back at validity while also looking to the horizon* [Invited Address]. Ciclo Formazione Metodologica (FORME), Dipartimento di Psicologia, Università Cattolica Del Sacro Cuore. https://brunozumbo.com/?page_id=31

Zumbo, B. D. (2023b). A dialectic on validity: Explanation-focused and the many ways of being human. *International Journal of Assessment Tools in Education, 10(Special Issue)*, 1-96. https://doi.org/10.21449/ijate.1406304

Zumbo, B. D. (2026a). Reliability as Projection in Operator-Theoretic Test Theory: Conditional Expectation, Hilbert Space Geometry, and Implications for Psychometric Practice. Educational and Psychological Measurement, 86(3), 433–469. https://doi.org/10.1177/00131644251389891

Zumbo, B. D. (2026b). From Linear Geometry to Nonlinear and Information-Geometric Settings in Test Theory: Bregman Projections as a Unifying Framework. Educational and Psychological Measurement, 86(4), 714-737. https://doi.org/10.1177/00131644251393483

Zumbo, B. D. (in press). Measurement Validity: Essential in Principle, Subordinated in Practice. *Quality of Life Research*.

Zumbo, B. D., & Kroc, E. (2019). A measurement is a choice and Stevens' scales of measurement do not help make it: A response to Chalmers. *Educational and Psychological Measurement, 79(6)*, 1184–1197. https://doi.org/10.1177/0013164419844305

Zumbo, B.D., Liu, Y., Wu, A.D., Shear, B.R., Olvera Astivia, O.L., & Ark, T.K. (2015). A methodology for Zumbo's third generation DIF analyses and the ecology of item responding. *Language Assessment Quarterly, 12(1)*, 136 151. https://doi.org/10.1080/15434303.2014.972559

Appendix

Lord and Novick's (1968) Derivation of Classical Test Theory

To my knowledge, a detailed description of Lord and Novick's (1968) motivation for and derivation of their measurement-error model, classical test theory, is not in the research literature. This appendix fills that need.

In Chapter 2 of their foundational text, Lord and Novick (1968) formally developed the axiomatic framework of Classical Test Theory (CTT). They derived the system by defining observed score (X) as the sum of an unobserved latent true score (T) and a random error component (E). A mathematician may ask the question whether the derivation of classical test theory in Lord and Novick's (1968) book presents a derivation by definition or a derivation by construction? The former is axiomatic, whereas the second is constructive. This question gets to the logical structure of Lord and Novick's theory.

The answer is somewhat subtle: The derivation in Lord and Novick (1968) is fundamentally a derivation by construction, not an axiomatic derivation by definition, although it contains definitional steps within that construction.

What is a derivation by definition?

An axiomatic derivation begins by postulating primitive objects and axioms. Everything else follows deductively. Nothing is "constructed"; the primitives are assumed. In this style of mathematics, one writes,

Let T be a random variable satisfying:

$$X = T + E,$$

with the axioms

$$\mathbb{E}(E) = 0,\ \mathrm{Cov}(T, E) = 0.$$

Many modern textbook presentations of classical test theory actually look like this. However, this is not what Lord and Novick did.

What Lord and Novick actually do

Lord and Novick do **not** begin this way. Instead, they begin with a **thought experiment**. They ask us to imagine infinitely many identical (or parallel forms) administrations of the same test to a single person. From that experiment, they construct a probability distribution. From that probability distribution, they define $T = E(X)$.

Notice the logical order: Repeated testing ➔ Probability distribution ➔ Expectation ➔ True Score ➔ Error ➔ Orthogonality ➔ Variance decomposition ➔ Reliability.

Everything is built sequentially. This logical sequence is the hallmark of a constructive derivation.

Why this matters mathematically

Suppose Lord and Novick had instead begun by assuming there exists a true score T.

Then

$$X = T + E$$

would be an axiom.

Instead, they ask, 'How should we define" true score"?' Their answer is: Construct a probability model using repeated testing. Only after this construction do they define $T = E(X)$.
Thus, the expectation is **constructed first**, and the true score is identified with it.

It is notable, however, that it is **partly definitional.** Once the repeated-testing model exists, Lord and Novick make **definitions**:

$$T := E(X),$$

and

$$E := X - T.$$

So there are definitional steps.

But the mathematical object being defined— $\mathbb{E}(X)$—is itself obtained from the earlier constructive framework. Thus, the definitions do not stand alone.

Comparison with the operator-theoretic approach

Interestingly, the operator-theoretic framework (e.g., Zumbo, 2016a) changes the nature of the construction. Instead of constructing repeated administrations, one begins with $(\Omega, \mathcal{F}, P)$ and a sub-σ-algebra $\mathcal{G}$. The true score is then:

$$T = E(X \mid \mathcal{G}) = P_{\mathcal{G}} X.$$

Now the sequence of the construction is: Probability space ➔ Information structure ➔ Conditional expectation operator ➔ Projection ➔ True score ➔ Error ➔ Reliability.

So the constructive idea remains, but the object being constructed changes from a **repeated-testing probability distribution** to an **orthogonal projection determined by an information structure**.

A philosophical interpretation

This distinction also has an ontological flavor.

- **Lord–Novick:** The ontology of the true score is **constructive**. A true score exists because it is constructed as the expectation induced by an idealized repeated-testing experiment.
- **Operator-theoretic framework:** the ontology is also constructive, but the construction is more abstract. The true score exists because it is the conditional expectation (equivalently, the orthogonal projection) determined by a specified sub-σ-algebra of relevant information.

Neither theory is axiomatic in the sense of simply postulating a true score. Both derive the true score from a prior mathematical construction. The difference lies in what is constructed: Lord and Novick construct a repeated-testing probability model, whereas the operator-theoretic framework constructs an information-dependent projection in a Hilbert space.